

Análise de Sentimentos de Tweets relacionados a Vacinas antes e durante a Pandemia da COVID-19 no Brasil

Ikaro A. de L. Silva¹, Taciana P. Falcão¹, Ermeson C. de Andrade¹

¹Departamento de Computação – Universidade Federal Rural de Pernambuco (UFRPE)
CEP 52.171-900 – Recife – PE – Brasil

{ikaro.alef, taciana.pontual, ermeson.andrade}@ufrpe.br

Abstract. *In early 2020, the COVID-19 disease spread rapidly around the world and one of the ways to fight it is the vaccine. Governments faced problems with fake news and anti-vaccination groups. Thus, it is necessary to understand the feelings of the population in order to propose efficient public policies. This article describes a sentiment analysis on vaccine-related tweets in Brazil from June 2020 to June 2021. The results revealed peaks in total tweets in January and May 2021, the predominance of positive tweets, and feelings of confidence, fear, submission and sadness. They are also associated with former President Jair Bolsonaro. The negative polarity was the least common, showing that the Brazilian population was receptive to vaccines.*

Resumo. *No início de 2020, a doença COVID-19 se alastrou rapidamente pelo mundo e uma das formas de combatê-la é a vacina. Governos enfrentaram problemas com notícias falsas e grupos contrários à vacinação. Assim, se faz necessário entender os sentimentos da população para propor políticas públicas eficientes. Este artigo descreve uma análise de sentimentos em tweets relacionados a vacinas no Brasil de junho de 2020 a junho de 2021. Os resultados revelaram picos do total de tweets em janeiro e maio de 2021, a predominância de tweets positivos e sentimentos de confiança, medo, submissão e tristeza. Também estão associados ao ex-presidente Jair Bolsonaro. A polaridade negativa foi a menos comum, mostrando que a população brasileira foi receptiva às vacinas.*

1. Introdução

Em dezembro de 2019, a Comissão de Saúde Municipal da cidade de Wuhan na China emitiu um relatório informando que foram detectados casos anormais de uma “pneumonia viral” desconhecida naquela mesma cidade [WHO 2020]. As primeiras descrições relatavam sintomas como febre, tosse, mialgia (dor muscular) e fadiga. A doença até então desconhecida se espalhou rapidamente e em poucas semanas casos foram detectados na Europa. Após novos estudos, o vírus dessa pneumonia foi identificado como da família dos coronavírus e nomeado *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2), ou síndrome respiratória aguda severa coronavírus 2. Em português a doença foi batizada de COVID-19 [Carvalho et al. 2021].

Diante da alta taxa de transmissão entre humanos, diversos governos e empresas iniciaram uma corrida contra o tempo para desenvolver um tratamento e a prevenção contra a COVID-19. Uma das formas de prevenção é a vacina com o vírus neutralizado. Enquanto o desenvolvimento de outras vacinas pode levar de 10 a 15 anos, as vacinas

contra a COVID-19 foram criadas em aproximadamente um ano após a detecção do novo vírus [Burgos et al. 2021]. Tal feito elevou a desconfiança de diversas pessoas acerca das vacinas [Troiano and Nardi 2021] gerando notícias falsas desencorajando a vacinação [Barifouse 2020]. Isso pode atrasar ou prejudicar a erradicação de novas e velhas doenças já eliminadas. Com isso em mente, governos e empresas veem a necessidade de compreender a opinião pública a fim de ajustar suas políticas de educação e conscientização sobre saúde. Há várias formas de entender o ponto de vista da população e uma delas é através das redes sociais. O Twitter é uma rede social fundada em 2006, onde os usuários compartilham mensagens e opiniões com até 280 caracteres. Seria humanamente impossível analisar individualmente cada *tweet*. Assim, é necessário recorrer à área de Mineração de Texto e utilizar uma técnica chamada Análise de Sentimentos para extrair e analisar informações do conteúdo dos *tweets* [Medhat et al. 2014].

A análise de sentimentos de *tweets* sobre vacinas pode fornecer uma visão valiosa sobre as opiniões e preocupações das pessoas em relação à imunização. Alguns estudos têm utilizado processamento de linguagem natural e classificação de texto para avaliar a polaridade de *tweets* sobre vacinas em geral, enquanto outros se concentram em eventos específicos, como a aprovação de uma nova vacina. No entanto, ainda há poucos estudos que abordam a análise de sentimentos de *tweets* sobre vacinação durante a pandemia de COVID-19 no Brasil. Portanto, este artigo pretende preencher essa lacuna na literatura e objetiva identificar e analisar o sentimento de usuários do Twitter no Brasil sobre as vacinas no Brasil durante a pandemia da COVID-19 através do conteúdo textual postado por eles nessa plataforma. Para isso, foram coletados aproximadamente um milhão de *tweets* entre 01/06/2020 e 11/06/2021. Os resultados revelaram que a maioria dos *tweets* tem polaridade positiva e sentimentos de confiança, tristeza e medo além de estarem associados ao governo do ex-presidente Jair Bolsonaro. A polaridade negativa foi a menos comum, mostrando que a população brasileira foi bem receptiva às vacinas.

Este artigo está dividido em: A Seção 2 apresenta conceitos básicos sobre as técnicas e as teorias envolvidas neste trabalho. A Seção 3 detalha os trabalhos relacionados. A Seção 4 demonstra os passos seguidos para o desenvolvimento e aquisição dos resultados desta pesquisa. A Seção 5 apresenta os resultados obtidos. Por fim, a Seção 6 apresenta as conclusões alcançadas neste estudo e também sugere trabalhos futuros.

2. Fundamentação

Nesta seção, abordaremos os conceitos básicos que foram explorados neste trabalho

2.1. Twitter

O Twitter é uma mídia social com cerca de 1,3 bilhão de contas cadastradas e 336 milhões de usuários ativos postando em torno de 500 milhões de *tweets* por dia [Clement 2020]. Criada em Março de 2006, o Twitter cumpre um papel importante para divulgação jornalística, científica e opinativa. Grandes fontes de informação como The New York Times, BBC, CNN e Nature possuem perfis oficiais e participam ativamente nesta mídia. Apesar da limitação de 280 caracteres por *tweet*, o número de usuários ativos é tão significativo que diversas pesquisas usam o Twitter como fonte de dados [Ahmed et al. 2017], especialmente os que envolvem mineração de texto [Karami et al. 2020]

2.2. Vacinação no Brasil

Em 1973, o governo do Brasil criou o Programa Nacional de Imunizações (PNI) sendo hoje, parte fundamental do Sistema Único de Saúde (SUS), criado em 1988. Reconhecido internacionalmente como um dos melhores programas de vacinação do mundo [Jaguraba 2018], o PNI também foi responsável por alcançar marcos significativos na saúde pública do Brasil, como a erradicação da poliomielite em 1994 e a eliminação da transmissão endêmica da febre amarela em 2018. Além disso, o PNI tem desempenhado um papel fundamental na campanha de vacinação contra a COVID-19.

Em 17 de janeiro de 2021, deu-se início à campanha de vacinação contra a COVID-19 no Brasil, tendo o foco principal os grupos de risco como profissionais de saúde, idosos, indígenas e quilombolas e mais tarde seguiu por faixa etária. Com um início gradual, essa campanha de vacinação enfrentou alguns problemas como conflitos políticos internacionais com China e Índia (grandes produtores de insumos farmacêuticos), conflitos internos entre o Governo Federal e o Ministério da Saúde, negligência por parte do Governo Federal e grupos negacionistas [Maciel et al. 2022]. Apesar do lento começo e das adversidades, quase um ano após o início da campanha, cerca de 70% da população brasileira já tinha recebido duas doses ou dose única da vacina [Mathieu et al. 2021].

2.3. Mineração de texto

Mineração de texto é uma subárea da Mineração de Dados que permite extrair novas informações, anteriormente desconhecidas, de diversas fontes de dados escritos. Desta forma, é possível forma novos fatos ou hipóteses que podem ser explorados mais a fundo por meios mais convencionais, como gráficos por exemplo [Hearst 2003]. Dentre as principais aplicações práticas da Mineração de texto, é possível citar a categorização de texto, a classificação de documentos, a análise de sentimentos, a sumarização de documento e a detecção de spam [Berry and Kogan 2010]. Neste trabalho, mais especificamente, utilizaremos a análise de sentimentos.

A Análise de Sentimentos (AS), também conhecida como Mineração de Opinião, se define como a busca e identificação de opiniões, atitudes e emoções de um autor sobre determinados indivíduos, eventos ou tópicos. De forma geral, a AS é um processo de classificação que pode ser dividido em três principais níveis: documento, sentença e aspecto. A partir de técnicas com abordagens baseada em léxico, aprendizagem de máquina e híbrida, pode-se identificar a intenção de um texto (se é uma opinião, reclamação, sugestão ou dúvida), a polaridade (positiva, negativa ou neutra), os sentimentos (alegria, tristeza, surpresa, raiva) e muitas outras informações [Medhat et al. 2014].

3. Revisão Bibliográfica

Nos últimos anos, inúmeros trabalhos foram desenvolvidos com foco na análise de sentimentos da Pandemia de COVID-19. Os autores de [Chen et al. 2020] descrevem uma base de dados multi-idioma sobre a COVID-19 no Twitter com a finalidade de disponibilizá-la para a comunidade acadêmica através do Github. Utilizando a API do Twitter e o Tweepy, coletaram 123 milhões de *tweets* com pelo menos 60% deles em inglês. A intenção é possibilitar outros estudos sobre a pandemia da COVID-19 em escala global, ajudando a entender os sentimentos das pessoas nesse contexto e combatendo a desinformação.

Em [Brum et al. 2020], os autores analisaram 56 milhões de *tweets* no contexto da pandemia da COVID-19 no período entre 23 de abril e 2 de julho de 2020, escritos em português. O trabalho visou caracterizar os *tweets* e os usuários extraíndo informações a respeito do volume e distribuição geográfica dos *tweets*, perfil dos usuários, além de realizar uma análise textual e de elementos do *tweet* como *hashtags*, menções e URLs. Essa caracterização pode ajudar outros trabalhos e entidades no combate à desinformação.

Em [Kouzy et al. 2020], uma pesquisa é conduzida para identificar a magnitude da desinformação e notícias falsas difundidas no Twitter no contexto da pandemia da COVID-19. O estudo coletou 673 *tweets*, em inglês, a partir de 14 diferentes *hashtags* e palavras-chave, resumiu e avaliou os *tweets* individualmente buscando indicativos de desinformação através da comparação com recursos previamente revisados por pares. Do total, 24,8% dos *tweets* continham desinformação e 17,4% incluíam informações não verificadas sobre o tema. Também observaram a relação de desinformação com o tipo do perfil (individual, verificado, entidade pública) e palavras-chave.

Recentemente, alguns trabalhos foram desenvolvidos com o foco em analisar os sentimentos relacionados a vacinação. Os pesquisadores de [Piedrahita-Valdés et al. 2021] coletaram e analisaram o sentimento de aproximadamente 1,5 milhão de *tweets* escritos em Inglês e Espanhol publicados entre 1 de junho de 2011 e 30 de abril de 2019 a respeito da hesitação sobre vacinas. O algoritmo implementado classificou 69,36% dos *tweets* como neutros, 21,78% como positivos e 8,86% como negativos. O estudo pode auxiliar governos a traçar estratégias de vacinação a partir do entendimento da opinião da população sobre vacinas.

Em [Farias and de Oliveira 2022], as autoras realizaram uma análise das polaridades demonstrados pela população brasileira quanto as vacinas para combater a COVID-19 através de métodos de mineração de dados e análise semântica com o uso do dicionário léxico para a língua portuguesa SentiLex-PT. Para isso, elas reuniram cerca de 5000 postagens públicas disponibilizadas pela API do Twitter e desenvolveram um algoritmo com a linguagem Python na plataforma Jupyter Notebook. O algoritmo criado apresentou acurácia de 68,5%, um valor adequado para tal modelo de pesquisa, segundo as autoras.

Considerando o contexto brasileiro, há estudos que analisam *tweets* sobre vacinas como [Farias and de Oliveira 2022] citado anteriormente e [Rodas et al. 2022] que analisou a polaridade da população brasileira quanto à vacina contra o COVID-19 em 70 mil *tweets*. Contudo, tais estudos contam com um número limitado de *posts* analisados ou não classificam os sentimentos dos usuários. Dessa forma, nosso trabalho apresenta uma análise mais abrangente considerando a polaridade, os sentimentos e os relacionamentos com palavras em aproximadamente 1 milhão de *tweets*. Assim, o trabalho fornece informações valiosas sobre a vacinação para grupos políticos e a população em geral, ajudando a ajustar campanhas de imunização e a compreender as opiniões sobre o assunto.

4. Metodologia

Esta pesquisa foi ordenada através dos seguintes passos: coletado de dados, tratamento de dados, classificação e análise. As próximas subsecções detalham cada um desses passos.

4.1. Coleta de dados

Para coleta dos *tweets*, foi utilizada a biblioteca Tweepy [Roesslein 2009] que tem como principal funcionalidade facilitar o acesso à API do Twitter através da linguagem de programação Python. Para obter o máximo de *tweets* relevantes para este estudo, foram selecionados os *tweets* em português que continham as palavras Brasil e pelo menos uma das seguintes palavras-chave: imunizado, imunizada, imunizar, vacinada, vacinado, vacina, vacinar, vacinação, pfizer, coronavac, astrazeneca (as três últimas são marcas de vacinas). Este trabalho buscou analisar os sentimentos dos brasileiros antes e durante a vacinação contra a COVID-19 no Brasil. Assim, consideramos os dados postados entre as datas 01/06/2020 e 11/06/2021, ou seja, aproximadamente seis meses antes da primeira vacina (17/01/2021) e seis meses depois, totalizando cerca de um milhão de *tweets*.

O produto dessa coleta de dados é um arquivo de aproximadamente 300 *Megabytes* com extensão CSV (*comma-separated values*) com as seguintes colunas: **created_at**, **date**, **time**, **username** e **tweet**. A primeira coluna auxilia apenas para confirmação da data, hora e fuso horário da criação de cada *tweet*. Por exemplo, o dado 2021-01-01 23:00:00 Hora Padrão de Greenwich denota que o *tweet* em questão foi criado no dia 01 de Janeiro de 2021 as 23 horas em ponto na Hora Padrão de Greenwich (GMT +0). Utilizando o padrão AAAA-MM-DD (ano, mês e dia), a coluna **date** representa a data em que o *tweet* foi postado. Com o formato 24h, a coluna **time** indica a hora, minutos e segundos que o *tweet* foi criado. A coluna **username** mostra o nome do usuário que criou a mensagem. Por último, a coluna **tweet** exprime o dado mais importante para essa pesquisa: o texto do usuário. A Figura 1 exemplifica o arquivo .CSV aberto no programa LibreOffice Calc com formatação em tabela.

1	created_at	date	time	username	tweet
2	2020-01-01 23:00:00 Hora Padrão de Greenwich	2020-01-01	23:00:00	minsauade	#RetrospectivaSaúde O BRASIL « SEGUE AVANÇANDO! 2019 foi um ano de grandes conquistas do @mir
3	2020-01-01 22:13:12 Hora Padrão de Greenwich	2020-01-01	22:13:12	k_aik_y	No Brasil é perigoso criança vacinada não poder voltar pra escola
4	2020-01-01 16:17:01 Hora Padrão de Greenwich	2020-01-01	16:17:01	lucable888	@greysanatombei @jairbolsonaro Mais uma farsante no mundo latino. É a maldição da Argentina este tipo de
5	2020-01-01 16:06:21 Hora Padrão de Greenwich	2020-01-01	16:06:21	chadegnomos	@whalebila Não é sofrência não HAHAHHA Eu postei pq tem, além da Ana em Madri, outros amigos em pai
6	2020-01-01 16:01:10 Hora Padrão de Greenwich	2020-01-01	16:01:10	marlonjcm031	@bazookaweb @joomikhail Ent a mamadeira de que racismo no Brasil n existe, q vacina faz mal e qu
7	2020-01-01 14:18:31 Hora Padrão de Greenwich	2020-01-01	14:18:31	anaf_garcia	A cobertura de vacinas no Brasil é bem MELHOR do que em Portugal. Aqui, criança não toma vacina contra c
8	2020-01-01 13:51:38 Hora Padrão de Greenwich	2020-01-01	13:51:38	rahmandes9	@Gessiel7 @cirogomes Ah mas foi a turminha dele q falava sobre o movimento anti vacinação. O sarampo tá
9	2020-01-01 11:58:37 Hora Padrão de Greenwich	2020-01-01	11:58:37	jaugustosrj	Que 2020 nos traga a certeza de que a Terra é redonda, que vacina funciona, e um Ministro da Educação que
10	2020-01-01 04:37:41 Hora Padrão de Greenwich	2020-01-01	04:37:41	leonardo08_ss	@dsathrin "Não vacinei meu filho e ela tá ótimo" 2019: volta do sarampo no Brasil O brasileiro vê a vacina e

Figura 1. Amostra de 10 linhas do arquivo .CSV contendo o produto da coleta de dados

4.2. Tratamento dos dados

Através da linguagem R em sua versão 4.1.2 e da IDE (*integrated development environment*, em português Ambiente Integrado de Desenvolvimento) RStudio versão 2021.09.2 Build 382, esta pesquisa tratou, classificou e analisou os dados obtidos na fase da coleta. Para tais fases de pesquisa, foram utilizadas as bibliotecas R *stringr*, *dplyr*, *stringi*, *textclean*, *syuzhet*, *ggplot2*, *tm*, *readr* e *reshape2*. Para realizar o pré-processamento dos dados brutos, primeiro, importou-se o arquivo CSV no RStudio através da função `read_csv`. Em seguida processou-se a coluna **tweet** para remover tudo o que não mostrou-se interessante para esta análise: *stopwords*, URLs, menções, acentos, emojis/emoticons, *hashtags*, pontuações, espaços em branco extras, além de transformar todas as letras em minúsculas. Logo após, os dados foram separados por mês e ano. Após essa separação, os dados estão prontos para a etapa de classificação. Apenas para a geração de nuvens de palavras, também removemos os termos que definiram a busca na subseção 4.1, pois eles são os termos com maior frequência nos dados e isso atrapalharia a análise.

4.3. Classificação

Com a função `get_nrc_sentiment()` da biblioteca `syuzhet`, analisamos cada um dos conjuntos de *tweets* separados por mês na etapa anterior. Essa função foi escolhida porque, de acordo com [Misuraca et al. 2020], a biblioteca possui uma das melhores acurácias em predição dos sentimentos. Esta função se baseia no NRC *Emotion lexicon* [Mohammad and Turney 2013]. Seguindo a teoria das emoções de Plutchik [Plutchik 1994], os autores destacam: “o NRC emotion lexicon consiste numa lista de palavras e suas respectivas associações com oito emoções (raiva, medo, antecipação, confiança, surpresa, tristeza, alegria e desgosto) e dois sentimentos de polaridade (negativo e positivo)”. A função citada retorna um *dataframe* no qual cada linha representa uma sentença do arquivo original. As colunas incluem uma para cada emoção, bem como a valência da polaridade positiva ou negativa. Levando em conta a definição de antecipação e surpresa segundo Plutchik, essas emoções mostraram pouca relevância relacionada ao tema. Por isso, para tornar a análise mais simples e objetiva, esta pesquisa considerou as emoções raiva, nojo (desgosto), medo, alegria, tristeza e confiança, além das polaridades supracitadas. Em um mesmo texto, a função pode atribuir valores iguais para as polaridades positiva e negativa. Por isso, adicionamos a polaridade neutra ao detectar tal situação a fim de eliminar essa ambiguidade e ter uma nova perspectiva sobre os dados.

4.4. Análise

Tendo em mente que o algoritmo utilizado pode atribuir mais de um sentimento e polaridade a um mesmo *tweet*, para facilitar a análise dos dados, foram gerados 2 gráficos de linha: um contendo a polaridade (positiva, negativa e neutra) e outro mostrando os sentimentos identificados ao longo do período analisado. Também geramos um gráfico de barras com a quantidade de *tweets* coletados em cada mês. Com esses gráficos, observamos três períodos bem definidos de divergência entre os demais e foi decidido que tais períodos seriam o foco de estudo. Esses subintervalos foram Outubro de 2020, Janeiro de 2021 e o período de Fevereiro a Maio de 2021. Também selecionamos os períodos de um mês antes e um mês depois da primeira vacina contra a Covid-19 no Brasil para gerar nuvens de palavras. Uma nuvem de palavras exibe os termos mais frequentes de um texto, variando o tamanho da palavra conforme sua frequência. As nuvens de palavras foram produzidas com a biblioteca R *wordcloud*. Com os gráficos e as nuvens gerados, foram realizadas pesquisas utilizando os dados obtidos a partir dos períodos escolhidos e as relacionamos com notícias e acontecimentos reportados pela mídia. Deste modo, procurou-se gerar uma associação entre as notícias e o movimento dos valores das emoções.

5. Resultados

A Figura 2 exibe, por mês, o total de *tweets* relacionados ao tema, recuperados na etapa de busca. Percebe-se um aumento expressivo de 358% entre Novembro e Dezembro de 2020, seguido de mais 67% em Janeiro de 2021 (mês que deu-se início à vacinação contra a Covid-19 no Brasil). Posteriormente, houve uma queda de 65% em Fevereiro de 2021, para logo depois continuar aumentando nos meses seguintes, atingindo um pico de 195 mil em Maio de 2021, até sofrer uma leve queda de 17% em Junho de 2021.

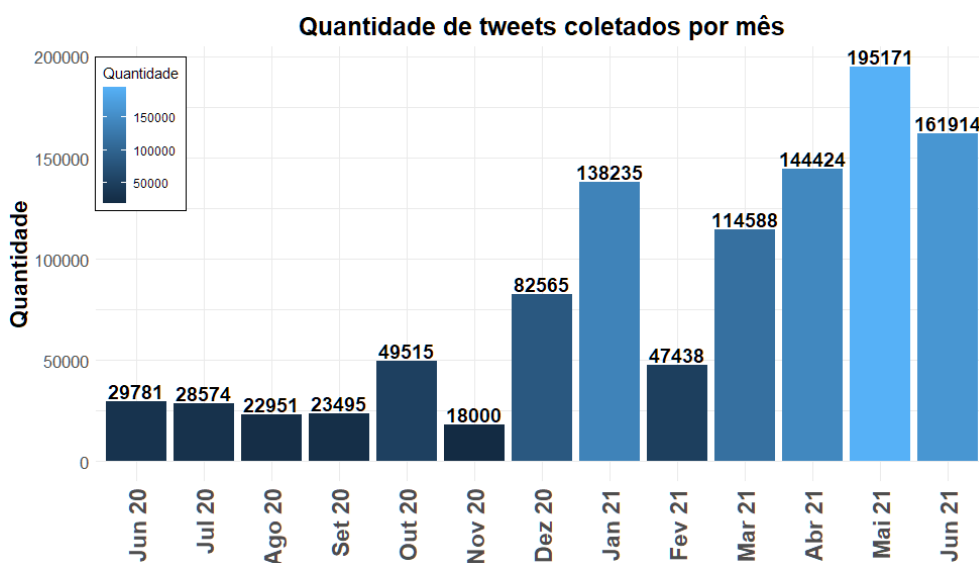


Figura 2. Gráfico de barras que exibe o total de *tweets* coletados por mês

A Figura 3 exibe o gráfico de polaridade de *tweets* relacionados à vacina no período de Junho de 2020 a Junho de 2021. De maneira geral, no período analisado, não foram observadas variações significativas na polaridade e nos sentimentos. A maioria dos dados foram classificados como positivos. Entretanto, observou-se uma crescente no número do total de *tweets*, indicando assim que o tema foi abordado por mais usuários e com mais ocorrências. Nota-se também, um pico em Outubro de 2020; pode-se correlacionar ao fato de que, no início desse mês, o Governo Federal lançou a Campanha Nacional de Vacinação contra a Poliomielite e de Multivacinação [Verdélío 2020], o que não está diretamente ligado com esse estudo mas cita o termo vacinação que está incluído na pesquisa de *tweets*. Também é possível relacionar ao crescente número de mortes por Covid-19, chegando aos 159 mil ao final do mês de Outubro de 2020 [G1 2020].

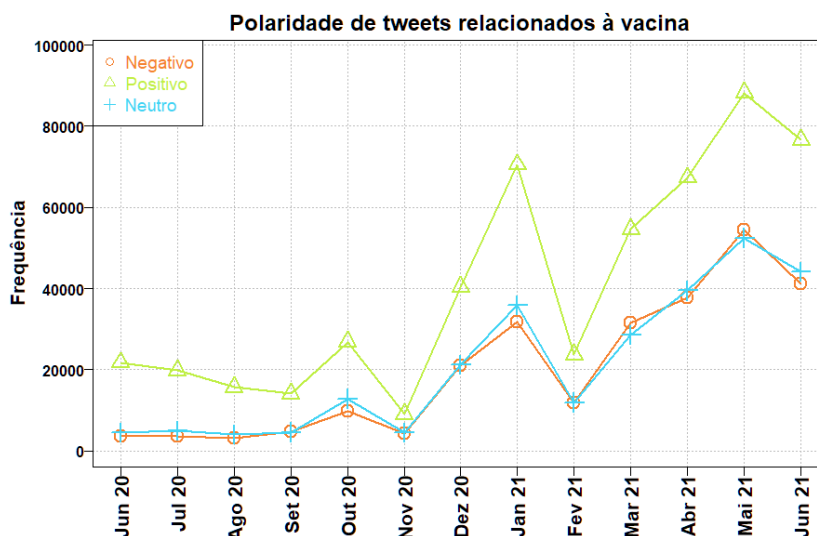


Figura 3. Gráfico de linha mostrando o avanço da polaridade sobre vacinas entre Junho de 2020 e Junho de 2021

Em Dezembro de 2020, o assunto voltou a ser bastante comentado na plataforma. No dia 16 de Dezembro de 2020, o Governo Federal anunciou o Plano Nacional de vacinação contra a Covid-19 [Peduzzi and Vilela 2020]. Este plano abrangia grupos prioritários, como profissionais da área de saúde, idosos, indígenas, pessoas com doenças graves e professores. O plano também estabeleceu acordo com três fabricantes de vacinas: Fiocruz/AstraZeneca, Covax Facility e Pfizer. Neste mesmo mês, as mortes pela doença cresceram cerca de 65% em relação ao mês anterior no Brasil [Rosa and Venaglia 2021].

Além disso, em Janeiro de 2021, iniciou-se a vacinação contra a Covid-19 no Brasil. Novamente, houve um aumento no número de *posts* sobre o tema. Entretanto, em Fevereiro de 2021, notou-se uma queda expressiva no número de *tweets*. É difícil definir a causa exata da queda, contudo, pode-se citar o Carnaval que apesar da falta de festividades de rua que foram restritas pelo governo, o período ainda foi bastante apreciado e comentado pelos brasileiros. Similarmente, pode-se citar o caso Flordelis que voltou a ser bastante comentado nesse mês. Por fim, no meses de Março de 2021 a Junho de 2021, o total de *tweets* voltou a crescer, mostrando o maior número em Maio de 2021. Isso pode ser consequência do fato de que, neste mês, a Covid-19 foi a maior causa de mortes no Brasil considerando a média mensal nos 5 anos anteriores [Mali and Mões 2021].

Ao todo, cerca de 50% dos dados processados apresentam polaridade positiva, indicando assim, opiniões positivas dos usuários acerca da vacinação. Em segundo lugar, com aproximadamente 25,4% do total, encontramos a polaridade neutra, perdendo para a polaridade negativa apenas nos meses de Março de 2021 e Maio de 2021. Em último lugar, encontra-se a polaridade negativa com 24,5%, refletindo que menos usuários expressaram opiniões negativas referentes à vacinação. Isso mostra que boa parte da população brasileira pensa que a vacina é algo positivo. Uma parcela menor da população, no entanto, julgam a vacinação como algo prejudicial.

A Figura 4 expõe o gráfico de linha com os sentimentos relacionados à vacina entre Junho de 2020 e Junho de 2021. Em todo o período, o sentimento de Confiança foi identificado em 66,1% dos *tweets*. De acordo com a Teoria dos Sentimentos de Plutchik [Plutchik 1994], esse sentimento nos diz que algo é seguro e benéfico. Isso indica que os usuários se sentem seguros sobre o que postaram, no sentido de que a vacina é algo confiável e que trará benefícios. Em segundo e terceiro lugar, encontrou-se os sentimentos de Medo (55,6%) e Tristeza (53,7%) alternando de lugar no período com pouquíssima diferença. Em outras palavras, isso mostra que os usuários se sentiram preocupados e infelizes ao discutir sobre as vacinas no Twitter, pois segundo Plutchik, o medo é a emoção que nos diz que algo é perigoso (nos deixando em alerta e preocupados) e a tristeza está relacionada com sensações de desamparo, perda e decepção. Ao se aprofundar na Teoria das Emoções supracitada, Plutchik sugere que a combinação de Confiança e Medo resulta no sentimento de Submissão que por sua vez pode ser encontrada em situações em que a pessoa se sente ameaçada ou pressionada a se submeter a uma figura de autoridade ou a um sistema. Dessa forma, os usuários demonstraram que se sentem ameaçados ou pressionados ao debater sobre as vacinas. Em seguida, verificou-se Raiva em 40,1% e Nojo (desgosto, aversão) em 29,2% dos *posts*. Por último, encontramos a emoção Alegria em 22,7% dos dados. Note que o algoritmo utilizado pode atribuir mais de um sentimento para um mesmo texto. Nesta análise, observou-se a linearidade na distribuição dos sentimentos, ou seja, não houve alterações significativas das posições no período analisado.

estejam afetando os sentimentos expressos no Twitter. Por isso, como trabalho futuro, avaliaremos outras fontes de dados para avaliar a opinião pública sobre a vacinação.

Referências

- Ahmed, W., Bath, P. A., and Demartini, G. (2017). Using twitter as a data source: An overview of ethical, legal, and methodological challenges. *The ethics of online research*, 2:79–107.
- Barifouse, R. (2020). Fake news sobre vacinas contra a covid-19 ameaçam combate à doença. <https://www.bbc.com/portuguese/geral-53795050>.
- Berry, M. W. and Kogan, J. (2010). *Text mining: applications and theory*. John Wiley & Sons.
- Brum, P. V., Teixeira, M. C., Miranda, R., Vimieiro, R., Meira Jr, W., and Pappa, G. L. (2020). A characterization of portuguese tweets regarding the covid-19 pandemic. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 177–184. SBC.
- Burgos, R. M., Badowski, M. E., Drwiega, E., Ghassemi, S., Griffith, N., Herald, F., Johnson, M., Smith, R. O., and Michienzi, S. M. (2021). The race to a covid-19 vaccine: Opportunities and challenges in development and distribution. *Drugs in context*, 10.
- Carvalho, T., Krammer, F., and Iwasaki, A. (2021). The first 12 months of covid-19: a timeline of immunological insights. *Nature Reviews Immunology*, 21(4):245–256.
- Chen, E., Lerman, K., and Ferrara, E. (2020). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273.
- Clement, J. (2020). Twitter-statistics & facts. *Statista*. Retrieved May, 2:2020.
- Farias, F. L. and de Oliveira, L. S. C. (2022). Mineração de textos e análise de sentimentos aplicados a postagens do twitter acerca das vacinas contra a covid-19. *Research, Society and Development*, 11(13):e364111335490–e364111335490.
- G1 (2020). Brasil soma mais de 159,5 mil mortos por covid; casos confirmados têm 4º dia de alta e somam 5,5 milhões. <https://g1.globo.com/bemestar/coronavirus/noticia/2020/10/30/casos-e-mortes-por-coronavirus-no-brasil-em-30-de-outubro-segundo-consorcio-de-veiculos-de-imprensa.ghtml>.
- Hearst, M. (2003). What is text mining. *SIMS, UC Berkeley*, 5.
- Jaguraba, M. (2018). Vacinação pni do brasil: um dos melhores do mundo. <https://www.vaticannews.va/pt/mundo/news/2018-07/vacinacao-pni-do-brasil-um-dos-melhores-do-mundo.html>.
- Karami, A., Lundy, M., Webb, F., and Dwivedi, Y. K. (2020). Twitter and research: A systematic literature review through text mining. *IEEE Access*, 8:67698–67717.
- Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E. W., and Baddour, K. (2020). Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. *Cureus*, 12(3).

- Maciel, E., Fernandez, M., Calife, K., Garrett, D., Domingues, C., Kerr, L., and Dalcolmo, M. (2022). A campanha de vacinação contra o sars-cov-2 no brasil e a invisibilidade das evidências científicas. *Ciência & Saúde Coletiva*, 27:951–956.
- Mali, T. and Mões, M. (2021). Covid É a maior causa de morte no brasil em maio de 2021; nos eua É a 5ª. <https://www.poder360.com.br/brasil/covid-e-a-maior-causa-de-morte-no-brasil-em-maio-de-2021-nos-eua-e-a-5a/>.
- Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., Giattino, C., and Rodés-Guirao, L. (2021). A global database of covid-19 vaccinations. *Nature human behaviour*, 5(7):947–953.
- Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113.
- Misuraca, M., Forciniti, A., Scepi, G., and Spano, M. (2020). Sentiment analysis for education with r: Packages, methods and practical applications. arxiv 2020. *arXiv preprint arXiv:2005.12840*.
- Mohammad, S. M. and Turney, P. D. (2013). Nrc emotion lexicon. *National Research Council, Canada*, 2:234.
- Peduzzi, P. and Vilela, P. (2020). Governo anuncia hoje plano nacional de vacinação contra covid-19. <https://agenciabrasil.ebc.com.br/saude/noticia/2020-12/governo-anuncia-hoje-plano-nacional-da-vacina-contracovid-19>.
- Piedrahita-Valdés, H., Piedrahita-Castillo, D., Bermejo-Higuera, J., Guillem-Saiz, P., Bermejo-Higuera, J. R., Guillem-Saiz, J., Sicilia-Montalvo, J. A., and Machío-Regidor, F. (2021). Vaccine hesitancy on social media: Sentiment analysis from june 2011 to april 2019. *Vaccines*, 9(1):28.
- Plutchik, R. (1994). *The psychology and biology of emotion*. HarperCollins College Publishers.
- Rodas, C. M., Barros, S. E. T., Souza, R. A. d. S., and Vidotti, S. A. B. G. (2022). Análise de sentimentos sobre as vacinas contra covid-19: um estudo com algoritmo de machine learning em postagens no twitter.
- Roesslein, J. (2009). tweepy documentation. *Online*] <http://tweepy.readthedocs.io/en/v3,5>.
- Rosa, A. and Venaglia, G. (2021). Mortes por covid-19 cresceram 65% em dezembro no brasil. <https://www.cnnbrasil.com.br/saude/mortes-por-covid-19-cresceram-65-em-dezembro-no-brasil/>.
- Troiano, G. and Nardi, A. (2021). Vaccine hesitancy in the era of covid-19. *Public health*, 194:245–251.
- Verdélío, A. (2020). Governo lança campanha nacional de multivacinação. <https://agenciabrasil.ebc.com.br/saude/noticia/2020-10/governo-lanca-campanha-nacional-de-multivacinacao>.
- WHO (2020). Pneumonia of unknown cause – china. <https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229>.