



# A Survey on Sentiment Analysis About COVID-19 Vaccines in Social Media

J. Jane Jenolin, M. Pramila, G. Radhika

1. Assistant Professor, Annai Vailankanni College of Engineering, Kanyakumari-629 401

## Abstract

Sentiment analysis, often referred to as "opinion mining," is a natural language processing (NLP) technique that determines the emotional undertone of a piece of writing. In March 2020, the World Health Organization (WHO) declared COVID-19 a global pandemic. As of late 2022, COVID-19 continues to be a global health issue. The introduction of COVID-19 vaccines has been crucial in combating the virus. According to the WHO, 50 vaccines have been approved worldwide by at least one country, with 11 receiving Emergency Use Listing (EUL). These include non-replicated viral vectors, inactivated vaccines, protein subunit vaccines, and RNA vaccines. This article reviews several studies that have conducted sentiment analysis on public opinions towards COVID-19 vaccines on social media platforms, particularly Twitter. Tweets reflect a range of emotions including eagerness, dissatisfaction, reservations, and apprehension regarding COVID-19 vaccines. Public opinion on COVID-19 vaccinations has varied significantly over time and across different regions. Research indicates that approximately 83% of people in the Philippines have had positive and enthusiastic sentiments towards vaccination, while negative sentiments were more common in Korea. In India, 78.5% of tweets about the COVID-19 vaccine were neutral or positive. Real-time sentiment analysis can help public health authorities develop localized vaccination education initiatives to address concerns about COVID-19 vaccines.

**Keywords:** Sentiment Analysis, COVID-19 Vaccine, Social Media, NLP, Twitter, Topic Modeling, LDA

## 1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), commonly known as COVID-19 or the new coronavirus, is a highly infectious respiratory infection. Discovered in December 2019 in Wuhan, China, it has since spread globally, affecting millions of people. The virus is primarily transmitted through the coughs and sneezes of infected individuals, but it can also be contracted by touching contaminated surfaces and then touching the eyes, nose, or mouth. In March 2020, the World Health Organization (WHO) declared COVID-19 a global pandemic, and it remains an ongoing issue as of late 2022.

Vaccines have been developed and distributed globally to prevent the spread of COVID-19. These vaccines represent a critical tool in the fight against the virus. As of the latest updates, 50 vaccines have been approved worldwide, with 11 receiving Emergency Use Listing (EUL) from the WHO. These include various types such as:



a) **mRNA Vaccines:** Examples include the Pfizer-BioNTech and Moderna vaccines, which use mRNA to instruct cells to produce a protein found on the virus's surface, triggering an immune response.

b) **Vector Vaccines:** Examples include the AstraZeneca and Johnson & Johnson vaccines, which use a weakened virus to deliver genetic material from SARS-CoV-2 into cells, stimulating an immune response.

c) **Protein Subunit Vaccines:** Examples include the Sinovac and Sinopharm vaccines, which use pieces of the virus's protein to provoke an immune response.

d) **Virus-like Particle Vaccines:** Bharat Biotech's COVAXIN is an example, using viral proteins to mimic the virus and induce an immune response.

Each of these vaccines has been shown to reduce the severity of COVID-19 infections and the risk of hospitalization or death. As of January 2023, approximately 5 billion people worldwide have been fully vaccinated against COVID-19.

Understanding public sentiment towards COVID-19 vaccinations is crucial. This can be achieved through social media sentiment analysis, which reflects emotions such as eagerness, dissatisfaction, reservations, and apprehension regarding the vaccines. Sentiment analysis is a systematic method of identifying, extracting, quantifying, and studying emotional states and subjective data using natural language processing, text analysis, and computational linguistics. It helps determine the emotional tone behind words and analyze opinions, attitudes, and emotions expressed in text, such as social media posts, customer reviews, surveys, and news articles. Social media sentiment analysis specifically focuses on analyzing sentiment expressed on platforms like Twitter, Facebook, and Instagram. The insights gained can assist in understanding public opinions, monitoring brand reputation, and evaluating the impact of various campaigns.

## **2. Techniques for Sentiment Analysis and Topic Modeling**

Sentiment Analysis and Topic Modeling are two essential techniques used in Natural Language Processing (NLP).

### **2.1 Sentiment Analysis**

Sentiment Analysis aims to determine the sentiment expressed in a text, typically categorized as positive, negative, or neutral. Several techniques are used for sentiment analysis:

- **Rule-Based Approaches:** These methods rely on a set of predefined rules or dictionaries to identify sentiment. Techniques include keyword matching, synonym matching, negation handling, and sentiment scoring. Each word in the text is assigned a sentiment score, which is then aggregated to determine the overall sentiment. While straightforward, this approach may struggle with complex or figurative expressions.



- **Machine Learning-Based Approaches:** These methods involve training machine learning algorithms, such as Naive Bayes, Support Vector Machines (SVM), and Deep Learning models, on large annotated datasets with known sentiments. The algorithms learn patterns and features to predict sentiment in new texts. Although effective for handling complex expressions, these approaches require substantial annotated data and may be influenced by the quality of the training data.
- **Hybrid Approaches:** Combining rule-based and machine learning methods, hybrid approaches enhance sentiment analysis precision. Rules are used for initial text preprocessing and sentiment identification, followed by machine learning techniques to refine predictions. This approach benefits from both domain knowledge and advanced pattern recognition.

### 2.1.1 NLP for Sentiment Analysis

Natural Language Processing (NLP) plays a crucial role in sentiment analysis, which involves several key steps:

- **Text Pre-processing:** Text must be cleaned and normalized before analysis. This involves removing punctuation, converting text to lowercase, stemming, and removing stop words.
- **Feature Engineering:** Text is converted into a numerical format for machine learning. Techniques include Bag-of-Words, n-grams, Term Frequency-Inverse Document Frequency (TF-IDF), and word embeddings.
- **Sentiment Classification:** The numerical representation is fed into a machine learning model for sentiment classification. This can be a simple model like Naive Bayes or a more complex model like a Deep Learning network.
- **Evaluation:** The performance of the sentiment classifier is assessed using metrics such as accuracy, precision, recall, and F1 score to ensure it predicts sentiment accurately.

NLP provides the tools and techniques necessary to process, analyze, and understand natural language text, facilitating effective sentiment analysis.

### 2.2 Topic Modeling

Topic Modeling involves identifying the underlying topics in a set of text documents or datasets to uncover hidden structures and organization. Key techniques include:

- **Latent Dirichlet Allocation (LDA):** LDA is a generative probabilistic model that treats documents as mixtures of topics and topics as mixtures of words. It estimates the probability of words in each topic and topics in each document to identify latent topics in the text. LDA is flexible and can handle large text collections to discover topics not explicitly mentioned.



- **Non-Negative Matrix Factorization (NMF):** NMF is a linear algebra-based technique that decomposes the term-document matrix into two matrices: one representing topics and the other representing word weights within topics. NMF is efficient and interpretable, producing coherent and meaningful topics from sparse data.
- **Hierarchical Dirichlet Process (HDP):** HDP extends LDA to discover an unlimited number of topics. It is a flexible, unsupervised method that can reveal a hierarchical structure of topics, capturing the organization of text.

### 2.2.1 LDA for Topic Modeling

Latent Dirichlet Allocation (LDA) is a widely used model for topic modeling in NLP. The process involves:

- **Defining a Document-Topic Matrix:** A matrix is created where rows represent documents and columns represent topics. Each element denotes the proportion of words in the document belonging to a specific topic.
- **Generating Topics:** Topics are generated by assuming probability distributions over words and topics. LDA determines which topics are most likely to generate each document and which words are most likely to belong to each topic.
- **Estimating Parameters:** The model parameters (word-topic and document-topic distributions) are estimated using an expectation-maximization (EM) algorithm variant.
- **Assigning Topics to Documents:** Topics are assigned to documents based on the estimated parameters, identifying the most likely topic for each word in the document.

## 3. Review of Literature

Sentiment analysis has recently gained prominence as a valuable research area, particularly in understanding emotional reactions to information. This review highlights the application of sentiment analysis to gauge public opinion on COVID-19 vaccines, emphasizing various methodologies and findings from recent studies.

### 3.1 Sentiment Analysis of COVID-19 Vaccines

The study in [2] analyzed tweets to assess public sentiment towards COVID-19 vaccinations. They collected over 4.5 million tweets from January 7, 2020, to January 3, 2021, and examined their evolution over time and across regions. The analysis revealed that positive sentiments were predominant and garnered more engagement. Despite this, vaccine rejection and reluctance were more prominently discussed, with notable differences across countries. Anti-vaccine propaganda was often propagated by Twitter bots and political activists, while pro-vaccine sentiments were shared by prominent individuals and groups.



The research in [3] utilized data from social media to analyze Indian users' attitudes towards COVID-19 vaccinations. They extracted 73,760 tweets in English and applied sentiment analysis using LDA models. The findings indicated that 35% of the tweets were positive, 17% negative, and 47% neutral. The major concerns among Indians included potential side effects and health risks associated with the vaccine during September to December 2020.

In [4], researchers examined anti-vaccine sentiments on social media platforms, including Instagram and Facebook, between November 16–18, 2020. They analyzed 9,705 English-language comments using QSR-NVivo 10. The study highlighted significant issues such as the perceived rush in vaccine production, concerns about long-term effects, and skepticism about vaccine safety and efficacy. Many respondents were opposed to receiving the COVID-19 vaccine, preferring natural immunity over vaccination.

The study presented in [5] focused on global attitudes towards COVID-19 vaccines by analyzing 2.7 million tweets from November 1, 2020, to January 31, 2021. Using the Valence Aware Dictionary and sEntiment Reasoner (VADER), they categorized tweets into positive, neutral, and negative sentiments. The analysis showed that 42.8% of tweets expressed positivity, 26.7% neutrality, and 30.3% negativity. Brazilian tweets had the worst sentiment scores, while those from the United Arab Emirates had the most positive sentiment.

In [6], researchers analyzed over 78 million vaccination-related tweets from December 1, 2020, to February 28, 2021, to explore regional and temporal variations in the discourse around COVID-19 vaccines in the United States. Their data-driven approach revealed a dynamic conversation with varying tones and topics. The discourse ranged from vaccine distrust and skepticism in the southern US to discussions on herd immunity, natural infections, and countering disinformation.

The study in [7] examined online discussions about COVID-19 vaccines using lexical-based sentiment analysis and topic modeling with LDA on 18,000 comments from 13 Reddit communities. This analysis aimed to uncover the themes and sentiments prevalent in these online discussions, providing insights into the public's views on the vaccine.

### **3. Review of Literature**

Sentiment analysis has become a critical tool for understanding public attitudes towards COVID-19 vaccines. This review synthesizes recent studies on sentiment analysis of vaccine-related discussions on social media platforms, highlighting methodologies and findings.

#### **3.1 Sentiment Analysis of COVID-19 Vaccines**

The study in [11] focused on Filipino sentiments regarding government initiatives on COVID-19 vaccination. By mining 11,974 tweets in English, Filipino, and Tagalog from March 2021, and applying Natural Language Processing (NLP) techniques with the Naïve Bayes classifier using RapidMiner, the researchers achieved an accuracy of 81.77%. The results indicated that 83% of tweets expressed



supportive and excited attitudes towards vaccination, while 9% were neutral and 8% unfavorable. This study reported a higher accuracy compared to recent sentiment analysis studies on Philippine Twitter data.

In [12], researchers analyzed 189,888 English tweets from India concerning adverse reactions to the COVID-19 vaccine. Employing both LDA and NLP methods, the study found that approximately 78.5% of the tweets were neutral or positive about side effects. The topic modeling revealed that concerns were primarily related to productivity at work and risk of death, reflecting a significant public discourse on potential vaccine side effects.

The study [13] used Twitter's API to collect tweets and hashtags about COVID-19 vaccines, focusing on India, the United States, the United Kingdom, and China. Natural Language Processing techniques and a decision tree algorithm were used for sentiment analysis. The study found that most responses were either ambiguous or favorable towards vaccination, although there remained a degree of hostility and anxiety about potential harms.

### 3.2 Summary of Research Articles on COVID-19 Vaccines

The following table summarizes research articles that conducted sentiment analysis on COVID-19 vaccines across different social media platforms:

Ref. No.	Data Source	Data Collection Timeline	Countries	Technique(s) Used	Findings
[2]	Twitter data with 4,552,652 tweets	7th Jan 2020 – 3rd Jan 2021	Global	VADER	Two-thirds of users had a positive impression of COVID-19 vaccinations.
[3]	Twitter data with 73,760 unique English tweets	Sep to Dec 2020	India	LDA model, NLP	Nearly 35% were in favor of COVID-19 vaccinations, and 17% were against it.
[4]	9705 Instagram and Facebook English comments	16th – 18th Nov 2020	US	QSR-NVivo 10 Software	Anti-COVID-19 vaccine sentiments were found.
[5]	Twitter data with 2,678,372 English tweets	1st Nov 2020 – 31st Jan 2021	Global	VADER, LDA	About 43% of the posts were in favor of vaccination.

[6]	Twitter data with 78.1 million tweets	1st Dec 2020 – 28th Feb 2021	US	Data-driven approach	COVID-19 vaccination discourse differs by region and changes over time.
[7]	18,000 Reddit community posts	1st Dec 2020 – 15th May 2021	US	Lexical-based sentiment analysis, LDA	Overall more positive than negative sentiments prevailed.
[8]	Twitter data with 220,844 tweets	12th Dec 2020 – 24th Nov 2021	Global	Proposed data analytics framework using LSTM	Optimistic attitudes about particular vaccination brands can be seen.
[9]	Twitter data with 221,922 tweets	1st Jan 2021 – 22nd Mar 2021	Global	Min-Hash5 local-sensitive hashing method, Botometer	Feelings regarding AstraZeneca were tainted by false information and negative coverage.
[10]	Twitter data with 13,414 Korean language tweets	23rd Feb 2021 – 22nd Mar 2021	South Korea	Spearman's correlation analysis, LDA	Negative sentiments about COVID-19 vaccines were prominent.
[11]	Twitter data with 11,974 tweets (en, fil, tl)	1st – 31st Mar 2021	Philippines	NLP, Naïve Bayes model, RapidMiner software	83% of tweets were pro-vaccination.
[12]	Twitter data with 189,888 English tweets	Mar – Apr 2021	India	NLP, LDA	78.5% of tweets were neutral or favorable regarding COVID-19 vaccination side effects.
[13]	Twitter data	Not Specified	India, US, UK, China	NLP, Decision tree algorithm	Most people are either ambivalent or enthusiastic regarding immunizations.

This summary captures the key methodologies and findings of the studies, providing a comprehensive overview of sentiment analysis on COVID-19 vaccines across various regions and platforms.

#### 4. Conclusion

The extensive use of Twitter for sentiment analysis in the reviewed studies highlights its effectiveness in raising public health awareness and influencing vaccination uptake. The findings underscore the



platform's potential to address vaccine hesitancy and resistance by leveraging positive messaging, mitigating negative content, and actively managing anti-vaccination accounts, including bots, based on public sentiment.

The analysis reveals a global trend where individuals are actively discussing their thoughts and concerns about COVID-19 vaccines on social media. This engagement varies significantly across different regions, reflecting diverse opinions and sentiments.

### Key Observations:

- **Prevalence of Twitter:** Twitter emerged as the predominant medium for sentiment analysis, illustrating its central role in public discourse on COVID-19 vaccines. The analysis across different languages, including English, Chinese, Korean, and Filipino, underscores the platform's broad reach.
- **Temporal and Regional Variations:** Sentiments about COVID-19 vaccines have evolved over time and differ by region. For example, sentiment in India and the Philippines was generally positive, while South Korea exhibited predominantly negative views at certain times. In the United States, sentiments shifted from negative to positive as vaccination campaigns progressed.
- **Methodological Insights:** Natural Language Processing (NLP) and Latent Dirichlet Allocation (LDA) were the most commonly used techniques, showing high accuracy in sentiment and topic analysis. This suggests that these methods are effective tools for understanding public sentiment.
- **Public Concerns and Preferences:** Concerns about vaccine side effects and health risks were prevalent. Optimistic attitudes towards specific vaccine brands were noted, and gender-specific studies revealed differing preferences for vaccine brands between men and women.

Overall, the study highlights the importance of using sentiment analysis to gauge public opinion and tailor vaccine education campaigns. By leveraging these insights, public health authorities can develop targeted strategies to enhance vaccine acceptance and address concerns more effectively.

### References

1. C. Arumugam and K. Nallaperumal, "EIAASG: Emotional Intensive Adaptive Aspect-Specific GCN for sentiment classification," *Knowledge-Based Systems*, vol. 260, p. 110149, Jan. 2023, doi: 10.1016/j.knosys.2022.110149.
2. S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, and S. Sharif, "An analysis of COVID-19 vaccine sentiments and opinions on Twitter," *International Journal of Infectious Diseases*, vol. 108, pp. 256–262, Jul. 2021, doi: 10.1016/j.ijid.2021.05.059.
3. S. Praveen, R. Ittamalla, and G. Deepak, "Analyzing the attitude of Indian citizens towards COVID-19 vaccine – A text analytics study," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 15, no. 2, pp. 595–599, Mar. 2021, doi: 10.1016/j.dsx.2021.02.031.





4. L. P. Wong, Y. Lin, H. Alias, S. A. Bakar, Q. Zhao, and Z. Hu, "COVID-19 Anti-Vaccine Sentiments: Analyses of Comments from Social Media," *Healthcare*, vol. 9, no. 11, p. 1530, Nov. 2021, doi: 10.3390/healthcare9111530.
5. S. Liu and J. Liu, "Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis," *Vaccine*, vol. 39, no. 39, pp. 5499–5505, Sep. 2021, doi: 10.1016/j.vaccine.2021.08.058.
6. S. C. Guntuku, A. M. Buttenheim, G. Sherman, and R. M. Merchant, "Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States," *Vaccine*, vol. 39, no. 30, pp. 4034–4038, Jul. 2021, doi: 10.1016/j.vaccine.2021.06.014.
7. C. A. Melton, O. A. Olusanya, N. Ammar, and A. Shaban-Nejad, "Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence," *Journal of Infection and Public Health*, vol. 14, no. 10, pp. 1505–1512, Oct. 2021, doi: 10.1016/j.jiph.2021.08.010.
8. K. T. Shahriar, M. N. Islam, Md. M. Anwar, and I. H. Sarker, "COVID-19 analytics: Towards the effect of vaccine brands through analyzing public sentiment of tweets," *Informatics in Medicine Unlocked*, vol. 31, p. 100969, 2022, doi: 10.1016/j.imu.2022.100969.
9. [9] D. Jemielniak and Y. Krempovych, "An analysis of AstraZeneca COVID-19 vaccine misinformation and fear mongering on Twitter," *Public Health*, vol. 200, pp. 4–6, Nov. 2021, doi: 10.1016/j.puhe.2021.08.019.
10. J.-G. Shim, K.-H. Ryu, S. H. Lee, E.-A. Cho, Y. J. Lee, and J. H. Ahn, "Text Mining Approaches to Analyze Public Sentiment Changes Regarding COVID-19 Vaccines on Social Media in Korea," *International Journal of Environmental Research and Public Health*, vol. 18, no. 12, p. 6549, Jun. 2021, doi: 10.3390/ijerph18126549.
11. C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, "Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes," *Information*, vol. 12, no. 5, p. 204, May 2021, doi: 10.3390/info12050204.
12. P. Sv, J. Tandon, Vikas, and H. Hinduja, "Indian citizen's perspective about side effects of COVID-19 vaccine – A machine learning study," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 15, no. 4, p. 102172, Jul. 2021, doi: 10.1016/j.dsx.2021.06.009.
13. P. Chinnasamy, V. Suresh, K. Ramprathap, B. J. A. Jebamani, K. Srinivas Rao, and M. Shiva Kranthi, "COVID-19 vaccine sentiment analysis using public opinions on Twitter," *Materials Today: Proceedings*, vol. 64, pp. 448–451, 2022, doi: 10.1016/j.matpr.2022.04.809.