

Sentiment Analysis for Monitoring Public Opinion on Social Media about Public Services

Giulia C. Bezerra, João Victor M. Vantil, Letícia T. M. Zoby

Instituto de Educação Superior de Brasília (IESB)
– Brasília – DF – Brasil

{giulia.bezerra, joao.vantil, leticia.zoby}@iesb.edu.br

Abstract. This paper presents a system for monitoring the public perspective on public services in Brazil, based on sentiment analysis of messages from the social network Twitter. The methodology involves four steps: data collection, pre-processing, analysis and visualization. Data collection is performed by a Python scraper, capable of bypassing the limitations of official APIs and extracting public tweets related to topics such as education, health and transportation. The data is structured in JSON and stored in the NoSQL MongoDB database. Pre-processing applies Natural Language Processing methods for text cleaning and normalization. Sentiment analysis is performed with the pre-trained lipaoMai/BERT model, which classifies texts as positive, negative or neutral. The results are displayed in an interactive dashboard developed with Angular and PrimeNG, powered by a RESTful API in Java Spring Boot. The platform highlights the potential of using Artificial Intelligence and text mining techniques in public management, by offering strategic support for decision-making based on up-to-date data that is representative of society's opinion.

Resumo. Este trabalho apresenta um sistema para monitoramento da perspectiva pública sobre serviços públicos no Brasil, a partir da análise de sentimentos de postagens da rede social Twitter. A metodologia envolve quatro etapas: coleta, pré-processamento, análise e visualização dos dados. A coleta é feita por um scraper em Python, capaz de contornar limitações das APIs oficiais e extrair tweets públicos relacionados a temas como educação, saúde e transporte. Os dados são estruturados em JSON e armazenados no banco NoSQL MongoDB. O pré-processamento aplica métodos de Processamento de Linguagem Natural para limpeza e normalização textual. A análise de sentimentos é elaborada com o modelo pré-treinado lipaoMai/BERT, que classifica os textos como positivos, negativos ou neutros. Os resultados são exibidos em um dashboard interativo desenvolvido com Angular e PrimeNG, alimentado por uma API RESTful em Java Spring Boot. A plataforma evidencia o potencial da utilização de Inteligência Artificial e técnicas de mineração de texto na gestão pública, ao oferecer apoio estratégico à formulação de decisões fundamentadas em dados atualizados e representativos da opinião da sociedade.

1. Introdução

A popularização das redes sociais provocou mudanças significativas na maneira pela qual a sociedade expressa opiniões, compartilha experiências e discute temas públicos. Dessa forma, a ampla disseminação de conteúdo digital, especialmente em plataformas como o Twitter (TWITTER, 2025), representa uma fonte de opinião pública em tempo real (BORGOGNO; CASALINI; IACOMINI, 2024). Com isso, a incorporação de ferramentas analíticas aos serviços públicos permite extrair *insights* relevantes sobre a percepção da população em relação a áreas essenciais como saúde, educação, transporte e segurança. Indubitavelmente, compreender esses sentimentos é primordial para que gestores públicos e formuladores de políticas possam

identificar demandas emergentes, avaliar a efetividade das ações governamentais e conduzir iniciativas mais alinhadas às prioridades sociais (ALVAREZ-MELIS; GITTENS; JAAKKOLA, 2023).

Considerando esse contexto, esta pesquisa propõe a criação de um sistema voltado à coleta, processamento e interpretação de sentimentos extraídos de postagens de redes sociais relacionadas a serviços públicos no Brasil. Todas as etapas de obtenção e tratamento dos dados foram conduzidas em conformidade com a Lei Geral de Proteção de Dados (LGPD), observando seus fundamentos, princípios e bases legais de tratamento (arts. 2º, 6º e 7º) (CARDOSO, 2025), a fim de preservar a privacidade e a integridade das informações coletadas.

A coleta foi realizada por meio de um *scraper* personalizado, que extraiu mais de dois mil *tweets* públicos publicados entre 2020 e 2025 contendo termos associados a distintas categorias de serviços públicos, como “educação”, “saúde” e “transporte”. As postagens capturadas passaram por uma etapa criteriosa de preparação, que incluiu a remoção de *hashtags*, menções, *links* e ruídos (textos muito curtos ou pouco relevantes).

Para a classificação dos sentimentos, utilizou-se o modelo *lipaoMai/BERT-sentiment-analysis-portuguese*, hospedado na plataforma Hugging Face (HUGGING FACE, 2023), capaz de rotular textos como positivos, negativos ou neutros com alta precisão (SOUZA; LIMA; COUTINHO, 2023). Os resultados foram estruturados em JSON — contendo metadados como ID do *tweet*, data, número de curtidas e *retweets*, categoria temática e sentimento identificado — e armazenados em um banco MongoDB (MONGODB, 2025), garantindo consultas rápidas e flexíveis para análises futuras (KREPS et al., 2023).

No *backend*, foi desenvolvida uma API RESTful utilizando Java com o *framework* Spring Boot (VMWARE, 2025), responsável por disponibilizar os dados processados ao *frontend*. Uma API (*Application Programming Interface*), ou Interface de Programação de Aplicações, é um conjunto de definições e protocolos que permite a comunicação entre diferentes *softwares*. Nesse contexto, a API RESTful atua como uma ponte entre o *frontend* (a interface visual) e o *backend* (o sistema onde os dados são armazenados e processados), recebendo requisições e retornando as informações solicitadas.

A interface visual foi construída com Angular (ANGULAR, 2025) e PrimeNG v19 (PRIMETEK INFORMATICS, 2025), proporcionando uma experiência moderna, intuitiva e responsiva. O *dashboard* exibe gráficos interativos e listagens categorizadas, facilitando a análise dos padrões de sentimento por período e por categoria de serviço público.

Dessa forma, este trabalho não só apresenta uma investigação acadêmica abrangente sobre o uso de Processamento de Linguagem Natural (PLN), Aprendizado de Máquina (AM) e Mineração de Texto, como também entrega uma ferramenta prática, funcional e extensível para acompanhar a opinião pública em ambientes digitais, contribuindo para a modernização da gestão pública por meio de uma escuta social eficaz e fundamentada em dados concretos e atualizados (ROKOSZ; BIEŃ, 2025).

2. Referencial Teórico

2.1 Mineração de Texto

A mineração de texto consiste na extração de informações relevantes a partir de grandes volumes de dados textuais não estruturados, convertendo textos escritos em linguagem comum em dados passíveis de análise (FELDMAN; SANGER, 2007). Para que isso seja possível, são realizadas etapas como a exclusão de palavras irrelevantes (*stopwords*), tokenização e a transformação do conteúdo textual em vetores. Nesse contexto, de acordo com Hotho, Nürnberg e Paaß (2005),

seus resultados são potencializados quando combinados com técnicas de AM e PLN. Segundo Aggarwal (2022), o avanço das bibliotecas *open source* e a integração com *big data* tornaram as técnicas de *text mining* ainda mais acessíveis e eficazes, especialmente para análises em tempo real.

2.2 Processamento de Linguagem Natural (PLN)

O PLN é o campo da IA que busca tornar possível a comunicação direta entre pessoas e computadores por meio da linguagem natural, e que envolve tarefas como análise sintática, semântica e pragmática, permitindo que sistemas compreendam textos de maneira mais semelhante ao entendimento humano (CAMBRIA; WHITE, 2014).

Em contraste com modelos mais antigos como o *bag-of-words*, que tratam cada palavra de forma isolada, sem considerar a sua ordem ou contexto, os *transformers* conseguem capturar a importância e o significado de cada palavra dentro da frase inteira. Essa capacidade de compreender o contexto permite que modelos como o BERT (*Bidirectional Encoder Representations from Transformers*) e o RoBERTa (*Robustly Optimized BERT Approach*) lidem melhor com ambiguidades e nuances da linguagem, resultando em um desempenho superior em diversas tarefas, como tradução de textos, sumarização de documentos e análise de sentimentos (JURAFSKY; MARTIN, 2023).

2.3 Aprendizado de Máquina (AM)

O AM refere-se ao desenvolvimento de algoritmos capazes de aprender padrões extraídos dos dados e tomar decisões sem regras explícitas (MITCHELL, 1997). Nesse sentido, modelos supervisionados como SVM (*Support Vector Machine* - Máquina de Vetores de Suporte), *Random Forest* (Floresta Aleatória) e RNC (Redes Neurais Convolucionais) têm mostrado excelente desempenho na tarefa de classificar textos (GÉRON, 2019). No entanto, segundo Bishop (2006), esses algoritmos requerem dados rotulados para aprendizado.

2.4 Análise de Sentimentos

A Análise de Sentimentos visa identificar emoções em textos, classificando-os como positivos, negativos ou neutros. Ao longo do tempo, segundo Zhang, Gui e Cambria (2022), ela evoluiu de uma simples classificação de polaridade para abordagens mais refinadas que incluem interpretação de emoções, sarcasmo e aspectos específicos de opinião (*aspect-based sentiment analysis*), sendo amplamente aplicada em cenários como marketing digital, política e suporte ao cliente. Em concordância, Liu (2020) destaca o papel dos modelos pré-treinados na identificação de sentimentos sutis, especialmente em textos informais como os publicados nas redes sociais.

2.5 Web Scraping

Web scraping consiste em um método utilizado para extração automatizada de informações contidas em páginas *web*, permitindo coletar dados estruturados ou semiestruturados que estão disponíveis de forma pública na internet. Nesse contexto, essa prática é amplamente aplicada em diversas áreas, como mineração de dados, análise de sentimentos, monitoramento de preços e tendências de mercado. Para isso, a extração ocorre por meio de *scripts* ou ferramentas que simulam a navegação humana, identifica padrões em elementos HTML (*HyperText Markup Language* - Linguagem de Marcação de Hipertexto) (como *tags*, classes e IDs) e extraem o conteúdo relevante, que posteriormente pode ser armazenado e analisado (MORIN; BOURDEAU, 2020). Quando se trata de páginas dinâmicas, técnicas mais avançadas são empregadas, como a utilização de bibliotecas que interagem com JavaScript (ECMA INTERNATIONAL, 2025), como Selenium (SELENIUM, 2025) ou Puppeteer (PUPPETEER,

2025). Apesar de sua eficiência, o *web scraping* exige atenção a questões éticas e legais, como o respeito aos termos de uso dos sites e à privacidade dos dados (KAPLAN, 2021).

2.6 Trabalhos Correlatos

O estudo de Silva et. al (2021) propõe uma análise de sentimentos sobre o Sistema Único de Saúde (SUS) a partir da coleta de 27,5 mil *tweets* em português, publicados entre dezembro de 2019 e outubro de 2020, com a finalidade de compreender a opinião pública antes e durante a pandemia de COVID-19. Para isso, os *tweets* foram classificados em sentimentos positivos, negativos e neutros, além de categorias emocionais como medo, raiva e alegria. Dentre os resultados, a maioria (36,4%) expressava sentimentos positivos, especialmente em momentos de maior pressão sobre o sistema, refletindo apoio ao SUS. Por outro lado, os sentimentos negativos estavam associados ao medo da sobrecarga hospitalar e não a uma rejeição direta ao serviço. Assim, o estudo evidencia a viabilidade de usar redes sociais para captar percepções públicas sobre políticas de saúde e contribui com metodologias replicáveis para o monitoramento de serviços públicos.

O trabalho de Correa (2024) apresenta uma investigação voltada à análise dos sentimentos manifestados por usuários do Twitter em relação ao Projeto de Lei nº 2630/2020, o PL das Fake News. Para isso, a pesquisa comprehende desde a obtenção de informações por meio de técnicas de *scraping* até o pré-processamento textual, rotulagem dos exemplos analisados e classificação supervisionada utilizando os algoritmos Naive Bayes, SVM e Decision Tree. Dentre esses métodos, o SVM com kernel linear apresentou o melhor desempenho, com acurácia de 86%. Complementarmente, a pesquisa utilizou nuvens de palavras, grafos de conexões entre usuários e gráficos temporais para explorar a dinâmica das opiniões no ambiente da plataforma. Logo, o estudo mostra como essas técnicas podem revelar percepções públicas sobre temas políticos e controversos nas redes sociais.

Já a pesquisa conduzida por Lopes (2023) analisa os sentimentos de usuários do Twitter sobre a temporada de 2022 da Fórmula 1, recorrendo a abordagens de extração textual e análise de sentimentos. Para isso, os dados coletados via API com palavras-chave relacionadas à F1, passaram pelo pré-processamento e foram classificados como positivos, negativos ou neutros com a biblioteca TextBlob (TEXTBLOB, 2013). Dessa forma, os resultados indicam que o uso de técnicas de análise de sentimentos em redes sociais é eficaz para compreender a opinião pública em tempo real, destacando o Twitter como uma fonte valiosa e o uso de PLN como uma ferramenta relevante em estudos esportivos.

Já Alves (2024) analisou sentimentos em mensagens sobre aplicativos de *delivery*, motivado pelo crescimento desse setor durante a pandemia. Para isso, o trabalho incluiu a extração de informações por meio de *web scraping* no Twitter, seguido de rotulagem dos dados, pré-processamento com técnicas de PLN e classificação utilizando diversos algoritmos, com o SVM apresentando o melhor desempenho. Por fim, o estudo destaca a análise de sentimentos como uma ferramenta estratégica para o setor, mas ressalta a necessidade de técnicas mais sofisticadas para lidar com a complexidade dos textos nas mídias sociais.

3. OpinaMAIS

O sistema desenvolvido neste trabalho foi denominado OpinaMAIS — Monitoramento da Atividade e Interação Social — em referência à sua proposta central de acompanhar, analisar e interpretar as manifestações da população sobre serviços públicos a partir de dados coletados em redes sociais.

A partir disso, a metodologia adotada neste trabalho foi organizada em quatro etapas principais, ilustradas nas Figura 1 (A – coleta e B – pré-processamento) e Figura 2 (C – análise e D – visualização). Dessa forma, o objetivo foi desenvolver um fluxo automatizado e contínuo que abrange desde a captação de conteúdos em redes sociais até a exibição de informações relevantes acerca da percepção pública em uma aplicação web interativa.

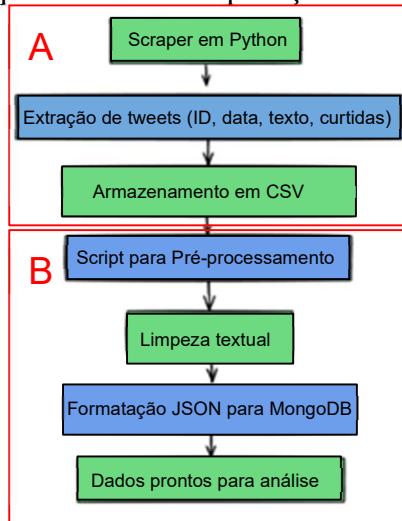


Figura 1 – Etapas de Coleta e Pré-Processamento de Dados

3.1 Coleta de Dados e Ferramentas Tecnológicas

A primeira etapa, representada na Figura 1A, envolveu a construção de um *scraper* personalizado em Python (PYTHON SOFTWARE FOUNDATION, 2023), utilizando as bibliotecas requests (REQUISITOS, 2025), BeautifulSoup (RICHARDSON, 2025) e Selenium (SELENIUM, 2025). Essa abordagem foi adotada devido às limitações impostas pelas APIs oficiais do Twitter, as quais restringem tanto o volume quanto a profundidade dos dados acessíveis para extração.

Dessa forma, o *scraper* foi programado para acessar postagens públicas vinculadas a temas como “educação”, utilizando consultas específicas exemplificadas na Tabela 1. A partir dessas postagens, são extraídos os seguintes campos: identificador do *tweet* (ID), data e horário da publicação, conteúdo textual, número de curtidas, *retweets* e visualizações (*analytics*). Em seguida, as informações coletadas são armazenadas em arquivos no formato CSV, a fim de viabilizar etapas posteriores de processamento e análise.

Tabela 1 - Parâmetros utilizados nas buscas para coleta de dados

Serviços Públicos	Parâmetros de Busca (<i>keywords e query</i>)
Educação	- ‘brasília AND educação AND pública’ - ‘educação’, ‘brasília’, ‘pública’
Infraestrutura	- ‘brasília AND infraestrutura AND pública’ - ‘infraestrutura’, ‘brasília’
Saneamento Básico	- ‘brasília AND saneamento AND público’ - ‘brasília AND limpeza AND urbana’ - ‘brasília’, ‘saneamento’, ‘limpeza’, ‘urbana’
Saúde	- ‘brasília AND saúde AND pública’ - ‘saúde’, ‘brasília’, ‘pública’

Segurança	<ul style="list-style-type: none"> - ‘brasília AND policiamento’ - ‘brasília AND segurança AND pública’ - ‘segurança’, ‘policamento’, ‘brasília’
Transporte	<ul style="list-style-type: none"> - ‘brasília AND transporte AND público’ - ‘transporte’, ‘público’, ‘brasília’

3.2 Pré-Processamento e Preparação dos Dados

No presente trabalho, o PLN foi aplicado no pré-processamento dos dados textuais, representado na Figura 1B, por meio da remoção de ruídos e normalização das informações. Essas etapas são essenciais, pois adequam os textos à entrada de algoritmos de classificação baseados em modelos de aprendizado automático, assegurando maior precisão na identificação e categorização de sentimentos.

Após a coleta, os dados são processados por um *script* em Python que executa diversas operações voltadas à melhoria da qualidade e à consistência das informações. As principais etapas de pré-processamento incluem:

- Limpeza textual: remoção de *links*, *hashtags*, menções, quebras de linha, excesso de espaços, além da exclusão de mensagens muito curtas (com menos de quatro palavras úteis). Também são tratados valores numéricos expressos em milhares (K) ou milhões (M), convertendo-os em números reais.
- Formatação para JSON: padronização de campos numéricos em tipos específicos (*\$numberInt*, *\$numberDouble*, *\$numberLong*), além da conversão de data/hora para o formato timestamp compatível com o MongoDB.

Portanto, essa etapa assegura que os dados estejam devidamente organizados para uso no modelo de classificação de sentimentos baseado em BERT e possam ser integrados a outras aplicações no fluxo da solução.

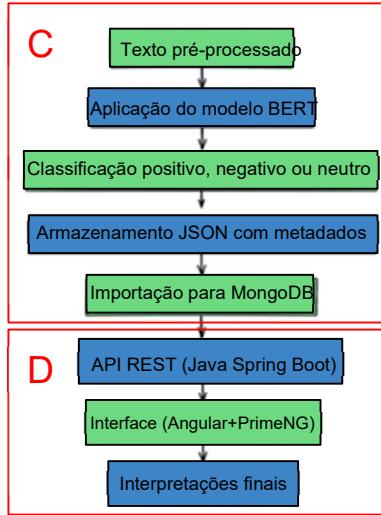


Figura 2 - Etapas de Análise e Visualização

3.3 Análise de Sentimentos

Neste estudo, a técnica de análise de sentimentos (Figura 2C) foi empregada para interpretar as percepções expressas pelos usuários sobre os serviços públicos em Brasília. Para isso, a análise foi realizada com base em uma abordagem supervisionada, por meio do uso do modelo *lipaoMai/BERT*, disponível na plataforma Hugging Face. Esse modelo, por sua vez,

atua diretamente sobre o texto já pré-processado e se baseia na arquitetura BERT, capaz de compreender o contexto de maneira bidirecional e aprofundada.

O *lipaoMai/BERT* é um modelo ajustado especificamente para o português brasileiro, com foco na tarefa de análise de sentimentos. Ele foi treinado a partir da arquitetura DistilBERT, uma versão compacta do BERT com seis camadas de transformadores, e posteriormente refinado para o português, tornando-o capaz de classificar frases e outros textos em positivo, negativo ou neutro.

A arquitetura subjacente segue o modelo BERT-base, composto por 12 camadas de transformadores com mecanismos de atenção que permitem ao modelo considerar o contexto à esquerda e à direita de cada palavra simultaneamente — o que é essencial para lidar com ambiguidades linguísticas e expressões idiomáticas. Esse funcionamento é resultado do treinamento original do BERT, baseado em duas tarefas principais: *Masked Language Modeling* (MLM), em que palavras são ocultadas aleatoriamente e o modelo aprende a prevê-las com base no contexto, e *Next Sentence Prediction* (NSP), que ensina o modelo a compreender relações entre frases consecutivas (DEVLIN, 2018). Esse processo resulta em uma representação contextualizada das palavras, tornando o modelo altamente eficaz em tarefas de compreensão de linguagem natural, como a análise de sentimentos.

No que tange à execução, o componente responsável por analisar sentimentos aplica o modelo a cada texto (respeitando o limite de 512 caracteres imposto pela arquitetura BERT) e retorna a classificação entre positivo, negativo ou neutro. Além disso, foi incorporada uma lógica adicional para rotular como neutras as mensagens com baixa confiança ($score < 0.70$) ou com conteúdo considerado insuficiente (menos de quatro palavras relevantes).

Posteriormente, cada entrada analisada recebe a classificação do sentimento e é armazenada no formato JSON, juntamente com seus metadados originais. Desse modo, os modelos de AM empregados neste projeto permitem detectar a polaridade dos textos processados, viabilizando a categorização automática das opiniões extraídas das plataformas sociais *online*.

3.4 Visualização dos Resultados

A etapa de visualização dos dados (Figura 2D) é concretizada por meio de um *dashboard web* interativo (Figura 3), desenvolvido com Angular e PrimeNG, consumindo uma API REST construída com Java Spring Boot. Nessa fase, o painel exibe os dados já enriquecidos com a classificação de sentimentos, permitindo sua análise de forma intuitiva e dinâmica. Além disso, a interface possibilita diferentes formas de exploração visual, como:

- Filtros interativos por ID, data, conteúdo textual, categoria, polaridade, entre outros (Figura 3A).
- Visualização temporal por período e categoria (Figura 4);

Dessa maneira, esse processo assegura que os *insights* obtidos sejam de fácil interpretação e aplicabilidade em estudos de opinião pública e gestão de serviços.



Figura 3 - Dashboard interativo do OpinaMAIS

Com o objetivo de proporcionar uma melhor compreensão da distribuição temática dos dados, a Figura 4A apresenta um gráfico de pizza com a proporção de *tweets* classificados por categoria de serviço público. Nessa visualização, observa-se um volume significativamente maior nas categorias educação e transporte, que juntas ultrapassam mais de duas mil postagens, evidenciando um maior interesse ou preocupação pública com esses setores.

Além disso, a exploração por categoria específica possibilita investigar como os sentimentos se distribuem internamente. A Figura 4B ilustra a polarização emocional (positiva, neutra ou negativa) para a categoria educação, utilizando um gráfico de barras dinâmico. Importante destacar que esse tipo de visualização está disponível para todas as categorias, por meio de um menu suspenso (*dropdown*)(Figura 4D).

Por fim, para acompanhar as variações temporais na quantidade de postagens, a Figura 4C apresenta um gráfico de linhas com a evolução mensal de *tweets* ao longo do ano de 2024. De forma complementar, essa visualização também pode ser filtrada por ano, facilitando a análise de tendências sazonais e eventos específicos.



Figura 4 - Gráficos interativos do OpinaMAIS

4. Resultados

A aplicação da metodologia proposta viabilizou, assim, o desenvolvimento de uma solução que engloba coleta, pré-processamento, análise de sentimentos e exibição visual de dados extraídos do Twitter/X. Nesse sentido, foram analisadas mais de duas mil postagens da região de Brasília

feitas entre 2020 e 2025, relacionadas a serviços públicos, com ênfase inicial na categoria “educação”. Em seguida, após o processamento, os dados foram enriquecidos com informações adicionais, como número de curtidas, *retweets* e data de publicação, sendo armazenados em um banco de dados MongoDB no formato JSON.

A etapa de análise de sentimentos empregou o modelo *lipaoMai/BERT*, uma versão *DistilBERT* já feito o *fine-tuning* para português e disponibilizada publicamente no Hugging Face. Esse ajuste foi realizado com *learning rate* de 2×10^{-5} , *batch size* de 16, *seed* 42, otimizador Adam e *scheduler* linear com *warm-up* de 3 %, durante duas épocas. Os resultados obtidos no conjunto de validação estão resumidos na Tabela 2.

Tabela 2 - Métricas do modelo lipaoMai/BERT

Métricas	Valor
Acurácia	0.9229
Precisão	0.9215
Revocação	0.9229
F1-score	0.9219
<i>Loss</i>	0.2375

Esses indicadores apontam um desempenho equilibrado entre as classes, reforçando a capacidade do modelo de identificar sentimentos positivos, negativos e neutros com pouca discrepância. Na aplicação deste estudo, cada texto (limitado a 512 caracteres, conforme a arquitetura BERT) é classificado como positivo, negativo ou neutro. Mensagens com confiança inferior a 0,70 ou com menos de quatro palavras relevantes são rotuladas como neutras para evitar decisões pouco confiáveis.

Por fim, os resultados são apresentados em uma interface dinâmica (Figura 3 e Figura 4), que permite a exploração dos dados por categoria, data, palavra-chave e polaridade. Além disso, os números consolidados de *tweets* classificados por polaridade para cada serviço público podem ser consultados na Tabela 3, oferecendo uma visão quantitativa complementar que apoia a análise estratégica da opinião pública ao longo do período estudado.

Tabela 3 – Quantidade de tweets classificados por polaridade

Serviço Público	Positivos	Negativos	Neutros
Educação	558	330	112
Infraestrutura	54	40	15
Saneamento Básico	45	35	8
Saúde	136	19	85
Segurança	143	185	65
Transporte	220	597	67
TOTAL	1156	1206	352

5. Conclusão

Este trabalho apresentou, por tanto, um modelo para acompanhar a percepção da sociedade sobre serviços públicos no ambiente digital, a partir da aplicação de técnicas de análise de sentimentos. A metodologia, que compreende desde a coleta automatizada de dados até sua exibição em um painel interativo, demonstrou viabilidade operacional, robustez técnica e potencial para gerar *insights* relevantes.

A utilização do modelo BERT voltado ao idioma português, por sua vez, confirmou a eficácia de abordagens supervisionadas no processamento de linguagem natural, especialmente quando precedidas por uma etapa rigorosa de preparação dos dados. Além disso, a integração entre as camadas de *backend* (Java Spring Boot), armazenamento (MongoDB) e *frontend* (Angular + PrimeNG) resultou em uma solução coesa, acessível e adaptável.

Como desdobramentos futuros, estão previstas as seguintes iniciativas:

- Ampliação do conjunto de dados para incluir outras plataformas digitais e fontes públicas, expandindo as possibilidades de análise;
- Inserção de novas localidades geográficas, com vistas a realizar análises comparativas entre municípios ou estados quanto à avaliação da população sobre os serviços ofertados;
- Integração com dados institucionais ou levantamentos oficiais de opinião, permitindo validação cruzada e aprofundamento analítico a partir de múltiplas perspectivas;
- Divulgação de um repositório público contendo o código-fonte e um conjunto de dados anonimizados, com foco em transparência e reproduzibilidade científica.

Dessa forma, a solução proposta avança no sentido de transformar a escuta social digital em uma ferramenta concreta de apoio à gestão pública e à formulação de políticas orientadas por dados.

Referências

- AGGARWAL, C. C. Machine learning for text. Cham: Springer, 2022. DOI: <https://doi.org/10.1007/978-3-031-01677-2>.
- ALVAREZ-MELIS, D.; GITTENS, A.; JAAKKOLA, T. Public opinion mining using social media in government decision-making. arXiv, 2023. Disponível em: <https://arxiv.org/abs/2302.17143>. Acesso em: 20 maio 2025.
- ALVES, J. P. R.; MENEZES, M. O. de. A análise de sentimentos e emoções de mensagens de redes sociais envolvendo serviços de entrega de alimentos. 2024. Disponível em: <https://adelpha-api.mackenzie.br/server/api/core/bitstreams/5967d6c5-3bdd-4e9f-963d-851d41c6f89f/content>. Acesso em: 19 maio 2025.
- ANGULAR. Angular: documentação oficial. 2025. Disponível em: <https://angular.io/docs>. Acesso em: 20 maio 2025.
- BISHOP, C. M. Pattern recognition and machine learning. New York: Springer, 2006.
- BORGOGNO, O.; CASALINI, A.; IACOMINI, E. Digital platforms and social listening: Using Twitter data for public interest analysis. Journal of Digital Government Research, v. 11, n. 1, p. 23–40, 2024.

- CAMBRIA, E.; WHITE, B. Jumping NLP curves: a review of natural language processing research. *IEEE Computational Intelligence Magazine*, v. 9, n. 2, p. 48–57, 2014.
- CARDOSO, Oscar. O Web Scraping Viola a Proteção de Dados Pessoais? | Jusbrasil. Disponível em: <<https://www.jusbrasil.com.br/artigos/o-web-scraping-viola-a-protacao-de-dados-pessoais/1152362639>>. Acesso em: 3 jun. 2025.
- CORREA, C. G. F. de O. Mineração de textos: análise de sentimentos de tweets sobre o Projeto de Lei nº 2630/2020. 2024. 61 f. Trabalho de Conclusão de Curso (Graduação em Gestão da Informação) – Universidade Federal de Uberlândia, Uberlândia, 2024. Disponível em: <https://repositorio.ufu.br/handle/123456789/43564>. Acesso em: 19 maio 2025.
- DEVLIN, Jacob et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*, 2018. Disponível em: <https://arxiv.org/abs/1810.04805>. Acesso em: 5 ago. 2025.
- ECMA INTERNATIONAL. ECMAScript Language Specification (JavaScript). 2023. Disponível em: <https://262.ecma-international.org/13.0/>. Acesso em: 20 maio 2025.
- FELDMAN, R.; SANGER, J. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge: Cambridge University Press, 2007.
- GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. 2. ed. Sebastopol: O'Reilly Media, 2019.
- HOTHO, A.; NÜRNBERGER, A.; PAAß, G. A brief survey of text mining. *LDV Forum*, v. 20, n. 1, p. 19–62, 2005.
- HUGGING FACE. BERT sentiment analysis portuguese. Hugging Face, 2023. Disponível em: <https://huggingface.co/lipaoMai/BERT-sentiment-analysis-portuguese>. Acesso em: 20 maio 2025.
- JURAFSKY, D.; MARTIN, J. H. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. 3. ed. [Online draft], 2023. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 19 maio 2025.
- KAPLAN, J. *Web Scraping with Python: Collecting Data from the Modern Web*. 2. ed. O'Reilly Media, 2021.
- KREPS, J. et al. MongoDB for Big Data and Real-Time Analytics. *Data Engineering Journal*, v. 9, n. 3, p. 45–62, 2023.
- LIU, B. *Sentiment analysis: mining opinions, sentiments, and emotions*. 2. ed. Cambridge: Cambridge University Press, 2020.
- LOPES, M. F. M. Análise de sentimentos em redes sociais durante a temporada 2022 da Fórmula 1: um estudo utilizando processamento de linguagem natural no Twitter. *Repositório Institucional da UTFPR*, 2023. Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/33486>. Acesso em: 19 maio 2025.
- MITCHELL, T. M. *Machine learning*. New York: McGraw-Hill, 1997.
- MONGODB INC. *MongoDB Documentation*. 2025. Disponível em: <https://www.mongodb.com/docs>. Acesso em: 20 maio 2025.
- MORIN, D.; BOURDEAU, D. *Data Science and Web Scraping Using Python*. Springer, 2020.
- PRIMETEK INFORMATICS. PrimeNG v19 – UI components for Angular. 2025. Disponível em: <https://primeng.org>. Acesso em: 20 maio 2025.

PUPPETEER. Puppeteer Documentation. 2025. Disponível em: <https://pptr.dev/>. Acesso em: 20 maio 2025.

PYTHON SOFTWARE FOUNDATION. Python Language Reference, versão 3.11. 2023. Disponível em: <https://docs.python.org/3/>. Acesso em: 20 maio 2025.

REQUISITOS. Requests: HTTP for Humans. 2025. Disponível em: <https://docs.python-requests.org/en/latest/>. Acesso em: 20 maio 2025.

RICHARDSON, L. BeautifulSoup Documentation. 2025. Disponível em: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. Acesso em: 20 maio 2025.

ROKOSZ, K.; BIEŃ, A. Sentiment analysis of digital public discourse as support for social policy. *Information Systems and e-Business Management*, v. 23, n. 1, p. 55–73, 2025.

SELENIUM. Selenium Documentation. 2025. Disponível em: <https://www.selenium.dev/documentation/>. Acesso em: 20 maio 2025.

SILVA, H. et al. Sentiment Analysis of Tweets Related to SUS Before and During COVID-19 pandemic. *IEEE Latin America Transactions*, v. 20, n. 1, p. 6–13, 2021. Disponível em: <https://latamt.ieeer9.org/index.php/transactions/article/view/5434>. Acesso em: 19 maio 2025.

SOUZA, F.; LIMA, F.; COUTINHO, T. Avaliação de modelos BERT para análise de sentimentos em português. *Revista Brasileira de Computação Aplicada*, v. 15, n. 2, p. 44–56, 2023.

TEXTBLOB. TextBlob: simplified text processing [software]. [S. l.], 2013. Disponível em: <https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis>. Acesso em: 4 ago. 2025.

TWITTER. Plataforma de microblog usada como fonte de dados para análise de sentimentos. Twitter, 2025. Disponível em: <https://twitter.com>. Acesso em: 20 maio 2025.

VMWARE. Spring Boot Documentation. 2025. Disponível em: <https://docs.spring.io/spring-boot/docs/current/reference/html/>. Acesso em: 20 maio 2025.

ZHANG, Y.; GUI, L.; CAMBRIA, E. Sentiment analysis: a survey. *Information Fusion*, v. 91, p. 202–240, 2022. DOI: <https://doi.org/10.1016/j.inffus.2022.01.004>.