COVID-19 vaccine sentiment analysis using public opinions on Twitter

P. Chinnasamy [a,*], V. Suresh [b], K. Ramprathap [c], B. Jency A. Jebamani [d], K. Srinivas Rao [e], M. Shiva Kranthi [f]

[a] Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, India
[b] Department of Computer Science and Engineering, Dr.N.G.P. Institute of Technology, Coimbatore, India
[c] Department of Management Studies, M.Kumarasamy College of Engineering, Karur, India
[d] Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India
[e] Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, India
[f] UG Student, Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, India

ARTICLE INFO

ABSTRACT

Twitter, as is well known, is one of the most active social media platforms, with millions of tweets posted every day, in which different people express their opinions on topics such as travel, economic concerns, political decisions, and so on. As a result, it is a useful source of knowledge. We offer Sentiment Analysis using Twitter Data for the research. Initially, our technology retrieves currently accessible tweets and hashtags about various types of covid vaccinations posted on Twitter through using Twitter's API. Following that, the imported Tweets are automatically configured to generate a collection of untrained rules and random variables. To create our model, we're utilizing, Tweepy, which is a wrapper for Twitter's API. Following that, as part of the sentiment analysis of new Messages, the software produces donut graphs.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Advanced Materials for Innovation and Sustainability.

1. Introduction

Passengers' thoughts on Twitter can be examined using sentiment analysis. In a study, sentiment analysis is used to investigate people's interests, assessments, and sentiments depends on what they wish to be written. Sentiment analysis is used as it has to deal with opinion, reality, and decisions that has to be taken. It is not uncommon for people to seek out the opinions of others when making decisions. Sentiment analysis is used in this study to look at ideas of people who wrote on Twitter. The goal is to see how commuter line passengers react to the Covid-19 being transmitted. The twitter data was evaluated. Furthermore, the information gathered was divided into three categories by IEEE 2020. This article is available for download and available, as well as complete text and data mining rights, as well as re-use and analysis rights for, namely Positive, Negative, and Neutral. After that, sentiment analysis can be used to determine people's true influence.

The NLTK approach will be employed in this investigation. Sentiment Analysis of Public Transportation on social media Using the NLTK was the method chosen based on past research. This study used sentiment analysis to evaluate public transit customers and got replies and feedback. Users' responses and feedback are extremely varied, and each has their own point of view. NLTK is used for classification because it is an optimal method for classifying text, and it was chosen to ease the difficulty of the material on internet, which can be quite extensive. NLTK was also chosen based on the results of a survey on algorithms that would be employed in this process. The output demonstrates that the NLTK is one of suitable methods for solving sentiment analysis challenges. Furthermore, the decision tree technique was considered above the others since it gave superior accuracy. The decision tree is also more optimal to utilize. Both methods are simple to use and produce accurate results. As a result, the nave Bayes method and the decision tree will be used in this study to discover which method produces superior accuracy. In this study, the writing system contains that summarizes the study's context and aims. The literature review highlights the hypotheses that have been to aid in the development of this study. The research procedure explains how the data was gathered. Various methodologies can be used to perform sentiment categorization. We can categorize these approaches in several ways. in the three sorts that follow: 1) A lexicon-based strategy; 2) a machine learning/deep learning strategy, and 3) a hybrid strategy.

* Corresponding author.
E-mail address: chinnasamyponnusamy@gmail.com (P. Chinnasamy).

2. Related works

Sentiment analysis is a technique for analyzing user's thoughts, feelings, and judgments. According to some viewpoints, sentiment analysis is used to determine what users believe based on information such as written opinions. The analysis was carried out on the written opinion, according to the two opinions. People frequently express and post their ideas on social media as a result of the digital era's development, which makes us unable to avoid it. When it comes to internet, Twitter is one of the most popular places for people to express themselves. This can be utilized as a data source for analysis. Because of its prominence, Twitter was chosen as a source of opinion mining. A concise explanation of beliefs and perceptions up to 150 characters long on Twitter. This can be utilized as information in sentiment analysis to generate information and determine people's true influence. Text mining, which is a way of extracting important information from a material, was used in this sentiment analysis study. Text mining is that entails document preparation. Preprocessing transforms text from unstructured documents into structured information. The structured information is subsequently classified using a data mining classification approach [1].

The second study [2] used Naive Bayes, Decision Trees, and Random Forest to analyze sentiment in social media. Positive, negative, and neutral sentiments are categorized based on Twitter data. As a result, Indonesian Twitter users are more neutral in their views. The three methods with the highest accuracy were compared using data from fast miner tools. With an accuracy of 86.43 percent, Nave Bayes is more accurate than the other method. The Decision Tree and Random Forest were both 82.91percent accurate.

The third study [3] looked at classification with the Naive Bayes Classifier and Decision Tree Algorithms. The Naive Bayes classification and decision tree method are suggested for this study as a more accurate and effective prediction in the assumption sort of brain tumor. The goal of this research is to show that the Decision Tree algorithm is more straightforward and straightforward than the Nave Bayes algorithm. The type of tumor has been identified using these two algorithms, and it allows for the examination of historical data from data sets, allowing neurologists to make educated guesses about the type of tumor. The Decision Tree algorithm is, without a doubt, faster and more accurate than the Naive Bayes algorithm. Furthermore, the fourth study uses Decision Tree, K-Nearest Neighbor, and Nave Bayes to analyze sentiment from tweets. A tweet from e-commerce is used as a case study in this study. Text mining, text preprocessing, categorization, and other approaches are used in this study. These strategies are used to develop sentiment analysis classification and analysis. Rapid miner is also utilized to aid in the comparison of analysis attitudes by utilizing three different categories inside the dataset. According to the findings of this study [4,5], the Nave Bayes algorithm has the maximum accuracy of 77 percent. The fifth study compares Naive Bayes, Decision Trees, and Neural Networks in classification. This study categorizes web pages through a of the three algorithms. According to the findings of this study [6], they improved Nave. The Bayes classifier not only out performs the traditional Nave classifier. Bayes, on the other hand, performs similarly well.

Khakharia et al [7], they have used Machine Learning to scan public tweets from Twitter connected to COVID-19 immunizations in order to determine the user's perspective on the vaccine (ML). The analysis of the Twitter data into positive or negative messages was done using three distinct machine learning algorithms such as Multinomial Nave Bayes (MNB), Support Vector Machine (SVM), and Logistic Regression (LR). The LR algorithm produced the best results in the analysis, with a 97.3 percent accuracy, while SVM had a 96.26 percent accuracy. In comparison to the other ML algorithms, the MNB had the lowest accuracy of 88 percent. Marcec and Likic [8] introduced vaccination drive analysis using the Twitter based sentiment analysis. Here they have used social networking sites like Twitter might very well prove to be a useful medium of communication on SARS-CoV-2 vaccinations feelings and beliefs which can be observed nearly instantly.

Alam et al [9], they have used valence aware dictionary for sentiment reasoned, a natural language processing (NLP) tool, was used to assess people's feelings towards certain vaccines (VADER). They were able to illustrate the entire scenario by grouping the received attitudes into three categories (positive, negative, and neutral). We found that 33.96 percent of the respondents were positive, 17.55 percent were negative, and 48.49 percent were neutral. Nezhad et al [10,11], investigated Iranian people's views on COVID-19 vaccinations by analyzing Iranian twitter posts. Discuss the various Iranian attitudes toward indigenous and imported COVID-19 vaccinations. They have used CNN-LSTM architecture, we determined emotions of recovered tweets. Furthermore, we looked into Iranian attitudes on the COVID-19 vaccine.

Samira et al [4], developed an analysis model based on the content of Twitter. They were able to determine popular attitude and thoughts about the COVID-19 vaccinations. They have discovered vaccination feelings and views in retweets after separation and analyzed their growth over time, geographical location, principal themes, hashtags, post analytics tools, as well as accounts features. Such tweets are collected from Twitter using a Twitter API authentication token in this investigation. NLP is used to receive and analyze the raw twitter posts. A supervised KNN clustering algorithm is then used to classify the produced data. The data is divided into three categories by the algorithm: positive, negative, and neutral. These categories pertain to the prevailing attitude of the persons whom Retweets are being analysed by Javed et al [12]. Ansari et al [13], they have analysed COVID-19 immunization retweets to offer an assessment of the government's sentiments to ongoing immunization drives. Investigations on sentiment classification were also conducted in order to reveal new knowledge regarding the effects of location and gender including [3,14,15].

As a result of the best approach that can be taken is that the Naive Bayes algorithm is used for selection for training web pages.

3. Proposed method

Panda's library was used to turn the data-set into a data frame for the purposes of this paper, making data manipulation and analysis easier. The Natural Language Toolkit (NLTK) is a computational linguistics package that provides easy-to-use interfaces to over 50 lexical resources. This library was used for text pre-processing (stemming, lemmatization, and stop word removal), tokenization, sentiment analysis (NLTK Sentiment Intensity Analyzer), and Emotional Analysis (Lexicon based) in this article. Matplotlib is a data visualization package that may be used to create static, animated, and interactive data visualizations. This library was used to plot time series graphs and donut chart for the purposes of this paper.

Tokenization is the process of dividing a large chunk of text into smaller tokens. In this case, tokens can be words, characters, or sub words. I utilized NLTK's treebank tokenizer to perform word level tokenization on tweets in this project. Text Pre-processing cleans the text data and removes superfluous elements such as stop words, hashtags, and emoticons, among other things, so that the text is ready to be given to the model.

These methods were used in this study. Because of the strong assumptions and great accuracy, the Naive Bayes method is used. The Decision Tree method was chosen because it is more straightforward and straightforward to utilize. The following are the explanations for the two methods:
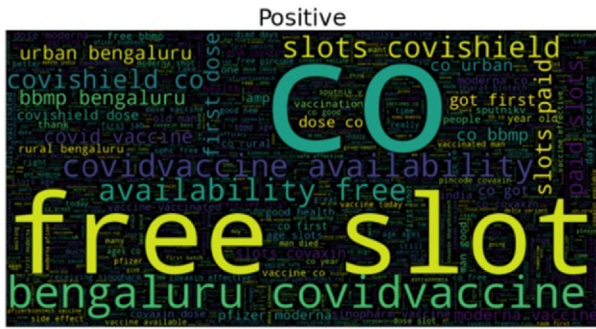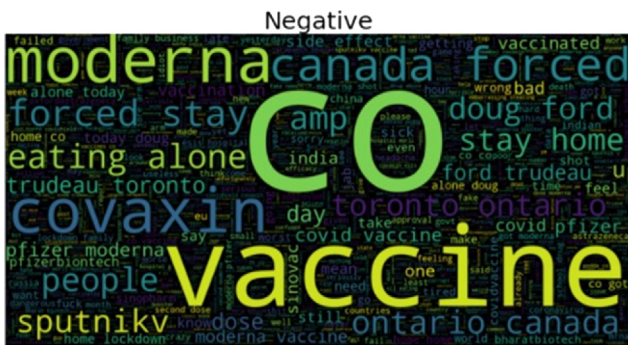
Fig. 1. The positive value.

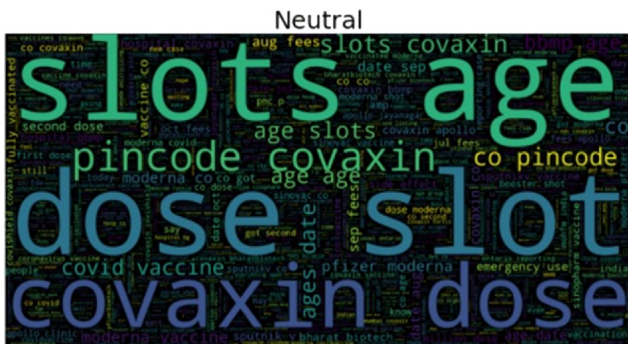**Fig. 2.** The negative value.

Fig. 3. The neutral value.

The proposed method starts with a data collection from different sources like twitter, Gaggle, GitHub etc. After that, we are applying tokenization methods on the dataset to identify the positive, negative and neutral values as shown in Figs. 1-3 respectively. Fig. 4 explain about the proposed system flow of our research. After tokenization, there are two processes going to happen like data preprocessing to remove the unwanted words from the data's, text processing is another task to identify the different values in the input data. Based on this output the results should passed into the cloud, then sentiment analysis is going to happen by applying the decision tree algorithm.

4. Performance evaluation

To summarize the findings, we can say that vaccines have had a beneficial impact on the population, with the majority of individuals being either neutral or positive about the various vaccines available. Individuals have a high level of trust in these vaccines, but there are still angry and fearful feelings in the community because many people are still infected with the deadly coronavirus, and many of these immunizations have had bad side effects. The majority of the tweets in this data set are from India, the United States, the United Kingdom, and China, as these are the nations that have been most afflicted by the virus or have been most involved in the development of coronavirus vaccines.

After applying the decision tree algorithm to analyze the covid'19 vaccines information. Here, we applied unigram and
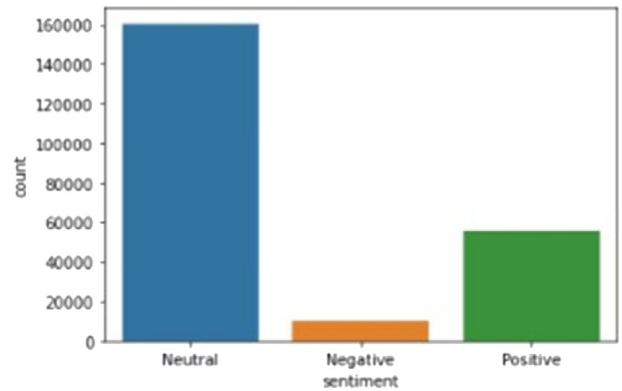

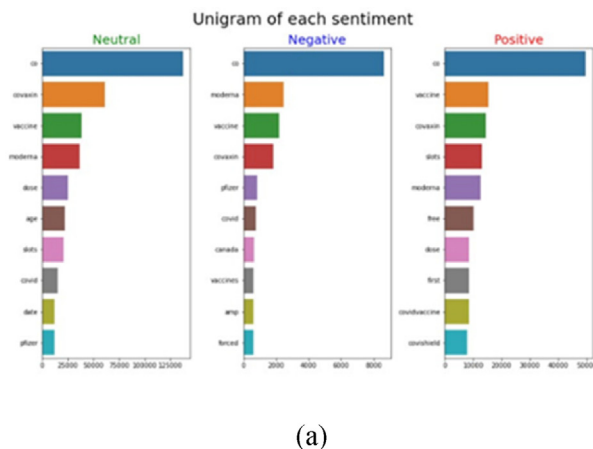
Fig. 5. (a) (b) the unigram and Bigram of each sentiment.

(a)

(b)

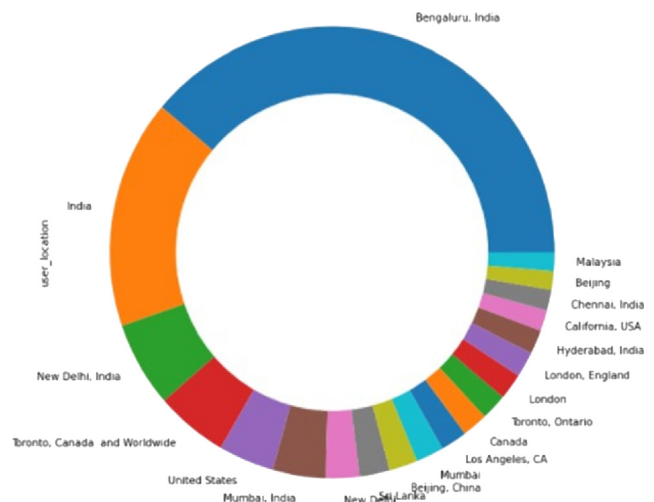**Fig. 4.** The architecture of the proposed method.

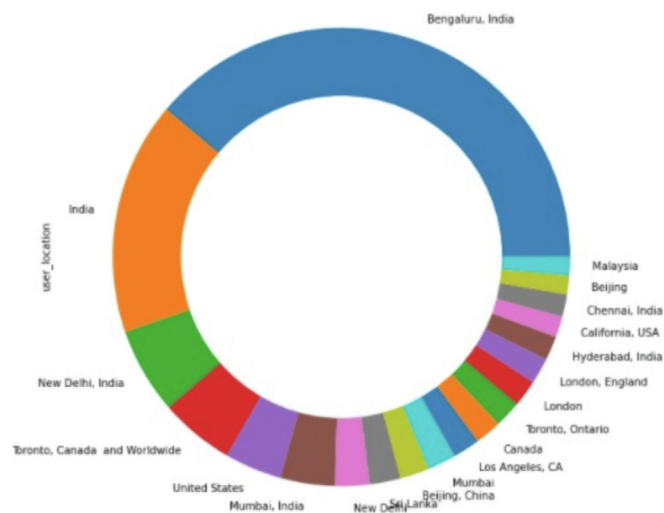**Fig. 6.** The proposed method analysis using decision tree.

Fig. 7. Different countries status of Covid'19 vaccines.

bigram techniques to apply for the each and every sentiment parameter as shown in Fig. 5.

Fig. 6 shows the proposed method analysis of sentiment on twitter-based account using decision tree algorithm. This method provides high accuracy (96%) compare to other methods like linear regression, classification etc. Fig. 7, clearly shows that the different countries of covid'19 vaccines details with sentiment analysis.

5. Conclusion

In recent months, there has been a growth in unfavorable opinion about both domestic and export vaccines in all over the globe, which is concerning. Because health authorities want to improve the adoption of COVID-19 immunizations in order to stop the epi-demic, they might use sites such as twitter to spread good messages and reduce negative ones.

P.Chinnasamy, Writing the papers and responsible for Implementation.

V.Suresh, Collection of DataSet.

K.Ramprathap - Proof Reading this article.

B Jency A Jebamani - Support for Plagiarism removal.

K. Srinivas Rao – Article writing.

Shiva Kranthi – Results and discussion section writing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. De Figueiredo, C. Simas, E. Karafillakis, P. Paterson, H.J. Larson, Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study, Lancet 396 (10255) (2020) 898–908, https://doi.org/10.1016/S0140-6736(20)31558-0.

[2] A. Hussain, A. Tahir, Z. Hussain, et al., Artificial intelligence-enabled analysis of public attitudes on facebook and twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational Study, J. Med. Internet Res. 2021;23(4):e26627. Published 2021 Apr 5. doi:10.2196/26627.

[3] K.S. Prasad, N.C.S. Reddy, B.N. Puneeth, A framework for diagnosing kidney disease in diabetes patients using classification algorithms, SN Comput. Sci. 1 (2020) 101, https://doi.org/10.1007/s42979-020-0096-7.

[4] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, S. Sharif, An analysis of COVID-19 vaccine sentiments and opinions on Twitter, Int. J. Infect. Diseases 108 (2021) 256–262.

[5] Nhan Cach Dang, María N. Moreno-García, Fernando De la Prieta, Sentiment analysis based on deep learning: a comparative study, Electronics 9(3) (2020) 483.

[6] Yong Shi et al., Survey on classic and latest textual sentiment analysis articles and techniques, Int. J. Inform. Technol. Decis. Making 18(04) (2019) 1243-1287.

[7] Aman Khakharia, Vruddhi Shah, Pragya Gupta, Sentiment analysis of COVID-19 vaccine tweets using machine learning (June 18, 2021). Available at http://dx.doi.org/10.2139/ssrn.3869531.

[8] R. Marcec, R. Likic, Using Twitter for sentiment analysis towards AstraZeneca/ Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines Postgraduate Medical, J. Publish. Online First 09 (2021), https://doi.org/10.1136/postgradmedj-2021-140685.

[9] K.N. Alam, M.S. Khan, A.R. Dhruba, M.M. Khan, J.F. Al-Amri, M. Masud, M. Rawashdeh, A. Korobeinikov, Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data, Comput. Math. Methods Med. 2021 (2021) 1–15.

[10] Z. Bokaee Nezhad, M.A. Deihimi, Twitter sentiment analysis from Iran about COVID 19 vaccine, Diabetes Metab Syndr. 16 (1) (2021), https://doi.org/10.1016/j.dsx.2021.102367 102367.

[11] Z. Bokaee Nezhad, M.A. Deihimi, Twitter sentiment analysis from Iran about COVID 19 vaccine, Diabetes Metabol. Syndrome: Clin. Res. Rev. 16 (1) (2022) 102367.

[12] F.M. Javed Mehedi Shamrat, Sovon Chakraborty, M. M. Imran, Jannatun Naeem Muna, Md. Masum Billah, Protiva Das, Md. Obaidur Rahman, Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm, Vol 23, No 1, pp. 463-470, 2021.

[13] M.T.J. Ansari, N.A. Khan, Worldwide COVID-19 Vaccines sentiment analysis through twitter content, Electron. J. Gen Med. (2021), https://doi.org/10.29333/ejgm/11316.

[14] K. Sai Prasad, O. Subhash Chander, G. Prabhakar Reddy, S. Gururaj, Artificial intelligence approach for classifying molecular dataset using density based technique with appropriate Euclidean distance measure, Mater. Today:. Proc. 4 (8) (2017) 8827–8836.

[15] K.S. Prasad, S. Pasupathy, P. Chinnasamy, A. Kalaiarasi, An approach to detect COVID-19 disease from CT scan images using CNN - VGG16 model, International Conference on Computer Communication and Informatics (ICCCI) 2022 (2022) 1–5, https://doi.org/10.1109/ICCCI54379.2022.9741050.