INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Twitter Sentiments and Opinions Analysis of COVID-19 Vaccine Regarding Effectiveness of Vaccine

Kritika Kumari[1]*, Sumesh Sood[2], Anshul Kalia[3], Santosh Saklani[1], Jyoti Negi[1]

1 Research Scholar, Department of Computer Science, Himachal Pradesh University, Shimla, Himachal Pradesh, India
2 Associate Professor, Department of Computer Science, Himachal Pradesh University, Shimla, Himachal Pradesh, India
3 Assistant Professor, Department of Computer Science, Himachal Pradesh University, Shimla, Himachal Pradesh, India

OPEN ACCESS

Received: 06-11-2023
Accepted: 14-01-2024
Published: 05-03-2024

Citation: Kumari K, Sood S, Kalia A, Saklani S, Negi J (2024) Twitter Sentiments and Opinions Analysis of COVID-19 Vaccine Regarding Effectiveness of Vaccine. Indian Journal of Science and Technology 17(11): 1051-1058. https://doi.org/10.17485/IJST/v17i11.2534

* **Corresponding author**.

kritikakhanna424@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: To make an extensive analysis of sentiment within the discourse surrounding COVID-19 vaccines on Twitter, employing Natural Language Processing (NLP) methodologies. **Method:** The research methodology encompasses data collection via the Twitter API, followed by sentiment analysis facilitated by the TextBlob library. Pre-processing stages are integrated to cleanse and standardize the Twitter data. Subsequently, sentiment analysis categorizes tweets into positive, negative, and neutral sentiments based on polarity scores. **Findings:** The findings, grounded in a dataset spanning from March 1, 2022, to April 30, 2022, comprising 61,934 tweets, unveil that 45.0% of tweets conveyed positive sentiment, 17.3% exhibited negativity, and 37.7% maintained neutrality. Moreover, an exploration of tweet subjectivity revealed that 70.1% of the content expressed subjectivity, while 29.9% conveyed objectivity. The research is augmented with visual representations, including word clouds and subjectivity-polarity graphs, that offer a more intuitive understanding of sentiment trends. **Novelty:** This study contributes to the expanding landscape of sentiment analysis and its application within the context of public health crises, empowering stakeholders with valuable knowledge to enhance vaccine acceptance and effectiveness. The tool used "Tweet Downloader" in data collection makes this study different from other reviewed studies.

Keywords: COVID-19 vaccine; Twitter sentiment analysis; Public perception; Natural Language Processing (NLP); Social media data

1 Introduction

The COVID-19 pandemic, stemming from the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has had a global impact, necessitating widespread public health interventions. Among these interventions, the development and deployment of COVID-19 vaccines have been pivotal in the fight against the virus. Concurrently, social media platforms, such as Twitter, have become significant channels

for disseminating information and gauging public sentiment regarding these vaccines.

Twitter, a microblogging site with over 300 million users, is increasingly being utilized to broadcast public health information and collect real-time health data using crowd-sensing techniques [1].

In 2020, Xue et al. reviewed Twitter data on how people feel and talk about COVID-19. To analyze tweets relating to the coronavirus, they employed machine-learning techniques. The maximum number of conspicuous important topics is 11, which are then divided into ten subtexts. Their Sentiment analysis findings show that uncertainty about the coronavirus tops all issues. They also looked at the importance and flaws of the study [2].

Yin et al. developed a novel approach to analyze the sentiment and subject matter of 13 million tweets related to COVID-19 in 2020. Their findings align with previous studies, indicating that there were slightly more positive tweets than negative ones during the two-week study period. Additionally, the study delved into the most popular topics discussed daily during the pandemic and identified the most common issues that people were concerned about during the research period [3].

Several studies based on collective sentiment were conducted throughout the COVID-19 period, and Twitter has emerged as a favorite tool for gathering research data [4,5].

An NLP tool called the valence-aware dictionary for sentiment reasoning was utilized by Alam et al. in 2021 to retrieve people's attitudes toward specific vaccines. By classifying the received attitudes into three groups (positive, negative, and neutral), they were able to show the full scenario [6].

Only a few research have examined global public sentiment over COVID-19 vaccine-related tweets. For instance, Dubey examined the emotions expressed in tweets about the COVID-19 vaccine that were only sent from India between January 14, 2021, and January 18, 2021 [7].

Rahul et al. (2021) conducted a study on the sentiment analysis and topic modeling of tweets about the COVID-19 vaccine from November 1 to December 16, 2020. They collected 980,557 tweets and used VADER and TextBlob to analyze the sentiment of 572,958 of them. TextBlob found that 48.3% of the tweets were positive, 36.1% were neutral, and 15.6% were negative, while VADER found that 47.3% of the tweets were positive, 28.6% were neutral, and 24.1% were negative [8].

The Naive Bayes algorithm, based on Bayesian probability, is used by Ritonga et al. in 2021. 6000 tweets from Indonesia using the phrase "vaccine COVID-19" in the algorithm showed 56% negative, 39% positive, and 1% neutral answers. Only two weeks of Indonesian tweets from January 2021 are included in this study project [9].

In 2021, Samira et al. created an analytical model based on Twitter content. They assess prevailing opinions and attitudes toward COVID-19 vaccines. They have identified the sentiments and opinions about vaccination in retweets following a breakup and have examined their development over time, geography, key themes, hashtags, post analytics tools, and account features. For this analysis, such tweets are gathered from Twitter utilizing a Twitter Application Programming Interface (API) authentication token [10].

According to a study conducted by Shamrat et al. in 2021, the general public's views on the Pfizer, Moderna, and AstraZeneca vaccines developed to fight COVID-19 were analyzed using NLP to preprocess unprocessed tweets, and KNN classification algorithms to classify the processed data. The study found that people have a more positive outlook towards Pfizer and Moderna vaccines than AstraZeneca vaccines, with rates of 47.29 and 46.16, respectively [4].

In 2022, Das and Kolya suggested a revolutionary method for achieving deep neural network accuracy in sentiment analysis of tweets about the coronavirus and future case rise predicting. From tweets about the coronavirus, they create a sizable Twitter collection. Also, they separated the data into training and test sets. They performed both predictive analytics and polarity categorization at the same time, and their prediction for the future increase in coronavirus cases was backed by statistical evidence. [11].

Marcec and Likic (2022), introduce drive analysis with sentiment analysis based on Twitter. They have used social networking sites like Twitter, which may prove to be a successful method of expressing sentiments and opinions concerning the SARS-COV-2 vaccine that may be seen practically quickly [12].

The goal of vaccination is to teach the immune system to remember the traits of the targeted viral pathogens and to be able to quickly and effectively mount an immunological response to counteract viral pathogen withdrawal in the future [13].

The provided literature reviews various studies that leverage Twitter data to analyze public sentiment and opinions related to COVID-19 and vaccination. The studies use diverse methodologies, including machine learning algorithms, sentiment analysis, and topic modeling While these studies collectively contribute to understanding public sentiment on Twitter regarding COVID-19 and vaccination, there are notable variations in their methodologies, geographic focus, timeframes, and specific aspects of analysis (e.g., sentiment, topics). The gap lies in the need for more standardized approaches, longer-term observations, and a global perspective to provide a comprehensive understanding of evolving public opinions over time and across different regions. Additionally, variations in sentiment analysis tools and algorithms contribute to differences in reported sentiment percentages across studies, emphasizing the importance of methodological consistency for meaningful comparisons. Future research could

focus on addressing these gaps to enhance the robustness and generalizability of findings in this dynamic and evolving field.

2 Methodology

In this paper, Twitter tweets regarding the COVID-19 vaccine were analyzed to show the positivity or negativity of sentiments using NLP. Twitter is regarded as a fantastic data mine because, unlike some other social media, every user's tweets are often public and extensive. This is a significant benefit when trying to gather a lot of data for sentiment analysis. The data was extracted from Tweet Downloader which is another tool of Twitter API, it required a secret access token to download the tweets. The study, based on a dataset spanning from March 1, 2022, to April 30, 2022, and consisting of 61,934 tweets, discloses that 45.0% of the tweets conveyed a positive sentiment, 17.3% displayed negativity, and 37.7% remained neutral. Additionally, an analysis of tweet subjectivity indicated that 70.1% of the content exhibited subjectivity, while 29.9% conveyed objectivity. The research is enhanced with visual aids such as word clouds and subjectivity-polarity graphs, providing a more intuitive grasp of sentiment trends. The gathered information was saved in CSV file format and analyzed using TextBlob, a sentiment analysis tool[14]. Once the data was available, sentiment analysis was carried out by invoking the TextBlob library. A well-liked Python text data processing package is called TextBlob[14]. The TextBlob library employs natural language processing and sophisticated machine learning techniques to examine each word in the documents within the corpus, identifying the overall sentiments expressed as positive, negative, or neutral. Operating on the bag-of-words model, TextBlob relies on a predefined dictionary that categorizes words as negative or positive. It assesses each word in the document, assigning individual scores, and the document's final sentiment score is determined through a pooling operation, which involves averaging all the individual sentiment scores[15]. It provides a straightforward API for exploring common NLP tasks including speaking part extraction, noun sentence extraction, feelings extraction, categorization and interpretation, and more. Figure 1 shows the various steps used to analyze the sentiments of the COVID-19 vaccine dataset.

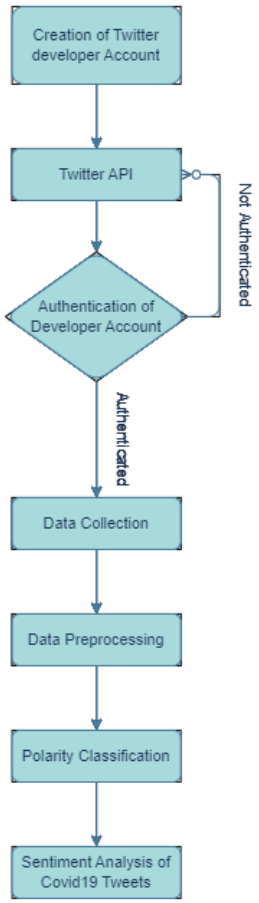

Fig 1. Flow chart for COVID-19 vaccine sentiment analysis

2.1 Data collection

The implementation is done on text data extracted from Twitter specifically the ones that are related to covid-19 vaccine. To download tweets "Tweet Downloader", is a Twitter API tool used, which downloads potentially large batches of tweets into CSV or JSON files. Figure 2 shows the data extraction method from Twitter.

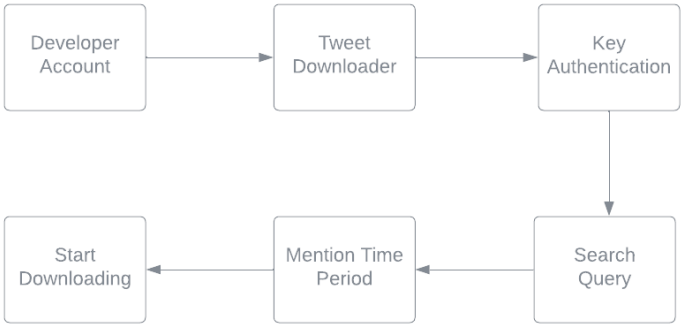

Fig 2. Twitter data extraction

2.2 Pre-processing dataset

Text mining must start with pre-processing. It could be used to remove undesired words from messages like punctuation, emails, phone numbers, and links. NLP approaches provide valuable services for that. These methods are used to move text from its native format to another reduced structure and minimize the size of textual data [16]. Figure 3 shows various steps used in pre-processing:

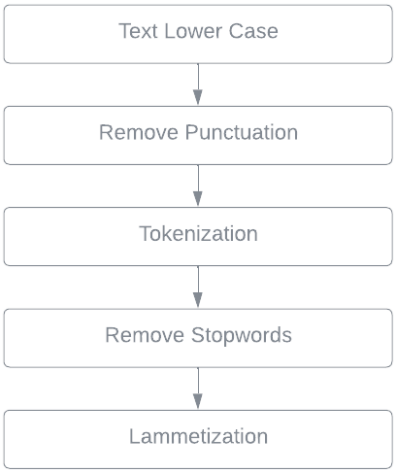

Fig 3. Data pre-processing techniques

2.2.1 Text Lower Case

This step involves converting all the text to lowercase. This is done to ensure consistency and to avoid treating the same word differently based on its capitalization. For example, "Hello" and "hello" would be considered the same word after this transformation.

2.2.2 Remove Punctuation

In this step, all punctuation marks (such as commas, periods, exclamation marks, etc.) are removed from the text. Punctuation is often irrelevant for many NLP tasks, and removing it simplifies the analysis.

2.2.3 Tokenization

Tokenization is the process of breaking down a text into smaller units called tokens. Tokens are usually words or subwords. This step is essential for further analysis because it helps to convert the raw text into a format that can be easily processed by machine learning models [17].

2.2.4 Remove Stopwords

Stopwords are common words (e.g., ”and,” ”the,” ”is”) that are often removed from text during preprocessing. These words don't carry much meaning on their own and are frequently used, so excluding them can help focus on more meaningful content and reduce the dimensionality of the data.

2.2.5 Lammetization

Lemmatization involves reducing words to their base or root form, known as a lemma. This is done to ensure that different inflected forms of a word are treated as the same word. For example, ”running” and ”ran” would be lemmatized to ”run.” This step is more advanced than stemming, which involves removing prefixes or suffixes to obtain the root form [4].

2.3 Polarity classification

The TextBlob library is used to calculate the polarity and subjectivity of the tweets. Using polarity scores, the tweets were classified as positive, negative, or neutral. A polarity score of -1 to 0 was classified as negative, a polarity score of 0 was classified as neutral, and a polarity score of 0 to 1 was classified as positive sentiment. The subjectivity ranges from [0.0, 1.0], where 0.0 is very objective and 1.0 is very subjective.

3 Sentiment Analysis Results

The tweets were collected from March 1, 2022, to April 30, 2022, from Twitter. The total number of gathered tweets was 61934, using TextBlob tweets categorized into three types that are positive sentiments, negative sentiments, and neutral sentiments. The result shows that the polarity of positive, negative, and neutral sentiments are 45.0%, 17.3%, and 37.7% respectively and the subjectivity of most subjective and objective sentiments are 70.1% and 29.9% respectively. Figure 4 shows the (a) value count of the polarity of the tweet in the bar graph and (b) distribution of COVID-19 vaccine sentiments (positive, negative, and neutral) percentage in a pie chart.

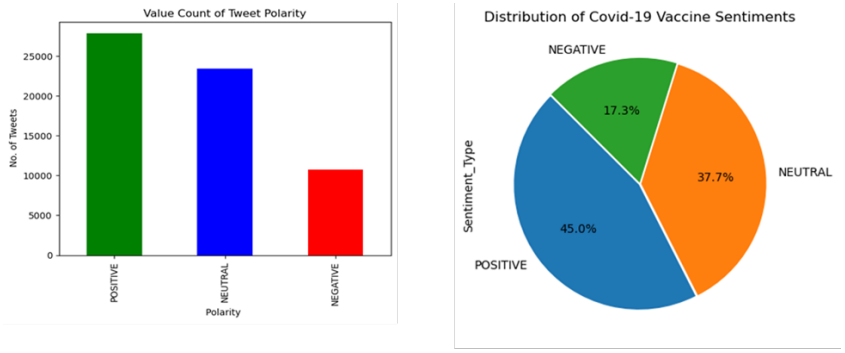

Fig 4. (a) Value count of Tweet Polarity, (b) Polarity distribution of Covid-19 vaccine sentiments

Figure 5 shows the (a) value count of tweet subjectivity and (b) distribution of COVID-19 vaccine sentiments (subjective and objective) percentage in a pie chart.

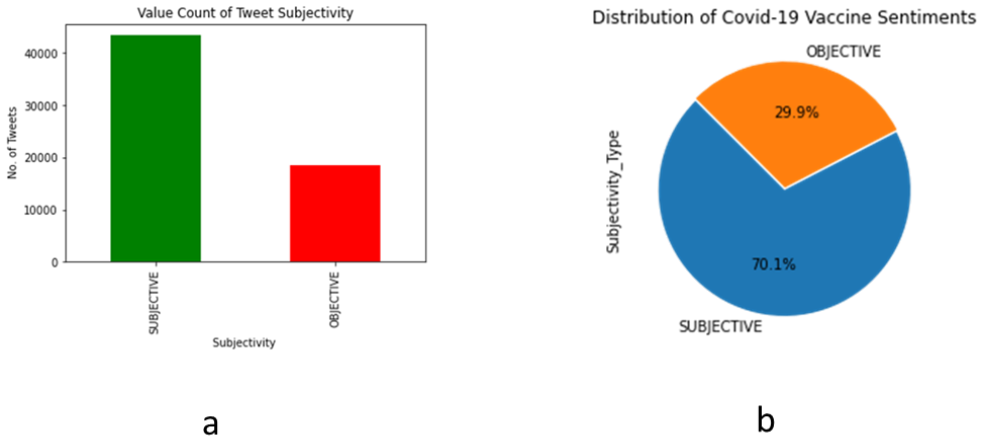

Fig 5. (a) Value count of Tweet Subjectivity, (b) Subjectivity distribution of Covid 19 vaccine sentiments

A word cloud is a visual representation of text data that displays the most frequently occurring words in a dataset. It is a popular way to visualize text–based data such as tweets. Figure 6 shows the python word cloud graphs of (a) overall tweets, (b) neutral tweets, (c) positive tweets, (d) and negative tweets.

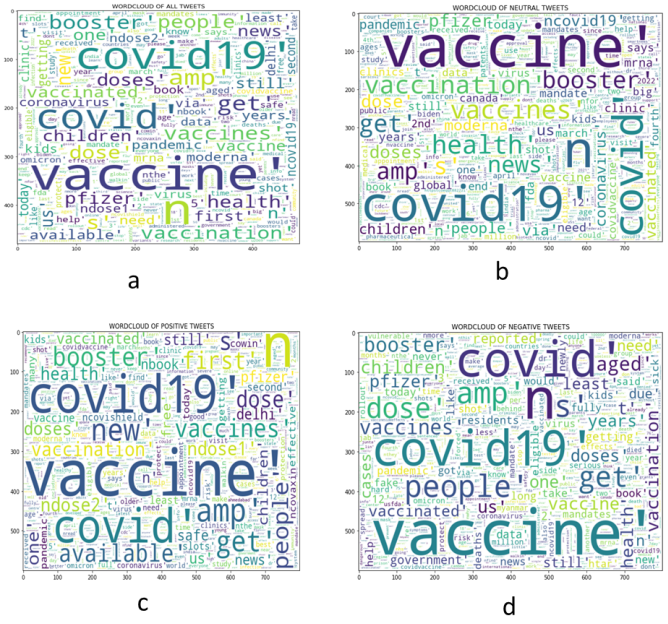

Fig 6. (a) All tweets, (b) Neutral tweets, (c) Positive tweets, (d) Negative tweets

The polarity and subjectivity graph of COVID-19 vaccine sentiments displays data points representing the sentiment of individual pieces of text on a two-dimensional plane. The X-axis represents polarity, with positive, negative, and neutral polarity. The Y-axis represents subjectivity, with high subjectivity at the top and low subjectivity at the bottom. The graph can provide insight into how people feel about the vaccine, how strongly they feel about it, and identify areas of concern. It shows a cluster of data points in the bottom-right quadrant indicating positive opinions without personal bias, a smaller cluster in the top-right quadrant indicating positive opinions with personal bias, a cluster in the top-left quadrant indicating negative opinions with personal bias, and a small cluster in the bottom-left quadrant indicating negative opinions without personal bias. Figure 7 shows the subjectivity and polarity graph of tweet sentiments.

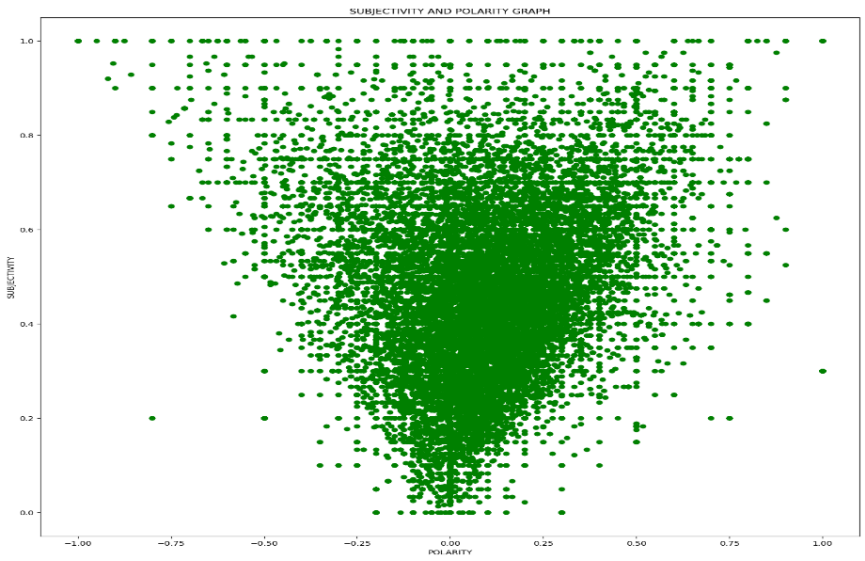

Fig 7. Subjectivity and Polarity Graph

4 Conclusion

Twitter, with its extensive user base, emerged as a vital platform for monitoring public sentiment and discourse regarding COVID-19 vaccines. Leveraging Natural Language Processing (NLP) techniques and data gathered from March 1, 2022, to April 30, 2022, this study meticulously analyzed 61,934 tweets containing the keyword ”covid vaccine.”

The findings revealed a nuanced landscape of public sentiment. Notably, 45.0% of tweets conveyed positive sentiments, showcasing optimism and support for COVID-19 vaccines. Conversely, 17.3% of tweets exhibited negative sentiments, reflecting concerns and skepticism. A substantial portion, 37.7%, maintained a neutral stance, reflecting a diverse range of opinions. Additionally, subjectivity analysis unveiled that 70.1% of tweets expressed subjective viewpoints, while 29.9% remained objective.

In future endeavors, this study will conduct a comparison of its results with findings from prior research by employing machine learning and deep learning algorithms. While recognizing the limitations of a confined geographic focus, short-term analysis, and methodological variations, the research uncovered valuable insights. Looking ahead, there are promising paths to enhance our comprehension of public sentiment on COVID-19 vaccines. Global expansion, longitudinal analyses, and advanced analytics can deepen understanding. Qualitative insights and community engagement strategies enrich the narrative for impactful health communication. Real-time monitoring, cross-platform analyses, and ethical considerations are vital for robust and reliable research practices. These avenues beckon a future of nuanced insights, contributing to a comprehensive understanding of evolving sentiments in the realm of public health on social media platforms.

References

1) Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An “Infodemic”: Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak. *Open Forum Infectious Diseases*. 2020;7(7):1–9. Available from: https://doi.org/10.1093/ofid/ofaa258.
2) Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLOS ONE*. 2020;15(9):1–12. Available from: https://doi.org/10.1371/journal.pone.0239441.
3) Yin H, Yang S, Li J. Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media. In: International Conference on Advanced Data Mining and Applications, ADMA 2020;vol. 12447 of Lecture Notes in Computer Science. Springer, Cham. 2021;p. 610–623. Available from: https://doi.org/10.1007/978-3-030-65390-3_46.
4) Shamrat FM, Chakraborty S, Imran MM, Muna JN, Billah MM, Das P, et al. Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*. 2021;23(1):463–470. Available from: http://doi.org/10.11591/ijeecs.v23.i1.pp463-470.
5) Agarwal B, Nayak R, Mittal N, Patnaik S. Deep learning-based approaches for sentiment analysis. 1st ed. Algorithms for Intelligent Systems;Singapore. Springer. 2020. Available from: https://doi.org/10.1007/978-981-15-1216-2.
6) Alam KN, Khan MS, Dhruba AR, Khan MM, Al-Amri JF, Masud M, et al. Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. *Computational and Mathematical Methods in Medicine*. 2021;2021:1–15. Available from: https://doi.org/10.1155/2021/4321131.

7) Dubey AD. Public Sentiment Analysis of COVID-19 Vaccination Drive in India. *SSRN Electronic Journal*. 2021;p. 1–5. Available from: https://dx.doi.org/10.2139/ssrn.3772401.
8) Rahul K, Jindal BR, Singh K, Meel P. Analysing Public Sentiments Regarding COVID-19 Vaccine on Twitter. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE. 2021;p. 488–493. Available from: https://doi.org/10.1109/ICACCS51430.2021.9441693.
9) Pristiyono, Ritonga M, Ihsan MAA, Anjar A, Rambe FH. Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. In: Annual Conference on Computer Science and Engineering Technology (AC2SET) ;vol. 1088 of IOP Conference Series: Materials Science and Engineering. IOP Publishing. 2021;p. 1–6. Available from: https://iopscience.iop.org/article/10.1088/1757-899X/1088/1/012045.
10) Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*. 2021;108:256–262. Available from: https://doi.org/10.1016/j.ijid.2021.05.059.
11) Das S, Kolya AK. Predicting the pandemic: sentiment evaluation and predictive analysis from large-scale tweets on Covid-19 by deep convolutional neural network. *Evolutionary Intelligence*. 2022;15(3):1913–1934. Available from: https://doi.org/10.1007/s12065-021-00598-7.
12) Marcec R, Likic R. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal*. 2022;98(1161):544–550. Available from: https://doi.org/10.1136/postgradmedj-2021-140685.
13) Ramadhan NG, Adhinata FD. Sentiment analysis on vaccine COVID-19 using word count and Gaussian Naïve Bayes. *Indonesian Journal of Electrical Engineering and Computer Science*. 2022;26(3):1765–1772. Available from: http://doi.org/10.11591/ijeecs.v26.i3.pp1765-1772.
14) Ansari MTJ, Khan NA. Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content. *Electronic Journal of General Medicine*. 2021;18(6):1–10. Available from: https://doi.org/10.29333/ejgm/11316.
15) Sv P, Lorenz JM, Ittamalla R, Dhama K, Chakraborty C, Kumar DVS, et al. Twitter-Based Sentiment Analysis and Topic Modeling of Social Media Posts Using Natural Language Processing, to Understand People's Perspectives Regarding COVID-19 Booster Vaccine Shots in India: Crucial to Expanding Vaccination Coverage. *Vaccines*. 2022;10(11):1–9. Available from: https://doi.org/10.3390/vaccines10111929.
16) Verma R, Chhabra A, Gupta A. A statistical analysis of tweets on covid-19 vaccine hesitancy utilizing opinion mining: an Indian perspective. *Social Network Analysis and Mining*. 2022;13(1):1–12. Available from: https://doi.org/10.1007/s13278-022-01015-2.
17) Chinnasamy P, Suresh V, Ramprathap K, Jebamani BJA, Rao KS, Kranthi MS. COVID-19 vaccine sentiment analysis using public opinions on Twitter. *Materials Today: Proceedings*. 2022;64(Part 1):448–451. Available from: https://doi.org/10.1016/j.matpr.2022.04.809.