

Análise do posicionamento dos usuários do Twitter acerca da Vacinação Infantil contra a COVID-19 no Brasil

Douglas Vitória¹, Hidelberg O. Albuquerque^{1,2}, Ellen Souza^{2,3},
Adriano L. I. Oliveira¹, Flávia Barros¹, Ricardo B. C. Prudêncio¹

¹Centro de Informática (CIn)
Universidade Federal de Pernambuco (UFPE) – Recife – PE – Brasil

²Unidade Acadêmica de Serra Talhada (UAST)
Universidade Federal Rural de Pernambuco (UFRPE) – Serra Talhada – PE – Brasil

³Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo (USP) – São Carlos – SP – Brasil

{damsv, hoa, alio, fab, rbcpr}@cin.ufpe.br, ellen.ramos@ufrpe.br

Abstract. *COVID-19 vaccination has become a hotly debated subject on social media since its start. However, childhood vaccination only started about a year after adult vaccination, which ended up postponing this specific debate until the end of 2021. This work, therefore, aims to analyze the position, whether favorable or contrary, of Brazilian Twitter users regarding the vaccination of children aged 5 to 11 years. Using Sentiment Analysis techniques, we noticed that most users were in favor of starting vaccination in the country. In addition, some factors that led users to stand in that way were pointed out.*

Resumo. *Com o início da vacinação contra a COVID-19, esta se tornou um assunto bastante debatido nas redes sociais. Porém, a vacinação infantil só veio a ter início cerca de um ano após a vacinação de adultos, o que acabou adiando esse debate específico para o fim de 2021. Este trabalho, portanto, visa analisar o posicionamento, se favorável ou contrário, dos usuários do Twitter no Brasil acerca da aplicação das vacinas nas crianças de 5 a 11 anos. Utilizando técnicas de Análise de Sentimentos, pôde-se perceber que a maior parte dos usuários se mostrou favorável ao início da vacinação no país. Também foram levantados pontos que levaram os usuários a se posicionar daquela forma.*

1. Introdução

Desde o início da pandemia de COVID-19, em março de 2020, as redes sociais se transformaram em uma grande fonte de informação acerca do Coronavírus, bem como passaram a receber uma enorme quantidade de opiniões sobre diversos aspectos relacionados a esse momento. Os assuntos abordados pelos usuários no Twitter, por exemplo, permearam desde tópicos diretamente ligados à COVID-19, como o aumento no número de casos/mortes e medidas de proteção e tratamento, até tópicos como política, esportes, economia e entretenimento, indiretamente afetados pela pandemia [Garcia and Berton 2021].

Um dos focos das discussões nas redes sociais sempre foi a vacinação contra a doença, pois, mesmo quando ainda não existiam vacinas aprovadas pelas autoridades

sanitárias, já havia a expectativa pelo início da imunização e pelo possível fim da pandemia. Esse debate intensificou-se ainda mais em dezembro de 2020, quando as primeiras pessoas foram imunizadas [Malagoli et al. 2021]. Porém, ao passo que a vacinação em adultos seguia avançando durante o ano de 2021, a imunização de crianças de 5 a 11 anos só teve início em alguns países no final desse ano e, em outros, nos primeiros meses de 2022, como é o caso do Brasil, cuja primeira dose de vacina infantil foi aplicada no dia 14 de janeiro¹. Sendo assim, é possível dizer que o debate específico sobre a vacinação infantil foi adiado, só adquirindo um foco maior conforme as vacinas infantis iam sendo aprovadas e o início da imunização se aproximava.

O presente trabalho visa analisar o posicionamento dos usuários do Twitter acerca da vacinação infantil contra a COVID-19 no Brasil. Inicialmente, foram coletados tweets publicados durante a semana de início da imunização no país. Os tweets coletados foram classificados utilizando técnicas de Análise de Sentimentos e posteriormente analisados com o objetivo de verificar o sentimento geral desse público e os pontos levantados nesse debate. Além disso, outra contribuição deste trabalho é a disponibilização do conjunto de dados construído, para pesquisas futuras.

O restante deste artigo está organizado da seguinte forma: a Seção 2 traz trabalhos relacionados ao tópico; o método utilizado é apresentado na Seção 3; as análises encontram-se na Seção 4; por fim, a Seção 5 apresenta as conclusões do trabalho.

2. Trabalhos Relacionados

Desenvolvendo um trabalho similar ao nosso, [Malagoli et al. 2021] realizaram uma análise do debate acerca da vacinação contra a COVID-19 no Twitter brasileiro, porém considerando o cenário no início da imunização no país. Eles investigaram o sentimento das pessoas quando comentavam sobre a vacinação, sem necessariamente identificar seu posicionamento, se a favor ou contra, quanto a ela. Os autores chegaram à conclusão de que muitos sentimentos negativos expressos pelos usuários eram motivados tanto pelo movimento antivacina quanto pela ansiedade pela demora no início da vacinação. Além do sentimento dos usuários, os autores analisaram fatores externos que impactaram nesse sentimento, bem como sua evolução ao longo do tempo.

Com um objetivo semelhante, [de Sousa and Becker 2021] utilizaram técnicas de modelagem de tópicos para levantar os pontos relacionados aos posicionamentos de usuários favoráveis e contrários à vacinação contra a COVID-19 no Brasil. Por meio de uma análise temporal, eles analisaram a evolução nos argumentos dos dois grupos no Twitter, notando que as manifestações antivacina se deram apenas no ano de 2020 e tinham como foco a obrigatoriedade da vacina e a desconfiança acerca dela. Já manifestações pró-vacina envolveram gratidão ao início da vacinação e críticas ao Governo.

Por fim, mostrando a diversidade de assuntos tratados pelos usuários do Twitter durante a pandemia, [Leite et al. 2021] e [Silva et al. 2021] realizaram, de maneira semelhante ao trabalho aqui apresentado, Análise de Sentimentos considerando, respectivamente, a adoção do ensino remoto pelas instituições brasileiras e as aplicações da medida de *lockdown* no Brasil. Os autores também utilizaram Aprendizagem de Máquina

¹<https://www.cnnbrasil.com.br/saude/crianca-indigena-de-sp-e-primeira-a-ser-vacinada-no-brasil-contra-a-covid-19/>

(AM) para classificar os conjuntos de tweets e analisaram diversos fatores que levaram os usuários a se posicionarem principalmente de maneira negativa ao ensino remoto e de maneira positiva à adoção de *lockdown*.

3. Construção e Avaliação do Classificador

Nesta seção, são apresentadas as etapas utilizadas para a construção do classificador de sentimentos, desde a anotação da base de treinamento até a avaliação do melhor classificador de Aprendizagem de Máquina (AM) a ser utilizado.

3.1. Base de dados

Foram coletados 10.251 tweets publicados entre os dias 11/01/2022 e 17/01/2022, correspondendo aos três dias anteriores e aos três dias posteriores à imunização da primeira criança brasileira, que ocorreu no dia 14/01. Uma aplicação Python foi desenvolvida para realizar a coleta dos tweets, na qual utilizou-se a biblioteca Tweepy² para acessar a Stream API do Twitter. Dois pares de termos foram utilizados para filtrar apenas tweets escritos em Português e que se referissem à vacinação infantil: (*vacina; vacinação*), e (*infantil; criança*). Assim, cada tweet coletado possuía pelo menos um dos termos de cada par.

Como foram utilizadas técnicas de Aprendizagem Supervisionada para classificação automática dos tweets, um conjunto de treinamento necessitou ser construído. Essa construção se deu pela anotação manual, realizada por dois anotadores, de 2.350 tweets em uma dentre três classes: A FAVOR, quando o tweet apresentava um posicionamento favorável à vacinação infantil; CONTRA, quando o usuário se posicionou contrário à aplicação da vacina em crianças; e NEUTRO, quando o tweet apresentava um fato/notícia, ou não apresentava uma opinião diretamente ligada à vacinação infantil.

O nível de concordância entre os dois anotadores permaneceu em torno de 91,3% e, nos casos em que eles discordaram quanto à classe do tweet, eles se reuniram para entrar em consenso. A Tabela 1 apresenta a distribuição dos tweets anotados por classe. Todo o conjunto de tweets coletados e a base anotada manualmente estão disponíveis³.

Table 1. Distribuição das classes na base de dados anotada.

Classe	Quantidade
A FAVOR	712
CONTRA	709
NEUTRO	929

3.2. Avaliação dos Classificadores

Para a classificação dos 7.901 tweets restantes, foi desenvolvida uma aplicação Python, utilizando as bibliotecas NLTK⁴ e scikit-learn⁵.

Previamente, três algoritmos de AM foram avaliados para a construção do classificador de sentimentos: *Multinomial Naïve Bayes* (MNB), *Support Vector Machine* (SVM)

²<https://www.tweepy.org>

³<https://bit.ly/3ywYVaV>

⁴<https://www.nltk.org>

⁵<https://scikit-learn.org/stable/>

e *Random Forest* (RF). Essa escolha se deu pelo fato de que estes são os três algoritmos mais utilizados para classificação de texto em Português [Souza et al. 2018]. Além disso, seis combinações de técnicas de pré-processamento também foram avaliadas: 1) unigram; 2) bigram; 3) unigram + bigram; 4) unigram + remoção de *stopwords*; 5) bigram + remoção de *stopwords*; 6) unigram + bigram + remoção de *stopwords*. Tendo a técnica TF-IDF sido utilizada para gerar o Modelo de Espaço Vetorial (VSM).

Os classificadores foram avaliados utilizando a técnica de validação-cruzada 10-fold, repetida 10 vezes. Os valores médios de acurácia e de f-measure obtidos em cada cenário podem ser visualizados na Tabela 2.

Table 2. Resultados da avaliação das técnicas a serem utilizadas.

Pré-processamento	Acurácia			F-measure		
	MNB	SVM	RF	MNB	SVM	RF
Unigram	0,7270	0,7328	0,6789	0,7289	0,7310	0,6723
Bigram	0,6627	0,6959	0,6723	0,6621	0,6906	0,6723
Unigram + bigram	0,7186	0,7265	0,6941	0,7161	0,7242	0,6859
Unigram + <i>stopwords</i>	0,7177	0,7173	0,6634	0,7204	0,7157	0,6617
Bigram + <i>stopwords</i>	0,5864	0,6204	0,6097	0,5885	0,6268	0,6034
Unigram + bigram + <i>stopwords</i>	0,6964	0,7029	0,6636	0,6978	0,7055	0,6502

Pelos resultados, pôde-se observar que o SVM obteve o melhor desempenho em quase todos os cenários, e, sendo assim, optou-se pela sua utilização na classificação dos demais tweets. Dentre as técnicas de pré-processamento, o uso apenas do modelo de unigram alcançou os melhores resultados, sendo o escolhido.

4. Análises

Utilizando o algoritmo SVM treinado com os 2.350 tweets anotados manualmente, pôde-se classificar os demais tweets que não haviam sido anotados. Após a classificação, foi obtida a seguinte distribuição de classes para os 10.251 tweets coletados:

- 4.203 apresentaram uma opinião A FAVOR da vacinação infantil;
- 2.154 apresentação uma opinião CONTRA a vacinação infantil;
- 3.894 foram classificados como NEUTRO.

Ignorando os tweets da classe NEUTRO, os quais correspondiam a notícias, fatos ou opiniões acerca de outros assuntos, a Figura 1 traz a proporção geral de tweets A FAVOR e CONTRA, mostrando que dois terços dos usuários que emitiram opinião acerca da vacinação infantil no período analisado se posicionaram a favor dela.

Outra análise pode ser feita com base na data de publicação dos tweets. A Figura 2 apresenta a distribuição entre as duas classes durante os sete dias da coleta. Com base nela, é possível concluir que a proporção de opiniões favoráveis foi maior em todos os momentos, porém essa proporção cresceu ainda mais depois do início oficial da vacinação infantil no Brasil (dia 14/01): até o dia 13/01, os tweets A FAVOR representavam cerca de 60% dos tweets diários, já a partir de 14/01, passaram a representar de 70% a 80%. Analisando os tweets, pôde-se perceber que isso se deveu ao fato de que, com o começo da imunização, mais pessoas foram ao Twitter se posicionar a favor da vacinação infantil.

Por fim, por meio das nuvens de palavras construídas a partir das duas classes de tweets opinativos, é possível analisar alguns pontos que levaram os usuários a terem

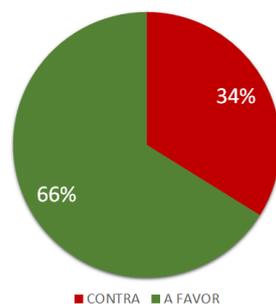


Figure 1. Gráfico com a porcentagem de tweets A FAVOR e CONTRA.

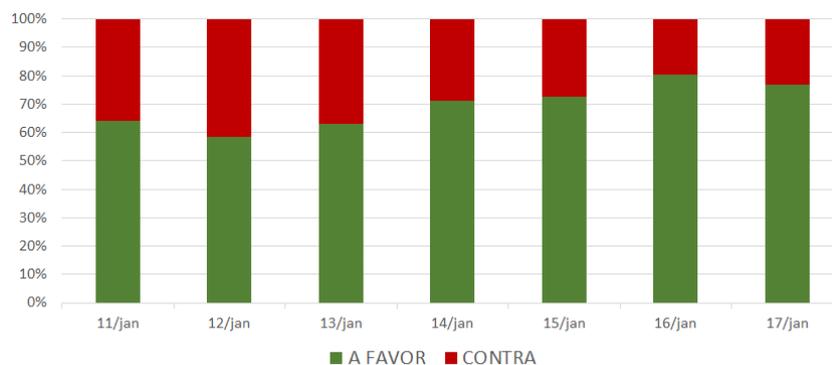


Figure 2. Distribuição do sentimento dos usuários ao longo do tempo.

opiniões favoráveis ou contrárias à vacinação infantil. Para a construção das nuvens de palavras deste estudo, as *stopwords* e os termos utilizados como filtro para a coleta dos tweets foram removidos, já que sua frequência é alta e não ajudariam na análise.

A nuvem A FAVOR (Figura 3A) mostra que, em muitos casos, a publicação de tweets favoráveis à vacinação infantil também conteve críticas ao governo brasileiro, como é possível notar pela presença de palavras como *presidente*, *bolsonaro* e *governo*, bem como às pessoas contrárias à vacina, denominadas *negacionistas*. Além disso, pôde-se observar um chamado à vacinação pela presença do imperativo *vamo*.

Já a nuvem CONTRA (Figura 3B) indica que as críticas e a resistência à vacinação infantil no Brasil se deram por dois fatores principais: 1) a expectativa de que os pais tivessem a obrigação legal de vacinar seus filhos (possível de ser notada pelos termos *obrigatória* e *obrigatoriedade*); e 2) o risco e possíveis *efeitos colaterais* da sua aplicação nas crianças, já que muitos usuários consideraram essa vacina infantil uma vacina ainda *experimental*. Ambas as análises das nuvens de palavras corroboram os achados de [de Sousa and Becker 2021], no seu trabalho de modelagem de tópicos para identificar os fatores das manifestações a favor e contra a vacinação de maneira geral.

5. Conclusão

Neste trabalho, foi construída uma aplicação de Análise de Sentimentos para classificar automaticamente tweets com o objetivo de analisar o posicionamento dos usuários do Twitter acerca do início vacinação infantil contra a COVID-19 no Brasil. Uma base de dados anotada também foi construída e disponibilizada.

