

Automatização do Reconhecimento do Bem-estar em Ambiente de Trabalho

Definição e implementação de um *pipeline* para um sistema de
detecção do bem-estar em trabalhadores de escritório



Célio Domingos de Faria Carvalho

Faculdade de Ciências e Tecnologia

Universidade Fernando Pessoa

Tese apresentada à Universidade Fernando Pessoa
como parte dos requisitos para obtenção do grau de
Doutor em Ciências da Informação, sob a orientação do
Prof. Doutor José Torres e Prof. Doutor Rui Moreira.

Porto 2021

(C) 2021

Célio Domingos de Faria Carvalho

”TODOS OS DIREITOS RESERVADOS”

Resumo

O reconhecimento automático de emoções, que refletem o bem-estar dos humanos, tem sido alvo de investigação em vários domínios. Considera-se nestes esforços que a relação entre pessoas e máquinas pode ser mais fácil, se estas conseguirem interpretar e comunicar emoções na interação com os humanos.

Contudo, cada pessoa é única e singular. Pode, por isso, reagir de forma diferente perante vários estímulos. Esta ambiguidade faz com que seja difícil reconhecer emoções de forma automática. Por este motivo, os investigadores procuram descobrir as melhores variáveis de contexto para identificar ou inferir emoções, usando, por exemplo, algoritmos de *Machine Learning* (ML).

O autor deste projecto acredita que a diversificação e combinação das modalidades de recolha e do cuidado no processamento dos dados, pode contribuir para o aumento da qualidade da identificação de emoções e determinar o bem-estar associado. Assim, conduziu-se uma experiência para recolher um *dataset* abrangente contendo dados de contexto fisiológico, social, cultural e religioso. Neste procedimento recolheram-se dados de 20 trabalhadores de escritório em 7 organizações, numa série de 10 iterações de recolha, utilizando 11 instrumentos & sensores. Realizaram-se diversas visitas de controlo durante a recolha para minimizar problemas na recolha (e.g. ajustando sensores, lembrando o preenchimento dos questionários). Invariavelmente, estes processos experimentais estão sujeitos a interferências, ruído e outros imprevistos. Por isso, na fase de pré-processamento dos dados, foi dada uma atenção especial ao tratamento e qualidade dos dados. Definiu-se uma estratégia e implementaram-se várias ferramentas de pré-processamento, com o objetivo de conferir a melhor qualidade possível aos dados recolhidos bem como de propriedades extraídas adicionais que foram geradas para enriquecer o *dataset* final.

Uma das principais contribuições desta tese, é o *dataset* multimodal pré-processado, com mais de 1500 horas de dados de contexto, recolhido em ambientes reais. Este *dataset* é inovador e pretende ser uma pedra fundamental para a investigação relacionada com a deteção automática do bem-estar em trabalhadores de escritório.

Adicionalmente, resultaram ainda outras contribuições igualmente importantes para além do *dataset*. Destas contribuições realça-se a criação de um relatório técnico muito exaustivo que descreve e analisa detalhadamente o estado da

arte no reconhecimento de emoções e bem-estar, bem como a criação de um conjunto de instrumentos para facilitar a avaliação da qualidade de dados e a geração do *dataset* final.

Abstract

The automatic recognition of emotions, which reflect the well-being of humans, has been the subject of research in various fields. It is considered in these efforts that the relationship between people and machines can be easier if machines can interpret and communicate emotions when interacting with humans.

However, each person is unique and singular and therefore may react differently to various stimuli. This ambiguity makes it difficult to recognize emotions automatically. For this reason, researchers are trying to discover the best context variables to identify or infer emotions, using, for example, Machine Learning (ML) algorithms.

The author of this project believes that diversification and combination of collection modalities and care in processing data, can contribute to increase the quality of emotion identification and determine the associated well-being. Therefore, an experiment was conducted to collect an embracing dataset containing physiological, social, cultural and religious context data. In this procedure, data was collected from 20 office workers, from 7 organizations, in a series of 10 collection iterations, using 11 instruments & sensors. Several control visits were made during the experiment to minimize collection problems (e.g. adjusting sensors, reminding filling out questionnaires). Invariably, these experimental processes are subject to interference, noise and other unforeseen events. Therefore, in the data pre-processing phase, special attention was paid to data processing and quality. A strategy was defined and several pre-processing tools were implemented, with the aim of giving the best possible quality to the collected data, as well as additional extracted properties that were generated to enrich the final dataset.

One of the main contributions of this thesis is the pre-processed multimodal dataset, with more than 1500 hours of context data, collected in real environments. This dataset is innovative and intends to be a cornerstone for research related to the automatic detection of well-being in office workers.

In addition, other equally important contributions beyond the dataset have resulted. These contributions include the creation of a very comprehensive technical report that describes and analyzes in detail the state of the art in emotion recognition and well-being, as well as the creation of a set of tools to facilitate the evaluation of data quality and the generation of the final dataset.

Résumé

La reconnaissance automatique des émotions, qui reflètent le bien-être des humains, a fait l'objet de recherches dans différents domaines. On considère dans ces efforts que la relation entre les personnes et les machines peut être plus facile si les machines peuvent interpréter et communiquer les émotions lorsqu'elles interagissent avec les humains.

Cependant, chaque personne est unique et singulière. Il ou elle peut donc réagir différemment à divers stimuli. Cette ambiguïté rend difficile la reconnaissance automatique des émotions. C'est pourquoi les chercheurs tentent de découvrir les meilleures variables de contexte pour identifier ou déduire les émotions, en utilisant, par exemple, des algorithmes d'apprentissage automatique.

L'auteur de ce projet pense que la diversification et la combinaison des modalités de collecte et le soin apporté au traitement des données peuvent contribuer à augmenter la qualité de l'identification des émotions et à déterminer le bien-être associé. Il a donc mené une expérience visant à collecter un ensemble complet de données contenant des données sur le contexte physiologique, social, culturel et religieux. Dans cette procédure, les données ont été collectées auprès de 20 employés de bureau de 7 organisations dans une série de 10 itérations de collecte à l'aide de 11 instruments & capteurs. Plusieurs visites de contrôle ont été effectuées pendant la collecte afin de minimiser les problèmes de collecte (par exemple, réglage des capteurs, rappel du remplissage des questionnaires, etc.) Invariablement, ces processus expérimentaux sont soumis à des interférences, au bruit et à d'autres événements imprévus. Par conséquent, lors de la phase de prétraitement des données, une attention particulière a été accordée au traitement et à la qualité des données. Une stratégie a été définie et plusieurs outils de prétraitement ont été mis en œuvre, dans le but de donner la meilleure qualité possible aux données collectées ainsi qu'aux propriétés supplémentaires extraites qui ont été générées pour enrichir le jeu de données final.

L'une des principales contributions de cette thèse est l'ensemble de données multimodales prétraitées, avec plus de 1500 heures de données contextuelles, collectées dans un environnement réel. Cet ensemble de données est innovant et se veut une pierre angulaire pour la recherche liée à la détection automatique du bien-être des employés de bureau.

En outre, d'autres contributions tout aussi importantes ont été apportées au-delà de l'ensemble de données. Parmi ces contributions, nous soulignons la

création d'un rapport technique très complet qui décrit et analyse en détail l'état de l'art en matière de reconnaissance des émotions et du bien-être, ainsi que la création d'un ensemble d'outils pour faciliter l'évaluation de la qualité des données et la génération de l'ensemble de données final.

Agradecimentos

O ciclo de estudos conducente ao grau de doutor é longo, obriga a muitos sacrifícios, e a prescindir de muitas coisas da vida pessoal, familiar e profissional. Agradeço a todos que tiveram paciência e me apoiaram para a concretização desta etapa.

Aos meus orientadores, obrigado por me terem aceite como orientando e me terem apoiado durante estes anos. Espero podermos continuar a trabalhar juntos no futuro.

Um agradecimento especial às organizações e trabalhadores de escritório que aceitaram participar na experiência de recolha de dados de contexto.

O meu obrigado especial à *Movisens* pelo empréstimo dos sensores EDAMove.

Índice

Índice	viii
Lista de Figuras	xi
Lista de Tabelas	xiii
Acrónimos	xvii
1 INTRODUÇÃO	1
1.1 Contexto	2
1.2 Objetivo	4
1.3 Notação utilizada	7
1.4 Contribuições	8
1.5 Plano de trabalho e estrutura da tese	12
1.6 Proteção de dados	14
2 ESTADO DA ARTE	16
2.1 Reconhecimento de emoções	17
2.1.1 Variáveis de contexto e propriedades de domínio	22
2.1.1.1 Expressão facial, expressão oral e postura corporal	23
2.1.1.2 Variáveis de contexto fisiológico	27
2.1.1.3 Variáveis de contexto social e psicológico	33
2.1.1.4 Outras variáveis	38
2.1.2 Instrumentos & sensores	38
2.1.2.1 Medição objetiva	39
2.1.2.2 Medição subjetiva	44
2.1.2.3 Outros instrumentos	47
2.1.3 Pré-processamento	47
2.1.3.1 Manutenção do <i>dataset</i>	48
2.1.3.2 Manutenção de sinal	51
2.1.3.3 Amostragem e meta-informação	53
2.1.3.4 Outras técnicas e instrumentos	55

2.1.4	Propriedades extraídas	56
2.1.4.1	Expressão facial, oral e postura corporal	57
2.1.4.2	Contexto fisiológico	59
2.1.4.3	Contexto social e psicológico	60
2.1.4.4	Outras	62
2.2	Estudo do bem-estar em trabalhadores de escritório	62
2.2.1	Estudo das emoções no contexto laboral	63
2.2.2	Reconhecimento automático de emoções no contexto laboral	64
2.3	Análise	67
3	EXPERIÊNCIA DE RECOLHA EM CONTEXTO LABORAL	71
3.1	Preparação	75
3.1.1	Seleção de variáveis de contexto	75
3.1.2	Seleção de Instrumentos & Sensores	82
3.1.3	Documentação	87
3.2	Execução	91
3.2.1	Constituição da Amostra	92
3.2.2	Recolha de dados	97
3.3	Análise	103
4	PRÉ-PROCESSAMENTO E PROPRIEDADES EXTRAÍDAS	109
4.1	Integração do dataset	112
4.1.1	Integração de dados <i>raw</i>	114
4.1.2	Problemas encontrados durante a integração	123
4.2	Técnicas de pré-processamento utilizadas	128
4.2.1	Estratégia utilizada para manutenção de dados	129
4.2.2	Técnicas aplicadas sobre os dados	131
4.3	Propriedades extraídas	138
4.3.1	Propriedades específicas do contexto	139
4.3.2	Outras propriedades extraídas	141
4.4	Ferramentas desenvolvidas DQVA e DQVA-PP	142
4.4.1	Tecnologias e arquitetura da ferramenta	146
4.4.2	Descrição agregada dos dados	149
4.4.3	Descrição visual dos dados	151
4.5	Análise	155
5	RESULTADOS E ANÁLISE	159
5.1	Estrutura do <i>dataset</i>	160
5.2	Arquitetura de vistas	167
5.3	Caracterização dos participantes	173

5.4	Caracterização dos sinais recolhidos	179
5.5	Exploração dos dados do <i>dataset</i>	188
5.6	Análise	193
6	CONCLUSÃO	195
6.1	Resumo das contribuições	198
6.2	Experiência adquirida e recomendações	198
6.3	Trabalhos futuros	201
7	ANEXOS	202
7.1	Questionário {SOCIOCULTURAL}	202
7.2	Questionário {WELLBEING1}	219
7.3	Questionário {WELLBEING2}	221
7.4	Questionário {SLEEPTIRED}	223
7.5	Questionário {CALLINTERACTIONS}	225
7.6	Questionário {LABOR}	227
7.7	Ficha de participante {SUBJECTFILE}	229
7.8	<i>Script</i> para recolha de dados {COLLECTIONSCRIPT}	232
7.9	Convite às organizações para participação em experiência	237
7.10	Sugestão de aceitação das organizações	240
7.11	Brochura de apresentação da experiência	242
7.12	Declaração de consentimento dos participantes	245
7.13	Autorização da Comissão de Ética	247
7.14	Autorização para utilização do {EMMBEP}	249
7.15	Autorização para utilização do {WHOQOLBREF}	252
7.16	<i>Backup</i> do questionário {WELLBEING1}	255
7.17	<i>Backup</i> do questionário {WELLBEING2}	257
7.18	<i>Backup</i> do questionário {SLEEPTIRED}	259
7.19	<i>Backup</i> do questionário {CALLINTERACTIONS}	261
7.20	<i>Backup</i> do instrumento {WINDOWSACTIVITY}	263
7.21	Dicionário de dados da base de dados do <i>dataset</i>	265
	Bibliografia	276

Lista de Figuras

1.1	<i>Pipeline</i> genérico para detecção e representação de emoções	6
1.2	Cronograma do projeto	13
3.1	Calendário de recolha de dados de contexto	97
3.2	Excerto do {COLLECTIONSCRIPT} com instruções de utilização da pulseira {MIBAND} e sensor {EDAMOVE}	101
3.3	Instrumentos utilizados para recolher dados de contexto	103
3.4	Excerto do <i>dataset raw</i> com os dados de uma organização exemplo	106
4.1	Pasta de ficheiros {MIBAND} após reorganização por instrumento de recolha	113
4.2	Dados a integrar na base de dados relacional	114
4.3	Arquitetura do DQVA	146
4.4	Exemplo de <i>query</i> criada dinamicamente pela <i>Stored Procedure</i>	147
4.5	Excerto do <i>recordset</i> devolvido pela <i>query</i> dinâmica (unidade: BPM)	148
4.6	Excerto da matriz (HR) por participante / dia de recolha (unidade: BPM)	148
4.7	Medidas estatísticas calculadas pelo {DQVA} (e.g. (HR) (unidade: BPM))	149
4.8	<i>Boxplot</i> do DQVA (e.g. sinal (TEMP) (unidade: °C))	152
4.9	Histograma do DQVA (e.g. sinal (TEMP) (unidade: °C))	152
4.10	Diagrama de Pareto do DQVA (e.g. sinal (TEMP) (unidade: °C))	153
4.11	<i>Scatterplot</i> do DQVA (representação de objetos, <i>fences</i> e <i>outliers</i>) (e.g. sinal (TEMP) (unidade: °C))	153
4.12	<i>Scatterplot</i> do DQVA (comparação de sinal entre participantes) (e.g. sinal (TEMP) (unidade: °C))	154
4.13	Visualização do {DQVA} com <i>extreme</i> e <i>mild outliers</i> (sinal (HR) do participante 1510) (unidade: BPM)	155
4.14	Visualização do {DQVA} com <i>gaps</i> de sinal (HR) em alguns participantes (unidade: BPM)	157
5.1	Excerto do modelo ER do <i>dataset</i> final	161
5.2	Excerto da hierarquia de <i>views</i> (dados da tabela ANDROIDSOCIAL_Calls)	168
5.3	<i>View</i> utilizada para impor a existência de, pelo menos, um objeto a cada SI	169
5.4	Excerto da <i>view</i> de nível 1 ANDROIDSOCIAL_SMSs_1_View	170

5.5	Excerto de uma <i>view</i> exemplo de nível 2	171
5.6	Excerto do resultado do <i>recordset</i> devolvido por uma <i>view</i> de nível 3 . . .	172
5.7	Distância de casa ao trabalho, qualidade (conforto) e última mudança de casa	173
5.8	Autoapreciação do suporte social dado e recebido, e percepção do estado de saúde dos elementos do domínios privado (e.g. conjuge e filhos) por participante	176
5.9	Vontade indicada pelos participantes para mudar de emprego e de profissão	176
5.10	Desempenho de papeis de relevo na sociedade, participação ativa (e.g. associações), e filiação em partidos políticos	179
5.11	Quantidade de respostas dadas pelos participantes aos questionários {WELLBEING1}, {WELLBEING2} e {SLEEPTIRED}	182

Lista de Tabelas

1.1	Resumo de contribuições efetuadas	9
2.1	Investigações consideradas no levantamento bibliográfico em AC	22
2.2	Resumo das variáveis de contexto da expressão facial	25
2.3	Resumo das variáveis de contexto da expressão oral	26
2.4	Resumo das variáveis de contexto da expressão gestual e postura corporal	26
2.5	Variável de contexto da atividade cerebral	28
2.6	Resumo das variáveis de contexto da atividade cardíaca	29
2.7	Resumo das variáveis de contexto da atividade respiratória	29
2.8	Resumo das variáveis de contexto relacionadas com a pele	31
2.9	Resumo das variáveis de contexto da atividade muscular	31
2.10	Resumo das variáveis de contexto da atividade ocular	32
2.11	Resumo das variáveis de contexto da atividade glandular	33
2.12	Resumo das variáveis de contexto social (domínio pessoal)	36
2.13	Resumo das variáveis de contexto social (outros domínios)	37
2.14	Resumo de instrumentos & sensores (técnicas não obstrutivas)	41
2.15	Resumo de instrumentos & sensores (técnicas pouco obstrutivas)	42
2.16	Resumo de instrumentos & sensores (técnicas obstrutivas)	44
2.17	Resumo de instrumentos & sensores (questionários formais)	46
2.18	Instrumentos & sensores (questionários informais)	47
2.19	Resumo de técnicas de manutenção do <i>dataset</i> (ruído, dados incompletos e inconsistentes)	49
2.20	Resumo de técnicas de manutenção do <i>dataset</i> (redução dimensional e redundância de objetos)	51
2.21	Resumo de técnicas de manutenção do <i>dataset</i> (conversão de sinal)	52
2.22	Resumo de técnicas de manutenção do <i>dataset</i> (normalização, amplificação e filtros)	52
2.23	Resumo de técnicas de manutenção do <i>dataset</i> (segmentação e geração de novo sinal)	53
2.24	Resumo de técnicas de amostragem e meta-informação (técnicas de amostragem)	55

2.25	Resumo de técnicas de amostragem e meta-informação (balanceamento e etiquetagem)	55
2.26	Resumo das propriedades genéricas	57
2.27	Resumo das propriedades específicas de contexto (expressão facial, oral e postura corporal)	58
2.28	Resumo das propriedades específicas do contexto (contexto fisiológico)	60
2.29	Resumo das propriedades específicas do contexto (contexto social e psicológico)	61
3.1	Âmbito e definição da amostra (perfil das organizações a convidar)	73
3.2	Âmbito e definição da amostra (perfil dos participantes a admitir)	74
3.3	Variáveis de contexto selecionadas (contexto fisiológico)	77
3.4	Variáveis de contexto selecionadas (contexto psicossocial)	82
3.5	Variáveis de contexto selecionadas (outras variáveis)	82
3.6	Instrumentos & sensores selecionados (medição objetiva)	85
3.7	Instrumentos & sensores selecionados (medição subjetiva)	87
3.8	Instrumentos & sensores selecionados (outros instrumentos)	87
3.9	Documentação de suporte à execução da experiência	91
3.10	Instrumentos & sensores utilizados, e variáveis de contexto recolhidas	105
3.11	Resumo da documentação de suporte à fase de execução da experiência	106
4.1	Instrumentos utilizados na fase de integração	116
4.2	<i>Scripts</i> criados para integração de informação recolhida manualmente	119
4.3	Mapeamento dos ficheiros com dados <i>raw</i> para a base de dados relacional	123
4.4	Problemas encontrados durante a fase de integração	128
4.5	Resumo das técnicas de pré-processamento utilizadas	137
4.6	Propriedades extraídas específicas de contexto	141
4.7	Outras propriedades extraídas	142
4.8	Resumo dos instrumentos utilizados na integração de dados	156
4.9	Instrumentos criados para avaliar a qualidade de dados	158
4.10	Outros instrumentos utilizados na fase de pré-processamento e na fase exploratória do <i>dataset</i> final	158
5.1	Tabelas do modelo ER do <i>dataset</i> final	163
5.2	Dicionário de dados parcial, da base de dados do <i>dataset</i>	167
5.3	Extrato de dados recolhidos do domínio pessoal dos participantes	175
5.4	Qualidade de vida e estado de saúde dos participantes	176
5.5	Informações relacionadas com a vida profissional dos participantes	177
5.6	Características físicas dos participantes, consumo de álcool, caféina e cigarros, participação em eventos e dias em contacto com a natureza	179

5.7	Iterações de recolha e número de horas recolhidas por participante	180
5.8	Quantidade de objetos de alguns sinais recolhidos por instrumento de recolha	182
5.9	Descrição de alguns sinais recolhidos presentes no <i>dataset</i>	188
5.10	Perguntas e análise de frequência das respostas do {WELLBEING1}	190
5.11	Perguntas e análise de frequência das respostas do {WELLBEING2}	190
5.12	Tabela de correlações do subconjunto de dados em análise (<i>Spearman</i>) . .	193

Acrónimos

Roman Symbols

AC *Affective Computing*

ANS *Autonomous Nervous System*

AVC *Acidente Vascular Cerebral*

BCI *Brain-Computer Interaction*

BMI *Body Mass Index*

BMR *Basal Metabolic Rate*

BPM *Batimentos Por Minuto*

CE *Comissão de Ética*

CSV *Comma-separated Values*

CV *Coeficiente de Variação*

DBMS *DataBase Management System*

DQVA *Data Quality Visual Analysis*

DQVA – PP *Data Quality Visual Analysis (Pre-processing)*

ER *Entidade Relação*

ETL *Extract, Transform and Load*

GPS *Global Position System*

HCI *Human-Computer Interaction*

IA *Inteligência Artificial*

IM *Instant Messaging*

<i>IMDb</i>	<i>Internet Movie Database</i>
<i>IQR</i>	<i>InterQuartile Range</i>
<i>JSON</i>	<i>JavaScript Object Notation</i>
<i>ML</i>	<i>Machine Learning</i>
<i>OMS</i>	<i>Organização Mundial de Saúde</i>
<i>PDF</i>	<i>Portable Document Format</i>
<i>RT</i>	<i>Relatório Técnico</i>
<i>s</i>	<i>Segundo</i>
<i>SCL</i>	<i>Skin Conductance Level</i>
<i>SCR</i>	<i>Skin Conductance Response</i>
<i>SDK</i>	<i>Software Development Kit</i>
<i>SI</i>	<i>Sampling Interval</i>
<i>SK</i>	<i>Skewness</i>
<i>SMS</i>	<i>Short Message Service</i>
<i>SNS</i>	<i>Sistema Nervoso Simpático</i>
<i>SO</i>	<i>Sistema Operativo</i>
<i>SQL</i>	<i>Structured Query Language</i>
<i>UFP</i>	<i>Universidade Fernando Pessoa</i>
<i>μS</i>	<i>Microsiemens</i>
<i>μSs</i>	<i>Microsiemens por segundo</i>
<i>VBA</i>	<i>Visual Basic for Applications</i>
<i>XLS</i>	<i>Excel Spreadsheet file</i>
<i>XLSX</i>	<i>Excel Microsoft Office Open XML Format Spreadsheet</i>
<i>XML</i>	<i>eXtensible Markup Language</i>

Capítulo 1

INTRODUÇÃO

As emoções acompanham as pessoas diariamente, contribuem amplamente para a subjetividade do Homem enquanto ser (Jones and Troen, 2007) e para o seu bem-estar, e não podem ser medidas diretamente porque cada pessoa pode senti-las ou manifestá-las de forma diferente (Kreibig, 2010; Murad and Malkawi, 2012). Apesar de existirem já vários estudos com o objetivo de detetar automaticamente emoções, o autor desta tese acredita que, a combinação de variáveis sociais, culturais e religiosas, com as fisiológicas, poderá contribuir de forma positiva para aferir o bem-estar de trabalhadores de escritório.

A deteção automática de emoções e o bem-estar associado, é uma tarefa tão complicada quanto definir o termo emoção, por não se conhecer de forma objetiva o que se procura. Apesar do estudo das emoções ter progredido muito nos últimos anos, continua a não haver uma resposta definitiva para a pergunta "O que é uma emoção?" (Scherer, 2005). Uma das definições mais referenciadas pela literatura considera como emoção "*The bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion*" (James, 1884). Apesar do significado de emoção estar longe de ser fechado (Mandryk and Atkins, 2007), com base nesta definição e nos resultados já conseguidos noutras investigações, é possível afirmar com alguma segurança, que as emoções manifestam-se como reações no corpo humano.

As emoções são intangíveis (Murad and Malkawi, 2012), influenciam as nossas decisões diárias, e representam um estado psicológico da mente enquanto combinação de experiência subjetiva, resposta física e comportamental (Lalitha et al., 2015; Nawasalkar et al., 2013). Investigações recentes têm mostrado que, além de estarem muito ligadas ao processo cognitivo, as emoções desempenham um papel preponderante na tomada de decisões, resolução de problemas, negociação e adaptação a imprevistos (Lisetti and Nazoz, 2004). Além disso, as emoções são muito importantes nos processos de comunicação não-verbal e nos estudos relacionados com o comportamento humano (Liu et al., 2010). No entanto, as emoções não podem ser medidas de forma simples (Murad and Malkawi, 2012). A relação entre os dados recolhidos de contexto e os estados emocionais humanos são ainda pouco conhecidos, sabendo-se apenas que a reação emocional é diferente en-

tre pessoas e estímulos (Raudonis, 2013). Além disso, nem todas as pessoas conseguem identificar de forma objetiva as emoções que sentem (cf. Alexitimia), enquanto outras apenas conseguem identificar se experienciaram emoções negativas ou positivas (Barrett et al., 2001).

As emoções são as bases da interação humana (Basu et al., 2016). O homem é um ser social e as emoções que sente influenciam e são influenciadas pela sua vida em grupo (Dickerson and Kemeny, 2004; Ekman, 1989). Por exemplo, perante uma tarefa em que o desempenho pessoal é posto à prova (e.g. falar em público, cálculo mental, etc.), as pessoas sentem a pressão da avaliação que possa ser feita pelos outros (Dickerson and Kemeny, 2004) provocando um aumento de cortisol e, conseqüentemente, um aumento de stress (Kirschbaum et al., 1993). No entanto, os humanos são seres complexos e podem exprimir emoções de forma ambígua. A tristeza provocada pelo fracasso pode (ou não) levar ao choro mas, algumas pessoas choram de felicidade em reação a um evento positivo, e outras sorriem enquanto experienciam uma emoção negativa (Basu et al., 2016). Neste caso, perante o fracasso, o sorriso não significaria felicidade mas sim desconforto, desilusão, mal-estar ou tristeza (Ekman, 1989). Esta ambiguidade humana torna muito difícil a deteção de emoções apenas analisando os dados de uma única modalidade (Basu et al., 2016), como fazem muitas das investigações relacionadas com a deteção emocional (Zhang et al., 2016).

Uma emoção pode ser vista como um constructo produzido de forma sistemática por processos cognitivos, sentimentos subjetivos, excitação fisiológica, tendências motivacionais e reações comportamentais (Kim and André, 2008). Apesar do aumento de precisão dos sistemas de inferência emocional resultante da diversificação da origem de dados de contexto (Bakhtiyari and Husain, 2014), o autor acredita que a correlação com dados de contexto social, cultural e religioso, poderá contribuir de forma positiva para a precisão de um sistema de deteção emocional e do bem-estar associado. Com o aumento da capacidade computacional e da tecnologia de sensores, é de esperar o crescimento do número de investigações relacionadas com a deteção emocional (Gogia et al., 2016). Tentar inferir emoções e bem-estar apenas com base em sensores fisiológicos, ou na expressão facial, oral, postura, etc., será algo muito redutor. Portanto, considera-se necessário correlacionar dados de várias modalidades incluindo dados de contexto subjetivo (cf. social, cultural, religioso, pessoal, etc.) com o objetivo de aumentar a fiabilidade dos sistemas de deteção emocional e aferição do bem-estar de cada indivíduo.

1.1 Contexto

A avaliação emocional tem sido uma vertente de investigação da área psicossocial e da saúde, desde há muitos anos. Foi a partir da década de 90 que o reconhecimento de emoções ganhou mais atenção por parte dos investigadores, tornando-se num importante

tópico de investigação até aos dias de hoje (Basu et al., 2016). Segundo Picard, o estudo das emoções moveu-se da psicologia para a área da computação, criando um novo campo de investigação chamado de *Affective Computing* (AC). Aliás, no seu livro “*Affective Computing*”, indica as bases para a criação de um sistema inteligente para deteção emocional de forma automática (Picard, 1995).

A investigação em AC centra-se na deteção de emoções com base nas expressões facial, oral e gestual, e noutros sinais fisiológicos do corpo humano (Gogia et al., 2016; Zhai and Barreto, 2006; Bos, 2010). A lista de áreas de aplicação inclui sistemas baseados em robótica (e.g. (Luefeng et al., 2016), (Angel and Bonarini, 2014)), deteção de níveis de alerta e atenção (e.g. condutores) (e.g. (Paschero et al., 2012), (Gutmann et al., 2015)), desenvolvimento de jogos (e.g. avaliação do nível de frustração dos jogadores) (e.g. (Mandryk and Atkins, 2007), (Hsu et al., 2012)), sistemas de eLearning para avaliação do estado emocional e de atenção dos alunos (e.g. (Caballe, 2015), (Gogia et al., 2016), (Zhai and Barreto, 2006)), desenho de interfaces de aplicações (e.g. (Lisetti and Nasoz, 2004), (Bakhtiyari and Husain, 2014)), reação ao contacto com determinada peça de informação (e.g. (Chandler and Cornes, 2012)), emoções no contexto de negócios online (e.g. (Jamshidnejad and Jamshidined, 2009)), etc.

Nos últimos anos, tem-se presenciado a um aumento de investigações da área do AC, talvez pela necessidade de transformar a relação entre as coisas (e.g. *hardware*, *software* e produtos em geral) e as pessoas, numa interação mais inteligente e natural (Picard and Klein, 2002). Os sistemas AC seguem genericamente três etapas: i) aquisição e processamento de sinais (e.g. expressões faciais, orais ou gestuais, sinais fisiológicos, etc.); ii) combinação mono ou multimodal dos sinais recolhidos (i.e. considerando sinais de apenas um ou de várias origens de contexto); e iii) classificação da emoção correspondente aos sinais recolhidos, processados e combinados (Gogia et al., 2016).

O AC é hoje um tópico importante de investigação. Desenvolve mecanismos para transformar os computadores em instrumentos mais empáticos e informados acerca do estado afetivo dos seus utilizadores (Bos, 2010; Zhai and Barreto, 2006; Basu et al., 2016).

São vários autores que consideram que a deteção automática de emoções poderá ter um impacto positivo na vida das pessoas. Por exemplo, a área da psicologia poderá beneficiar, com menos subjetividade, de dados contínuos e menos diferidos no tempo; a saúde poderá ser avaliada com informação complementar à fisiológica; poderá ser mais fácil detetar delitos como atos de delinquência e atentados terroristas; e será mais fácil desenhar produtos especializados em provocar ou transmitir emoções no mundo virtual (Murad and Malkawi, 2012). Em resumo, poderá ser possível criar sistemas inteligentes do ponto de vista afetivo, informados ao nível emocional, capazes de perceber e reagir às emoções dos utilizadores.

As organizações são cada vez mais dinâmicas e exigem cada vez mais dos seus trabalhadores, nomeadamente dos que trabalham no escritório. O stress relacionado com o

cumprimento de prazos faz com que muitas vezes o local de trabalho seja um ambiente agressivo onde se experienciam emoções negativas. No entanto, a realização pessoal resultante do reconhecimento por parte dos colegas de trabalho e superiores hierárquicos, o bom espírito de equipa e cumprimento de objetivos (e.g. concretização de negócios, volumes de faturação, cumprimento de prazos), etc., serão fatores que poderão ajudar no contra balanceamento emocional com emoções positivas.

O autor pretende estudar a deteção automática do bem-estar em trabalhadores de escritório. Os escritórios inteligentes, no futuro, serão dotados de tecnologia programada para promover a adequação do local de trabalho às necessidades individuais de cada pessoa. A deteção automática de emoções no escritório, pode permitir que estes espaços se reconfigurem de forma autónoma, com o objetivo de melhorar o bem-estar no trabalho. Serão beneficiários desta tecnologia os próprios trabalhadores, mas também as organizações, que verão uma melhoria nos seus indicadores de produtividade e de retenção de pessoas nos seus quadros. O projeto descrito nesta tese, visa preparar um *dataset* final que pretende ser uma contribuição fundamental para a deteção automática do bem-estar em trabalhadores de escritório. No futuro, pretende-se utilizá-lo para testar o desempenho na classificação do bem-estar em trabalhadores de escritório.

A motivação em estudar o ambiente emocional dos escritórios resulta: i) da necessidade de se estudar o ambiente em que trabalham as pessoas com base na monitorização das suas variáveis de contexto; e ii) da ligação próxima que o autor tem ao nível profissional com o mundo empresarial.

1.2 Objetivo

Nesta investigação acredita-se que será possível criar sistemas de escritório mais inteligentes e informados acerca das emoções dos indivíduos, para que no futuro possa ser melhorado o bem-estar dos trabalhadores.

A OMS define “saúde” como sendo o estado de completo bem-estar físico, social e mental (OMS, 2016). Olhando para o estado emocional como uma forma de avaliar o equilíbrio triangular da saúde defendida pela OMS, neste trabalho considera-se que uma deteção de emoções mais eficiente, poderá promover uma melhoria nos processos de avaliação do bem-estar dos trabalhadores que são sujeitos ao *pressing* administrativo dos escritórios. Neste contexto, acredita-se vir a ser possível, no futuro, criar sistemas capazes de reagir de forma automática reconfigurando o ambiente dos locais onde se desempenham tarefas administrativas, com base na monitorização emocional dos seus trabalhadores. Esta inteligência afetiva dos sistemas poderá promover uma melhoria do bem-estar e qualidade de vida dos trabalhadores, rumo a uma humanização dos sistemas e a um melhor relacionamento entre pessoas e máquinas.

Um dos grandes desafios do *Affective Computing* (AC) é precisamente a identificação

de emoções humanas. Muitas das investigações focam-se na captura de dados de contexto fisiológico (Alabdulkarim, 2015). Como muitas das alterações fisiológicas são mensuráveis de forma objetiva através de sensores, os investigadores são tentados a procurar bases de correlação para a deteção emocional automática, baseando-se apenas nesta origem de dados. Apesar dos vários estudos já existentes na área do AC, acredita-se que a combinação de variáveis sociais, culturais e religiosas com as fisiológicas, poderá contribuir de forma positiva para a qualidade dos resultados obtidos na deteção automática de emoções e do bem-estar associado. Não desconsiderando a importância dos dados fisiológicos, os dados sociais, culturais e religiosos, poderão também contribuir de forma muito positiva para a precisão de um sistema de inferência emocional e bem-estar. Ao nível social, a interação entre pessoas pode desencadear emoções positivas como a alegria, ou negativas como a vergonha. A obtenção de conhecimento ou competências pode suscitar emoções positivas como satisfação ou negativas como stress. Ao nível religioso as pessoas poderão sentir emoções mais positivas pelo cumprimento dos desígnios do seu credo ou negativas inerentes a sentimentos de culpa ou pecado (Fredrickson, 2002).

Assim, uma pessoa que trabalha sozinha num escritório poderá vivenciar emoções diferentes daquelas que trabalham num espaço partilhado com colegas. Se a empresa onde trabalha oferece espaços promotores de momentos de lazer ou políticas de flexibilidade de trabalho laboral, os colaboradores poderão sentir-se melhor no local de trabalho por causa da qualidade de vida proporcionada pelo ambiente onde passam grande parte do seu tempo. Um funcionário de uma organização que promove a valorização dos currículos do seu pessoal, sentir-se-á provavelmente mais satisfeito e motivado por sentir que os seus superiores promovem o aumento das suas competências através de formação contínua. Um trabalhador de uma empresa que proporciona ao seu pessoal atividades extra-profissionais como passeios, caminhadas, e atividades ao ar-livre, uma pessoa que goza períodos de férias repartidos, lê regularmente, vai ao teatro ou ao cinema com regularidade, e até participa em cerimónias religiosas, sentirá provavelmente emoções diferentes daqueles que não trabalham num ambiente de trabalho tão agradável ou não têm uma vida social e religiosa tão ativa e inclusiva.

Neste contexto, conduziu-se uma experiência de recolha de dados, com o objetivo de preparar um *dataset* final pré-processado, que pretende ser uma contribuição fundamental para a deteção automática do bem-estar em trabalhadores de escritório. Neste estudo recolheram-se variáveis de contexto de várias modalidades, que foram posteriormente pré-processadas de forma a garantir a qualidade do *input* para os algoritmos de *Machine Learning* (ML) e respetiva classificação. Além de se recolherem dados de contexto fisiológico, decidiu-se também utilizar dados de contexto social e religioso por constituírem uma parte fundamental de vida dos trabalhadores.

Os investigadores tendem a utilizar dados de contexto fisiológico por serem involuntários (i.e., são difíceis de mascarar e controlar pelos participantes), e porque tendem a

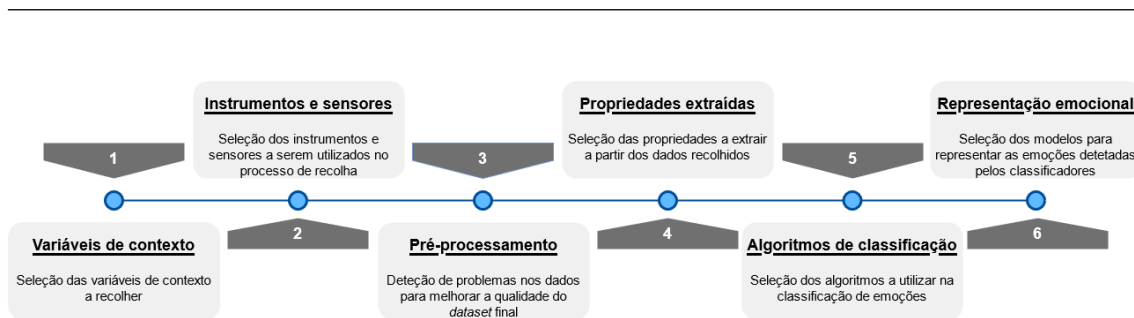


Figura 1.1: *Pipeline* genérico para detecção e representação de emoções

representar pontos objetivos de dados, já que são recolhidos através de sensores (Rani and Sarkar, 2006; Nawasalkar et al., 2013). No entanto, acredita-se que os fatores sociais e religiosos também influenciarão o bem-estar dos trabalhadores de escritório como, por exemplo: i) satisfação com o emprego e profissão; ii) intensidade (ritmo) de trabalho; iii) quantidade e qualidade das interações sociais; iv) suporte social dado e recebido; v) qualidade do sono; vi) sentimento de cansaço; vii) faltas e atrasos ao trabalho; viii) gozo de férias; ix) atividade física; x) consumo de bebidas alcoólicas e tabaco. Neste sentido, o *dataset* recolhido e processado neste projeto, será uma contribuição essencial para a criação futura de sistemas inteligentes do ponto de vista afetivo, informados ao nível emocional, capazes de perceber e reagir às emoções dos funcionários para melhorar o bem-estar no trabalho.

Os estudos relacionados com a deteção automática de emoções seguem uma sequência genérica de etapas transversal às diferentes investigações acerca do tema (ver Figura 1.1): i) seleção das variáveis de contexto a recolher; ii) escolha dos instrumentos e sensores suporte à recolha; iii) pré-processamento do *dataset* (e.g. remoção de ruído, redimensionamento de dados, etc.); iv) seleção das propriedades a extrair; v) escolha dos algoritmos a utilizar na classificação de emoções; e vi) definição dos modelos representativos das emoções detetadas.

Este modelo genérico de investigação pode ser assumido como referência para o processo de desenvolvimento de um sistema de deteção emocional e bem-estar, já que a sua especificação resulta da condensação das etapas seguidas pelos projetos da área de reconhecimento automático de emoções, de forma padronizada. Com base neste modelo acreditamos ser possível criar um sistema inteligente e informado ao nível dos afetos, i.e., um sistema capaz de reconhecer automaticamente as emoções dos trabalhadores de escritório e, desta forma, contribuir para o desenvolvimento futuro de componentes que possam reagir, reconfigurando o ambiente com o objetivo de promover um melhor bem-estar no local de trabalho.

Neste trabalho consideramos que as emoções são reações inconscientes a acontecimentos percebidos e que se manifestam no equilíbrio das componentes de saúde, tal como definida pela OMS (cf. bem-estar físico, social e mental (OMS, 2016)). Desta forma, é convicção do investigador ser possível criar sistemas informados do ponto de

vista emocional, i.e., sistemas capazes de monitorizar as emoções através da avaliação física, social, cultural e religiosa de trabalhadores em ambiente de escritório.

Com esta tese pretende-se, por isso, abordar a seguinte questão científica: "Será possível criar um sistema para recolha, integração e pré-processamento de dados de contexto abrangentes, com o objetivo de detetar automaticamente o bem-estar em trabalhadores de escritório, para que futuramente se possa usar esta informação na promoção da adaptação dinâmica do ambiente de trabalho?"

1.3 Notação utilizada

Durante esta tese foi feito um levantamento bibliográfico dos trabalhos realizados na área do AC. Este levantamento é resumido na secção 2.1. Como complemento, foi também publicado um Relatório Técnico (Carvalho et al., 2018) que relata, com maior detalhe, o levantamento do estado da arte efetuado na área da deteção automática de emoções. Este Relatório Técnico deu um importante suporte à preparação e execução da experiência de recolha, e também à execução da fase de pré-processamento do *dataset* recolhido.

O conhecimento obtido da análise das várias publicações consideradas no levantamento do estado da arte, foi representado com *tokens* representativos de informação, categorizados em quadros-resumo ao longo deste documento e do Relatório Técnico. Nestes *tokens* é utilizada a seguinte notação:

- As variáveis de contexto são apresentadas entre parêntesis (e.g. (EDA) refere-se à variável de contexto *Electrodermal Activity*).
- As propriedades de domínio (propriedades extraídas diretamente a partir do sinal de entrada, cujo método é confundível com o utilizado na aquisição do próprio sinal original) (ver secção 2.1.1 para mais informações), são apresentadas com a notação *OriginalSignal(DomainProperty)* para representar o encadeamento de sinais e respetivas propriedades de domínio (e.g. (ECG(HR)), significa propriedade de domínio (HR) extraída do sinal de (ECG)).
- Os instrumentos & sensores são identificados com chavetas (e.g. {MIBAND} refere-se à pulseira MiBand (XiaoMi, n.d.), {EDAMOVE} refere-se ao sensor EDAMove da (movisens GmbH, 2018c)).
- As técnicas de pré-processamento (incluindo encadeamento de técnicas), são representadas com parêntesis retos (e.g. (EDA) [-NOISE] significa a utilização da técnica [-NOISE] (remoção de ruído) aplicado ao sinal (EDA) (dados *raw* da atividade eletrodermal), (EDA) [LPF [DY/DX]], refere-se ao sinal resultante da aplicação da técnica [DY/DX] (primeira derivada), ao sinal resultante da técnica [LPF] (filtro passa-baixa), cujo *input* foram os dados (EDA) recolhidos do contexto).

- As propriedades extraídas (incluindo encadeamento de propriedades) são indicadas entre os sinais menor e maior (e.g. (PUPIL) <DIAMETER> refere-se à propriedade <DIAMETER> (diâmetro) da variável de contexto (PUPIL) (que significa pupila do olho), (TEMP) <MEAN> refere-se à propriedade <MEAN> (média), da variável representativa da temperatura ambiente (TEMP)).

1.4 Contribuições

Esta investigação pretende contribuir para a humanização dos sistemas e promover uma melhor interação entre pessoas e máquinas. Na tentativa de obter resposta à questão científica apresentada na secção 1.2, conduziu-se uma experiência com o objetivo principal de construir um *dataset* final amplamente rico, integrado, pré-processado e anotado.

A Tabela 1.1 resume as contribuições efetuadas no âmbito deste projeto de investigação.

Id	Descrição	Capítulo
C01	Relatório Técnico, publicado no Repositório Institucional da Universidade Fernando Pessoa (UFP) (Carvalho et al., 2018).	2
C02	Planeamento e execução da experiência que recolheu uma grande quantidade de dados em ambiente real.	3
C03	Instrumentos criados para apoiar no processo de recolha de dados de contexto dos participantes: i) {WINDOWSACTIVITY}, dados acerca da utilização do computador; e ii) {ANDROIDANALYTICS}, aplicação de questionários e recolha de dados relacionados com a interação social. ¹	3
C04	Questionários informais para recolha de outros dados: i) {SOCIOCULTURAL}, questionário para recolha de dados de cariz social e cultural; ii) {WELLBEING1} e {WELLBEING2} para recolha de dados <i>ground-truth</i> acerca da perceção do bem-estar; iii) {SLEEPTIRED} questionário para recolha de dados relacionados com o sono e cansaço; iv) {CALLINTERACTIONS} questionário para classificação das (CALL); e v) {LABOR}, ficheiro para recolha de dados de natureza laboral, relacionados com cada participante. ¹	3
C05	Foi também criada a documentação de suporte necessária à fase de execução da experiência: i) {SUBJECTFILE}, ficha de participação em experimencia; ii) {COLLECTIONSCRIPT} <i>script</i> a executar a cada <i>sprint</i> de recolha; iii) convite a enviar às organizações para participação na experiência; iv) <i>template</i> de resposta de aceitação por parte das organizações (sugestão de resposta); v) brochura de apresentação da experiência; vi) declaração de consentimento a assinar por cada participante antes de participar na experiência; e vii) formulários <i>backup</i> em papel para o {WELLBEING1}, {WELLBEING2}, {SLEEPTIRED}, {CALLINTERACTIONS} e {WINDOWSACTIVITY}.	3
C06	<i>Dataset raw</i> , conjunto de sub-pastas e ficheiros estruturados por organização, com os dados recolhidos de contexto dos trabalhadores de escritório que participaram na experiência dessa organização (i.e. sem qualquer pré-processamento).	3

¹ Os questionários {WELLBEING1}, {WELLBEING2}, {SLEEPTIRED} e {CALLINTERACTIONS}, foram aplicados pelo {ANDROIDANALYTICS}.

C07	Instrumentos criados para utilização na fase de integração: i) {WINDOWSACTIVITYIMPORT} para integração dos dados recolhidos pelo {WINDOWSACTIVITY}; ii) {ANDROIDANALYTICSIMPORT} para integração dos dados recolhidos pelo {ANDROIDANALYTICS}; e iii) {MISCELLANEOUSIMPORT} para integração dos demais ficheiros produzidos pelos instrumentos & sensores utilizados na experiência.	4
C08	Instrumentos criados para análise da qualidade de dados: i) {DQVA} para análise da qualidade dos dados durante o processo de integração; e ii) {DQVA-PP}, para análise da qualidade durante e após a aplicação de técnicas de pré-processamento.	4
C09	<i>Dataset</i> final integrado, pré-processado e anotado.	5

Tabela 1.1: Resumo de contribuições efetuadas

O projeto iniciou com o levantamento do estado da arte que culminou com a elaboração do Relatório Técnico (Carvalho et al., 2018), depositado no Repositório Institucional da Universidade Fernando Pessoa (UFP) (C01). Este relatório resume o trabalho relacionado na área do AC, recolhido durante a revisão bibliográfica efetuada sobre o tema. Nele, sugere-se um conjunto de *tokens* para melhor categorização da informação. Propõe-se também uma sistematização da informação, através da categorização desses *tokens* em quadros resumo, para permitir uma análise mais fácil e resumida das investigações abordadas.

Na experiência conduzida pelo autor, recolheu-se dados de origem multimodal em ambiente real, durante vários dias, de 20 participantes trabalhadores de escritório de 7 organizações distintas. A recolha foi efetuada em várias iterações porque só existiam disponíveis dois sensores EDAMove (movisens GmbH, 2018d). A segunda contribuição que se apresenta, diz respeito à preparação e execução da experiência em si (C02). Tratou-se de um processo complexo e bastante exigente. Na fase da preparação foi necessário criar um conjunto de ferramentas e documentação para que a recolha de dados corresse de forma organizada e sem percalços. A fase de execução foi pautada por diversas reuniões e visitas de acompanhamento. Além das reuniões com responsáveis das organizações e candidatos a participantes na experiência, fizeram-se também várias visitas de acompanhamento nas iterações de recolha. O objetivo principal das visitas, era garantir o cumprimento das regras comunicadas aos participantes, para promover a qualidade dos dados recolhidos (e.g. lembrar o preenchimento dos questionários, verificar a correta utilização dos instrumentos & sensores). Cada iteração terminava com a organização e arquivo dos dados produzidos pelos vários instrumentos & sensores. O processo de preparação e execução é abordado com mais detalhe na secção 3.

Para apoiar no processo de recolha foram desenvolvidos os instrumentos a seguir enumerados (C03):

- {WINDOWSACTIVITY}, aplicação criada para recolher dados de utilização do computador de trabalho dos participantes, tendo por base o registo de eventos do próprio

Sistema Operativo (SO) (e.g. *login*, *logout*).¹.

- {**ANDROIDANALYTICS**}, aplicação criada para recolher dados das chamadas e SMS enviados e recebidos. Este instrumento foi também utilizado para aplicar os questionários {**WELLBEING1**}, {**WELLBEING2**}, {**SLEEPTIRED**} e {**CALLINTERACTIONS**}.¹

Além destes instrumentos, criou-se também um conjunto de questionários tendo em vista a recolha de *ground-truth* e mais dados de contexto dos participantes (**C04**):

- {**SOCIOCULTURAL**}, questionário para recolher dados de contexto sócio-cultural (modelo no Anexo 7.1).
- {**WELLBEING1**}, questionário para recolher dados *ground-truth* acerca da perceção do bem-estar, ao final do dia (modelo no Anexo 7.2).
- {**WELLBEING2**}, questionário para recolha dados *ground-truth* acerca da perceção do bem-estar, aleatoriamente durante o dia (modelo no Anexo 7.3).
- {**SLEEPTIRED**}, questionário para recolher dados acerca da perceção da qualidade do sono e sensação de descanso no início de cada dia de recolha (modelo no Anexo 7.4).
- {**CALLINTERACTIONS**}, questionário para recolha e classificação das chamadas efetuadas e recebidas (modelo no Anexo 7.5).
- {**LABOR**}, questionário para recolha de dados acerca de períodos de férias, faltas e atrasos ao trabalho, dos participantes na experiência (ficheiro preenchido pelos departamentos de recursos humanos das várias organizações) (modelo no Anexo 7.6).

Foram também preparados os seguintes documentos de suporte necessários ao acompanhamento da execução da experiência (**C05**):

- {**SUBJECTFILE**}, ficha de participação em experiência com informação da organização e do participante: i) dados gerais de cada participante (cf. género, idade, peso, altura); ii) dias em que a recolha foi efetuada; número da pulseira MiBand (XiaoMi, n.d.) e do sensor {**EDAMove**} (movisens GmbH, 2018c) utilizados; e iii) pulsos do participante utilizados em cada um dos sensores (ver modelo no Anexo 7.7).
- {**COLLECTIONSCRIPT**}, *script* a executar a cada *sprint* de recolha: angariação de participantes; tarefas a executar no dia anterior ao início da recolha (e.g. preenchimento da ficha de participação, instalação de {**WINDOWSACTIVITY**} e {**ANDROIDANALYTICS**}, atribuição e formação acerca de sensores); tarefas a executar no primeiro dia de recolha; tarefas a executar no último dia de recolha; etc. (ver modelo no Anexo 7.8).
- Convite às organizações para participação na experiência (carta a endereçar a cada organização candidata a participar na experiência) (ver Anexo 7.9).

¹Consulte a secção 3.1 para mais detalhes acerca deste instrumento

-
- Proposta de resposta de aceitação por parte das organizações, tendo em vista facilitar o processo administrativo de responder e, essencialmente, servir de guia para a redação do conteúdo (ver Anexo 7.10);
 - Brochura de apresentação da experiência a disponibilizar aos responsáveis das organizações e candidatos a participante, para ajudar na decisão individual de participar na experiência (ver Anexo 7.11).
 - Declaração de consentimento a assinar por cada participante antes de participar na experiência (ver Anexo 7.12).
 - Formulários *backup* em papel para {WELLBEING1} (ver Anexo 7.16), {WELLBEING2} (ver Anexo 7.17), {SLEEPTIRED} (ver Anexo 7.18), {CALLINTERACTIONS} (ver Anexo 7.19), {WINDOWSACTIVITY} (ver Anexo 7.20). Estes formulários *backup* em papel foram criados antecipadamente, para permitir continuar com a experiência caso surgisse alguma impossibilidade na utilização do *smartphone* ou computador de trabalho.

O objetivo principal da fase experimental, foi obter um *dataset raw* multimodal, com dados de contexto fisiológico, social e religioso (C06). O *dataset raw* que se apresenta é composto por 46 variáveis de contexto.

Na fase de pré-processamento, foi também necessário criar alguns instrumentos com o objetivo de tornar possível o processo de integração (C07) (ver secção 4.1):

- {WINDOWSACTIVITYIMPORT}, aplicação criada para integração dos dados *sqlite* (SQLite Consortium, 2021) recolhidos através da aplicação {WINDOWSACTIVITY}.
- {ANDROIDANALYTICSIMPORT}, aplicação criada para integração dos dados *sqlite* recolhidos através da aplicação {ANDROIDANALYTICS}.
- {MISCELLANEOUSIMPORT}, projeto *Extract, Transform and Load* (ETL), implementado em *Microsoft SQL Integration Services* (MSSIS) (Microsoft, 2018), para integração dos dados de questionários e ficheiros *Comma-separated Values* (CSV) e *Microsoft Excel Open XML Spreadsheet* (XLSX) (Microsoft, n.d.b) que constituem o *dataset raw*.

Uma das preocupações com os algoritmos de classificação é a qualidade dos dados de *input* (Gama et al., 2012; Kandel et al., 2012; Chu et al., 2016; Jeatrakul et al., 2010; Bakhtiyari et al., 2015; Tae et al., 2019; Urges, 2021). Como se utilizaram vários instrumentos, a recolha ocorreu em ambiente real em várias organizações, e durante vários dias, a avaliação de qualidade dos dados na fase de pré-processamento revestiu-se de especial importância para a experiência. Só com um *dataset raw* de qualidade, se conseguirá promover a posterior qualidade da classificação e conseqüente confiança nos resultados obtidos. Foi neste contexto, que se decidiu criar uma ferramenta utilizada na fase de pré-processamento para analisar os dados recolhidos de contexto: *Data Quality Visual Analysis* {DQVA} (C08) (ver secção 4.4):

-
- **{DQVA}**, para apoiar o processo de análise de qualidade de dados durante a integração dos mesmos.
 - **{DQVA-PP}**, variante do **{DQVA}** para carregamento de dados com base nas *views* criadas no pré-processamento (ver secção 5.2 para mais informações), para apoiar no processo de análise de qualidade de dados, durante e após a aplicação de técnicas de pré-processamento.

Por fim, o *dataset* final pré-processado e anotado surge como uma das principais contribuições (C09). Este *dataset* foi desenvolvido com base numa arquitetura hierárquica de *views* para permitir enriquecer o *dataset* com propriedades extraídas de forma normalizada, sem sobreposição os dados *raw* originais. As várias *views* implementadas, permitem uma extração de dados mais flexível e personalizável. O *dataset* final é apresentado com maior detalhe na secção 5.1. Este *dataset* poderá ser usado futuramente para o desenvolvimento e avaliação de componentes que permitam detetar automaticamente o bem-estar em ambientes de escritório.

1.5 Plano de trabalho e estrutura da tese

O principal objetivo definido para este projeto foi a obtenção de um *dataset* final, integrado, pré-processado, anotado, e contendo dados recolhidos em contexto de trabalho de ambientes reais. Pretendia-se que o *dataset* fosse constituído com dados multimodais (cf. contexto fisiológico, social e religioso).

Um dos problemas habituais nas experiências que envolvem pessoas é a angariação de participantes. A forte ligação do autor a várias organizações do mundo empresarial e, mais concretamente, o convívio quase diário deste com os funcionários de escritório dessas organizações, permitiu antecipar uma forte probabilidade de se conseguir várias pessoas auto-propostas a participar na experiência.

Neste contexto, auscultou-se a receptividade de várias organizações e de vários participantes por organização. Como o tempo de empréstimo dos sensores {EDAMove} (movisens GmbH, 2018d) era limitado, havia que escolher entre recolher em poucas organizações vários participantes, ou diversificar a recolha por várias organizações. Optou-se pela segunda por se considerar ser preferível diversificar os ambientes de recolha (i.e. vários escritórios).

A Figura 1.2 apresenta o cronograma resumido do projeto. Antevia-se uma fase de recolha complexa e longa, porque se pretendia ir mais longe no número de participantes do que as demais investigações existentes. Quando se iniciou esta fase, houve a perceção que se avizinharia também uma fase de pré-processamento exigente e demorada, fruto da recolha de dados em ambiente real e da utilização de vários instrumentos & sensores.

Os trabalhos iniciaram-se com a análise em detalhe do trabalho já realizado noutras

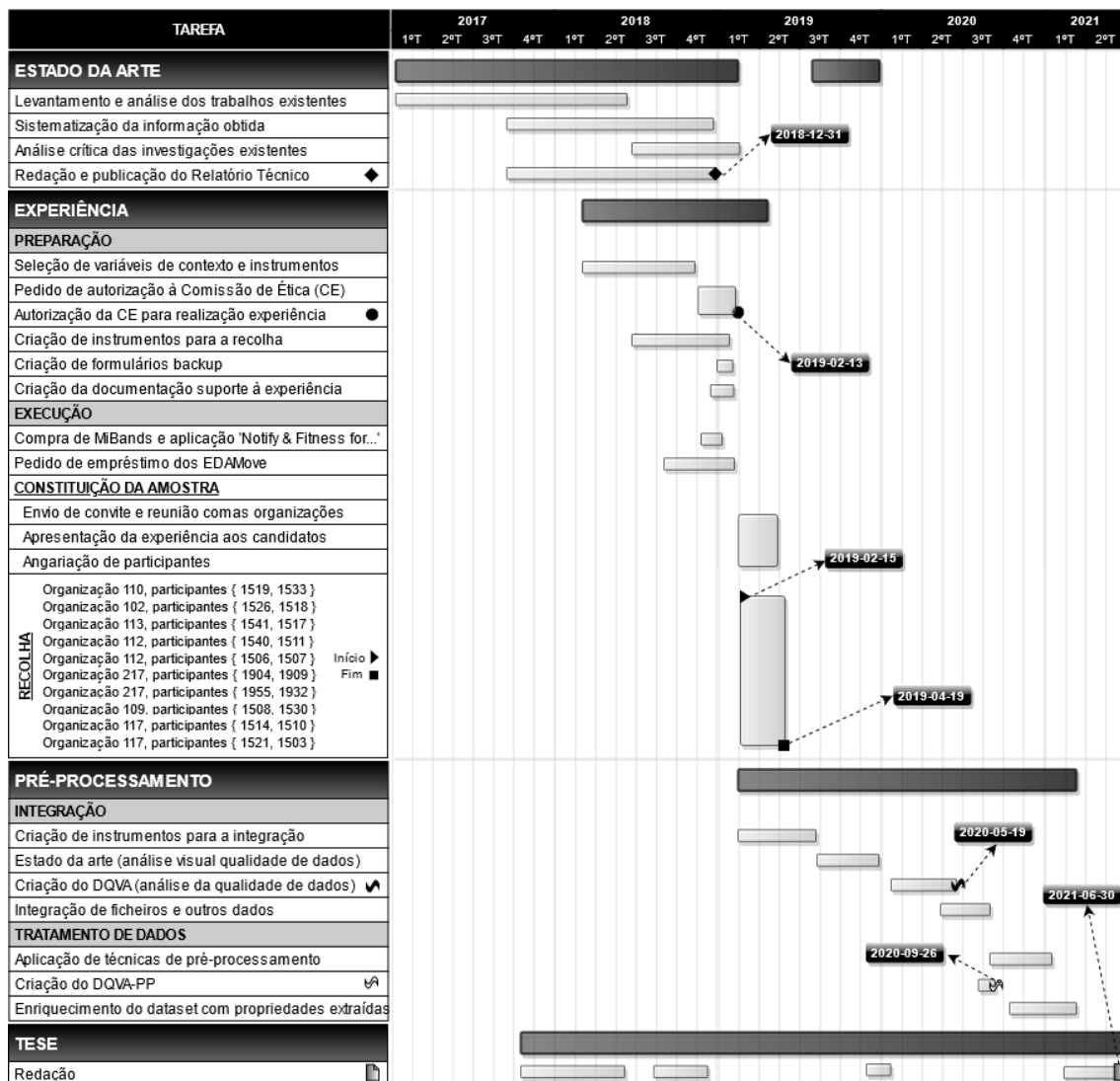


Figura 1.2: Cronograma do projeto

investigações. Desta fase resultou um Relatório Técnico (C01), depositado no Repositório Institucional da Universidade Fernando Pessoa (UFP). Além de permitir conhecer o trabalho já realizado por outros investigadores, este relatório permitiu reter o conhecimento adquirido através da utilização de *tokens* sintetizadores de informação, categorizados em quadros-resumo. Permitiu também resumir a sequência de etapas transversalmente utilizada pelos investigadores do AC (Figura 1.1). Este modelo genérico de investigação foi assumido como referência para o processo de desenvolvimento da experiência conduzida pelo autor.

A fase experimental iniciou-se logo após a aprovação do projeto pela Comissão de Ética (CE). Até essa data, foram criadas as ferramentas necessárias para apoiar o processo de recolha (ver secção 3.1.2) e a documentação de suporte à recolha de dados propriamente dita (ver secção 3.1.3). Como é visível no cronograma (Figura 1.2), as tarefas relacionadas com a constituição da amostra e a recolha de dados de contexto em ambi-

ente real, decorreram durante mais de dois meses. A fase da integração foi mais longa do que a fase experimental. Além dos instrumentos que se tiveram que criar para permitir a integração de dados numa base de dados relacional (ver secção 4.1), também se tiveram que criar instrumentos para facilitar a análise da qualidade de dados do *dataset* durante a fase da integração e, posteriormente, na fase da aplicação de técnicas de pré-processamento para enriquecimento do *dataset* (ver secção 4.4).

O restante conteúdo deste documento será organizado em função da ordem dos trabalhos executados no projeto. Cada capítulo termina com uma descrição e análise resumida das contribuições apresentadas nesse mesmo capítulo.

O capítulo 2 apresenta o resumo das investigações analisadas remetendo para o Relatório Técnico (Carvalho et al., 2018) onde são explanadas com mais detalhe (contribuições: C1). No Capítulo 3 descreve-se o planeamento e execução da experiência realizada (contribuições: C2, C3, C4, C5 e C6). O Capítulo 4 descreve o processo de pré-processamento efetuado desde a integração até à aplicação de técnicas de pré-processamento para melhorar a qualidade aos dados (contribuições: C7 e C8). No Capítulo 5 discutem-se os resultados obtidos e descreve-se o *dataset* final (contribuições: C9). Termina-se com o Capítulo 6 onde é apresentada uma conclusão do trabalho.

1.6 Proteção de dados

A privacidade é um tema de grande importância para o autor. Defende-se o respeito pelos dados pessoais dos participantes, adotando procedimentos que assegurem a privacidade e segurança dos dados tratados. O autor compromete-se a utilizar meios técnicos e organizativos adequados à promoção da confidencialidade e segurança de todos os dados recolhidos.

Apenas foram recolhidos os dados necessários, adequados e pertinentes à experiência. Respeitou-se o princípio da minimização da recolha. Antes da recolha todos os participantes foram informados acerca dos dados que iriam ser recolhidos e a respetiva justificação para o tratamento, respeitando o princípio da transparência. Os dados não serão utilizados para outros fins que não os previamente comunicados aos respetivos titulares, respeitando o princípio da lealdade e da confiança depositada pelos participantes na investigação e no investigador. Apenas foram convidadas organizações que demonstraram preocupação e respeito pelos princípios da privacidade. Não foram, nem serão, comunicados dados pessoais às organizações (entidades empregadoras) antes, durante ou após a experiência. Os dados recolhidos nunca serão partilhados com terceiros e ficarão apenas guardados em recursos que estejam sob controlo do autor e para fins de investigação.

Na experiência apresentada nesta tese foi utilizado o modelo de auto-proposição com o objetivo de promover a auto-seleção pelos próprios candidatos. Não existiu nenhum processo de *screening* nem de verificação de cumprimento dos critérios de exclusão. Os

participantes na experiência auto-propuseram-se, i.e., voluntariaram-se livremente. O convite à participação dos trabalhadores foi efetuado, num primeiro momento, através de uma brochura disponibilizada publicamente (ver Anexo 7.11) e, posteriormente, num evento público coletivo onde foi apresentado o projeto oralmente. Neste evento, além de se esclarecerem as dúvidas colocadas, foi explicado o modelo da auto-proposição. Foi também apresentado o objetivo da experiência e justificada a necessidade dos dados a recolher. Procurou-se adicionalmente demonstrar o compromisso para com a privacidade de todos. Por fim, foi também abordada a necessidade de obtenção de consentimento informado e a possibilidade deste ser retirado, a qualquer momento, sem necessidade de justificação.

As declarações de consentimento foram recolhidas antes dos respetivos *sprints* de recolha (*template* no Anexo 7.12). Esta experiência foi autorizada pela Comissão de Ética (CE) da UFP em 2019-02-13 (ver Anexo 7.13).

Capítulo 2

ESTADO DA ARTE

Medir emoções é um dos maiores problemas da ciência afetiva e do estudo da apazibilidade. A grande quantidade de fatores que afetam o estado emocional como a fisiologia, experiência subjetiva, antecedentes e comportamento (Babiker et al., 2013), fazem com que as emoções não possam ser medidas diretamente (Nawasalkar et al., 2013), porque cada pessoa pode reagir de forma diferente perante o mesmo estímulo emocional (Raudonis, 2013). A forma como se experiencia uma emoção varia em função do tempo, do espaço, e da componente social e cultural de cada pessoa (Kim and André, 2008; Dong et al., 2011).

A criação de um sistema para reconhecimento automático de emoções é um desafio (Gunes and Piccardi, 2007). Apesar do estudo da cognição humana ter progredido muito nos últimos anos, a ideia do que é exatamente uma emoção está longe de ser fechada (Mandryk and Atkins, 2007). Existem já muitas investigações na área emocional, mas continua a não existir uma resposta definitiva para a pergunta: "o que é uma emoção?" (Scherer, 2005). A detecção automática de emoções será uma tarefa tão complicada quanto definir o termo - emoção - por não se conhecer o que se procura de forma objetiva.

A entropia que envolve o tema emocional é perceptível na literatura. Bastará analisar a quantidade e diversidade de termos referidos como emoções, para perceber a dificuldade existente na definição do que se sente. Apesar de alguns termos serem facilmente diferenciáveis (e.g. alegria e tristeza, bem-estar e depressão), outros parecem redundar-se não se percebendo as diferenças ou a fronteira que os distingue: alegria e felicidade, entusiasmo e excitação, medo e preocupação, etc. Existem outros termos ainda que, apesar de serem referenciados como emoções, parecem ser mais consequências emocionais do que emoções propriamente ditas (e.g. choro, isolamento, agitação).

O reconhecimento emocional é uma área de investigação interdisciplinar que envolve vários campos do conhecimento: ciências da computação, ciências cognitivas, e psicologia (Aracena et al., 2016). A capacidade dos sistemas reconhecerem as emoções das pessoas e adequar as suas ações em função desse reconhecimento, é uma área de investigação do *Human-Computer Interaction* (HCI) (Gunes and Piccardi, 2007; Jerritta et al., 2011).

Dotar os computadores e robots com a capacidade de entender emoções, é uma área de estudo do *Affective Computing* (AC) (Gogia et al., 2016; Zhai and Barreto, 2006). O reconhecimento automático de emoções, facilitará a interação entre máquinas e pessoas (Jerritta et al., 2011; Alabdulkarim, 2015; Rani and Sarkar, 2006).

O reconhecimento de emoções tem sido alvo de investigação em vários domínios: i) na indústria automóvel (e.g. os projetos de Paschero e Gutmann estudaram as emoções durante a condução (Paschero et al., 2012; Gutmann et al., 2015)); ii) cinematográfica e jogos (e.g. o estudo de Guillotel pretendia recolher continuamente a resposta emocional durante a visualização de um filme (Guillotel et al., 2013)); iii) *Human-Robot Interaction* (HRI) (e.g. Luefeng et al. apresentaram robots que conseguiam identificar e reagir em tempo real às emoções das pessoas (Luefeng et al., 2016)); iv) *Brain-Computer Interaction* (BCI) (e.g. Liu et al. propuseram um sistema em tempo real para detetar emoções a partir do eletroencefalograma (EEG) (Liu et al., 2010)); v) HCI (e.g. Jamshidnejad et al. estudaram o reconhecimento emocional para registar as emoções dos utilizadores enquanto navegavam e compravam online (Jamshidnejad and Jamshidined, 2009)); etc.

Os sistemas em geral, apesar de mais autónomos, interagem cada vez mais com os humanos (Rani et al., 2003). Neste contexto, os investigadores em AC, estudam formas de desenvolver sistemas capazes de reconhecer e exprimir afetos, com o objetivo de melhorar o relacionamento entre as pessoas e as máquinas. São de esperar novas linhas de investigação interdisciplinares que olhem para o HCI, não só como uma forma de promover a melhor experiência de utilização possível, mas também para satisfazer as necessidades do utilizador ao nível emocional (Bos, 2010).

O autor deste documento, está particularmente interessado nas investigações relacionadas com a deteção emocional em intersecção com as linhas de investigação do AC, por causa da experiência em curso que pretende preparar um *dataset* final para, no futuro, detetar automaticamente o bem-estar em trabalhadores de escritório. Assim, neste capítulo, faz-se o levantamento do estado da arte relacionado com a deteção automática de emoções em dois eixos: i) trabalhos já realizados relacionados com a deteção automática de emoções (cf. variáveis de contexto normalmente recolhidas, instrumentos & sensores utilizados, técnicas de pré-processamento aplicadas, e propriedades extraídas); e ii) revisão da literatura relacionada com as investigações incidentes sobre os trabalhadores de escritório, com foco principal nas relacionadas com a deteção automática do bem-estar.

2.1 Reconhecimento de emoções

Esta secção resume as obras estudadas na área do AC. A quantidade de investigações relacionadas com a deteção automática de emoções tem crescido e continuará a crescer nos próximos anos. Esta probabilidade motivou o autor a preparar um documento dinâmico que pudesse ser complementado com nova informação de forma rápida e flexível. A

ideia era facilitar a categorização de nova informação, ao longo das várias iterações de atualização do estado da arte. Esse documento, assume atualmente a forma de um Relatório Técnico (RT), e está disponível para consulta no Repositório Institucional da Universidade Fernando Pessoa (UFP) (Carvalho et al., 2018).

O estado da arte desta tese é um resumo do RT. A estrutura do RT foi pensada para permitir o registo de forma sistemática, sintética e organizada, da informação obtida a partir da revisão do estado da arte de trabalhos realizados na deteção automática de emoções. A estratégia seguida na organização do documento passava por: i) ancorar a informação obtida a secções representativas das diferentes etapas do *pipeline* genérico para deteção e representação de emoções (Figura 1.1); e ii) categorizar essa informação em quadros-resumo, através de *tokens* representativos do conteúdo, para facilitar a procura e leitura da informação sistematizada.

Como já anteriormente referido, os estudos relacionados com a deteção automática de emoções, seguem uma sequência genérica e transversal de etapas. Para facilitar a leitura deste documento, esta secção foi organizada em função desse modelo genérico que é transversalmente assumido nas investigações para deteção emocional. Inicia-se com o resumo das variáveis de contexto e propriedades de domínio utilizadas pelos autores (secção 2.1.1). Depois são apresentados os instrumentos & sensores utilizados na recolha das variáveis de contexto (secção 2.1.2). Posteriormente são resumidas as técnicas de pré-processamento utilizadas (secção 2.1.3). Conclui-se enumerando as propriedades extraídas mais utilizadas nas obras estudadas (secção 2.1.4).

Este levantamento bibliográfico serviu de suporte à preparação e execução da experiência do autor, que pretendia contribuir com um *dataset* multimodal, integrado e pré-processado, peça fundamental para detetar o bem-estar em trabalhadores de escritório.

A Tabela 2.1, resume as investigações estudadas no âmbito do reconhecimento automático de emoções. Como cada investigação é referenciada várias vezes ao longo das próximas sub-secções, preferiu-se apresentar antecipadamente cada projeto envolvido neste levantamento. Estas investigações serão à frente referenciadas para dar exemplos: i) de variáveis de contexto e/ou propriedades de domínio recolhidas; ii) instrumentos & sensores utilizados; iii) técnicas de pré-processamento aplicadas; e iv) propriedades extraídas a partir dos dados recolhidos.

Investigação	Resumo	Emoções
(Perdiz et al., 2017)	Apresentam uma <i>framework</i> que combina a utilização do Electromiografia (EMG) e do Electroculografia (EOG) para detetar emoções.	Fúria, tristeza, alegria e neutral.
(Lee et al., 2016)	Propõem um novo método dinâmico de reconhecimento, com base em expressões faciais, para resolver o problema dos desencontros temporais em sequências de vídeo (e.g. duração das transições).	Fúria, desprezo, desgosto, medo, alegria, tristeza e surpresa.

(Eckert et al., 2016)	Apresentam um algoritmo rápido e simples para <i>emotion-awareness</i> , capaz de executar em <i>realtime</i> e em <i>background</i> , para ser reutilizado por outras aplicações (e.g. jogos).	Fúria, desgosto, medo, tristeza, alegria e surpresa.
(Matlovic et al., 2016)	Compararam a deteção de emoções utilizando dados recolhidos pelo Eletroencefalograma (EEG) e Atividade Eletrodermal (EDA), com ferramentas já existentes que fazem reconhecimento emocional através das expressões faciais.	Alegria, surpresa, tristeza, medo, desgosto, fúria e estado neutral.
(Gogia et al., 2016)	Propõem um sistema com recolha multimodal para deteção emocional, a ser utilizado em ambiente de aprendizagem.	Estado de atenção.
(Zhang et al., 2016)	Apresentam uma base de dados emocional multimodal, anotada, e multi-étnica.	Alegria, surpresa, tristeza, susto, embaraço, medo, dor, fúria e desgosto.
(Sano and Eng, 2016)	Estudam o impacto de vários fatores (e.g. interação social), no sono, stress, disposição, e bem-estar em geral.	Stress, disposição e bem-estar.
(Zhao et al., 2016)	Inferirem emoções humanas usando a reflexão dos sinais <i>wireless</i> em substituição do eletrocardiograma (ECG) tradicional.	Alegria, prazer, tristeza e fúria.
(Zenonos et al., 2016)	Reconhecem a disposição das pessoas catalogando-a em cinco níveis de intensidade. Os investigadores vêm a disposição como o resultado das emoções sentidas.	Excitação, alegria, calma, cansaço, tédio, tristeza, stress e fúria.
(Basu et al., 2016)	Apresentam um sistema simples de reconhecimento emocional considerando 9 sinais fisiológicos com recolha não-invasiva.	<i>Valence e arousal</i> .
(Aracena et al., 2016)	Apresentam uma abordagem de reconhecimento emocional com base na informação cedida pela pupila do olho.	Estados emocionais: positivo, neutral e negativo.
(Adams and Robinson, 2015)	Apresentam um sistema capaz de reconhecer estados afetivos complexos de forma categórica.	Receio, fúria, vergonha, tédio, deceção, desgosto, animação, frustração, alegria, mágoa, interesse, brincadeira, orgulho, tristeza, matreirice, surpresa, preocupação e neutral.
(Turan et al., 2015)	Propõem um algoritmo adaptativo de seleção baseado em descritores, capaz de determinar as duas melhores propriedades de cada classe de expressão, com os objetivos de as fundir e assim conseguir um melhor reconhecimento.	Fúria, desgosto, medo, surpresa, tristeza e alegria.
(Korkmaz and Atasoy, 2015)	Investigam o conteúdo emocional existente no discurso oral (através da leitura de enunciados). Os autores utilizam os <i>Mel Frequency Cepstral Coefficients (MFCC)</i> .	Alegria, fúria, tristeza e neutral.
(Lalitha et al., 2015)	Apresentam um método para identificar as propriedades <i>Discrete Wavelet Transform (DWT)</i> , porque as consideram as mais adequadas a um reconhecimento emocional com maior precisão.	Ansiedade, desgosto, alegria, tédio, tristeza, fúria e neutral.
(Singh et al., 2015)	Propõem um sistema para detetar emoções e o estado mental de uma pessoa através da sua postura corporal.	Disposição, confusão e dúvida, calma e estado neutral, preguiça ou desinteresse, e fúria.
(Murali et al., 2015)	Apresentam um dispositivo vestível pequeno e leve, para monitorização da saúde física e emocional.	<i>Arousal e valence</i> .
(Jaques et al., 2015)	Utilizaram um algoritmo de <i>Machine Learning (ML)</i> , para distinguir os alunos alegres dos não alegres. O objetivo do estudo era prever a depressão.	Felicidade.

(Cruz et al., 2015)	Apresentam um sistema para reconhecimento automático da expressão emocional da face com base na movimentação dos olhos.	Emoções humanas.
(Saha et al., 2014)	Criaram um sistema para classificar emoções com base em gestos do corpo.	Fúria, medo, alegria, tristeza e relaxamento.
(Matiko et al., 2014)	Classificam emoções como positivas ou negativas, através de um <i>fuzzy algorithm</i> , utilizando o EEG como sinal de entrada.	<i>Valence</i> .
(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014)	Propõem um sistema não obstrutivo para reconhecimento do stress com base em métricas de comportamento (e.g. utilização do smartphone, perfil de personalidade, tempo).	Stress.
(Agrawal et al., 2013)	Criaram um sistema para detecção de fadiga e desatenção por parte do condutor, com base nos gestos da face e nas emoções deduzidas desses gestos.	Alegria, fúria, tristeza e surpresa.
(Soleymani et al., 2013)	Detetam a agradabilidade (<i>valence</i>) de forma contínua durante a visualização de um vídeo, com base na análise simultânea do EEG e da expressão facial.	<i>Valence</i> .
(Vermun et al., 2013)	Criaram um sistema mais preciso para reconhecimento do estado afetivo e cognitivo com base em gestos.	Alegria, ansiedade, desgosto, irritação, interesse, nível de alerta, tédio e fadiga.
(Kusserow et al., 2013)	Monitorizam o stress em oradores, músicos, atletas olímpicos e nas pessoas em geral.	Stress.
(Alzoubi et al., 2013)	Estudaram a variação no tempo das propriedades fisiológicas que permitem detetar o afeto.	Afeto, <i>arousal</i> e <i>valence</i> .
(Nawasalkar et al., 2013)	Propõem um sistema para reconhecer os efeitos de música clássica e rock no corpo humano. Durante a sua experiência, os autores provocam as seguintes emoções: alegria, surpresa, tristeza e excitação.	Stress.
(Sano and Picard, 2013b)	O objetivo do estudo era encontrar sinais fisiológicos ou de comportamento que funcionem como marcadores de stress.	Stress.
(Raudonis, 2013)	Desenvolveram um sistema de reconhecimento de emoções com base nos movimentos do olho.	Neutral, desgosto, diversão e interesse.
(Kawai et al., 2013)	Investigaram a variação do diâmetro da pupila com o objetivo de conseguir medir o afeto positivo ou negativo de pessoas que não conseguem mover a cabeça ou mãos para comunicar de outra forma.	Afetos positivos e negativos.
(Babiker et al., 2013)	Identificaram as diferenças de diâmetro da pupila resultantes dos estados emocionais individuais positivos ou negativos.	Emoções positivas e negativas.
(LiKamWa et al., 2013)	Desenvolveram uma aplicação para <i>smartphone</i> (Mood-Scope) capaz de inferir a disposição do seu utilizador.	Disposição.
(Murad and Malkawi, 2012)	Detetaram emoções humanas através de um modelo <i>neurofuzzy</i> .	Tristeza, alegria, medo, embaraço, desgosto, ansiedade, raiva, suspense, alívio, orgulho e diversão.
(Chang et al., 2012)	Propõem um sistema baseado em sinais fisiológicos para reconhecer a fúria.	Fúria.
(Bauer and Lukowicz, 2012)	Apresentam um sistema para detetar situações de stress através de sensores existentes no <i>smartphone</i> .	Stress.
(Yang and Bhanu, 2011)	Criaram um <i>avatar</i> representativo das emoções de um vídeo ou sequência de imagens.	Fúria, medo, alegria, alívio e tristeza.

(Dhall et al., 2011)	Propõem um método de reconhecimento automático de emoções.	Fúria, medo, alegria, alívio e tristeza.
(Mokhayeri et al., 2011)	Detetam o stress através da dimensão da pupila.	Stress.
(Hernandez et al., 2011)	Estudam o stress num ambiente de <i>call center</i> .	Stress.
(Lane et al., 2011)	Propõem uma aplicação (BeWell) para promover o estilo de vida saudável com base no comportamento diário do utilizador.	Bem-estar.
(Lane et al., 2010)	Propõem um sistema eficiente para detetar a fadiga em condutores em tempo real, com base no estado do olho.	Fadiga.
(Bos, 2010)	Criaram um sistema de baixo custo e de fácil instalação para reconhecimento emocional (representam as emoções reconhecidas no modelo bidimensional <i>arousal/valence</i>).	<i>Arousal e valence</i> .
(Liu et al., 2010)	Criaram um sistema de reconhecimento de emoções em tempo real, e representam as emoções reconhecidas através de um <i>avatar</i> .	Medo, frustração, tristeza, alegria, agradabilidade e satisfação.
(Setz et al., 2010)	Distinguem o stress de excesso de trabalho cognitivo em ambiente de escritório	Stress.
(Kim and André, 2008)	Investigaram o potencial dos sinais fisiológicos no reconhecimento de emoções durante a audição de música.	<i>Valence e arousal</i> .
(Lichtenstein, Antje; Oehme, 2008)	Estudam emoções durante a visualização de filmes através da recolha de parâmetros fisiológicos.	<i>Valence e arousal</i> .
(Bradley et al., 2008)	Avaliaram o efeito hedónico do <i>valence</i> e do <i>arousal</i> emocional nas respostas da pupila.	<i>Valence e arousal</i> .
(Gunes and Piccardi, 2007)	Criaram um sistema para reconhecimento afetivo, com base na expressão facial e nos gestos da parte superior do corpo.	Ansiedade, fúria, desgosto, medo, alegria e incerteza.
(Castellano et al., 2007)	Tentaram descobrir quais as pistas do movimento do corpo que melhor classificam emoções, e criaram um modelo para reconhecimento emocional com base na análise de vídeo.	Fúria, alegria, prazer e tristeza.
(Mandryk and Atkins, 2007)	Detetaram e quantificaram emoções no contexto de tecnologias afetivas (e.g. jogos digitais), para evitar a medição apenas no final através de questionários.	Tédio, desafio, excitação, frustração e diversão.
(Sebe et al., 2006)	Conseguiram mais precisão na inferência de estados emocionais através da combinação das pistas dadas pela expressão facial e informação vocal.	Alegria, surpresa, fúria, desgosto, medo, tristeza, interesse, tédio, confusão e frustração.
(Zhai and Barreto, 2006)	Apresentaram um sistema para deteção de stress com base em sinais fisiológicos recolhidos por sensores não intrusivos.	Stress.
(Healey and Picard, 2005)	Apresentaram métodos para recolher e analisar dados fisiológicos durante a condução real, com o objetivo de determinar o nível de stress.	Stress.
(Herbon et al., 2005)	Encontraram variáveis fisiológicas capazes de mapear para emoções de forma não ambígua, para utilização em sistemas digitais.	<i>Arousal e valence</i> .
(Partala et al., 2005)	Desenvolveram um sistema automático que, em tempo real, conseguisse estimar emoções a partir do EMG na face.	<i>Valence</i> .
(van Eck et al., 2005)	Analisaram os efeitos de stress percebido com estados de humor e eventos diários stressantes, nos níveis de cortisol na saliva.	Stress e disposição (humor).

(Busso et al., 2004)	Combinaram os dados da expressão facial e oral com o objetivo de melhorar a precisão no reconhecimento de emoções.	Tristeza, fúria, alegria e estado neutral.
(Lisetti and Nazoz, 2004)	Desenvolveram um sistema HCI sensível às emoções dos utilizadores, capaz de responder às emoções do utilizador em função do contexto e da aplicação.	Tristeza, raiva, surpresa, medo, frustração e diversão.
(Kim et al., 2004)	Desenvolveram um sistema de reconhecimento de emoções independente do utilizador.	Tristeza, fúria, stress e surpresa.
(Haag et al., 2004)	Apresentaram um procedimento para treinar computadores para reconhecer emoções a partir de vários tipos de biossensores.	<i>Arousal e valence.</i>
(Partala and Surakka, 2003)	Estudaram a variação do tamanho da pupila com a aplicação de estímulos sonoros.	<i>Arousal e valence.</i>
(Harmer et al., 2003)	Estudaram o efeito da serotonina em distúrbios emocionais.	Alegria, tristeza, medo, fúria e desgosto.
(Nwe et al., 2001)	Propõem um sistema para classificar automaticamente emoções com base no discurso oral.	Fúria, antipatia, medo, alegria, tristeza e surpresa.
(Buchanan and Lovallo, 2001)	Estudaram o impacto do cortisol na memória humana, utilizando imagens para estimular a <i>arousal</i> emocional.	<i>Arousal.</i>
(Healey et al., 2000)	Estudaram o stress durante a condução.	Stress durante a condução, raiva, ódio, tristeza, amor, alegria, reverência, e estado neutral.
(Vrijkotte et al., 2000)	Estudaram o stress no trabalho.	Stress.
(Ritz et al., 2000)	Estudaram o impacto de diferentes estados emocionais e stress na <i>Oscillatory resistance</i> (ROS) em asmáticos e não asmáticos.	Ansiedade, fúria, depressão, alegria, felicidade, contentamento, neutral, e stress.
(Chen et al., 1998)	Integraram áudio e vídeo para promover um melhor reconhecimento de emoções.	Alegria, tristeza, fúria, antipatia, surpresa e medo.
(Healey and Picard, 1998)	Apresentaram um método para reconhecimento de padrões em dados de contexto, representativos de estados afetivos do utilizador.	Fúria, ódio, aflição, amor familiar, amor romântico, alegria e reverência (respeito profundo). <i>Arousal e Valence.</i>
(Sinha, 1996)	Detetaram emoções no ser humano, utilizando EMG nos músculos da face e o EOG para monitorizar a movimentação vertical e horizontal do olho.	Medo, alegria, tristeza e raiva.
(Vrana, 1993)	Estudaram as diferenças entre as emoções classificadas fisiologicamente e as reportadas pelos participantes. Diferenciaram as emoções negativas e estudaram as reações musculares da face.	Desgosto, fúria, prazer e alegria.
(Sinha et al., 1992)	Estudaram os padrões da atividade cardiovascular durante diferentes estados emocionais.	Alegria, tristeza, medo e fúria.

Tabela 2.1: Investigações consideradas no levantamento bibliográfico em AC

2.1.1 Variáveis de contexto e propriedades de domínio

A relação entre os dados recolhidos de contexto com os estados emocionais do Homem são ainda pouco conhecidos (Raudonis, 2013). A intangibilidade das emoções torna im-

possível a sua medição direta e objetiva (Murad and Malkawi, 2012). A reação emocional perante o mesmo estímulo é diferente de pessoa para pessoa, porque cada uma as pode sentir ou manifestar de forma distinta (Murad and Malkawi, 2012), e algumas pessoas nem conseguem identificar quais as emoções que sentem (Alexitimia) (Barrett et al., 2001).

A subjetividade emocional será um dos fatores que mais contribui para a complexidade do Homem enquanto ser. Como as emoções não podem ser medidas diretamente (Murad and Malkawi, 2012), os investigadores recolhem dados de contexto das pessoas para as tentar descobrir (Kreibig, 2010). Este sinais de entrada podem ser provenientes de várias origens: reações fisiológicas, expressões ou postura corporal (e.g. expressão facial), relato de sentimentos subjetivos, etc. (Sim et al., 2007; Scherer and Ekman, 1984).

Nesta secção apresentam-se as variáveis de contexto recolhidas pelas investigações em análise. As investigações da área do AC recolhem dados de várias modalidades para tentar reconhecer emoções. Após análise da literatura, decidiu-se categorizar as variáveis de contexto em três grandes grupos: i) expressão facial, oral, gestual e postura corporal; ii) variáveis de contexto fisiológico; e iii) variáveis de contexto social e psicológico. O conteúdo desta secção segue esta categorização para apresentar as variáveis de contexto mais utilizadas pelos investigadores.

As propriedades de domínio são também consideradas nesta secção. Resolveu-se pela sua anexação às variáveis de contexto, quando o método de determinação é confundível com o de aquisição. Assim, as propriedades extraídas diretamente a partir do sinal de entrada, cujo método de obtenção seja confundível com o utilizado na aquisição do próprio sinal, serão incluídas nesta secção e não na secção das propriedades extraídas (secção 2.1.4). Por e.g., o (HR) é normalmente assumido pelos autores como uma variável de contexto. No entanto, é extraído a partir de outros sinais de entrada (e.g. Eletrocardiograma (ECG), Fotopletismografia (PPG)). Como é normalmente calculado pelo próprio instrumento de recolha, o (HR) pode ser considerado simultaneamente variável de contexto ou propriedade extraída. Como os investigadores as consideram normalmente variáveis de contexto, decidiu-se pela inclusão destas propriedades nesta secção.

2.1.1.1 Expressão facial, expressão oral e postura corporal

As emoções têm um papel importante nos processos de comunicação e entendimento do ser Humano (Lopes et al., 2003; Raudonis, 2013). O principal meio de comunicação das pessoas é a voz (Nwe et al., 2001). No entanto, além da componente verbal, as pessoas utilizam a componente não-verbal para enriquecer o conteúdo, adornando-o com carga emocional, através de expressões faciais, postura corporal, gestos, etc. (Singh et al., 2015; Nawasalkar et al., 2013).

Dada a importância das emoções nos processos de comunicação, a interação entre homens e dispositivos poderá ser mais natural, se os computadores conseguirem perceber

e responder utilizando também linguagem não-verbal, ou seja, adicionando carga emocional às mensagens transmitidas (Busso et al., 2004; Castellano et al., 2007).

São várias as investigações que se focam na recolha de dados da expressão facial, oral e postura corporal, para estudar formas de reconhecer automaticamente emoções. O motivo poderá estar relacionado com maior facilidade no acesso aos dados a recolher, mas também com o facto de se saber que as pessoas utilizam a comunicação não-verbal, para adicionar conteúdo emocional às mensagens que transmitem (Ekman et al., 2015; Birdwhistell, 1970; Nwe et al., 2001). No entanto, os humanos conseguem controlar as expressões faciais e entoação (Liu et al., 2010). Como tendem a ocultar os seus sentimentos para evitar a exposição ao grupo, tendem também a suprimir ou a adulterar o conteúdo emocional das suas expressões e da sua voz (Kim et al., 2004). Contudo existem também outros desafios à recolha de dados deste contexto: i) as expressões faciais, corporais e entoação de voz, não são sensíveis à idade, género e cultura das pessoas (Rani and Sarkar, 2006); ii) tem que se acautelar a presença de barba, bigode, óculos, etc., aquando da recolha da expressão facial (Wang et al., 2010; Lai et al., 2008); iii) o ruído ambiente, ou distância entre as pessoas e instrumentos de recolha, pode prejudicar a qualidade dos dados recolhidos na expressão oral (Sim et al., 2007); iv) a recolha de dados da postura corporal pode ser comprometida com questões de ordem cultural (Basu et al., 2016); etc.

Expressão facial

A expressão facial é uma das expressões mais utilizadas pelo Homem para manifestar o que sente (Chandler and Cornes, 2012; Recognition, 2009). Existem já muitas investigações que utilizam este sinal de *input* para detetar automaticamente emoções. No entanto, continuam a surgir novos estudos, na tentativa de melhorar a qualidade dos resultados obtidos no reconhecimento (Matlovic et al., 2016).

Existem dois tipos de abordagens utilizadas no reconhecimento emocional através da expressão facial: as baseadas em geometria dos componentes da face, e as que são baseadas na aparência através de descritores que serão apresentados na secção das propriedades extraídas (secção 2.1.4).

As abordagens baseadas nas propriedades geométricas usam informação acerca da dinâmica dos elementos da face que alteram consoante o estado emocional e o movimento das pessoas (Turan et al., 2015). O sistema *Facial Action Coding System* (FACS) é um standard que define os contornos para o reconhecimento de emoções através da expressão facial (Fasel and Luetten, 2003). A divisão da face em zonas, torna possível a deteção de movimento interno localizado a partir da monitorização muscular (Ekman, Paul; Friesen, 1978). Os investigadores Eckert et al. (Eckert et al., 2016) apresentaram o conceito de *Combined Action Units* (CAU) que agrupam *Action Units* (AU) do sistema (FACS). Por exemplo, o movimento muscular no grupo AU7, AU23 e AU24 significará fúria, e

atividade no grupo AU9 e AU10 poderá significar desgosto (Eckert et al., 2016).

A Tabela 2.2 apresenta algumas variáveis de contexto e investigações, que recolhem dados da expressão facial.

Variável	Investigações (exemplo)
Bochechas (CHEEKS)	(Adams and Robinson, 2015) (Gunes and Piccardi, 2007)
Queixo (CHIN)	(Adams and Robinson, 2015)
Sobrancelhas (EYEBROWS)	(Lee et al., 2016) (Adams and Robinson, 2015)
Pálpebras (EYELIDS)	(Gunes and Piccardi, 2007) (Sebe et al., 2006)
Olhos (EYES)	(Turan et al., 2015) (Agrawal et al., 2013)
Face (FACE)	(Turan et al., 2015) (Yang and Bhanu, 2011)
Testa (FOREHEAD)	(Gunes and Piccardi, 2007) (Busso et al., 2004)
Carranca (FROWN)	(Chen et al., 1998)
Cabeça (HEAD)	(Perdiz et al., 2017) (Gogia et al., 2016)
Lábios (LIPS)	(Vermun et al., 2013) (Gunes and Piccardi, 2007)
Mandíbula (JAW)	(Lee et al., 2016) (Adams and Robinson, 2015)
Boca (MOUTH)	(Agrawal et al., 2013) (Vermun et al., 2013)
Nariz (NOSE)	(Soleymani et al., 2013) (Lee et al., 2016)
Pele (SKIN)	(Agrawal et al., 2013)
Ruga (WRINKLES)	(Lee et al., 2016) (Adams and Robinson, 2015)
<i>Combined Action Units</i> (CAU)	(Eckert et al., 2016)
<i>Facial Action Coding System</i> (FACS)	(Lee et al., 2016) (Zhang et al., 2016)

Tabela 2.2: Resumo das variáveis de contexto da expressão facial

Expressão oral

A espécie humana comunica através da voz. No entanto, decoramos os sons que emitimos com aditivos (e.g. visuais) como forma de enriquecer as mensagens com carga emocional (Lopes et al., 2003).

As emoções que o Homem sente enquanto fala tendem a influenciar o tom, velocidade, qualidade, e articulação das palavras (Cahn, 1990). Por este motivo, estas características acústicas são fortes candidatos a variáveis de contexto a utilizar num sistema de reconhecimento emocional (Haritaoglu et al., 2001).

Um dos desafios colocados aos investigadores que tentam detetar emoções a partir da expressão oral, é a escolha dos dados de contexto a recolher. As propriedades devem ser independentes do orador, e espelhar a parte emocional do conteúdo oral (Shegokar and Sircar, 2016). Neste contexto, uma das variáveis mais recolhidas é a entoação (SPEECH), por poder transportar conteúdo emocional (Rani and Sarkar, 2006). Nos últimos anos vários investigadores têm tentado combinar a recolha da expressão facial com a oral, com o objetivo de melhorar a precisão do reconhecimento emocional (Jerritta et al., 2011).

A Tabela 2.3 apresenta algumas variáveis de contexto e investigações que utilizam dados da expressão oral.

Variável	Investigações (exemplo)
<i>Pitch</i> ¹ (PITCH)	(Lalitha et al., 2015) (Busso et al., 2004)
Entoação (SPEECH)	(Korkmaz and Atasoy, 2015) (Lalitha et al., 2015)
Volume (intensidade) (VOLUME)	(Korkmaz and Atasoy, 2015) (Lalitha et al., 2015)

Tabela 2.3: Resumo das variáveis de contexto da expressão oral

Expressão gestual e postura

A expressão gestual é importante na detecção automática de emoções, porque é um indicador do estado mental das pessoas (Saha et al., 2014). No entanto, apesar de os gestos transmitirem carga emocional durante o processo de comunicação, são muito dependentes do contexto cultural e social de cada pessoa, levantando problemas à generalização dos sistemas (Rehm et al., 2008).

A recolha de dados de contexto com base na expressão gestual, envolve essencialmente informação dos membros superiores do corpo (Saha et al., 2014). Contudo, alguns investigadores, também recolhem dados de contexto relacionados com a postura corporal nas suas investigações (e.g. Vermun et al. conjugou dados recolhidos da expressão gestual com a postural corporal para identificar estados afetivos e cognitivos em sala de aula (Vermun et al., 2013)).

A Tabela 2.4 apresenta as variáveis de contexto relacionadas com a expressão gestual e postura, utilizadas pelos investigadores deste levantamento.

Variável	Investigações (exemplo)
Braços ² (ARMS)	(Castellano et al., 2007)
Cotovelos (ELBOWS)	(Saha et al., 2014)
Dedos (FINGERS)	(Gunes and Piccardi, 2007)
Punho ³ (FISTS)	(Gunes and Piccardi, 2007)
Cabeça (HEAD)	(Saha et al., 2014)
Mãos (HANDS)	(Singh et al., 2015) (Gunes and Piccardi, 2007)
Quadril ⁴ (HIP)	(Vermun et al., 2013)
Joelhos ⁵ (KNEES)	(Vermun et al., 2013)
Pescoço ⁶ (NECK)	(Gunes and Piccardi, 2007)
Palmas da mão (PALMS)	(Gunes and Piccardi, 2007)
Ombros (SHOULDERS)	(Singh et al., 2015) (Vermun et al., 2013)
Coluna vertebral (SPIN)	(Saha et al., 2014)
Pulsos (WRISTS)	(Saha et al., 2014)

Tabela 2.4: Resumo das variáveis de contexto da expressão gestual e postura corporal

¹Perceção da qualidade de som

²Movimentação dos braços

³Mão fechada

⁴Movimentação do quadril

⁵Movimentação dos joelhos

⁶Movimentação das mãos junto ao pescoço

2.1.1.2 Variáveis de contexto fisiológico

As variáveis de contexto fisiológico são muito utilizadas pelos investigadores da área do AC. Apesar da possível obstrutividade causada pela necessidade de medir através de sensores, os investigadores parecem preferir recolher dados deste contexto, porque muitas das alterações fisiológicas são possíveis de medir de forma objetiva, através de sensores (Kim and André, 2008; Alabdulkarim, 2015).

Além disso, grande parte da informação recolhida tem origem no *Autonomous Nervous System* (ANS). Por este motivo, os autores consideram que estes dados espelham de forma mais realista as emoções sentidas, atendendo a que as pessoas não podem manipular as variáveis recolhidas. Apesar de os dados serem de natureza mais subjetiva, a impossibilidade de manipulação destes dados, faz com que os sistemas fiquem mais blindados ao mascaramento social (Kim and André, 2008).

As emoções afetam o sistema cardiovascular, respiratório, eletrodermal e muscular (Basu et al., 2016). Além disso, é possível recolhe-los em tempo real sem que as pessoas se tenham que expressar, alargando assim o âmbito de aplicação (e.g. doentes, trabalhadores a operar máquinas) (Poole and Ball, 2005). No entanto, existem vários desafios que os investigadores que utilizam dados de contexto fisiológico têm que enfrentar nas suas investigações: i) obstrutividade causada pela utilização de alguns sensores (Barea et al., 2011); ii) ruído presente no sinal recolhido (e.g. influência da atividade muscular no sinal (EEG)); iii) problemas de generalização, porque os stressores atuam de forma diferente entre indivíduos, causando diferentes reações fisiológicas de pessoa para pessoa (Biondi and Picardi, 1999); etc.

Atividade cerebral

A literatura estabelece uma correlação forte entre a atividade cerebral e as emoções sentidas pelos humanos (Bos, 2010). As variáveis de contexto produzidas pelo cérebro humano, são um importante *input* nos sistemas de deteção emocional, porque o hipotálamo é responsável por processar os sinais chegados ao cérebro e desencadear as respetivas respostas fisiológicas (e.g. aumentar o (HR), provocar uma resposta de condutância da pele (SCR)) (Kandel et al., 2000).

Apesar das várias investigações que recolhem o (EEG) existem vários desafios específicos à utilização deste sinal (e.g. a propriedade alfa do (EEG) é impactada pelos movimentos musculares da face (Bos, 2010)). Além disso o funcionamento do cérebro humano precisa ainda de mais investigação. Por e.g., os investigadores Niemic et al. e Canli et al., defendem que uma maior atividade na região frontal esquerda do cérebro indica uma emoção positiva, e a maior atividade na zona anterior do lóbulo direito está relacionada com uma emoção negativa (Niemic et al., 2002; Canli et al., 1998). No entanto os trabalhos (Liu et al., 2010) e (Lane et al., 1997) não sustentam estas conclusões, e investigam as hipóteses contrárias.

A Tabela 2.5 apresenta algumas das investigações analisadas neste levantamento de estado da arte, que utilizam o sinal (EEG).

Variável	Investigações (exemplo)
<i>Electroencephalogram</i> (EEG)	(Matlovic et al., 2016) (Matiko et al., 2014)

Tabela 2.5: Variável de contexto da atividade cerebral

Atividade cardíaca

A atividade cardíaca é uma importante fonte de informação para os investigadores que estudam as emoções. A literatura sustenta a relação emocional com os sinais recolhidos a partir desta origem.

São várias as investigações que correlacionam a atividade cardíaca com emoções. Por e.g., a frequência cardíaca (HR) varia com o medo, fúria e susto (Lisetti and Nasoz, 2004). A variabilidade da frequência cardíaca (HRV), que reflete as alterações entre batimentos no (HR) (Mokhayeri et al., 2011), é conhecido por diminuir com a ansiedade e aumentar com a diversão, stress mental e frustração (Murad and Malkawi, 2012). O (HRV) é muito utilizado pelos investigadores (e.g. os investigadores Muaremi et al utilizaram-no para medir o stress (Muaremi et al., 2014)). O volume sistólico (SV) diminui com as emoções negativas (e.g. desgosto) (Kreibig, 2010). O período pré-ejeção (PEP)¹ é conhecido por aumentar com a tristeza aguda, e diminuir com a alegria e com a fúria (Murad and Malkawi, 2012). A pressão sanguínea (BP) varia com o stress, medo e fúria (Wang et al., 2014). O aumento da pressão arterial sistólica (SBP) está relacionado com o medo, ansiedade e com a fúria (Murad and Malkawi, 2012). A pressão arterial diastólica (DBP) é conhecida por aumentar com a fúria (Sinha et al., 1992) ansiedade e desgosto, e diminuir com a tristeza aguda (Kreibig, 2010).

A Tabela 2.6 resume as variáveis mais utilizadas pelos autores que recolhem dados de contexto da atividade cardíaca.

Variável	Investigações (exemplo)
<i>Blood Pressure</i> (BP)	(Zhang et al., 2016) (Murad and Malkawi, 2012)
<i>Blood Volume Pulse</i> (BVP)	(Chang et al., 2012) (Zhai and Barreto, 2006)
<i>Diastolic Blood Pressure</i> (DBP)	(Zhang et al., 2016) (Vrijkotte et al., 2000)
<i>Electrocardiogram</i> (ECG)	(Zhao et al., 2016) (Zenonos et al., 2016)
<i>Heart rate</i> (HR)	(Basu et al., 2016) (Kusserow et al., 2013)
<i>Heart Rate Variability</i> (HRV)	(Zenonos et al., 2016) (Alzoubi et al., 2013)
<i>Inter-Beat Interval</i> (IBI)	(Lichtenstein, Antje; Oehme, 2008)
<i>Impedance Cardiogram</i> (ICG)	(Murali et al., 2015) (Sinha et al., 1992)
<i>Non-Invasive Blood Pressure</i> (NIBP)	(Murali et al., 2015) (Nawasalkar et al., 2013)
<i>Pre-Ejection Period</i> (PEP)	(Murali et al., 2015) (Murad and Malkawi, 2012)
<i>Photoplethysmography</i> (PPG)	(Zenonos et al., 2016) (Mokhayeri et al., 2011)

¹Tempo que decorre entre uma *Q-wave ECG onset* e um *B-point* do cardiograma de impedância (ICG) (Stephens et al., 2010)

<i>Pulse Rate</i> (PR)	(Zhang et al., 2016) (Basu et al., 2016)
<i>RootMeanSquare of Sucessive Differences</i> (RMSSD)	(Zenonos et al., 2016) (Vrijkotte et al., 2000)
<i>Systolic Blood Pressure</i> (SBP)	(Zhang et al., 2016) (Murad and Malkawi, 2012)
<i>Standard Deviation of NN Intervals</i> (SDNN)	(Zhao et al., 2016) (Zenonos et al., 2016)
<i>Stroke Volume</i> (SV)	(Murad and Malkawi, 2012) (Sinha et al., 1992)

Tabela 2.6: Resumo das variáveis de contexto da atividade cardíaca

Atividade respiratória

A monitorização contínua da respiração pode ser importante para a prevenção de problemas pulmonares e cardíacos (Ravichandran et al., 2015). A precisão da medição é maior se incidir sobre a troca de gases efetuada pelos pulmões. No entanto, esta técnica tem o problema da obstrutividade porque dificulta a atividade normal das pessoas (Healey et al., 2000). Por este motivo, os investigadores utilizam técnicas alternativas como a medição através da impedância, ou a medição da expansão da cavidade peitoral (Nawasalkar et al., 2013; Healey et al., 2000).

A atividade respiratória é usada na saúde e na psicologia (Ravichandran et al., 2015). Os investigadores recolhem dados desta origem para vários fins: i) investigar o stress (e.g. (Healey and Picard, 2005)); ii) estudar o sono (e.g. (Long et al., 2014)); iii) analisar o riso (e.g. (Fukumto and Nagamatsu, 2016)); etc. No entanto existem também várias investigações que utilizam a respiração como fonte de dados para estudar as emoções (e.g. (Alzoubi et al., 2013), (Murali et al., 2015)).

Utiliza-se o *token* (RESP), quando as variáveis de contexto recolhidas da respiração, não são totalmente identificadas pelos autores.

A Tabela 2.7 apresenta algumas variáveis e investigações que recolhem dados de contexto da atividade respiratória.

Variável	Investigações (exemplo)
<i>Breath Depth</i> (RDEP)	(Lichtenstein, Antje; Oehme, 2008)
<i>Respiration</i> (RESP)	(Zhang et al., 2016) (Zhao et al., 2016)
<i>Oscillatory resistance</i> (ROS)	(Murad and Malkawi, 2012) (Ritz et al., 2000)
<i>Respiration rate</i> (RR)	(Basu et al., 2016) (Murali et al., 2015)
<i>Tidal volume</i> (VT)	(Murad and Malkawi, 2012) (Ritz et al., 2000)

Tabela 2.7: Resumo das variáveis de contexto da atividade respiratória

Pele

A pele é o maior órgão do corpo humano, e tem funções importantes na proteção do corpo e regulação da temperatura corporal (Silva, 2014). São essencialmente três as variáveis de contexto que são estudadas pelos investigadores: a atividade eletrodermal (EDA), temperatura da pele (ST), e a condutância não específica da pele ($nSRR$).

A variável de contexto (EDA), refere-se à capacidade de condutância da pele, e é referida pela literatura com vários termos (e.g. *galvanic skin response* (GSR), *electrodermal response* (EDR)) (Healey and Picard, 2005; Sano and Eng, 2016; Boucsein, 2012). A capacidade de condução elétrica da pele aumenta momentaneamente quando as pessoas são sujeitas a um estímulo excitante. Nestes casos, a condutância da pele tem um pico seguido de uma diminuição exponencial (Jaques et al., 2015). Como é possível medir a condutância através de sensores, os investigadores recolhem estas variações, e usam esses dados como *input* nos seus sistemas de deteção emocional.

O (EDA) é um importante indicador da excitação psicológica ou fisiológica. Além disso, a condutância da pele reflete a inervação sudomotora e a atividade da glândula do suor que aumentam com a atividade do Sistema Nervoso Simpático (SNS) (Taylor et al., 2015). Como o SNS é influenciado pelo hipotálamo e pelo sistema límbico, e estas são estruturas relacionadas com as emoções, o (EDA) é, assim, uma variável de contexto interessante na deteção automática de emoções (Taylor et al., 2015).

Apesar de existirem vários projetos que utilizam o (EDA), existem desafios a enfrentar pelos investigadores. O (EDA) é influenciável pela temperatura ambiente (TEMP) (Chandler and Cornes, 2012), e podem existir variações causadas por outros fatores que não o emocional (e.g. exercício físico) (Chen et al., 2014; Jaques et al., 2015).

A relação estabelecida pela literatura entre o (EDA) e o reconhecimento emocional é forte. O (EDA) correlaciona-se melhor com o *Arousal* (calma *versus* excitação) do que com o *Valence* (negativo *versus* positivo) (Iwasaki et al., 2010). Aumenta com a frustração (Lisetti and Nasoz, 2004), varia com emoções negativas (e.g. tédio) (Giakoumis et al., 2010), aumenta mais com a raiva do que com o medo e com a alegria (Levenson et al., 1990), etc.

Outra variável recolhida pelos investigadores a partir da pele é a temperatura da pele (ST). A variação da (ST) acontece por causa do fluxo sanguíneo em consequência da resistência vascular e da pressão arterial (Kim et al., 2004). Esta variável é também importante no estudo emocional, mas é, normalmente, utilizada em conjunto com outras variáveis por causa da possível influência de fatores externos (e.g. temperatura ambiente (TEMP) (Chandler and Cornes, 2012)). A (ST) não deve ser confundida com a temperatura corporal (Sano and Eng, 2016). A descida da temperatura interna do corpo é normalmente precedida de um aumento da (ST) (Sarabia et al., 2008) e, durante o ciclo circadiano, existem alturas em que o comportamento de ambas é inverso (Gradisar and Lack, 2004; Č et al., 2016).

A Tabela 2.8 apresenta algumas investigações que utilizam variáveis recolhidas da

pele.

Variável	Investigações (exemplo)
<i>Electrodermal activity</i> (EDA)	(Matlovic et al., 2016) (Zhang et al., 2016)
<i>Skin temperature</i> (ST)	(Zhang et al., 2016) (Zenonos et al., 2016)

Tabela 2.8: Resumo das variáveis de contexto relacionadas com a pele

Atividade muscular

O principal método de recolha da atividade muscular é a Electromiografia (EMG). O (EMG) regista as voltagens elétricas quando os músculos contraem (Lichtenstein, Antje; Oehme, 2008). Este sinal pode ser recolhido através de pequenas agulhas ou através da colocação de eléctrodos sobre a pele na zona dos músculos (Haag et al., 2004; Chandler and Cornes, 2012).

Apesar dos desafios na utilização da atividade muscular nos sistemas de deteção emocional (e.g. ruído capturado pelos eléctrodos pela fala ou tosse, recolha invasiva através de agulhas) (Mandryk and Atkins, 2007), o sinal (EMG) é utilizado em várias áreas de investigação: entretenimento (Mandryk et al., 2006); reabilitação física (Ehrampoosh et al., 2016); terapia para pessoas com Acidente Vascular Cerebral (AVC) (Thielbar et al., 2016); etc. No entanto também é uma variável de contexto utilizada na deteção automática de emoções. Através do (EMG) é possível distinguir o *valence* positivo do negativo (Partala et al., 2005) (e.g., o sorrir consegue ser reconhecido através da atividade dos músculos da face (Sloan, 2004)). O (EMG) permite detetar o susto (Haag et al., 2004), correlaciona-se fortemente com o desafio (Mandryk and Atkins, 2007) e consegue discriminar o desgosto da fúria (Lisetti and Nasoz, 2004).

A Tabela 2.9 apresenta algumas investigações que utilizam o (EMG) como variável de contexto.

Variável	Investigações (exemplo)
<i>Electromyogram</i> (EMG)	(Perdiz et al., 2017) (Sano and Eng, 2016)

Tabela 2.9: Resumo das variáveis de contexto da atividade muscular

Atividade ocular

A monitorização dos olhos é utilizada no estudo de problemas psicológicos e físicos (Raudonis, 2013) (e.g. a forma de olhar pode ser vista como um indicador de ansiedade ou outro tipo de perturbações (Schulze et al., 2013)). A monitorização da atividade ocular é, normalmente, feita através do Eletrooculograma (EOG). O sinal é recolhido através da colocação de eléctrodos na zona dos olhos para analisar a diferença de voltagem entre a

córnea e a retina (Zhang and Sawchuk, 2013). A utilização do (EOG) permite recolher várias variáveis de contexto como o movimento e piscar dos olhos (Morris and Miller, 1996).

O (EOG) é utilizado em várias áreas de investigação: cansaço (e.g. (Morris and Miller, 1996)), sono (e.g. (Estrada et al., 2006)), segurança (e.g. (Hossain et al., 2016)), etc. No entanto também é utilizado pelas investigações do reconhecimento automático de emoções onde são utilizadas essencialmente dois componentes dos olhos: a pupila (PUPIL) e a posição do olhar (contemplação) (GAZE). O carácter involuntário da reação do diâmetro da (PUPIL) faz com que este seja um bom *input* para os sistemas de deteção emocional (Babiker et al., 2013). Existe uma boa correlação com o *valence* (Kawai et al., 2013) e *arousal* emocional (Aracena et al., 2016; Bradley et al., 2008). No entanto, existem também desafios lançados às investigações que utilizam este sinal de entrada: i) o movimento do olho e o tamanho da pupila são diferentes entre pessoas, dependem da carga cognitiva e estado mental a cada momento (Raudonis, 2013); ii) a pupila está constantemente a adaptar-se à variação da luz (Aracena et al., 2016); iii) o piscar de olhos e movimentos instantâneos do olhar têm que ser considerados na análise de dados (Raudonis, 2013); etc.

A Tabela 2.10 apresenta alguns estudos que recolhem dados de contexto da atividade ocular.

Variável	Investigações (exemplo)
<i>Electrooculogram</i> (EOG)	(Perdiz et al., 2017) (Sano and Eng, 2016)
<i>(Gaze)</i> ¹ (PUPIL)	(Aracena et al., 2016) (Raudonis, 2013)
<i>(Pupil)</i> (PUPIL)	(Kawai et al., 2013) (Babiker et al., 2013)

Tabela 2.10: Resumo das variáveis de contexto da atividade ocular

Atividade glandular

A atividade glandular pode contribuir com dados interessantes para um sistema de deteção emocional. No entanto, os investigadores tendem a não usar estas variáveis, por causa do método de recolha (urina, sangue ou saliva) (Omar, 2006; Frazão, 2016). Falta ainda evolução tecnológica, para permitir que estas variáveis de contexto possam ser recolhidas através de métodos menos invasivos.

O cortisol (CORT) é importante para a saúde física e psicológica das pessoas (Dickerson and Kemeny, 2004), e é conhecida como a hormona reguladora do stress (Smeets et al., 2009). Existem várias áreas de investigação que utilizam o (CORT): agressão social (e.g. (Terburg et al., 2009)); humor (e.g. (van Eck et al., 2005)); memória (e.g. (Buchanan and Lovallo, 2001)); etc. No entanto, também é considerado em estudos relacionados com as emoções. O corpo humano ativa o (CORT) perante situações negativas como o medo,

¹Posição do olhar (contemplação)

ameaça, hipótese de falhar, ou impedimento de atingir objetivos (Blascovich and Tomaka, 1996).

A melatonina (MELAT) é uma hormona importante para regular o ritmo biológico do corpo. É conhecida como a hormona reguladora do sono (Markho, 2016) e, por esse motivo, é bastante utilizada em investigações relacionadas com a disposição (humor) e com o ciclo circadiano (Curcio et al., 2016). A produção da (MELAT) está relacionada com a alternância da luz do dia e da escuridão da noite (Č et al., 2016). A luz funciona como fator de treino do período de sono no ciclo circadiano no corpo humano (Czeisler et al., 1986). Por esse motivo, são vários os investigadores que estudam o impacto da luz artificial dos dispositivos eletrônicos no sono (Sroykham and Wongsawat, 2013).

A serotonina (SEROT) está relacionada com o bem-estar e com o prazer (Fonte, 2015). É conhecida como a hormona do humor (Markho, 2016) e, por esse motivo, é utilizada em investigações relacionadas com a disposição, cognição, etc. (Visser et al., 2011). A menor presença da (SEROT) no organismo está relacionada com a depressão e apatia (Markho, 2016). A maior presença no organismo, aumenta o sentimento de domínio (assertividade), diminui a hostilidade, reduz o sentimento de afetos negativos, facilita a tomada de decisões, etc. (Merens et al., 2007). A (SEROT) está presente em várias investigações relacionadas com o estudo emocional: disposição (humor) (e.g. (Merens et al., 2007)); ansiedade (e.g. (Harmer et al., 2004)); etc.

A Tabela 2.11 apresenta algumas das investigações em análise, que utilizaram estas variáveis de contexto nas suas experiências.

Variável	Investigações (exemplo)
<i>Cortisol</i> (CORT)	(van Eck et al., 2005)
<i>Melatonin</i> (MELAT)	(Sano and Eng, 2016)
<i>Serotonin</i> (SEROT)	(Harmer et al., 2003)

Tabela 2.11: Resumo das variáveis de contexto da atividade glandular

2.1.1.3 Variáveis de contexto social e psicológico

O homem é um ser social e coletivo. As emoções que sente influenciam e são influenciadas pela sua vida em grupo (Ekman, 1989; Dickerson and Kemeny, 2004). As relações sociais são o principal instigador das emoções (Kemper, 1991, 1978; Lazarus and Launier, 1978), e as emoções são respostas a eventos do contexto (Arnold, 1970).

Existem vários fatores que podem afetar o estado emocional do ser Humano: questões fisiológicas, envolvente social, experiência subjetiva, antecedentes, cultura, etc. (Dong et al., 2011). Assim, cada pessoa pode reagir de forma distinta perante o mesmo estímulo emocional (Raudonis, 2013), e a mesma pessoa pode, em momentos diferentes, reagir de forma diferente a estímulos semelhantes (Dong et al., 2011; Kim and André, 2008). Apesar das emoções serem a base da interação humana, medi-las é um desafio para a

ciência afetiva (Basu et al., 2016). Os humanos são seres complexos e podem exprimir emoções de forma ambígua. A tristeza provocada pelo fracasso pode (ou não) levar ao choro mas, algumas pessoas choram de felicidade em reação a um evento positivo, e outras sorriem enquanto experienciam uma emoção negativa (Basu et al., 2016).

São várias as áreas investigações da área emocional que utilizam variáveis de contexto social nos seus projetos: qualidade de vida (e.g. (Pais-Ribeiro, 1999)); bem-estar físico e psicológico (e.g. (Heitzmann and Kaplan, 1988)); humor (disposição) (Moturu et al., 2011); depressão (e.g. (George et al., 1989)); alegria e bem-estar (e.g. (Seligman, 2011)); etc.

O suporte social é um conceito da psicologia da saúde (Dunbar et al., 1998) e é um determinante importante para o bem-estar ao longo da vida (Kahn and Antonucci, 1980). Um indivíduo com suporte social é alguém que tem pessoas em quem pode confiar, que gostam de si, valorizam e se preocupam consigo (Sarason et al., 1987). A literatura defende uma relação estreita entre a saúde e o suporte social (Wallston et al., 1983). Um bom suporte social: reduz a sensação de mal-estar (Sarason et al., 1985); protege da depressão (Peirce et al., 2000); alivia o stress (Pais-Ribeiro, 1999); aumenta a satisfação com a vida (Hohaus and Berah, 1996); promove um envelhecimento saudável (Antonucci, 2001); etc.

Kahn e Antonucci defenderam que cada pessoa é rodeada de uma rede pessoal de pessoas das quais dá e recebe suporte social. Segundo os autores, a qualidade e composição dessa rede é moldada em função de vários fatores como o tempo, idade, personalidade, papéis desempenhados no grupo, trabalho, etc. (Kahn and Antonucci, 1980). Antonucci representa este conjunto de relacionamentos através de um conjunto de círculos concêntricos. O mais interno representa o próprio indivíduo e os restantes três as pessoas que pertencem à sua rede pessoal a quem dá e de quem recebe suporte social. A ordenação de círculos representa a proximidade dos elementos da rede num determinado momento da vida, sendo o externo o que representa os membros mais afastados. O círculo mais interno (excluindo o que representa o próprio indivíduo) representa os familiares e amigos próximos (e.g. filhos), o seguinte representa os restantes familiares e amigos, e o círculo mais externo, os vizinhos, colegas de trabalho, etc. Este conceito de Antonucci é conhecido como modelo *convoy* (Antonucci, 2001).

Pais Ribeiro propõe quatro fatores para a caracterizar o suporte social de cada indivíduo em relação à sua rede social: i) satisfação com os amigos; ii) satisfação com suporte social íntimo; iii) satisfação com a família; e iv) atividades sociais, que contempla a satisfação com as atividades realizadas que envolvem pessoas (Pais-Ribeiro, 1999).

O modelo *convoy* de Antonucci (Antonucci, 2001) e a investigação de Pais-Ribeiro (Pais-Ribeiro, 1999), inspiraram a organização do conteúdo desta secção. Decidiu-se dividir em domínios e categorias as variáveis de contexto social utilizadas nos sistemas de deteção automática de emoções. Os domínios representam a proximidade e relevância social de cada variável, e a categorização resulta de um agrupamento das variáveis em

função da natureza, pertinência e papel na promoção do relacionamento com os outros elementos da rede. O domínio pessoal representa o contexto individual de cada pessoa e é composto pelas categorias demográfica, psicossocial e rotina diária. As categorias interação, reconhecimento e integração, representam o suporte social dado e recebido do indivíduo, nos domínios privado, próximo e público. Agrupam-se os três domínios num único (i.e. outros domínios), porque as suas variáveis de contexto psicossocial podem estar presentes transversalmente em qualquer nível de relacionamento dentro da rede.

Domínio pessoal (demografia, psicossocial e rotina diária)

Nesta secção incluem-se as variáveis que representam o contexto pessoal de cada indivíduo. Neste domínio existem três categorias de variáveis de contexto: informação demográfica, dados de ordem psicossocial, e dados relacionados com a rotina diária das pessoas.

Na categoria demográfica incluem-se os dados relacionados com a informação pessoal de cada indivíduo. Fazem parte desta categoria variáveis de contexto como área de estudo do indivíduo (SCHOOLA), ano escolar (atual ou último) (SCHOOLY), com quem vive (e.g. país) (LIVING), etc.

A categoria psicossocial agrupa as variáveis relacionadas com aspetos gerais da vida. Estas variáveis de contexto são, normalmente, recolhidas através de questionários e inclui várias naturezas de informação: sono (SLEEP) (e.g. tempo, hora de acordar ou deitar); quantidade de sestas feitas durante o dia (NAP); autoapreciação subjetiva da saúde (HEALTH), humor (MOOD), e emoções em geral (EMOTIONS); existência de eventos marcantes recentes (LIFEEVENTS) (e.g. morte de alguém chegado, divórcio), ou complicações na vida (DIFFICULTIES) (e.g. problemas no trabalho ou em casa, problemas de dinheiro ou de família); etc.

Agrupados na categoria rotina diária encontram-se as variáveis relacionadas com as atividades do dia-a-dia das pessoas: utilização de aplicações (APPS), utilização de equipamentos eletrónicos (ELECTR), experiência na utilização da tecnologia (TECHEXPERT); etc.

A Tabela 2.12 apresenta algumas das variáveis de contexto utilizadas nas investigações em análise. A tabela agrupa as variáveis pelas respetivas categorias.

Variável	Investigações (exemplo)
Categoria: demográfica	
<i>Age</i> (AGE)	(Sano and Eng, 2016) (Herbon et al., 2005)
<i>Ethnicity</i> (ETHNICITY)	(Sano and Eng, 2016) (Lisetti and Nasoz, 2004)
<i>Gender</i> (GENDER)	(Herbon et al., 2005) (Lisetti and Nasoz, 2004)
<i>Living situation</i> (LIVING)	(Sano and Eng, 2016)
<i>School year</i> (SCHOOLY)	(Sano and Eng, 2016)
<i>Study area</i> (SCHOOLA)	(Sano and Eng, 2016)
Categoria: psicossocial	
<i>Level of alert felt</i> (ALERT)	(Jaques et al., 2015) (Sano and Picard, 2013b)

<i>Level of anger felt</i> (ANGER)	Van Eck et al. (van Eck et al., 2005)
<i>Level of Anxiety felt</i> (ANXIETY)	(van Eck et al., 2005) (Harmer et al., 2003)
<i>Level of Calm felt</i> (CALM)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Level of Depression felt</i> (DEPRESSION)	(Lane et al., 2011) (van Eck et al., 2005)
<i>Life difficulties</i> (DIFFICULTIES)	(van Eck et al., 2005)
<i>Emotions felt</i> (EMOTIONS)	(Matlovic et al., 2016) (Zhang et al., 2016)
<i>Level of energy felt</i> (ENERGY)	(Jaques et al., 2015) (Harmer et al., 2003)
<i>Level of happiness felt</i> (HAPPY)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Level of health felt</i> (HEALTH)	(Sano and Eng, 2016) (Sano and Picard, 2013b)
<i>Recent life events</i> (LIFEEVENTS)	(van Eck et al., 2005)
<i>Mood status felt</i> (MOOD)	(Zenonos et al., 2016) (Kusserow et al., 2013)
<i>Nap data</i> (NAP)	(Jaques et al., 2015) (Sano and Picard, 2013b)
<i>Personality data</i> (PERSON)	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014) (Vrijkotte et al., 2000)
<i>Sleep data</i> (SLEEP)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Level of stress felt</i> (STRESS)	(Sano and Eng, 2016) (Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014)
<i>Level of tiredness felt</i> (TIRED)	(Sano and Picard, 2013b)
<i>Level of wellbeing felt</i> (WELLBEING)	(Lane et al., 2011) (van Eck et al., 2005)
Categoria: rotina diária	
<i>Apps usage</i> (APPS)	(Sano and Eng, 2016) (LiKamWa et al., 2013)
<i>Browser usage</i> (BROWSER)	(LiKamWa et al., 2013)
<i>Electronic devices usage</i> (ELECTR)	(Sano and Picard, 2013b)
<i>Geolocalization</i> (LOCAL)	(Sano and Picard, 2013b) (Bauer and Lukowicz, 2012)
<i>Screen on and off</i> (SCREEN)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Technological experience</i> (TECHEXPERT)	(Herbon et al., 2005)

Tabela 2.12: Resumo das variáveis de contexto social (domínio pessoal)

Outros domínios (interação, reconhecimento e integração)

Nesta secção apresentam-se as variáveis de contexto dos domínios: privado (i.e. amigos íntimos e familiares de primeira linha), próximo (i.e. restante família e amigos), e público (e.g. vizinhos, colegas). Como as variáveis de contexto podem existir em cada um dos três domínios, optou-se por agrupar todas as variáveis num único domínio a que se chamou "Outros domínios".

As variáveis de contexto deste agrupamento, são categorizadas em função do suporte social dado e recebido. A categoria interação contém as variáveis relacionadas com a manutenção do contacto entre elementos da rede. O reconhecimento agrupa as características ou propriedades que os outros elementos da rede interpretam como sinal de prestígio ou competências (qualidades). Por fim, a integração, contém as variáveis relacionadas com as atividades sociais ou comportamentos resultantes da integração na rede (i.e. comportamentos adotados para promoção da integração entre elementos da rede).

Fazem parte da categoria interação as variáveis relacionadas com a interação entre o indivíduo e os outros elementos da rede: correio eletrónico (EMAIL), interações cara-a-

cara (FTF), autoapreciação da qualidade das interações sociais (SOCIAL), proximidade com outras pessoas (PROXIMITY), etc.

Na categoria do reconhecimento agrupam-se as variáveis de contexto que representam admiração (respeito) dos outros elementos da rede: desempenho académico (ACADGR); grau académico (ACADDG); atividade física (PHYSI); tempo de estudo (ACADST) (e.g. tempo dedicado a um projeto académico); atividade académica curricular (ACADCL) (e.g. presença e assiduidade na atividade letiva); etc.

A categoria da integração inclui variáveis de contexto que representam comportamentos adotados ou promovidos para adaptação (integração) à rede: consumo de bebidas com cafeína (CAFFEI); consumo de cigarros (SMOKING); consumo de álcool (ALCOH); atividade académica extracurricular (ACADEX); etc.

A Tabela 2.13 apresenta algumas das variáveis de contexto das três categorias abordadas nesta secção.

Variável	Investigações (exemplo)
Categoria: interação	
<i>Email</i> (EMAIL)	(Sano and Eng, 2016) (LiKamWa et al., 2013)
<i>Phonecalls</i> (CALL)	(Jaques et al., 2015) (Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014)
<i>Proximity interaction</i> (PROXIMITY)	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014) (Bauer and Lukowicz, 2012)
<i>Quality of social interactions</i> (SOCIAL)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>SMS messages</i> (SMS)	(Jaques et al., 2015) (Sano and Picard, 2013b)
<i>Time of talk</i> (TALK)	(Lane et al., 2011)
Categoria: reconhecimento	
<i>Academic curricular activity</i> (ACADCL)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Academic degree</i> (ACADDG)	(Sano and Eng, 2016) (Vrijotte et al., 2000)
<i>Academic grades</i> (ACADGR)	(Sano and Eng, 2016)
<i>Academic Study Time</i> (ACADST)	(Jaques et al., 2015)
<i>Body Mass Indice</i> (BMI)	(Vrijotte et al., 2000)
<i>Height</i> (HEIGHT)	(Vrijotte et al., 2000)
<i>Physical activity</i> (PHYSI)	(Lane et al., 2011) (van Eck et al., 2005)
<i>Waist size</i> (WAIST)	(Vrijotte et al., 2000)
<i>Weight</i> (WEIGHT)	(Vrijotte et al., 2000)
<i>Years of work</i> (WORKYEARS)	Vrijotte et al. (Vrijotte et al., 2000)
Categoria: integração	
<i>Academic Extracurricular Activity</i> (ACADEX)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Alcohol intake</i> (ALCOH)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Caffeine intake</i> (CAFFEI)	(Jaques et al., 2015) (Sano and Picard, 2013b)
<i>Drugs intake</i> (DRUGS)	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Information about smoking</i> (SMOKING)	(van Eck et al., 2005) (Vrijotte et al., 2000)

Tabela 2.13: Resumo das variáveis de contexto social (outros domínios)

2.1.1.4 Outras variáveis

Além das variáveis de contexto apresentadas anteriormente, existem outras que também são utilizadas pelos investigadores no reconhecimento automático de emoções. Nesta secção consideram-se também as variáveis que são recolhidas apenas para apoiar no pré-processamento (e.g. deteção de ruído).

Os dados do acelerómetro (ACC) são utilizados em várias investigações: estudo de padrões de sono (e.g. (Ancoli-Israel et al., 2003)), contagem de passos (e.g. (Sano and Eng, 2016)), deteção de atividades humanas (e.g. (Jaques et al., 2015)), etc. Os dados do (ACC) podem também ser utilizados diretamente pelos algoritmos de deteção de emoções: LiKamWa et al. utilizaram o (ACC) no estudo acerca da disposição (LiKamWa et al., 2013); Garica-Ceja et al. estudaram o stress utilizando o (ACC) (Garica-Ceja et al., 2015); Castellano et al. estudaram emoções a partir do movimento e, por isso, recolheram o (ACC) (Castellano et al., 2007); etc.

As variações do tempo (WEATHER) são também utilizadas como variáveis de contexto. Bogomolov et al. utilizaram o (WEATHER) na sua investigação acerca do stress (Bogomolov, Trento, Lepri, Pianesi, Kessler, Kessler and Pentland, 2014); Denissen et al., tentaram estabelecer uma relação entre a disposição e o (WEATHER) (Denissen et al., 2008); etc.

Alguns autores consideram a temperatura (TEMP) e a luz ambiente (LIGHT) como variáveis de mascaramento, porque podem influenciar outros dados de contexto recolhidos (Martinez-Nicolas et al., 2013). A (TEMP) pode influenciar o *core body temperature* (Wakamura and Tokura, 2002) e, por esse motivo, é utilizada pelos investigadores para discriminar o aumento de (ST) causado por origem emocional ou ambiental (Jaques et al., 2015). A (TEMP) e a (LIGHT) são variáveis também utilizadas em estudos relacionados com as emoções: Sander et al. estudaram o sono e o bem-estar utilizando a (TEMP); Harb et al. estudaram a relação entre a luz natural no espaço de trabalho e sintomas depressivos (Harb et al., 2014); etc.

2.1.2 Instrumentos & sensores

Como as emoções não podem ser medidas diretamente, os investigadores recorrem a instrumentos & sensores para recolher dados de contexto, com o objetivo de os utilizar na deteção automática de emoções (Sano and Eng, 2016). Nesta secção apresentam-se alguns dos instrumentos & sensores, das investigações em análise, que são utilizados para recolher dados de contexto. Os restantes podem ser consultados no Relatório Técnico disponível no Repositório Institucional da Universidade Fernando Pessoa (UFP) (Carvalho et al., 2018).

As máquinas são surdas e mudas (Haritaoglu et al., 2001). Como não conseguem comunicar de forma natural com as pessoas através de meios verbais e não-verbais, também

não conseguem identificar diretamente as emoções a partir das expressões humanas. Relativamente aos dados recolhidos do contexto fisiológico, apesar do seu carácter mais objetivo, o seu significado emocional não é direto. No entanto, conhece-se a influência bidirecional desses dados nas emoções (i.e. as emoções podem influenciar os dados de contexto, e esses dados podem ser influenciados pelas emoções) (Rani and Sarkar, 2006). No contexto social e psicológico, também não é possível detetar diretamente emoções a partir dos dados recolhidos. Os computadores são socialmente ignorantes (Pentland, 2005). São ferramentas lógicas e racionais, logo, incompatíveis com a natureza irracional e ilógica das emoções do ser humano (Picard, 2000). Os dados com teor emocional são deduzidos de forma indireta a partir da informação proveniente de outros sistemas (e.g. email, SMS) ou recolhidos através de questionários. Os questionários são muito utilizados na avaliação emocional (Fulton and Medlock, 2003), e para recolha de *ground-truth* (Mandryk and Atkins, 2007).

Neste contexto, alguns investigadores optam por recolher dados de várias modalidades, incluindo através de questionários, com o objetivo de aumentar a precisão dos sistemas (Rani and Sarkar, 2006). No entanto, os instrumentos de medição objetiva, apesar de recolherem dados com maior precisão, podem ser intrusivos ou obstrutivos à rotina normal dos utilizadores, podendo enviesar os próprios dados que se pretendem recolher (e.g. eléctodos para recolha do (HR)) (Guinot Jimeno et al., 2011). Por outro lado, a medição obtida pela aplicação de questionários ou realização de entrevistas, poderá facilitar a generalização por causa da maior abstração das particularidades individuais de cada pessoa (e.g. particularidades fisiológicas, culturais), mas o carácter subjetivo das respostas dos respondentes, e a possível influência dos profissionais que aplicam o questionário ou conduzem a entrevista, podem conduzir a medições imprecisas (Johnston et al., 2009).

Nesta secção, decidiu-se agrupar os instrumentos & sensores pelo tipo de medição que efetuam: medição objetiva, e medição subjetiva. Utiliza-se a notação apresentada na secção 1.3, para representar instrumentos & sensores (e.g. {FUNF}).

2.1.2.1 Medição objetiva

Nesta secção são apresentados os instrumentos & sensores que permitem recolher dados através da medição objetiva.

Os dados obtidos através da medição objetiva tendem a representar pontos objetivos de dados de contexto (Rani and Sarkar, 2006) (e.g. um valor (ST) representa objetivamente uma medição quantitativa, uma (CALL) representa uma interação social real). Apesar da precisão destes dados, não é possível identificar diretamente emoções apenas tendo por base os valores recolhidos. No entanto, estes dados podem fornecer informação importante para um sistema de inferência do estado cognitivo e emocional das pessoas (Schumm et al., 2010).

Para recolher dados de contexto social e psicológico para medição objetiva, é necessário instalar meios físicos ou lógicos de recolha que podem importunar as tarefas diárias do utilizador. Esta obstrutividade pode conduzir a uma recolha enviesada dos dados de contexto, devido à possível interferência na rotina das pessoas e por causa da consciência do utilizador da presença desses dispositivos de recolha (Ouwkerk et al., 2008). A intrusividade também é um aspeto a ter em conta. A recolha de dados privados implica a necessidade de assegurar de forma ainda mais eficiente as questões de ordem ética e de segurança. Apesar da importância da intrusividade, é na obstrutividade que recai o estudo nesta secção, por causa da necessidade de assegurar a qualidade dos dados recolhidos.

No âmbito deste documento, será utilizado o termo obstrutividade para representar o possível bloqueio das atividades normais das pessoas, causado pela utilização de instrumentos de recolha. A organização desta secção é feita em função desse conceito: técnicas não obstrutivas, técnicas pouco obstrutivas, e técnicas obstrutivas. Esta categorização baseia-se na avaliação subjetiva do autor deste documento, acerca da obstrutividade causada por cada dispositivo presente na literatura em revisão.

Técnicas não obstrutivas

A obstrutividade dos dispositivos pode implicar diretamente a qualidade dos dados. Por esse motivo, é uma característica importante a ter em conta, no momento de selecionar os instrumentos & sensores a utilizar numa experiência. Quanto menos obstrutivos forem os dispositivos, maior será a qualidade e fidelidade dos dados recolhidos (Ouwkerk et al., 2008).

Nesta secção apresentam-se os instrumentos & sensores que não implicam obstrutividade para o utilizador, ou que apresentam um grau tão baixo que pode ser ignorado (e.g. instalação de uma nova aplicação {APP} no *smartphone* para recolher dados em segundo plano, sem nenhuma interação do utilizador). Também são considerados nesta secção os instrumentos & sensores cuja implementação real (i.e. não laboratorial), poderá ser feita de forma não obstrutiva ou significar uma obstrutividade tão baixa que pode ser ignorada. Por fim, incluem-se também aqui bases de dados ou outros recursos disponíveis a serem utilizados pelos investigadores como dados de contexto {EXISTINGDATA} (e.g. Bogomolov et al. utilizou o *Wolfram Alpha* (Wolfram Alpha, 2013) para obter dados diariamente acerca do (WEATHER) (Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014), Hernandez et al., utilizou os registos existentes de (CALL) dos colaboradores na sua experiência (Hernandez et al., 2011)).

Uma das estratégias seguidas para recolher dados de forma não obstrutiva, é a utilização de instrumentos já existentes no quotidiano das pessoas. Os *smartphones* são plataformas pervasivas para *sensing* oportunista de comportamentos e de opiniões (Madan et al., 2012), que ocupam um lugar fixo na vida do ser humano (Eckert et al., 2016). Como são amplamente utilizados, são a ferramenta mais adequada à recolha não obstrutiva de dados

de contexto social ou psicofisiológicos (Rachuri et al., 2010). São várias as investigações que utilizam o *smartphone*: i) o *MoodScope* {MOODSCOPE} infere a disposição do utilizador (recolhendo (SMS), (EMAIL), etc.) (LiKamWa et al., 2013); ii) o {FUNF} é uma plataforma que recolhe dados de contexto social (e.g. (CALL), (SMS)) (Aharony and Gardner, 2011); etc. Além destas, existem outras aplicações referenciadas nas investigações que não são totalmente identificadas pelos autores (APP) (e.g. (Eckert et al., 2016)) ou que são modificadas a partir de outras já existentes (e.g. (Sano and Eng, 2016)). Outra estratégia utilizada para diminuir ou eliminar a obstrutividade, é utilizar dispositivos de vídeo {VIDEO}, áudio {AUDIO}, e imagens {PICTURES}. Apesar de poderem significar alguma obstrutividade em laboratório, numa implementação real poderiam ser quase transparentes para o utilizador (e.g. (Lee et al., 2016)).

São vários os instrumentos & sensores utilizados nas investigações analisadas: i) o {FACEREADER} é um sistema de deteção emocional com base no reconhecimento facial (Noldus, 2017); ii) o {KINECT} é um sensor de movimento criado para permitir a interação com jogos digitais (Leyvand et al., 2011); etc.

A Tabela 2.14 resume alguns dos instrumentos presentes na literatura em revisão.

Instrumento	Investigações (exemplo)
Aplicação (<i>software</i>) {APP}	(Eckert et al., 2016) (Sano and Eng, 2016)
Áudio {AUDIO}	(Korkmaz and Atasoy, 2015) (Lalitha et al., 2015)
<i>BeWell</i> ¹ {BEWELL}	(Lane et al., 2011)
<i>EQ-Radio</i> ² {EQ-RADIO}	(Zhao et al., 2016)
Dados existentes {EXISTINGDATA}	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014) (Hernandez et al., 2011)
<i>FaceReader</i> {FACEREADER}	(Matlovic et al., 2016)
<i>Face Tracker</i> ³ {FACETRACKER}	(Adams and Robinson, 2015)
<i>Funf</i> {FUNF}	(Sano and Eng, 2016)
<i>Kinect Sensor</i> {KINECT}	(Gogia et al., 2016) (Saha et al., 2014)
<i>MoodScope</i> {MOODSCOPE}	(LiKamWa et al., 2013)
Imagens {PICTURES}	(Eckert et al., 2016) (Turan et al., 2015)
<i>Shore</i> ⁴ {SHORE}	(Matlovic et al., 2016)
Vídeo {VIDEO}	(Lee et al., 2016) (Gogia et al., 2016)

Tabela 2.14: Resumo de instrumentos & sensores (técnicas não obstrutivas)

Técnicas pouco obstrutivas

Quanto menor for o grau de obstrutividade dos instrumentos & sensores, mais será a predisposição das pessoas para aceitarem a recolha de dados do seu contexto (Schumm et al., 2010). Neste contexto, considera-se que o grau de obstrutividade entre dispositivos obstrutivos não é igual. A utilização de uma cinta peitoral para medição da (RESP), não

¹Monitorização do bem-estar, atividade física, etc. (Lane et al., 2011)

²Permite recolher o (HR) tendo por base a reflexão do sinal RF do WiFi (Zhao et al., 2016)

³*Cambridge Face Tracker*, para reconhecimento facial a partir de video (Thomas et al., 2016)

⁴Sistema para deteção emocional através de reconhecimento facial (Fraunhofer IIS, 2017)

obstrui o contexto do utilizador da mesma forma que a utilização de uma máscara na face para avaliar a quantidade de gases trocados pelos pulmões (Healey et al., 2000). Existem ainda outros dispositivos que, apesar de terem sido apenas utilizados em ambiente experimental, aparentam estar próximo de uma versão implementável na vida real (e.g. Zhang et al. utilizaram óculos para recolha do (EOG), com transmissão via WiFi (Zhang and Sawchuk, 2013)). Nesta secção são apresentados os instrumentos & sensores cuja utilização pode ser esquecida durante a recolha. Dispositivos que se vão tornando transparentes ao longo do tempo (i.e. tecnologia que desaparece no *background*) (Weiser and Brown, 1996).

Foram vários os instrumentos & sensores classificados como pouco obstrutivos: i) o *Zephyr* {BIOHARNESS} é um dispositivo WiFi que permite a recolha do sinal (ECG) e dados relacionados com a respiração (Zephyr Technology, 2012); ii) a {EMOTIONBOARD} mede o (EDA) nos dedos, e movimento de dedos e mãos (Schumm et al., 2008); a pulseira *Toshiba Silmee W20/W21* {SILMEEW2X} consegue medir o (HR), (ST), (SLEEP), etc. (Toshiba, 2015); o *Toshiba Silmee Bar Type* {SILMEEBTYPE} é um pequeno dispositivo para colocar no peito para recolher o (ECG), (PPG) e (BP) (Silmee, n.d.); etc.

Existem também dispositivos específicos criados pelos investigadores: i) Herbon et al. recolheram dados da (PUPIL) utilizando uma câmara montada num capacete (Herbon et al., 2005); ii) Kusserow et al. desenvolveram também vários dispositivos para monitorizar violoncelistas, atletas, etc. (Kusserow et al., 2013); etc.

A Tabela 2.15 resume algumas das técnicas pouco obstrutivas utilizadas pelos investigadores em análise.

Instrumento	Investigações (exemplo)
<i>Affective Q</i> ^{1 2} {AFFECTIVAQ}	(Sano and Eng, 2016) (Jaques et al., 2015)
<i>Zephyr BioHarness</i> {BIOHARNESS}	(Basu et al., 2016)
<i>Brainquiry PET</i> ³ {BQPET}	(Bos, 2010)
<i>Emotion-Board</i> {EMOTIONBOARD}	(Setz et al., 2010)
<i>Jawbone</i> ⁴ {JAWBONE}	(Lisetti and Nasoz, 2004)
<i>Neurosky MindWave</i> ⁵ {MINDWAVE}	(Gogia et al., 2016)
<i>Toshiba Silmee Bar Type</i> {SILMEEBTYPE}	(Zenonos et al., 2016)
<i>Toshiba Silmee W20/W21</i> {SILMEEW2X}	(Zenonos et al., 2016)

Tabela 2.15: Resumo de instrumentos & sensores (técnicas pouco obstrutivas)

Técnicas obstrutivas

As técnicas obstrutivas alteram ou interferem com os hábitos das pessoas. Nesta secção

¹Recolhe o (EDA), (ST) e (ACC), e comunica via WiFi (Affective, 2014).

²Atualmente a Affective é uma empresa que trabalha o *Affective Computing* (Affective, 2021)

³Recolhe o (EEG), (EMG) e (ECG), e comunica via WiFi (Brainquiry, 2017).

⁴Pulseira que recolhe (EDA), (HR), (ST), etc. (Jawbone, 2017).

⁵Recolhe o (EEG) e comunica via WiFi (Neurosky, 2017).

abordam-se os instrumentos & sensores que perturbam a normalidade na execução das atividades diárias das pessoas (Chen et al., 2013). Estes dispositivos foram classificados como obstrutivos porque a sua implementação real teria um impacto grande no dia-a-dia dos utilizadores. Dificilmente perderiam a noção da sua utilização, mesmo que os utilizassem de forma regular durante muito tempo (e.g. utilização de eléctrodos diariamente para recolher o (ECG) ou (EMG)). Incluem-se também nesta secção os instrumentos que implicam ações manuais dos utilizadores {MANUAL} (e.g. contagem de medicamentos tomados).

Nem sempre é possível identificar claramente os instrumentos utilizados pelos autores. Alguns trabalhos apenas identificam os sinais que recolhem, não especificando a marca e modelo do dispositivo utilizado. Estes instrumentos são identificados de forma genérica através da notação *Undefined*<variável> (e.g. o token {Undefined (ECG)} identifica um dispositivo que recolhe o sinal (ECG)).

Neste levantamento bibliográfico identificaram-se vários instrumentos & sensores obstrutivos: i) o {EPOC} da *Emotiv* é um *headset* composto por 16 eléctrodos colocados na cabeça (Ramirez et al., 2015); ii) o kit *Orion Diagnostica* {ORION} mede as concentrações de (CORT) na saliva; iii) o sistema *Biosemi* {ACTIVEII} recolhe vários sinais com eléctrodos (Biosemi, n.d.); iv) o {FLEXCOMP} da *Infiniti*, é um sistema de aquisição e monitorização de sinais de contexto fisiológico (Thought Technology Ltd., 2016); v) os sistemas da ADInstruments {FE116}, {ML135}, {ML309}, e {ML870}, são utilizados na recolha de vários sinais de contexto (e.g. (EDA), (ECG), (ST)) (ADInstruments, n.d.); e vi) os sistemas *Coulbourn* {S71-22} e {S75-01}, utilizados para recolher o (EDA) e o (ECG) (Coulbourn, 2017); o *BioPac* {MP150} que processa vários sinais fisiológicos como o (ECG) (Linton Instrumentation, 2011); o *BioPac* {NIBP100D} que recolhe continuamente o (BP), (SDP) (Biopac Systems Inc, 2017); o *T-Sens GSR* {TGSR} que recolhe o (EDA) (TEA, 2017); etc.

A Tabela 2.16 apresenta alguns instrumentos utilizados pelos investigadores, que são exemplos de técnicas obstrutivas.

Instrumento	Investigações (exemplo)
<i>Biosemi Active II</i> {ACTIVEII}	(Soleymani et al., 2013)
<i>EyeLink 1000</i> ¹ {EL1000}	(Aracena et al., 2016)
<i>Emotiv Epoc</i> {EPOC}	(Matlovic et al., 2016) (Liu et al., 2010)
<i>ADInstruments FE-116 GSR</i> {FE116}	(Basu et al., 2016)
<i>TNO Biomedical Inst.IV</i> ² {FINAPRESS4}	(Ritz et al., 2000)
<i>FlexComp Infiniti</i> {FLEXCOMP}	(Healey and Picard, 2005)
<i>Koralewski Health Lab</i> ³ {HEALTHLAB}	(Lichtenstein, Antje; Oehme, 2008)
Recolha manual {MANUAL}	(Harmer et al., 2003)
<i>ADInstruments ML-135</i> {ML135}	(Basu et al., 2016)
<i>ADInstruments ML-309</i> {ML309}	(Basu et al., 2016)

¹Recolhe dados acerca do olhar e da pupila (SR Research Ltd, 2013).

²Recolhe o (HR), (BP), (SBP) e (DBP), de forma contínua (Imholz et al., 1988).

³Recolhe vários dados de contexto (e.g. (RESP)) (Koralewski, n.d.)

<i>ADInstruments ML-870</i> {ML870}	(Basu et al., 2016) (Chang et al., 2012)
<i>g.MOBILab+</i> ¹ {MOBILAB}	(Cruz et al., 2015)
<i>BioPac MP150</i> {MP150}	(Zhang et al., 2016) (Alzoubi et al., 2013)
<i>BioPac NIBP100D</i> {NIBP100D}	(Zhang et al., 2016)
<i>Procomp Infiniti</i> ² {PROCOMP}	(Kim and André, 2008) (Mandryk and Atkins, 2007)
<i>Coulbourn S71-22</i> {S71-22}	(Bradley et al., 2008) (Vrana, 1993)
<i>Coulbourn S75-01</i> {S75-01}	(Bradley et al., 2008) (Vrana, 1993)
<i>T-Sens GSR</i> {TGSR}	(Matlovic et al., 2016)

Tabela 2.16: Resumo de instrumentos & sensores (técnicas obstrutivas)

2.1.2.2 Medição subjetiva

Os investigadores das ciências comportamentais utilizam questionários e observação como técnicas de medição (Rachuri et al., 2010). Considera-se que estes instrumentos recolhem dados de forma subjetiva, por estarem sujeitos à subjetividade dos respondentes e dos profissionais que aplicam os questionários ou observam as pessoas. Esta subjetividade pode conduzir a medições erradas (Johnston et al., 2009). No entanto, são aceites pela literatura como forma de medir emoções e são bastante utilizados na avaliação emocional (Babiker et al., 2013). A medição subjetiva que efetuam, recolhem dados de contexto de forma mais compatível e comparável entre indivíduos, porque o inquirido responde no contexto da sua especificidade pessoal, num nível de abstração superior.

Além de permitirem mais facilmente a generalização, os questionários recolhem informação direta daquilo que se pretende medir (Mandryk et al., 2006). Na avaliação emocional, os questionários podem ser agrupados em dois tipos: i) os de recolha verbal, que são constituídos por um conjunto de perguntas escritas (palavras); e ii) os de recolha não-verbal cujas perguntas são colocadas de forma visual, com o objetivo de diminuir à subjetividade da interpretação textual (e.g. o *Self-Assessment Manikin* {SAM} mede a reação emocional a um determinado estímulo através de um conjunto de imagens representativas de estados emocionais) (Herbon et al., 2005). Os questionários são utilizados na recolha de dados de contexto, mas também para apoio no processo de seleção de candidatos, e na recolha de dados *ground-truth*.

A questão da obstrutividade também poderia ser levantada na análise dos questionários. No entanto, como a sua aplicação é escalar no tempo e, normalmente são respondidos no início ou fim de ciclos experimentais, decidimos não tratar a obstrutividade destes instrumentos. No entanto, alguns investigadores optam por digitalizar os questionários com o objetivo de facilitar o processo de resposta pelos respondentes.

¹Recolhe (EEG), (EMG), (EOG), etc. (G.tec, 2017)

²Recolhe (EMG), (EDA), (ECG), etc. (Infiniti, 2008).

São muitos os investigadores que utilizam questionários estatisticamente validados, cujo reconhecimento científico é comprovado pela adoção noutras investigações. No entanto, alguns investigadores recorrem à utilização de questionários de forma mais informal. Normalmente são criados para recolher dados específicos de contexto e não têm uma boa sustentação estatística para aferir a sua validade (e.g. Sano et al. criaram um questionário informal porque os formais elegíveis a serem usados, eram demasiado longos (Sano and Eng, 2016)). Decidiu-se por dividir o conteúdo desta secção, em função da formalidade de cada questionário. Inicia-se pela apresentação dos questionários formais identificados nas investigações em análise e, posteriormente, analisam-se alguns exemplos de utilização de questionários informais.

Questionários formais

Nesta secção analisam-se os questionários estatisticamente validados, i.e. aqueles cuja reutilização por vários investigadores lhes providencia validade científica.

São muitos os questionários formais, do tipo verbal, utilizados pelos investigadores: i) o *Positive And Negative Affect Schedule – Expanded* {PANAS-X} é uma versão expandida do {PANAS} que acrescenta um segundo nível de avaliação da disposição com a medição de 11 emoções específicas consideradas pelos psicólogos como emoções básicas (cf. medo, tristeza, culpa, hostilidade, timidez, fadiga, surpresa, jovialidade (bom-humor), autoconfiança, atenção e serenidade) (Watson and Clark, 1999); ii) o *Perceived Stress Scale* {PSS} mede a perceção do (STRESS) (Cohen et al., 1983); iii) o *State-Trait Anxiety Inventory* {STAI} mede a (ANXIETY) (Spielberger, 1983); iv) o *State-Trait Anger Scale* {STAS} e a *State-Trait Anger Expression Inventory* {STAXI} são escalas para medir a (ANGER) (Spielberger, 2010); v) o *Big Five Inventory Personality Test* {BFIPT} é um inventário que mede a personalidade (PERSON) (John, O. P., & Srivastava, 1999); etc.

Os investigadores também utilizam questionários não-verbais: i) o *Self-Assessment Manikin* {SAM} é utilizado para medir (EMOTIONS) através de imagens representativas de estados emocionais (cf. *happy/unhappy*, *excited/calm* e *controlled/uncontrolled*) (Bradley and Lang, 1994); ii) o *Circumplex Model of Affect* {CIRCUMPLEX} mede (EMOTIONS) através de uma representação circular desenhada no plano bidimensional tendo por base o respetivo *arousal* e *valence* (Russell, 1980); e iii) o {AFFECTGRID}, criado com base no {CIRCUMPLEX}, para avaliar o afeto também no espaço bidimensional (cf. *pleasure/displeasure* e *arousal/sleepiness*) (Russell et al., 1989).

Por fim, também são utilizados questionários na fase de seleção de candidatos: i) o *Toronto Alexithymia Scale* {TAS} avalia a capacidade de interpretação de emoções a partir de imagens (Taylor et al., 1985; Bagby et al., 1994); ii) o *Marlowe-Crowne Social Desirability Scale* {MCSDS} avalia a desejabilidade social, (necessidade de aprovação social) (Crowne and Marlowe, 1960); etc.

A Tabela 2.17 resume alguns dos questionários presentes na literatura em análise.

Instrumento	Investigações (exemplo)
<i>Affect Grid</i> {AFFECTGRID}	(Alzoubi et al., 2013)
<i>Big Five Inventory Personality Test</i> {BFIPT}	(Sano and Eng, 2016) (Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014)
<i>Circumplex Model of Affect</i> {CIRCUMPLEX}	(LiKamWa et al., 2013)
<i>Long-term Difficulties Inventory</i> ¹ {LDI}	(Harmer et al., 2003)
<i>Marlowe-Crowne Social Desirability Scale</i> {MCSDS}	(Ritz et al., 2000)
<i>Positive And Negative Affect Schedule – Exapanded Form</i> {PANAS-X}	(Babiker et al., 2013)
<i>Perceived Stress Scale</i> {PSS}	(Sano and Eng, 2016) (van Eck et al., 2005)
<i>Self-Assessment Manikin</i> {SAM}	(Chang et al., 2012) (Lichtenstein, Antje; Oehme, 2008)
<i>State-Trait Anxiety Inventory</i> {STAI}	(Sano and Eng, 2016) (van Eck et al., 2005)
<i>Spielberger State-Trait Anger Scale</i> {STAS}	(van Eck et al., 2005)
<i>Toronto Alexithymia Scale</i> {TAS}	(Ritz et al., 2000) (Sinha, 1996)

Tabela 2.17: Resumo de instrumentos & sensores (questionários formais)

Questionários informais

Apesar de menos utilizados, os questionários informais permitem recolher dados de contexto mais específicos. Nesta secção são apresentados os questionários cuja validade não foi possível verificar tendo por base evidências científicas ou reutilização por outros investigadores. Incluem-se nesta categoria os dados recolhidos a partir de diários ou outro tipo de registos informais por parte das pessoas.

Os investigadores considerados nesta revisão bibliográfica, utilizam questionários informais {INFORMAL} essencialmente para a recolha de dados acerca das emoções sentidas (EMOTIONS): i) Matlovic et al. criaram um questionário informal com as questões “*How strong was the emotion that you felt? (arousal)*”, “*How positive was the emotion that you felt? (valence)*” e “*What emotion did you feel the most?*” (Matlovic et al., 2016); ii) Z. Zhang et al. criaram um formulário para que os respondentes pudessem indicar as emoções sentidas e a intensidade desse sentimento na experiência que conduziu (Zhang et al., 2016); iii) Lisetti et al. solicitaram aos participantes a identificação da emoção sentida aquando da visualização de uma cena de vídeo (era permitida a inserção de uma emoção não presente na lista de opções) (Lisetti and Nasoz, 2004); iv) Rajita Sinha criou um questionário em que pedia ao participante a descrição de uma experiência passada, que provocasse a emoção sentida por cada imagem visualizada (Sinha, 1996); etc.

Os questionários informais são também utilizados para recolher outros dados de contexto psicossocial: i) Sano et al. utilizaram o *smartphone* para colocar questões acerca do sono (SLEEP), stress (STRESS), cansaço (TIRED), etc. (Sano and Picard, 2013b); ii) Sano et al. criaram um questionário para recolher a intensidade sentida de (HEALTH),

¹ Avaliar as dificuldades diárias (Hendriks et al., 1990; Rosmalen et al., 2012).

(ALERT), etc. através de *sliders* [0:100] (Sano and Eng, 2016); iii) Hernandez et al. pediram aos participantes que respondessem à questão “*How was the last call ?*”, utilizando uma escala de *Likert*, para medir o (STRESS) (Hernandez et al., 2011); etc.

A Tabela 2.18 apresenta alguns exemplos de investigações onde são utilizados questionários informais.

Instrumento	Investigações (exemplo)
Questionário informal {INFORMAL}	(Harmer et al., 2003) (Healey and Picard, 2005)

Tabela 2.18: Instrumentos & sensores (questionários informais)

2.1.2.3 Outros instrumentos

Nesta secção são apresentados outros instrumentos utilizados indiretamente no processo de recolha de dados de contexto. Incluem-se também nesta secção os instrumentos de recolha entrevista {INTERVIEW} e a observação {OBSERVATION}, por terem uma utilização residual nas investigações em análise.

Na literatura analisada, identificaram-se vários instrumentos utilizados de forma secundária na recolha de dados de contexto: i) o *Tobii* {TSTUDIO} é uma aplicação utilizada nas investigações que recolhem dados de contexto ocular (Tobii Technology AB, 2015)); ii) o {LABCHART} da *ADInstruments* é uma aplicação que suporta a recolha de dados e monitorização em tempo real, dos dados recolhidos pelos seus sensores (ADInstruments, n.d.); iii) o {BIOEXPLORER} é uma aplicação utilizada na análise de dados do (EEG) (Brainquiry, 2017); iv) o {CAPTIV} é uma ferramenta que facilita o tratamento de dados de vídeo (TEA, 2016); v) os *Software Development Kit* (SDK) do {KINECT} e do {EPOC}, permitem fazer integrações ao nível do *software* {KINECT-SDK} {EPOC-SDK} (Microsoft, 2017a; Emotiv, 2017); etc.

2.1.3 Pré-processamento

A qualidade dos dados é importante por causa do impacto que tem no desempenho algorítmico (Kandel et al., 2012). Os resultados dos algoritmos de classificação, dependem da qualidade dos dados utilizados como *input*. Por esse motivo, os investigadores utilizam técnicas para avaliar a qualidade dos seus dados e assumem medidas para mitigar o impacto de anomalias no processamento dos algoritmos (Chu et al., 2016). Nesta secção abordam-se as técnicas de pré-processamento utilizadas nas investigações relacionadas com o reconhecimento automático de emoções. O objetivo principal destas técnicas é aumentar a qualidade dos dados para otimizar a eficiência dos algoritmos, e criar modelos mais fiéis à realidade (Mokhayeri et al., 2011).

As técnicas de pré-processamento permitem identificar e reduzir a ocorrência de problemas nos *datasets*. Gama et al. categorizou as tarefas de pré-processamento em vários grupos: i) integração de dados de várias fontes; ii) técnicas de amostragem para otimização do conteúdo do *dataset* (objetos e atributos); iii) balanceamento do número de objetos entre classes por causa da sensibilidade de alguns algoritmos ao processamento de conjuntos de dados com classes desbalanceadas; iv) limpeza de dados (e.g. remoção de ruído, reconstrução de dados em falta, incompletos ou inconsistentes); v) redução dimensional para promover a eficiência dos algoritmos (e.g. redução de atributos redundantes, irrelevantes ou inconsistentes); e vi) transformação de dados (e.g. normalização de sinal, conversão de tipos) (Gama et al., 2012).

Tendo por base os tipos de anomalia de Gama et al. e as investigações em análise, decidiu-se organizar esta secção em função da seguinte categorização: i) manutenção do *dataset* que inclui as técnicas relacionadas com a qualidade dos dados; ii) manutenção de sinal que engloba técnicas como a normalização, segmentação e geração de novo sinal; e iii) amostragem e meta-informação onde constam as técnicas de amostragem, balanceamento e etiquetagem de dados. Utiliza-se a notação apresentada na secção 1.3 (e.g. remoção *outliers* [-OUTLIERS]).

2.1.3.1 Manutenção do *dataset*

As técnicas de manutenção têm como objetivo aumentar a qualidade dos dados do *dataset*. Nesta secção abordam-se as técnicas relacionadas com a prevenção de redundância de informação e correção de problemas no conjunto de dados. São vários os problemas que os dados de um *dataset* podem apresentar: valores inesperados ou inconsistentes (i.e. valores impossíveis ou contraditórios entre atributos), valores redundantes (i.e. valor de um atributo pode ser deduzido de outro), dados incompletos (atributos sem valor), etc. (Gama et al., 2012). Tendo por base a literatura em análise, organizou-se o conteúdo desta secção nas seguintes sub-secções: i) ruído, dados incompletos e inconsistentes; e ii) redução dimensional e redundância de objetos.

Ruído, dados incompletos e inconsistentes

A recolha de dados de contexto pode originar a presença de dados no *dataset* que não pertencem à distribuição que os gerou. São várias as razões para existência de ruído nos dados: i) eletrostática; ii) interferências causadas por outros sinais ou sensores (e.g. movimento dos olhos ou músculos causam ruído no sinal (EEG)); e iii) inserção incorreta de dados por parte dos utilizadores; etc. Estes problemas devem ser tratados antes da utilização dos dados pelos algoritmos (Jerritta et al., 2011).

São várias as técnicas utilizadas para a remoção do ruído dos dados [-NOISE]: i) remoção de picos de sinal [-PEAK]; ii) remoção do piscar de olhos em {VIDEO} [-EYEBLINK]; iii) remoção do movimento rápido dos olhos (*saccades*) [-SACCADE]; etc. Ou-

tro problema são os dados incompletos. Este problema pode resultar, por e.g., do esquecimento ou impossibilidade de preenchimento por parte do utilizador. Por este motivo, os investigadores determinam o nível de confiança dos dados [RELIABILITY] e utilizam técnicas de pré-processamento para resolver estes problemas: i) descarte dos dados [DISCARDATA]; ii) geração de dados sintéticos [FAKEDATA]; iii) assunção da inexistência de dados [NULL]; etc. O *dataset* também pode conter dados inconsistentes (e.g. pessoa pesa 130kg com três anos de idade). Para detetar estes casos, os investigadores avaliam a consistência [CONSISTENCY] dos dados, e utilizam técnicas para resolver o problema: i) ajustam manualmente valores [MANDADJUST]; ii) inserem manualmente valores em falta [MANINSERT]; iii) procuram e removem *outliers* [-OUTLIERS]; e iv) utilizam técnicas de mitigação [MITIGATION] para atenuar efeitos externos do sinal (e.g. diminuir o efeito da atividade física no sinal (EDA) recolhido).

A Tabela 2.19 apresenta algumas técnicas utilizadas pelos autores para tratamento do ruído, dados incompletos ou inconsistentes.

Técnica pré-processamento	Investigações (exemplo)
Ruído	
Remover piscar de olhos [-EYEBLINK]	(Gogia et al., 2016) (Aracena et al., 2016)
Diminuir ou remover ruído [-NOISE]	(Sano and Eng, 2016) (Zhao et al., 2016)
Remover picos do sinal [-PEAK]	(Lichtenstein, Antje; Oehme, 2008)
Remover movimentos rápidos do olho [-SACCADE]	(Aracena et al., 2016)
Dados incompletos	
Descartar objetos [DISCARDATA]	(Jaques et al., 2015) (Babiker et al., 2013)
Gerar dados sintéticos [FAKEDATA]	(Lee et al., 2016) (Bradley et al., 2008)
Assumir inexistência de dados [NULL]	(Jaques et al., 2015)
Confiabilidade de dados ¹ [RELIABILITY]	(Healey and Picard, 2005) (Picard et al., 2001)
Dados inconsistentes	
Consistência de dados ² [CONSISTENCY]	(LiKamWa et al., 2013)
Remover valores discrepantes [-OUTLIERS]	(Babiker et al., 2013) (van Eck et al., 2005)
Ajuste manual de dados [MANADJUST]	(Kawai et al., 2013) (Setz et al., 2010)
Inserir manualmente dados [MANINSERT]	(Lane et al., 2011) (Chen et al., 1998)
Atenuar interferências de sinal [MITIGATION]	(Jaques et al., 2015) (Kawai et al., 2013)

Tabela 2.19: Resumo de técnicas de manutenção do *dataset* (ruído, dados incompletos e inconsistentes)

Redução dimensional e redundância de objetos

A dimensão e redundância de dados num *dataset* tem implicações no processamento algorítmico. Os algoritmos de classificação são sensíveis à presença de atributos pouco relevantes ou redundantes num conjunto de dados (Gilad-bachrach, 2004). Nesta secção abordam-se as técnicas utilizadas para lidar com o problema da redundância de atributos e de objetos.

¹Análise do nível de confiabilidade dos dados

²Análise do nível de consistência de dados

A aumento do número de atributos aumenta o número de combinações possíveis de dados. O elevado número de atributos é conhecido como a maldição da dimensionalidade, refere-se ao impacto resultante da introdução de novos atributos num conjunto de dados, e pode influenciar o desempenho do processamento algorítmico (Raschka, 2014). Com o objetivo de promover o desempenho e as taxas de acerto dos algoritmos de classificação, os investigadores utilizam técnicas como o [AGGREGATION] para combinar vários atributos dependentes num único, e selecionam manualmente atributos [MANSELECT] para serem descartados por serem irrelevantes [DISCARDDATA] para o processamento algorítmico.

O processo para resolver o problema da redundância dimensional, pode envolver alguns algoritmos: i) o *Linear Discriminant Analysis* [LDA] é um algoritmo supervisionado que projeta um *dataset* com menos dimensões fazendo uma separação otimizada por classes para evitar o *overfitting* e diminuir o esforço de processamento (Raschka, 2014); ii) o *Principal Component Analysis* [PCA] é um algoritmo não supervisionado que converte um conjunto de objetos possivelmente correlacionáveis num novo *dataset* de dados não correlacionados (o objetivo do algoritmo é encontrar os *principal components* responsáveis pela variância dos dados) (Pearson, 1901); iii) o *Sequential Forward Selection* [SFS] é um algoritmo rápido que vai adicionando propriedades a um conjunto inicialmente vazio, fazendo uma procura *bottom-up* heurística em profundidade nas propriedades disponíveis (Marcano-Cedeno et al., 2010); iv) o *Sequential Backward Selection* [SBS] similar ao [SFS] mas faz uma procura *top-down* e remove uma propriedade a cada iteração (Yıldırım et al., 2014); etc.

Também é possível encontrar atributos redundantes tendo por base outras técnicas: i) o [SCATTER] *plot* permite analisar a interdependência de duas variáveis, visualmente através dos padrões formados no espaço bidimensional cartesiano (Phinyomark et al., 2012); ii) as técnicas de correlação [CORRELATION] permitem analisar a relação existente entre variáveis (Crossman, 2017); iii) o $[X^2]$ permite avaliar a dependência entre duas variáveis (Cambridge University, 2008; Adhikari, 2016); etc.

A redundância de objetos não é tão utilizada pelos investigadores. No entanto, quando existem objetos redundantes, alguns investigadores removem-nos [-DUPLICATE] e, se o mesmo valor de sinal estiver repetido, os investigadores podem considerar apenas um dos objetos (e.g. um objeto por intervalo de tempo [DISTINCTOBJ]).

A Tabela 2.20 resume algumas das técnicas de pré-processamento utilizadas para redução dimensional e tratar a redundância de objetos.

Técnica pré-processamento	Investigações (exemplo)
	Redução dimensional
Agregação do conteúdo ¹ [AGGREGATION]	(Perdiz et al., 2017) (Busso et al., 2004)
Correlação entre variáveis [CORRELATION]	(Sano and Picard, 2013b)

¹Combinação de vários atributos dependentes num único

<i>Linear Discriminant Analysis</i> [LDA]	(Perdiz et al., 2017)
Seleção manual ¹ [MANSELECT]	(Basu et al., 2016) (Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014)
<i>Principal Component Analysis</i> [PCA]	(Zhang et al., 2016) (Sano and Picard, 2013b)
<i>Sequential Backward Selection</i> [SBS]	(Kim and André, 2008) (Busso et al., 2004)
<i>Scatter plot</i> ² [SCATTER]	(Phinyomark et al., 2012) (Matiko et al., 2014)
<i>Sequential Forward Selection</i> [SFS]	(LiKamWa et al., 2013)
<i>Chi-square</i> [X2]	(Alzoubi et al., 2013)
Redundância de objetos	
Apenas objetos distintos [DISTINCTOBJ]	(Sano and Eng, 2016)
Remoção de objetos duplicados [-DUPLICATE]	(Gogia et al., 2016)

Tabela 2.20: Resumo de técnicas de manutenção do *dataset* (redução dimensional e redundância de objetos)

2.1.3.2 Manutenção de sinal

Os algoritmos de classificação podem não conseguir utilizar os dados no seu formato original. Podem ser sensíveis a determinados tipos de dados ou à existência de intervalos numéricos diferentes entre atributos (Gama et al., 2012). Nesta secção são apresentadas técnicas de pré-processamento responsáveis por adequar os sinais ao processamento de algoritmos. Resolveu-se agrupar as técnicas identificadas em três categorias: i) conversão de sinal e de tipos de dados; ii) normalização, amplificação e filtros; e iii) segmentação e geração de novo sinal.

Conversão de sinal

Por vezes os investigadores têm necessidade de converter o sinal, para os adequar ao processamento nos sistemas (Singh et al., 2015). A categoria conversão de sinal agrupa um conjunto de técnicas de pré-processamento bastante utilizadas pelos investigadores: i) conversão do domínio do tempo para o domínio das frequências (e.g. transformadas de *Fourier* [FOURIER]) (Bracewell, 2014); ii) transformada de *Z* [ZTRANSFORM] utilizada na análise de sinais digitais discretos (Smith, 1997); iii) o *Hilbert-Huang Transform* [HHT] utilizado na análise de sinal não linear e não estacionário (Tan, 2016); etc.

A conversão de tipos de dados é também considerada nesta secção [QUALI-QUANTI]. A conversão de um atributo pode ser necessária por vários motivos, nomeadamente pelo facto do novo tipo ser mais adequado à utilização num determinado algoritmo. Alguns algoritmos apenas processam valores numéricos (e.g. *Support Vector Machine* (SVM)), e outros são mais adequados para trabalhar com valores qualitativos como nos modelos Bayesianos (Gama et al., 2012).

A Tabela 2.21 apresenta algumas das técnicas utilizadas pela literatura em revisão.

¹Seleção manual de atributos irrelevantes

²Análise visual da redundância entre atributos

Técnica pré-processamento	Investigações (exemplo)
Conversão de sinal	
Transformada de <i>Fourier</i> [FOURIER]	(Korkmaz and Atasoy, 2015) (Bos, 2010)
<i>Hilbert-Huang Transform</i> [HHT]	(Nawasalkar et al., 2013)
Conversão de vídeo em imagens [VIDEO-PICS]	(Singh et al., 2015) (Saha et al., 2014)
Transformada de <i>Z</i> [ZTRANSFORM]	(Herbon et al., 2005)
Conversão de tipo de dados	
Bidimensional para discreto ¹ [2D-DISCRETE]	(Liu et al., 2010)
Qualitativos <-> quantitativos [QUALI-QUANTI]	(Vrana, 1993)

Tabela 2.21: Resumo de técnicas de manutenção do *dataset* (conversão de sinal)

Normalização, amplificação e filtros

A normalização [NORM] de dados, é uma das técnicas de pré-processamento mais utilizadas pelos investigadores. Esta técnica é utilizada em vários contextos: preparar dados ao processamento dos algoritmos, uniformizar unidades de medida, padronizar limites entre atributos, etc. (Gama et al., 2012). A literatura também relata outras formas de normalização mais específicas (e.g. normalização de tamanho de imagens [IMGSIZE]).

Nesta revisão bibliográfica, identificaram-se também técnicas de pré-processamento relacionadas com a amplificação [SIGAMP] e filtros [filtros] de sinal: *Low-Pass Filter* [LPF]; *Band-Pass Filter* [BPF]; etc.

A Tabela 2.22 apresenta algumas técnicas utilizadas pelos investigadores.

Técnica pré-processamento	Investigações (exemplo)
Normalização de sinal	
Tamanho de imagens [IMGSIZE]	(Mokhayeri et al., 2011) (Wang et al., 2010)
Normalização de dados [NORM]	(Perdiz et al., 2017) (Lee et al., 2016)
Amplificação de sinal	
Amplificação de sinal [SIGAMP]	(Perdiz et al., 2017) (Korkmaz and Atasoy, 2015)
Filtros	
Filtro passa-banda [BPF]	(Perdiz et al., 2017) (Cruz et al., 2015)
Filtro passa-alta ² [HPF]	(Chang et al., 2012) (Setz et al., 2010)
Filtro passa-baixa [LPF]	(Sano and Eng, 2016) (Zhao et al., 2016)

Tabela 2.22: Resumo de técnicas de manutenção do *dataset* (normalização, amplificação e filtros)

Segmentação e geração de novo sinal

Nesta secção abordam-se as técnicas de pré-processamento relacionadas com a segmentação de dados e a geração de novo sinal. A segmentação também pode ser vista como geradora de sinal. No entanto, decidiu-se separar as técnicas cujo único propósito é

¹Conversão do espaço bidimensional para discreto (e.g. níveis *arousal/valente* para emoções)

²*High-Pass Filter* [HPF]

apenas a geração de novo sinal.

São várias as técnicas de segmentação utilizadas nas investigações em análise: i) repartir o sinal em vários sinais mais pequenos [INTERVALSPLIT] (e.g. Liu et al. dividiram o sinal (EEG) recolhido em vários intervalos de tempo (Liu et al., 2010)); ii) separação do sinal em bandas [BANDS] (e.g. Soleymani et al. segmentaram o sinal (EEG) nas bandas *theta*, *alpha*, *beta* e *gamma* (Soleymani et al., 2013)); iii) Matlovic et al. também segmentaram o sinal do (EEG) para extrair as bandas *alpha* e *beta*, mas utilizando o *Discrete Wavelet Transform* [DWT] (Matlovic et al., 2016); etc.

Os investigadores também utilizam técnicas para geração de novo sinal tendo por base os dados do *dataset* inicial: i) primeira e segunda derivada [DY/DX] [D2Y/DX2]; ii) dedução de locais geográficos habituais [USUALPLACES]; iii) trajetos regularmente percorridos [PATHSTAKEN]; iv) deteção de atividades diárias [ADL]; etc;

Incluem-se nesta secção as técnicas de integração ou fusão de dados [INTEGRATION] (e.g. Jaques et al. integraram as coordenadas do GPS, com os dados do WiFi, e do telemóvel (antenas do operador) para criar um registo único de localização dos utilizadores (Jaques et al., 2015); Soleymani et al. fundiram dados provenientes de várias modalidades num único *dataset* (Soleymani et al., 2013)).

A Tabela 2.23 identifica algumas das técnicas, relacionadas com a segmentação e geração de novo sinal, presentes na literatura em análise.

Técnica pré-processamento	Investigações (exemplo)
Segmentação de sinal	
Segmentação em bandas [BANDS]	(Zenonos et al., 2016) (Soleymani et al., 2013)
<i>Discrete Wavelet Transform</i> [DWT]	(Matlovic et al., 2016) (Lalitha et al., 2015)
Realce de formas ¹ [-IMGBKG]	(Singh et al., 2015) (Saha et al., 2014)
Divisão em intervalos de tempo [INTERVALSPLIT]	(Korkmaz and Atasoy, 2015) (Cruz et al., 2015)
Geração de novo sinal	
Identificação de tarefas diárias. [ADL]	(Sano and Eng, 2016) (Sano and Picard, 2013b)
Segunda derivada [D2Y/DX2]	(Zhao et al., 2016) (Korkmaz and Atasoy, 2015)
Primeira derivada [DY/DX]	(Sano and Eng, 2016) (Sano and Picard, 2013b)
Integração ou fusão de dados [INTEGRATION]	(Jaques et al., 2015) (Kusserow et al., 2013)
Trajetos percorridos [PATHSTAKEN]	(Jaques et al., 2015)
Locais frequentes [USUALPLACES]	(Bauer and Lukowicz, 2012)

Tabela 2.23: Resumo de técnicas de manutenção do *dataset* (segmentação e geração de novo sinal)

2.1.3.3 Amostragem e meta-informação

Os investigadores utilizam amostras em representação do *dataset* completo. Mesmo que os conjuntos de dados sejam grandes, não significa que todos os objetos e atributos sejam necessários ao processamento algorítmico. Aliás, grandes quantidades de dados,

¹E.g. remoção do fundo para cálculo do declive dos ombros (Singh et al., 2015)

podem se prejudiciais para alguns algoritmos. Podem conduzir à saturação de memória e quebras de desempenho (Phinyomark et al., 2012). Nesta secção abordam-se as técnicas de amostragem que permitem a criação de sub-conjuntos de dados com o objetivo de promover o desempenho dos algoritmos. Aborda-se também nesta secção, o balanceamento entre classes e o enriquecimento dos objetos com meta-informação (i.e. etiquetagem).

Técnicas de amostragem

O equilíbrio entre a quantidade de dados e o desempenho computacional é um desafio para os processos de indução (Gama et al., 2012). Os algoritmos de classificação podem conseguir melhores taxas de acerto com *datasets* maiores. No entanto, quanto mais forem os dados, mais informação existe para processar, e maior será o esforço de processamento que pode conduzir a uma degradação no desempenho dos sistemas (Zhao et al., 2016). É por este motivo que os investigadores utilizam amostras em representação do *dataset* completo.

Uma amostra é um sub-conjunto de dados de tamanho n , representativo do conjunto de dados original de tamanho N (Pocinho, 2009). A utilização de uma amostra em representação do *dataset* original, reduz a quantidade de informação a processar e pode promover o desempenho algorítmico (Zhao et al., 2016). A amostra deve ser pequena, e representativa do *dataset* original do ponto de vista da sua distribuição estatística (Investopedia, 2017).

Em processamento de sinal, uma das técnicas utilizadas é reduzir a taxa de amostragem (*downsampling*) [DECIMATION] (Lyons, 2004). No entanto, existem também técnicas probabilísticas para efetuar a amostragem de um conjunto de dados: i) amostragem aleatória simples [RANDOM] em que os objetos a considerar na amostra são sorteados; ii) amostragem estratificada [STRATIFIED] em que os objetos são inicialmente divididos em extratos e , posteriormente, são escolhidos objetos de cada um desses extratos; e iii) amostragem sistemática [SYSTEMATIC] em que é selecionado um objeto a cada intervalo r , iniciando-se num objeto sorteado k . Por fim, alguns investigadores selecionam objetos com base num critério por si definido, constituindo uma amostra intencional [INTENTIONAL] (IPLeiria, 2009).

Os autores nem sempre indicam a utilização de técnicas de amostragem nas suas investigações. A Tabela 2.24 apresenta algumas investigações em análise que relatam a utilização destas técnicas.

Técnica pré-processamento	Investigações (exemplo)
Diminuir da taxa de amostragem ¹ [DECIMATION]	(Kim et al., 2004)
Amostragem personalizada ² [INTENTIONAL]	(Vrijkotte et al., 2000)
Amostragem aleatória [RANDOM]	(Zhang et al., 2016)

¹*Downsampling*

²O critério de amostragem é definido pelo investigador

Amostragem extratificada [STRATIFIED]	(Mandryk and Atkins, 2007) (Vrijkotte et al., 2000)
A seleção é feita a cada [SYSTEMATIC]	(Aracena et al., 2016) (Babiker et al., 2013)

Tabela 2.24: Resumo de técnicas de amostragem e meta-informação (técnicas de amostragem)

Balanceamento e etiquetagem

A composição das amostras deve atender ao balanceamento entre classes. Quando treinados com um conjunto de dados desbalanceado, os algoritmos tendem a classificar na classe com maior número de objetos. A solução para este problema passa por aumentar o número de objetos recolhendo mais dados de contexto ou gerando dados artificiais para balancear as classes [+ARTIFICIALDATA] (Gama et al., 2012).

A etiquetagem de dados permite aos investigadores acrescentar meta-informação ao *dataset* [LABELING]. São vários os autores que anotam os seus dados com informação adicional (e.g. Gogia et al. anotaram o sinal (EEG) recolhido com informação acerca do estado de atenção dos participantes aquando da visualização de VIDEO (Gogia et al., 2016)).

A Tabela 2.25 apresenta algumas investigações que utilizaram técnicas de balanceamento de classes e etiquetagem de dados.

Técnica pré-processamento	Investigações (exemplo)
	Balanceamento
Gerar dados sintéticos [+ARTIFICIALDATA]	(Gogia et al., 2016) (Wang et al., 2010)
	Etiquetagem
Etiquetagem de dados [LABELING]	(Gogia et al., 2016) (Zhang et al., 2016)

Tabela 2.25: Resumo de técnicas de amostragem e meta-informação (balanceamento e etiquetagem)

2.1.3.4 Outras técnicas e instrumentos

Nesta secção apresentam-se outras técnicas não apresentadas nas secções anteriores. Essencialmente são técnicas auxiliares à aplicação de outras técnicas de pré-processamento, e técnicas para representação de dados. São também apresentados nesta secção alguns instrumentos utilizados na fase do pré-processamento.

Os investigadores utilizam técnicas de suporte para tratamento de vídeo, áudio e imagem: i) algoritmos para detetar regiões em imagens [FINDREGION]; ii) técnicas de *clustering* para identificação de zonas em imagem [CLUSTERING]; iii) recorte de imagens [CROP]; iv) deteção de piscar de olhos [BLINKDETECT]; v) deteção de movimento [MOTIONDETECT]; vi) deteção de picos de sinal [PEAKDETECT]; etc.

Ao nível da representação de dados, os investigadores utilizam técnicas como *Sparse Representation* [SPARSEREP], histogramas [HISTOGRAM], etc.

2.1.4 Propriedades extraídas

A escolha dos dados a serem utilizados como *input* nos algoritmos de classificação, é um passo importante nos sistemas de reconhecimento emocional, porque a precisão dos algoritmos depende da qualidade desses dados (Adams and Robinson, 2015). As propriedades extraídas são mais apropriadas ao processo de classificação, e podem ser vistas como *inputs* complementares ou de substituição aos atributos do *dataset* original (Haag et al., 2004). Nesta secção abordam-se as propriedades extraídas utilizadas pelos autores nos sistemas de reconhecimento emocional.

Nesta revisão bibliográfica identificaram-se quatro categorias de propriedades extraídas: i) propriedades de domínio; ii) propriedades específicas de contexto; iii) as propriedades matemáticas e de análise de sinal; e iv) outras propriedades mais genéricas. Nesta secção são analisadas, principalmente, as propriedades específicas de contexto, porque aumentam a variedade e qualidade do *dataset* (Matlovic et al., 2016). As propriedades de domínio já foram identificadas na secção 2.1.1 pelo facto do método de obtenção ser confundível com o utilizado na própria recolha. As propriedades matemáticas e de análise de sinal, e as propriedades genéricas, não são abordadas em detalhe por se considerar serem de conhecimento geral (e.g. média <MEAN>). No entanto, como são bastante referenciadas pela literatura, resumem-se na Tabela 2.26. A organização de conteúdos desta secção, segue a estratégia de enquadramento das propriedades extraídas, com as respetivas origens das variáveis de contexto.

Propriedade extraída	Descrição
Medidas de tendência central	
<MEAN> <MEDIAN> <MODE>	Média, mediana e moda.
<MAV> <MAVS>	Média dos valores absolutos e versão pesada por <i>windows functions</i> .
<MAV1D> <MAV2D>	Média dos valores absolutos da primeira e da segunda diferença.
<Q1> <Q3> <IQR>	Quartil 1, quartil 3 e <i>InterQuartile Range</i> .
Medidas de dispersão e distribuição	
<VAR> <COVAR>	Variância e covariância.
<STD> <AD>	Desvio padrão e desvio absoluto.
<MIN> <MAX> <AMP> <CV>	Mínimo, máximo, amplitude e coeficiente de variação.
<SKEWNESS> <KURTOSIS>	Obliquidade, curtose.
Medidas geométricas	
<ANGLE>	Ângulo.
<DIAMETER> <RADIUS> <AREA>	Diâmetro, raio e área.
<DISTANCE> <SIZE> <SLOPE>	Distância, tamanho e declive.
<CENTROID> <SHAPE>	Centro geométrico de <i>clusters</i> e forma.
Análise de sinal (domínio do tempo)	
<i>Waveform length</i> <WL>	Complexidade de um sinal.
<i>Willison Amplitude</i> <WAMP>	Quantas vezes dois segmentos adjacentes excederam um <i>threshold</i> .
<i>Auto-Regressive</i> <AR>	Descreve cada amostra como uma combinação de várias amostras.

Análise de sinal (domínio de frequência)	
<i>Power Spectral Density</i> <PSD>	Mostra a distribuição da <VAR> em função da frequência.
<PKF> <PLF> <PHF>	Pico de frequência, e pico de baixa e alta frequência.
<DMF> <LF/HF>	Frequência dominante, rácio entre baixa e alta frequência.
<LFMF/HF>	Rácio da soma da baixa e média frequência com a alta frequência.
<MAG> <ENERGY> <POWER> <STRENGTH>	Magnitude, energia, <i>power</i> e força.
<MAINPEAK> <RISETIME> <RECOVERYTIME> <HALFRECOVERYTIME>	Pico principal tempo de subida do sinal, tempo de recuperação e metade do tempo de recuperação.
<INTEGRAL>	Integral da função.
<1DIFF> <2DIFF>	Primeira e segunda diferença (aprox. [DY/DX] e [D2Y/DX2]).
<ZEROCROSSINGS>	Número de <i>zero crossings</i> .
<CHANGES> <TPOINTS>	Alterações e pontos de viragem do sinal.
Outras	
<RATIO> <SUM> <COUNT>	Relação entre valores, soma e quantidade de ocorrências.
<DURATION> <PERIODS>	Duração e agrupamento de dados em períodos (segmentos).
<LATENCY> <ELAPSEDTIME>	Latência e tempo decorrido entre ações.
<DISPLACEMENT> <MOTION>	Deslocamento e movimento.
<SPEED> <DIRECTION>	Velocidade e direção.
<LOCATION> <POSITION>	Localização e posição.
<INTENSITY> <THRESHOLD>	Intensidade e limiar (limite).
<LEFT> <CENTER> <RIGHT>	Esquerda, centro e direita.
<BEGIN> <END>	Início e fim.
<FIRST> <SECOND> <THIRD>	Primeiro, segundo e terceiro.
<TIMING> <LASTTIME>	Momento ou última vez que uma tarefa ou evento ocorreu.
<PLANNING> <REGULARITY>	Planeamento e regularidade de uma ação ou evento.
<QUALITY> <QUANTITY>	Qualidade e quantidade.
<LEVEL> <HEIGHT>	Nível e altura.
<PHASES> <STATUS>	Fase e estado.
<GROUP> <CATEGORY>	Agrupamento e categoria.
<WHY> <USUAL>	Motivo e tarefa ou ação mais usual.

Tabela 2.26: Resumo das propriedades genéricas

2.1.4.1 Expressão facial, oral e postura corporal

As propriedades extraídas relacionadas com as expressões e postura corporal, utilizadas pelos investigadores, são retiradas essencialmente do {VIDEO}, {AUDIO} e {PICTURES}. Nesta secção apresentam-se as propriedades extraídas mais utilizadas pelos investigadores, a partir dos dados de contexto recolhidos da expressão facial, oral e postura corporal.

Na expressão facial e postura corporal são extraídas propriedades a partir de {VIDEO} e {PICTURES}. São várias as propriedades identificadas nesta secção: i) testa ou nariz enrugado <WRINKLEDFH> <WRINKLEDNOSE>; ii) abertura <MOUTHOPENING>, alongamento <MOUTHSTRETCH> e extremidades da boca <MOUTHCORNERS>; iii) a íris <IRIS> e extremidades dos olhos <EYECORNERS>; iv) cruzamento de braços <CROSSEDARMS> e batidas dos dedos <FINGERTAPPING>; v) costas direitas <ERECTBACK> e postura

sentada <SITTINGPOSTURE>; vi) *clusters* <CLUSTERS> (e.g. Lee et al. utilizaram {CLUSTERING} para reconhecimento de partes da face para depois calcular o <DISPLACEMENT> geométrico de zonas (Lee et al., 2016)); etc.

Alguns investigadores utilizam descritores por serem menos sensíveis à intensidade luminosa e menos dependentes da pose das pessoas nas imagens e vídeos, no tratamento dos dados da expressão facial (Turan et al., 2015): i) o *Local Binary Pattern* <LBP> é o descritor relativo à textura com base na intensidade; ii) o *Local Phase Quantization* <LPQ>, descritor relativo à textura com base no *blur*; etc. (Shen et al., 2007).

Na expressão oral os *Mel Frequency Cepstral Coefficients* <MFCC>, são muito usados por sistemas de reconhecimento automático de discurso oral e de oradores, porque imitam o funcionamento do ouvido humano (Cryptography, 2009). Os <MFCC> são ideais para processar as propriedades relacionadas com os problemas de reconhecimento emocional (Korkmaz and Atasoy, 2015). São ainda utilizadas outras propriedades extraídas como o *Teager Energy* <TEAE>, *Harmonic to Noise Ratio* <HNR>, *zero Crossing Rate* <ZCR>, etc. (Lalitha et al., 2015).

A Tabela 2.27 apresenta algumas das propriedades extraídas utilizadas nas investigações em análise.

Propriedade extraída	Investigações (exemplo)
Expressão facial e postura corporal	
<i>Clusters</i> em imagens ou vídeo <CLUSTERS>	(Lee et al., 2016) (Busso et al., 2004)
Cruzamento dos braços <CROSSEDARMS>	(Vermun et al., 2013)
Costas direitas <ERECTBACK>	(Vermun et al., 2013)
Extremidades dos olhos <EYECORNERS>	(Soleymani et al., 2013)
Batidas dos dedos <FINGERTAPPING>	(Gunes and Piccardi, 2007)
Íris do olho <IRIS>	(Soleymani et al., 2013)
Extremidades da boca <MOUTHCORNERS>	(Gunes and Piccardi, 2007)
Abertura da boca <MOUTHOPENING>	(Vermun et al., 2013)
Alongamento da boca <MOUTHSTRETCH>	(Gunes and Piccardi, 2007)
Postura sentada <SITTINGPOSTURE>	(Vermun et al., 2013)
Testa enrugada <WRINKLEDFH>	(Gunes and Piccardi, 2007)
Nariz enrugado <WRINKLEDNOSE>	(Gunes and Piccardi, 2007)
Expressão facial	
<i>Local Binary Pattern</i> <LBP>	(Lee et al., 2016) (Turan et al., 2015)
<i>Local Phase Quantization</i> <LPQ>	(Yang and Bhanu, 2011) (Dhall et al., 2011)
Expressão oral	
<i>Harmonic to Noise Ratio</i> <HNR>	(Lalitha et al., 2015)
<i>Mel Frequency Cepstral Coefficients</i> <MFCC>	(Korkmaz and Atasoy, 2015) (Nwe et al., 2001)
<i>Teager Energy</i> <TEAE>	(Lalitha et al., 2015)
<i>Zero Crossing Rate</i> <ZCR>	(Korkmaz and Atasoy, 2015) (Lalitha et al., 2015)

Tabela 2.27: Resumo das propriedades específicas de contexto (expressão facial, oral e postura corporal)

2.1.4.2 Contexto fisiológico

Nesta secção apresentam-se algumas propriedades específicas de contexto, relacionadas com as variáveis recolhidas do contexto fisiológico.

Ao nível da atividade cerebral, são várias as propriedades extraídas do sinal (EEG) utilizadas pelos autores da literatura em revisão: i) as ondas <ALPHA>, <BETA>, <THETA>, <DELTA> e <GAMMA> (e.g. Soleymani et al. extraíram estas propriedades a partir do <PSD> (Soleymani et al., 2013)); ii) o nível de meditação <MEDITATION> e atenção <ATTENTION> (e.g. Gogia et al. extraíram estas propriedades do sinal (EEG) (Gogia et al., 2016)); iii) a atividade dos hemisférios <HEMISPHERE> do cérebro (e.g. Matiko et al. mediram a atividade cerebral através do sinal (EEG) (Matiko et al., 2014)); etc.

Relativamente à atividade cardíaca, são também várias as propriedades extraídas pelos investigadores: i) o *Heart Rate Variability Index* <HRVI> que representa a divisão dos *inter-beat intervals* pela altura do histograma de todos os (IBI) (Zenonos et al., 2016); ii) o *Heart Rate Variability Triangular Index* <HRVTI> é uma propriedade extraída do (HRV) (Mehta et al., 2002), que representa o integral da densidade dos intervalos N-N dividida pelo máximo da densidade de distribuição (Electrophysiology, 1996); iii) o *Triangular Interpolation of N-N Interval Histogram* <TINN> é o tamanho base da distribuição medida como a base de um triângulo aproximando a distribuição do intervalo N-N (Jarkovska et al., 2016); etc.

A partir da atividade muscular, é possível extrair várias propriedades: i) atividade <MUSCLE>, contrações <CONTRACTIONS> e tensão <TENSION> muscular; ii) a atividade dos músculos *corrugator supercilii muscle* (pequeno músculo na zona da sobrancelha) <EMGCOR>, *zygomaticus major muscle* (músculo que se estende desde o osso zigomático até o ângulo da boca) <EMGZYG> e *Levator labii superioris* (o músculo levantador do lábio superior) <LEVATOR>; iii) o sorrir <SMILING> e o ar carrancudo (mal-humorado) <FROWNING>; etc.

Aos investigadores também extraem propriedades da atividade ocular: i) dilatação <DILATION> (aumento ou diminuição da (PUPIL)); ii) diâmetro <DIAMETER> da (PUPIL); ¹ iii) fixação do olhar <FIXATION> iv) movimentação horizontal <EOGH> e vertical dos olhos <EOGV>; etc.

A Tabela 2.28 apresenta algumas das propriedades extraídas de contexto fisiológico, utilizadas nas investigações em análise.

Propriedade extraída	Investigações (exemplo)
	Atividade cerebral
Ondas <i>alpha</i> <ALPHA>	(Matlovic et al., 2016) (Sano and Eng, 2016)
Nível de atenção <ATTENTION>	(Gogia et al., 2016)
Ondas <i>beta</i> <BETA>	(Matlovic et al., 2016) (Sano and Eng, 2016)
Ondas <i>delta</i> <DELTA>	(Murad and Malkawi, 2012) (Liu et al., 2010)

¹Os autores referem-se ao tamanho da (PUPIL) como diâmetro <DIAMETER> ou tamanho <SIZE>. Escolheu-se por unificar este tamanho no *token* <DIAMETER>.

Ondas <i>gamma</i> <GAMMA>	(Soleymani et al., 2013) (Liu et al., 2010)
Atividade hemisférica <HEMISPHERE>	(Matiko et al., 2014) (Soleymani et al., 2013)
Nível de meditação <MEDITATION>	(Gogia et al., 2016)
Ondas <i>theta</i> <THETA>	(Sano and Eng, 2016)
Atividade cardíaca	
<i>Heart Rate Variability Index</i> <HRVI>	(Zenonos et al., 2016)
<i>Heart Rate Variability Triangular Index</i> <HRVTI>	(Zhao et al., 2016)
<i>Triangular interp. of N-N Interval Histogram</i> <TINN>	(Zhao et al., 2016) (Zenonos et al., 2016)
Atividade muscular	
Contrações musculares <CONTRACTIONS>	(Lichtenstein, Antje; Oehme, 2008)
<i>Corrugator supercilii muscle</i> ¹ <EMGCOR>	(Perdiz et al., 2017) (Basu et al., 2016)
<i>Zygomaticus major muscle</i> ¹ <EMGZYG>	(Perdiz et al., 2017) (Basu et al., 2016)
Franzir de sobranceiras, ar carrancudo <FROWNING>	(Mandryk and Atkins, 2007) (Partala et al., 2005)
<i>Levator labii superioris</i> ¹ <LEVATOR>	(Vrana, 1993)
Atividade muscular <MUSCLE>	(Lichtenstein, Antje; Oehme, 2008)
Sorrir <SMILING>	(Mandryk and Atkins, 2007) (Partala et al., 2005)
Tensão muscular <TENSION>	(Vrana, 1993)
Atividade ocular	
Dilatação da pupila <DILATION>	(Aracena et al., 2016) (Babiker et al., 2013)
Movimento horizontal do olho <EOGH>	(Cruz et al., 2015)
Movimento vertical do olho <EOGV>	(Cruz et al., 2015)
Fixação do olhar <FIXATION>	(Babiker et al., 2013) (Partala and Surakka, 2003)

Tabela 2.28: Resumo das propriedades específicas do contexto (contexto fisiológico)

2.1.4.3 Contexto social e psicológico

Nesta secção apresentam-se as propriedades extraídas dos dados provenientes do contexto social e psicológico. O instrumento mais utilizado para recolher estes dados são os questionários. As perguntas colocadas são quase sempre de resposta fechada para diminuir a subjetividade. As respostas fechadas são normalmente respondidas através de escalas de *Likert*, ou através de um conjunto de etiquetas, nomes ou símbolos (e.g. positivo, negativo; pouco, médio, muito). São várias as propriedades possíveis de extrair destas respostas (e.g. <COUNT>, <MODE>), e os autores nem sempre indicam quais as propriedades que utilizam. Assim decidiu-se agrupar as propriedades possíveis de extrair com o *token* <CATEGORICAL> atendendo à natureza categórica destes dados.

Os investigadores extraem várias propriedades de contexto relativamente aos dados das (CALL), (SMS) e (EMAIL): i) comunicações recebidas <INCOMING>, efetuadas <OUT GOING> e (CALL) não atendidas <MISSEDCALL>; ii) interações apenas iniciadas e não concretizadas <INITIATEDONLY>; iii) número interlocutores <INTERLOCUTORS>; etc. A localização (LOCAL) é tratada pelos investigadores através de várias propriedades: i) a latitude <LATITUDE> e a longitude <LONGITUDE> do sistema *Global Position System* (GPS); ii) as *Regions of Interest* <ROI> (e.g. Bauer et al. utilizaram as localizações

¹Atividade muscular

habitualmente visitadas pelos participantes (Bauer and Lukowicz, 2012)); etc. Ao nível da variável (SLEEP) os investigadores também extraem algumas propriedades específicas de contexto: i) número de despertares inesperados durante o sono <AWAKENINGS>; ii) forma de acordar <WAKEUPKIND> (e.g. acordar espontâneo, despertador); etc. Também são extraídas propriedades do (STRESS): i) último evento stressante <LASTSTRESS>; ii) recompensa (e.g. salário, estatuto) <REWARD>; etc. A partir da variável de contexto (PERSON) são consideradas como propriedades as várias dimensões recolhidas através do {BFIPT}: i) a extroversão <NEUROTICISM>; ii) a afabilidade <AGREEABLENESS>; etc.

Existem ainda outras propriedades extraídas de dados relacionados com variáveis de contexto relativas à interação social: i) o número de dispositivos nas imediações <DEVICES> (variável de contexto (PROXIMITY)); ii) qualidade das interações sociais <INTERACTIONS> e a afinidade com essas pessoas <AFFINITY>, etc.

A Tabela 2.29 apresenta algumas das propriedades específicas de contexto utilizadas nas investigações em análise, do contexto social e psicológico.

Propriedade extraída	Investigações (exemplo)
	(CALL), (SMS) e (EMAIL)
Comunicações recebidas <INCOMING>	(Sano and Eng, 2016) (Jaques et al., 2015)
Chamadas apenas iniciadas <INITIATEDONLY>	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014)
Número de interlocutores <INTERLOCUTORS>	(Sano and Eng, 2016) (Jaques et al., 2015)
Chamadas não atendidas <MISSEDCALL>	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014) (Sano and Picard, 2013b)
Comunicações efetuadas <OUTGOING>	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014) (Sano and Picard, 2013b)
	(LOCAL)
Latitude e longitude <LATITUDE>	(Jaques et al., 2015)
Regions of Interest <ROI>	(Bauer and Lukowicz, 2012)
	(SLEEP)
Despertares durante o sono <AWAKENINGS>	(Sano and Eng, 2016)
Tipo de acordar <WAKEUPKIND>	(Sano and Eng, 2016)
	(STRESS)
Último evento stressante <LASTSTRESS>	(van Eck et al., 2005)
Recompensa <REWARD>	(Vrijkotte et al., 2000)
	(PERSON)
Agradabilidade <AGREEABLENESS>	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014) (Sano and Picard, 2013b)
Neuroticismo <NEUROTICISM>	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014) (Sano and Picard, 2013b)
	Outras
Afinidade <AFFINITY>	(Sano and Eng, 2016)
Número de dispositivos detetado <DEVICES>	(Bogomolov, Lepri, Ferron, Pianesi and Pentland, 2014)
Qualidade das interações sociais <INTERACTIONS>	(Sano and Eng, 2016)

Tabela 2.29: Resumo das propriedades específicas do contexto (contexto social e psicológico)

2.1.4.4 Outras

Por fim, nesta secção apresentam-se algumas propriedades extraídas de outras variáveis de contexto recolhidas pelos investigadores.

A variável de contexto (WEATHER) é recolhida através de vários instrumentos & sensores, que contribuem com propriedades para a variável final: humidade <HUMIDITY>; visibilidade <VISIBILITY>; vento <WIND>; pressão <PRESSURE>; etc.

Existem casos mais específicos de propriedades de contexto: i) Matiko et al. criaram uma propriedade oscilatória <MATIKO> que é calculada com base no <MIN> e <MAX> do sinal (EEG), e que informa como é que o <POWER> do sinal está relacionado com as ativações e inativações de regiões do cérebro (Matiko et al., 2014); e ii) Castellano et al. extraiu o <ACC> com base na diferença entre *frames* de {VIDEO}, utilizando o {EYESWEB} (Castellano et al., 2007), fazendo com que o *token* ACC seja simultaneamente uma variável recolhida do contexto (i.e. (ACC)) e uma propriedade específica de contexto (i.e. <ACC>) dependentemente da forma como é obtida.

2.2 Estudo do bem-estar em trabalhadores de escritório

O bem-estar nos trabalhadores de escritório é um tema importante para as organizações. Um trabalhador que não se sente bem no seu local de trabalho, produz menos, e pode influenciar negativamente o desempenho da organização onde trabalha (Bruntwood, 2021). Nesta secção faz-se uma revisão da literatura das investigações relacionadas com os ambientes de trabalho com foco nos trabalhadores de escritório. Analisam-se alguns projetos que estudam as emoções no contexto laboral e, em concreto, o reconhecimento automático de emoções, com ênfase no bem-estar.

Existe cada vez mais consciência empresarial para as vantagens em zelar pelo bem-estar dos colaboradores (MAPIQ, 2021). Um ambiente de trabalho perigoso e tóxico, dissuade os trabalhadores de inovar e prejudica a reputação das empresas (Croman, 2018). Os empregadores devem promover o bem-estar das pessoas. Manter os trabalhadores satisfeitos, é sinónimo de maior produtividade e inovação no trabalho (Croman, 2018). Algumas empresas promovem o bem-estar oferecendo melhores instalações aos seus trabalhadores, disponibilizando alimentos mais saudáveis, patrocinando melhores serviços de saúde, etc. Outras, vão mais além, e implementam escritórios inteligentes para permitir que os funcionários possam ter maior controlo sobre o ambiente em que trabalham e, assim, consigam adequar o local de trabalho às suas necessidades individuais (The Economist, 2021). Um escritório inteligente é um espaço dotado de tecnologia que dá às pessoas controlo sobre o ambiente, conectando-as ao edifício. Estes edifícios inteligentes, aumentam o bem-estar dos trabalhadores, e melhoram significativamente o desempenho dos negócios (MAPIQ, 2021). Em futuros empregos, as pessoas darão preferência às

organizações que promovam o seu bem-estar (Ginger, 2021).

Este projeto, visa criar um *dataset* final com dados recolhidos de contexto. Estes dados serão utilizados no futuro, para testar o desempenho de algoritmos de classificação na deteção do bem-estar em trabalhadores de escritório. Com base nesta informação automática, poderá vir a ser possível criar sistemas que melhorem o conceito de escritório inteligente. Sistemas autónomos que se reconfiguram para adaptar o ambiente de trabalho das pessoas com o objetivo de promover o seu bem-estar e, conseqüentemente, aumentar a produtividade e a criatividade. A organização do conteúdo desta secção está orientada à proximidade das investigações analisadas, à tecnologia. Inicialmente resumem-se algumas investigações relacionadas com o estudo emocional no contexto laboral sem atender à sua natureza tecnológica. Posteriormente analisam-se alguns trabalhos relacionados com o reconhecimento automático de emoções, com ênfase no estudo do bem-estar, em trabalhadores de escritório.

2.2.1 Estudo das emoções no contexto laboral

Um ambiente de trabalho mais amigável e mais apelativo, traduz-se num espaço de trabalho menos stressante e mais produtivo. Os trabalhadores de escritório, precisam de calma e conforto no seu ambiente de trabalho, para darem o melhor do seu rendimento (Kohll, 2021). Nesta secção apresentam-se algumas investigações relacionadas com as diversas áreas do contexto laboral, em que as emoções são estudadas.

A literatura reporta vários trabalhos de investigação em várias áreas relacionadas com a atividade laboral: i) satisfação no trabalho (e.g. Rothmann et al. estudaram a relação entre a satisfação no trabalho, stress ocupacional, *burnout* e o bem-estar (Rothmann, 2008)); ii) impacto do trabalho na vida pessoal e familiar (e.g. Fritz et al. estudaram o impacto do desligar do trabalho, na felicidade, saúde e produtividade (Fritz et al., 2010)); iii) absentismo no trabalho (e.g. Duijts et al. e Michie et al. identificam alguns preditores que levam ao absentismo por questões de saúde de ordem psicológica (Duijts et al., 2007; Michie and Williams, 2003)); iv) sobrecarga e pressão laboral (e.g. Holden et al. estudaram a quebra no desempenho profissional causada pelo excesso de horas de trabalho das pessoas (Holden et al., 2010); Züger et al. estudaram a frustração causada pelas interrupções no trabalho (Züger et al., 2018)); v) *burnout* (e.g. Lim et al. (Lim et al., 2010), Truchot et al. (Truchot, 2012), e Sonnenschein (Sonnenschein, 2007)); etc.

A literatura defende que as emoções no local de trabalho têm impacto direto na saúde das pessoas (Zapf, 2002). O trabalho em ambiente de escritório, quando é mentalmente desafiante, provoca stress nos colaboradores (Nkurikiyeyezu, Shoji, Yokokubo and Lopez, 2019). O excesso de stress pode diminuir a produtividade e impactar negativamente o bem-estar e a saúde das pessoas (Umematsu et al., 2021). Apesar de existirem várias investigações que estudam as emoções no contexto laboral, é no stress que se

concentra o maior número de investigações (e.g. (Alberdi et al., 2016), (Sanchez et al., 2017), (Zapf, 1993), (Wijnen et al., 2020)). No entanto, também existem trabalhos que investigam outras emoções no contexto do trabalho: i) Taris et al. estudaram o impacto da felicidade na produtividade (Taris and Schreurs, 2009); ii) Zaman et al. e Taylor et al. estudaram a utilização do email no trabalho e o impacto da sua utilização na saúde, stress, e produtividade (Zaman et al., 2019; Taylor et al., 2008); iii) Arnetz et al. estudaram a relação entre o reconhecimento dos empregadores e sintomas psicossomáticos dos funcionários (Arnetz and Wiholm, 1997); etc.

O bem-estar dos trabalhadores é essencial para o sucesso das organizações a longo prazo. Estudos mostram uma ligação direta entre os níveis de produtividade, saúde em geral e o bem-estar dos trabalhadores (ILO, 2021). A literatura relata vários trabalhos relacionados com o estudo do bem-estar no local de trabalho: i) Duyan et al. e Agarwal et al. investigaram a correlação entre a qualidade de vida no local de trabalho e o bem-estar (Duyan, 2013; Agarwal et al., 2019)); ii) Boyd et al. estudaram o bem-estar mental tendo por base a sobrecarga de tarefas, prazos apertados, falta de apoio profissional, e ambiente hostil no trabalho, enquanto fatores causadores de stress, depressão e ansiedade (Boyd et al., 2008); iii) Robertson et al. estudaram o bem-estar, produtividade e felicidade no trabalho (Robertson and Cooper, 2011); iv) Soh et al. investigaram a relação do bem-estar com a personalidade e a perceção do suporte obtido por parte da organização onde se trabalha (Soh et al., 2016); v) Jackson et al. estudaram o bem estar relacionado com o trabalho, tentando relacionar o *burnout*, saúde e tarefas profissionais (Jackson and Rothmann, 2005); vi) Klatt et al. estudaram o bem-estar no local de trabalho relacionado com os efeitos do stress crónico (Klatt et al., 2016); etc.

Existem também investigações que focam a sua atenção no estudo das emoções, especificamente nos trabalhadores de escritório: i) Fostervold et al. estudaram o impacto da luz direta e indireta na saúde, bem-estar, e desempenho cognitivo (Fostervold and Nersveen, 2008); ii) Vimalanathan et al. estudaram o impacto do ambiente dos escritórios no desempenho, produtividade, saúde e bem-estar (Vimalanathan and Babu, 2014); iii) Anderson et al. estudaram a relação do bem-estar afetivo e o teletrabalho (Anderson et al., 2015); iv) Thenmozhi et al. avaliaram o stress administrativo em trabalhadores de escritório (Thenmozhi et al., 2014); v) Bridger et al., na sua experiência, concluíram que a frustração pelo não cumprimento de objetivos foi o fator que mais contribuiu negativamente para o bem-estar mental dos trabalhadores de escritório (Bridger and Brasher, 2011); etc.

2.2.2 Reconhecimento automático de emoções no contexto laboral

O trabalho de escritório pode ser exigente e stressante. Os trabalhadores têm múltiplas tarefas a executar num curto espaço de tempo, muitas são complexas e exigem um ele-

vado grau de responsabilidade. Nesta secção faz-se o levantamento do estado da arte do reconhecimento automático de emoções, em trabalhadores de escritório, com foco no bem-estar.

São vários os investigadores que recorrem à componente tecnológica do escritório para estudarem emoções dos trabalhadores. Zaman et al. investigaram a influência dos emails no stress e na produtividade dos trabalhadores de escritório (Zaman et al., 2019). Koldijk et al. recolheram *logs* de computadores, expressões faciais, postura corporal e dados fisiológicos, com o objetivo de treinar modelos para inferir estados mentais relacionados com o stress (Koldijk et al., 2018). Brouwer et al. fizeram um estudo para monitorização do estado mental dos trabalhadores de escritório (o projeto envolvia a recolha do (HR) e expressões faciais em ambiente real, e o preenchimento do *valence-arousal affect grid* (Russell et al., 1989) para *ground-truth*) (Brouwer et al., 2018). Setz et al. criaram um sistema para detetar o stress através do (EDA). O objetivo do projeto era distinguir o stress, da carga cognitiva normal do ambiente do escritório. Os investigadores criaram um ambiente em laboratório para simular um escritório em que os participantes desempenhavam tarefas administrativas. Em simultâneo, os participantes eram provocados com stressores para gerar stress mental e stress psicossocial (Setz et al., 2010). Nkurikiyeyezu et al. alegaram que os vários sistemas já propostos na literatura para reconhecimento do stress não seriam implementáveis no mundo real. Defenderam que esses sistemas teriam um problema de generalização porque apenas conseguiriam inferir o stress em utilizadores conhecidos. Para resolver o problema, defendem que devem ser utilizados mecanismos mais genéricos (Nkurikiyeyezu, Yokokubo and Lopez, 2019). Noutro trabalho, os mesmos autores investigaram a possibilidade de distinguir as variações no (HRV) provocadas pelo desconforto termal, das provocadas pelo stress relacionado com o trabalho no escritório (Nkurikiyeyezu, Shoji, Yokokubo and Lopez, 2019).

O investimento feito pelas organizações no bem-estar e na saúde dos colaboradores, é um investimento com retorno (International Labour Organization, 2021). A promoção do bem-estar dos trabalhadores, aumenta a sua produtividade e promove a sua saúde (ILO, 2021; IJsselsteijn et al., 2006). São vários os projetos relacionados com o reconhecimento automático de emoções que estudam o bem-estar.

Zhao et al. pretendiam avaliar o impacto de um ambiente controlado na capacidade de concentração dos trabalhadores após um evento stressante. Para isso, criaram uma sala protótipo em laboratório para conseguirem controlar a luz, som e vídeo. Na sua experiência utilizaram vários instrumentos para recolher dados de contexto: i) cinta de peito *Zephyr* {BIOHARNESS} três para recolher dados como o (HR), (ECG), (RR), (ACC), etc.; ii) a pulseira *Empatica E4* para recolher (EDA), (ST), (BVP), etc.; e iii) a aplicação *Intraface* para recolha dados de vários pontos faciais (e.g. nariz, olhos, lábios). O objetivo do estudo era conseguir promover a produtividade e o bem-estar no trabalho (Zhao et al.,

2017).

Alonso et al. defenderam que as novas tecnologias que têm sido adicionadas ao ambiente de escritório não contribuíram para reduzir o stress nos trabalhadores, porque se focam essencialmente na recolha de dados fisiológicos. Alegam que os sistemas não atendem ao contexto mental, espiritual e ao bem-estar emocional de cada pessoa e, por esse motivo, defendem ser necessário criar novas soluções que previnam o stress e promovam o bem-estar nos trabalhadores de escritório. O objetivo dos autores era detetar o stress recorrendo a instrumentos já utilizados pelos participantes. Inspirados nos gestos espontâneos feitos pelos trabalhadores com a caneta quando lhes é dada uma tarefa mentalmente exigente, desenvolveram um dispositivo para ajudar os trabalhadores a relaxar e voltar ao estado de equilíbrio após um momento de stress (Alonso et al., 2008).

Saha et al. estudaram a produtividade e o bem-estar no trabalho. Com esse objetivo, recolheram dados de histórico das redes sociais para criação do contexto social individual dos participantes da sua experiência. Posteriormente, relacionaram esses dados com informação demográfica, dados acerca da personalidade dos participantes e informação acerca do bem-estar. Os investigadores contribuíram com uma *framework* que permite a recolha de dados sociais de forma passiva (Saha et al., 2019).

Chandraprabha et al. criaram um sistema que ajuda os responsáveis das organizações a tomarem decisões para promover o nível de concentração e a produtividade dos trabalhadores. O sistema mantém o empregador a par do bem-estar dos seus trabalhadores, recolhendo a expressão facial através das *webcam* dos computadores, e inferindo as seguintes emoções em tempo real: felicidade, tristeza, surpresa, medo, desgosto e fúria (Chandraprabha et al., 2021).

Kimani et al. defendem que os temas da produtividade e do bem-estar têm sido pouco estudados, e apresentam um sistema de conversação baseado em *chat*, para ajudar os trabalhadores no planeamento e execução das suas tarefas diárias. Os investigadores pretendiam avaliar o impacto que existiria no número de tarefas concluídas diariamente, se os trabalhadores: i) tivessem apoio no agendamento e priorização de tarefas; ii) fossem lembrados das pausas para descanso; e iii) se a utilização das redes sociais fosse monitorizada (Kimani et al., 2019).

Koldijk desenvolveu uma ferramenta para ajudar as pessoas a lidar com os determinantes negativos do seu bem-estar. Recolheram dados relacionados com: i) a utilização do computador através da atividade do teclado e rato; ii) a conversação através da informação do email e do telefone; iii) dados fisiológicos (cf. (RESP), (ECG), (SCR) e (BVP)); e iii) outra informação como localização (GPS), nível de atividade (acelerómetros), e outra informação áudio e vídeo. O investigador concluiu que a sua investigação poderá fornecer pistas para melhorar o bem-estar no trabalho através de sensorização não obstrutiva (Koldijk, 2012).

Umematsu et al. recolheram o (HR), (EDA), (ST) e (ACC), com o objetivo de estudar

formas de melhorar o bem-estar dos trabalhadores de escritório, prevenindo os níveis de stress das pessoas (Umematsu et al., 2021).

Robles-Granda et al. estudaram a predição do bem-estar, considerando os desafios associados ao processamento de dados incompletos e com ruído. Propõem um análise preditiva do bem-estar tendo por base o comportamento físico, psicológico, estado e características psicológicas, e desempenho no trabalho. Os autores recolheram dados relativos à interação social a partir do telefone e das redes sociais, rotinas relacionadas com o trabalho, e saúde e bem-estar tanto no escritório como em casa. Criaram uma aplicação para a recolha de dados, e utilizaram *beacons* para identificar se as pessoas estavam em casa ou no trabalho. Os estados psicológicos das pessoas foram obtidos a partir dos *posts* feitos em redes sociais (cf. *Facebook* e *LinkedIn*), em conjunto com a informação recolhida com a aplicação de telemóvel. Os investigadores utilizaram a pulseira *Garmin Vivosmart 3* para recolher: (HR), contagem de passos, número de pisos que subiu, calorias, (PHYSI), (SLEEP), duração da luz, (STRESS), etc. (Robles-Granda et al., 2021).

2.3 Análise

Neste capítulo fez-se o levantamento do estado da arte relacionado com: i) a deteção automática de emoções na generalidade e em particular com ii) o estudo aplicado no mundo laboral, focando as emoções e o bem-estar dos trabalhadores de escritório.

A revisão bibliográfica, aqui apresentada, do reconhecimento automático de emoções, é um resumo do Relatório Técnico (RT) criado em (Carvalho et al., 2018). Por questões de economia de espaço, este resumo apresenta apenas alguns dos *tokens* criados no RT. A forma como o relatório foi organizado, permite facilmente incluir novo conteúdo à medida que uma nova investigação é analisada e o seu conteúdo é catalogado. O RT é um complemento ao estado da arte descrito neste documento, está disponível para consulta no Repositório Institucional da UFP, e constitui a contribuição C01 desta tese.

O levantamento bibliográfico efetuado com o RT, serviu de suporte à preparação da experiência conduzida nesta tese no âmbito da área de AC. O objetivo da experiência focou-se na recolha de um *dataset* multimodal em ambiente real, para ser utilizado na deteção automática do bem-estar de trabalhadores de escritório. O conhecimento obtido tendo por base a revisão da literatura efetuada, sustentou a seleção das variáveis a recolher, instrumentos & sensores a usar, técnicas possíveis de aplicar na fase de pré-processamento, e propriedades a extrair a partir dos dados pré-processados, com o objetivo de enriquecer o *dataset* final.

A ambiguidade humana torna muito difícil a deteção de emoções apenas analisando os dados de uma única modalidade como sugerem muitas investigações (Basu et al., 2016; Zhang et al., 2016). Os estudos que têm demonstrado melhor precisão no reconhecimento emocional, diversificam métodos e técnicas, mas também diversificam os tipos de dados

de contexto recolhidos (Bakhtiyari and Husain, 2014). É neste sentido que se acredita que a diversificação com dados de várias modalidades, pode contribuir para uma maior precisão dos algoritmos no reconhecimento automático de emoções.

São várias as investigações que se focam na recolha de dados da expressão facial, oral e postura corporal para inferir emoções. No entanto, os problemas na generalização das soluções resultantes destas investigações, causados pela dependência da cultura, género e idade (Rani and Sarkar, 2006), a erraticidade provocada pelo possível mascaramento (e.g. fingimento) do utilizador (Chen et al., 1998), e a ambiguidade na interpretação de emoções a partir da forma como as pessoas se exprimem não-verbalmente (e.g. choro de alegria ou sorriso de vergonha) (Ekman, 1989), tornam estes sinais, *per se*, menos fiáveis para *input* de um sistema automático de deteção emocional.

As variáveis escolhidas habitualmente pelos investigadores para deteção emocional, são as que foram demonstrando alguma taxa de sucesso na correlação com as emoções (Nasoz et al., 2004). Por exemplo, existem sensores capazes de medir de forma objetiva sinais do contexto fisiológico. Portanto, são muitos os investigadores que recolhem dados desta modalidade. No entanto, não é possível mapear os sinais recolhidos diretamente em emoções (Kim and André, 2008). Outra justificação para que tantos investigadores recolham dados deste contexto, será o facto de serem menos suscetíveis de manipulação pelas pessoas (Van Der Vloed and Berentsen, 2009). Contudo, existe o problema da possível obstrutividade dos instrumentos & sensores para efetuar a recolha, e a necessidade de tratar o ruído dos sinais (e.g. ruído da atividade muscular ou ocular no (EEG)) (Soleymani et al., 2013). A atividade glandular parecem também oferecer dados muito interessantes para a inferência emocional, atendendo à relação que a literatura estabelece entre a (SEROT) e o bem-estar (Fonte, 2015), o (CORT) e o stress (Kirschbaum et al., 1993) e a (MELAT) com o humor (Curcio et al., 2016). No entanto, será a intrusividade dos métodos de recolha que desencorajam a utilização destes dados.

A solidez da relação do suporte social com o bem-estar e a saúde em geral, tornam interessantes os sinais recolhidos deste contexto (Sarason et al., 1985; Seligman, 2011; Hohaus and Berah, 1996). No entanto, não existe uma forma transversalmente aceite, e de recolha automática, para medir o suporte social (Pais-Ribeiro, 1999; Heitzmann and Kaplan, 1988), tornando difícil a seleção de variáveis de contexto a recolher desta modalidade. Contudo, a relevância dos dados provenientes do contexto social e psicológico, e a expectativa que existe por parte do autor no peso que podem representar para um sistema de deteção automática de emoções, motiva a diversificação do número de variáveis recolhidas deste contexto, com o objetivo de aumentar as hipóteses de descobrir correlações com o bem-estar e melhorar os resultados já obtidos noutras investigações.

Num futuro próximo, as emoções poderão vir a ser reconhecidas por muitos dispositivos (El Kaliouby, 2005). O projeto apresentado nesta tese pretende contribuir para esta evolução, criando um *dataset* com dados multimodais, recolhidos em ambiente real

de escritório, com o objetivo de poderem ser utilizados por um sistema de detecção automática de emoções. Serão beneficiários desta tecnologia os próprios trabalhadores porque poderão ser desenvolvidos mecanismos automatizando a melhoria do seu bem-estar, mas também as organizações que poderão beneficiar com um aumento de produtividade dos seus trabalhadores.

A literatura contém vários trabalhos realizados na detecção automática de emoções. São também bastantes os trabalhos de investigação que estudam as emoções no contexto laboral. Contudo, são poucas as que investigam o reconhecimento automático de emoções no ambiente de escritório, e menos ainda as que estudam o bem-estar como objetivo principal. A maioria das investigações que trabalha a detecção automática de emoções em ambiente de escritório, estudam o bem-estar sob a perspetiva do benefício para as entidades empregadoras (e.g. impacto na produtividade e criatividade). Não obstante ser importante estudar o bem-estar dos trabalhadores na perspetiva do benefício das organizações, neste trabalho pretende-se estudar o bem-estar na perspetiva das pessoas, olhando as vantagens para a organização como uma consequência indireta, e não como o objetivo principal.

Nesta investigação recolheu-se dados em ambiente real. A maioria das investigações recolhem dados em ambiente controlado. Nestes casos, os investigadores implementam medidas para promover o sucesso do processo de recolha (e.g. restringir a locomoção dos participantes para garantir a correta colocação dos elétrodos durante todo o tempo). Apesar de se conseguir melhor qualidade nos dados recolhidos, esses dados poderão ser contaminados pela influencia direta dos investigadores no controlo do processo. Além disso, os participantes da experiência não estão no seu ambiente real de trabalho: i) existirá uma maior predisposição mental para o experiência; ii) os participantes estarão mais atentos às suas ações e reações; e iii) será difícil reproduzir em laboratório o ritmo, responsabilidade e stress, resultante das múltiplas tarefas e interações sociais existentes em simultâneo num escritório.

O objetivo deste projeto focou-se na criação de um *dataset* multimodal cujos dados recolhidos espelhassem de forma alargada o contexto e comportamento dos participantes em ambiente real. São poucos os investigadores que recolhem dados do contexto social e psicológico. Não desconsiderando a importância dos dados fisiológicos, os dados sociais, culturais e religiosos, poderão contribuir de forma muito positiva para a precisão de um sistema de detecção emocional. Ao nível cultural, a aquisição de novos conhecimentos ou competências poderá desencadear emoções positivas (e.g. satisfação ou orgulho) ou negativas (e.g. stress ou frustração). Por outro lado, a falta de novos conhecimentos poderá conduzir ao sentimento de ansiedade por falta de segurança num determinado tema, preocupação ou susto perante a hipótese de despedimento por falta de um título académico, ou chatice pela consciência que deverá continuar os seus estudos como forma de não perder o emprego. As pessoas mais envolvidas em iniciativas religiosas sentirão emoções mais positivas (e.g. bem-estar) por estarem a cumprir com as leis da sua religião

(Fredrickson, 2002). No entanto convivem com temas como a morte podendo conduzir ao sentimento de emoções negativas (Kim-Prieto and Diener, 2009). Portanto, a avaliação do bem-estar de uma pessoa não pode deixar de parte informação importante como esta, devendo correlacionar os dados de contexto social, cultural e religioso, com os recolhidos a partir dos sensores fisiológicos que, apesar de oferecerem dados objetivamente mensuráveis, espelham apenas uma parte da realidade do contexto do utilizador.

A experiência apresentada nesta tese diferencia-se das demais pelos motivos atrás descritos, e contribui com um *dataset* com dados do contexto psicossocial e fisiológico de 20 participantes. A experiência é descrita em detalhe no capítulo 3.

Capítulo 3

EXPERIÊNCIA DE RECOLHA EM CONTEXTO LABORAL

As organizações são cada vez mais dinâmicas e exigentes com os seus trabalhadores. O stress relacionado com o aglomerar de tarefas e cumprimento de prazos faz com que, muitas vezes, o local de trabalho seja um ambiente agressivo onde se experienciam emoções negativas. No entanto, a realização pessoal resultante do reconhecimento por outros, o bom espírito de equipa, e o cumprimento de objetivos, pode ajudar no contra balanceamento emocional com emoções positivas. Neste capítulo apresenta-se a experiência conduzida em contexto laboral. O *dataset raw* multimodal apresentado nesta secção, foi recolhido em ambiente real, engloba várias variáveis de contexto, e constitui a contribuição principal deste capítulo.

As emoções são respostas a eventos de contexto (Arnold, 1970). Perante o mesmo estímulo emocional, as pessoas podem reagir de forma diferente, porque o impacto desse estímulo depende da experiência passada, personalidade, profissão, cultura, etc. (Bos, 2010). A mesma pessoa pode, também ela, reagir de forma diferente em momentos distintos, porque o seu estado de espírito e contexto também mudam no tempo (Raudonis, 2013). Assim, considerando a dificuldade em identificar diretamente emoções, e o carácter singular de cada pessoa (Murad and Malkawi, 2012), a alternativa será tentar descobrir as melhores variáveis de contexto para classificar emoções. Acredita-se que, a utilização de dados de contexto de várias modalidades, pode melhorar a precisão na deteção automática do bem-estar.

Nesta experiência, recolheu-se dados de origem multimodal em ambiente real, durante vários dias, de 20 participantes trabalhadores de escritório de 7 organizações distintas. Recolheram-se 46 variáveis de contexto, através de 11 instrumentos & sensores. Pretendia-se utilizar poucos instrumentos para manter o mais possível a normalidade do contexto pessoal e profissional, e seleccionar apenas os instrumentos que apresentassem um baixo nível de obstrutividade. As possíveis restrições ao movimento e desconforto na utilização de sensores pelos participantes, poderiam interferir nos dados a recolher (e.g.

suor) (Tran et al., 2007; Barea et al., 2011; Wu et al., 2008). De preferência, preferia-se utilizar instrumentos já usados pelos participantes no seu dia-a-dia (e.g. computador e *smartphone*), para diminuir ao nível de interferência no ambiente e ritmo habitual do escritório. No entanto, e apesar da vontade em não utilizar instrumentos obstrutivos, optou-se por recolher (EDA) através de elétrodos colados na palma da mão dos participantes utilizando o sensor EDAMove {EDAMOVE} (movisens GmbH, 2018d). As evidências presentes na literatura que relacionam o (EDA) com as emoções são tão fortes que tornavam este sinal essencial para a experiência. As alternativas eram limitadas e inacessíveis por causa do preço. Assim, dada a importância do (EDA) na deteção emocional, e o facto de se ter conseguido uma parceria com a Movisens (Movisens, 2020) para empréstimo de sensores {EDAMOVE}, decidiu-se recolher este sinal afixando elétrodos na palma da mão dos participantes.

O autor tem uma ligação próxima com vários escritórios do mundo empresarial. O objetivo final desta experiência era contribuir com um *dataset* que pudesse vir a ser usado para classificar automaticamente o bem-estar dos seus trabalhadores. Com base nesta classificação, poderá ser possível que os sistemas possam promover ações de reconfiguração autónoma do ambiente do escritório, com o objetivo de melhorar o bem-estar. Neste contexto, definiram-se os considerandos base que deram suporte ao processo de definição do âmbito e das regras para constituição da amostra da experiência. A vontade de recolher dados que espelhassem o contexto real dos trabalhadores de escritório, significaria recolher dados de vários participantes, pertencentes a várias organizações, com o objetivo de diversificar o conteúdo das variáveis recolhidas. A pretensão de recolher um grande *dataset* multimodal, significaria utilizar múltiplos instrumentos & sensores, durante vários dias, por participante. Atendendo à disponibilidade de apenas dois sensores {EDAMOVE} a processo de recolha seria longo e necessariamente distribuído no tempo, podendo existir inclusive várias iterações de recolha na mesma organização. Além do {EDAMOVE} poderia haver a necessidade de adquirir outros instrumentos & sensores para recolher dados de contexto de outras modalidades. Além disso, considerando que os participantes usariam vários sensores incluindo um com elétrodos, e durante vários dias, sentia-se a necessidade de recompensar os participantes pela sua participação. Resumindo, o número de participantes e organizações a envolver na experiência, teria que atender aos fatores tempo, esforço e investimento, sem prejuízo da qualidade do *dataset* resultante do processo de recolha, por ser este o objetivo principal da experiência.

Neste contexto, definiu-se o âmbito e perfil das organizações e participantes a convidar para a experiência. O convite a cada organização foi feito por escrito (ver Anexo 7.10) e existiu uma ou mais reuniões para esclarecimento de dúvidas. A aceitação por parte das organizações foi formalizada num documento com conteúdo sugerido pelo investigador (ver Anexo 7.10). A Tabela 3.1 apresenta o perfil das organizações que foram convidadas a participar na experiência.

Perfil das organizações convidadas	
Localização	Norte de Portugal.
Atividade	Organizações do setor comercial ou industrial, cuja atividade envolvesse uma forte componente administrativa.
Dimensão	Foram convidadas a participar na experiência organizações com, no mínimo, duas pessoas alocadas a trabalhar no escritório, e que interagissem com pessoas internas e externas.
Critérios para seleção e/ou exclusão de organizações	
Regras	Para poderem participar na experiência, as organizações tinham que: <ul style="list-style-type: none"> i permitir que os trabalhadores que se autopropunham a participar na experiência, pudessem transportar sensores durante os dias definidos para a recolha de dados; ii aceitar a instalação de aplicações nos computadores e <i>smartphones</i> da organização, para a recolha de dados estatísticos relacionados com a utilização, interação social, e para aplicação de questionários; ¹ e iii autorizar que os colaboradores pudessem dedicar alguns minutos diários do tempo de trabalho, na resposta a questionários breves.

Tabela 3.1: Âmbito e definição da amostra (perfil das organizações a convidar)

A Tabela 3.2 apresenta o perfil dos candidatos a participante da experiência. O convite foi feito, num primeiro momento, através da disponibilização pública, não endereçada, de uma brochura a apresentar o projeto e a experiência (ver Anexo 7.11) e, posteriormente, através de um evento público onde o projeto foi apresentado verbalmente. Ninguém foi forçado a presenciar o evento de apresentação pública, nem convidado direta ou pessoalmente para estar presente. Durante essa apresentação, foi explicado o modelo da auto-proposição para angariação de participantes (ver secção 1.6) e retiradas todas as dúvidas colocadas pelos presentes.

Estudo de coorte	
Tamanho da amostra	n = [20;50]
Perfil dos participantes convidados	
Género	Qualquer.
Idade	[18; 68] anos.
Outros	Foram aceites participantes por auto-proposição que cumpriam os critérios seguintes: <ul style="list-style-type: none"> i trabalhadores da organização há, pelo menos, 12 meses seguidos; ² ii pessoas que realizavam tarefas administrativas em ambiente de escritório; iii trabalhadores que desempenhavam tarefas que envolviam a interação com outras pessoas internas e externas à organização; iv utilizadores com computador ³ e <i>smartphone</i> Android individual. ⁴
Critérios de exclusão	

¹Apenas foram recolhidos dados estatísticos e nunca conteúdo específico (i.e. texto).

²Necessário para existir histórico recente de gozo de férias.

³Necessário para instalar aplicações para recolher dados de utilização.

⁴Por questões de compatibilidade e restrições de acesso ao conteúdo do SO.

Funções motoras	Candidatos sem independência física, porque seriam recolhidos dados relacionados com a atividade física (e.g. quantidade de passos percorridos).
Saúde	Candidatos com doenças da memória e funções cognitivas (e.g. Alzheimer, Parkinson).
Eventos marcantes recentes	Existência de eventos marcantes ou complicações da vida há menos de 6 meses (e.g. divórcio). A existência destes eventos, poderia enviesar o <i>dataset</i> recolhido.
Regras	Para poderem participar na experiência, os participantes tinham que aceitar: <ul style="list-style-type: none"> i assinar uma declaração de consentimento escrito a demonstrar a sua participação voluntária, livre, e informada na experiência; ii a utilização de instrumentos & sensores para a recolha de dados fisiológicos (e.g. elétrodos, pulseiras); iii a monitorização estatística de dados relacionados com o seu trabalho, relação laboral com a organização, e atividade social (emails e SMS), através da instalação de aplicações no computador e <i>smartphone</i>; e iv responder a questionários criado pelo autor acerca das suas rotinas diárias, bem-estar, disposição e saúde em geral.

Tabela 3.2: Âmbito e definição da amostra (perfil dos participantes a admitir)

O ambiente real é o mais indicado para recolher dados de contexto. As pessoas abstraem-se da sua participação na experiência por estarem absorvidas nas suas tarefas e responsabilidades diárias. Este alheamento dos participantes, por um lado é bom porque permite a recolha dos dados realísticos que se pretendiam recolher do ambiente real mas, por outro, significa que os participantes deixam de estar atentos às questões logísticas da experiência (e.g. resposta aos questionários no momento previsto). Além deste problema, o acréscimo de tarefas decorrentes da própria experiência, pode contribuir para o enviesamento dos dados recolhidos por representarem mais tarefas a realizar (e.g. a necessidade de reafixar um elétrodo ou de responder a um questionário no meio de uma tarefa de trabalho, poderá aumentar os níveis de stress).

Portanto, a recolha de dados em ambiente real é um desafio que levanta várias questões ao processo de investigação. Era expectante que existissem problemas nos dados recolhidos, previa-se uma fase de pré-processamento importante para a qualidade dos dados do *dataset*. As preocupações existentes com a qualidade dos dados acabaram por se verificar legítimas. Nas visitas surpresa efetuadas a cada iteração de recolha, houve a necessidade de recolocar elétrodos do {EDAMOVE}, redimensionar a largura da pulseira MiBand {MIBAND} (TechFunology, 2016; XiaoMi, n.d.) para garantir a proximidade suficiente com a pele para a correta leitura do (PPG), apelar ao preenchimento atempado dos questionários, etc. Antevendo-se uma fase exigente de pré-processamento, ainda durante a fase de recolha, começou a planear-se a integração e análise da qualidade dos dados, que será apresentada no capítulo 4.

Na primeira parte deste capítulo descreve-se o processo de preparação da experiência. Nesta secção apresentam-se as variáveis de contexto recolhidas, os instrumentos & sensores selecionados e também os criados pelo autor, e a documentação criada para suporte

à experiência. Na segunda parte do capítulo é relatada a execução da experiência, descrevendo o processo de constituição da amostra e da recolha de dados. Termina-se com uma análise onde são identificadas as contribuições efetuadas neste capítulo.

3.1 Preparação

A fase de recolha da experiência apresentada neste capítulo, durou um pouco mais de dois meses. Recolheu-se dados de contexto de várias modalidades, de vários participantes, trabalhadores de escritório de várias organizações. Os constrangimentos impostos pelo número reduzido de sensores {EDAMOVE}, fazia prever um processo longo e complexo, que impunha um bom planeamento prévio para que nada falhasse durante a execução. Nesta secção relata-se a fase de preparação da experiência, que conduziu ao planeamento antecipado de toda a logística necessária à recolha de dados de contexto.

A preparação da experiência iniciou com a seleção das variáveis de contexto a recolher, e os instrumentos & sensores a utilizar nessa recolha. Após esta seleção, procedeu-se ao pedido de autorização à CE para realização da experiência. Obteve-se autorização em 2019-02-13 e, durante o tempo em que se esperava pela aprovação, foram sendo adiados outros trabalhos como: i) criação de instrumentos necessários para a recolha; ii) criação de formulários *backup* para recolher dados aquando da indisponibilidade do computador ou *smartphone*; iii) criação de outra documentação suporte ao processo de recolha; etc.

3.1.1 Seleção de variáveis de contexto

As investigações da área do AC utilizam variáveis de contexto de várias modalidades no reconhecimento emocional: i) expressão facial, oral, gestual e postura corporal; ii) variáveis de contexto fisiológico (e.g. pele, atividade cerebral, cardíaca, respiratória, muscular, ocular e glandular, etc.); e iii) variáveis de contexto social e psicológico (e.g. informação demográfica, psicossocial, rotinas diárias, interação social, etc.). Nesta secção são apresentadas as variáveis de contexto e propriedades de domínio, escolhidas para recolha nesta experiência.

Uma emoção é um constructo produzido de forma sistemática por excitação fisiológica, processos cognitivos, sentimentos subjetivos, comportamentos, etc. (Kim and André, 2008). Neste contexto, acredita-se que a correlação das variáveis do contexto social, cultural e religioso, com as fisiológicas, pode contribuir de forma positiva para a precisão de um sistema de deteção emocional. A literatura sugere que a recolha multimodal pode contribuir para o aumento da precisão dos algoritmos de classificação (Bakhtiyari and Husain, 2014). Tentar classificar emoções tendo por base apenas dados de uma modalidade, poderá ser algo muito redutor do significado subjetivo do que é ser um ser humano.

Além de se pretender recolher dados sociais e psicológicos, decidiu-se recolher também dados de contexto fisiológico. A ideia passava por incluir no *dataset* variáveis já amplamente relacionadas pela literatura com as emoções. Os dados *ground-truth* relativos ao bem-estar, foram recolhidos através de questionários informais desenvolvidos pelo autor, tendo por base a *Échelle de Mesure des Manifestations du Bien-Être Psychologique* {EMMBEP} (Monteiro et al., 2012) (ver secção 3.1.2). Cada variável de contexto escolhida, foi individualmente ponderada. O custo de intrusão na vida pessoal dos participantes foi pesado com a contribuição que significaria para o *dataset raw* final. Selecionaram-se apenas as variáveis que poderiam ser úteis a um sistema de deteção automática de emoções e, mais concretamente, ao reconhecimento do bem-estar em trabalhadores de escritório.

São vários os investigadores que recolheram um elevado número de variáveis de contexto nos suas experiências. Sano & Picard, por e.g., recolheram dados do sono, disposição, personalidade, stress, níveis de cansaço, interações com dispositivos eletrónicos e *smartphone*, chamadas efetuadas e recebidas, consumos de álcool e cafeína, etc. (Sano and Picard, 2013a); LikamWa et al. analisaram a utilização dos navegadores de internet, utilização de aplicações e localização no *smartphone*, utilizaram também informação de mensagens escritas, email e chamadas telefónicas (LiKamWa et al., 2013); e Sano & Eng recolheram uma grande e diversa quantidade de dados como a etnia, raça, anos de escola, personalidade, sono, saúde, disposição, felicidade, nível de alerta, nível de energia, ansiedade, utilização de dispositivos móveis, utilização de eletrodomésticos, aplicações utilizadas nos *smartphone*, localizações de GPS, chamadas e mensagens do *smartphone*, emails enviados e recebidos, interações face a face, etc. (Sano and Eng, 2016). Apesar da grande quantidade de variáveis recolhidas do contexto social e psicológico que se decidiu recolher, acredita-se que nem todas sejam necessárias podendo algumas ser descartadas aquando do pré-processamento (e.g. atributos redundantes).

Optou-se por não recolher dados relacionados com as expressões facial, oral e postura corporal, por causa das reservas levantadas pela literatura a este sinal (e.g. problemas de generalização, mascaramento social, ambiguidade na interpretação não-verbal). Apesar de muitos trabalhos utilizarem estes sinais, e apesar da vantagem de poderem ser recolhidos de forma não obstrutiva através de {VIDEO} ou {PICTURES}, o autor considera que os problemas levantados aos dados desta modalidade, poderiam por em causa os resultados obtidos na classificação com base nestes sinais.

Ao nível do contexto fisiológico, a relação sustentada pela literatura do (EDA) com as emoções, fazem desta variável um sinal essencial para a experiência. O (EDA) está associado ao SNS tornando impossível a sua manipulação por parte dos participantes (Poh et al., 2010; Sano and Picard, 2013b). Apesar de ser influenciável por fatores externos como a (TEMP) e a atividade física, considerou-se importante recolher esta variável de contexto, por causa da relevância que teria para *dataset raw* final pretendido. A Tabela 3.3 apresenta as variáveis de contexto fisiológico recolhidas nesta experiência.

Variável	Descrição
Pele	
(EDA)	A atividade eletrodermal (EDA) mede a condutância da pele (Healey and Picard, 2005; Sano and Eng, 2016; Boucsein, 2012). É uma variável interessante no processo de detecção emocional, porque é um indicador de excitação psicológica ou fisiológica (Murad and Malkawi, 2012). A condutância da pele reflete a inervação sudomotora e a atividade da glândula do suor que aumentam com a atividade SNS (Taylor et al., 2015; Poh et al., 2010).
(ST)	A temperatura da pele (ST) é utilizada no pré-processamento para tratamento do ruído do sinal (EDA).
Atividade cardíaca	
(HR)	A frequência cardíaca (HR) reflete atividade emocional (Mandryk and Atkins, 2007; Winton et al., 1984). Por e.g., o (HR) varia com o medo, fúria e susto (Vrana et al., 1986; Lisetti and Nasoz, 2004; Haag et al., 2004).

Tabela 3.3: Variáveis de contexto selecionadas (contexto fisiológico)

A expectativa existente na contribuição dos dados de contexto social e psicológico, no aumento da precisão dos algoritmos de reconhecimento automático de emoções, levou o autor à recolha de várias variáveis deste contexto. A relação defendida pela literatura entre o suporte social dado e recebido com o bem-estar, contribuíram para pretensão de incluir várias variáveis deste contexto no *dataset* com o qual se pretendia contribuir. Assim, decidiu-se por recolher variáveis relacionadas com informação demográfica e psicossocial. Decidiu-se também recolher dados relacionados com a interação social, dados que permitissem aferir acerca do reconhecimento social de outras pessoas pelo participante, e outra informação que evidenciasse atividades sociais ou comportamentos adquiridos resultantes da vida em sociedade.

As variáveis de contexto social e psicológico recolhidas, são apresentadas na Tabela 3.4. Para facilitar o enquadramento de recolha das várias variáveis, são apresentadas algumas propriedades de contexto possíveis de extrair de cada sinal. O objetivo é justificar a recolha de cada variável, apresentando alguns benefícios decorrentes dessa sua recolha.

Variável	Descrição
Domínio pessoal / demografia	
(AGE)	Idade.
(GENDER)	Género.
(ETHNICITY)	Etnia.
(CIVILSTATUS)	Estado civil.
(RELIGION)	Religião praticada.
(LIVING)	Pessoa ou pessoas com quem vive.
(ACADDG)	Níveis previstos no Quadro Nacional de Qualificações. (portaria 782/2009, de 23 de julho) ¹

¹Nível 1 (2º ciclo do ensino básico); nível 2 (3º ciclo do ensino básico); nível 3 (ensino secundário para ingresso no ensino superior); nível 4 (ensino secundário com estágio profissional de, no mínimo, 6 meses); nível 5 (qualificação de nível pós-secundário não superior); nível 6 (licenciatura); nível 7 (mestrado); e nível 8 (doutoramento).

(DEPENDENTS) **Informação dos dependentes (DEPENDENTS) que inspirem atenção, cuidados ou outras preocupações diárias de carácter permanente.** ¹ O símbolo # significa o número do dependente.

(DEPENDENT#) <AGE>	Idade.
(DEPENDENT#) <KINSHIP>	Grau de parentesco.
(DEPENDENT#) <SAMEHOUSE>	Vive na mesma habitação de forma permanente.
(DEPENDENT#) <MONEYDEPENDENCY>	Dependência financeira.
(DEPENDENT#) <DAILYTASKSDEPENDENCY>	Dependência nas tarefas do dia-a-dia.
(DEPENDENT#) <SPECIALNEEDS>	Necessidades especiais.
(DEPENDENTS) <COUNT>	Número total de dependentes.

(HOME) **Informação relacionada com o local onde vive (HOME), distância e acessibilidade ao local de trabalho (HOME-WORK) (HOME-WORK).**

(HOME) <TYPE>	Tipo de habitação.
(HOME) <ENVIRONMENT>	Ambiente habitacional.
(HOME) <OWNER>	Titularidade da habitação.
(HOME-WORK) <DISTANCE>	Distância de casa ao trabalho (número de quilómetros).
(HOME-WORK) <DURATION>	Duração de casa ao trabalho (número de minutos).
(HOME-WORK) <TRANSPORTATION>	Meio de transporte utilizado para chegar ao trabalho.
(HOME) <HOUSEMOVE<LASTTIME>>	Última vez que trocou de casa (número de anos).
(HOME) <QUALITY>	Qualidade da habitação atual (conforto proporcionado).

(NEIGHBORS) **Informação relacionada com vizinhos (NEIGHBORS).**

(NEIGHBORS) <FRIENDS<COUNT>>	Existência de vizinhos amigos.
(NEIGHBORS) <NEUTRAL<COUNT>>	Existência de vizinhos neutros.
(NEIGHBORS) <PROBLEMATIC<COUNT>>	Existência de vizinhos problemáticos (e.g. ruidosos, conflituosos).

(LOANS) **Informação relacionada com responsabilidades bancárias ou financeiras.**

(LOANS) <HOUSEMORTGAGE>	Existência de crédito a habitação.
(LOANS) <CARLOAN>	Existência de crédito automóvel.
(LOANS) <OTHERLOANS>	Existência de outros créditos

(WORK ABSENCES) **Informação relacionada com faltas ao trabalho. (WORKABSENCES)² O símbolo # significa o número da falta.**

(WORKABSENCES) <COUNT, LASTTIME, REGULARITY>	Número de faltas, última falta e regularidade de faltas (média de dias entre faltas).
(WORKABSENCE#) <TIMING, PLANNING, TYPE>	Data de cada falta, informação se foi uma falta planeada com antecedência (mínimo um dia), e o tipo de falta (e.g. doença própria, greve).

¹Primeiros quatro dependentes que representam maior grau de dependência (e.g. filho, pai acamado).

²Dados relativos ao último mês.

(WORKDELAYS) Informação relacionada com atrasos ao trabalho (WORKDELAYS).² O símbolo # significa o número do atraso.

(WORKDELAYS) <COUNT, LASTTIME, REGULARITY>	Número de atrasos, último atraso e regularidade de atrasos (média de dias entre atrasos).
(WORKDELAYS) <TIMMING, PLANNING, PERIODS, WHY>	Data de cada atraso, informação se foi um atraso planeado com antecedência (mínimo duas horas), período de trabalho (e.g. manhã, tarde), e o motivo do atraso (e.g. dificuldades em acordar, trânsito).

(WORKLATES) Informação relacionada com trabalho horas-extra trabalhadas (WORKLATES).² O símbolo # significa o número da hora-extra trabalhada.

(WORKLATES) <COUNT, LASTTIME, REGULARITY>	Número de horas extra trabalhadas, última vez que fez horas-extra e regularidade com que faz horas extra (média entre dias com horas-extra).
(WORKLATE#) <TIMMING, PLANNING>	Data de cada hora-extra, informação se a hora-extra de trabalho foi prevista antecipadamente (mínimo um dia).

(VACATIONS) Informação relacionada com férias gozadas nos últimos 12 meses (VACATIONS). O símbolo # significa o número do dia de gozo de férias.

(VACATIONS) <COUNT, LASTTIME, REGULARITY>	Número de dias de férias gozados, última vez de gozo, e regularidade com que goza férias.
(VACATION#) <TIMMING, PLANNING, TYPE>	Data de início e fim de cada dia de férias gozado, gozo previsto com antecedência ou não, e tipo de férias (e.g. anos anteriores, do próprio ano).

(MISSED VACATIONS) Informação relacionada com férias não gozadas nos últimos 12 meses (MISSEDVACATIONS). O símbolo # significa o número do dia de férias não gozadas.

(MISSEDVACATIONS) <COUNT, LASTTIME, REGULARITY>	Número de dias de férias não gozados, última dia de férias previsto mas não gozado, e regularidade com que acontece falhas no gozo de férias.
(MISSEDVACATION#) <TIMMING, PLANNING, TYPE>	Data de início e fim de cada dia de férias não gozado, previsão antecipada de não gozo, e tipo de férias (e.g. anos anteriores, do próprio ano).

Domínio pessoal / psicossocial

(SOCIAL SUPPORT) Dados relacionados com suporte social (SOCIALSUPPORT) dado e recebido, e saúde (HEALTH) familiar.

(SOCIALSUPPORT) <GIVEN, RECEIVED>	Autoapreciação do suporte social recebido.
(HEALTH) <CATEGORICAL>	Autoapreciação geral de saúde dos elementos do domínio privado (e.g. filhos).

(WORKLIFE) Dados relacionados com o emprego e profissão exercida.

(WORKLIFE) <PROFESSION<EXCHANGE>>	Vontade de mudar de profissão (atividade profissional desempenhada).
(WORKLIFE) <JOB<EXCHANGE>>	Vontade de mudar de emprego atual (organização ou empresa onde trabalha).
(WORKLIFE) <FULFILLMENT>	Autoapreciação do sentimento de realização profissional.

(SLEEP) **Dados relacionados com o sono (SLEEP) e com o cansaço (TIRED) no momento de acordar.**
(TIRED)

(SLEEP) <FALLASLEEP<TIMMING>>	Hora de deitar.
(SLEEP) <WAKE<TIMMING>>	Hora de acordar.
(SLEEP) <WAKE<DURATION>>	Duração do despertar.
(SLEEP) <DURATION>	Duração total de sono.
(SLEEP) <DEEPSLEEP<DURATION>>	Duração do sono profundo.
(SLEEP) <LIGHTSLEEP<DURATION>>	Duração do sono leve.
(SLEEP) <QUALITY>	Autoapreciação da noite de sono.
(TIRED) <LEVEL>	Autoapreciação do cansaço após despertar.

(WELLBEING) **Informação acerca do bem-estar (WELLBEING) dados utilizados para *ground truth*.**

(WELLBEING) <CATEGORICAL>	Autoapreciação da sensação de bem-estar.
---------------------------	--

(QUALITYOF LIFE) **Informação relativa à qualidade de vida (QUALITYOFLIFE).**

(QUALITYOFLIFE) <CATEGORICAL>	Autoapreciação da qualidade de vida, saúde e outras áreas da vida.
-------------------------------	--

Domínio pessoal / rotina

(WORK INTENSITY) **Dados do último mês relacionados com o tempo e intensidade de trabalho (WORKINTENSITY).**

(WORKINTENSITY) <WORKSHIFTS <FIRST <BEGIN <TIMMING>>>>	Hora de chegada ao primeiro período do dia.
(WORKINTENSITY) <WORKSHIFTS <FIRST <END <TIMMING>>>>	Hora de saída do trabalho do primeiro período do dia.
(WORKINTENSITY) <WORKSHIFTS <SECOND <BEGIN <TIMMING>>>>	Hora de chegada ao segundo período do dia.
(WORKINTENSITY) <WORKSHIFTS <SECOND <END <TIMMING>>>>	Hora de saída do trabalho do segundo período.
(WORKINTENSITY) <WORKSHIFTS <ELAPSED TIME>>	Duração do intervalo entre períodos de trabalho.
(WORKINTENSITY) <COMPUTERSUCCESSLOGIN<COUNT>>	Quantidade de autenticações bem-sucedidas efetuadas no computador.
(WORKINTENSITY) <COMPUTERFAILEDLOGIN<COUNT>>	Quantidade de tentativas falhadas de autenticação no computador.
(WORKINTENSITY) <COMPUTERLOCKOUTS<COUNT>>	Quantidade de bloqueios automáticos de ecrã ocorridos no computador.

(ELECTR) **Informação relacionada com a utilização do *smartphone* (ELECTR).**

(ELECTR) <USAGE<LASTTIME>>	Hora da última utilização do dia antes de dormir.
(ELECTR) <USAGE<LASTTIME <DURATION>>>	Tempo de utilização antes de adormecer .
(ELECTR) <USAGE<COUNT>>	Quantidade de utilizações diárias.

Outros domínios / interação

(CALL) (SMS)	Informação relativa à quantidade e qualidade das interações sociais ocorridas.
(CALL) <INCOMING<COUNT, DURATION>	Chamadas telefónicas recebidas.
(CALL) <OUTGOING<COUNT, DURATION>	Chamadas telefónicas efetuadas.
(CALL) <MISSEDCALL<COUNT>	Chamadas telefónicas recebidas não atendidas.
(CALL) <INITIATEDONLY<COUNT>	Chamadas telefónicas apenas iniciadas (menos de três segundos).
(SMS) <<INCOMING, WORDS, CHARS>><COUNT>>	SMS recebidos.
(SMS) <<OUTGOING, WORDS, CHARS>><COUNT>>	SMS enviados.
(CALL, SMS) <INTERLOCUTORS<COUNT>>	Quantidade de interlocutores distintos.
(CALL, SMS) <AFFINITY<COUNT>>	Afinidade com os interlocutores.
(INTERACTIONS) <CATEGORICAL>>	Autoapreciação da qualidade das interações sociais do dia.

Outros domínios / reconhecimento

(PHYSI)	Número de passos percorridos (PHYSI) <STEPS <COUNT>>.
(HEIGHT)	Altura.
(WEIGHT)	Peso (WEIGHT) e índice de massa corporal (WEIGHT, HEIGHT) <BMI>.
(WORKLIFE)	Informação acerca da vida profissional e desempenho de papéis na sociedade.
(WORKLIFE) <DURATION>	Número de anos de trabalho (todos os empregos).
(WORKLIFE) <JOB<DURATION>>	Número de anos do emprego atual.
(WORKLIFE) <JOB<COUNT>>	Número de empregos desde o início da vida ativa.
(WORKLIFE) <PROFESSION<DURATION>>	Número de anos de exercício da profissão atual.
(SOCIETY)	Informação acerca de aspetos relacionados com a vida em sociedade.
(SOCIETY) <POLITICALMEMBER>	Filiação em partidos políticos.
(SOCIETY) <ACTIVEPARTICIPATION>	Participação ativa em partidos políticos, religião, ou outras associações.
(SOCIETY) <RELEVANTROLES>	Desempenho de papéis sociais de relevo em organizações sociais, políticas ou religiosas (e.g. junta de freguesia, grupo desportivo, etc.).
(CAFFEI)	Número aproximado de tomas diárias de café (CAFFEI) <COUNT>.
(SMOKING)	Número aproximado de cigarros consumidos por dia (SMOKING) <COUNT>.
(ALCOH)	Consumo de álcool (e.g. refeições, amigos, nunca) (ALCOH) <CATEGORICAL>.
(DIET)	Faz ou já fez dieta para manutenção do aspeto físico (DIET) <CATEGORICAL>.
(SOCCULTURAL EVENTS)	Número de eventos culturais a que presenciou no último mês (SOCCULTURALEVENTS) <COUNT>.

(NATURE)	Número de dias despendidos na natureza no último mês (NATURE) <DURATION>.
----------	---

Tabela 3.4: Variáveis de contexto selecionadas (contexto psicossocial)

Por fim, foram também recolhidas outras variáveis de contexto, utilizadas no pré-processamento do sinal (EDA).

Variável	Descrição
(TEMP)	A temperatura ambiente (TEMP) e os dados do acelerómetro (ACC) em conjunto com a
(ACC)	(ST) apresentado na Tabela 3.3, foram utilizadas no tratamento do ruído do sinal (EDA).

Tabela 3.5: Variáveis de contexto selecionadas (outras variáveis)

3.1.2 Seleção de Instrumentos & Sensores

Perante a incapacidade de medir as emoções diretamente (Nawasalkar et al., 2013), os investigadores recorrem a instrumentos & sensores para recolher dados de contexto a serem correlacionados pelos sistemas de reconhecimento automático de emoções (Sano and Eng, 2016). Os recentes avanços tecnológicos dos sensores tem permitido aumentar a precisão dos algoritmos no reconhecimento emocional (Wu et al., 2007). Nesta secção são apresentados os instrumentos & sensores selecionados para recolher dados de contexto na experiência descrita nesta tese.

Com o surgimento de mais e melhores sensores, e de dispositivos com maior capacidade de processamento, espera-se um crescimento de investigações que recorram à medição objetiva, para tentar aumentar a precisão dos sistemas de reconhecimento emocional (Gogia et al., 2016). Apesar da medição objetiva proporcionada pelos instrumentos & sensores, não é possível estabelecer uma relação direta entre esses dados e as emoções (Rani and Sarkar, 2006). No entanto, estes dados podem ser utilizados com o objetivo de inferir emoções e, mais concretamente, o bem-estar em trabalhadores de escritório.

Os sistemas são cada vez mais caros e complexos (Caballe, 2015). A necessidade de diversificar os dados de contexto a correlacionar, torna inevitável a utilização de um maior número de dispositivos na recolha de dados de contexto. Contudo, são várias os fatores a ter em conta no momento da seleção dos dispositivos: obstrutividade e influência nos dados a recolher; escolha de questionários aferidos para a população alvo; investimento e constrangimentos orçamentais; questões de ordem ética e de segurança para os participantes e organizações; etc. Por outro lado, há que pesar a relevância dos dados com que cada instrumento pode contribuir para o objetivo final pretendido, neste caso, um *dataset* final de boa qualidade.

Pretendia-se utilizar a menor quantidade possível de instrumentos & sensores para manter a normalidade do contexto pessoal e profissional dos participantes na experiência.

Além disso, pretendia-se utilizar apenas instrumentos que apresentassem um baixo nível de obstrutividade para evitar constrangimentos aos utilizadores, mudanças de comportamento, e enviesamento dos próprios dados recolhidos (e.g. stress provocado pela utilização de sensores, preenchimento de longos questionários) (Guinot Jimeno et al., 2011; Tran et al., 2007; Barea et al., 2011; Wu et al., 2008). De preferência, preferia-se utilizar instrumentos já usados pelos participantes no seu dia-a-dia, de forma a evitar interferências no fluxo de trabalho. Pretendia-se que a utilização de instrumentos & sensores fosse o mais despercebida possível, não obstruindo a vida e a atividade normal dos participantes. O objetivo era evitar a contaminação dos dados recolhidos por via da interferência no contexto, e promover a qualidade ao *dataset raw* final com o qual se pretendia contribuir.

No entanto, e apesar da vontade em não utilizar instrumentos obstrutivos, optou-se por recolher dados (EDA) através de eléctrodos colados na palma da mão dos participantes utilizando o sensor {EDAMove} (movisens GmbH, 2018d,c). A relação estabelecida pela literatura entre o (EDA) e as emoções é tão forte, que não se poderia deixar de fora a recolha desta variável de contexto. Apesar de existirem alternativas, o custo de aquisição era considerável. No entanto, durante a pesquisa de opções de recolha do (EDA), foi possível estabelecer um contacto frutífero com a *Movisens* (Movisens, 2020) para empréstimo (sem custos) de dois sensores {EDAMove} (movisens GmbH, 2018c). Apesar destes sensores utilizarem eléctrodos para recolher dados de contexto, optou-se pela sua utilização atendendo ao investimento que teria que ser feito nas alternativas. No entanto, o tempo de empréstimo era limitado pressionando assim o calendário de recolha. Além disso, a preocupação relacionada com a qualidade dos dados recolhidos começava a evidenciar-se. A utilização de eléctrodos num processo de recolha em ambiente real, poderia impactar negativamente a qualidade dos dados recolhidos. Antevia-se uma fase de pré-processamento muito importante para a qualidade do *dataset* final. A fase de pré-processamento é descrita em detalhe no capítulo 4.

Os questionários são muito utilizados na avaliação emocional (Fulton and Medlock, 2003; Mandryk and Atkins, 2007) porque recolhem informação direta do que se pretende medir (Johnston et al., 2009). São também muito utilizados para recolha de *ground-truth* (Fulton and Medlock, 2003; Mandryk and Atkins, 2007). No entanto, apesar do carácter subjetivo das perguntas e respostas poder conduzir a medições erradas (Caballe, 2015), são aceites pela literatura como forma de medir emoções (Babiker et al., 2013). Alguns autores optam pela substituição da recolha do *ground-truth*, pela utilização de bases de dados de imagens, sons ou vídeos pré-etiquetados com as emoções que provocam (e.g. *International Affective Digitized Sounds* (IADS) (Bradley and Lang, 1999), *International Affective Picture System* (IAPS) (Lang et al., 2005)). Contudo, a estratégia de induzir emoções com recursos pré-etiquetados com o conteúdo emocional também não está isenta de problemas. As pessoas percecionam as imagens e sons de forma distinta entre si, porque a perceção depende da experiência passada e da personalidade de cada um e, a

forma de ponderar esta informação com os resultados da indução, é precisamente através da recolha de dados individuais com os questionários que se pretendiam evitar.

O acréscimo de tarefas decorrentes da própria experiência, pode contribuir para o enviesamento dos dados recolhidos por representarem mais tarefas a realizar. Neste contexto, decidiu-se por não utilizar os questionários formais. Desenharam-se pequenos questionários informais a aplicar através do *smartphone*, com o objetivo de diminuir ao mínimo o constrangimento na vida normal dos participantes. Optou-se por respostas simples (cf. escolha de um ícone em três opções de resposta), na tentativa de evitar que a própria resposta ao questionário significasse um aumento de stress para o participante. Os dois questionários informais, desenvolvidos pelo autor, para recolha do *ground-truth* acerca da perceção do bem-estar, foram inspirados na escala {EMMBEP} aferida para a população portuguesa (Monteiro et al., 2012).

As tabelas seguintes apresentam os instrumentos & sensores utilizados na experiência apresentada neste capítulo. A organização do conteúdo em cada tabela, respeita o carácter objetivo ou subjetivo dos dados recolhidos por cada instrumento & sensor.

A Tabela 3.6 apresenta os instrumentos & sensores, utilizados nesta experiência, que recolheram dados de contexto de carácter objetivo.

Inst. / sensor	Descrição
	Técnicas não obstrutivas
{WINDOWS ACTIVITY}	O {WINDOWSACTIVITY} foi desenvolvido pelo autor em C#, para recolher dados de utilização do computador de trabalho de cada participante. ¹ A aplicação utilizou o sistema de eventos do sistema operativo (<i>Windows</i>) para extrair informação acerca da utilização do computador (e.g. eventos <i>login</i> , <i>logout</i>). Estes dados iriam permitir determinar o número de tentativas bem e mal-sucedidas no sistema, deduzir a hora de chegada e saída de cada período de trabalho, intervalo entre períodos (e.g. almoço), etc. A recolha destes dados e a utilização desta aplicação, implicou: i) a definição de uma palavra-passe complexa para acesso à sessão de trabalho (duas maiúsculas + duas minúsculas + dois números + três símbolos); ii) configuração do bloqueio automático da sessão de trabalho após cinco minutos de inatividade; e iii) a programação das <i>Group Policies</i> para ativar o <i>log feeding</i> dos eventos que se pretendiam extrair através do {WINDOWSACTIVITY} no último dia de recolha. Estas configurações foram revertidas no final de cada iteração de recolha. Os procedimentos de configuração e reversão de configurações, fazem parte do {COLLECTIONSCRIPT} (ver Anexo 7.8). Aplicação: A aplicação foi executada no último dia de recolha para exportação dos dados do registo de eventos do sistema operativo para um ficheiro <i>sqlite</i> (SQLite Consortium, 2021). Variáveis recolhidas: (WORKINTENSITY).

¹ Apenas foram recolhidos dados estatísticos referentes aos dias de recolha.

{ANDROID ANALYTICS}	<p>A atividade no <i>smartphone</i> foi recolhida através da aplicação {ANDROIDANALYTICS}, desenvolvida pelo autor em Android. ¹ Os dados recolhidos por esta aplicação permitiria obter dados das (CALL), (SMS), bloqueios de ecrã, etc. No final de cada chamada, a aplicação apresentava o {CALLINTERACTIONS} para catalogar a interação social acabada de ocorrer para: i) classificar o interlocutor (cf. cônjuge ou filhos, amigo chegado, restante família e conhecidos, colegas de trabalho, outros interlocutores de trabalho); e ii) recolher a Autoapreciação acerca da qualidade da interação social (cf. negativa, neutra, positiva) (ver anexo Anexo 7.5). O {ANDROIDANALYTICS} também permitiu aplicar de forma mais simples os questionários: i) {WELLBEING1}, {WELLBEING2} e {SLEEPTIRED}. A digitalização destes questionários pretendia evitar esquecimentos de preenchimento, e diminuir alguma obstrutividade causada pelo seu preenchimento. A instalação e desinstalação desta aplicação também fazem parte do {COLLECTIONSCRIPT} (ver Anexo 7.8). Aplicação: A aplicação foi instalada no início do período de recolha, e desinstalada no último dia, após a extração da base de dados <i>sqlite</i> (SQLite Consortium, 2021) com os dados recolhidos. Variáveis recolhidas: (CALL), (SMS), (AFFINITY), (INTERACTIONS), (ELECTR) e (SCREEN).</p>
Técnicas pouco obstrutivas	
{MIBAND}	<p>A pulseira <i>XiaoMI Mi Band</i> {MIBAND} foi utilizada para recolher dados de contexto cardíaco, atividade física e sono. A {MIBAND} é um produto comercial utilizado normalmente para monitorizar a quantidade de passos e a qualidade do sono. Permite recolher o número de passos do utilizador, frequência cardíaca (HR) via (PPG), e permite também monitorizar o sono do utilizador (SLEEP) (TechFunology, 2016; XiaoMi, n.d.). Aplicação: Utilização diária nos dias de recolha. Variáveis recolhidas: (HR), (SLEEP) e (PHYSI).</p>
Técnicas obstrutivas	
{EDAMOVE}	<p>O sensor {EDAMOVE} é afixado como uma pulseira na mão não-dominante. Foi utilizado principalmente para recolher dados de contexto acerca da pele (movisens GmbH, 2018d). O {EDAMOVE} é um sensor bastante utilizado noutras investigações e é referenciado por vários documentos científicos: Herlan et al. utilizaram este sensor num estudo acerca dos padrões do (EDA) nas diferentes fases do sono (Herlan et al., 2018), Basacik et al. estudaram a vertente fisiológica da sobrecarga cognitiva durante a condução de comboios utilizando o {EDAMOVE} (Basacik et al., 2015); Kärner utilizou o {EDAMOVE} para estudar a reatividade fisiológica na resolução de problemas (Kärner, 2017); etc. Pesa 31 gramas, e é composto por um sensor de recolha e de dois elétrodos Ag/AgCl a serem colocados na parte interior da mão na zona onde termina o pulso. O dispositivo pode guardar dados até quatro semanas. Aplicação: Utilização diária nos dias de recolha, na mão não-dominante. Variáveis recolhidas: (EDA), (ST), (TEMP) e (ACC).</p>

Tabela 3.6: Instrumentos & sensores selecionados (medição objetiva)

A Tabela 3.7 apresenta os instrumentos escolhidos para a recolha de dados subjetivos.

Inst. / sensor	Descrição
Questionários formais	
{WHOQOLBREF}	<p>O <i>World Health Organization Quality-Of-Life assessment</i> {WHOQOLBREF} (Lewandowski et al., 2015), foi criado pela Organização mundial de Saúde (OMS) para avaliar a qualidade de vida das pessoas (Group, 1998; Who, n.d.) (Rehabilitation Institute of Chicago, 2014). O {WHOQOLBREF} avalia a perceção individual no contexto da sua cultura, valores, objetivos pessoais, e preocupações. O instrumento avalia os domínios: saúde física, saúde psicológica, relações sociais e ambiente. Será utilizada a versão portuguesa (Vaz Serra et al., 2006; Universidade de Coimbra, n.d.). A autorização para utilização do questionário pode ser consultada no Anexo 7.15. Aplicação: Questionário respondido uma vez, antes do início da experiência. Variáveis recolhidas: (QUALITYOFLIFE).</p>

{EMMBEP}	A escala <i>Échelle de Mesure des Manifestations du Bien-Être Psychologique</i> {EMMBEP}, permite medir a manifestação do bem-estar psicológico dos respondentes (Monteiro et al., 2012). Foi adaptada para Portugal em 2012, e já foi utilizada em várias investigações portuguesas (e.g. (Ferreira, 2013), (Gonçalves, 2011), (Monteiro, 2008)). A estrutura do {EMMBEP} é constituída por seis fatores que traduzem a felicidade, sociabilidade, controlo de si e dos acontecimentos, envolvimento social, autoestima e equilíbrio. A autorização para utilização deste questionário pode ser consultada no Anexo 7.14. Aplicação: Este questionário não foi aplicado nesta experiência. Serviu apenas de base para criar os questionários informais {WELLBEING1} e o {WELLBEING2} para recolher a perceção do bem-estar. Seleccionaram-se as questões que se considerou melhor permitirem a recolha da perceção do bem-estar dos participantes. Variáveis recolhidas: <i>Não aplicável.</i>
Questionários informais	
{SOCIO CULTURAL}	O questionário {SOCIOCULTURAL} recolheu dados gerais acerca de cada participante. Recolheu informação genérica como idade, género, etnia, etc., e informação mais específica como dados referentes a dependentes, dados relacionados com a casa, trabalho, créditos, vizinhos, participação na sociedade e eventos culturais, etc. (ver Anexo 7.1). Aplicação: Questionário respondido uma vez, antes do início da experiência. Variáveis recolhidas: (AGE), (GENDER), (ETHNICITY), (CIVILSTATUS), (RELIGION), (LIVING), (ACADDDG), (DEPENDENTS), (HOME), (HOME-WORK), (NEIGHBORS), (LOANS), (SOCIALSUPPORT), (HEALTH), (WORKLIFE), (WEIGHT), (HEIGHT), (SOCIETY), (CAFFEI), (SMOKING), (ALCOH), (DIET), (SOCULTURALEVENTS) e (NATURE).
{WELLBEING1}	O questionário {WELLBEING1} permitiu recolher o <i>ground-truth</i> (em conjunto com o {WELLBEING2}) (ver Anexo 7.2). Recolheu informação acerca da perceção de bem-estar após um dia de trabalho. O questionário foi aplicado pela aplicação {ANDROIDANALYTICS}. A hora de resposta podia ser definida pelo participante. As perguntas que compõem o questionário foram inspiradas no {EMMBEP}. Questões: i) <i>Relacionei-me facilmente com as pessoas à minha volta;</i> ii) <i>Estive capaz de enfrentar situações difíceis de uma forma positiva;</i> iii) <i>Senti que os outros gostavam de mim e me apreciavam;</i> iv) <i>Senti-me satisfeito com o que fui capaz de alcançar, senti-me orgulhoso de mim próprio;</i> e v) <i>A minha vida foi bem equilibrada entre as minhas atividades familiares, pessoais e académicas.</i> Resposta (opções): { <i>discordo, nem discordo nem concordo, concordo</i> }. Aplicação: Respondido pelo participante ao final de cada dia de recolha, momentos antes de deitar. Variáveis recolhidas: (WELLBEING).
{WELLBEING2}	O questionário {WELLBEING2} complementou o {WELLBEING1} na recolha de <i>ground-truth</i> (ver Anexo 7.3). Foi aplicado pela aplicação {ANDROIDANALYTICS}. O {WELLBEING2} recolhia informação acerca da perceção do bem-estar, pelo menos uma vez durante o dia. A aplicação despoletava o pedido de preenchimento uma vez por dia automaticamente, mas o utilizador poderia registar o seu bem-estar as vezes que pretendesse se abrisse a aplicação. No futuro, o {ANDROIDANALYTICS} será evoluído, para aplicar este questionário sempre que detete um evento que possa impactar o bem-estar do utilizador. As perguntas também foram inspiradas do {EMMBEP}. Questões: i) <i>Senti-me emocionalmente equilibrado;</i> ii) <i>Senti-me bem, em paz comigo próprio;</i> e iii) <i>Senti-me confiante.</i> Resposta (opções): { <i>discordo, nem discordo nem concordo, concordo</i> }. Aplicação: Respondido pelo participante, pelo menos uma vez durante o dia. Variáveis recolhidas: (WELLBEING).
{SLEEPTIRED}	Este questionário recolheu a Autoapreciação subjetiva acerca da qualidade do sono e do sentimento de cansaço, após cada noite de sono. Este questionário foi aplicado pelo {ANDROIDANALYTICS} numa hora definida previamente pelo utilizador (ver Anexo 7.4). Questões: i) <i>Dormi muito bem e sinto que o meu sono foi totalmente reparador;</i> e ii) <i>Sinto-me totalmente descansado após esta noite de sono.</i> Resposta (opções): { <i>discordo, nem discordo nem concordo, concordo</i> }. Aplicação: Respondido pelo participante ao início de cada manhã. Variáveis recolhidas: (TIRED) e (SLEEP).

{CALL INTERACTIONS}	Este questionário permitiu recolher a qualidade emocional, e identificar o tipo de inter-locutor a cada (CALL). Questões: i) <i>Identifique o tipo de interlocutor</i> ; e ii) <i>Identifique a qualidade emocional da interação social</i> ¹ . Resposta (opções): questão i) <i>Cônjuge, filho, ou amigo próximo; Restantes familiares e amigos; Outros (e.g. colegas de trabalho, supervisores, clientes, fornecedores, vizinhos)</i> ; ¹ e questão ii) { <i>Interação negativa, Interação neutra, Interação positiva</i> }. Aplicação: aplicado pelo {ANDROIDANALYTICS} a cada (CALL) efetuada ou recebida (ver Anexo 7.5). Variáveis recolhidas: (CALL) e (SOCIAL).
{LABOR}	O {LABOR} foi criado para recolher dados relacionados com a atividade laboral de cada participante, como faltas e atrasos ao trabalho, férias gozadas e não gozadas, etc. (ver Anexo 7.6). Aplicação: Questionário respondido pelo departamento de recursos humanos da organização (ou alguém indicado pelo empregador), respondido uma vez por participante durante o período de recolha. Variáveis recolhidas: (WORKABSENCES), (WORKDELAYS), (WORKLATES), (VACATIONS) e (MISSEDVACATIONS).

Tabela 3.7: Instrumentos & sensores selecionados (medição subjetiva)

A Tabela 3.8 apresenta outros instrumentos utilizados na recolha de dados de contexto.

Inst. / sensor	Descrição
{SENSOR MANAGER}	Aplicação <i>Movisens Sensor Manager</i> {SENSORMANAGER}, foi utilizada para transferir os dados recolhidos do sensor {EDAMOVE} para o computador (movisens GmbH, 2018e).
{DATA ANALYSER}	Aplicação <i>Movisens Data Analyser</i> {DATAANALYSER}, foi utilizada para pré-processar os ficheiros provenientes do {SENSORMANAGER}, disponibilizando o resultado nos formatos <i>Comma-Separated Values</i> (CSV) e Unisens (Kirst, 2010). Os dados são agregados através da média, pelo próprio instrumento, no <i>Sampling Interval</i> (SI) pretendido pelo utilizador. Nesta experiência utilizou-se um $SI = 1h$. (movisens GmbH, 2018a,b). O autor decidiu fazer o pré-processamento dos dados dos sensores {EDAMOVE} no final de cada iteração de recolha, atendendo ao tempo limitado de disponibilidade do material emprestado pela Movisens (Movisens, 2020).
{NOTIFY FITNESS}	Aplicação <i>Notify & Fitness for MiBand</i> {NOTIFYFITNESS}, foi utilizada para transferir os dados recolhidos da pulseira {MIBAND} para o computador (OneZeroBit, 2021).
{GOOGLE FORMS}	O <i>Google Forms</i> {GOOGLEFORMS}, aplicação <i>web</i> utilizada para criar e aplicar o questionário {SOCIOCULTURAL}. Os dados foram retirados da aplicação através da exportação para CSV (Google, 2021).

Tabela 3.8: Instrumentos & sensores selecionados (outros instrumentos)

3.1.3 Documentação

Nesta experiência recolheu-se dados de contexto de vários participantes, trabalhadores de diferentes organizações, e utilizando várias instrumentos & sensores. O processo de recolha decorreu durante durante mais de dois meses, e foi dividido em múltiplas iterações de recolha por causa do número limitado de sensores. A quantidade de participantes, organizações, instrumentos, e variáveis de contexto, faziam antever uma fase de execução complexa que deveria ser convenientemente planeada. Neste contexto, havia a

¹As opções para resposta foram inspiradas no modelo *convoy* de Antonucci (Antonucci, 2001).

necessidade de criar um conjunto de documentação para apoiar a fase de execução. Nesta secção apresentam-se os documentos de suporte criados pelo autor, para apoiar e controlar a logística associada ao processo de recolha de dados de contexto.

A ficha de participação em experiência {`SUBJECTFILE`}, foi criada com o objetivo de registar os dados gerais de cada participante, e os sensores atribuídos a cada um. A estratégia passava por iniciar novas iterações de recolha, enquanto fosse estivessem disponíveis os sensores {`EDAMOVE`} emprestados pela *Movisens* (Movisens, 2020). O *template* deste documento pode ser consultado no Anexo 7.7.

Atendendo ao número de instrumentos & sensores, parametrizações necessárias, e procedimentos a executar no final de cada período de recolha de cada participante, optou-se por controlar o fluxo das iterações de recolha através de um *script* criado antecipadamente {`COLLECTIONSCRIPT`}. Pretendia-se definir a sequência de tarefas a executar para que fossem constantes ao longo processo de recolha. O resultado final da aplicação deste *script*, foi um *dataset raw* organizado uniformemente entre participantes, organizações e iterações de recolha. Este *script* pode ser consultado no Anexo 7.8.

O processo de recolha iniciou com o envio do convite às organizações a solicitar a participação na experiência. O Anexo 7.9 apresenta o conteúdo da carta enviada às organizações. O documento apresenta em traços gerais a experiência a realizar e os objetivos da experiência.

Colocou-se também à disposição das organizações, um modelo de resposta que poderiam utilizar para responder afirmativamente à participação na experiência. Pretendia-se facilitar os processos administrativos na elaboração da resposta, mas também era importante que o conteúdo do documento autorizasse a recolha de dados de contexto, e reafirmasse a participação desinteressada por parte da organização. O *template* entregue às organizações como proposta de resposta ao convite de participação na experiência, pode ser consultado no Anexo 7.10.

O modelo da auto-proposição tem a vantagem de evitar a resposta a perguntas constrangedoras, mas cria algum distanciamento entre os investigadores e os candidatos. O *screening*, na opinião do autor, além de validar o cumprimento dos critérios de seleção, promove alguma confiança entre candidatos e investigadores. Como nesta experiência não se utilizou o *screening* tradicional, sentiu-se a necessidade de criar um documento adicional que as pessoas pudessem levar para casa, para ajudar na decisão acerca da sua participação. Alguns candidatos, por sua própria iniciativa, colocaram dúvidas em privado ao investigador. No entanto, a disponibilização da brochura poderá ter ajudado na decisão de auto-proposição de vários candidatos. A brochura de apresentação da experiência pode ser consultada no Anexo 7.11.

A preocupação com os dados pessoais foi uma constante desde o início da preparação da experiência. A declaração de consentimento informado por parte dos participantes, é a formalização do carácter voluntário de participação na experiência, em que os partici-

pantes declaram ter percebido o propósito, e os dados a serem recolhidos do seu contexto. A declaração de consentimento pode ser consultada no Anexo 7.12. Só foram recolhidos dados dos participantes, após terem entregue ao investigador a respetiva declaração de consentimento assinada.

Atendendo aos longos períodos de recolha, e considerando que poderiam surgir imprevistos que pudessem por em causa a recolha de dados, o autor criou alternativas a serem utilizadas caso surgissem impedimentos na utilização dos instrumentos principais. Apenas foram criadas versões dos instrumentos com viabilidade de recolha manual (e.g. é possível recolher o bem-estar através de um questionário em papel, mas não é viável recolher o (HR) através de um registo manual). Os instrumentos *backup* criados podem ser consultados nos Anexos: 7.16, 7.17, 7.18, 7.19, e 7.20.

A Tabela 3.9 resume a documentação de suporte criada para apoiar a fase de execução da recolha de dados de contexto.

Documento	Descrição
{SUBJECT FILE}	A ficha de participação na experiência {SUBJECTFILE} foi criada com o objetivo de melhorar a gestão dos dados dos participantes e os instrumentos & sensores por eles utilizados. O {SUBJECTFILE} é a folha de rosto do processo individual de cada participante na experiência. Conteúdo: i) dados gerais do participante e dias de recolha; ii) configurações efetuadas no sistema operativo do computador do participante (utilizado para registar quais as configurações que deveriam ser revertidas após o processo de recolha); iii) identificação da pulseira {MIBAND} e pulso de utilização; e iv) identificação do sensor {EDAMOVE} e pulso (mão) onde foram colados os elétrodos. Anexo: 7.7.
{COLLECTION SCRIPT}	O <i>script</i> de suporte à recolha de dados dos participantes {COLLECTIONSCRIPT}, permitiu que, ao longo do processo de recolha, o procedimento fosse único e transversal a todos os participantes. O objetivo era garantir a consistência da organização do <i>dataset raw</i> no final, entre os vários participantes. O {COLLECTIONSCRIPT} descreve a sequência de tarefas a executar em cada iteração de recolha de dados de contexto. Cada <i>sprint</i> iniciou com a preparação do processo de recolha (dia anterior ao início de recolha), e terminou com a extração dos dados dos instrumentos & sensores e com o armazenamento da informação na estrutura de pastas definida. Conteúdo: i) tarefas executadas no ato público de apresentação do projeto para angariação de participantes (e.g. explicar o modelo da auto-proposição); ii) tarefas para o dia anterior ao início da recolha (e.g. preenchimento do {SUBJECTFILE}, configurações de SO, instalação de aplicações, <i>reset</i> de dados da iteração anterior, pedido de preenchimento do {LABOR}); iii) apoio ao arranque do processo no primeiro dia de recolha (e.g. validar a correta utilização de elétrodos, verificar o execução das aplicações); iv) no último dia, recolha dos instrumentos & sensores, reversão das configurações no SO, recolha do ficheiro {LABOR}, etc.; e v) exportação e arquivo de dados dos vários instrumentos & sensores, respeitando a arquitetura de pastas e ficheiros previamente definida. Anexo: 7.8.

Convite às organizações	<p>O convite foi endereçado por escrito às organizações, e dava a conhecer a experiência a realizar. O conteúdo da carta pretendia transmitir credibilidade à experiência, e convidar as organizações a participar. O objetivo era tentar agendar uma reunião com os decisores, com o objetivo de abordar o tema com maior detalhe, promover a reunião pública com os trabalhadores de escritório para divulgar a experiência, e iniciar o processo de angariação de participantes através da sua auto-proposição. Conteúdo: i) pressupostos de aceitação à participação pelas organizações; ii) participação voluntária dos trabalhadores de escritório, utilizando o método da auto-seleção através da auto-proposição; iii) assinatura prévia da declaração de consentimento informado pelos participantes; iv) variáveis de contexto a recolher e privacidade no tratamento de dados pessoais durante todo o processo; v) instrumentos & sensores a utilizar e a necessidade de instalar aplicações no computador e <i>smartphone</i> para recolher dados estatísticos; vi) resultados esperados com a experiência e vantagens que poderiam significar para o ambiente de escritório, e para os seus trabalhadores; e vii) importância do desinteresse da organização em conhecer os dados recolhidos dos seus trabalhadores, individualmente ou em conjunto. Anexo: 7.9.</p>
Aceitação das organizações	<p>O <i>template</i> entregue às organizações como sugestão de resposta afirmativa à participação na experiência, pretendia atingir vários objetivos. Antes de mais, por questões de documentação futura, pretendia-se que a resposta afirmativa à participação fosse efetuada por escrito. As organizações poderiam optar por responder oralmente, algo que o autor pretendia evitar. Com a proposta de um modelo, estar-se-ia a facilitar o processo administrativo de responder, e a propor o conteúdo pretendido no documento de resposta. Além disso, desta forma, ficava garantido que a redação autorizaria a recolha de dados, e ficaria reforçada a ideia de que os responsáveis não teriam acesso aos dados recolhidos. Conteúdo: i) aceitação da recolha de dados de contexto após consentimento dos trabalhadores; ii) autorização para a instalação de aplicações nos computadores e <i>smartphones</i>; iii) aceitação para cedência de tempo de trabalho para preenchimento de questionários breves; iv) aceitação do não acesso aos dados recolhidos de cada trabalhador. Anexo: 7.10.</p>
Brochura de apresentação	<p>A brochura de apresentação da experiência foi criada com o objetivo de ajudar os candidatos na decisão de se auto-proporem como participantes na experiência. O autor tinha o receio que os candidatos não colocassem as suas dúvidas durante a reunião pública e viram na criação desta brochura uma forma de passar mais informação às pessoas. A brochura foi disponibilizada antes da reunião pública e as pessoas poderem ficar com um exemplar para si. Conteúdo: i) importância da privacidade para o autor e a não partilha de informação com ninguém; ii) explicação do modelo da auto-proposição, como forma de evitar perguntas constrangedoras, através da promoção da auto-exclusão; iii) pressupostos para admissão (e.g. utilização de sensores, instalação de aplicações no <i>smartphone</i> e computador, inexistência de eventos marcantes há menos de 6 meses); iv) dados a serem recolhidos; e v) relevância da experiência na promoção do bem-estar no trabalho de escritório. Anexo: 7.11.</p>
Declaração de consentimento informado	<p>A preocupação com a privacidade foi algo que o autor sentiu desde o momento que iniciou a seleção das variáveis de contexto a recolher. Houve sempre a preocupação de pesar a relevância dos dados para o <i>dataset</i> pretendido. Na brochura de apresentação e na reunião com os participantes, o autor foi sempre transparente relativamente aos dados que pretendiam recolher e a razão pela qual seriam recolhidos. Para formalizar o entendimento por parte dos participantes, solicitou-se a cada um que assinasse de forma consciente e informada a declaração de consentimento. Conteúdo: i) assunção por parte do candidato que entendeu os detalhes da experiência e que foram satisfeitas as suas dúvidas; e ii) conhecimento por parte do candidato que pode desistir da sua participação, a qualquer momento, mesmo depois de assinar a declaração de consentimento. Anexo: 7.12.</p>

Formulários <i>backup</i>	Foram criados vários formulários <i>backup</i> para permitir a continuação do processo de recolha caso ocorresse alguma indisponibilidade dos instrumentos principais de recolha (e.g. inoperabilidade do computador ou <i>smartphone</i>). Os seguintes instrumentos foram objeto de disponibilização de versão em papel: {WELLBEING1}, {WELLBEING2}, {SLEEPTIRED}, {CALLINTERACTIONS} e {WINDOWSACTIVITY}. Conteúdo: Dados de contexto possíveis de recolher através do registo manual em papel. Anexo: 7.16, 7.17, 7.18, 7.19 e 7.20.
---------------------------	---

Tabela 3.9: Documentação de suporte à execução da experiência

3.2 Execução

Depois da preparação da experiência, iniciou-se a fase de recolha de dados de contexto, com o pedido de empréstimo dos {EDAMOVE} e com a aquisição das pulseiras {MIBAND}. Nesta secção apresenta-se a fase de execução da experiência, descreve-se o processo de constituição da amostra e da recolha de dados de contexto dos participantes.

A disponibilidade dos {EDAMOVE} era o fator que mais pressionava o cronograma de recolha. Eram apenas dois sensores e tinham que ser devolvidos após 6 semanas da data de receção. Nesse período de tempo, teria que se conhecer o funcionamento dos dispositivos, aprender a utilizar as aplicações para programação e exportação de dados, e, principalmente, efetuar a recolha de dados de contexto no maior número possível de participantes. Apesar de se ter conseguido prorrogar a devolução dos {EDAMOVE} duas semanas, só se conseguiu fazer 10 iterações de recolha com duração média de três dias.

No início desta fase, adquiriram-se também as duas pulseiras {MIBAND} a serem utilizadas no processo de recolha. Adquiriu-se também duas aplicações {NOTIFYFITNESS}, a instalar em dois *smartphones* do autor. Esta aplicação permitiu a transferência de dados para o computador, em formato CSV, dos vários dados de contexto recolhidos pela {MIBAND} (e.g. (HR)).

Os constrangimentos impostos ao cronograma de recolha pelos sensores {EDAMOVE} obrigava a uma otimização do tempo. Assim, decidiu-se por executar em paralelo as tarefas relacionadas com a constituição da amostra e recolha de dados de contexto. À medida que organizações aceitavam participar na experiência, e existiam participantes já com declarações de consentimento assinadas, o processo de recolha naquela organização era calendarizado e executado, enquanto que outras organizações e participantes decidiam se queriam ou não participar.

3.2.1 Constituição da Amostra

A preocupação do autor com os dados pessoais foi permanente durante toda a experiência. Atendeu-se sempre às questões relacionadas com a privacidade das pessoas, mas também com os dados das organizações. Nesta secção aborda-se o processo que conduziu à constituição da amostra. Durante as reuniões que ocorreram com os decisores das organizações e com os trabalhadores de escritório (candidatos a participarem na experiência), houve sempre a preocupação em demonstrar a importância da privacidade, com foco especial nos dados pessoais e no direito que cada pessoa tem ao recato da sua vida privada.

Neste contexto, decidiu-se não efetuar o *screening* através de entrevistas e questionários. Considerando a natureza dos dados a recolher e dos critérios de exclusão definidos para constituição da amostra (ver Tabela 3.2), pretendia-se evitar a colocação de questões invasivas e suscitar respostas com conteúdos constrangedores que fossem gatilhos para sentimentos menos agradáveis. A preocupação com a privacidade das pessoas, conduziu à decisão de se utilizar a estratégia da auto-proposição na angariação de participantes para a experiência. A ideia passava por promover que cada candidato pudesse decidir sozinho acerca da sua predisposição para a participação na experiência, e se voluntariasse livremente, sem qualquer pressão, convite ou interpelação.

Assim, foram os próprios candidatos a ajuizar a existência ou não de eventos marcantes recentes na sua vida, que pudessem por em causa os dados recolhidos. O modelo da auto-proposição permitiu que os candidatos se auto-excluísem pela via da não auto-proposição. O peso que tem o mesmo evento marcante para duas pessoas, pode ser diferente. Da mesma forma que não existe uma forma objetiva de avaliar as emoções das pessoas, também não há uma forma fiel de avaliar o impacto de um evento marcante, recente ou longínquo, no estado emocional atual de alguém. Um divórcio que ocorreu há pouco tempo, pode ter promovido uma melhoria instantânea do bem-estar. No entanto, um divórcio que ocorreu há muito tempo também pode ainda repercutir efeitos emocionais negativos no presente. A morte de alguém chegado há pouco tempo, pode ter significado um alívio na carga emocional das pessoas por causa de um longo período de sofrimento que chegou ao fim, mas também pode ter sido um evento tão marcante ocorrido há já vários meses ou anos, que ainda está presente no estado emocional atual. Além disso, o impacto do mesmo evento marcante pode afetar de forma diferente as pessoas, dependentemente do seu contexto. O impacto da morte do pai no bem-estar do filho, depende da proximidade da relação entre ambos. Se não se relacionavam há vários anos, o efeito negativo no bem-estar será menor do que se tivessem um relacionamento mais próximo e frequente.

A constituição da amostra, como já referido anteriormente, decorreu em paralelo com a recolha de dados de contexto. As reuniões com os decisores das organizações e com os

trabalhadores de escritório ocorriam ao mesmo tempo que, noutra organização, se procedia à recolha de dados de contexto de outros participantes. O objetivo era ter sempre dois participantes a aguardar para fazer a recolha, para evitar longos tempos de intervalo entre iterações de recolha e, assim, otimizar o tempo de disponibilidade dos {EDAMOVE}. A constituição da amostra foi dividida em duas atividades principais: i) envio do convite às organizações, seguida da reunião com os seus decisores; e ii) angariação de participantes para a experiência, durante ou depois da reunião pública de apresentação aos trabalhadores de escritório da organização.

O primeiro contacto com as organizações foi efetuado através do envio do convite escrito, apresentado na secção 3.1.3. Este convite dava a conhecer a experiência em traços gerais, e pretendia, essencialmente, agendar uma reunião para debater o assunto com maior detalhe. Esta reunião serviu para esclarecer as dúvidas dos decisores, explicar os dados a recolher e os objetivos pretendidos, e perceber se a organização estaria disposta a cumprir com todas as regras de participação:

- Os responsáveis da organização teriam que autorizar que os participantes dispusessem de alguns minutos do seu tempo de trabalho, durante os dias de recolha, para preencher questionários rápidos não relacionados com a atividade laboral.
- A organização teria que permitir a instalação de aplicações nos computadores e *smart-phones*, com prévio conhecimento dos utilizadores, para a recolha de dados estatísticos (e.g. entradas no SO, registo de número de chamadas telefónicas e SMS). As aplicações seriam instaladas no dia anterior ao início de recolha, e desinstaladas no último dia, no final da recolha. O autor reforçou que apenas se iria recolher dados estatísticos, e nunca seriam recolhidos dados da empresa ou relacionados com o negócio.
- Os participantes teriam que utilizar elétrodos na palma da mão durante os dias de recolha, e poderiam ter necessidade de os reafixar algumas vezes durante o dia. O autor expressou a sua opinião que a utilização de elétrodos não afetaria de forma relevante o desempenho das tarefas dos trabalhadores de escritório.
- Iria ser necessário que o departamento de recursos humanos procedesse ao preenchimento de um questionário, acerca de férias gozadas e não-gozadas, faltas e atrasos, e trabalho suplementar, dos trabalhadores participantes na experiência.
- A organização teria que autorizar a distribuição de uma brochura e a realização de uma reunião pública para apresentação da experiência aos trabalhadores de escritório.
- A participação na reunião pública e na experiência, nunca poderia ser imposta a nenhum trabalhador, nem, por outro lado, a organização poderia recusar que um trabalhador específico participasse na reunião ou na experiência.
- Os dados recolhidos nunca seriam partilhados com a organização direta ou indiretamente, incluindo dados individuais ou dados agregados anónimos (vários participan-

tes). O autor assegurou também que nunca divulgaria os dados com outra pessoa ou entidade, e que só seriam tornados públicos dados anónimos agregados, i.e., nunca referenciando nem organizações, nem participantes.

Ainda durante a reunião com os decisores da organização, foi também apresentado o modelo da auto-proposição dos candidatos. Os responsáveis poderiam estar preocupados com a necessidade de se colocar perguntas invasivas da privacidade aos seus colaboradores. Foi explicado que, através do modelo de auto-proposição, evitar-se-ia a colocação de perguntas constrangedoras aos candidatos. Este modelo promoveria a auto-seleção das pessoas através da sua própria triagem. Na reunião pública de divulgação da experiência, cada candidato seria convidado a auto-avaliar se vivenciou eventos marcantes recentes da sua vida, que pudessem inviabilizar a sua participação na experiência. Em caso positivo, não se deveriam voluntariar como participantes na experiência. Ser-lhes-ia garantido durante a reunião, que ninguém lhes iria perguntar o motivo de não se terem auto-proposto.

No final da reunião, procedeu-se à entrega do *template* de resposta aos decisores da organização apresentado na secção 3.1.3, para facilitar o trabalho administrativo no caso de resposta afirmativa. O autor explicou tratar-se apenas de um modelo e que não teria que ser utilizado na resposta, mas salientaram a necessidade de incluir na redação a autorização da recolha de dados, e o desinteresse por parte da organização no acesso aos dados recolhidos. Todas as organizações aceitantes utilizaram o *template* para responder. Quase todas as organizações tomaram a sua decisão de participar na experiência ainda durante a reunião. Algumas organizações agendaram a reunião pública de divulgação da experiência aos seus trabalhadores de escritório no próprio dia. Nos casos em que a reunião ficava agendada para um dia seguinte, solicitou-se autorização para divulgação antecipada da experiência, através da disponibilização de brochuras nos espaços públicos próximos do escritório da organização.

O evento de angariação de participantes foi a reunião pública com os trabalhadores de escritório. As reuniões foram sempre realizadas após a formalização de aceitação por parte das organizações. Cada reunião iniciou com a entrega da brochura aos participantes, se assim o pretendessem, e se não tivessem já um exemplar. De seguida era apresentada a experiência detalhadamente, desenvolvendo os tópicos seguintes: i) contextualização da experiência na área do AC; ii) objetivo pretendido com a experiência e relevância do reconhecimento automático de emoções nos escritórios do futuro; iii) variáveis de contexto a recolher e preocupação com a privacidade e tratamento de dados pessoais; iv) instrumentos & sensores a utilizar na recolha e possíveis constrangimentos que poderiam significar no dia-a-dia dos participantes; e v) questionários a preencher e importância de o fazer nos momentos em que fossem solicitados.

Depois da apresentação da experiência foi também apresentado o modelo da auto-proposição como forma de evitar entrevistas ou questionários para angariar participantes.

Foi explicado que se pretendia evitar a colocação de questões constrangedoras, potencialmente provocadoras de mau-estar, ou que lembrassem fases menos boas da vida. Os participantes ficaram conscientes que ninguém os iria tentar persuadir a participar na experiência, nem nunca seriam interpelados acerca das razões da sua decisão. Ficou claro que a participação seria voluntária, e sempre proposta pelo próprio candidato. Para que cada um pudesse refletir acerca da sua vontade ou não de participar, foram discutidas em detalhe as regras principais para admissão de candidatos (ver Tabela 3.2). Para poderem participar, os trabalhadores de escritório tinham que:

- Desempenhar tarefas de escritório na organização que envolvessem interação com pessoas internas e externas à organização, há pelo menos 12 meses seguidos.
- Utilizar computador individual e autorizar a recolha de dados estatísticos relacionados com utilização do SO (e.g. *login*, *logout*).
- Utilizar *smartphone* Android individual, e autorizar a recolha de dados estatísticos relacionados com chamadas telefónicas efetuada e recebidas, SMS enviados e recebidos, e número de bloqueios e desbloqueios de ecrã.
- Assinar uma declaração de consentimento, demonstrando a sua participação livre, voluntária e informada, na experiência.
- Aceitar utilizar sensores para monitorização fisiológica, nomeadamente utilizando elétrodos colados na palma da mão (e.g. (HR), (EDA), (ST)).
- Aceitar responder a questionários, quando solicitado pelo investigador ou pela aplicação instalada no *smartphone*.
- Por fim, os candidatos não poderiam ter eventos marcantes ou complicações da vida recentes que pudessem influenciar os dados a recolher de contexto (e.g. morte de alguém chegado, divórcio, problemas graves com a justiça, problemas financeiros ou com a família). Foi dada a referência temporal dos últimos 6 meses, mas pediu-se que fossem os próprios candidatos a pesar o impacto desses eventos marcantes no seu contexto emocional presente.

Terminou-se a apresentação das regras de admissão dos candidatos, discutindo a subjetividade do conceito de eventos marcantes e complicações da vida recentes. Explicou-se aos candidatos que só cada um individualmente poderia avaliar o impacto que um determinado evento ainda tinha na sua vida e no seu bem-estar. O autor remeteu para a brochura para informações adicionais acerca da experiência, e colocou-se à disposição para responder a qualquer pergunta que entendessem, em público ou em privado. Ficou claro que teriam que ser os candidatos a demonstrar interesse em participar na experiência, porque ninguém lhes iria perguntar se queriam ou não participar, durante ou após aquele evento público.

Depois de apresentadas as regras para admissão de candidatos na experiência, foi abordada a necessidade de os candidatos assinarem uma declaração de consentimento a autorizar a recolha de dados de contexto, e da possibilidade que tinham em desistir da experiência a qualquer momento, mesmo após iniciar a fase de recolha de dados de contexto. Garantiu-se que os dados nunca seriam divulgados a terceiros, nem seriam partilhados com a organização. Os dados seriam sempre mantidos anónimos em sistemas controlados pelo investigador, e apenas seriam tornados públicos dados agregados, sem nunca especificar as organizações ou participantes onde foram recolhidos.

A maioria das questões foram levantadas publicamente durante a reunião. Apenas dois participantes de duas organizações distintas, solicitaram em privado ajuda na tentativa de perceber se os seus casos seriam enquadráveis no conceito de eventos marcantes recentes. Ambos queriam participar na experiência, mas tinham dúvidas se os seus casos específicos deveriam impedir a sua auto-proposição. O pai de um deles tinha falecido recentemente, e o outro tinha o avô hospitalizado com um quadro clínico reservado. Aconselhou-se os dois participantes a refletirem acerca da carga emocional que estes eventos poderiam representar para si no momento atual, e tomarem a sua decisão com total tranquilidade, sem pressas nem pressões.

Havia receio que alguns candidatos não colocassem as suas dúvidas durante ou após a reunião. O *screening* através de entrevistas e questionários tem a vantagem de as pessoas criarem alguma empatia com os investigadores, tornando mais fácil a colocação de questões. Nesta experiência não houve qualquer reunião individual, e existia o receio de que este distanciamento fizesse com que as pessoas não se sentissem à vontade para colocar as suas dúvidas e até para participar na experiência. Foi esse o motivo que levou o autor a criar a brochura de apresentação da experiência já apresentada na secção 3.1.3 e a promover que, cada participante, ficasse com o seu exemplar.

Alguns candidatos expressaram a sua vontade de participar ainda durante a reunião. No entanto, a grande maioria, expressou a sua vontade de participar nos dias que se seguiram. Atendendo a que existiam apenas dois sensores {EDAMOVE}, deu-se prioridade de agendamento às organizações com quantidade par de participantes, com o objetivo de recolher dados de contexto do maior número possível de participantes. Nunca foram recolhidos dados de contexto, antes dos participantes assinarem a sua declaração de consentimento informado.

Apenas foram admitidas na experiência, organizações que demonstraram preocupação e respeito pela privacidade das pessoas e dos dados pessoais. Todas as organizações respeitaram as regras previamente definidas. A validação do cumprimento de regras de admissão de participantes, só incidiu sobre aquelas que não punham em causa a privacidade dos candidatos (e.g. trabalhar com computador e *smartphone*, interagir com pessoas internas e externas). Não existiram desistências de participantes após se terem auto-proposto.

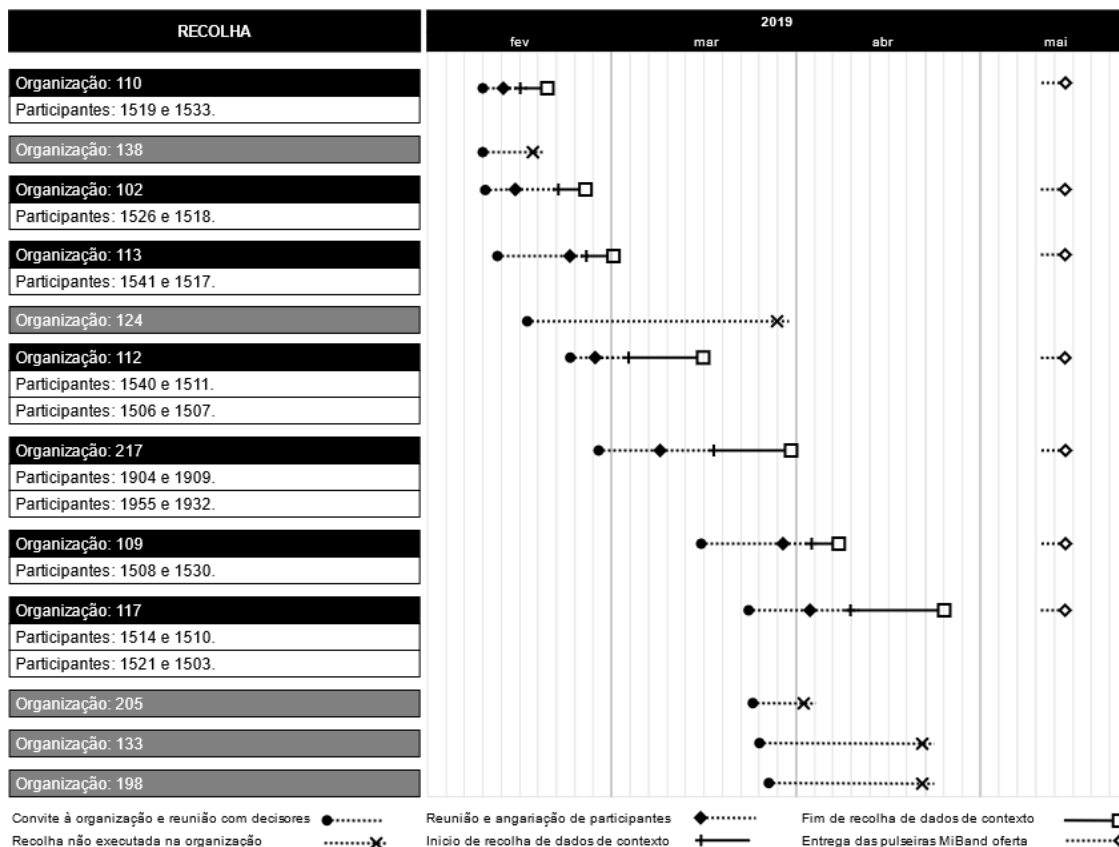


Figura 3.1: Calendário de recolha de dados de contexto

3.2.2 Recolha de dados

Atendendo ao número limitado de sensores {EDAMOVE}, existiram várias iterações de recolha. Como já referido, entre reuniões com decisores das organizações e reuniões públicas de apresentação da experiência aos trabalhadores de escritório, executava-se a recolha de dados de contexto noutra organização que já tinha passado por essas fases. Pretendia-se recolher dados de contexto do maior número possível de participantes, considerando o período de empréstimo dos {EDAMOVE}. O objetivo passava por otimizar o calendário de recolha, tentando evitar longos períodos de inatividade dos sensores. O agendamento da recolha de dados de contexto dos participantes era efetuado, assim que era validado o documento de formalização de aceitação emitido pela organização, e eram verificadas as declarações de consentimento informado assinadas pelos participantes. Nesta secção descreve-se o processo de recolha de dados de contexto. Resumidamente, esse processo começa com a preparação dos sensores para o início do período de recolha, segue com as iterações de recolha em que os dados são recolhidos dos vários participantes das várias organizações, e termina com a organização final do *dataset raw* com o qual se pretendia contribuir.

Iniciou-se a recolha de dados de contexto logo após a primeira organização e respetivos participantes terem formalizado a sua decisão para participar na experiência. Neste

contexto, a lista final de organizações e participantes a recolher, só foi conhecida no fim de todas as iterações de recolha. Por esse motivo, o calendário de recolha foi sendo construído à medida que organizações e participantes iam aceitando participar na experiência. Assim, o calendário de recolha apresentado na Figura 3.1, não resume o planeamento prévio das várias iterações de recolha, mas sim, descreve a ordem das organizações e participantes pela qual foram recolhidos os dados de contexto.

A fase de recolha de dados foi bastante exigente. Além da pressão exercida pelo tempo limitado de disponibilidade dos {EDAMOVE}, teve que se executar repetidamente as várias tarefas de logística das iterações de recolha, tendo em vista a obtenção de um *dataset raw* que considerasse dados de qualidade do maior número possível de participantes. Assim, e durante mais de dois meses, o investigador teve que, cumulativamente: i) enviar convites e promover pelo agendamento de reuniões com as organizações para apresentação detalhada da experiência e esclarecimento de dúvidas; ii) reunir com os decisores dessas organizações com o objetivo de promover a decisão de participação na experiência, e avaliar o comprometimento destes com o cumprimento das várias regras definidas; iii) conduzir reuniões públicas para apresentação da experiência aos trabalhadores de escritório com o objetivo de angariar participantes; iv) visitar a organização nos dias seguintes à apresentação pública da experiência, sob o pretexto de deixar mais brochuras, com o objetivo de ser visto pelos trabalhadores de escritório e, por essa via, mostrar de forma passiva a receptividade à aceitação de participantes para a experiência e/ou disponibilidade para esclarecimento de dúvidas; e v) executar iterativamente os vários procedimentos apresentados nesta secção, e que estão relacionados com a recolha de dados de contexto, (e.g. higienização e programação de sensores para nova iteração de recolha; instalação e desinstalação de aplicações; reversão de configurações no SO; exportação de dados dos sensores; visitas de acompanhamento pré-agendadas e de surpresa).

Por este motivo, ainda na fase de preparação, foi criado o {COLLECTIONSCRIPT} apresentado na secção 3.1.3. Antevia-se um período de trabalho longo e exigente, considerando a execução das tarefas da experiência, mas também os afazeres da vida pessoal e responsabilidades da vida profissional do autor. O {COLLECTIONSCRIPT} permitiu que as várias iterações de recolha ao longo do tempo ocorressem de forma similar, e produzissem um *dataset* no final com uma organização consistente. O instrumento descreve a sequência de tarefas a executar no processo de recolha, iniciando com a reunião de apresentação pública da experiência aos trabalhadores de escritório. Enumera as tarefas a executar desde o início de cada *sprint*, que passaram pela preparação dos instrumentos & sensores para a recolha, até à exportação dos dados recolhidos pelos sensores, no último dia, e já em laboratório. A descrição da recolha de dados de contexto apresentada nesta secção, segue a organização de tarefas do {COLLECTIONSCRIPT}.

A fase de recolha iniciou com os testes feitos aos vários instrumentos & sensores a utilizar. Nesta fase, pretendia-se testar o funcionamento do {EDAMOVE} e da {MIBAND}, mas

também das aplicações envolvidas na exportação dos dados recolhidos por estes dispositivos para o computador (cf. {SENSORMANAGER}, {DATAANALYZER} e {NOTIFYFITNESS}).

Começou-se com a atualização dos *firmware* dos dois {EDAMOVE} e duas {MIBAND} que seriam utilizados nas iterações de recolha, com o objetivo de uniformizar o comportamento na recolha e exportação, em ambas unidades de cada sensor. Fez-se também uma reflexão acerca da melhor forma de manter colados os elétrodos do {EDAMOVE} nas mãos dos participantes. Como já foi referido anteriormente, a recolha de dados através do elétrodos, em ambiente real, era uma fonte de preocupação porque poderia por em causa a qualidade dos dados. Por esse motivo, testaram-se várias estratégias para evitar a descolagem inadvertida dos elétrodos, sem que tal significasse um aumento da obstrutividade para os participantes, nem influenciasse os dados recolhidos pelo sensor. Após vários testes, decidiu-se sugerir aos participantes que, caso sentissem dificuldade em manter os elétrodos afixados, utilizassem uma rede tubular elástica em forma de luva. O objetivo era proteger os cabos que ligavam os elétrodos à pulseira do {EDAMOVE} afixada no pulso, resguardando também um pouco os elétrodos colados na mão. Esta estratégia, poderia também ir recordando os participantes da presença de um "objeto estranho" colado na mão e, assim, promover algum cuidado extra no desempenho das tarefas administrativas.

Os dados dos sensores iriam ser descartados no início de cada *sprint* de recolha. Pretendia-se garantir que, a cada iteração, os dados exportados de cada sensor apenas pertenciam ao participante que recolheu naquela iteração. Assim, era importante definir um procedimento: i) que garantisse a exportação total, em formato legível, dos dados recolhidos pelo sensor para o computador; e ii) que, antes de cada iteração, ficasse assegurada a expurgação de todos os dados recolhidos na iteração anterior. Ainda assim, e como segunda garantia, na exportação de dados da {MIBAND}, feita através do {NOTIFYFITNESS}, foram aplicados filtros de data e hora de início, para exportação apenas dos dados referentes à iteração pretendida. O {SENSORMANAGER} não permitia a aplicação de filtros para a exportação de dados do {EDAMOVE}, mas também não permitia a calendarização de um novo ciclo de recolha sem limpar previamente todos os dados da iteração de recolha anterior. Após estes testes foi possível fechar os procedimentos a executar no início e fim de cada iteração de recolha. Estes procedimentos foram incluídos no {COLLECTIONSCRIPT} para garantir a sua execução, a cada interação de recolha, sem esquecimentos.

Cada *sprint* iniciava no dia anterior ao primeiro dia de recolha, com uma verificação da documentação necessária à participação na experiência (i.e. aceitação formal da organização, declarações de consentimento informado assinadas por parte dos participantes). Os {EDAMOVE} eram programados para iniciar a recolha às 9:30 do dia seguinte e terminar no último dia de recolha previsto, às 17:30. As pulseiras {MIBAND} eram reprogramadas de origem e os dados das aplicações {NOTIFYFITNESS} instaladas nos *smartphones* do investigador, eram também eliminados. Os sensores eram higienizados e entregues aos participantes para início de utilização no dia seguinte. Aquando da entrega era dada formação

de utilização dos instrumentos & sensores (e.g. manter o (PPG) da pulseira {MIBAND} em contacto permanente com a pele, manter os eléctodos colados na mão, não molhar o {EDAMOVE}). Além disso, disponibilizava-se sempre o *cradle* para carregamento da bateria, rede tabular retalhada a ser utilizada opcionalmente como luva, gel e anéis adesivos de reserva para afixar os eléctodos do {EDAMOVE} à mão. Também eram disponibilizados carregadores de bateria das {MIBAND} aos participantes.

Ainda antes da recolha, o investigador preenchia o {SUBJECTFILE} para: i) identificar, de forma pseudonimizada, cada participante; ii) identificar os dias de recolha de dados de contexto; iii) configurações de SO efetuadas no computador de trabalho; iv) identificação do sensor {EDAMOVE} atribuído e mão utilizada para recolha; e v) identificação da {MIBAND} atribuída e pulso utilizado. De seguida, era instalada a aplicação {WINDOWSACTIVITY} para recolher dados relacionados com a utilização computador de trabalho (e.g. *login*, *logout*). Era definida uma nova senha de acesso à sessão de trabalho, ativadas as regras para registo dos eventos do SO pretendidos a serem guardados, e configurado o bloqueio automático da sessão de trabalho em cinco minutos de inatividade. Por fim, era instalada a aplicação {ANDROIDANALYTICS} no *smartphone* com o objetivo de aplicar questionários digitalizados (cf. {WELLBEING1}, {WELL BEING2}, {SLEEPTIRED} e {CALLINTERACTIONS}), e guardar dados de contexto relacionados com a interação social (e.g. (CALL), (SMS)). As baterias do {EDAMOVE} e da {MIBAND} ficavam em carregamento até o dia seguinte, i.e. início do período de recolha.

Durante a formação prévia à recolha, tentava-se consciencializar os participantes da importância dos dados recolhidos pelos vários instrumentos & sensores. Salientou-se a importância de estar sempre atento à correta afixação dos eléctodos, à necessidade do contacto permanente do (PPG) da {MIBAND} com a pele, à impossibilidade de poderem utilizar os sensores noutra mão que não a definida inicialmente, etc. Pediu-se também que respondessem aos questionários aplicados pelo {ANDROIDANALYTICS} sempre que o *smartphone* o solicitasse. Recomendou-se que personalizassem as horas em que o {WELLBEING1} e o {SLEEPTIRED} deveria ser despoletado pelo *smartphone*. Além disso, lembrou-se que poderiam responder ao {WELLBEING2} várias vezes ao dia se assim o entendessem. Informou-se também os participantes que deveriam utilizar os formulários *backup* em papel, caso ocorresse um problema com a aplicação {WINDOWSACTIVITY} ou {ANDROIDANALYTICS}. Por fim, solicitou-se aos participantes o preenchimento do {SOCIOCULTURAL}, e ao departamento de recursos humanos o preenchimento do {LABOR} considerando apenas os trabalhadores participantes na experiência.

O primeiro dia de recolha era acompanhado pelo investigador sempre que possível. Validava-se o bom funcionamento das aplicações {WINDOWSACTIVITY} e {ANDROIDANALYTICS}, a afixação dos eléctodos do {EDAMOVE} e a largura da pulseira {MIBAND}. Solicitou-se novamente o preenchimento dos questionários solicitados pelo *smartphone*, resumiram-se as boas práticas debatidas na formação do dia anterior (ver Figura 3.2), e solicitava-se

INÍCIO DA RECOLHA	
8. Entrega e explicação de funcionamento dos instrumentos a utilizar	
8.1. {MIBAND} (PHYSY, SLEEP, HR)	
8.1.1	Identificar os dados recolhidos e a importância que têm para o experimento----- <input type="checkbox"/>
8.1.2	Utilizar sempre, mesmo durante o banho e durante os períodos de sono----- <input type="checkbox"/>
8.1.3	Não trocar de pulso----- <input type="checkbox"/>
8.1.4	A pulseira tem que ser utilizada justa ao pulso (i.e. não muito apertada nem muito solta)----- <input type="checkbox"/>
8.1.5	Deve ser colocada às 09:00 no 1º dia e retirada do pulso no último dia às 17:30----- <input type="checkbox"/>
8.2. {EDAMOVE} (EDA, TEMP, ACC)	
8.2.1	Identificar os dados recolhidos e a importância que têm para o experimento----- <input type="checkbox"/>
8.2.2	Utilizar sempre incluindo durante o sono----- <input type="checkbox"/>
8.2.3	✗ Nunca utilizar o sensor em ambientes com humidade (e.g. banho)----- <input type="checkbox"/>
8.2.4	✗ O contacto dos elétrodos com a humidade dificulta a aderência dos adesivos à pele----- <input type="checkbox"/>
8.2.5	Demonstrar a colocação dos elétrodos para quando for necessário reafixar (e.g. banho)----- <input type="checkbox"/>
8.2.6	■ Piscar a vermelho 1x/2seg = sensor ativo, aguarda início período de recolha----- <input type="checkbox"/>
8.2.7	■ Piscar a vermelho 1x/1seg = sensor ativo e a registar dados recolhidos----- <input type="checkbox"/>
8.2.8	■ Piscar a vermelho 3x/1seg = sensor em erro, necessário fazer manutenção----- <input type="checkbox"/>
8.2.9	■ Piscar azul rapidamente = bateria precisa de ser recarregada----- <input type="checkbox"/>
8.2.10	■ Nenhum led ligado = sensor desligado (modo que assume após o período de recolha)----- <input type="checkbox"/>
8.2.11	🕒 Deve ser colocada às 09:00 no 1º dia e retirada no último dia às 17:30----- <input type="checkbox"/>

Figura 3.2: Excerto do {COLLECTIONSCRIPT} com instruções de utilização da pulseira {MIBAND} e sensor {EDAMOVE}

ao participante o preenchimento do {SOCIOCULTURAL} (caso ainda não o tivesse feito). O departamento de recursos humanos era interpelado acerca do preenchimento do {LABOR} e, caso não tivesse ainda preenchido o ficheiro de Excel, solicitava-se que o fizesse até ao final do período de recolha na organização.

Até ao último dia de recolha, eram feitas várias visitas não calendarizadas, para validar o cumprimento das boas práticas pelos participantes. Foram várias as vezes que se teve que retificar a afixação dos elétrodos do {EDAMOVE} e sugerir a utilização das redes tubulares como luva. Nalguns casos, a largura das pulseiras {MIBAND} também teve que ser ajustada para garantir o contacto do sensor do (PPG) com a pele. Sempre que era feita uma visita, lembrava-se a importância de cumprir com as regras de utilização dos instrumentos & sensores, e com a necessidade de responder a todos os questionários despoletados pelo *smartphone*. Os problemas encontrados durante estas visitas, confirmaram a preocupação com a qualidade dos dados. Por este motivo, ainda durante a recolha de dados de contexto, começou-se a planear a fase de pré-processamento que se antevia, também, muito exigente.

No último dia de recolha solicitava-se a resposta ao {WHOQOLBREF}, instrumento para avaliar a qualidade de vida das pessoas. Às 17:30, os dados de utilização do computador eram exportados através do {WINDOWSACTIVITY}, a aplicação era desinstalada, e as configurações específicas efetuadas no computador eram revertidas para o seu estado original antes do início da recolha. De igual forma, os dados dos questionários {WELLBEING1}, {WELLBEING2}, {SLEEPTIRED} e {CALLINTERACTIONS} eram exporta-

dos, assim como a informação estatística relacionada com as chamadas efetuadas e recebidas, SMS enviados e recebidos, e bloqueios de ecrã do *smartphone*. No final, o {ANDROIDANALYTICS} era também desinstalado. Por fim todo o material era recolhido: pulseiras {MIBAND}, sensores {EDAMOVE}, formulários *backup*, ficheiro {LABOR}, carregadores, etc.

A última fase do *script* de recolha de dados, diz respeito à exportação de dados do {EDAMOVE} e da {MIBAND}. A exportação de dados do {EDAMOVE} foi efetuada através do {SENSORMANAGER}. Depois, os dados foram pré-processados com o {DATAANALYSER}, que disponibilizou os resultados em CSV e *Unisens* (Kirst, 2010), com agregação $SI = 1h$. Os dados resultantes do {DATAANALYSER} foram estruturados em pastas. Cada pasta identifica o código da organização, número de participante, instrumento de recolha, e o formato dos dados dos ficheiros contidos (e.g. a pasta 117_1503_EdaMove_CSV contém os dados recolhidos pelo {EDAMOVE} em formato CSV, do participante 1503 pertencente à organização 117). Os dados da {MIBAND} foram exportados através da aplicação {NOTIFY FITNESS}. Foram vários os ficheiros resultantes desta exportação: i) ???_????_MiBand_steps.csv, é um ficheiro que contém informação acerca do número de passos do participante ???; ii) os ficheiros ???_????_MiBand_sleep*.csv contém informação acerca do sono; iii) os dados relacionados com o (HR) estão guardados nos ficheiros ???_????_MiBand_hr*.csv; e iv) o ficheiro ???_????_MiBand_calories.csv contém dados acerca das calorias consumidas por cada participante.

Durante o período de recolha desta experiência, não se registaram quaisquer avarias nos instrumentos & sensores utilizados. As aplicações utilizadas na exportação de dados funcionaram corretamente em todas as iterações. Não se detetaram problemas na utilização das aplicações desenvolvidas pelo autor. No entanto, três participantes tiveram que registar a sua atividade no computador através do formulário *backup* do {WINDOWS ACTIVITY}: i) as políticas de segurança dos computadores dos participantes 1506 e 1507 da organização 112 impostas pelo controlador de domínio, não permitiram a configuração de *policies* para o registo dos eventos pretendidos; e ii) o SO do participante 1909 da organização 217, estava danificado e não registava os eventos pretendidos, mesmo após as configurações necessárias. Apesar do {ANDROIDANALYTICS} ter sido corretamente instalado nos *smartphones* dos participantes 1518 e 1526 da organização 102, e participante 1506 da organização 112, o *Android* não despoletava os questionários como era suposto. Assim, decidiu-se por utilizar os formulários *backup* em papel para responder aos questionários, e também para registar as interações sociais ocorridas. Os dados recolhidos através de formulários em papel foram digitalizados manualmente pelo investigador. Todos os dados extraídos dos instrumentos & sensores, e questionários respondidos em papel, ficaram sempre na posse do investigador, e de forma anonimizada.

A fase de execução terminou com: i) a devolução do {EDAMOVE} à Movisens (Movisens, 2020); ii) exportação dos dados do {SOCIOCULTURAL} para CSV a partir do



Figura 3.3: Instrumentos utilizados para recolher dados de contexto

{GOOGLEFORMS}; e iii) organização final do *dataset raw* com o qual se pretendia contribuir (C06).

3.3 Análise

Neste capítulo apresentou-se a experiência conduzida em contexto laboral. Pretendia-se contribuir com um *dataset raw* multimodal exaustivo, recolhendo dados de vários trabalhadores de escritório, de diferentes organizações. Este *dataset* foi recolhido em ambiente real e engloba variáveis de contexto fisiológico e psicossocial. Esta secção, resume as várias contribuições feitas neste capítulo.

A experiência relatada neste documento foi exigente e complexa. Foi recolhida uma grande quantidade de dados de contexto em pouco mais de dois meses. Os constrangimentos levantados pela existência de, apenas, dois {EDAMOVE} que teriam de ser devolvidos após poucas semanas, obrigou a uma preparação antecipada de todo o processo de recolha, com o objetivo de diminuir ao mínimo as probabilidades de surgirem problemas aquando da recolha. Nesta experiência utilizaram-se 11 instrumentos de recolha (ver Figura 3.3): i) quatro instrumentos de medição objetiva (representados a cor-de-laranja) (cf. {EDAMOVE}, {MIBAND}, {WINDOWSACTIVITY}, {ANDROIDANALYTICS}); ii) dois questionários formais (representados a verde) ¹ (cf. {EMMBEP} e {WHOQOLBREF}); iii) 6 questionários informais (representados a azul) (cf. {SOCIOCULTURAL}, {WELLBEING1}, {WELLBEING2}, {SLEEPTIRED}, {CALLINTERACTIONS}, {LABOR}); iv) 6 documentos de apoio à execução da experiência (cf. {SUBJECTFILE}, {COLLECTIONSCRIPT}, convite às organizações, modelo para aceitação por parte das organizações, brochura de apresentação da experiência, e declaração de consentimento informado a assinar pelos participantes); e v) vários for-

¹O {EMMBEP} não foi considerado na contagem do número de instrumentos utilizados, porque foi apenas utilizado como base para a criação dos questionários informais {WELLBEING1} e {WELLBEING2}.

mulários de *backup* a utilizar em caso de falha dos instrumentos principais de recolha. O número reduzido de sensores fez com que o processo de recolha durasse várias semanas. Como só existiam dois sensores {EDAMOVE} disponíveis, só era possível recolher dados de contexto em dois participantes em simultâneo. Além disso, o facto de terem que ser devolvidos passadas poucas semanas, impunha uma pressão adicional na recolha de dados de contexto do maior número possível de participantes. Neste contexto, optou-se por, ao mesmo tempo que se iam recolhendo dados de participantes de uma organização: i) enviar convites para participação na experiência a outras organizações; ii) conduzir reuniões com decisores para esclarecimento de dúvidas; e iii) promover reuniões de apresentação da experiência com trabalhadores de escritório, para angariação de participantes. O planeamento e execução desta experiência constitui a primeira contribuição deste capítulo (C02).

Dos quatro instrumentos de medição objetiva já referidos, dois foram criados pelo autor: i) o {WINDOWSACTIVITY} que permitiu a recolha de dados relacionados com a utilização do computador, tendo por base os dados dos eventos do SO; e ii) o {ANDROIDANALYTICS}, instalado no *smartphone*, que permitiu recolher dados relacionados com a interação social dos participantes, e aplicar os questionários {WELLBEING1} {WELLBEING2}, {SLEEPTIRED} e {CALLINTERACTIONS}. As aplicações foram instaladas no dia anterior ao início de recolha, e desinstaladas no último dia, após exportação dos dados para bases de dados *sqlite*. As duas aplicações constituem a contribuição C03.

O autor também criou 6 questionários informais: i) o {SOCIOCULTURAL} permitiu a recolha de dados de natureza social, cultural e religiosa; ii) o {WELLBEING1} e {WELLBEING2} recolheram dados *ground-truth* acerca da perceção do bem-estar, e a sua criação foi baseada no {EMMBEP}; iii) o {SLEEPTIRED} permitiu a recolha de dados relacionados com o sono e sentimento de cansaço; iv) o {CALLINTERACTIONS} possibilitou a classificação das chamadas telefónicas efetuadas e recebidas, no que diz respeito à qualidade da interação social e tipo de interlocutores envolvidos nessas interações; e v) o {LABOR}, recolheu dados relacionados com a atividade laboral de cada participante (e.g. férias, faltas). Estes 6 questionários consubstanciam a contribuição C04.

A Tabela 3.10 resume os instrumentos & sensores utilizados nesta experiência, e cruza-os com as variáveis de contexto recolhidas por cada um.

Instrumento / sensor	Variáveis de contexto fisiológico	Variáveis de contexto psicossocial	Contribuição
Medição objetiva			
{EDAMOVE}	(EDA, ST, TEMP, ACC)		—
{MIBAND}	(HR)	(PHYSI, SLEEP) ¹	—
{WINDOWSACTIVITY}		(WORKINTENSITY)	C03
{ANDROIDANALYTICS}		(CALL, SMS, ELECTR, SCREEN)	C03
Medição subjectiva			

¹ O (SLEEP) foi recolhido através dos instrumentos {MIBAND} e {SLEEPTIRED}.

{EMMBEP}	Questionário não aplicado diretamente. ¹	—
{WHOQOLBREF}	(QUALITYOFLIFE)	—
{SOCIOCULTURAL}	(AGE, GENDER, ETHNICITY, CIVILSTATUS, RELIGION, LIVING, ACADDG, DEPENDENTS, HOME, HOME-WORK, NEIGHBORS, LOANS, SOCIALSUPPORT, HEALTH, WORKLIFE, WEIGHT, HEIGHT, SOCIETY, CAFFEI, SMOKING, ALCOH, DIET, SOCCULTURALEVENTS, NATURE)	C04
{WELLBEING1} ²	(WELLBEING)	C04
{WELLBEING2} ²	(WELLBEING)	C04
{SLEEPTIRED} ²	(SLEEP ¹ , TIRED)	C04
{CALLINTERACTIONS} ²	(SOCIAL)	C04
{LABOR}	(WORKABSENCES, WORKDELAYS, WORKLATES, VACATIONS, MISSEDVACATIONS)	C04

Tabela 3.10: Instrumentos & sensores utilizados, e variáveis de contexto recolhidas

Criou-se também um conjunto de documentação de suporte à fase de execução da experiência: i) {SUBJECTFILE} ficha de participação na experiência; ii) {COLLECTIONSCRIPT} *script* a executar a cada *sprint* de recolha; iii) convite a enviar às organizações para participação na experiência; iv) *template* de resposta de aceitação por parte das organizações (sugestão de resposta); v) brochura de apresentação da experiência; vi) declaração de consentimento a assinar por cada participante antes de participar na experiência; e vii) formulários *backup* em papel para o registo manual em caso de impossibilidade de utilização do {ANDROIDANALYTICS} (cf. {WELLBEING1}, {WELLBEING2}, {SLEEPTIRED}, {CALLINTERACTIONS}), e registo manual da atividade do computador, em caso de impossibilidade de utilizar o {WINDOWSACTIVITY}. Este conjunto de documentos de apoio perfazem a contribuição C05.

A Tabela 3.11 resume a documentação criada para dar suporte à fase de execução da experiência.

Documento	Descrição	Contribuição
{SUBJECTFILE}	Ficha de participação na experiência.	C05
{COLLECTIONSCRIPT}	<i>Script</i> de suporte à recolha de dados de contexto.	C05
Convite às organizações	Convite enviado às organizações para pedir a participação na experiência.	C05
Aceitação das organizações	Modelo entregue às organizações como sugestão de resposta afirmativa ao convite para participar na experiência.	C05

¹ O autor inspirou-se no {EMMBEP} para criar o {WELLBEING1} e {WELLBEING2}.

² Questionário aplicado pelo {ANDROIDANALYTICS}.

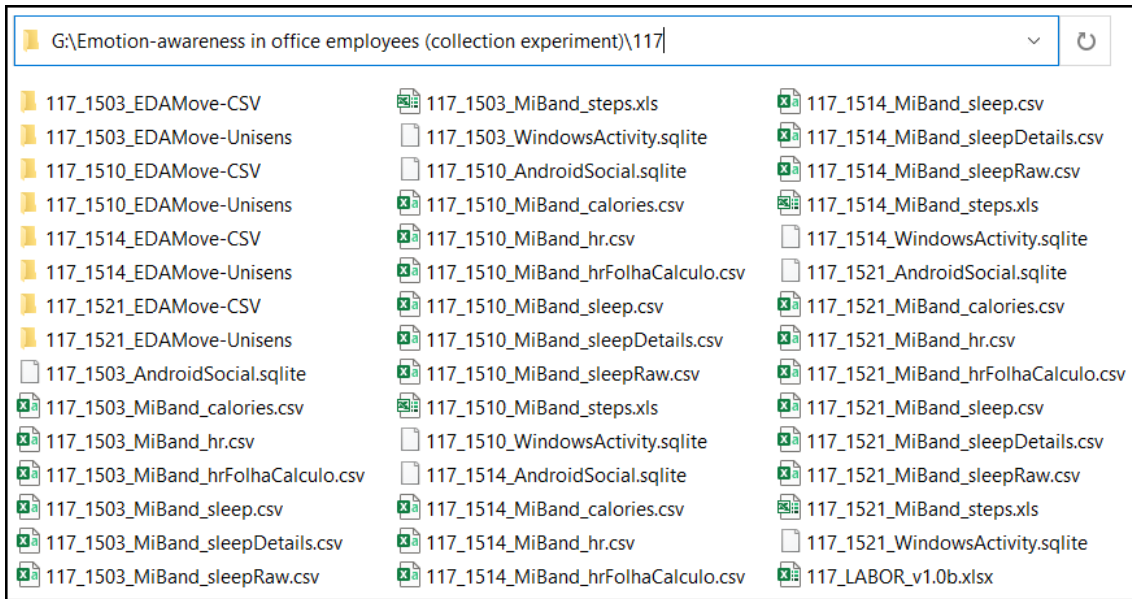


Figura 3.4: Excerto do *dataset raw* com os dados de uma organização exemplo

Brochura de apresentação	A brochura de apresentação da experiência foi criada com o objetivo de divulgar a experiência, e ajudar os candidatos na decisão de se auto-proporem como participantes.	C05
Declaração de consentimento	Declaração assinada pelos participantes para registar o consentimento informado de dados de contexto.	C05
Formulários <i>backup</i>	Vários formulários criados para permitir a continuação do processo de recolha, caso ocorresse alguma indisponibilidade de utilização dos instrumentos principais.	C05

Tabela 3.11: Resumo da documentação de suporte à fase de execução da experiência

A última contribuição reportada neste capítulo é o *dataset raw* resultante da experiência de recolha de dados de contexto em ambiente real, com dados multimodais referente ao contexto dos trabalhadores de escritório. Pretendia-se que os dados espelhassem o comportamento dos participantes, através de variáveis recolhidas num ambiente o mais próximo do real possível. Quando a recolha de dados é feita num ambiente controlado (e.g. laboratório), os investigadores gerem o contexto, implementando medidas para promover o sucesso do processo de recolha. Apesar de se conseguir melhor qualidade, os dados recolhidos poderão ser influenciados pela ação direta dos investigadores. Além disso, os participantes do experiência não estão no seu ambiente real de trabalho: i) existirá uma maior predisposição mental para o experiência; ii) os participantes estarão mais atentos às suas ações e reações; e iii) será difícil reproduzir em laboratório o ritmo e stress resultante das múltiplas tarefas e interações sociais existentes em simultâneo num escritório. Neste contexto, optou-se por recolher dados em ambiente real. Apesar desta decisão poder ter impacto na qualidade dos dados recolhidos, preferiu-se seguir este caminho porque seria difícil simular o real ambiente de escritório em laboratório.

Os dados exportados pelos vários instrumentos & sensores foram organizados e arquivados em pastas por organização. A título de exemplo, a Figura 3.4 apresenta a estrutura de ficheiros dos dados *raw* recolhidos dos participantes da organização 117. Nesta organização foram recolhidos dados de contexto dos participantes 1503, 1510, 1514 e 1521. O nome dado às sub-pastas e ficheiros dentro de cada pasta representativa da organização, identifica o número do participante, e o instrumento que recolheu os dados de contexto (e.g. 113_1541_EDAMove-Csv, representa os dados recolhidos pelo {EDAMOVE} em formato CSV, do participante 1541 que pertence à organização 113). Em alguns casos, é também identificada a variável de contexto daquele ficheiro em específico (e.g. nos ficheiros 102_1518_MiBand_sleep*.csv, estão guardados os dados recolhidos pela {MIBAND} acerca do (SLEEP) do participante 1518, funcionário de escritório da organização 102). O {LABOR} foi criado sob a forma de um ficheiro de Excel e foi preenchido pelo departamento de recursos humanos de cada organização. Neste caso, para facilitar o processo de preenchimento, o mesmo ficheiro contém informação respeitante aos vários funcionários de escritório da organização que participaram na experiência. As várias dimensões de informação (e.g. férias gozadas e não gozadas, faltas) são diferentes folhas de cálculo dentro do mesmo ficheiro. Por este motivo, o nome do ficheiro do {LABOR} é mais simples, porque identifica apenas o código da organização (e.g. 217_LABOR, significa dados de contexto relacionados com o trabalho de todos os participantes na experiência da organização 217). O ficheiro CSV do {SOCIOCULTURAL}, exportado do {GOOGLEFORMS}, contém os dados de todos os participantes. Como é um ficheiro comum a todas as organizações, foi arquivado no mesmo nível das pastas representativas das organizações.

Sobre o *dataset raw* apresentado como contribuição neste capítulo, não foi efetuado qualquer pré-processamento, com a exceção dos dados recolhidos pelos {EDAMOVE}. Como se tinha de devolver o material emprestado pela Movisens (Movisens, 2020), preferiu-se pré-processar os sinais recolhidos pelos sensores, logo após cada iteração de recolha. O pré-processamento foi feito através do {DATAANALYZER} (movisens GmbH, 2018a), que disponibilizou os dados em formato CSV e Unisens. Durante este processamento, os dados foram agregados através da média ($Sample\ Interval(SI) = 1h$).

Como as aplicações foram instaladas no dia anterior ao início da recolha, e nem sempre foi possível estar presente às 17:30 do último dia, o *dataset* contém dados de contexto que estão fora do intervalo temporal definido para recolha. Estes dados existem nos ficheiros que compõem o *dataset raw*. No entanto, foram inativados com a anotação DATA_OUTSIDE_COLLECTION_INTERVAL, no *dataset* final pré-processado apresentado como contribuição no Capítulo 4. O *dataset raw* constitui a última contribuição reportada neste capítulo (C06).

Esta experiência pretendia recolher dados de contexto de trabalhadores de escritório, com o objetivo de criar um *dataset raw* multimodal, parte fundamental para qualquer sis-

tema de deteção automática do bem-estar. Em resultado desta deteção, poderá também vir a ser possível criar sistemas que se reconfigurem de forma autónoma, para adaptar o ambiente de trabalho das pessoas, contribuindo para a tecnologia dos escritórios inteligentes. Os trabalhadores de escritório serão os principais beneficiários desta tecnologia, mas as organizações também beneficiarão, porque poderão oferecer um melhor ambiente de trabalho aos seus trabalhadores.

Capítulo 4

PRÉ-PROCESSAMENTO E PROPRIEDADES EXTRAÍDAS

A qualidade dos dados é importante para os resultados dos algoritmos. Por esse motivo, os investigadores utilizam técnicas de pré-processamento para mitigar o impacto de anomalias nos *datasets* e, assim, contribuírem para a eficiência algorítmica (Urges, 2021; Gama et al., 2012; Kandel et al., 2012; Chu et al., 2016; Jeatrakul et al., 2010; Rençberoğlu, 2019). A precisão dos sistemas de detecção emocional tem vindo a aumentar fruto da diversificação de dados de contexto recolhidos, evolução tecnológica ao nível de sensores, aparecimento de novos algoritmos Inteligência Artificial, etc. No entanto, para este aumento também tem contribuído a evolução das técnicas utilizadas no pré-processamento (Bakhtiyari and Husain, 2014; Gama et al., 2012). Neste capítulo descreve-se o pré-processamento do *dataset raw* recolhido na experiência de recolha de dados em contexto laboral, apresentada no Capítulo 3. Inicialmente é descrita a forma como os dados foram integrados, segue-se uma apresentação das técnicas utilizadas para melhorar a qualidade dos dados, e termina-se com a enumeração das propriedades extraídas escolhidas para enriquecer o *dataset* final. Ainda neste capítulo, apresenta-se o *Data Quality Visual Analysis* {DQVA}, instrumento criado pelo autor, para facilitar a exploração e análise de qualidade dos dados (Carvalho et al., 2021).

A experiência de recolha de dados de contexto descrita no Capítulo 3, pretendeu criar um *dataset* final com dados de contexto fisiológico, social e religioso. No entanto, além de se ter recolhido dados em ambiente real, existiam vários fatores que poderiam influenciar diretamente a qualidade dos dados recolhidos: i) o processo de recolha envolvia vários instrumentos & sensores; ii) um desses instrumentos era obstrutivo e utilizava elétrodos afixados na mão para recolher dados fisiológicos relacionados com a pele (cf. {EDAMOVE}); iii) o processo de recolha ocorreu no local de trabalho, durante vários dias, concomitantemente com um ritmo de trabalho intenso; iv) dada a pressão com que trabalham os trabalhadores de escritório, existia a possibilidade de saturação ou esquecimento de utilização de sensores por parte dos participantes; v) além disso, receava-se

que os questionários não fossem respondidos por falta de tempo (priorização das tarefas e solicitações do trabalho de escritório); e vi) receava-se pela falta do cuidado necessário para o bom preenchimento do {LABOR} porque os dados seriam preenchidos por pessoas que não se haviam voluntariado para a experiência (recursos humanos da organização).

Neste contexto, antevia-se a probabilidade de existirem problemas para resolver no *dataset raw*. No decurso das várias visitas às organizações durante a fase de recolha, e apesar dos vários pedidos feitos aos participantes para cumprimento das boas práticas, os receios existentes verificaram-se: i) houve a necessidade de reafixar elétrodos à mão dos participantes; ii) teve que se ajustar as pulseiras {MIBAND} para garantir a proximidade suficiente com a pele para a correta leitura do (PPG); iii) houve a necessidade de pedir a resposta a questionários em atraso; iv) aparentemente alguns {LABOR} poderiam não ter sido preenchidos com rigor necessário porque, em várias situações, o preenchimento só foi concretizado por imposição dos responsáveis das organizações após pedidos reiterados por parte do investigador; etc.

Durante as visitas, constatou-se a enorme pressão de trabalho com que vivem diariamente a generalidade dos trabalhadores de escritório. Um ritmo, por vezes, muito agressivo que contrasta com a necessidade de se atender às várias responsabilidades assumidas para participar numa experiência científica, mas também que reforça a necessidade de se estudar o bem-estar nos trabalhadores de escritório. A recolha de dados em ambiente real é um desafio. Recolher dados de contexto no próprio escritório onde trabalham os trabalhadores, era o ambiente ideal para recolher os dados que se pretendiam. Os participantes alheavam-se da experiência por causa das múltiplas tarefas e responsabilidades, permitindo assim recolher os dados realísticos pretendidos. No entanto, este alheamento significava que os participantes deixavam de estar atentos às questões logísticas da experiência (e.g. correta colocação de elétrodos, resposta atempada a questionários). Além deste problema, tentou-se reduzir ao mínimo, a quantidade de tarefas extra a serem desempenhadas pelos participantes, resultantes da sua participação na experiência. Estas tarefas poderiam significar mais stress, causando enviesamento nos dados que se pretendiam recolher. Aliás, foi por este motivo que, ainda na fase de preparação da experiência, se decidiu utilizar questionários informais mais pequenos. Decidiu-se por não utilizar os formulários formais, atendendo a que, normalmente, contém várias perguntas (e.g. o {EMMBEP} é composto por 25 perguntas). Optou-se por digitalizar os questionários criados, aplicando-os através do *smartphone*, para diminuir ao mínimo o constrangimento na vida normal dos participantes. A resposta às perguntas passava apenas por clicar num ícone entre três opções de resposta disponíveis. O objetivo era tentar evitar que a própria resposta aos questionários pudesse significar um aumento de stress. Pretendia-se também que, a simplicidade na resposta, promovesse o preenchimento assíduo dos questionários entre os vários afazeres normais de trabalho dos participantes.

A qualidade dos dados das várias variáveis recolhidas, no *dataset raw*, depende do

fator humano. O processo de integração de todos os dados numa única base de dados relacional, permitiu ter a noção dos problemas, e decidir as melhores técnicas de pré-processamento a aplicar sobre os dados existentes. Um dos receios incidia sobre a qualidade dos dados recolhidos através do {EDAMOVE} por causa da utilização dos elétrodos. No entanto, havia também preocupação com os dados recolhidos pelos questionários, nomeadamente pelo {LABOR}, cujo preenchimento, na maioria das vezes, foi efetuado por pessoas que mostravam indisponibilidade para o fazer de boa vontade. Além disso, desconhecia-se a quantidade de informação recolhida pelos vários instrumentos & sensores (e.g. desconhecia-se a quantidade e qualidade de dados recolhidos pelo {ANDROIDANALYTICS}, no que diz respeito aos dados relacionados com as chamadas telefónicas e mensagens escritas).

Considerando a qualidade pretendida para o *dataset* final, a fase de pré-processamento seria exigente e rigorosa. Tinha-se consciência que, só com um *dataset* de qualidade, se conseguiria promover a eficiência dos algoritmos e confiar nos resultados obtidos. Esta fase revestia-se de especial importância para o objetivo final pretendido: no futuro, utilizar o *dataset* como *input* para reconhecer automaticamente o bem-estar em trabalhadores de escritório.

A estratégia desenhada para a fase de pré-processamento, passava pela integração de todos os dados recolhidos de contexto numa base de dados relacional. A cada ficheiro componente do *dataset raw* corresponderia uma tabela na base de dados, que referenciaria cada registo com o respetivo participante e organização. Cada variável recolhida corresponderia a um atributo nessa tabela. Perante a possibilidade de ser necessário aplicar várias técnicas nas diversas variáveis de contexto recolhidas, documentou-se todas as técnicas de pré-processamento aplicadas sobre o *dataset*. Além disso, implementou-se uma estratégia de anotação para tornar mais fácil o processo de identificação dos dados impactados, e das técnicas aplicadas para mitigação dos problemas encontrados. A ideia passava por tornar possível a edição de dados, anotando as alterações, sem alterar os valores originais recolhidos do contexto. Assim, sobre a estrutura de tabelas da base de dados relacional, criou-se uma hierarquia de *views*. A cada nível de *view* corresponde um nível de abstração no pré-processamento. A *view* de nível mais baixo apresenta os dados *raw* pré-processados e anotados, e, no nível mais alto, são calculadas propriedades extraídas para cada variável. Além de se pretender aplicar técnicas de pré-processamento sobre os dados *raw* sem que tal significasse a alteração de dados originais, pretendia-se também que fosse possível executar consultas à base de dados relacional a qualquer nível de pré-processamento pretendido. A hierarquia de *views* é descrita com maior detalhe na secção [4.2.1](#).

Considerando a grande quantidade de dados recolhidos, encontrar problemas pelas vias habituais como pesquisa manual de valores, cálculo de medidas estatísticas, etc., seria um processo bastante demorado e exposto ao erro humano. Descobrir problemas nos

dados pode consumir muito tempo, e os métodos automatizados para cálculo e deteção podiam ajudar bastante na fase de pré-processamento (Kandel et al., 2012). O autor sentia a necessidade de agilizar este processo, utilizando um instrumento que promovesse a análise dos dados de forma mais célere e eficiente. Foi neste contexto que se desenvolveu o {DQVA}. Além de calcular as várias medidas estatísticas normalmente utilizadas (e.g. *average*, *median*, *InterQuartile Range (IQR)*, *skewness*), o {DQVA} também contém recursos para analisar os dados de forma visual: permite comparar participantes e organizações; verificar visualmente a existência de *outliers* e *gaps* de sinal; facilita a análise da distribuição dos dados através de histogramas; etc.

O {DQVA} permitiu analisar individualmente cada variável, com o objetivo de conhecer os dados por organização, participante ou conjunto de participantes. Os cálculos feitos pelo {DQVA} incidem sobre os valores integrados nas tabelas da base de dados. Desta forma, era mais fácil analisar os valores durante o processo de integração do *dataset raw*. Criou-se também uma variante do {DQVA} para analisar os dados de cada variável pré-processada. Com base no {DQVA-PP} foi possível analisar o comportamento dos dados, utilizando como origem as *views* implementadas na base de dados. A utilização do {DQVA} e do {DQVA-PP}, promoveu a qualidade dos dados do *dataset* final apresentado como contribuição principal desta tese no Capítulo 5.

Nesta secção descreve-se o pré-processamento efetuado sobre os dados *raw* recolhidos na experiência relatada no Capítulo 3. O único processamento sobre os dados que foi feito antes desta fase, foi o efetuado pelo {DATAANALYSER} que pré-processou os dados recolhidos do {EDAMOVE} com $SI = 1h$. De seguida, descreve-se o processo de integração de ficheiros e de outros dados (e.g. formulários em papel). Apresentam-se os instrumentos criados pelo autor, e descreve-se os problemas encontrados durante esta fase. Posteriormente apresentam-se as técnicas de pré-processamento utilizadas para mitigar os problemas nos dados, e descrevem-se as propriedades extraídas adicionadas ao *dataset* final, focando essencialmente as propriedades específicas de contexto. Por fim, apresenta-se o {DQVA} e a variante {DQVA-PP}. Conclui-se com uma análise final, onde são resumidas as contribuições feitas neste capítulo.

4.1 Integração do dataset

A estratégia para a fase de pré-processamento, passava pela integração de todos os dados que compunham o *dataset raw* recolhido na experiência de recolha apresentada no Capítulo 3, numa única base de dados relacional. Atendendo à quantidade de dados em causa, a utilização de um *DataBase Management System (DBMS)* facilitaria a organização dos dados, e a exploração de informação através da execução de *queries*. Além disso, tornar-se-ia possível implementar um mecanismo assente em *views*, que iria permitir a edição de dados e aplicar técnicas de pré-processamento, sem alteração dos va-

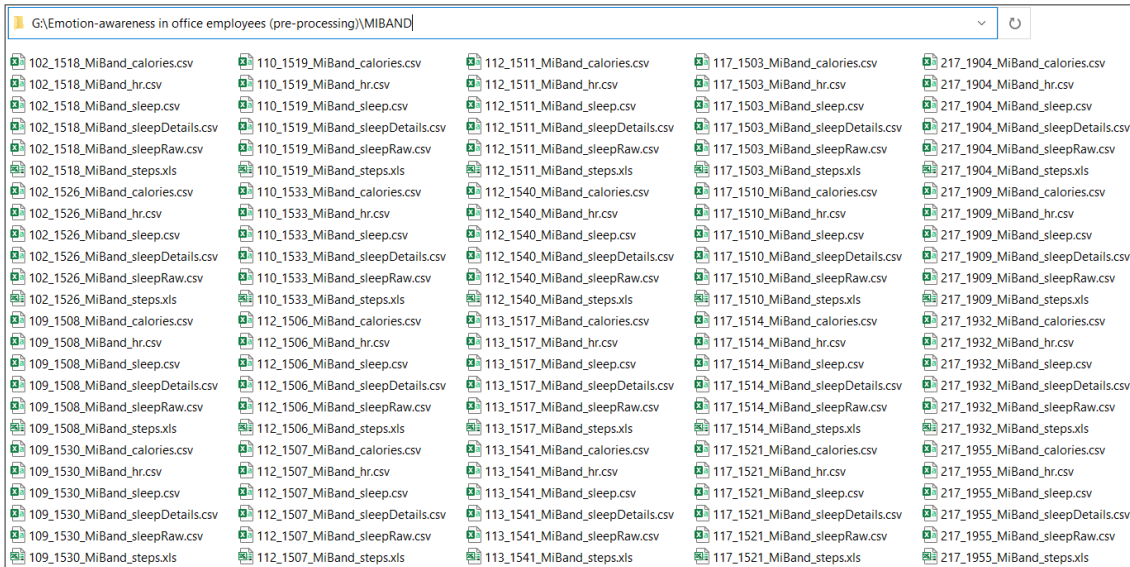


Figura 4.1: Pasta de ficheiros {MIBAND} após reorganização por instrumento de recolha

lores originalmente recolhidos. Nesta secção aborda-se a [INTEGRATION] do *dataset raw*, e os problemas encontrados durante o processo. Apresentam-se também as ferramentas desenvolvidas para *Extract Transform and Load* (ETL) dos dados.

O *dataset raw* a integrar era composto por vários ficheiros de dados *raw*. Estes ficheiros, gerados no processo de exportação de dados de cada instrumento no final de cada iteração de recolha (ver secção 3.2.2), foram arquivados nas pastas correspondentes a cada organização. Apesar de esta estrutura ser a mais natural para o arquivo de dados durante o fase de recolha, não era a estrutura mais adequada à execução da fase de pré-processamento. Assim, a integração de dados iniciou com uma reorganização dos ficheiros componentes do *dataset raw*. A nova estrutura dispunha os ficheiros por instrumento de recolha, em detrimento da estrutura original por organização (ver Figura 3.4). Esta reorganização pretendia facilitar a execução dos vários instrumentos ETL utilizados no pré-processamento, para integrar os dados na base de dados relacional. A título de exemplo, a Figura 4.1 apresenta a pasta de todos os ficheiros *raw* recolhidos pela {MIBAND} após reorganização.

Como já referido, a estratégia passava pela integração do *dataset raw* numa base de dados relacional. Utilizou-se o *Microsoft SQL Server 2017* {MSSQL} (SQL Server Team, 2017), mas poder-se-ia ter utilizado qualquer outra tecnologia de bases de dados relacionais, desde que permitisse implementar a hierarquia de *views* pretendida, e a criação de funções. O conteúdo de cada ficheiro do *dataset raw*, foi integrado numa tabela espelho com estrutura equivalente, no modelo relacional. Cada tabela tem uma chave primária composta pelo código da organização, número do participante, e um campo data com precisão igual ou superior ao milissegundo, para guardar a data e hora de cada introspeção de contexto efetuada pelo instrumento ou sensor (foi utilizado o tipo de dados *datetime2*

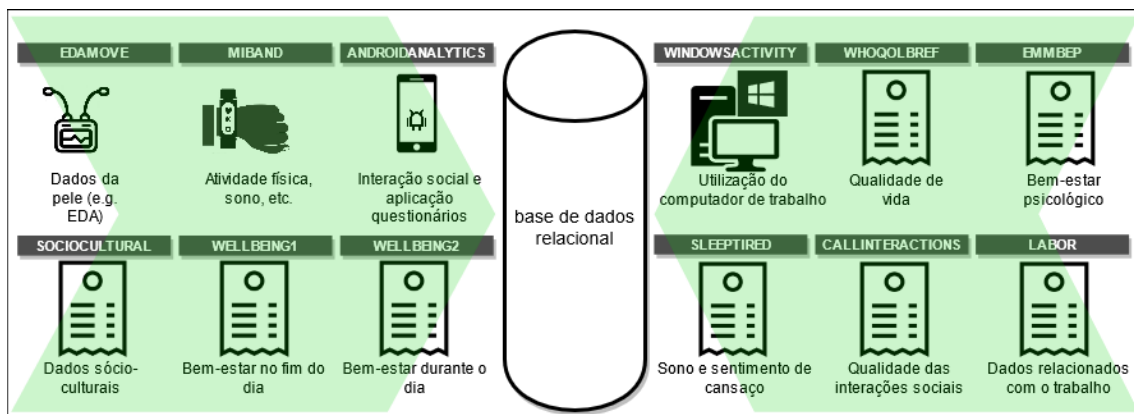


Figura 4.2: Dados a integrar na base de dados relacional

do {MSSQL} que tem precisão de 100 nanosegundos).

Durante a integração, alguns objetos em duplicado foram descartados para evitar a violação de chave primária na base de dados relacional. Este problema é apresentado na secção dos problemas encontrados durante a integração (ver secção 4.1.2).

De seguida descreve-se o processo de integração dos dados componentes do *dataset raw* apresentado no Capítulo 3, e os vários instrumentos utilizados no processo. Termina-se descrevendo-se os problemas encontrados no processo de integração.

4.1.1 Integração de dados *raw*

A fase de pré-processamento iniciou com a integração dos dados recolhidos do contexto. O *dataset raw* recolhido era composto por ficheiros com dados estruturados, e documentos em suporte papel. Grande parte da informação a integrar eram ficheiros com dados, mas também haviam alguns questionários e formulários *backup* que precisavam de ser inseridos na base de dados integradora. Nesta secção, apresentam-se os instrumentos e *scripts*, que foram utilizados na integração dos dados do *dataset raw*, na base de dados relacional integradora.

Como referido no Capítulo 3, durante a recolha de dados de contexto foram utilizados vários instrumentos & sensores. Os dados exportados de cada um desses instrumentos, foram guardados em ficheiros de dados (e.g. CSV). Os questionários e formulários *backup* preenchidos em papel, foram guardados também no *dataset raw*, em formato *Portable Document Format* (PDF). A Figura 4.2 resume os vários instrumentos & sensores origem dos dados integrados na base de dados relacional. Nesta fase de integração, utilizaram-se genericamente duas estratégias: i) integração dos ficheiros de dados através de instrumentos implementados para esse efeito; e ii) integração dos questionários e documentação em papel, através da criação de *scripts* em *Structured Query Language* (SQL). Estes *scripts* foram editados manualmente para inserção direta de dados nas tabelas da base de dados.

No processo de integração dos ficheiros na base de dados integradora utilizaram-se dois tipos de soluções: i) foram desenvolvidas aplicações para integração dos dados existentes em ficheiros *sqlite* (e.g. dados provenientes do {ANDROIDANALYTICS}); e ii) foi desenvolvido um projeto de ETL em *Microsoft SQL Server Integration Services* {MSSIS} (Microsoft, 2018), com o objetivo de carregar os ficheiros CSV, *Excel Spreadsheet file* (XLS) e *Excel Microsoft Office Open XML Format Spreadsheet* (XLSX), produzidos pelos vários instrumentos & sensores utilizados durante a fase de recolha (ver Tabela 4.1). Durante o processo de integração não existiu qualquer agregação nem pré-processamento de dados, com a exceção de alguns registos que existiam em duplicado no *dataset raw* e, por isso, foram descartados para evitar a violação de chaves primárias (ver 4.1.2).

No processo de exportação de dados de cada instrumento & sensor utilizado na fase de recolha, foram produzidos um ou mais ficheiros com dados *raw*. Cada um desses ficheiros foi integrado numa tabela espelho na base de dados relacional. A maioria dos ficheiros produzidos eram do tipo CSV e XLS?. No entanto, existiam também ficheiros *sqlite* para integrar. Estes ficheiros foram produzidos pelos instrumentos {WINDOWSACTIVITY} e {ANDROIDANALYTICS}). Como a ferramenta de ETL {MSSIS} não conseguia tratar corretamente os dados provenientes deste tipo de ficheiros, optou-se por criar duas pequenas aplicações para efetuar esta integração: i) o {WINDOWSACTIVITYIMPORT} importou os dados dos ficheiros *sqlite* produzidos pelo {WINDOWSACTIVITY}; e ii) o {ANDROIDANALYTICSIMPORT} importou os ficheiros do {ANDROIDANALYTICS}. Os restantes ficheiros foram integrados através do projeto ETL {MISCELLANEOUSIMPORT}. Neste projeto, implementado em {MSSIS}, foram definidos 9 *packages* (procedimentos ETL): i) o *package EDAMOVE_Results* carregou os dados dos ficheiros ???_???_EDAMove-Unisens/Results.xlsx existentes dentro da pasta do sensor {EDAMOVE}; ii) os *packages MIBAND_**, carregaram os dados provenientes dos ficheiros *MiBand_*.x* existentes dentro da pasta da pulseira {MIBAND}; e iii) o último *package* carregou os dados do ficheiro SocioCultural_v1.0a.xlsx existente dentro da pasta do {SOCIOCULTURAL}. O mapeamento dos dados *raw* com a base de dados integradora, é resumido através da Tabela 4.3).

A Tabela 4.1 resume os instrumentos ETL utilizados, e os outros instrumentos envolvidos no processo de integração de dados na base de dados relacional.

Instrumento	Descrição
<i>Instrumentos Extract Transform & Load (ETL)</i>	
{WINDOWSACTIVITYIMPORT}	O {WINDOWSACTIVITYIMPORT} foi desenvolvido pelo autor em C# para fazer a integração das bases de dados <i>sqlite</i> produzidas pelo {WINDOWSACTIVITY} dos vários participantes da experiência de recolha. A aplicação abre cada um dos ficheiros *WindowsActivity*.sqlite existentes na pasta do instrumento {WINDOWSACTIVITY}, e integra os seus dados na tabela WINDOWSACTIVITY_LogEntries do {MSSQL}.

{ANDROIDANALYTICS IMPORT}	O {ANDROIDANALYTICSIMPORT}, também desenvolvido pelo autor em C#, faz a integração dos ficheiros das bases de dados <i>sqlite</i> produzidas pelo {ANDROIDANALYTICS}. Cada ficheiro <i>sqlite</i> contém várias tabelas de dados: <i>calls</i> , <i>sleep_tired</i> , <i>sms</i> , <i>toogle_screen</i> , <i>well_being1</i> e <i>well_being2</i> . A aplicação abre todos os ficheiros *AndroidSocial*.sqlite existentes na pasta do instrumento {ANDROIDANALYTICS}, e integra o conteúdo das várias tabelas origem, nas tabelas espelho equivalentes do {MSSQL} (e.g. o conteúdo da tabela <i>calls</i> do {ANDROIDANALYTICS} é transferido para a tabela ANDROIDSOCIAL_Calls do {MSSQL}; o conteúdo da tabela <i>well_being1</i> é transferido para a tabela ANDROIDSOCIAL_WellBeing1).
{MISCELLANEOUS IMPORT}	O projeto ETL {MISCELLANEOUSIMPORT}, foi desenvolvido pelo autor em {MSSIS}, e faz a integração dos ficheiros provenientes dos instrumentos de recolha: {SOCIOCULTURAL}, {EDAMOVE} e {MIBAND}. A solução abre cada um dos ficheiros destes instrumentos de recolha, e transfere todo o seu conteúdo para as tabelas correspondentes no {MSSQL} (e.g. o conteúdo dos ficheiros *MiBand.hr*.csv é integrado no {MSSQL} na tabela MIBANDIBAND_HR; o conteúdo dos ficheiros *MiBand.steps*.xls é integrado na tabela MIBAND_Steps). O código da organização e o número do participante, são extraídos do nome dos vários ficheiros.
Outros instrumentos	
{MSSQL}	O <i>Microsoft SQL Server 2017</i> {MSSQL} foi o DBMS escolhido pelo autor para implementação da base de dados relacional. Nesta base de dados foram integrados todos os dados <i>raw</i> recolhidos na experiência de recolha em contexto laboral descrita no Capítulo 3. O {MSSQL} pode ser substituído por outro motor de bases de dados relacionais que permita a implementação de camadas de <i>views</i> . Se esse novo DBMS não permitir a criação de funções (retorno de valor escalar), as várias <i>views</i> terão que ser editadas para transformar em constantes os valores representativos da inexistência de valor (i.e. <i>NULL</i>), em vez das chamadas que são feitas às <i>Scalar-valued Functions</i> criadas pelo autor: <i>GV_DT()</i> , <i>GV_FLT()</i> , <i>GV_STR()</i> e <i>GV_T()</i> .
{MSSIS}	O <i>Microsoft SQL Server Integration Services</i> {MSSIS} foi a ferramenta ETL escolhida pelo autor para desenvolver os procedimentos de integração dos ficheiros CSV, XLS e XLSX, do <i>dataset raw</i> . Poderia utilizar-se qualquer outra ferramenta de ETL compatível com os ficheiros que se pretendiam carregar e que permitisse ligações de dados com o DBMS escolhido, neste caso o {MSSQL}.
{MSSMS}	O <i>Microsoft SQL Server Management Studio</i> {MSSMS}, é um ambiente integrado para gestão de {MSSQL}. O autor utilizou-o para criação do <i>schema</i> da base de dados (e.g. tabelas, <i>views</i> , <i>functions</i>), e para editar e executar os <i>scripts</i> para integração da documentação em suporte de papel. Pode ser substituído por qualquer outra ferramenta equivalente que permita ligação ao {MSSQL}.
{VS2017}	O <i>Microsoft Visual Studio 2017</i> {VS} foi o ambiente de desenvolvimento que facilitou a criação do {WINDOWSACTIVITYIMPORT} e {ANDROIDANALYTICSIMPORT} em linguagem C#.

Tabela 4.1: Instrumentos utilizados na fase de integração

Durante a experiência foram recolhidos vários dados de contexto em papel. Como já referido, estes dados foram inseridos na base de dados integradora através de *scripts* editados manualmente. Além dos questionários {WHOQOLBREF}, pretendia-se integrar também os dados do {LABOR}, e a restante documentação em papel que fazia parte do *dataset raw*. Os *scripts* SQL, criados para a inserção manual dos dados recolhidos pelos questionários

e formulários em papel, foram executados através do {MSSMS}. O autor preferiu a edição de *scripts* de integração em detrimento da inserção direta na base de dados por dois motivos: i) desta forma, posteriormente, seria possível integrar os dados com outro DBMS se necessário; e ii) durante a fase de integração agora apresentada, a integração destes dados poderia ser repetida várias vezes sem a preocupação com dados editados diretamente na base de dados. Os registos inseridos através dos *scripts*, foram diferenciados dos demais com a definição do valor `false` no atributo `AutomaticIntegration`. Este atributo existe em todas as tabelas espelho da base de dados integradora.

Os dados do {WHOQOLBREF} foram vertidos no *script* `WHOQOLBREF_integration.sql` que existe dentro da pasta do instrumento. Optou-se por integrar as respostas de todas as perguntas. Os resultados dos vários domínios do instrumento e classificação da qualidade de vida de cada participante, são calculados como propriedades extraídas específicas de contexto (ver secção 4.6), em *runtime* pelas *views* (Rimas, 2020).

O {LABOR} foi preenchido pelos departamentos de recursos humanos das organizações. Era um ficheiro Excel que pretendia recolher dados acerca das férias, faltas, atrasos e trabalho suplementar. Assim, para recolher cada uma destas informações, o respondente deveria preencher cinco folhas de cálculo dentro desse ficheiro: *faltas ao trabalho*, *atrasos ao trabalho*, *férias gozadas*, *férias não gozadas*, e *trabalho suplementar*. Considerando a baixa quantidade de dados a processar, preferiu-se criar *scripts* para inserção dos dados na base de dados, em vez de criar vários *packages* no {MISCELLANEOUSIMPORT} para a integração desta informação. Neste contexto, foram criados vários *scripts* para integração dos dados de cada uma das folhas pertencentes ao ficheiro `???.Labor_v1.0b.xlsx`. Estes ficheiros existem dentro da pasta de pré-processamento (ver Tabela 4.2).

Os participantes 1506, 1507, 1518, 1526 das organizações 112 e 102, registaram em papel o {WELLBEING1}, {WELLBEING2}, {SLEEPTIRED} e {CALLINTERACTIONS}. Não se conseguiu instalar e/ou executar com sucesso a aplicação {ANDROIDANALYTICS} no *smartphone* destes participantes. Assim, na indisponibilidade de recolher os dados com instrumento originalmente criado para o efeito, solicitou-se o preenchimento dos formulários *backup*. Estes dados recolhidos em papel foram inseridos na base de dados através de *scripts*. Os ficheiros *script* existem na pasta do instrumento (ver Tabela 4.2).

Também não foi possível implementar a recolha de dados de contexto com o instrumento {WINDOWSACTIVITY} nos participantes 1506, 1507 e 1909. As alterações feitas às *Group Policies* do SO dos participantes 1506 e 1507 não produziam efeito. Não era, assim, possível recolher os dados de utilização do computador tendo por base o sistema de registo de eventos do SO. No participante 1909 não se conseguiu que o administrador de domínio, em tempo útil, procedesse às configurações de políticas de registo de eventos necessárias para recolher dados do computador daquele utilizador. Nestes casos, recolheu-se dados relacionados com a utilização do computador, através de formulários *backup*. Os dados destes formulários foram inseridos na base de dados através do *script*

WindowsActivity_LogEntries_ManualRecords_integration.sql que existe na pasta do instrumento original de recolha.

A Tabela 4.2 resume os vários *scripts* criados para integração dos dados recolhidos em suporte papel.

Nome do <i>script</i>	Descrição
{WHOQOLBREF}	
WHOQOLBREF_integration.sql	O <i>script</i> inclui instruções de inserção das várias respostas de cada participante ao {WHOQOLBREF}. Os resultados por domínios e o resultado final da qualidade de vida por participante, são calculados pela <i>view</i> Participants_2_View. O ficheiro existe na pasta de pré-processamento do /WHOQOLBREF.
{LABOR}	
Labor_Vacations_ManualRecords_integration.sql	Contém dados relativos às férias gozadas pelos participantes. Foi criado tendo por base os dados provenientes da folha de cálculo <i>férias gozadas</i> existente no ficheiro ???_LABOR_v1.0b.xlsx. O <i>script</i> existe na pasta de pré-processamento /LABOR.
Labor_MissedVacations_ManualRecords_integration.sql	O <i>script</i> contém informação a férias não gozadas. Foi criado tendo por base os dados provenientes da folha de cálculo <i>férias não gozadas</i> existente no ficheiro ???_LABOR_v1.0b.xlsx. O <i>script</i> existe na pasta de pré-processamento /LABOR.
Labor_WorkAbsences_ManualRecords_integration.sql	O ficheiro <i>script</i> contém informação relativa a faltas. Foi criado tendo por base os dados provenientes da folha de cálculo <i>faltas ao trabalho</i> existente no ficheiro ???_LABOR_v1.0b.xlsx. O <i>script</i> existe na pasta de pré-processamento /LABOR.
Labor_WorkDelays_ManualRecords_integration.sql	O <i>script</i> contém informação relativa aos atrasos ocorridos ao trabalho. Foi criado tendo por base os dados provenientes da folha de cálculo <i>atrasos ao trabalho</i> existente no ficheiro ???_LABOR_v1.0b.xlsx. O <i>script</i> existe na pasta de pré-processamento /LABOR.
Labor_WorkLates_ManualRecords_integration.sql	Contém informação relativa ao trabalho suplementar de cada participante (i.e. horas extra). Foi criado tendo por base os dados provenientes da folha de cálculo <i>trabalho suplementar</i> existente no ficheiro ???_LABOR_v1.0b.xlsx. O <i>script</i> existe na pasta de pré-processamento /LABOR.
{ANDROIDANALYTICS}	
AndroidSocial_WellBeing1_PaperRecords_integration.sql	O <i>script</i> insere na base de dados, os dados relativos ao bem-estar recolhido através do instrumento <i>backup</i> do {WELLBEING1}. O <i>script</i> existe na pasta de pré-processamento /ANDROIDANALYTICS (instrumento original de recolha).
AndroidSocial_WellBeing2_PaperRecords_integration.sql	O ficheiro contém as instruções para inserir os dados relativos ao bem-estar recolhido através do instrumento <i>backup</i> do {WELLBEING2}. O <i>script</i> existe na pasta de pré-processamento /ANDROIDANALYTICS (instrumento original de recolha).
AndroidSocial_SleepTired_PaperRecords_integration.sql	O <i>script</i> insere na base de dados, os dados relativos ao sono e sentimento de cansaço, recolhidos através do instrumento <i>backup</i> do {SLEEPTIRED}. O <i>script</i> existe na pasta de pré-processamento /ANDROIDANALYTICS (instrumento original de recolha).
AndroidSocial_Calls_PaperRecords_integration.sql	O <i>script</i> insere na base de dados os dados recolhidos pelos formulários <i>backup</i> do {CALLINTERACTIONS}. O ficheiro existe também na pasta de pré-processamento /ANDROIDANALYTICS.

{WINDOWSACTIVITY}	
WindowsActivity_LogEntries _ManualRecords_integracion.sql	Este <i>script</i> contém informação relativa à utilização do computador, registada em formulários manuais em substituição da utilização do {WINDOWSACTIVITY}. O ficheiro existe também na pasta de pré-processamento /WINDOWSACTIVITY.

Tabela 4.2: *Scripts* criados para integração de informação recolhida manualmente

A Tabela 4.3 resume a proveniência dos dados integrados na base de dados relacional. São indicados os vários ficheiros origem de dados dos vários instrumentos & sensores utilizados na fase de recolha. Na tabela destino são apenas enumerados alguns atributos como exemplo. Os ficheiros `???_????_MiBand_hrFolhaCalculo.csv` não foram integrados por conterem dados redundantes com os ficheiros `102_1518_MiBand_hr.csv`. Cada pasta de pré-processamento *Unisens* do {EDAMOVE} contém 42 ficheiros gerados pelo {DATAANALYSER} no final de cada iteração de recolha, aquando da exportação de dados dos sensores. Estes ficheiros contém dados *raw* acerca dos eixos do acelerómetro, altitude, inclinação, aceleração, etc. Como não eram relevantes diretamente com o propósito de recolha, o autor optou por integrar apenas os dados do ficheiro `Results.xlsx`, porque resume os dados do (EDA) que se pretendia incluir no *dataset*. Os restantes dados do ficheiro não serão usados neste projeto. Só foram integrados os atributos do {WHOQOLBREF} que se considerou necessários para o fim pretendido para o *dataset* (e.g. o campo *Distrito* não foi integrado). Por fim, os dados respeitantes ao terceiro e quarto dependente recolhidos pelos {SOCIOCULTURAL} não foram integrados, porque não haver dados a integrar.

Ficheiro	Tabela destino (base de dados)	Instrumento (integração)		
{WINDOWSACTIVITY}				
???_????_WindowsActivity.sqlite				
— Contém informação relacionada com a utilização do computador de trabalho. Recolha: SO. Exportação: {WINDOWSACTIVITY}. Ficheiros: um por participante. Conteúdo de cada ficheiro: cada registo corresponde a um evento ocorrido no SO (e.g. <i>UserLogonSucess</i> , <i>WorkstationUnlocked</i>).				
	<table border="1"> <tr> <td>WINDOWSACTIVITY_LogEntries Atributos: { <i>UserName</i>, <i>Type</i> }</td> <td>{WINDOWSACTIVITYIMPORT}</td> </tr> </table>	WINDOWSACTIVITY_LogEntries Atributos: { <i>UserName</i> , <i>Type</i> }	{WINDOWSACTIVITYIMPORT}	
WINDOWSACTIVITY_LogEntries Atributos: { <i>UserName</i> , <i>Type</i> }	{WINDOWSACTIVITYIMPORT}			
WindowsActivity_LogEntries_ManualRecords_integracion.sql				
— Contém informação relacionada com a utilização do computador de trabalho. Recolha: formulário <i>backup</i> (papel). Exportação: N/A. Ficheiros: um ficheiro <i>script</i> contém os dados de todos os participantes. Conteúdo de cada ficheiro: cada registo corresponde a um evento ocorrido no SO.				
	<table border="1"> <tr> <td>WINDOWSACTIVITY_LogEntries</td> <td>(<i>Script</i>)</td> </tr> </table>	WINDOWSACTIVITY_LogEntries	(<i>Script</i>)	
WINDOWSACTIVITY_LogEntries	(<i>Script</i>)			
{ANDROIDANALYTICS}				
???_????_AndroidSocial.sqlite				

— Contém informação relacionada chamadas telefónicas, SMSs, dados dos questionários aplicados pelo *smartphone*, etc. **Recolha:** {`ANDROIDANALYTICS`}. **Exportação:** {`ANDROIDANALYTICS`}. **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada ficheiro contém as tabelas *calls*, *sleep_tired*, *sms*, *toogle_screen*, *well_being1* e *well_being2*.

ANDROIDSOCIAL.WellBeing1 Atributos: { <i>EasyRelations, DifficultSituations, TheyLikeMe, ProudMySelf, RightBalance, Stress-Begin, StressEnd</i> }	{ <code>ANDROIDANALYTICSIMPORT</code> }
ANDROIDSOCIAL.WellBeing2 Atributos: { <i>EmotionalyBalanced, PeaceWithMyself, ImConfident</i> }	{ <code>ANDROIDANALYTICSIMPORT</code> }
ANDROIDSOCIAL.SleepTired Atributos: { <i>SleepQuality, FeelingTired</i> }	{ <code>ANDROIDANALYTICSIMPORT</code> }
ANDROIDSOCIAL.Calls Atributos: { <i>InterlocutorType, Interaction, Duration</i> }	{ <code>ANDROIDANALYTICSIMPORT</code> }
ANDROIDSOCIAL.SMSs Atributos: { <i>InterlocutorType, Words, Chars</i> }	{ <code>ANDROIDANALYTICSIMPORT</code> }
ANDROIDSOCIAL.ToogleScreen Atributos: { <i>ScreenIsOn</i> }	{ <code>ANDROIDANALYTICSIMPORT</code> }

AndroidSocial.WellBeing1.PaperRecords.integration.sql

— Contém dados das respostas do {`WELLBEING1`} **Recolha:** formulário *backup* (papel). **Exportação:** N/A. **Ficheiros:** um ficheiro *script* contém os dados de todos os participantes. **Conteúdo de cada ficheiro:** cada registo corresponde a um resposta ao questionário.

ANDROIDSOCIAL.WellBeing1	(<i>Script</i>)
---------------------------------	-------------------

AndroidSocial.WellBeing2.PaperRecords.integration.sql

— Contém dados das respostas do {`WELLBEING2`} **Recolha:** formulário *backup* (papel). **Exportação:** N/A. **Ficheiros:** um ficheiro *script* contém os dados de todos os participantes. **Conteúdo de cada ficheiro:** cada registo corresponde a uma resposta ao questionário.

ANDROIDSOCIAL.WellBeing2	(<i>Script</i>)
---------------------------------	-------------------

AndroidSocial.SleepTired.PaperRecords.integration.sql

— Contém dados das respostas do {`SLEEPTIRED`} **Recolha:** formulário *backup* (papel). **Exportação:** N/A. **Ficheiros:** um ficheiro *script* contém os dados de todos os participantes. **Conteúdo de cada ficheiro:** cada registo corresponde a uma resposta ao questionário.

ANDROIDSOCIAL.SleepTired	(<i>Script</i>)
---------------------------------	-------------------

AndroidSocial.Calls.PaperRecords.integration.sql

— Contém dados das respostas do {`CALLINTERACTIONS`} **Recolha:** formulário *backup* (papel). **Exportação:** N/A. **Ficheiros:** um ficheiro *script* contém os dados de todos os participantes. **Conteúdo de cada ficheiro:** cada registo corresponde a uma resposta ao questionário.

ANDROIDSOCIAL.Calls	(<i>Script</i>)
----------------------------	-------------------

{MIBAND}

—**???_???_MiBand_calories.csv**

— Contém informação acerca do consumo diário de calorias e *Basal Metabolic Rate* (BMR) (*Harris-Benedict Equation*). **Recolha:** {MIBAND}. **Exportação:** {NOTIFYFITNESS}. **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada registo contém dados de um dia de recolha.

MIBAND_Calories Atributos: { BMR, Steps_Calories, Exercises_Calories }	{MISCELLANEOUSIMPORT} Package: MiBand_Calories
---	---

???_???_MiBand_hr.csv

— Contém a frequência cardíaca registada a cada minuto. **Recolha:** {MIBAND}. **Exportação:** {NOTIFYFITNESS}. **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada registo contém dados de um minuto de recolha.

MIBAND_HR Atributos: { HeartRate }	{MISCELLANEOUSIMPORT} Package: MiBand_HR
---	---

???_???_MiBand_sleep.csv

— Contém informação diária acerca do sono, e estatísticas da frequência cardíaca. **Recolha:** {MIBAND}. **Exportação:** {NOTIFYFITNESS}. **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada registo contém dados de um dia de recolha.

MIBAND_Sleep Atributos: { TotalSleep, DeepSleep, LightSleep, HR_Avg, HR_Max, HR_Min }	{MISCELLANEOUSIMPORT} Package: MiBand_Sleep
--	--

???_???_MiBand_sleepDetails.csv

— Cada registo corresponde a um período de sono do participante (cf. sono leve, sono profundo). **Recolha:** {MIBAND}. **Exportação:** {NOTIFYFITNESS}. **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada registo contém dados de um dia de recolha.

MIBAND_SleepDetails Atributos: { Type, Duration }	{MISCELLANEOUSIMPORT} Package: MiBand_SleepDetails
--	---

???_???_MiBand_sleepRaw.csv

— Contém informação *raw* para cálculo dos períodos de sono. **Recolha:** {MIBAND}. **Exportação:** {NOTIFYFITNESS}. **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada registo contém dados de um minuto de recolha.

MIBAND_SleepRaw Atributos: { Intensity, Steps, HR }	{MISCELLANEOUSIMPORT} Package: MiBand_SleepRaw
--	---

???_???_MiBand_steps.xls

— Contém informação acerca dos passos dados pelo participante. **Recolha:** {MIBAND}. **Exportação:** {NOTIFYFITNESS}. **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada registo contém dados de 10 minutos de recolha.

MIBAND_Steps Atributos: { <i>StepsDayAccumulation, Steps</i> }	{MISCELLANEOUSIMPORT} Package: MiBand_Steps
---	--

{EDAMOVE}

???_???_EDAMove-Unisens/Results.xlsx

— Contém informação acerca do (EDA) e outros dados agregados pela média com $SI = 1h$. **Recolha:** {EDAMOVE}. **Exportação:** {SENSORMANAGER} (pré-processamento feito no final de cada iteração de recolha através do {DATAANALYSER}). **Ficheiros:** um por participante. **Conteúdo de cada ficheiro:** cada registo contém dados de um hora de recolha.

EDAMOVE Results Atributos: { <i>EdaArousalMean, EdaSclMean, EdaScrAmplitudesMean, EdaScrCount, EdaScrEnergiesMean, EdaScrHalfRecoveryTimesMean, EdaScrRiseTimesMean</i> }	{MISCELLANEOUSIMPORT} Package: EDAMOVE_Results
--	---

{WHOQOLBREF}

WHOQOLBREF.integration.sql

— Contém as respostas às perguntas do questionário {WHOQOLBREF}. **Recolha:** questionário em papel. **Exportação:** N/A. **Ficheiros:** um com os dados de todos os participantes. **Conteúdo de cada ficheiro:** dados de todas as respostas do questionário.

WHOQOLBREF Atributos: { <i>Q01, Q02, Q..., Q26</i> }	(Script)
---	----------

{SOCIOCULTURAL}

SocioCultural.v1.0a.xlsx

— Contém as respostas às perguntas do questionário {SOCIOCULTURAL}. **Recolha:** questionário {GOOGLEFORMS}. **Exportação:** {GOOGLEFORMS}. **Ficheiros:** um com os dados de todos os participantes. **Conteúdo de cada ficheiro:** dados de todas as respostas do questionário.

SOCIOCULTURAL Atributos: { <i>Age, Gender, CivilStatus, Religion, Home_Type, HomeWork_Distance, ...</i> }	{MISCELLANEOUSIMPORT} Package: SOCIOCULTURAL
--	---

{LABOR}

Labor.Vacations_ManualRecords.integration.sql

— Contém informação acerca do gozo de férias de cada participante. **Recolha:** {LABOR}. **Exportação:** N/A. **Ficheiros:** um ficheiro *script* que contém os dados de todos os ficheiros {LABOR} de todas organizações. **Conteúdo de cada ficheiro:** dados de férias gozadas de todos os participantes.

LABOR_Vacations Atributos: { <i>LastDate, Type, Planned</i> }	(Script)
--	----------

Labor.MissedVacations_ManualRecords.integration.sql

— Contém informação acerca das férias não gozadas de cada participante. **Recolha:** {LABOR}. **Exportação:** N/A. **Ficheiros:** um ficheiro *script* que contém os dados de todos os ficheiros {LABOR} de todas organizações. **Conteúdo de cada ficheiro:** dados de férias não gozadas de todos os participantes.

LABOR_MissedVacations Atributos: { <i>LastDate, Type, Planned</i> }	(<i>Script</i>)
Labor.WorkAbsences_ManualRecords_integration.sql — Contém informação acerca das faltas de cada participante. Recolha: {LABOR}. Exportação: N/A. Ficheiros: um ficheiro <i>script</i> que contém os dados de todos os ficheiros {LABOR} de todas organizações. Conteúdo de cada ficheiro: dados de faltas de todos os participantes.	
LABOR_WorkAbsences Atributos: { <i>Type, NumberOfHours, Planned</i> }	(<i>Script</i>)
Labor.WorkDelays_ManualRecords_integration.sql — Contém informação acerca dos atrasos de cada participante. Recolha: {LABOR}. Exportação: N/A. Ficheiros: um ficheiro <i>script</i> que contém os dados de todos os ficheiros {LABOR} de todas organizações. Conteúdo de cada ficheiro: dados de atrasos de todos os participantes.	
LABOR_WorkDelays Atributos: { <i>Delay, Planned, Cause</i> }	(<i>Script</i>)
Labor.WorkLates_ManualRecords_integration.sql — Contém informação acerca do trabalho suplementar de cada participante. Recolha: {LABOR}. Exportação: N/A. Ficheiros: um ficheiro <i>script</i> que contém os dados de todos os ficheiros {LABOR} de todas organizações. Conteúdo de cada ficheiro: dados de trabalho suplementar de todos os participantes.	
LABOR_WorkLates Atributos: { <i>NumberOfHours, Planned</i> }	(<i>Script</i>)

Tabela 4.3: Mapeamento dos ficheiros com dados *raw* para a base de dados relacional

4.1.2 Problemas encontrados durante a integração

A qualidade dos dados do *dataset* foi uma preocupação desde o início da experiência relatada no Capítulo 3. Recolheu-se dados em ambiente real, utilizando vários instrumentos & sensores, e um deles era obstrutivo porque utilizava elétrodos para recolher dados de contexto. Durante as visitas de acompanhamento à recolha, as evidências confirmavam a probabilidade de existirem problemas nos dados a integrar. Nesta secção resumem-se os problemas encontrados durante a fase de integração do *dataset raw*, e as técnicas de pré-processamento aplicadas para os mitigar.

A preocupação com a qualidade dos dados recolhidos, surgiu ainda na fase de planeamento com a seleção do {EDAMOVE} para recolher dados de contexto em ambiente real. Tratava-se de um instrumento obstrutivo atendendo a que se utilizavam elétrodos colados na mão não-dominante para efetuar a recolha do (EDA). Além disso, iria recolher-se dados em ambiente real de escritório, utilizando vários instrumentos & sensores em si-

multâneo. Posteriormente, durante as visitas de acompanhamento na fase recolha, esse receio confirmou-se: i) houve a necessidade de reafixar os elétrodos do {EDAMOVE}; ii) teve que se ajustar as pulseiras {MIBAND} para garantir a proximidade suficiente à pele; iii) houve a necessidade de lembrar a resposta aos questionários aplicados pelo {ANDROIDANALYTICS}; iv) teve que se reconfigurar o bloqueio automático da sessão de trabalho em dois participantes; v) um desses participantes tinha também redefinido a senha de acesso para uma palavra simples, teve que ser reconfigurada novamente; vi) outro participante retirou a {MIBAND} durante algum tempo, porque não tinha percebido que era à prova de água; vii) tornava-se evidente a falta de boa vontade por parte do departamento de recursos humanos, para uma resposta cuidada ao {LABOR}; viii) com o passar do tempo de recolha, o cansaço dos participantes era visível; etc. Conheciam-se, portanto, a forte probabilidade de existirem problemas nos sinais recolhidos (e.g. *gaps*, *outliers*). Foi nesse contexto que se decidiu criar o {DQVA} apresentado na secção 4.4. Era necessária uma ferramenta que permitisse acompanhar o processo de integração de cada variável de contexto, para avaliar a qualidade dos dados integrados. A utilização dessa ferramenta tornaria mais simples e rápido o processo de integração, e promoveria a qualidade dos dados do *dataset* final.

Durante a fase de integração do *dataset raw*, detetaram-se genericamente três categorias distintas de problemas: i) existência de registos em duplicado dentro dos ficheiros a integrar, que provocavam violações de chaves primárias durante o processo de integração; ii) problemas no preenchimento de alguns campos (e.g. ausência de valores por falta de preenchimento, utilização incorreta de unidades de medida, utilização de caracteres incorretos); e iii) falta de consistência e coerência dos dados preenchidos pelos departamentos de recursos humanos no questionário {LABOR}. Para viabilizar o processo de integração, teve que se editar alguns dados dos ficheiros de origem. No entanto, os ficheiros que compõem o *dataset raw* recolhido no Capítulo 3, mantêm-se intactos desde a respetiva exportação no final de cada iteração de recolha.

O problema com maior número de ocorrências na fase de integração do *dataset raw*, foi a existência de registos em duplicado. Nestes casos, e após verificação de se tratar efetivamente de registos repetidos, aplicou-se a técnica [DISTINCTOBJ], considerando-se apenas a ocorrência mais antiga de cada conjunto de objetos repetidos. Uma das vantagens em utilizar uma base de dados relacional, é a imposição pelo DBMS da integridade referencial implementada pelo modelo Entidade Relação (ER). A unicidade de registos imposta pelas chaves primárias, garante a impossibilidade de repetição de registos. Neste contexto, a deteção de registos em duplicado nos ficheiros origem foi natural. Como na base de dados relacional integradora foi implementado um modelo ER que considerava como chave primária composta os atributos *OrganizationId* (chave estrangeira da tabela *Organizations*), *ParticipantId* (chave estrangeira da tabela *Participants*), e *Date* (data e hora com precisão superior ao milissegundo), a tentativa de inserção de

registos em duplicado durante o processo de integração, despoletava imediatamente um erro de violação de chave primária. Além disso, as *constraints* definidas nas relações entre tabelas, promove a consistência e fiabilidade dos dados guardados no modelo relacional. A tentativa de integração de códigos de organizações ou número de participantes inexistentes, era impedida pelo próprio DBMS que disparava um erro a informar a indicação de um valor inválido para chave estrangeira. Quando as ferramentas de ETL tentavam guardar um valor vazio (i.e. *NULL*), eram aplicadas as configurações definidas no modelo ER, impondo a indicação de um valor para os atributos *non-nullable*.

Menos frequentemente, houve a necessidade de preencher dados em falta (e.g. o participante 1521 não preencheu o campo idade no {WHOQOLBREF}), normalizar unidades de medida (e.g. a altura dos participantes 1503, 1517, 1519, 1904 e 1955, foi convertida para metros), etc. Só se efetuaram ajustes aos dados nos casos em que não existiam dúvidas tratar-se de um lapso no preenchimento pelos participantes, e quando o autor conhecia os valores corretos a preencher ou conseguiam deduzi-los a partir de outros (e.g. os participantes 1519 e 1533 identificaram-se com um valor incorreto na inicialização da aplicação {ANDROIDANALYTICS} e este valor foi retificado durante a integração, porque se tinha utilizado o número de participante no nome do ficheiro). Também surgiram alguns casos de registos incompletos. Nestes casos, os registos foram descartados (e.g. alguns registos do participante 1541 relacionados com o bloqueio e desbloqueio do *smartphone*, foram desconsiderados por estarem incompletos).

O último problema a relatar da fase da integração refere-se ao instrumento {LABOR}, mais concretamente à qualidade dos dados recolhidos. Nas várias visitas efetuadas às organizações para acompanhamento das iterações de recolha, foi possível constatar a falta de boa vontade do departamento de recursos humanos para responder de forma profissional às várias tabelas do ficheiro Excel. Este problema aconteceu em várias organizações e, em algumas delas, a devolução do {LABOR} preenchido só foi possível após a intervenção dos responsáveis. Neste contexto, o autor estava especialmente expectantes quanto à qualidade e quantidade dos dados recolhidos por este instrumento. Durante a integração constatou-se a parca informação prestada pelos departamentos de recursos humanos. Além disso, alguma informação era contraditória e inconsistente (e.g. registos demonstravam que o mesmo trabalhador estava de férias e a efetuar trabalho suplementar no mesmo dia). Assim, e na tentativa de não perder esta informação importante para o *dataset* com que se pretendia contribuir, solicitou-se a revisão do conteúdo do ficheiro a várias organizações. Esta revisão foi feita em conjunto com o autor, e os ficheiros {LABOR} do *dataset raw* revistos, foram substituídos pela nova versão.

Os dados do {LABOR} foram integrados na base de dados integradora. No entanto, mesmo após a revisão do conteúdo, receia-se pela validade dos dados recolhidos por este instrumento. Algumas retificações do {LABOR} foram feitas com base em suposições ou recurso à memória dos respondentes, dada a inexistência de outra fonte de informação

mais precisa. Aliás, durante esta revisão do {LABOR} na presença do autor, foi perceptível que nem sempre as organizações mantêm registos atualizados de controlo de presenças, férias, etc. Algumas permitem a substituição arbitrária de faltas por dias ou horas de férias, outras não mantêm registos dos atrasos ocorridos por considerarem que o benefício não justifica o trabalho, outras ainda, mantêm a prática de trabalho suplementar sem qualquer registo oficial, etc.

A Tabela 4.4 resume alguns problemas encontrados durante a integração em cada instrumento. São apenas descritas as técnicas de pré-processamento utilizadas na resolução de problemas durante a fase de integração do *dataset*. As restantes técnicas aplicadas nesta fase, serão abordadas na secção 4.2.

Instrumento	Dados impactados pelas técnicas aplicadas
{WINDOWS ACTIVITY}	<p>???_????_WindowsActivity.sqlite</p> <ul style="list-style-type: none"> • O SO do participante 1541 registou os eventos <i>UserLogonSucess</i> e <i>WorkstationUnlocked</i>, no mesmo instante. Atendendo a que não se tratava do mesmo evento, aumentou-se em um milissegundo à hora do evento <i>WorkstationUnlocked</i>, com o objetivo de evitar o erro de violação de chave primária — [MANADJUST].
{ANDROID ANALYTICS}	<p>???_????_AndroidSocial.sqlite</p> <ul style="list-style-type: none"> • Os participantes 1519 e 1533, inseriram erradamente o código da organização e/ou o número de participante, aquando da primeira utilização do {ANDROIDANALYTICS}. Os dados extraídos do nome do ficheiro no processo de ETL, foram confrontados com os dados guardados no conteúdo de cada ficheiro. No caso de falha, os dados foram corrigidos manualmente — [CONSISTENCY, MANDADJUST]. • Os participantes 1519, 1521, 1541, 1904, 1909 e 1932, tinham vários registos em duplicado na tabela <i>calls</i>. Como se tratavam de objetos repetidos, considerou-se apenas o primeiro de cada participante — [DISTINCTOBJ]. • O participante 1541, tinha registos repetidos na tabela <i>sms</i>. Foi considerado apenas o objeto mais antigo — [DISTINCTOBJ]. • O participante 1510, tinha duas respostas ao questionário {SLEEPTIRED} no mesmo dia (tabela <i>sleep_tired</i>). Foi considerada apenas a primeira resposta — [DISTINCTOBJ]. • Os participantes 1511, 1519, 1514 e 1904, tinham registos em duplicado na tabela <i>toogle_screen</i>. Foi considerado o objeto mais antigo de cada conjunto repetido — [DISTINCTOBJ]. • O participante 1541 tinha objetos com dados incompletos na tabela <i>toogle_screen</i>. Os registos foram descartados — [DISCARDDATA].

{MIBAND}	<p>???_????_MiBand_calories.csv</p> <ul style="list-style-type: none"> Os registos do ficheiro origem terminavam com um ponto e vírgula (;), indicando erradamente a existência de um último atributo final. Estes caracteres foram eliminados manualmente. <p>???_????_MiBand_hr.csv</p> <ul style="list-style-type: none"> O participante 1503 apresentava registos em duplicado. Foi utilizado apenas o objeto mais antigo a cada conjunto de repetições — [DISTINCTOBJ]. <p>???_????_MiBand_steps.csv</p> <ul style="list-style-type: none"> Os passos registados pela {MIBAND} é um contador acumulador diário. No entanto, como são disponibilizados valores parciais desse acumulador a cada 10 minutos, preferiu-se calcular e registar o valor do aumento a cada objeto integrado. O objetivo era facilitar os posteriores processos de agregação — [NORM]. <p>???_????_MiBand_sleepRaw.csv</p> <ul style="list-style-type: none"> As datas <i>POSIX</i> deste ficheiro foram exportadas pelo {NOTIFYFITNESS} com menos três dígitos à direita. Implementou-se uma tarefa em ETL para a correta conversão para o tipo de dados <i>datetime2</i> do {MSSQL} — [NORM].
{WHOQOLBREF}	<p>WHOQOLBREF_integration.sql</p> <ul style="list-style-type: none"> O participante 1514 não preencheu todos os campos de informação geral do {WHOQOLBREF}. Como se tinha recolhido esta informação também com o {SOCIOCULTURAL} preencheu-se os dados em falta no <i>script</i> de integração — [MANINSERT]. Os participantes 1506 e 1507 são programadores mas não indicaram a profissão. Como a profissão é conhecida do autor, foi manualmente corrigida para <i>IT</i> — [MANINSERT].
{SOCIO CULTURAL}	<p>SocioCultural_v1.0a.xlsx</p> <ul style="list-style-type: none"> O participante 1519, na pergunta "<i>Número de empregos desde o início da vida ativa</i>", respondeu 18 porque entendeu como anos de trabalho (o participante em causa tinha 35 anos) — [RELIABILITY, MANADJUST]. O participante 1521 não preencheu a idade. Foi calculada com base na data de nascimento inserida no {WHOQOLBREF} — [MANINSERT]. Os participantes 1508, 1511 e 1533 registaram um peso ligeiramente diferente do que tinham informado no preenchimento do {SUBJECTFILE}. Os valores do {SOCIOCULTURAL} foram retificados para os indicados no {SUBJECTFILE}, porque foram estes os utilizados no processamento do {DATANALYSER} para geração dos ficheiros <i>raw</i> do {EDAMOVE} — [RELIABILITY, CONSISTENCY, MANADJUST]. Os participantes 1503, 1517, 1519, 1904 e 1955, inseriram a altura em centímetros e não em metros. Como preencheram o {SOCIOCULTURAL} no <i>smartphone</i>, terão tido dificuldade em inserir o dígito separador decimal. Os valores foram convertidos para metros — [NORM]. A distância de casa ao trabalho em quilómetros registada pelo participante 1514 foi arredondada às unidades — [NORM].

{LABOR}	<p>Labor_Vacations_ManualRecords_integration.sql</p> <p>Labor_MissedVacations_ManualRecords_integration.sql</p> <p>Labor_WorkAbsences_ManualRecords_integration.sql</p> <p>Labor_WorkDelays_ManualRecords_integration.sql</p> <p>Labor_WorkLates_ManualRecords_integration.sql</p> <ul style="list-style-type: none"> • Durante a análise de qualidade de dados do {LABOR}, detetaram-se problemas de consistência e coerência de dados. Com o objetivo de mitigar o problema, solicitou-se a revisão dos dados respondidos na tentativa de lhes dar mais consistência — [RELIABILITY, CONSISTENCY, MITIGATION].
---------	--

Tabela 4.4: Problemas encontrados durante a fase de integração

4.2 Técnicas de pré-processamento utilizadas

As técnicas de pré-processamento visam mitigar o impacto das anomalias dos *datasets* no processamento algorítmico. Os investigadores dedicam muito tempo na preparação dos dados com o objetivo de melhorar a sua qualidade porque, só com um *dataset* de qualidade, podem obter bons resultados dos algoritmos de classificação (Urges, 2021; Gama et al., 2012). Nesta secção descrevem-se as técnicas de pré-processamento utilizadas para melhorar a qualidade do *dataset* após a fase de integração.

A fase de pré-processamento foi longa e exigente. Iniciou com a [INTEGRATION] do *dataset raw* recolhido (ver secção 4.1) e terminou com a disponibilização dos dados pré-processados, sobre os quais foram extraídas as propriedades descritas na secção 4.3. Desde muito cedo havia a perceção que poderiam existir problemas nos dados do *dataset raw*. Decidiu-se recolher dados em ambiente real, utilizando vários instrumentos & sensores, e um deles recolheria dados através de eléctrodos colados na mão dos participantes. Durante as visitas de acompanhamento efetuadas nas várias iterações de recolha, detetaram-se problemas na utilização dos sensores, reforçando a possibilidade de existirem anomalias nos dados. Durante a integração dos dados, ficou claro que seria necessário avaliar a qualidade de dados de cada sinal de cada participante e, foi neste contexto, que se desenvolveu o {DQVA}. O {DQVA} tornou mais simples e rápido o processo de análise de cada sinal. Este instrumento é apresentado em detalhe na secção 4.4.

Foi na fase de preparação da experiência de recolha do Capítulo 3 que se decidiu recolher dados em ambiente real. Muitas investigações do AC preferiram recolher dados em ambiente controlado, talvez por ser mais fácil promover o sucesso do processo de recolha. No entanto, apesar de se conseguir recolher um *dataset* com menos problemas, talvez os dados não representem tão realisticamente o contexto de um ambiente real de escritório. Será difícil simular em laboratório o ritmo e stress resultante da azáfama de tarefas e interações sociais, existentes em simultâneo num escritório real. Foi neste contexto que

se preferiu de conduzir a experiência de recolha em laboratório. No entanto, a decisão de recolher em ambiente real tem consequências, sendo uma das mais importantes, o possível impacto negativo na qualidade dos dados recolhidos. Durante a integração do *dataset* constatou-se a existência de problemas nos dados. Alguns sinais precisariam de ser reparados porque apresentavam *gaps* provocados por interrupções momentâneas de recolha. Sem surpresas, o sinal recolhido através de elétrodos era aquele que requeria mais atenção. Eram visíveis as intermitências do sinal, e teriam que ser aplicadas técnicas para a reconstrução dos dados em falta.

Inicia-se esta secção com uma descrição da estratégia seguida para a fase de pré-processamento, e conclui-se resumindo as técnicas de pré-processamento utilizadas e alguns casos do *dataset* onde foram aplicadas.

4.2.1 Estratégia utilizada para manutenção de dados

A fase de pré-processamento iniciou com a integração de dados descrita na secção 4.1. Os vários ficheiros componentes do *dataset raw* resultantes da experiência de recolha descrita no Capítulo 3, foram integrados numa base de dados relacional, e os dados integrados ficaram disponíveis para a aplicação de outras técnicas de pré-processamento, tendo em vista a promoção da sua qualidade. Nesta secção descreve-se a estratégia seguida para o pré-processamento dos dados, na fase posterior à integração.

A estratégia desenhada para a fase de pré-processamento, iniciada com a [INTEGRATION] apresentada na secção 4.1, tinha como objetivo principal dar qualidade ao *dataset* final, contribuição principal nesta tese. Durante a integração, apenas foram aplicadas as técnicas necessárias para viabilizar o processo. Os dados foram integrados numa base de dados relacional {MSSQL} e, cada ficheiro componente do *dataset raw* foi integrado numa tabela espelho de estrutura equivalente. Sabia-se que existiriam anomalias que poderiam existir nos dados recolhidos e, por esse motivo, decidiu-se criar o {DQVA} para promover a avaliação mais rápida da qualidade dos dados *raw* durante a integração. Posteriormente, criou-se também a variante {DQVA-PP} para análise da qualidade dos dados pré-processados. O {DQVA} e o {DQVA-PP}, ajudaram a fase de pré-processamento, porque permitiram avaliar a qualidade dos dados de forma mais simples e rápida (ver secção 4.4).

Como já referido, durante a integração constatou-se a existência de várias anomalias nos dados que teriam que ser mitigadas no pré-processamento. Considerando a quantidade de dados existente no *dataset* e as várias técnicas que poderiam ter que ser aplicadas, decidiu-se definir uma estratégia para permitir a edição de dados sem alteração dos valores originais, com o objetivo de melhor documentar todas as alterações ocorridas no *dataset*. Neste contexto, decidiu-se implementar alterações no modelo ER da base de dados, para anotar todas as técnicas utilizadas referenciando os dados impactados. O objetivo passava por contribuir com um *dataset* integrado, pré-processado e anotado. Assim,

para cada atributo (*var*), representativo de uma variável de contexto recolhida existente em tabelas espelho na base de dados integradora, implementaram-se dois atributos adicionais: i) o atributo *var_EditedValue*, por defeito vazio (i.e. *NULL*), onde foi inserido o valor editado, nos casos em que existiu pré-processamento sobre o conteúdo de *var*; e ii) o atributo *var_AnnotationId*, onde se guardou o identificador correspondente à anotação que melhor descrevia a técnica aplicada ao valor do atributo *var*. O atributo *var* contém os dados originalmente recolhidos de contexto. A estrutura do *dataset*, é analisado com maior detalhe na secção 5.1.

Sobre a estrutura de tabelas da base de dados {*MSSQL*}, desenhou-se uma arquitetura de *views* composta por vários níveis. No nível 1 de *views* são apresentados todos os objetos existentes na tabela, incluindo os inativados. Neste nível, os atributos *var* são substituídos pelos *var_EditedValue*, caso tenha ocorrido algum processamento sobre o valor original. Este nível de *views*, permite abstrair as camadas de nível superior acerca da origem dos valores que irão consumir. O nível 2 de *views* apresenta apenas os objetos ativos das *views* homólogas do nível 1. Neste nível, os objetos são agregados e as anotações não são apresentadas por serem desnecessárias para os níveis superiores. Ainda no nível 2, é imposta a existência de um registo a cada *SI*, através da junção à esquerda com a *view* *_MandatorySignalPKsView*, que gera artificialmente combinações de *Date*, *OrganizationId* e *ParticipantId*, para cada *SI = 1h*. O objetivo é forçar existência de *gaps* nos *recordsets* extraídos a partir desta *view* quando, para uma determinada hora, não existe valor. A representação destes *gaps*, pode ser personalizada através das *Scalar-valued Functions* *GV_DT()*, *GV_FLT()*, *GV_STR()* e *GV_T()*, implementadas pelo autor no {*MSSQL*}. A agregação é feita através da soma de valores aditivos, e média dos não aditivos. Os valores categóricos são representados pela moda. No nível 3 foi criada uma *view* para cada variável recolhida. Cada uma destas *views* calcula várias medidas estatísticas da variável respetiva. Neste nível são também calculadas as *mild* e *extreme outlier fences* para a variável. As medidas estatísticas calculadas serão apresentadas na secção propriedades extraídas. O último nível é uma *view* global que apresenta todos os dados e propriedades extraídas das *views* de nível 3, com o objetivo de facilitar a exploração de dados multi-atributo e/ou propriedade extraída. A arquitetura de *views* é abordada com maior pormenor na secção 5.2.

A extração de dados pode incidir sobre qualquer nível de abstração. Caso se pretenda aceder aos dados originais importados dos ficheiros do *dataset raw*, as *queries* devem referenciar as tabelas do modelo relacional. Caso se pretenda aceder aos dados pré-processados e analisar a respetiva anotação que descreve a técnica aplicada, a extração deve incidir sobre as *views* de nível 1. Se a ideia é analisar apenas os objetos ativos agregados com *SI = 1h*, os dados devem ser extraídos a partir do nível 2. Se, por outro lado, se pretende aceder às propriedades extraídas de cada variável, deve consumir-se as *views* de nível 3. A *view* de nível 4, permite explorar, simultaneamente, qualquer conjunto de atri-

butos ou propriedades extraídas sem necessidade de haver a preocupação com a junção de tabelas ou *views*. Contudo, todas as *views* contêm os atributos necessários para permitir a junção através do código da organização, número do participante, e um campo data (i.e. atributos da chave primária composta). Assim, é possível executar *queries* que permitem a extração de dados de tabelas e/ou *views* dos vários níveis, em separado ou em conjunto, utilizando junções.

As técnicas aplicadas sobre o *dataset*, cujo resultado pode ser analisado consumindo dados a partir das *views* de nível 1 e 2, serão apresentadas na secção 4.2.2. As propriedades extraídas calculadas pelas *views* de nível 3 são apresentadas na secção 4.3. Por fim, o *dataset* final composto pelos dados pré-processados e propriedades extraídas, obtidos pela consulta da *view* de nível 4, será detalhado no Capítulo 5.

A tabela *Organizations* e *Participants*, foi preenchida manualmente. Durante a edição dos *scripts* para inserção dos dados recolhidos em papel (ver secção 4.1.1), utilizou-se a técnica [NULL] para representar valores em falta ou ilegíveis.

A arquitetura utilizada no desenho e construção da base de dados não foi otimizada. As questões de desempenho e gestão de recursos, não eram o foco neste trabalho. O autor instrumentalizou o {MSSQL}, apenas para conseguir organizar os dados, com o objetivo de facilitar a sua extração e análise, em função do nível de abstração (*view*) pretendido.

4.2.2 Técnicas aplicadas sobre os dados

Definida a estratégia a seguir para aumentar a qualidade dos dados, iniciou-se a avaliação individual de cada sinal integrado na base de dados. Nesta secção apresentam-se as técnicas de pré-processamento utilizadas para melhorar a qualidade dos dados integrados.

A fase de pré-processamento demorou vários meses. Iniciou com a integração e com a criação dos instrumentos de suporte (ver secção 4.1), e terminou com a disponibilização do *dataset* final integrado e anotado que será apresentado como contribuição no Capítulo 5. O autor pretendia contribuir com um *dataset* de qualidade. Atendendo à quantidade de dados e aos problemas que existiriam para resolver, decidiu criar uma ferramenta que facilitasse o processo de avaliação da qualidade dos dados. A deteção de anomalias e a avaliação da qualidade dos dados são temas muito presentes na literatura. Existem vários trabalhos relacionados com o tema de limpeza de dados: i) Chandola et al. propuseram um agrupamento de técnicas enquanto resumem a literatura relacionada com a deteção de problemas nos dados (Chandola et al., 2009; Prasad et al., 2009); ii) Elmagarmid et al. abordaram técnicas para deteção de registos duplicados, apresentando alguns algoritmos utilizados para este fim (Elmagarmid et al., 2007); iii) Huhtala et al. estudaram as dependências das relações entre os dados e apresentaram um algoritmo para detetar dependências em bases de dados (Huhtala et al., 1999); iv) Hellerstein resume um con-

junto de métodos possíveis de utilizar no tratamento de erros em atributos qualitativos em grandes bases de dados (Hellerstein, 2008); v) Hodge et al. apresentaram um resumo de técnicas para a deteção de *outliers* (Hodge and Austin, 2004); vi) Tsay estudou a deteção de *outliers* e a variação de dados em séries temporais, aplicando métodos simples em vários *datasets* reais (Tsay, 1988); vii) Sukhobok et al. sequenciaram um conjunto de operações de limpeza e de transformação numa aplicação *web* (Sukhobok et al., 2018), etc. O {DQVA} e {DQVA-PP} permitiram detetar as anomalias dos dados de forma mais simples e rápida. Após selecionada a variável de contexto pretendida, eram calculadas várias medidas estatísticas usualmente utilizadas na fase exploratória dos dados. Além disso era apresentado um conjunto de visualizações que permitiam, de forma mais fácil, analisar o conteúdo da variável e encontrar problemas nos dados (Carvalho et al., 2021). Acredita-se ter contribuído para a qualidade do *dataset* final com a utilização destas ferramentas. Atendendo ao tempo que iria demorar a fase de pré-processamento, a exposição ao erro humano era um fator adicional de preocupação. A utilização de uma ferramenta que tornasse mais simples e rápido o processo de avaliação da qualidade dos dados, uniformizaria a utilização das várias técnicas aplicadas, e diminuiria a exposição ao erro. O {DQVA} e {DQVA-PP} são apresentados na secção 4.4.

Durante a fase de integração, constatou-se a existência de várias anomalias nos dados. Os dados de contexto que mais preocupavam, eram os que tinham sido recolhidos pelo sensor {EDAMOVE} e pelo (PPG) da {MIBAND}. Tinha ficado claro durante as visitas de acompanhamento, que existiriam falhas no sinal recolhido por estes dois sensores. Foram várias as vezes que houve a necessidade de reafixar elétrodos à mão dos participantes, e reajustar a largura da pulseira {MIBAND} para garantir a correta leitura do (PPG). O autor tinha também alguma expectativa relativamente aos dados relacionados questionários aplicados pelo {ANDROIDANALYTICS}. Durante a exploração de dados percebeu-se também que alguns participantes poderiam ter falhado na resposta a alguns questionários. Neste contexto, uma das técnicas mais importantes utilizadas na fase de pré-processamento, foi a interpolação de sinal [INTERPOLATION]. Salienta-se a utilização desta técnica de pré-processamento, porque terá sido aquela em que mais tempo investiu.

O autor utilizou sempre o bom senso na aplicação da técnica [INTERPOLATION]. Pretendia-se dar qualidade aos dados, sem afetar a sua validade. Cada sinal de cada participante foi analisado individualmente. Caso existissem falhas de sinal em valores reais, os objetos saudáveis à volta da zona com problemas eram considerados na solução. Se o problema se limitava até o máximo de dois objetos, os dados eram reconstruídos com base na mediana dos valores dos dois objetos anteriores e dois objetos posteriores. Caso contrário, os objetos com anomalia eram descartados. Na interpolação de valores inteiros utilizou-se o mesmo algoritmo, e o resultado da mediana foi arredondado à unidade. Nos questionários apenas se utilizou a interpolação nos casos em que faltava apenas um dia de resposta. As respostas desse dia foram reconstruídas com as respostas do dia anterior.

Nos casos em que o dia em falta era o primeiro, utilizou-se as respostas do dia seguinte. Como o fim de cada iteração de recolha acontecia às 17:30, o questionário {WELLBEING1} não foi respondido pelos participantes, atendendo a que apenas o iriam fazer à noite antes de deitar. Nestes casos, utilizaram-se os valores do dia anterior.

O autor utilizou também a técnica [-OUTLIER] em alguns objetos recolhidos pelo {EDAMOVE} e pela {MIBAND}. Utilizou-se a estratégia já apresentada para a falha de sinal de valores reais. Os valores *outlier* foram substituídos até ao máximo de dois objetos seguidos. Nos restantes casos, os valores foram descartados.

Houve também a necessidade de classificar ou reclassificar [MANCLASSIFICATION] valores durante o pré-processamento. Preferiu-se utilizar textos em Inglês representativos dos dados em vez de valores (e.g. inteiros), com o objetivo de facilitar a leitura do *dataset* (e.g. as respostas dos questionários aplicados pelo {ANDROIDANALYSTICS} foram guardadas com os valores { 1, 2, 3 }, e foram reclassificadas para { *I disagree, Neither disagree nor agree, I agree* }; no sinal (CALL) do {ANDROIDANALYSTICS}, preferiu-se identificar o *Type* com os termos { *MissedCall, InitiatedOnly, Incoming, Outgoing* } do que { 1, 2, 3, 4 }). O processo de anonimização do número de telefone dos interlocutores das (CALL) e (SMS), foi incluído na técnica [MANCLASSIFICATION].

Nem sempre houve a disponibilidade para estar presente no início de cada iteração de recolha. Por esse motivo, várias vezes teve que instalar e disponibilizar os instrumentos & sensores aos participantes no dia anterior ao início de recolha. De igual modo, nem sempre foi possível estar presente no momento final. Neste contexto, houve a necessidade de aplicar a técnica [DISCARDATA] aos dados anteriores ao momento inicial de recolha e posteriores ao final.

A Tabela 4.5 resume algumas das técnicas utilizadas na fase de pré-processamento, e apresenta alguns exemplos do *dataset* onde foram aplicadas. As anotações de *dataset* fruto da aplicação de técnicas de pré-processamento, são indicadas nos textos resumo da tabela, se existirem. Os registos (CALL) e (SMS) referentes a publicidade não foram descartados. Considerou-se que estas interações sociais fazem parte das rotinas dos trabalhadores de escritório e, por esse motivo, podem influenciar o stress e o bem-estar das pessoas.

Técnica	Dados impactados pelas técnicas aplicadas
[CONSISTENCY]	<div style="text-align: right;">{WINDOWSACTIVITY}</div> <ul style="list-style-type: none"> Durante a integração, foi verificada a consistência entre os códigos das organizações e número dos participantes, existente dentro dos ficheiros ???_???_WindowsActivity.sqlite e os extraídos do nome de cada ficheiro.

[MANADJUST]	<ul style="list-style-type: none"> • O SO do participante 1541 registou os eventos <i>UserLogonSucess</i> e <i>WorkstationUnlocked</i>, no mesmo instante temporal. Editou-se o evento <i>WorkstationUnlocked</i> para um milissegundo mais tarde para evitar a violação de chave primária. • Por lapso, os ficheiros <i>sqlite</i> do {WINDOWSACTIVITY} registaram o nome e não o código das organizações. Estes valores foram anonimizados. ¹
[MANINSERT]	<ul style="list-style-type: none"> • Os registos de utilização do computador dos participantes 1506, 1507 e 1909, foram recolhidos em papel. Posteriormente foram integrados através de um <i>script</i>.
{ANDROIDANALYTICS}	
[CONSISTENCY]	<ul style="list-style-type: none"> • A consistência entre os códigos das organizações e dos participantes existentes nos ficheiros ???_???_AndroidSocial.sqlite, e os retirados de cada nome do ficheiro, foi verificada durante a integração.
[DISCARDATA]	<ul style="list-style-type: none"> • O participante 1508 tinha vários registos incompletos e iguais na tabela <i>calls</i>. Não era possível reconstruir os dados em falta. • O participante 1541 tinha objetos com dados incompletos na tabela <i>toogle_screen</i>. Não era possível reconstruir os dados em falta. • Alguns participantes responderam aos questionários antes do início do período de recolha (e.g. participante 1519, 1533, 1541). Annotation: DATA_OUTSIDE_COLLECTION_INTERVAL.
[DISTINCTOBJ]	<ul style="list-style-type: none"> • Os participantes 1519, 1521, 1541, 1904, 1909 e 1932, tinham registos em duplicado na tabela <i>calls</i>. Considerou-se apenas o primeiro de cada conjunto de repetidos. • Também existiam registos em duplicado na tabela <i>sms</i> (e.g. participante 1541). Considerou-se apenas o primeiro registo de cada conjunto de duplicados. • O participante 1510 tinha duas respostas ao questionário {SLEEPTIRED} no mesmo dia (tabela <i>sleep_tired</i>). Foi considerada apenas a primeira resposta ao questionário. • Os participantes 1511, 1519, 1514 e 1904, tinham registos em duplicado na tabela <i>toogle_screen</i>. Foi considerado o objeto mais antigo de cada conjunto repetido. • A aplicação do participante 1514 registou várias respostas ao {WELLBEING2}. Os participantes poderiam responder várias vezes a este questionário durante o dia. No entanto, como os registos tinham apenas alguns segundos de diferença, foi considerada apenas o primeira resposta de cada conjunto. Annotation: TOO_MANY_ANSWERS.
[INTERPOLATION]	<ul style="list-style-type: none"> • O participante 1521 não respondeu ao questionário {SLEEPTIRED} do segundo dia de recolha. Annotation: FORGETFULNESS_OF_THE_PARTICIPANT. • o participante 1508 não respondeu ao questionário {WELLBEING1} no primeiro dia de recolha. Annotation: FORGETFULNESS_OF_THE_PARTICIPANT. • o participante 1518 não respondeu ao questionário {SLEEPTIRED} no último dia de recolha. Annotation: FORGETFULNESS_OF_THE_PARTICIPANT. • o participante 1518 não respondeu ao questionário {WELLBEING2} no último dia de recolha. Annotation: FORGETFULNESS_OF_THE_PARTICIPANT.

¹Por questões de privacidade, esta alteração impactou também os ficheiros existentes no *dataset raw*.

[MANADJUST]	<ul style="list-style-type: none"> Os participantes 1519 e 1533, inseriram o código da organização e/ou o número de participante de forma incorreta na primeira utilização da aplicação. Os valores do atributo <i>OSVersion</i> foram ajustados para facilitar a leitura (incluiu-se o nome do SO e utilizou-se apenas o número principal da versão) (e.g. Android 9). Annotation: VALUE_RECLASSIFICATION.
[MANCLASSIFICATION]	<ul style="list-style-type: none"> Todos os números de telefone das (CALL) e (SMS) foram anonimizados. Os tipos de interlocutor foram reclassificados de { 1, 2, 3 } para { <i>Spouse-Sons-CloseFriends, RemainingFamily-OtherFriends, Others, Not-classified</i> }. Os tipos de interação foram reclassificados de { 1, 2, 3 } para { <i>Negative, Neutral, Positive</i> }. Os tipos de (CALL) foram reclassificados de { 1, 2, 3, 4 } para { <i>MissedCall, InitiatedOnly, Incoming, Outgoing</i> }. As (CALL) do tipo <i>incoming</i> e <i>outgoing</i>, com duração igual ou inferior a 3", foram reclassificadas para <i>InitiatedOnly</i>. Annotation: CALL_ANSWERED_DURING_HANGING_UP. As respostas dos questionários foram reclassificadas de { 1, 2, 3 } para { <i>I disagree, Neither disagree nor agree, I agree</i> }.
[MANINSERT]	<ul style="list-style-type: none"> A integração dos dados registados em papel foi efetuada através de <i>scripts</i> (e.g. participantes 1506, 1507). Annotation: COLLECTION_THROUGH_PAPER.
{MIBAND}	
[-OUTLIER]	<ul style="list-style-type: none"> Existiam vários valores <i>mild</i> e <i>extreme outlier</i> no sinal (HR) e registo de passos. Alguns <i>extreme outlier</i> foram interpolados outros foram descartados (e.g. participantes 1510, 1514, 1529, 1904). Annotation: EXTREME_OUTLIER_DISCARDED.
[DISCARDDATA]	<ul style="list-style-type: none"> O número de passos dos participantes 1517 e 1909, foram reiniciados às 10:00 no primeiro dia recolha. Os participantes tinham colocado as pulseiras antes da hora de início da experiência.
[DISTINCTOBJ]	<ul style="list-style-type: none"> O participante 1503 tinha registos em duplicado no ficheiro <code>???_???_MiBand_hr.csv</code>. Considerou-se apenas o primeiro de cada conjunto de repetidos.
[NORM]	<ul style="list-style-type: none"> O número de passos registados pela {MIBAND} é um contador diário. Cada registo do <code>???_???_MiBand_steps.csv</code> representa o acumulado a cada 10'. Integrou-se o número de passos dados em cada 10' para facilitar a agregação posterior. As datas <i>POSIX</i> do ficheiro <code>???_???_MiBand_sleepRaw.csv</code> foram exportadas pelo {NOTIFYFITNESS} com menos três dígitos à direita. Implementou-se uma tarefa específica de ETL para a conversão correta para o tipo <i>datetime2</i> do {MSSQL}.
[RELIABILITY]	<ul style="list-style-type: none"> Detetou-se que os participantes 1517 e 1909, tinham dado 3054 e 1907 passos respetivamente, minutos depois do início da recolha. Confirmou-se que os participantes tinham colocado a {MIBAND} antes das 9:30, e tinham feito ginástica matinal.
{EDAMOVE}	
[-OUTLIER]	<ul style="list-style-type: none"> O <i>dataset</i> continha valores <i>mild</i> e <i>extreme outlier</i> (e.g. (TEMP)). Os valores <i>extreme outlier</i> foram interpolados ou descartados (e.g. participante 1518). Annotation: EXTREME_OUTLIER_DISCARDED.

[DISCARDDATA]	<ul style="list-style-type: none"> • O participante 1518 não utilizou o sensor no último dia, porque participou numa cerimónia e não queria ser vista com os elétrodos. Annotation: UNEXPECTED_SENSOR_REMOVAL. • Alguns participantes removeram o sensor durante algum tempo para executar tarefas incompatíveis com a utilização de elétrodos (e.g. o participante 1530 retirou o sensor para dar banho ao filho). Annotation: SENSOR_TEMPORARILY_REMOVED. • As falhas de sinal que excederam as 2h foram descartadas. Annotation: HEAVY_ELECTRODES_FIXING_PROBLEM.
[INTERPOLATION]	<ul style="list-style-type: none"> • Os ficheiros <code>???_????_EDAMove-Unisens/Results.xlsx</code>, gerados pelo <code>{DATAANALYSER}</code>, não continham valores na última hora de recolha em alguns participantes (e.g. 1519, 1526). Annotation: ALGORITHM_FEATURE. • Reconstruíram-se as falhas de sinal relativas aos momentos de exercício físico e momentos de higiene pessoal (períodos sem recolha inferiores a 2h) (e.g. participantes 1503, 1508). Annotation: PERSONAL_HYGIENE_PHYSICAL_ACTIVITY. • As falhas não justificadas de sinal até 1h, foram interpoladas. Annotation: LIGHT_ELECTRODES_FIXING_PROBLEM. • As falhas não justificadas de sinal até 2h (seguidas), foram também interpoladas (nos restantes casos, os dados foram descartados). Annotation: MEDIUM_ELECTRODES_FIXING_PROBLEM. • Os objetos com valores sem sentido (ruído), foram interpolados também com a mediana (e.g. participantes 1506 e 1518). Annotation: OTHER_UNSPECIFIED_LIGHT_PROBLEMS. • O participante 1519 removeu o sensor durante um curto espaço de tempo para realizar uma tarefa incompatível com a utilização do sensor. Annotation: SENSOR_TEMPORARILY_REMOVED.
[MANCLASSIFICATION]	<ul style="list-style-type: none"> • O atributo <i>NonWearSleepWake</i> foi reclassificado de <code>{ 0, 1, 2 }</code> para <code>{ Wake, Sleep, Not worn }</code>.
<code>{WHOQOLBREF}</code>	
[MANINSERT]	<ul style="list-style-type: none"> • Os dados gerais do participante 1514, foi preenchida a partir do <code>{SOCIOCULTURAL}</code>. • A profissão dos participantes 1506 e 1507 foi atribuída pelo autor, porque não foi preenchida pelos participantes.
[MANCLASSIFICATION]	<ul style="list-style-type: none"> • O atributo <i>IsCurrentlySick</i> foi reclassificado de <code>{ 0, 1 }</code> para <code>{ Currently not sick, Currently sick }</code>.
[NORM]	<ul style="list-style-type: none"> • As profissões foram uniformizadas: <i>Comercial</i>, <i>IT</i>, <i>Office Worker</i>, <i>Manager</i> ou <i>Others</i>. As profissões "Comercial" e "Comercial Têxtil", foram mapeadas para <i>Comercial</i>. As profissões "Programador", "Técnico informática" e "Administrador Sistemas", foram reclassificadas como <i>IT</i>. As profissões "Assistente Administrativa", "Contabilista Certificado", "Técnica Contabilidade", "Administrativa", foram reclassificadas para <i>Office Worker</i>. As profissões "Encarregado Geral" e "Empresário Têxtil", foram reclassificadas como <i>Manager</i>. Todas as outras profissões (e.g. "Técnica Produção", "Técnico") foram reclassificadas no grupo <i>Others</i>.
<code>{SOCIOCULTURAL}</code>	

[MANADJUST]	<ul style="list-style-type: none"> • O participante 1519, na pergunta "Número de empregos desde o início da vida ativa", respondeu 18 porque entendeu como anos de trabalho (o participante em causa tinha 35 anos). Perguntou-se ao participante o valor correto. • Os participantes 1508, 1511 e 1533 registaram um peso ligeiramente diferente do que tinham informado no preenchimento do {SUBJECTFILE}. Os valores do {SOCIOCULTURAL} foram retificados para os indicados no {SUBJECTFILE}, porque foram estes os utilizados no processamento do {DATANALYSER} para geração dos ficheiros <i>raw</i> do {EDAMOVE}.
[MANINSERT]	<ul style="list-style-type: none"> • O participante 1521 não preencheu a idade. Foi calculada com base na data de nascimento inserida no {WHOQOLBREF}.
[NORM]	<ul style="list-style-type: none"> • A unidade de medida dos valores do atributo "Altura" foram normalizados para metros (e.g. participantes 1519, 1503, 1517, 1904 e 1955). • A distância de casa ao trabalho em quilómetros registada pelo participante 1514 foi arredondada às unidades.
[MAN CLASSIFICATION]	<ul style="list-style-type: none"> • O atributo relativo a dependência financeira dos dependentes, foi reclassificado para { <i>Money independent, Money dependent</i> }. • O atributo relativo às necessidades especiais pelos dependentes, foi reclassificado para { <i>No special needs required, Special needs required</i> }. • O atributo relativo à distância entre a casa e o trabalho, foram reclassificados para { <i>Workplace is very near my home, Workplace is near my home, Workplace is little far from my home, Workplace is far from my home</i> }. • O atributo relativo a duração da vida ativa, foi reclassificado para { <i>Worklife begun very recently, Worklife begun recently, Worklife begun some time ago</i> }. • O atributo relativo aos papéis de relvo na sociedade, foi reclassificado para { <i>Does not have an important social role on organizations, Do have an important social role on organizations</i> }.
[RELIABILITY]	<ul style="list-style-type: none"> • Durante a análise dos dados, detetou-se que o participante 1519 tinha introduzido um valor impossível no campo onde era solicitado o número de empregos desde o início da vida ativa. Aplicou-se a técnica [MANADJUST] para corrigir o valor. • Na comparação de dados gerais do {SOCIOCULTURAL} com o {SUBJECTFILE}, detetou-se que haviam ligeira diferenças dos pesos entre ambos os instrumentos. Aplicou-se a técnica [MANADJUST] para retificar os valores do {SOCIOCULTURAL}.
{LABOR}	
[NORM]	<ul style="list-style-type: none"> • Os atrasos ao trabalho superiores a 60 minutos foram normalizados com o valor 999999.

Tabela 4.5: Resumo das técnicas de pré-processamento utilizadas

4.3 Propriedades extraídas

As propriedades extraídas são mais apropriadas ao processo de classificação e podem ser vistas como *inputs* complementares ou de substituição aos atributos do *dataset* original (Haag et al., 2004). As propriedades podem ser extraídas diretamente do sinal recolhido pelos instrumentos & sensores, ou do sinal resultante da aplicação de uma técnica de pré-processamento. Depois de melhorar a qualidade dos dados através da aplicação de técnicas de pré-processamento, é necessário encontrar as propriedades e medidas estatísticas ideais para o reconhecimento automático de emoções (Jerritta et al., 2011). Nesta secção são apresentadas as propriedades extraídas do *dataset* pré-processado que foram escolhidas para enriquecer o *dataset* final.

O desempenho dos algoritmos depende da qualidade dos dados pré-processados, mas também das propriedades extraídas escolhidas pelos autores para incluir no *dataset* final (Urges, 2021; Matlovic et al., 2016; Adams and Robinson, 2015; Gama et al., 2012). Neste contexto, considera-se que a seleção de propriedades a extrair, é uma fase importante para os objetivos pretendidos com este projeto.

A escolha das propriedades a extrair, iniciou-se durante a definição da estratégia assente em *views* seguida na fase de pré-processamento (ver secção 4.2.1). O objetivo passava por incluir a maior quantidade possível de propriedades extraídas no *dataset* final. A ideia era não limitar a escolha de atributos a utilizar futuramente na deteção automática do bem-estar. Neste contexto, a estratégia definida para a implementação, pressuponha a homogeneização algorítmica entre as *views* do mesmo nível. Além de facilitar a leitura e manutenção futura, a implementação inicial das *views* dos vários sinais também seria mais rápida, atendendo a que se resumiria a um processo de adaptação de código para cada caso.

Os algoritmos definidos para a implementação das *views* de nível 2 e nível 3, iriam implementar o cálculo das propriedades específicas de contexto e das restantes propriedades (e.g. medidas estatísticas). Como a implementação das *views* resultou da propagação do mesmo algoritmo de forma transversal em cada nível, foram calculadas as mesmas medidas estatísticas para todos os sinais. A implementação das propriedades extraídas aconteceu em simultâneo com a aplicação de técnicas de pré-processamento. À medida que um sinal era pré-processado com a edição e anotação de dados nas várias tabelas da base de dados, eram aplicadas as restantes técnicas de pré-processamento nas *views* de nível 1 e 2 se necessário, e definidas as propriedades extraídas nas *views* de nível 2 e 3.

Na Tabela 3.4 enumerou-se algumas propriedades possíveis de extrair dos dados recolhidos de contexto social e psicológico. No âmbito deste trabalho apenas se extraiu um subconjunto dessas, que podem ser consultadas na Tabela 4.6. As medidas estatísticas calculadas sobre os valores contínuos, e as medidas estatísticas calculadas sobre os valores categóricos, podem ser consultadas na Tabela 4.7.

Inicia-se esta secção descrevendo algumas das propriedades específicas de contexto selecionadas, e termina-se resumindo as restantes propriedades extraídas consideradas no *dataset* final.

4.3.1 Propriedades específicas do contexto

As propriedades específicas de contexto aumentam a variedade e qualidade do *input* dos sistemas, e podem promover a eficiência dos classificadores na deteção automática de emoções por causa da estreita relação com o contexto de recolha (Matlovic et al., 2016). Nesta secção apresentam-se as propriedades extraídas que são específicas ao contexto de recolha. Como são dependentes do contexto de onde são extraídas, são apresentadas em função do instrumento ou sensor original de recolha.

Foram extraídas propriedades específicas de contexto de várias variáveis recolhidas. O `{WHOQOLBREF}` calcula de forma subjetiva o indicador de qualidade de vida. O instrumento avalia os domínios: saúde física, saúde psicológica, relações sociais e ambiente. Além disso calcula também um indicador global acerca da qualidade de vida. Como já referido na fase de integração, apenas foram integradas as respostas às questões colocadas aos participantes. Os resultados dos quatro domínios e do indicador de qualidade de vida, são propriedades específicas de contexto calculadas pela *view* de nível 2 `Participants_2_View`.

As propriedades específicas de contexto do sinal (EDA) recolhido pelo `{EDAMOVE}`, foram calculadas aquando do pré-processamento efetuado pelo `{DATAANALYSER}` ainda durante o processo de recolha de dados de contexto. O autor utilizou os instrumentos disponibilizados pela Movisens (Movisens, 2020) para recolher e pré-processar os dados recolhidos pelo sensor. Apesar de identificadas nesta secção, estas propriedades específicas de contexto foram calculadas no final de cada iteração de recolha. Tinha que se devolver os sensores `{EDAMOVE}` e, por este motivo, preferiu efetuar o pré-processamento destes dados logo após o processo de recolha. Desta forma, se algum problema ocorresse durante a exportação de dados efetuada pelo `{SENSORMANAGER}` e/ou pré-processamento pelo `{DATAANALYSER}`, poder-se-ia repetir o processo.

A Tabela 4.6 resume algumas das propriedades específicas de contexto calculadas, agrupando-as pelo instrumento ou sensor original de recolha.

Propriedade	Descrição
	<code>{ANDROIDANALYTICS}</code>
<code>(CALL) <DURATION<MEAN>></code>	Duração média das chamadas telefónicas. ¹
<code>(CALL) <DURATION<SUM>></code>	Duração total das chamadas telefónicas. ¹
<code>(CALL) <INCOMING<DURATION<SUM>>></code>	Duração total das chamadas telefónicas recebidas.

¹ São apenas consideradas as chamadas recebidas e efetuadas (i.e. tipo *incoming* e *outgoing*).

(CALL) <OUTGOING<DURATION<SUM>>>	Duração total das chamadas telefônicas efetuadas.
(CALL) <INTERLOCUTORS<COUNT>>	Número de interlocutores (distintos).
(CALL) <INTERLOCUTORS<MODE>>	Interlocutor mais frequente.
(CALL) <COUNT>	Número total de chamadas efetuadas e recebidas.
(CALL) <INCOMING<COUNT>>	Número total de chamadas recebidas.
(CALL) <OUTGOING<COUNT>>	Número total de chamadas efetuadas.
(CALL) <MISSED<COUNT>>	Número total de chamadas perdidas.
(CALL) <INITIATEDONLY<COUNT>>	Número total de chamadas apenas iniciadas.
(SMS) <COUNT>	Número total de chamadas apenas iniciadas.
(SMS) <WORDS<COUNT>>	Média do número de palavras por mensagem.
(SMS) <CHARS<COUNT>>	Média do número de caracteres por mensagem.
(SMS) <INTERLOCUTORS<COUNT>>	Número de interlocutores (distintos).
(SMS) <INTERLOCUTORS<MODE>>	Interlocutor mais frequente.
(ELECTR) <USAGE<COUNT>>	Número de <i>screen on/off</i> do <i>smartphone</i> .
(WELLBEING) <EASYRELATIONS<MODE>>	Resposta mais frequente à pergunta <i>"Relacionei-me facilmente com as pessoas à minha volta"</i> .
(WELLBEING) <DIFFICULTSITUATIONS<MODE>>	Resposta mais frequente à pergunta <i>"Estive capaz de enfrentar situações difíceis de uma forma positiva"</i> .
(WELLBEING) <THEYLIKEME<MODE>>	Resposta mais frequente à pergunta <i>"Senti que os outros gostavam de mim e me apreciavam"</i> .
(WELLBEING) <PROUDMYSELF<MODE>>	Resposta mais frequente à pergunta <i>"Senti-me satisfeito com o que fui capaz de alcançar, senti-me orgulhoso de mim próprio"</i> .
(WELLBEING) <RIGHTBALANCE<MODE>>	Resposta mais frequente à pergunta <i>"A minha vida foi bem equilibrada entre as minhas atividades familiares, pessoais e académicas"</i> .
(WELLBEING) <EMOTIONALLYBALANCED<MODE>>	Resposta mais frequente à pergunta <i>"Senti-me emocionalmente equilibrado"</i> .
(WELLBEING) <PEACEWITHMYSELF<MODE>>	Resposta mais frequente à pergunta <i>"Senti-me bem, em paz comigo próprio"</i> .
(WELLBEING) <IMCONFIDENT<MODE>>	Resposta mais frequente à pergunta <i>"Senti-me confiante"</i> .
(SLEEP) <QUALITY<MODE>>	Resposta mais frequente à pergunta <i>"Dormi muito bem e sinto que o meu sono foi totalmente reparador"</i> .
(TIRED) <LEVEL<MODE>>	Resposta mais frequente à pergunta <i>"Sinto-me totalmente descansado após esta noite de sono"</i> .
{EDAMOVE}	
(EDA) <AROUSAL<MEAN>>	Média dos valores de <i>arousal</i> . Estes valores são calculados com base nas <i>Skin Conductance Responses</i> (SCR) (movisens GmbH, 2018a).
(EDA) <SCL<MEAN>>	Média dos <i>Skin Conductance Level</i> (SCL). (movisens GmbH, 2018a).
(EDA) <SCR<AMP<MEAN>>>	Média de amplitudes dos SCR (movisens GmbH, 2018a).
(EDA) <SCR<COUNT>>	Número de SCR detetadas (movisens GmbH, 2018a).
(EDA) <SCR<ENERGY<MEAN>>>	Média da energia das SCR detetadas (movisens GmbH, 2018a).

(EDA) <SCR<HALFRECOVERYTIME<MEAN>>>	Média dos tempos de recuperação até metade da amplitude de cada SCR detetada (movisens GmbH, 2018a).
(EDA) <SCR<RISETIME<MEAN>>>	Média dos tempos de subida de cada SCR detetada (movisens GmbH, 2018a).
(TEMP) <MEAN>	Média da temperatura ambiente (movisens GmbH, 2018a).
{WHOQOLBREF}	
(QUALITYOFLIFE) <PHYSICALDOMAIN>	Pontuação do domínio físico (Rimas, 2020).
(QUALITYOFLIFE) <PSYCHOLOGICALDOMAIN>	Pontuação do domínio psicológico (Rimas, 2020).
(QUALITYOFLIFE) <SOCIALRELATIONSHIPS>	Pontuação do domínio relações sociais (Rimas, 2020).
(QUALITYOFLIFE) <ENVIRONMENT>	Pontuação do domínio ambiente (Rimas, 2020).
(QUALITYOFLIFE)	Indicador global da qualidade de vida (Rimas, 2020).
{SOCIOCULTURAL}	
(DEPENDENTS) <COUNT>	Número de dependentes.
(DEPENDENTS) <CATEGORY<COUNT>>	Número de dependentes por categorias (crianças e adultos).
(WEIGHT, HEIGHT) <BMI>	Índice de massa corporal (Who, 2021a).
(WEIGHT, HEIGHT) <BMI<CATEGORY>>	Categoria do índice de massa corporal (Who, 2021a).
(WEIGHT, HEIGHT) <BMR>	<i>Basal Metabolic Rate</i> (BMR) (Who, 2021b).

Tabela 4.6: Propriedades extraídas específicas de contexto

4.3.2 Outras propriedades extraídas

Além das propriedades específicas de contexto, foram também extraídas outras propriedades como medidas estatísticas, etc. Nesta secção são apresentadas as outras propriedades calculadas, não apresentadas na secção 4.3.1.

No grupo das outras propriedades, incluem-se as propriedades de origem matemática, análise de sinal e outras propriedades genéricas. Estas propriedades não são descritas em detalhe neste documento, porque são de domínio geral e são também bastante referenciadas na literatura. Os *tokens* utilizados na sua referência são auto-descritivos.

O cálculo das outras propriedades é da responsabilidade das *views* de nível 3. Como já referido na secção 4.3, a implementação das *views* dos vários sinais, resultou da propagação e adaptação de código, para cada sinal analisado e pré-processado. Assim, os vários atributos do *dataset* deram origem às mesmas medidas estatísticas. Neste contexto, as outras propriedades extraídas apresentadas nesta secção, não são apresentadas em função do instrumento & sensor original dos valores, mas sim em função do tipo de dados de cada variável de contexto origem.

As *views* de nível 3 calculam também medidas estatísticas para as propriedades específicas de contexto (e.g. para (CALL) <INCOMING<DURATION<SUM>>> são calculadas também as medidas estatísticas <MEAN>, <MAX>, <VAR>, <STD>, etc.).

A tabela 4.7 apresenta as outras propriedades calculadas para os valores contínuos e valores categóricos.

Propriedade	Descrição
{Valores contínuos}	
<NORM>	Valor normalizado. ¹
<MEAN>	Média individual e de todos os participantes.
<MEDIAN>	Mediana individual e de todos os participantes.
<MIN>	Valor mínimo individual e de todos os participantes.
<MAX>	Valor máximo individual e de todos os participantes.
<AMP>	Amplitude individual e de todos os participantes.
<VAR>	Variância individual e de todos os participantes.
<STD>	Desvio padrão individual e de todos os participantes.
<CV>	Coefficiente de variação individual e de todos os participantes.
<Q1>	Quartil 1 individual e de todos os participantes.
<Q3>	Quartil 3 individual e de todos os participantes.
<IQR>	Amplitude interquartil individual e de todos os participantes.
<UPPEROUTERFENCE>	Cerca superior externa (determinação de <i>extreme outliers</i>).
<UPPERINNERFENCE>	Cerca superior interna (determinação de <i>mild outliers</i>).
<LOWERINNERFENCE>	Cerca inferior interna (determinação de <i>mild outliers</i>).
<LOWEROUTERFENCE>	Cerca inferior externa (determinação de <i>extreme outliers</i>).
<OUTLIERCANDIDATE>	Objeto candidato a <i>mild</i> ou <i>extreme outlier</i> .
Valores categóricos	
<MODE>	Moda individual e de todos os participantes.

Tabela 4.7: Outras propriedades extraídas

4.4 Ferramentas desenvolvidas DQVA e DQVA-PP

O *Data Quality Visual Analysis* {DQVA} foi criado pelo autor para apoiar a fase de integração de dados do *dataset raw* apresentado no Capítulo 3. Os recursos que disponibiliza permitiram que a integração fosse feita de forma mais simples e rápida (Carvalho et al., 2021). Criou-se também a variante {DQVA-PP} que permitiu analisar a qualidade dos dados pré-processados. Nesta secção apresenta-se o {DQVA} e {DQVA-PP}, ferramentas que se acredita, terem facilitado a fase de pré-processamento e terem contribuído para a qualidade final do *dataset* integrado, apresentado no Capítulo 5 como contribuição.

São muitos os investigadores que recorreram à estatística descritiva para conseguirem detetar anomalias nos dados, utilizando técnicas quantitativas da estatística (Chu et al., 2016; Prokoshyna et al., 2016). Resumem-se algumas dessas estratégias: i) Mishra et al. utilizaram a mediana e o desvio absoluto mediano para encontrar *outliers* nos seus

¹Valor calculado com a fórmula $z = (x - mean)/std$.

dados, e calcularam o mínimo e máximo para normalizar o (EDA), (ST) e (HR), em cada participante (Mishra et al., 2020); ii) Krishnan et al. desenvolveram o projeto *Sample-Clean* que continha rotinas de limpeza de dados utilizando várias medidas estatísticas (e.g. contagens, médias, somas) (Krishnan et al., 2015); iii) Wang et al. desenvolveram uma *framework* para processamento com *queries* sobre dados sem qualidade e, na fase de limpeza de dados da amostra, utilizaram agregadores estatísticos como média, soma, contagem, variância, média geométrica, etc. (Wang et al., 2014); iv) Prokoshyna et al. criaram uma *framework* para detetar problemas na qualidade dos dados (Prokoshyna et al., 2016); v) Volkovs et al. apresentaram uma aplicação para *continuous data cleaning* e utilizaram medidas estatísticas para analisar mudanças na distribuição de dados (Volkovs et al., 2014); etc.

Apesar de serem muitos os investigadores que utilizam técnicas quantitativas na análise da qualidade dos dados, estas medidas apenas representam as características gerais desses dados (Anscombe, 1973). A agregação feita através dos cálculos pode ocultar comportamentos do sinal, escondendo informação que pode ser importante acerca dos próprios dados. Além disso, a representação agregada, pode conduzir a conclusões incorretas acerca dos objetos que compõem o sinal, porque dão apenas uma visão global dos dados (Anscombe, 1973). Não obstante a importância dos valores agregados na descrição dos dados, a utilização de gráficos para análise de padrões visuais no sinal pode contribuir com mais informação sobre os próprios dados. Apesar da exatidão dos cálculos numéricos, os gráficos permitem uma perceção mais abrangente sobre os dados, e promovem a descoberta de mais informação acerca desses mesmos dados (Anscombe, 1973) (e.g. os histogramas e *scatterplots* podem revelar a existência de valores discrepantes numa distribuição (Kandel et al., 2012)).

Neste contexto, existem investigações que incidem sobre o estudo dos dados pela vertente visual. São vários os investigadores utilizam ferramentas genéricas para análise visual e tratamento dos dados (e.g. Orange (University of Ljubljana, 2020), Weka (University of Waikato, 2020), Knime (KNIME AG, 2020), SPSS (Field, 2013; Hinton, 2014; Arbuckle, 2010), GGobi (Swayne et al., 2003), Tableau (Tableau Software LLC, 2020), OpenRefine (CS&S, 2020) (Google, 2020), Excel (Microsoft, n.d.b)). A literatura relata também várias investigações onde foram desenvolvidos instrumentos visuais para análise de dados: i) Carr et al. abordaram o problema da utilização do *scatterplot* na representação de grandes quantidades de dados (Carr et al., 1987); ii) Kosara et al. apresentaram um novo tipo de histogramas chamado de *TimeHistogram* como uma extensão do histograma normal, que tem em conta o momento temporal dos dados (Kosara et al., 2004); iii) Rosario et al. abordaram o problema da representação de variáveis nominais através de ferramentas de análise visual desenhadas para variáveis numéricas (Rosario et al., 2003, 2004); iv) Unwin et al. abordaram estratégias para representar visualmente grandes *datasets* de dados (Unwin et al., 2006); v) Guo utilizou instrumentos visuais para

detetar padrões em dados com muitas dimensões (Guo, 2003); vi) Kandel et al. criaram um sistema para manipular dados visualmente, e poder inferir transformações de forma automática (Kandel et al., 2011); vi) Zhang et al. propuseram um sistema para reparar de forma autónoma os dados de *time series* (Zhang et al., 2017); vii) Tae et al. apresentaram uma *framework* integradora de técnicas para promover o pré-processamento dos dados *input* de algoritmos de *machine learning* (Tae et al., 2019); viii) Krisnan et al. propuseram um sistema para limpeza iterativa de dados para utilizar no treino de modelos estatísticos (Krishnan et al., 2016); ix) Rekatsinas et al. propuseram uma *framework* para reparação de dados através de inferência probabilística (Rekatsinas et al., 2017); x) He et al. criaram uma ferramenta interativa para limpar dados através de instruções SQL (He et al., 2016); xi) a *framework Voila* permitia detetar anomalias de forma interativa em dados provenientes de *streaming* (Cao et al., 2018); e xii) o *BigDancing* baseava-se num conjunto de regras de qualidade para deteção de problemas, e tenta aplicar reparações nos dados com o objetivo de automatizar o processo de limpeza (Khayyat et al., 2015).

O {DQVA} disponibiliza um conjunto de medidas agregadas e de gráficos gerados automaticamente, para promover a análise e descoberta de mais informação sobre o *dataset*. A literatura relata várias investigações que utilizam gráficos nos seus projetos: i) o *Profiler* era uma ferramenta de análise visual que permitia detetar anomalias e apoiar no processo de limpeza de dados em formato tabular, resumindo os dados em visualizações (Kandel et al., 2012); ii) Asghari et al. criaram um sistema para limpeza semiautomática de *datasets* através de um conjunto de ferramentas que permitem que o utilizador identifique padrões nos dados (e.g. utilizando expressões regulares) (Asghari et al., 2018); iii) A *framework* de Thankachan identificava a estrutura e tipos de dados de forma automática, e permitia detetar anomalias nos dados (Thankachan, 2018); iv) o EventThread3 de S. Guo et al. facilitava a exploração interativa dos dados (Guo et al., 2019); e v) o Nadeef promovia a deteção e limpeza de dados através da definição de regras pré-definidas (Dallachiesat et al., 2013). Os trabalhos de Asghari et al., S. Guo et al., e Dallachiesat et al., focaram-se essencialmente na deteção e reparação de dados. Em termos gerais, os seus recursos gráficos existem para apresentar erros ou estatísticas de erros. O Nadeef utiliza *dashboards* com baixo nível de personalização, para permitir que *domain experts* possam interagir visualmente com os dados. O EventThread3 de S. Guo et al. foi criado para detetar problemas em dados referentes a sequências de eventos. O sistema de Asghari et al., apesar de integrar com {MSSQL}, não está vocacionado para analisar visualmente vários sinais do mesmo atributo, além de não oferecer as medidas estatísticas de agregação que se pretendiam calcular neste experiência. Os recursos gráficos do Profiler de Kandel et al. representavam apenas valores agregados. A *framework* Thankachan et al. permitiu que fosse feita análise de dados com base em métodos estatísticos e métodos visuais através de gráficos. No entanto, além de não estar claro o nível de personalização da componente gráfica, estes instrumentos são utilizados apenas para representar pontos problemáticos

nos dados, não sendo possível representar o sinal não agregado para comparação entre participantes e organizações. Além disso, não é possível definir medidas de agregação adicionais, e não é claro se a *framework* conseguia integrar com bases de dados relacionais. Esta integração não era vital, porque poderia ser feita de forma indireta (e.g. através da exportação para ficheiros XML, JSON, CSV). No entanto, dado o número elevado de atributos do *dataset*, a integração direta com a base de dados onde havia sido integrado o *dataset*, poderia facilitar o processo de análise do grande número de atributos recolhidos.

Resumindo, as plataformas genéricas como o Orange (University of Ljubljana, 2020), SPSS (Field, 2013), e também o Excel (Microsoft, n.d.b), não ofereciam o nível de automação para a análise visual que se precisava para analisar os sinais do *dataset* individualmente e por organização. Além disso, as ferramentas desenvolvidas pelos outros investigadores eram demasiado específicas para serem reutilizadas neste projeto. Foi neste contexto que se decidiu criar uma ferramenta que calculasse e apresentasse automaticamente as medidas agregadas e visualizações que simplificassem o processo de análise de dados através de *data insights*. O {DQVA}, além de promover a descoberta de *outliers* e falhas de sinal, permitia também que os sinais dos diferentes participantes e organizações pudessem ser comparados através de quadros resumo e visualizações. Cada sinal de cada participante ou organização é representado como uma *time-series*. Além disso, é possível escolher quais os participantes e organizações que se pretendem comparar visualmente, sendo possível escolher vários participantes ou várias organizações em simultâneo. Na impossibilidade de se encontrar uma ferramenta que permitisse a personalização pretendida, escolheu-se uma folha de cálculo porque oferecia, de forma nativa, a possibilidade de utilizar medidas estatísticas e visualizações. Os únicos requisitos na seleção da folha de cálculo, era a possibilidade de integrar com bases de dados relacionais, e permitir a representação gráfica de várias séries temporais em simultâneo, porque havia a necessidade de apresentar meta-dados visualmente como se fossem séries normais de dados. Optou-se pelo Excel, mas poderia ter-se escolhido qualquer outra folha de cálculo que cumprisse com os requisitos necessários.

Além das habituais medidas da estatística descritiva utilizadas na avaliação da qualidade dos dados (e.g. *average, median, trimmed mean, midhinge, trimean, midrange, maximum, minimum, range, IQR, skewness, kurtosis*), o {DQVA} permite também analisar e explorar os dados de forma visual: i) permite comparar os dados de cada variável entre os participantes filtrados; ii) avaliar o comportamento dos dados de cada variável entre organizações onde se procedeu à recolha; iii) verificar a existência de *mild* e *extreme outliers*, e respetivas *mild* e *extreme outlier fences*; iv) analisar visualmente a existência de *gaps* de sinal (i.e. missing data); e v) conhecer a distribuição dos dados em classes através de um histograma de frequências e de um diagrama de *Pareto*. Além da análise visual, os dados são também agregados com a média por dia relativo de recolha, por participante e

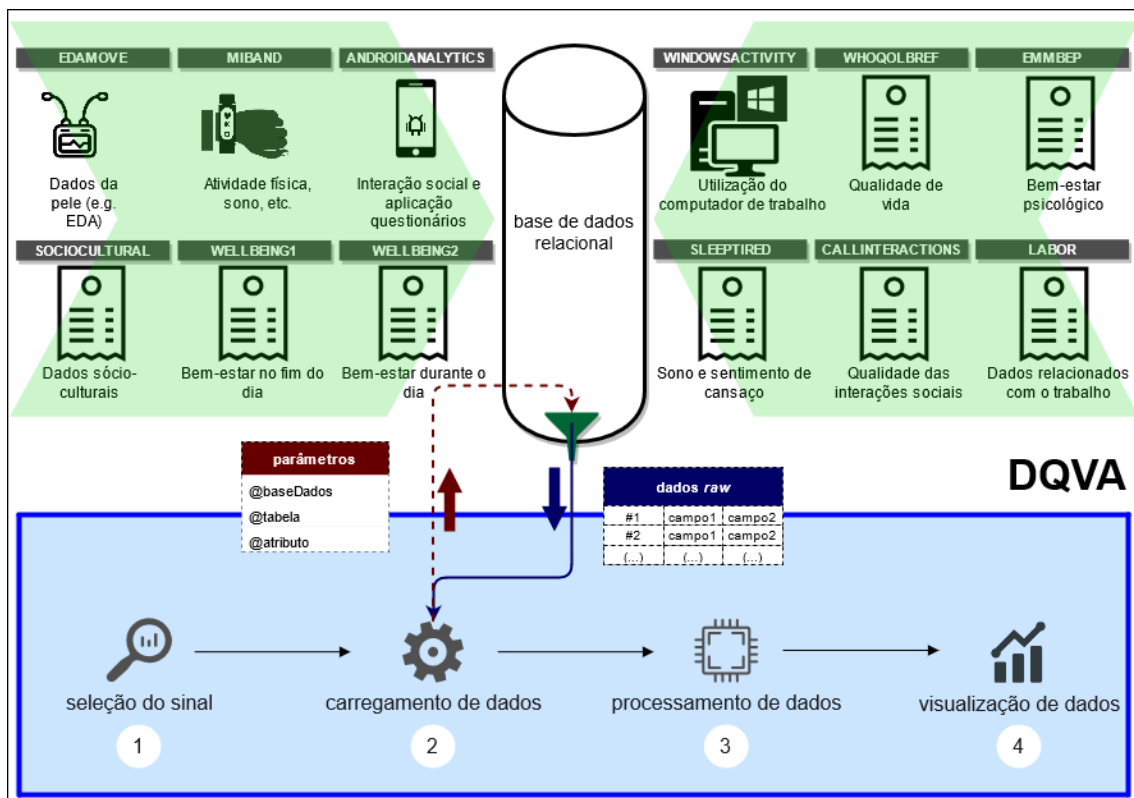


Figura 4.3: Arquitetura do DQVA

por organização.

A variante $\{DQVA-PP\}$ do $\{DQVA\}$ não é apresentada em detalhe, porque a diferença principal entre ambos os instrumentos é a origem dos dados que processa. Enquanto que o $\{DQVA\}$ opera sobre os dados *raw* integrados, o $\{DQVA-PP\}$ processa sobre os dados provenientes das *views* que abstraem o resultado da aplicação das técnicas de pré-processamento já relatadas neste capítulo.

De seguida apresenta-se a arquitetura do $\{DQVA\}$. Posteriormente abordam-se os recursos que oferece ao nível da descrição agregada, e ao nível da descrição visual dos dados.

4.4.1 Tecnologias e arquitetura da ferramenta

O $\{DQVA\}$ foi criado para analisar a qualidade dos dados durante a fase de integração do *dataset raw*, recolhido na experiência de recolha de dados em contexto laboral apresentada no Capítulo 3. A avaliação da qualidade e limpeza dos dados, tem um papel importante na fase de pré-processamento. A eficiência dos algoritmos depende da qualidade dos *datasets*. Nesta secção apresenta-se a arquitetura e as tecnologias utilizadas no desenvolvimento do $\{DQVA\}$.

A arquitetura atual da solução apresentada na Figura 4.3 é composta por dois com-

```

SELECT
    pks.OrganizationId,
    pks.ParticipantId,
    dbo.GetRoundedDatetime2(r.[Date]) AS DateTimeReal,
    pks.[Date] AS [TimeRelative],
    r.[TempMean] AS [Value]
FROM EXP001 [dbo]. _MandatorySignalPKsView pks
    INNER JOIN [dbo].Participants p ON
        (
            pks.OrganizationId = p.OrganizationId AND pks.ParticipantId = p.Id
        )
    LEFT JOIN EXP001 [dbo]. EDAMOVE Results r ON
        (
            pks.[Date] = DATEADD(SECOND, DATEDIFF(SECOND, CAST(p._d1 AS DATE), dbo.GetRoundedDatetime2(r.[Date])), '')
            AND
            pks.OrganizationId = r.OrganizationId
            AND
            pks.ParticipantId = r.ParticipantId
        )
WHERE dbo.GetRoundedDatetime2(COALESCE(r.[Date], CAST('1900-01-01' AS DATETIME2))) <= dbo.GetRoundedDatetime2(p._d2)
    AND dbo.GetRoundedDatetime2(COALESCE(r.[Date], CAST('2099-12-31' AS DATETIME2))) >= dbo.GetRoundedDatetime2(p._d1)
ORDER BY pks.OrganizationId, pks.ParticipantId, pks.[Date]

```

Figura 4.4: Exemplo de *query* criada dinamicamente pela *Stored Procedure*

ponentes principais: i) o *Microsoft SQL Server 2017* {MSSQL} (SQL Server Team, 2017) que desempenha o papel de DBMS no sistema, onde foi implementada a base de dados relacional; e ii) o {DQVA} que permitiu a análise estatística agregada e visual da qualidade dos dados, após carregamento a partir do DBMS. Além destes componentes, também são representadas as várias origens dos dados que foram integrados na base de dados relacional (e.g. {EDAMOVE}, {MIBAND}, {ANDROID ANALYTICS}, {WELLBEING1}). As setas verdes representam os instrumentos responsáveis pela integração dos dados no {MSSQL} (ver secção 4.1.1).

Como já referido, o {DQVA} foi desenvolvido em Excel, num ficheiro com permissão para execução de código *Visual Basic for Applications* (VBA) (Burnett, 2009). Este código apenas apoia o utilizador da ferramenta no processo de escolha do instrumento e da variável de contexto a analisar. A operação do {DQVA} inicia com a seleção: i) da base de dados relacional; ii) tabela onde se encontram os dados; e iii) o atributo cujos dados de sinal se pretende analisar. O carregamento de dados é efetuado pelo *Microsoft Power Query* (Microsoft, n.d.a). O carregamento de dados é efetuado com recurso a uma *Stored Procedure* (Microsoft, 2017b) alojada dentro do {MSSQL}.

Aquando da execução, o *Power Query* entrega ao DBMS, três parâmetros de entrada (cf. @baseDados, @tabela e @atributo) que são utilizados internamente pela *Stored Procedure* para a construção dinâmica da *query* a executar sobre a base de dados relacional (ver retângulos amarelo, verde e vermelho, da Figura 4.4). Após a execução da *query*, o *recordset* respetivo é devolvido à folha de cálculo no formato de uma série temporal.

A *Stored Procedure* foi desenhada de forma a impor a existência de, no mínimo, um objeto a cada SI. Desta forma, seria possível evidenciar de forma visual a existência de *gaps* no sinal. Esta imposição de dados foi estabelecida com uma junção à esquerda com a *view* `_MandatorySignalPKsView` (ver retângulo preto da Figura 4.4). Além de

	OrganizationId	ParticipantId	DateTimeReal	TimeRelative	Value
43681	117	1521	2019-02-19 10:00:00.0000000	1900-01-01 10:00:00.000	89
43682	117	1521	2019-02-19 11:00:00.0000000	1900-01-01 11:00:00.000	65

Figura 4.5: Excerto do *recordset* devolvido pela *query* dinâmica (unidade: BPM)

	1503	1506	1507	1508	1510	1511	1514	1517	1518	15
day 1, 09:00	73,04545455	80,16666667	71,94736842	77,36363636	80,30000000	71,75000000	76,35000000	80,96428571	75,71428571	84,396226
day 1, 10:00	79,64814815	69,85294118	66,06666667	84,40540541	79,21428571	76,30769231	85,30000000	71,08163265	71,54285714	83,878787
day 1, 11:00	84,09090909	65,21621622	60,27027027	79,67741935	78,26530612	68,65625000	82,33333333	64,81250000		78,905660
day 1, 12:00	75,89795918	72,00000000	64,48648649	76,35897436	88,78000000	76,42307692	83,83018868	75,30612245	71,70000000	80,851063
day 1, 13:00	75,83333333	73,77083333	63,74418605	76,30000000	82,90909091	73,53658537	87,28947368	71,51020408	71,62162162	75,100000
day 1, 14:00	76,41509434	64,77500000	54,82608696	74,48275862	78,21621622	72,78125000	78,94736842	72,84000000	80,56818182	78,186046
day 1, 15:00	78,14814815	67,30188679	64,13888889	83,78260870	81,84444444	66,46666667	83,44444444	64,14814815	77,40540541	78,933333
day 1, 16:00	73,65217391		61,38709677	86,81250000	80,97297297	68,05263158	77,80555556	68,20000000	75,04761905	85,936170
day 1, 17:00	76,46511628	76,40740741	63,08888889	77,17391304	77,56097561	80,33333333	80,35000000	65,12000000	72,82926829	91,840000

Figura 4.6: Excerto da matriz (HR) por participante / dia de recolha (unidade: BPM)

garantir a existência de, pelo menos, um objeto a cada hora de recolha, a *view* facilita também a normalização das datas para o dia relativo de recolha (cf. *day 1*, *day 2*, *day n*). Como a recolha foi feita em várias iterações (ver Capítulo 3), com a normalização para uma data relativa ao dia de recolha, tornava-se possível comparar os sinais dos diferentes participantes, na mesma série temporal.

A Figura 4.5 apresenta um excerto do *recordset* devolvido pela *Stored Procedure* quando selecionado o sinal (HR) do instrumento {MIBAND}. A coluna *OrganizationId* e *ParticipantId*, identificam anonimamente a organização e participante. As colunas *DateTimeReal* e *TimeRelative*, apresentam, respetivamente, a data real e relativa de recolha. A coluna *Value* contém o valor da variável a analisar, neste caso o (HR), em batimentos por minuto (BPM). Este *recordset* é apresentado pelo {DQVA} de três formas distintas: i) dados originais (i.e. sem qualquer processamento); ii) dados agregados pela média, por participante / SI; e iii) dados agregados pela média, por organização / SI. A título de exemplo, a Figura 4.6 apresenta um excerto da matriz representativa do sinal agregado (HR), recolhido pela {MIBAND}, de alguns participantes. As células em branco representam *gaps* no sinal (e.g. no primeiro dia de recolha, não existem dados (HR) dos participantes 1506 e 1518, na hora 16 e 11 respetivamente).

O {DQVA} integra com o DBMS em dois momentos: i) aquando da seleção do instrumento e variável pretendida para analisar, porque as opções disponíveis para seleção, são lidas da base de dados; e ii) e quando é executada a *Stored Procedure* para obter o *recordset* com os objetos a analisar. Apesar de criado originalmente para ser ligado ao {MSSQL}, o {DQVA} pode ser utilizado com outras origens de dados, bastando configurar o mecanismo de ligação e acesso a dados. O {DQVA} continuará a executar, independentemente da origem dos dados, desde que seja respeitada a estrutura do *recordset* apresentado na Figura 4.5.

GENERAL MEASURES		
	Count (all records)	60049
	Count (object records)	59574
	Count (gap records)	475
TENDENCY MEASURES		
	Average (AVG)	75,64942760
	Median (MED)	72,00000000
	Mode	70,00000000
	(*1) Trimmed Mean (cutoff = 20%)	74,18313051
	(*2) Midhinge, $MH = (Q1 + Q3) / 2$	74,00000000
	(*3) Trimean, $TM = (Q1 + (2Q2) + Q3) / 4$	73,00000000
	(*4) Midrange, $MR = (MAX + MIN) / 2$	116,50000000
	Q1	65,00000000
	Q2	72,00000000
	Q3	83,00000000
DISPERSION MEASURES		
	Minimum (MIN)	40,00000000
	Maximum (MAX)	193,00000000
	Range (RNG)	153,00000000
	(*5) Interquartile range, $IQR = Q3 - Q1$	18,00000000
	Variance (VAR) (sample)	254,68220999
	Standard deviation (STD) (sample)	15,95876593
	(*6) Coefficient of variation, $CV = STD / AVG$	0,21095686
OUTLIER FENCES (*7)		
	Upper Outer Fence = $Q3 + (3,0 * IQR)$	137,00000000
	Upper Inner Fence = $Q3 + (1,5 * IQR)$	110,00000000
	Lower Inner Fence = $Q1 - (1,5 * IQR)$	38,00000000
	Lower Outer Fence = $Q1 - (3,0 * IQR)$	11,00000000
SKEWNESS AND KURTOSIS (*8)		
	Adjusted Fisher-Pearson standardized coeff (SK)	0,95713570
	Skewness (SK) (Pearson) $SK = (3(AVG - MED)) / STD$	0,68603568
	Kurtosis (KU)	1,17607191

Figura 4.7: Medidas estatísticas calculadas pelo {DQVA} (e.g. (HR) (unidade: BPM))

Além desta representação em que os dados são apresentados de forma agregada por pistas representativas dos participantes e organizações de recolha, são também calculadas medidas estatísticas que descrevem de forma agregada os dados, e gráficos que tornam possível efetuar uma análise visual dos dados considerando a totalidade dos objetos que compõem o sinal. Estes componentes do {DQVA} serão apresentados nas secções seguintes.

4.4.2 Descrição agregada dos dados

Nesta secção apresentam-se as medidas estatísticas disponibilizadas pelo {DQVA}. Estas medidas descrevem apenas as características gerais dos dados. No entanto, e apesar destes valores agregados poderem ocultar comportamentos mais específicos do sinal (Anscombe, 1973), podem complementar a análise visual dos dados. Depois de carregado o *dataset* a partir da base de dados relacional, o {DQVA} gera os seguintes tipos de medidas quantitativas (ver Figura 4.7): i) medidas de tendência central; ii) medidas de dispersão; iii) assimetria e curtose; e iv) *outlier fences*. As imagens apresentadas nesta secção como exemplo, referem-se ao sinal (HR) recolhido pela {MIBAND}.

As medidas de tendência central representam a tendência de agrupamento de dados em redor do valor central (Dodge, 2006). O {DQVA} apresenta as medidas de tendência

central mais utilizadas (cf. média, mediana e moda), mas também outras medidas que tentam descrever a tendência dos dados na distribuição: i) *Trimmed Mean*, medida similar à média que remove uma pequena percentagem dos valores maiores e menores, com o objetivo de tentar reduzir os efeitos de *outliers* (Kenton, 2020); ii) *Midhinge* que representa a média do primeiro e terceiro quartil (Q1) e (Q3) (Taylor, 2019); iii) *Trimean* que é definida como uma média pesada, e não é muito afetada pelos *outliers* da distribuição (Stephanie, 2017); iv) *Midrange*, média calculada com base no máximo e mínimo da distribuição (medida influenciada pelos *outliers*) (Stephanie, 2013b); e v) são calculados também o primeiro (Q1), segundo (Q2) e terceiro (Q3) quartil (que ajudam na percepção da propagação de valores na distribuição, e representam, respetivamente, os 25% de valores mais baixos, os seguintes 25% de valores até à mediana, e os seguintes 25% de valores mais altos (Stephanie, 2013a; Minitab, 2019).

Além de medidas que tentam resumir a centralidade de dados, o {DQVA} também calcula algumas medidas que permitem analisar a dispersão dos dados na distribuição. Estas medidas tentam descrever resumidamente a disposição dos valores na distribuição (Toppr, n.d.; Stephanie, 2014). As seguintes medidas são apresentadas pelo DQVA: i) máximo e mínimo; ii) amplitude; iii) *InterQuartile Range* (IQR) que tenta representar a variabilidade dos dados subtraindo o Q3 ao Q1 para representar os 50% de objetos centrais da distribuição (Glen, n.d.); iv) variância que indica a variância dos valores em torno da média da distribuição (quão longe está cada valor do conjunto longe da média) (Hayes, 2019); v) desvio padrão que mede a dispersão em relação à média (quanto maior for o desvio padrão, mais espalhados estarão os valores no conjunto (Hargrave and Wetfall, 2020)); e vi) coeficiente de variação, que representa o rácio do desvio padrão na média, e é útil para comparar o grau de variação entre várias séries de dados, mesmo que tenham médias muito díspares (Hayes and Wetfall, 2020).

O {DQVA} também calcula algumas medidas relacionadas com a assimetria e curtose. Estas medidas tentam retratar de forma agregada o aspeto da curva representativa da distribuição em análise. A assimetria indica a distorção da curva representativa da distribuição em comparação com uma curva base representativa da distribuição normal. A assimetria de uma distribuição normal é zero (Chen, 2019). A curtose descreve a distribuição dos dados à volta da média (Kenton, 2019). A curtose mede o achatamento da distribuição em comparação com uma distribuição normal. Considerando o pico da distribuição num gráfico representativo da distribuição, a curtose descreve como os dados dessa distribuição se comportam à volta do pico (Casella and Berger, 2017; Kenton, 2019). O {DQVA} apresenta as seguintes medidas estatísticas: i) *Adjusted Fisher-Pearson standardized coeff*, fórmula normalmente utilizada pelos programas como o Excel para cálculo da assimetria (Doanne and Seward, 2011; Drefahl, 2016)); ii) *Skewness*, cálculo da assimetria utilizando a fórmula de *Pearson* tendo por base a mediana (Glen, 2014); e iii) *kurtosis*, cujo valor positivo significa existirem muitos dados nas caudas da linha

(chama-se uma curva leptocúrtica e é uma curva mais afunilada do que a distribuição normal), um valor negativo significa que existem poucos dados nas caudas da linha (chama-se curva platicúrtica e é uma curva mais achatada do que a distribuição normal), e um valor de zero, que significa que os dados estão distribuídos normalmente (chama-se curva mesocúrtica e é uma curva com achatamento similar a uma distribuição normal) (Kenton, 2019; Taylor, 2017).

Além das medidas agregadas já referidas, são também calculados valores das *outer* e *inner fences*. Estas medidas, são calculadas tendo por base o IQR, e são utilizadas para permitir a análise visual de *mild* e *extreme outliers* no {DQVA} (Glen, 2017).

Apesar da exatidão dos valores agregados na descrição dos dados, a utilização de visualizações pode permitir a deteção de mais informação acerca dos dados da distribuição (Anscombe, 1973). A secção seguinte apresenta os instrumentos visuais disponibilizados pelo {DQVA} que podem contribuir para um melhor conhecimento dos dados do sinal em análise.

4.4.3 Descrição visual dos dados

Apesar da importância dos valores agregados na descrição dos dados, a análise gráfica através de visualizações, pode facilitar a deteção de padrões visuais e permitir a descoberta de comportamentos de sinal que seriam difíceis de detetar através da análise de medidas estatísticas agregadas (Anscombe, 1973). Nesta secção apresentam-se algumas das visualizações geradas automaticamente pelo {DQVA}. Tendo por base os valores carregados da base de dados relacional, o {DQVA} disponibiliza visualmente: i) um *boxplot*; ii) um histograma de frequências e um diagrama de *Pareto*; iii) um *scatterplot* com a representação dos objetos na linha do tempo (dia relativo de recolha); e iv) outras visualizações representativas do sinal recolhido, por participante e organização, onde são visíveis *gaps*, e *extreme* e *mild outliers* do sinal. As imagens apresentadas nesta secção como exemplo, referem-se ao sinal (TEMP) recolhido pelo {EDAMOVE}.

O *boxplot* é utilizado normalmente para descrever a distribuição graficamente através da representação das medidas agregadas: mínimo, Q1, Q3 e máximo. O *boxplot* também representa a mediana e o IQR, e contém informação importante acerca da existência de *outliers* na distribuição (Galarnyk, 2018). A oferta nativa deste tipo de gráfico na folha de cálculo selecionada para o {DQVA} (i.e. Excel), apenas representa alguns dos dados que se pretendia representar visualmente no *boxplot*. Assim, e tendo por base um gráfico combinado, decidi criar um *boxplot*, contendo a informação já referida e, adicionalmente, a *outer* e *inner fences*. A Figura 4.8 apresenta o *boxplot* do {DQVA}.

O histograma de frequências apresenta visualmente com que frequência cada valor ocorre num conjunto de dados. Além de tornar perceptível visualmente as classes de dados mais dominantes na distribuição, também é possível analisar a distribuição dos valores

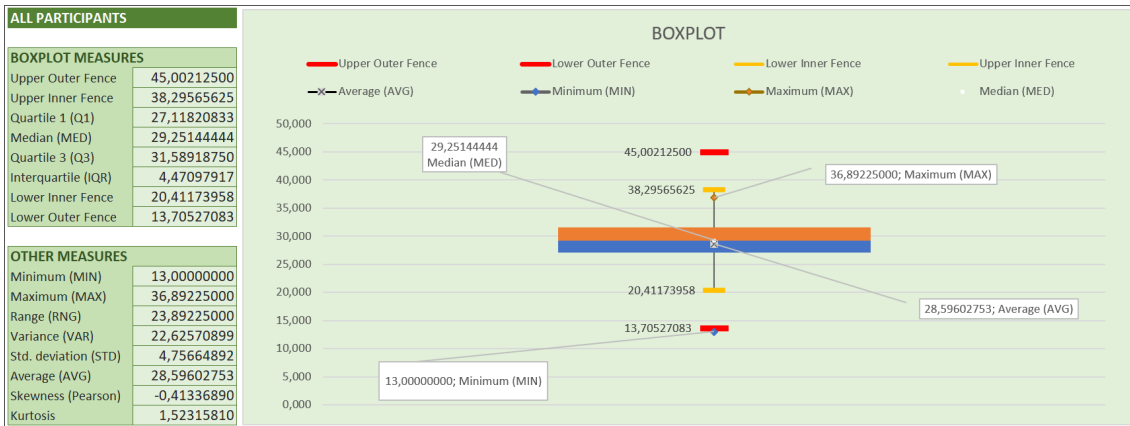


Figura 4.8: *Boxplot* do DQVA (e.g. sinal (TEMP) (unidade: °C))

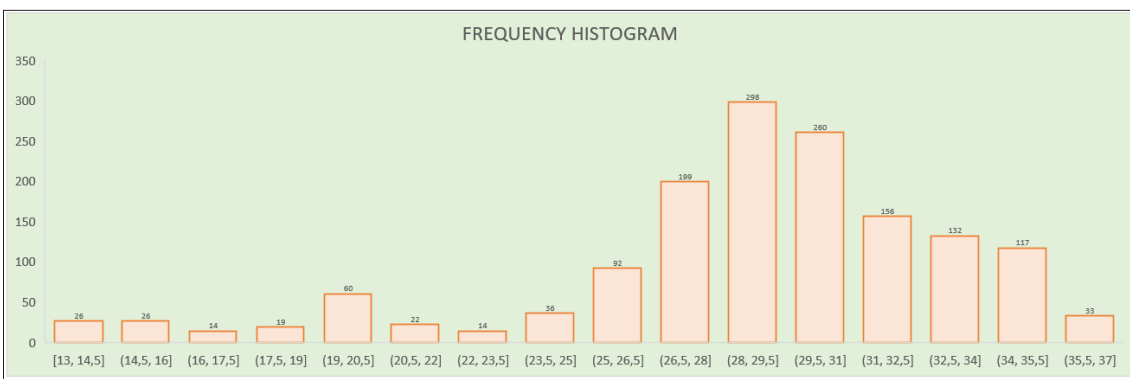


Figura 4.9: Histograma do DQVA (e.g. sinal (TEMP) (unidade: °C))

entre as várias classes, i.e., a sua assimetria e curtose (Tague, 2005). As classes do histograma não são sobrepostas e têm normalmente dimensão igual entre si (Pearson, 1895; Taylor, 2020). A Figura 4.9 apresenta o histograma de frequências do {DQVA}.

O diagrama de *Pareto* ordena as frequências acumuladas, permitindo analisar mais facilmente a contribuição de cada classe para a totalidade do conjunto de dados (Wilkinson, 2006). Pode ser entendido como um gráfico representativo do “*running total*” das frequências (maths, n.d.). Adiciona-se cada frequência às anteriores para obter a total cumulativo parcial, e representa-se esse acumulado numa série representada com uma linha contínua, cujo declive a cada classe documenta a contribuição dessa classe para os 100% de valores da distribuição. O diagrama de *Pareto* tenta representar o princípio de Vilfredo Pareto 80/20, em que 80% das consequências advêm de 20% das causas (Nisonger, 2008). A Figura 4.10 apresenta um diagrama de *Pareto* produzido pelo {DQVA}. A ordenação descendente de classes permite saber quais as (TEMP) mais comuns na distribuição, e a curva cumulativa de frequências, permite ter uma melhor perceção da contribuição de cada classe para o conjunto de dados global.

O *scatterplot* é um dos tipos de gráficos estatísticos mais versátil e polimórfico. O

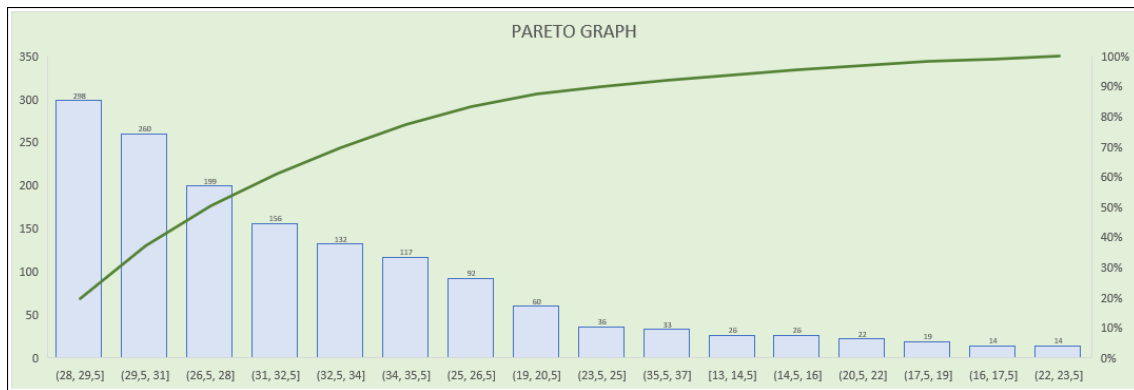


Figura 4.10: Diagrama de Pareto do DQVA (e.g. sinal (TEMP) (unidade: °C))

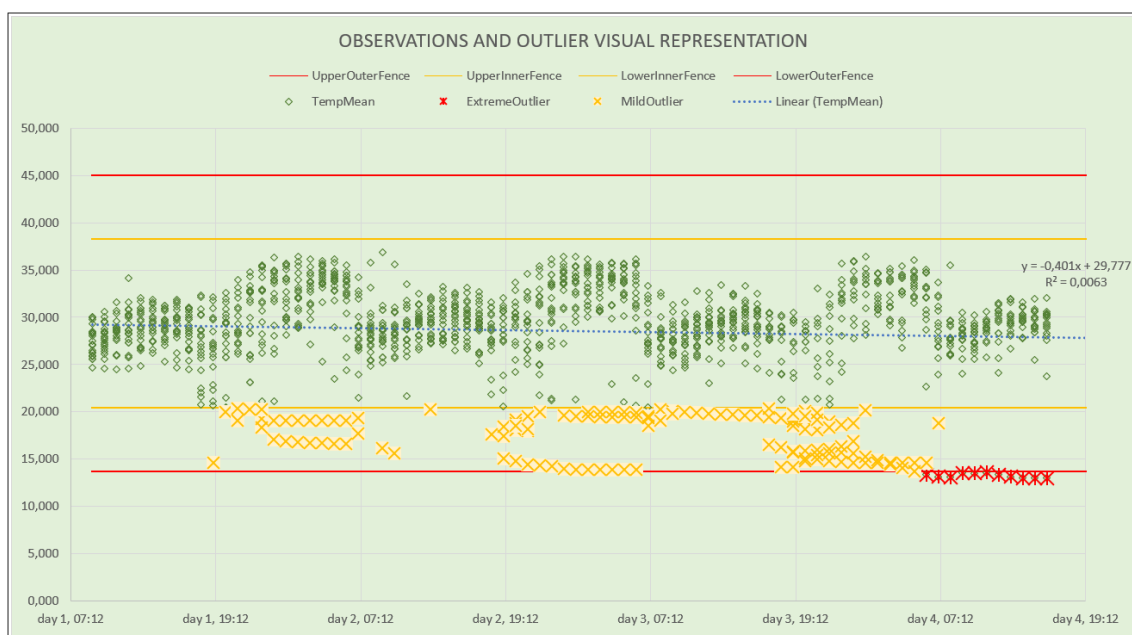


Figura 4.11: Scatterplot do DQVA (representação de objetos, fences e outliers) (e.g. sinal (TEMP) (unidade: °C))

scatterplot utiliza o plano cartesiano para apresentar os valores dos objetos da distribuição. Uma das grandes vantagens em utilizar o scatterplot é a possibilidade de ser possível avaliar a regularidade de valores visualmente (Friendly and Denis, 2005). A scatterplot da Figura 4.11, representa visualmente os objetos, cercas e outliers, existentes na variável de contexto em análise. O gráfico é composto por 7 séries de dados: i) os pequenos losangos a verde representam os valores dos objetos; ii) as linhas a vermelho e amarelo representam as cercas que definem o ponto a partir do qual os objetos podem ser considerados extreme outlier ou mild outlier; e iii) as cruces a vermelho e amarelo representam objetos candidatos a extreme outliers e a mild outliers. O efeito empilhado dos losangos verdes que compõem a série principal, justifica-se com o facto de os dados provenientes do {EDAMOVE} terem já sido agregados com $SI = 1h$ pelo {DATAANALYSER}.

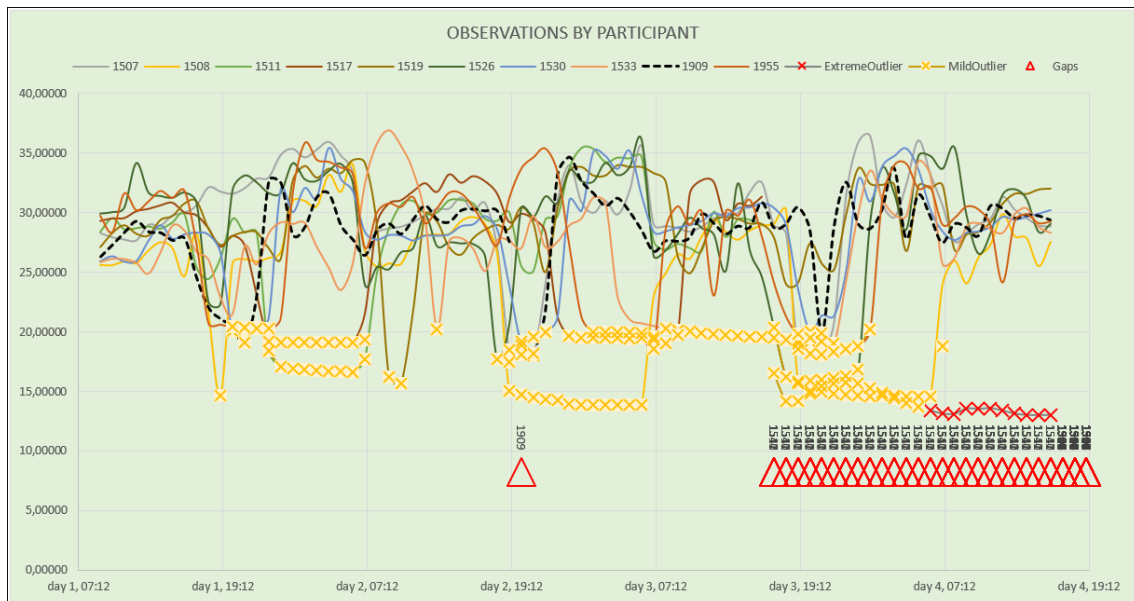


Figura 4.12: Scatterplot do DQVA (comparação de sinal entre participantes) (e.g. sinal (TEMP) (unidade: °C))

A visualização da Figura 4.12 compara o sinal recolhido entre vários participantes. Para facilitar a leitura, apenas são representados 10 dos 20 participantes da experiência. São várias as séries que compõem o gráfico: i) cada linha representa um participante; ii) os *outliers* são também representados com cruces vermelhas e amarelas como na Figura 4.11; e iii) os *gaps* de sinal são evidenciados através dos triângulos vermelhos. A título de exemplo, foram apagados temporariamente os dados das 19:00 do 2º dia de recolha do participante 1909. A linha respetiva foi reformatada para tracejado a preto, com o objetivo de a destacar e assim facilitar a análise dessa série de dados. Por questões de agenda de participantes e organizações, não foi possível recolher dados de contexto no último dia de recolha de quatro participantes. Os *gaps* de sinal visíveis no gráfico a partir das 17:30 do 3º dia de recolha (sequência de triângulos vermelhos à direita do gráfico) evidenciam essa ausência de dados provocada pela interrupção da recolha.

O gráfico representativo dos objetos por participante é personalizável. Além de ser possível selecionar quais os participantes (séries) a visualizar, também é possível mostrar ou ocultar: i) a série representativa dos *gaps* de sinal; ii) a série representativa dos *mild outliers*; e iii) a série representativa dos *extreme outliers*.

O {DQVA} também contém uma visualização que permite analisar os dados de uma variável por organização. Os dados dos participantes de cada organização são agregados pela média a cada SI. Nesta visualização, também é possível representar os *gaps* e os *outliers*.

Os diferentes recursos calculados pelo {DQVA} (cf. medidas estatísticas agregadas e visualizações), facilitaram o processo de deteção de anomalias nos dados nesta fase de

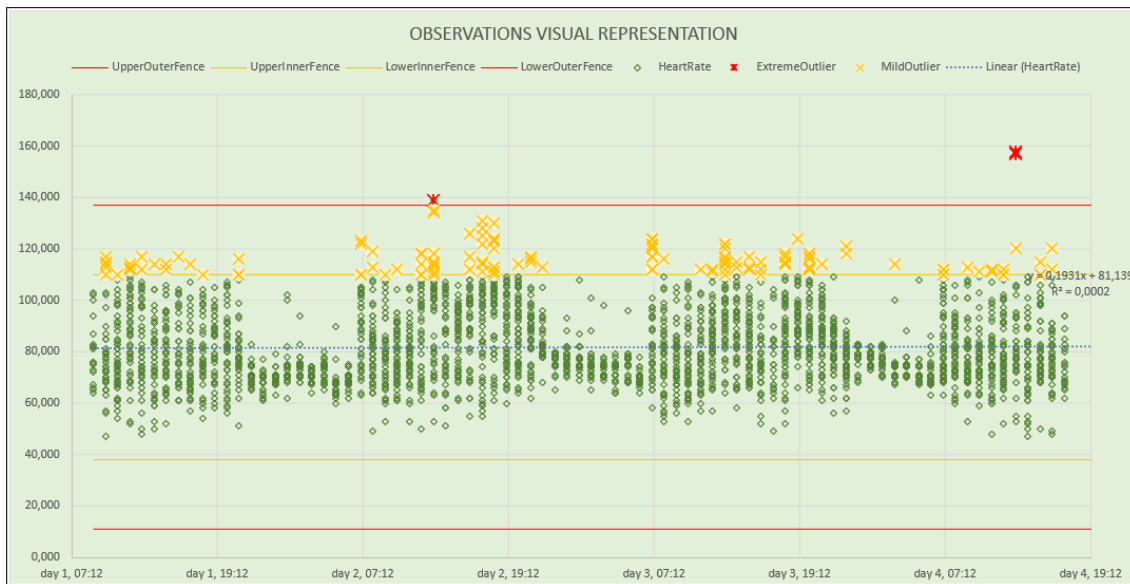


Figura 4.13: Visualização do $\{DQVA\}$ com *extreme* e *mild outliers* (sinal (HR) do participante 1510) (unidade: BPM)

integração do *dataset* final. A flexibilidade de se utilizar uma folha de cálculo, permitiu incluir ou excluir séries dados, para: i) comparar facilmente sinais entre participantes e organizações; ii) detetar *gaps* de sinal em cada participante; e iii) identificar de forma mais rápida quais os pontos de dados candidatos a *outlier*.

4.5 Análise

Neste capítulo apresentou-se a fase de pré-processamento dos dados recolhidos na experiência. Descreveu-se o processo de integração dos dados e os problemas encontrados durante a integração, depois resumiram-se as técnicas aplicadas sobre os dados e as propriedades extraídas. Por fim, apresentou-se o $\{DQVA\}$ e o $\{DQVA-PP\}$, ferramentas criadas pelo autor para avaliar de forma mais rápida e simples a qualidade dos dados (Carvalho et al., 2021).

Nesta fase de pré-processamento, foi dada uma atenção especial à qualidade dos dados. Havia consciência da importância da qualidade dos dados do *dataset* para a eficiência algorítmica. Neste contexto, o objetivo principal na fase do pré-processamento, foi conferir qualidade aos dados do *dataset*. Pretendia-se integrar todos os dados componentes do *dataset raw* recolhido na experiência de recolha apresentada no Capítulo 3, aplicar técnicas de pré-processamento para mitigar os problemas que pudessem existir, e adicionar propriedades extraídas para enriquecer o *dataset* final. No percurso, pretendia-se anotar o *dataset* com o objetivo de documentar as técnicas aplicadas.

O período de recolha de dados de contexto estendeu-se por mais de dois meses. Além disso, apesar da vontade em não se utilizar instrumentos obstrutivos no processo

de recolha, acabou por se recolher o sinal (EDA) através de elétrodos colados na mão dos participantes. Dada a importância defendida na literatura deste sinal no reconhecimento de emoções, e a indisponibilidade de alternativas de recolha menos obstrutivas, decidiu-se recolher o (EDA) através do sensor {EDAMOVE}, aumentando a preocupação com a qualidade dos dados. Considerando a qualidade pretendida para o *dataset* final, a fase de pré-processamento avizinhava-se exigente e rigorosa. Por esse motivo, fizeram-se várias visitas de acompanhamento em cada iteração de recolha. Por várias vezes, alguns procedimentos de utilização dos instrumentos pelos participantes tiveram que ser revistos, o que aumentava o receio na provável existência de problemas de dados a resolver na fase de pré-processamento.

O pré-processamento iniciou com a integração de todos os ficheiros do *dataset raw* numa base de dados relacional. Pretendia-se utilizar as *constraints* do DBMS para dar consistência aos dados. Além disso, a estratégia de implementar uma hierarquia de *views* para permitir abstrair a aplicação de técnicas de pré-processamento e cálculo de propriedades extraídas de uma forma standardizada, poderia promover que a fase de pré-processamento fosse mais rápida e menos exposta ao erro. Esta estratégia permitiria pré-processar os dados anotando as alterações efetuadas, sem alterar os dados originalmente recolhidos de contexto. O *dataset raw* era composto por vários ficheiros de diferentes tipos. Para os integrar, desenvolveu-se um projeto ETL {MISCELLANEOUSIMPORT} em {MSSIS}, e dois projetos para importação dos ficheiros *sqlite* produzidos pelas aplicações {WINDOWSACTIVITY} e pelo {ANDROIDANALYTICS}: o {WINDOWSACTIVITYIMPORT}, e o {ANDROIDANALYTICSIMPORT}. Estes projetos são resumidos na Tabela 4.8 e perfazem a contribuição C07.

Documento	Descrição	Contribuição
{WINDOWSACTIVITYIMPORT}	Instrumento criado pelo autor para integração dos dados dos ficheiros *WindowsActivity*.sqlite, recolhidos de contexto pelo {WINDOWSACTIVITY}.	C07
{ANDROIDANALYTICSIMPORT}	Instrumento criado pelo autor para integração dos dados dos ficheiros *AndroidSocial*.sqlite, recolhidos de contexto pelo {ANDROIDANALYTICS}.	C07
{MISCELLANEOUSIMPORT}	Projeto ETL implementado em {MSSIS} para integração de ficheiros CSV e XLS?, componentes do <i>dataset raw</i> .	C07

Tabela 4.8: Resumo dos instrumentos utilizados na integração de dados

Foram vários os problemas encontrados no *dataset raw*. Existiam registos em duplicado, valores em falta, etc. No entanto, talvez o problema detetado de maior relevância, tenha sido com os dados recolhidos através do instrumento {LABOR} (ver secção 4.1.2). A indisponibilidade e boa vontade demonstrada pelas pessoas dos departamentos de recursos no preenchimento do questionário, resultou em informação contraditória e inconsistente. Na tentativa de corrigir o problema para não se perder estes dados, solicitou-se a revisão da informação prestada. No entanto, a forma imprecisa com que os dados foram

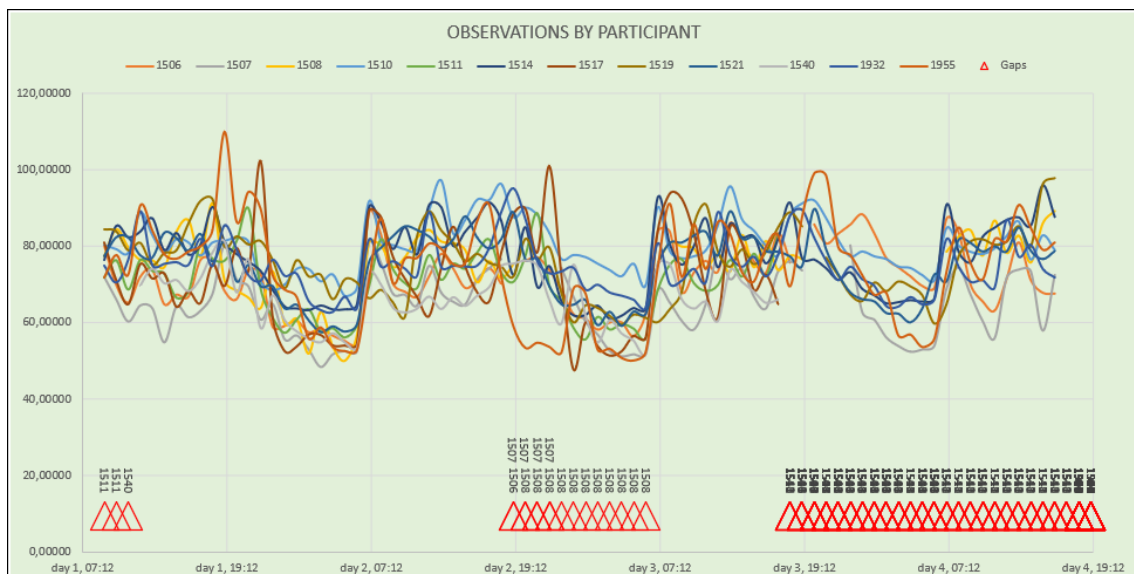


Figura 4.14: Visualização do $\{DQVA\}$ com *gaps* de sinal (HR) em alguns participantes (unidade: BPM)

retificados, aumentou ainda mais as dúvidas quanto à validade dos dados recolhidos pelo $\{LABOR\}$, e desincentivaram o autor em utilizá-los. Estes dados foram integrados na base de dados e serão disponibilizados no *dataset* final integrado e anotado. No entanto, não serão utilizados pelo autor no futuro, para detecção automática do bem-estar em trabalhadores de escritório.

Considerando a quantidade de dados recolhidos, encontrar os problemas poderia ser um processo sujeito a erro humano e bastante demorado. Foi nesse contexto que se decidiu criar o $\{DQVA\}$. Precisava-se de uma ferramenta que tornasse mais simples e rápido o processo de avaliação da qualidade dos dados do *dataset*. O $\{DQVA\}$ calcula automaticamente as medidas estatísticas de agregação comumente utilizadas pelos investigadores, mas também disponibiliza instrumentos visuais para facilitar a descoberta de mais informação acerca dos próprios dados (e.g. comparação de comportamento do sinal entre participantes). Além de calcular as várias medidas estatísticas normalmente utilizadas, permite também analisar os dados de forma visual: i) comparar o sinal em análise entre participantes e organizações; ii) verificar a existência de *outliers*; iii) detetar falhas de sinal; iv) analisar a distribuição através de histogramas; etc.

A utilização do $\{DQVA\}$ revelou-se muito importante durante a fase de pré-processamento. Permitiu detetar anomalias nos dados, comparar padrões de sinal entre participantes e organizações, detetar *outliers* e falhas, etc. O $\{DQVA\}$ promoveu a descoberta de *outliers* e falhas de sinal, permitiu que os sinais dos diferentes participantes e organizações, pudessem ser comparados através de quadros resumo e visualizações, etc. A título de exemplo, através da visualização da Figura 4.13, foi possível perceber a existência de dois *extreme outliers* e vários *mild outliers* no sinal (HR) do participante 1510. Outro

exemplo, detetou-se problemas na utilização da {MIBAND} pelos participantes 1506, 1507 e 1508, através da visualização apresentada na Figura 4.14. Nesta visualização, são filtrados apenas os sinais (HR) de alguns participantes, com o objetivo de facilitar a leitura.

Criou-se também o {DQVA-PP}. Este instrumento é uma variante do {DQVA}. A diferença entre ambos é que o {DQVA} carrega os dados para processamento a partir das tabelas, e o {DQVA-PP} carrega os dados a partir das *views*. O {DQVA} e {DQVA-PP} são resumidos na Tabela 4.9 e constituem a contribuição C08.

Documento	Descrição	Contribuição
{DQVA}	Instrumento criado pelo autor para avaliar a qualidade dos dados durante a fase de integração do <i>dataset</i> na base de dados relacional. Os dados carregados pelo {DQVA} tem origem nas tabelas da base de dados.	C08
{DQVA-PP}	Variante do {DQVA}, o {DQVA-PP} carrega dados a partir das <i>views</i> e, por isso, permitiu avaliar a qualidade dos dados durante e após a aplicação de técnicas de pré-processamento.	C08

Tabela 4.9: Instrumentos criados para avaliar a qualidade de dados

Foram ainda utilizados outros instrumentos, como ferramentas de suporte à criação de *views* e exploração de dados. A Tabela 4.10 resume os instrumentos utilizados como suporte à fase de pré-processamento e fase exploratória de dados.

Documento	Descrição
{SSMS}	<i>Microsoft Management Studio 2017</i> , utilizado na criação de <i>views</i> , e na exploração de dados através de <i>queries</i> .
{SPSS}	<i>Statistical Package for the Social Sciences</i> , utilizado no pré-processamento como suporte à exploração de dados, e na análise exploratória de dados relatada na secção 5.5.

Tabela 4.10: Outros instrumentos utilizados na fase de pré-processamento e na fase exploratória do *dataset* final

Capítulo 5

RESULTADOS E ANÁLISE

O objetivo principal desta tese era contribuir com um *dataset* multimodal pré-processado, preparado para ser utilizado, no futuro, na deteção automática do bem-estar em trabalhadores de escritório. Este capítulo apresenta o *dataset* final resultante da fase de pré-processamento e adição de propriedades extraídas, descrita no Capítulo 4. São apresentados detalhes relacionados com a estrutura do modelo relacional da base de dados integradora, e são dados mais pormenores acerca da arquitetura de *views* para facilitar a exploração de dados. Neste capítulo é também feita a caracterização do perfil dos participantes e dos dados. Apresenta-se também o resultado da análise exploratória de dados, efetuada sobre o *dataset* final.

Neste capítulo apresenta-se com maior detalhe, a base de dados criada para integrar os dados, e a estratégia de *views* implementada para facilitar o pré-processamento e exploração dos dados. Os participantes e os dados são caracterizados, e é feita uma análise exploratória sobre os mesmos. Optou-se por integrar os dados numa base de dados *Microsoft SQL Server 2017* {MSSQL}, mas poderia ter-se escolhido qualquer outra tecnologia de bases de dados relacionais, que permitisse a implementação de várias camadas *views*. Recorreu-se a uma base de dados relacional com o objetivo de utilizar os mecanismos promotores de consistência da informação durante a fase de integração e, também, para facilitar a extração de dados. A utilização de *queries* SQL para extrair dados, pode incidir sobre qualquer nível de *views*, relacionando várias tabelas e/ou *views* em simultâneo. Nesta implementação de base de dados, não foram consideradas as questões de eficiência na gestão de recursos do DBMS e da base de dados.

O *dataset* final apresentado como contribuição neste capítulo, não engloba todos os dados presentes no *dataset raw*. Algumas variáveis não foram consideradas importantes para os fins pretendidos neste projeto (e.g. os dados do (ACC) recolhidos pelo {EDAMOVE} foram utilizados pelo {DATAANALYSER} para processamento dos dados do (EDA)). Também não se incluem no *dataset* final os sinais redundantes (e.g. o número de passos foi recolhido pela {MIBAND}, e pelo {EDAMOVE}). Os dados do {LABOR} foram integrados no *dataset*, mas não serão utilizados no futuro. Como já referido na secção 4.5,

o autor desinteressou-se por estes dados por causa das dúvidas acerca da sua validade. Os questionários foram preenchidos de forma pouco consistente, e foram retificados de forma pouco rigorosa.

O *dataset* final é materializado num ficheiro *backup* do {MSSQL}, e num ficheiro CSV. Atendendo à grande quantidade de dados do *dataset*, recomenda-se a sua exploração através de *queries* SQL incidentes sobre a base de dados relacional. Os dados constantes do *dataset* final, foram anonimizados. A tabela `Organizations` e `Participants` contém um campo `Name` com abreviaturas genéricas para ajudar o autor no tratamento de dados. Estas abreviaturas ajudaram na fase de pré-processamento para contextualizar o processo de recolha. No entanto, quando lidos por terceiros, não identificam ninguém em concreto.

Este capítulo é composto por seis secções. A primeira apresenta com maior detalhe a estrutura do *dataset* final apresentado como contribuição deste capítulo. A segunda aborda pormenores das *views* implementadas apresentando alguns excertos de dados. As secções três e quatro caracterizam os participantes e os dados. A quinta secção apresenta os resultados da análise exploratória sobre os dados do *dataset*. Por fim, a análise final deste capítulo, resume o processo que conduziu à disponibilização do *dataset* final, última contribuição desta tese.

5.1 Estrutura do *dataset*

Durante a fase de integração de dados, os vários ficheiros componentes do *dataset raw*, foram integrados na base de dados relacional integradora. Cada registo integrado faz referencia à organização e participante respetivo, e regista também o momento em que cada objeto foi recolhido de contexto. Além destes atributos, cada registo guarda também os valores recolhidos, e outros dados relacionados com a estratégia criada para o pré-processamento anotado do *dataset*. Nesta secção, aborda-se em detalhe o *dataset* final, apresentado como contribuição neste capítulo.

A primeira fase do pré-processamento foi a integração do *dataset raw* na base de dados relacional. Todos os dados recolhidos de contexto foram integrados em tabelas espelho de estrutura equivalente aos ficheiros criados no processo de exportação dos dados recolhidos por cada instrumento & sensor, no final de cada iteração de recolha. Optou-se por utilizar uma base de dados relacional para integrar os dados: i) por causa dos mecanismos de consistência que oferece através das várias *constraints* que impõe (e.g. integridade referencial dos dados); e ii) para facilitar a exploração de dados através de *queries* SQL que permitem o relacionamento entre tabelas e *views*. A escolha de uma base de dados relacional produziu frutos logo na fase de integração, quando não permitiu a integração de registos duplicados existentes nos ficheiros com dados *raw* (ver secção 4.4). No entanto, também facilitou a implementação da estratégia assente em *views*, para abstração da aplicação das técnicas de pré-processamento.

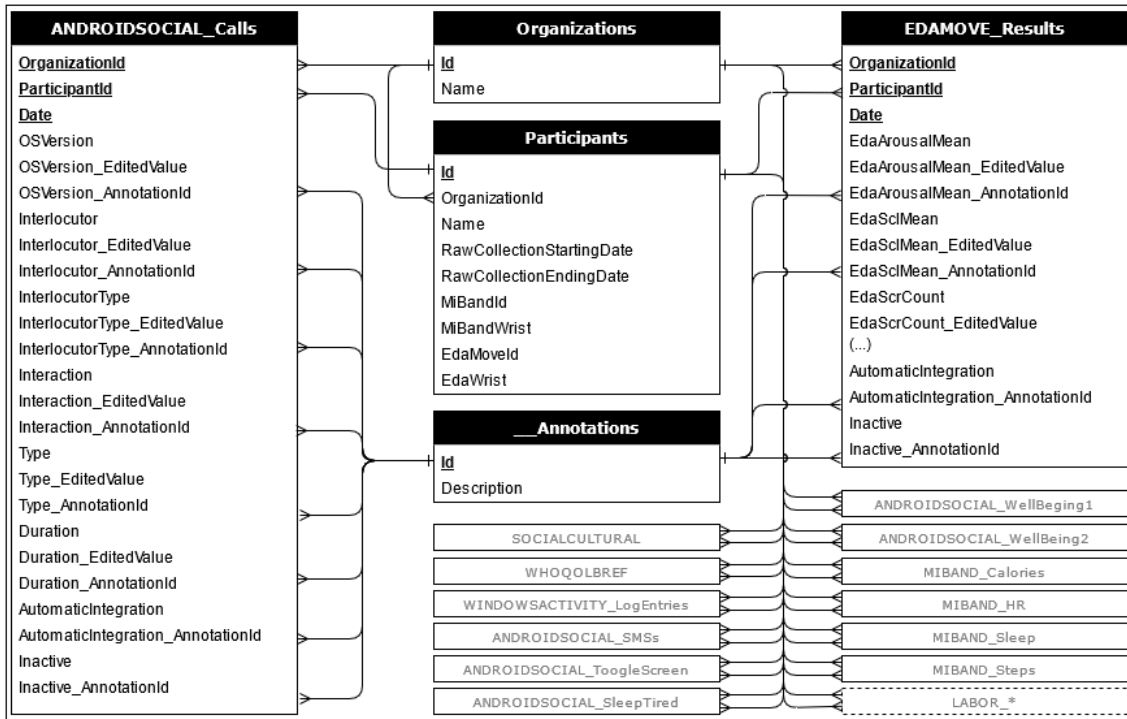


Figura 5.1: Excerto do modelo ER do *dataset* final

O modelo ER da base de dados relacional, cujo extrato é apresentado pela Figura 5.1, contempla três tipos de tabelas: i) a tabela *Organizations* e *Participants* que identificam, respetivamente, as organizações onde existiram iterações de recolha, e os participantes dos quais se recolheu dados de contexto; ii) a tabela *_Annotations* onde foram criadas as várias *tags* de anotações possíveis de referenciar, aquando da aplicação de uma qualquer técnica de pré-processamento que implicasse a edição dos dados originalmente recolhidos de contexto; e iii) as restantes tabelas que referenciam as anteriores, e que registam os dados recolhidos de contexto. A estratégia de anotações seguida pelo autor, foi aplicada em todas as tabelas que contêm variáveis recolhidas de contexto, com a exceção das recolhidas através dos instrumentos {*WHOQOLBREF*} e {*SOCIOCULTURAL*}. Os problemas existentes nestes questionários, além de residuais, foram resolvidos durante a sua integração na base de dados relacional. A Tabela 5.1 em conjunto com o modelo ER parcial representado na Figura 5.1, fazem um resumo das várias tabelas que compõem o *dataset* final, apresentado neste capítulo.

Para facilitar a leitura da Figura 5.1, várias tabelas são representadas no modelo ER parcial, apenas pelo seu nome (e.g. *SOCIOCULTURAL*, *WHOQOLBREF*). A estrutura completa pode ser conhecida em detalhe no dicionário de dados apresentado no Anexo 7.21. As tabelas *LABOR_** são representadas a tracejado. Como já referido, apesar dos dados terem sido integrados, sugere-se a sua utilização com cautela, por causa dos problemas ocorridos durante a recolha (ver secção 4.1.2).

Tabela	Descrição
Organizations	Cada registo corresponde a uma organização onde se recolheram dados de contexto. Os registos desta tabela foram inseridos manualmente.
Participants	Cada registo corresponde a um participante na experiência de recolha de dados de contexto. Os registos desta tabela foram inseridos manualmente.
...Annotations	Cada registo corresponde a uma <i>tag</i> possível de ser referenciada para anotação da aplicação de uma técnica de pré-processamento.
WINDOWSACTIVITY_LogEntries	Cada registo contém informação relacionada com um evento registado pelo SO, relativo à utilização do computador de trabalho.
ANDROIDSOCIAL.WellBeing1	Cada registo corresponde a uma resposta, de um participante, ao questionário {WELLBEING1}.
ANDROIDSOCIAL.WellBeing2	Cada registo corresponde a uma resposta, de um participante, ao questionário {WELLBEING2}.
ANDROIDSOCIAL.SleepTired	Cada registo corresponde a uma resposta, de um participante, ao questionário {SLEEPTIRED}.
ANDROIDSOCIAL.Calls	Cada registo representa uma chamada telefónica efetuada ou recebida no <i>smartphone</i> , e a qualidade de interação social ocorrida através do preenchimento do {CALLINTERACTIONS}
ANDROIDSOCIAL.SMSs	Cada registo representa um SMS enviado ou recebido pelo <i>smartphone</i> de um participante.
ANDROIDSOCIAL.ToogleScreen	Cada registo representa um <i>Screen on</i> ou <i>Screen off</i> do <i>smartphone</i> de um participante.
MIBAND.Calories	Cada registo contém o gasto calórico diário de cada participante.
MIBAND.HR	Cada registo contém a frequência cardíaca de um participante a cada minuto.
MIBAND.Sleep	Contém informação acerca do sono dos participante. Cada registo corresponde a uma noite ou grande período de sono (e.g. na noite do segundo dia de recolha do participante 1503, existem dois períodos de sono).
MIBAND.SleepDetails	Cada registo contém os períodos de sono leve e profundo, de cada participante. Esta tabela contém o detalhe de cada registo da tabela <i>MIBAND.Sleep</i> .
MIBAND.SleepRaw	Cada registo contém informação para cálculo dos períodos de sono e classificação como leve ou profundo. Os registos representam informação minuto a minuto por participante.
MIBAND.Steps	Cada registo contém o número de passos dados a cada 10 minutos por cada participante.
EDAMOVE.Results	Cada registo contém a média de $SI = 1h$ de várias variáveis de contexto, recolhidas pelo instrumento {EDAMOVE}, de um participante. O pré-processamento dos dados recolhidos por este instrumento, foi feito pelo {DATAANALYSER} no final de cada iteração (ver secção 4.3.1) (movisens GmbH, 2018a).
WHOQOLBREF	Cada registo contém a resposta de um questionário aplicado a um participante.
SOCIOCULTURAL	Cada registo contém a resposta de um questionário aplicado a um participante.

LABOR*	Os dados dos questionários {LABOR}, apesar de integrados, não serão utilizados pelo autor (ver secção 4.5).
--------	---

Tabela 5.1: Tabelas do modelo ER do *dataset* final

A tabela `Participants`, além de identificar de forma anónima os trabalhadores de escritório que aceitaram participar na experiência de recolha de dados de contexto, contém também outros atributos que contextualizam o processo de recolha: i) os atributos `RawCollectionStartingDate` e `RawCollectionEndingDate`, indicam a data de início e fim de recolha de dados de contexto de cada participantes; ii) os valores de `MiBandId` e `MiBandWrist`, identificam a pulseira e o pulso utilizado para recolha; e iii) o `EdaMoveId` e `EdaWrist`, identificam o sensor {EDAMOVE} e pulso (mão) utilizado para recolha.

As tabelas cujo nome inicia com o nome de um instrumento, são tabelas que contêm objetos recolhidos de contexto por aquele instrumento ou sensor de recolha. As chaves primárias destas tabelas são compostas pelos atributos: i) `OrganizationId` que referencia a tabela `Organizations`; ii) `ParticipantId` que referencia a tabela `Participants`; e iii) `Date` que regista o momento de recolha dos dados pelo instrumento & sensor, com precisão de até 100 nanosegundos. Além dos atributos componentes da chave primária composta, cada tabela contém também atributos para registo dos valores recolhidos de contexto. Para cada um destes valores, existe um segundo atributo com o mesmo nome mas com o sufixo `_EditedValue`, e um terceiro com o sufixo `_AnnotationId`. Por defeito, o conteúdo destes atributos é `NULL`. Apenas contêm um valor diferente de `NULL`, os atributos que foram objeto da aplicação de técnicas de pré-processamento que implicaram a edição do valor originalmente recolhido de contexto. Nos atributos principais ficou sempre guardado o valor recolhido através do instrumento & sensor, sem qualquer pré-processamento. Os atributos com sufixo `_EditedValue`, se diferente de `NULL`, contêm os valores resultantes da aplicação das técnicas de pré-processamento, caso essas técnicas tenham impactado os valores recolhidos de contexto. Os atributos `*_AnnotationId` fazem parte da estratégia da anotação do *dataset*. Caso tenha sido aplicada uma técnica sobre o valor recolhido de contexto (i.e. caso exista um valor no atributo `*_EditedValue`), a *tag* existente no atributo `*_AnnotationId` correspondente, descreve de forma sucinta o tipo de processamento efetuado naquele valor. Assim, o *dataset* final contém os dados *raw*, os dados pré-processados, e as anotações.

Os atributos `AutomaticIntegration`, `AutomaticIntegration.AnnotationId`, `Inactive` e `Inactive.AnnotationId`, existem em todas as tabelas que contêm dados recolhidos de contexto. Os dois primeiros foram utilizados para distinguir os objetos recolhidos pelos instrumentos originais de recolha, dos que foram recolhidos através de formulários *backup*. Os atributos `Inactive*`, foram utilizados para inativar e anotar o motivo da inativação de objetos (e.g. `[DISCARDATA]`, `[-OUTLIER]`).

A Tabela 5.2 apresenta um excerto do dicionário de dados da base de dados {MSSQL}

onde existe o *dataset* final. O dicionário de dados completo, pode ser consultado no Anexo 7.21. Para facilitar a leitura, não se representam os atributos *_EditedValue e *_AnnotationId. A existência destes atributos é indicada à frente do nome do atributo original. A letra β indica que, para o atributo indicado, existe o segundo atributo _EditedValue. A letra Φ faz o mesmo para o atributo _AnnotationId.

Como já referido anteriormente, o pré-processamento dos dados recolhidos pelos sensores {EDAMOVE}, foi feito através da aplicação {DATAANALYSER} após cada iteração de recolha. Os dados integrados na base de dados relacional, estavam contidos no ficheiro EDAMOVE_Results.xlsx criado durante esse pré-processamento. O processamento do {DATAANALYSER} agregou os objetos com $SI = 1h$ através da média. Justifica-se assim o nome da tabela do modelo relacional (i.e. EDAMOVE_Results), e também o facto de, alguns atributos desta tabela, conterem a palavra *mean* no seu nome.

Atributo	Tipo	Descrição
Organizations		
PK Id	Inteiro	Código da organização.
Name	Texto	Nome pseudonimizado da organização.
Participants		
PK Id	Inteiro	Número do participante
FK OrganizationId	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
Name	Texto	Nome pseudonimizado do participante.
RawCollectionStartingDate	Data / hora	Início do período de recolha.
RawCollectionEndingDate	Data / hora	Fim do período de recolha.
MiBandId	Inteiro	Id da pulseira {MIBAND} utilizada na recolha.
MiBandWrist	Texto	{ <i>Left, Right</i> }
EdaMoveId	Inteiro	Id do sensor {EDAMOVE} utilizado na recolha.
EdaWrist	Texto	{ <i>Left, Right</i> }
_Annotations		
PK Id	Texto	Tag a utilizar nas anotações.
Descrição	Texto	Descrição da anotação.
WINDOWSACTIVITY_LogEntries		
PK Date	Data / hora	Data e hora de recolha de contexto.
PK FK OrganizationId	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK ParticipantId	Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
Computer	$\beta \Phi$ Texto	Identificação do computador de recolha.
AutomaticIntegration	Φ Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
Inactive	Φ Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_WellBeing1		
PK Date	Data / hora	Data e hora de recolha de contexto.

¹ Se *false*, o objeto foi recolhido através de formulário *backup* (substituição do instrumento principal).

² Se *false*, o objeto não é considerado no *dataset* final (e.g. foi descartado).

PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	EasyRelations	$\beta \Phi$	Inteiro	Resposta à questão " <i>Relacionei-me facilmente com as pessoas à minha volta</i> ".
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_WellBeing2				
	EmotionallyBalanced	$\beta \Phi$	Inteiro	Resposta à questão " <i>Senti-me emocionalmente equilibrado</i> ".
	PeaceWithMyself	$\beta \Phi$	Inteiro	Resposta à questão " <i>Senti-me bem, em paz comigo próprio</i> ".
	ImConfident	$\beta \Phi$	Inteiro	Resposta à questão " <i>Senti-me confiante</i> ".
ANDROIDANALYTICS_SleepTired				
	SleepQuality	$\beta \Phi$	Inteiro	Resposta à questão " <i>Dormi muito bem e sinto que o meu sono foi totalmente reparador</i> ".
	FeelingTired	$\beta \Phi$	Inteiro	Resposta à questão " <i>Sinto-me totalmente descansado após esta noite de sono</i> ".
ANDROIDANALYTICS_Calls				
	Interlocutor	$\beta \Phi$	Texto	Número de telefone do <i>smartphone</i> de recolha (anonimizado).
	InterlocutorType	$\beta \Phi$	Texto	{ <i>Spouse-Sons-CloseFriends, RemainingFamily-OtherFriends, Others, Not-classified</i> }.
	Interaction	$\beta \Phi$	Texto	{ <i>Negative, Neutral, Positive</i> }.
	Type	$\beta \Phi$	Texto	{ <i>MissedCall, InitiatedOnly, Incoming, Outgoing</i> }
	Duration	$\beta \Phi$	Inteiro	Duração de cada chamada telefónica.
ANDROIDANALYTICS_SMSs				
	Interlocutor	$\beta \Phi$	Texto	Número de telefone do <i>smartphone</i> de recolha (anonimizado).
	InterlocutorType	$\beta \Phi$	Texto	{ <i>Spouse-Sons-CloseFriends, RemainingFamily-OtherFriends, Others, Not-classified</i> }.
	Type	$\beta \Phi$	Texto	{ <i>Incoming, Outgoing</i> }.
	Word	$\beta \Phi$	Inteiro	Número de palavras da mensagem.
	Char	$\beta \Phi$	Inteiro	Número de caracteres da mensagem.
ANDROIDANALYTICS_ToogleScreen				
	ScreenIsOn	$\beta \Phi$	Booleano	Representa um evento de ligar o ecrã do <i>smartphone</i> (<i>true = on, false = off</i>).
MIBAND_Calories				
	BMR	$\beta \Phi$	Inteiro	<i>Basal Metabolic Rate</i> (BMR) (<i>Harris-Benedict Equation</i>).
	Steps_Calories	$\beta \Phi$	Inteiro	Gasto de calorias durante a atividade de caminhar.
	Exercises_Calories	$\beta \Phi$	Inteiro	Gasto de calorias durante o exercício físico.
	Total	$\beta \Phi$	Inteiro	Gasto calórico diário.
MIBAND_HR				
	HeartRate	$\beta \Phi$	Inteiro	Frequência cardíaca.

MIBAND_Sleep			
TotalSleep	$\beta \Phi$	Inteiro	Total de minutos de sono.
DeepSleep	$\beta \Phi$	Inteiro	Total de minutos de sono profundo.
LightSleep	$\beta \Phi$	Inteiro	Total de minutos de sono leve.
MIBAND_Steps			
StepsDayAccumulation	$\beta \Phi$	Inteiro	Acumulado diário de passos (registo a cada 10 minutos).
Steps	$\beta \Phi$	Inteiro	Número de passos percorridos nos últimos 10 minutos.
EDAMOVE_Results			
EdaArousalMean	$\beta \Phi$	Real	Média dos valores <i>Arousal</i> , calculados com base nas SCR (movisens GmbH, 2018a).
EdaSclMean	$\beta \Phi$	Real	Média dos SCL (movisens GmbH, 2018a).
EdaScrAmplitudesMean	$\beta \Phi$	Real	Média das amplitudes dos SCR (movisens GmbH, 2018a).
EdaScrCount	$\beta \Phi$	Real	Média do número de SCR detetadas (movisens GmbH, 2018a).
EdaScrEnergiesMean	$\beta \Phi$	Real	Média da energia das SCR detetadas (movisens GmbH, 2018a).
EdaScrHalfRecoveryTimesMean	$\beta \Phi$	Real	Média dos tempos de recuperação até metade da amplitude de cada SCR detetada (movisens GmbH, 2018a).
EdaScrRiseTimesMean	$\beta \Phi$	Real	Média dos tempos de subida de cada SCR detetada (movisens GmbH, 2018a).
WHOQOLBREF			
DateOfBirth		Data	Data de nascimento.
Schooling		Texto	Nível de escolaridade.
Job		Texto	Profissão.
MaritalStatus		Texto	Estado civil.
IsCurrentlySick		Booleano	Estado de saúde (<i>true</i> = está doente).
Q01		Inteiro	Resposta dada à questão 01.
Qn		Inteiro	Resposta dada à questão...
SOCIOCULTURAL			
Age		Inteiro	Idade.
Gender		Texto	{ <i>Male, Female, Undefined</i> }.
Ethnicity		Texto	{ <i>Caucasian, Yellow, Black, Other</i> }.
CivilStatus		Texto	{ <i>Single, Married, Union of fact</i> }.
Religion		Texto	{ <i>Catholic, Evangelical, Jehovah's Witness, Anglican, Islamic, Other, None</i> }.
Living		Texto	{ <i>Alone, Direct family (spouse, sons), Indirect family (brothers or parents)</i> }.
Dependent1_Kinship		Texto	{ <i>Mother, Father, Son or daughter (stepson, adopted), Spouse, Other</i> }.
Dependent1_Age		Inteiro	Idade do dependente.
Dependent1_SameHouse		Booleano	<i>True</i> , se o dependente vive na mesma habitação que o participante.
Dependent1_MoneyDependency		Booleano	<i>True</i> , se o dependente depende financeiramente do participante.
Dependent1_DailyTasksDependency		Booleano	<i>True</i> , se o dependente depende do participante para realizar tarefas diárias.

Dependent1_SpecialNeeds	Booleano	<i>True</i> , se o dependente necessita de cuidados especiais.
HomeType	Texto	{ <i>Apartment, Dwelling, Farmhouse, Other</i> }.
Home_Environment	Texto	{ <i>Town house, Country house</i> }.
HomeWork_Distance	Inteiro	Distância de casa ao trabalho (km).
Neighbors_Problematic	Booleano	<i>True</i> , se tem vizinhos problemáticos.
Loans_HouseMorgage	Booleano	<i>True</i> , se o participante tem crédito a habitação.
Loans_CarLoan	Booleano	<i>True</i> , se tem crédito para compra de automóvel.
SocialSupport_Received	Inteiro	Autoapreciação do suporte social recebido.
WorkLife_Job_Count	Inteiro	Número de empregos desde início da vida ativa.
Society_ActiveParticipation	Booleano	<i>True</i> , o participante participa ativamente em partidos políticos, religião ou associações.
Society_RelevantRoles	Booleano	<i>True</i> , o participante desempenha papéis sociais de relevo em organizações sociais, políticas ou religiosas (e.g. juntas de freguesia, grupos desportivos).
CaffeineConsumption	Inteiro	Número aproximado de tomas de café diariamente.
AlcoholConsumption	Texto	{ <i>With friends, Only with meals, Never, Other</i> }.
DietPhysicalAppearance	Booleano	<i>True</i> , se o participante faz ou fez dieta para manutenção do aspeto físico.

Tabela 5.2: Dicionário de dados parcial, da base de dados do *dataset*

5.2 Arquitetura de vistas

Após a integração de todos os dados *raw* na base de dados relacional, utilizou-se uma estratégia assente em *views*, para abstrair a aplicação das várias técnicas de pré-processamento e cálculo de propriedades extraídas. O nível mais baixo de *views* permite consultar os dados pré-processados e anotados. Nos níveis mais altos podem ser consultados os dados pré-processados e as propriedades extraídas. Nesta secção apresenta-se com maior detalhe, a estratégia de *views* implementada durante a fase de pré-processamento.

A hierarquia de *views* criada, tinha o objetivo de implementar vistas em camadas sobre os dados, para abstrair o resultado da aplicação das várias técnicas durante a fase do pré-processamento, e cálculo de propriedades extraídas. A ideia passava por tornar possível a edição dos dados *raw* integrados, sem que isso implicasse a sobreposição dos valores originais. Além disso, essas alterações ficariam anotadas no próprio *dataset*. Desta forma, poder-se-ia consultar os dados originais recolhidos de contexto, os dados pré-processados (i.e. editados pela aplicação de técnicas de pré-processamento), e a anotação que melhor descreveria a alteração efetuada em cada um dos valores. A utilização de *views* facilitaria também a exploração de dados. O mecanismo de relacionamento do DBMS entre tabelas e/ou *views* dos vários níveis, iria facilitar a exploração de dados através de *queries* SQL. Num primeiro momento, as *views* iriam apoiar na análise de dados durante e após

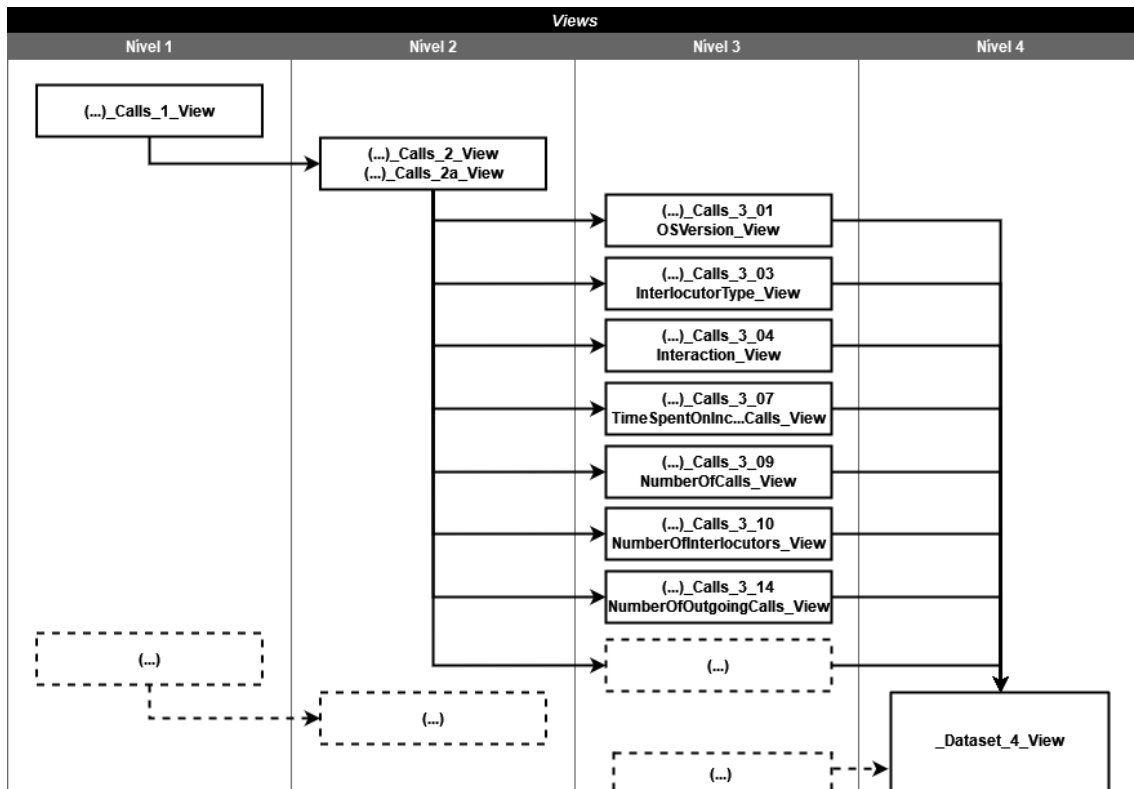


Figura 5.2: Excerto da hierarquia de *views* (dados da tabela ANDROIDSOCIAL_Calls)

a aplicação das técnicas de pré-processamento (e.g. {DQVA-PP}). Posteriormente, poderiam ser úteis para explorar os dados do *dataset* final, e seleção de atributos a utilizar como *input* no processamento algorítmico.

Outra vantagem que se pretendia obter através da estratégia das *views*, estava relacionada com a uniformização de código. O facto de se utilizar a mesma estratégia de forma transversal nas várias *views* de cada nível, iria facilitar a exploração de dados do *dataset* final. Assim, quando alguém pretendesse explorar o *dataset* apresentado como contribuição neste capítulo, bastaria estudar a implementação de uma das *views* de cada camada, para conhecer, de forma genérica, o conteúdo e organização das restantes. Considerando a dimensão do *dataset*, esta estratégia poderia promover a utilização destes dados noutras investigações. Além desta vantagem, a implementação homogénea das *views* da mesma camada, iria aumentar o ritmo na fase de pré-processamento. A uniformidade algorítmica, facilitaria a leitura e manutenção no futuro, mas também iria promover que a implementação das várias *views*, sobre os restantes sinais, fosse mais rápida. A criação de uma *view*, seria pouco mais do que uma tarefa de adaptação de código à realidade dos dados de cada sinal.

A Figura 5.2 apresenta um excerto da arquitetura de *views* implementada. Para facilitar a leitura, representa-se apenas algumas *views* da hierarquia iniciada na tabela ANDROID SOCIAL_Calls. Nas *views* de nível 1 são apresentados todos os objetos pré-processados.

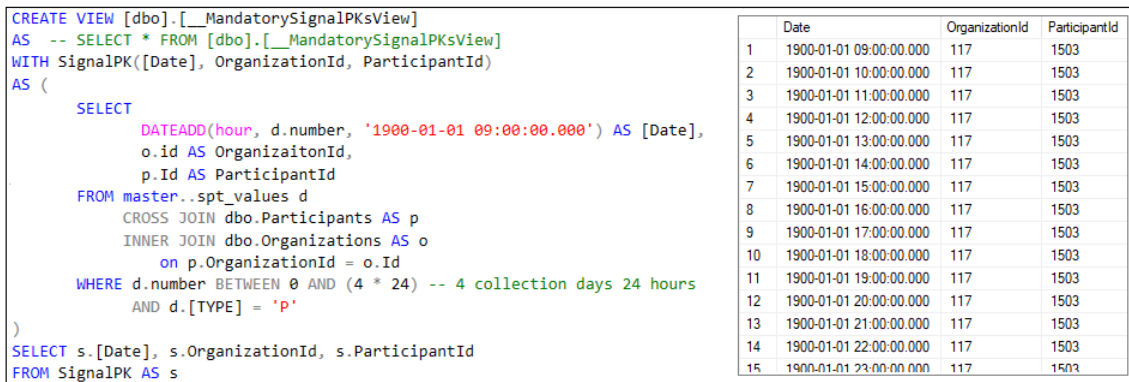


Figura 5.3: View utilizada para impor a existência de, pelo menos, um objeto a cada SI

Neste nível, os valores originais são substituídos pelos editados durante pré-processamento, caso tenham sido aplicadas técnicas que implicassem a sua edição. São também apresentadas as anotações efetuadas no *dataset* que justificam, de forma sucinta, a substituição do valor original pelo editado. Nas *views* de nível 2 e 3, são calculadas as propriedades específicas de contexto e outras propriedades extraídas (e.g. medidas estatísticas). As propriedades específicas de contexto calculadas pelas *views*, podem ser consultadas na secção 4.3.1. Como a implementação das *views* resultou da propagação do mesmo algoritmo de forma transversal em cada nível, nas *views* de nível 3, foram calculadas as mesmas propriedades estatísticas para os vários sinais que compõem o *dataset* final. As medidas estatísticas calculadas para os sinais contínuos e para os sinais categóricos, podem ser consultadas na secção 4.3.2. As *views* de nível 2 apresentam os objetos existentes após o pré-processamento (e.g. após aplicação de técnicas [DISCARD DATA]). A imposição da existência de dados a cada SI é feita por *views* específicas também de nível 2, que impõe uma junção à esquerda com a *view* `__MandatorySignalPKsView`. Esta *view* gera artificialmente combinações de chaves primárias compostas a cada $SI = 1h$ (cf. `Date + OrganizationId + ParticipantId`) (ver Figura 5.3). Como já referido no Capítulo 4, atendendo a que existiram várias iterações de recolha desfasadas no tempo, houve a necessidade de normalizar as datas de recolha para dias relativos. Esse motivo justifica que as datas geradas na *view* `__MandatorySignalPKsView` façam referência a janeiro de 1900. Por fim, a *view* de nível 4 permite explorar um subconjunto de valores disponibilizados pelas *views* de nível 3, ancorados de forma tabular em cada SI.

As *views* de nível 1 apresentam todos os objetos existentes nas tabelas onde foram integrados os dados do *dataset raw*. Como já referido na secção 5.1, os ficheiros com os dados *raw* foram integrados em tabelas espelho na base de dados relacional e, para cada atributo representativo de um valor recolhido de contexto, foram criados dois atributos adicionais: um com o sufixo `*_EditedValue`, e outro com o sufixo `*_AnnotationId`. Desta forma, foi possível implementar a aplicação de técnicas de pré-processamento de edição de dados, sem alterar os dados recolhidos de contexto. As *views* de nível 1, tem

```

CREATE VIEW [dbo].[ANDROIDSOCIAL_SMSs_1_View]
AS -- SELECT * FROM [dbo].[ANDROIDSOCIAL_SMSs_1_View] r ORDER BY r.[ParticipantId], r.[DateTimeRealRounded]
SELECT
    dbo.GetRoundedDatetime2(r.[Date]) AS [DateTimeRealRounded],
    DATEADD(SECOND, DATEDIFF(SECOND, CAST(p.[_d1] AS DATE), dbo.GetRoundedDatetime2(r.[Date])), '') AS [DateTimeRelative],
    FORMAT(r.[Date], 'ddd', 'en-EN') AS [WeekDay],
    r.[OrganizationId],
    r.[ParticipantId],
    (...)
    -- Type { Incoming, Outgoing }
    --r.[Type] AS [Type_RAW],
    COALESCE(r.[Type_EditedValue], r.[Type]) AS [Type__Tag],
    COALESCE(r.[Type_AnnotationId], '') AS [Type_AnnotationId],

    -- Word (number of words)
    --r.[Word] AS [Word_RAW],
    COALESCE(r.[Word_EditedValue], r.[Word]) AS [Word],
    COALESCE(r.[Word_AnnotationId], '') AS [Word_AnnotationId],
    (...)
    r.[AutomaticIntegration],
    COALESCE(r.[AutomaticIntegration_AnnotationId], '') AS [AutomaticIntegration_AnnotationId],
    r.[Inactive],
    COALESCE(r.[Inactive_AnnotationId], '') AS [Inactive_AnnotationId]
FROM [dbo].[ANDROIDSOCIAL_SMSs] r
INNER JOIN [dbo].[Participants_2_View] p ON (r.ParticipantId = p.Id)

```

Figura 5.4: Excerto da *view* de nível 1 ANDROIDSOCIAL_SMSs_1_View

como origem de dados as tabelas da base de dados. O objetivo destas *views* é apresentar o resultado da aplicação de técnicas de pré-processamento, substituindo os dados *raw* pelos dados editados pelas técnicas de pré-processamento, nos casos em que a aplicação dessas técnicas significou alteração dos valores originais. As *views* de nível 1, apresentam todos os registos, incluindo os inativos. Estas *views* permitem abstrair as camadas de cima, da origem dos valores que irão consumir (i.e. dados originais recolhidos de contexto, ou dados editados por uma técnica de pré-processamento). Neste contexto, para explorar os dados originais (i.e. importados dos ficheiros do *dataset raw*), as *queries* devem incidir sobre as tabelas do modelo de dados. Caso se pretenda aceder aos dados pré-processados e anotados, a extração deve ter como origem as *views* de nível 1. A título de exemplo, apresenta-se um excerto da *view* ANDROID_SMSs_1_View na Figura 5.4, onde são perceptíveis as instruções para inclusão das anotações no *recordset* a ser apresentado, e substituição do valor original pelo editado, caso exista, utilizando a função do `COALESCE()` do {MSSQL}.

As *views* de nível 2 têm como origem de dados as *views* de nível 1, e apresentam apenas os objetos ativos. Nesta camada, os objetos são agregados com $SI = 1h$. A agregação dos valores contínuos é feita através da média, e os valores categóricos são agregados com a moda. Nas *views* de nível 2, são também calculadas propriedades específicas de contexto. Como exemplo, apresenta-se na Figura 5.5 um excerto da *view* de nível 2 onde são calculadas algumas propriedades extraídas: (WEIGHT, HEIGHT) <BMI, BMR>, e (QUALITYOFLIFE) <PHYSICALDOMAIN, PSYCHOLOGICALDOMAIN, SOCIALRELATIONS HIPS, ENVIRONMENT>. Na figura, é também perceptível a aplicação da técnica [MANCLAS SIFICATION] sobre o atributo *IsCurrentlySick*. Como os dados do {EDAMOVE} foram integrados do ficheiro EDAMOVE_Results.xlsx, e este foi disponibilizado pelo {DATAANALYS

```

(... )
-- others (extracted properties)
[dbo].[GetBMI] (sc.[Weight], sc.Height) AS BMI,
[dbo].[GetBMICategory]([dbo].[GetBMI] (sc.[Weight], sc.Height)) AS BMI_Category,
[dbo].[GetBMR] (sc.Gender, sc.Age, sc.[Weight], sc.Height) AS BMR,
-- WHOQOLBREF (personal data)
COALESCE(wqb.[Schooling], dbo.[GV_STR] ('EMPTY')) AS [Schooling],
COALESCE(wqb.[Job], dbo.[GV_STR] ('EMPTY')) AS [Job],
CASE WHEN wqb.[IsCurrentlySick] IS NULL THEN dbo.[GV_STR] ('EMPTY')
      WHEN wqb.[IsCurrentlySick] = 0 THEN 'Currently not sick'
      ELSE dbo.[GV_STR] ('EMPTY')
END AS [CurrentlySick],
(... )
-- WHOQOLBREF (QoL domains)
((6 - wqb.[Q03]) + (6 - wqb.[Q04]) + wqb.[Q10] + wqb.[Q15] + wqb.[Q16] + wqb.[Q17] + wqb.[Q18]) * 100.0 / 35 AS [QualityOfLife_PhysicalDomain],
((wqb.[Q05] + wqb.[Q06] + wqb.[Q07] + wqb.[Q11] + wqb.[Q19]) + (6 - wqb.[Q26])) * 100.0 / 30 AS [QualityOfLife_PsychologicalDomain],
(wqb.[Q20] + wqb.[Q21] + wqb.[Q22]) * 100.0 / 15 AS [QualityOfLife_SocialRelationships],
(wqb.[Q08] + wqb.[Q09] + wqb.[Q12] + wqb.[Q13] + wqb.[Q14] + wqb.[Q23] + wqb.[Q24] + wqb.[Q25]) * 100.0 / 40 AS [QualityOfLife_Environment],
-- QualityOfLife
(wqb.[Q01] + wqb.[Q02]) * 100.0 / 10 AS [QualityOfLife]
FROM [dbo].[Participants] p
INNER JOIN [dbo].[Organizations] o ON (p.OrganizationId = o.Id)
LEFT JOIN [dbo].[SOCIOCULTURAL] sc ON (p.OrganizationId = sc.OrganizationId AND p.Id = sc.ParticipantId)
LEFT JOIN [dbo].[WHOQOLBREF] wqb ON (p.OrganizationId = wqb.OrganizationId AND p.Id = wqb.ParticipantId)

```

Partici	BMI	BMI_Categ	BMR	Currently Sick	...Physical...	...Psycholog	...Social...	...Environme	QualityOfLife
1507	25.43261	Pre-obesity	1855.8	Currently not s	82.857142	86.666666	93.333333	75.000000	90.000000
1508	26.39581	Pre-obesity	1821.31	Currently not s	85.714285	80.000000	86.666666	75.000000	80.000000
1510	21.82991	Normal weig	1350.8	Currently not s	82.857142	76.666666	80.000000	65.000000	80.000000
1511	24.50891	Normal weig	1749.3	Currently not s	82.857142	76.666666	60.000000	75.000000	90.000000
1514	24.15881	Normal weig	1707.71	Currently not s	94.285714	90.000000	100.000000	82.500000	90.000000
1517	22.10021	Normal weig	1329.1	Currently not s	94.285714	76.666666	66.666666	87.500000	90.000000
1518	20.57611	Normal weig	1294.21	Currently not s	57.142857	53.333333	60.000000	60.000000	80.000000
1519	29.41171	Pre-obesity	1888.7	Currently sick	40.000000	43.333333	66.666666	67.500000	40.000000

Figura 5.5: Excerto de uma *view* exemplo de nível 2

ER} com a agregação esperada, é redundante voltar a agregar os mesmos dados nas *views* de nível 2. No entanto, atendendo a que as questões de gestão de recursos e desempenho do DBMS, não eram o foco dos trabalhos, optou-se por manter a coerência algorítmica entre *views*. As anotações da origem de dados (i.e. *views* de nível 1) não são disponibilizadas para as camadas acima, por serem desnecessárias ao processamento das *views* de nível superior. Tendo por base as *views* de nível 2, é possível explorar os dados de duas formas: i) apenas os objetos resultantes do pré-processamento; ou ii) estes mesmos objetos ancorados a cada $SI = 1h$.

No nível 3 foi criada uma *view* para cada variável recolhida. Cada uma destas *views* calcula as restantes propriedades extraídas dessa variável (e.g. medidas estatísticas). São também calculadas as *mild* e *extreme outlier fences* para cada sinal. Cada valor de variável recolhida de contexto é testado para ver se está fora das cercas calculadas. Se for o caso, o objeto é etiquetado como candidato a *mild* ou *extreme outlier*. Para a implementação das *views* deste nível seguiu-se duas estratégias distintas, mediante o tipo de dados origem da variável em questão. No caso de se tratar de uma variável com valores contínuos, a *view* calcula medidas como média, máximo, IQR, etc. (ver Tabela 4.7). No caso de valores categóricos, atendendo à natureza dos seus valores, a medida calculada é a moda. Em ambas as estratégias foram calculadas medidas estatísticas globais e por participante. Como a estrutura das *views* é equivalente dentro da mesma camada, as propriedades extraídas calculadas das várias variáveis de contexto, são idênticas. Pretendia-se incluir no *dataset* final o maior número possível de *features* para permitir que o *dataset* pudesse ser reutilizado também por outros investigadores. No futuro, o autor pretende utilizar algoritmos de ML com o objetivo de tentar descobrir quais as *features* do *dataset* que conseguem melhores resultados na deteção automática do bem-estar em trabalhadores de escritório. A Figura 5.6 apresenta um extrato do *recordset* obtido pela execução da *view*

DateTimeRelati	Day	ParticId	EdaScrCount	Partic_Avg	Glob_Avg	Partic_Max	Global_Rng	Global_Std	Particip_Median	Global_Median	Participant_CV	UpperOuter	UpperInn
1900-01-02 15	Wed	1503	476	239	209	485	663	169,12824	252	182	0,0066265163	1277	815
1900-01-02 15	Wed	1506	212	136	209	499	663	169,12824	105	182	0,0078985563	1277	815
1900-01-02 15	Wed	1507	113	136	209	412	663	169,12824	110	182	0,0097610547	1277	815
1900-01-02 15	Wed	1508	394	301	209	499	663	169,12824	315	182	0,0098876170	1277	815
1900-01-02 15	Wed	1510	482	277	209	524	663	169,12824	357	182	0,0051377936	1277	815
1900-01-02 15	Thu	1511	389	329	209	577	663	169,12824	354	182	0,0066690204	1277	815
1900-01-02 15	Wed	1514	91	66	209	242	663	169,12824	51	182	0,0169391626	1277	815
1900-01-02 15	Thu	1517	267	293	209	431	663	169,12824	312,5	182	0,0107720204	1277	815
1900-01-02 15	Fri	1518	391	264	209	510	663	169,12824	316	182	0,0062074846	1277	815
1900-01-02 15	Sat	1519	429	242	209	496	663	169,12824	253,5	182	0,0060250520	1277	815
1900-01-02 15	Wed	1521	123	39	209	162	663	169,12824	27	182	0,0246343502	1277	815
1900-01-02 15	Fri	1526	386	198	209	432	663	169,12824	256	182	0,0064712727	1277	815
1900-01-02 15	Wed	1530	220	128	209	336	663	169,12824	134	182	0,0111993060	1277	815
1900-01-02 15	Sat	1533	413	215	209	553	663	169,12824	223,5	182	0,0058667809	1277	815
1900-01-02 15	Thu	1540	152	88	209	230	663	169,12824	86,5	182	0,0140077246	1277	815
1900-01-02 15	Thu	1541	385	182	209	445	663	169,12824	217,5	182	0,0072437238	1277	815
1900-01-02 15	Wed	1904	465	186	209	515	663	169,12824	165	182	0,0069585818	1277	815
1900-01-02 15	Wed	1909	414	251	209	468	663	169,12824	308,5	182	0,0061298207	1277	815
1900-01-02 15	Wed	1932	467	356	209	625	663	169,12824	434	182	0,0050176229	1277	815
1900-01-02 15	Wed	1955	360	342	209	664	663	169,12824	404	182	0,0045665930	1277	815

Figura 5.6: Excerto do resultado do *recordset* devolvido por uma *view* de nível 3

EDAMOVE_Results_3_06EdaScrCount_View. Na impossibilidade de apresentar todos os dados do conjunto devolvido, são apenas apresentados os registos das 15 horas do segundo dia de recolha, de todos os participantes. Também por questões de espaço, são apenas apresentadas algumas colunas com parte das propriedades extraídas calculadas pela *view*.

A *view* *_Dataset_4_View*, é a *view* global de nível 4. Esta *view* apresenta todos as variáveis de contexto e propriedades extraídas das *views* de nível 3. A utilização desta *view* facilita a exploração de vários atributos e propriedades extraídas em simultâneo, sem haver a necessidade de definir as relações entre as várias *views*. No entanto, todas as tabelas e *views* contêm os atributos necessários para permitir a junção personalizada. Assim, a exploração de dados pode incidir sobre a *view* global de nível 4, ou pode ser criada outra *query* específica para extração de dados. Além disso, a organização de código da *view* *_Dataset_4_View*, foi também pensada para facilitar a criação de outras *views* personalizadas via clonagem. Assim, a criação de um subconjunto de registos e/ou atributos, pode resumir-se à duplicação e personalização da *view* de nível 4 proposta pelo autor, aplicando filtros na clausula *where* da *query* SQL, ou comentando (retirando) atributos da cláusula *select*, que já foram dispostos de forma a facilitar o processo. Além de apresentar todos os dados pré-processados e propriedades extraídas, esta *view* ancora toda a informação a cada SI.

A estratégia de *views* apresentada nesta secção, foi implementada em todas as tabelas com dados *raw* recolhidos de contexto. No entanto, atendendo à especificidade dos dados das tabelas *Participants*, *Organizations*, *SOCIOCULTURAL* e *WHOQOLBREF*, o pré-processamento e cálculo de propriedades extraídas das variáveis recolhidas por estes instrumentos, foi efetuado na *view* *Participants_2_View*.

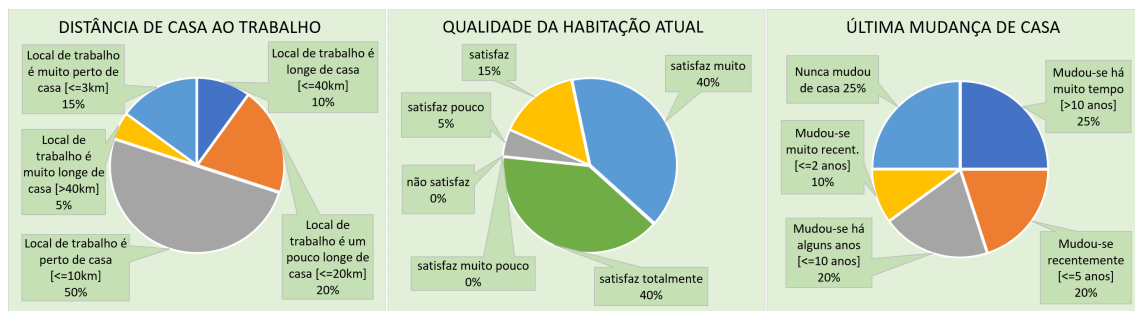


Figura 5.7: Distância de casa ao trabalho, qualidade (conforto) e última mudança de casa

5.3 Caracterização dos participantes

O *dataset* final apresentado neste capítulo foi recolhido em ambiente real, e ao longo de mais de dois meses. Foram recolhidos dados de contexto de 20 participantes pertencentes a 7 organizações distintas. Nesta secção faz-se a caracterização dos participantes na experiência.

Os participantes nesta experiência, são trabalhadores de escritório, com idades compreendidas entre os 21 e os 45 anos (média: 31) (AGE). A grande maioria são de religião católica e de etnia caucasiana (RELIGION) (ETHNICITY). 11 participantes são do género masculino e 9 são do género feminino (GENDER). Relativamente ao seu grau académico (ACADDG): i) 35% tem licenciatura (Nível 6)¹; ii) 25% tem um curso pós-secundário não superior (Nível 5); iii) 15% terminou os estudos no ensino secundário (Nível 3); iv) também 15% dos participantes terminou o ensino básico (Nível 2); e v) 10% terminou os seus estudos no ensino secundário com estágio profissional.

A maioria dos participantes (75%) vive na cidade, e desses 50% vive numa moradia (HOME) <ENVIRONMENT, TYPE>. 55% tem ao seu encargo dependentes sendo que, de entre esses, 36% tem à sua responsabilidade mais do que um dependente e 64% são menores de idade (DEPENDENTS). 55% vive com os filhos e/ou conjuge, e os restantes participantes vivem com outros familiares (e.g. irmãos e/ou pais) (LIVING). 50% dos participantes têm créditos, desses, 60% têm um crédito para compra de casa, e 20% têm mais que um crédito <LOANS>. 75% dos participantes, indica ter vizinhos amigos e 40% reporta a existência de vizinhos problemáticos (conflituosos) (NEIGHBORS).

De entre os participantes da experiência, 45% vivem em casa própria, outros 45% vivem em casa de família, e 10% vivem em casa arrendada (HOME) <OWNER>. A Figura 5.7 resume: i) a distância em quilómetros de casa ao trabalho dos participantes (HOME-WORK) <DISTANCE>; ii) a qualidade / conforto percebido pelos participantes

¹Os níveis de qualificação especificados, são os previstos no Quadro Nacional de Qualificações (definidos pela portaria 782/2009, de 23 de julho): Nível 1 (2º ciclo do ensino básico); nível 2 (3º ciclo do ensino básico); nível 3 (ensino secundário para ingresso no ensino superior); nível 4 (ensino secundário com estágio profissional de, no mínimo, 6 meses); nível 5 (qualificação de nível pós-secundário não superior); nível 6 (licenciatura); nível 7 (mestrado); e nível 8 (doutoramento).

relativamente à casa onde moram (HOME) <QUALITY>; e iii) informação relacionada com a última troca de casa (HOME) <HOUSEMOVE<LASTTIME>>. A maior parte dos participantes vive perto do local de trabalho (HOME-WORK) <DISTANCE>, e pontua bastante bem a qualidade (conforto) da casa atual (HOME) <QUALITY>. Uma grande parte dos participantes mudou de casa há mais de 5 anos (HOME) <HOUSEMOVE<LASTTIME>>, e indica viver perto ou mesmo muito perto do local de trabalho. 80% dos participantes gostam da qualidade da sua casa, sendo que metade deles indica mesmo estarem totalmente satisfeitos. Não existem participantes que indicam viver numa casa que não satisfaz os seus critérios de qualidade mas, 5% indica satisfazer pouco e 15% classificam a sua habitação atual como satisfaz.

A Tabela 5.3 resume os dados demográficos dos participantes na experiência. Além dos dados gerais de cada participante, é apresentada informação: i) acerca dos dependentes que representam preocupações diárias de carácter permanente (DEPENDENTS); ii) local e tipo de habitação (e.g. moradia na cidade, apartamento no campo, moradia no campo) (HOME) <ENVIRONMENT, TYPE>; iii) coabitação do participante (cf. esposa e/ou filhos ou com os pais e/ou irmãos) (LIVING); iv) existência de créditos (e.g. compra de habitação, carro, créditos pessoais) (LOANS); e v) da existência de vizinhos amigos e/ou problemáticos (NEIGHBORS).

Part.	Idade	Género	Hab.Lit.	Est.Civil	Depend.	Habitação	Coabitação	Créditos	Vizinhos
1503	34	♂	Nível 2	Solteiro					
1506	31	♂	Nível 5	Solteiro	—			—	
1507	21	♂	Nível 5	Solteiro	—				
1508	31	♂	Nível 6	Solteiro	—			—	
1510	21	♀	Nível 3	Solteiro	—			—	
1511	23	♂	Nível 5	Solteiro	—			—	—
1514	41	♂	Nível 4	Casado					
1517	40	♀	Nível 6	Solteiro					
1518	26	♀	Nível 6	Solteiro	—			—	
1519	35	♂	Nível 3	Solteiro					
1521	42	♂	Nível 2	Casado					
1526	36	♀	Nível 5	Casado					
1530	45	♀	Nível 6	Casado					
1533	23	♀	Nível 6	Solteiro	—			—	—
1540	38	♀	Nível 2	Casado					
1541	33	♂	Nível 6	Solteiro	—			—	

1904	32	♀	Nível 6	Casado					
1909	27	♂	Nível 5	Solteiro				—	
1932	29	♀	Nível 3	Solteiro				—	
1955	27	♂	Nível 4	Solteiro	—			—	—

— género masculino; — género feminino; — dependente menor de idade (e.g. criança);
 — dependente maior de idade (e.g. idoso); — vive numa moradia na cidade; — vive numa moradia no campo; — vive num apartamento na cidade; — vive num apartamento no campo; — vive noutro tipo de habitação na cidade; — vive noutro tipo de habitação no campo; — coabita com a esposa e/ou filhos; — coabita com os pais e/ou irmãos; — existência de créditos habitação; — existência de créditos para compra de casa; — existência de outro tipo de créditos (e.g. crédito pessoal); — existência de vizinhos amigos; e — existência de vizinhos problemáticos.

Tabela 5.3: Extrato de dados recolhidos do domínio pessoal dos participantes

Relativamente aos dados recolhidos da categoria psicossocial, 85% dos participantes apresenta um indicador global de qualidade de vida igual ou superior a 80 (QUALITYOFLIFE). A qualidade de vida medida pelo {WHOQOLBREF} considera dados de quatro domínios: físico, psicológico, relações sociais e ambiente. Cada um destes domínios é composto por facetas da qualidade de vida que sumarizam o respetivo domínio da qualidade de vida a que pertencem. O indicador global de qualidade de vida, é calculado com base nos resultados obtidos destes domínios (Vaz Serra et al., 2006). A Tabela 5.4 apresenta alguns dos dados recolhidos referentes à categoria psicossocial. Além do indicador global da qualidade de vida (QUALITYOFLIFE), são apresentados os quatro domínios base ao cálculo do indicador, e informação acerca do estado de saúde do participante na recolha.

Part.	Doente	Qualidade Vida (físico)	Qualidade Vida (psicológ.)	Qualidade Vida (social)	Qualidade Vida (ambiente)	Qualidade de Vida
1503	—	85,7	83,3	73,3	67,5	80
1506	—	88,6	86,7	93,3	77,5	80
1507	—	82,9	86,7	93,3	75,0	90
1508	—	85,7	80,0	86,7	75,0	80
1510	—	82,9	76,7	80,0	65,0	80
1511	—	82,9	76,7	60,0	75,0	90
1514	—	94,3	90,0	100,0	82,5	90
1517	—	94,3	76,7	66,7	87,5	90
1518	—	57,1	53,3	60,0	60,0	80
1519	✓	40,0	43,3	66,7	67,5	40
1521	—	77,1	70,0	53,3	70,0	80
1526	—	60,0	73,3	66,7	70,0	60
1530	—	88,6	86,7	80,0	80,0	80
1533	—	82,9	90,0	86,7	75,0	50

1540	—	71,4	86,7	100,0	72,5	80
1541	—	100,0	93,3	100,0	85,0	90
1904	—	91,4	83,3	93,3	80,0	80
1909	—	71,4	70,0	66,7	67,5	80
1932	—	82,9	70,0	73,3	75,0	80
1955	—	91,4	96,7	80,0	82,5	80

✓ — participante sentia-se doente no momento da recolha de dados de contexto.

Tabela 5.4: Qualidade de vida e estado de saúde dos participantes

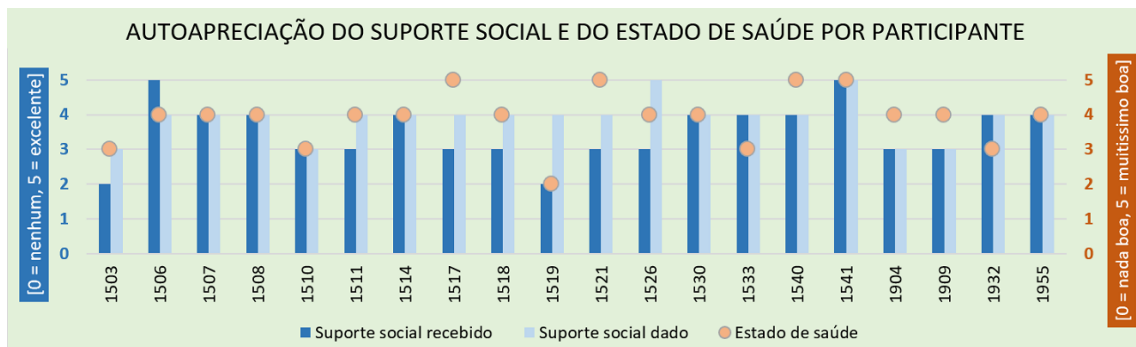


Figura 5.8: Autoapreciação do suporte social dado e recebido, e percepção do estado de saúde dos elementos do domínio privado (e.g. conjuge e filhos) por participante

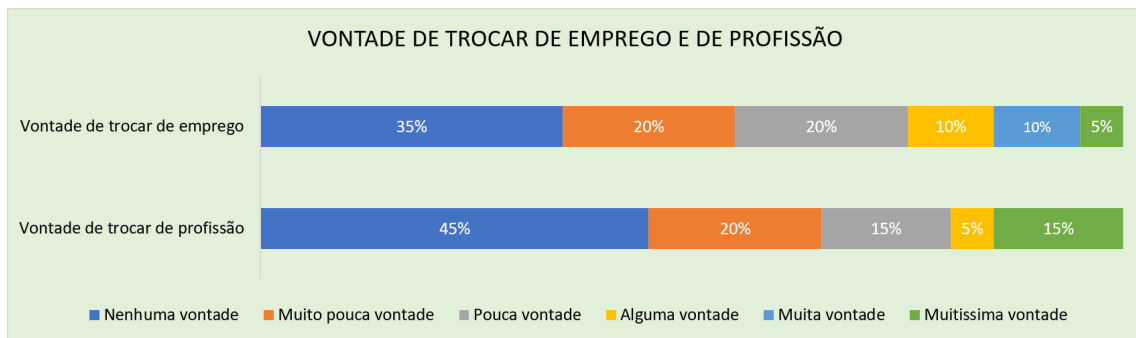


Figura 5.9: Vontade indicada pelos participantes para mudar de emprego e de profissão

A Figura 5.8 resume a Autoapreciação indicada pelos participantes, relativamente ao suporte social dado e recebido, e a percepção relativa ao estado de saúde geral dos elementos do seu domínio privado (e.g. conjuge e filhos) (SOCIALSUPPORT) <GIVEN, RECEIVED>. 35% dos participantes indicam dar mais suporte social do que o que recebe, e apenas um participante (5%) indica receber mais do que o que dá. Ao nível da percepção de saúde, 75% dos participantes classificam a saúde dos elementos do seu domínio privado como muitíssimo boa ou muito boa, havendo apenas 1 participante que indica uma percepção não tão positiva (HEALTH).

São muitos os participantes que reportam a vontade de manter o emprego e a profissão atual. 55% indica muito pouca ou nenhuma vontade de trocar emprego, e 45% pretende manter a sua profissão. Apenas 15% expressam grande vontade em trocar de emprego e de profissão. A Figura 5.9 resume a vontade dos participantes da experiência em trocar de emprego e de profissão.

Na categoria reconhecimento agrupam-se as características ou propriedades que as outras pessoas interpretam como sinal de prestígio ou competências (qualidades). Dos trabalhadores que participaram na experiência (WORKLIFE) <JOB>: i) 40% são administrativos e executam tarefas de expediente; ii) 25% é pessoal informático que faz *helpdesk* interno e/ou externo (e.g. clientes); iii) 20% são vendedores que fazem apoio comercial e vendas a clientes a partir do escritório; iv) 10% executa funções de gerência; e v) 5% desempenham outras funções no escritório. 35% dos participantes estão no emprego atual há 10 anos ou mais, e desses, 57% nunca mudaram de emprego.

A Tabela 5.5 apresenta a profissão de cada participante, a duração da sua vida ativa (número de anos de trabalho considerando todos as profissões e empregos), número de anos a exercer a profissão atual, número de anos no emprego atual, e quantidade de empregos até ao momento da participação na experiência.

Part.	Profissão	Vida ativa (anos)	Profissão atual (anos)	Emprego atual (anos)	Empregos (quantidade)
1503	Gerente	19	12	10	7
1506	Informático	3	1	1	8
1507	Informático	1	1	1	1
1508	Comercial	8	2	5	3
1510	Administrativo	5	1	1	3
1511	Informático	4	3	2	3
1514	Comercial	21	5	1	4
1517	Administrativo	21	16	16	3
1518	Administrativo	4	2	2	2
1519	Informático	15	15	15	18
1521	Gerente	24	18	18	1
1526	Administrativo	18	18	18	1
1530	Comercial	22	20	3	5
1533	Comercial	1	1	1	3
1540	Administrativo	20	20	20	1
1541	Administrativo	15	15	15	1
1904	Administrativo	9	9	9	1
1909	Informático	9	1	2	4
1932	Outro	11	1	1	2
1955	Administrativo	9	8	8	2

Tabela 5.5: Informações relacionadas com a vida profissional dos participantes

A Tabela 5.6 resume alguns dos dados recolhidos de cada participante, referentes à categoria reconhecimento: i) a altura é indicada em metros (m) (HEIGHT); ii) o peso é indicado em quilogramas (kg) (WEIGHT); iii) o *Body Mass Index* (BMI) é calculado e categorizado tendo por base as fórmulas da Organização Mundial da Saúde (OMS) (Who, 2021a) (WEIGHT, HEIGHT) <BMI>; iv) na coluna da dieta são identificados os participantes que fazem ou já fizeram dieta para manutenção do aspeto físico (DIET); v) é também referido o número aproximado de cafés consumidos por dia (#/dia) (CAFFEEI); vi) o número aproximado de cigarros consumidos por dia (#/dia) (SMOKING); vii) a quantidade de eventos culturais presenciados no último mês (SOCCULTURALEVENTS) <COUNT>; e viii) os dias despendidos em contacto com a natureza no último mês (NATURE) <DURATION>.

Part.	Altura (m)	Peso (kg)	BMI	Dieta	Álcool	Caféina (#/dia)	Tabaco (#/dia)	Cultura (#/mês)	Natureza (#/mês)
1503	1,75	116		—		1	0	2	2
1506	1,80	68		✓		3	20	0	5
1507	1,74	77		✓		1	0	0	0
1508	1,73	79		✓		5	0	1	4
1510	1,63	58		✓		4	0	0	0
1511	1,69	70		✓		3	0	0	2
1514	1,69	69		—		1	0	1	3
1517	1,62	58		✓		3	0	0	2
1518	1,62	54		✓		0	2	0	2
1519	1,70	85		✓		4	10	0	0
1521	1,64	70		—		2	0	0	0
1526	1,56	56		✓		1	0	0	0
1530	1,69	61		—		3	0	0	3
1533	1,65	56		—		2	0	0	1
1540	1,56	60		—		2	0	0	0
1541	1,80	76		—		0	0	2	4
1904	1,61	80		—		2	—	1	1
1909	1,71	63		—		5	20	0	0
1932	1,65	60		✓		0	0	0	3
1955	1,80	85		✓		5	5	1	1











 — abaixo do peso normal;  — peso normal;  — pré-obesidade;  — obesidade classe I;  — obesidade classe II;  — obesidade classe III;  — consome álcool apenas com amigos;  — consome álcool apenas às refeições;  — não consome álcool.  — outro;

Tabela 5.6: Características físicas dos participantes, consumo de álcool, caféina e cigarros, participação em eventos e dias em contacto com a natureza



Figura 5.10: Desempenho de papéis de relevo na sociedade, participação ativa (e.g. associações), e filiação em partidos políticos

A maioria dos participantes apresentam peso normal (65%) sendo que, 25% estão num estado de pré-obesidade, e 10% tem obesidade de classe I ou classe II. 11 participantes (55%) fazem dieta para manutenção do aspeto físico. 75% consomem álcool e, desses, 80% consomem-no em convívio com amigos. A maioria dos participantes toma até três cafés por dia e não fuma (75%). Apenas seis participantes (30%) participaram em eventos culturais no último mês anterior à experiência. Relativamente ao número de dias em contacto com a natureza no mês anterior à recolha de dados: 35% não teve qualquer contacto com a natureza, 50% indica ter tido contacto de três dias ou menos, e os restantes 15% estiveram em contacto quatro ou cinco dias.

Apenas 25% dos participantes desempenham papéis sociais de relevo em organizações sociais, políticas ou religiosas, e indicam participar ativamente em partidos políticos, religião ou outras associações. Apenas 20% dos participantes estão filiados num partido político. A Figura 5.10 apresenta visualmente a participação na vida em sociedade pelos participantes na experiência relatada neste documento.

5.4 Caracterização dos sinais recolhidos

O *dataset* final apresentado como contribuição neste capítulo contém dados de várias variáveis de contexto fisiológico e psicossocial. Nesta secção faz-se a caracterização dos dados que compõem o *dataset* final.

Na experiência relatada neste documento recolheram-se dados de contexto de várias modalidades, de vários participantes pertencentes a 7 organizações distintas. Pretendia-se recolher dados de contexto do maior número de participantes possível. No entanto, apesar de existirem mais candidatos, o tempo limitado de empréstimo dos dois sensores {EDAMOVE}, impossibilitou que fossem admitidos mais participantes e, consequentemente, impediu a recolha de mais dados de contexto. Com o objetivo de otimizar o processo de recolha e maximizar a quantidade de dados recolhidos, deu-se sempre prioridade à recolha em organizações com quantidade par de participantes. Executaram-se 10 iterações de recolha e, em cada iteração, recolheu-se dados de dois participantes.

A Tabela 5.7 apresenta alguns detalhes acerca das várias iterações de recolha executadas: i) organização onde trabalhava cada participante; ii) número da iteração de recolha que representa também a ordem pela qual foram recolhidos os dados de contexto de cada participante; iii) data de início e; iv) data de fim de recolha de cada iteração; v) dias da semana em que a recolha foi efetuada; e v) a quantidade de horas de dados recolhidas de cada participante.

Part.	Org.	Iteração (#)	Início da recolha (data)	Fim da recolha (data)	Período (de semana)	#Horas recolhidas
1519	110	1	2019-02-15 09:30	2019-02-18 17:30	sex — seg	80
1533						80
1518	102	2	2019-02-21 09:30	2019-02-24 17:30	qui — dom	80
1526						80
1517	113	3	2019-02-27 09:30	2019-03-01 17:30	qua — sex	56
1541						56
1511	112	4	2019-03-06 09:30	2019-03-08 17:30	qua — sex	56
1540						56
1506	112	5	2019-03-12 09:30	2019-03-15 17:30	ter — sex	80
1507						80
1904	217	6	2019-03-19 09:30	2019-03-22 17:30	ter — sex	80
1909						80
1932	217	7	2019-03-26 09:30	2019-03-29 17:30	ter — sex	80
1955						80
1508	109	8	2019-04-02 09:30	2019-04-05 17:30	ter — sex	80
1530						80
1510	117	9	2019-04-09 09:30	2019-04-12 17:30	ter — sex	80
1514						80
1503	117	10	2019-04-16 09:30	2019-04-19 17:30	ter — sex	80
1521						80

Tabela 5.7: Iterações de recolha e número de horas recolhidas por participante

Quando se iniciou a fase de execução da recolha, pretendia-se recolher dados de forma a que o *dataset* contivesse dados referentes aos dias de trabalho de sexta e segunda-feira. A ideia passava por recolher dados que espalhassem o estado emocional dos participan-

tes aquando do fim e no início da semana de trabalho. Pretendia-se, mais tarde, estudar a correlação desta informação com os restantes dados recolhidos. No entanto, logo na primeira iteração de recolha foi perceptível que seria necessário acompanhar de perto os processos de recolha pelas razões já referidas na secção 3.2.2. Como esse acompanhamento implicaria visitas ao domicílio de cada participante durante o fim-de-semana, optou-se por recolher dados apenas durante os dias de trabalho.

Em média recolheu-se mais do que 75 horas de dados de contexto por participante. O *dataset* final pré-processado apresentado como contribuição neste documento, contém mais de 1500 horas de dados recolhidos de contexto. No entanto, os vários sinais não contribuem de igual forma para a dimensão do conjunto de dados final. São os dados de origem fisiológica que mais contribuem para o tamanho do *dataset*. A forma de introspeção periódica e automática de contexto do {EDAMOVE} e {MIBAND}, justifica que sejam os dados recolhidos por estes sensores, os que mais contribuem para o tamanho do *dataset* final. A quantidade de dados recolhidos pelos instrumentos {ANDROIDANALYTICS} e {WINDOWSACTIVITY} foi menor. Ao contrário dos dispositivos de recolha de dados fisiológicos, a forma de recolha destes instrumentos não implica a introspeção constante de dados de contexto. Estes instrumentos diferenciam-se por recolherem dados de forma oportunista, apenas quando existem interações sociais no *smartphone* ou utilização do computador de trabalho. A natureza de recolha, também justifica a diferença na quantidade de dados relacionados com a interação social, recolhidos pelo {ANDROIDANALYTICS}. As diferenças existentes no número de (CALL) e (SMS) entre os vários participantes, poderão ser justificadas pela diversidade dos meios de comunicação existentes atualmente. Alguns participantes utilizaram bastante o contacto verbal via telefone, mas outros privilegiaram alternativas escritas para o fazer. O número de (CALL) varia muito de participante para participante, mas a variação é ainda mais evidente nas (SMS). Durante as várias visitas efetuadas às organizações foi possível constatar a utilização do correio eletrónico e ferramentas de *Instant Messaging* (IM) na comunicação interna e externa (e.g. *WhatsApp*, *Skype*, *Facebook*). Em experiências futuras, deverá ser ponderado o alargamento das origens de recolha a estes meios de comunicação.

A Tabela 5.8 resume o número de objetos de alguns sinais do *dataset*. A caracterização de dados feita nesta tabela, incide sobre os dados pré-processados não agregados. Como já referido, os dados do (EDA) foram pré-processados após cada iteração de recolha (ver secção 4.3.1). Assim, o número de objetos indicado na coluna do instrumento {EDAMOVE}, já se refere à quantidade de objetos resultante da agregação $SI = 1h$. O número de objetos indicados no sinal (SLEEP) [RAW] e (WORKINTENSITY) [RAW] referem-se à quantidade total de registos existente no *dataset* pré-processado. Os sinais (WORKINTENSITY) <COMPUTERSUCCESSLOGIN, COMPUTERFAILEDLOGIN, COMPUTERLOCKOUT>, são subconjuntos do conjunto de dados respetivo. A propriedade específica de contexto (ELECTR) <USAGE>, refere-se à quantidade de *tooglescreens* do *smartphone* de cada participante.

Instrumento	Sinal	#Objetos
{EDAMOVE}	(EDA) <AROUSAL<MEAN>>	1289 ¹
	(EDA) <SCL<MEAN>>	1289 ¹
	(EDA) <SCR<AMP<MEAN>>>	1289 ¹
	(EDA) <SCR<COUNT>>	1289 ¹
	(EDA) <SCR<ENERGY<MEAN>>>	1289 ¹
	(EDA) <SCR<HALFRECOVERYTIME<MEAN>>>	1289 ¹
	(EDA) <SCR<RISETIME<MEAN>>>	1289 ¹
	(TEMP) <MEAN>	1331 ¹
{MIBAND}	(HR)	59574
	(SLEEP) [RAW]	84054
	(SLEEP) <DEEPSLEEP>	317
	(SLEEP) <LIGHTSLEEP>	366
	(PHYSI) <STEPS>	8814
	(PHYSI) <CALORIES>	75
{WINDOWSACTIVITY}	(WORKINTENSITY) [RAW]	10682
	(WORKINTENSITY) <COMPUTERSUCCESSLOGIN>	417
	(WORKINTENSITY) <COMPUTERFAILEDLOGIN>	99
	(WORKINTENSITY) <COMPUTERLOCKOUTS>	334
{ANDROIDANALYTICS}	(CALLS) [RAW]	706
	(SMS)	111
	(ELECTR) <USAGE>	1722

Tabela 5.8: Quantidade de objetos de alguns sinais recolhidos por instrumento de recolha

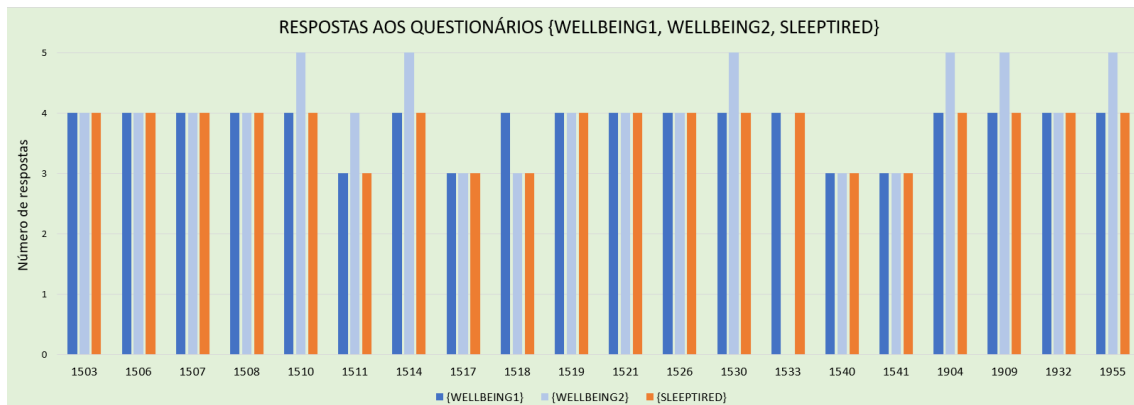


Figura 5.11: Quantidade de respostas dadas pelos participantes aos questionários {WELLBEING1}, {WELLBEING2} e {SLEEPTIRED}

A Figura 5.11 apresenta a quantidade de respostas dadas pelos participantes aos questionários {WELLBEING1}, {WELLBEING2} e {SLEEPTIRED}. Estes instrumentos recolheram as variáveis de contexto (WELLBEING), (SLEEP) e (TIRED). O {WELLBEING1} foi aplicado uma vez no final de cada dia de trabalho, e o {WELLBEING2} foi aplicado

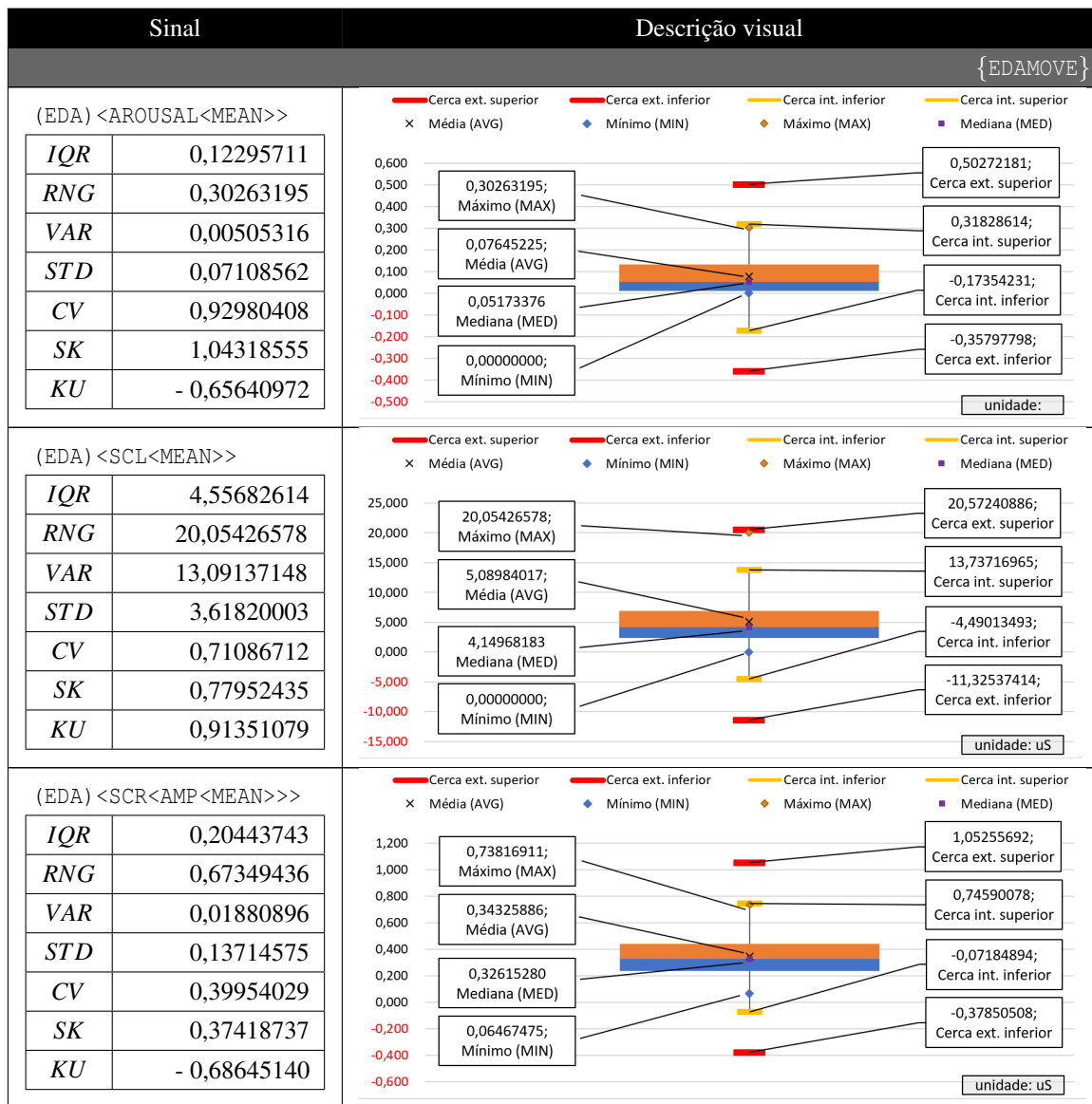
¹ Quantidade refere-se ao número de registos resultantes da agregação (ver secção 4.3.1).

uma vez durante o dia com horário aleatório. O questionário {SLEEPTIRED} recolheu a autoavaliação subjetiva acerca da qualidade do sono e sentimento de cansaço após cada noite de sono. Como é perceptível na Figura 5.11, existem mais respostas ao questionário {WELLBEING2} do que ao {WELLBEING1}. Esta diferença justifica-se porque, apesar de o {ANDROIDANALYTICS} despoletar apenas uma vez por dia o {WELLBEING2}, os participantes podiam registar várias respostas a este questionário no mesmo dia se assim o entendessem. Considerando que estes questionários foram criados especificamente para a experiência, importa analisar a capacidade que têm para medir. Neste contexto, utilizou-se o coeficiente Alfa de Cronbach (α) para estimar a confiabilidade de cada questionário (Cronbach, 2004): i) o {WELLBEING1} apresenta um coeficiente $\alpha = 0.875$; o ii) questionário {WELLBEING2} $\alpha = 0.802$; e iii) o {SLEEPTIRED} apresenta um coeficiente $\alpha = 0.859$. Desta forma, assume-se que os questionários estarão a medir o pretendido.

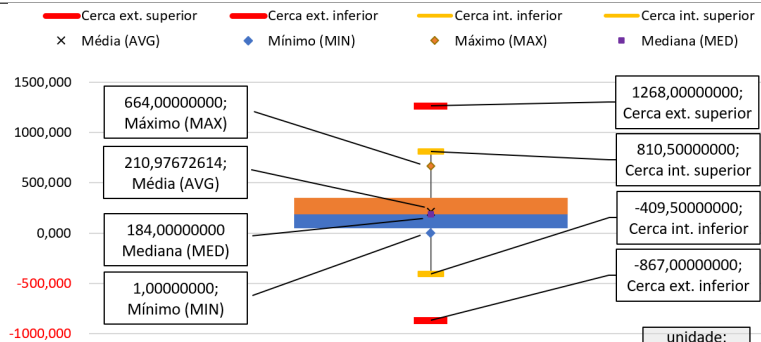
A Tabela 5.9 faz a descrição de alguns dos dados que pertencem ao *dataset* final pré-processado. Os valores indicados foram extraídos a partir das *views* de nível 3 respetivas, e contém a agregação $SI = 1h$: i) as contagens são representadas pelo número de ocorrências (e.g. (WORKINTENSITY)<COMPUTERLOCKOUTS <COUNT>>); ii) os valores aditivos foram agregados com a soma (e.g. (PHYSI)<STEPS>); iii) os não aditivos com a média (e.g. (HR)); e iv) os valores categóricos são representados pela moda do SI. Nos sinais (EDA) mantém-se a designação original atribuída pelo {DATAANALYSER} durante a agregação (i.e. <MEAN>). O coeficiente de variação (CV) é definido como o rácio entre o desvio padrão e a média ($CV = STD/AVG$), e expressa a extensão da variabilidade dos dados em relação à média. A *skewness* (SK) é indicada pelo coeficiente de assimetria de Pearson (Doanne and Seward, 2011). A diferença entre a média e a mediana (ou moda) da fórmula de cálculo do SK, indica quão longe a distribuição se encontra da simetria (distribuição normal) $SK = 0$. Um valor $SK < 0$ indica uma assimetria negativa ou à esquerda. O valor $SK > 0$ indica uma assimetria positiva ou à direita, significando que a cauda da distribuição é mais denunciada à direita. Quanto mais uma distribuição se afasta da distribuição normal, maior será o valor de assimetria. A curtose (KU) indica o achatamento da curva em relação à distribuição normal. Uma distribuição normal tem $KU = 0$ (distribuição mesocúrtica). Um valor $KU > 0$ indica caudas mais pesadas do que a distribuição normal, tornando a curva mais apontada (distribuição leptocúrtica). Um valor $KU < 0$ indica o oposto (distribuição platicúrtica).

Pela análise da tabela, é perceptível a existência de objetos candidatos a *outlier* por estarem além das cercas exteriores calculadas (e.g. (HR), (PHSI)<STEPS>, (CALL)<INCOMING<DURATION<SUM>>>)). No entanto, foram apenas removidos os objetos com valor inquestionavelmente fora do intervalo possível para o sinal (e.g. erros de leitura). Cada sinal foi analisado individualmente durante o pré-processamento para fazer esta avaliação específica. Para facilitar a leitura, algumas das visualizações apresentados na Tabela 5.9, não representam todo o conjunto original de objetos do sinal respetivo (e.g.

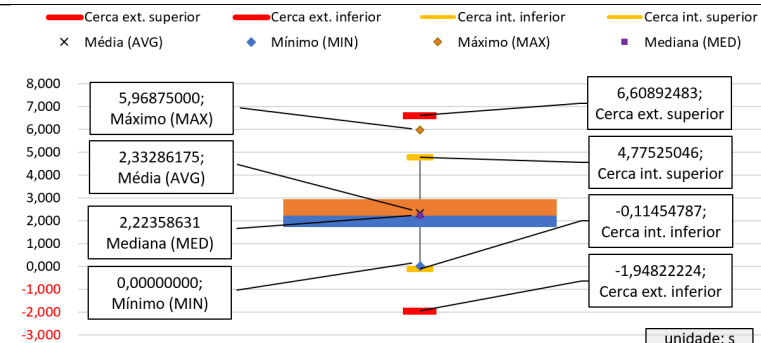
(CALL) <INCOMING<DURATION<SUM>>>, (CALL) <OUTGOING<COUNT>>, (CALL) <MISSED <COUNT>>, (SMS) <COUNT>). Como se tratam de sinais de recolha oportunista, não existem objetos para todos os SI do *dataset*. Esta inexistência é representada no *dataset* final com o valor a zero (e.g. se não existem chamadas perdidas numa determinada hora, o valor de (CALL) <MISSED<COUNT>> do SI correspondente é zero). Como o número de ocorrências de valores zero é expressiva, a inclusão desta frequência nas visualizações dificultaria bastante a análise do sinal. Assim, nestes casos, a visualização descreve as restantes frequências diferentes de zero, e é anotado o número de objetos representados na visualização e o total de objetos do conjunto (e.g. a visualização do sinal (CALL) <MISSED<COUNT>> descreve um subconjunto de 81 de 1524 objetos, que é o número de SI representados no *dataset* com valor diferente de zero).



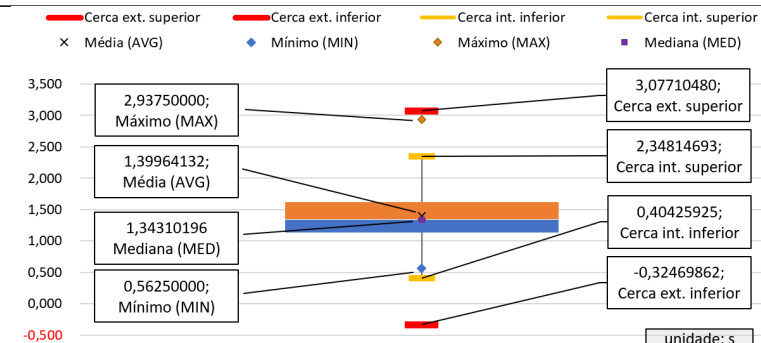
(EDA) <SCR<COUNT>>	
<i>IQR</i>	305,0000000
<i>RNG</i>	663,0000000
<i>VAR</i>	28543,05225294
<i>STD</i>	168,94689181
<i>CV</i>	0,80078450
<i>SK</i>	0,47902733
<i>KU</i>	-1,08322767



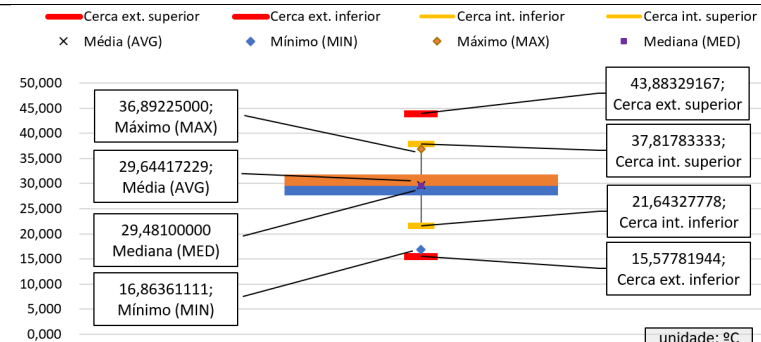
(EDA) <SCR<HALFRECOVERY TIME<MEAN>>>	
<i>IQR</i>	1,22244958
<i>RNG</i>	5,96875000
<i>VAR</i>	1,02498627
<i>STD</i>	1,01241606
<i>CV</i>	0,43398031
<i>SK</i>	0,32380592
<i>KU</i>	0,82980195



(EDA) <SCR<RISETIME<MEAN>>>>	
<i>IQR</i>	0,48597192
<i>RNG</i>	2,37500000
<i>VAR</i>	0,10682149
<i>STD</i>	0,32683557
<i>CV</i>	0,23351380
<i>SK</i>	0,51897074
<i>KU</i>	1,27842202

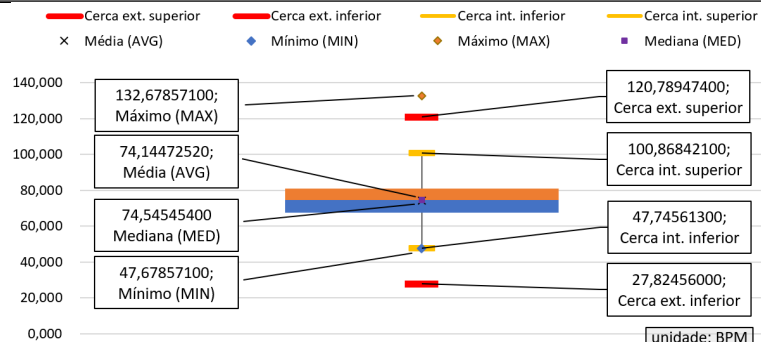


(TEMP) <MEAN>	
<i>IQR</i>	4,04363889
<i>RNG</i>	20,02863889
<i>VAR</i>	10,13208711
<i>STD</i>	3,18309395
<i>CV</i>	0,10737672
<i>SK</i>	0,15378649
<i>KU</i>	0,65285124



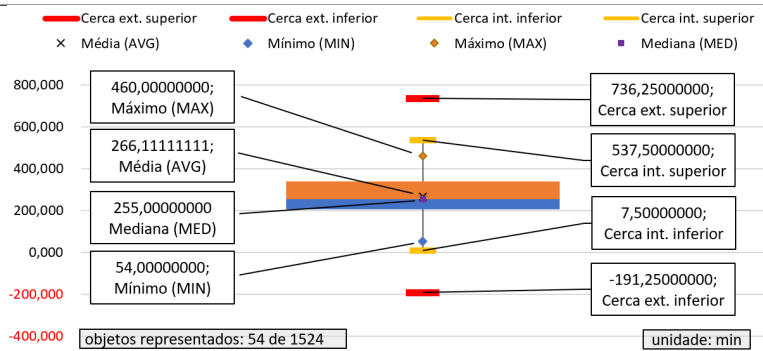
{MIBAND}

(HR)	
<i>IQR</i>	13,28070200
<i>RNG</i>	85,00000000
<i>VAR</i>	101,66855114
<i>STD</i>	10,08308242
<i>CV</i>	0,13599190
<i>SK</i>	-0,11922806
<i>KU</i>	0,56934425



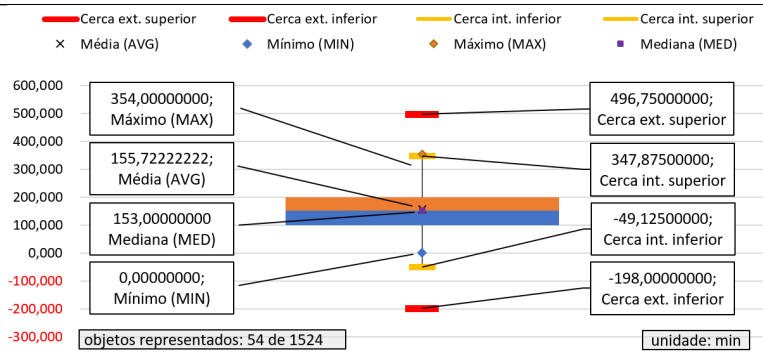
(SLEEP) <LIGHTSLEEP<DURATION>>

IQR	132,50000000
RNG	406,00000000
VAR	9730,89308176
STD	98,64528920
CV	0,37069211
SK	0,33791105
KU	-0,45927795



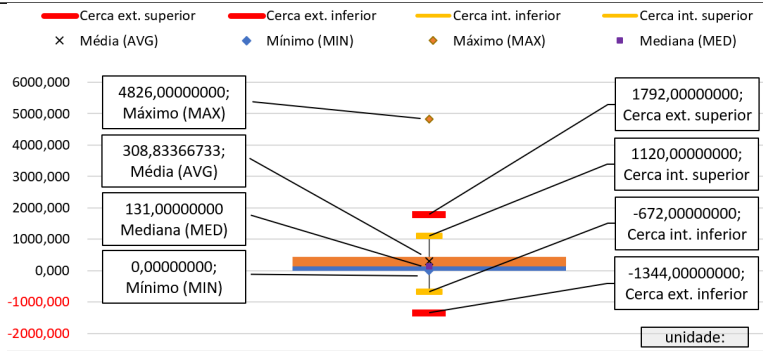
(SLEEP) <DEEPSLEEP<DURATION>>

IQR	99,25000000
RNG	354,00000000
VAR	6125,14779874
STD	78,26332346
CV	0,50258288
SK	0,10434858
KU	0,23295919



(PHYSI) <STEPS>

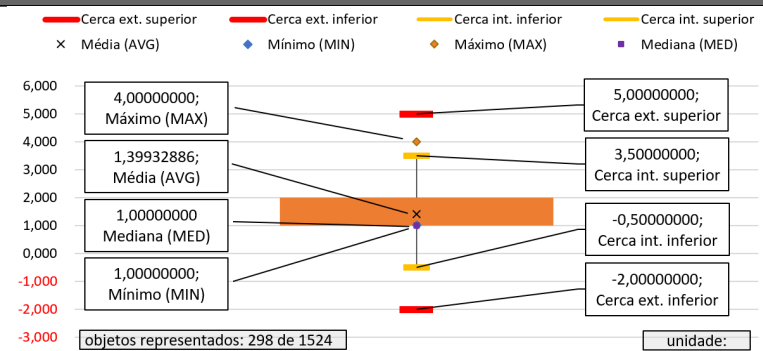
IQR	448,00000000
RNG	4826,00000000
VAR	195394,01977484
STD	442,03395772
CV	1,43130107
SK	1,20692312
KU	15,17870199



{WINDOWSACTIVITY}

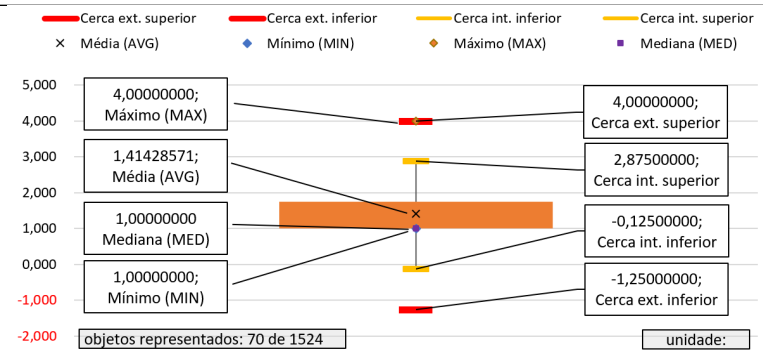
(WORKINTENSITY) <COMPUTERSUCCESSLOGIN<COUNT>>

IQR	1,00000000
RNG	3,00000000
VAR	0,47636318
STD	0,69019069
CV	0,49322980
SK	1,73573275
KU	3,02685261

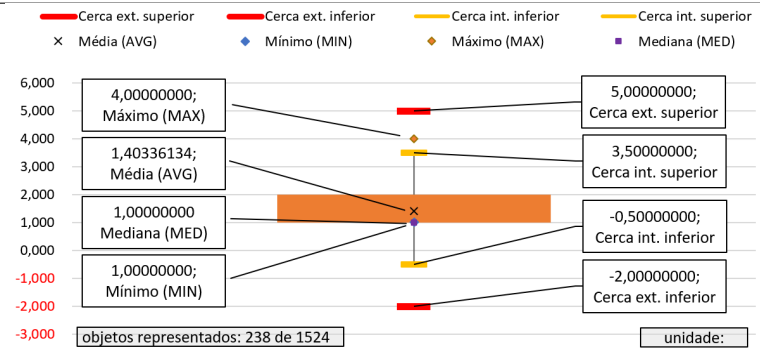


(WORKINTENSITY) <COMPUTERFAILEDLOGIN<COUNT>>

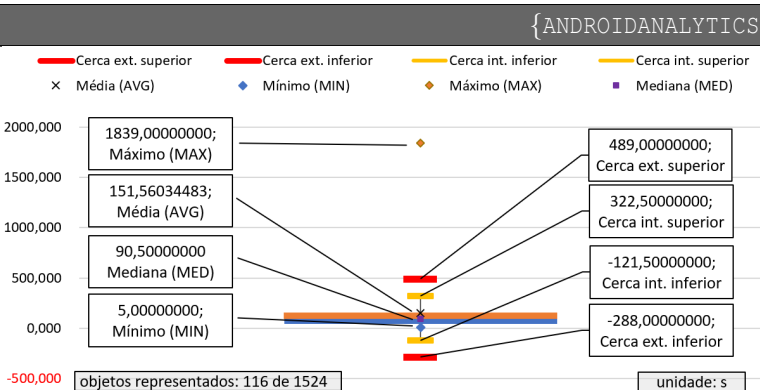
IQR	0,75000000
RNG	3,00000000
VAR	0,70993789
STD	0,84257812
CV	0,59576231
SK	1,47506458
KU	4,16661019



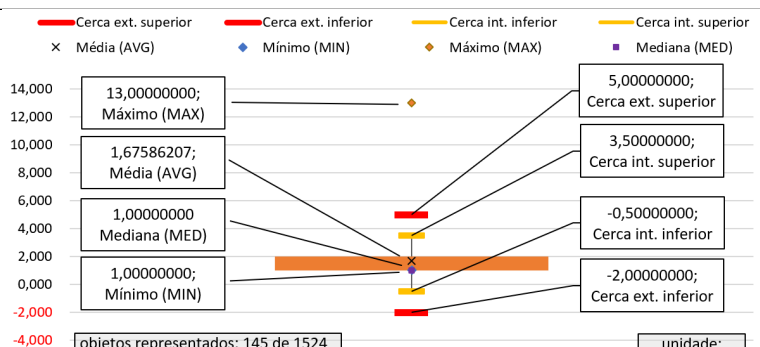
(WORKINTENSITY) <COMPUTERLOCKOUTS<COUNT>>	
<i>IQR</i>	1,00000000
<i>RNG</i>	3,00000000
<i>VAR</i>	0,41045279
<i>STD</i>	0,64066589
<i>CV</i>	0,45652240
<i>SK</i>	1,88879109
<i>KU</i>	1,94493699



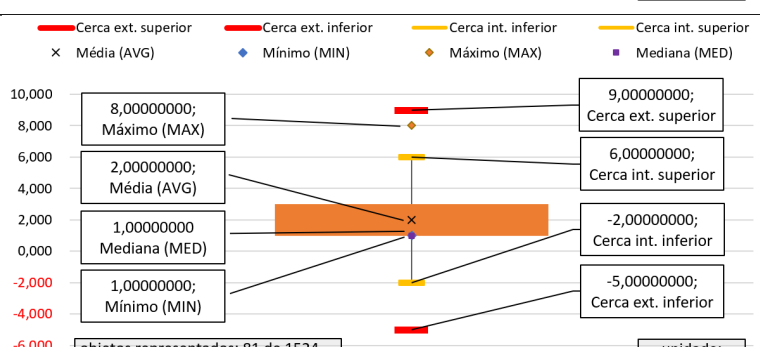
(CALL) <INCOMING<DURATION<SUM>>>	
<i>IQR</i>	111,00000000
<i>RNG</i>	1834,00000000
<i>VAR</i>	49113,17023988
<i>STD</i>	221,61491430
<i>CV</i>	1,46222229
<i>SK</i>	0,82657359
<i>KU</i>	31,04579092



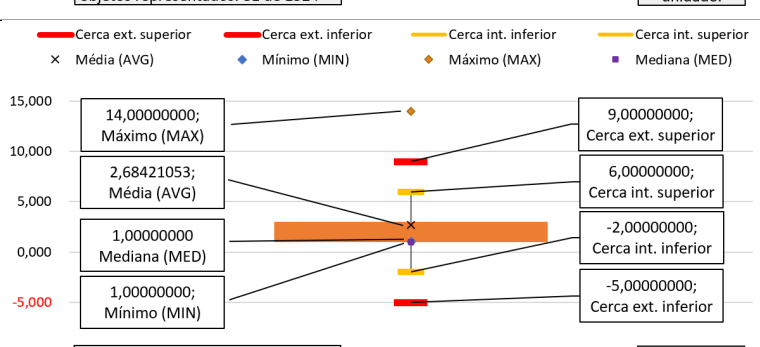
(CALL) <OUTGOING<COUNT>>>	
<i>IQR</i>	1,00000000
<i>RNG</i>	12,00000000
<i>VAR</i>	1,76226054
<i>STD</i>	1,32750161
<i>CV</i>	0,79213059
<i>SK</i>	1,52737005
<i>KU</i>	37,09785448



(CALL) <MISSED<COUNT>>>	
<i>IQR</i>	2,00000000
<i>RNG</i>	7,00000000
<i>VAR</i>	2,27500000
<i>STD</i>	1,50831031
<i>CV</i>	0,75415516
<i>SK</i>	1,98898063
<i>KU</i>	3,16849719



(SMS) <COUNT>>	
<i>IQR</i>	2,00000000
<i>RNG</i>	13,00000000
<i>VAR</i>	8,27596017
<i>STD</i>	2,87679686
<i>CV</i>	1,07174785
<i>SK</i>	1,75633937
<i>KU</i>	6,16642481



{ANDROIDANALYTICS}

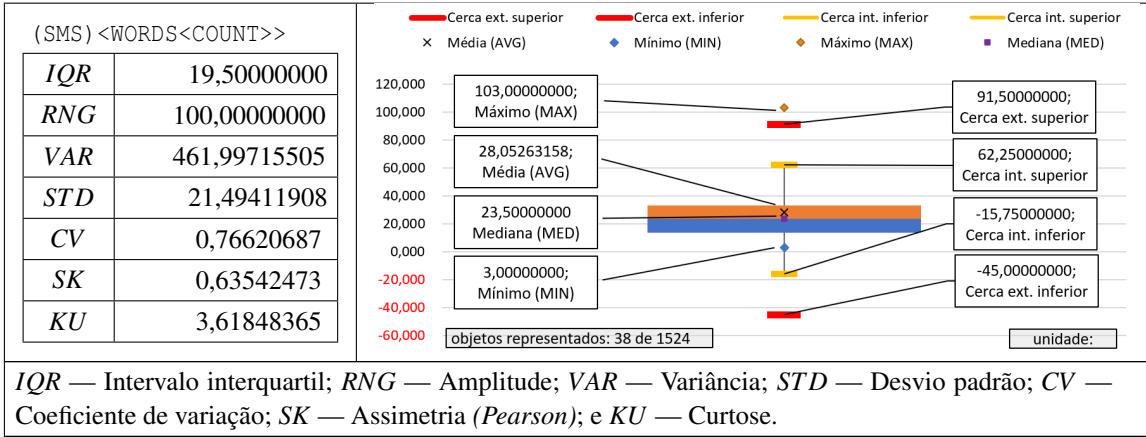


Tabela 5.9: Descrição de alguns sinais recolhidos presentes no *dataset*

5.5 Exploração dos dados do *dataset*

O *dataset* final apresentado neste capítulo, será utilizado no futuro para inferir o bem-estar através de algoritmos de ML. Pretende-se descobrir as variáveis de contexto e/ou propriedades extraídas mais adequadas para a deteção automática do bem-estar em trabalhadores de escritório. Nesta secção faz-se uma análise exploratória, tendo em vista a deteção de possíveis correlações nos dados do *dataset*. Estas correlações, poderão vir a fornecer pistas importantes para o processo de seleção de *features* a submeter no futuro à classificação algorítmica para a deteção automática do bem-estar.

Os classificadores são algoritmos de aprendizagem supervisionada que classificam a categoria de novos objetos, tendo por base a aprendizagem obtida durante a fase de treino. Durante esta fase, os algoritmos aprendem a classificar dados em classes, tendo em conta o valor dos atributos *ground-truth*, no contexto dos restantes valores recolhidos. Durante o processo de treino, os dados *ground-truth* funcionam como *targets* para o classificador. Após a fase de treino, o classificador processa novos objetos tendo por base o *dataset* pré-classificado resultante da fase de treino. Foi neste contexto que, na fase de preparação da experiência, se decidiu recolher dados a serem utilizados posteriormente pelos algoritmos como *ground-truth* (ver secção 3.1).

Os dados *ground-truth* da experiência relatada nesta tese, foram recolhidos através dos instrumentos {WELLBEING1} e {WELLBEING2}. Estes questionários foram aplicados durante o processo de recolha dos restantes dados de contexto de cada participante (ver secção 3.2). As perguntas destes questionários foram inspiradas na escala adaptada para Portugal da medição de manifestação do bem-estar psicológico (i.e. *Échelle de Mesure des Manifestations du Bien-Être Psychologique* {EMMBEP}) (Monteiro et al., 2012). A escala completa é composta por 25 questões agrupadas em 6 sub-escalas (factores): fe-

licidade, sociabilidade, controlo de si e dos acontecimentos, envolvimento social, autoestima, e equilíbrio.

A Tabela 5.10 resume as perguntas escolhidas pelo autor para o questionário da percepção do bem-estar {WELLBEING1}. A tabela apresenta também a frequência de respostas para cada pergunta. As respostas dos participantes baseavam-se na escolha de uma das seguintes três opções: i) 🗨️ discordo; ii) 🤝 nem concordo nem concordo; e iii) 👍 concordo. A tabela categoriza as perguntas em função da sub-escala original do {EMMBEP}. Para este questionário, o autor selecionou questões das sub-escalas: sociabilidade, controlo de si e dos acontecimentos, autoestima, e equilíbrio. O {WELLBEING1} também solicitou a cada participante a identificação do período do dia em que mais stress sentiu. Estes dados não são apresentados por não se tratarem de dados *ground-truth*.

Pergunta / Respostas									
SOCIABILIDADE									
Relacionei-me facilmente com as outras pessoas (W1Q1).									
<table border="1"> <thead> <tr> <th>Resposta</th> <th>Frequência (%)</th> </tr> </thead> <tbody> <tr> <td>1 - discordo</td> <td>2,60%</td> </tr> <tr> <td>2 - nem concordo nem discordo</td> <td>21,10%</td> </tr> <tr> <td>3 - concordo</td> <td>76,30%</td> </tr> </tbody> </table>		Resposta	Frequência (%)	1 - discordo	2,60%	2 - nem concordo nem discordo	21,10%	3 - concordo	76,30%
Resposta	Frequência (%)								
1 - discordo	2,60%								
2 - nem concordo nem discordo	21,10%								
3 - concordo	76,30%								
CONTROLO DE SI E DOS ACONTECIMENTOS									
Estive capaz de enfrentar situações difíceis de uma forma positiva (W1Q2).									
<table border="1"> <thead> <tr> <th>Resposta</th> <th>Frequência (%)</th> </tr> </thead> <tbody> <tr> <td>1 - discordo</td> <td>1,30%</td> </tr> <tr> <td>2 - nem concordo nem discordo</td> <td>28,90%</td> </tr> <tr> <td>3 - concordo</td> <td>69,70%</td> </tr> </tbody> </table>		Resposta	Frequência (%)	1 - discordo	1,30%	2 - nem concordo nem discordo	28,90%	3 - concordo	69,70%
Resposta	Frequência (%)								
1 - discordo	1,30%								
2 - nem concordo nem discordo	28,90%								
3 - concordo	69,70%								
AUTOESTIMA									
Senti que os outros gostavam de mim e me apreciavam (W1Q3).									
<table border="1"> <thead> <tr> <th>Resposta</th> <th>Frequência (%)</th> </tr> </thead> <tbody> <tr> <td>1 - discordo</td> <td>3,90%</td> </tr> <tr> <td>2 - nem concordo nem discordo</td> <td>30,30%</td> </tr> <tr> <td>3 - concordo</td> <td>65,80%</td> </tr> </tbody> </table>		Resposta	Frequência (%)	1 - discordo	3,90%	2 - nem concordo nem discordo	30,30%	3 - concordo	65,80%
Resposta	Frequência (%)								
1 - discordo	3,90%								
2 - nem concordo nem discordo	30,30%								
3 - concordo	65,80%								
Senti-me satisfeito(a) com o que fui capaz de alcançar, senti-me orgulhoso(a) de mim próprio (W1Q4).									
<table border="1"> <thead> <tr> <th>Resposta</th> <th>Frequência (%)</th> </tr> </thead> <tbody> <tr> <td>1 - discordo</td> <td>1,30%</td> </tr> <tr> <td>2 - nem concordo nem discordo</td> <td>31,60%</td> </tr> <tr> <td>3 - concordo</td> <td>67,10%</td> </tr> </tbody> </table>		Resposta	Frequência (%)	1 - discordo	1,30%	2 - nem concordo nem discordo	31,60%	3 - concordo	67,10%
Resposta	Frequência (%)								
1 - discordo	1,30%								
2 - nem concordo nem discordo	31,60%								
3 - concordo	67,10%								
EQUILÍBRIO									
A minha vida foi bem equilibrada entre as minhas atividades familiares, pessoais e académicas (W1Q5).									

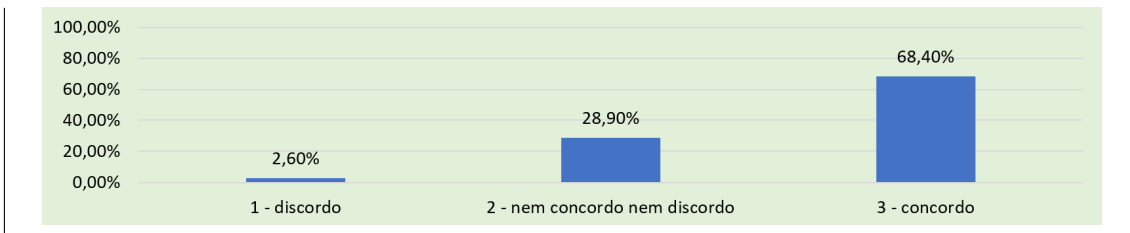


Tabela 5.10: Perguntas e análise de frequência das respostas do {WELLBING1}

A Tabela 5.11 apresenta as três perguntas escolhidas pelo autor para o questionário {WELLBING2}, e a frequência de resposta para cada uma. As opções de resposta disponibilizadas aos participantes, foram as mesmas utilizadas no questionário {WELLBING1}. Para este questionário foram selecionadas questões das seguintes sub-escalas: felicidade, e autoestima.

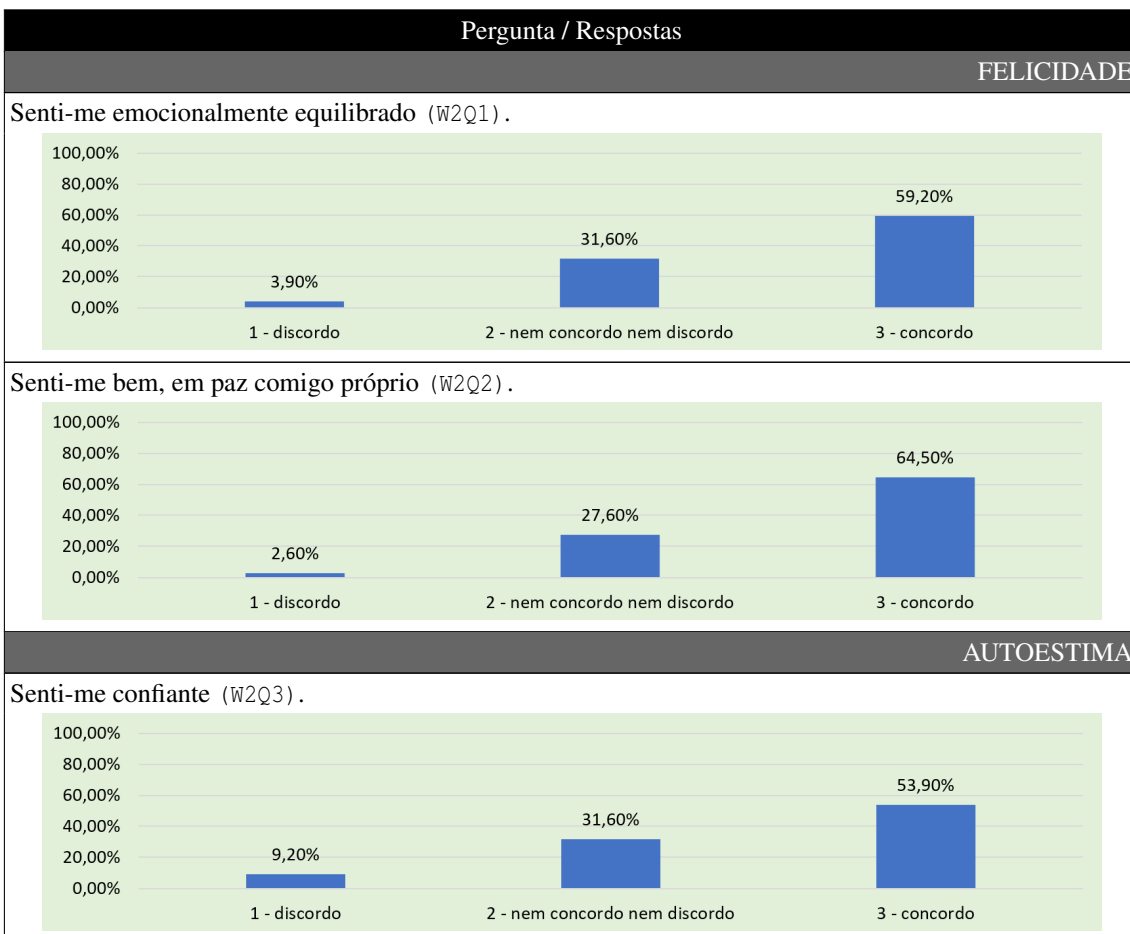


Tabela 5.11: Perguntas e análise de frequência das respostas do {WELLBING2}

Na análise exploratória de dados efetuada nesta secção, foi utilizado um subconjunto de dados do *dataset* final. Neste subconjunto, foram incluídas algumas das *features* consideradas pelo autor como candidatos à futura classificação algorítmica do bem-estar. As respostas aos questionários {WELLBING1} e {WELLBING2}, foram também incluídas no subconjunto por serem estas as variáveis para as quais se pretendia procurar correlações

(*ground-truth* do bem-estar). Nesta análise preliminar do *dataset*, decidiu-se agregar os dados por participante para cada dia de recolha ($SI = 1d$). No caso dos valores quantitativos utilizaram-se as funções de agregação média e desvio padrão. Para as medidas aditivas foi também calculada a soma dos valores. Não foi possível calcular o desvio padrão para todas as variáveis por falta de variância para alguns participantes em alguns dias (e.g. (SLEEP) <DEEPSLEEP<DURATION>>).

As respostas dos questionários {WELLBEING1} não foram sujeitas a qualquer agregação, atendendo a que apenas existia uma resposta para cada participante no mesmo dia. No caso do {WELLBEING2} foi utilizada a média dos valores das respostas, considerando a possibilidade de existirem múltiplas respostas ao questionário para o mesmo dia do mesmo participante. Como as opções de resposta para ambos os questionários eram as mesmas, foram ainda extraídas três propriedades adicionais com base nestes dados: i) soma dos valores das respostas de cada objeto do {WELLBEING1} ($W1SUM = W1Q1 + W1Q2 + W1Q3 + W1Q4 + W1Q5$); ii) soma dos valores das respostas de cada objeto do {WELLBEING2} ($W2SUM = W2Q1 + W2Q2 + W2Q3$); e iii) soma das somas das respostas por objeto ($WSUM = W1SUM + W2SUM$). Do resultado da agregação $SI = 1d$ resultou um subconjunto com 76 objetos ($N = 76$). Foi com base neste subconjunto de dados que se realizou uma análise de correlações, com o objetivo de perceber quais as *features* com maior grau de relacionamento com os dados *ground-truth* recolhidos através dos questionários. Considerando que não estavam reunidas as condições para a aplicação do coeficiente de correlação de *Pearson*, optou-se por utilizar o coeficiente de correlação de *Spearman*. Além de alguns sinais não seguirem a distribuição normal, as respostas dos questionários são valores ordinais e não quantitativos, pressupostos para a utilização do coeficiente de correlação de *Pearson* (Leard, 2021a).

O coeficiente de *Spearman* (ρ) é uma medida não-paramétrica que mede a força e direção da relação linear existente entre um par de variáveis (Leard, 2021b). São três os requisitos para se utilizar este coeficiente: i) as variáveis devem ser do tipo ordinal (e.g. escalas de *Likert*) e quantitativo (e.g. peso, altura); ii) as duas variáveis devem formar um par com dados do mesmo objeto; e iii) deve existir uma relação monotónica entre as duas variáveis (i.e. o valor de ambas as variáveis crescem ou decrescem em conjunto, ou uma cresce e a outra decresce). O (ρ) varia entre $[-1;1]$ e indica a relação existente entre as duas variáveis. O valor absoluto representa a força da relação, e o sinal do valor indica a direção dessa relação (Leard, 2021b). Se ambas aumentam ou diminuem, o coeficiente é positivo ($\rho > 0$). Se uma tende a aumentar enquanto que a outra tende a diminuir, o valor de ρ será um valor negativo ($\rho < 0$). Se não existir qualquer correlação entre ambas as variáveis, o índice é nulo ($\rho = 0$). Quanto mais o valor se afastar do zero, mais forte será a relação entre as variáveis em análise.

A Tabela 5.12 apresenta os coeficientes de correlação de *Spearman* obtidos no teste executado sobre as variáveis escolhidas para o subconjunto de dados utilizado nesta análise

exploratória. A tabela apresenta apenas as correlações em que $\rho > 0,5$. Conjuntamente com o coeficiente é também indicado o nível de significância (p), que indica quão significativo é o valor de correlação calculado em cada teste. Considera-se um valor estatisticamente significativo quando $p < 0,05$, que corresponde a um nível de confiança de 95%. A tabela indica também a quantidade de objetos considerados em cada teste (N). Apesar de o subconjunto ser composto por 76 objetos resultantes da agregação $SI = 1d$, alguns atributos não tem valor em alguns objetos.

Não são apresentados valores para a quarta questão do {WELLBEING1} (i.e. W1Q4), porque não foram detetadas correlações com o coeficiente de *Spearman* nas respostas para esta questão. O coeficiente de correlação mais elevado obtido no teste realizado, foi o resultante da relação entre a propriedade W2SUM e a média de agregação do sinal (SMS) <COUNT>. Neste caso, houve uma correlação forte e positiva entre ambas as variáveis, que foi estatisticamente significativa ($\rho = 0,730$; $p = 0,001$; $N = 16$).

Sinal		W1Q1	W1Q2	W1Q3	W1Q5	W2Q1	W2Q2	W2Q3	W1SUM	W2SUM	WSUM
(SMS) <WORDS <COUNT>> — Σ	ρ									0,510	
	p	-	-	-	-	-	-	-	-	0,043	-
	N									16	
(SMS) <WORDS <COUNT>> — μ	ρ					0,520				0,564	
	p	-	-	-	-	0,039	-	-	-	0,023	-
	N					16				16	
(SMS) <CHARS <COUNT>> — μ	ρ		0,504							0,516	
	p	-	0,046	-	-	-	-	-	-	0,041	-
	N		16							16	
(SMS) <COUNT> — Σ	ρ					0,549	0,534	0,614		0,681	
	p	-	-	-	-	0,028	0,033	0,011	-	0,004	-
	N					16	16	16		16	
(SMS) <COUNT> — μ	ρ					0,511	0,584	0,681		0,730	
	p	-	-	-	-	0,043	0,018	0,004	-	0,001	-
	N					16	16	16		16	
(WORKINTENSITY) <COMPUTERFAILED LOGIN<COUNT>> — σ	ρ			0,509					0,527		0,643
	p	-	-	0,026	-	-	-	-	0,020	-	0,004
	N			19					19		18
(CALL) <INCOMING <DURATION>> — Σ	ρ							0,527			
	p	-	-	-	-	-	-	0,001	-	-	-
	N							37			
(CALL) <INCOMING <DURATION>> — μ	ρ							0,502			
	p	-	-	-	-	-	-	0,002	-	-	-
	N							37			
(CALL) <INCOMING <DURATION>> — σ	ρ							0,527			
	p	-	-	-	-	-	-	0,010	-	-	-
	N							23			
(CALL) <MISSED CALL<COUNT>> — Σ	ρ	-0,579		-0,520					-0,513		-0,549
	p	0,001	-	0,003	-	-	-	-	0,004	-	0,002
	N										

	N	30		30					30		30
(CALL) <MISSED CALL<COUNT>> — μ	ρ										-0,513
	p	-	-	-	-	-	-	-	-	-	0,004
	N										30
(CALL) <INCOMING <COUNT> — σ	ρ	-0,503		-0,658	-0,533				-0,582		
	p	0,014	-	0,001	0,009	-	-	-	0,004	-	-
	N	23		23	23				23		
μ — média; σ — desvio padrão; e Σ — soma.											

Tabela 5.12: Tabela de correlações do subconjunto de dados em análise (*Spearman*)

5.6 Análise

Neste capítulo apresentou-se a estrutura do *dataset* final e a arquitetura de *views* que foi implementada para dar suporte à fase de pré-processamento e posterior exploração de dados. Fez-se uma caracterização dos participantes e dos dados presentes no *dataset* final, e foram também apresentados alguns resultados da análise exploratória de dados efetuada sobre o *dataset* final apresentado como contribuição deste capítulo.

No início da fase de pré-processamento, havia alguma expectativa relativamente à quantidade e qualidade dos dados recolhidos. A fase de recolha tinha sido bastante exigente. Tinha-se recolhido dados de vários participantes durante mais de dois meses, utilizando vários instrumentos & sensores de recolha. Além disso, um dos sensores utilizava elétrodos colados na mão dos participantes, para recolher dados de contexto durante o desempenho as suas funções habituais como trabalhadores de escritório. Na fase de pré-processamento, teve que se resolver vários problemas relacionados com a qualidade dos dados. O autor pretendia contribuir com um grande *dataset* multimodal, contendo dados realísticos e com qualidade. Por este motivo, recolheu-se dados em ambiente real, e cada sinal foi tratado individualmente durante a fase de pré-processamento.

Na opinião do investigador, a escolha do ambiente onde é feita a recolha de dados, implica diretamente a qualidade dos dados. A probabilidade de se obter um *dataset* com mais qualidade, será maior quando se recolhe dados em ambiente controlado. Quando se recolhe em laboratório, os investigadores gerem o contexto implementando medidas para promover o sucesso da experiência. No entanto, apesar de se conseguir melhor qualidade de dados, a informação desses dados poderá estar contaminada pela influência direta dos investigadores no processo. Por outro lado, a recolha em ambiente real poderá recolher dados de comportamento mais natural. Os participantes poderão abstrair-se da sua participação na experiência, não estando tão atentos às suas ações e reações. Contudo, apesar de serem estes os dados que se pretendem recolher, os participantes também não estarão tão atentos às questões logísticas da recolha. Decidiu-se recolher dados em

ambiente real. Optou-se por priorizar a qualidade da informação presente nos dados relativamente à qualidade dos dados recolhidos. No entanto, sabia-se que tal decisão poderia implicar uma fase de pré-processamento mais longa e exigente. Em consequência, uma das técnicas de pré-processamento de relevo utilizadas, foi a reconstrução de sinal através da [INTERPOLATION]. As falhas de sinal resultaram precisamente do facto dos participantes se terem abstraído da participação na experiência, e se terem esquecido, por várias vezes, de atender às questões logísticas da recolha (e.g. verificar a correta afixação dos elétrodos na mão).

O *dataset* final contém dados pré-processados de várias modalidades. Muitos dos trabalhos presentes na literatura da área do AC, recolhem dados de poucos participantes durante um curto espaço de tempo, e em ambiente controlado. Além disso, são vários os que se focam na recolha de dados de uma só modalidade. O *dataset* final apresentado neste capítulo foi recolhido em ambiente real, ao longo de vários meses. Contém dados de 20 participantes recolhidos através de vários instrumentos & sensores. Posteriormente foram pré-processados, numa fase extensa e exigente atendendo à quantidade dos dados recolhidos. As operações de pré-processamento sobre os dados foram anotadas e, essas anotações, fazem parte do *dataset* final.

Acredita-se estar a contribuir com um *dataset* que espelha o ritmo e stress próprio de um ambiente de escritório. Considera-se ter-se conseguido capturar do contexto, o real comportamento resultante das múltiplas tarefas e interações sociais existentes em simultâneo num escritório. O autor acredita que, os resultados obtidos na análise exploratória de dados apresentada neste capítulo, poderão antever bons resultados quando o *dataset* for usado como *input* em algoritmos ML, para classificar o bem-estar em trabalhadores de escritório.

O *dataset* é disponibilizado em formato CSV e no formato de base de dados *Microsoft SQL Server 2017*. Atendendo à estrutura do *dataset* e arquitetura de *views*, recomenda-se a exploração de dados efetuando *queries* SQL diretamente à base de dados. O *dataset* final integrado, pré-processado e anotado, materializa a contribuição C09 desta tese.

Capítulo 6

CONCLUSÃO

Nesta tese procurou descrever-se o projeto que culminou com a contribuição de um *dataset* multimodal recolhido em ambiente real e posteriormente pré-processado e anotado. Este *dataset* contém dados de contexto de 20 trabalhadores de escritório de 7 organizações distintas. Acreditamos que este *dataset* poderá ser uma contribuição fundamental para a criação de um sistema de deteção automática de emoções. Beneficiarão com esta tecnologia as próprias pessoas que trabalham no escritório, mas também as organizações que poderão ver melhorados os seus indicadores de produtividade e de retenção de trabalhadores nos seus quadros.

A deteção automática de emoções é importante para melhorar a interação entre pessoas e máquinas (Basu et al., 2016). Dada a importância das emoções nos processos de comunicação dos humanos, a interação entre homens e dispositivos será mais natural se os sistemas ubíquos conseguirem perceber e adaptar-se às emoções transmitidas (Busso et al., 2004; Castellano et al., 2007). No entanto, apesar de existir bastante investigação com o objetivo de tornar mais produtiva a relação entre homem e máquina, a interação entre ambos está ainda longe de ser natural, entre outros, devido à fraca comunicação emocional (Korkmaz and Atasoy, 2015). A aferição de emoções é um dos maiores desafios dos investigadores da área do AC, devido à existência de vários fatores que afetam o estado emocional do ser humano (e.g. fisiologia, experiência subjetiva, antecedentes) (Babiker et al., 2013). Cada pessoa pode reagir de forma diferente a um mesmo estímulo emocional e até a mesma pessoa, pode reagir de forma diferente em função do momento e do espaço onde se encontra (Kim and André, 2008; Dong et al., 2011). As emoções não são passíveis de medição direta (Raudonis, 2013; Nawasalkar et al., 2013). Assim, uma das formas utilizadas pelos investigadores para as tentar inferir, é recorrendo a dados recolhidos do contexto que posteriormente podem ser utilizados por classificadores de emoções (Raudonis, 2013).

Na literatura existem várias investigações que tentam detetar automaticamente emoções. No entanto, os trabalhos existentes apresentam problemas ao nível da exatidão de resultados. Talvez um dos fatores que contribui para esta ineficiência, seja a inexistência

de uma definição clara e objetiva do que são realmente as emoções. Será complicado dizer que se encontrou algo que não se consegue definir, porque também não se sabe exatamente onde e o que se procurar de forma objetiva. No entanto, e apesar destas limitações, existe consenso que as emoções influenciam e são influenciadas pelo contexto das pessoas. Por esse motivo, os investigadores recorrem a algoritmos de classificação para tentar inferir emoções, utilizando como *input* dados recolhidos de contexto dos participantes nas suas experiências. A maioria dos investigadores do AC recolhem dados durante um curto espaço de tempo e em ambiente controlado. São poucos os que recolhem dados de várias modalidades, e ainda menos os que consideram dados de contexto social, cultural e religioso. Neste trabalho considerou-se importante recolher dados alargados a estes contextos para abranger informação que pode influenciar significativamente a deteção de emoções e, mais concretamente, o bem-estar dos trabalhadores de escritório. Ao nível social, a interação entre pessoas pode desencadear emoções positivas como a alegria, ou negativas como a vergonha. A obtenção de conhecimento ou competências pode suscitar emoções positivas como satisfação ou negativas como stress. Ao nível religioso as pessoas poderão sentir emoções mais positivas pelo cumprimento dos desígnios do seu credo ou negativas inerentes a sentimentos de culpa ou pecado (Fredrickson, 2002). Por estas razões, conduziu-se uma experiência de recolha de dados em ambiente real e de forma mais alargada, procurando obter um *dataset* completo, com recurso a 11 instrumentos & sensores. O *dataset* final contém dados de contexto fisiológico, social, cultural e religioso. Consideramos que este *dataset* constitui uma contribuição fundamental nesta área de AC, e pretendemos vir a utilizá-lo brevemente como *input* em vários algoritmos ML, com o intuito de automatizar a classificação do bem-estar em trabalhadores de escritório.

A construção de um *dataset* variado foi definida como uma prioridade, desde a fase de planeamento da experiência de recolha apresentada no Capítulo 3. Pretendia-se diferenciar este projecto de outras investigações, apresentando um *dataset* multimodal, recolhido em ambiente real, com dados de qualidade, obtidos de vários participantes, trabalhando em diferentes instituições. Apesar de previamente planeada, a recolha de dados de contexto foi bastante exigente. Por indisponibilidade de alternativas viáveis, um dos instrumentos utilizados na recolha utilizava elétrodos para recolher dados de contexto. Quando se decidiu recolher dados utilizando um instrumento obstrutivo, a preocupação com a qualidade final dos dados aumentou de forma significativa. Pretendia-se recolher um *dataset* extenso e realístico, mas que também tivesse boa qualidade. O objetivo era potenciar a eficiência algorítmica, e sentir confiança nos resultados obtidos, quando se utilizasse o *dataset* como *input* para classificar o bem-estar em trabalhadores de escritório. Neste contexto, efetuaram-se várias visitas de acompanhamento em todas as iterações de recolha. Foi necessário ajustar várias vezes os procedimentos, recordar os participantes das boas práticas na utilização dos instrumentos & sensores, e lembrar a resposta atempada e cuidada aos questionários. A recolha em ambiente real acarreta desafios que não

se colocam nas recolhas em laboratório. Logo à partida, será muito mais fácil promover a qualidade dos dados recolhidos através de elétrodos em laboratório, do que em ambiente real, uma vez que os participantes realizam as tarefas normais de expediente, no ritmo stressante característico de um escritório. Portanto, quando na fase de planeamento se descartou a recolha de dados em laboratório, pretendia-se conseguir recolher dados mais realísticos em ambiente laboral. Contudo, a probabilidade de existirem anomalias na recolha dos dados seria exponencialmente maior, o que implicaria necessariamente uma fase de pré-processamento mais longa e exigente.

No final do período de recolha, havia grande expectativa relativamente à qualidade e quantidade de dados recolhidos. A estratégia para a fase de pré-processamento, relatada no Capítulo 4, iniciou-se com a integração de todos os dados recolhidos numa base de dados relacional. A cada ficheiro componente do *dataset raw*, corresponderia uma tabela espelho na base de dados e, sobre essas tabelas, iriam ser criadas *views* para apoio do pré-processamento e cálculo das propriedades extraídas. Além disso, pretendia-se implementar um mecanismo para documentar as técnicas aplicadas sobre os dados. Neste contexto, complementou o modelo de dados com atributos adicionais, para permitir a edição anotada de dados, sem alteração dos dados originalmente recolhidos de contexto. Sabia-se que existiriam vários problemas para resolver, mas não se conhecia a sua dimensão. Durante a fase de integração, constatou-se a existência de várias anomalias, e as mais relevantes estavam relacionadas com algumas falhas de sinal nos dados recolhidos pelos sensores {EDAMOVE} e {MIBAND}. Assim, de entre as técnicas de pré-processamento utilizadas, uma das mais relevantes foi a [INTERPOLATION] para reconstrução de sinal.

O pré-processamento é uma fase importante para os sistemas de deteção automática de emoções (Kandel et al., 2012; Aggarwal, 2013). É nesta fase que os dados são trabalhados com o objetivo de melhorar a qualidade (Urges, 2021; Gama et al., 2012). Considerando a quantidade de dados recolhidos, foi necessário tornar o processo de análise dos dados mais rápido e menos suscetível ao erro humano. Utilizar técnicas exploratórias de dados, e análise de medidas estatísticas na tentativa de encontrar problemas, tornaria o processo muito longo e ainda mais exigente. Segundo Andrew Ng, grande parte do tempo utilizado pelos investigadores em sistemas de ML, é despendido na preparação dos dados (Urges, 2021). Foi neste contexto que se criou o {DQVA} e o {DQVA-PP}, instrumentos apresentados na secção 4.4. Sentia-se a necessidade de agilizar o processo de pré-processamento, desenhando uma ferramenta que permitisse analisar dados de forma mais simples e rápida. Além disso, considerando a quantidade de dados existentes no *dataset raw*, a análise de qualidade de dados utilizando apenas valores agregados, poderia encobrir padrões no sinal e dificultar a deteção de anomalias nos dados.

O resultado final da fase de pré-processamento é o *dataset* pré-processado adicionado das propriedades extraídas mencionadas na secção 4.3. Este *dataset* é descrito no Capítulo 5, e é a contribuição principal desta tese.

Na secção seguinte faz-se um resumo das contribuições deste trabalho. Posteriormente, resume-se a experiência adquirida com recomendações para outros trabalhos, e termina-se abordando os trabalhos futuros.

6.1 Resumo das contribuições

Apesar de todas as contribuições apresentadas serem importantes e necessárias à condução dos trabalhos deste projeto, algumas serão mais relevantes do que outras. Nesta secção, destacam-se as contribuições mais importantes efetuadas nesta tese.

De entre as várias contribuições considera-se que as mais relevantes serão: i) o *dataset* final pré-processado e anotado apresentado como contribuição no Capítulo 5; e ii) o *dataset raw* que contém os dados recolhidos em ambiente real, contribuição do Capítulo 3. O *dataset* final poderá ser uma contribuição fundamental para investigações que estudem o reconhecimento automático de emoções. Pretende-se vir a utilizar este *dataset* na deteção automática do bem-estar em trabalhadores de escritório.

Destacam-se ainda outras duas contribuições com relevo: i) o Relatório Técnico (RT) publicado no Repositório Institucional da Universidade Fernando Pessoa, apresentado como contribuição no Capítulo 2 (Carvalho et al., 2018); e ii) o {DQVA} que permitiu tornar mais simples e rápida a fase de pré-processamento deste projeto, contribuição do Capítulo 4 (Carvalho et al., 2021). O RT é um documento dinâmico que permite o registo sistemático, sintético e organizado da informação das investigações relacionadas com o AC. O {DQVA} e a variante {DQVA-PP} facilitam a avaliação da qualidade dos dados através de visualizações (ver secção 4.4).

As restantes contribuições desta tese, podem ser consultadas na Tabela 1.1

6.2 Experiência adquirida e recomendações

Durante a fase da recolha de dados de contexto, ocorreram vários problemas com impacto nos dados recolhidos e que devem ser considerados no futuro. O primeiro está relacionado com o facto de se ter recolhido dados de contexto com elétrodos, em ambiente real. As alternativas implicavam custos financeiros consideráveis e, como se tinha conseguido o empréstimo de dois {EDAMOVE}, optou-se por recolher dados do (EDA) com este sensor. Assim, um dos problemas mais relevantes na fase de pré-processamento, foi a necessidade de reconstrução dos sinais recolhidos por estes sensores. Em projetos futuros deve ser reequacionada a utilização de instrumentos & sensores obstrutivos, principalmente quando se tratar de uma recolha em ambiente real com duração de vários dias. Fica-se também com a percepção de que é no primeiro dia que os participantes menos se esquecem das questões logísticas associadas ao processo de recolha. A qualidade dos da-

dos recolhidos é tendencialmente maior no início do período de recolha e degrada-se de forma progressiva ao longo dos restantes dias. Recolher dados em ambiente real permite tempos de recolha mais alargados, atendendo a que as pessoas continuam com a sua vida de forma próxima ao normal. No entanto, verificou-se que, recolher dados durante muito tempo, também pode prejudicar a qualidade do *dataset* devido ao desgaste que pode gerar nos participantes. Este fator será o principal responsável, pela menor atenção dada pelos participantes às questões logísticas (e.g. verificar a correta afixação de elétrodos, resposta atempada a questionários).

O segundo problema a destacar está relacionado com a recolha de dados de contexto laboral. A dificuldade de colaboração dos departamentos de recursos humanos para preenchimento do {LABOR}, foi evidente nas várias incoerências detetadas nos dados preenchidos. Durante a revisão destes dados, foi possível constatar a forma pouco rigorosa com que foram inseridos. Uma solução poderá passar por envolver o pessoal dos recursos humanos no projeto de outra forma (e.g. convidar a participar nas reuniões de apresentação do projeto aos responsáveis). Outra opção poderá passar por solicitar o preenchimento do questionário aos próprios participantes. No entanto, como são solicitados dados históricos (e.g. atrasos ocorridos nos últimos 30 dias), teria que se iniciar o período de recolha mais cedo, para que os participantes pudessem ir fazendo os registos necessários ao longo do tempo. No início da experiência, achou-se que seria mais lógico solicitar estes dados ao departamento de recursos humanos, por ser da sua competência a manutenção desta informação atualizada. No entanto, foi perceptível que nem sempre as organizações mantêm registos atualizados relativos ao controlo de presenças, férias gozadas e não gozadas, etc.

Outra questão a rever é a recolha da interação social dos participantes através das (CALL) e (SMS). Existe uma grande disparidade de número de chamadas telefónicas e SMS enviados e recebidos, entre os participantes. Estas diferenças podem ter várias justificações que vão desde a maneira de ser das pessoas até às políticas de trabalho de cada organização (e.g. timidez dos funcionários, organização de trabalho das equipas e das próprias pessoas, autonomia, tipo e ritmo de trabalho). No entanto, durante as várias visitas de acompanhamento efetuadas em cada iteração de recolha, foi possível perceber a diversidade de formas que as pessoas utilizam para comunicar interna e externamente. Na interação social através da voz é essencialmente utilizado o *smartphone* pessoal, mas também é utilizado o telefone fixo, o *smartphone* de outros, e a comunicação face-a-face. Ao nível escrito a diversidade é ainda maior. Além do SMS, são também utilizados outros instrumentos como o email e ferramentas de *Instant Messaging* (IM) (e.g. *Rocket Chat WhatsApp, Skype, Snapchat, Slack, Telegram, Facebook Messenger*). A quantidade de ferramentas utilizadas por utilizador também varia muito. Alguns utilizadores apenas usam uma ou outra ferramenta de forma esporádica, outros utilizam várias destas ferramentas, cumulativamente e de forma habitual (e.g. alguns utilizadores acedem a

várias contas de email utilizando clientes de email diferentes ou o próprio *browser*). Será difícil conseguir recolher dados de todas estas fontes de informação. O levantamento prévio acerca dos instrumentos utilizados pelas organizações e participantes na recolha, implicaria necessariamente que se fizesse o trabalho de angariação de participantes num momento muito anterior ao início do período de recolha. No entanto, receia-se que as pessoas e organizações, acabem por desmotivar quanto à participação na experiência, fruto do compasso de espera necessário à preparação e desenvolvimento dos instrumentos de recolha. Talvez a melhor alternativa, seja alargar a recolha a outros instrumentos, e se possível solicitar aos participantes que dêem preferência à comunicação através destes.

Anotaram-se ainda outros problemas relacionados com a estabilidade e permissões dos SO. Em projetos futuros, sugere-se que também o administrador do sistema esteja presente na reunião de apresentação da experiência aos responsáveis da organização. A ideia é que se sinta envolvido na experiência e colabore na instalação e configuração de aplicações quando necessário. Ao nível da estabilidade dos SO, o problema maior é a imprevisibilidade. Não sendo possível prever antecipadamente os problemas deste género, sugere-se o registo em papel utilizando documentos *backup* de suporte como se fez neste projeto.

Da mesma forma que se ficou com a perceção que os participantes atendem mais às questões logísticas no início do período de recolha, termina-se também com a sensação de que os questionários são respondidos de forma mais cuidada e assertiva, ao início de cada manhã. Talvez exista maior predisposição para executar tarefas relacionadas com questões logística no início de cada dia. Verifica-se que a acumulação de tarefas e de stress, durante e na fase final do dia, prejudica a qualidade dos dados que são recolhidos através dos questionários.

No decurso da utilização do {DQVA}, foram identificadas algumas propostas de melhoria a considerar na próxima versão da ferramenta. Uma das propostas passará por diferenciar a representação de *gap* de sinal resultante da falta de dados (e.g. problema no sensor de recolha), da falta de sinal causada pelo fim do processo recolha (e.g. finalização da fase de recolha mais cedo do que os restantes participantes). Esta diferenciação na representação poderá facilitar a identificação dinâmica do último momento de recolha, da ausência de sinal por problemas na recolha. Esta alteração pode facilitar a representação visual dos rótulos dos dados, aquando da representação de várias etiquetas no mesmo ponto temporal (ver Figura 4.12). Quando as etiquetas identificativas dos participantes dos *gaps* de sinal no fim antecipado de recolha aparecem sobrepostas, é necessário deslocá-las manualmente para as conseguir ler. Todos os *gaps* estão representados numa única série transversal a todos os participantes, não sendo possível selecionar, por exemplo, a apresentação dos *gaps* de um participante em específico. Numa versão atualizada da ferramenta os *gaps* de cada participante serão incluídos numa série autónoma para apresentação ou ocultação individual.

Por fim, o autor considera que seria interessante estudar os vários sinais recolhidos no contexto dos valores basais de cada participante para cada sinal (emocionalmente neutros) (e.g. o intervalo de valores basais do (HR) para cada pessoa é diferente). Por isso, a análise de cada sinal só pode ser corretamente interpretada, no contexto dos valores basais dos respetivos sinais, para a mesma pessoa. É convicção do autor que, se forem utilizados estes valores para cada participante/sinal, os resultados obtidos serão melhores, por causa da possibilidade de comparação de valores no contexto de cada um.

6.3 Trabalhos futuros

Uma vez que a área de AC está em constante evolução, pretende-se continuar a atualizar o Relatório Técnico com os trabalhos que forem surgindo, relacionados em particular com a deteção automática de emoções. Pretende-se também utilizar o *dataset* final pré-processado, contribuição principal desta tese, para treinar e testar vários algoritmos ML, comparando a sua exatidão na classificação do bem-estar em trabalhadores de escritório. Por fim, pretende-se continuar a disseminar os resultados desta tese, publicando não só os desenvolvimentos teóricos na base da revisão bibliográfica, como também resultados mais recentes relacionados com os desenvolvimentos das ferramentas produzidas na tese. Adicionalmente, pretende-se explorar também a publicação de resultados associados à natural utilização de classificadores para prever automaticamente o bem-estar em trabalhadores de escritório.

Capítulo 7

ANEXOS

7.1 Questionário {SOCIOCULTURAL}

QUESTIONÁRIO SOCIOCULTURAL

1. ID organização:

2. Participante (número de identificação):

Avançar para a pergunta 3

DADOS GERAIS

3. Idade:

4. Género:

Marcar apenas uma oval.

- feminino
 masculino
 outro

5. Etnia:

Marcar apenas uma oval.

- caucásica
 amarela
 negra
 outra

6. Estado civil:

Marcar apenas uma oval.

- solteiro/a
- casado/a
- divorciado/a
- união de facto
- outro

7. Religião praticada:

Marcar apenas uma oval.

- católica
- evangélica
- testemunha de jeová
- anglicana
- islâmica
- outra
- nenhuma

8. Com quem vive:

Marcar apenas uma oval.

- sozinho/a
- família direta (cônjuge, filhos)
- família indireta (irmãos ou pais)
- outro

9. Nível de qualificação:

Marcar apenas uma oval.

- nível 1 (2º ciclo do ensino básico)
- nível 2 (3º ciclo do ensino básico)
- nível 3 (ensino secundário para ingresso no ensino superior)
- nível 4 (ensino secundário com estágio profissional de, no mínimo, 6 meses)
- nível 5 (qualificação de nível pós-secundário não superior)
- nível 6 (licenciatura)
- nível 7 (mestrado)
- nível 8 (doutoramento)

DEPENDENTES

Informação, até 4 pessoas, que inspirem atenção, cuidados ou outras preocupações diárias, de carácter permanente (e.g. filhos, enteados, adotados, pais, tios, etc.).

10. Tem dependentes:

Marcar apenas uma oval.

- sim
- não *Avançar para a pergunta 38*

DEPENDENTE

11. Grau de parentesco:

Marcar apenas uma oval.

- filho/a (enteado/a, adotado/a)
- cônjuge
- pai
- mãe
- outro

12. Idade:

13. Vive na mesma habitação de forma permanente:

Marcar apenas uma oval.

sim

não

14. Dependência financeira:

Marcar apenas uma oval.

sim

não

15. Dependência nas tarefas do dia-a-dia:

Marcar apenas uma oval.

sim

não

16. Tem necessidades especiais:

Marcar apenas uma oval.

sim

não

17. Inserir outro dependente:

Marcar apenas uma oval.

- sim
 não *Avançar para a pergunta 38*

DEPENDENTE

18. Grau de parentesco:

Marcar apenas uma oval.

- filho/a (enteado/a, adotado/a)
 cônjuge
 pai
 mãe
 outro

19. Idade:

20. Vive na mesma habitação de forma permanente:

Marcar apenas uma oval.

- sim
 não

21. Dependência financeira:

Marcar apenas uma oval.

- sim
 não

22. Dependência nas tarefas do dia-a-dia:

Marcar apenas uma oval.

sim

não

23. Tem necessidades especiais:

Marcar apenas uma oval.

sim

não

24. Inserir outro dependente:

Marcar apenas uma oval.

sim

não *Avançar para a pergunta 38*

DEPENDENTE

25. Grau de parentesco:

Marcar apenas uma oval.

filho/a (enteado/a, adotado/a)

cônjuge

pai

mãe

outro

26. Idade:

27. Vive na mesma habitação de forma permanente:

Marcar apenas uma oval.

sim

não

28. Dependência financeira:

Marcar apenas uma oval.

sim

não

29. Dependência nas tarefas do dia-a-dia:

Marcar apenas uma oval.

sim

não

30. Tem necessidades especiais:

Marcar apenas uma oval.

sim

não

31. Inserir outro dependente:

Marcar apenas uma oval.

sim

não *Avançar para a pergunta 38*

DEPENDENTE

32. Grau de parentesco:

Marcar apenas uma oval.

- filho/a (enteado/a, adotado/a)
- cônjuge
- pai
- mãe
- outro

33. Idade:

34. Vive na mesma habitação de forma permanente:

Marcar apenas uma oval.

- sim
- não

35. Dependência financeira:

Marcar apenas uma oval.

- sim
- não

36. Dependência nas tarefas do dia-a-dia:

Marcar apenas uma oval.

- sim
- não

37. Tem necessidades especiais:

Marcar apenas uma oval.

sim

não

HABITAÇÃO

Informação relacionada com o local onde vive, distância e acessibilidade ao local de trabalho.

38. Tipo de habitação:

Marcar apenas uma oval.

apartamento

moradia

quinta

outra

39. Ambiente habitacional:

Marcar apenas uma oval.

cidade

campo

40. Qualidade da habitação atual (conforto proporcionado):

Marcar apenas uma oval.

0 1 2 3 4 5

não satisfaz satisfaz totalmente

41. Titularidade da habitação:

Marcar apenas uma oval.

- própria
 arrendada
 de família

42. Distância de casa ao trabalho (número de km) :

43. Duração de casa ao trabalho (número de minutos):

44. Meio de transporte utilizado para chegar ao trabalho:

Marcar apenas uma oval.

- automóvel
 motociclo
 bicicleta
 transportes públicos
 outros

45. Duração de residência na habitação atual (número de anos):

VIZINHANÇA

46. Existência de vizinhos amigos:

Marcar apenas uma oval.

sim

não

47. Existência de vizinhos neutros:

Marcar apenas uma oval.

sim

não

48. Existência de vizinhos problemáticos:

Marcar apenas uma oval.

sim

não

CRÉDITOS

Informação relacionada com encargos financeiros suportados (e.g. responsabilidades bancárias)

49. Existência de crédito à habitação:

Marcar apenas uma oval.

sim

não

50. Existência de crédito automóvel:

Marcar apenas uma oval.

sim

não

51. Existência de outros créditos (e.g. pessoais):

Marcar apenas uma oval.

sim

não

FAMÍLIA E AMIGOS

Dados relacionados com suporte social dado e recebido

52. Autoapreciação do suporte social recebido:

Marcar apenas uma oval.

	0	1	2	3	4	5	
nenhum	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	excelente

53. Autoapreciação do suporte social dado:

Marcar apenas uma oval.

	0	1	2	3	4	5	
nenhum	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	excelente

54. Autoapreciação geral de saúde dos elementos do domínio privado (e.g. cônjuge e filhos):

Marcar apenas uma oval.

	0	1	2	3	4	5	
nada boa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	muitíssimo boa

VIDA PROFISSIONAL

55. Vontade de mudar de profissão (atividade profissional desempenhada):

Marcar apenas uma oval.

	0	1	2	3	4	5	
nenhuma vontade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	muitíssima vontade

56. Vontade de mudar de emprego atual (empresa ou organização onde trabalha):

Marcar apenas uma oval.

	0	1	2	3	4	5	
nenhuma vontade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	muitíssima vontade

ASPETOS FÍSICOS

57. Peso:

58. Altura:

VIDA ATIVA

Informação acerca da vida profissional e desempenho de papéis na sociedade.

59. Número de anos de trabalho (todos os empregos):

60. Número de anos do emprego atual:

61. Número de anos de exercício da profissão atual:

62. Número de empregos desde o início da vida ativa:

VIDA EM SOCIEDADE

63. Filiação em partidos políticos:

Marcar apenas uma oval.

sim

não

64. Participação ativa em partidos políticos, religião ou outras associações:

Marcar apenas uma oval.

sim

não

65. Desempenho de papéis sociais de relevo em organizações sociais, políticas ou religiosas (e.g. junta de freguesia, grupo desportivo, etc.):

Marcar apenas uma oval.

sim

não

VIDA SOCIAL

Informação acerca de hábitos sociais e participação na sociedade.

66. Número aproximado de tomas diárias de café:

67. Número aproximado de cigarros consumidos por dia:

68. Consumo de álcool:

Marcar apenas uma oval.

regularmente

às refeições

com amigos

nunca

outro

69. Faz ou já fez dieta para manutenção do aspeto físico:

Marcar apenas uma oval.

sim

não

04/06/2021

QUESTIONÁRIO SOCIOCULTURAL

70. Número de idas a eventos culturais no último mês:

71. Número de dias despendidos na natureza no último mês:

Este conteúdo não foi criado nem aprovado pela Google.

Google Formulários

7.2 Questionário {WELLBEING1}

QUESTIONÁRIO DO BEM-ESTAR
{WELLBEING1} aplicado pelo {ANDROIDANALYTICS}

As perguntas que compõem o questionário foram inspiradas na escala adaptada para Portugal da medida de manifestação de bem-estar psicológico (i.e. “*Échelle de Mesure des Manifestations du Bien-Être Psychologique*”) {EMMBEP} (Sara; Monteiro, Tavares, Pereira, & Universidade de Aveiro, 2012).

PERGUNTAS

SOCIABILIDADE

» Relacionei-me facilmente com as outras pessoas. [1|2|3]

CONTROLO DE SI E DOS ACONTECIMENTOS

» Estive capaz de enfrentar situações difíceis de uma forma positiva. [1|2|3]

AUTOESTIMA

» Senti que os outros gostavam de mim e me apreciavam. [1|2|3]

» Senti-me satisfeito(a) com o que fui capaz de alcançar, senti-me orgulhoso(a) de mim próprio. [1|2|3]

EQUILÍBRIO

» A minha vida foi bem equilibrada entre as minhas atividades familiares, pessoais e académicas. [1|2|3]

RESPOSTAS [1|2|3]



1 - DISCORDO



2 - NEM DISCORDO NEM CONCORDO



3 - CONCORDO

STRESS

» Qual o período do dia em que sentiu mais stress (não preencha caso não tenha sentido stress)?

Início: ____:____ Fim: ____:____

7.3 Questionário {WELLBEING2}

QUESTIONÁRIO DO BEM-ESTAR
{WELLBEING2} aplicado pelo {ANDROIDANALYTICS}

As perguntas que compõem o questionário foram inspiradas na escala adaptada para Portugal da medida de manifestação de bem-estar psicológico (i.e. “*Échelle de Mesure des Manifestations du Bien-Être Psychologique*”) {EMMBEP} (Sara; Monteiro, Tavares, Pereira, & Universidade de Aveiro, 2012).

PERGUNTAS

FELICIDADE

» Senti-me emocionalmente equilibrado. [1|2|3]

AUTOESTIMA

» Senti-me confiante. [1|2|3]

FELICIDADE

» Senti-me bem, em paz comigo próprio. [1|2|3]

RESPOSTAS



1 - DISCORDO



2 - NEM DISCORDO NEM CONCORDO



3 - CONCORDO

7.4 Questionário {SLEEPTIRED}

QUESTIONÁRIO DO SONO E SENTIMENTO DE CANSAÇO {SLEPTIRED} aplicado pelo {ANDROIDANALYTICS}

As perguntas que compõem o questionário foram criadas pelos autores e têm como objetivo avaliar, de forma subjetiva, a autoapreciação da qualidade da noite de sono anterior e do sentimento de cansaço após o despertar.

PERGUNTAS

SONO

» Dormi muito bem e sinto que o meu sono foi totalmente reparador. [1|2|3]

CANSAÇO

» Sinto-me totalmente descansado após esta noite de sono. [1|2|3]

RESPOSTAS



1 - DISCORDO



2 - NEM DISCORDO NEM CONCORDO



3 - CONCORDO

7.5 Questionário {CALLINTERACTIONS}

QUESTIONÁRIO PARA CLASSIFICAÇÃO DAS CHAMADAS TELEFÓNICAS {CALLINTERACTIONS} aplicado pelo {ANDROIDANALYTICS}

Este documento resume em tópicos o conteúdo do formulário despoletado no final de cada chamada recebida ou efetuada pelos participantes. As opções disponibilizadas na classificação do interlocutor foram criadas tendo por base o modelo *convoy* de Antonucci que representa o conjunto de relacionamentos do ser humano através de um conjunto de círculos concêntricos. No modelo de Antonucci, o círculo mais interno representa o próprio indivíduo e os restantes três o *convoy* de relacionados, ou seja, as pessoas importantes ao nível do suporte social para o indivíduo (Kahn & Antonucci, 1980) (Antonucci, 2001).

RELATIVAMENTE À CHAMADA ACABADA DE OCORRER,
por favor, classifique:

O INTERLOCUTOR (apenas se ainda não classificado anteriormente)

- Cônjuge, filho, ou amigo próximo.
- Restantes familiares e amigos.
- Outros (e.g. colegas de trabalho, supervisores, clientes, fornecedores, vizinhos, etc.).

A QUALIDADE EMOCIONAL DA INTERAÇÃO SOCIAL



INTERAÇÃO NEGATIVA



INTERAÇÃO NEUTRA



INTERAÇÃO POSITIVA

7.6 Questionário {LABOR}

QUESTIONÁRIO LABORAL {LABOR}
dados preenchidos pelo departamento de Recursos Humanos
(ficheiro EXCEL)

Este ficheiro recolhe dados do contexto laboral, acerca das férias gozadas e não gozadas, horas extra, faltas e atrasos ao trabalho.

DADOS A PREENCHER

FALTAS AO TRABALHO DO ÚLTIMO MÊS

(arredondamento para a unidade superior)

PARTICIPANTE	DATA	TIPO	HORAS	PREVISÃO
145	2018-12-31	Greve laboral	4	Prevista com 1 dia de antecedência
		Doença (própria)		Não prevista antecipadamente
		Doença (filhos, cônjuge, etc.)		Prevista com 1 dia de antecedência
		Assistência familiar (filhos, cônjuge, etc.)		Prevista com 2 dias de antecedência
		Greve laboral		Prevista com mais de 2 dias de antecedência
		Outro		

ATRASOS AO TRABALHO DO ÚLTIMO MÊS

(arredondamento para a unidade superior)

PARTICIPANTE	DATA	PERÍODO	ATRASO	PREVISÃO	MOTIVO
145	2018-12-31	Primeiro período de trabalho	20 minutos	Previsto com menos de 2 horas	Imprevisto familiar
		Segundo período de trabalho		Não previsto antecipadamente	Dificuldades em acordar
				Previsto com menos de 2 horas	Imprevisto familiar
				Previsto com menos de 1 dia de antecedência	Trânsito
				Previsto com mais de 1 dia de antecedência	Outro

FÉRIAS GOZADAS NOS ÚLTIMOS 12 MESES

(arredondamento para a unidade superior)

PARTICIPANTE	INÍCIO	FIM	TIPO	PREVISÃO
145	2018-12-31	2018-12-31	Atrasadas (anos anteriores ou próprio ano)	Gozo não previsto com antecedência
			Atrasadas (anos anteriores ou próprio ano)	Gozo não previsto com antecedência
			Próprio ano	Gozo previsto há menos de 1 dia
			Adiantadas (ano seguinte)	Gozo previsto há menos de 1 semana
				Gozo previsto há menos de 1 mês
				Gozo previsto há mais de 1 mês

FÉRIAS NÃO GOZADAS NOS ÚLTIMOS 12 MESES

(arredondamento para a unidade superior)

PARTICIPANTE	INÍCIO	FIM	TIPO	PREVISÃO
145	2018-12-31	2018-12-31	Atrasadas (anos anteriores ou próprio ano)	Não gozo não previsto com antecedência
			Atrasadas (anos anteriores ou próprio ano)	Não gozo não previsto com antecedência
			Próprio ano	Não gozo previsto há menos de 1 dia
			Adiantadas (ano seguinte)	Não gozo previsto há menos de 1 semana
				Não gozo previsto há menos de 1 mês
				Não gozo previsto há mais de 1 mês

TRABALHO SUPLEMENTAR DO ÚLTIMO MÊS

(arredondamento para a unidade superior)

PARTICIPANTE	DATA	HORAS	PREVISÃO
145	2018-12-31	5 ou mais	Previsto com 1 dia de antecedência
		1	Não previsto antecipadamente
		2	Previsto com 1 dia de antecedência
		3	Previsto com 2 dias de antecedência
		4	Previsto com mais de 2 dias de antecedência
		5 ou mais	

7.7 Ficha de participante {SUBJECTFILE}

EMOTION-AWARENESS IN OFFICE EMPLOYEES



FICHA DE PARTICIPAÇÃO EM EXPERIÊNCIA {SUBJECTFILE} (v1.0)

Este registo contém os dados gerais de suporte à experiência para cada participante. Este registo deve manter-se pseudonimizado.

ELEGIBILIDADE

NOTA PRÉVIA INICIAL RELATIVA À ANGARIAÇÃO POR AUTO PROPOSIÇÃO

A angariação de participantes para esta experiência, baseia-se no modelo de auto proposição. Pretende-se evitar a colocação de perguntas que podem ter respostas constrangedoras. Não se auto proponha como participante caso tenha estado exposto a algum evento emocionalmente marcante nos últimos 6 meses (e.g. divórcio, morte de um familiar próximo ou de um amigo querido, a perda de uma habitação, o surgimento de uma doença grave em alguém da família próxima, etc.). Caso já tenha exprimido vontade de participar nesta experiência pode mudar de opinião. Apesar de ser muito importante e agradeceremos a sua participação, ninguém lhe vai perguntar a razão da sua desistência. Obrigado pela sua participação.

EXPERIMENTO

ID ORGANIZAÇÃO	ID PARTICIPANTE	DIA 1	DIA 2	DIA 3	DIA 4

DADOS GERAIS

GÉNERO <input type="radio"/>	<input type="radio"/> Masculino <input type="radio"/> Feminino <input type="radio"/> Outro
IDADE (anos)	
PESO (kg)	
ALTURA (cm)	

{WINDOWSACTIVITY}

WINDOWS GROUP POLICY SETTINGS	VALOR A REPOR	VALOR TEMPORÁRIO
[Configuração do Computador > Definições do Windows > Definições de segurança > Configuração políticas auditoria avançada > Políticas de Auditoria de Sistema > Controlo Detalhado]		
Auditar a criação de processos	Não configurado	Êxito
Auditar a terminação de processos	Não configurado	Êxito
[Configuração do Computador > Definições do Windows > Definições de segurança > Configuração políticas auditoria avançada > Políticas de Auditoria de Sistema > Início de Sessão/Fim de Sessão]		
Auditar início de sessão	Não configurado	Êxito e Falha
Auditar fim de sessão	Não configurado	Êxito
Auditar outros eventos de Início de sessão/Fim de sessão...	Não configurado	Êxito
Auditar Início de Sessão Especial	Não configurado	Sem auditoria
[Configuração do Computador > Definições do Windows > Definições de segurança > Configuração políticas auditoria avançada > Políticas de Auditoria de Sistema > Início de Sessão de Conta]		
Auditar Validação de Credenciais	Não configurado	Êxito e Falha
Auditar Serviço de Autenticação Kerberos	Não configurado	Êxito e Falha

EMOTION-AWARENESS IN OFFICE EMPLOYEES

FICHA DE PARTICIPAÇÃO EM EXPERIÊNCIA {SUBJECTFILE} (v1.0)



{MIBAND}

DADOS GERAIS	
ID (mac address)	
FIRMWARE	
COLOCAÇÃO ☉	<input type="radio"/> Pulso esquerdo <input type="radio"/> Pulso direito
PASSOS	
STEP LENGTH (cm)	
FREQUÊNCIA CARDÍACA	
BATIMENTO CARDÍACO	
SLEEP ASSISTANCE	

{EDAMOVE}

DADOS GERAIS	
ID (serial number)	
FIRMWARE	1.12.0
COLOCAÇÃO ☉	<input type="radio"/> Pulso esquerdo <input type="radio"/> Pulso direito
CONFIGURAÇÕES DE RECOLHA	
PRESS (Hz)	8 Hz
TEMP (Hz)	1 Hz
TEMP_MEAN (Hz)	1/60Hz
ACC (Hz)	64 Hz
MOVEMENT_ACC (Hz)	1/60Hz
STEPS (Hz)	1/60Hz
EDA (Hz)	32 Hz
EDA_SCL_MEAN (Hz)	1/60Hz

7.8 *Script* para recolha de dados {COLLECTIONSCRIPT}

EMOTION-AWARENESS IN OFFICE EMPLOYEES



SCRIPT DE SPRINT PARA RECOLHA DE DADOS DE CONTEXTO

Este documento descreve a sequência de tarefas a executar num ciclo (*sprint*) de recolha de dados de contexto. Cada *sprint* inicia com a preparação do processo de recolha (dia anterior ao início de recolha) e termina com a extração dos dados guardados pelos instrumentos (e.g. sensores, formulários, etc.) e respetivo arquivo para posterior processamento (dia imediatamente a seguir ao fim do período de recolha).

SEMANA ANTERIOR AO DA RECOLHA

1. ANGARIAÇÃO DE PARTICIPANTES

- 1.1. Ato público de apresentação do projeto e angariação de candidatos
- 1.1.1 Apresentação do projeto e entrega da brochura-----
- 1.1.2 Explicar modelo de auto proposição pelos candidatos -----
- 1.1.3 Explicar como os dados pessoais serão pseudonimizados e posteriormente anonimizados-----
- 1.1.4 Apresentar o processo de recolha de dados -----

DIA ANTERIOR AO INICIO DA RECOLHA (PREPARAÇÃO DA RECOLHA)

2. {SUBJECTFILE}

- 2.1. Atribuição de códigos aos participantes (pseudonimização) -----
- 2.2. Pedido de preenchimento do formulário de caracterização de cada participante {SUBJECTFILE}-----

3. RECOLHA DA DOCUMENTAÇÃO

- 3.1. Documento de aceitação de participação em experimento pela organização (Anexo III)-----
- 3.2. Consentimentos assinados pelos participantes-----

4. {WINDOWSACTIVITY} <https://tinyurl.com/WindowsActivity>

- 4.1. Definir nova senha de acesso ao SO (2 maiúsculas + 2 minúsculas + 2 números + 3 símbolos)-----
- 4.2. Configurar o bloqueio automático de ecrã para 5 minutos de inatividade-----
- 4.3. Programar *Group Policies* (*gpedit.msc*) necessárias para extração de dados -----
- 4.4. Testar o correto funcionamento do {WINDOWSACTIVITY}-----

5. {ANDROIDANALYTICS} <https://tinyurl.com/AndroidAnalytics>

- 5.1. Instalar e testar o funcionamento da APP -----

6. {MIBAND}

- 6.1. Recarregar bateria -----
- 6.2. Configurações gerais do NOTIFY & FITNESS FOR MI BAND (OneZeroBit) {NOTIFY&FITNESS}
- 6.2.1 Definir o pulso a utilizar (cf. esquerdo ou direito) -----
- 6.3. Passos
- 6.3.1 Executar a opção “Reiniciar passos” -----
- 6.3.2 Definir o “*Step length*” 74.52 -----
- 6.4. Sono (dados de...)
- 6.4.1 Executar “Eliminar dados do sono” -----
- 6.4.2 Configurar “*Sleep as Android*” em Definições / batimento cardíaco a cada 2 minutos -----
- 6.4.3 Ativar a “Monitorização cardíaca” para garantir uma melhor estatística de sono -----
- 6.5. Frequência cardíaca
- 6.5.1 Executar “Apagar dados antigos” -----
- 6.5.2 Executar “Apagar” -----
- 6.5.3 Configurar “Batimento Cardíaco” para “1 minuto” -----

7. {EDAMOVE}

- 7.1. Recarregar bateria -----
- 7.2. Reset aos sensores-----
- 7.3. Definir novo período de recolha (data e hora de início de fim) ^{*1)}-----

EMOTION-AWARENESS IN OFFICE EMPLOYEES



SCRIPT DE SPRINT PARA RECOLHA DE DADOS DE CONTEXTO

INÍCIO DA RECOLHA

8. Entrega e explicação de funcionamento dos instrumentos a utilizar

8.1. {MIBAND} (PHYSY, SLEEP, HR)

- 8.1.1 Identificar os dados recolhidos e a importância que têm para o experimento-----
- 8.1.2 Utilizar sempre, mesmo durante o banho e durante os períodos de sono-----
- 8.1.3 Não trocar de pulso-----
- 8.1.4 A pulseira tem que ser utilizada justa ao pulso (i.e. não muito apertada nem muito solta)-----
- 8.1.5 Deve ser colocada às 09:00 no 1º dia e retirada do pulso no último dia às 17:30-----

8.2. {EDAMOVE} (EDA, TEMP, ACC)

- 8.2.1 Identificar os dados recolhidos e a importância que têm para o experimento-----
- 8.2.2 Utilizar sempre incluindo durante o sono-----
- 8.2.3 ✘ Nunca utilizar o sensor em ambientes com humidade (e.g. banho)-----
- 8.2.4 ✘ O contacto dos elétrodos com a humidade dificulta a aderência dos adesivos à pele-----
- 8.2.5 Demonstrar a colocação dos elétrodos para quando for necessário reafixar (e.g. banho)-----
- 8.2.6 ■ Piscar a vermelho 1x/2seg = sensor ativo, aguarda início período de recolha-----
- 8.2.7 ■ Piscar a vermelho 1x/1seg = sensor ativo e a registar dados recolhidos-----
- 8.2.8 ■ Piscar a vermelho 3x/1seg = sensor em erro, necessário fazer manutenção-----
- 8.2.9 ■ Piscar azul rapidamente = bateria precisa de ser recarregada-----
- 8.2.10 ■ Nenhum led ligado = sensor desligado (modo que assume após o período de recolha)-----
- 8.2.11 ⓘ Deve ser colocada às 09:00 no 1º dia e retirada no último dia às 17:30-----

8.3. {ANDROIDANALYTICS}

- 8.3.1 O GoogleForms por vezes utiliza a vírgula ou ponto como separador decimal-----
- 8.3.2 Identificar os dados recolhidos e a importância que têm para o experimento-----
- 8.3.3 Recolha de dados acerca das interações sociais (interlocutores e qualidade de interação)-----
- 8.3.4 Importância do preenchimento do questionário de bem-estar {WELLBEING1}-----

8.4. {SLEEPTIRED}

- 8.4.1 Se não pedir o formulário, colocar a hora de recolha 2 minutos no futuro-----
- 8.4.2 Apresentar o questionário e solicitar o preenchimento a cada manhã-----
- 8.4.3 Configurar naquele instante um lembre recorrente para lembrar preenchimento-----

8.5. {LABOR}

- 8.5.1 Apresentar o questionário, identificar a importância dos dados recolhidos-----
- 8.5.2 Solicitar o preenchimento até ao último dia de recolha-----
- 8.5.3 Os dados acerca de faltas, trabalho suplementar e atrasos são relativos ao último mês-----
- 8.5.4 Os dados relativos às férias gozadas e não gozadas são do último ano-----

8.6. {SOCIOCULTURAL} <https://tinyurl.com/SocioCultural2019>

- 8.6.1 Apresentar o questionário, identificar a importância dos dados recolhidos-----
- 8.6.2 Demonstrar o preenchimento com organização 000 e participante 0000-----
- 8.6.3 Solicitar o preenchimento até às 24:00-----

EMOTION-AWARENESS IN OFFICE EMPLOYEES

SCRIPT DE SPRINT PARA RECOLHA DE DADOS DE CONTEXTO



PRIMEIRO DIA DE RECOLHA

9. APOIO AO ARRANQUE

9.1. {MIBAND} e {EDAMOVE}

9.1.1 Verificar a correta colocação -----

9.1.2 Validar os indicadores luminosos para garantir que a recolha está a ser efetuada -----

9.2. {SOCIOCULTURAL}

9.2.1 Verificar se foi preenchido por todos os participantes -----

9.3. {ANDROIDANALYTICS}

9.3.1 Verificar o correto funcionamento da APP -----

9.3.2 Esclarecer as dúvidas relativamente ao tipo de interlocutor -----

9.4. {SLEEPTIRED}

9.4.1 Acompanhar o preenchimento do primeiro SLEEPTIRED -----

9.5. {LABOR}

9.5.1 Validar se já se procedeu ao início do preenchimento / reforçar o pedido -----

9.6. OUTROS

9.6.1 Esclarecimento das dúvidas que existam -----

9.6.2 Lembrar a necessidade de preenchimento sério e consciente dos questionários -----

EMOTION-AWARENESS IN OFFICE EMPLOYEES

SCRIPT DE SPRINT PARA RECOLHA DE DADOS DE CONTEXTO



ÚLTIMO DIA DE RECOLHA (NA ORGANIZAÇÃO)

10. FIM DO PERÍODO DE RECOLHA

- 10.1. Agradecimento
- 10.2. {WHOQLBREF}
- 10.2.1. Pedido de preenchimento do questionário -----
- 10.3. {WINDOWSACTIVITY}
- 10.3.1. Exportar dados ficheiro (000_0000_WindowsActivity.sqlite) -----
- 10.3.2. Repor as configurações originais das *Group Policies* (gpedit.msc) -----
- 10.3.3. Desinstalar a aplicação -----
- 10.4. {ANDROIDANALYTICS}
- 10.4.1. Exportar dados ficheiro (000_0000_AndroidSocial.sqlite) -----
- 10.4.2. Desinstalar a aplicação -----
- 10.5. {MIBAND} e {EDAMOVE}
- 10.5.1. Recolher pulseiras, sensores, gel e anéis adesivos não utilizados, etc. -----
- 10.6. {LABOR}
- 10.6.1. Recolher 1 ficheiro com dados de todos os participantes -----

ÚLTIMO DIA DE RECOLHA (EM LABORATÓRIO)

- 10.7. {MIBAND}
- 10.7.1. Exportação de dados do {NOTIFY&FITNESS} (aplicar sempre o filtro entre datas) *2) -----
- 10.7.2. Passos (PHYSI)
- 10.7.2.1. Executar "Exportar dados" (000_0000_MiBand_steps.csv) -----
- 10.7.3. Sono (SLEEP)
- 10.7.3.1. Executar "Exportar dados" (000_0000_MiBand_sleep.csv) -----
- 10.7.3.2. Executar "Exportar dados detalhados" (000_0000_MiBand_sleepDetails.csv) -----
- 10.7.3.3. Executar "Export raw data" (000_0000_MiBand_sleepRaw.csv) -----
- 10.7.4. Frequência cardíaca (HR)
- 10.7.4.1. Executar "Exportar dados folha de cálculo" (000_0000_MiBand_hrFolhaCalculo.csv) -----
- 10.7.4.2. Executar "Exportar dados" (000_0000_MiBand_hr.csv) -----
- 10.7.5. Calorias (...)
- 10.7.5.1. Executar "Exportar dados" (000_0000_MiBand_calories.csv) -----
- 10.8. {EDAMOVE}
- 10.8.1. Exportação de dados utilizando o "SensorManager" da movisens -----
- 10.8.1.1. Executar o "Save data" no formato Unisens (000_0000_EDAMove-Unisens*.*) -----
- 10.8.1.2. Executar o "Save data" no formato CSV (000_0000_EDAMove-CSV*.*) -----

OPORTUNAMENTE

11. FIM DO SPRINT DE RECOLHA

- 11.1. MiBand3 para oferta -----

*1) A calendarização tem obrigatoriamente de ser definida no máximo até dois dias e apenas permite recolha até 5 dias sem recarga (desligar o bluetooth) (definir período de recolha com início na sexta-feira 9:30 e recolher durante 3 dias e 8 horas).

*2) A aplicação nem sempre apaga os dados anteriores pelo que tem que se garantir que se considera apenas os dados pertinentes para *sprint* de recolha.

7.9 Convite às organizações para participação em experiência

Investigadores: Célio Carvalho, José Torres e Rui Moreira
Universidade Fernando Pessoa (Porto)

Exm^{os} Senhores
Gerentes/Administradores da ORGANIZAÇÃO XXXXXX

Assunto: Convite à participação em experiência científica.

No âmbito da investigação "*Emotion-awareness in office employees*" pretende-se conduzir uma experiência em ambiente de trabalho. Para tal objetivo, precisamos da autorização da V/ para a recolha de dados de contexto dos trabalhadores de escritório.

Apresenta-se, de seguida, o projeto em desenvolvimento.

O *Affective Computing* (AC) é uma área de investigação centrada na deteção de emoções das pessoas com base nas expressões facial, oral e gestual, e noutros sinais fisiológicos do corpo humano. No entanto, acreditamos que a correlação de variáveis sociais, culturais e religiosas com outros já explorados (e.g. sinais fisiológicos), poderá contribuir para melhorar a qualidade dos resultados já obtidos na deteção automática de emoções. Esta informação emocional poderá permitir que no futuro sejam criados ambientes de trabalho mais inteligentes através da sua reconfiguração automática. Esta inteligência afetiva poderá promover uma melhoria do bem-estar e qualidade de vida das pessoas no seu trabalho, rumo a uma humanização dos sistemas e a um melhor relacionamento com a tecnologia.

É neste contexto que contactamos a V/ empresa, solicitando autorização para recolher dados de contexto dos trabalhadores de escritório durante 7 dias. Pretendemos inicialmente realizar uma sessão pública de apresentação da experiência aos trabalhadores, convidando-os à participação voluntária numa recolha de dados previamente por si consentida.

A participação dos trabalhadores é voluntária e será utilizado o método da autoexclusão para evitar respostas constrangedoras no processo de triagem. Assim, na sessão pública com os funcionários, serão enumerados e justificados os critérios de exclusão com o objetivo que apenas se auto proponham as pessoas elegíveis à participação na experiência (e.g. morte recente de algum familiar próximo, divórcio recente, etc.).

A V/ participação pressupõe a aceitação: i) do uso de sensores não obstrutivos pelos participantes durante o trabalho (e.g. pulseiras); ii) da recolha de dados acerca do fluxo de trabalho (e.g. quantidade diária de *logon* e *logoff* do sistema operativo); iii) da recolha de dados estatísticos acerca das chamadas e SMS ocorridas nos *smartphones* (quantidade de chamadas efetuadas, recebidas e perdidas, quantidade mensagens recebidas e enviadas, etc.); iv) da autorização para que os participantes possam despende de alguns segundos do seu dia no preenchimento de questionários curtos (e.g. qualidade de uma interação social no final de uma

chamada telefónica, classificação subjetiva momentânea do bem-estar numa escala de *likert*, etc.); e v) da recolha de informação laboral junto dos recursos humanos (e.g. férias marcadas, último período gozado, faltas, atrasos, etc.).

Apenas serão recolhidos os dados necessários, adequados e pertinentes à experiência, respeitando o princípio da minimização da recolha. Os dados serão apenas utilizados no contexto da experiência já apresentada respeitando o princípio da transparência. Todos os dados recolhidos dos participantes serão mantidos anónimos e confidenciais mesmo em relação à entidade patronal. Os dados das organizações serão também mantidos anónimos.

Este estudo foi aprovado pela Comissão de Ética (CE) da Universidade Fernando Pessoa (UFP). Só serão recolhidos dados após: i) aceitação formal da organização para participação na experiência; e ii) o consentimento escrito, livre e informado de cada participante. Os dados recolhidos nunca serão vendidos ou partilhados com terceiros e ficarão apenas guardados em recursos que estejam sob controlo dos investigadores. No final do estudo, todos os dados pessoais serão apagados.

Os resultados do estudo serão divulgados através de documentos científicos públicos. Será sempre preservada a privacidade dos participantes, mesmo nos assuntos que possam interessar, direta ou indiretamente, à direção ou chefias da organização

Assim, solicitamos o agendamento de uma reunião (cerca de 30 minutos), para: i) o esclarecimento de dúvidas que possam existir acerca da experiência; ii) a apresentação da brochura do projeto a divulgar junto dos trabalhadores ainda antes da sessão pública de esclarecimento para angariação de participantes; iii) assinatura da declaração de aceitação à participação por parte da organização; e iv) agendamento da sessão pública de angariação de participantes.

Agradecemos desde já a vossa disponibilidade para a participação neste projeto.

Com os melhores cumprimentos,

Os Investigadores

7.10 Sugestão de aceitação das organizações

Exm^{as} Senhores

Investigadores: Célio Carvalho, José Torres e Rui Moreira

Universidade Fernando Pessoa (Porto)

Assunto: Resposta ao V/ convite à participação em experimento científico.

Vimos pela presente comunicar que aceitamos o convite que nos foi dirigido por V.(s) Exa.(s), para participarmos na V/ experiência científica. Autorizamos a recolha de dados de contexto dos trabalhadores de escritório, desde que estes tenham sido previamente informados dos dados que irão ser recolhidos e que tenham consentido essa recolha por escrito.

Confirmamos também que não nos opomos ao preenchimento de questionários breves e à utilização de sensores pelos participantes durante o seu horário laboral. Autorizamos a recolha e tratamento de dados estatísticos dos dispositivos de trabalho através da instalação de aplicações nos computadores e *smartphone* da empresa. Relativamente aos dados relacionados com a atividade laboral de cada participante solicito o envio dos formulários respetivos para que sejam devidamente preenchidos.

Apreciamos e partilhamos a preocupação demonstrada com os dados pessoais. Realçamos a vontade que sejam respeitados todos os valores e princípios defendidos pelo Regulamento Geral de Proteção de Dados (RGPD), e entendemos os motivos subjacentes à necessidade de não nos ser possível aceder aos dados individuais recolhidos por se tratar de informação pessoal de cada trabalhador.

Estamos ao V/ dispor para agendar a sessão pública de apresentação e angariação de participantes através da auto proposição.

Com os melhores cumprimentos,

A Gerência / Administração

7.11 Brochura de apresentação da experiência

PRIVACIDADE

A privacidade é um tema de grande importância para os autores.

Apenas recolheremos os dados necessários, adequados e pertinentes à experiência que se pretende conduzir, respeitando o princípio da minimização da recolha.

No momento de angariação de participantes, todos serão informados acerca dos dados que serão recolhidos e a respetiva justificação para o tratamento, respeitando o princípio da transparência.

Os dados não serão utilizados para outros fins que não os previamente comunicados aos respetivos titulares, respeitando o princípio da lealdade e da confiança depositada pelos participantes na investigação e nos investigadores.

Os dados recolhidos nunca serão vendidos ou partilhados com terceiros e ficarão apenas guardados em recursos que estejam sob controlo dos investigadores. No final da experiência, todos os dados pessoais serão apagados.

ADMISSÃO DE PARTICIPANTES

Convidamos todos os trabalhadores de escritório a participar nesta experiência. Será efetuado um ato público onde serão apresentados todos os pormenores da experiência a realizar. O modelo de auto-proposição será adotado para evitar a necessidade de resposta a perguntas que possam ser consideradas constrangedoras e que respeitem os pressupostos da admissão.

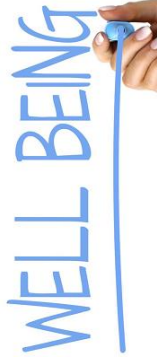
PRESSUPOSTOS DA ADMISSÃO

- i) Utilização de sensores (e.g. pulseira);
- ii) Autorização para recolha de dados estatísticos de fluxo de trabalho na empresa, quantidades de chamadas e SMS do *smartphone* pessoal e da empresa;
- iii) Disponibilidade para preencher questionários (dados socioculturais, qualidade de vida, bem-estar, etc.);
- iv) Inexistência de deficiências motoras porque vai ser recolhido o número de passos diário;
- v) Inexistência de doenças da memória e funções cognitivas (e.g. Alzheimer, etc.);
- vi) Inexistência de eventos marcantes há menos de 6 meses (e.g. morte de alguém chegado, divórcio, problemas com a família, ou problemas graves com a justiça ou financeiros).

Investigadores: Célio Carvalho,
José Torres e Rui Moreira



UNIVERSIDADE
FERNANDO
PESSOA



EMOTIONAL-AWARENESS IN OFFICE EMPLOYEES

As organizações são cada vez mais dinâmicas e exigem cada vez mais dos seus trabalhadores, nomeadamente dos que trabalham no escritório. O *stress* relacionado com o cumprimento de prazos faz com que muitas vezes o local de trabalho seja um ambiente agressivo onde se experienciam emoções negativas. No entanto, a realização pessoal, resultante do reconhecimento, bom espírito de equipa, cumprimento de objetivos poderá ajudar no contra balanceamento emocional com emoções positivas.

*Sessão de esclarecimento à
participação voluntária na
investigação científica*

CONVITE PARTICIPAÇÃO EM EXPERIÊNCIA

A Organização Mundial de Saúde (OMS) define “saúde” como sendo o estado de completo bem-estar físico, social e mental. Olhando para o estado emocional como uma forma de avaliar o equilíbrio triangular da saúde defendida pela OMS, neste projeto considera-se que uma deteção de emoções mais eficiente, poderá promover uma melhoria nos processos de avaliação do bem-estar dos trabalhadores sujeitos ao *pressing* administrativo dos escritórios.

Acreditamos que a deteção de emoções, além de se basear na análise de sinais fisiológicos, deve também atender aos dados de contexto social, cultural e até religioso.

É neste contexto que gostaríamos de vos convidar a participar nesta experiência, permitindo-nos recolher dados fisiológicos, sociais, culturais e religiosos, durante o período de 7 dias.

A participação nesta experiência é livre, voluntária e não remunerada. No final da experiência, será oferecida a pulseira Xiaomi Mi Band.

AFFECTIVE COMPUTING (AC)

O *Affective Computing* (AC) é uma área de investigação centrada na deteção de emoções com base nas expressões facial, oral e gestual, e noutros sinais fisiológicos. Acreditamos que a correlação de variáveis sociais, culturais e religiosas com sinais fisiológicos do corpo humano, poderá contribuir para melhorar a qualidade dos resultados já obtidos na deteção automática de emoções.

Estamos convictos ser possível a criação de sistemas conscientes do ponto de vista emocional, capazes de monitorizar as emoções através da avaliação física, social, cultural e religiosa de trabalhadores em ambiente de escritório com o objetivo de melhorar o seu ambiente de trabalho e bem-estar.

Esta informação emocional poderá permitir que sejam criados ambientes de trabalho mais inteligentes, promovendo uma melhoria do bem-estar e qualidade de vida das pessoas no seu trabalho, rumo a uma humanização dos sistemas e a um melhor relacionamento com a tecnologia.

QUE DADOS SERÃO RECOLHIDOS?

Sensores não obstrutivos nem intrusivos:

- i) **EDA**, sensor que mede atividade eletrodermal através de eléctrodos colocados sobre a pele (Movisens EDAMove); e
- ii) **HR** e **HRV**, sensor que mede atividade física e cardíaca através de uma pulseira (Xiaomi Mi Band).

Questionários:

- i) **Dados socioculturais** (e.g. estado civil, etnia, habitação, vizinhança, existência de créditos [S/N], etc.);
- ii) **Dados laborais** (férias e faltas);
- iii) **Sono** (e.g. nível de cansaço matinal);
- iv) **Interações sociais** (e.g. chamada telefónica positiva ou negativa, etc.);
- v) **Qualidade de vida** (e.g. satisfação com a própria saúde, etc.);
- vi) **Bem-estar**, (e.g. autoapreciação do nível de bem-estar, sentimento de confiança, etc.).

CONFIDENCIALIDADE

Os dados pessoais recolhidos serão mantidos sob confidencialidade mesmo em relação à entidade patronal. Este estudo foi aprovado pela Comissão de Ética (CE) da Universidade Fernando Pessoa (UFP). So serão recolhidos dados após:

- i) Aceitação formal da organização para participação na experiência; e
- ii) Consentimento escrito, livre e informado de cada participante.

7.12 Declaração de consentimento dos participantes

DECLARAÇÃO DE CONSENTIMENTO INFORMADO

Emotion-awareness in office employees:

Eu, abaixo-assinado, _____,
compreendi a explicação que me foi fornecida acerca da participação na investigação que se
tenciona realizar, bem como do estudo em que serei incluído. Foi-me dada oportunidade de
fazer as perguntas que julguei necessárias, e de todas obtive resposta satisfatória.

Tomei conhecimento de que a informação ou explicação que me foi prestada versou os objetivos
e os métodos. Além disso, foi-me afirmado que tenho o direito de recusar a todo o tempo a
minha participação no estudo, sem que isso possa ter como efeito qualquer prejuízo pessoal.

Foi-me ainda assegurado que os registos em suporte papel e/ou digital (sonoro e de imagem)
serão confidenciais e utilizados única e exclusivamente para o estudo em causa, sendo
guardados em local seguro durante a pesquisa e destruídos após a sua conclusão.

Por isso, consinto em participar no estudo em causa.

Data: ____/_____/20__

Assinatura do participante no projeto: _____

O Investigador responsável:

Nome: Célio Carvalho

Assinatura:

Comissão de Ética da Universidade Fernando Pessoa

7.13 Autorização da Comissão de Ética



Universidade Fernando Pessoa
www.ufp.pt

Exmo. Senhor
Prof. Doutor Feliz Gouveia
Diretor da FCT

Porto, 13 de Fevereiro de 2019


Exmo. Senhor Prof. Doutor,

A Comissão de Ética, depois de reapreciado o projeto de investigação de Cêlio Domingos de Faria Carvalho, intitulado "Emotion-awareness in office employees", e depois de analisada a resposta ao nosso primeiro parecer que nos foi enviada pelos investigadores responsáveis, decidiu aprovar o projeto tendo em consideração os seguintes aspetos:

- a) Os investigadores comprometem-se a garantir a liberdade de participação dos funcionários das empresas que serão contactadas;
- b) Os investigadores comprometem-se a garantir que o consentimento será de facto livre, informado e esclarecido, e que os participantes podem desistir da sua participação, sem qualquer prejuízo para os mesmos;
- c) Os investigadores comprometem-se a analisarem os dados em termos estatísticos e não ao nível do conteúdo;
- d) Os investigadores comprometem-se a não fornecerem às empresas dados que estas eventualmente peçam sobre os seus funcionários, garantindo que as empresas terão acesso apenas aos resultados do estudo na sua globalidade;
- e) Os investigadores terão de garantir que o levantamento de dados é realizado apenas em contexto laboral, assegurando que os sensores serão usados apenas enquanto os participantes trabalham. Este aspeto é relevante, dado que o objetivo do estudo é assegurar um contexto de trabalho de maior qualidade para os funcionários, não havendo portanto qualquer razão para os sensores serem usados no domicílio ou fora do contexto laboral.

Com os melhores cumprimentos.

A Presidente da
Comissão de Ética da UFP


Susana Teixeira Magalhães



Fundação Ensino e Cultura "Fernando Pessoa"

MFC 502 051 602 - Reg. Comercial n.º 26 Conservatória do Registo Comercial do Porto

REITORIA - [Faculdade de Ciências Humanas e Sociais] - [Faculdade de Ciências e Tecnologia] Praça 9 de Abril, 349 - 4249-004 Porto-Portugal - T +351 22 207 1300 - F +351 22 520 8269 - geral@ufp.pt

[Faculdade de Ciências da Saúde] - [Escola Superior de Saúde] R. Carlos da Maia, 216 - 4200-150 Porto - Portugal - T +351 22 507 4630 - F +351 22 507 4631 - R. Delfim Maia, 314 - 4200-253 Porto - Portugal

T +351 22 509 6371 - geral.asaude@ufp.pt UNIDADE de Ponte de Lima - Casa da Garrida - R. Conde de Berrandós - 4990-078 Ponte de Lima - Portugal - T +351 258 741 026 - F +351 258 741 412 - geral@ufp.pt

7.14 Autorização para utilização do {EMMBEP}

RE: EMMBEP - Pedido de autorização para utilização

Sara Monteiro <smonteiro@ua.pt>

ter 20-11-2018 13:13

Para: Célio Carvalho <cdcarvalho@ipca.pt>; José Tavares <jtav@ua.pt>; Anabela Pereira <anabelapereira@ua.pt>;

Cc: Jose Manuel Torres <jtorres@ufp.edu.pt>; Rui Silva Moreira <rmoreira@ufp.edu.pt>;

Caro Dr. Célio Carvalho:

Conforme solicitado, dou a minha permissão para utilizarem a Escala de Medida de Manifestação de Bem-Estar Psicológico no vosso estudo.

Indico também referências importantes a consultar, com destaque para a última:

- Massé, R., Poulin, C., Dassa, C., Lambert, J., Bélair, S., & Battaglini, A. (1998). Élaboration et validation d'un outil de mesure du bien-être psychologique: L'É.M.M.B.E.P. *Revue Canadienne de Santé Publique*, 89(5), 352-357.
- Monteiro, S. (2008). Optimismo e vinculação na transição para o ensino superior: Relação com sintomatologia psicopatológica, bem-estar e rendimento académico. Tese de doutoramento, Universidade de Aveiro, Aveiro.
- Monteiro, S., Tavares, J., & Pereira, A. (2006). Estudo das características psicométricas da escala de medida de manifestação de bem-estar. In I. Leal, J. L. Ribeiro e S. N. Jesus (Eds.), *Actas do 6.º Congresso Nacional de Psicologia da Saúde, Bem-Estar e Qualidade de Vida* (pp. 53-58). Lisboa: ISPA Edições.
- Monteiro, S., Tavares, J., & Pereira, A. (2012). Adaptação portuguesa da escala de medida de manifestação de bem-estar psicológico com estudantes universitários – EMMBEP. *Psicologia, Saúde & Doenças*, 13(1), 66-77.

Estou disponível para outras informações que necessite.

Com os melhores cumprimentos,
Sara Monteiro

Sara Otília Marques Monteiro
Prof. Auxiliar Convidada / Assistant Professor
Psicóloga / Psychologist (Céd. OPP nº4830)
Especialista em Psicologia Clínica e da Saúde / Especialista em Psicologia da Educação
Univ. de Aveiro / Univ. of Aveiro.
Dep. de Educação e Psicologia / Dep. of Education and Psychology
Campus Universitário de Santiago. 3810-193 Aveiro, Portugal.
Tel. / Phone: (+351) 234 370 353

De: Célio Carvalho [cdcarvalho@ipca.pt]
Enviado: quinta-feira, 15 de Novembro de 2018 20:04
Para: Sara Monteiro; José Tavares; Anabela Pereira
Cc: Jose Manuel Torres; Rui Silva Moreira
Assunto: EMMBEP - Pedido de autorização para utilização

Exmo.(s) Senhor(es)
» Prof. Doutora Sara Monteiro,
» Prof. Doutor José Tavares, e

» Prof. Doutora Anabela Pereira

Sou aluno de doutoramento na Universidade Fernando Pessoa e os meus orientadores são os professores: Doutor José Torres e Doutor Rui Moreira.

Estamos a preparar uma investigação para reconhecimento automático de emoções em trabalhadores de escritório. Pretendemos recolher dados de contexto fisiológico, social e laboral, para classificar emoções utilizando algoritmos de Inteligência Artificial. No entanto, precisamos também de recolher informação relacionada com o bem-estar dos participantes do experimento. É neste contexto que solicitamos autorização para utilização da Escala de Medida de Bem-estar Psicológico (EMMBEP) (versão adaptada por V.(s) Exa.(s) para Portugal).

Assumimos o compromisso de referenciar o vosso trabalho nas publicações que efetuarmos neste projeto.

Obrigado pelo apoio.

Cumprimentos,
Célio Carvalho

IPCA - Instituto Politécnico do Cávado e do Ave
Escola Superior de Tecnologia

7.15 Autorização para utilização do {WHOQOLBREF}

Re: Fw: WHOQOL-BREF - Pedido de autorização para utilização

Adriano Vaz-Serra <adriano.vazserra@gmail.com>

sex 16-11-2018 15:25

Para:Célio Carvalho <cdcarvalho@ipca.pt>;

Exma Sra Dr Célio Carvalho,
Desde já fica autorizado a usar o instrumento **WHOQOL-Bref**.
Desejo os melhores êxitos para a sua actividade.
Aceite os melhores cumprimentos de,
Adriano Vaz Serra

Célio Carvalho <cdcarvalho@ipca.pt> escreveu no dia quinta, 15/11/2018 à(s) 19:48:

Reencaminho para outro endereço de e-mail do Prof. Doutor Adriano Vaz Serra (obtido no www.degois.pt),
atendendo a que o endereço anterior foi rejeitado.

Obrigado mais uma vez.

Cumprimentos,
Célio Carvalho

IPCA - Instituto Politécnico do Cávado e do Ave
Escola Superior de Tecnologia

De: Célio Carvalho
Enviado: 15 de novembro de 2018 19:29
Para: adrianovs@netvisao.pt; mccanavarro@fpce.uc.pt
Cc: Jose Manuel Torres; Rui Silva Moreira
Assunto: WHOQOL-BREF - Pedido de autorização para utilização

Exmo.(s) Senhor(es)
» Prof. Doutor Adriano Vaz Serra, e
» Prof. Doutora Maria Cristina Canavarro.

Sou aluno de doutoramento na Universidade Fernando Pessoa e os meus orientadores são os professores:
Doutor José Torres e Doutor Rui Moreira.

Estamos a preparar uma investigação para reconhecimento automático de emoções em trabalhadores de escritório. Pretendemos recolher dados de contexto fisiológico, social e laboral, para classificar emoções utilizando algoritmos de Inteligência Artificial. No entanto, precisamos também de recolher informação relacionada com a qualidade de vida dos participantes no experimento. É neste contexto que solicitamos a autorização para utilização do WHOQOL-BREF (versão adaptada por V.(s) Exa.(s) para Portugal).

Assumimos o compromisso de referenciar o vosso trabalho nas publicações que efetuarmos neste projeto.

Obrigado pelo apoio.

Cumprimentos,
Célio Carvalho

IPCA - Instituto Politécnico do Cávado e do Ave
Escola Superior de Tecnologia

7.16 *Backup* do questionário {WELLBEING1}

EMOTION-AWARENESS IN OFFICE EMPLOYEES

{WELLBEING1}




QUESTIONÁRIO DO BEM-ESTAR

ID ORGANIZAÇÃO	ID PARTICIPANTE	DIA

LEGENDA: [1 – discordo, 2 – nem discordo nem concordo, 3 -concordo]




SOCIABILIDADE

» Relacionei-me facilmente com as outras pessoas.

 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>




CONTROLO DE SI E DOS ACONTECIMENTOS

» Estive capaz de enfrentar situações difíceis de uma forma positiva.




 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

AUTOESTIMA

» Senti que os outros gostavam de mim e me apreciavam.




 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

» Senti-me satisfeito(a) com o que fui capaz de alcançar, senti-me orgulhoso(a) de mim próprio.

 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

EQUILÍBRIO

» A minha vida foi bem equilibrada entre as minhas atividades familiares, pessoais e académicas.

 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

STRESS

» Qual o período do dia em que sentiu mais stress (não preencha caso não tenha sentido stress)?

Hora inicial	Hora final
--------------	------------

7.17 *Backup* do questionário {WELLBEING2}

EMOTION-AWARENESS IN OFFICE EMPLOYEES

{WELLBEING2}




QUESTIONÁRIO DO BEM-ESTAR (ESPONTÂNEO)

ID ORGANIZAÇÃO	ID PARTICIPANTE	DIA

LEGENDA: [1 – discordo, 2 – nem discordo nem concordo, 3 -concordo]




EQUILÍBRIO

» Senti-me emocionalmente equilibrado.




 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

AUTOESTIMA

» Senti-me bem, em paz comigo próprio.

 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

» Senti-me confiante.

 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7.18 *Backup* do questionário {SLEEPTIRED}

EMOTION-AWARENESS IN OFFICE EMPLOYEES

{SLEEPTIRED}






QUESTIONÁRIO DO SONO E SENTIMENTO DE CANSAÇO PELA MANHÃ

ID ORGANIZAÇÃO	ID PARTICIPANTE	DIA

LEGENDA: [1 – discordo, 2 – nem discordo nem concordo, 3 -concordo]




SONO

» Dormi muito bem e sinto que o meu sono foi totalmente reparador.

 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

CANSAÇO

» Sinto-me totalmente descansado após esta noite de sono.

 1	 2	 3
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Obrigado pela sua resposta.

Os investigadores.

7.19 *Backup* do questionário {CALLINTERACTIONS}

EMOTION-AWARENESS IN OFFICE EMPLOYEES

{CALLINTERACTIONS}



Neste documento devem ser classificadas as chamadas telefónicas ocorridas.

ID ORGANIZAÇÃO	ID PARTICIPANTE	DIA

LEGENDA

(CLASSIFICAÇÃO DO INTERLOCUTOR)

 Cônjuge, filho ou amigo próximo (e.g. filhos e amigo muito próximo)	 Restantes familiares e amigos (e.g. pai, mãe, irmãos e sobrinhos)	 Outras pessoas (e.g. colegas de trabalho, clientes e vizinhos)
---	---	--

LEGENDA

(CLASSIFICAÇÃO DA QUALIDADE EMOCIONAL DA INTERAÇÃO SOCIAL)

 Interação negativa	 Interação neutra	 Interação positiva
------------------------	----------------------	------------------------

REGISTO DE INTERAÇÕES SOCIAIS

HORÁRIO LABORAL		
	INTERLOCUTOR	QUALIDADE
1	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
2	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
3	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
4	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
5	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
6	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>

HORÁRIO PRÉ OU PÓS-LABORAL		
	INTERLOCUTOR	QUALIDADE
1	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
2	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
3	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
4	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
5	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>
6	 <input type="radio"/> <input type="radio"/> <input type="radio"/>	 <input type="radio"/> <input type="radio"/> <input type="radio"/>

7.20 *Backup* do instrumento {WINDOWSACTIVITY}

{WINDOWSACTIVITY} backup
dados preenchidos pelo participante
(ficheiro EXCEL)

Este ficheiro recolhe dados das várias entradas e saídas do sistema operativo, durante os dias de recolha.

DADOS A PREENCHER

UTILIZAÇÃO DO COMPUTADOR NOS DIAS DE RECOLHA
(arredondamento ao minuto)

	Dia 1		Dia 2			Dia 3			Dia 4		
Hora	Login	Logout	Hora	Login	Logout	Hora	Login	Logout	Hora	Login	Logout
10:28	x			x			x			x	
10:52		x			x			x			x
11:07	x			x			x			x	
		x			x			x			x
	x			x			x			x	
		x			x			x			x

7.21 Dicionário de dados da base de dados do *dataset*

Atributo	Tipo	Descrição
Organizations		
PK Id	Inteiro	Código da organização.
Name	Texto	Nome pseudonimizado da organização.
Participants		
PK Id	Inteiro	Número do participante
FK OrganizationId	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
Name	Texto	Nome pseudonimizado do participante.
RawCollectionStartingDate	Data / hora	Início do período de recolha.
RawCollectionEndingDate	Data / hora	Fim do período de recolha.
MiBandId	Inteiro	Id da pulseira {MIBAND} utilizada na recolha.
MiBandWrist	Texto	{ <i>Left, Right</i> }
EdaMoveId	Inteiro	Id do sensor {EDAMOVE} utilizado na recolha.
EdaWrist	Texto	{ <i>Left, Right</i> }
__Annotations		
PK Id	Texto	<i>Tag</i> a utilizar nas anotações.
Descrição	Texto	Descrição da anotação.
WINDOWSACTIVITY_LogEntries		
PK Date	Data / hora	Data e hora de recolha de contexto.
PK OrganizationId FK	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK ParticipantId FK	Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
Computer	$\beta \Phi$ Texto	Identificação do computador de recolha.
OSVersion	$\beta \Phi$ Texto	Versão do SO do computador de recolha.
UserName	$\beta \Phi$ Texto	Nome de utilizador da sessão de trabalho do SO.
Type	$\beta \Phi$ Texto	Tipo de evento registado pelo SO.
RecordId	$\beta \Phi$ Inteiro	Número dado ao evento pelo sistema de eventos do SO.
LogonId	$\beta \Phi$ Texto	Identificador da sessão de trabalho do SO.
Info	$\beta \Phi$ Texto	Informação adicional do evento (e.g. processo criado).
Raw	$\beta \Phi$ Texto	Documento XML com informação <i>raw</i> do evento.
AutomaticIntegration	Φ Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
Inactive	Φ Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_WellBeing1		
PK Date	Data / hora	Data e hora de recolha de contexto.
PK OrganizationId FK	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK ParticipantId FK	Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
EasyRelations	$\beta \Phi$ Inteiro	Resposta à questão " <i>Relacionei-me facilmente com as pessoas à minha volta</i> ".

¹ Se *false*, o objeto foi recolhido através de formulário *backup* (substituição do instrumento principal).

² Se *false*, o objeto não é considerado no *dataset* final (e.g. foi descartado).

	DifficultSituations	$\beta \Phi$	Inteiro	Resposta à questão "Estive capaz de enfrentar situações difíceis de uma forma positiva".
	TheyLikeMe	$\beta \Phi$	Inteiro	Resposta à questão "Senti que os outros gostavam de mim e me apreciavam".
	ProudMySelf	$\beta \Phi$	Inteiro	Resposta à questão "Senti-me satisfeito com o que fui capaz de alcançar, senti-me orgulhoso de mim próprio".
	RightBalance	$\beta \Phi$	Inteiro	Resposta à questão "A minha vida foi bem equilibrada entre as minhas atividades familiares, pessoais e académicas".
	StressBegin	$\beta \Phi$	Hora	Hora de início do momento mais stressante do dia.
	StressEnd	$\beta \Phi$	Hora	Hora de fim do momento mais stressante do dia.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_WellBeing2				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	EmotionallyBalanced	$\beta \Phi$	Inteiro	Resposta à questão "Senti-me emocionalmente equilibrado".
	PeaceWithMyself	$\beta \Phi$	Inteiro	Resposta à questão "Senti-me bem, em paz comigo próprio".
	ImConfident	$\beta \Phi$	Inteiro	Resposta à questão "Senti-me confiante".
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_SleepTired				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	SleepQuality	$\beta \Phi$	Inteiro	Resposta à questão "Dormi muito bem e sinto que o meu sono foi totalmente reparador".
	FeelingTired	$\beta \Phi$	Inteiro	Resposta à questão "Sinto-me totalmente descansado após esta noite de sono".
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_Calls				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	OSVersion	$\beta \Phi$	Texto	Versão do SO do <i>smartphone</i> de recolha.
	Interlocutor	$\beta \Phi$	Texto	Número de telefone do <i>smartphone</i> de recolha (anonimizado).

	InterlocutorType	$\beta \Phi$	Texto	{ <i>Spouse-Sons-CloseFriends, RemainingFamily-Other Friends, Others, Not-classified</i> }.
	Interaction	$\beta \Phi$	Texto	{ <i>Negative, Neutral, Positive</i> }.
	Type	$\beta \Phi$	Texto	{ <i>MissedCall, InitiatedOnly, Incoming, Outgoing</i> }
	Duration	$\beta \Phi$	Inteiro	Duração de cada chamada telefónica.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_SMSs				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	OSVersion	$\beta \Phi$	Texto	Versão do SO do <i>smartphone</i> de recolha.
	Interlocutor	$\beta \Phi$	Texto	Número de telefone do <i>smartphone</i> de recolha (anoni- mizado).
	InterlocutorType	$\beta \Phi$	Texto	{ <i>Spouse-Sons-CloseFriends, RemainingFamily-Other Friends, Others, Not-classified</i> }.
	Type	$\beta \Phi$	Texto	{ <i>Incoming, Outgoing</i> }.
	Word	$\beta \Phi$	Inteiro	Número de palavras da mensagem.
	Char	$\beta \Phi$	Inteiro	Número de caracteres da mensagem.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
ANDROIDANALYTICS_ToogleScreen				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	OSVersion	$\beta \Phi$	Texto	Versão do SO do <i>smartphone</i> de recolha.
	ScreenIsOn	$\beta \Phi$	Booleano	Representa um evento de ligar o ecrã do <i>smartphone</i> (<i>true = on, false = off</i>).
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
MIBAND_Calories				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	BMR	$\beta \Phi$	Inteiro	<i>Basal Metabolic Rate</i> (BMR) (<i>Harris-Benedict Equa- tion</i>).
	Steps_Calories	$\beta \Phi$	Inteiro	Gasto de calorias durante a atividade de caminhar.
	Exercises_Calories	$\beta \Phi$	Inteiro	Gasto de calorias durante o exercício físico.
	Total	$\beta \Phi$	Inteiro	Gasto calórico diário.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹

	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
MIBAND_HR				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	HeartRate	$\beta \Phi$	Inteiro	Frequência cardíaca.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
MIBAND_Sleep				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	End	$\beta \Phi$	Data / hora	Fim da noite de sono (ou grande período de sono). O início da noite ou grande período de sono foi o registado no atributo <i>Date</i> .
	TotalSleep	$\beta \Phi$	Inteiro	Total de minutos de sono.
	DeepSleep	$\beta \Phi$	Inteiro	Total de minutos de sono profundo.
	LightSleep	$\beta \Phi$	Inteiro	Total de minutos de sono leve.
	HR_Avg	$\beta \Phi$	Inteiro	Média do (HR) no período de sono.
	HR_Max	$\beta \Phi$	Inteiro	Máximo do (HR) no período de sono.
	HR_Min	$\beta \Phi$	Inteiro	Mínimo do (HR) no período de sono.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
MIBAND_SleepDetails				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	End	$\beta \Phi$	Data / hora	Fim do período de sono (leve e profundo). O início de período foi o registado no atributo <i>Date</i> .
	Type	$\beta \Phi$	Texto	{ <i>Deep</i> , <i>Light</i> }
	Duration	$\beta \Phi$	Inteiro	Minutos de duração do período de sono.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
MIBAND_SleepRaw				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .

	Category	$\beta \Phi$	Texto	<i>Questionou-se o fabricante do {NOTIFYFITNESS} acerca do significado dos valores deste atributo e ainda não se obteve resposta.</i>
	Intensity	$\beta \Phi$	Inteiro	<i>Idem.</i>
	Steps	$\beta \Phi$	Inteiro	Número de passos registados naquele minuto (grão do registo).
	HR	$\beta \Phi$	Inteiro	(HR) registado naquele minuto (grão do registo).
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
MIBAND_Steps				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	StepsDayAccumulation	$\beta \Phi$	Inteiro	Acumulado diário de passos (registo a cada 10 minutos).
	Steps	$\beta \Phi$	Inteiro	Número de passos percorridos nos últimos 10 minutos.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
EDAMOVE_Results				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	ActivityEnergyExpenditure	$\beta \Phi$	Real	<i>Activity Energy Expenditure</i> (AEE), é a quantidade média de energia gasta por tipo de atividade física (movisens GmbH, 2018a).
	Altitude	$\beta \Phi$	Real	Média da altitude.
	EdaArousalMean	$\beta \Phi$	Real	Média dos valores <i>Arousal</i> , calculados com base nas SCR (movisens GmbH, 2018a).
	EdaSclMean	$\beta \Phi$	Real	Média dos SCL (movisens GmbH, 2018a).
	EdaScrAmplitudesMean	$\beta \Phi$	Real	Média das amplitudes dos SCR (movisens GmbH, 2018a).
	EdaScrCount	$\beta \Phi$	Real	Média do número de SCR detetadas (movisens GmbH, 2018a).
	EdaScrEnergiesMean	$\beta \Phi$	Real	Média da energia das SCR detetadas (movisens GmbH, 2018a).
	EdaScrHalfRecoveryTimesMean	$\beta \Phi$	Real	Média dos tempos de recuperação até metade da amplitude de cada SCR detetada (movisens GmbH, 2018a).
	EdaScrRiseTimesMean	$\beta \Phi$	Real	Média dos tempos de subida de cada SCR detetada (movisens GmbH, 2018a).
	InclinationDown	$\beta \Phi$	Real	Inclinação do eixo do corpo para baixo (movisens GmbH, 2018a).
	InclinationForward	$\beta \Phi$	Real	Inclinação do eixo do corpo para a frente (movisens GmbH, 2018a).
	InclinationRight	$\beta \Phi$	Real	Inclinação do eixo do corpo para a direita (movisens GmbH, 2018a).

	MET	$\beta \Phi$	Real	<i>Metabolic Equivalent Of a Task</i> (MET), representa o gasto de energia enquanto a realizar uma atividade física tendo por base uma atividade de referência (de Almeida Mendes et al., 2018) (Ashok et al., 2017). O MET é a relação entre a taxa metabólica no desempenho de uma atividade, e a taxa metabólica do período de descanso (WHO, 2019).
	MovementAcceleration	$\beta \Phi$	Real	Contém o <i>Movement Acceleration Intensity</i> (MAI), métrica relacionada com a atividade física, cujos valores se correlacionam com a intensidades dos movimentos do corpo (movisens GmbH, 2018a).
	NonWearSleepWake	$\beta \Phi$	Inteiro	{ 0 = <i>Worn</i> , 1 = <i>Not worn</i> } (movisens GmbH, 2018a).
	StepCount	$\beta \Phi$	Inteiro	Média do número de passos percorridos (movisens GmbH, 2018a).
	TempMean	$\beta \Phi$	Real	Temperatura média ambiente (movisens GmbH, 2018a).
	TotalEnergyExpenditure	$\beta \Phi$	Real	<i>Total Energy Expenditure</i> (TEE), é o total de energia gato por uma pessoa. $TEE = BMR + AEE$ (movisens GmbH, 2018a).
	VerticalSpeed	$\beta \Phi$	Real	Um valor positivo representa um movimento ascendente e um negativo um movimento descendente (movisens GmbH, 2018a).
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
WHOQOLBREF				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	Age		Inteiro	Idade.
	Gender		Texto	Género.
	DateOfBirth		Data	Data de nascimento.
	Schooling		Texto	Nível de escolaridade.
	Job		Texto	Profissão.
	MaritalStatus		Texto	Estado civil.
	IsCurrentlySick		Booleano	Estado de saúde (<i>true</i> = está doente).
	Q01		Inteiro	Resposta dada à questão 01.
	Q02		Inteiro	Resposta dada à questão...
	Q03		Inteiro	Idem.
	Q04		Inteiro	Idem.
	Q05		Inteiro	Idem.
	Q06		Inteiro	Idem.
	Q07		Inteiro	Idem.
	Q08		Inteiro	Idem.
	Q09		Inteiro	Idem.
	Q10		Inteiro	Idem.
	Q11		Inteiro	Idem.
	Q12		Inteiro	Idem.

	Q13		Inteiro	Idem.
	Q14		Inteiro	Idem.
	Q15		Inteiro	Idem.
	Q16		Inteiro	Idem.
	Q17		Inteiro	Idem.
	Q18		Inteiro	Idem.
	Q19		Inteiro	Idem.
	Q20		Inteiro	Idem.
	Q21		Inteiro	Idem.
	Q22		Inteiro	Idem.
	Q23		Inteiro	Idem.
	Q24		Inteiro	Idem.
	Q25		Inteiro	Idem.
	Q26		Inteiro	Idem.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
SOCIOCULTURAL				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	Age		Inteiro	Idade.
	Gender		Texto	<i>Male, Female, Undefined</i> .
	Ethnicity		Texto	{ <i>Caucasian, Yellow, Black, Other</i> }.
	CivilStatus		Texto	{ <i>Single, Married, Union of fact</i> }.
	Religion		Texto	{ <i>Catholic, Evangelical, Jehovah's Witness, Anglican, Islamic, Other, None</i> }.
	Living		Texto	{ <i>Alone, Direct family (spouse, sons), Indirect family (brothers or parents)</i> }.
	Acaddg		Texto	{ <i>Level 3 (Secondary education (to university access)), Level 6 (University graduation), Level 5 (Post-secondary graduation not university), Level 4 (Secondary education (with professional internship of 6 months minimum)), Level 2 (Basic education (3rd cycle))</i> }.
	Dependent1_Kinship		Texto	{ <i>Mother, Father, Son or daughter (stepson, adopted), Spouse, Other</i> }.
	Dependent1_Age		Inteiro	Idade do dependente.
	Dependent1_SameHouse		Booleano	<i>True</i> , se o dependente vive na mesma habitação que o participante.
	Dependent1_MoneyDependency		Booleano	<i>True</i> , se o dependente depende financeiramente do participante.
	Dependent1_DailyTasksDependency		Booleano	<i>True</i> , se o dependente depende do participante para realizar tarefas diárias.
	Dependent1_SpecialNeeds		Booleano	<i>True</i> , se o dependente necessita de cuidados especiais.

Dependent2_Kinship	Texto	{ <i>Mother, Father, Son or daughter (stepson, adopted), Spouse, Other</i> }.
Dependent2_Age	Inteiro	Idade do dependente.
Dependent2_SameHouse	Booleano	<i>True</i> , se o dependente vive na mesma habitação que o participante.
Dependent2_MoneyDependency	Booleano	<i>True</i> , se o dependente depende financeiramente do participante.
Dependent2_DailyTasksDependency	Booleano	<i>True</i> , se o dependente depende do participante para realizar tarefas diárias.
Dependent2_SpecialNeeds	Booleano	<i>True</i> , se o dependente necessita de cuidados especiais.
Dependent3_Kinship	Texto	{ <i>Mother, Father, Son or daughter (stepson, adopted), Spouse, Other</i> }.
Dependent3_Age	Inteiro	Idade do dependente.
Dependent3_SameHouse	Booleano	<i>True</i> , se o dependente vive na mesma habitação que o participante.
Dependent3_MoneyDependency	Booleano	<i>True</i> , se o dependente depende financeiramente do participante.
Dependent3_DailyTasksDependency	Booleano	<i>True</i> , se o dependente depende do participante para realizar tarefas diárias.
Dependent3_SpecialNeeds	Booleano	<i>True</i> , se o dependente necessita de cuidados especiais.
Dependent4_Kinship	Texto	{ <i>Mother, Father, Son or daughter (stepson, adopted), Spouse, Other</i> }.
Dependent4_Age	Inteiro	Idade do dependente.
Dependent4_SameHouse	Booleano	<i>True</i> , se o dependente vive na mesma habitação que o participante.
Dependent4_MoneyDependency	Booleano	<i>True</i> , se o dependente depende financeiramente do participante.
Dependent4_DailyTasksDependency	Booleano	<i>True</i> , se o dependente depende do participante para realizar tarefas diárias.
Dependent4_SpecialNeeds	Booleano	<i>True</i> , se o dependente necessita de cuidados especiais.
HomeType	Texto	{ <i>Apartment, Dwelling, Farmhouse, Other</i> }.
Home_Environment	Texto	{ <i>Town house, Country house</i> }.
Home_Quality	Inteiro	Qualidade da habitação onde vive.
Home_Owner	Texto	{ <i>My own, Rented, Family home</i> }.
HomeWork_Distance	Inteiro	Distância de casa ao trabalho (km).
HomeWork_Duration	Inteiro	Duração de casa ao trabalho (minutos).
HomeWork_Transportation	Texto	{ <i>Car, Moto, Bike, Public transportation, Other</i> }.
Home_HouseMove_LastTime	Inteiro	Duração de residência na habitação atual (anos).
Neighbors_Friends	Booleano	<i>True</i> , se o participante tem vizinhos amigos.
Neighbors_Neutral	Booleano	<i>True</i> , se tem vizinhos neutros.
Neighbors_Problematic	Booleano	<i>True</i> , se tem vizinhos problemáticos.
Loans_HouseMorgage	Booleano	<i>True</i> , se o participante tem crédito a habitação.
Loans_CarLoan	Booleano	<i>True</i> , se tem crédito para compra de automóvel.
Loans_OtherLoans	Booleano	<i>True</i> , se o participante tem outros créditos (e.g. pessoais).
SocialSupport_Received	Inteiro	Autoapreciação do suporte social recebido.
SocialSupport_Given	Inteiro	Autoapreciação do suporte social dado.

	HealthPerception		Inteiro	Autoapreciação geral de saúde dos elementos do domínio privado.
	WorkLife_Profession_Exchange		Inteiro	Vontade de mudar de profissão.
	WorkLife_Job_Exchange		Inteiro	Vontade de mudar de emprego atual.
	Weight		Real	Peso.
	Height		Real	Altura.
	WorkLife_Duration		Inteiro	Número de anos de trabalho (vida ativa).
	WorkLife_Job_Duration		Inteiro	Número de anos no emprego atual.
	WorkLife_Profession_Duration		Inteiro	Número de anos de exercício da profissão atual.
	WorkLife_Job_Count		Inteiro	Número de empregos desde início da vida ativa.
	Society_PoliticalMember		Booleano	<i>True</i> , o participante é filiado em partidos políticos.
	Society_ActiveParticipation		Booleano	<i>True</i> , o participante participa ativamente em partidos políticos, religião ou associações.
	Society_RelevantRoles		Booleano	<i>True</i> , o participante desempenha papéis sociais de relevo em organizações sociais, políticas ou religiosas (e.g. juntas de freguesia, grupos desportivos).
	CaffeineConsumption		Inteiro	Número aproximado de tomas de café diariamente.
	SmokingConsumption		Inteiro	Número aproximado de cigarros consumidos por dia.
	AlcoholConsumption		Texto	{ <i>With friends, Only with meals, Never, Other</i> }.
	DietPhysicalAppearance		Booleano	<i>True</i> , se o participante faz ou fez dieta para manutenção do aspeto físico.
	SocioCulturalEvents_Count		Inteiro	Número de idas a eventos culturais no último mês.
	Nature_Duration		Inteiro	Número de dias despendidos na natureza no último mês.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
LABOR_Vacations				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .
	LastDate		Data / hora	Fim do período de férias. O início das férias foi registado no atributo <i>Date</i> .
	Type		Texto	{ <i>Own year, Delayed (previous years), Early years (next year)</i> }.
	Planned		Texto	{ <i>Not Scheduled in advance, Scheduled less than 1 day in advance, Scheduled less than 1 week in advance, Scheduled less than 1 month in advance, Scheduled more than 1 month in advance</i> }.
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
LABOR_MissedVacations				
PK	Date		Data / hora	Data e hora de recolha de contexto.
PK FK	OrganizationId		Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .
PK FK	ParticipantId		Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .

	LastDate	Data / hora	Fim do período de férias. O início das férias foi registrado no atributo <i>Date</i> .	
	Type	Texto	{ <i>Own year, Delayed (previous years), Early years (next year)</i> }.	
	Planned	Texto	{ <i>Not anticipated in advance, Anticipated less than 1 day in advance, Anticipated less than 1 week in advance, Anticipated less than 1 month in advance, Anticipated more than 1 month in advance</i> }.	
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
LABOR_WorkAbsences				
PK	Date	Data / hora	Data e hora de recolha de contexto.	
PK FK	OrganizationId	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .	
PK FK	ParticipantId	Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .	
	Type	Texto	{ <i>Illness (own), Illness (children, spouse, etc.), Family assistance (children, spouse, etc.), Labor strike, Other</i> }.	
	NumberOfHours	Inteiro	Número de horas de falta.	
	Planned	Texto	{ <i>Not planned, Scheduled 1 day in advance, Scheduled 2 days in advance, Scheduled with more than 2 days in advance</i> }.	
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
LABOR_WorkDelays				
PK	Date	Data / hora	Data e hora de recolha de contexto.	
PK FK	OrganizationId	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .	
PK FK	ParticipantId	Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .	
	WorkingPeriod	Texto	{ <i>First, Second</i> }.	
	Delay	Inteiro	Número de minutos de atraso.	
	Planned	Texto	{ <i>Not planned, Scheduled with less than 2 hours in advance, Scheduled with less than 1 day in advance, Scheduled with more than 1 day in advance</i> }.	
	Cause	Texto	{ <i>Difficulty in waking up, Unexpected family event, Traffic, Other</i> }.	
	AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
	Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²
LABOR_WorkLates				
PK	Date	Data / hora	Data e hora de recolha de contexto.	
PK FK	OrganizationId	Inteiro	Código da organização. Chave estrangeira da tabela <i>Organizations</i> .	
PK FK	ParticipantId	Inteiro	Número do participante. Chave estrangeira da tabela <i>Participants</i> .	
	NumberOfHours	Inteiro	Número de horas de trabalho suplementar.	

Planned		Texto	{ <i>Not Scheduled in advance, Scheduled with 1 day in advance, Scheduled with 2 days in advance, Scheduled with more than 2 days in advance</i> }.
AutomaticIntegration	Φ	Booleano	Indica como é que o objeto foi recolhido de contexto. ¹
Inactive	Φ	Booleano	Indica se o objeto pertence ao <i>dataset</i> final. ²

Tabela 7.1: Dicionário de dados completo, da base de dados do *dataset*

Bibliografia

- Adams, A. and Robinson, P. (2015). Automated recognition of complex categorical emotions from facial expressions and head motions, *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015* pp. 355–361. [19](#), [25](#), [41](#), [56](#), [138](#)
- Adhikari, P. R. (2016). How is chi test used for feature selection in machine learning?
URL: <https://www.quora.com/How-is-chi-test-used-for-feature-selection-in-machine-learning> [50](#)
- ADInstruments (n.d.). LabChart for Research — ADInstruments.
URL: <https://www.adinstruments.com/products/labchart/labchart-for-research> [43](#), [47](#)
- Affectiva (2021). Affectiva. [42](#)
- Affectiva, I. (2014). Affectiva — How tight does the sensor have to be? (W...
URL: <http://qsensor-support.affectiva.com/customer/en/portal/articles/1539222-how-tight-does-the-sensor-have-to-be-what-is-the-optimum-fit-for-the-q-sensor-> [42](#)
- Agarwal, S., Garg, P. and Rastogi, R. (2019). Testing the reciprocal relationship between quality of work life and subjective well-being: A path analysis model, *International Journal of Project Organisation and Management* **11**(2): 140–153. [64](#)
- Aggarwal, C. C. (2013). *Outlier analysis*. [197](#)
- Agrawal, U., Giripunje, S. and Bajaj, P. (2013). Emotion and gesture recognition with soft computing tool for drivers assistance system in human centered transportation, *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013* pp. 4612–4616. [20](#), [25](#)
- Aharony, N. and Gardner, W. (2011). Funf: Open Sensing Framework.
URL: <http://www.funf.org> [41](#)
- Alabdulkarim, A. (2015). Towards hand-gesture frustration detection in interactive systems, *Proceedings - 2014 3rd International Conference on User Science and Engineering: Experience. Engineer. Engage, i-USER 2014* pp. 153–157. [5](#), [17](#), [27](#)

- Alberdi, A., Aztiria, A. and Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. [64](#)
- Alonso, M. B., Keyson, D. V. and Hummels, C. C. (2008). Squeeze, rock, and roll; Can tangible interaction with affective products support stress reduction?, *TEI'08 - Second International Conference on Tangible and Embedded Interaction - Conference Proceedings* pp. 105–108. [66](#)
- Alzoubi, O., Fossati, D., D’Mello, S. and Calvo, R. A. (2013). Affect detection and classification from the non-stationary physiological data, *Proceedings - 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013* **1**: 240–245. [20](#), [28](#), [29](#), [44](#), [46](#), [51](#)
- Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W. and Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms., *Sleep* **26**(3): 342–392. [38](#)
- Anderson, A. J., Kaplan, S. A. and Vega, R. P. (2015). The impact of telework on emotional experience: When, and for whom, does telework improve daily affective well-being?, *European Journal of Work and Organizational Psychology* **24**(6): 882–897. **URL:** <http://www.tandfonline.com/doi/full/10.1080/1359432X.2014.966086> [64](#)
- Angel, M. F. and Bonarini, A. (2014). Studying People ’ s Emotional Responses to Robot ’ s Movements in a Small Scene. [3](#)
- Anscombe, F. J. (1973). Graphs in Statistical Analysis, *The American Statistician* **27**(1): 17. **URL:** <https://www.jstor.org/stable/2682899?origin=crossref> [143](#), [149](#), [151](#)
- Antonucci, T. C. (2001). Social relations: an examination of social networks, social support, and sense of control, *Handbook of the psychology of aging* (November): 427–453. [34](#), [87](#)
- Aracena, C., Basterrech, S., Snael, V. and Velasquez, J. (2016). Neural Networks for Emotion Recognition Based on Eye Tracking Data, *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015* . [16](#), [19](#), [32](#), [43](#), [49](#), [55](#), [60](#)
- Arbuckle, J. (2010). *Ibm Spss Amos 19: User’s Guide*. [143](#)
- Arnetz, B. B. and Wiholm, C. (1997). Technological stress: Psychophysiological symptoms in modern offices, *Journal of Psychosomatic Research*. [64](#)

- Arnold, M. B. (1970). Perennial problems in the field of emotion, *Feelings and emotions : The Loyola Symposium*, Vol. 7, Academic Press. New York, pp. 169–185. [33](#), [71](#)
- Asghari, M., Sierra-Sosa, D. and Elmaghraby, A. (2018). A Semi-Automatic System for Data Management and Cleaning, *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, number March 2019, IEEE, pp. 393–397.
URL: <https://ieeexplore.ieee.org/document/8642787/> [144](#)
- Ashok, P., Kharche, J. S., Raju, R. and Godbole, G. (2017). Metabolic equivalent task assessment for physical activity in medical students, *National Journal of Physiology, Pharmacy and Pharmacology* **7**(3): 236–239. [270](#)
- Babiker, A., Faye, I. and Malik, A. (2013). Pupillary behavior in positive and negative emotions, *IEEE ICSIPA 2013 - IEEE International Conference on Signal and Image Processing Applications* pp. 7–11. [16](#), [20](#), [32](#), [44](#), [46](#), [49](#), [55](#), [60](#), [83](#), [195](#)
- Bagby, M., Parker, J. D. a. and Taylor, G. J. (1994). the Twenty-Item Item Selection Toronto and Cross-Validation Structure, *Journal of Psychosomatic research* **38**(1): 23–32. [45](#)
- Bakhtiyari, K. and Husain, H. (2014). Fuzzy Model on Human Emotions Recognition, *arXiv preprint arXiv:1407.1474* pp. 77–82.
URL: <http://arxiv.org/abs/1407.1474> [2](#), [3](#), [68](#), [75](#), [109](#)
- Bakhtiyari, K., Taghavi, M. and Husain, H. (2015). Hybrid affective computing—keyboard, mouse and touch screen: from review to experiment, *Neural Computing and Applications* **26**(6): 1277–1296. [11](#)
- Barea, R., Boquete, L., Rodriguez-Ascariz, J. M., Ortega, S. and López, E. (2011). Sensory system for implementing a human-computer interface based on electrooculography, *Sensors* **11**(1): 310–328. [27](#), [72](#), [83](#)
- Barrett, L. F., Gross, J., Christensen, T. C. and Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation, *Cognition & Emotion* **15**(6): 713–724. [2](#), [23](#)
- Basacik, D., Waters, S. and Reed, N. (2015). Detecting Cognitive Underload in Train Driving: A Physiological Approach, *The Fifth International Rail Human Factors Conference* . [85](#)
- Basu, S., Bag, A., Mahadevappa, M., Mukherjee, J. and Guha, R. (2016). Affect detection in normal groups with the help of biological markers, *12th IEEE International Confer-*

- ence Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015* pp. 1–6. [2](#), [3](#), [19](#), [24](#), [27](#), [28](#), [29](#), [34](#), [42](#), [43](#), [44](#), [51](#), [60](#), [67](#), [195](#)
- Bauer, G. and Lukowicz, P. (2012). Can smartphones detect stress-related changes in the behaviour of individuals?, *2012 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2012* (March): 423–426. [20](#), [36](#), [37](#), [53](#), [61](#)
- Biondi, M. and Picardi, A. (1999). Psychological stress and neuroendocrine function in humans: the last two decades of research., *Psychotherapy and psychosomatics* **68**(3): 114–50.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/10224513> [27](#)
- Biopac Systems Inc (2017). MP System Comparison: MP150 vs. MP100 — BIOPAC.
URL: <https://www.biopac.com/knowledge-base/mp-system-comparison-mp150-vs-mp100/> [43](#)
- Biosemi (n.d.). Analog Input Box, AIB.
URL: <https://www.biosemi.com/aib.htm> [43](#)
- Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body-motion Communication*, University of Pennsylvania press.
URL: <https://books.google.com/books?id=NvzCSAAACAAJ&pgis=1> [24](#)
- Blascovich, J. and Tomaka, J. (1996). The Biopsychosocial Model of Arousal Regulation, *Advances in Experimental Social Psychology*, Vol. 28, pp. 1–51.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S006526010860235X> [33](#)
- Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F. and Pentland, A. S. (2014). Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits, *Proceedings of the ACM International Conference on Multimedia - MM '14*, pp. 477–486.
URL: <http://dl.acm.org/citation.cfm?doid=2647868.2654933> [20](#), [36](#), [37](#), [40](#), [41](#), [46](#), [51](#), [61](#)
- Bogomolov, A., Trento, I.-P., Lepri, B., Pianesi, F., Kessler, F. B., Kessler, F. B. and Pentland, A. S. (2014). Daily Stress Recognition from Mobile Phone Data , Weather Conditions and Individual Traits, pp. 477–486. [38](#)
- Bos, D. (2010). EEG-based Emotion Recognition, *IEEE Transactions on Biomedical Engineering* **57**(7): 1798–1806. [3](#), [17](#), [21](#), [27](#), [42](#), [52](#), [71](#)

Boucsein, W. (2012). Electrodermal Activity, *Electrodermal Activity* **3**(1973): 3–67.

URL: [http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Electrodermal+Activity#0%5B1-4614-1126-0](http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Electrodermal+Activity#0%5B1-4614-1126-0%5D) 30, 77

Boyd, R., Hunt, A. and Ortiz, R. (2008). A review of cost-effectiveness literature on: public health interventions that promote mental wellbeing in the workplace, (February). 64

Bracewell, R. N. (2014). The Fourier Transform, *Diagnostic Ultrasound Imaging: Inside Out* **260**(6): 765–783.

URL: <http://linkinghub.elsevier.com/retrieve/pii/B978012396487800029X> 51

Bradley, M. and Lang, P. (1999). International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings. 83

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential, *Journal of Behavior Therapy and Experimental Psychiatry* **25**(1): 49–59.

URL: <http://www.sciencedirect.com/science/article/pii/0005791694900639> 45

Bradley, M. M., Miccoli, L., Escrig, M. A. and Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation, *Psychophysiology* **45**(4): 602–607.

URL: <http://www.ncbi.nlm.nih.gov/pubmed/18282202>
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3612940> 21, 32, 44, 49

Brainquiry (2017). BioExplorer Software License – brainquiry – Wireless, high quality signal EEG, EMG, ECG.

URL: <http://www.brainquiry.com/bioexplorer-software-license/> 47

Branquiry (2017). brainquiry – Wireless, high quality signal EEG, EMG, ECG.

URL: <http://www.brainquiry.com/> 42

Bridger, R. S. and Brasher, K. (2011). Cognitive task demands, self-control demands and the mental well-being of office workers, *Ergonomics* **54**(9): 830–839.

URL: <https://www.tandfonline.com/doi/full/10.1080/00140139.2011.596948> 64

Brouwer, A. M., van de Water, L., Hogervorst, M., Kraaij, W., Schraagen, J. M. and Hogenelst, K. (2018). Monitoring mental state during real life office work, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10727 LNCS, Springer, pp. 18–29.

URL: http://link.springer.com/10.1007/978-3-319-91593-7_3 65

- Bruntwood (2021). How office environment can impact your employees' mental health.
URL: <https://bruntwood.co.uk/blog/how-office-environment-can-impact-your-employees-mental-health/> 62
- Buchanan, T. W. and Lovallo, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans, *Psychoneuroendocrinology* 26(3): 307–317. 22, 32
- Burnett, T. (2009). Getting Started with VBA in Outlook 2010.
URL: <http://msdn.microsoft.com/en-us/library/ee814736.aspx> 147
- Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U. and Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information, *Proceedings of the 6th international conference on Multimodal interfaces - ICMI '04* p. 205.
URL: <http://portal.acm.org/citation.cfm?doid=1027933.1027968> 22, 24, 25, 26, 50, 51, 58, 195
- Č, J. Š., Škoda, J., Krbal, M. and Wasserbauer, V. Č. (2016). Comparison of Light Sources for Household Use due Circadian Effect, 1: 2–5. 30, 33
- Caballe, S. (2015). Towards a Multi-modal Emotion-Awareness e-Learning System, *Proceedings - 2015 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2015* pp. 280–287. 3, 82, 83
- Cahn, J. E. (1990). Generating Expression in Synthesized Speech, p. 177. 25
- Cambridge University (2008). Feature selection Chi2 Feature selection.
URL: <https://nlp.stanford.edu/IR-book/html/htmledition/feature-selectionchi2-feature-selection-1.html> 50
- Canli, T., Desmond, J. E., Zhao, Z., Glover, G. and Gabrieli, J. D. E. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI, *Neuroreport* 9(14): 3233–3239. 27
- Cao, N., Lin, C., Zhu, Q., Lin, Y. R., Teng, X. and Wen, X. (2018). Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data, *IEEE Transactions on Visualization and Computer Graphics* 24(1): 23–33. 144
- Carr, D. B., Littlefield, R. J., Nicholson, W. L. and Littlefield, J. S. (1987). Scatterplot Matrix Techniques for Large N, *Journal of the American Statistical Association* 82(398): 424. 143

- Carvalho, C., Moreira, R. S. and Torres, J. M. (2021). Data Quality Visual Analysis (DQVA) A tool to process and pinspot raw data irregularities, *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, pp. 1036–1045.
URL: <https://ieeexplore.ieee.org/document/9375961/> 109, 132, 142, 155, 198
- Carvalho, C., Torres, J. and Moreira, R. (2018). Revisão de contribuições para um pipeline de reconhecimento e consciência de emoções, *Technical report*, Universidade Fernando Pessoa (ISUS), Porto.
URL: <http://hdl.handle.net/10284/9911> 7, 8, 9, 14, 18, 38, 67, 198
- Casella, G. and Berger, R. L. (2017). *Inferência Estatística: Tradução da 2ª edição norte-americana*, Cengage Brasil. 150
- Castellano, G., Villalba, S. D. and Camurri, A. (2007). Recognising Human Emotions from Body Movement and Gesture Dynamics, *Affective Computing and Intelligent Interaction* pp. 71–82.
URL: http://link.springer.com/10.1007/978-3-540-74889-2_7 21, 24, 26, 38, 62, 195
- Chandler, C. and Cornes, R. (2012). Biometric Measurement of Human Emotions, **4**(2). 3, 24, 30, 31
- Chandola, V., Banerjee, A. and Kumar, V. (2009). Anomaly detection, *ACM Computing Surveys* **41**(3): 1–58.
URL: <https://dl.acm.org/doi/10.1145/1541880.1541882> 131
- Chandraprabha, K. S., Shwetha, A. N., Kavitha, M. and Sumathi, R. (2021). Real time-Employee Emotion Detection system (RtEED) using Machine Learning, *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, IEEE, pp. 759–763.
URL: <https://ieeexplore.ieee.org/document/9388510/> 66
- Chang, C.-Y., Lin, Y.-M. and Zheng, J.-Y. (2012). Physiological Angry Emotion Detection Using Support Vector Regression, *2012 15th International Conference on Network-Based Information Systems* pp. 592–596.
URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6354890> 20, 28, 44, 46, 52
- Chen, J. (2019). Skewness.
URL: <https://www.investopedia.com/terms/s/skewness.asp> 150
- Chen, L. S., Huang, T. S., Miyasato, T. and Nakatsu, R. (1998). Multimodal Human Emotion Expression Recognition, *Third International Conf Automatic Face and Gesture Recognition* **13**(7): 366–371. 22, 25, 49, 68

Chen, W., Jaques, N., Taylor, S., Sano, A., Fedor, S. and Picard, R. W. (2014). WAVELET-BASED MOTION ARTIFACT REMOVAL FOR ELECTRODERMAL ACTIVITY Electrodermal Activity Motion Artifact in EDA Our Method A . Stationary wavelet transform C . Inverse wavelet transform EDA data containing motion artifacts was obtained from a. [30](#)

Chen, Z., Lin, M., Chen, F., Lane, N., Cardone, G., Wang, R., Li, T., Chen, Y., Choudhury, T. and Cambell, A. (2013). Unobtrusive Sleep Monitoring using Smartphones, *Proceedings of the ICTs for improving Patients Rehabilitation Research Techniques* . **URL:** <http://eudl.eu/doi/10.4108/icst.pervasivehealth.2013.252148> [43](#)

Chu, X., Ilyas, I. F., Krishnan, S. and Wang, J. (2016). Data cleaning: Overview and emerging challenges, *Proceedings of the ACM SIGMOD International Conference on Management of Data 26-June-20*: 2201–2206. [11](#), [47](#), [109](#), [142](#)

Cohen, S., Kamarck, T. and Mermelstein, R. (1983). A Global Measure of Perceived Stress. [45](#)

Coulbourn (2017). Coulbourn Instruments.

URL: <http://www.coulbourn.com/> [43](#)

Croman, J. (2018). How Artificial Intelligence Can Improve Your Health and Productivity. [62](#)

Cronbach, A. D. (2004). Alfa de Cronbach, *Encontro da Associação Nacional de Pós-graduação e Pesquisa em Administração* pp. 1–4. [183](#)

Crossman, A. (2017). Correlation Analysis in Sociological Research.

URL: <https://www.thoughtco.com/what-is-correlation-analysis-3026696> [50](#)

Crowne, D. P. and Marlowe, D. (1960). Marlowe-Crowne Social Desirability Scale, *Journal of Consulting Psychology* **24**(4): 349–354.

URL: <http://dx.doi.org/10.1037/t05257-000> [45](#)

Cruz, A., Garcia, D., Pires, G. and Nunes, U. (2015). Facial Expression Recognition based on EOG toward Emotion Detection for Human-Robot Interaction., *Biosignals* **2**: 31–37.

URL: <http://www.dblp.org/rec/bibtex/conf/biostec/CruzGPN15%5Cnpapers2://publication/uuid/F0A7D08E8-4967-B20B-7EE34D518E7D> [20](#), [44](#), [52](#), [53](#), [60](#)

Cryptography, P. (2009). Mel Frequency Cepstral Coefficient (MFCC) tutorial.

URL: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/> [58](#)

CS&S (2020). OpenRefine.

URL: <http://openrefine.org/> <https://github.com/OpenRefine> 143

Curcio, G., Piccardi, L., Ferlazzo, F., Giannini, A. M., Burattini, C. and Bisegna, F. (2016). LED lighting effect on sleep, sleepiness, mood and vigor, *EEEIC 2016 - International Conference on Environment and Electrical Engineering* pp. 0–4. 33, 68

Czeisler, C. A., Allan, J. S., Strogatz, S. H., Ronda, J. M., Sanchez, R., Rios, C. D., Freitag, W. O., Richardson, G. S. and Kronauer, R. E. (1986). Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle, *Science* **233**(4764): 667–671. 33

Dallachiesat, M., Ebaid, A., Eldawy, A., Elmagarmid, A., Ilyas, I. F., Ouzzani, M. and Tang, N. (2013). NADEEF: A commodity data cleaning system, *Proceedings of the ACM SIGMOD International Conference on Management of Data* pp. 541–552. 144

de Almeida Mendes, M., da Silva, I., Ramires, V., Reichert, F., Martins, R., Ferreira, R. and Tomasi, E. (2018). Metabolic equivalent of task (METs) thresholds as an indicator of physical activity intensity, *PLoS ONE* **13**(7). 270

Denissen, J. J. a., Butalid, L., Penke, L. and van Aken, M. a. G. (2008). The effects of weather on daily mood: a multilevel approach, *Emotion (Washington, D.C.)* **8**(5): 662–667.

URL: <http://dx.doi.org/10.1037/a0013497> 38

Dhall, A., Asthana, A., Goecke, R. and Gedeon, T. (2011). Emotion recognition using PHOG and LPQ features, *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011* pp. 878–883. 21, 58

Dickerson, S. S. and Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research, *Psychological bulletin* **130**(3): 355–391.

URL: http://anthro.vancouver.wsu.edu/media/Course_files/anth-260-edward-h-hagen/acute-stressors-and-cortisol-responses-a-theoretical-integration-and-synthesis-of-laboratory-research-1.pdf%5Cnhttp://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=a5d7d532-7f97-2, 32, 33

Doanne, D. P. and Seward, L. E. (2011). Measuring Skewness. 150, 183

Dodge, Y. (2006). *The Oxford Dictionary of Statistical Terms*, Oxford University Press. 149

- Dong, W., Lepri, B. and Pentland, A. (2011). Modeling the co-evolution of behaviors and social relationships using mobile phone data, *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia - MUM '11* pp. 134–143.
URL: <http://dl.acm.org/citation.cfm?doid=2107596.2107613> 16, 33, 195
- Drefahl, A. (2016). Commonly used skewness coefficients for sampled data, pp. 1–5. 150
- Duijts, S. F., Kant, I., Swaen, G. M., van den Brandt, P. A. and Zeegers, M. P. (2007). A meta-analysis of observational studies identifies predictors of sickness absence, *Journal of Clinical Epidemiology* **60**(11): 1105–1115. 63
- Dunbar, M., Ford, G. and Hunt, K. (1998). Why is the receipt of social support associated with increased psychological distress? An examination of three hypotheses, *Psychology and health* **13**(3): 37–41.
URL: <http://www.tandfonline.com/doi/abs/10.1080/08870449808407308> 34
- Duyan, E. C. (2013). Measuring Work Related Quality of Life and Affective Well-being in Turkey, *Mediterranean journal of social sciences* .
URL: http://www.mcser.org/index.php?option=com_content&view=article&id=926:measuring-work-related-quality-of-life-and-affective-well-being-in-turkey&catid=50:mjss-vol-4-no1-2013&Itemid=135 64
- Eckert, M., Gil, A., Zapatero, D., Meneses, J. and Martínez Ortega, J. F. (2016). Fast facial expression recognition for emotion awareness disposal, *IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin 2016-October*: 179–183. 19, 24, 25, 40, 41
- Ehrampoosh, A., Yousefi-koma, A. and Mohtasebi, S. (2016). EMG-Based Estimation of Shoulder Kinematic Using Neural Network and Quadratic Discriminant Analysis, pp. 471–476. 31
- Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion.
URL: <https://www.paulekman.com/wp-content/uploads/2013/07/The-Argument-And-Evidence-About-Universals-In-FacialExpressi.pdf> 2, 33, 68
- Ekman, P., Levenson, R. W. and Friesen, W. V. (2015). American Association for the Advancement of Science Autonomic Nervous System Activity Distinguishes among Emotions, *Science* **221**(4616): 1208–1210.
URL: <http://www.jstor.org/about/terms.html> . *JSTOR 's Terms and Conditions of Use provides , in part , that unless you have obtained page of such transmission . journals and scholarly literature from around the world . The Archive is supported by libraries* , 24

Ekman, Paul; Friesen, W. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. 24

El Kaliouby, R. (2005). Mind-reading machines: automated inference of complex mental states, *Bell Laboratories memorandum* (March): 185.

URL: <http://www.mendeley.com/research/mindreading-machines-automated-inference-of-complex-mental-states/> 68

Electrophysiology, T. F. o. t. E. S. o. C. t. N. A. S. (1996). Heart Rate Variability : Standards of Measurement, Physiological Interpretation, and Clinical Use, *Circulation* **93**(5): 1043–1065.

URL: <http://circ.ahajournals.org/cgi/doi/10.1161/01.CIR.93.5.1043> 59

Elmagarmid, A. K., Ipeirotis, P. G. and Verykios, V. S. (2007). Duplicate record detection: A survey, *IEEE Transactions on Knowledge and Data Engineering* . 131

Emotiv (2017). EMOTIV SDK and Apps for Developers and proprietary research.

URL: <https://www.emotiv.com/developer/> 47

Estrada, E., Nazeran, H., Barragan, J., Burk, J. R., Lucas, E. A. and Behbehani, K. (2006). EOG and EMG: Two important switches in automatic sleep stage classification, *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings* pp. 2458–2461. 32

Fasel, B. and Luetttin, J. (2003). Automatic facial expression analysis: a survey, *Pattern Recognition* **36**(1): 259–275.

URL: <http://www.sciencedirect.com/science/article/pii/S0031320302000523%5Cnhttp://linkinghub.elsevier.com/locate/0031320302000523> 24

Ferreira, C. (2013). Relação Com O Bem-Estar Psicologico. 86

Field, A. (2013). Discovering statistics using IBM SPSS statistics. 143, 145

Fonte, R. (2015). Fazer um exame ao nível de serotonina.

URL: <http://www.vladman.net/blog/fazer-um-exame-ao-nivel-de-serotonina> 33, 68

Fostervold, K. and Nersveen, J. (2008). Proportions of direct and indirect indoor lighting — The effect on health, well-being and cognitive performance of office workers, *Lighting Research & Technology* **40**(3): 175–200.

URL: <http://journals.sagepub.com/doi/10.1177/1477153508090917> 64

Fraunhofer IIS (2017). SHORE™ – pioneering facial analysis.

URL: <https://www.iis.fraunhofer.de/en/ff/bsy/tech/bildanalyse/shore-gesichtsdetektion.html> 41

- Frazão, A. (2016). Serotonina - Funções e como Aumentar.
URL: <https://www.tuasaude.com/serotonina/> 32
- Fredrickson, B. (2002). How does religion benefit health and well-being? Are positive emotions active ingredients?, *Psychological Inquiry* **13**(3): 209–213.
URL: <http://www.jstor.org/stable/1449332> 5, 70, 196
- Friendly, M. and Denis, D. (2005). The early origins and development of the scatterplot, *Journal of the History of the Behavioral Sciences* **41**(2): 103–130. 153
- Fritz, C., Yankelevich, M., Zarubin, A. and Barger, P. (2010). Happy, healthy, and productive: The role of detachment from work during nonwork time., *Journal of Applied Psychology* **95**(5): 977–983.
URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0019462> 63
- Fukumto, M. and Nagamatsu, R. (2016). Feedback of Laughter by Detecting Variation in Respiration Amplitude for Augmenting Laughter, *2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)* pp. 139–142.
URL: <http://ieeexplore.ieee.org/document/7794453/> 29
- Fulton, B. and Medlock, M. (2003). Beyond Focus Groups : Getting More Useful Feedback from Consumers. 39, 83
- Galarnyk, M. (2018). Understanding Boxplots.
URL: <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51> 151
- Gama, J., Carvalho, A., Faceli, K., Lorena, A. and Oliveira, M. (2012). *Extração de Conhecimento de Dados - Data Mining*, 2 edn, Edições Sílabo, Lda., Lisboa. 11, 48, 51, 52, 54, 55, 109, 128, 138, 197
- Garica-Ceja, E., Osmani, V. and Mayora, O. (2015). Automatic Stress Detection in Working Environments from Smartphones' Accelerometer Data: A First Step, *Biomedical and Health Informatics, IEEE Journal of (to appear)*: 1–8. 38
- George, L. K., Blazer, D. G., Hughes, D. C. and Fowler, N. (1989). Social support and the outcome of major depression, *The British Journal of Psychiatry* **154**(4): 478–485.
URL: <http://bjp.rcpsych.org/cgi/doi/10.1192/bjp.154.4.478> 34
- Giakoumis, D., Vogianou, A., Kosunen, I., Moustakas, K., Tzovaras, D. and Hassapis, G. (2010). Identifying Psychophysiological Correlates of Boredom and Negative Mood Induced During HCI, *Bio-inspired Human-Machine Interfaces and Healthcare Applications* pp. 3–12.

- URL:** http://www.iti.gr/moustak/BIOSTEC10_Correlates.pdf%5Cnhttp://eprints.ucl.ac.uk/20225/%5Cn
30
- Gilad-bachrach, R. (2004). Margin Based Feature Selection - Theory and Algorithms. 49
- Ginger (2021). 2021 Workforce Attitudes Toward Mental Health.
URL: <https://go.ginger.io/third-annual-workforce-attitudes-study> 63
- Glen, S. (2014). Pearson's Coefficient of Skewness.
URL: <https://www.statisticshowto.com/pearsons-coefficient-of-skewness/> 150
- Glen, S. (2017). Upper and Lower Fences.
URL: <https://www.statisticshowto.com/upper-and-lower-fences/> 151
- Glen, S. (n.d.). Interquartile Range (IQR): What it is and How to Find it.
URL: <https://www.statisticshowto.com/probability-and-statistics/interquartile-range/>
150
- Gogia, Y., Singh, E., Mohatta, S. and Sreejith, V. (2016). Multi-modal Affect Detection for Learning Applications, *2016 IEEE Region 10 Conference (TENCON)*, pp. 3747–3751. 2, 3, 17, 19, 25, 41, 42, 49, 51, 55, 59, 60, 82
- Gonçalves, E. (2011). Caracterização do voluntariado hospitalar em contexto oncológico. 86
- Google (2020). google-refine.
URL: <https://code.google.com/archive/p/google-refine/> 143
- Google (2021). Google Forms.
URL: <https://www.google.com/forms/about/> 87
- Gradisar, M. and Lack, L. (2004). Relationships between the circadian rhythms of finger temperature, core temperature, sleep latency, and subjective sleepiness., *Journal of biological rhythms* **19**(2): 157–163. 30
- Group, T. W. (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group, *Psychol Med* **28**(3): 551–558.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/9626712> 85
- G.tec (2017). g.tec medical engineering.
URL: <http://www.gtec.at/Products/Hardware-and-Accessories/g.MOBIIlab-Specs-Features> 44
- Guillotet, P., Fleureau, J., Orlac, I. and Silveira, F. (2013). On the fly user's emotion capture, *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013* pp. 713–714. 17

- Guinot Jimeno, F. A., Yuste Bielsa, S., Cuadros Fernández, C., Lorente Rodríguez, A. I. and Mercadé Bellido, M. (2011). Objective and subjective measures for assessing anxiety in paediatric dental patients, *European Journal of Paediatric Dentistry* **12**(4): 239–244. [39](#), [83](#)
- Gunes, H. and Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures, *Journal of Network and Computer Applications* **30**(4): 1334–1345. [16](#), [21](#), [25](#), [26](#), [58](#)
- Guo, D. (2003). Coordinating Computational and Visual Approaches for Interactive Feature Selection and Multivariate Clustering, *Information Visualization* . [144](#)
- Guo, S., Jin, Z., Chen, Q., Gotz, D., Zha, H. and Cao, N. (2019). Visual Anomaly Detection in Event Sequence Data, *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019* pp. 1125–1130. [144](#)
- Gutmann, M., Grausberg, P., Kyamakya, K. and Klagenfurt, A.-a. U. (2015). Detecting Human Driver ' s Physiological Stress and Emotions Using Sophisticated One-Person Cockpit Vehicle Simulator, pp. 15–18. [3](#), [17](#)
- Haag, A., Goronzy, S., Schaich, P. and Williams, J. (2004). Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System, *Affective dialogue systems* **i**: 36–48. [22](#), [31](#), [56](#), [77](#), [138](#)
- Harb, F., Hidalgo, M. P. and Martau, B. (2014). Lack of exposure to natural light in the workspace is associated with physiological, sleep and depressive symptoms, *Chronobiology International* **1**(1): 1–8. [38](#)
- Hargrave, M. and Wetfall, P. (2020). Standard Deviation Definition.
URL: <https://www.investopedia.com/terms/s/standarddeviation.asp> [150](#)
- Haritaoglu, I., Cozzi, A., Koons, D., Flickner, M., Zotkin, D. and Yacoob, Y. (2001). ATTENTIVE TOYS IBM Almaden Research , San Jose , CA 95120 , USA Computer Vision Laboratory University of Maryland , College Park , MD 20742 , USA. [25](#), [38](#)
- Harmer, C. J., Bhagwagar, Z., Perrett, D. I., Völlm, B. A., Cowen, P. J. and Goodwin, G. M. (2003). Acute SSRI administration affects the processing of social cues in healthy volunteers., *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **28**(1): 148–52.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/12496951> [22](#), [33](#), [36](#), [43](#), [46](#), [47](#)
- Harmer, C. J., Shelley, N. C., Cowen, P. J. and Goodwin, G. M. (2004). Increased positive versus negative affective perception and memory in healthy volunteers following selec-

- tive serotonin and norepinephrine reuptake inhibition, *American Journal of Psychiatry* **161**(7): 1256–1263. [33](#)
- Hayes, A. (2019). Variance.
URL: <https://www.investopedia.com/terms/v/variance.asp> [150](#)
- Hayes, A. and Wetfall, P. (2020). Coefficient of Variation (CV).
URL: <https://www.investopedia.com/terms/c/coefficientofvariation.asp> [150](#)
- He, J., Veltri, E., Santoro, D., Li, G., Mecca, G., Papotti, P. and Tang, N. (2016). Interactive and Deterministic Data Cleaning, pp. 893–907. [144](#)
- Healey, J. A. and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors, *IEEE Transactions on Intelligent Transportation Systems* **6**(2): 156–166. [21](#), [29](#), [30](#), [43](#), [47](#), [49](#), [77](#)
- Healey, J. a., Picard, R. W. and Smith, A. C. (2000). Wearable and Automotive Systems for A ect Recognition from Physiology Accepted by Wearable and Automotive Systems for A ect Recognition by, p. 158. [22](#), [29](#), [42](#)
- Healey, J. and Picard, R. (1998). Digital processing of affective signals, *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)* **6**: 3749–3752. [22](#)
- Heitzmann, C. A. and Kaplan, R. M. (1988). Assessment of methods for measuring social support., *Health Psychology* **7**(1): 75–109. [34](#), [68](#)
- Hellerstein, J. M. (2008). Quantitative Data Cleaning for Large Databases, *United Nations Economic Commission for Europe* p. 42.
URL: <http://db.cs.berkeley.edu/jmh/cleaning-unece.pdf%5Cnpapers2://publication/uuid/DC7173AB-6B26-4B8B-AEC3-4C7E65CEEFED> [132](#)
- Hendriks, A. A. J., Ormel, J. and van de Willige, G. (1990). Long-term difficulties measured by a self-report questionnaire and semi-structured interview: a comparison of methods [in Dutch], *Gedrag en Gezondheid* **18**: 273–283. [46](#)
- Herbon, A., Peter, C., Markert, L. and Meer, E. V. D. (2005). Emotion studies in HCI – a new approach, *Proceedings of the 2005 HCI International Conference* (1986).
URL: http://www.prometei.de/fileadmin/prometei.de/publikationen/Herbon_etal_2005.pdf
[21](#), [35](#), [36](#), [42](#), [44](#), [52](#)
- Herlan, A., Ottenbacher, J., Schneider, J., Riemann, D. and Feige, B. (2018). Electrodermal activity patterns in sleep stages and their utility for sleep versus wake classification, *Journal of Sleep Research* p. e12694.
URL: <http://doi.wiley.com/10.1111/jsr.12694> [85](#)

- Hernandez, J., Morris, R. R. and Picard, R. W. (2011). Call Center Stress Recognition with Person-Specific Models, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6974 LNCS(PART 1)**: 125–134. [21](#), [40](#), [41](#), [47](#)
- Hinton, P. (2014). *SPSS Explained*. [143](#)
- Hodge, V. J. and Austin, J. (2004). A survey of outlier detection methodologies. [132](#)
- Hohaus, L. and Berah, E. (1996). Stress, achievement, marriage and social support: Effects on the psychological well-being of physicians entering mid-life/ mid-career, *Psychology and Health* **11(5)**: 715–731.
URL: <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0030499781&partner=40&rel=R4.5.0> [34](#), [68](#)
- Holden, L., Scuffham, P. A., Hilton, M. F., Vecchio, N. N. and Whiteford, H. A. (2010). Work performance decrements are associated with Australian working conditions, particularly the demand to work longer hours, *Journal of Occupational and Environmental Medicine* **52(3)**: 281–290. [63](#)
- Hossain, M. S., Huda, K., Rahman, S. M. S. and Ahmad, M. (2016). Implementation of an EOG based security system by analyzing eye movement patterns, *Proceedings of 2015 3rd International Conference on Advances in Electrical Engineering, ICAEE 2015* pp. 149–152. [32](#)
- Hsu, Y. J., Shih, J. L. and Chen, C. H. (2012). Emotion labyrinth: Learning to rationalize emotions through 3D game environment, *Proceedings of the 2012 IIAI International Conference on Advanced Applied Informatics, IIAIAI 2012* (1998): 153–158. [3](#)
- Huhtala, Y., Kärkkäinen, J., Porkka, P. and Toivonen, H. (1999). TANE: An efficient algorithm for discovering functional and approximate dependencies, *Computer Journal* . [131](#)
- IJsselsteijn, W., De Kort, Y., Midden, G., Eggen, B. and Van Den Hoven, E. (2006). Persuasive technology for human well-being: Setting the scene, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [65](#)
- ILO (2021). Workplace well-being.
URL: https://www.ilo.org/global/topics/safety-and-health-at-work/areasofwork/workplace-health-promotion-and-well-being/WCMS_118396/lang-en/index.htm [64](#), [65](#)

- Imholz, B. P. M., Montfrans, G. A. V., Settels, J. J., Hoeven, G. M. A. V. D., Karemaker, J. M. and Wieling, W. (1988). Continuous non-invasive blood pressure monitoring: Reliability of finapres device during the valsalva manoeuvre. 43
- Infiniti, P. (2008). ProComp Infiniti, (800): 1–45. 44
- International Labour Organization (2021). Workplace health promotion and well-being.
URL: <https://www.ilo.org/global/topics/safety-and-health-at-work/areasofwork/workplace-health-promotion-and-well-being/lang-en/index.htm>
 65
- Investopedia (2017). Sampling.
URL: <http://www.investopedia.com/terms/s/sampling.asp> 54
- IPLeiria (2009). Técnicas de Amostragem - WikiEducacao.
URL: http://wiki.ued.ipleiria.pt/wikiEducacao/index.php/Técnicas_de_Amostragem 54
- Iwasaki, K., Miyaki, T. and Rekimoto, J. (2010). AffectPhone: A Handset Device to Present User's Emotional State with Warmth / Coolness, *B-Interface* 1: 83–88. 30
- Jackson, L. and Rothmann, S. (2005). Work-related well-being of educators in a district of the Norm-West Province. 64
- James, W. (1884). H.—WHAT IS AN EMOTION ? By Professor, *Mind* 9: 188–205.
URL: <http://mind.oxfordjournals.org/content/os-IX/34/188> 1
- Jamshidnejad, A. and Jamshidined, A. (2009). Facial emotion recognition for human computer interaction using a fuzzy model in the e-business, *2009 Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA)* (July): 202–204.
URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5224214> 3, 17
- Jaques, N., Taylor, S., Azaria, A., Ghandeharioun, A., Sano, A. and Picard, R. (2015). Predicting students' happiness from physiology, phone, mobility, and behavioral data, *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015* pp. 222–228. 19, 30, 35, 36, 37, 38, 42, 49, 53, 61
- Jarkovska, D., Valesova, L., Chvojka, J., Benes, J., Svirglerova, J., Florova, B., Nalos, L., Matejovic, M. and Stengl, M. (2016). Heart rate variability in porcine progressive peritonitis-induced sepsis, *Frontiers in Physiology* 6(JAN). 59
- Jawbone (2017). UP by Jawbone — Fitness trackers for a healthier you.
URL: <https://jawbone.com/?rf=bodymedia> <https://jawbone.com/> 42

Jeatrakul, P., Wong, K. W. and Fung, C. C. (2010). Data cleaning for classification using misclassification analysis, *Journal of Advanced Computational Intelligence and Intelligent Informatics* **14**(3): 297–302. [11](#), [109](#)

Jerritta, S., Murugappan, M., Nagarajan, R. and Wan, K. (2011). Physiological signals based human emotion Recognition: a review, *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on* pp. 410–415.

URL: <http://ieeexplore.ieee.org/ielx5/5755141/5759825/05759912.pdf?tp=&arnumber=5759912&isnu>
[16](#), [17](#), [25](#), [48](#), [138](#)

John, O. P., & Srivastava, S. (1999). Big Five Inventory (BFI), *Handbook of personality: Theory and research* **2**: 102–138.

URL: <http://www.fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Personality.pdf>
[45](#)

Johnston, D. W., Propper, C. and Shields, M. A. (2009). Comparing Subjective and Objective Measures of Health : Evidence from Hypertension for the Income / Health Gradient Comparing Subjective and Objective Measures of Health : Evidence from Hypertension for the Income / Health Gradient, **2737**(2737): 540–552. [39](#), [44](#), [83](#)

Jones, C. M. and Troen, T. (2007). Biometric valence and arousal recognition, *Proceedings of the 2007 conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artifacts and environments - OZCHI '07* pp. 191–194.

URL: <http://portal.acm.org/citation.cfm?doid=1324892.1324929> [1](#)

Kahn, R. L. and Antonucci, T. C. (1980). Convoys over the life course: Attachment, roles, and social support, *Life-span Development and Behavior* (3): 253–286. [34](#)

Kandel, E. R., Schwartz, J. H. and Jessell, T. M. (2000). *Principles of Neural Science*, Vol. 4, McGraw-hill New York.

URL: <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0071120009%5Cnhttp://www.amazon.de/exec/obidos/redirect?tag=citeulike01-21&path=ASIN/0071120009%5Cnhttp://www.amazon.fr/exec/obidos/redirect?tag=citeulike06-21&path=ASIN/00> [27](#)

Kandel, S., Paepcke, A., Hellerstein, J. and Heer, J. (2011). Wrangler: Interactive visual specification of data transformation scripts, *Conference on Human Factors in Computing Systems - Proceedings* pp. 3363–3372. [144](#)

Kandel, S., Parikh, R., Paepcke, A., Hellerstein, J. M. and Heer, J. (2012). Profiler: Integrated statistical analysis and visualization for data quality assessment, *Proceedings of*

- the Workshop on Advanced Visual Interfaces AVI* pp. 547–554. [11](#), [47](#), [109](#), [112](#), [143](#), [144](#), [197](#)
- Kärner, T. (2017). A mixed-methods study of physiological reactivity to domain-specific problem solving: methodological perspectives for process-accompanying research in VET, *Empirical Research in Vocational Education and Training* **9**(1): 10.
URL: <http://ervet-journal.springeropen.com/articles/10.1186/s40461-017-0054-3> [85](#)
- Kawai, S., Takano, H. and Nakamura, K. (2013). Pupil diameter variation in positive and negative emotions with visual stimulus, *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013* pp. 4179–4183. [20](#), [32](#), [49](#)
- Kemper, T. D. (1978). *A social interactional theory of emotion*, Vol. 61 (5), Wiley New York. [33](#)
- Kemper, T. P. (1991). Predicting Emotions from Social Relations, *Social Psychology Quarterly* **54**(4): 330–342.
URL: <http://www.jstor.org/stable/2786845> [33](#)
- Kenton, W. (2019). Kurtosis.
URL: <https://www.investopedia.com/terms/k/kurtosis.asp> [150](#), [151](#)
- Kenton, W. (2020). Trimmed Mean Definition.
URL: https://www.investopedia.com/terms/t/trimmed_mean.asp [150](#)
- Khayyat, Z., Ilyasz, I. F., Jindal, A., Madden, S., Ouzzani, M., Papotti, P., Quiané-Ruiz, J. A., Tang, N. and Yin, S. (2015). BigDancing: A system for big data cleansing, *Proceedings of the ACM SIGMOD International Conference on Management of Data 2015-May*: 1215–1230. [144](#)
- Kim, J. and André, E. (2008). Emotion recognition based on physiological changes in music listening, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(12): 2067–2083. [2](#), [16](#), [21](#), [27](#), [33](#), [44](#), [51](#), [68](#), [75](#), [195](#)
- Kim, K. H., Bang, S. W. and Kim, S. R. (2004). Emotion recognition system using short term monitoring of physiological signals, *Medical Biological Engineering and computing* **42**(Journal Article): 419–427. [22](#), [24](#), [30](#), [54](#)
- Kim-Prieto, C. and Diener, E. (2009). Religion as a source of variation in the experience of positive and negative emotions, *The Journal of Positive Psychology* **4**(6): 447–460. [70](#)
- Kimani, E., Rowan, K., McDuff, D., Czerwinski, M. and Mark, G. (2019). A Conversational Agent in Support of Productivity and Wellbeing at Work, *2019 8th International*

- Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, pp. 1–7.
URL: <https://ieeexplore.ieee.org/document/8925488/> 66
- Kirschbaum, C., Pirke, K. M. and Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—a tool for investigating psychobiological stress responses in a laboratory setting.
URL: <http://p113367.typo3server.info/uploads/media/lit9304.pdf%5Chttp://www.ncbi.nlm.nih.gov/pub>
 2, 68
- Kirst, M. (2010). Unisens 2.0, *Signals* . 87, 102
- Klatt, M. D., Wise, E. and Fish, M. (2016). Mindfulness and Work-Related Well-Being, *Mindfulness and Buddhist-Derived Approaches in Mental Health and Addiction*, pp. 313–336. 64
- KNIME AG (2020). Knime.
URL: <https://www.knime.com/> 143
- Kohll, A. (2021). How Your Office Space Impacts Employee Well-Being.
URL: <https://www.forbes.com/sites/alankohll/2019/01/24/how-your-office-space-impacts-employee-wellbeing/?sh=43d959c764f3> 63
- Koldijk, S. (2012). Automatic recognition of context and stress to support knowledge workers, *ACM International Conference Proceeding Series*. 66
- Koldijk, S., Neerincx, M. A. and Kraaij, W. (2018). Detecting Work Stress in Offices by Combining Unobtrusive Sensors, *IEEE Transactions on Affective Computing* 9(2): 227–239.
URL: <https://ieeexplore.ieee.org/document/7572141/> 65
- Koralewski (n.d.). Koralewski Elektronik.
URL: <http://www.koralewski.de/english/index.php> 43
- Korkmaz, O. E. and Atasoy, A. (2015). Emotion Recognition from Speech Signal Using Mel-Frequency Cepstral Coefficients, pp. 1254–1257. 19, 26, 41, 52, 53, 58, 195
- Kosara, R., Bendix, F. and Hauser, H. (2004). TimeHistograms for Large, Time-Dependent Data, *Joint Eurographics - IEEE TCVG Symposium on Visualization (Vis-Sym)*. 143
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review, *Biological Psychology* 84(3): 394–421.
URL: <http://dx.doi.org/10.1016/j.biopsycho.2010.03.010> 1, 23, 28

- Krishnan, S., Wang, J., Franklin, M. J., Goldberg, K., Kraska, T., Milo, T. and Wu, E. (2015). SampleClean: Fast and Reliable Analytics on Dirty Data, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* pp. 59–75.
URL: <http://sites.computer.org/debull/A15sept/p59.pdf> 143
- Krishnan, S., Wang, J., Wu, E., Franklin, M. J. and Goldberg, K. (2016). ActiveClean: Interactive data cleaning for statistical modeling, *Proceedings of the VLDB Endowment* 9(12): 948–959. 144
- Kusserow, M., Amft, O. and Troster, G. (2013). Monitoring Stress Arousal in the Wild, *IEEE Pervasive Computing* 12(2): 28–37.
URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6253197> 20, 28, 36, 42, 53
- Lai, H., Ramanathan, V. and Wechsler, H. (2008). Reliable face recognition using adaptive and robust correlation filters, *Computer Vision and Image Understanding* 111(3): 329–350.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S1077314208000088> 24
- Lalitha, S., Mudupu, A., Nandyala, B. V. and Munagala, R. (2015). Speech emotion recognition using DWT, *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, Vol. - 6, IEEE, pp. 1–4.
URL: <http://ieeexplore.ieee.org/document/7435630/> 1, 19, 26, 41, 53, 58
- Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T. and Campbell, A. T. (2010). A survey of mobile phone sensing, *IEEE Communications Magazine* 48(9): 140–150. 21
- Lane, N., Mohammad, M., Lin, M. and Yang, X. (2011). Bewell: A smartphone application to monitor, model and promote wellbeing, ... *computing technologies ...*
URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.298.2259&rep=rep1&type=pdf> 21, 36, 37, 41, 49
- Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J. and Schwartz, G. E. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion, *Neuropsychologia* 35(11): 1437–1444. 27
- Lang, P., Bradley, M. and Cuthbert, B. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual., *Tech. Rep. A-6* . 83
- Lazarus, R. S. and Launier, R. (1978). Stress-Related Transactions between Person and Environment, *Perspectives in Interactional Psychology*, Springer, pp. 287–327.
URL: http://link.springer.com/10.1007/978-1-4613-3997-7_12 33

- Leard (2021a). Pearson's Product-Moment Correlation using SPSS Statistics.
URL: <https://statistics.laerd.com/spss-tutorials/pearsons-product-moment-correlation-using-spss-statistics.php> 191
- Leard (2021b). Spearman's Rank-Order Correlation using SPSS Statistics.
URL: <https://statistics.laerd.com/spss-tutorials/spearmans-rank-order-correlation-using-spss-statistics.php> 191
- Lee, S. H., Member, S., Ro, Y. M. and Member, S. (2016). Partial Matching of Facial Expression Sequence Using Over-Complete Transition Dictionary for Emotion Recognition, *7*(4): 389–408. 18, 25, 41, 49, 52, 58
- Levenson, R. W., Ekman, P. and Friesen, W. V. (1990). Voluntary Facial Action Generates Emotion Specific Autom. 30
- Lewandowski, C. M., Co-investigator, N. and Lewandowski, C. M. (2015). WHOQOL-Bref (versão em Português de Portugal), *The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference* 1: 1689–1699. 85
- Leyvand, T., Meekhof, C., Wei, Y. C., Sun, J. and Guo, B. (2011). Kinect identity: Technology and experience, *Computer* 44(4): 94–96. 41
- Lichtenstein, Antje; Oehme, A. K. S. J. T. (2008). *Comparing Two Emotion Models for Deriving Affective States from Physiological Data*.
URL: <http://www.ulb.tu-darmstadt.de/tocs/59142804.pdf> 21, 28, 29, 31, 43, 46, 49, 60
- LiKamWa, R., Liu, Y., Lane, N. D. and Zhong, L. (2013). MoodScope: Building a Mood Sensor from Smartphone Usage Patterns, *Proceeding of the 11th annual international conference on Mobile systems, applications, and services - MobiSys '13* p. 389.
URL: <http://dl.acm.org/citation.cfm?doid=2462456.2464449> 20, 36, 37, 38, 41, 46, 49, 51, 76
- Lim, N., Kim, E. K., Kim, H., Yang, E. and Lee, S. M. (2010). individual and work-related factors influencing burnout of mental health professionals: a meta-analysis, *Journal of Employment Counseling* 47(2): 86–96.
URL: <http://doi.wiley.com/10.1002/j.2161-1920.2010.tb00093.x> 63
- Linton Instrumentation (2011). Biopac Biopac MP100 - BIOPAC MP100 MP100 : Linton Instruments.
URL: http://www.lintoninst.co.uk/Products/tabid/63/ProdID/487/Language/en-US/CatID/127/Biopac_MP100_.aspx 43

- Lisetti, C. L. and Nasoz, F. (2004). Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals, *Journal of Applied Signal Processing* **11**: 1672–1687. [1](#), [3](#), [22](#), [28](#), [30](#), [31](#), [35](#), [42](#), [46](#), [77](#)
- Liu, Y., Sourina, O. and Nguyen, M. K. (2010). Real-time EEG-based human emotion recognition and visualization, *Proceedings - 2010 International Conference on Cyberworlds, CW 2010* pp. 262–269. [1](#), [17](#), [21](#), [24](#), [27](#), [43](#), [52](#), [53](#), [59](#), [60](#)
- Long, X., Fonseca, P., Foussier, J., Haakma, R. and Aarts, R. M. (2014). Sleep and wake classification with actigraphy and respiratory effort using dynamic warping, *IEEE Journal of Biomedical and Health Informatics* **18**(4): 1272–1284. [29](#)
- Lopes, P. N., Salovey, P. and Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships, *Personality and Individual Differences* **35**(3): 641–658. [23](#), [25](#)
- Luefeng, C., Min, W. U., Mengtian, Z., Jinhua, S. H. E. and Kaoru, H. (2016). Dynamic Emotion Understanding Using FCM Based SVR in Human-Robot Interaction, pp. 7064–7069. [3](#), [17](#)
- Lyons, R. G. (2004). *Understanding Digital Signal Processing*, Vol. 40.
URL: [http://books.google.co.uk/books?id=8osoAQAAAJ%5Cnhttp://doi.wiley.com/10.1002/1521-3773\(20010316\)40:6%3C9823::AID-ANIE9823%3E3.3.CO;2-C](http://books.google.co.uk/books?id=8osoAQAAAJ%5Cnhttp://doi.wiley.com/10.1002/1521-3773(20010316)40:6%3C9823::AID-ANIE9823%3E3.3.CO;2-C) [54](#)
- Madan, A., Cebrian, M., Moturu, S., Farrahi, K. and Pentland, A. S. (2012). Sensing the health state of a community, *IEEE Pervasive Computing* **11**(4): 36–45.
URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6072198 [40](#)
- Mandryk, R. L. and Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies, *International Journal of Human Computer Studies* **65**(4): 329–347. [1](#), [3](#), [16](#), [21](#), [31](#), [39](#), [44](#), [55](#), [60](#), [77](#), [83](#)
- Mandryk, R. L., Atkins, M. S. and Inkpen, K. M. (2006). A continuous and objective evaluation of emotional experience with interactive play environments, *Proceedings of the SIGCHI conference on Human Factors in computing systems CHI 06* **2**(October 2016): 1027. [31](#), [44](#)
- MAPIQ (2021). The ultimate smart office solutions guide.
URL: <https://www.mapiq.com/smart-office-solutions> [62](#)
- Marcano-Cedeno, A., Quintanilla-Dominguez, J., Cortina-Januchs, M. G. and Andina, D. (2010). Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network, *IECON 2010 - 36th Annual Conference*

- on *IEEE Industrial Electronics Society*, IEEE, pp. 2845–2850.
URL: <http://ieeexplore.ieee.org/document/5675075/> 50
- Markho, F. (2016). Major Influences in Households and Business Spaces – Wi-Fi , Telecommunication Masts Outputs and Electrical Pollution, (Epe): 20–22. 33
- Martinez-Nicolas, A., Ortiz-Tudela, E., Rol, M. A. and Madrid, J. A. (2013). Uncovering Different Masking Factors on Wrist Skin Temperature Rhythm in Free-Living Subjects, *PLoS ONE* 8(4): e61142.
URL: <http://dx.plos.org/10.1371/journal.pone.0061142> 38
- maths, R. (n.d.). Histograms and Cumulative Frequency.
URL: <https://revisionmaths.com/advanced-level-maths-revision/statistics/histograms-and-cumulative-frequency> 152
- Matiko, J. W., Beeby, S. P. and Tudor, J. (2014). Fuzzy logic based emotion classification, *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 4389–4393.
URL: <http://ieeexplore.ieee.org/document/6854431/> 20, 28, 51, 59, 60, 62
- Matlovic, T., Gaspar, P., Moro, R., Simko, J. and Bielikova, M. (2016). Emotions Detection Using Facial Expressions Recognition and EEG, *11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)* pp. 18–23. 19, 24, 28, 31, 36, 41, 43, 44, 46, 53, 56, 59, 138, 139
- Mehta, S. K., Super, D. M., Salvator, A., Fradley, L. G., Connuck, D. and Kaufman, E. S. (2002). Heart rate variability by triangular index in infants exposed prenatally to cocaine., *Annals of noninvasive electrocardiology : the official journal of the International Society for Holter and Noninvasive Electrocardiology, Inc* 7(4): 374–8.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/12431317> 59
- Merens, W., Willem Van der Does, A. J. and Spinhoven, P. (2007). The effects of serotonin manipulations on emotional information processing and mood, *Journal of Affective Disorders* 103(1-3): 43–62. 33
- Michie, S. and Williams, S. (2003). Reducing work related psychological ill health and sickness absence: A systematic literature review, *Occupational and Environmental Medicine* 60(1): 3–9. 63
- Microsoft (2017a). Kinect – Desenvolvimento de aplicações do Windows.
URL: <https://developer.microsoft.com/pt-pt/windows/kinect> 47

Microsoft (2017b). Stored Procedures (Database Engine).

URL: <https://docs.microsoft.com/en-us/sql/relational-databases/stored-procedures/stored-procedures-database-engine?view=sql-server-ver15> 147

Microsoft (2018). SQL Server Integration Services.

URL: <https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver15> 11, 115

Microsoft (n.d.a). Get & Transform in Excel (Power Query).

URL: <https://support.office.com/en-us/article/get-transform-in-excel-881c63c6-37c5-4ca2-b616-59e18d75b4de?omkt=en-001&ui=en-US&rs=en-001&ad=US> 147

Microsoft (n.d.b). Microsoft Excel.

URL: <https://www.microsoft.com/pt-pt/microsoft-365/excel> 11, 143, 145

Minitab (2019). What are quartiles?

URL: <https://support.minitab.com/en-us/minitab/18/help-and-how-to/graphs/how-to/boxplot/interpret-the-results/quartiles/> 150

Mishra, V., Pope, G., Lord, S., Lewia, S., Lowens, B., Caine, K., Sen, S., Halter, R. and Kotz, D. (2020). Continuous Detection of Physiological Stress with Commodity Hardware, *ACM Transactions on Computing for Healthcare* **1**(2): 1–30. 143

Mokhayeri, F., Akbarzadeh-T, M. R. and Toosizadeh, S. (2011). Mental stress detection using physiological signals based on soft computing techniques, *2011 18th Iranian Conference of Biomedical Engineering, ICBME 2011*. 21, 28, 47, 52

Monteiro, S. (2008). Optimismo e vinculação na transição para o ensino superior: relação com sintomatologia psicopatológica, bem-estar e rendimento académico, p. 456.

URL: <http://ria.ua.pt/bitstream/10773/1116/1/2010000583.pdf> 86

Monteiro, S., Tavares, J., Pereira, A. and Universidade de Aveiro, P. (2012). Adaptação portuguesa da escala de medida de manifestação de bem-estar psicológico com estudantes universitários – embeb, *Psicologia, Saúde & Doenças* **13**(1): 66–77. 76, 84, 86, 188

Morris, T. L. and Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight, *Biological Psychology* **42**(3): 343–360. 32

Moturu, S. T., Khayal, I., Aharony, N., Pan, W. and Pentland, A. S. (2011). Using Social Sensing to Understand the Links Between Sleep, Mood, and Sociability.

URL: <http://hd.media.mit.edu/tech-reports/TR-670.pdf> 34

Movisens (2020). movisens.

URL: <https://www.movisens.com/en/> 72, 83, 87, 88, 102, 107, 139

movisens GmbH (2018a). DataAnalyzer User manual. 87, 107, 140, 141, 162, 166, 269, 270

movisens GmbH (2018b). DataAnalyzer, Versuib 1.11.19 (29, April 2018).

URL: https://www.movisens.com/wp-content/downloads/DataAnalyzer_User_Manual.pdf 87

movisens GmbH (2018c). EdaMove 3.

URL: <https://www.movisens.com/en/products/edamove-3/> 7, 10, 83

movisens GmbH (2018d). EdaMove 3 – EDA and Activity Sensor.

URL: <https://www.movisens.com/en/products/eda-and-activity-sensor-move-3/> 9, 12, 72, 83, 85

movisens GmbH (2018e). movisens SensorManager.

URL: <https://www.movisens.com/en/products/sensormanager/> 87

Muaremi, A., Bexheti, A., Gravenhorst, F., Arnrich, B. and Tröster, G. (2014). Monitoring the Impact of Stress on the Sleep Patterns of Pilgrims using Wearable Sensors, *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* pp. 3–6. 28

Murad, U. and Malkawi, M. (2012). Artificial neuro fuzzy logic system for detecting human emotions, *IEEE CITS 2012 - 2012 International Conference on Computer, Information and Telecommunication Systems* pp. 1–13. 1, 3, 20, 23, 28, 29, 59, 71, 77

Murali, S., Rincon, F. and Atienza, D. (2015). A wearable device for physical and emotional health monitoring, *2015 Computing in Cardiology Conference (CinC)* pp. 121–124.

URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7408601> 19, 28, 29

Nasoz, F., Alvarez, K., Lisetti, C. L. and Finkelstein, N. (2004). Emotion recognition from physiological signals using wireless sensors for presence technologies, *Cognition, Technology & Work* 6(1): 4–14.

URL: <http://dx.doi.org/10.1007/s10111-003-0143-x><http://www.springerlink.com/content/cl7j99lrc> 68

Nawasalkar, R. K., Lawange, H. R., Gupta, S. D., Butey, P. K. and Email, W. S. (2013). Study of comparison of human bio-signals for emotion detection using HCI, *Interna-*

- tional Journal of Emerging Trends & Technology in Computer Science* **2**(2): 449–452.
1, 6, 16, 20, 23, 28, 29, 52, 82, 195
- Neurosky (2017). EEG Headsets — NeuroSky Store.
URL: <http://neurosky.com/biosensors/eeg-sensor/biosensors/> <http://neurosky.com/> 42
- Niemic, C. P., Kirk, A., Brown, W. and Ph, D. (2002). Studies of Emotion: A Theoretical and Emperical Review of Psychophysiological Studies of Emotion, *Journal of Undergraduate Research* pp. 15–18. 27
- Nisonger, T. E. (2008). The “80/20 Rule” and Core Journals, *The Serials Librarian* **55**(1-2): 62–84.
URL: <http://www.tandfonline.com/doi/abs/10.1080/03615260801970774> 152
- Nkurikiyeyezu, K., Shoji, K., Yokokubo, A. and Lopez, G. (2019). Thermal comfort and stress recognition in office environment, *HEALTHINF 2019 - 12th International Conference on Health Informatics, Proceedings; Part of 12th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2019 (Biostec)*: 256–263. 63, 65
- Nkurikiyeyezu, K., Yokokubo, A. and Lopez, G. (2019). Importance of individual differences in physiological-based stress recognition models, *2019 15th International Conference on Intelligent Environments (IE)* . 65
- Noldus (2017). Facial expression recognition software: FaceReader.
URL: <http://www.noldus.com/human-behavior-research/products/facereader> 41
- Nwe, L., Wei, S., Silva, D., Liyanage, C., Silva, D., Speech, L. C., Nwe, T. L., Member, S., Wei, F. S. and Member, S. (2001). Speech based emotion classification. 22, 23, 24, 58
- Omar, D. A. (2006). EMOÇÕES - Artigos.com.
URL: <http://www.artigos.com/artigos/13851-emocoas> 32
- OMS (2016). WHO — World Health Organization.
URL: http://www.who.int/features/factfiles/mental_health/en/ 4, 6
- OneZeroBit (2021). Notify and Fitness for MiBand.
URL: <https://mibandnotify.com/> 87
- Ouwerkerk, M., Pasveer, F. and Langereis, G. (2008). Unobtrusive sensing of psychophysiological parameters, *Probing Experience* pp. 163–193.
URL: http://link.springer.com/chapter/10.1007/978-1-4020-6593-4_15 40

Pais-Ribeiro, J. L. (1999). Escala de satisfação com o suporte social (ESSS), *Análise Psicológica* **3**(17): 547–558.

URL: http://www.scielo.oces.mctes.pt/scielo.php?pid=S0870-82311999000300010&script=sci_arttext%5Cnhttps://repositorio-aberto.up.pt/bitstream/10216/5544/2/80971.pdf 34, 68

Partala, T. and Surakka, V. (2003). Pupil size variation as an indication of affective processing, *International Journal of Human Computer Studies* **59**(1-2): 185–198. 22, 60

Partala, T., Surakka, V. and Vanhala, T. (2005). Person-independent estimation of emotional experiences from facial expressions, *Proceedings of the 10th international conference on Intelligent user interfaces - IUI '05* p. 246.

URL: <http://portal.acm.org/citation.cfm?doid=1040830.1040883> 21, 31, 60

Paschero, M., Del Vescovo, G., Benucci, L., Rizzi, A., Santello, M., Fabbri, G. and Mascioli, F. M. F. (2012). A real time classifier for emotion and stress recognition in a vehicle driver, *IEEE International Symposium on Industrial Electronics* pp. 1690–1695. 3, 17

Pearson, K. (1895). X. Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material, *Philosophical Transactions of the Royal Society of London. (A.)* **186**: 343–414.

URL: <https://royalsocietypublishing.org/doi/10.1098/rsta.1895.0010> 152

Pearson, K. (1901). Principal components analysis.

URL: http://www.botany.hawaii.edu/bot644/Manly_multivariate.pdf%5Cnpapers2://publication/uuid/82298A-4D21-A76D-CE842083AA61 50

Peirce, R. S., Frone, M. R., Russell, M., Cooper, M. L. and Mudar, P. (2000). A longitudinal model of social contact, social support, depression, and alcohol use, *Health Psychol* **19**(1): 28–38. 34

Pentland, A. (2005). Socially aware computation and communication, *Computer* **38**(3): 33–40. 39

Perdiz, J., Pires, G. and Nunes, U. J. (2017). Emotional state detection based on EMG and EOG biosignals: A short survey, *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)*, IEEE, pp. 1–4.

URL: <http://ieeexplore.ieee.org/document/7889451/> 18, 25, 31, 32, 50, 51, 52, 60

Phinyomark, A., Phukpattaranont, P. and Limsakul, C. (2012). Feature reduction and selection for EMG signal classification, *Expert Systems with Applications* **39**(8): 7420–7431.

URL: <http://www.sciencedirect.com/science/article/pii/S0957417412001200> 50, 51, 54

- Picard, R. and Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications, *Interacting with Computers* **14**(Spitz 1945): 141–169. [3](#)
- Picard, R. W. (1995). Affective Computing, *MIT press* (321): 1–16.
URL: papers3://publication/uuid/9C02FCAE-FE2E-4D2C-9707-766804777DC9 [3](#)
- Picard, R. W. (2000). *Affective Computing*, MIT Press. [39](#)
- Picard, R. W., Vyzas, E. and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(10): 1175–1191. [49](#)
- Pocinho, M. (2009). Estatística: teoria e exercícios passo a passo. Volume I, *Ismt I*: 1–82.
URL: http://docentes.ismt.pt/m_pocinho/Sebenta_estatistica_I.pdf [54](#)
- Poh, M.-Z., Swenson, N. C. and Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assesment of electrodermal activity, *IEEE Transactions on Biomedical Engineering* **57**(5): 1243–1252. [76](#), [77](#)
- Poole, A. and Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects, *Encyclopedia of Human-Computer Interaction* pp. 211–219. [27](#)
- Prasad, N. R., Almanza-Garcia, S. and Lu, T. T. (2009). Anomaly detection, *Computers, Materials and Continua* **14**(1): 1–22. [131](#)
- Prokoshyna, N., Szlichta, J., Chiang, F., Miller, R. J. and Srivastava, D. (2016). Combining quantitative and logical data cleaning, *Proceedings of the VLDB Endowment* **9**(4): 300–311. [142](#), [143](#)
- Rachuri, K. K., Mascolo, C., Rentfrow, P. J. and Longworth, C. (2010). EmotionSense : A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research, *International Studies* **10**: 281—290.
URL: <http://www.cl.cam.ac.uk/cm542/papers/Ubicomp10.pdf> [41](#), [44](#)
- Ramirez, R., Ramirez, R. and Vamvakousis, Z. (2015). Detecting emotion from EEG signals using the emotive epoc device Detecting Emotion from EEG Signals Using the Emotive Epoc Device, (October): 175–184. [43](#)
- Rani, P. and Sarkar, N. (2006). A new approach to implicit human-robot interaction using affective cues, *Mobile Robots, Towards New Applications* **3866113145**(December): 233–252.
URL: <http://cdn.intechweb.org/pdfs/57.pdf> [6](#), [17](#), [24](#), [25](#), [39](#), [68](#), [82](#)

Rani, P., Science, C., Sarkar, N., Smith, C. a. and Adams, J. a. (2003). Affective Communication for Implicit Human-, *Proceedings. ICRA'03. IEEE International Conference* pp. 2382–2387. [17](#)

Raschka, S. (2014). Linear Discriminant Analysis.

URL: http://sebastianraschka.com/Articles/2014_python_lda.html [50](#)

Raudonis, V. (2013). Evaluation of Human Emotion from Eye Motions, *International Journal of Advanced Computer Science and Applications* **4**(8): 79–84. [2](#), [16](#), [20](#), [22](#), [23](#), [31](#), [32](#), [33](#), [71](#), [195](#)

Ravichandran, R., Saba, E., Chen, K. Y., Goel, M., Gupta, S. and Patel, S. N. (2015). WiBreathe: Estimating respiration rate using wireless signals in natural settings in the home, *2015 IEEE International Conference on Pervasive Computing and Communications, PerCom 2015* **16**: 131–139. [29](#)

Recognition, F. E. (2009). Facial Expression Recognition, *Analysis* **77**(6): 1507–1513.

URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2861580&tool=pmcentrez&rendertype=full> [24](#)

Rehabilitation Institute of Chicago (2014). WHO Quality of Life-BREF (WHOQOL-BREF), *Rehabmeasures.Org* .

URL: <http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=937> [85](#)

Rehm, M., Bee, N. and André, E. (2008). Wave like an Egyptian: accelerometer based gesture recognition for culture specific interactions, *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction* pp. 13–22.

URL: <http://dl.acm.org/citation.cfm?id=1531514.1531517> [26](#)

Rekatsinas, T., Chuy, X., Ilyasy, I. F. and Ré, C. (2017). HoloClean: Holistic data repairs with probabilistic inference, *Proceedings of the VLDB Endowment* **10**(11): 1190–1201. [144](#)

Rençberoğlu, E. (2019). Fundamental Techniques of Feature Engineering for Machine Learning.

URL: <https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114> [109](#)

Rimas (2020). WHOQOL-Bref.

URL: <http://rimas.uc.pt/instrumentos/100/> [117](#), [141](#)

- Ritz, T., Steptoe, A., DeWilde, S. and Costa, M. (2000). Emotions and Stress Increase Respiratory Resistance in Asthma, *Psychosomatic Medicine* **62**(3): 401–412.
URL: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00006842-200005000-00014> 22, 29, 43, 46
- Robertson, I. and Cooper, C. (2011). *Well-Being Productivity and Happiness at Work*. 64
- Robles-Granda, P., Lin, S., Wu, X., Martinez, G. J., Mattingly, S. M., Moskal, E., Striegel, A., Chawla, N. V., D’Mello, S., Gregg, J., Nies, K., Mark, G., Grover, T., Campbell, A. T., Mirjafari, S., Saha, K., De Choudhury, M. and Dey, A. K. (2021). Jointly Predicting Job Performance, Personality, Cognitive Ability, Affect, and Well-Being, *IEEE Computational Intelligence Magazine* **16**(2): 46–61. 67
- Rosario, G. E., Rundensteiner, E. A., Brown, D. C. and Ward, M. O. (2003). Mapping nominal values to numbers for effective visualization, *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*. 143
- Rosario, G. E., Rundensteiner, E. A., Brown, D. C., Ward, M. O. and Huang, S. (2004). Mapping nominal values to numbers for effective visualization, *Information Visualization* . 143
- Rosmalen, J. G. M., Bos, E. H. and de Jonge, P. (2012). Validation of the Long-term Difficulties Inventory (LDI) and the List of Threatening Experiences (LTE) as measures of stress in epidemiological population-based cohort studies, *Psychological Medicine* **42**(12): 2599–2608.
URL: http://www.journals.cambridge.org/abstract_S0033291712000608 46
- Rothmann, S. (2008). Job satisfaction, occupational stress, burnout and work engagement as components of work-related wellbeing, *SA Journal of Industrial Psychology* **34**(3): 11–16. 63
- Russell, J. A. (1980). A circumplex model of affect., *Journal of Personality and Social Psychology* **39**(6): 1161–1178.
URL: <http://content.apa.org/journals/psp/39/6/1161> 45
- Russell, J. A., Weiss, A. and Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal., *Journal of Personality and Social Psychology* **57**(3): 493–502.
URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.57.3.493> 45, 65
- Saha, K., Bayraktaroglu, A. E., Campbell, A. T., Chawla, N. V., De Choudhury, M., D’Mello, S. K., Dey, A. K., Gao, G., Gregg, J. M., Jagannath, K., Mark, G., Martinez, G. J., Mattingly, S. M., Moskal, E., Sirigiri, A., Striegel, A. and Yoo, D. W. (2019). Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior

- and Wellbeing, *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, pp. 1–8.
URL: <https://dl.acm.org/doi/10.1145/3290607.3299065> 66
- Saha, S., Datta, S., Konar, A. and Janarthanan, R. (2014). A study on emotion recognition from body gestures using Kinect sensor, *Communications and Signal Processing (ICCSP), 2014 International Conference on* pp. 56–60.
URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6949798> 20, 26, 41, 52, 53
- Sanchez, W., Martinez, A. and Gonzalez, M. (2017). Towards Job Stress Recognition Based on Behavior and Physiological Features, pp. 311–322.
URL: http://link.springer.com/10.1007/978-3-319-67585-5_33 64
- Sano, A. and Eng, B. (2016). Measuring College Students' Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones, (2003): 182. 19, 30, 31, 32, 33, 35, 36, 37, 38, 41, 42, 45, 46, 47, 49, 51, 52, 53, 59, 60, 61, 76, 77, 82
- Sano, A. and Picard, R. W. (2013a). Recognition of sleep dependent memory consolidation with multi-modal sensor data, *2013 IEEE International Conference on Body Sensor Networks, BSN 2013* . 76
- Sano, A. and Picard, R. W. (2013b). Stress Recognition using Wearable Sensors and Mobile Phones, *Humaine Association Conference on Affective Computing and Intelligent Interaction Stress* pp. 671–676. 20, 35, 36, 37, 46, 50, 51, 53, 61, 76
- Sarabia, J. A., Rol, M. A., Mendiola, P. and Madrid, J. A. (2008). Circadian rhythm of wrist temperature in normal-living subjects. A candidate of new index of the circadian system, *Physiology and Behavior* **95**(4): 570–580.
URL: <http://dx.doi.org/10.1016/j.physbeh.2008.08.005> 30
- Sarason, I. G., Sarason, B. R., Potter, E. H. and Antoni, M. H. (1985). Life events, social support, and illness., *Psychosomatic medicine* **47**(2): 156–63.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/17992965> 34, 68
- Sarason, I. G., Sarason, B. R., Shearin, E. N. and Pierce, G. R. (1987). Assessing social support: The Social Support Questionnaire, *Journal of Social and Personal Relationships* **4**(1): 497–510. 34
- Scherer, K. R. (2005). What are emotion? And how can they be measured?, *Social Science Information Sur Les Sciences Sociales* **44**(4): 695–729. 1, 16
- Scherer, K. R. and Ekman, P. (1984). On the nature and function of emotion: A component process approach, in Press Psychology (ed.), *Approaches To Emotion*, Taylor &

- Francis, New Jersey, pp. 294–317.
URL: <https://books.google.com/books?hl=fr&lr=&id=k0mhAwAAQBAJ&pgis=1> 23
- Schulze, L., Renneberg, B. and Lobmaier, J. S. (2013). Gaze perception in social anxiety and social anxiety disorder, *Frontiers in Human Neuroscience* 7.
URL: <http://journal.frontiersin.org/article/10.3389/fnhum.2013.00872/abstract> 31
- Schumm, J., B??chlin, M., Setz, C., Arnrich, B., Roggen, D. and Tr??ster, G. (2008). Effect of movements on the electrodermal response after a startle event, *Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare 2008, PervasiveHealth* (June 2017): 315–318. 42
- Schumm, J., Setz, C., B??chlin, M., B??chler, M., Arnrich, B. and Tr??ster, G. (2010). Unobtrusive physiological monitoring in an airplane seat, *Personal and Ubiquitous Computing* 14(6): 541–550. 39, 41
- Sebe, N., Cohen, I., Gevers, T. and Huang, T. S. (2006). Emotion recognition based on joint visual and audio cues, *Proceedings - International Conference on Pattern Recognition* 1: 1136–1139. 21, 25
- Seligman, M. E. (2011). *Flourish: A visionary view understanding of happiness and well-being*, Nicholas Brealey Publishing, Sydney. 34, 68
- Setz, C., Arnrich, B., Schumm, J., Marca, R. L., Tr, G. and Ehlert, U. (2010). Using a Wearable EDA Device, *Technologynology* 14(2): 410–417. 21, 42, 49, 52, 65
- Shegokar, P. and Sircar, P. (2016). Continuous Wavelet Transform based Speech Emotion Recognition, (720). 25
- Shen, L. L., Bai, L. and Fairhurst, M. (2007). Gabor wavelets and General Discriminant Analysis for face identification and verification, *Image and Vision Computing* 25(5): 553–563. 58
- Silmee, P. (n.d.). Silmee Bar type. 42
- Silva, J. (2014). Portugal vai entrar para o grupo dos países "super-idosos" em 2020 - Observador.
URL: <http://observador.pt/2014/08/07/envelhecimento-da-populacao-moodys/> 30
- Sim, K.-b., Jang, I.-h. and Park, C.-h. (2007). The Development of Interactive Feature Selection and GA Feature Selection Method for Emotion Recognition, in B. Apolloni, R. J. Howlett and L. Jain (eds), *Database*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 73–81.
URL: http://dx.doi.org/10.1007/978-3-540-74829-8_10 23, 24

- Singh, S., Sharma, V., Jain, K. and Bhall, R. (2015). EDBL-algorithm for detection and analysis of emotion using body language, *Next Generation Computing Technologies (NGCT), 2015 1st International Conference on* (September): 820–823. [19](#), [23](#), [26](#), [51](#), [52](#), [53](#)
- Sinha, R. (1996). Multivariate Response Patterning of Fear and Anger, *Cognition & Emotion* **10**(2): 173–198. [22](#), [46](#)
- Sinha, R., Lovallo, W. R. and Parsons, O. a. (1992). Cardiovascular differentiation of emotions., *Psychosomatic medicine* **54**(4): 422–435. [22](#), [28](#), [29](#)
- Sloan, D. M. (2004). Emotion regulation in action: Emotional reactivity in experiential avoidance, *Behaviour Research and Therapy* **42**(11): 1257–1270. [31](#)
- Smeets, T., Dziobek, I. and Wolf, O. T. (2009). Social cognition under stress: Differential effects of stress-induced cortisol elevations in healthy young men and women, *Hormones and Behavior* **55**(4): 507–513. [32](#)
- Smith, S. W. (1997). The z-Transform, *The Scientist and Engineer's Guide to Digital Signal Processing* pp. 605–630. [51](#)
- Soh, M., Zarola, A., Palaiou, K. and Furnham, A. (2016). Work-related well-being, *Health Psychology Open* **3**(1). [64](#)
- Soleymani, M., Asghari-esfeden, S., Pantic, M. and Fu, Y. (2013). Continuous emotion detection using EEG signals and facial expressions, *IEEE Conference on Multimedia and Expo (ICME)* **231287**(231287): 3–8. [20](#), [25](#), [43](#), [53](#), [58](#), [59](#), [60](#), [68](#)
- Sonnenschein, M. (2007). *Sick with burnout : clarified through electronic diaries*.
URL: <http://dspace.library.uu.nl/bitstream/handle/1874/21596/?sequence=4> [63](#)
- Spielberger, C. (1983). Manual for the State-Trait Anxiety Inventory (STAI), *Consulting Psychologists Press* pp. 4–26. [45](#)
- Spielberger, C. D. (2010). State-Trait Anger Expression Inventory, *The Corsini Encyclopedia of Psychology* .
URL: <http://doi.wiley.com/10.1002/9780470479216.corpsy0942> [45](#)
- SQL Server Team (2017). Microsoft SQL Server 2017.
URL: <https://www.microsoft.com/pt-pt/sql-server/sql-server-2017> [113](#), [147](#)
- SQLite Consortium (2021). SQLite.
URL: <https://www.sqlite.org/index.html> [11](#), [84](#), [85](#)

- SR Research Ltd (2013). SR Research - EyeLink II.
URL: <http://www.sr-research.com/eyelinkII.html> 43
- Sroykham, W. and Wongsawat, Y. (2013). Effects of LED-backlit computer screen and emotional selfregulation on human melatonin production, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* pp. 1704–1707. 33
- Stephanie (2013a). What are Quartiles?
URL: <https://www.statisticshowto.com/what-are-quartiles/> 150
- Stephanie (2013b). What is the Midrange?
URL: <https://www.statisticshowto.com/midrange-definition-use/> 150
- Stephanie (2014). Dispersion / Measures of Dispersion: Definition.
URL: <https://www.statisticshowto.com/dispersion/> 150
- Stephanie (2017). Trimean (Tukey’s Trimean).
URL: <https://www.statisticshowto.com/trimean/> 150
- Stephens, C. L., Christie, I. C. and Friedman, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis, *Biological Psychology* **84**(3): 463–473.
URL: <http://dx.doi.org/10.1016/j.biopsycho.2010.03.014> 28
- Sukhobok, D., Nikolov, N. and Roman, D. (2018). Tabular Data Anomaly Patterns, *Proceedings - 2017 International Conference on Big Data Innovations and Applications, Innovate-Data 2017* **2018-Janua**: 25–34. 132
- Swayne, D. F., Temple Lang, D., Buja, A. and Cook, D. (2003). GGobi: Evolving from XGobi into an extensible framework for interactive data visualization, *Computational Statistics and Data Analysis* . 143
- Tableau Software LLC (2020). Tableau.
URL: <https://www.tableau.com/> 143
- Tae, K. H., Roh, Y., Oh, Y. H., Kim, H. and Whang, S. E. (2019). Data Cleaning for Accurate, Fair, and Robust Models, *Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning - DEEM’19*, ACM Press, New York, New York, USA, pp. 1–4.
URL: <http://dl.acm.org/citation.cfm?doid=3329486.3329493> 11, 144
- Tague, N. R. (2005). *The Quality Toolbox, Second Edition (e-book)*, Wisconsin. 152

- Tan, A. (2016). Hilbert-Huang Transform.
URL: <https://www.mathworks.com/matlabcentral/fileexchange/19681-hilbert-huang-transform?requestedDomain=www.mathworks.com> 51
- Taris, T. W. and Schreurs, P. J. (2009). Well-being and organizational performance: An organizational-level test of the happy-productive worker hypothesis, *Work & Stress* 23(2): 120–136.
URL: <http://www.tandfonline.com/doi/abs/10.1080/02678370903072555> 64
- Taylor, C. (2017). KURT Function for Kurtosis in Excel.
URL: <https://www.thoughtco.com/kurt-function-for-kurtosis-in-excel-3126625> 151
- Taylor, C. (2019). What Is the Midhinge?
URL: <https://www.thoughtco.com/what-is-the-midhinge-3126246> 150
- Taylor, C. (2020). Histogram Classes.
URL: <https://www.thoughtco.com/different-classes-of-histogram-3126343> 152
- Taylor, G. J., Ryan, D. and Bagby, M. (1985). Toward the Development of a New Self-Report Alexithymia Scale, *Psychotherapy and Psychosomatics* 44(4): 191–199.
URL: <http://www.karger.com/doi/10.1159/000287912> 45
- Taylor, H., Fieldman, G. and Altman, Y. (2008). E-mail at work: A cause for concern? The implications of the new communication technologies for health, wellbeing and productivity at work, *Journal of Organisational Transformation & Social Change* 5(2): 159–173.
URL: http://www.tandfonline.com/doi/full/10.1386/jots.5.2.159_1 64
- Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A. and Picard, R. (2015). Automatic Identification of Artifacts in Electrodermal Activity Data, *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* pp. 1934–1937. 30, 77
- TEA (2016). Captiv - Skilful and adptable software.
URL: <http://teaergo.com/site/en/products/manufacturers/tea/captiv-software> 47
- TEA (2017). Wireless sensor - GSR. 43
- TechFunology (2016). Xiaomi Mi Band 2 Review — TechFunology.com.
URL: <https://www.techfunology.com/electronics/smartwatches/xiaomi-mi-band-2-review/> 74, 85
- Terburg, D., Morgan, B. and van Honk, J. (2009). The testosterone-cortisol ratio: A hormonal marker for proneness to social aggression, *International Journal of Law and*

Psychiatry **32**(4): 216–223.

URL: <http://dx.doi.org/10.1016/j.ijlp.2009.04.008> 32

Thankachan, K. (2018). Automating anomaly detection for exploratory data analytics, *Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017*, IEEE, pp. 711–715.

URL: <https://ieeexplore.ieee.org/document/8365228/> 144

The Economist (2021). Wellbeing in the smart office.

URL: <https://empoweringspaces.economist.com/wellbeing-in-the-smart-office/> 62

Thenmozhi, R., Kabali, B. and Kapali, S. C. (2014). Effect of administrative stress on mental health in a group of office workers-Across sectional study, *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, IEEE, pp. 1–3.

URL: <http://ieeexplore.ieee.org/document/7043570/> 64

Thielbar, K. O., Triandafilou, K. M., Fischer, H. C., O’Toole, J. M., Listenberger, M. L., Ochoa, J. M., Stoykov, M. E. and Kamper, D. G. (2016). Benefits of using a voice and EMG-driven actuated glove to support occupational therapy for stroke survivors, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(3): 1–1.

URL: <http://ieeexplore.ieee.org/document/7470432/> 31

Thomas, P. B. M., Baltrušaitis, T., Robinson, P. and Vivian, A. J. (2016). The Cambridge Face Tracker: Accurate, Low Cost Measurement of Head Posture Using Computer Vision and Face Recognition Software, *Translational Vision Science & Technology* **5**(5).

URL: <http://tvst.arvojournals.org/article.aspx?doi=10.1167/tvst.5.5.8> 41

Thought Technology Ltd. (2016). FlexComp System with/ BioGraph Infiniti Software - T7555M.

URL: <http://thoughttechnology.com/index.php/flexcomp-system-with-biograph-infiniti-software-t7555m.html> 43

Tobii Technology AB (2015). tobii pro studio.

URL: <https://www.tobii.com/product-listing/tobii-pro-studio/>
<http://www.tobii.com/product-listing/tobii-pro-studio/> 47

Toppr (n.d.). Measures of Dispersion.

URL: <https://www.toppr.com/guides/business-mathematics-and-statistics/measures-of-central-tendency-and-dispersion/measure-of-dispersion/> 150

Toshiba (2015). Toshiba Silmee™ W20/W21.

URL: https://www.toshiba.co.jp/about/press/2015_08/pr_j1701.htm 42

- Tran, T. Q., Boring, R. L., Dudenhoefter, D. D., Hallbert, B. P., Keller, M. D. and Anderson, T. M. (2007). Advantages and disadvantages of physiological assessment for next generation control room design, *IEEE Conference on Human Factors and Power Plants* (1): 259–263. [72](#), [83](#)
- Truchot, D. (2012). Burnout and depression differential diagnosis: Back to basics, *EHPHS 2012* . [63](#)
- Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series, *Journal of Forecasting* . [132](#)
- Turan, C., Lam, K.-m. and He, X. (2015). Facial Expression Recognition with Emotion-Based Feature Fusion, (December): 1–6. [19](#), [24](#), [25](#), [41](#), [58](#)
- Umematsu, T., Sano, A. and Picard, R. W. (2021). Improving Wellbeing for Office Workers, *media.mit.edu/* .
URL: <https://www.media.mit.edu/projects/improving-well-being-for-office-workers/overview/> [63](#), [67](#)
- Universidade de Coimbra (n.d.). WHOQOL-BREF (versão portuguesa). [85](#)
- University of Ljubljana (2020). Orange.
URL: <https://orange.biolab.si/> [143](#), [145](#)
- University of Waikato (2020). WEKA - The workbench for machine learning.
URL: <https://www.cs.waikato.ac.nz/ml/weka/index.html> [143](#)
- Unwin, A., Theus, M. and Hofmann, H. (2006). *Graphics of Large Datasets*, Statistics and Computing, Springer New York, New York, NY.
URL: <http://link.springer.com/10.1007/0-387-37977-0> [143](#)
- Urges, A. N. (2021). Big Data To Good Data: Andrew Ng Urges ML Community To Be More Data-Centric And Less Model-Centric.
URL: <https://analyticsindiamag.com/big-data-to-good-data-andrew-ng-urges-ml-community-to-be-more-data-centric-and-less-model-centric/> [11](#), [109](#), [128](#), [138](#), [197](#)
- Van Der Vloed, G. and Berentsen, J. (2009). Measuring emotional wellbeing with a non-intrusive bed sensor, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **5727 LNCS(PART 2)**: 908–911. [68](#)
- van Eck, M., Berkhof, H., Nicolson, N. and Sulon, J. (2005). The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol., *Psychosomatic medicine* **58**(5): 447–458.

URL: [papers2://publication/uuid/7F45144B-27F3-4949-B97F-A0964D4D288C](#) 21, 32, 33, 36, 37, 46, 49, 61

Vaz Serra, A., Cristina Canavarro, M., Simões, M., Pereira, M., Gameiro, S., João Quartilho, M., Rijo, D., Carona, C. and Paredes, T. (2006). Estudos Psicométricos do Instrumento de Avaliação da Qualidade de Vida da Organização Mundial de Saúde (WHOQOL-Bref) para Português de Portugal, *Psiquiatria Clínica* **27**(271): 41–49. 85, 175

Vermun, K., Senapaty, M., Sankhla, A., Patnaik, P. and Routray, A. (2013). Gesture-based affective and cognitive states recognition using kinect for effective feedback during E-learning, *Proceedings - 2013 IEEE 5th International Conference on Technology for Education, T4E 2013* pp. 107–110. 20, 25, 26, 58

Vimalanathan, K. and Babu, T. R. (2014). The effect of indoor office environment on the work performance, health and well-being of office workers, *Journal of Environmental Health Science and Engineering* . 64

Visser, A. K. D., van Waarde, A., Willemsen, A. T. M., Bosker, F. J., Luiten, P. G. M., den Boer, J. A., Kema, I. P. and Dierckx, R. A. J. O. (2011). Measuring serotonin synthesis: from conventional methods to PET tracers and their (pre)clinical implications., *European journal of nuclear medicine and molecular imaging* **38**(3): 576–91.

URL: <http://www.ncbi.nlm.nih.gov/pubmed/21113591>
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3034914> 33

Volkovs, M., Chiang, F., Szlichta, J. and Miller, R. J. (2014). Continuous data cleaning, *Proceedings - International Conference on Data Engineering* pp. 244–255. 143

Vrana, S. R. (1993). The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG, *Psychophysiology* **30**(3): 279–286.

URL: <http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1993.tb03354.x/abstract> 22, 44, 52, 60

Vrana, S. R., Cuthbert, B. N. and Lang, P. J. (1986). Fear imagery and text processing. 77

Vrijkotte, T. G., van Doornen, L. J. and de Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability., *Hypertension (Dallas, Tex. : 1979)* **35**(4): 880–6. 22, 28, 29, 36, 37, 54, 55, 61

Wakamura, T. and Tokura, H. (2002). Circadian rhythm of rectal temperature in humans under different ambient temperature cycles, *Journal of Thermal Biology* **27**(5): 439–447.

URL: <http://linkinghub.elsevier.com/retrieve/pii/S0306456502000141> 38

- Wallston, B. S., Devellis, B. M. and Devellis, R. F. (1983). Social Support and Physical Health, *Health Psychology* **2**(4): 367–391.
URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-6133.2.4.367> 34
- Wang, H., Zhou, L. B. and Ying, Y. (2010). A novel approach for real time eye state detection in fatigue awareness system, *2010 IEEE Conference on Robotics, Automation and Mechatronics, RAM 2010* pp. 528–532. 24, 52, 55
- Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D. and Campbell, A. T. (2014). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* pp. 3–14. 28, 143
- Watson, D. and Clark, L. A. (1999). The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form, *Department of Psychological & Brain Sciences Publications* .
URL: http://ir.uiowa.edu/psychology_pubs/11 45
- Weiser, M. and Brown, J. S. (1996). Designing Calm Technology. 42
- WHO (2019). What is Moderate-intensity and Vigorous-intensity Physical Activity?
URL: https://www.who.int/dietphysicalactivity/physical_activity_intensity/en/ 270
- Who (2021a). Body mass index - BMI.
URL: <http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi> 141, 178
- Who (2021b). EQUATIONS FOR THE PREDICTION OF BASAL METABOLIC RATE.
URL: <http://www.fao.org/3/AA040E/AA040E15.htm#an1> 141
- Who (n.d.). WHO Quality of Life-BREF (WHOQOL-BREF).
URL: http://www.who.int/substance_abuse/research_tools/whoqolbref/en/ 85
- Wijnen, B. F., Lokkerbol, J., Boot, C., Havermans, B. M., van der Beek, A. J. and Smit, F. (2020). Implementing interventions to reduce work-related stress among health-care workers: an investment appraisal from the employer’s perspective, *International Archives of Occupational and Environmental Health* **93**(1): 123–132.
URL: <https://doi.org/10.1007/s00420-019-01471-y> 64
- Wilkinson, L. (2006). Revising the Pareto Chart, *The American Statistician* **60**(4): 332–334.
URL: <http://www.tandfonline.com/doi/abs/10.1198/000313006X152243> 152

- Winton, W. M., Putnam, L. E. and Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion, *Journal of Experimental Social Psychology* **20**(3): 195–216. [77](#)
- Wolfram Alpha (2013). Wolfram—Alpha: Computational Knowledge Engine.
URL: <http://www.wolframalpha.com/> [40](#)
- Wu, W. H., Batalin, M. A., Au, L. K., Bui, A. A. T. and Kaiser, W. J. (2007). Context-aware sensing of physiological signals, *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings* pp. 5271–5275. [82](#)
- Wu, W. H., Bui, A. A. T., Batalin, M. A., Au, L. K., Binney, J. D. and Kaiser, W. J. (2008). MEDIC: Medical embedded device for individualized care, *Artificial Intelligence in Medicine* **42**(2): 137–152.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S0933365707001510> [72](#), [83](#)
- XiaoMi (n.d.). Mi Band 2 - Mi Global Home.
URL: <https://www.mi.com/en/miband2/> [7](#), [10](#), [74](#), [85](#)
- Yang, S. and Bhanu, B. (2011). Facial expression recognition using emotion avatar image, *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011* pp. 866–871. [20](#), [25](#), [58](#)
- Yıldırım, A. A., Özdoğan, C. and Watson, D. (2014). Parallel Data Reduction Techniques for Big Datasets, *Big Data Management, Technologies, and Applications*, IGI Global, pp. 72–93.
URL: <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-4699-5.ch004> [50](#)
- Zaman, S., Wesley, A., Silva, D. R. D. C., Buddharaju, P., Akbar, F., Gao, G., Mark, G., Gutierrez-Osuna, R. and Pavlidis, I. (2019). Stress and productivity patterns of interrupted, synergistic, and antagonistic office activities, *Scientific data* **6**(1): 264.
URL: <http://dx.doi.org/10.1038/s41597-019-0249-5> [64](#), [65](#)
- Zapf, D. (1993). Stress-oriented Analysis of Computerized Office Work, *European Work and Organizational Psychologist* **3**(2): 85–100. [64](#)
- Zapf, D. (2002). Emotion work and psychological well-being. A review of the literature and some conceptual considerations, *Human Resource Management Review* **12**(2): 237–268. [63](#)
- Zenonos, A., Khan, A., Kalogridis, G., Vatsikas, S., Lewis, T. and Sooriyabandara, M. (2016). HealthyOffice: Mood recognition at work using smartphones and wearable

- sensors, *2016 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops 2016* . 19, 28, 29, 31, 36, 42, 53, 59, 60
- Zephyr Technology (2012). BioHarness 3.0 User Manual, *zephyranywhere.com* . 42
- Zhai, J. and Barreto, A. (2006). Stress detection in computer users based on digital signal processing of noninvasive physiological variables, *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE (Ic)*: 1355–1358. 3, 17, 21, 28
- Zhang, A., Song, S., Wang, J. and Yu, P. S. (2017). Time series data cleaning, *Proceedings of the VLDB Endowment* **10**(10): 1046–1057.
URL: <http://dl.acm.org/citation.cfm?doid=3115404.3115410> 144
- Zhang, M. and Sawchuk, A. A. (2013). Human daily activity recognition with sparse representation using wearable sensors, *IEEE Journal of Biomedical and Health Informatics* **17**(3): 553–560. 32, 42
- Zhang, Z., Girard, J. M., Wu, Y., Zhang, X., Liu, P., Ciftci, U., Canavan, S., Reale, M., Horowitz, A., Yang, H., Cohn, J. F., Ji, Q. and Yin, L. (2016). Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 3438–3446. 2, 19, 25, 28, 29, 31, 36, 44, 46, 51, 54, 55, 67
- Zhao, M., Adib, F. and Katabi, D. (2016). Emotion recognition using wireless signals, *22nd Annual International Conference on Mobile Computing and Networking*, <http://home.jelena.org/Downloads/fneng-07-00021.pdf> orking, *MobiCom 2016* (CONF CODENUMBER): 95–108.
URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84994165777&partnerID=40&md5=6a458969d083b4780979a1946a0aef2> 19, 28, 29, 41, 49, 52, 53, 54, 60
- Zhao, N., Azaria, A. and Paradiso, J. (2017). Mediated Atmospheres: A Multimodal Mediated Work Environment, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**(2): 1–23. 65
- Züger, M., Müller, S. C., Meyer, A. N. and Fritz, T. (2018). Sensing interruptibility in the office: A field study on the use of biometric and computer interaction sensors, *Conference on Human Factors in Computing Systems - Proceedings*. 63