

Machine Learning based Sentiment Analysis of COVID-19 Vaccination Drive in India

Smitesh D. Patravali*

Department of CSE, SDMCET, Dharwad, Karnataka.
Corresponding Author Email. smitesh.patravali@gmail.com

Siddu P. Algur

Vice Chancellor

Vijayanagara Sri Krishnadevaraya University, Ballari, Karnataka

Nihal S. Algur

Bosch Global Software Technologies Pvt. Ltd, Bangalore

(Received January 23, 2023, Revised March 25, 2023, Accepted April 14, 2023)

Abstract: Over 95 million people have been affected by the COVID-19, and two million people have deceased as a result. The lockdown affected most of the world and the need of immunization began. After Bharat Biotech and Serum Institute of India (SII) received formal approval from the Central Drugs and Standards Committee (CDSCO), India, COVAXIN and COVISHIELD were released to assure the citizens immunity by intensifying vaccination campaign. The COVID-19 vaccination, on the other hand, had its critics in the community. The purpose of this study was to look at how Indian tweets about these two vaccines made people feel. The analysis reveals that, despite some negative sentiments, majority of people were supportive of the vaccines. TextBlob outperformed the NRC Lexicon approach when it came to sentiment classification, by to a comparison study.

Keywords: TextBlob, NRC Lexicon, Sentiment, Vaccine, COVID-19.

1. Introduction

The pandemic virus has resulted in a global health crisis, necessitating never-ending pains to reduce health risk. To lessen the likelihood of COVID-19 infection and death, fast-paced vaccination campaigns were underway^{1,2}. People expressed their opinions on social media platforms regarding vaccination risks despite warning from government. These opinions can be analyzed for inference and have vaccination acceptance policies. The

proposed research of how people felt about the COVID-19 vaccine was voluntary. Pharmaceutical companies were invited to produce vaccines by the World Health Organization. Currently, India has approved vaccines manufactured by pharmaceutical companies like COVISHIELD, COVAXIN, and others. Studying people's behavior patterns can assist in developing appropriate policies to increase vaccination acceptance of COVID-19. The actors involved in the process, the strategies developed by governments, and the related authorities will have a noteworthy impact on the behavioral patterns of vaccine adoption. The achievement of the vaccination drives will be greatly influenced by the skill to gather information in real time to make dynamic decisions³. Semantic study was carried on data extracted from Facebook and tweets that have been geo-tagged to identify trends in particular countries or cities^{4,5}. The Chinese government utilized information from 'SinaWeibo' and other online entertainment stages to provide the public with the preferred information about the pandemic⁶. The authors^{7,8} demonstrated need of NLP techniques to determine a population's opinions, thoughts, and feelings about events.

For various natural language processing methods on raw text, TextBlob^{9,10} is a lexicon-based method with a Python library that provides training interface to process text data, analyze sentiments, mine noun phrases, create POS tags, interpret, and classify. The TextBlob library accompanies different in-constructed capabilities that help the assignment of language handling. It can be used with Spanish, English, Arabic, and other languages. NLTK and pattern processing are used in conjunction with the TextBlob for emotion analysis¹¹⁻¹³. Its dictionary has roughly 2918 lexicons. The polarity computation is either based on facts or individual opinions. The sentiment score is returned by the sentiment analyzer. The NRC Emotion Lexicon is a list of English words and their associations with eight emotions like surprise, anger, fear, sadness, joy, anticipation, trust, and disgust. Crowdsourcing was used to manually create the annotations.

The following research questions are outlined in this study for the purpose of analyzing people's sentiments regarding the vaccination:

Q1: On Twitter, what opinions do individuals have regarding the COVID-19 vaccination?

Q2: How successful is the proposed method for determining the tweet emotion?

The following are some of the contributions:

1. A method for conducting a study of people's feelings regarding the COVID vaccine is proposed.

2. To classify opinion into positive and negative by utilizing TextBlob and NRC based approaches.

The remaining part of the article is prepared as follows: The proposed effort is discussed in section II. The experimentation is elaborated in section III followed by results. Finally, the article is concluded in section IV.

2. Proposed Methodology

Dataset overview and pre-processing

The Twitter developer account was used to mine COVID vaccine tweets. The tweets were filtered using various geolocations and specific keywords like #Covid19, #Vaccine, #Corona #Vaccine, #covidvaccine,#coronavaccine, #corona vaccination, and #covid19 vaccination. Data preprocessing removes unrelated information like punctuation, stop words, @username and # characters to advance the training process of the models. The NLTK library was utilized for preprocessing.

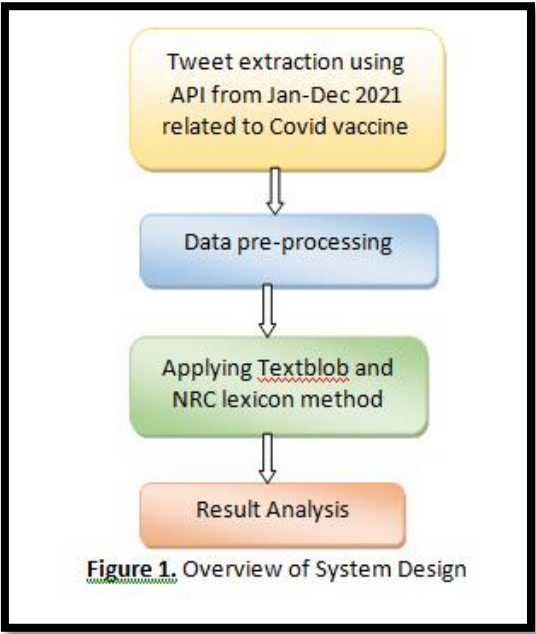


Figure 1. Proposed system model

Algorithm 1. Sentiment analysis of vaccines using TextBlob.

Input: COVID vaccine tweets # India
Result: To evaluate the polarity of obtained tweet
Initialization (for every tweet in the total set)
 Determine the polarity score.

```
Check state:
    if the polarity score is greater than 0 then
        Tweet is +ve;
    else
        Tweet is -ve;
End of condition
End of loop
```

Algorithm2. Sentiment analysis of vaccines using the NRC Lexicon.

```
Input: COVID vaccine tweets # India
Result: To evaluate the polarity of obtained tweet
Initialization (for every tweet in the total set)
    Determine the polarity score.
Check state:
    if the polarity score is greater than 0 then
        Tweet is +ve;
    else
        Tweet is -ve;
End of condition
End of loop
```

3. Result Analysis

A. COVID-19 Vaccine based Sentiment Analysis using TextBlob approach

In month of January, it was observed that COVISHIELD had 61% tweets are associated with positive score of sentiment and 39% with negative sentiment score. Further, it was observed that COVAXIN had 28% tweets are associated with positive score of sentiment and 72% with negative sentiment score. In month of February, it was observed that COVISHIELD had 60% tweets are associated with positive score of sentiment and 40% with negative sentiment score. Further, it was observed that COVAXIN had 18% tweets are associated with positive score of sentiment and 82% with negative sentiment score. In month of March, it was observed that COVISHIELD had 69% tweets are associated with positive score of sentiment and 31% with negative sentiment score. Further, it was observed that COVAXIN had 38% tweets are associated with positive score of sentiment and 62% with negative sentiment score. In month of April, it was observed that COVISHIELD had 65% tweets are associated with positive score of sentiment and 35% with negative sentiment score. Further, it was observed that COVAXIN had 40% tweets are associated with positive score of sentiment and 60% with negative sentiment score.

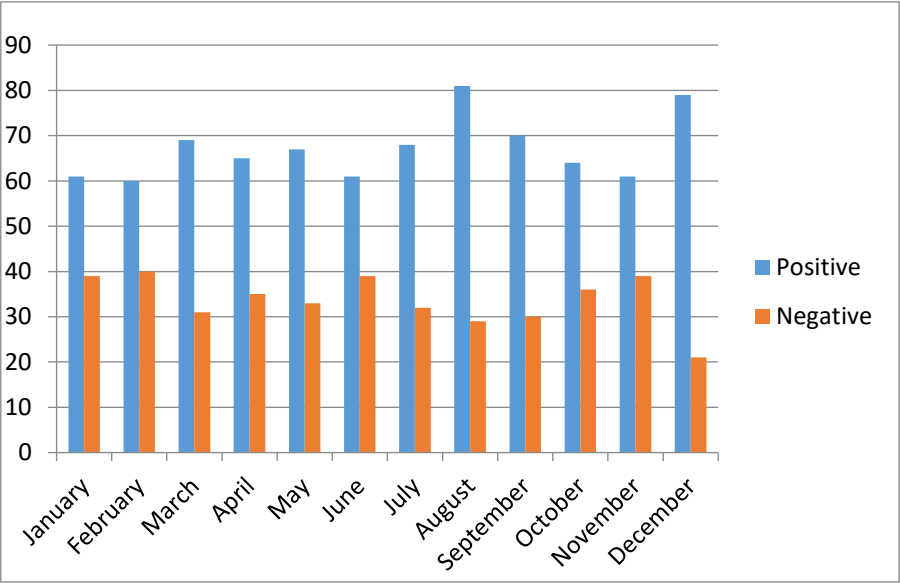


Figure 2. Textblob analysis of sentiments from 23-25 from Jan-Dec 2021 for COVISHIELD

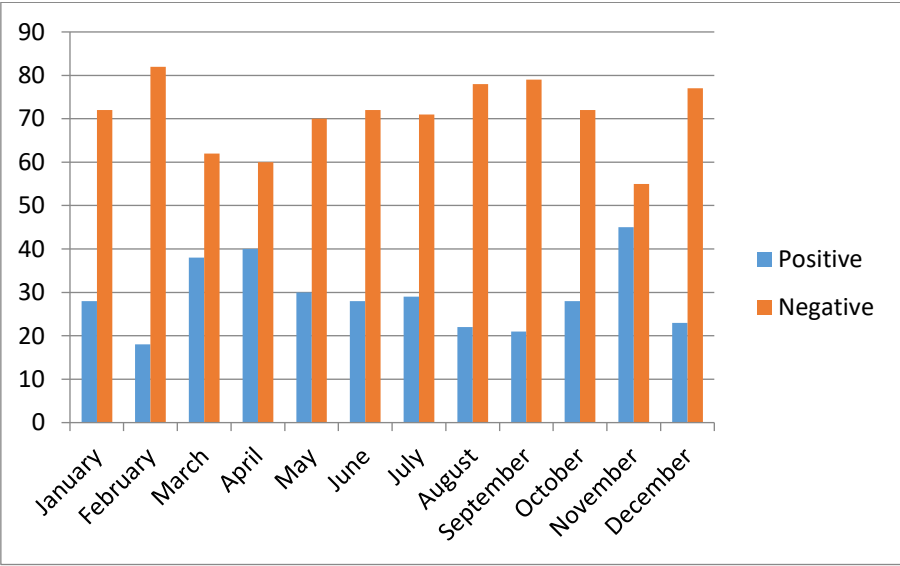


Figure 3. Textblob analysis of sentiments from 23-25 from Jan-Dec 2021 for COVAXIN

In month of May, it was observed that COVISHIELD had 67% tweets are associated with positive score of sentiment and 33% with negative sentiment score. Further, it was observed that COVAXIN had 30% tweets are associated with positive score of sentiment and 70% with negative sentiment score. In

month of June, it was observed that COVISHIELD had 61% tweets are associated with positive score of sentiment and 39% with negative sentiment score. Further, it was observed that COVAXIN had 28% tweets are associated with positive score of sentiment and 72% with negative sentiment score. In month of July, it was observed that COVISHIELD had 68% tweets are associated with positive score of sentiment and 32% with negative sentiment score. Further, it was observed that COVAXIN had 29% tweets are associated with positive score of sentiment and 71% with negative sentiment score. In month of August, it was observed that COVISHIELD had 81% tweets are associated with positive score of sentiment and 29% with negative sentiment score. Further, it was observed that COVAXIN had 22% tweets are associated with positive score of sentiment and 78% with negative sentiment score.

In month of September, it was observed that COVISHIELD had 70% tweets are associated with positive score of sentiment and 30% with negative sentiment score. Further, it was observed that COVAXIN had 21% tweets are associated with positive score of sentiment and 79% with negative sentiment score. In month of October, it was observed that COVISHIELD had 64% tweets are associated with positive score of sentiment and 36% with negative sentiment score. Further, it was observed that COVAXIN had 28% tweets are associated with positive score of sentiment and 72% with negative sentiment score. In month of November, it was observed that COVISHIELD had 61% tweets are associated with positive score of sentiment and 39% with negative sentiment score. Further, it was observed that COVAXIN had 45% tweets are associated with positive score of sentiment and 55% with negative sentiment score. In month of December, it was observed that COVISHIELD had 79% tweets are associated with positive score of sentiment and 21% with negative sentiment score. Further, it was observed that COVAXIN had 23% tweets are associated with positive score of sentiment and 77% with negative sentiment score.

B. COVID-19 Vaccine based Sentiment Analysis using NRC Lexicon approach

In month of January, it was observed that COVISHIELD had 54% tweets are associated with positive score of sentiment and 43% with negative sentiment score. Further, it was observed that COVAXIN had 33% tweets are associated with positive score of sentiment and 53% with negative sentiment score. In month of February, it was observed that COVISHIELD had 52% tweets are associated with positive score of sentiment and 44% with negative sentiment score. Further, it was observed that COVAXIN had 35% tweets are associated with positive score of sentiment and 52% with negative sentiment score. In month of March, it was observed that COVISHIELD had 55% tweets are

associated with positive score of sentiment and 45% with negative sentiment score. Further, it was observed that COVAXIN had 32% tweets are associated with positive score of sentiment and 55% with negative sentiment score. In month of April, it was observed that COVISHIELD had 51% tweets are associated with positive score of sentiment and 42% with negative sentiment score. Further, it was observed that COVAXIN had 31% tweets are associated with positive score of sentiment and 51% with negative sentiment score.

In month of May, it was observed that COVISHIELD had 56% tweets are associated with positive score of sentiment and 41% with negative sentiment score. Further, it was observed that COVAXIN had 33% tweets are associated with positive score of sentiment and 54% with negative sentiment score. In month of June, it was observed that COVISHIELD had 52% tweets are associated with positive score of sentiment and 44% with negative sentiment score. Further, it was observed that COVAXIN had 35% tweets are associated with positive score of sentiment and 55% with negative sentiment score. In month of July, it was observed that COVISHIELD had 51% tweets are associated with positive score of sentiment and 47% with negative sentiment score. Further, it was observed that COVAXIN had 37% tweets are associated with positive score of sentiment and 57% with negative sentiment score. In month of August, it was observed that COVISHIELD had 55% tweets are associated with positive score of sentiment and 48% with negative sentiment score. Further, it was observed that COVAXIN had 34% tweets are associated with positive score of sentiment and 53% with negative sentiment score.

In month of September, it was observed that COVISHIELD had 52% tweets are associated with positive score of sentiment and 44% with negative sentiment score. Further, it was observed that COVAXIN had 36% tweets are associated with positive score of sentiment and 55% with negative sentiment score. In month of October, it was observed that COVISHIELD had 56% tweets are associated with positive score of sentiment and 42% with negative sentiment score. Further, it was observed that COVAXIN had 31% tweets are associated with positive score of sentiment and 54% with negative sentiment score. In month of November, it was observed that COVISHIELD had 55% tweets are associated with positive score of sentiment and 41% with negative sentiment score. Further, it was observed that COVAXIN had 33% tweets are associated with positive score of sentiment and 57% with negative sentiment score. In month of December, it was observed that COVISHIELD had 56% tweets are associated with positive score of sentiment and 44% with negative sentiment score. Further, it was observed that COVAXIN had 35% tweets are associated with positive score of sentiment and 56% with negative sentiment score.

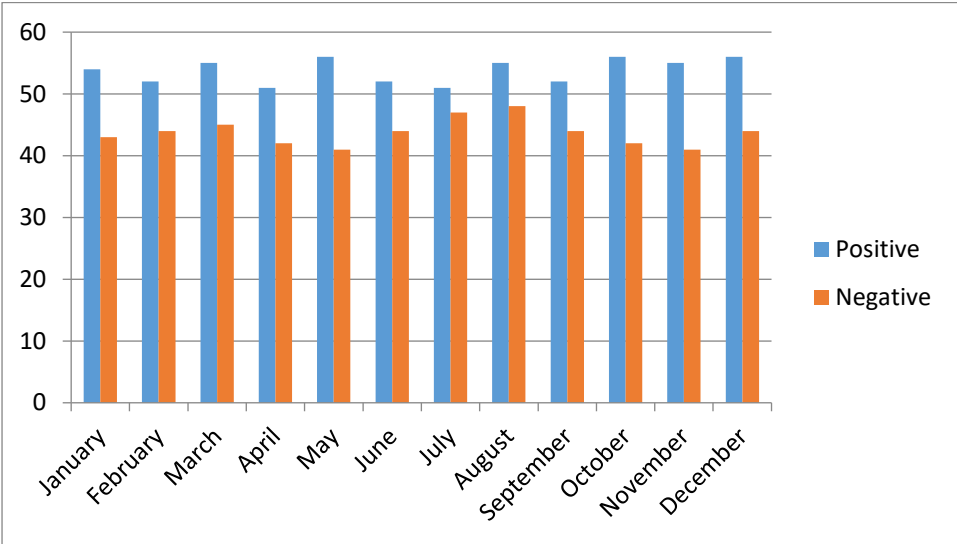


Figure 4. NRC lexicon analysis of sentiments from 23-25 from Jan-Dec 2021 for COVISHIELD

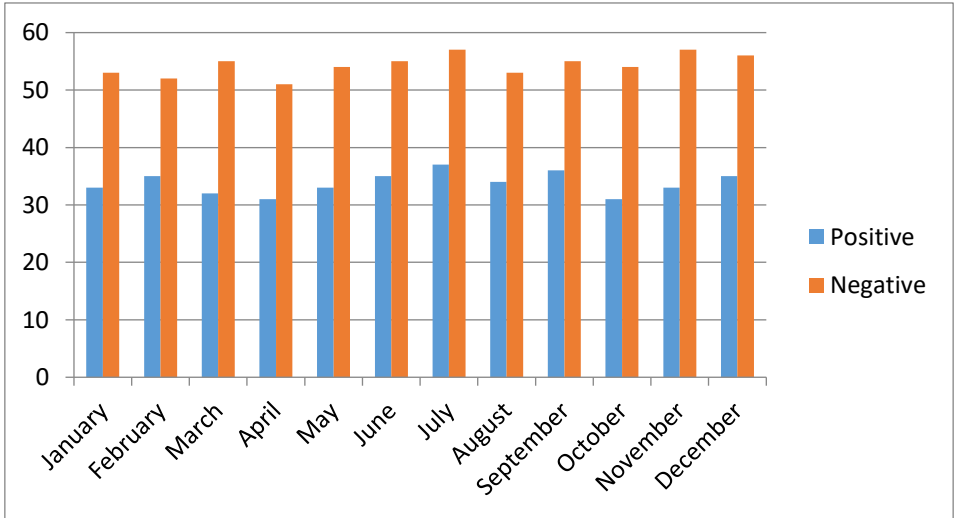


Figure 5. NRC lexicon analysis of sentiments from 23-25 from Jan-Dec 2021 for COVAXIN

Table 1. Accuracy for the dataset

Model	Accuracy
Textblob	89%
NRC Lexicon	79%

4. Conclusion

WHO encouraged the rapid vaccination for the entire population to reduce the risk of spread and fatalities. People expressed their concerns and reservations regarding the potential side effects despite the recommendations to take the vaccine from government officials, medical professionals, and social workers. Most tweets in the collected dataset are positive or supportive of the COVID-19 vaccination. Two methods strengthen the study: sentiment analysis using TextBlob and the NLP lexicon-based method. The results show that the TextBlob method outperformed the NRC Lexicon method when comparing positive and negative tweets. When compared to tweets about COVAXIN, the experiment revealed that tweets about COVISHIELD had more positive sentiments. Feelings like euphoria and trust were more on account of COVISHIELD than of the COVAXIN. A system that can analyze social trends and sentiment in real time for efficient decision-making can be considered as part of future scope.

Acknowledgement: We thank our institutions S D M College of Engineering and Technology, Dharwad and Vijayanagara Sri Krishnadevaraya University, Ballari for providing for their timely support and guidance.

References

1. World Health Organization. *Behavioural Considerations for Acceptance and Uptake of COVID-19 Vaccines: WHO Technical Advisory Group on Behavioural Insights and Sciences for Health; Meeting Report*, 15 October 2020; World Health Organization: Geneva, Switzerland, 2020.
2. I. Allagui; Breslow, H. social media for public relations: Lessons from four effective cases, *Public Relat. Rev.*, 42, (2016), 20-30.
3. C. Valentini; Is using social media “good” for the public relations profession? A critical reflection, *Public Relat. Rev.*, 41 (2015), 170-177.
4. World Health Organization. *Guidance on Developing a National Deployment and Vaccination Plan for COVID-19 Vaccines: Interim Guidance*, 16 November 2020; World Health Organization: Geneva, Switzerland, 2020.
5. O.B. Driss, S. Mellouli and Z. Trabelsi; From citizens to government policy-makers: Social media data analysis, *Gov. Inf. Q.* 36, (2019), 560-570.
6. T. Yigitcanlar, N. Kankanamge, A. Preston, P.S. Gill, M. Rezayee, M. Ostadnia, B. Xia and G. Ioppolo; How can social media analytics assist authorities in pandemic-related policy decisions? Insights from Australian states and territories, *Health Inf. Sci. Syst.* 8, (2020), 1-21.

7. Q. Liao, J. Yuan, M. Dong, L. Yang, R.;Fielding, W.W.T. Lam; Public engagement, and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data, *J. Med. Internet Res.* 22, (2020), e18796.
8. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede; Lexicon-based methods for sentiment analysis, *Comput. Linguist.*, 37, (2011), 267-307.
9. B. Liu; Opinion mining and sentiment analysis, *Web Data Mining* (2011), 459-526.
10. Y. Garg, N. Chatterjee; Sentiment analysis of twitter feeds, *Proceedings of the International Conference on Big Data Analytics, New Delhi, India*, 20-23 December 2014; Springer: Berlin/Heidelberg, Germany, (2014), 33-52.
11. I. El Alaoui, Y. Gahi, R. Messoussi, Y. Chaabi, A. Todoskoff, A. Kobi; A novel adaptable approach for sentiment analysis on bigsocial data, *J. Big Data*, 5, (2018), 12.
12. J.F. Sánchez-Rada, Iglesias, C.A. CRANK; A Hybrid Model for User and Content Sentiment Classification Using Social Contextand Community Detection, *Appl. Sci.*, 10, (2020), 1662.
13. S. Loria; TextBlob Documentation. Release 0'15, 2, (2018), 269.