Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 04 Issue 02
2023

Research Article
https://doi.org/10.56979/402/2023

Omicron Tweet Sentiment Analysis Using Ensemble Learning

Muhammad Khurram Iqbal[1], Kamran Abid[1], M fuzail[1], Salah u din Ayubi[2], Naeem Aslam[1]

[1]Department of Computer Science, NFC Institute of Engineering and Technology Multan, Pakistan
[2]Department of Information Technology, Faculty of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan.
*Corresponding Author: Muhammad Khurram Iqbal. Email: khurram.iqbal29@gmail.com

Received: January 01, 2023 **Accepted:** March 05, 2023 **Published:** March 29, 2023.

Abstract: In 2019, the COVID-19 pandemic took the world by storm, resulting in far-reaching impacts on education, economics, and health. As the coronavirus epidemic progressed, new mutations such as the Beta, Delta, and Omicron variants developed, causing fear and anxiety among the public. According to World meter, approximately 6 million people have so far died from COVID-19 and its variants. On November 24, 2021, the "SARS-CoV-2" omicron strain was first observed in South Africa and has since spread to over 57 countries. This study provides an analysis of the sentiment and behavior of people toward the omicron variant. We propose a method for performing sentiment analysis on Twitter data related to the omicron strain. Natural Language Processing techniques are used in Python to extract optimized features from the omicron tweets, creating a dataset that is then used by Machine Learning tools to train various models. The dataset is used to classify user emotional behavior into Neutral, Negative, and Positive categories using six machine learning classifiers: Naive Bayes, Random Forest, Decision Tree, Support Vector Machine, Voting and Stacking techniques Classifier. Accurately measuring the results of the predictions is the goal. According to the study's findings, when compared to other classifiers, the ensemble voting classifier had a performance accuracy of 85.33% and the ensemble stacking classifier had a performance accuracy of 87.5%.

Keywords: Machine Learning; Sentiment Analysis; Twitter, Omicron; COVID-19; Tweets; NLP; Big data.

1. Introduction

It is crucial to learn how to interpret the tone and sentiment of a writing, especially in the commercial world and during political decision-making. Reviews, rating and recommendation systems, product advertising, and other areas all use sentiment analysis. Businesses can save time and improve processes by automating the human labelling and categorization of evaluations. Sentiment analysis is beneficial for a variety of purposes, including evaluating consumer patterns, comparing the products of rival companies, and determining the sentiment of a product. Additionally, it can be employed in research, where categorized data provides richer understanding than raw data. It can be used in a variety of fields, including "health, stock markets, disasters, political themes, social topics, and more."` Internet users now have a platform to express their opinions and give their thoughts on a variety of topics and events thanks to social media. People today have a wide variety of internet channels at their disposal where they can voice their opinions on a wide range of subjects, events, or aspects of their personal lives [1]. People can now communicate their thoughts and information more easily because to the widespread use of social media. For analysts and researchers looking for information to aid in making strategic decisions, this data is an invaluable resource [2]. The majority of people today consider other people's points of view and openly express their agreement or disagreement with an argument. Twitter, a microblogging system, has become a global phenomenon and is used as a broadcast medium to share real-time information about what one is doing, thinking, or feeling [3]. Twitter has roughly 600 million active users, generating around 58 million tweets each day. Sentiment analysis is a useful technique for quickly gauging people's opinions from a massive amount of text data [4]. Opinions expressed on the internet are now a highly important resource for conveying sentiments and producing sentiment analysis data, enabled by the growing potency of social media networks [5][6]. "COVID-19 is a novel viral disease, initially observed in China in December 2019, and has since infected over 24.80 crore people worldwide by November 2021, causing an estimated 50.20 lakhs deaths, [9]. Numerous methods have been used to reduce the spread of illnesses, including company closures, self-quarantines, travel restrictions, and social isolation policies that have profoundly altered the social structures of societies all over the world [10]. "Several countries have established vaccination protocols to protect their citizens against the virus's devastating diseases [11]. Social media sites like Twitter, Instagram, Facebook, WhatsApp, and others assist gather useful COVID-19 information in this epidemic situation [12]. Content about "medical services, epidemic signs, and the communities affected by COVID-19 outbreaks can all be found on social media [13]. "Compared to other sites, Twitter is particularly effective in sharing informative messages with a length of 280 characters. Active users tweet multiple insightful information about the location and travel history of the patients, cases recovered, suspected, and confirmed, and the symptoms of the patients such as body pains, running nose, headache, fever, and cold [14]. The COVID-19-related tweets are labeled as 'informative' tweets, and the irrelevant user tweets are labeled as" 'uninformative' tweets" [14].

The recently found coronavirus that causes the infectious sickness known as Coronavirus Disease (COVID-19), which was initially identified in Wuhan, Hubei Province, China in December 2019, needs to be brought up in class (W.H.O., 2020). The virus has infected 213 nations across six continents (World ometer, 2020). The huge zoonotic virus family known as coronaviruses is responsible for a variety of ailments, including the common cold and respiratory conditions (Lab Manager, 2020). Sentiment analysis, or opinion mining, is a powerful tool for learning more about any given topic. It is widely used in research to gain an in-depth understanding of the data beyond just raw numbers. It can be used to analyze opinions on product reviews, "medicine, stock markets, disasters, political topics like elections, social topics like cyberbullying and rape, and more" (Mäntylä Graziotin & Kuutilaa, 2018)

This research seeks to study the Sentiment Analysis of COVID-19 Tweets. From this project, we want to attain the following objectives: to establish a reliable source for analysis, evaluations, and attitudes of the positive and negative reactions within the tweets through opinion summarization systems, to create a mutual understanding of the pandemic between the community and the nation, to track, monitor, and understand. Tweets on COVID-19 for a deeper understanding of the audience, to stay updated with what's being said about the pandemic, to discover new trends in the medical and research domains and use them to form effective COVID-19 strategic plans, to monitor and handle people's grievances, and to gain in-depth information for strategic analysis and determine the best algorithms for accurate evaluation. This research seeks to study the Sentiment Analysis of COVID-19 Tweets. From this project, we want to attain the following objectives: to establish a reliable source for analysis, evaluations, and attitudes of the positive and negative reactions within the tweets through opinion summarization systems, to create a mutual understanding of the pandemic between the community and the nation, to track, monitor, and understand.

The objectives of this study are to use machine learning (ML) models to analyze people's sentiments, To Train and build ML models to evaluate the performance of omicron tweets And to Compare the accuracy and confusion matrix results of different Ensemble models. The purpose of this project is to determine the sentiment of tweets about Omicron Corporation by using ML models such as SVM and DL Models. This project can help Omicron Corporation to better understand the sentiment of their customers and develop strategies to improve their customers' sentiments. This project can also provide insight into how customers feel about Omicron Corporation and its products and services, which can help Omicron Corporation to improve its products and services. This project can also help Omicron Corporation to identify potential issues with their products and services, and take corrective actions to address them.

In this paper, the aim is to learn how to create a professional approach to sentiment analysis and classifying the emotional behavior of Twitter users. Section 1 describes the introduction of the topic. Section 2 discusses some related works on sentiment analysis and describes the different techniques used to find the sentiment of user's tweets. Section 3 will cover our suggested approach to sentiment classification, using fundamental NLP methods and ML classifiers. Section 4 will explain the results of our suggested models, and present the experimental findings. Finally, the conclusion of the study will draw our study to a close.

2. Literature Work

In order to identify sentiment, Singh and colleagues used a BERT model to analyze Twitter data. The data was loaded into the BERT model for sentiment analysis and classification based on the content of the tweets [1]. The model's performance was then assessed using an SVM classifier, and on the collected dataset, an accuracy of about 94% was realized. The most popular subjects addressed on Twitter during and after the initial COVID-19 pandemic outbreak was investigated by Manal Abdulaziz et al. Latent Dirichlet Allocation (LDA) was used for theme mining, while a lexicon-based method was used for sentiment analysis. The study discussed how people felt during the first wave of the epidemic about many subjects. An English-language dataset of 600,000 tweets was used; 80% of the data were used to train the model and 20% were used to test it. With the aid of sentiment analysis, the sentiment related to the most popular themes was investigated and provided in the manuscript [2,3,4]. ML is a powerful tool used in fields from computer vision to text understanding and has proven to be effective in constructing effective models for pattern recognition. As machine learning becomes increasingly important and successful in engineering and sciences, it is important to understand the multi-step process of learning it entails [15]. The initial phase involves collecting and categorizing data into training and test sets, attributing labels and qualities to the training sets, and creating a model that depicts the relationship between the qualities and the label. Finally, the model is applied to various test sets using algorithms like classification and ranking to evaluate its accuracy. Different terms are used in machine learning that is important to understand. Sentiment analysis is a process that involves understanding people's attitudes toward certain topics. It is typically divided into three categorization levels: DL, SL, and AL. DL sentiment analysis focuses on analyzing the whole document and categorizing it as either positive or negative. SL sentiment analysis looks at each sentence separately, as if it were a brief document. Finally, AL sentiment analysis studies the emotions conveyed by individual words. By analyzing sentiment, we can gain insights into people's opinions and feelings [17]. Learning about the rise of social media starts with understanding the Internet and mobile technologies that make it possible. These technologies provide the technical framework, from content creation and distribution to interactive communication. Social media has become an essential part of the information ecosystem. Research on social media has grown substantially due to a surge of interest from multiple fields. Social media analytics is the process of designing and assessing informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data, often in service of a specific application [18].

Nair et al. investigated the rise in COVID-19-related tweets, which included both favorable and negative as well as neutral commentary. They were prompted to conduct sentiment analysis as a result of the increased social media activity in order to better understand the varying public perceptions on COVID-19. The extremes of opinion would be categorized by the conventional sentiment analysis techniques as either positive, negative, or neutral tweets. They suggested using Logistic Regression sentiment analysis, "VADER sentiment analysis, and BERT sentiment analysis to increase precision [5]. These suggested research methods can be categorized based on the domain and are more responsive to sentiment expressions on social media. "Despite the fact that three separate algorithms were used, all preprocessing and subsequent steps—with the exception of the sentiment analysis algorithm—remained the same. The three sentiment analysis techniques can be compared with the aid of this uniform processing. In the study by Vijay et al., an analysis of Twitter data from every Indian state from November 2019 to May 2022 was done. The results of the dataset's gathering revealed that the Indian community's general attitude was favorable [6,11]. This was caused by certain states receiving more tweets on the greater rate of COVID-19 positive patients. An innovative approach to analyzing sentiment in Twitter datasets was put forth in the work by Hassan Saif et al. It required connecting the represented notion and the positive/negative sentiment and adding semantics as extra features to the training set. Results showed that when compared to the baselines of unigrams and POS features, this technique increased the average F harmonic accuracy score by 6.5% and 4.8% for both positive and negative sentiment, respectively. In comparison to a system based on topic analysis with sentiment-bearing, precision with lower Recall and F scores were also seen when categorizing positive sentiment, and improved precision with higher Recall and F scores when classifying negative sentiment. In Patel Ravikumar, a novel method for sentiment analysis of Twitter data was put out. The most popular research approaches for sentiment analysis to extract pertinent information from linguistic data are NLP and input mining methods [8]. The created strategy used ML techniques to analyze Twitter data to determine general sentiment. Twitter was used to gather the data set for this study while the 2014 FIFA World Cup of Soccer was taking place in Brazil. Throughout this time, people expressed their opinions, sentiments, and attitudes on the game, the promotion, and the players. Based on the emotive terms in the user tweets, NLP techniques were used to filter and evaluate the data in order to estimate the sentiment polarity. The Python programming language and the freely accessible NLTK toolkit for academic and research purposes were used in this procedure. The created system used WordNet and its POS to extract emotional terms from phrases, which were then assigned sentiment polarity using either lexicon-based techniques or the Sent Word Net Dictionary [12]. The NB and SVM ML algorithms were used to further explore the given polarity, and the generated data were then assessed. A DL method for analyzing public mood was introduced in Jintao Ling. He described how Chinese citizens' use of microblogging as a major venue for sharing their opinions on current affairs. When the coronavirus outbreak occurred, the number of linked entries on microblogs immediately rose, giving an excellent opportunity to gauge public opinion. o finds out how people in various countries expressed their emotions, Dubey (2020) looked at "Tweets connected to coronavirus from March 11 to March 31, 2020. In comparison to other nations like Italy, Spain, and Belgium, he discovered that people in France, Switzerland, the Netherlands, and the USA displayed more hostility and mistrust. To acquire deeper insights, additional analysis that combines sentiment analysis and topic modelling can be done. According to a 2020 study by Medford et al on-Twitter activity, content, and feelings related to the US Presidential Election, there were shifts in both content and mood throughout the first three months of the election. Common applications of sentiment analysis include stock market forecasting, market trend analysis, product flaw analysis, and crisis management. The given table shows the previous studies work techniques, tools and languages [13]. This analysis typically involves a pre-set vocabulary, with each word assigned a value to indicate whether it is positive or negative. Sentences are broken down so each word can be identified, and then its assigned value is used to determine the sentiment of the overall text. The survey in the Dube's study discovered that immigration, healthcare, employment, and climate change were the primary subjects of Twitter activity and sentiment around the 2020 US Presidential Election. The survey also discovered that from January to March 2020, there was a notable rise in online conversations concerning these subjects. After researching the methods already in use, Anton and Andrey created a model for automatic sentiment analysis [15]. No matter the language or format of the material, they were able to create a system that can be applied to anything. They also demonstrated that their model's accuracy was on par with the very best, which is excellent news for those attempting to study text.

Table 1. Comparison of different research work

Ref.	Year	Dataset	Methods	Algorithm	Tools	Accuracy	Precision	Recall	F1-Score
(Hassan Saif et al.)	2017	Three different Twitter datasets	NLP, ML	SVM, NB	Python	75	77	75	74
(Patel Ravikumar et al.)	2020	Twitter dataset for the 2014 FIFA World Cup	NLP, POS, ML	SVM, KNN	Weka, Python	85	82	73	86
(Jintao Ling et al.)	2020	1 million blog postings	NLP, DL	BERT	Python	75.13	75	73	71
(Mohammad Mahyoob et al.)	2022	18,737 tweets from Twitter	SentiStrength software		SentiStrength	71	N/A	N/A	N/A
(Lokesh Mandloi and Ruchi Patel)	2020	Twitter dataset	ML algorithms	NB, SVM,	Python	86	88	80	81

(Doaa Mohey El-Din Mohamed Hussein)	2016	Online papers	ML, SAOOP	KNN, NB, RF	Python	83	76	71	80
(Deepika Vatsa and Ashima Yadav)	2022	Two-month data set of tweets	Linear Discriminant Analysis	N/A	Python	84	N/A	N/A	N/A
(Sanjeev Verma)	2023	Public services dataset	ML, HCI	NB	HCI	86	78	76	82
(Oksana Tokarchuk et al.)	2022	TripAdvisor online reviews.	Statistical analysis	N/A	SPSS	87	N/A	N/A	N/A

3. Materials and Methods

Over Constructed methodology Initially focused on cleaning and extracting the features from omicron Tweets data by using freely available nltk library of python for the preprocessing of dataset with NLP techniques. Secondly, identified optimized and most informative features in the dataset to enhanced the performance of the models. Thirdly some ML models (Naïve bayes, Random Forest, Decision Tree, Support Vector Machine and apply two ensemble techniques (Voting and Stacking) on these models to improve the accuracy) are applied to evaluate the created dataset. Finally, Ensemble ML techniques Voting and Stacking are Applied and enhanced the accuracy of models.

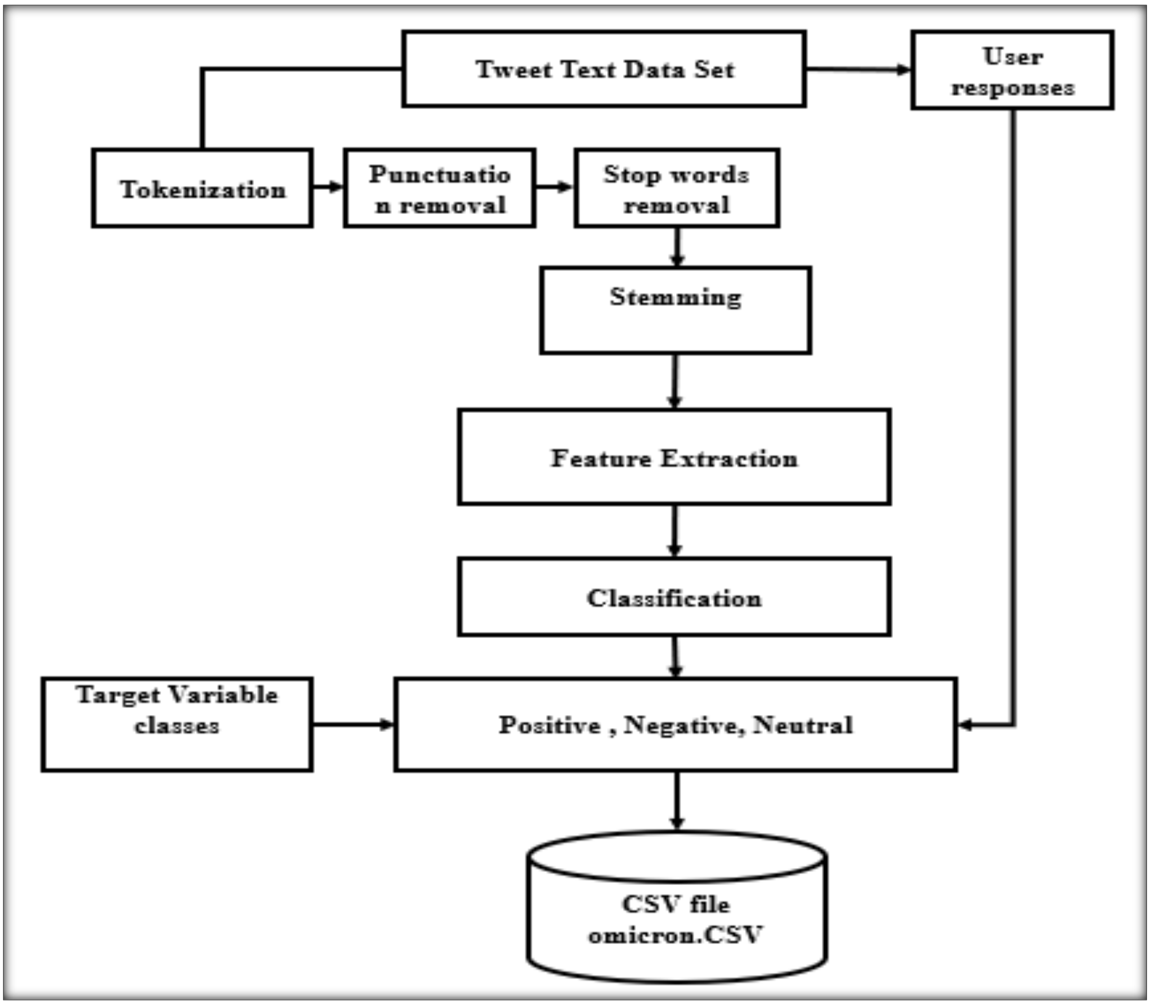

Figure 1. Proposed Methodology

The creation of the dataset is an ongoing process and will continue to be updated with new data. Firstly, we selected the dataset from Kaggle website. We then cleaned the data and removed stop words. Next, we analyzed the data by applying a sentiment analysis to it. After that, we created a word cloud for each category of tweets. Lastly, we put together our final dataset. Target variable is classified into three classes, ("positive, negative, and neutral").

3.1 Preprocessing

The preprocessing of the dataset is the first step in a process called text mining. In the preprocessing of the dataset, four tasks are done. They are tokenization, removal, filtering and formatting. The preprocessing of the dataset is needed before any other steps can be done to the data. Preprocessing of omicron tweets by using machine learning is a computational process that transforms raw data into a desired output. In this process, the raw data is broken down into smaller sets of data and then the data is analyzed and structured in a way that makes sense of the information. For this process, the machine learning algorithm will learn from the raw data and from previous processed data to make sense of it. A machine learning algorithm is an algorithm that learns from examples**.** In preprocessing, we use the tweets text to detect the sentiment of tweets.

3.2 Tokenization

Tokenization is the process of dividing a text into manageable chunks known as tokens. It is a way to convert raw text into a language-independent representation, so that the text can be processed by algorithms. Tokenization is often used in natural language processing, where it is typically performed on sentences or paragraphs. The word "tokenization" is used to describe a specific process of removing the inherent bias and emotional content from an audio file. This can be done by using a noise-reduction algorithm to remove the background noise, then using speech recognition software to isolate the speech. This process is commonly used in language translation, though can be applied to other industries as well.

3.3 Removal of Stop Words and punctuations

The words "a," "an," and "the" etc. are called stop words. They are words that have no meaning in a sentence. Stop words are removed from a text, so they don't appear. This is because it is more important to remove the stop words than it is to keep the original text**.** After the removal of Stop words, the punctuation removal process is done by using the NLTK library in Python.

3.4 Stemming

Stemming is the process of removing all the words like (suffixes+prefixes+other word elements until only the lemma, or root is left. The preprocessing of omicron Text is finalized after this step. Stemming is the last stage of preprocessing of the dataset.

Table 2. Preprocessing of text

Stages	Preprocessing Steps
1	Original Tweet Text: " im done with bitter bitches its a wrap for that . if you a angry bird theres a app for that "
2	Tokenized Tweet Text: '"', 'im', 'done', 'with', 'bitter', 'bitches', 'its', 'a', 'wrap', 'for', 'that', '.', 'if', 'you', 'a', 'angry', 'bird', 'theres', 'a', 'app', 'for', 'that', '"'
3	Tweet Text after Elimination of stop word: '"', 'im', 'done', 'bitter', 'bitches', 'wrap', '.', 'angry', 'bird', 'theres', 'app', '"'
4	Tweet Text after Elimination of Punctuation: 'im', 'done', 'bitter', 'bitches', 'wrap', 'angry', 'bird', 'theres', 'app
5	Tweet Text after stemming: im done bitter bitch wrap angri bird there app

The above table 2, shows the preprocessing stages that collect the original tweet text and end with tweet text after stemming in stage 5.

Figure 2 shows the features of the dataset. There are 12 features are containing in the dataset we select the text column and all the processing steps are applied on this column.

A	B	C	D	E	F	G	H	I	J	K	L	M
	tweet_id	date	text	user_nam	user_loca	user_desc	user_crea	user_follc	user_frier	user_favo	hashtags	source
0	1.47E+18	2021-11-2!	Will Boris	James robertson			2013-04-2:	303	188	84059		Twitter fo
1	1.47E+18	2021-11-2!	Omicron	Gene Bulr	East Coast	#ChristFol	2014-11-1(	99	162	745	['Omicron	Twitter fo
2	1.47E+18	2021-11-2!	How To St	Andrew Arcie Galen(		I AM	2021-01-2:	114	3223	4830		Twitter fo
3	1.47E+18	2021-11-2!	Gold Coas	myGC.con	Gold Coas	The Gold (	2009-03-3(	12232	3191	271	['GoldCoa	Zapier.cor
4	1.47E+18	2021-11-2!	What Is	Ghost Of (	Poughkee	Freelanc	2011-10-1'	603	960	819		Twitter fo
5	1.47E+18	2021-11-2!	Angeliqu	becbarr		Love my G	2020-11-1:	114	461	1714		Twitter fo
6	1.47E+18	2021-11-2!	Oil recouр	Calgary He	Calgary, A	Latest loc:	2008-05-1(	201011	802	659		Echobox
7	1.47E+18	2021-11-2!	Canada	IndiaToda	India	Brings you	2009-02-0:	5864435	246	4445	['Canada',	Twitter W
8	1.47E+18	2021-11-2!	Ontario	rizalimak	Toronto, (	Ismaili Mu	2015-06-2:	51	612	4071		Twitter fo
9	1.47E+18	2021-11-2!	Johnson, I	Bromio	bro city M	dude man	2021-05-3(	32	41	1162	['Omicron	Twitter fo
10	1.47E+18	2021-11-2!	Omicron c	Attilio Cot	Italia	Cardiovas	2012-02-1!	363	1429	7368	['Omicron	Twitter fo
11	1.47E+18	2021-11-2!	COVID-	Sparky Bri	The blue	Paddling	2021-10-0:	58	403	1033		Twitter W
12	1.47E+18	2021-11-2!	Benita sa	Î•Î»ÎµÎ½Î±	Australia		2012-02-1(	128	505	35151		Twitter fo
13	1.47E+18	2021-11-2!	While	Apple or è	Japan	Japanese	2021-09-2:	2	36	68		Twitter W
14	1.47E+18	2021-11-2!	Omicron	Michael H	Pennsylva	A dad. A h	2008-12-3(	198	301	7106		Twitter fo
15	1.47E+18	2021-11-2!	With the s	TMX Capital		Blockchai	2021-01-2:	1037	7	2		Twitter W
16	1.47E+18	2021-11-2!		ThePlan	UK Somev	James 1:1	2021-01-2:	447	451	3073		Twitter fo
17	1.47E+18	2021-11-2!	Could the	Matthew I	Brisbane,	Professor	2009-06-2(	17072	18751	60557		Twitter fo
18	1.47E+18	2021-11-2!	South	SEO Web '	Los Angel	SEO Web '	2007-09-0!	777	478	13854		Twitter W
19	1.47E+18	2021-11-2!	"The NT's	Josh	Brisbane,	Oi Oi Oi ð	2015-10-1	56	281	976		Twitter W
20	1.47E+18	2021-11-2!	In other n	Paul Reim	Schanzenfeld MB		2012-11-0:	99	65	1096		Twitter fo
21	1.47E+18	2021-11-2!	Omicron	Siri Rathо	Bengaluru	Keep Smil	2021-08-2:	20	2	7		Microsoft
22	1.47E+18	2021-11-2!	Canada's f	Krista	Born Mari	Host of 'M	2011-02-0	599	1185	6683		Twitter W
23	1.47E+18	2021-11-2!	The omic	KoenmaJr	Cornwall England		2017-03-0:	78	272	1024		Twitter W
24	1.47E+18	2021-11-2!	Why is ev	Postâ˜€ï¸,A	Woman O	"Post-Am	2019-09-0(	7597	287	174556		Twitter fo

Figure 2. Description of dataset

After the preprocessing of data, the evaluation of model is done the methodology for the evaluation of ML models are given below.

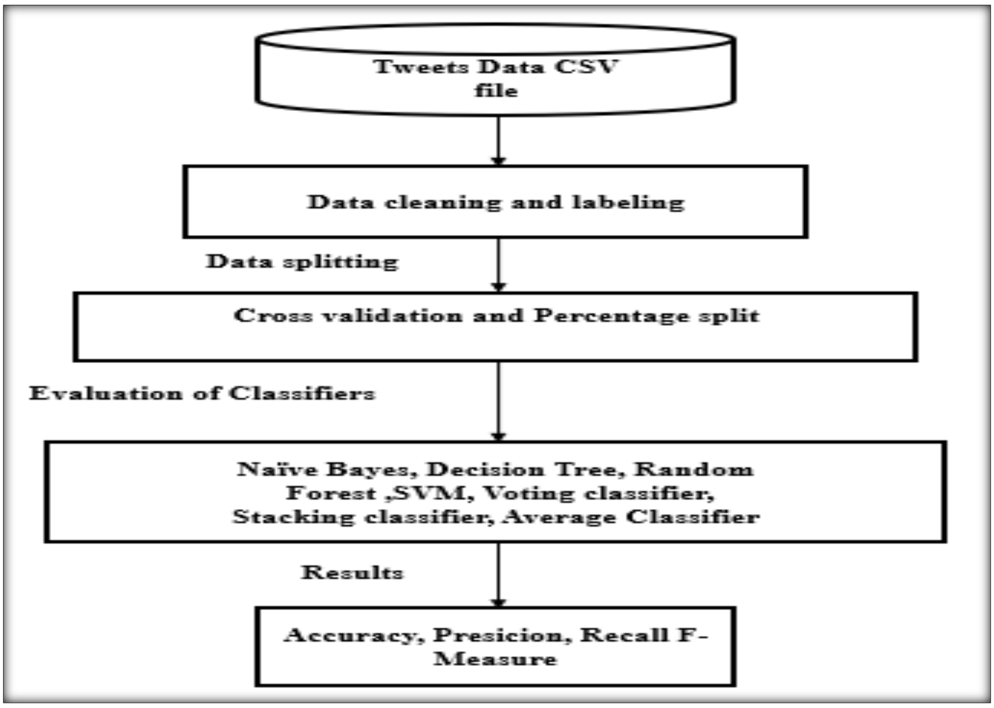

Figure 3. Evaluation of ML models

The Figure 2 represents the flow of our evaluation of our selected ML models. The dataset consists of CSV files and data mining techniques are applied on data such as Data cleaning data normalization. The prediction results representation methods used in this work are accuracy, precision, recall, F1 score the proposed model consists of the following steps: Data cleaning Data normalization Application of learning techniques on the data.

4. Results

For the machine learning model Evaluation, we have used the Six classifiers in this study, Decision tree, Random Forest, Naïve bayes, SVC, Ensemble Voting and Ensemble Stacking. For all of the used models we used training testing split evaluation method that available in the python and we used 75 percent of data for the training and remaining 25 percent for the testing.

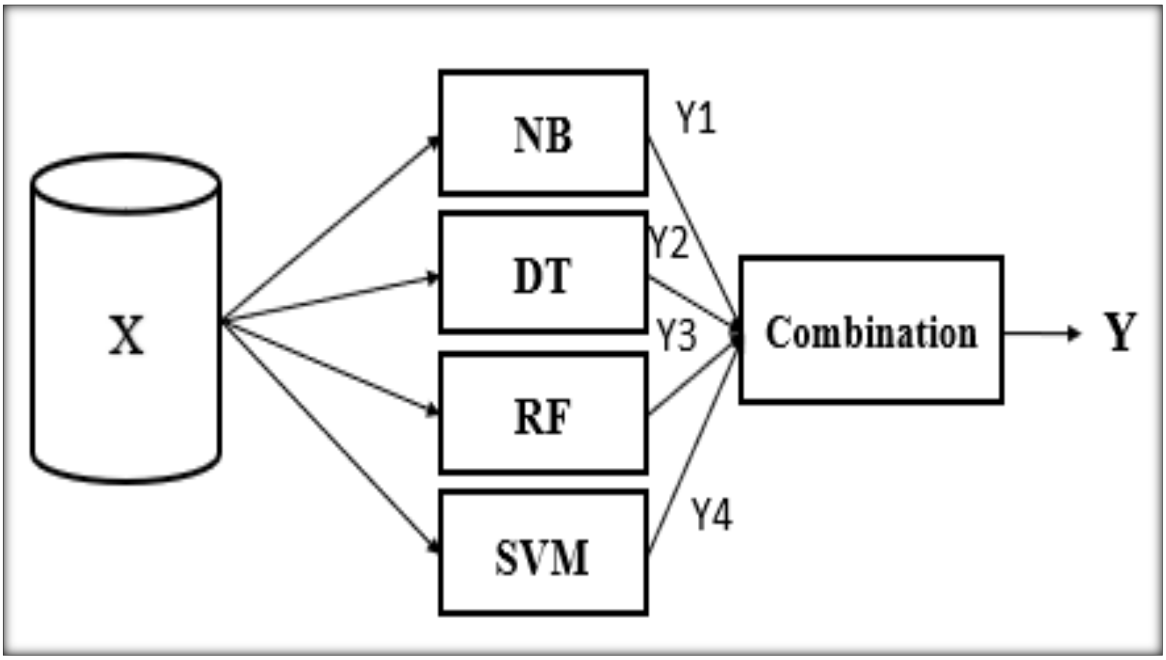

Figure 4. Ensemble models

The stacking and voting classifier combine the prediction results of decision tree classifier, random forest classifier, Naïve bayes classifier, and SVC classifier in one ensemble model. In this study we have performed comparative analysis of ensemble model and single model. The finding of this study represents that the stacking classifier give a highest accuracy of 88.06% than voting classifier the following mentioned below graph represent training and testing accuracy of voting and stacking classifier. The below diagram represents that the voting and stacking classifier gives the highest 100% percent accuracy and Stacking classifier gives highest accuracy 88.05%.

Table 3. Accuracy

Classifiers	Accuracy%
DT	81.25
RF	83.14
NB	62.31
SVC	86.06
VC	85.33
SC	88.06

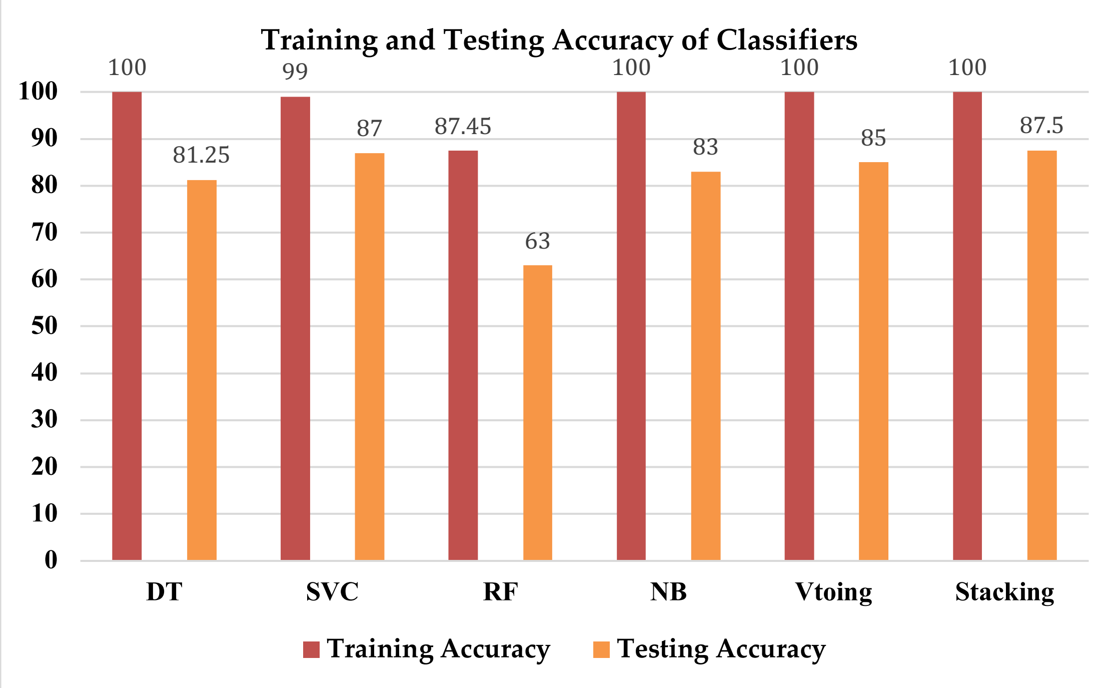

Figure 5. Accuracy Comparison

The Voting classifier is applied on dataset and gives the following accuracy in two forms: training and testing. The following table represents the Precision recall F1 score of Voting classifier. The class neutral shows higher precision (90%), recall (88%), f1score (89%).

Table 4. Confusion matrix of Voting Classifier

Class	Precision %	Recall %	F1-Score %
Negative (0)	85	70	77
Positive (1)	81	92	86
Neutral (2)	90	88	89

The Stacking classifier is applied on dataset and gives the following accuracy in two forms: training and testing. The following table represents the Precision recall F1 score of Stacking classifier. The class neutral shows higher precision (89%), recall (91%), f1score (90%).

Table 5. Confusion Matrix of Stacking Classifier

Class	Precision %	Recall %	F1-Score %
Negative (0)	90	80	84
Positive (1)	86	90	88
Neutral (2)	89	91	90

5. Conclusions

The first step of sentiment analysis is to gathered the relevant Twitter dataset. In this analysis, we gathered the dataset from the omicron. After the pre-processing of data, the feature extraction process is done by collecting the relevant features. In this research, we included only the text feature. The dataset used in this thesis is based on the omicron tweets that took all through the third wave of the corona. The dataset has 8073 tweets from different users. In the purposed method, we used NLP strategies in python language to extract optimized characteristics from the omicron tweets and created a dataset that recognizes by using the gadget-learned equipment to train the models. In this work, we have used six classifiers: Decision tree, Random Forest, Naïve Bayes, Support vector machine, voting classifier, and Stacking classifiers. We used to google PyCharm to develop machine learning models. We have analyzed in this study that the ensemble meta-classifiers give strong prediction results as compared to the single classifier model results. The voting classifiers give 85.33 accuracy. The stacking classifier gives 88.06 performance accuracy.

References

1. Gupta P., Kumar S., Suman R.R., Kumar V. Sentiment analysis of lockdown in India during COVID-19: a case study on twitter. IEEE Trans. Comput. Soc. Syst. 2020
2. William, P., Shrivastava, A., Chauhan, P. S., Raja, M., Ojha, S. B., & Kumar, K. (2023). Natural Language Processing Implementation for Sentiment Analysis on Tweets. In Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022 (pp. 317-327). Singapore: Springer Nature Singapore.
3. An, R., Yang, Y., Batcheller, Q., & Zhou, Q. (2023). Sentiment Analysis of Tweets on Soda Taxes. Journal of Public Health Management and Practice, 10-1097.
4. Chouhan, K., Yadav, M., Rout, R. K., Sahoo, K. S., Jhanjhi, N., Masud, M., & Aljahdali, S. (2023). Sentiment analysis with tweets behaviour in twitter streaming api. Comput Syst Sci Eng, 45(2), 1113-1128.
5. Aggrawal P., Jolly B.L.K., Gulati A., Sethi A., Kumaraguru P., Sethi T. Psychometric analysis and coupling of emotions between state bulletins and twitter in India during COVID-19 infodemic. Front. Commun. 2021;6
6. Osakwe Z.T., Ikhapoh I., Arora B.K., Bubu O.M. Identifying public concerns and reactions during the COVID-19 pandemic on twitter: a text-mining analysis. Public Health Nurs. 2021;38(2):145–151.
7. Lamsal R. Design and analysis of a large-scale COVID-19 tweets dataset. Appl. Intell. 2021;51(5):2790–2804.
8. Al-Shaher M.A. A hybrid deep learning and NLP based system to predict the spread of Covid-19 and unexpected side effects on people. Period. Eng. Nat. Sci. (PEN) 2020;8(4):2232–2241.
9. Sharma S., Sharma S. Analyzing the depression and suicidal tendencies of people affected by COVID-19's lockdown using sentiment analysis on social networking websites. J. Stat. Manag. Syst. 2021;24(1):115–133.
10. Kwan J.S.L., Lim K.H. Proceedings of the 26th International Conference on Intelligent User Interfaces. 2021. TweetCOVID: a system for analyzing public sentiments and discussions about COVID-19 via twitter activities; pp. 58–60.
11. Rustam F., Khalid M., Aslam W., Rupapara V., Mehmood A., Choi G.S. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLoS One. 2021;16(2)
12. Xue J., Chen J., Chen C., Zheng C., Li S., Zhu T. Public discourse and sentiment during the COVID 19 pandemic: using latent Dirichlet allocation for topic modeling on twitter. PLoS One. 2020;15(9).
13. Bello, A., Ng, S. C., & Leung, M. F. (2023). A BERT Framework to Sentiment Analysis of Tweets. Sensors, 23(1), 506.
14. Mostafa, R., Mehedi, M. H. K., Alam, M. M., & Rasel, A. A. (2023, March). Bidirectional LSTM and NLP Based Sentiment Analysis of Tweets. In Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022) (pp. 647-655). Cham: Springer Nature Switzerland.
15. Afroz N., Boral M., Sharma V., Gupta M. Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) IEEE. 2021. Sentiment analysis of COVID-19 nationwide lockdown effect in India; pp. 561–567.
16. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining.", LREc, vol. 10, pp. 1320--1326, 2010.
17. Elzayady H, Badran KM, Salama GI. Arabic Opinion Mining Using Combined CNN-LSTM Models. International Journal of Intelligent Systems & Applications. 2020 Aug 1;12(4).
18. Asur, Sitaram, and Bernardo A. Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
19. Gandhe, K., Varde, A. S., & Du, X. (2018). Sentiment analysis of twitter data with hybrid learning for recommender applications. In, 2018. 2018 9th IEEE annual ubiquitous computing, Electronics & Mobile Communication Conference (UEMCON), New York City, NY, USA (pp. 57–63). https://doi.org/10.1109/UEMCON.2018.8796661.
20. Lab-Manager. (2020, 03 16). COVID-19: A History of Coronavirus. Retrieved from Lab Manager: https://www.labmanager.com/lab-health-and-safety/covid-19- a -history-of-coronavirus-22021.
21. M. Singh, A.K. Jakhar, S Pandey Sentiment analysis on the impact of coronavirus in social life using the BERT model. Soc. Netw. Anal. Min., 11 (2021), p. 33, 10.1007/s13278-021-00737-z
22. International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 16, Number 2 (2020), pp. 87–104 © Research India Publications
23. Shaheen M, Awan SM, Hussain N, Gondal ZA. Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques. International Journal of Modern Education & Computer Science. 2019 Jul 1;11(7).
24. Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." LREc. Vol. 10. 2010.
25. H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of Twitter," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7649 LNCS, no. PART 1, pp. 508–524, 2012, doi: 10.1007/978-3-642-35176-1_32.
26. R. Patel, "Sentiment Analysis on Twitter Data Using Machine Learning by Ravikumar Patel A thesis submitted in partial fulfillment of the requirements for the degree of MSc Computational Sciences the Faculty of Graduate Studies," 2017.
27. J. Ling, "Coronavirus public sentiment analysis with BERT deep learning," Information, Commun. Soc., vol. 22, no. 13, pp. 2037–2038, 2019, DOI: 10.1080/1369118x.2019.1620824.
28. M. Mahyoob, J. Algaraady, M. Alrahiali, and A. Alblwi, "Sentiment Analysis of Public Tweets Towards the Emergence of SARS-CoV-2 Omicron Variant: A Social Media Analytics Framework," Eng. Technol. Appl. Sci. Res., vol. 12, no. 3, pp. 8525–8531, 2022, doi: 10.48084/etasr.4865.
29. L. Mandloi R. Patel, "Twitter sentiments analysis using machine learning methods," 2020 International Conference for Emerging Technology, INCET 2020.

30. D. Hussein, "Analyzing scientific papers based on sentiment analysis," Syst. Dep. Fac. Comput, no. June 2016, 2016, Available: https://www.researchgate.net/profile/Doaa-Mohey-El.
31. D. Vatsa, D. Vatsa, D. Vatsa, and A. Yadav, "An analytical insight of discussions and sentiments of Indians on Omicron-driven third wave of COVID- 19 using twitter data An analytical insight of discussions and sentiments of Indians on," pp. 0–15, 2022.
32. B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," Knowledge-Based Syst., vol. 235, p. 107643, 2022, doi: 10.1016/j.knosys.2021.107643.
33. S. Verma, "Sentiment analysis of public services for smart society: Literature review and future research directions," Gov. Inf. Q., vol. 39, no. 3, p. 101708, 2022, doi: 10.1016/j.giq.2022.101708.