RESEARCH ARTICLE

Opinion Severity Prediction about Covid-19 Vaccination using Twitter Sentiments Analysis

Maaz Khan[1] **and Inayat Khan**[2,*]

[1] Institute of Management Sciences, Peshawar, Pakistan
[2] Department of Computer Science, University of Engineering and Technology Mardan, Mardan 23200, Pakistan

Abstract

Public opinion on COVID-19 vaccines strongly affects vaccination uptake. This study analyzes the sentiment of tweets about COVID-19 vaccines. Tweets from December 2021 were collected with the Twitter API, stored in a CSV file, and preprocessed with natural language processing techniques. A Naïve Bayes (NB) classifier was used for supervised classification and reached an accuracy of 90%. The study has three objectives: 1) data extraction and preprocessing, 2) sentiment analysis of people's views, and 3) evaluation of the NB classifier. Of the tweets, 36.82% are positive, 12.97% are negative, and 50.21% are neutral. The large neutral share suggests that many people may still be uncertain about vaccination and points to the need for targeted awareness programs. In addition, k-means clustering was used to determine severity levels, which produced high-positive, low-positive, high-negative, and low-negative clusters. On the same dataset, the NB classifier outperformed SVM, multinomial NB, random forest, and gradient boosting in accuracy. The findings can help authorities and public health initiatives address prevailing doubts and encourage vaccination, and they can support the control and prevention of future pandemics.

Keywords: COVID-19 vaccines, Tweets, sentiment analysis (SA), Naïve Bayes (NB), natural language processing (NLP), k-means clustering.

1 Introduction

With the arrival of the internet, digitalization has created many opportunities and challenges, including for the awareness, prevention, and control of COVID-19, where sentiment analysis of online content has become an important tool [1]. Social media platforms such as Twitter and Facebook allow people to collect and analyze a variety of information on trending topics with different tools and technologies. A quick view of public sentiment about a trending event such as COVID-19 is very useful for developing programs that promote social awareness and prevention [1].

Misleading information on social media has made COVID-19 a source of despair, stress, and worry, in addition to being a communicable disease that spreads through contact and through small droplets produced when people talk, cough, or sneeze. Sharing

Submitted: 22 Jun 2025
Revised: 27 Nov 2025
Accepted: 24 Jan 2026
Published: 23 November 2026

Vol. 2, **No.** 4, 2026.
10.62762/TACS.2025.224386

***Corresponding author:**
✉ Inayat Khan
inayatkhan@uetmardan.edu.pk

Citation
Khan, M., & Khan, I. (2026). Opinion Severity Prediction about Covid-19 Vaccination using Twitter Sentiments Analysis. *ICCK Transactions on Advanced Computing and Systems*, 2(4), 291–299.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

misleading information on platforms such as Twitter and Facebook fuels vaccine hesitancy [2].

Social media sites where people share their thoughts about a trending topic, such as Twitter (now X), generate a large number of pandemic-related tweets [3]. Such trending data can be mined with multiple tools and techniques to obtain a quick view of public opinion on topics such as COVID-19 [4] and Omicron [5]. Public opinion reveals patterns in trending topics that can support social awareness and prevention programs [6]. Over time, the number of COVID-19 vaccine brands and types has grown, each claiming a different level of efficacy, and each has its own difficulties, challenges, and complications. Social media in general, and Twitter in particular, show a visible division in society on this issue [7]. People around the globe are trying to speed up vaccination so that a large share of the population can be vaccinated early. However, many people are unwilling to cooperate because they hold different opinions about COVID-19 vaccination, which can be positive, negative, or neutral. Sentiment analysis (SA) has been used to analyze public sentiment towards COVID-19 vaccines [8]. SA is often performed on information extracted from Twitter to investigate public opinion on COVID-19 vaccination [9]. Using Twitter API access, public opinions can be extracted from unstructured data and categorized into three classes: i) positive sentiment, where the person agrees with vaccination; ii) negative sentiment, where the person does not agree with vaccination; and iii) neutral sentiment, where the person is doubtful about vaccination or only reports vaccine-related information [10].

Many studies have analyzed public sentiment during the COVID-19 period, and Twitter is one of the most popular data sources for such research [1, 11]. Chakraborty et al. [12] collected COVID-19-related tweets and analyzed their sentiment with deep learning classifiers. Pristiyono et al. [9] collected tweets manually through the Twitter API with a data crawler, using vaccine-related keywords, and classified them with the NB algorithm, which works on small as well as large datasets [10]; most of their tweets were negative. Twitter sentiment analysis has also been used to track vaccine hesitancy [13]. Deep learning [14] and other machine learning approaches [15] have also been used for SA. However, deep learning has limitations such as high computational cost and the black-box problem [16].

In this study, public tweets were collected from Twitter by scraping with the Twitter API, using the keywords 'Covid-19 Vaccine' and 'Covid-19 Vaccination' and hashtags such as #covidvaccines, #covidvaccination, #pfizer, #coronavaccine, #moderna, #AstraZeneca, and #Sinopharm. The collected data were analyzed to find whether public opinion towards COVID-19 vaccination is neutral, positive, or negative, and to estimate the awareness level of the public. Sentiment labels were derived from TextBlob polarity and subjectivity scores, and the NB classifier was then used to classify the tweets into neutral, negative, and positive classes. In addition, k-means clustering was used to form "high positive", "low positive", "high negative", and "low negative" clusters that indicate the severity of the tweets. This research can help authorities plan awareness, control, and prevention measures for future pandemics.

The rest of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the proposed approach, Sections 4 and 5 present the results and the classifier comparison, and Section 6 concludes the study.

2 Related Work

This section reviews previous work on sentiment analysis. Over the past few years there has been much discussion of the causes, consequences, and trends of the coronavirus pandemic. This part covers tweet SA with different machine learning (ML) and natural language processing (NLP) techniques. Extracting the necessary information from raw data is difficult and requires ML and NLP tools and techniques [15].

SA identifies people's thoughts, perceptions, viewpoints, ideas, and feelings, and it is used here to recognize people's sentiments and beliefs about trending topics on social media platforms [17]. SA has been applied to monitor vaccination sentiment and infectious disease dynamics [18], and to analyze public opinion on specific COVID-19 vaccines [19, 20].

Xue et al. [21] studied how people feel about COVID-19 and share their thoughts on Twitter, using ML techniques to evaluate COVID-19 tweets. They identified eleven salient topics, which were grouped into ten themes, and their sentiment analysis showed that fear of the unknown nature of the coronavirus dominated all topics. They also discussed the importance and limitations of the study.

Hung et al. [22] applied machine learning to English tweets originating in the United States during one month (20 March to 19 April 2020). They classified

902,138 tweets as positive (48.2%), neutral (20.7%), or negative (31.1%), and identified five dominant themes: healthcare environment, emotional support, business economy, social change, and psychological stress. These themes can help authorities understand the public response to the coronavirus and manage the pandemic.

Pristiyono et al. [9] assessed public perception of COVID-19 vaccines in Indonesia in January 2021. About 6000 tweets were crawled with vaccine-related keywords, and the NB algorithm classified them as negative (56%), positive (39%), and neutral (1%).

Samuel et al. [23] presented a textual analysis of COVID-19 tweets to investigate public sentiment, tracking the progression of fear associated with the rapid spread of infections as COVID-19 approached peak levels in the United States. They compared two ML classification methods (NB and logistic regression) and discussed the methods, implications, limitations, and opportunities of their approach. A limitation of their work is that it focused mainly on fear sentiment.

Kaur et al. [24] collected tweets using the hashtag keywords COVID-19, coronavirus, deaths, new cases, and recovered. They designed a Hybrid Heterogeneous Support Vector Machine (H-SVM) to classify sentiment as positive, negative, or neutral, and they compared its performance with RNN and SVM.

Pokharel [25] focused on the sentiment of Nepalese citizens towards the coronavirus epidemic. Tweets were collected with the hashtags #COVID-19 and #coronavirus from users whose location was Nepal, and sentiment was computed with Python libraries (Tweepy and TextBlob) without training ML classifiers. Sattar and Arifuzzaman [6] analyzed tweets about seven COVID-19 vaccine types and combined the sentiment results with a forecast of the vaccinated population in the USA. Dubey [26] analyzed tweets from twelve countries in March 2020 and reported mostly positive and hopeful sentiment, together with fear, sadness, and disgust.

3 Materials and Methods

This section presents the proposed sentiment analysis approach and its empirical evaluation. The study has three objectives: 1) data extraction and preprocessing, 2) analysis of people's sentiments, and 3) evaluation of the performance of the ML classifier. The NLP and ML tools used are described below.

For this study, tweets were first obtained from Twitter with the Tweepy library and saved in CSV format. The collected tweets include emojis, special characters, retweets, and hyperlinks, so they require preprocessing. NLP tools were used to preprocess the data so that it can be used in a supervised classification system. After special characters were removed, the text was tokenized, lemmatized, and normalized. Polarity and subjectivity were then computed with TextBlob [27]. A supervised NB algorithm was used to classify the polarity classes of the dataset. After that, k-means clustering was used to form "High-Positive", "Low-Positive", "High-Negative", and "Low-Negative" clusters that indicate the severity of the tweets. After cleaning (see Section 3.3), the dataset of 13,000 tweets was divided into an 80–20 split, i.e., 10,400 tweets for training the model and 2,600 tweets for independent testing.

3.1 Proposed Solution Framework

A methodology framework gives the research a structured plan, which supports research design and keeps the work clear and focused. Figure 1 shows the overall proposed methodology.

In this study, public tweets were collected from Twitter, and the implementation is based on tweets related to "COVID-19 vaccines". A Twitter API account is required to extract tweets. The authentication object is created after the Twitter keys and API credentials are stored in variables, and the access token and access token secret are then configured. Figure 2 shows the setup.

Tweets were extracted through the Twitter API account with the Python library Tweepy. Hashtags such as #covidvaccines, #covidvaccination, #pfizer, #coronavaccine, #moderna, #AstraZeneca, and #Sinopharm were used to find related tweets. In total, 13,109 related tweets were retrieved (Table 1). Only the textual data were extracted and stored in CSV format as the dataset. Figure 3 shows the data extraction setup.

Twitter is widely used to express concerns about COVID-19 [3], so it was selected as the data source. Sixteen seed keywords were chosen for the search query. The keywords and the retrieved data are presented in Table 1.

3.2 Data Annotation

During data collection, tweets were retrieved from Twitter and annotated. Data annotation labels text so that machines can learn from it, and it is a key

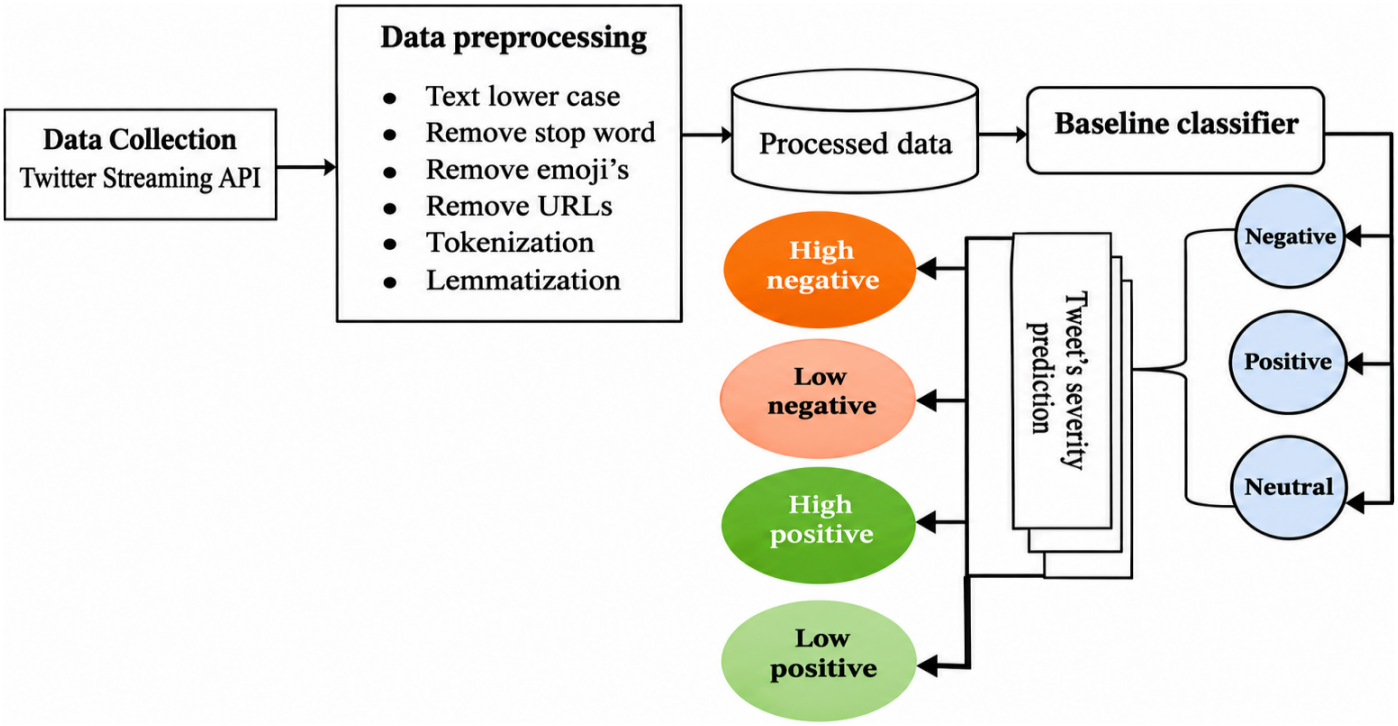


Figure 1. Proposed methodology framework.

Table 1. Related Words & Fetched Rows.

S.No.	Dataset	
	Keyword	**Rows Fetched**
1	AstraZeneca, AstraZeneca vaccine	2014
2	Corona vaccine, covid-19 vaccine, Covid vaccination, covidvaccine	6627
3	Moderna, modernavaccine	1219
4	Pfizer, Pfizer vaccines	2518
5	Sinopharm, Sinopharm vaccine	301
6	Sinovac, Sinovac vaccine	290
7	Sputnik, sputnik vaccine	140
Total		**13109**

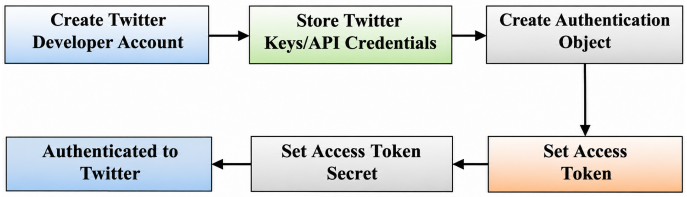


Figure 2. Twitter Authentication.

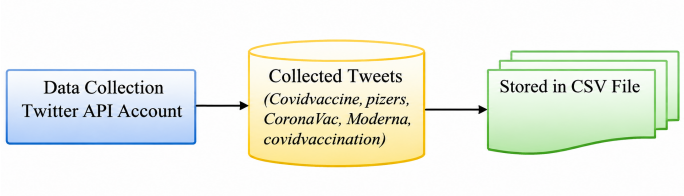


Figure 3. Data extraction.

component of supervised learning.

TextBlob can determine the polarity and subjectivity of a statement. The polarity range is [-1,1], with '-1' denoting negative emotion, '1' denoting positive emotion, and '0' denoting neutral [27]. TextBlob divides tweets into three polarities: (i) neutral, (ii) negative, and (iii) positive. Tweets with positive polarity expressed excitement at the arrival of COVID-19 vaccines in the nation, a desire to get vaccinated, or hope that vaccines would work. People who agreed with or supported the COVID-19 vaccination program were classified as positive polarity (polarity > 0). Users who were undecided about getting vaccinated, or who only reported

vaccine-related news, were classified as neutral (polarity $= 0$). Tweets with negative polarity (polarity < 0) were opposed to COVID-19 vaccination. Negative tweets showed negative emotions, debates, and opposition to immunization, as well as dissatisfaction with adverse side effects after vaccination. Some annotated tweets are shown in Table 2.

Table 2. Polarity and their tweets.

Labels	Tweets	Polarity
Positive	Lebanon on getting first vaccines in February. Zimbabwe gets the first batch '*Sinopharm*' jabs China. '*Pfizer*' jab effective symptomatic covid study.	>0
Negative	alleged making billions corona vaccine. access cost hesitation: (political) economics vaccine distribution. Portugal: nurse dies days getting Pfizer covid-vaccine	<0
Neutral	'*Pfizer-BioNTech*' covid vaccine approved for emergency use by the European Union drug regulator. Pakistan receives second batch covid vaccine soon. Israel re-opens restaurants and bars country fully vaccinated.	=0

3.3 Preprocessing Dataset

A text processing method under NLP is used to clean and prepare text for model building. Tweets contain noise in many forms, such as emoticons, punctuation marks, numbers, and special characters, so the preparation stage is critical when working with unstructured data [28]. Raw data must go through preprocessing before it can be used with ML algorithms [29], and the choice of preprocessing steps affects the performance of NB classifiers [30]. Standard NLP steps were applied [29], and all text was converted to lowercase. After that, duplicate, unrelated, and unimportant rows were removed. The search often returns tweets that are not relevant to the topic, such as advertising or news links (Table 3). These rows are unimportant for the proposed study, so they were removed from the dataset. Figure 4 gives a visual representation of the preprocessing method.

Figure 5 gives an overview of the SA. The X-axis shows the sentiment categories (positive, negative, and neutral), and the Y-axis shows the weight of each category as a bar graph. The neutral bar is the highest because about half of the tweets are neutral. The positive bar is the second highest, and the negative bar is the lowest.

Table 3. Unrelated Rows.

S.No.	Unrelated Rows Irrelevant Data Example
1	Adamu Garba to the Twitter family after the federal government of #nigeria (fgn). Create an account on koo instead of crowwe.
2	If you watch this video, you'll trade candlesticks like a pro. Save the date 12th June, and click to subscribe to the channel.
3	Environment ecology please click the below link and watch the video.

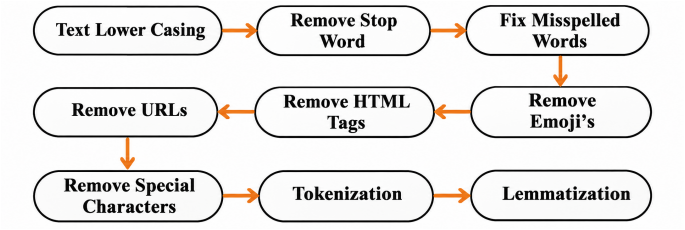


Figure 4. Data Preprocessing.

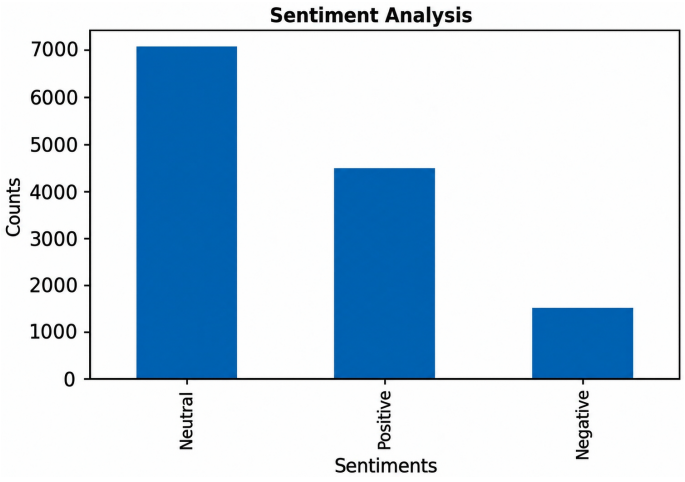


Figure 5. Overall sentiments.

3.4 Clustering

In this study, k-means clustering was used to cluster the sentiments by severity. The tweets were already classified into positive, negative, and neutral, but this classification does not show the severity of the tweets. Therefore, k-means clustering was applied to group the sentiments into high positive, low positive, high negative, and low negative. Most of the tweet clusters are close to the neutral side, which suggests that many people are doubtful: they neither fully agree nor fully disagree with vaccination. The clusters are described below and shown in Figures 6 and 7:

- **High positive**: these tweets show that people strongly agree with COVID-19 vaccination, so they need less awareness effort. This is a positive sign for vaccination. The high positive range is above

0.5 up to 1, as shown in Figure 6.

- **Low positive:** the polarity of these tweets is positive but low. People in this category need awareness about COVID-19 vaccination because they are close to neutral and could move towards high positive sentiment. More highly positive people are needed so that the world population can be fully vaccinated and the pandemic can end. The low positive range is above 0 up to 0.5, as shown in Figure 6.
- **High negative:** these people strongly disagree with COVID-19 vaccination. It is necessary to learn about their reasons, including their views on COVID-19 cases and deaths, and to find out why they are on the negative side. This knowledge helps to make them aware of the importance of vaccination and to move them towards the positive side, which is important for ending the pandemic. The high negative range is below -0.5 down to -1, as shown in Figure 6.
- **Low negative:** this cluster is also on the negative side, but these people need only a little effort to move towards the positive side. The low negative range is below 0 down to -0.5, as shown in Figure 6.
- **Neutral:** this cluster also needs awareness to move towards the positive side. Neutral people may move towards either side; a move towards the positive side is good, whereas a move towards the negative side is dangerous. Therefore, efforts are needed to move these people towards the positive side, as shown in Figure 6.

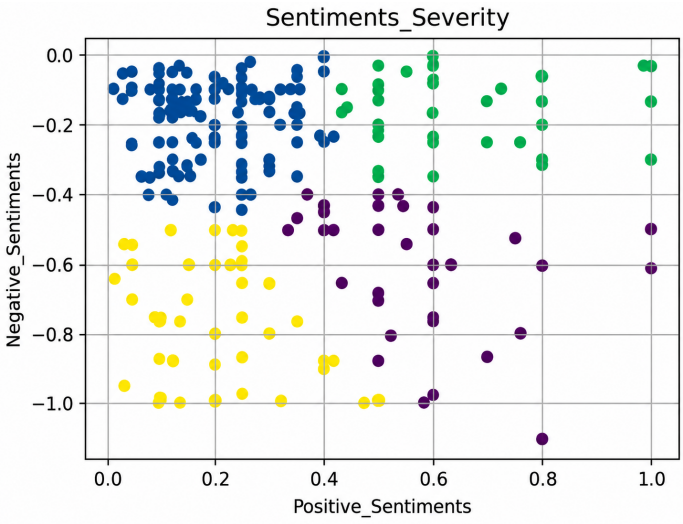


Figure 6. Clustering of sentiments on the basis of severity.

Figure 7 also shows that both sides of the data are close to neutral. The data up to 0.5 is higher than the data up to 1. Positive tweets account for about 37% in total (about 26% low positive and 10% high positive), and negative tweets account for about 13% in total (about 10% low negative and 2% high negative); the components are rounded and may not add up exactly to the totals. Severity corresponds to the polarity value, whereas the height of a bar in the histogram shows the density of tweets in that polarity range.

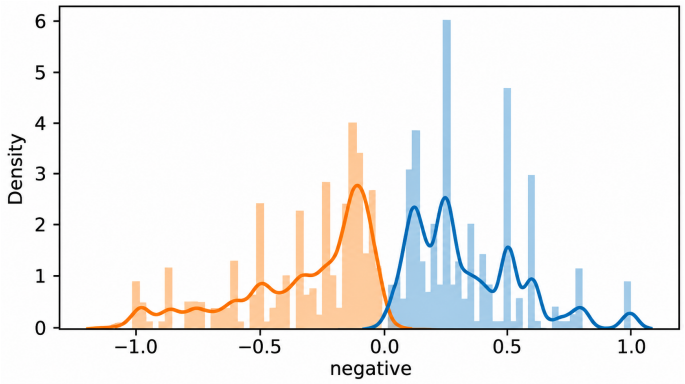


Figure 7. Histogram showing the clustering data.

4 Results and Discussion

This section presents and discusses the overall results of the proposed study. The results of five algorithms, namely Naïve Bayes (NB), support vector machine (SVM), multinomial NB (MNB), gradient boosting (GB), and random forest (RF), are compared in Section 5. In the overall dataset, 36.82% of the tweets were positive and 12.97% were negative. Some negative tweets mentioned side effects such as fever, chills, and dizziness, and a heavy and uncomfortable injection site.

Most of the tweets (50.21%) were neutral. This suggests that many people may still be in doubt about vaccination or may simply be sharing information. To prevent and control COVID-19, government institutions should urgently develop awareness programs to clear public doubts so that people can be vaccinated as early as possible. Sentiment analysis can help authorities plan awareness, control, and prevention measures for future pandemics.

5 Comparison of all Classifiers Accuracy

Table 4 lists the accuracy of five ML classifiers (NB, SVM, MNB, RF, and GB) on the same dataset. NB has the highest accuracy (90%) and is listed at the top of the table. Using several classifiers on the same dataset makes it possible to compare their performance. NB

achieved the highest accuracy, 90%, compared with SVM 87.1%, MNB 86.5%, GB 84%, and RF 81%. The F1-scores are 0.90 for NB, SVM, and MNB, 0.88 for GB, and 0.81 for RF, so NB, SVM, and MNB are comparable on this metric.

Table 4. Comparison of five classifiers.

Classifier Name	Accuracy	Precision (%)	Recall (%)	F1-Score
Naïve Bayes Classifier	**90%**	89.2	**89.5**	**0.90**
Support Vector Machine (SVM)	87.1%	88	88	0.90
Multinomial Naïve Bayes	86.5%	88	87	0.90
Gradient Boosting Classifier	84%	**90**	89	0.88
Random Forest	81%	87	82	0.81

NB is fast and performs well even on small datasets [10]. MNB reached a lower accuracy than NB in this study (Table 4). GB and RF reached lower accuracy than NB on this dataset (Table 4). Because the sentiment labels were derived from TextBlob, the reported accuracy measures agreement with the TextBlob labels rather than with manually annotated ground truth.

Figure 8 compares the accuracy of the five ML classifiers. NB has the highest accuracy (90%).

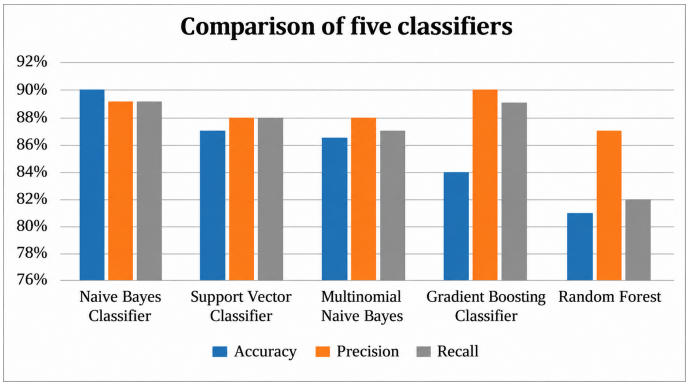


Figure 8. Accuracy of the five classifiers.

6 Conclusion

This study explored public opinion on COVID-19 vaccines. During the pandemic, people have frequently used social media to voice their opinions about COVID-19 and its vaccines, which makes Twitter one of the most important information sources. Tweets were collected through a Twitter API account. NLP tools and techniques were used to prepare the data for a supervised classification system. After preprocessing, polarity and subjectivity were computed with TextBlob, and a supervised NB algorithm was used to classify the polarity classes of the dataset. K-means clustering was used to form "High-positive", "Low-positive", "High-negative", and "Low-negative" clusters that indicate the severity of the tweets. Positive tweets account for about 37% in total (about 26% low positive and 10% high positive), and negative tweets account for about 13% in total (about 10% low negative and 2% high negative). The purpose of clustering was to find the severity of the tweets on both sides.

Five algorithms (NB, SVM, MNB, GB, and RF) were tested on the same dataset to compare accuracy, precision, recall, and F1-score and to compare their performance. NB achieved the highest accuracy, 90%, compared with SVM 87.1%, MNB 86.5%, GB 84%, and RF 81%. The F1-scores are 0.90 for NB, SVM, and MNB, 0.88 for GB, and 0.81 for RF. About half of the tweets (50.21%) are neutral, which may indicate that many people are still unsure whether to get vaccinated. Government institutions therefore need to develop awareness programs that clear public doubts, so that people can be vaccinated as early as possible. Sentiment analysis can help authorities plan such programs and the control and prevention of future pandemics.

Data Availability Statement

Data will be made available on request.

Funding

This research received no external funding.

Conflicts of Interest

Inayat Khan served as an Associate Editor of the *ICCK Transactions on Advanced Computing and Systems* at the time of manuscript submission. To ensure the integrity of the peer-review process, Inayat Khan was not involved in the editorial handling, peer review, or decision-making process for this manuscript, which was handled independently by another editor. The remaining author declares no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

This study used only the text of publicly available

tweets collected through the Twitter API in accordance with the platform's terms of service. No interaction with human participants took place and no private information was accessed. Usernames and other personal identifiers were removed, and example tweets are shown without user identifiers. Therefore, ethical approval and informed consent were not required.

References

[1] Jalil, Z., Abbasi, A., Javed, A. R., Badruddin Khan, M., Abul Hasanat, M. H., Malik, K. M., & Saudagar, A. K. J. (2022). COVID-19 related sentiment analysis using state-of-the-art machine learning and deep learning techniques. *Frontiers in public health, 9*, 812735. [CrossRef]

[2] Muric, G., Wu, Y., & Ferrara, E. (2021). COVID-19 vaccine hesitancy on social media: Building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. *JMIR Public Health and Surveillance, 7*(11), e30642. [CrossRef]

[3] Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. *Journal of Medical Internet Research, 22*(4), e19016. [CrossRef]

[4] Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. *Journal of medical Internet research, 22*(11), e20550. [CrossRef]

[5] Mahyoob, M., Algaraady, J., Alrahiali, M., & Alblwi, A. (2022). Sentiment analysis of public tweets towards the emergence of SARS-CoV-2 Omicron variant: A social media analytics framework. *Engineering, Technology & Applied Science Research, 12*(3), 8525–8531. [CrossRef]

[6] Sattar, N. S., & Arifuzzaman, S. (2021). COVID-19 vaccination awareness and aftermath: Public sentiment analysis on Twitter data and vaccinated population prediction in the USA. *Applied Sciences, 11*(13), 6128. [CrossRef]

[7] Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., & Sharif, S. (2021). An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases, 108*, 256–262. [CrossRef]

[8] Amjad, A., Qaiser, S., Anwar, A., & Ali, R. (2021, September). Analysing public sentiments regarding COVID-19 vaccines: a sentiment analysis approach. In *2021 IEEE international smart cities conference (ISC2)* (pp. 1-7). IEEE. [CrossRef]

[9] Pristiyono, Ritonga, M., Ihsan, M. A. A., Anjar, A., & Rambe, F. H. (2021, February). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1088, No. 1, p. 012045). IOP Publishing. [CrossRef]

[10] Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes. *Information, 12*(5), 204. [CrossRef]

[11] Barkur, G., Vibha, & Kamath, G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry, 51*, 102089. [CrossRef]

[12] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing, 97*, 106754. [CrossRef]

[13] Piedrahita-Valdés, H., Piedrahita-Castillo, D., Bermejo-Higuera, J., Guillem-Saiz, P., Bermejo-Higuera, J. R., Guillem-Saiz, J., Sicilia-Montalvo, J. A., & Machío-Regidor, F. (2021). Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines, 9*(1), 28. [CrossRef]

[14] Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. *IEEE Access, 8*, 181074–181090. [CrossRef]

[15] Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. *Procedia Computer Science, 87*, 44–49. [CrossRef]

[16] Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics, 9*(3), 483. [CrossRef]

[17] Ansari, M. T. J., & Khan, N. A. (2021). Worldwide COVID-19 vaccines sentiment analysis through Twitter content. *Electronic Journal of General Medicine, 18*(6), em329. [CrossRef]

[18] Salathé, M., & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS Computational Biology, 7*(10), e1002199. [CrossRef]

[19] Marcec, R., & Likic, R. (2022). Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal, 98*(1161), 544–550. [CrossRef]

[20] Lyu, J. C., Han, E. L., & Luli, G. K. (2021). COVID-19 vaccine-related discussion on Twitter: Topic modeling and sentiment analysis. *Journal of Medical Internet Research, 23*(6), e24435. [CrossRef]

[21] Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet

Allocation for topic modeling on Twitter. *PLoS ONE, 15*(9), e0239441. [CrossRef]

[22] Hung, M., Lauren, E., Hon, E. S., Birmingham, W. C., Xu, J., Su, S., ... & Lipsky, M. S. (2020). Social network analysis of COVID-19 sentiments: Application of artificial intelligence. *Journal of Medical Internet Research, 22*(8), e22590. [CrossRef]

[23] Samuel, J., Ali, G. G. M. N., Rahman, M. M., Esawi, E., & Samuel, Y. (2020). COVID-19 public sentiment insights and machine learning for tweets classification. *Information, 11*(6), 314. [CrossRef]

[24] Kaur, H., Ul Ahsaan, S., Alankar, B., & Chang, V. (2021). A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets. *Information Systems Frontiers, 23*(6), 1417–1429. [CrossRef]

[25] Pokharel, B. P. (2020). Twitter sentiment analysis during COVID-19 outbreak in Nepal. *SSRN Electronic Journal*. [CrossRef]

[26] Dubey, A. D. (2020). Twitter sentiment analysis during COVID-19 outbreak. *SSRN Electronic Journal*. [CrossRef]

[27] Loria, S. (n.d.). TextBlob: Simplified text processing [Software documentation]. Retrieved from https://textblob.readthedocs.io/

[28] Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science, 17*, 26–32. [CrossRef]

[29] Jianqiang, Z., & Xiaolin, G. (2017). Comparison research on text pre-processing methods on Twitter sentiment analysis. *IEEE Access, 5*, 2870–2879. [CrossRef]

[30] Palomino, M. A., & Aider, F. (2022). Evaluating the effectiveness of text pre-processing in sentiment analysis. *Applied Sciences, 12*(17), 8765. [CrossRef]



Maaz Khan received the BS degree in computer science from the University of Peshawar and the MS degree in computer science from the Institute of Management Sciences, Peshawar. His research interests include data mining, text mining, and sentiment analysis. (Email: maazkhanszic@gmail.com)