

MINERAÇÃO DE TEXTOS: ANÁLISE DE SENTIMENTO UTILIZANDO TWEETS REFERENTES À VACINAÇÃO DA COVID-19 EM CRIANÇAS

Lara Selena Gonçalves Scaranello, Carlos Magnus Carlson Filho*

e-mail:

laraselena0911@gmail.com; carlos.carlson@fatec.sp.gov.br

Resumo: O artigo aborda a importância da análise de sentimentos no Twitter, agora conhecido como X, relacionada à vacinação infantil da Covid-19. Com o avanço da internet e o aumento do volume de informações, a mineração de dados e a análise de sentimentos se tornaram essenciais para lidar com esse contexto. O estudo tem como objetivo identificar e avaliar os sentimentos dos usuários do Twitter por meio de seus *tweets*, buscando compreender as opiniões em relação à vacinação em crianças. São explorados conceitos como o uso do Twitter como fonte de informação, mineração de textos, análise de sentimentos e o contexto da pandemia da Covid-19.

Palavras-chave: Análise de sentimentos. Vacinação infantil. Twitter como fonte de informação. Pandemia da Covid-19. Mineração de dados em redes sociais.

Abstract: The article discusses the importance of sentiment analysis on Twitter, now referred to as X, in relation to childhood Covid-19 vaccination. With the advancement of the internet and the increase in the volume of information, data mining and sentiment analysis have become essential to deal with this context. The study aims to identify and evaluate the sentiments of Twitter users through their tweets, seeking to comprehend opinions regarding child vaccination. Concepts such as the use of Twitter as a source of information, text mining, sentiment analysis, and the context of the Covid-19 pandemic are explored.

Keywords: Sentiment analysis. Childhood vaccination. Twitter as a source of information. Covid-19 pandemic. Data mining in social networks.

1 Introdução

A evolução da internet desencadeou uma profunda transformação na forma como as pessoas se comunicam, tornando-se uma necessidade fundamental em suas vidas e redefinindo as interações sociais (NAKAGAKI; SARPONG, 2021). Segundo dados da Pesquisa Nacional por Amostra de Domicílios, o acesso à internet abrange cerca de 90% das residências no Brasil (BRASIL, CASA CIVIL, 2022), destacando o amplo alcance dessa conectividade na realidade atual.

É produzido diariamente um grande volume de informações por profissionais e usuários de redes sociais, ocasionando dificuldades para a recuperação e na diferenciação das informações no ambiente da internet. Conseguir conhecimento útil nesse ambiente é uma tarefa cada vez mais desafiadora, pois é impossível analisar manualmente esse enorme volume de dados, mas o desenvolvimento de novas tecnologias tem possibilitado a análise e visualização

de informações a partir de uma grande quantidade de dados. (MAGALHÃES *et al.*, 2022).

Com o avanço das Tecnologias de Comunicação e Informação (TIC) e da Inteligência Artificial (IA) é possível obter meios e mecanismos para a obtenção e análise dos dados digitais. De acordo com Liu (2010), a categorização de textos é dada como fatos ou opiniões, enquanto isso, na mineração de dados textuais ou Análise de Sentimentos essa categorização dos sentimentos em texto é realizada por meio de plataformas digitais, como as redes sociais. A Análise de Sentimentos é uma área do Processamento de Linguagem Natural (PLN), que contém sistemas que diferenciam e extraem opiniões dentro de textos.

O aumento do acesso à internet e o desenvolvimento de plataformas de publicações aprimoraram os meios em que usuários produzem conteúdos informativos, dessa maneira é muito importante a criação de pesquisas e ferramentas que analisem os conteúdos que são produzidos nessas plataformas. Algoritmos de *Machine Learning* (ML) aperfeiçoam esses processos de recuperação de informações e permitem que haja avanços na automatização nesses processos. “O Aprendizado de Máquina é a área de maior destaque em análise de *Big Data*, pois permite a criação de modelos computacionais para processamento automático ou semiautomático.” (MAGALHÃES *et al.*, 2022).

Segundo Magalhães *et al.*, (2022) a Análise de Sentimentos é bastante significativa, pois possui uma enorme diversidade de informações que ela pode nos mostrar. Conhecer as opiniões alheias de um assunto pode influenciar de forma positiva a nossa tomada de decisões e permitir a percepção de tendências ou comportamentos sociais de forma geral.

Tendo em vista que usuários preferem receber e compartilhar informações de leitura e entendimento fácil, a rede social Twitter¹ é um enorme canal de rápida divulgação de opiniões e informações, além de permitir que seus usuários tenham contato direto entre empresas e influenciadores digitais. Recuero e Zago (2011) referem-se ao Twitter como sendo uma ferramenta de divulgação de informações através da publicação de mensagens ou *tweets* de até 280 caracteres para os seguidores, e, também, permite a comunicação entre usuários por meio de *replies* (forma pública) ou por *messages* (forma privada).

O Twitter também se tornou uma ferramenta para jornalistas e empresas, onde os usuários podem expor suas opiniões sobre diversos produtos e fatos. Segundo Murthy (2018), isso fez com que empresas se motivassem a estudar os sentimentos dos usuários consumidores no Twitter por meio das *hashtags*, que possuem o símbolo de cerquilha (#) e é acompanhado de tópicos ou palavras-chave de um determinado assunto, seu uso faz com que as publicações do mesmo assunto sejam encontradas com mais facilidade.

Em 2019, a Organização Mundial da Saúde (OMS) recebeu um alerta sobre vários casos de pneumonia na China, sendo causada por uma nova cepa de coronavírus, que até então nunca tinha sido detectada em humanos (ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE – OPAS, 2021). O vírus SARS-CoV-2, da família dos coronavírus é o causador da doença Covid-19, que, no início da pandemia ficou conhecido como o “novo coronavírus” por ser uma variante desconhecida, ele desencadeia uma doença respiratória muito agressiva com uma alta

¹Durante o período desta pesquisa, a plataforma de mídia social conhecida como Twitter passou por uma mudança de marca e agora é conhecida como X. No entanto, para fins de consistência e clareza, todas as referências à plataforma neste documento continuarão a usar o nome original, Twitter.

taxa de transmissão, que logo se espalhou para todo o planeta, não dando tempo de que medidas de contenção efetivas fossem descobertas antes da propagação da doença para os continentes (BUTANTAN, 2021). Em 30 de janeiro de 2020, a OMS declarou o surto do novo coronavírus como uma Emergência de Saúde Pública de Importância Internacional (ESPII) (OPAS, 2020) e a pandemia da Covid-19 acabou se tornando um dos assuntos mais recorrentes dos meios de comunicação e das redes sociais, como o Twitter.

De acordo com o Painel Coronavírus Brasil, até o dia 12 de maio de 2023 haviam sido registrados 702.116 óbitos pela Covid-19 (CORONAVÍRUS BRASIL, 2023). Enquanto isso uma pesquisa realizada pela Fiocruz, indicia que entre 01 de janeiro e 11 de outubro de 2022 foram totalizados 314 óbitos de crianças na faixa etária de 6 meses a 5 anos (FIOCRUZ, 2022).

Com o surgimento das vacinas contra à Covid-19, os profissionais da saúde apontam como é importante que a população se vacine, porém as desinformações sobre o tema continuaram a crescer e a se espalhar pelas redes. Segundo dados do Ministério da Saúde mais de 69 milhões de brasileiros não tomaram a primeira dose de reforço contra a Covid-19 (BRASIL, MINISTÉRIO DA SAÚDE, 2022).

A fim de conter o compartilhamento de informações falsas e enganosas, o Twitter divulgou, em 16 de dezembro de 2020, o método que seria utilizado em relação aos compartilhamentos de informações sobre a vacina da Covid-19. Para o próprio Twitter o mundo estava iniciando um processo de vacinação e os seus usuários o procuravam para comunicação e obtenção de informações.

Essa pesquisa visa utilizar técnicas de mineração de dados para extrair informações em formato de texto do Twitter, com foco nas opiniões dos usuários sobre a vacinação da Covid-19 em crianças. A análise dos dados coletados pode proporcionar uma melhor compreensão da percepção pública sobre este tópico importante. Segundo Fosciera (2012), “[...] o tratamento computacional da opinião vem representando um desafio na atualidade, quando temos à disposição um grande número de informações relevantes que nos permite entender o ponto de vista dos outros”.

Para alcançar esses objetivos, é necessário a realização de uma série de etapas. Primeiro, os dados precisam ser extraídos do Twitter usando técnicas de mineração de dados. Em seguida, esses dados passam por um processo de limpeza para garantir que sejam precisos e relevantes para a pesquisa. Finalmente, é realizada uma análise de sentimento para identificar e categorizar as opiniões expressas nos *tweets*.

Os resultados dessa pesquisa podem fornecer *insights* valiosos sobre a opinião pública em relação à vacinação da Covid-19 em crianças. Essas informações podem ser úteis para os formuladores de políticas, profissionais de saúde e o público em geral, ajudando a informar as decisões relacionadas à vacinação.

2 Justificativa

A pandemia da Covid-19 é a mais grave e a maior do século até agora, causando impactos em várias áreas como a saúde e a economia e afetando todos os países desenvolvidos e em desenvolvimento, ameaçando desfazer anos de crescimento sustentável (ROY; GHOSH,

2021). Essa crise epidemiológica destacou a importância da administração da saúde pública na prática e nos estudos (HATCHER, 2020).

Analisar os sentimentos gerados em redes sociais é um tema em crescimento e que está ganhando muito destaque, superando limitações de pesquisas tradicionais para entendimento da população, permitindo um ponto de vista global (SATTAR; ARIFUZZAMAN, 2021).

Dados divulgados pela Fiocruz em parceria com a Organização Mundial da Saúde (OMS), mostram que a taxa de vacinação infantil vem sofrendo uma queda brusca, sendo de 93,1% para 71,49%, deixando a população infantil exposta a doenças que já não eram preocupação, e, também, à Covid-19 (FIOCRUZ, 2022).

3 Objetivos

Este trabalho tem como objetivo principal identificar e avaliar o sentimento dos usuários do Twitter, através de suas postagens (*tweets*), relacionando as sensações identificadas com os acontecimentos reais que estamos vivenciando com a Covid-19 e sua vacinação em crianças. Para atingir esse objetivo, esta pesquisa se concentrará em:

- Apresentar o papel da rede social *Twitter* (X) como fonte de informações e os desafios na pandemia da Covid-19, especificamente relacionados à vacinação da doença em crianças;
- Aplicar a técnica de mineração mais relevante para extração dos resultados;
- Analisar e classificar os sentimentos negativos ou positivos dos usuários após a aplicação da mineração dos dados.

4 Fundamentação Teórica

Neste tópico serão apresentadas as explicações sobre alguns dos principais conceitos explorados neste trabalho.

4.1 Twitter

O Twitter, atualmente conhecido como X, é uma mídia social que está concentrada nos principais assuntos do momento, seus usuários podem compartilhar suas opiniões em textos curtos de até 280 caracteres, chamados *tweets*. Segundo sua política de dados “O Twitter é público e os *tweets* são imediatamente visíveis e pesquisáveis por qualquer pessoa em todo o mundo” (TWITTER, 2023). Porém, é possível que os usuários limitem a permissão de quem pode ver seus *tweets*, para os que não concordam com a regra.

Cada usuário do Twitter possui um nome único composto pelo caractere “@” precedido pelo nome ou apelido, os assuntos podem ser marcados pelo símbolo de cerquilha (#) seguido pelos tópicos. Essas palavras que identificam os assuntos são as *hashtags*, e os mais comentados no momento na plataforma ficam em destaque para chamar a atenção de outros usuários para a conversa. Todas as informações dos *tweets* podem ser utilizadas na filtragem de pesquisa pelos usuários.

Uma possibilidade de interação entre os usuários é o *retweet*, onde um conteúdo que já foi postado é republicado no perfil que quem usou esse recurso. Também é permitido ao usuário “curtir” algum conteúdo na rede, indicando que gostou daquela publicação e realizar comentários nas postagens de outros perfis.

Em 2022, a rede social Twitter foi comprada por Elon Musk, que no final de julho de 2023 trouxe algumas reformulações à rede social, por exemplo: a partir do dia 24 de julho de 2023 o passarinho azul deixou de ser o logotipo para ser trocado por um “X”, que agora também é o nome da rede social (O GLOBO, 2023). Outra mudança que ocorreu foi a alteração do nome dos *tweets* para *posts*, mas para a continuação do desenvolvimento da pesquisa ainda usaremos a palavra *tweet*, pois muitos usuários ainda usam o nome antigo e estão mais familiarizados com o termo.

4.2 Mineração de Texto

A mineração de textos é uma subárea especializada dentro do campo da mineração de dados. Seu principal objetivo é extrair informações significativas de textos que podem estar em vários formatos, como estruturados (por exemplo: planilhas e documentos HTML) ou não estruturados (por exemplo: textos planos). Na atual era digital, os dados mais comumente encontrados na internet são os não estruturados, isso torna a mineração de textos uma ferramenta essencial para analisar e interpretar esses dados. A mineração de textos proporciona que as organizações extraiam informações valiosas, que possam transformar em *insights* valiosos (HUPDATA, 2022).

As principais aplicações da mineração de textos são: categorização, que faz a identificação de temas-chave em um documento, essa técnica é útil para a compreensão do conteúdo geral de um grande volume de textos; outra aplicação é a sumarização, que envolve a criação de um resumo conciso das informações principais contidas em um texto, isso é útil quando se lida com grandes volumes de texto e se deseja obter uma compreensão rápida do conteúdo sem ter que ler todo o texto (GUPTA; LEHAL, 2009).

4.3 Análise de Sentimentos

A análise de sentimentos é uma aplicação emergente e fascinante da mineração de textos que se concentra na identificação e extração de informações subjetivas dos dados do texto. Essa técnica pode determinar o sentimento geral de um texto (por exemplo, se ele é positivo, negativo ou neutro), e, ainda, identificar emoções específicas expressas no texto, como ironia ou sarcasmo. A análise de sentimentos é bastante usada em mídias sociais e análises de opiniões para entender os sentimentos do público sobre um determinado assunto. O primeiro passo para a aplicação da análise de sentimentos é determinar o tema a ser estudado. Em seguida, é necessário realizar o tratamento inicial dos dados, sendo ele a divisão do texto em blocos para análise. Após a divisão, os *tokens* (unidades básicas individuais que possuem os valores necessários para a realização da análise) são obtidos através de um processo de filtragem e aprimoramento. Por fim, uma pontuação de sentimento é atribuída a cada bloco de acordo

com os *tokens* de cada um. Por meio dessa pontuação é possível classificar se o sentimento presente no bloco de texto é positivo, negativo ou neutro (AWRAHMAN; ALATAS, 2017).

4.4 Covid-19

A Covid-19, causada pelo coronavírus SARS-CoV-2, é uma infecção respiratória aguda que se caracteriza pela sua gravidade e alta transmissibilidade. O vírus foi descoberto em 2019 na China, em pacientes que apresentavam um tipo de pneumonia até então desconhecida (OPAS, 2021).

Os coronavírus compõem uma vasta família de vírus encontrados em várias espécies de animais, incluindo seres humanos. No entanto, é raro o caso em que um coronavírus de origem animal consiga infectar um ser humano (BRASIL. MINISTÉRIO DA SAÚDE, 2021).

Devido à sua alta transmissibilidade, a Covid-19 rapidamente se espalhou pelo mundo, e, em janeiro de 2020, a OMS declarou o surto do novo coronavírus como uma ESPII (OPAS, 2020).

A pandemia global desafiou os sistemas de saúde e exigiu respostas rápidas e eficazes dos governos e da comunidade científica. A busca por tratamentos eficazes e a implementação de medidas preventivas, como a vacinação em massa, tornaram-se prioridades em todo o mundo (THE WORLD BANK, 2020).

Dessa maneira, é possível perceber que a análise de sentimentos e as opiniões expressas nas redes sociais, como o Twitter, podem fornecer *insights* valiosos sobre a percepção do público em relação à pandemia e às medidas adotadas para combatê-la, reforçando a importância da mineração de dados e da análise de sentimentos neste cenário.

4.5 Vacina

As vacinas são as formas mais eficazes do combate a doenças que podem ser prevenidas e no combate a epidemias. A OMS estima que 2 a 3 milhões de mortes são evitadas por ano graças à vacinação (PFIZER, 2021).

Em setembro de 2022 a OMS recomendou o uso das vacinas contra a Covid-19 que foram validadas por ela, como Lista de Uso Emergencial. O primeiro programa de vacinação se deu no início de dezembro de 2020. As vacinas precisam garantir e atender aos padrões aceitáveis de qualidade, segurança e eficácia a partir de dados de ensaios clínicos, fabricação e controle de qualidade. De acordo com regulamentações e legislações nacionais, cada país tem a autoridade de emitir autorizações do uso emergencial para os produtos de saúde (OMS, 2022).

Em 17 de janeiro de 2021 ocorreu a primeira vacinação contra a Covid-19 no Brasil, 40 dias após a primeira vacinação no mundo, que ocorreu no Reino Unido. De acordo com especialistas em Saúde Pública, se as vacinas houvessem sido adquiridas mais cedo a saúde pública brasileira poderia ter evitado o colapso ocorrido, assim como o número de internações e mortes (FIOCRUZ, 2022). O início da vacinação infantil contra a Covid-19 foi no dia 14 de janeiro de 2022, com crianças entre 5 e 11 anos (FMUSP, 2022).

A vacinação da Covid-19 em crianças é separada por faixa etária. Crianças de 6 meses a 4 anos e 11 meses recebem a Pfizer Baby, de 3 a 4 anos tomam CoronaVac com o reforço da

Pfizer Baby e as crianças de 5 a 11 anos recebem a Pfizer Pediátrica (SECRETARIA DE SAÚDE DO DISTRITO FEDERAL, 2023).

4.6 Web Scraping

Web Scraping é uma técnica que permite a extração automatizada de dados de *sites* públicos na internet. Esses dados podem ser baixados ou armazenados para diversas finalidades. A técnica é essencial para extrair grandes quantidades de dados de maneira automatizada, economizando tempo em comparação com a cópia manual. Os “raspadores da *web*” são programas que visitam sites, escaneiam páginas e extraem informações (AWARI, 2022).

Esse tipo de “garimpo” da internet extrai informações relevantes de determinado *site* para serem analisadas. Esses dados serão usados no aprimoramento da tomada de decisões com maior chance de acerto e sucesso. O mesmo processo pode ser feito de forma manual, mas quando se fala de *Web Scraping* a ideia é automatizar o trabalho usando *bots*, dessa maneira é possível coletar um número muito maior de dados em um pequeno tempo (ROCKCONTENT, 2021).

Como o assunto é captar dados de outros *sites*, é importante tomar cuidado com os limites dessa prática, tanto em sentido legal quanto moral, pois esse processo se trata de uma coleta de dados estruturados da *web*, e, em geral, ele é usado por pessoas, empresas e, preocupantemente, criminosos, que procuram usar a enorme quantidade de dados disponíveis publicamente para tomar decisões mais inteligentes ou cometer crimes (CANALTECH, 2021).

4.7 BERT

O BERT (*Bidirectional Encoder Representation from Transformers*) é um modelo de Processamento de Linguagem Natural (PLN), proposto em 2018 pelo Google AI Language. Trata-se de uma rede neural capaz de compreender e aprender as formas de expressão da linguagem humana. Essa técnica é baseada no modelo *Transformer*, que permite o entendimento da relação entre palavras em uma frase, e não uma por uma em ordem. O BERT compreende completamente o contexto da palavra, ou seja, as que vêm antes e depois e suas relações, dessa maneira é possível entender os conteúdos de *sites* e intenções de usuários. Esse modelo é utilizado na análise de dados, com aplicações que incluem pesquisa de mercado, análise de comportamento de clientes em mídias sociais e extração de dados de *sites* de comércio eletrônico para análise da concorrência (ROCKCONTENT, 2020).

No BERT, os *encoders* são unidades fundamentais para o processamento da informação textual, pois compõem a arquitetura do modelo e são responsáveis por gerar representações vetoriais para as palavras ou *tokens* de uma sentença. Esses *encoders* utilizam mecanismos de autoatenção para compreender o contexto das palavras em uma sequência de texto, produzindo representações vetoriais (*embeddings*) para cada palavra, levando em consideração o contexto ao redor (MEDIUM, 2021).

Sendo assim, o BERT tem uma grande capacidade de entender o contexto da informação e gerar representações vetoriais, isso o torna muito eficiente para a análise de sentimentos.

4.8 Python 3

O Python 3 é uma linguagem de programação versátil e poderosa, com ela é possível trabalhar de forma rápida e integrar sistemas de maneira eficaz. É a versão mais recente da linguagem, lançada em 3 de dezembro de 2008 (PYTHON, 2023).

Python 3 é amplamente utilizado em domínios de aplicação, incluindo desenvolvimento web, computação científica e numérica, educação, interfaces gráficas, desenvolvimento de softwares e aplicações de negócios (PYTHON, 2023).

Essa linguagem de programação é muito usada em mineração de textos e análise de sentimentos, pois oferece técnicas de criação e modelos de *Machine Learning* com estruturas de dados de alto nível e eficientes e uma abordagem simples (REAL PYTHON, 2020).

4.9 VS Code

O *Visual Studio Code*, mais conhecido como VS Code, é um editor de código-fonte desenvolvido pela *Microsoft* e suporta muitas linguagens de programação. Essa ferramenta é conhecida por sua velocidade e eficiência, sendo o preferido para uso diário (VISUAL STUDIO CODE, 2023).

Ele disponibiliza recursos como a integração ao Git, ou seja, permite o controle de versões sem sair do editor; mudança de modo de linguagem; personalização, permitindo a mudança do *layout*; modo *zen*, onde o editor oculta todos os menus e janelas, não permitindo distrações; depuração; conclusão inteligente de código; entre outros. O VS Code ainda apresenta uma enorme biblioteca de extensões que permite adicionar novas linguagens, temas, depuradores e se conectar a serviços adicionais (VISUAL STUDIO CODE, 2023).

5 Trabalhos Similares

A pesquisa intitulada Análise de sentimentos dos conteúdos gerados pelos usuários no Twitter a partir das comunicações dos políticos em relação a vacina da Covid-19, desenvolvido por Alex Matias Gomes na Universidade Federal do Rio Grande do Norte em 2022 tem como objetivo analisar as comunicações de políticos brasileiros e os conteúdos gerados pelos usuários sobre a vacina contra a Covid-19 no Twitter. Através de comentários na rede social após as comunicações dos políticos serão classificados os sentimentos dos usuários.

De forma geral, sem fazer a segmentação por grupos de políticos, os resultados apontaram que o sentimento predominante dos CGU foi o negativo em 35,5% dos tweets, seguido pelos tweets com sentimento positivo com 34% e tweets com sentimento neutro com 30,5% (GOMES, 2022).

A pesquisa intitulada Análise de sentimentos sobre as vacinas contra covid-19: um estudo com algoritmo de *machine learning* em postagens no *twitter*, desenvolvido por Cecilio Merlotti Rodas, Suellen Elise Timm Barros, Rodrigo Ananias da Silva Souza, Silvana Aparecida Borsetti Gregorio Vidotti, publicado como artigo de mestrado na revista Saúde

Digital e Tecnologias Educacionais em 2022 tem como objetivo realizar uma análise de sentimentos através de postagens no *Twitter* (X) em relação às vacinas da Covid-19 no período de 17 de janeiro a 6 de março de 2021, utilizando a linguagem de programação *Python3* e aplicando bibliotecas específicas para a coleta, manipulação e visualização dos tweets e dos dados. “O sentimento que a informação produz nas pessoas pode ter impacto direto ou indireto em sua vida e na sociedade em geral apresentando significativa relevância, pois o sentimento pode ser um fator importante na experiência dos usuários” (RODAS et al., 2022).

A pesquisa intitulada *Análise de opiniões utilizando técnicas de mineração de dados em redes sociais. Estudo de caso: Twitter*, desenvolvido por Tatiane Gomes dos Santos pelo Centro Universitário de Anápolis (UniEVANGÉLICA) em 2017 tem como objetivo apresentar os resultados obtidos da análise e validação do uso das principais técnicas de Mineração de Dados no Twitter para a classificação de opiniões para um domínio específico definido durante o processo. Demonstra-se através da validação das principais técnicas de mineração de dados os seus aspectos e diferenças de aplicação. Investigando qual técnica se evidenciou a melhor execução para a classificação das opiniões nas mídias sociais e verificar os efeitos para a classificação, obtendo uma análise das opiniões extraídas das redes sociais e uma maior assertividade na classificação dos dados.

6 Metodologia

Esse estudo foi realizado em duas fases principais: a mineração dos dados por meio de *Web Scraping* e a análise de sentimentos. Ambas as etapas foram executadas utilizando a linguagem de programação *Python 3*, escolhida por suas técnicas e aplicações robustas para o Aprendizado de Máquina, e o ambiente de desenvolvimento VS Code.

Na primeira fase, foi usado bibliotecas como *Selenium*, que possibilita automatizar ações na *web*, ou seja, permite o controle do comportamento de uma página e uma interação automática com ela; a *Time*, que juntamente com seu método *sleep* força o navegador a fechar depois de um determinado tempo; a *Pandas*, que apresenta tecnologias de manipulação, leitura e visualização de dados; a *csv* para ler e escrever arquivos *.CSV*; a *OS* para a utilização das funcionalidades dependentes do sistema operacional; a *Codecs* para trabalhar com codificação de texto e conjuntos de caracteres e ao *RE* para a utilização de expressões regulares.

Para a mineração de dados da rede social Twitter, foi desenvolvido um algoritmo de *Web Scraping* utilizando a biblioteca *Selenium* para coletar dados de *tweets* relacionados a palavras-chaves pré-definidas. A coleta dos dados foi realizada anualmente, 2021, 2022 e 2023, e em intervalos de 10 dias. Após a coleta, os *tweets* foram limpos e armazenados em um arquivo *.CSV* para análise posterior.

Na segunda fase, foi realizado uma análise de sentimentos nos *tweets* coletados, para isso as bibliotecas usadas foram: *Transformers* para o fornecimento de APIs e ferramentas de treinamento de uma base de dados, ela inclui os métodos *BertTokenizer* para tokenizar o texto de entrada para o modelo BERT e a *BertForSequenceClassification* para tarefas de classificação de sequência, *Torch* para um aprendizado profundo, *Matplotlib* para o desenho de gráficos e *Seaborn* para a visualização dos dados.

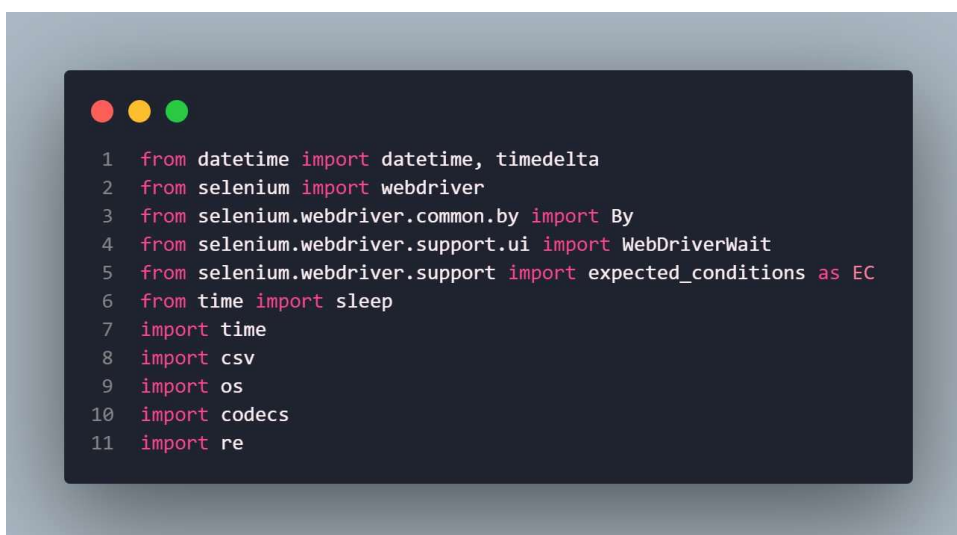
Para a análise de sentimentos, foi desenvolvido um algoritmo que utiliza a técnica BERT com um modelo e *token* pré-treinados. Cada *tweet* foi processado pelo algoritmo e classificado como positivo ou negativo, refletindo o sentimento dos usuários em relação à vacinação da Covid-19 em crianças.

7 Desenvolvimento

A primeira etapa da pesquisa consiste na coleta dos *tweets* relacionados à vacinação da Covid-19 em crianças através de um *Web Scraping*

O código começa com a configuração inicial, ou seja, com as importações das bibliotecas utilizadas, a função para a limpeza dos *tweets* para a próxima etapa, onde são removidas URLs, menções a usuários e *hashtags* e a configuração do *driver* do *Selenium* para a utilização do navegador *Google Chrome*.

Figura 1 Importação das bibliotecas



```
1 from datetime import datetime, timedelta
2 from selenium import webdriver
3 from selenium.webdriver.common.by import By
4 from selenium.webdriver.support.ui import WebDriverWait
5 from selenium.webdriver.support import expected_conditions as EC
6 from time import sleep
7 import time
8 import csv
9 import os
10 import codecs
11 import re
```

Fonte: Código desenvolvido pela autora

Figura 2 Função da limpeza dos *tweets* e configuração do *driver* do *Selenium*

```
1 def clean_tweet(tweet_text):
2     # Remove URLs
3     tweet_text = re.sub(r"http\S+|www\S+|https\S+", '', tweet_text)
4     # Remove menções a usuários (@username)
5     tweet_text = re.sub(r'@\w+', '', tweet_text)
6     # Remove hashtags (#hashtag)
7     tweet_text = re.sub(r'#\w+', '', tweet_text)
8     return tweet_text
9
10 # Configura o driver do Selenium para o Chrome
11 options = webdriver.ChromeOptions()
12 options.binary_location = r'C:\Program Files\Google\Chrome\Application\chrome.exe'
13
14 # Cria o driver Chrome
15 driver = webdriver.Chrome(options=options)
16
17 # URL para a abertura da página de login do twitter
18 url1 = 'https://twitter.com/login'
19
20 # Abre a página no navegador
21 driver.get(url1)
```

Fonte: Código desenvolvido pela autora

A próxima etapa envolve a autenticação e a coleta de dados no Twitter. Para isso foi utilizada a biblioteca *Selenium* para acessar a página de *login* do Twitter e realizar a autenticação usando um nome de usuário e senha pré-definidos. Após a autenticação bem-sucedida, as datas de início e fim para a coleta dos *tweets*, com intervalos de 10 dias são definidas. Este processo garante que a coleta de dados relevantes dentro de um período específico.

Figura 3 Autenticação na página de *login*

```
1 # LOGIN
2 # inserir usuário
3 sleep(3)
4 usuario = driver.find_element(By.XPATH, "//*[@name='text']")
5 usuario.send_keys("@Lara2R167598")
6 avanca = driver.find_element(By.XPATH, "//span[contains(text(),'Avançar')]")
7 avanca.click()
8
9 # inserir senha
10 sleep(3)
11 senha = driver.find_element(By.XPATH, "//*[@name='password']")
12 senha.send_keys("LSRR@16@09")
13 entrar = driver.find_element(By.XPATH, "//span[contains(text(),'Entrar')]")
14 entrar.click()
15 sleep(3)
```

Fonte: Código desenvolvido pela autora

Figura 4 Definição das datas de início e fim e o intervalo de pesquisa

```
1 # Data de início
2 data_inicio = datetime(2023, 7, 1)
3
4 # Intervalo de 10 dias
5 intervalo = timedelta(days=10)
6
7 # Data de término (10 dias após a data de início)
8 data_fim = data_inicio + intervalo
9
10 # Data final
11 data_final_desejada = datetime(2023, 10, 31)
```

Fonte: Código desenvolvido pela autora

A etapa subsequente envolve a codificação do *loop* para a coleta dos *tweets*. Inicialmente, é realizada a iteração dos intervalos de datas definidos para a coleta de dados. Em seguida, a URL para buscar os *tweets* específicos relacionados à vacinação da Covid-19 em crianças é construída. Após a construção da URL, é feita a navegação até a página correspondente ao intervalo de datas especificado. Nesta etapa, a página é rolada para baixo para permitir o carregamento de mais *tweets*.

Uma vez que a página é carregada, os elementos que contêm os *tweets* são identificados e sua extração é realizada. Os *tweets* coletados são então armazenados em uma lista para análise posterior. Este processo é repetido até que todos os intervalos de datas tenham sido percorridos e os *tweets* correspondentes coletados.

Figura 5 Lista para armazenamento dos *tweets*

```
1 # Inicializa uma lista para armazenar todos os tweets coletados
2 lista_total = []
3 count = 0
```

Fonte: Código desenvolvido pela autora

Figura 6 Loop da coleta de tweets

```
1 while data_inicio < data_final_desejada:
2
3     # Constroe a URL com as datas atuais
4     url = f'https://twitter.com/search?q=vacinação covid crianças since:
5         {data_inicio:%Y-%m-%d} until:{data_fim:%Y-%m-%d}&src=typed_query'
6
7     # Abre a página no navegador
8     driver.get(url)
9
10    # Rola a página para baixo para carregar mais tweets
11    for _ in range(20): # Rolando 5 vezes
12        driver.execute_script("window.scrollTo(0, 1000);")
13        time.sleep(1)
14
15    # Espera até que os tweets sejam carregados na página
16    print(f"Esperando por elementos ... {data_inicio:%Y-%m-%d}")
17    time.sleep(10)
18
19    # Encontra e extrai os tweets
20    tweets = driver.find_elements(By.CSS_SELECTOR, "div[data-testid='tweetText']")
21
22    lista = []
```

Fonte: Código desenvolvido pela autora

Figura 7 Continuação do loop da coleta de tweets

```
1 for index, tweet in enumerate(tweets):
2     try:
3         tweet_text = clean_tweet(tweet.text)
4         lista.append(tweet_text)
5     except Exception as e:
6         print(f"Erro ao extrair tweet: {e}")
7
8     # Adiciona os tweets coletados nesta rodada à lista total
9     lista_total.extend(lista)
10
11    # Avança a data de início e término para o próximo intervalo
12    data_inicio += intervalo
13    data_fim += intervalo
```

Fonte: Código desenvolvido pela autora

Finalmente, os *tweets* coletados são armazenados para análise posterior. Primeiro, é criado o diretório necessário para armazenar o arquivo .CSV. Em seguida, após a limpeza e

formatação adequada dos *tweets*, eles são escritos no arquivo .CSV. Esse processo garante que os dados coletados estão prontos e disponíveis para análises futuras.

Figura 8 Criação do diretório onde o arquivo .csv será salvo

```
1 # Diretório onde vai salvar o arquivo CSV
2 diretorio_destino = r'D:\ADS_FATEC\6_perodo\TG_II'
3
4 # Arquivo CSV
5 nome_arquivo_csv = 'tweets_2023.csv'
6
7 # Verifica se o diretório de destino existe
8 # se não existir, cria
9 if not os.path.exists(diretorio_destino):
10     os.makedirs(diretorio_destino)
11
12 # Caminho completo para o arquivo CSV
13 caminho_arquivo_csv = os.path.join(diretorio_destino, nome_arquivo_csv)
14
15 # Verifica se o arquivo CSV já existe
16 if os.path.exists(caminho_arquivo_csv):
17     modo_arquivo = 'a' # Modo de adição (append)
18 else:
19     modo_arquivo = 'w' # Modo de escrita (create)
```

Fonte: Código desenvolvido pela autora

Figura 9 Armazenamento dos *tweets*

```
1 # Salva todos os tweets coletados no arquivo CSV
2 with codecs.open(caminho_arquivo_csv, modo_arquivo, encoding='utf-8-sig') as csvfile:
3     csv_writer = csv.writer(csvfile)
4     for tweet in lista_total:
5         try:
6             # Remove quebras de linha por espaços em branco
7             tweet_sem_quebra_de_linha = tweet.replace('\n', ' ')
8             # Escreve o tweet no arquivo CSV
9             csv_writer.writerow([tweet_sem_quebra_de_linha])
10        except Exception as e:
11            print(f"Erro ao escrever tweet no arquivo CSV: {e}")
12
13 # Fecha o navegador
14 driver.quit()
```

Fonte: Código desenvolvido pela autora

Após a conclusão do processo de *Web Scraping*, a codificação para a análise de sentimentos utilizando o modelo BERT é iniciada. Para começar, as bibliotecas necessárias são importadas. Como os códigos de cada processo estão em arquivos diferentes, algumas

bibliotecas podem ser repetidas. Isso garante que todas as ferramentas necessárias para realizar a análise de sentimentos de maneira eficaz estão presentes.

Figura 10 Importação das bibliotecas para a análise de sentimentos

```
1 import os
2 import pandas as pd
3 from transformers import BertTokenizer, BertForSequenceClassification
4 import torch
5 import matplotlib.pyplot as plt
6 import seaborn as sns
```

Fonte: Código desenvolvido pela autora

A sequência do processo começa com a abertura do arquivo que foi salvo no código anterior. Em seguida, os dados não classificados são lidos linha por linha e adicionados a uma nova lista.

Figura 11 Leitura do CSV salvo no *Web Scraping*

```
1 # Carrega os dados do CSV com tweets não classificados
2 lines = []
3 file_path = 'D:/ADS_FATEC/6_período/TG_II/tweets_2023.csv'
4
5 # Abre o arquivo CSV e lê os dados
6 with open(file_path, 'r', encoding='utf-8-sig') as file:
7     for line in file:
8         # Adiciona cada linha à lista, removendo qualquer espaço em branco adicional
9         lines.append(line.strip())
```

Fonte: Código desenvolvido pela autora

O passo subsequente envolve a configuração do modelo BERT. Para isso, é preciso especificar o modelo pré-treinado que será utilizado, o qual é fornecido pelo próprio BERT.

Uma vez tendo o modelo pré-treinado, o tokenizador e o modelo de classificação de sequência são carregados a partir da biblioteca *Transformers*.

Figura 12 Configuração do modelo BERT

```
1 # Escolhe o modelo pré-treinado
2 model_name = "bert-base-uncased"
3
4 # Carrega o tokenizador e o modelo
5 tokenizer = BertTokenizer.from_pretrained(model_name)
6 model = BertForSequenceClassification.from_pretrained(model_name)
```

Fonte: Código desenvolvido pela autora

Para cada *tweet* não classificado da lista criada anteriormente o algoritmo tokeniza o texto do *tweet* utilizando o tokenizador e logo em seguida realiza a inferência com o modelo pré-treinado para a obtenção das probabilidades de cada classe de sentimento. Assim, o sentimento do *tweet* com base na classe com maior probabilidade é determinado. Após esse processo, a contagem de *tweets* classificados como “positivo” ou “negativo” é atualizada e a informação é armazenada na lista de *tweets* classificados.

Figura 13 Configuração das listas para a análise de sentimentos

```
1 # Lista para armazenar os sentimentos classificados
2 sentiments = []
3
4 # Inicializa dicionários para contar os sentimentos
5 contagem_sentimentos = {'Positivo': 0, 'Negativo': 0}
6
7 # Cria uma lista para armazenar os tweets classificados
8 classified_tweets = []
```

Fonte: Código desenvolvido pela autora

Figura 14 Classificação dos *tweets* através do modelo BERT

```
1 # Para cada tweet não classificado
2 for tweet in lines:
3     # Tokeniza o texto
4     inputs = tokenizer(tweet, return_tensors="pt", padding=True, truncation=True)
5
6     # Realiza a inferência com o modelo
7     outputs = model(**inputs)
8     logits = outputs.logits
9
10    # Calcula as probabilidades
11    probabilities = torch.softmax(logits, dim=1)
12
13    # Obtem o índice da classe com a maior probabilidade
14    predicted_class = torch.argmax(probabilities, dim=1).item()
15
16    # Interpreta o resultado
17    sentiment_classes = ["Positivo", "Negativo"]
18    sentiment = sentiment_classes[predicted_class]
19
20    sentiments.append(sentiment)
21
22    # Atualiza a contagem de acordo com o sentimento
23    contagem_sentimentos[sentiment] += 1
24
25    # Salva o tweet classificado com seu sentimento em uma lista
26    classified_tweets.append((tweet, sentiment))
```

Fonte: Código desenvolvido pela autora

Figura 15 Verifica se o arquivo .csv para salvar *tweets* classificados existe

```
1 # Verifica se o arquivo 'tweets_classificados.csv' já existe e, se existir, apaga
2 output_file_path = 'D:/ADS_FATEC/6_perodo/TG_II/tweets_classificados_2023.csv'
3 if os.path.exists(output_file_path):
4     os.remove(output_file_path)
```

Fonte: Código desenvolvido pela autora

Figura 16 Salva os *tweets* classificados no arquivo .csv

```
1 # Salva os tweets classificados em um arquivo CSV
2 classified_data = pd.DataFrame(classified_tweets, columns=['Tweets', 'Sentimento'])
3 classified_data.to_csv(output_file_path, index=False, encoding='utf-8-sig')
```

Fonte: Código desenvolvido pela autora

Após a classificação dos *tweets*, a análise da distribuição dos sentimentos é realizada. Isso envolve o cálculo da quantidade e da porcentagem de *tweets* classificados como “positivo” ou “negativo”. Para visualizar essa distribuição, um gráfico de barras é criado utilizando as bibliotecas *Matplotlib* e *Seaborn*.

Figura 17 Geração dos gráficos

```
1 # Mostra o total de positivo e negativo
2 total_positivo = contagem_sentimentos['Positivo']
3 total_negativo = contagem_sentimentos['Negativo']
4
5 # Calcula as porcentagens
6 total_tweets = total_positivo + total_negativo
7 percent_positivo = (total_positivo / total_tweets) * 100
8 percent_negativo = (total_negativo / total_tweets) * 100
9
10 # Cria um gráfico de barras para visualizar a distribuição dos sentimentos
11 sentiment_labels = ['Positivo', 'Negativo']
12 sentiment_percentages = [percent_positivo, percent_negativo]
13
14 plt.figure(figsize=(6, 4))
15 sns.barplot(x=sentiment_labels, y=sentiment_percentages)
16 plt.title('Distribuição de Sentimentos (Porcentagem)')
17 plt.xlabel('Sentimento')
18 plt.ylabel('Porcentagem')
19 plt.ylim(0, 100)
20 plt.show()
```

Fonte: Código desenvolvido pela autora

8 Resultados e Discussões

A vacinação infantil contra a Covid-19, que abrange crianças entre 5 e 11 anos, teve início no dia 14 de janeiro de 2022. A discussão sobre a vacinação infantil, no entanto, já era relevante desde o início da imunização em adultos, em 17 de janeiro de 2021. Nesse contexto, a coleta de dados dos *tweets* para a análise de sentimentos foi realizada, abrangendo os anos de 2021, 2022 e 2023, proporcionando uma visão abrangente das opiniões e sentimentos do público durante esse período crucial.

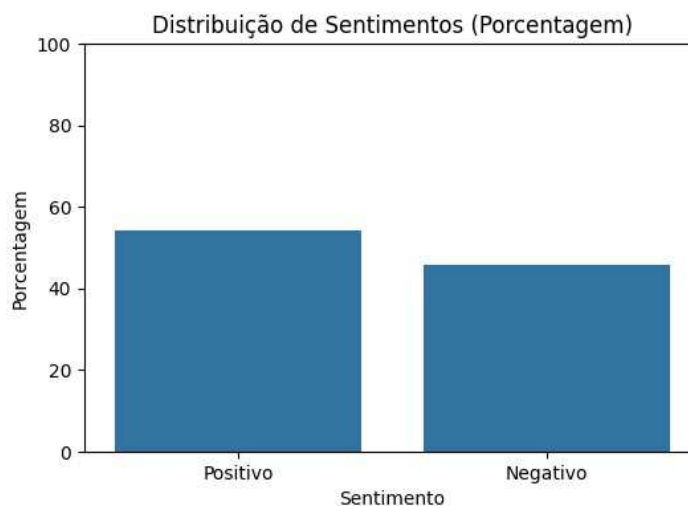
No ano de 2021, iniciou-se a coleta de *tweets*, resultando em um total de 340. Este foi o primeiro ano de coleta, estabelecendo assim a linha de base para comparações futuras.

Em 2022, houve um aumento significativo na coleta de *tweets*. O total coletado foi de 435, um aumento notável em relação ao ano anterior. Isso indica um maior engajamento no Twitter durante este ano.

No entanto, em 2023, até a data de 31 de outubro, a coleta de *tweets* diminuiu drasticamente para 180. Isso pode ser devido a vários fatores, incluindo possíveis mudanças no comportamento do usuário.

Na análise de sentimentos para o ano de 2021 observou-se uma distribuição equilibrada das opiniões. Os sentimentos positivos e negativos estão próximos quanto ao número de publicações no Twitter. Os sentimentos positivos são ligeiramente superiores aos negativos, conforme indicado no gráfico abaixo.

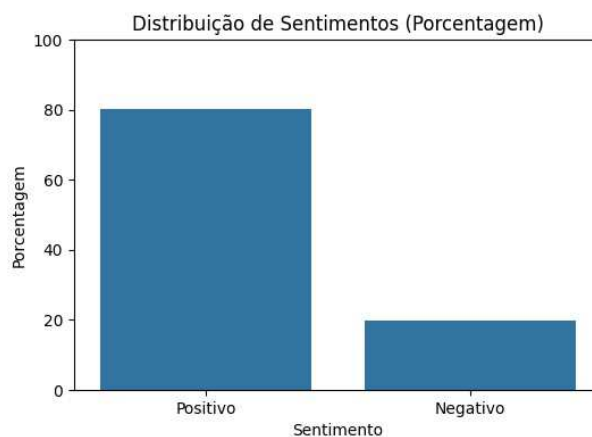
Figura 18 Análise de Sentimentos em 2021



Fonte: Gráfico desenvolvido pela autora

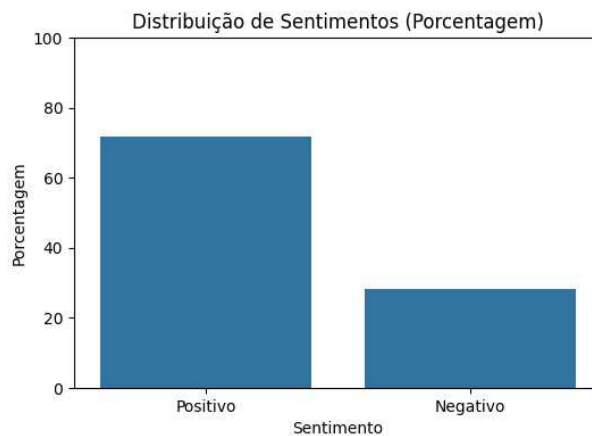
Em 2022, houve um expressivo aumento nos sentimentos positivos em relação os negativos, sugerindo uma possível mudança na percepção e aceitação da vacinação infantil da Covid-19.

As possíveis influências para esse aumento podem incluir: campanhas de conscientização e esclarecimentos sobre a eficácia e segurança da vacinação infantil e a divulgação de resultados científicos relacionados à efetividade das vacinas.

Figura 19 Análise de Sentimentos em 2022

Fonte: Gráfico desenvolvido pela autora

Embora haja, em 2023, uma diminuição nos sentimentos positivos em relação a 2022, sua predominância se mantém, indicando um cenário de estabilidade na percepção sobre a vacinação infantil.

Figura 20 Análise de Sentimentos em 2023

Fonte: Gráfico desenvolvido pela autora

A evolução temporal dos sentimentos pode fornecer *insights* valiosos sobre a percepção pública e a eficácia das estratégias de informação e divulgação. O aumento nos sentimentos positivos em 2022 e sua continuidade em 2023 são indicativos de uma mudança na percepção pública em favor da vacinação infantil. Isso é um testemunho do poder das redes sociais como uma plataforma para moldar opiniões e atitudes públicas.

9 Conclusões

A análise de sentimentos é uma ferramenta valiosa que permite entender as opiniões e atitudes da sociedade em relação a questões críticas, não apenas em situações de pandemia como a Covid-19, mas em todos os casos. Isso é particularmente relevante quando se trata da vacinação contra a Covid-19 em crianças. Ela fornece aos pesquisadores, profissionais de saúde e formuladores de políticas a capacidade de monitorar as mudanças nas percepções do público ao longo do tempo, identificar preocupações predominantes e combater a desinformação ou mal-entendidos sobre a vacinação infantil.

Embora seja especialmente vital durante uma pandemia, onde as opiniões e comportamentos do público podem afetar diretamente a saúde e o bem-estar da comunidade, sua aplicação é igualmente relevante e necessária em todos os aspectos da sociedade para informar e orientar ações eficazes.

Essa análise de sentimentos poderá ser expandida para além da pandemia da Covid-19 e vacinação em crianças, reconhecendo sua utilidade em todos os aspectos da sociedade. O objetivo é utilizar essa ferramenta para monitorar percepções do público sobre diversas questões críticas, permitindo a identificação de preocupações predominantes e o combate à desinformação.

O monitoramento contínuo das mudanças nas percepções do público também é previsto, respondendo a mudanças ao longo do tempo. Dessa forma, é possível estar sempre preparado para abordar efetivamente qualquer desafio que possa surgir.

Referências

AWARI. **O que é Web Scraping e para que serve essa técnica?** Disponível em: <https://awari.com.br/web-scraping/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=O%20que%20é%20Web%20Scraping%20e%20para%20que%20serve%20essa%20técnica>. Acesso em: 15 de agosto de 2023.

AWRAHMAN B.; ALATAS B. **Sentiment Analysis and Opinion Mining within Social Networks using Konstanz Information Miner**. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(1):15-22, 2017.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação: Conceitos e tecnologia das máquinas de busca**. Tradução de Leandro Krug Wives, Viviane Pereira Moreira. 2. ed. Porto Alegre: Bookman, 2013.

BRASIL. AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA – ANVISA. **Covid-19: Anvisa aprova vacina da Pfizer para crianças entre 6 meses e 4 anos**. Disponível em: <<https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2022/covid-19-anvisa-aprova-vacina-da-pfizer-para-criancas-entre-6-meses-e-4-anos>>. Acesso em: 8 de maio de 2023.

BRASIL. CASA CIVIL. 90% dos lares brasileiros já tem acesso à internet no Brasil, aponta pesquisa. Disponível em: <<https://www.gov.br/casacivil/pt-br/assuntos/noticias/2022/setembro/90-dos-lares-brasileiros-ja-tem-acesso-a-internet-no-brasil-aponta-pesquisa#:~:text=Conectividade-,90%25%20dos%20lares%20brasileiros%20já%20tem%20acesso,internet%20no%20Brasil%2C%20aponta%20pesquisa&text=Em%202021%2C%20o%20número%20de,mais%20do%20que%20em%202019>>. Acesso em: 8 de maio de 2023.

BRASIL. MINISTÉRIO DA SAÚDE. Mais de 69 milhões de brasileiros não tomaram a primeira dose de reforço contra a Covid-19. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/noticias/2022/novembro/mais-de-69-milhoes-de-brasileiros-nao-tomaram-a-primeira-dose-de-reforco-contr-a-covid-19#:~:text=Mais%20de%2069%20milhões%20de%20brasileiros%20ainda%20não%20voltaram%20aos,mas%20ainda%20não%20se%20vacinaram>>. Acesso em: 11 de maio de 2023.

BRASIL. MINISTÉRIO DA SAÚDE. O que é a Covid-19? Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus/o-que-e-o-coronavirus>>. Acesso em 11: de maio de 2023.

CANALTECH. O que é web scraping e como ocorre? Disponível em: <<https://canaltech.com.br/seguranca/o-que-e-web-scraping/>>. Acesso em: 20 de agosto de 2023.

CORONAVÍRUS BRASIL. Painel Coronavírus. Disponível em: <<https://covid.saude.gov.br>>. Acesso em: 12 de maio de 2023.

FACULDADE DE MEDICINA DA UNIVERSIDADE DE SÃO PAULO (FMUSP). Primeira vacinação do público infantil no Brasil contra COVID-19 começou no HCFMUSP. Disponível em: <<https://www.fm.usp.br/fmusp/noticias/-primeira-vacinacao-do-publico-infantil-no-brasil-contr-a-covid-19--comecou-no-hcfmusp>>. Acesso em: 13 de junho de 2023.

FIOCRUZ. Brasil celebra um ano da vacina contra a COVID-19. Disponível em: <<https://portal.fiocruz.br/noticia/brasil-celebra-um-ano-da-vacina-contr-a-covid-19>>. Acesso em 11: de maio de 2023.

FIOCRUZ. Covid-19: Brasil registra uma morte por dia entre crianças de 6 meses a 5 anos em 2022. Disponível em: <<https://portal.fiocruz.br/noticia/covid-19-brasil-registra-uma-morte-por-dia-entre-criancas-de-6-meses-5-anos-em-2022#:~:text=Compartilhar%3A,nessa%20faixa%20etária%20no%20período>>. Acesso em: 10 de maio de 2023.

FIOCRUZ. Vacinação infantil sofre queda brusca no Brasil. Disponível em: <<https://portal.fiocruz.br/noticia/vacinacao-infantil-sofre-queda-brusca-no-brasil>>. Acesso em: 05 de abril de 2023.

FOSCHIERA, S. M. P. **A semântica da emoção: um estudo contrastivo a partir da FrameNet e da roda das emoções**. 2012. 290 p. Tese (Doutorado em Linguística Aplicada) - Universidade do Vale do Rio dos Sinos, São Leopoldo, 2012.

GOMES, A. **ANÁLISE DE SENTIMENTOS DOS CONTEÚDOS GERADOS PELOS USUÁRIOS NO TWITTER A PARTIR DAS COMUNICAÇÕES DOS POLÍTICOS EM RELAÇÃO A VACINA DA COVID-19**. Mestre em Administração, pelo Programa de Pós-Graduação em Administração—Universidade Federal do Rio Grande do Norte.: [s.n.], 2022.

GUPTA, V.; LEHAL, G. **A survey of text mining techniques and applications**. Journal of emerging technologies in web intelligence, 1(1):60-76, 2009.

HASHTAG. **GUIA PARA AUTOMAÇÃO WEB NO PYTHON UTILIZANDO O SELENIUM**. Disponível em: <<https://www.hashtagtreinamentos.com/automacao-web-no-python#:~:text=O%20Selenium%20Python%20é%20uma,extremamente%20útil%20para%20programadores%20web>>. Acesso em: 16 de agosto de 2023.

HATCHER, W. **A Failure of Political Communication Not a Failure of Bureaucracy: The Danger of Presidential Misinformation During the COVID-19 Pandemic**. American Review of Public Administration, v. 50, n. 6–7, p. 614–620, 2020.

HOME HOST. **Como usar Selenium Python para melhorar os testes web**. Disponível em: <<https://www.homehost.com.br/blog/pythondjango/selenium-python/>>. Acesso em: 16 de agosto de 2023.

HUGGING FACE. **Transformers**. Disponível em: <<https://huggingface.co/docs/transformers/index>>. Acesso em: 15 de agosto de 2023.

HUPDATA: Data Analysis Solutions. **Mineração de Texto. O que é? Como aplicar?** Disponível em: <<https://hupdata.com/mineracao-de-texto-o-que-e-como-aplicar/>>. Acesso em: 05 de abril de 2023.

INSTITUTO BUTANTAN. **Qual a diferença entre SARS-CoV-2 e Covid-19? Prevalência e incidência são a mesma coisa? Mortalidade e letalidade?** Disponível em: <<https://butantan.gov.br/covid/butantan-tira-duvida/tira-duvida-noticias/qual-a-diferenca-entre-sars-cov-2-e-covid-19-prevalencia-e-incidencia-sao-a-mesma-coisa-e-mortalidade-e-letalidade>>. Acesso em: 10 de maio de 2023.

LINKEDIN. **Twitter agora é X: o que está por trás da mudança**. Disponível em: <<https://www.linkedin.com/news/story/twitter-agora-é-x-o-que-está-por-trás-da-mudança-5706852/>>. Acesso em: 10 de agosto de 2023.

LIU, B. **Sentiment analysis and subjectivity**. In: INDURKHYA, N.; DAMERAU, F. Handbook of Natural Language Processing. 2. ed. [Londres]: Chapman and Hall/CRC, 2010. p. 1-38.

MAGALHÃES, L.; CARVALHO, R.; MORAES, I., *et al.* **Análise de sentimentos utilizando o algoritmo naive bayes multinominal.** In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. Anais [...]. Florianópolis: Ancib, Universidade Federal de Santa Catarina, 2019.

MATPLOTLIB. **Tutorials.** Disponível em: <https://matplotlib.org/stable/tutorials/index.html>. Acesso em: 25 de agosto de 2023.

MEDIUM. **Usando o BERT para Análise de Sentimentos.** Disponível em: <https://arthuremanuel-carosia.medium.com/usando-o-bert-para-análise-de-sentimentos-363f76eee15e>. Acesso em: 04 de agosto de 2023.

MURTHY, D. **Twitter: “social communication in the twitter age”.** In: MURTHY, D. Twitter: social communication in the twitter age (digital media and society). 2. ed. Medford: Polity Press, Apple Books, 2018. 220 p.

NAKAGAKI, M.; SARPONG, E. **Superando os desafios da COVID-19 e da exclusão digital em áreas rurais para alcançar as metas de acesso universal.** Panorama Setorial da Internet, v. 13, n. 1, p. 1–9, 2021.

O GLOBO. **O X da questão: como é que vamos nos referir agora ao Twitter?** Disponível em: <https://oglobo.globo.com/economia/tecnologia/noticia/2023/08/06/o-x-da-questao-como-e-que-vamos-nos-referir-agora-ao-twitter.ghtml>. Acesso em: 10 de agosto de 2023.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **Coronavirus disease (COVID-19): Vaccines.** Disponível em: [https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-\(covid-19\)-vaccines](https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-vaccines). Acesso em: 12 de maio de 2023.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE (OPAS). **Histórico da pandemia de COVID-19.** Disponível em: <https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>. Acesso em: 11 de maio de 2023.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE (OPAS). **OMS declara Emergência de Saúde Pública de Importância Internacional por surto de novo coronavírus.** Disponível em: <https://www.paho.org/pt/news/30-1-2020-who-declares-public-health-emergency-novel-coronavirus>. Acesso em: 11 de maio de 2023.

PFIZER. **Saiba tudo sobre vacinas.** Disponível em: <https://www.pfizer.com.br/sua-saude/vacinacao/tudo-sobre-vacinas>. Acesso em: 12 de maio de 2023.

PYTHON. **Applications for Python.** Disponível em: <https://www.python.org/about/apps/>. Acesso em: 10 de junho de 2023.

PYTHON. **About Python.** Disponível em: <https://www.python.org/about/>. Acesso em: 10 de junho de 2023.

PYTHON. **The Python Standard Library.** Disponível em: <https://docs.python.org/3/library/index.html>. Acesso em: 10 de agosto de 2023.

PYTHON. Tutorial do Python. Disponível em: <<https://docs.python.org/pt-br/3/tutorial/index.html>>. Acesso em: 10 de junho de 2023.

PYTORCH. Pytorck documentation. Disponível em: <<https://pytorch.org/docs/stable/index.html>>. Acesso em: 20 de agosto de 2023.

REAL PYTHON. Sentiment Analysis: First Steps With Python's NLTK Library. Disponível em: <<https://realpython.com/python-nltk-sentiment-analysis/>>. Acesso em: 10 de junho de 2023.

RECUERO, R.; ZAGO, G. A Economia do retweet: redes, difusão de informações e capital social no Twitter. Contracampo, Niterói, v. 24, n. 1, p. 19-43, jul. 2011.

ROCHA, M.; CORTEZ, P.; NEVES, J. M. Análise inteligente de dados: algoritmos e implementação em Java. Lisboa: FCA, 2008. 204 p.

ROCKCONTENT. Descubra o que é Web Scraping e como isso pode ajudar a sua estratégia digital. Disponível em: <<https://rockcontent.com/br/blog/web-scraping/>>. Acesso em: 20 de agosto de 2023.

ROCKCONTENT. Tudo sobre o BERT: o novo algoritmo do Google que promete revolucionar as SERPs. Disponível em: <<https://rockcontent.com/br/blog/bert/>>. Acesso em: 04 de agosto de 2023.

RODAS, C; BARROS, S; SOUZA R., et al. Análise de sentimentos sobre as vacinas contra covid-19: um estudo com algoritmo de machine learning em postagens no twitter. Revista de Saúde Digital e Tecnologias Educacionais. [online], volume 7, número especialIII. Editor responsável: Luiz Roberto de Oliveira. Fortaleza, fevereiro de 2022, p. 24-44.

ROY, S.; GHOSH, P. A comparative study on distancing, mask and vaccine adoption rates from global twitter trends. Healthcare (Switzerland), v. 9, n. 5, p. 1–14, 2021.

SANTOS, T. ANÁLISE DE OPINIÕES UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS EM REDES SOCIAIS. ESTUDO DE CASO: TWITTER. Bacharel em Engenharia de Computação—Centro Universitário de Anápolis— UniEVANGÉLICA: [s.n.], 2017.

SATTAR, N. S.; ARIFUZZAMAN, S. Covid-19 vaccination awareness and aftermath: Public sentiment analysis on twitter data and vaccinated population prediction in the USA. Applied Sciences (Switzerland), v. 11, n. 13, 2021.

SCIELO BRASIL. Inteligência Artificial e os rumos do processamento do português brasileiro. Disponível em: <<https://www.scielo.br/j/ea/a/63sbv5qSnnrqg8WpVwpgXzD/>>. Acesso em: 01 de agosto de 2023.

SEABORN. An introduction to seaborn. Disponível em: <<https://seaborn.pydata.org/tutorial/introduction>>. Acesso em: 25 de agosto de 2023.

SECRETARIA DE SAÚDE DO DISTRITO FEDERAL. **Vacinação Infantil contra a COVID-19.** Disponível em: <<https://www.saude.df.gov.br/vacinacao-infantil-contr-a-covid-19#:~:text=A%20vacinação%20infantil%20contra%20a,a%20segunda%20e%20a%20terceira>>. Acesso em: 13 de junho de 2023.

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, 2022.

SOMBRA, T. **Reconhecimento de padrões em rede social científica: Aplicação do algoritmo Naive Bayes para classificação de Papers no Mendeley.** 2018. 198 p. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro e Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018.

THE WORLD BANK. **Covid-19 no Brasil: Impactos e Respostas de Políticas Públicas.** Disponível em: <<https://www.worldbank.org/pt/country/brazil/publication/covid-19-in-brazil-impacts-policy-responses>>. Acesso em: 15 de maio de 2023.

TWITTER PRIVACIDADE. **Política de privacidade twitter.** Disponível em: <https://twitter.com/pt/privacy/previous/version_15>. Acesso em: 10 de maio de 2023.

VISUAL STUDIO CODE. **Getting Started.** Disponível em: <<https://code.visualstudio.com/Docs>>. Acesso em: 20 de agosto de 2023.