

# MACHINE LEARNING-BASED SENTIMENT ANALYSIS OF TWITTER COVID-19 VACCINATION

## Abstract

The COVID-19 pandemic has caused significant fear, anxiety, and complex emotions or feelings in a large number of people. A global vaccination campaign to end the SARS-CoV-2 epidemic is now in progress. People's feelings have become more complex and varied since the introduction of vaccinations against coronavirus. The use of social media platforms such as Twitter enables users to communicate with one another and share information and perspectives on a wide variety of topics, spanning from local to international concerns, from global to personal. Twitter will prove to be a helpful source of information that can be tracked regarding views and sentiments regarding the SARS-CoV-2 vaccination. To better understand public views, concerns, and emotions that may influence the achievement of herd immunity targets and limit the pandemic's impact, this study uses deep learning to identify the themes and sentiments in the public about COVID-19 immunization on Twitter. Moreover, this paper consists of a detailed explanation of the sentiment analysis with their challenges, classification, approaches, applications, and VADER.

**Keywords:** COVID-19, Vaccination, Sentiment Analysis, Opinion Mining, Social Media, Twitter, VADER, Machine-learning.

## Authors

### **Santosh Kumar**

Associate Professor

Department of Computer Science and Engineering

Arya College of Engineering and IT  
Jaipur, Rajasthan, India.

### **Dr. Vishal Shrivastava**

Professor

Department of Computer Science and Engineering

Arya College of Engineering and IT  
Jaipur, Rajasthan, India.

### **Satish Chandra Sudhansu**

M.Tech. Scholar

Department of Computer Science and Engineering,

Arya College of Engineering and IT  
Jaipur, Rajasthan, India.

## I. INTRODUCTION

The current COVID-19 outbreak has had significant repercussions for the healthcare industry, and as a direct consequence, our fundamental understanding of safety has been disrupted. Isolation from others has the potential to halt or significantly delay the propagation of the corona virus. At this time, it is necessary to take precautions such as washing your hands frequently, using a mask, and avoiding close personal contact as much as possible. However, these can only reduce the risk of transmitting the corona virus; they cannot eliminate the risk. In light of the current circumstances, vaccination has emerged as the only strategy capable of battling and possibly wiping out corona virus. Pfizer's early research on mRNA vaccines involved more than 40,000 participants, while a more recent immunization study included 30,000 participants. Together, these numbers represent the number of persons who received the vaccine. In both studies, the efficiency of the vaccine was measured at 94% on average, and there was not a single fatality associated with either study. Another viral vector-based vaccine called Johnson & Johnson is efficient against corona virus which increases response of the immune system of those who get it. It has a success rate of 85% and there are no noticeable bad effects associated with getting it [1][2].

A significant number of analysts have made use of ML techniques in order to investigate how people talk about COVID-19 vaccinations online. The rise of social media may be one contributor to the decline in the number of people ready to be vaccinated. Analyzing the messages that were communicated to the public in this context reveals the public's opinion on the matter of COVID-19 immunizations. At this time, if it has been determined that 90% of vaccinations are successful, then immunizations will be done starting in the U.K. from December 8, 2020. In this study, tweets pertaining to vaccines were analyzed in order to have a better understanding of their effect. In this study, Twitter was used for increasing vaccination compliance, decreasing vaccination reluctance and resistance, and enhancing vaccination acceptability [3]. It is possible that the spread of misinformation about vaccinations may be mitigated if public health workers had a more nuanced awareness of the perspectives and attitudes around the topic. Authorities could use Twitter to actively promoting the use of vaccines among the general public while reducing vaccine hesitancy the general public. This would be a way of influencing people's attitudes around vaccination. People's views and beliefs regarding the matter were altered as a result of various attacks made against vaccines during the outbreak.

The development and approval of the COVID-19 immunization have given people renewed optimism that the pandemic can be put an end to and that normal life can be resumed. Unfortunatously, there is a substantial obstacle in the way of getting vaccination rates, and that obstacle is hesitation over-vaccination, which is occasionally motivated by misinformation. One of the most difficult components of machine learning is processing data in a way that detects emotions using methods that allow us to evaluate whether people have positive or negative perspectives on a topic. Although social media and micro blogging sites are excellent sources of information, their primary function is to allow users to communicate thoughts and beliefs that are uniquely their own.

## II. BACKGROUND STUDY OF COVID-19

The 2019 corona virus disease (COVID-19) outbreak was initially discovered in Wuhan, China in December of this year. It then rapidly grew to become a global epidemic that affected millions of people worldwide. The novel severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has been attributed to COVID-19, and clinical manifestations of the virus have ranged from asymptomatic and mild symptoms related to flu to pneumonia, acute respiratory distress syndrome (ARDS), and in some cases, death. COVID-19 is associated with SARS-CoV-2. It is hoped that the spread of the virus could be contained by social isolation, the use of masks, the development of novel antiviral medications, and the production of an efficient vaccine. Establishing herd immunity through natural immunity through diseases is possible, but doing so could have catastrophic consequences, as was seen in Sweden, where authorities believed that infecting up to 60% of the population would be enough to protect the more vulnerable population through herd immunity. This plan backfired; however, as there are at least five times as many COVID-19-related fatalities per million people in Sweden than there are in Germany. As a result, the production of an efficient vaccine is of the utmost importance and is regarded as the only viable option for achieving herd immunity.

Researchers from all around the world are actively toiling away around the clock in an effort to produce a vaccine that is effective against COVID-19. There are now around 200 potential vaccines that are in the process of going through various stages of development. These vaccines include the AZD1222 vaccine developed by AstraZeneca and Oxford and the mRNA-1273 vaccine developed by Moderna. There are presently 30 vaccines being tested in clinical studies. Although it is possible that production capacity will not be sufficient for meeting the global demand for vaccines, it would be advantageous if a small selection of vaccines were available for emergency use for population sections that are more vulnerable. The goal is to achieve global vaccine distribution for stopping and limiting the impact of Covid-19.

In order to successfully create a vaccine that is both safe and efficient, it is essential that all stages of trials and testing be carried out with extreme caution to prevent serious adverse effects. Accelerating COVID-19 Therapeutic Interventions and Vaccinations (ACTIV), the Gavi alliance, the World Health Organization (WHO), and the Bill & Melinda Gates Foundation (BMGF) must collaborate to secure appropriate financing for vaccines and a coordinated response to the ongoing COVID-19 epidemic. This review provides a summary of the immune response and biology demonstrated by previous corona virus infections and SARS-CoV-2, the impact of Twitter as an information-providing service, and the potential problems that may occur as a result of speeding up the production of vaccines. The development of vaccines could take around 15 years, but with advanced technologies and the urgent need for vaccines, the development time could be reduced to one and half years or less or less. This could potentially raise issues regarding safety and efficacy affecting public acceptance of vaccines.

**1. Application Scenarios on COVID-19 Data:** As a result of having to deal with the aftermath of COVID-19, many people are experiencing a wide range of mental health problems. During COVID-19, many researchers worked to analyse public opinion[4].

- **Mental Health Analysis of Students during the Lockdown:** The practice of social distance, which resulted in less encounters between people, was implemented in an effort to halt the progression of COVID. A number of nations went into lockdown, which included closing their airspace as well as educational and other institutions. As a result of the lockdown, people, particularly students, had to remain a significant distance from their houses, be confined within their dormitories, and cease their educational activities. This results in students experiencing anxiety and tension. During the lockdown, students expressed their feelings via social media platforms, and researchers attempted to investigate those feelings. Data from Twitter was analyzed in order to gain a better understanding of those feelings.
- **Reopening after COVID-19:** As a result of the corona virus, billions of individuals all over the world have been affected in some way. It has generated economic turmoil all around the world which is a roadblock to reopening. The permanent stagnation of the economy poses a risk to the continued existence of any nation. People are being forced to reopen companies and get back to living their usual lives as a result of these factors. The researchers focused their efforts on determining what kinds of businesses people are considering reopening after COVID-19.
- **Restaurant Reviews:** Customers have the ability to voice their opinions and provide feedback regarding the quality of the products or services provided by a variety of companies in today's digital world. These reviews are beneficial to other consumers who are about to utilize the service or purchase the goods since they assist them make judgments. The internet reviews are tied to the overall star rating, which in turn influences the amount of money that the restaurant makes. People were especially concerned about the spread of COVID during COVID, so special SOPs were announced for eateries during COVID. As a result, numerous eateries received poor evaluations due to their chilly outdoor areas and their delayed service. Researchers studied the comments that individuals had made about restaurants, which assisted restaurant management in preserving the high quality of both the cuisine and the atmosphere.
- **Racial Sentiments and Vaccine Sentiments:** The production of a vaccine against COVID may prove useful in stopping the disease's further spread. As a result, a great number of industries are putting their efforts into developing various types of vaccines. However, the key necessity in order to reduce COVID with vaccines is for people to be willing to accept and take their vaccinations. In the event that people are unwilling to get themselves, there will be a significant obstacle in the way of the control of COVID. In, researchers investigated the opinions of the general public regarding vaccinations. Additionally, COVID was responsible for feelings of discrimination across international borders, which led to an increase in people's racist attitudes.

### III. SENTIMENT ANALYSIS

Opinion mining and sentiment analysis are both terms that refer to the same thing: the automated computer study of people's attitudes, emotions, and expressions in relation to a certain objective. Any person, event, or subject matter could be taken as the target object. Analysis of sentiment and opinion mining are interchangeable phrases that can be used in the same context. However, according to the findings of some studies, these two terms refer to somewhat distinct mental images. Sentiment analysis involves identifying and analysing the emotion conveyed in a piece of text, while opinion mining involves obtaining and analysing

people's opinions on an entity. Therefore, the purpose of sentiment analysis is to automate the process of discovering opinions, determining the emotions that those opinions represent, and then categorizing the polarity of those emotions. In many different spheres, the consideration of public opinion is an absolutely necessary step in the decision-making process. Before purchasing a certain item, a person who is interested in making a purchase could find it helpful to inquire about other people's experiences with the item in question. In the real world, companies and organizations often solicit client opinions on the quality of their goods and services. In recent years, applications of sentiment analysis have grown over a wide variety of fields, including ad placements, trend prediction, and recommendation systems as well as politics and healthcare. The past several years have seen a meteoric rise in the prevalence of social media on the internet, such as online reviews, comments, forums, blogs, and comments on social networking sites. The majority of organizations are basing their choices on the contents of these reports. As a result of the vast amounts of data that are readily available to the public today, modern enterprises no longer need to rely on opinion polls, surveys, or focus groups. The necessity of checking each unique website makes the work of mining opinions a challenging and difficult one. Finding the relevant websites and gleaning the opinions contained within them can be an extremely challenging task for a human reader.

As a result, automated sentiment analysis is something that is desperately needed. The majority of companies are relying on their own in-house research and analytic systems to learn what their customers think. Opinion mining and sentiment analysis are typically carried out utilizing one of two methods: 1) An approach based on machine learning 2) Derived from a lexicon. The strategy that is based on machine learning makes use of a number of different supervised and unsupervised learning algorithms in order to classify sentiment. For the purpose of sentiment classification, lexicon-based algorithms make use of a dictionary containing words connected to specific domains that convey a range of emotions. It is possible to learn whether a certain term is positively or negatively connoted by consulting a dictionary and detecting the polarity of the words by comparing the word in the sentence with the words in the dictionary.

The most important part of opinion mining is identifying the type of sentence. Sentences have to be classified either subjective or objective. Researchers are using both supervised and unsupervised learning techniques for providing different methods of sentiment analysis. In general, —Sentiment Analysis includes advanced processes. It has a totally different set of tasks: Subjective or objective analysis, opinion extraction, sentiment classification (supervised or unsupervised). Labelling any text document or sentence as subjective or objective can be done using subject-level analysis. Sentiment classification includes probing the sentiment polarity of the filtered sentences. All sentences or texts are divided into positive, negative, or neutral types depending on the emotions we get from the texts or sentences[5].

- 1. Valence Aware Dictionary for Sentiment Reasoning (VADER):** The NLTK module VADER (Valence Aware Dictionary for Sentiment Reasoning) generates sentiment scores from the words in a document. It's a sentiment analyzer based on rules, where words are assigned positive or negative labels based on their semantic orientation. The VADER method is built on lexicons and makes use of gold-standard heuristics in addition to English-language sentiment lexicons. Lexicographies are subjected to human inspection

and verification. They make use of qualitative methods in order to increase the effectiveness of the emotion analyzer [6]. The VADER corpus is the result of pooling together several different data sets. The polarity of the emotions was provided by the initial corpus, in contrast to VADER, which contains an additional element that shows the intensity of that polarity score. Its corpus has over 7500 dictionaries, including slang and abbreviations. Scores might range from -4.0 to +4.0. These values indicate an attitude threshold, with scores below 4 suggesting negative sentiments and scores over +4 indicating good sentiments.

VADER uses grammatical and syntactic rules in addition to a sentiment vocabulary to indicate the severity and polarity of the sentiments being expressed. Utilizing a wide variety of language features, such as emoticons and acronyms, the VADER lexicon comprises over 7500 sentiment qualities. Since the emotional weight of a word is established by taking grammatical constraints into account, word sentiment scores might vary.

- 2. Related Work:** COVID-19 vaccine topic from Twitter was studied in [9]. It was found that people had different feelings about the Chinese vaccine compared to those about vaccines made in other countries, and the value of those feelings could be affected by the number of deaths and cases reported in the daily news as well as the nature of the most pressing problems in the communication network.

The positive appeals of recent news on the safety of the COVID-19 vaccination and the government's proactive risk communication were reflected in the finding that positive views outlasted negative feelings for 56 days (62.20%), was found in [10]. There was also a considerable correlation between positive vaccination attitudes and rising vaccination rates, which was statistically significant.

In this study, Qorib et al. [11] the reluctance over the COVID-19 vaccine is analysed using three different approaches of emotion computation: Azure Machine Learning, VADER, and TextBlob. demonstrates that people's resistance to the COVID-19 vaccination lessens over time, which suggests that the general population may eventually feel more optimistic and happy about becoming vaccinated against COVID-19..

Yousefinaghani et al. [3] shown that there is a difference, albeit a little one, between the frequency of positive and negative emotions, with the former being the more common polarity and eliciting more responses. More time was spent talking about people's fears and reservations about vaccinations than actually learning about them, according to the study's findings. It found that some anti-vaccination accounts were run by Twitter bots or political activists, whereas pro-vaccination accounts tended to be associated with more authoritative figures or organizations.

In [12] when compared to states in other regions of the United States, states in the South showed a much higher incidence of negative tweets, but states with higher incomes reported a lower prevalence of negative tweets. Due to the fact that our data indicate the existence of negative vaccine attitudes as well as geographic variation in these opinions, it is necessary to customize our efforts to promote vaccinations, particularly in the southern portion of the United States.

Bokaee Nezhad and Deihimi[13]revealed a statistically significant difference (although a little one) in the amount of people in Iran who had a favourable view of domestic and imported vaccines, with the latter having the more prevalent positive polarity. The number of people who are worried about vaccines, both at home and abroad, has increased noticeably in recent months. Conclusions There were no statistically significant differences in the percentage of Iranians who had a positive and unfavorable perspective about immunization.

By analyzing the tone of 2,678,372 tweets on the COVID-19 vaccine posted between November 1, 2020 and January 31, 2021, researchers in [14] found that 42.8 percent were positive and 30.3 percent were negative. The public's mood and the number of tweets both spiked after Pfizer announced that the first COVID-19 immunization had attained 90% efficacy, and both continued to rise until the end of December, when they finally settled at a neutral emotion. Furthermore, people's perspectives varied depending on where they were located.

**1. Proposed Methodology:** This section provides an explanation of the processes that were utilised to examine the efficacy of the suggested machine-learning strategy in the classification of the sentiments contained within COVID-19 tweets. In the present investigation, we created a model based on machine learning for analysing the sentiment of Vaccination Responses in reaction to an increase in the number of newly reported cases of COVID-19. We utilised an open-source dataset titled "All COVID-19 Vaccines Tweets," which contained each and every tweet ever sent out regarding the COVID-19 vaccine. The collection contains all COVID-19 vaccine-related tweets ever sent anywhere in the world. There was feeling behind every tweet, whether or not it was pro-vaccine. Our initial task was to determine the overall sentiment polarity of all the tweets. The positive, negative, or neutral nature of a tweet is indicated by the polarity of its expression. What the tweeter really wants to say is made clear here. With this data, we can assess the global effect of the COVID-19 vaccine. To determine the general tone of the tweets, we coded a Python tool. We've been using the Python Tweepy package for this purpose. Tweepy is a free and easy-to-use module for Python that acts as a gateway between your Python app and Twitter's API. This was used to collect the polarity values of the emotions. Now, the polarity value of emotion can be either zero, a positive number, or a negative number. We only considered tweets that expressed either good or negative emotions in our analysis. After carrying out the aforementioned steps, we had a dataset that could be used by our algorithms. When processing the data, we used the train-test split function, which is a key part of the machine-learning method. Train and test data were split 75%:25%. Reality's text data is a mess. Consequently, certain pre-processing operations have to be executed prior to feeding the dataset into the ML algorithms. Data pre-processing includes operations such as stemming, stop-word removal, tokenization, de-tokenization, URL removal, punctuation removal, URL removal, removal of incidents, removal of double spaces, and so on. All of these steps are essential components of the data preparation method. Then, analyse the results with VADER sentiment. Three different machine learning algorithms (NB, LR, and VC) were utilised to analyse the data.

All of the processes involved in the study approach are unmasked and discussed in this part very briefly:

- **Data Gathering:** Collecting relevant data is the starting point for the planned work. The "COVID-19 All Vaccines Tweets" dataset was used throughout this study.
- **Data Pre-processing:** The quality of the data pre-processing used in constructing an ML model is directly related to the effectiveness of that model.

The NLTK (Natural Language Toolkit) Python package was used in this part to prepare the text. The text can be pre-processed in a number of different ways. Reduced text size, elimination of URLs, removal of punctuation, tokenization, stop-word removal, and stemming are all pre-processing techniques.

- **Sentiment Analysis:** Opinions, attitudes, emotions, and views can be automatically mined from audio, text, tweets, and database sources via a process called Sentiment Analysis (SA), which employs NLP (Natural Language Processing). In a SA, the text is parsed for positive, negative, and neutral sentiments. Other names for SA include opinion mining, assessment extraction, and evaluation extraction. Though they are often used interchangeably, "opinion," "sentiment," "view," and "belief" each have distinct meanings [13].
2. **VADER (Valence Aware Dictionary for Sentiment Reasoning):** VADER is a lexicon-based method that makes use of heuristics considered to be the gold standard in the field, as well as lexicons of sentiment expressed in the English language. Lexicons are checked and verified by humans. To improve the efficacy of the emotion analyzer, they employ qualitative methods [14]. [15] According to [15], VADER can produce sentiment analysis results that are on par with those produced by human raters. The VADER corpus is the result of pooling together several different data sets. The polarity of the emotions was provided by the initial corpus, in contrast to VADER, which contains an additional element that shows the intensity of that polarity score. Its corpus has about 7500 lexicons, the sum of which includes slang and abbreviations. The possible range of scores is from -4.0 (being the lowest) to 4.0 (being the highest). Scores below -4 indicate negative sentiments, and scores above +4 indicate positive sentiments, hence these values serve as a threshold for attitudes. The results from the VADER look like this (neg, neu, pos, compound). In this situation, the compound score is calculated as the average lexical score of the entire text or a single sentence, and can vary from 0 (no score) to 1 (perfect score).
- **Dataset Splitting:** After performing above procedures, the data is split in two sets for training and testing in the ratio of 75% and 25% respectively.
  - **Classification Technique:** To classify unlabeled data, ML has produced a number of classification methods that make use of different approaches. In this research, we employed ML-based classification strategies as NB, LR, and Voting Classifier.



## IV. RESULTS ANALYSIS AND DISCUSSIONS

In this section, the results of the analysis that was done in the previous section are presented, and then an assessment of the performance is made. Python, a general-purpose, open-source programming language and interactive platform used for data analysis, scientific visualisation, and scientific computation, was utilised for the simulation. Tensorflow, SciPy, NumPy, Pandas, Matplotlib, SciKit-learn, Keras, Pytorch, Scrappy, and Theano are only some of the excellent Python libraries for data science that were utilised.

**1. Performance Evaluation Metrics:** Measures of how well a classifier performs are controversial, with no one method currently dominating the field. Its effectiveness is measured with the Confusion Matrix, as well as precision, accuracy, and recall. We will then go on to explore the various metrics used to assess the classifier's performance.

**2. Confusion Matrix:** The capacity of a classifier to discriminate between tuple instances of different classes can be evaluated using a confusion matrix. It keeps track of both the actual and expected categories generated by a given method. Matrix data is frequently employed in evaluating the efficiency of such systems.

- **Accuracy:** The ratio of accurate predictions to total predictions is a measure of accuracy. The accuracy of a classifier can be evaluated using the matrix that was shown earlier, as is shown down below:

$$Accuracy = \frac{TP}{TP + FP + FN + TN} \dots (4.1)$$

- **Precision:** Precision is utilised to circumvent the constraint of Accuracy. The precision indicates the proportion of positive predictions that were accurate.

$$Precision = \frac{TP}{TP + FP} \dots (4.2)$$

- **Recall:** Recall seeks to assess the proportion of actual positives that were wrongly detected.

$$Recall = \frac{TP}{TP + FN} \dots (4.3)$$

- **F1 Score:** The F1 score is calculated by finding the harmonic mean of the recall and precision scores.

$$F1\ Score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \dots (4.4)$$

**3. Simulation Results and Discussion:** The results of our tests and an evaluation of our methodology are presented and analysed in this section. We begin by contrasting the efficacy of several approaches to sentiment analysis using information collected from the hashtag #vaccines on Twitter. In this part, we examine the accuracy of Naive Bayes, a voting classifier, and a logistic regression classifier on the COVID-19 Vaccines Twitter dataset after the training phase is complete, using the tweets as examples (2). Here, we present the classification results achieved by various techniques.

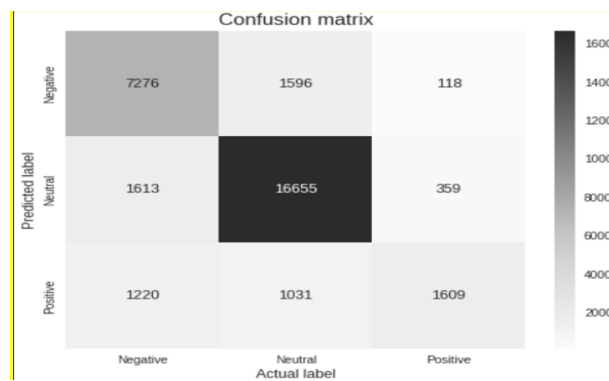
**4. Results of Naïve Bayes Classifier:** These are the outcomes of running the Naive Bayes algorithm on the sample data:

	precision	recall	f1-score	support
0	0.72	0.81	0.76	8990
1	0.86	0.89	0.88	18627
2	0.77	0.42	0.54	3860
accuracy			0.81	31477
macro avg	0.78	0.71	0.73	31477
weighted avg	0.81	0.81	0.80	31477

Accuracy : 0.8114

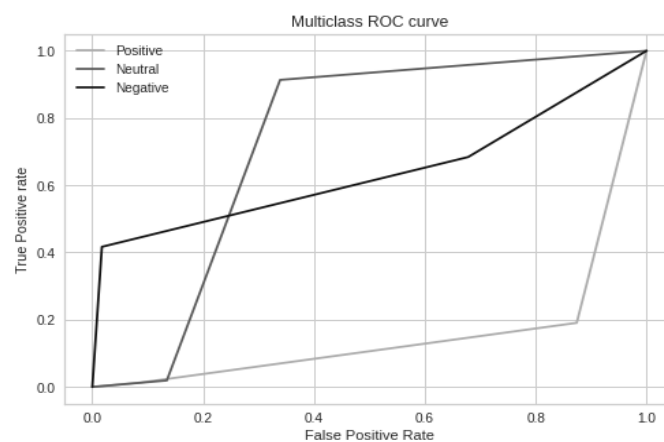
**Figure 1: NBC Classification Report**

The image above is a report generated by NB Machin's learning classifier while using the covid vaccine dataset for classification purposes. Do the NBCs in three different areas of figure drawing? With a f1-score of 72%, 81%, and 76% for positive data, 86%, 89%, and 88% for neutral data, and 77%, 42%, and 54% for negative data, the overall NBC accuracy is 81%.



**Figure 2 : NBC Confusion Matrix**

Figure 2 presents the NBC confusion matrix. In this graph, the actual label from the dataset is displayed along the x-axis, while the predicted label is displayed along the y-axis. The NBC model is 81 percent accurate. As an illustration, 1609 of the immunisation records are considered positive, 16650 are classified as neutral, and 7276 are considered negative.



**Figure 3 : NBC Multi-class ROC Curve**

The Naive Bayes machine learning classifier's multiclass ROC curve is depicted in Figure 3. Methods exist, such as the Receiver Operating Characteristics (ROC) graph, for organising classifiers and demonstrating algorithmic efficiency on training data. values of positive, neutral, and negative R-squared correlations. The largest ROC Area indicates that the POSITIVE class performs better than the other classes in the dataset, so we can conclude that this data set is positively biased. Improved performance and accuracy were also achieved while classifying weighted values for ROC area data determined using Naive Bayes.

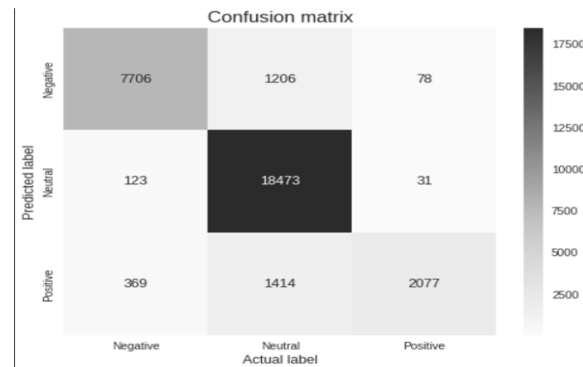
- **Results of Proposed Voting Classifier:** Results of sentiment analysis on the dataset may be predicted with 90% accuracy using a Voting classification method.

	precision	recall	f1-score	support
0	0.94	0.86	0.90	8990
1	0.88	0.99	0.93	18627
2	0.95	0.54	0.69	3860
accuracy			0.90	31477
macro avg	0.92	0.80	0.84	31477
weighted avg	0.90	0.90	0.89	31477

Accuracy : 0.8977

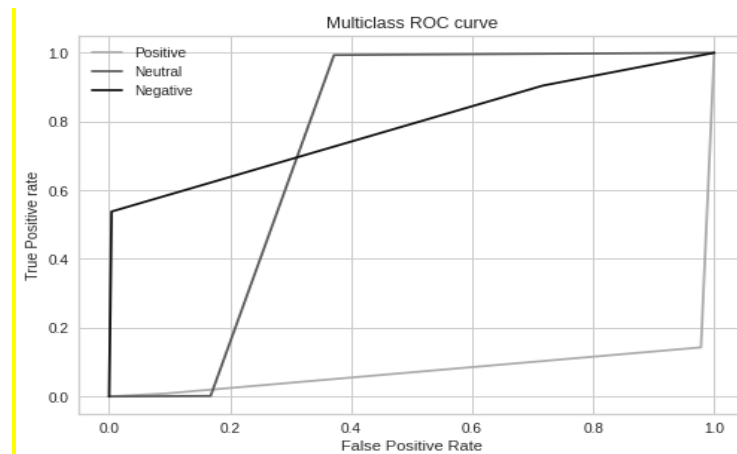
**Figure 4:** Voting Classifier Classification Report

Figure 4 above is a report generated by Voting Machin's learning classifier on the covid vaccine dataset. There are three different labels displayed for the findings. Precision, recall, and f1-score for the positive data, labelled as 0, are 94%, 86%, and 90%, respectively; for the neutral data, labelled as 1, they are 88%, 99%, and 93%; and for the negative data, labelled as 2, they are 95%, 54%, and 69%, for an overall VC accuracy of 90%.



**Figure 5:** Voting Classifier Confusion Matrix

A representation of the confusion matrix for the Voting classifier is shown in Figure 5. The x-axis shows the actual label from the dataset, while the y-axis shows the predicted label. The image depicts 2077 successful vaccination records, 1847 neutral label records, and 7706 failure immunisation instances with an accuracy of 90%.



**Figure 6:** Voting Classifier Multiclass ROC curve

Multiclass ROC curve of voting machine learning classifier is shown in Figure 6. On the x-axis of the graph, is the percentage of false positives, and on the y-axis, the percentage of true positives in the data set is displayed.

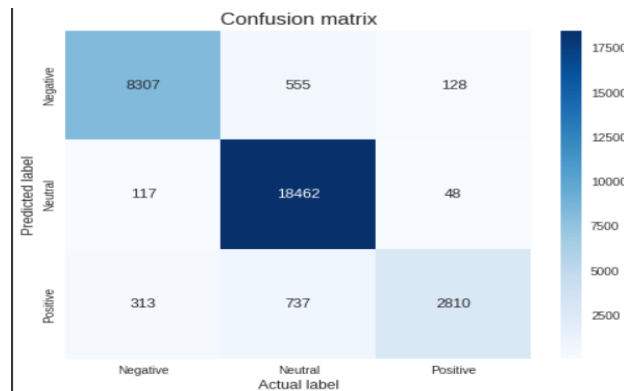
- **Results of Proposed Logistics Regression Classifier:** The dataset is subjected to a third analysis, this time using a Logistic Classifier. To conduct sentiment analysis on Twitter data, we apply this method to the dataset and find that it yields 94% accurate results. Positivity, negativity, and a neutral 0 are all valid values for expressing probabilities.

	precision	recall	f1-score	support
0	0.95	0.92	0.94	8990
1	0.93	0.99	0.96	18627
2	0.94	0.73	0.82	3860
accuracy			0.94	31477
macro avg	0.94	0.88	0.91	31477
weighted avg	0.94	0.94	0.94	31477

Accuracy : 0.9397

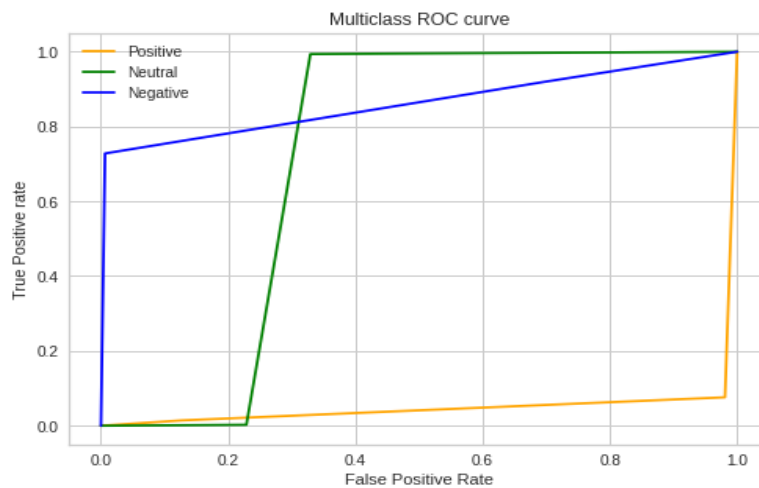
**Figure 7:** Logistic Regression Classification Report

The classification produced by the LR Machine Learning classifier while employing the covid vaccination dataset is shown in Figure 7. Three different types of results are displayed. Precision, recall, and f1-score for the positive data (labelled as 0) are 95%, 92%, and 94%, respectively; for the neutral data (labelled as 1), they are 93%, 99%, and 96%; and for the negative data (labelled as 2), they are 94%, 73%, and 82%, for an overall LR accuracy of 94%.



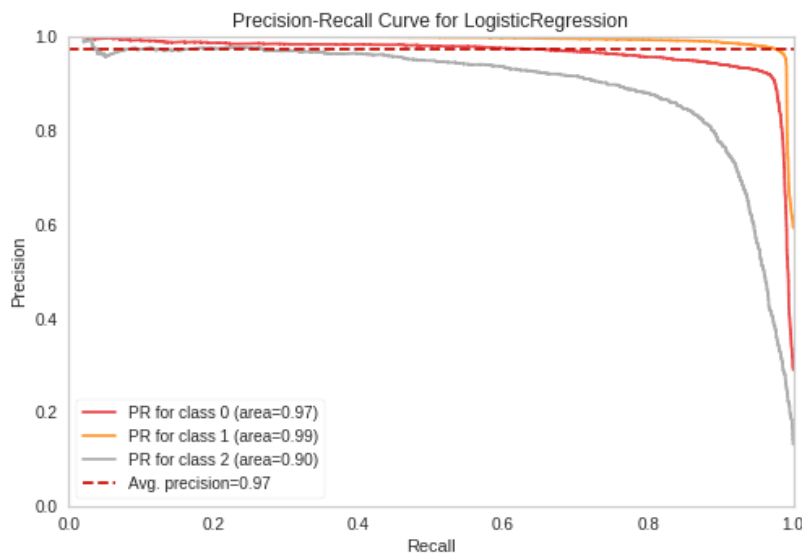
**Figure 8:** LRC Confusion Matrix

The LRC confusion matrix is shown in Figure 8. In terms of making forecasts, the LRC model has a 94% accuracy. There are favourable tweets about vaccinations in 2810 records, no opinions in 18462, and negative tweets about vaccinations in 8307.



**Figure 9:** LRC Multi-class ROC curve

The LR machine learning classifier's multiclass ROC curve is displayed in Figure 9. The x-axis of the graph shows the percentage of false positives while the y-axis shows the percentage of real positives. in the data set.

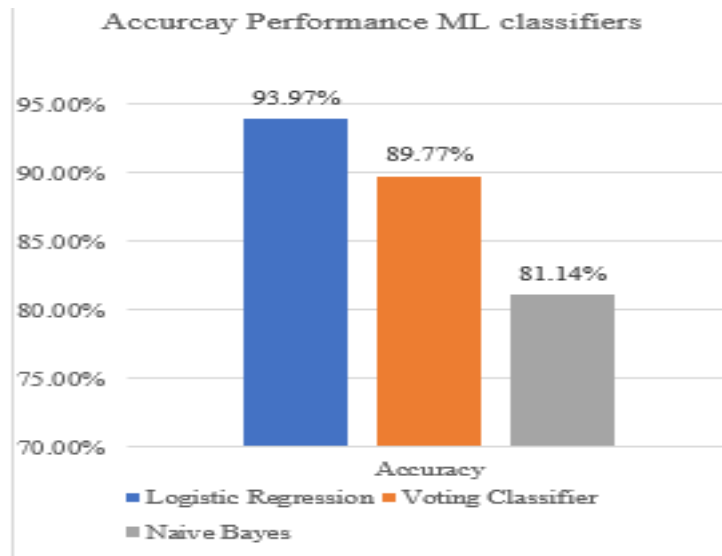


**Figure 10:** LRC PR curve

Figure 10 displays the PR curve for the LRC model, which outperforms the other models. The correlation between precision and recall is depicted by a straight line called a Precision-Recall (PR) curve. It can also be written as:  $TP/(TP+FN)$  on the y-axis,  $TP/(TP+FP)$  on the x-axis, and so on for the entire PR curve. Besides its more common name, "positive predictive value," "precision" describes how likely something is to be correct (PPV). The figure shows a PR curve of 0.97 percent for the positive class, 0.99 percent for the neutral class, 0.90 percent for the negative review class, and 0.97 percent for the over-average precision of the LR model on calls.

**Table 1: Machine learning classifier performance in terms of Accuracy**

Classifier	Accuracy
Logistic Regression	93.97 %
Voting Classifier	89.77 %
Naive Bayes	81.14 %



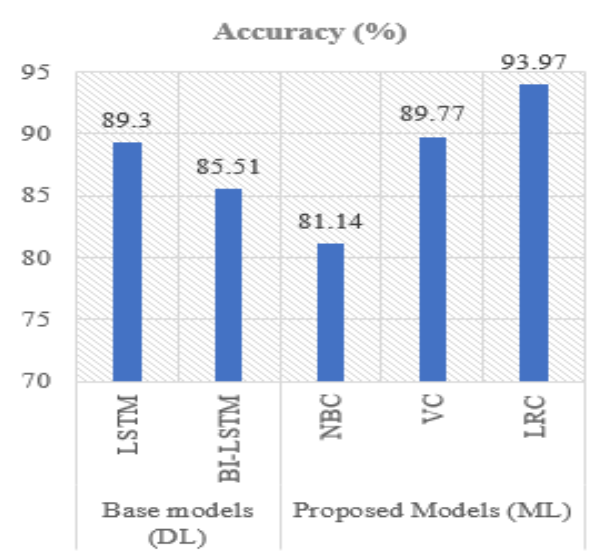
**Figure 11:** Bar Graph of Proposed model's accuracy

The accuracy of the categorization findings is displayed in Figure 11. The proposed Nave Bayes model achieves an accuracy of 81.14 percent, the voting classifier achieves an accuracy of 89.77 percent, and the third-best logistic regression model achieves an accuracy of 93.94 percent.

- **Comparative Results**

**Table 2 :** Comparative analysis of base and propose models

Parameter performance	Base models (DL)		Proposed Models (ML)		
	LSTM	BI-LSTM	NBC	VC	LRC
Accuracy (%)	89.30	85.51	81.14	89.77	93.97



**Figure 12:** Bar Graph of Accuracy Comparison between Base and Proposed Models'

Figure 12 compares the original LSTM (89.30%) and Bi-LSTM (85.51%) model to the selected Naive Bayes (81.14%), VC (89.77%), and LR (93.97%) suggested models for evaluating the vaccine's sentiment analysis. In light of Twitter data, it seems that the former is the optimal approach. The logistic regression model excels at classification compared to other methods (93.97 percent).

## V. CONCLUSION AND FUTURE SCOPE

The global SARS-CoV-2 corona virus disease pandemic (COVID-19) poses a serious risk to public health. The pandemic has, without a doubt, changed the way we look at the world. A number of people who have received the COVID-19 vaccine have resorted to Twitter to discuss their experience. We offer a tool that can analyse Twitter data for sentiment, which can then be used in research.

To determine the user's perspective on ML, we analysed public tweets on COVID-19 vaccines using Machine Learning. Our study shows that machine learning techniques can be successfully used for sentiment analysis tasks. Simple natural language processing (NLP)-based sentiment analysis techniques were developed using the positive, negative, and neutral emotion polarities. Few machine learning (ML) algorithms exist now. Using a voting classifier, logistic regression, and naive Bayes, we evaluated the accuracy of our predictions and analyses. The results of the network visualisation demonstrate that in order to combat the infodemic and increase vaccination rates, local-government health organisations and healthcare professionals need to be aware of the current state-of-the-art techniques in applying sentiment analysis. All the models had excellent F-1 scores, confusion matrices, precision, recall and precision; the Logistics Regression scored 93.97%, the voting classifier scored 89.77% and the NB Classifier scored 81.14 percent. Large numbers of people have decided to get vaccinated, but there are still many who are reluctant to do so. Some because they are either unsure of the process, terrified of needles, or both.



Medical researchers will gain insight from this study as they learn more about the difficulties of the immunisation process. A clear image of the vaccines' efficacy can be obtained by vaccine producers, health ministries, governments across nations, and agencies such as World Health Organization. Those involved will have a clearer picture of what has to be done to restore faith in immunisations. We hope that our efforts, however small, can help frontline workers in the fight against this novel corona virus and save lives.

## REFERENCES

- [1] K. N. Alam et al., "Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data," *Comput. Math. Methods Med.*, 2021, doi: 10.1155/2021/4321131.
- [2] T. M. Lincoln et al., "Taking a machine learning approach to optimize prediction of vaccine hesitancy in high income countries," *Sci. Rep.*, 2022, doi: 10.1038/s41598-022-05915-3.
- [3] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, and S. Sharif, "An analysis of COVID-19 vaccine sentiments and opinions on Twitter," *Int. J. Infect. Dis.*, 2021, doi: 10.1016/j.ijid.2021.05.059.
- [4] A. Umair and E. Masciari, "A Survey of Sentimental Analysis Methods on COVID-19 Research," *CEUR Workshop Proc.*, vol. 3194, pp. 167–174, 2022.
- [5] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- [6] A. Amin, I. Hossain, A. Akther, and K. M. Alam, "Bengali VADER: A Sentiment Analysis Approach Using Modified VADER," 2019. doi: 10.1109/ECACE.2019.8679144.
- [7] S. Gouthami and N. P. Hegde, "a Survey on Challenges and Techniques of Sentiment Analysis," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 06, pp. 4510–4515, 2021.
- [8] A. Alsaeedi and M. Z. Khan, "A study on sentiment analysis techniques of Twitter data," *Int. J. Adv. Comput. Sci. Appl.*, 2019, doi: 10.14569/ijacsa.2019.0100248.
- [9] H. Xu, R. Liu, Z. Luo, and M. Xu, "COVID-19 vaccine sensing: Sentiment analysis and subject distillation from twitter data," *Telemat. Informatics Reports*, vol. 8, p. 100016, 2022, doi: <https://doi.org/10.1016/j.teler.2022.100016>.
- [10] A. R. Rahmanti et al., "Social media sentiment analysis to monitor the performance of vaccination coverage during the early phase of the national COVID-19 vaccine rollout," *Comput. Methods Programs Biomed.*, vol. 221, p. 106838, 2022, doi: <https://doi.org/10.1016/j.cmpb.2022.106838>.
- [11] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotaе, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst. Appl.*, vol. 212, p. 118715, 2023, doi: <https://doi.org/10.1016/j.eswa.2022.118715>.
- [12] R. Sun and H. Budhwani, "Negative sentiments toward novel coronavirus (COVID-19) vaccines," *Vaccine*, vol. 40, no. 48, pp. 6895–6899, 2022, doi: <https://doi.org/10.1016/j.vaccine.2022.10.037>.
- [13] Z. Bokaeе Nezhad and M. A. Deihimi, "Twitter sentiment analysis from Iran about COVID 19 vaccine," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 16, no. 1, p. 102367, 2022, doi: <https://doi.org/10.1016/j.dsx.2021.102367>.
- [14] S. Liu and J. Liu, "Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis," *Vaccine*, vol. 39, no. 39, pp. 5499–5505, 2021, doi: <https://doi.org/10.1016/j.vaccine.2021.08.058>.