

Douglas Farias Cordeiro

Posicionamentos Jornalísticos sobre a Vacinação da COVID-19: Estudo comparativo
entre Brasil e Portugal

Relatório apresentado à Universidade Fernando Pessoa como parte dos requisitos para cumprimento do programa de pós-doutoramento em Jornalismo, sob a supervisão do Prof. Doutor Jorge Pedro Sousa.

Universidade Fernando Pessoa
Porto 2023

RESUMO

O tema da investigação é a identificação do posicionamento de portais de notícias digitais em peças jornalísticas frente a questões políticas relacionadas às campanhas de vacinação contra a COVID-19 no ano de 2021 no Brasil e em Portugal. A pesquisa parte da hipótese de reverberação de posicionamento político por parte da imprensa, no caso específico, de canais de notícias na Internet, para com os conteúdos publicados em peças jornalísticas frente às campanhas e estratégias de enfrentamento da COVID-19 por meio da vacinação. Se realiza um estudo de caso comparativo entre Brasil e Portugal, no qual são elencados dez portais de notícias para cada um dos países. A seleção dos portais segue um critério de popularidade baseado em indicadores disponíveis em motores de busca. A partir disso é construído um banco de dados contendo as peças jornalísticas, as quais foram recuperadas por meio do uso de solução de raspagem de dados (*web scraping*). São utilizadas soluções computacionais de processamento de linguagem natural (PLN) e inteligência artificial para rotulagem dos textos em relação aos posicionamentos apresentados (análise de sentimentos). Além disso, também são realizadas análises utilizando PLN quanto às temáticas que permeiam as peças jornalísticas.

ABSTRACT

This research aims to identify the positioning of digital news portals in journalistic pieces in the face of political issues related to vaccination campaigns against COVID-19 in the year 2021 in Brazil and Portugal. The research considers the hypothesis of reverberation of political positioning by the press, in the specific case of news channels on the Internet, towards the contents published in journalistic pieces in the face of the campaigns and strategies to face COVID-19 through vaccination. A comparative case study is carried out between Brazil and Portugal, in which ten news portals for each country are listed. The selection of portals follows a criterion of popularity based on indicators available in search engines. From this, a database is built containing the journalistic pieces, which were retrieved through the use of a data scraping solution (web scraping). Natural language processing (NLP) and artificial intelligence computational solutions are used to label the texts in relation to the positions presented (sentiment analysis). In addition, analyzes are also carried out using PLN regarding the themes that permeate the journalistic pieces.

DEDICATÓRIA

De boas palavras transborda o meu coração. Ao Rei consagro o que compus.

AGRADECIMENTOS

Na presença de Cristo Jesus, rendo graças a meu Deus pela vida de cada um de vocês que estiveram comigo em todo esse tempo. Graça e paz.

ÍNDICE

INTRODUÇÃO	12
CAPÍTULO 1 - PERCURSOS METODOLÓGICOS	14
1.1. KDD	14
1.2. Conjunto de Dados	17
1.3. Coleta de dados	24
1.4. Pré-processamento e transformação de dados	26
1.5. Processamento de linguagem natural	27
1.6. Classificação temática	29
1.7. Análise de sentimentos	30
CAPÍTULO 2 – JORNALISMO EM TEMPOS DE COVID-19	33
2.1 Importância do jornalismo na saúde pública	34
2.2 Pluralidade de visões e discursos	35
CAPÍTULO 3 – VACINAÇÃO, JORNALISMO E POSICIONAMENTOS	37
3.1 Análise exploratória das peças noticiosas	37
3.2 Temas e significados	42
3.3 Posicionamentos e produção de sentidos	57
CONCLUSÃO	63
REFERÊNCIAS BIBLIOGRÁFICAS	64

LISTA DE ABREVIATURAS

KDD	10
LDA	26
PLN	24
APCT	18
BERT	21
LeIA	28

LISTA DE FIGURAS

Figura 1 - Arquitetura geral da solução	16
Figura 2 - Captura de tela de busca parametrizada no motor de busca Google	26
Figura 3 - Exemplo de ocorrência de menções a outras notícias	28

LISTA DE GRÁFICOS

Gráfico 1 - Variação de popularidade de sites brasileiros a partir do Google Trends	18
Gráfico 2 - Variação de popularidade de sites portugueses a partir do Google Trends	22
Gráfico 3 - Distribuição quantitativa percentual de notícias capturadas (Brasil)	38
Gráfico 4 - Série temporal de notícias capturadas (Brasil)	39
Gráfico 5 - Nuvem de palavras segmentadas por jornal (Brasil)	40
Gráfico 6 - Distribuição quantitativa percentual de notícias capturadas (Portugal)	41
Gráfico 7 - Série temporal de notícias capturadas (Portugal)	41
Gráfico 8 - Nuvem de palavras segmentadas por jornal (Portugal)	42
Gráfico 9 - Análise de coerência de tópicos (Brasil)	43
Gráfico 10 - Distribuição temática de tópicos (Brasil)	44
Gráfico 11 - Nuvens de palavras segmentadas por tópicos (Brasil)	45
Gráfico 12 - Análise de coerência de tópicos (Portugal)	51
Gráfico 13 - Distribuição temática de tópicos (Portugal)	52
Gráfico 14 - Nuvens de palavras segmentadas por tópicos (Portugal)	52
Gráfico 15 - Infográfico de posicionamento jornalístico (Brasil)	58
Gráfico 16 - Variações de sentimentos por jornal (Brasil)	59
Gráfico 17 - Infográfico de posicionamento jornalístico (Portugal)	60
Gráfico 18 - Variações de sentimentos por jornal (Portugal)	61

LISTA DE QUADROS

Quadro 1 - Jornais brasileiros considerados no corpo de pesquisa	17
Quadro 2 - Jornais portugueses considerados no corpo de pesquisa	21
Quadro 3 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 1 (Brasil)	45
Quadro 4 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 2 (Brasil)	47
Quadro 5 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 3 (Brasil)	48
Quadro 6 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 4 (Brasil)	49
Quadro 7 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 5 (Brasil)	50
Quadro 8 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 1 (Portugal)	53
Quadro 9 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 2 (Portugal)	54
Quadro 10 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 3 (Portugal)	55
Quadro 11 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 4 (Portugal)	56

LISTA DE TABELAS

Tabela 1 - Ranking de popularidade de motores de busca em março de 2023	25
Tabela 2 - Sentimento médio percentual (Brasil)	58
Tabela 3 - Sentimento médio percentual (Portugal)	60

INTRODUÇÃO

O Sars-Cov-2 (novo coronavírus), responsável pela doença denominada de COVID-19, foi responsável por uma das maiores crises sanitárias, em nível global, da idade contemporânea. Esse cenário exigiu a aplicação de intervenções não farmacológicas por parte das autoridades sanitárias, tais como o isolamento social e o uso de máscaras (Tabai, Santos e Coqueiro, 2022). Junto a todas estas questões, o fenômeno da desinformação, a alta circulação de *fake news* e a infodemia influenciaram fortemente a leitura situacional por parte da sociedade e, conseqüentemente, o posicionamento quanto às medidas de enfrentamento à COVID-19 (Sullcahuamán *et alli.*, 2022).

A democratização de acesso às tecnologias de informação e comunicação, assim como os avanços tecnológicos alcançados no âmbito da Internet, resultaram na imersão de um cenário onde os meios tradicionais de comunicação deixaram ser fontes exclusivas de informação, entrando em cena as redes sociais, os aplicativos de mensagens e os inumeráveis sítios que fornecem informações sem necessariamente uma garantia de fonte (Catalina-García, Sousa e Sousa, 2019). Neste contexto, o trabalho desempenhado por jornalistas tornou-se fundamental, tanto para a circulação de informações essenciais ao enfrentamento da COVID-19, quanto no combate à desinformação. Entretanto, é importante destacar que diversos estudos apontam alterações e adequações da produção jornalística frente às necessidades decorrentes da pandemia (Lopes *et alli.*, 2021), impactando também na maneira com que as informações são noticiadas. Para além disso, há também de se pontuar as relações intrínsecas entre posicionamentos e discursos adotados pelos veículos jornalísticos, o que, a certo modo, podem interferir no significado e na produção de sentidos decorrentes do consumo das peças noticiosas.

Neste cenário, se torna de grande relevância a compreensão destes possíveis posicionamentos dos veículos de notícias quanto às estratégias de enfrentamento da COVID-19 por parte das autoridades sanitárias, no sentido de mapear e identificar as dinâmicas entre notícias e sociedade, no sentido da importância atribuída por parte dos canais aos temas abordados, assim como na convergência dos veículos jornalísticos enquanto fontes informacionais à população.

De maneira geral, a adoção de mecanismos de enfrentamento contra a COVID-19, principalmente por meio de campanhas vacinais, foi uma prioridade estratégica por parte de autoridades sanitárias. Neste cenário, se busca realizar um mapeamento da mídia jornalística, brasileira e portuguesa, através de um estudo de caso comparativo, sobre um conjunto de peças jornalísticas, veiculadas em formato digital, que referenciem a pandemia da COVID-19 e as campanhas vacinais, de maneira a identificar possíveis posicionamentos, assim como a produção de sentidos, essencialmente no contexto de menções a autoridades políticas. As análises compreendem a utilização de instrumentos metodológicos baseados no uso de soluções computacionais, tanto no que se refere à obtenção de dados quanto à própria geração de resultados que possam apoiar o processo de identificação dos conteúdos veiculados e os possíveis posicionamentos.

Caracterizado como um documento de domínio público, leva-se em conta que o jornal, a partir de sua multiplicidade e diversidade de falas, reflete visões sociais, culturais e políticas, as quais, em certa forma, vão de encontro a visões de certos grupos ou mesmo se enquadram enquanto recurso propagandístico. A partir disso, a investigação também busca estabelecer uma interdisciplinarização entre os estudos comunicacionais no contexto do problema a ser investigado com soluções inovadoras de inteligência artificial e mineração de dados, de maneira a romper barreiras principalmente na análise e avaliação de grandes volumes de dados, neste caso, peças noticiosas.

CAPÍTULO 1 - PERCURSOS METODOLÓGICOS

A presente investigação possui uma abordagem quantitativa-qualitativa, sendo o viés quantitativo voltado, principalmente, à geração de resultados a partir da exploração de métodos de análise estatística, e o viés qualitativo com foco na identificação das informações circulantes e dos sentimentos associados a tais informações por parte dos portais de notícias, analisados na perspectiva das abordagens de comunicação e nas relações com possíveis posicionamentos sobre as estratégias de enfrentamento da COVID-19 por parte das autoridades sanitárias. As análises são instrumentalizadas por meio do processo Descoberta do Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* - KDD), proposto por Fayyad, Shapiro-Piatetsky e Smith (1996), com aplicação de soluções de classificação temática e análise de sentimentos.

1.1. KDD

No contexto das investigações na área de comunicação, embora ainda existam desafios quanto ao uso de soluções computacionais inteligentes, a inovação em termos metodológicos têm sido explorada em temas que englobam questões que vão desde a análise de redes sociais (Cordeiro *et alli.*, 2022; Toraman, Şahinuç e Yilmaz, 2022) à aplicações voltadas à compreensão da circulação de informação por parte de peças noticiosas (Taj, Shaikh e Fatemah Meghji, 2019; Cordeiro *et alli.*, 2021). As necessidades que orbitam o uso de soluções computacionais emergem a partir das perspectivas oriundas do cenário *big data* ao qual o jornalismo digital assumiu ao longo das primeiras décadas do século XXI (Fernandes, Moro e Cortez, 2023).

Big Data, de forma geral, pode ser descrito como um fenômeno para o qual atividades, processos e rotinas são envoltas à geração e ao consumo de dados (Sagiroglu e Sinanc, 2013). Os avanços tecnológicos alcançados no contexto das tecnologias de

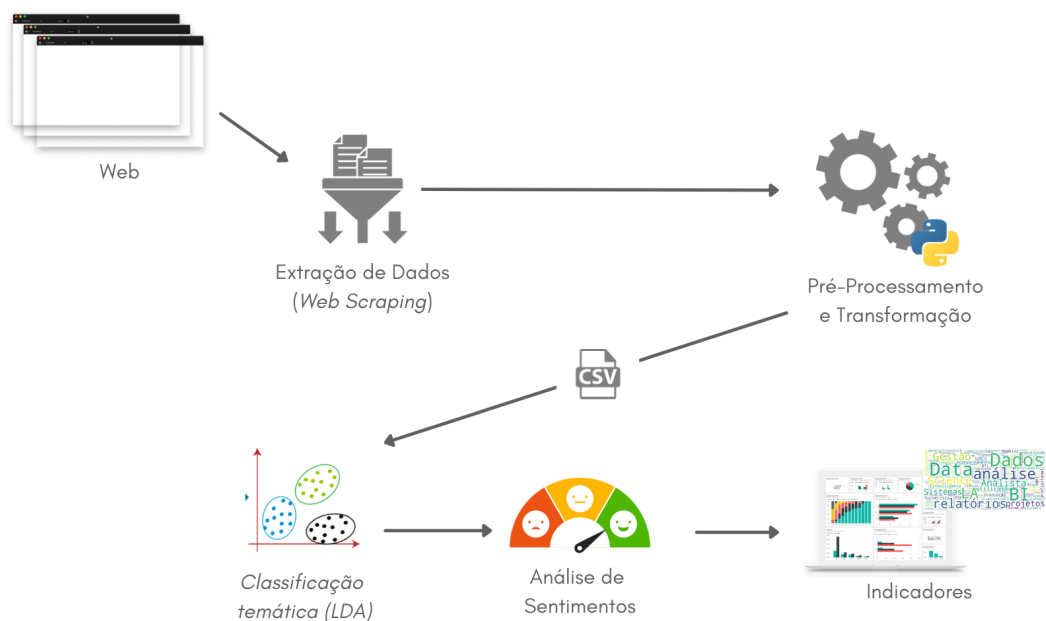
informação e comunicação, amparados por uma alta capacidade de processamento e armazenamento de dados, possibilitaram um ambiente onde o uso de aplicações digitais se tornou algo comum e necessário. A digitalização observada nos diversos setores da sociedade passaram a ser moldadas para lidar com ecossistemas massivos de dados, os quais representam oportunidades de geração de informação e descoberta do conhecimento. Cabe ressaltar que, conforme pontuado por Nguyen (2018), investigações em contexto de *big data* devem ser pautadas por características fundamentais de grandes conjuntos de dados, conhecidas como os 5 V's do *big data*:

- **Valor:** está relacionado à relevância e importância dos dados para o contexto de análise de dados do problema investigado. O valor dos dados está associado à assertividade, utilidade e significado das informações geradas frente às necessidades de análise;
- **Variedade:** está associada à diversidade de fontes de dados disponíveis. A utilização de dados pode requerer diferentes fontes de dados, os quais podem ser estruturados ou não-estruturados, assim como ser apresentados em diferentes formatos (dados numéricos, textos, imagens, som e vídeo);
- **Velocidade:** está relacionada às necessidades de acesso e processamento de dados em soluções de análise de dados. Determinados tipos de projetos demandam soluções que possibilitem a extração de grandes volumes de dados em tempo reduzido ou em tempo real, com sucessiva geração de informações;
- **Veracidade:** está relacionada às necessidades de qualidade de dados. A imprecisão de dados pode gerar informações incorretas ou irrelevantes, podendo gerar um cenário de conclusões incorretas sobre determinados assuntos;
- **Volume:** está relacionado à quantidade de dados gerados, coletados e processados. Fora de um contexto *big data* é comum a análise de dados a partir de conjuntos amostrais limitados, porém em um cenário *big data* é possível, por meio do uso de tecnologias computacionais inteligentes, explorar volumes massivos de dados, proporcionando resultados mais profundos e significativos.

A interdisciplinaridade entre big data, ciência de dados e inteligência artificial é algo amplamente explorado no contexto acadêmico e no contexto empresarial. Neste sentido, estratégias voltadas à geração de informações por meio da aplicação de algoritmos inteligentes são exploradas juntamente com soluções de visualização de dados, possibilitando interpretação e comunicação de resultados. Sistematizar estes procedimentos é algo fundamental para garantir resultados mais assertivos e alinhados com a resolução de problemas de investigação, o que pode ser realizado por meio do uso de processos de descoberta do conhecimento, como é o caso do KDD.

O KDD é composto por cinco etapas principais: seleção de dados, pré-processamento, transformação, mineração de dados e avaliação/interpretação. Cada uma das etapas é realizada de maneira a fornecer insumos para a etapa subsequente, permitindo, além disso, a realização de verificações e análises que busquem garantir a qualidade dos resultados gerados. Esta interatividade é considerada como uma das grandes vantagens do KDD (Vieira e Cordeiro, 2023). A Figura 1 apresenta a arquitetura geral da solução explorada para a presente investigação. Devido às características específicas dos dados de investigação, provenientes de portais digitais de notícias, foram realizadas adaptações pontuais nas tarefas do KDD, essencialmente considerando, neste caso, um paralelismo nas tarefas de pré-processamento e transformação de dados.

Figura 1 - Arquitetura geral da solução



Fonte: autor.

1.2. Conjunto de Dados

Para a composição da base de dados foram selecionados dez diferentes jornais com veiculação digital de notícias, sendo cinco brasileiros e cinco portugueses. A escolha foi baseada na relevância em termos de popularidade de busca e acesso, a partir da observação de dados na ferramenta Google Trends considerando os anos de 2021 e 2022. O Google Trends é uma ferramenta que possui forma de acesso aberto e gratuito e possibilita o monitoramento de interesse e popularidade de termos ou sites específicos para um período passível de delimitação. Além disso, a ferramenta ainda possui a vantagem de possibilitar uma filtragem associada a regiões, permitindo uma personificação de análises de interesse. No caso do presente estudo, foram realizadas avaliações por meio da ferramenta para sites de notícias no contexto de Brasil e Portugal.

Para o Brasil foram considerados os sites de notícias elencados no Quadro 1, os quais incluem plataformas provenientes de jornais que originalmente foram criados com foco na mídia impressa e plataformas nativas digitais. Todos os sites possuem mecanismo próprio de busca indexada direta, com exceção do jornal Poder360 que utiliza solução integrada do buscador Google. É importante ressaltar tal característica pelo fato de que a identificação das notícias que abordam a temática de interesse da presente investigação necessitam de identificação prévia para que possam ser recuperadas por solução computacional de extração de dados.

Quadro 1 - Jornais brasileiros considerados no corpo de pesquisa

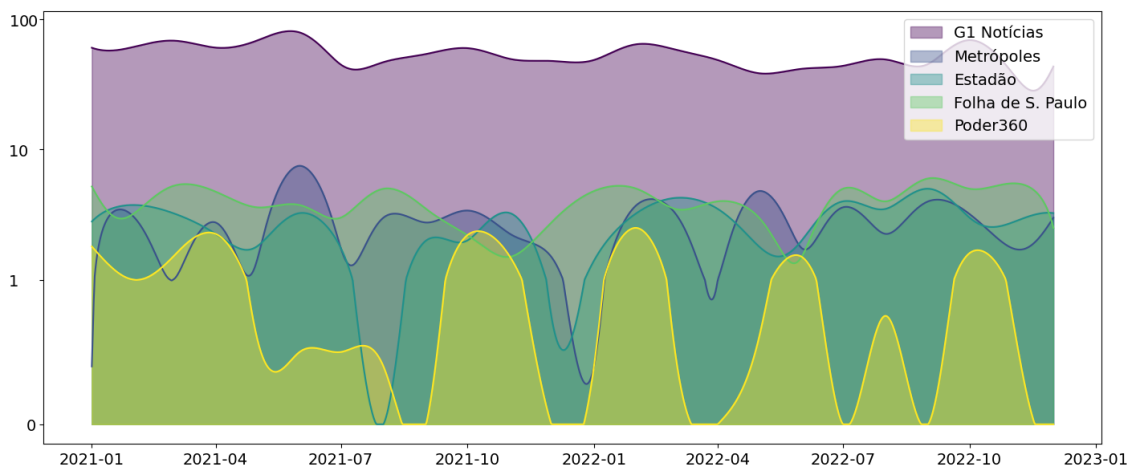
Jornal	Endereço eletrônico	Acessibilidade informacional
Estadão	https://www.estadao.com.br/	Ferramenta de pesquisa própria
Folha	https://www.folha.uol.com.br/	Ferramenta de pesquisa própria
G1 Notícias	https://g1.globo.com/	Ferramenta de pesquisa própria
Metrópoles	https://www.metropoles.com/	Ferramenta de pesquisa

		própria
Poder360	https://www.poder360.com.br/	Ferramenta de pesquisa própria limitada (consolidação de dados via Google)

Fonte: dados de pesquisa.

A partir dos resultados observados junto à ferramenta Google Trends para o período que compreende os anos de 2021 e 2022 foi gerada a visualização apresentada no Gráfico 1 referente à popularidade de busca dos jornais brasileiros considerados. O portal de notícias G1 apresenta popularidade notavelmente superior aos demais jornais digitais, o que tornou necessária a utilização de estratégia de visualização baseada em segmentação descontínua de eixo, isto é, variação proporcional entre segmentos delimitados no gráfico. Os jornais Estadão, Folha de S. Paulo, Metrôpoles e Poder360 possuem popularidade relativamente similar na ferramenta de monitoramento, com pequenas variações observadas em períodos específicos, o que não desqualifica a importância dos mesmos em relação ao interesse e acesso.

Gráfico 1 - Variação de popularidade de sites brasileiros a partir do Google Trends



Fonte: adaptado de Google Trends, 2023.

O jornal Estadão, conhecido também como O Estado de São Paulo, foi fundado na segunda metade do século XIX, tendo o seu primeiro exemplar publicado em 4 de janeiro de 1875. À época o jornal possuía o nome A Província de São Paulo, o qual foi mantido até o final do período monárquico brasileiro, quando passou a se chamar O

Estado de São Paulo (Berbel, 2017). O jornal é historicamente considerado como exemplo de inovação desde seus primeiros anos, sendo conhecido pelo seu pioneirismo na venda avulsa de exemplares em 1876 (Barsotti, 2015), assim como no lançamento de uma edição extra noturna, em 1915, a qual foi apelidado de Estadinho pelo seu tamanho reduzido e culminou no surgimento do nome Estadão ao jornal (Santos, 2018). Em 20 de fevereiro de 1995 o Estadão passou a ter uma versão digital presente no portal americano World News, e em 2000 inaugurou oficialmente o domínio *estadao.com.br* na Internet (Gonçalves, Capoano, 2017). Em março de 2021 o Estadão alcançou números recordes em audiência na Internet, considerando sua penetração através de seu site e de plataformas de redes sociais, ultrapassando o número de 47 milhões de usuários únicos (Estadão, 2021).

O jornal A Folha de S. Paulo remete ao início da década de 1920, quando um grupo de jornalistas criou o jornal Folha da Noite, com a proposta de ser oposição ao então principal jornal paulista da época, o Estadão, através de um viés mais opinativo e próximo da população urbana (Albuquerque, 2018). O sucesso do jornal levou ao lançamento de outros dois jornais adjacentes: o Folha da Manhã e, posteriormente, o Folha da Tarde. Historicamente os três jornais ficaram conhecidos como defensores do progresso e voz de oposição e enfrentamento a problemas recorrentes no Brasil, tais como a corrupção e o crime organizado. Embora populares, os jornais tiveram suas primeiras décadas marcadas com problemas administrativos e econômicos, o que levou à fusão dos três jornais em um único, no início da década de 1960, o qual foi batizado como Folha de S. Paulo (Souza, 2018). No âmbito digital, o jornal Folha de S. Paulo ficou conhecido por seu pioneirismo, se tornando, em 1983, o primeiro jornal da América do Sul a possuir uma redação informatizada (Costa, 2001). Assim como o Estadão, o jornal Folha de S. Paulo inaugurou suas atividades na Internet em 1995, alcançando em 2023 aproximadamente 28 milhões de visitantes únicos e uma média de 200 milhões de visitas mensais (Folha de S. Paulo, 2023).

No contexto do ciberespaço, o jornal Poder360, considerado como o jornal nativo digital mais antigo em operação no Brasil, tem sua criação remetida ao início do século XXI, quando, especificamente no ano 2001, o jornalista Fernando Rodrigues, então da Folha de S. Paulo, deu início a um projeto jornalístico no formato de blogue com foco em assuntos políticos (Bittencourt, 2018). A página se tornou o portal de

notícias Poder360 em 2016, com disponibilização de conteúdos em formato aberto e livre, tendo seu faturamento proveniente da comercialização de newsletters, publicidade e serviços de pesquisas de opinião (Belda, Gonçalves, 2021). Em 2022 o jornal Poder360 alcançou o total de 83,4 milhões de visitantes únicos, com 354,5 milhões de páginas visualizadas (Poder360, 2023).

Diferente dos jornais apresentados anteriormente, o G1 nasceu com a proposta de ser um portal de notícias digital, focado na integração e disponibilização de conteúdos jornalísticos provenientes de diferentes empresas do Grupo Globo (Hummel, Volcan, 2017). O G1 foi lançado em 2006, em uma época em que a Internet, devido a limitações em termos de velocidade de acesso, assim como com relação a questões de acessibilidade a equipamentos de tecnologia de informação e comunicação, acabava por gerar um cenário de limitação ao jornalismo digital, que em sua maioria se revelava como um recorte da mídia impressa. O modelo inicial do G1 foi consideravelmente semelhante ao de um jornal impresso, com sua página inicial semelhante à primeira página de um jornal e uma divisão baseada em editorias, o que foi sendo adaptado ao longo dos anos para um formato integrado que explora as linhas nativas digitais, tais como a notícia online, redes sociais, podcasts, mas traz também o jornalismo televisivo para dentro de sua estrutura (Memória Globo, 2022). O portal de notícias G1 possui um público médio de aproximadamente 55 milhões de usuários únicos ao mês, sendo considerado líder no segmento de portais de notícias no Brasil.

Na mesma linha de veículos jornalísticos nativos digitais, o jornal Metrôpoles foi fundado em 2015, na capital federal brasileira, tendo foco principal em notícias do Distrito Federal. A criação do jornal possui uma forte ligação com o encerramento da revista Veja Brasília em formato impresso, sendo grande parte da redação inicial do jornal composta por jornalistas provenientes desta revista (Pires, 2015). A redação do jornal Metrôpoles passou por reformulações que ocasionaram na transformação do site para um portal de notícias em 2017, expandindo a cobertura noticiosa para o contexto nacional, desde quando mantém parceria com outros veículos jornalísticos. O Metrôpoles é considerado um dos portais de notícias brasileiros com maior audiência na Internet, possuindo um público médio mensal de aproximadamente 86 milhões de usuários únicos (Metrôpoles, n. d.).

Para os jornais digitais portugueses foram considerados os cinco elencados no Quadro 2, os quais, de semelhante forma aos abordados no contexto do Brasil, apresentam jornais nativos digitais e jornais que foram criados com foco na mídia impressa. Foi identificado nos jornais Correio da Manhã e Público mecanismo próprio de busca de notícias, enquanto nos jornais Diário de Notícias, Jornal de Notícias e Observador foi identificada a utilização de ferramenta de pesquisa integrada do motor de busca Google.

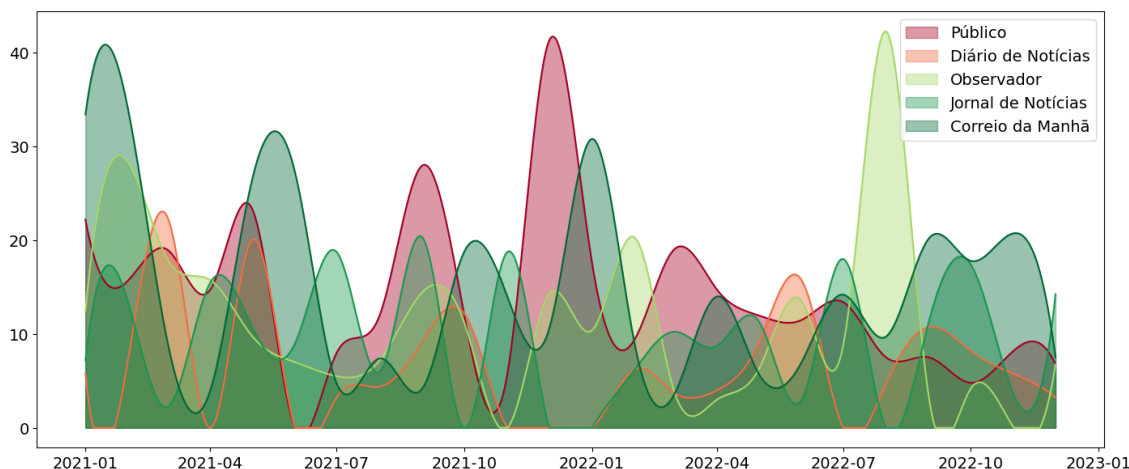
Quadro 2 - Jornais portugueses considerados no corpo de pesquisa

Jornal	Endereço eletrônico	Acessibilidade informacional
Correio da Manhã	https://www.cmjornal.pt/	Ferramenta própria de pesquisa
Diário de Notícias	https://www.dn.pt/	Ferramenta de pesquisa própria limitada (consolidação de dados via Google)
Jornal de Notícias	https://www.jn.pt/	Ferramenta de pesquisa própria limitada (consolidação de dados via Google)
Observador	https://www.observador.pt/	Ferramenta de pesquisa própria limitada (consolidação de dados via Google)
Público	https://www.publico.pt/	Ferramenta própria de pesquisa

Fonte: dados de pesquisa.

Em termos de popularidade, o interesse de busca observado por avaliação de dados de monitoramento da ferramenta Google Trends revela uma proximidade entre os cinco jornais, com variações pontuais observadas ao longo do período analisado, isto é, os anos de 2021 e 2022. Cabe ressaltar que os jornais Correio da Manhã, Observador e Público apresentaram picos pontuais de popularidade, porém mantendo, de forma geral, valores próximos ao longo de toda a série temporal.

Gráfico 2 - Variação de popularidade de sites portugueses a partir do Google Trends



Fonte: adaptado de Google Trends, 2023.

O jornal português Diário de Notícias é um dos mais tradicionais do país, tendo sido fundado em Lisboa ainda no século XIX, no ano de 1864, por Eduardo Coelho, jornalista e escritor, e Tomás Quintino Antunes, escritor e empresário, conhecido como 1º Conde de Marçal. O jornal é também conhecido por ter abrigado contribuições de importantes nomes da literatura portuguesa, como Ramalho Ortigão, Eça de Queirós, Pinheiro Chagas e Cesário Verde. Ao longo de sua história, um dos momentos mais impactantes ocorreu junto ao 25 de novembro de 1975, quando o jornal sofreu uma ruptura estrutural, cerceada por fatores político-ideológicos (Gomes, 2012). Décadas mais tarde, em 1995, o Diário de Notícias deu início à sua versão digital, com uma proposta inicial de apresentar as notícias consideradas mais significativas da versão impressa e um dossiê temático sobre as eleições (Bastos, 2011). De acordo com dados da Associação Portuguesa para o Controlo de Tiragem e Circulação (APCT)¹, no ano de 2022 o jornal Diário de Notícias obteve o oitavo lugar em número de circulação paga no contexto digital.

Pouco tempo após a criação do Diário de Notícias em Lisboa, o Jornal de Notícias foi fundado no Porto, em 1888. Juntamente com os jornais O Comércio do Porto e O Primeiro de Janeiro se consolidou como um dos grandes jornais portugueses, sendo dos três o único a sobreviver às transformações ocorridas na imprensa ao longo da segunda metade do século XX (Lima, 2017). O Jornal de Notícias foi o primeiro

¹ Dados disponíveis em: <https://www.apct.pt/analise-simples>. Acesso em: 15 de maio de 2023.

jornal português a se lançar na Internet, fato este ocorrido em 1995, sendo ainda o primeiro jornal a possuir uma redação digital (Bastos, 2011). Em termos de circulação digital paga, conforme dados da APCT, o Jornal de Notícias em 2022 alcançou a terceira posição. Em número totais, o Jornal de Notícias alcançou em Outubro de 2022 o primeiro lugar em número de visitas digitais no segmento de jornais portugueses (JN, 2022).

Fundado em 1979 com a proposta de se diferenciar de outros jornais por meio de uma linguagem mais popular e acessível, o Correio da Manhã alcançou sucesso por meio do comprometimento de explorar questões apelativamente mais humanas (Brito, 2010). O Correio da Manhã iniciou suas atividades na Internet em 1998, com um rol de informações relativamente limitado em comparação aos seus concorrentes (Bastos, 2011). Embora o Correio da Manhã, em uma pesquisa realizada em 2019, tenha se destacado por ser o jornal português com menor índice de confiança por parte de seus leitores, lidera os rankings de popularidade no contexto digital (CM, 2020).

O jornal Público teve seu primeiro número publicado em 1990, e no ano seguinte passou a integrar o World Media Network, o qual consiste de uma associação mundial de jornais de referência, a exemplo do espanhol El País, o francês Libération e o italiano La Stampa (Oyama, 2013). Em 1995 o Público se tornou o segundo jornal português a se lançar na Internet e, em 1999, se tornou o primeiro a disponibilizar notícias em tempo real, por meio de um serviço batizado de “Última Hora” (Duarte, 2013). De acordo com dados da APCT, em 2022 o Público foi o jornal com maior circulação digital paga.

O jornal Observador é um jornal português nativo digital, criado com a proposta de ser um jornal inteiramente digital com publicações em tempo contínuo. Em seu primeiro mês de funcionamento o jornal registrou um público de aproximadamente seiscentos visitantes, cenário que foi consideravelmente alterado nos anos seguintes. Em março de 2016 o jornal registrou 42,8 milhões de visualizações em suas páginas (Observador, 2016). O jornal apresenta seu conteúdo de forma gratuita ao público, sendo financiado essencialmente por publicidade (Conde, Magalhães e Moreira, 2018). Marques (2020) destaca que, considerando o digital como eixo central da produção jornalística do Observador, o jornal apresentou uma aparente conotação política com associação à direita em Portugal.

1.3. Coleta de dados

A partir da delimitação dos veículos jornalísticos considerados como fontes dos dados analisados na presente investigação, fez-se necessário a delimitação de uma estratégia para coleta sistematizada de dados. Neste ponto se destaca um dos grandes desafios dos processos de geração de informação e descoberta do conhecimento: a seleção de dados. Determinados projetos com foco na análise de dados possuem a disponibilização prévia dos dados necessários ou ferramentas de conexão direta com bases de dados de interesse. Entretanto, trazendo para o cenário das investigações no contexto das ciências sociais aplicadas, é comum se esbarrar em desafios quanto à obtenção de dados necessários, os quais podem orbitar questões típicas de um ambiente Big Data, tais como a multiplicidade de fontes, a variação de formatos e a disponibilidade e acesso a soluções computacionais que possibilitem uma coleta sistematizada e massiva de dados.

As fontes de dados consideradas se referem a um conjunto de páginas de acesso aberto na Internet. Diante disso, foi construída uma solução computacional baseada na utilização da linguagem de programação Python para raspagem de dados. A escolha da linguagem de programação Python está baseada em sua flexibilidade e diversidade de funções adequadas para tarefas de análise de dados (Vanderplas, 2023). Neste sentido, o primeiro desafio é a obtenção de quais são as notícias de interesse para a composição da base de dados. Embora os sites jornalísticos considerados possuam ferramentas nativas ou terceirizadas para a pesquisa de notícias, verificações prévias realizadas mostraram limitação destas ferramentas, principalmente no sentido listagem temporal das notícias, ou seja, apresentação indexada de notícias mais antigas em suas ferramentas de busca. A partir disso, foi considerada a listagem de notícias com base em busca direta por meio da indexação realizada pelo motor de busca Google.

De acordo com dados de monitoramento da Internet, o motor de busca Google lidera o ranking de popularidade de utilização, respondendo por uma fatia que ultrapassa os 90% de interesse em todo o mundo (Tabela 1). Além disso, o Google responde por diversos avanços no desenvolvimento e aplicação de soluções inovadoras de inteligência artificial no âmbito do processamento de linguagem natural, utilizadas

no aprimoramento de suas tarefas de indexação e disponibilização de conteúdos, tais como Word2Vec (Mikolov *et alli.*, 2013a; Mikolov *et alli.*, 2013b) e BERT (Devlin *et alli.*, 2018), os quais proporcionam resultados mais assertivos quanto a interesses de busca.

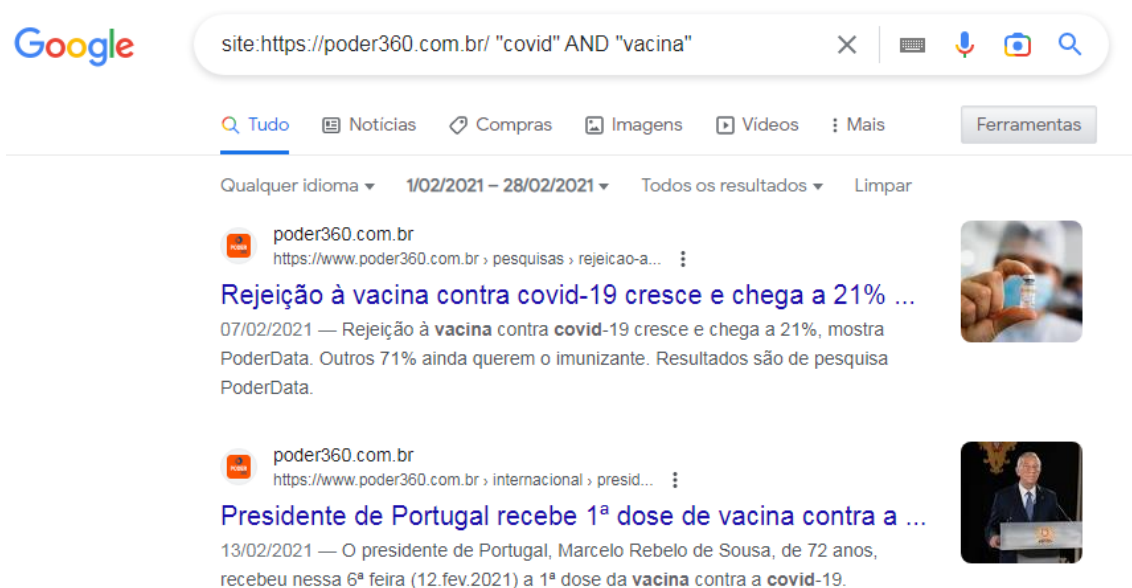
Tabela 1: Ranking de popularidade de motores de busca em março de 2023

Motor de busca	Participação
Google	90,60%
Yahoo	3,25%
Bing	3,12%
Yandex	0,46%
Naver	0,43%
Outros	2,13%

Fonte: Similarweb (2023).

De forma geral, a solução de obtenção de dados está baseada no conceito de *web scraping* (raspagem de dados), o qual se refere à construção de robôs que automatizam o processo de acesso e extração de conteúdos de interesse armazenados na Internet (Broucke e Baesens, 2018). O motor de busca Google possui ferramentas que possibilitam a filtragem de resultados com base em parametrização de delimitadores específicos, tais como site, termos, período temporal e uso de operadores lógicos (Hrabiec-Hojda, Trzeciakowska, 2019). A Figura 2 apresenta uma visualização de uma pesquisa no buscador onde se especifica o site do jornal Poder360 para o período de fevereiro de 2021, com os termos “covid” e “vacina”. A utilização do operador lógico *AND* proporciona que os resultados alcançados na busca contenham ambos os termos de pesquisa. A partir disso, em uma primeira etapa de extração de dados, foi realizada a construção de uma solução computacional que mapeia, por meio do motor de busca Google, os endereços eletrônicos para as notícias veiculadas nos sites de interesse, consolidando-os em uma base de dados estruturada a ser utilizada posteriormente para acesso às notícias.

Figura 2 - Captura de tela de busca parametrizada no motor de busca Google



Fonte: Google, 2023.

A segunda etapa da obtenção dos dados se refere à efetiva obtenção dos conteúdos textuais veiculados. Como estratégia de obtenção de tais dados foi construída uma solução computacional que acessa, para cada um dos sites, os endereços eletrônicos obtidos na primeira etapa de extração de dados e por meio de técnicas de *web scraping* coleta e consolida os dados na base de dados. É importante ressaltar que cada um dos sites possui uma estrutura particular, o que exige ajustes quanto às parametrizações utilizadas para identificação dos conteúdos dentro de suas respectivas páginas.

1.4. Pré-processamento e transformação de dados

Originalmente o processo KDD prevê a execução das tarefas de pré-processamento e transformação de dados de maneira isolada e sequencial. Entretanto, variações nos requisitos referentes aos procedimentos de análise de dados podem requerer adequações que possibilitem simplificação de ações e alinhamento com as necessidades de dados e informações. A coleta de dados por meio de *web scraping* torna necessária a consolidação dos dados imediatamente após a sua identificação, sendo comum que os dados de origem, em seu formato bruto, não possuam uma

estrutura padrão para além de sua apresentação e organização nas páginas eletrônicas. Neste sentido, as tarefas de pré-processamento e transformação de dados, as quais tratam, consecutivamente, da limpeza e estruturação de dados para os procedimentos de mineração de dados, podem ser realizadas de maneira paralela.

Na presente investigação, os dados obtidos por meio da aplicação da solução de *web scraping* são consolidados em um formato semi-estruturado que possibilita o processamento por meio de ferramentas computacionais. Para tanto, a estratégia de consolidação de dados está pautada no formato *csv* (*comma separated values*), contemplando os seguintes atributos: nome do jornal, país, data de publicação, título da notícia e texto da notícia. Entre as vantagens de utilização do padrão *csv* se destacam o fato de ser um formato aberto e processável por diferentes tipos de tecnologias, quer sejam proprietárias ou abertas, o que corrobora para uma simplificação dos processamentos computacionais necessários em análise de dados, assim como para uma melhor organização e qualificação dos dados (FERNANDES; CORDEIRO, 2016).

Durante a consolidação dos dados para a estrutura *csv* são realizadas verificações quanto ao conteúdo textual recuperado, o que se refere ao pré-processamento de dados. Os conteúdos textuais recuperados podem apresentar ocorrências de menções a publicidades ou chamadas a notícias correlatas dentro do espaço de corpo das notícias, como mostra o exemplo da Figura 3. A limpeza de dados é realizada por meio da aplicação de rotinas baseadas em expressões regulares (Fitzgerald, 2012), as quais identificam tais ocorrências com base em padrões textuais e realizam a sua remoção.

1.5. Processamento de linguagem natural

As evoluções tecnológicas vividas ao longo do século XX e reverberadas no século XXI proporcionaram um cenário de abundante volume de textos em linguagem natural. Embora exista um considerável conteúdo implícito nas bases de dados textuais, explorá-lo é um grande desafio, principalmente considerando limitações temporais (Chowdhary, 2020). A utilização de soluções baseadas em processamento de linguagem natural (PLN) possui como objetivo a realização deste tipo de tarefa com precisão e em tempo consideravelmente menor ao que um ser humano faria. A aplicação de PLN para solução de problemas de análise de dados textuais possibilita extrapolar estratégias

clássicas de análise de conteúdo que se apropriam, de forma geral, de conjuntos amostrais limitados (Tixier *et alli.*, 2016; Lucy *et alli.*, 2020). Neste sentido, se observa a consolidação direta de uma das características da análise de dados em um contexto Big Data, isto é, o volume de dados. Tal característica acaba por possibilitar um aprofundamento no que se refere à identificação de temas, sentidos e significados implícitos.

Figura 3 - Exemplo de ocorrência de menções a outras notícias

The image shows a screenshot of a news article from the website Metrôpoles. At the top, there is a navigation bar with a 'Menu' icon, a notification bell, the 'METRÔPOLES' logo, and a search bar labeled 'Busca'. The main text of the article discusses researchers from the University of Texas and Pfizer testing a vaccine with RNA messenger technology, highlighting three mutations in the spike protein: D614G, E484K, and N501Y. Below the main text, there is a section titled 'Mais sobre o assunto' which recommends four related articles. Each recommendation includes a small image, a category label (e.g., 'Brasil', 'Saúde'), and a headline. The recommended articles are: 1. 'Pfizer solicita à Anvisa registro definitivo de vacina contra a Covid-19' (Brazil); 2. 'Covid-19: Reino Unido estuda combinação das vacinas de Oxford e Pfizer' (Saúde); 3. 'Entenda por que mutações do coronavírus se tornaram uma preocupação mundial' (Saúde); 4. 'Mutações britânicas do coronavírus deve ser predominante nos EUA até março' (Saúde).

Os testes foram feitos com amostras de sangue de 20 voluntários participantes dos testes clínicos da vacina. Esse material é composto por anticorpos obtidos

Fonte: Metrôpoles², 2023.

Soluções baseadas em PLN podem ser aplicadas em uma diversidade de problemas, incluindo sumarização textual, reconhecimento de voz, tradução automática, extração de informação, agrupamento e classificação de textos, geração de texto e

² Captura de tela da notícia “Vacina da Pfizer funciona contra 3 variantes do coronavírus, diz estudo”. Disponível em: <https://www.metropoles.com/saude/vacina-da-pfizer-funciona-contras-3-variantes-do-coronavirus-diz-estudo>. Acesso em: 3 de maio de 2023.

análise de sentimentos (Khurana *et alli.*, 2023). No contexto de investigações relacionadas à compreensão de conteúdos circulantes em peças noticiosas, soluções para agrupamento e classificação textual (classificação temática) e análise de sentimentos são de especial relevância, uma vez que possibilitam a compreensão de temas abordados a partir de grandes volumes de notícias, assim como possíveis posicionamentos relativos a determinados assuntos.

1.6. Classificação temática

No contexto de processamento de linguagem natural, a classificação temática é uma tarefa de grande importância e interesse, uma vez que possibilita a identificação dos possíveis conteúdos para grupos distintos de elementos, sejam estes frases, parágrafos ou documentos, que possuam similaridade entre si. Existem diferentes abordagens que possibilitam a análise de conteúdo por meio de técnicas baseadas em aprendizado de máquina, as quais podem ser baseadas em tarefas de agrupamento ou ainda na modelagem de tópicos (Yu, Jannasch-Pennell e DiGangi, 2011; Xing, Lee e Shibani, 2020; Saheb, Dehghani e Saheb, 2022).

O agrupamento textual, conhecido também pela palavra clusterização derivada do original da língua inglesa *clustering*, pode ser realizado por diferentes técnicas, as quais possuem em comum o cálculo de determinados indicadores que possibilitam a mensuração da distância de similaridade entre dois diferentes elementos, de maneira a agrupar elementos a partir de características em comum, tais como palavras, estruturas ou estilos (Vajalla *et alli.*, 2020). No contexto da classificação temática, grandes volumes de dados textuais podem ser agrupamentos por meio de técnicas de aprendizado de máquina e posteriormente avaliados quanto aos seus possíveis temas através da aplicação, por exemplo, de técnicas de visualização de dados. Diferentes investigações apresentam a aplicação de técnicas de agrupamento textual, tais como o algoritmo *k-Means*, com a posterior avaliação de conteúdos por meio da visualização de nuvem de palavras e análise dos elementos textuais centrais que compõem os grupos identificados (Garg e Rani, 2017; Cordeiro *et alli.*, 2023).

Outra estratégia para a classificação temática é o uso de técnicas de modelagem de tópicos, as quais possuem foco na identificação e extração de assuntos ou tópicos

implícitos ao conjunto de dados textuais a ser analisado. A aplicação de modelagem de tópicos é voltada para a identificação de tendências e é explorada por técnicas como: *Latent Dirichlet Allocation* (LDA) (Blei, Ng e Jordan, 2003), *Non-negative Matrix Factorization* (NMF) (Dhillon e Sra., 2005) e *Correlated Topic Model* (CTM) (Blei e Laffert, 2007).

A técnica LDA é considerada como uma das estratégias mais populares e utilizadas no contexto de modelagem de tópicos (Jelodar et alli, 2019). De maneira geral a técnica parte da premissa de que documentos textuais seguem uma estrutura composta por combinações de tópicos, e cada tópico também é composto por uma combinação de termos, os quais normalmente possuem relações entre si, quer sejam em termos sintáticos ou semânticos (Tong e Zhang, 2016). Um tópico pode ser compreendido como um padrão recorrente em diferentes documentos, o que pode ser utilizado como insumo para a identificação de relações entre tais documentos (Wu et alli, 2017).

A aplicação da técnica LDA no contexto de classificação temática de textos jornalísticos têm sido explorada em diferentes contextos, tais como saúde e qualidade de vida (Lee, 2020; Wahyudi *et alli.*, 2021), política (Bagozzi e Schrodte, 2012) e economia (Che et alli, 2021). Na presente investigação a técnica LDA é utilizada para identificar as classes temáticas (tópicos) do conjunto de notícias extraídos dos dez sites jornalísticos considerados para composição da base de dados amostral. A partir disso, os documentos textuais devidamente associados aos seus grupos são avaliados por meio de análise de sentimentos.

1.7. Análise de sentimentos

A análise de sentimentos pode ser descrita como uma técnica de PLN que possui como objetivo a identificação e quantificação da polaridade emocional associada a um determinado elemento textual, comumente baseada nos rótulos positivo, neutro e negativo (Liu e Zhang, 2012). As aplicações de análise de sentimento envolvem investigações em diferentes áreas do conhecimento, com foco em questões como compreensão de mercado, análise política, análise de redes sociais e identificação de comportamentos sociais (Soong *et alli.*, 2019). Os algoritmos de análise de sentimentos podem se basear em diferentes estratégias, tanto supervisionadas quanto

não-supervisionadas, a depender do problema a ser resolvido e da disponibilidade de dados.

Soluções computacionais de análise de sentimentos supervisionadas requerem a disponibilidade de uma amostra de dados previamente rotulada, isto é, dados textuais que possuam um rótulo de sentimento associado, de modo que estes possam ser utilizados por algoritmos inteligentes em seu processo de aprendizado (Li *et alli.*, 2019). A utilização de soluções supervisionadas é interessante principalmente sob o ponto de vista das possibilidades de validação de resultados de inferência computacional alcançada. Entretanto, a disponibilidade de dados rotulados, em determinados contextos, é um grande desafio, uma vez que requer uma avaliação e anotação em grandes volumes de dados, o que acaba tornando necessária a utilização de métodos não-supervisionados.

No contexto de métodos de análise de sentimentos não-supervisionados, uma das principais estratégias são os algoritmos baseados em léxicos. De forma geral, um léxico pode ser descrito como um dicionário de termos únicos com classificação de polaridade, os quais são criados por meio de anotação manual ou através de análise automática de grandes volumes de dados (Hussein, 2018). A partir disso, os modelos de análise de sentimentos baseados em léxico realizam cálculos sobre os elementos textuais de maneira a avaliar as pontuações atribuídas aos termos que os compõem, fornecendo uma pontuação entre um valor extremamente negativo e um valor extremamente positivo. Entre as vantagens das soluções de análise de sentimentos baseadas em léxicos se destacam a simplicidade de implementação e a velocidade de processamento quando comparada a outras técnicas (Taj, Shaikh e Meghji, 2019).

Entre as soluções de análise de sentimentos baseadas em léxicos está o VADER (*Valence Aware Dictionary and sEntiment Reasoner*), proposto por Hutto e Gilbert (2014), o qual se destaca como uma interessante ferramenta, uma vez que possui uma maior capacidade de lidar com variações textuais que extrapolam a norma culta da escrita, sendo de grande relevância para investigações no âmbito das ciências sociais aplicadas, tais como em análise de notícias ou análise de redes sociais (Khemani e Adgaonkar, 2021; Castellanos, Xie e Brenner, 2021). Embora o VADER possua como foco a língua inglesa, a ramificação (*fork*) conhecida como LeIA (Léxico para Inferência Adaptada) possibilita sua aplicação para a língua portuguesa, possibilitando a aplicação em dados textuais provenientes de mídias digitais (Almeida, 2018), como é o

caso da presente investigação, tornando possível a geração de insumos para a realização de análises aprofundadas no sentido de identificação de possíveis polarizações e posicionamentos nos conteúdos constantes das notícias analisadas.

CAPÍTULO 2 – JORNALISMO EM TEMPOS DE COVID-19

O jornalismo na saúde é algo de grande importância, tendo seus estudos remetidos à interdisciplinarização entre as Ciências da Comunicação e as Ciências da Saúde, sendo visto como um campo de investigação em crescimento (Lee e McElroy, 2019). Diversas investigações buscam compreender como o jornalismo pode influenciar o comportamento individual e as dinâmicas comunicacionais e informacionais da população (Hinnant, Len-Ríos e Ho, 2012). Ao mesmo tempo, cabe ressaltar que a circulação de notícias imprecisas podem provocar o surgimento de expectativas irrealistas por parte do público, assim como contribuir para a geração de falsas esperanças e medos, corroborando para o aumento de dilemas ou ainda para a circulação de desinformação (Waisbord, 2018).

Os veículos jornalísticos apresentam diversas informações e conhecimentos relacionados à saúde nos mais diversos níveis, desde questões relacionadas à prevenção e diagnóstico, até a abordagem do tratamento de doenças, reabilitação e consequências decorrentes das mesmas (Hallin *et alli.*, 2013). Frente às necessidades informacionais atreladas às questões de saúde provocadas pela crise sanitária da pandemia da COVID-19, se observou uma grande demanda e necessidade por informações precisas, relevantes, rápidas e imparciais (Mohammadi *et alli.*, 2021). Cabe ainda ressaltar que o volume e o formato da cobertura jornalística podem influenciar não apenas a sociedade como um todo, mas ainda as autoridades responsáveis pela formulação de políticas públicas (Bou-Karroum *et alli.*, 2017).

No que tange ao contexto de Portugal, Lopes *et alli.* (2020) afirmam que a pandemia da COVID-19 acabou por se mostrar um importante marco para o jornalismo português, onde veículos jornalísticos se consolidaram enquanto frentes de enfrentamento à pandemia, com discursos voltados para a circulação de informações de apoio ao não contágio e às medidas de segurança sanitária. No Brasil, de maneira

ligeiramente similar, o jornalismo enquanto fonte informacional se viu envolta a manifestações por parte da sociedade em busca de informações fidedignas e apartidárias, voltadas à uma orientação da verdade (Varão e Ferreira, 2020). E para além disso, há também a questão da intersubjetividade e o dever social associado ao jornalismo em meio a um cenário de infodemia e desinformação (Rêgo, Santos e Rocha, 2020).

2.1 Importância do jornalismo na saúde pública

Os avanços alcançados no âmbito das tecnologias de informação e comunicação, sobretudo nas soluções comunicacionais associadas ao uso da Internet, emergiram um cenário onde veículos de comunicação eletrônica se tornaram uma das principais fontes de informação sobre a saúde pública (Heldman, Schindelar e Weaver, 2013; Thompson, 2019). Ações como a busca de sintomas associados a determinadas doenças ou indicações de tratamento se tornaram comuns. Neste sentido, o jornalismo eletrônico possui um papel de grande importância na disponibilidade e circulação de informações relevantes e de credibilidade no contexto da saúde pública, atuando, inclusive, no combate à infodemia (Cunha e Moreira, 2023).

A saúde pública tem como objetivo central a organização de serviços e sistemas de saúde por meio da exploração de conhecimentos, os quais podem ser provenientes da área médica ou de outras áreas, tais como em tarefas que remetem à gestão e distribuição de medicamentos, desenvolvimento de imunizantes, ou mesmo em ações voltadas à disponibilização de informações (Hortelan *et alli.*, 2019; Sousa e Fernandes, 2020). O jornalismo eletrônico se destaca por sua responsabilidade inerente à disponibilização de informações confiáveis, amparadas por evidências científicas.

Entre as vantagens alcançadas por meio do jornalismo eletrônico destacam-se a sua capacidade de circulação de maneira rápida e abrangente, além da possibilidade de interação entre jornalistas e população por meio de ferramentas disponíveis em sites e portais jornalísticos (Ferrari, 2014). Entretanto, paralelamente surgem desafios e problemas relativos às próprias características que foram imputadas por esse cenário de cibercultura, isto é, o ambiente proveniente dos fenômenos, costumes e práticas advindos da acessibilidade digital, protagonizados por uma fluidez e velocidade

informativa (Durmus, 2021). Entre os principais desafios encontrados, se destacam a credibilidade de informações, assim como a associação interpretativa a determinados fatos atrelada a possíveis posicionamentos presentes nos conteúdos jornalísticos.

Conforme abordado por Malinverni, Cuenca e Brigagão (2012), é necessário que as práticas jornalísticas estejam atreladas às suas implicações no fazer cotidiano da sociedade, devendo ainda estarem sensíveis às complexidades dos processos informativos, principalmente no contexto de saúde pública, não se valendo da legitimada bandeira da liberdade de expressão para a construção de significados. Junto a isso, o jornalismo eletrônico, face ao seu importante papel na saúde pública, deve se valer do papel crítico dos jornalistas na verificação de informações e no estabelecimento de colaborações com profissionais da saúde, de modo a prover informações relevantes, auxílio na comunicação da ciência e apoio na promoção de medidas preventivas.

2.2 Pluralidade de visões e discursos

Embora sejam notáveis as movimentações decorrentes do uso de tecnologias emergentes, a exemplo da aplicação de ferramentas de inteligência artificial no contexto da prática jornalística (Zandomênic, 2022), a mudança de paradigmas em uma macrovisão social é algo mais complexo de se vislumbrar. Sousa (2006) pontua que neste cenário, um ponto de partida é a necessidade de informação de qualidade por parte da sociedade, o que, eventualmente, pode estar relacionado à uma pluralidade de visões e discursos, quer sejam a partir do próprio delineamento descritivo apropriado pelo profissional do jornalismo ou ainda por parte do veículo de comunicação (Reginato, 2018).

O jornalismo enquanto objeto de estudo multifacetado e complexo acaba por tornar necessária a realização de análise mais profundas quanto à pluralidade intrínseca do fenômeno noticioso, podendo demandar análises que permeiam a produção e circulação noticiosa, assim como levem em conta as estruturas que moldam o próprio consumo da informação (Silva, 2009), o que pode estar relacionado a agendas políticas, interesses comerciais e viés ideológico (Cota e Maciel, 2016).

No âmbito do jornalismo eletrônico, considerando as já citadas vantagens de acesso à informação características do uso de tecnologias da informação e comunicação,

a discussão pública e a natureza plural do fenômeno noticioso acabam por proporcionar um ambiente diversificado de visões, corroborando para análises mais profundas e críticas sobre um determinado problema (Peeters e Maesele, 2023). Tal característica enriquece o debate público e se alinha ao próprio exercício da cidadania (Miguel e Biroli, 2010). Entretanto, ao mesmo tempo, em um universo plural, o acesso individualizado e restrito pode ir de encontro à formação de opiniões e visões enviesadas para com o ponto de vista do jornalista ou do veículo de comunicação (Helberger, 2020).

A responsabilidade de profissionais do jornalismo ao apresentar de forma credível e ética a informação, ainda que sob diferentes perspectivas, é algo fundamental para construção de um espaço de controvérsias legítimas, possibilitando contradiscursos e debate público. Tudo isso converge na disponibilidade informacional à sociedade de uma maneira que extrapola a mera repetição e reprodução acrítica de discursos (Budarick, 2023). Todavia, o volume informacional dos ambientes digitais ainda carece de mecanismos mais direcionados e profundos que permitam compreender toda a dinâmica plural intrínseca à circulação de informação no contexto do jornalismo eletrônico.

CAPÍTULO 3 – VACINAÇÃO, JORNALISMO E POSICIONAMENTOS

A realização de análises sobre conjuntos amostrais de peças noticiosas é algo de grande relevância, uma vez que possibilita uma compreensão aprofundada sobre como a mídia abordou temáticas de interesse da sociedade, quer seja em termos quantitativos como qualitativos. Neste capítulo são apresentados e comentados os resultados alcançados no contexto da análise de peças noticiosas relacionadas à vacinação contra a COVID-19 a partir do uso de soluções computacionais inteligentes.

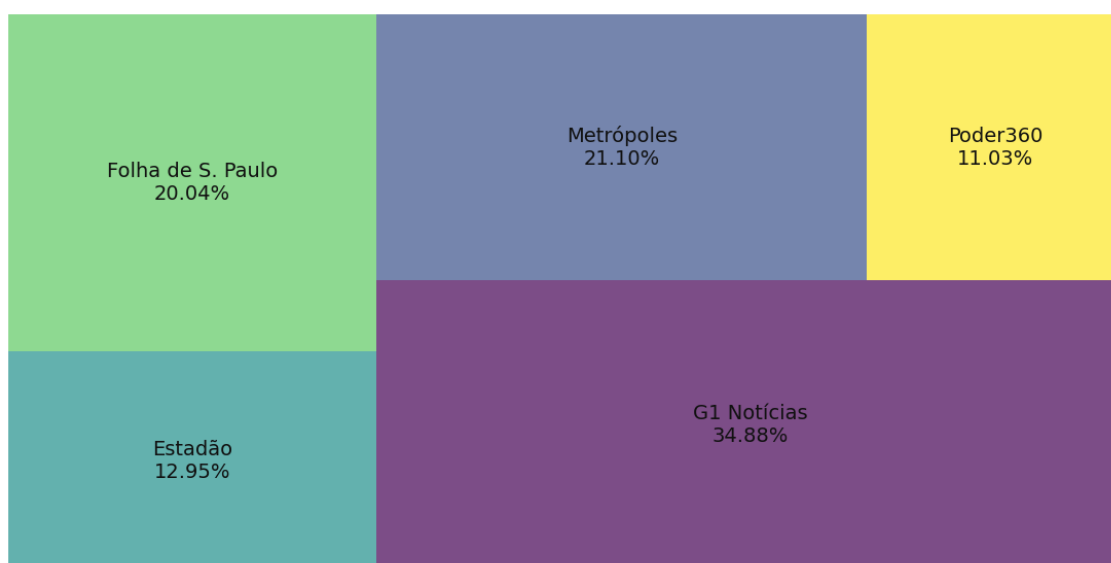
3.1 Análise exploratória das peças noticiosas

Através da aplicação das soluções de extração de dados construídas, foram identificadas, em sua totalidade, 23.679 peças noticiosas para o conjunto amostral de jornais brasileiros e portugueses considerados. Conforme detalhado na Seção 1.2, o conjunto de dados se refere a notícias recuperadas em dez sites jornalísticos a partir da indexação no motor de busca Google com referência aos termos “vacina” e “covid”. Após o tratamento do conjunto de dados, foram aplicadas estratégias de análise exploratória de dados, com o propósito de se conhecer de forma amplificada a distribuição das notícias quanto aos jornais e quanto ao período de publicação em suas plataformas digitais.

No contexto do Brasil, foram recuperadas 20.098 notícias. O Gráfico 3 apresenta a distribuição quantitativa das notícias em relação aos sites jornalísticos. Em termos numéricos, o portal G1 apresenta aproximadamente um terço do total de notícias, seguido pelos jornais Folha de S. Paulo e Metrôpoles. Os jornais Estadão e Poder360 apresentaram os menores percentuais. É importante destacar que a avaliação numérica foi realizada de maneira isolada, isto é, não considerou o proporcional de publicações de notícias relacionadas aos termos de interesse com o total de publicações realizadas pelos jornais.

De modo a complementar a compreensão quantitativa sobre as publicações de notícias, o Gráfico 4 apresenta a série temporal histórica. O G1 Notícias, portal com o maior volume de publicações, se manteve dentro um valor aproximado contínuo de trezentas publicações mensais. Cabe ressaltar que o G1 Notícias é um portal de notícias generalista que, além da publicação de notícias produzidas diretamente pela redação do jornal, integra notícias provenientes de empresas regionais ligadas ao Grupo Globo, responsável pela manutenção do portal.

Gráfico 3 - Distribuição quantitativa percentual de notícias capturadas (Brasil)

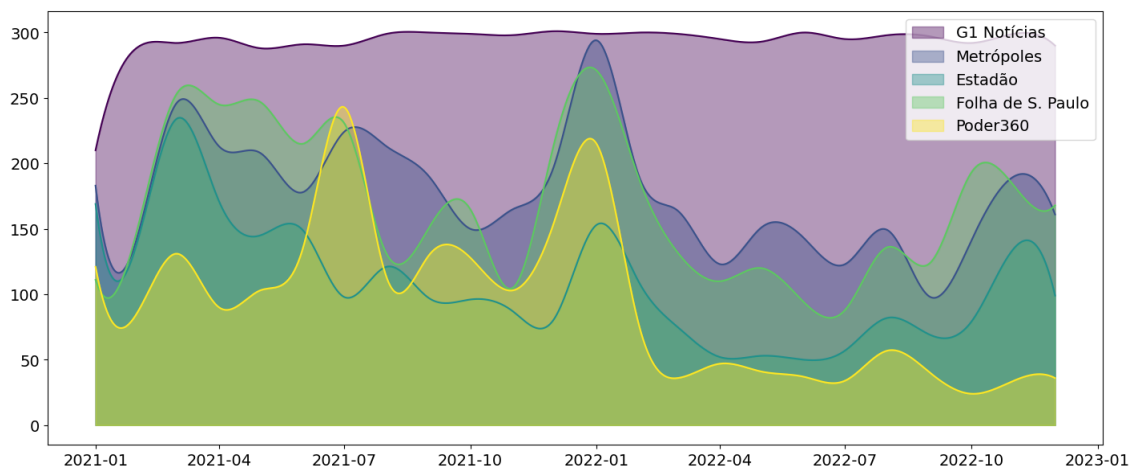


Fonte: dados de pesquisa.

Ainda com relação à distribuição temporal de publicações de notícias no Brasil, o jornal Metrôpoles, segundo em termos de volume, apresentou flutuações ao longo dos anos de 2021 e 2022 que acompanharam eventos específicos associados à vacinação, tais como o anúncio do desenvolvimento de uma vacina nacional e o registro recorde diário no número de mortes decorrentes da COVID-19 ocorridos em março de 2021. Um comportamento numérico semelhante pode ser observado para os demais jornais, principalmente em relação ao ano de 2022, onde se observam dois picos principais, o primeiro em janeiro e o segundo em dezembro. Em janeiro de 2022 o Brasil presenciou novos recordes no número de registros diários de casos de COVID-19, associados a uma nova onda de infecção e à nova variante Ômicron (Agência Brasil, 2022). Em dezembro de 2022 ocorreu a diplomação do presidente Luiz Inácio Lula da Silva, assim

como atos e ações decorrentes das eleições ocorridas, os quais reverberaram na temática vacinação e COVID-19.

Gráfico 4 - Série temporal de notícias capturadas (Brasil)



Fonte: dados de pesquisa.

O Gráfico 5 apresenta as nuvem de palavras segmentadas para cada um dos jornais brasileiros, possibilitando uma observação sobre os principais termos envolvidos às notícias publicadas. Em todas as imagens é possível observar uma recorrência dos termos “caso” e “doença”, referenciando de forma direta à pandemia da COVID-19. Entretanto, variações pontuais de termos indicam a exploração de temáticas mais específicas, tais como o destaque para o termo “bolsonaro” na nuvem de palavras do jornal Metrôpoles, se referindo ao então presidente brasileiro Jair Messias Bolsonaro, ou ainda o termo “usf”, sigla para Unidade de Saúde Familiar (USF), com considerável presença na nuvem de palavras do portal G1 Notícias.

No contexto de Portugal foram recuperadas 3.581 peças noticiosas, sendo a composição apresentada no Gráfico 6. Os jornais Público e Diário de Notícias respondem por aproximadamente dois terços do volume de notícias recuperadas. Sendo o jornal Correio da Manhã o que possui a menor parcela de peças noticiosas identificadas. De forma complementar, o Gráfico 7 apresenta a série temporal histórica, onde é possível observar um comportamento de diminuição no volume de publicações ao longo dos anos de 2022 e 2021. Todos os jornais apresentaram picos numéricos de publicação ao longo dos primeiros meses de 2021, altura em que foram realizadas as

primeiras ações de vacinação no país, e um segundo pico em dezembro de 2021 e janeiro de 2022, associado à identificação da variante Ômicron.

Gráfico 5 - Nuvem de palavras segmentadas por jornal (Brasil)

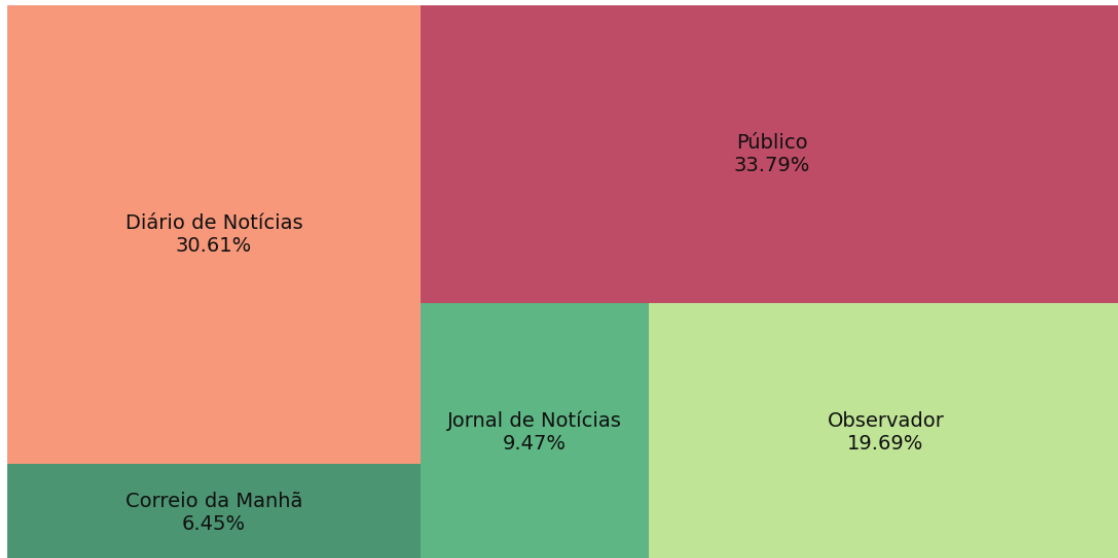


Fonte: dados de pesquisa.

O Gráfico 8 apresenta as nuvens de palavras segmentadas por jornais para o contexto de Portugal. De maneira semelhante ao que é observado nos dados dos jornais brasileiros, os termos “caso” e “dose” são recorrentes, se referenciando de maneira direta à doença, tais como em notícias informativas apresentando boletins de casos e informações sobre disponibilização e aplicações da vacina. Outros termos observados apresentam também proximidade direta com a pandemia da COVID-19 e as ações de

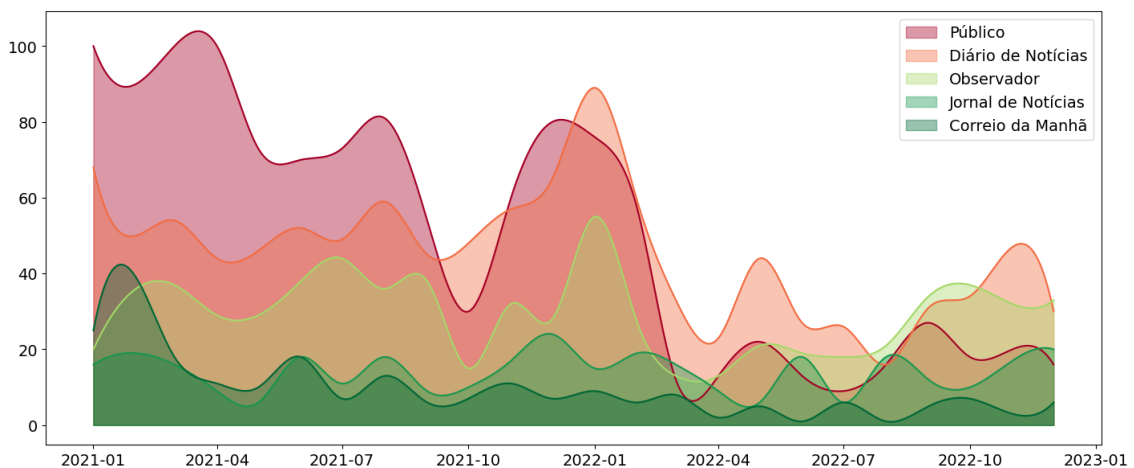
enfrentamento da doença por meio da vacinação, sendo necessárias explorações mais aprofundadas para identificação de possíveis variações temáticas abordadas (Seção 3.2).

Gráfico 6 - Distribuição quantitativa percentual de notícias capturadas (Portugal)



Fonte: dados de pesquisa.

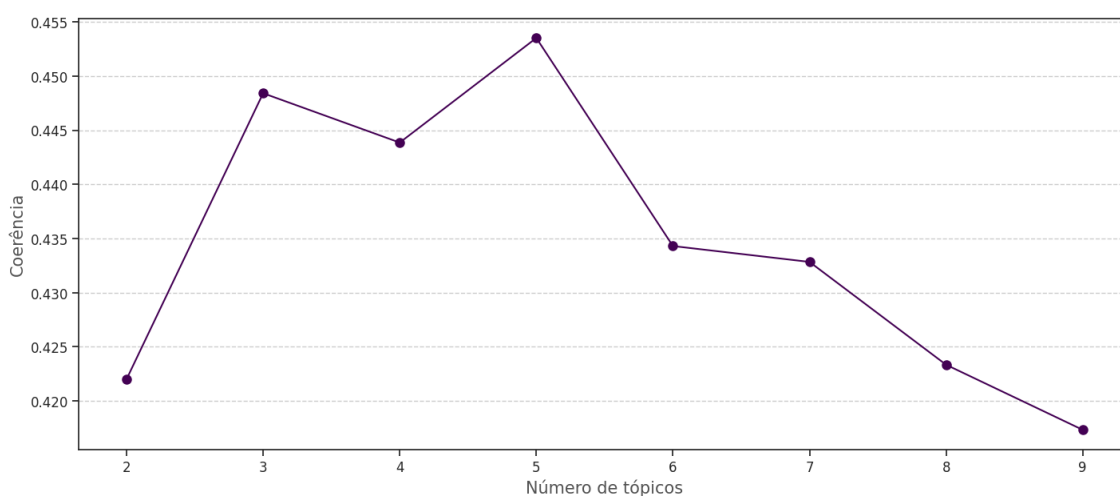
Gráfico 7 - Série temporal de notícias capturadas (Portugal)



Fonte: dados de pesquisa.

O Gráfico 9 apresenta a análise de coerência para o conjunto de notícias brasileiras. É possível observar que o valor 5 foi identificado como o de maior coerência. A partir disso, foi realizada a identificação de tópicos usando o LDA para cinco grupos temáticos. O Gráfico 10 apresenta a distribuição de grupos, onde se observa a existência de três grandes áreas temáticas, a primeira composta pelos conjuntos de tópicos sinalizados como 1, 2 e 5, a segunda referente ao conjunto de tópicos 4 e, finalmente, o conjunto de tópicos 3. A distância visual entre os conjuntos representa a similaridade temática, isto é, quanto mais próximo estão dois conjuntos, mais semelhantes são as temáticas que possuem, tal como pode ser observado nos conjuntos 1 e 2, os quais possuem uma considerável intersecção. Para cada um dos tópicos foram geradas nuvens de palavras de maneira a possibilitar uma melhor compreensão dos conteúdos circulantes e suas temáticas. O Gráfico 11 apresenta as nuvens de palavras segmentadas por tópicos.

Gráfico 9 - Análise de coerência de tópicos (Brasil)



Fonte: dados de pesquisa.

O tópico 1 possui considerável destaque para os termos “gente” e “dose”, acompanhado por outros termos, com menor destaque, como “família”, “vídeo”, “casa” e “vacinado”. Esse conjunto de termos faz menção a notícias mais específicas sobre procedimentos vacinais realizados na população, sobretudo noticiando casos de destaque. O modelo computacional utilizado para identificação de tópicos possibilita a

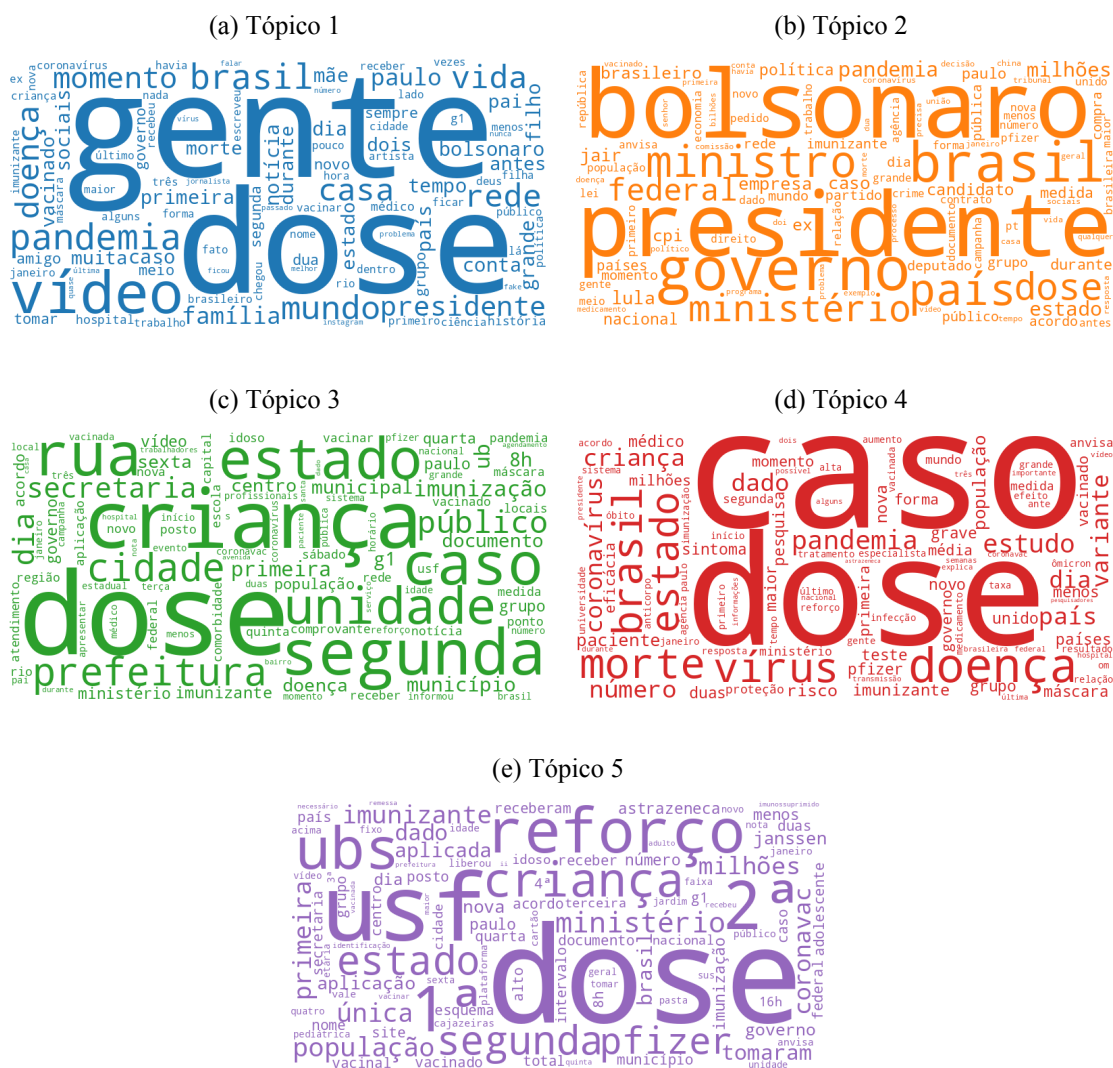
mensuração de representativa dos documentos textuais, isto é, as notícias, associando um percentual que permite recuperar elementos centrais do corpus textual. A partir disso, o Quadro 3 apresenta um conjunto das dez notícias de maior representatividade para o tópico 1, onde se observa a noticiabilidade da vacinação em pessoas, sobretudo, notadamente conhecidas pela sociedade.

Gráfico 10 - Distribuição temática de tópicos (Brasil)



Fonte: dados de pesquisa.

Gráfico 11 - Nuvens de palavras segmentadas por tópicos (Brasil)



Fonte: dados de pesquisa.

Quadro 3 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 1 (Brasil)

Título da notícia	Jornal	Data de publicação
Léo Santana recebe 1ª dose de vacina contra Covid: “Viva a vida”	Metrópoles	03/08/2021
Tadeu Schmidt se emociona ao receber primeira dose da vacina contra covid-19	Estadão	30/06/2021

Posicionamentos Jornalísticos sobre a Vacinação da COVID-19

Gretchen é vacinada contra a Covid-19 e comemora: “Obrigada, SUS”	Metrópoles	28/03/2021
Denise Fraga se emociona e chora ao tomar vacina contra covid-19: 'Viva quem se preocupa com o povo'	Estadão	11/06/2021
Sertanejo autor de 'Viva a Vida' se emociona ao tomar vacina contra Covid e faz nova versão de hit de 1988	G1	24/02/2021
Sílvio Santos e Ary Fontoura tomam segunda dose da vacina contra covid-19	Estadão	11/03/2021
Creone, do Trio Parada Dura, se vacina contra Covid-19: “Esperança”	Metrópoles	07/03/2021
Aos 90 anos, Silvio Santos toma dose de reforço da vacina contra a covid	Poder360	28/10/2021
Fã homenageia Paulo Gustavo ao receber 2ª dose de vacina contra a Covid-19: 'Não teve oportunidade'	G1	17/08/2021
Lolita Rodrigues, ícone da TV, é vacinada contra a Covid aos 91 anos	Metrópoles	15/02/2021

Fonte: dados de pesquisa.

Conforme pode ser observado no Gráfico 10, os conjuntos de tópicos 1 e 2 possuem uma alta intersecção. Entretanto, a nuvem de palavras do tópico 2 (Gráfico 11) apresenta termos que distam daqueles observados no tópico 1, tais como: “bolsonaro”, “presidente” e “governo”, remetendo a questões relacionadas a ações, decisões e manifestações associadas ao governo, de forma geral, e ao presidente Jair Bolsonaro. O Quadro 4 apresenta as dez notícias centrais do tópico 2, onde se observa principalmente a exploração da repercussão de falas de autoridades e ações ocorridas no contexto do avanço do processo vacinal no Brasil. A proximidade entre os conjuntos de tópicos 1 e 2 se dá principalmente pela ocorrência de termos compartilhados de forma específica e manifestações por parte de pessoas sobre políticas e ações do governo.

Quadro 4 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 2 (Brasil)

Título da notícia	Jornal	Data de publicação
Presidente da China pede reconhecimento mútuo de vacinas contra Covid-19	G1	30/10/2021
PF vê crime de Bolsonaro por associar vacina anticovid à aids	Poder360	28/12/2022
Quebra de patente de vacinas contra Covid esbarra em resistência do governo e de farmacêuticas	Folha de S. Paulo	12/08/2021
Qual a dificuldade do Brasil para importar insumos e vacinas contra a covid?	Estadão	22/01/2021
Chanceler diz que Brasil mantém posição contrária à quebra de patente de vacinas contra a Covid	G1	06/05/2021
Presidente da Anvisa diz que prazo de cinco dias para aprovar vacinas contra a Covid-19 é 'irreal'	G1	05/02/2021
Vídeo: “Melhor vacina é pegar a doença”, diz Onyx sobre Covid, em debate	Metrópoles	28/10/2022
Atraso para aquisição de vacinas ganha destaque na minuta do relatório da CPI da Covid	G1	19/10/2021
Governo Bolsonaro só gastou 9% da verba emergencial liberada para vacinas contra a Covid-19	Folha de S. Paulo	13/02/2021
Governo ignorou 10 emails da Pfizer sobre vacinas em 1 mês, mostram documentos da CPI da Covid	Folha de S. Paulo	20/05/2021

Fonte: dados de pesquisa.

O tópico 3 apresenta em seu conjunto de termos, representado pela nuvem de palavras (Gráfico 10-c), palavras como: “dose”, “criança”, “prefeitura”, “unidade” e “secretaria”. Se observa neste contexto menção às ações de vacinação contra a COVID-19, com informações sobre pontos de vacinação e testes de infecção. O Quadro 5 apresenta, em maiores detalhes, as dez notícias mais representativas deste tópico, onde se observa, inclusive, a menção à vacinação de crianças e ações de prefeituras.

Quadro 5 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 3 (Brasil)

Título da notícia	Jornal	Data de publicação
Feira de Santana: vacinação contra Covid-19 está disponível à noite em Unidades de Saúde da Família; veja endereço dos postos	G1	29/11/2022
Veja quais são os novos 15 pontos de vacinação contra a Covid-19 em Fortaleza	G1	30/05/2021
Saiba onde fazer testes gratuitos de Covid e se vacinar contra a doença no Recife nesta sexta-feira	G1	04/02/2022
Veja onde adultos e crianças podem se vacinar contra a Covid-19 na capital	G1	02/05/2022
Vacinação contra Covid-19 para crianças de 11 anos em Salvador começa no sábado; veja locais e horários	G1	14/01/2022
Saiba onde fazer testes gratuitos de Covid e se vacinar contra a doença no Recife nesta terça-feira	G1	01/02/2022
Maceió inicia agendamento online para vacinação contra Covid-19	G1	25/05/2021
Saiba onde se vacinar contra Covid, gripe e sarampo em Natal	G1	07/06/2022
Prefeitura atualiza relação de postos de vacinação contra a Covid-19 em Belém	G1	01/03/2021
Novo 'Centro de Vacinação Covid-19' é inaugurado em Caruaru	G1	28/06/2021

Fonte: dados de pesquisa.

A Figura 10-d apresenta a nuvem de palavras para o tópico 4. Neste tópico se observam termos mais relacionados à doença no contexto clínico, centralizados por “caso” e “dose”, e acompanhados por termos como “estudo”, “pandemia”, “morte”, “vírus” e “dado”. Os avanços científicos alcançados ao longo da pandemia da COVID-19 em termos de intervenções médicas, desenvolvimento de vacinas e estudos comportamentais foram notáveis (Tzenios, Chahine e Tazanio, 2023), sendo isso consideravelmente explorado e noticiada pela mídia (Chan, Cheung e Erduran, 2023; Valenti *et alli.*, 2023). O Quadro 6 apresenta o recorte com as dez notícias de maior

representatividade para o grupo, onde se pode observar a ocorrência de notícias sobre a realização de estudos científicos abordando sintomas, riscos e associações a outras doenças.

Quadro 6 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 4 (Brasil)

Título da notícia	Jornal	Data de publicação
Estudo sugere que causa de coágulos pós vacina da Covid é genética	Metrópoles	07/04/2022
Infecção por Covid complementa vacina e cria superproteção, diz estudo	Metrópoles	27/01/2022
Estudo de Harvard mostra que vacinados transmitem Covid 2 dias a menos	Metrópoles	07/12/2021
Vacina anticovid afeta o ciclo menstrual, diz estudo	Poder360	18/07/2022
Vacinas contra Covid-19 são seguras para grávidas, aponta estudo	Metrópoles	25/11/2021
Raio-x mostra que pulmões de vacinados sofrem menos com a Covid	Metrópoles	02/02/2022
Vacina protege melhor contra variantes da covid que infecção	Poder360	17/02/2022
Covid: mais de 60% das reações à vacina são psicológicas, diz estudo	Metrópoles	19/01/2022
Risco de desenvolver coágulo raro é de 8 a 10 vezes maior depois da Covid do que depois de vacina, dizem cientistas de Oxford	G1	15/04/2021
Vacina contra Covid pode auxiliar o tratamento de câncer, diz pesquisa	Metrópoles	14/11/2022

Fonte: dados de pesquisa.

O conjunto de notícias identificado pelo tópico 5 trata sobretudo de questões relacionadas a notícias informativas sobre o processo de vacinação no Brasil. A nuvem de palavras (Gráfico 10-e) apresenta destaque para os termos “dose”, “usf”, “ubs”, “reforço” e “criança”. Tais termos, acompanhados de outros de menor frequência na nuvem de palavras, demonstram a tendência de conteúdos sobre questões relacionadas

ao processo vacinal e às unidades de saúde. No Quadro 7 é possível observar as dez notícias de maior representatividade para o tópico, onde se notam, essencialmente, boletins e informações sobre o avanço da vacinação no país.

Quadro 7 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 5 (Brasil)

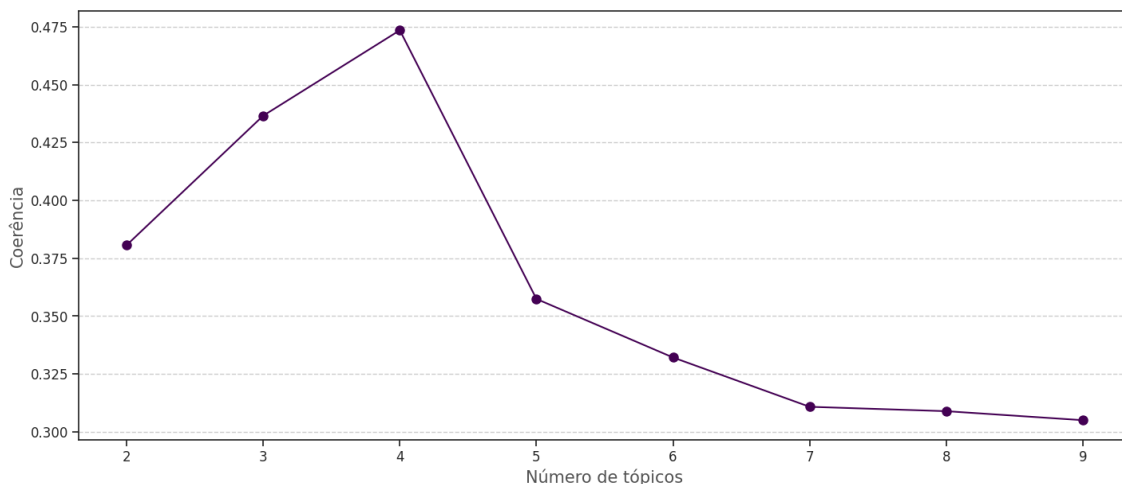
Título da notícia	Jornal	Data de publicação
Brasil tem mais de 75 milhões de pessoas com vacinação completa contra a covid-19	Estadão	14/09/2021
Vacinação no Brasil: mais de 30 milhões de pessoas tomaram as duas doses ou dose única de vacinas contra a Covid	G1	10/07/2021
Brasil aplicou mais de 160 milhões de doses de vacinas contra a covid	Poder360	13/08/2021
Vacinação contra a Covid: mais de 26% da população tomou as doses necessárias e está imunizada; 58,65% tomou a primeira dose	G1	23/08/2021
Vacinação contra a Covid: mais de 39% da população está com esquema vacinal completo; 67% tomou a 1ª dose	G1	22/09/2021
Brasil tem 80 milhões de pessoas totalmente vacinadas contra a covid	Poder360	18/09/2021
Vacinação contra a Covid: mais de 143,3 milhões de pessoas estão totalmente imunizadas	G1	31/12/2021
Brasil aplicou 322 milhões de doses da vacina contra a covid-19	Poder360	16/12/2021
Covid: Brasil chega a 18 milhões de vacinados com dose de reforço	Poder360	07/12/2021
Brasil tem 61,2% da população com o 1º ciclo vacinal completo contra covid	Poder360	24/11/2021

Fonte: dados de pesquisa.

De maneira semelhante ao que foi realizado com o conjunto amostral de notícias brasileiras, foram aplicadas as estratégias de identificação temática com base no uso de soluções computacionais inteligentes para as notícias portuguesas. Inicialmente se realizou a análise de coerência, a qual indicou o número de tópicos igual a 4, conforme

está apresentado no Gráfico 12. A partir disso, foi realizada a inferência de tópicos, sendo os quatro conjuntos apresentados no Gráfico 13, onde é possível observar que os conjuntos 3 e 4 encontram-se relativamente próximos, enquanto os conjuntos 1 e 2 distam consideravelmente entre si e dos demais.

Gráfico 12 - Análise de coerência de tópicos (Portugal)

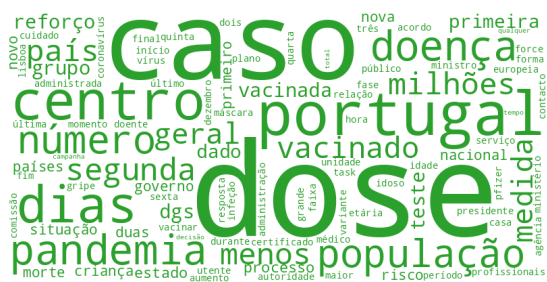


Fonte: dados de pesquisa.

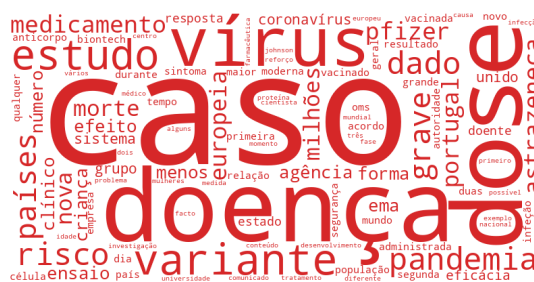
O Gráfico 14 apresenta as nuvens de palavras segmentadas por tópicos para o conjunto amostral de notícias portuguesas. É interessante observar uma grande variação nos termos presentes em cada uma das nuvens de palavras, sendo que apenas os conjuntos de tópicos 3 e 4 compartilham algumas das palavras principais, com destaque para o termo “caso”.

O conjunto de tópicos 1 apresenta forte relação com temas relacionados a autoridades governamentais (Gráfico 13-a) e questões internacionais, onde se observam termos como: “presidente”, “governo”, “estado”, “ministro”, “internacional” e “mundo”. Um ponto que chama a atenção é a notável ocorrência do termo “bolsonaro”. O Quadro 8 apresenta as dez notícias mais representativas, onde é possível notar a presença de notícias que remetem a ações e falas do presidente brasileiro Jair Bolsonaro durante o período pandêmico. Para além disso se observam ainda notícias no contexto internacional, tanto relacionadas a autoridades governamentais quanto a desportistas. Entre o recorte de notícias de maior representatividade se observa apenas uma notícia

(c) Tópico 3



(d) Tópico 4



Fonte: dados de pesquisa.

Quadro 8 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 1 (Portugal)

Título da notícia	Jornal	Data de publicação
Bolsonaro relacionou as vacinas da covid-19 com a sida e foi suspenso do YouTube	Público	26/10/2021
João Costa abdica dos Jogos Olímpicos por falta de vacina contra a covid-19	Público	14/05/2021
Juíza do Supremo recusa-se a arquivar investigação a Bolsonaro sobre compra de vacina da covid	Público	30/03/2022
Covid-19: EUA estudam possível certificado de vacinação para viagens ao estrangeiro	Público	28/05/2021
André Ventura lamenta "vergonha de Presidente" e promete ser último vacinado contra a Covid-19	Correio da Manhã	22/01/2021
É o segundo homem mais rápido depois de Bolt mas não tem dúvidas: "Prefiro falhar Jogos do que receber vacina da Covid-19"	Observador	01/03/2021
No último briefing na Casa Branca, Fauci apela aos americanos para tomarem reforço da vacina anti-Covid	Diário de Notícias	22/11/2022
UE exportou mais de mil milhões de vacinas contra a covid-19 e doou 87 milhões	Público	18/10/2021
Covid-19. Amnistia Internacional acusa farmacêuticas de recusar aumento da oferta de vacinas	Público	22/09/2021
Com receio do Ano Novo, China está a vacinar 50 milhões de pessoas contra a covid-19	Público	22/01/2021

Fonte: dados de pesquisa.

O conjunto de tópicos 2 revela termos principais que vão de encontro a questões diretas da pandemia de COVID-19, tais como: “doente”, “médico”, “situação”, “sistema” e “serviço”, conforme pode ser observado no Gráfico 13-b. O Quadro 9 apresenta as dez notícias de maior representatividade, para o qual se nota uma concentração de notícias que relatam a vacinação de indivíduos, onde há uma apresentação de casos denotados como exemplos da eficácia e importância do imunizante, e situações que foram de encontro à quebra de regras delimitadas no contexto do esquema vacinal.

Quadro 9 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 2 (Portugal)

Título da notícia	Jornal	Data de publicação
Familiares da assistente do IPO que morreu após vacina contra a Covid-19 esperam resultado da autópsia	Correio da Manhã	05/01/2021
GNR de Viseu surdo há um ano depois de tomar vacina Covid da AstraZeneca	Correio da Manhã	19/01/2022
Administrador de hospital vacina filha e mulher contra Covid-19	Correio da Manhã	01/02/2021
Realiza-se esta quarta-feira funeral de auxiliar do IPO que morreu após vacina da Covid-19	Correio da Manhã	06/02/2021
Médicos alertam para sinais da DPOC e defendem reforço na vacina contra a covid-19	Público	16/11/2021
Covid-19. Há 126 funcionários da Segurança Social suspeitos de terem sido vacinados antes do tempo	Público	28/01/2021
Padre vacinado contra a Covid por dirigir dois lares em Tavira	Correio da Manhã	03/01/2021
Mulher portuguesa vacinada contra a Covid-19 aos 111 anos	Correio da Manhã	13/01/2021
Covid-19: chefe das Forças Armadas espanhol demite-se após vacinar-se antes do tempo	Público	23/01/2021

Posicionamentos Jornalísticos sobre a Vacinação da COVID-19

Mulher morre com covid após esconder dos médicos que o certificado de vacinação era falso	Diário de Notícias	11/12/2021
---	--------------------	------------

Fonte: dados de pesquisa.

O conjunto referente ao tópico 3 traz “caso” e “dose” como termos principais, seguidos de termos como “vacinado”, “população”, “número” e “grupo”, conforme é apresentado no Gráfico 13-c. Este conjunto é composto por notícias que possuem o foco na informação sobre o avanço da campanha vacinal em Portugal, tanto em termos numérico quanto na divulgação de calendários (Quadro 10).

Quadro 10 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 3 (Portugal)

Título da notícia	Jornal	Data de publicação
Covid-19. Quase metade dos portugueses com vacinação completa	Público	21/07/2021
Mais de dois milhões já receberam as vacinas contra a Covid-19 e gripe	Observador	16/11/2022
Covid-19: centros vão encher-se no fim-de-semana para vacinar mais de 100 mil jovens com 16 e 17 anos	Público	13/08/2021
Covid-19: Portugal continua com 84% da população totalmente vacinada. Norte tem maior percentagem	Público	05/10/2021
DGS anuncia que pessoas com mais de 80 anos vão receber quarta dose da vacina contra a Covid-19 a partir de segunda-feira	Observador	12/05/2022
Covid-19: vacina de reforço disponível em "casa aberta" para maiores de 75 anos	Observador	04/11/2022
Covid-19: Portugal já administrou mais de 800 mil doses de reforço da vacina	Público	23/11/2021
Covid-19: quarta dose da vacina deverá ser administrada antes do Outono	Público	21/04/2022
Covid-19. Idosos começam hoje a receber segundo reforço da vacina	Público	16/05/2022

Posicionamentos Jornalísticos sobre a Vacinação da COVID-19

Covid-19. Quase 70% dos portugueses já têm uma dose de vacina, 57% têm vacinação completa	Público	03/08/2021
---	---------	------------

Fonte: dados de pesquisa.

Por fim, o conjunto de tópicos 4, o qual foi identificado com uma proximidade relativa ao conjunto de tópicos 3, apresenta destaque para os termos “caso”, “doença”, “vírus” e “dose”, circundado por termos como “estudo”, “medicamento”, “variante” e “risco” (Gráfico 13-d). O Quadro 11 apresenta o recorte das dez notícias de maior representatividade para o conjunto de tópicos 4, onde se nota uma concentração de notícias sobre estudos remetendo, principalmente, a questões que extrapolam o próprio cerne da doença em si, assim como questões relacionadas à eficácia da vacina face às variantes identificadas ao longo do período pandêmico.

Quadro 11 - Dez notícias com maior contribuição na identificação do conjunto de tópicos 4 (Portugal)

Título da notícia	Jornal	Data de publicação
Estudo. Vacinas não afetam fertilidade dos casais, mas Covid reduz a dos homens	Observador	22/01/2022
Covid-19: Pfizer-BioNTech dizem que vacina é eficaz face às mutações do vírus	Público	28/01/2021
Cientistas testam com êxito vacina contra a sida baseada na técnica mRNA da covid-19	Diário de Notícias	09/12/2021
Covid-19: vacinas são menos eficazes contra a variante da África do Sul, mas parecem prevenir casos graves	Público	07/02/2021
Covid-19: estudo indica que vacinas podem ter causado alterações na menstruação	Público	21/07/2022
Covid-19: EMA inicia estudo de vacina da Pfizer adaptada à variante Ómicron	Público	15/06/2022
Poderão as vacinas contra a Covid-19 afetar o ciclo menstrual? Imunologista defende que se investigue	Observador	16/09/2021
Covid-19: estudo do Insa sugere menor eficácia das vacinas ARNm contra variante Delta	Público	24/08/2021

Está confirmado. Estudo pioneiro identifica alterações na menstruação por causa das vacinas contra a Covid-19	Observador	21/07/2022
Pericardite e miocardite: inflamações cardíacas associadas à covid-19 e à vacina	Jornal de Notícias	12/08/2021

Fonte: dados de pesquisa.

3.3 Posicionamentos e produção de sentidos

A partir da realização dos procedimentos de identificação temática, foram realizadas investigações sobre os posicionamentos e sentidos com base nos sentimentos identificados por meio da aplicação de uma solução computacional de análise de sentimento baseada no *fork* LeIA (Seção 1.7). Os resultados obtidos foram quantificados por meio de uma visualização comparativa entre os jornais e através da apresentação dos indicadores gerais de sentimento segmentados por jornal. Cabe ressaltar que as análises de sentimentos na presente investigação possuem o foco nos conteúdos jornalísticos que remetem a autoridades governamentais e relações com a vacinação contra a COVID-19. Neste sentido, toma-se como base a própria filtragem realizada para a consolidação do corpus textual, a qual considera apenas notícias que contenham os termos “vacina” e “covid”, juntamente com os resultados obtidos por meio do processo de identificação temática (tópicos).

No contexto das notícias brasileiras se considerou o conjunto de tópicos 2, a partir do qual se fez a inferência de sentimentos. O Gráfico 15 apresenta uma visão geral do posicionamento dos jornais brasileiros analisados. Para construção do gráfico se considerou a média aritmética da soma dos sentimentos relativos a cada uma das notícias que compõem o corpus textual. É possível observar que todos os jornais apresentaram uma tendência majoritária a um tom negativo, com destaque para o jornal Folha de S. Paulo. A Tabela 2 apresenta os percentuais gerais obtidos para os jornais brasileiros.

Tabela 2 - Sentimento médio percentual (Brasil)

Jornal	Sentimento médio
Folha de S. Paulo	-7,51%
Metrópoles	-6,94%
Estadão	-6,77%
G1 Notícias	-6,42%
Poder360	-5,34%

Fonte: dados de pesquisa.

Gráfico 15 - Infográfico de posicionamento jornalístico (Brasil)

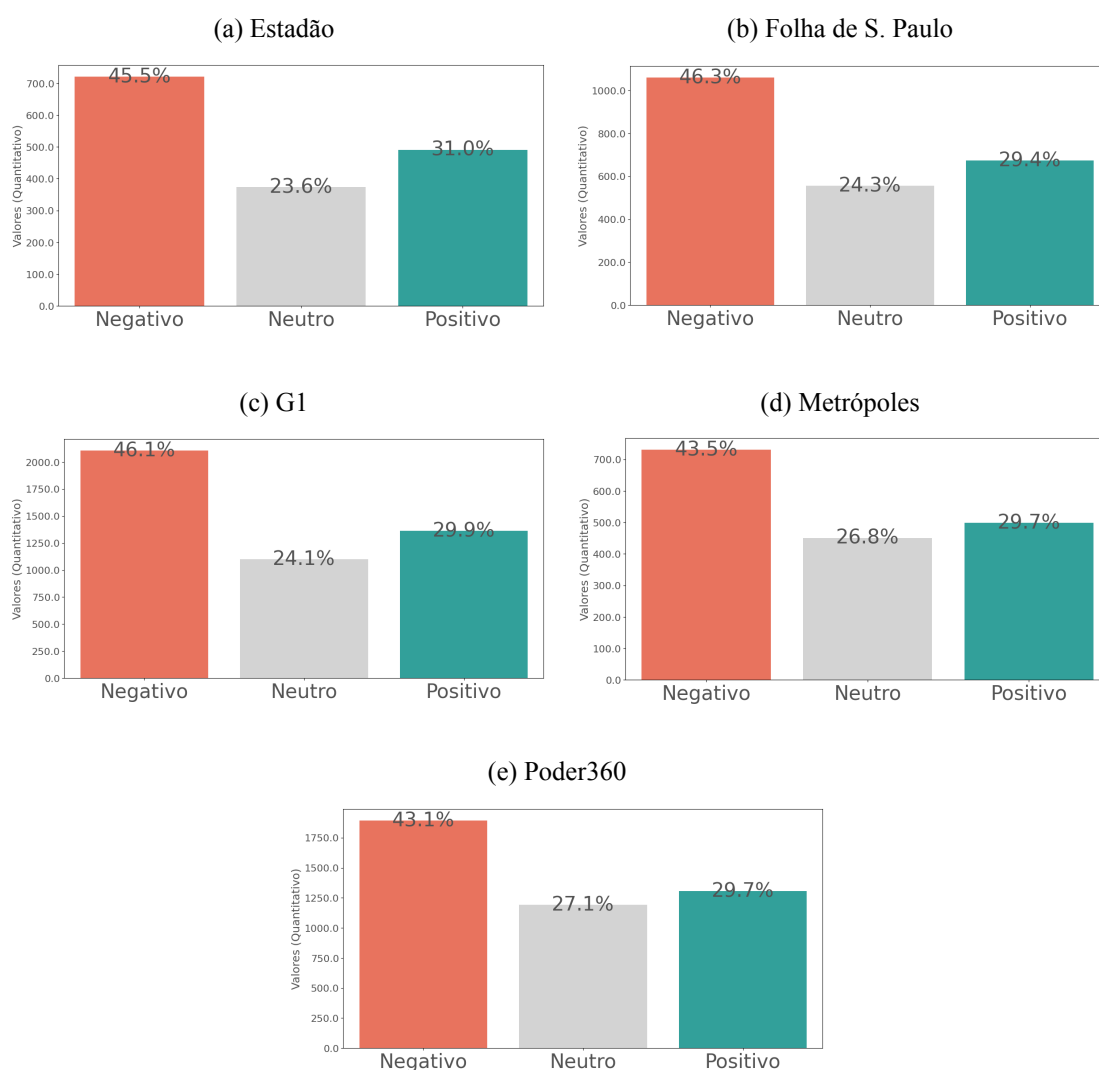


Fonte: dados de pesquisa.

O Gráfico 16 apresenta, de forma segmentada, o percentual de sentimentos para cada um dos jornais brasileiros. Para todos os jornais o percentual de notícias com sentimento inferido como negativo ficou próximo de 50% e como positivo com valores próximos a 30%. Em parte, a expressão anunciada pelo considerável número de notícias identificadas com sentimento negativo vai de encontro a, conforme abordado por Mello (2021), um embate entre o então presidente Jair Bolsonaro e a mídia brasileira, quer seja na comunicação da informação, ou ainda na argumentação que eventualmente evoca a emoção ou mesmo a visão ideológica.

O grande ponto é que, sendo a mídia favorável ou não a Bolsonaro, para a compreensão teórica do espetáculo a política atualmente está completamente inserida numa lógica de paixões, emoções e dramatização. Bolsonaro e a imprensa brasileira possuem todos os mecanismos (atores e cenários, por que não dizer) necessários à lógica do espetáculo. A mídia não controla Bolsonaro e Bolsonaro nem tanto a controla. Mas as disputas e dramatizações estão presentes em todos os seus episódios. (Cioccarri e Persichetti, 2021, p. 212).

Gráfico 16 - Variações de sentimentos por jornal (Brasil)



Fonte: dados de pesquisa.

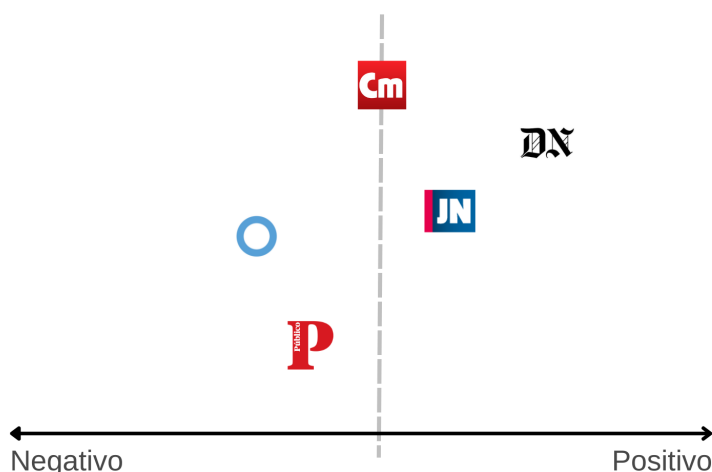
Por outro lado, as análises de sentimento realizadas no contexto das notícias portuguesas revelaram um cenário de polaridade entre os jornais, indo do jornal Observador com um maior registro de notícias classificadas com sentimento negativo para o Diário de Notícias com um retrato oposto, apontando para um conteúdo classificado com sentimentos positivos, o que pode ser notado na Tabela 3 e no Gráfico 17.

Tabela 3 - Sentimento médio percentual (Portugal)

Jornal	Sentimento médio
Observador	-19,93%
Público	-8,88%
Correio da Manhã	0,00%
Jornal de Notícias	5,88%
Diário de Notícias	11,59%

Fonte: dados de pesquisa.

Gráfico 17 - Infográfico de posicionamento jornalístico (Portugal)



Fonte: dados de pesquisa.

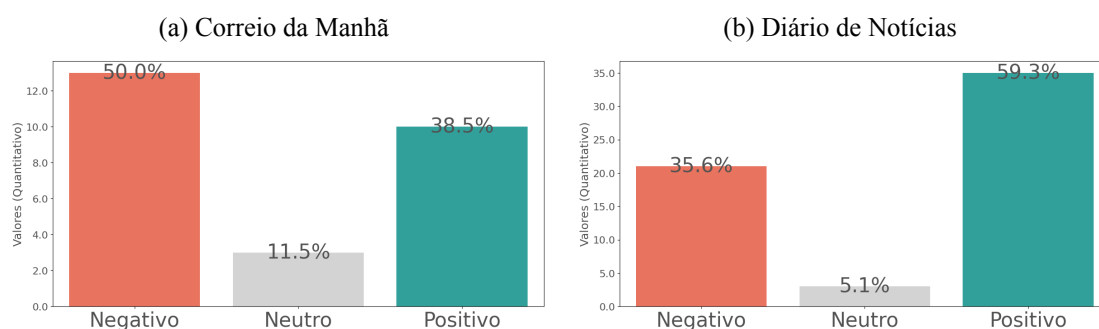
O Gráfico 18 apresenta a distribuição percentual quantitativa de sentimento para os jornais portugueses. Observa-se que os jornais Correio da Manhã, Observador e

Público apresentam percentual de notícias classificadas com sentimento negativo superior a 50%. É importante pontuar que na análise geral ponderada, o Correio da Manhã apresentou uma tendência central. Por outro lado, os jornais Diário de Notícias e Jornal de Notícias se destacam pelo quantitativo de notícias com classificação de sentimentos positivos, ambos com valores superiores a 50%.

Quanto às notícias que tendem ao sentimento negativo se destacam falas de autoridades sobre fatos ocorridos em Portugal em decorrência da crise sanitária da COVID-19, se referindo a questões como saúde pública e economia, tal como em notícia publicada pelo jornal Observador sobre a disponibilização de informações relativas à disponibilização de informações sobre o trânsito livre de pessoas: “Costa lamenta a “falta de clareza” na informação das autoridades portuguesas, a qual “não se sabe muito bem onde se pode encontrar”, estando dispersa nos sites do Turismo de Portugal e do Ministério dos Negócios Estrangeiros (MNE).”³.

Do outro lado se destacam notícias com foco em manifestações sobre os avanços das ações de enfrentamento da COVID-19 e a retomada de atividades sociais e econômicas, tal como a afirmação do primeiro-ministro António Costa abordada em notícia do jornal Diário de Notícias: “O primeiro-ministro, António Costa, disse hoje que a indústria portuguesa irá contribuir certamente para o "esforço coletivo" acordado entre todos os Estados-membros no sentido de aumentar a capacidade de produção de vacinas contra a covid-19 na União Europeia”⁴.

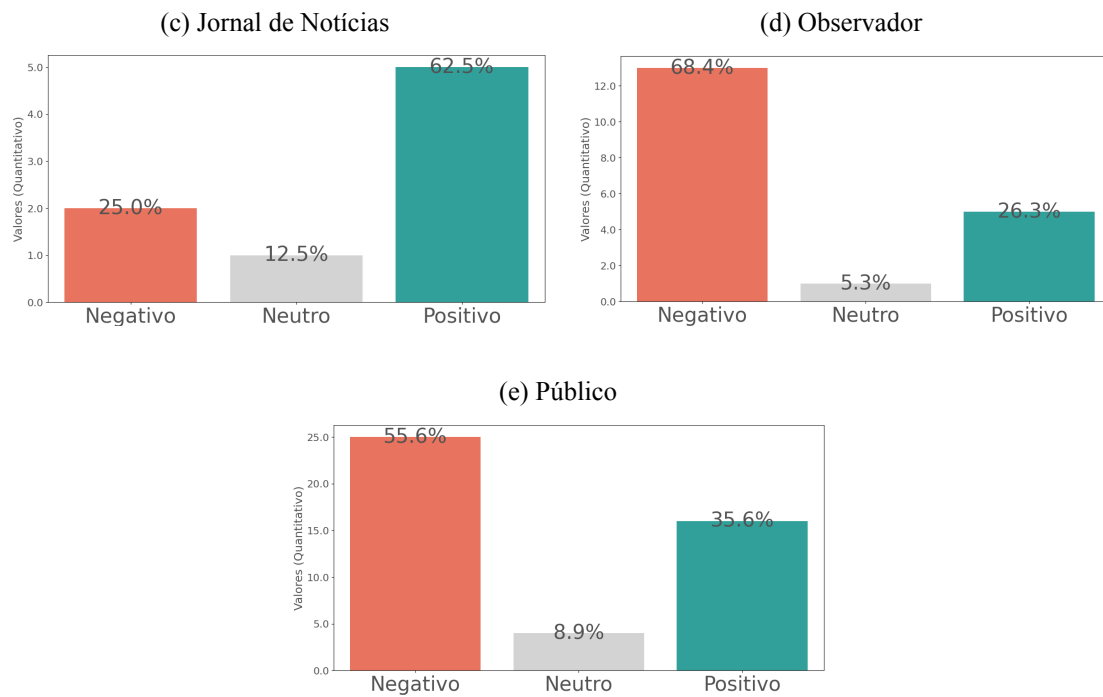
Gráfico 18 - Variações de sentimentos por jornal (Portugal)



³ Dezenas de portugueses impedidos de viajar por causa de novas regras. Observador. <https://observador.pt/2021/12/17/dezenas-de-portugueses-impedidos-de-viajar-por-causa-de-novas-regras/>. Acesso em: 17 de maio de 2023.

⁴ Indústria portuguesa contribuirá para esforço coletivo da UE no aumento da produção de vacinas, disse Costa. Diário de Notícias. Disponível em: <https://www.dn.pt/politica/industria-portuguesa-contribuira-para-esforco-coletivo-da-ue-no-aumento-da-producao-de-vacinas-disse-costa-13395870.html>. Acesso em: 17 de maio de 2023.

Posicionamentos Jornalísticos sobre a Vacinação da COVID-19



Fonte: dados de pesquisa.

CONCLUSÃO

A importância de se compreender os sentimentos implícitos na circulação de conteúdos e informações em jornais eletrônicos é algo de grande importância, principalmente ao se tratar de questões relacionadas à saúde pública. Nesta investigação foram apresentados resultados comparativos que abordaram jornais eletrônicos brasileiros e portugueses, destacando um levantamento de temas e sentimentos associados aos textos veiculados. A aplicação de soluções computacionais para extração e análise de dados se revelou consideravelmente favorável ao volume de documentos recuperados, possibilitando a sumarização de dados e a geração de indicadores úteis para o direcionamento e construção de análises mais orientadas.

Notadamente os jornais brasileiros considerados na composição do corpus textual revelaram uma tendência de maior volume de textos com classificação de sentimentos negativos, o que, de certa forma, vai de encontro a questões que estiveram dentro dos holofotes da mídia ao longo do período pandêmico, a destacar manifestações e ações ligadas ao então presidente do Brasil, Jair Bolsonaro, assim como medidas e estratégias governamentais. Por outro lado, no conjunto de dados dos jornais portugueses se observa uma polarização de sentimento entre os jornais considerados, que tanto revela um equilíbrio saudável ao debate público e formação de opinião, quando reafirma visões e posicionamentos explícitos dos veículos de comunicação.

Cabe ressaltar que o delineamento dos resultados alcançados abrem a oportunidade para aprofundamento posterior da investigação, tanto por meio da aplicação de métodos de análise de conteúdo no conjunto de dados textuais categorizados por tópicos, assim como em um ponto de vista orientado pelos sentimentos identificados e as polarizações presentes dentro de um mesmo veículo de comunicação e na composição comparativa dos jornais eletrônicos explorados, o que pode ser realizado, por exemplo, através de uma análise de cartografia de controvérsias.

REFERÊNCIAS BIBLIOGRÁFICAS

- Agência Brasil (2022). *Retrospectiva 2022: confirma as principais notícias de janeiro*. Agência Brasil. Disponível em <<https://agenciabrasil.ebc.com.br/geral/noticia/2022-09/retrospectiva-2022-confirma-principais-noticias-de-janeiro>>. [Consultado em 17/05/2023].
- Albuquerque, A. (2018). Journalism and Multiple Modernities: The Folha de S. Paulo Reform in Brazil. *Journalism Studies*, 20(11), pp. 1546-1562. <https://doi.org/10.1080/1461670X.2018.1528881>.
- Alemida, R. (2018). *LeIA - Léxico para Inferência Adaptada*. Disponível em: <<https://github.com/rafjaa/LeIA>>. [Consultado em 10/03/2023].
- Bagozzi, B. E. e Schrodtt, P. A. (2012). The dimensionality of political news reports. *In: European Political Science Association Meetings*, Berlin.
- Bastos, H. (2011). Para uma história do ciberjornalismo em Portugal : das origens às múltiplas plataformas. *In: Actas do Congresso Mundial de Comunicação Ibero-Americana*, São Paulo. Disponível em <<http://hdl.handle.net/10216/57427>>. [Consultado em 15/05/2023].
- Barsotti, A. (2015). Extra! Extra! As origens da primeira página moderna no jornalismo brasileiro. *In: XXXVIII Congresso Brasileiro de Ciências da Comunicação*, Rio de Janeiro: Intercom. Disponível em <<https://www.portalintercom.org.br/anais/nacional2015/resumos/R10-2787-1.pdf>>. [Consultado em 18/05/2023]
- Belda, R. F., Gonçalves, J. (2021). Perspectivas para o mercado de jornalismo: uma análise das tendências de inovação em gestão de mídia. *Revista Observatório*, 7(4), pp. a3pt. <https://doi.org/10.20873/uft.2447-4266.2021v7n4a3pt>.
- Berbel, D. B. (2017). *Controvérsias Científicas e Tecnológicas no Jornal “A Província De São Paulo”*: 1875-1889. Tese de Doutorado, Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Universidade Federal de São Carlos.
- Bittencourt, M. C. A. (2018). Jornalismo, inovação e empreendedorismo: questões sobre modelos de negócio em contexto de crise. *Liberio*, 41, pp. 74-87.

Blei, D. M., Ng, A. Y. e Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning*, 3, pp. 993-1022.

Blei, D. M. e Laffert, J. D. (2007). A correlated topic model of science. *Ann. Appl. Stat.*, 1(1), pp. 17-35. <https://doi.org/10.1214/07-AOAS114>.

Bou-Karroum, L. *et alli*. (2017). Using media to impact health policy-making: an integrative systematic review. *Implementation Science*, 12(1), pp. 1-15.

Brito, J. M. (2010). *As estratégias de comunicabilidade de um jornal popular*. O caso do diário português - Correio da Manhã. Dissertação de Mestrado, Ciências da Comunicação, Universidade Nova de Lisboa.

Broucke, S. e Baesens, B. (2018). *Practical Web Scraping for Data Science: Best Practices and Examples with Python*. Apress, New York, USA.

Budarick, J. (2023). Media, democracy and pluralism: exploring a radical response to the crisis of journalism. *Journalism Studies*.
<https://doi.org/10.1080/1461670X.2023.2173959>.

CM (2020). *CM Online aumenta liderança digital*. Correio da Manhã. Disponível em <<https://www.cmjornal.pt/tv-media/detalhe/lideranca--cm-online-aumenta-audiencia-digital-e-continua-a-ser-o-site-mais-lido>>. [Consultado em 15/05/2023].

Castellanos, E., Xie, H. e Brenner, P. (2021). Global News Sentiment Analysis. In: Yang, Z., von Briesen, E. (eds) *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*. CSSSA 2020. Springer Proceedings in Complexity, Springer, Cham.
https://doi.org/10.1007/978-3-030-77517-9_9.

Catalina-Garcia, B., Sousa, J. P. e Sousa, L. S. C. S. (2019). Consumo de noticias y percepción de fake news entre estudiantes de Comunicación de Brasil, España y Portugal. *Revista de Comunicación*, 18(2), pp. 93-115.
<http://dx.doi.org/10.26441/rc18.2-2019-a5>.

Chan, H.-Y., Cheung, K. K. C. e Erduran, S. (2023). Science communication in the media and human mobility during the COVID-19 pandemic: a time series and content. *Public Health*, 218, pp. 106-113. <https://doi.org/10.1016/j.puhe.2023.03.001>.

Che, S., Nan, D., Kamphuis, P. e Kim, J. H. (2021). A comparative analysis of attention to facial recognition payment between China and South Korea: a news analysis using latent dirichlet allocation. In: Stephanidis, C., Antona, M., Ntoa, S. (eds). *HCI International 2021 - Posters*. HCII 2021. Communications in Computer and Information Science, vol 1420. Springer, Cham.
https://doi.org/10.1007/978-3-030-78642-7_11.

Chowdhary, K. R. (2020). *Fundamentals of Artificial Intelligence*. Springer, Berlim.

Cioccari, D. e Persichetti, S. (2021). Armas, ódio, medo e espetáculo em Jair Bolsonaro. *Revista Alterjor*, 18(2), pp. 201-214.

Conde, A. S., Magalhães, A. e Moreira, F. T. (2018). Uma visão da internet das coisas na imprensa nacional: o jornal Observador. *Internet Latent Corpus Journal*, 8(1), pp. 14-30. Disponível em <<http://revistas.ua.pt//index.php/ilcj/index>>. [Consultado em 15/05/2023].

Cordeiro, D. F. *et alii*. (2021). Análise de dados aplicada ao levantamento da cobertura noticiosa sobre agricultura familiar no G1 Notícias. *Panorama*, 11(2), pp. 19-25.

Cordeiro, D. F. *et alii*. (2022). Cartografando comentários e sentimentos no perfil de Jair Bolsonaro no Instagram acerca da Covid-19. *Galáxia*, 47, pp. 1-24.

Cordeiro, D. F. *et alii*. (2023). Incêndios na Amazônia a partir de uma perspectiva de circulação de tweets. *Revista Observatório*, no prelo.

Costa, G. M. F. Q. (2001). As novas tecnologias na imprensa brasileira. In: *XXIV Congresso Brasileiro de Ciências da Comunicação*. Campo Grande: INTERCOM.

Cota, V. R. e Maciel, R. M. (2016). Tomando partido: um caso de influência de ideologia política na divulgação da neurociência. *Revista da Biologia*, 15(1), pp. 74-80. <https://doi.org/10.7594/revbio.15.01.09>.

Cunha, A. R. e Moreira, T. M. M. (2023). Relevância de fontes acadêmicas da Universidade Estadual do Ceará para o jornalismo de saúde no combate às fake news: estudo de caso. *Reciis – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, 17(1), pp. 84-93. <https://doi.org/10.29397/reciis.v17i1.3479>.

Devlin, J., Chang, M., Lee, K., Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv*, 2018. DOI: <https://doi.org/10.48550/arXiv.1810.04805>.

Dhillon, I. S. e Sra, S. (2005). Generalized nonnegative matrix approximations with Bregman divergences. In: *Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS'05)*. MIT Press, Cambridge, MA, USA, pp. 283-290.

Duarte, A. I. C. (2013). *O Público e o Jornalismo de Referência*. Dissertação de Mestrado. Universidade do Porto.

Durmus, K. (2021). The Impact of Cyber Culture on New Media Consumers. In: David, S., Anand, R. S., Jeyakrishnan, V. e Niranjanamurthy, M. *Security Issues and Privacy Concerns in Industry 4.0 Applications*. John Wiley & Sons. <https://doi.org/10.1002/9781119776529.ch12>.

Estadão (2021). 'Estadão' bate recorde de audiência na internet em março. Disponível em

<<https://www.estadao.com.br/economia/estadao-bate-recorde-de-audiencia-na-internet-em-marco/>>. [Consultado em 18/05/2023].

Fernandes, I., Moro, S. e Cortez, P. (2023). Data science, machine learning and big data in digital journalism: a survey of state-of-the-art, challenges and opportunities. *Expert Systems with Applications*, 221, 119795. <https://doi.org/10.1016/j.eswa.2023.119795>.

Ferrari, P. (2014). *A Força da Midia Social: interface e linguagem jornalística no ambiente digital*. Estação das Letras e Cores, São Paulo.

Fitzgerald, M. (2012). *Introdução às Expressões Regulares*. Novatec, São Paulo.

Folha de S. Paulo. *Circulação e Audiência*, 2023. Disponível em <<https://www1.folha.uol.com.br/institucional/circulacao.shtml>>. [Consultado em 03/05/2023].

G1. *Sobre o g1*, n. d. Disponível em <<https://g1.globo.com/institucional/sobre-o-g1.ghtml>>. [Consultado em 04/05/2023].

Gargi, N. e Rani, R. (2017). Analysis and visualization of Twitter data using k-means clustering. In: *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India. <https://doi.org/10.1109/ICCONS.2017.8250547>.

Gomes, P. M. (2012). O Diário de Notícias e o 25 de Novembro de 1975: um momento de mudanças. *Comunicação Pública*, 17(11), pp. 23-40. <https://doi.org/10.4000/cp.178>.

Gonçalves, J. e Capoano, E. (2017). Reorganização de redações no Brasil: os casos dos ecossistemas endógenos do Estadão e da sucursal do Huffington Post. *Comunicologia - Revista De Comunicação Da Universidade Católica De Brasília*, 10(1), pp. 17-28. <https://doi.org/10.24860/comunicologia.v10i1.8127>.

Hallin, D. C., Brandt, M. e Briggs, C. L. (2013). Biomedicalization and the public sphere: newspaper coverage of health and medicine, 1960s e 2000s. *Social Science and Medicine*, 96, pp. 121-128.

Helberger, N. (2020). The political power of platforms: how current attempts to regulate misinformation amplify opinion power. *Digital Journalism*, 8(6), pp. 842-854. <https://doi.org/10.1080/21670811.2020.1773888>.

Heldman, A. B., Schindelar, J. e Weaver, J. B. (2013). Social media engagement and public health communication: implications for public health organizations being truly “social”. *Public Health Reviews*, 35, 13. <https://doi.org/10.1007/BF03391698>.

Hidayat, W. e Yaqin, A. (2019). Business Trends Based on News Portal Websites for Analysis of Big Data Using K-Means Clustering. In: *2019 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, 2019, pp. 445-450. <https://doi.org/10.1109/ICOIACT46704.2019.8938413>.

Hinnant, A., Len-Ríos, M. E. e Oh, H. J. Are Health Journalists' Practices Tied to Their Perceptions of Audience? An Attribution and Expectancy-Value Approach. *Health Communications*, 27(3), pp. 234-243.

Hortelan, M. S., Almeida, M. L., Fumincelli, L., Zilly, A., Nihei, O. K., Peres, A. M., Antônio Sobrinho, R. e Pereira, P. E. (2019). Papel do gestor de saúde pública em região de fronteira: scoping review. *Acta Paul. Enferm.*, 32(2), pp. 229-236. <https://doi.org/10.1590/1982-0194201900031>.

Hrabiec-Hojda, P. Trzeciakowska, J. (2019). *Google Hacking*. Cracóvia, Infomedia Group.

Hummel, L. F., Volcan, T. O. (2017). Uma análise da cobertura jornalística do portal G1 na operação “Carne Fraca”. In: *XVIII Congresso de Ciências da Comunicação na Região Sul*. Caxias do Sul, INTERCOM. Disponível em: <<https://www.portalintercom.org.br/anais/sul2017/resumos/R55-0432-1.pdf>>. [Consultado em 04/05/2023].

Hussein, D. M. E. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, 30(4), pp. 330-338. <https://doi.org/10.1016/j.jksues.2016.04.002>.

Hutto, C. J. e Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI.

Jelodar, H. *et alii*. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78, pp. 15169-15211. <https://doi.org/10.1007/s11042-018-6894-4>.

JN (2022). *JN volta a ser o jornal online mais lido em Portugal*. Jornal de Notícias. Disponível em <<https://www.jn.pt/artes/media/jn-volta-a-ser-o-jornal-online-mais-lido-em-portugal-15333074.html>>. [Consultado em 15/05/2023].

Khemani, B. e Adgaonkar, A. (2021). A Review on Reddit News Headlines with NLTK tool. In: *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*. <https://dx.doi.org/10.2139/ssrn.3834240>.

Khurana, D., Koli, A., Khatter, K. *et alli*. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82, pp. 3713-3744. <https://doi.org/10.1007/s11042-022-13428-4>.

Lee, S. (2020). Topic modeling of newspaper articles on government 'Senior job program' via latent dirichlet allocation. *Journal of Digital Convergence*, 18(10), pp. 537-546. <https://doi.org/10.14400/JDC.2020.18.10.537>.

- Lee, N. Y. e McElroy, K. (2019). Online comments: The nature of comments on health journalism. *Computers in Human Behaviors*, 92, pp. 282-287.
- Li, Z., Fan, Y., Jiang, B., Lei, T. e Liu, W. (2019). A survey on sentiment analysis and opinion mining for social multimedia. *Multimed Tools and Applications*, 78, pp. 6939-6967. <https://doi.org/10.1007/s11042-018-6445-z>.
- Lima, H. (2017). A evolução editorial do Jornal de Notícias e a inclusão de elementos de jornalismo popular. *Revista Portuguesa de História da Comunicação*, 1. Disponível em: <https://hdl.handle.net/10216/110514>. Acesso em: 15 de maio de 2023.
- Liu, B. e Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal, C., Zhai, C. (eds) *Mining Text Data*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_13.
- Lopes, F. et alli. (2021). Covid-19: uma pandemia que reconfigura o jornalismo? *Media & Jornalismo*, 21(39), pp. 57-75.
- Lopes, F. et alli. (2020). COVID-19: quando o jornalismo se assume como uma frente de combate à pandemia. In: Martins, M. e Rodrigues, E. *A Universidade do Minho em tempos de pandemia: Tomo III: Projeções*. Braga, UMinho Editora.
- Lucy, L., Demszky, D., Bromley, P. e Jurafsky, D. (2020). Content analysis of textbooks via natural language processing: findings on gender, race, and ethnicity in Texas U.S. history textbooks. *AERA Open*, 6(3), pp. 1-27. <https://doi.org/10.1177/2332858420940312>.
- Malinverni, C., Cuenca, A. M. B. e Brigagão, J. I. M. (2012). Epidemia midiática: produção de sentidos e configuração social da febre amarela na cobertura jornalística, 2007-2008. *Physis - Revista de Saúde Coletiva*, 22(3), pp. 853-872. <https://doi.org/10.1590/S0103-73312012000300002>.
- Marques, J. D. R. (2020). *Jornais e Agências: A Relação entre o Observador e a Lusa*. Dissertação de Mestrado, Universidade Nova de Lisboa.
- Mello, R. A. (2021). Bolsonaro e o jornalismo em conflito midiático. *Revista de Estudos da Linguagem*, 29(4), pp. 2485-2508. <http://dx.doi.org/10.17851/2237-2083.29.4.2485-2508>.
- Memória Globo. *G1: O portal de notícias da Globo é líder de audiência no jornalismo digital*. Disponível em: <https://memoriaglobo.globo.com/produtos-digitais/g1/noticia/g1.ghtml>. [Consultado em 04/05/2023].
- Metrópoles. *Quem somos*, n. d. Disponível em <https://www.metropoles.com/quem-somos>. [Consultado em 05/05/2023].

Miguel, L. F. e Biroli, F. (2010). produção da imparcialidade: a construção do discurso universal a partir da perspectiva jornalística. *Revista Brasileira de Ciências Sociais*, 25(73), pp. 59-74. <https://doi.org/10.1590/S0102-69092010000200004>.

Mikolov, T., Chen, K., Corrado, G. e Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv*. <https://doi.org/10.48550/arXiv.1301.3781>.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. e Dean, J. (2013b). Distributed Representations of Words and Phrases and their Compositionality. *arXiv*. <https://doi.org/10.48550/arXiv.1310.4546>.

Mohammadi S. *et alli*. (2021). Why medical journalism wins public health journalism: systems thinking recommendations for health-promoting media. *Health Education*, 121 (2), pp. 161-173.

Santos, N. (2021). Fontes de informação nas redes pró e contra o discurso de Bolsonaro sobre o Coronavírus. *E-Compós*, 24. <https://doi.org/10.30962/ec.2210>.

Nguyen, T. L. (2018). A Framework for Five Big V's of Big Data and Organizational Culture in Firms. In: *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, pp. 5411-5413. <https://doi.org/10.1109/BigData.2018.8622377>.

Observador (2016). *Saimos do Netscope. E mostramos os nossos números*. Observador. Disponível em <<https://observador.pt/2016/04/13/saimos-do-netscope-mostramos-os-numeros/>>. [Consultado em 15/05/2023].

Oyama, R. H. (2013). *O processo de produção de notícias, a convergência digital e a integração entre os dois meios*. Um estudo de caso do jornal Público. Dissertação de Mestrado. Universidade Nova de Lisboa.

Peeters, M. e Maesele, P. (2023) The interpretive interface: where journalistic and pluralistic discourses (don't) meet. *Digital Journalism*. <https://doi.org/10.1080/21670811.2023.2174150>.

Pires, C. *Notícias do cárcere: Luiz Estevão vira empresário de Mídia*. Piauí, n. 107, ago. de 2015. Disponível em <<https://piaui.folha.uol.com.br/materia/noticias-do-carcere/>>. [Consultado em 05/05/2023].

Poder360 (2023). *Poder360 dobra audiência e bate recorde em 2022*. Disponível em <<https://www.poder360.com.br/economia/poder360-dobra-audiencia-e-bate-recorde-em-2022/>>. [Consultado em 03/05/2023].

Reginato, G. D. (2018). As finalidades do jornalismo: percepções de veículos, jornalistas e leitores. *Famecos*, 25(3), ID29349. <http://dx.doi.org/10.15448/1980-3729.2018.3.29349>.

- Rêgo, M. G. S., Santos, R. N. e Rocha, H. C. L. (2020). A Intersubjetividade Jornalística e a Cobertura Sobre a Pandemia do Coronavírus no Brasil. *Revista Comunicando*, 9(1), pp. 80-96. <https://doi.org/10.58050/comunicando.v9i1.52>.
- Sagioglu, S. e Sinanc, D. (2013). Big data: a review. *In: 2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, USA, pp. 42-47. <https://doi.org/10.1109/CTS.2013.6567202>.
- Saheb, T., Dehghani, M. e Saheb, T. (2022). Artificial intelligence for sustainable energy: A contextual topic modeling and content analysis. *Sustainable Computing: Informatics and Systems*, 35, 100699. <https://doi.org/10.1016/j.suscom.2022.100699>.
- Santos, J. F. A. (2018). *Júlio de Mesquita Filho e o projeto de ensino superior paulista: seus escritos, sua atuação (1920-1938)*. Dissertação de Mestrado, Pontifícia Universidade Católica de São Paulo.
- Similarweb. *Participação de Mercado dos Motores de Busca*. Similarweb, 2023. Disponível em <<https://www.similarweb.com/pt/engines/>>. [Consultado em 05/05/2023].
- H. -C. Soong, H. -C, Jalil, N. B. A., Kumar Ayyasamy, R. e Akbar, R. (2019). The Essential of Sentiment Analysis and Opinion Mining in Social Media : Introduction and Survey of the Recent Approaches and Techniques. *In: 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Malaysia, 2019, pp. 272-277. <https://doi.org/10.1109/ISCAIE.2019.8743799>.
- Sousa, C. e Fernandes, V. C. (2020). Aspectos históricos da saúde pública no Brasil: revisão integrativa da literatura. *Journal of Management & Primary Health Care*, 12, e1. <https://doi.org/10.14295/jmpfhc.v12.579>.
- Sousa, J. P. (2006). Reflexões sobre um horizonte possível para o jornalismo impresso generalista de qualidade. *In: Actas das Jornadas Internacionais de Jornalismo*, Universidade Fernando Pessoa. pp. 48-60.
- Souza, I. I. (2018). Os jornais O Estado de São Paulo e Folha de São Paulo diante da política externa independente (1961-1964). *Monções: Revista de Relações Internacionais da UFGD*, 7(13), pp. 366-395. <https://doi.org/10.30612/rmufgd.v7i13.8730>.
- Sullcahuamán, E. F. D. *et alli*. (2022). Infodemia en tiempos de pandemia de la covid 19: revisión sistemática. *Ciência Latina - Revista Multidisciplinar*, 6(4), pp. 3879-3896.
- Tabai, B. J., Santos, T. B. e Coqueiro, J. M. (2022). Quando não é possível deixar de informar: o processo de trabalho de jornalistas durante a pandemia da Covid-19. *Saúde Debate*, 46, pp. 93-104.
- Taj, S., Shaikh, B. B. e Meghji, A. F. (2019). Sentiment Analysis of News Articles: A Lexicon based Approach. *In: 2nd International Conference on Computing, Mathematics*

and Engineering Technologies (iCoMET), Sukkur, Pakistan, pp. 1-5.
<https://doi.org/10.1109/ICOMET.2019.8673428>.

Thompson, E. E. (2019). Communicating a health risk/crisis: exploring the experiences of journalists covering a proximate epidemic. *Science Communication*, 41(6), pp. 707–731. <https://doi.org/10.1177/1075547019878875>.

Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B. e Bowman, D. Automated content analysis for construction safety: a natural language processing system to extract precursors and outcomes from unstructured injury reports. *Automation in Construction*, 62, pp. 45-56. <https://doi.org/10.1016/j.autcon.2015.11.001>.

Tong, Z. e Zhang, H. (2016). A text mining research based on lda topic modelling. The In: *Sixth International Conference on Computer Science, Engineering and Information Technology* (CCSEIT 2016), Vienna, Austria, pp. 201-210.
<https://doi.org/10.5121/csit.2016.60616>.

Tzenios, N., Chahine, M. e Tazanios, M. (2023). Better Strategies For Coronavirus (COVID-19) Vaccination. *Special Journal of the Medical Academy and Other Life Sciences*, 1(2). <https://doi.org/10.58676/sjmas.v1i2.11>.

Valenti, A., Mirabile, M., Cannone, E., Boccuni, F., Dionisi, P., Fortuna, G., Gagliardi, D., Vizzaccaro, R. e Iavicoli S. (2023). The impact of covid-19 pandemics on the development of health risk communication: challenges and opportunities. *International Journal of Environmental Research and Public Health*, 20(1), 645.
<https://doi.org/10.3390/ijerph20010645>

Vanderplas, J. (2023). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly, Sebastopol, USA.

Vajalla, S., Majumder, B., Gupta, A. e Surana, H. (2020). *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. O'Reilly Media.

Varão, R. e Ferreira, F. V. (2020). Jornalismo como Instância de Confiabilidade de Informações durante a Pandemia da Covid-19. In: Oliveira, H. M. G e Gadini, S. *Jornalismo em tempos de pandemia do novo coronavírus*. Aveiro, RIA Editorial, pp. 373-398.

Vieira, L. M. e Cordeiro, D. F. (2023). The dark side of anti-vaccination: analysis of a brazilian anti-vaccine Facebook group. *FAMECOS*. (no prelo).

Waisbord, S. (2018) Truth is What Happens to News. *Journalism Studies*, 19(13), pp. 1866-1878. <https://doi.org/10.1080/1461670X.2018.1492881>.

Wahyudi, M. D. R. *et alli*. (2021). Topic modeling of online media news titles

during COVID-19 emergency response in Indonesia using the latent dirichlet allocation (lda) algorithm. *Telematika*, 14(2), pp. 101-111.
<http://dx.doi.org/10.35671/telematika.v14i2.1225>.

Wu *et alli*. (2017). MLLDA: Multi-level LDA for modelling users on content curation social networks. *Neurocomputing*, 236, pp. 73-81.
<https://doi.org/10.1016/j.neucom.2016.08.114>.

Xing, W., Lee, W. e Shibani, A. Identifying patterns in students' scientific argumentation: content analysis through text mining using Latent Dirichlet Allocation. *Educational Technology Research and Development*, 68, pp. 2185-2214.
<https://doi.org/10.1007/s11423-020-09761-w>.

Yu, C. H., Jannasch-Pennell, A. e DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *Qualitative Report*, 16(3), pp. 730-744. Disponível em:
<https://eric.ed.gov/?id=EJ926322>. Acesso em: 7 de maio de 2023.

Zandomênicó, R. (2022). Inteligência artificial e jornalismo: implicações na redação de notícias e na aquisição do conhecimento. *Pauta Geral - Estudos Em Jornalismo*, 9(2), pp. 23-38. <https://doi.org/10.5212/RevistaPautaGeral.v.9.21397>.