VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2023, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 11, Number 1, January-March 2023 *pp:67-75*

Sentiment Analysis of Omicron Tweets by using Machine Learning Models

Unaiza Fazal[1], Muhibullah Khan[1], Muhammad Sajid Maqbool[2*], Hadia Bibi [2], Rubaina Nazeer[3]

[1]Department of computer science, NFC Institute of Engineering and Technology Multan, Pakistan
[2]Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan
[3]Department of Information Sciences, University of Education, Lahore, Pakistan
*Corresponding author email: sajidmaqbool7638@gmail.com

ABSTRACT
The COVID-19 epidemic has been affecting a lot of individuals worldwide since 2019. It is emerging as an infectious disease that set off a disaster with far-reaching effects on things like education, economics, and health. During the coronavirus outbreak, new COVID-19 mutations such the Beta, Delta, and Omicron variants emerged, terrifying and alarmed the population. Around 6 million people reportedly died as a result of COVID-19 variations, according to World Meter. The SARS-CoV-2 omicron strain was initially identified in South Africa on November 24, 2021, and it has since spread to more than 57 nations. In this essay, we examine how people feel and act toward the omicron variation. On Omicron, we proposed an approach for determining sentiment analysis for tweets from Twitter. The analysis of Twitter data's sentiment has a lot of potential. In the intended methodology, we extract the best characteristics from the Omicron tweets using NLP techniques in Python, resulting in a dataset that can be used to train the Models. The produced dataset was employed by four ML Classifiers, including "Naive Bayes (NB), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM)", to accurately categorise users' emotional behavior into three categories: neutral, negative, and positive. The Class Neutral receives the best score and the Class Negative receives the lowest score based on the accuracy of the forecast level.

KEYWORDS:
Machine Learning, Sentiment Analysis, Twitter, Omicron, COVID-19, Tweets, NLP, Big data

JOURNAL INFO
HISTORY: Received:February 02, 2023 Accepted: March 27, 2023 Published: March 31, 2023

1 INTRUDUCTION

Web users now have a platform to express and share their ideas on a variety of subjects and events thanks to the popularity of social media. [1] People now have access to a wide range of social media platforms through which they can voice their opinions on a variety of issues, incidents, or aspects of their own lives. Social media users now actively share their personal thoughts and facts with the world, which has led to an increase in the sharing of opinions in our period. This data is a gold mine for an analyst or researcher looking for valuable information for strategic decision-making [2]. Most individuals nowadays evaluate other people's viewpoints and openly express their agreement or disagreement with an argument. Now a days, a microblogging system, Twitter has become a worldwide phenomenon. Twitter is used as broadcast medium which is a way to share quickly where one is: what he is doing: what he is thinking: what he feels right now. It becomes a communication channel and has nearly 600 million users around the globe who actively used the service and they post around 58 million tweets daily [3]. Sentiment analysis is a useful method for quickly obtaining people's opinions from massive amounts of text data. Opinions on the internet have become a thing of a highly significant asset for conveying thoughts and generating Sentiment analysis data, thanks to the rising strength of social media networks for conveying ideas of current events and the rapid dissemination of information via the internet. COVID-19 arose as an infectious illness that triggered a worldwide catastrophe that had far-reaching consequences in areas such as education, economics and health. New COVID-19 mutations, such as the Beta, Delta, and Omicron variants, developed during the coronavirus epidemic causing terror and anxiety among the public. On November 24, 2021, the SARS-CoV-2 omicron strain was discovered for the first time in South Africa, and it has since spread to over 57 countries [5]. . In this research work we Purposed a method for finding Sentiment Analysis for Twitter Tweets on the Omicron. A lot of work can be done on the sentiment analysis of Twitter data. Our suggested dataset is based on the omicron-Tweets[1] that take during the third wave of the corona. The dataset 8073 tweets of different users from the date 2021-11-29 to 2021-12-02 the duration of two months. In the purposed methodology, we use NLP techniques in python language to extract optimized feature from the omicron-tweets and create a dataset that understand by the Machine Learning tools to train the Models. We construct ML models from a variety of classifiers using the PyCharm tool. The produced dataset was employed by four

[1] "https://www.kaggle.com/datasets/shivamb/omicron-covid19-variant-tweets?resource=download"

ML Classifiers, including NB, RF, DT, and SVM, to accurately categories users' emotional behaviour into three categories: neutral, negative, and positive. The rest of this work is structured as follows: The second section examines different NLP-based language recognition techniques to predict the sentiments of tweets.

The suggested NLP and ML based sentiment detection technique is introduced in Section 3. The effectiveness of the system is examined in Section 4. The conclusion for automatically detecting sentiuments is discussed in section five.

2 LITERATURE REVIEW

The objective of the current study is to expand on prior research that has shown the existence of both positive and negative attitudes toward the Omicron variation. Hassan Saif et al. [1] proposed research work and explain that Twitter sentiment research gives businesses a quick and efficient approach to tracking public opinion on their brand, operation, executives, etc. They present a novel method in this research for including semantics as additional features in the training set for sentiment analysis and adding the semantic concept (such as "Apple product") for each retrieved entity (such as the "iPhone") from tweets as an additional feature and calculate the correlation between the represented concept and the positive/negative sentiment. Patel Ravikumar [2] introduced a new technique for the sentiment analysis of Twitter data. NLP and Input Mining approaches are the most concentrated research terminology utilized for sentiment analysis to extract this unknown information from the linguistic data. Jintao Ling [3] a DL technique is introduced to public sentiment analysis. In this research work, he explains that In China the microblog has emerged as a major forum for the expression of views on current affairs. When a coronavirus outbreak occurs suddenly, the volume of linked posts on microblogs typically increases right once, offering a fantastic opportunity to gauge how the general public feels about the situation. Mohammad Mahyoob et al. [4] a new Twitter sentiment analysis technique is proposed during the omicron by using SentiStrength software. They express that While many COVID-19 variations have a significant negative impact on the lives of millions of people worldwide, a new COVID-19 variant known as "SARS-CoV-2 Omicron" arose. This study examines how the public feels about the spread of the SARS-CoV-2 Omicron strain on Twitter. The suggested method depends on text analytics of Twitter data considering the primary topics of tweets, retweets, and hashtags, the restriction of the pandemic, the effectiveness of COVID-19 vaccinations, transmissible variations, and the increase in infection.

Lokesh Mandloi and Ruchi Patel [5] presented that the process of determining whether a text's sentiments are favorable, negative, or neutral is called sentiment analysis. It is often referred to as the mining of opinions or material polarity. Numerous people were interested in social media platforms' expansion and development. Social media sites like Twitter allow users to write tweets with 280 characters or less. Tweets only have a certain number of characters, making sentiment analysis simple. Doaa Mohey El-Din Mohamed Hussein [6] suggests a novel method dubbed "Sentiment Analysis of Online Papers" for examining online reviews in the context of scientific research (SAOOP). By enabling researchers to report the overall rating for the publications, SAOOP seeks to support researchers and save their time and effort. A hybrid paradigm and new standards for judging scientific papers are introduced by SAOOP. This improved Part-of-Speech and Bag-of-Words hybrid model. SAOOP increases accuracy by resolving several sentiment problems. Each research paper is subjected to two evaluations as part of SAOOP: Sentiment score and System score. Online sentiments are rated using a sentiment score. System score is a brand-new parameter for topic evaluation. The suggested method offers remedies to some of the issues with sentiment to increase accuracy and identify these difficulties by searching the online reviews of scientific studies. Bi-polar terms, negation, world knowledge, topic domain traits, and developing a large lexicon are some of these difficulties. The suggested method assesses the implicit and explicit negative sentiment polarity of words and sentences. It also recognizes ambiguous or ambivalent words in attitudes and extracts topic elements, keywords, and attributes to support the judgment. It can be difficult to absorb world knowledge or to name notable scientists. In order to address this issue, the suggested technique generates a solution using similarities and differences algorithms and the hierarchal database model in nouns. On the basis of negative strength, they suggest a novel algorithm for evaluating negative challenges. System score is the second consideration when evaluating an important paper. This score is based on three important factors that are relevant to scientific papers. These factors are 1) the publication location, 2) the number of citations, and 3) the date of publication. Deepika Vatsa and Ashima Yadav [7] proposed a study on the impact of omicron in India in which they explain that Microblogging has emerged as one of the most important tools for expressing and exchanging viewpoints on everyday happenings. Online public health issues are being monitored through digital means. “Twitter is a highly well-known source for tweets about the public's attitude during the COVID outbreak. Many studies have utilized tweets to track public opinion on issues including the coronavirus vaccine, mental health issues, doctor treatment, the effects of lockdowns, etc. However, the first and second waves of the pandemic were mostly excluded from these studies”. In this research work, they examine the effects of the pandemic's third wave, which began in India in December 2021, to achieve this by gathering a two-month data set of tweets that discussed COVID-19 and had "IN" as the country code between December 2021 and January 2022 and used the Latent Dirichlet Allocation (LDA) method for topic modeling and assigned the most appropriate topic words to each tweet. Additionally, we used sentiment labels for each tweet and examined how different themes were distributed among the

various sentiment labels. The proposed study was able to assess people's opinions and feelings regarding many topical debates thanks to this. The investigation revealed that "precautionary measures" such as "get well fast," "keep safe," wearing masks, etc., and "vaccine," where people have debated its effectiveness and immunization drive in India, were the two most talked-about issues. Bin Liang et al. [8] proposed a new DL technique for sentiment analysis of tweets. They explained that the goal of aspect-based sentiment analysis, a fine-grained sentiment analysis task, is to identify the polarity of the sentiment toward a certain feature. Recently, aspect-based sentiment analysis has seen widespread use of graph neural networks over dependency trees To be more precise, we investigate a novel method of building graph neural networks by incorporating emotive information from SenticNet to improve the dependency graphs of phrases Sanjeev Verma [9]. The innovative affective enhanced graph model is based on it and considers both the interdependence of contextual words and aspect words as well as the affective information between opinion words and the aspect. Experimental results on numerous open benchmark datasets show that our suggested model can outperform cutting-edge approaches. Giuseppe D'Aniello et al. [11] This exercise employs the Syuzhet sentiment scoring package, which includes four sentiment dictionaries and a way for accessing the sentiment extraction tool built by Stanford's NLP department. The package includes four sentiment dictionaries as well as a way for accessing the Stanford NLP group's comprehensive, but computationally costly, sentiment extraction tool [12]. Bellegarda has published research that employs advanced dimension reduction techniques (variations of latent semantic analysis) to automatically find emotion phrases and achieves significant gains in categorizing newspaper headlines into distinct emotion categories and shift away from categorizing sentences from the writer's point of view and instead assign mental states to items referenced in the text. Their work focuses on polarity, but research into assigning emotions to items stated in text is also a viable subject for future research. [15]. Sautera et al. explored six emotions, which have been the subject of a lot of recent research. These feelings are a subset of the eight proposed by Plutchik: joy, sorrow, anger, fear, disgust, and surprise. Complex emotions, for example, receive less attention. Politeness, rudeness, embarrassment, formality, persuasion, deceit, confidence, and disbelief are all topics covered. For these feelings, they created a game-based annotation project [26]. Francisco and Gervás used lexicons of terms linked with the three categories to identify sentences in fairy tales with tags for pleasantness, activation, and dominance [16]. In today's world, social media is becoming an inextricable aspect of our lives, changing the way we create, share, and consume information. Hundreds of millions of individuals use social media sites like Facebook, Twitter, and YouTube to share, tweet, like, and post information. It's nearly difficult to picture our lives without social media, which is one of the most powerful technology innovations in recent years. Existing research has employed sentiment analysis for a number of objectives, including decision

Table 1. Comparison of some Research work

Ref.	Dataset	Technique	Tools
Hassan Saif et al. (2019)	three separate Twitter datasets	NLP, ML	Python
Patel Ravikumar (2012)	Twitter dataset for the 2014 FIFA World Cup of Soccer	NLP, POS, ML	Weka, Python
Jintao Ling (2019)	1 million blog postings	NLP, BERT, DL	Python
Mohammad Mahyoob et al. (2022)	18,737 tweets from Twitter	SentiStrength software	SentiStrength
Lokesh Mandloi and Ruchi Patel (2020)	Twitter dataset	ML algorithms	Python
Doaa Mohey El-Din Mohamed Hussein (2016)	Online papers	ML, SAOOP	N/A
Deepika Vatsa and Ashima Yadav (2022)	two-month data set of tweets	LDA	N/A
(Bin Liang et al. (2022)	multiple public benchmark datasets	BL	Python, Google Colab
Sanjeev Verma (2022)	Public services dataset	ML, HCI	HCI
Oksana Tokarchuk et al., (2022)	TripAdvisor online reviews.	Statistical analysis	SPSS
Giuseppe D'Aniello et al., (2022)	product reviews	aspect-based sentiment analysis (ABSA)	Statistical tools

Table 2. Preprocessing of Tweet Text

"Tweet Text: The largest language in the world as a mother tongue is modern Chinese, which is speaking 70 million people
Tokenized Text: 'The', 'largest', 'language', 'in', 'the', 'world', 'as', 'a', 'mother', 'tongue', 'is', 'modern', 'Chinese,', 'which', 'is', 'speaking', '70', 'million', 'people,'
Text after Removal of stop word: The', 'largest', 'language', 'world', 'mother', 'tongue', 'modern', 'Chinese,', 'speaking', '70', 'million', 'people,'
Text after Removal of Punctuation: the', 'largest', 'language', 'world', 'mother', 'tongue', 'modern', 'chinese,', 'speaking', '70', 'million', 'people
Text after Stemming: the largest language world mother tongue modern chinese speak 70 million people"

assistance, education, politics, opinion mining, data visualization, healthcare, and hate crimes, as well as the significance of education, gender sensitivity, and motivation. Many social media monitoring systems and trend analysis apps use opinion mining (also known as sentiment analysis). It uses computational linguistics, natural language processing, and other text analytics technologies to extract user attitudes or views from text sources at any granularity (words or phrases up to entire documents). Subjective data on people, goods, services, and other entities is used to assist a variety of tasks, including Stock market forecasting, market trend analysis, product flaw analysis, and crisis management [17]. They have a pre-set vocabulary in traditional dictionary-based analysis, and each word has a value, indicating whether the word's influence is good or bad. As a result, the sentences are dissected so that each word can be identified, and then the provided value is assigned to the impact that that word also has, according to our dictionary. Positive, neutral, and negative emotions may all be characterized [27]. Table 1 shows different available dataset of tweets and techniques

Table 3. Comparison of results with previous works

Ref.	Year	Dataset	Methods	Algoritham	Tools	Accuracy	Presicion	Recall	F1-Score
(Mohammad Mahyoob et al.)	2022	18,737 tweets from Twitter	SentiStrength software		SentiStrength	71	N/A	N/A	N/A
(Hassan Saif et al.)	2017	Three different Twitter datasets	NLP, ML	SVM, NB	Python	75	77	75	74
(Jintao Ling et al.)	2020	1 million blog postings	NLP, DL	BERT	Python	75.13	75	73	71
(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
(Patel Ravikumar et al.)	2020	Twitter dataset for the 2014 FIFA World Cup	NLP, POS, ML	SVM, KNN	Weka, Python	85	82	73	86
(Lokesh Mandloi and Ruchi Patel)	2020	Twitter dataset	ML algorithms	NB, SVM,	Python	86	88	80	81
(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
(Proposed)	2023	Tweets dataset	NLP and ML	NB, SVM, DT and RF	Python	89.8	89	82	85

that applied on these dataset by many researcher. Refeerence and used tools are also associated in this table. Table 3 lists different research works related to our domain. Precision, Recall, F1-Score, Accuracy, used tools and techniques, used algorithms, methods, dataset, references and years of all the mentioned studies are given. Some studies only focused on precision and accuracy by using ML models but needs to improve the results. our proposed model gives better result as compare to the existing research works.

3 METHODOLOGY

On Omicron, we aimed to find a method for sentiment analysis of tweets from Twitter. Our proposed dataset is based on the omicron-Tweets that occur during the third wave of COVID-19. The intended technique classifies user emotional behavior into (Neutral, Negative, and Positive) on the basis of various features using four machine learning algorithms: Naive Bays (NB), Random Forest (RF), Decision Tree (DT), and SVM. Our methodology devided into two main steps, first step is data creation methology and its flow is diplay in figure 1 and second step is performance evaluation methodology and its flow is display in figure 2.

3.1 DATA CREATION METHODOLOGY

Our selected dataset contains many features such as (Id, Date, Text, User, etc.) The creation of the dataset includes many stages that display in the Figure 1.

3.1.1 PREPROCESSING

Table 2 show one example of text with all preprocessing steps from simple text to processed text. Fiogure 1 depict that preprocessing contains four steps in order to create dataset which requires for sentiment analysis. Tokenization, removal of stop words and punctuation, and stemming are included in this step.

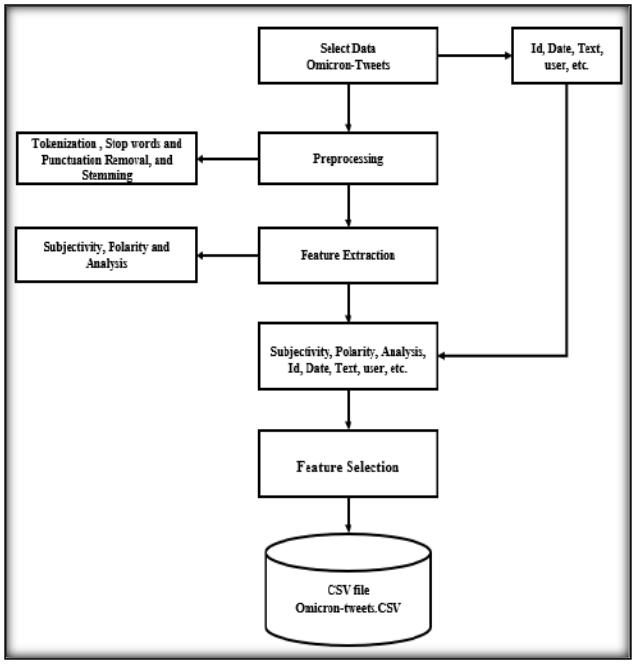

Figure 1 data creation

3.1.2 TOKENIZATION

Cutting the raw text into manageable chunks is known as tokenization. The original text is tokenized, or broken up into tokens like words and sentences.

3.1.3 REMOVAL OF STOP WORDS

After the Tokenization of Tweets Text, the removal of stop words is done by using the NLTK Library of Python. Stop words are a class of regularly used terms in all languages. English stop words that include non-technical terms.

3.1.4 REMOVAL OF PUNCTUATION

After the removal of Stop words, the punctuation removal process is done by using the NLTK library in Python.

3.1.5 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.

3.2 PERFORMANCE ANALYSIS

Performance Analysis is done by using the four machine learning models NB, RF, DT, and SVM. For evaluating the performance of ML models, we used created dataset omicron-tweets.csv file. In this methodology the cleaning and labeling of the created dataset are done and then Firstly, we split the dataset into training and testing (Testing data=75 % and training data= 25 %). Secondly, the selected models fit and can be trained on the training dataset and tested on the testing dataset. Thirdly the accuracy of each model is extracted and recorded for comparison. A confusion matrix for each model can also be plotted and created. Accuracy, Precision, Recall, and F-1 scores of each model are recorded. Finally, a complete comparative analysis is done between the Accuracy and Confusion matrix of models (NB, RF, DT, and SVM). Figure 2 display all of the evaluation fraomework phases.

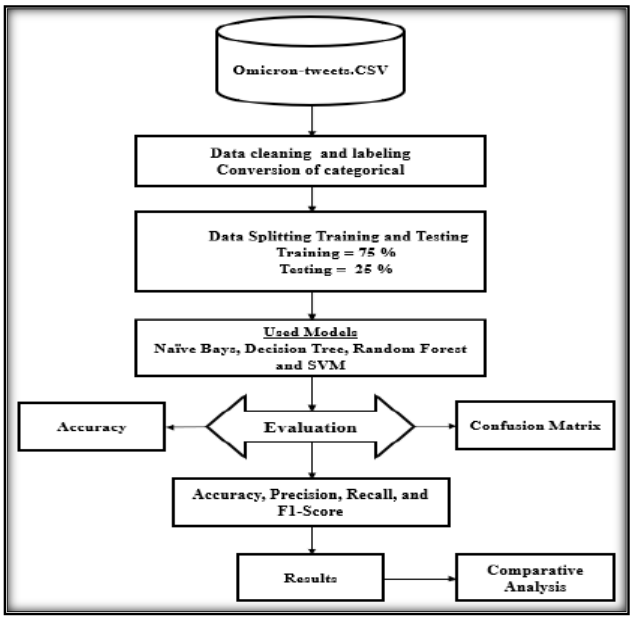

Figure 2 performance evaluation framework

3.3 MACHINE LEARNING MODELS

In our proposed Methodology we use four ML Models (NB, RF, DT, and SVM) to evaluate the sentiment analysis system.

3.3.1 EVALUATION METRICS

Four Evaluation metrics are used in our purpose Accuracy of Models, Precision of models on each, recall of models on each, and the F-1 score.

Accuracy: we use the following formula to calculate the accuracy. This is simple performance metric that tells that how many instances are correctly predicted [28]. It is given by

$$Accuracy = \frac{TP+FP}{TP+FP+TN+FN} \quad (1)$$

Precision: we used This statistic to describes that how much the predictions are near to each other. In other words, smallest the standard deviation, higher the precision.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall: Recall is used to find the true positive rate or sensitivity. Recall looks at the number of false negatives that included in confusion matrix. Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score: The F1 score of the models helps to measure precision and recall at the same time.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

4 RESULTS AND DISCUSSION

This study used a semi-structured data collection of Twitter data that was made accessible. The data set includes a "text" field where user-generated tweets are utilized for the study. These tweets may contain noise in addition to incomplete and erroneous linguistic data. To apply the proposed methodology to our data set it is necessary to clean the irregular dataset and apply the proper punctuation and stemming. After this, the actual meaning and emotions should be evaluated. To apply the cleaning and filtering of the data, we use algorithms in the python programming language. Four ML models are applied to the omicron-tweets dataset to evaluate the model. All four models are evaluated on the 4 matrices and the comparative analysis is done in this section.

4.1.1 ACCURACY OF MODELS

Table 4 is containing the accuracies of all four ML models. Each model has three types of accuracies one is for Cross-Validation and the other is for Training and last is for Testing.

Table 4 Accuracy

Models	Training	Testing	Cross-Validation
NB	87	59.4	63
RF	100	89	81
SVM	100	85	83
DT	100	90.8	81

Figure 3 depict the accuracy of different models. The models SVM, RF, and DT Classifier have Highest Accuracy on Training Data with an accuracy of 100 and the lowest Accuracy is given on the model NB Classifier.

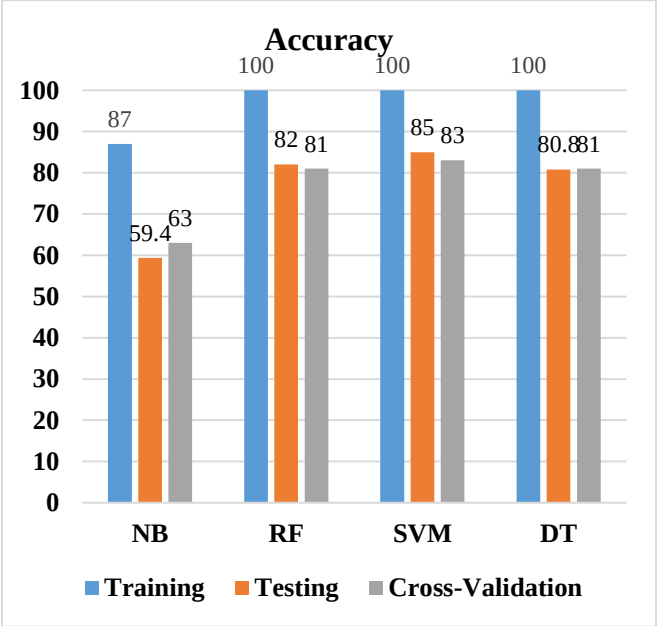

Figure 3 Accuracy of Models

In the above Graph accuracy ML models are compared on three matrices CV, Training, and testing. The Maximum accuracy in the Cross training is 100 % which is calculated by the RF and DT Classifier and the minimum Accuracy of CV is on the model NB with an accuracy of 59.4 percent.

4.1.2 PRECISION OF MODELS

The precision Score of all selected four ML Models is given class class wise in table 5. NB perform well in the class Negative with an accuracy of 100 Percent and the lowest performance in Positive behavior. RF gives Maximum

precision on the Behavior of Negative with a score of 100 % and gives minimum precision on Positive behavior with a score of 63 %. The SVM Classifier gives Maximum accuracy at Neutral behavior with the precision of 93 percent and the Behavior Positive has minimum precision with score of 83 percent.

Table 5 Precision of Models

Models	Negative	Positive	Neutral
NB	46	63	69
RF	100	76	83
SVM	90	84	93
DT	77	83	81

Figure 4 shows that DT classifier gives maximum Precision at class Neutral and Positive with an accuracy of 83 and 81 % respectively and minimum performance at class Negative with an accuracy of 77 %.

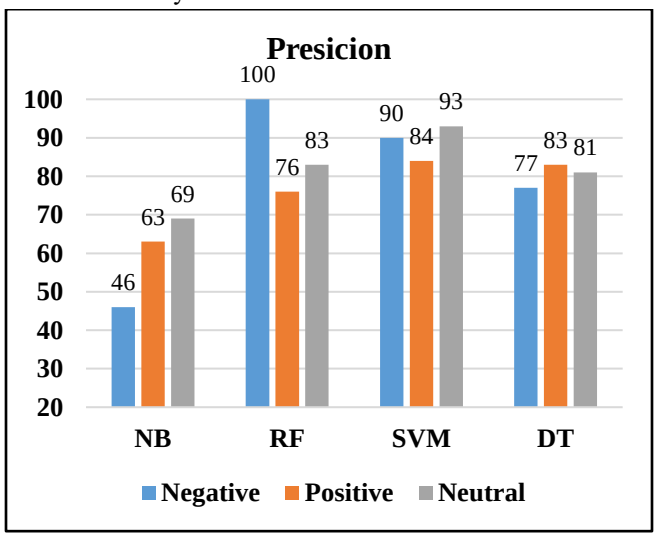

Figure 4 Precision

4.1.3 RECALL OF MODELS

Recall of all four ML Models is given in Table 6 with each class (Neutral, Positive and Negative). Recall of NB Classifier is maximum at the Negative emotional behavior of the users with a performance of 69 percent and the minimum recall is in the Positive emotional behavior of users with a performance of 52 percent. RF Classifier give maximum Recall on Positive and Neutral emotional behavior of users with a score of 90 % and gives minimum Recall on the class Negative with poor performance of 56 %. The SVM Classifier gives Highest score at class Neutral with Recall of 96% and the Positive behavior has low Recall with score of 92 percent.

Table 6 Recall

Models	Negative	Positive	Neutral
NB	69	52	61
RF	56	90	90
SVM	58	92	96
DT	68	81	88

DT Classifier gives maximum Recall at class Neutral with score of 88 percent and minimum score at class Negative.

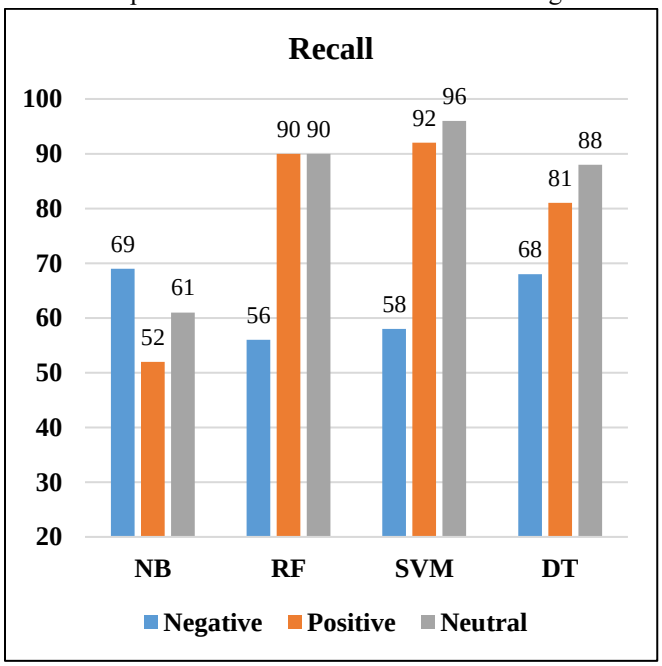

Figure 5 Recall

The in Figure 5 contains the values of Recall with each class and shows that the lowest Recall is calculated by the on Negative Class and the highest Recall is calculated by the SVM with accuracy of 99 % respectively.

4.1.4 F-1 SCORE

F-1 Score of selected ML models are given in Table 7 with each class of emotional behavior. F-1 Score of NB Classifier Shows that the highest F1 Score at the class Neutral with an accuracy of 65 and the lowest F-1 Score in Positive and Negative emotional behavior of the users with an accuracy of 57 and 55 percent respectively. RF Classifier give a maximum F-1 Score on Neutral with a performance of 86 % and gives poor results on the Negative with an accuracy 71%. The SVM Classifier gives maximum performance at class

Neutral with Score of 82 percent and the class Positive has poor score of 80 percent.

Table 7 F-1 Score

Models	Negative	Positive	Neutral
NB	55	57	65
RF	71	82	86
SVM	81	80	82
DT	72	82	84

DT Classifier gives maximum F-1 Score at Neutral with 84 percent and poor performance at class Negative with accuracy of 72 percent.

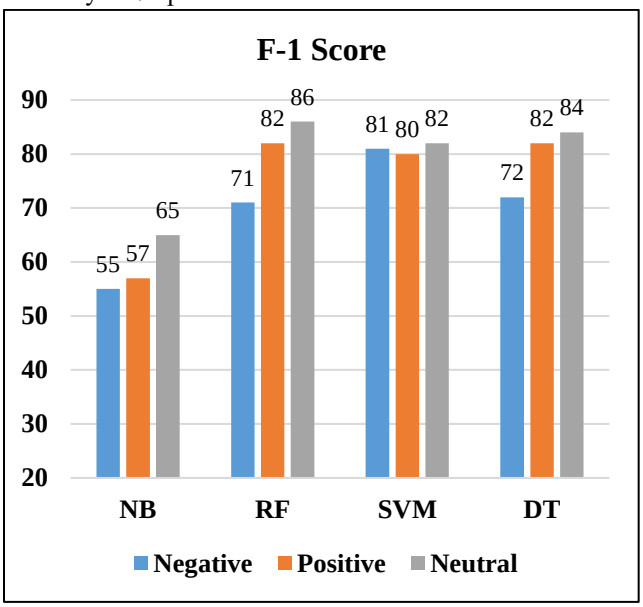

Figure 6 F1-score

5 CONCLUSION

Twitter was used as a broadcast medium which was a way to share quickly where one was: what he was doing: what he was thinking: and what he felt right now. In this researched worked we purposed a method for finding sentiment analysis for Twitter tweets on the omicron. A lot of work could be done on the sentiment analysis of Twitter data. Our suggested dataset was based on the omicron tweets that took during the third wave of the corona. The dataset has 8073 tweets of different users from the date 2021-11-29 to 2021-12-02 the duration of two months. "In the purposed methodology, we used NLP techniques in python language to extract optimized features from the omicron tweets and created a dataset that understands by the machine-learned tools to train the models. We used PyCharm tool to build ML models from various classifiers. Four ml classifiers such as naïve bays (NB), random forest (RF), decision tree (DT), and support vector machine (SVM) used the created dataset to classify the emotional behavior of users into (neutral, negative, and positive) on the basis of some features to measure the accuracy".

The SVM and RF classifier better performed when the model is trained using all the features. The accuracy of SVM classifier is 89.8 % and the accuracy of RF classifier is 82%. According to the precision of the prediction level, the highest score is given by the Class Neutral and the lowest score is given by the class Negative.

In the future, to improve the results our created dataset is evaluated by using different Deep learning techniques as well as the Bidirectional Encoding Representation by Transformer (BERT) technique is used to create the dataset again and then predict the sentiments of users.

CREDIT AUTHOR STATEMENT

Unaiza Fazal: Methodology, Software**. Muhibullah Khan:** Data curation, Supervision **Muhammad Sajid Maqbool**: Conceptualization, Investigation, Coreesponding, Writing- Original draft preparation. **Hadia Bibi:** Visualization, Validation **Rubaina Nazeer:** Software, review

COMPLIANCE WITH ETHICAL STANDARDS

It is declare that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] L. Mandloi R. Patel, "Twitter sentiments analysis using machine learning methods," 2020 International Conference for Emerging Technology, INCET 2020.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] O. Tokarchuk, J. C. Barr, and C. Cozzio, “How much is too much? Estimating tourism carrying capacity in urban context using sentiment analysis,” Tour. Manag., vol. 91, no. January, p. 104522, 2022, doi: 10.1016/j.tourman.2022.104522.

[11] G. D’Aniello, M. Gaeta, and I. La Rocca, KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis, no. 0123456789. Springer Netherlands, 2022.

[12] F. Yousaf, S. Iqbal, N. Fatima, T. Kousar and M. S. M. Rahim, "Multi-class disease detection using deep learning and human brain medical imaging," Biomedical Signal Processing and Control, vol. 85, pp. 104875, 2023.

[13] M. S. Maqbool, I. Hanif, S. Iqbal, A. Basit and A. Shabbir, "Optimized Feature Extraction and Cross-Lingual Text Reuse Detection using Ensemble Machine Learning Models," in IEEE [insert name of conference or journal], 2022, pp. [insert page numbers].

[14] A. Akbik, S. Schweter, D. Blythe, and R. Vollgraf, “F LAIR : An Easy-to-Use Framework for State-of-the-Art NLP,” pp. 54–59, 2019.

[15] R. Rosu, A. S. Stoica, P. S. Popescu, and M. C. Mihaescu, “NLP based Deep Learning Approach for Plagiarism Detection,” Int. Joural User-System Interact., vol. 13, no. 1, pp. 48–60, 2020, doi: 10.37789/ijusi.2020.13.1.4.

[16] C. Li and C. Wu, “A new semi-supervised support vector machine learning algorithm based on active learning,” Proc. 2010 2nd Int. Conf. Futur. Comput. Commun. ICFCC 2010, vol. 3, pp. 638–641, 2010, doi: 10.1109/ICFCC.2010.5497471.

[17] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[18] I. Lopez Torres, “Omicron Tweets Sentiment Analysis,” SSRN Electron. J., 2022, doi: 10.2139/ssrn.3987756.

[19] D. M. E. D. M. Hussein, “A survey on sentiment analysis challenges,” J. King Saud Univ. - Eng. Sci., vol. 30, no. 4, pp. 330–338, 2018, doi: 10.1016/j.jksues.2016.04.002.

[20] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 4, pp. 1–25, 2018, doi: 10.1002/widm.1253.

[21] D. M. E. D. M. Hussein, “A survey on sentiment analysis challenges,” J. King Saud Univ. - Eng. Sci., vol. 30, no. 4, pp. 330–338, 2018, doi: 10.1016/j.jksues.2016.04.002.

[22] R. Prabowo and M. Thelwall, “Sentiment analysis: A combined approach,” J. Informetr., vol. 3, no. 2, pp. 143–157, 2009, doi: 10.1016/j.joi.2009.01.003.

[23] P. Goncalves, B. Fabrício, A. Matheus, and C. Meeyoung, “Comparing and Combining Sentiment Analysis Methods Categories and Subject Descriptors,” Proc. first ACM Conf. Online Soc. networks, pp. 27–38, 2013.

[24] N. Thakur and C. Y. Han, “An Exploratory Study of Tweets about the SARS-CoV-2 Omi- cron Variant : Insights from Sentiment Analysis , Language In- terpretation , Source Tracking , Type Classification , and Embed- ded URL Detection,” vol. 2, no. May, 2022, doi: 10.20944/preprints202205.0238.v1.

[25] M. S. Maqbool, I. Hanif, S. Iqbal, A. Basit and A. Shabbir, "Optimized Feature Extraction and Cross-Lingual Text Reuse Detection using Ensemble Machine Learning Models," in Proceedings of the IEEE [insert name of conference or symposium], pp. [insert page numbers], 2022.

[26] A. Srivastava, V. Singh, and G. S. Drall, “Sentiment analysis of twitter data: A hybrid approach,” Int. J. Healthc. Inf. Syst. Informatics, vol. 14, no. 2, pp. 1–16, 2019, doi: 10.4018/IJHISI.2019040101.

[27] R. Marcec and R. Likic, “Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines,” Postgrad. Med. J., pp. 544–550, 2021, doi: 10.1136/postgradmedj-2021-140685.

[28] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168-177.

[29] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), 2010, pp. 1320-1326.

[30] M. Thelwall, K. Buckley and G. Paltoglou, "Sentiment in Twitter events," Journal of the American Society for Information Science and Technology, vol. 63, no. 1, pp. 163-173, 2012.

[31] A. Bakliwal, P. Arora, S. Madhappan, N. Kapre, M. Singh and V. Varma, "Mining sentiments from tweets," in Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, 2012, pp. 11-18.

[32] M. Bouazizi and T. Ohtsuki, "Sentiment analysis in twitter: From classification to quantification of sentiments within tweets," in Proceedings of the 2016 IEEE Global Communications Conference (GLOBECOM), 2016, pp. 1-6.

[33] R. Chandrasekaran, V. Mehta, T. Valkunde and E. Moustakas, "Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study," Journal of Medical Internet Research, vol. 22, no. 10, e22624, 2020.

.