




VADER vs. BERT: A Comparative Performance Analysis for Sentiment on Coronavirus Outbreak

Subrata Saha¹, Md. Imran Hossain Showrov²(✉) , Md. Motinur Rahman¹,
and Md. Ziaul Hasan Majumder¹

¹ Institute of Electronics, Bangladesh Atomic Energy Commission, Dhaka, Bangladesh
² Institute of Computer Science, Bangladesh Atomic Energy Commission, Dhaka, Bangladesh
showrov.cse@gmail.com

Abstract. Nowadays, social media is getting more attention. Opinions on a wide range of topics are constantly expressed and distributed via a variety of social media platforms including Twitter. These ever-increasing subjective data are unquestionably a wealth of information for any decision-making process. The field of Sentiment Analysis has emerged to automate the analysis of such data which combines Natural Language Processing (NLP) and Machine Learning (ML). A branch of artificial intelligence (AI), machine learning uses algorithms to improve prediction and analysis for a variety of real-world issues. In this paper, initially, we crawled tweets from Twitter about COVID-19 and Omicron, and two different datasets were created. We applied VADER and BERT for finding the sentiment polarity of the tweets. Then 5 popular supervised ML algorithms were used to analyze the classification performance. From research, we found that most of the tweets expressed negative sentiments and in the classification task, Support Vector Machine outperformed other algorithms with an accuracy of 92% using BERT on the omicron dataset. We also found that BERT improves the classification performance except for the Naive Bayes algorithm.

Keywords: sentiment analysis · covid19 · omicron · VADER · BERT

1 Introduction

‘Microblogging’ is a broad term for the practice of posting very concise status updates which made it very popular due to its services such as Twitter. Twitter is one of the most popular microblogging platforms, where enrolled users can easily update their messages and follow others. Twitter’s texts (called tweets) include user’s views, thoughts, and remarks on particular subjects. Tweet length, which is limited to about 280 characters, makes it simple to share [1]. People also tweet about a variety of topics, including news, movies, product brands, political systems, and so on. As a result, many of these contents inevitably express the users’ sentiments. This data source presents a fantastic opportunity for businesses and academics to conduct community thinking research [2, 3].

Sentiment analysis is a subfield of opinion mining; which includes the process of classification of texts based on the feelings that the text appears to express. Sentiment analysis usually categorizes texts as neutral, positive, or negative. For example, “This novel was awesome!” is classified as positive, whereas “This novel was boring” is classified as negative, and “This is a novel written by Rabindranath Tagore” is classified as neutral [4]. Twitter sentiment analysis is an important topic because the research findings can provide us with a wealth of previously unknown information such as the stock market’s trends, and what people think about cryptocurrency or bitcoin.

The Chinese government first reported unidentified pneumonia cases in December 2019. Since then, the pandemic, known as COVID-19, has spread throughout the world. As the disease spread, people all over the world began to use social media including Twitter to voice their concerns about COVID-19. This wealth of data has enabled researchers to analyze the pandemic using Twitter data. For instance, since the beginning of the pandemic, studies have been conducted to examine the tweets of various world leaders and their messages to the public [5]. Several variants have emerged since the beginning of the COVID-19 pandemic in December 2019. Some variants have spread globally, contributing significantly to the cyclical waves of infection that occur in different regions. Among many variants, the highly infectious Delta variant of concern (VOCs) eventually displaced all other variants in most parts of the world. While Delta continued to transmit at high levels in the Northern hemisphere in October 2021, Delta waves were starting to disappear in different places of southern Africa. At the same time, Omicron, a new variant of COVID-19 emerged. Since then, and until the beginning of 2022, Omicron has dominated around the world [6]. Later on 7th January 2022, virologist Leondios announced the identification of ‘Deltacron’ which has similarities to both Delta and Omicron [7].

Analysis of Twitter data sentiment can be accomplished through twitter’s tweet extraction. Then implement some sentiment classification algorithms to all these tweets to determine whether a particular tweet conveys either a favorable, negative, or neutral feeling. The mental health of people around the world can be studied based on this analysis, whether they still feel secure or are in the excessive worry stage. The neutral and positive sentiment values indicate safe conditions, and the negative sentiment values indicate a worrying condition [8].

In this paper, we used two popular sentiment analyzers namely VADER and BERT for sentiment analysis after crawling data from Twitter using the hashtags “covid-19” and “omicron” for a 7 days tenure. At first, we made two separate datasets for covid and omicron related tweets, preprocessed the data, and then analyzed the sentiments. Finally, supervised algorithms were used over the datasets for prediction.

The remaining paper is organized as follows. Section 2 represents the previous works related to sentiment analysis and covid19 outbreak. Section 3 has outlined the proposed technique. Section 4 provides information about the detailed setup for the experiment and obtained results. The paper is concluded by the conclusion and future works in Sect. 5.

2 Related Work

One of the earliest works on sentiment analysis was done by Go et al. [9] where they utilized positive emoticons such as “:)” “:-)” for positive tweets and negative emoticons such as “:(” “:-(” for negative tweets. To collect sentiment data, they employed distant learning. They developed the model using Naive Bayes, MaxEnt, and Support Vector Machines (SVM) and showed that SVM outperforms others. They also combined parts-of-speech (POS) characteristics with a Unigram, Bigram model and the unigram model surpasses all others.

Shihab Elbagir et al. [10] performed tweets sentiment classification by using Valence Aware Dictionary for Sentiment Reasoner (VADER) and Natural Language Toolkit (NLTK). In their study, they proposed a multiclass classification system for analyzing tweets related to the 2016 US election. From their study, 29% of the tweets expressed positive, 22.89% expressed negative, 46.7% expressed neutral, and 1.41% expressed highly negative opinions. By modifying the VADER, to allow Bengali sentiment polarity detection, Amin et al. [11] examined Bengali sentiment analysis in their study. Based on VADER’s English polarity lexicon, they produced a Bengali version. Additionally, they changed English VADER’s features so that it can categorize Bengali text feelings without the requirement for Bengali to English translators like Google Translator or MyMemory Translator. Their study improved upon the present model’s effectiveness for Bengali text sentiment evaluation.

Mao and Liu [12] Presented a technique for Automated Humor Identification and Grading based on Bidirectional Encoder Representations from Transformers (BERT). In this paper, they used corpus for tweets and predicted whether or not a particular tweet was a joke, as well as assigned a score to it. They used BERT and pre-trained the Humor Analysis based on the Human Annotation task, HAHA task, for this purpose. The mean square error was used to calculate the score. This method was suitable for tasks involving multilingual text classification.

Shamrat et al. [13] worked on sentiment analysis over a dataset compilation of tweets about the COVID-19 vaccine using NLP and KNN algorithm for classification. From the analysis, they discovered that Pfizer, Moderna, and AstraZeneca received 47.29%, 46.16%, and 40.08% positive feelings, respectively.

Kaur et al. [14] used R as the programming language to analyze tweets. They used several keywords named COVID-19, coronavirus, new cases, recovered cases, death cases, etc. to collect Twitter data. Hybrid Heterogeneous Support Vector Machine (H-SVM) was employed in their work to do sentiment classification. They also evaluated the effectiveness of H-SVM with the Recurrent Neural Network (RNN) and the Support Vector Machine (SVM). N. I. Mahbub et al. [15] presented a model for sentiment analysis and context learning on covid19 dataset. They used different machine learning algorithms and they found that the random forest algorithm is the most convenient one with an accuracy of 93%.

In another study, A. D. Dubey [16] took into account COVID-19 related tweets from twelve different countries, and sentiment analysis was done after text preprocessing. The results of the study concluded that although the majority of people throughout the world took a positive and hopeful approach, there were instances of fear and sadness.

However, compared to the other eight nations, France, Switzerland, the Netherlands, and the United States have shown more skepticism and antagonism.

3 Methodology

This section illustrates the methodology of the proposed model for determining how individuals feel about the ongoing Coronavirus variants in a global context, as shown in Fig. 1.

The overall process consists of three parts. At first, data is collected from Twitter about different covid19 variants, using Tweepy which is a Python library for accessing the Twitter API [17], followed by filtering and preprocessing tweet datasets to fit them into a machine learning algorithm. In the second part, sentiment analysis is performed using VADER and BERT models to get the polarity score of each individual tweet. Finally, the prediction of tweet sentiments is done using several ML algorithms and the performances are compared.

3.1 Data Acquisition and Pre-processing

The crucial component of data analysis is to ensure that the data is understandable by machines. Machines can only recognize 1's and 0's, not text, photos, or movies. It takes several steps to convert our text data to binary numbers. So Preprocessing the data is a must, and it requires cleaning the data and also involves converting raw data into a machine recognizable format. Regular expressions (@, https, # etc.) were removed along with the punctuation. After applying the case-folding for the letters, we employed tokenization and stemming.

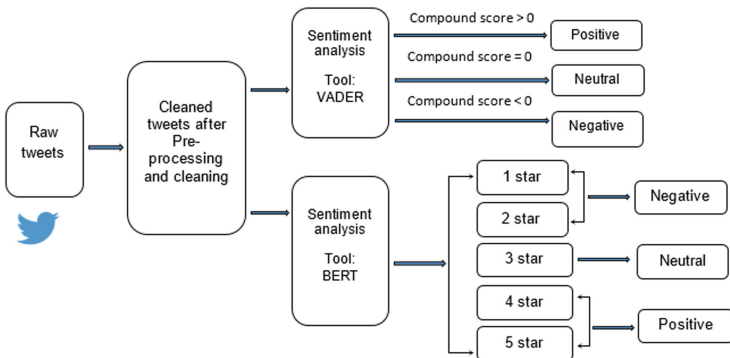


Fig. 1. Block diagram showing the Sentiment Analysis process

3.2 Sentiment Analysis

We used two main types of methods for sentiment or emotion analysis, Lexicon-based and Deep learning Based.

Lexicon-Based Method VADER: The detection of sentiment polarity (negative, positive, neutral, and complex) in tweets is done using VADER, a lexicon-based program that analyzes Twitter sentiments and categorizes tweets according to vocabulary. VADER’s dictionary differs from conventional dictionaries because it has acronyms, contractions, emoticons, and slang words that are frequently utilized in informal online interactions like those on Twitter. Moreover, VADER takes into account degree modifiers that affect sentiment intensity. In our study, the emotion of each tweet was evaluated using the VADER’s complex score, which ranges from -1 to 1 (-1 represents severe negative, and 1 implies extreme positive) [18]. A tweet was categorized as positive if its compound score was more than 0 (zero), negative if it had a compound score below 0 (zero), and neutral if it had a compound score of 0 (zero). These are the conventional cutoff points derived from the literature.

VADER examines a text for any known emotional components and modifies the intensity and polarisation in accordance with the rules to determine the sentimental score of the full text. Then, VADER adds the feature scores and uses the function to normalize the final score to $(-1, 1)$ as shown in Eq. 1.

$$x = \frac{x}{\sqrt{x^2 + \alpha}} \quad (1)$$

where x is the sum of sentimental scores, and α is the normalization constant and by default, its value is set to 15 [19, 20].

Deep Learning Based Method BERT: Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained word embedding model which is based on transformer-encoded architecture. Without needing extensive task-specific architectural changes, the pre-trained BERT model may be improved just with one more output layer to provide cutting-edge models for a number of tasks, such as question answering and language inference. It provides brand-new, cutting-edge findings on eleven natural language processing tasks [21, 22]. The “nlptown/bert-base-multilingual-uncased-sentiment” model, which has been optimized for sentiment analysis in six distinct languages, including English, is used in this work. The sentiment level provided by this model ranges from 1 to 5 , where 1 represents an extremely negative and 5 represents an extremely positive sentiment. In our study, we converted scores $1, 2$ to negative, 3 to neutral, and $4, 5$ to positive.

3.3 Model Building

Finally, we employed 5 popular supervised ML algorithms viz. Naive Bayes (NB), Random Forest (RF), Gradient boosting (GBC), XGBoost (XGB) and Support Vector Machine (SVM) to classify sentiments of tweets.

4 Experiment Setup and Results

4.1 Dataset

At first, Excel's built-in option was used to remove duplicate records. 835 duplicate entries were found and removed. We have collected 10 thousand tweets for #covid19 (COVID19_Dataset) and 10 thousand tweets for #omicron (Omicron_Dataset). We began by using regular expressions to eliminate mentions (@), Retweets (RT), links (https), and hashtags. Then, numbers, special characters, punctuations, and emoticons were eliminated from tweets. Finally, all characters were converted to lowercase. After that, Tokenization was done, which converts texts into tiny units. Tokenization assists in understanding the meaning by looking at the word order in the text. For stemming, Porter Stemmer was used in our study. This was accomplished by removing the suffix to create stems [23, 24]. The fully preprocessed tweets were then added to our existing data frame of tweets dataset and placed in a new panda's column named "cleaned tweets".

4.2 Word Embedding

As computers cannot analyze text data in its raw formats, data needs to be processed for training the ML models, so word embedding is used. These techniques are used to mathematically represent the words. Word Embedding technologies such as One Hot Encoding, Bag-of-words, TF-IDF, and Word2Vec are commonly utilized. Sentences from tweets are transformed into numerical vectors by these techniques. Depending on the condition, amount, and purpose of the data to be processed, one of these procedures (or multiple) is used. In this paper, TF-IDF (Term frequency-Inverse document frequency) was used as the word embedding technique. The statistical tool TF-IDF assesses a word's relevance to a document within a collection of documents. There are two terms that make up the TF-IDF weight.

Term Frequency (TF): TF measures how frequently a term occurs in a document and is calculated using Eq. 2.

$$TF = \frac{\text{the frequency of a phrase } (t) \text{ appears in a document}}{\text{total word count in the document}} \quad (2)$$

Inverse Document Frequency (IDF): IDF evaluates a term's significance. When computing the TF, each term is given an equal amount of weight. However, it is widely acknowledged that some words, such as "is", "of", and "that", may be used repeatedly but have little importance. In order to scale up the unusual words while scaling down the frequent ones, we must do so. by using Eq. 3 to calculate.

$$IDF = \log \log e \left(\frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}} \right) \quad (3)$$

TF-IDF: Tf-IDF can be calculated using Eq. 4.

$$TF - IDF(t, d) = TF(t, d) * IDF(t) \quad (4)$$

4.3 Experimental Results and Analysis

For the covid dataset, first, we used the VADER library for emotion analysis and then applied the selected machine learning algorithms to compare their accuracies. Later, we used the BERT mode on the same dataset for analyzing emotions. Finally, we employed machine learning algorithms in order to compare their performances. The whole process is repeated on the omicron dataset as well.

Figure 2(a) shows the comparison of tweet classifications between VADER and BERT on the covid dataset, where -1 , 0 , and $+1$ denote negative, neutral, and positive sentiment. The covid dataset contains 10000 tweets. 3890, 2390, and 3720 tweets of them are detected by VADER as negative, neutral, and positive tweets respectively. From the same dataset, BERT detects 7250 as negative, 490 as neutral, and 2260 as tweets expressing positive sentiments respectively.

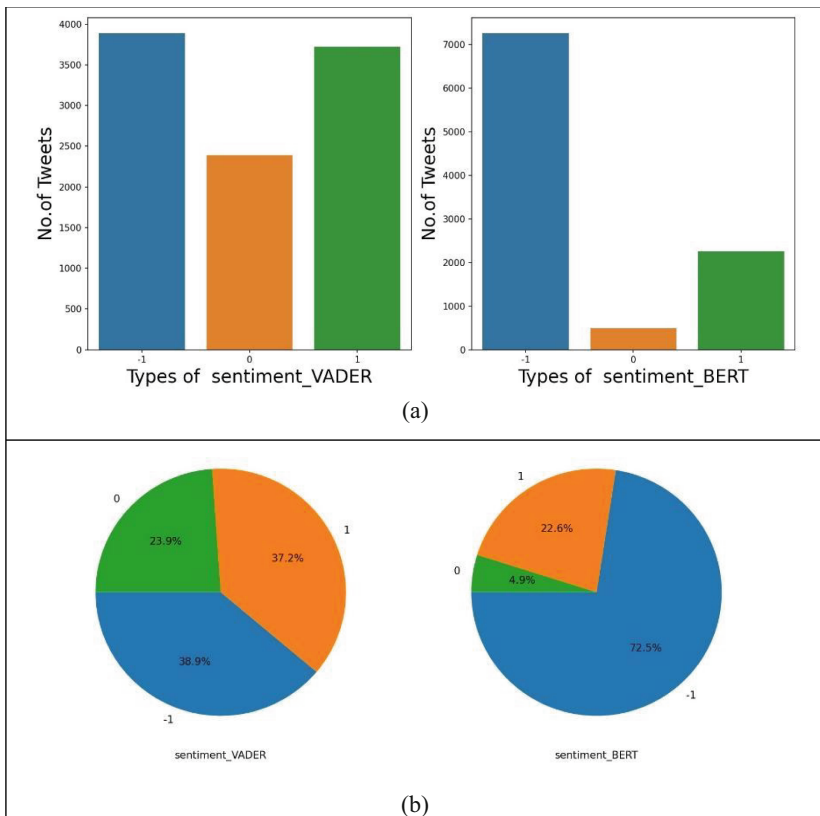


Fig. 2. (a) Comparison of the total number of tweets between VADER and BERT (b) comparison of percentages of tweets for each class between VADER and BERT on COVID19_Dataset (1 = Positive, 0 = Neutral, -1 = Negative sentiment).

Comparison of tweet classifications between VADER and BERT on the omicron dataset is illustrated in Fig. 3(a). In this case, VADER detects 2270 tweets as negative,

4010 tweets as neutral and 3710 tweets as positive tweets and BERT classifies 5850, 810, and 3350 tweets as negative, neutral, and positive sentiments respectively.

Both VADER and BERT classify most tweets as negative for the covid dataset but this trend is not followed in the omicron dataset. Here, VADER classifies the least number of tweets as negative and BERT detects the maximum number of tweets as negative.

VADER vs. BERT for COVID19 Dataset: After employing VADER on COVID19_Dataset, 38.9% of tweets are classified as negative whereas 23.9% and 37.2% of tweets are categorized as neutral and positive respectively and BERT finds 72.5% of tweets as negative, 4.9% tweet as neutral and 22.6% tweet as positive as shown in Fig. 2(b). Performances of machine learning algorithms are depicted in Table 1 and Table 2.

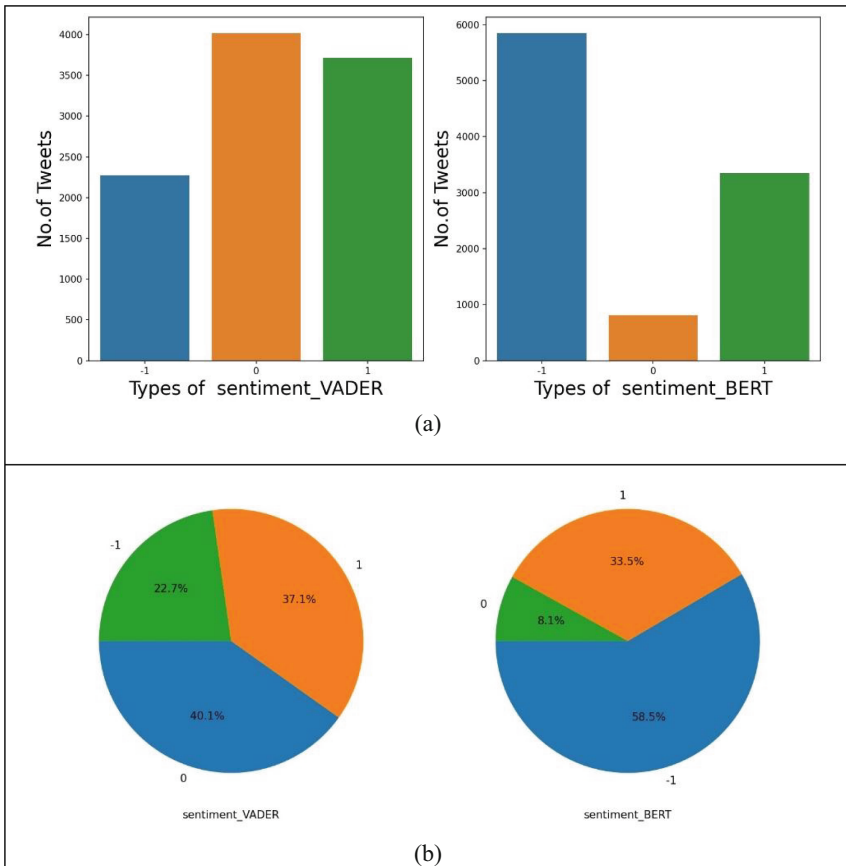


Fig. 3. (a) Comparison of the total number of tweets between VADER and BERT (b) comparison of percentages of tweets for each class between VADER and BERT on Omicron_Dataset (1 = Positive, 0 = Neutral, -1 = Negative sentiment)

Table 1. Result obtained from COVID19_Dataset after applying different ML algorithms using VADER (-1 = Negative, 0 = Neutral, 1 = Positive sentiment).

Algorithm	Sentiment	Precision	Recall	F1 measure	Accuracy
NB	-1	00.91	00.85	00.88	00.85
	0	00.86	00.73	00.79	
	1	00.79	00.91	00.85	
RF	-1	00.96	00.86	00.91	00.87
	0	00.78	00.89	00.83	
	1	00.87	00.87	00.87	
GBC	-1	00.95	00.73	00.83	00.78
	0	00.87	00.60	00.71	
	1	00.65	00.94	00.77	
XGB	-1	00.95	00.69	00.80	00.75
	0	00.85	00.58	00.69	
	1	00.63	00.93	00.75	
SVM	-1	00.95	00.90	00.92	00.89
	0	00.90	00.81	00.85	
	1	00.84	00.94	00.89	

Table 2. Result obtained from COVID19_Dataset after applying different ML algorithms employing BERT (-1 = Negative, 0 = Neutral, 1 = Positive sentiment).

Algorithm	Sentiment	Precision	Recall	F1 measure	Accuracy
NB	-1	00.82	00.97	00.89	00.84
	0	01.00	00.20	00.33	
	1	00.85	00.52	00.65	
RF	-1	00.88	00.99	00.93	00.89
	0	00.97	00.63	00.76	
	1	00.95	00.65	00.77	
GBC	-1	00.84	00.99	00.91	00.85
	0	00.95	00.53	00.68	
	1	00.91	00.50	00.64	
XGB	-1	00.82	00.99	00.90	00.83
	0	00.98	00.47	00.63	

(continued)

Table 2. (continued)

Algorithm	Sentiment	Precision	Recall	F1 measure	Accuracy
SVM	1	00.88	00.43	00.58	00.90
	-1	00.88	00.99	00.94	
	0	01.00	00.59	00.74	
	1	00.97	00.68	00.80	

For the covid19 dataset, SVM gives the highest accuracies of 89% and 90% and XGB gives the lowest accuracies of 75% and 83% for VADER and BERT respectively. RF gives the second best performances in both cases. Also, BERT improves accuracy for all the algorithms except Naive Bayes.

Figure 4 shows the word clouds of the most frequent words for VADER on covid19 dataset. From 4(a) “unpaid leave”, “leave”, “identity crisis”, “crisis” are prominent negative tweet words that indicate people’s distress towards possible loss of employment. With time more studies and analyses have been performed and people’s optimism has been growing which is indicated by positive words such as “insights”, “analytics”, “covid vaccine” as shown in Fig. 4(b). Also “name”, “coverage”, “beijing man”, “olympics” are some neutral words displayed in Fig. 4(c).

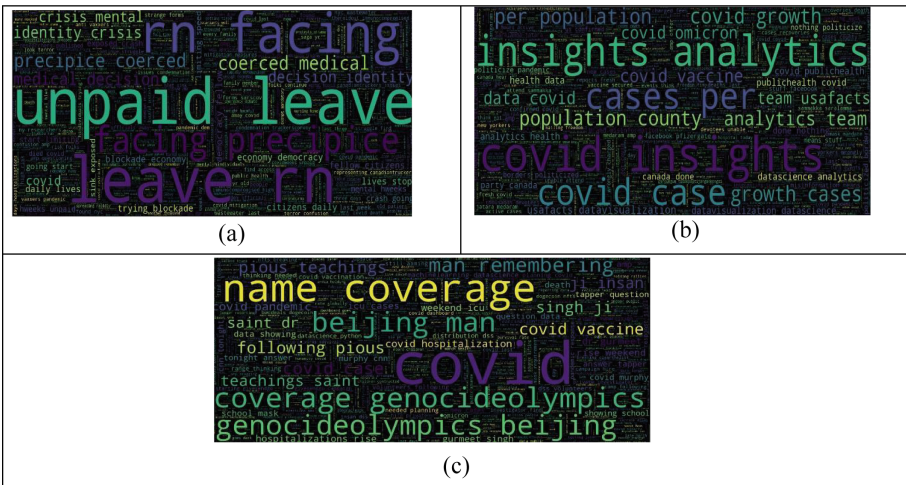


Fig. 4. Wordcloud for (a) Negative, (b) Postive and (c) Neutral sentiment using VADER library on COVID19_Dataset

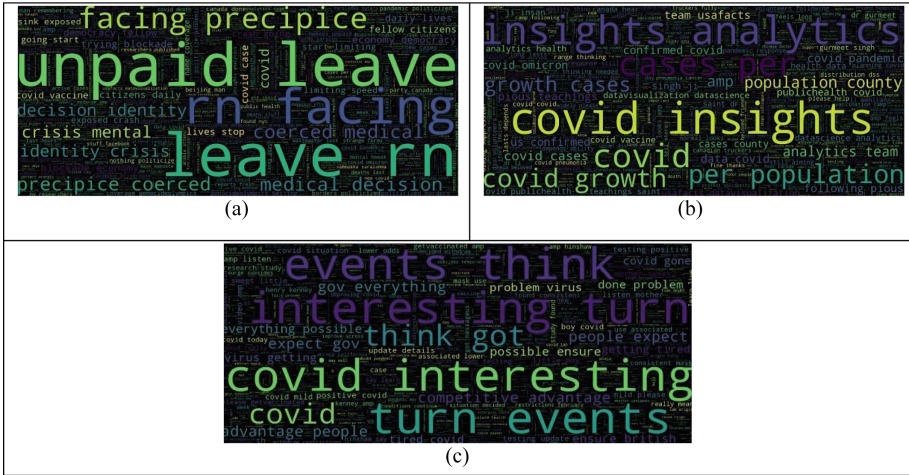


Fig. 5. Wordcloud for (a) Negative, (b) Positive and (c) Neutral sentiment using BERT model on COVID19 dataset

Figure 5 presents the word clouds for BERT. Negative and positive important words are almost similar to VADER as shown in Fig. 5(a) and (b). “Turn events”, “interesting”, “think” are some significant neutral words depicted in Fig. 5(c).

VADER vs. BERT for Omicron Dataset: After employing VADER on Omicron Dataset, 22.7% of tweets are classified as negative whereas 40.1% and 37.1% of tweets are categorized as neutral and positive respectively. On the contrary, BERT classifies 58.5% of tweets as negative, 8.1% as neutral, and 33.5% as positive shown in Fig. 3(b). The result after applying machine learning algorithms is shown in Table 3 and Table 4.

Table 3. Result obtained from Omicron_Dataset after applying different machine learning algorithms using VADER (-1 = Negative, 0 = Neutral, 1 = Positive sentiment).

Algorithm	Sentiment	Precision	Recall	F1 measure	Accuracy
NB	-1	00.81	00.82	00.81	00.85
	0	00.92	00.82	00.87	
	1	00.81	00.89	00.84	
RF	-1	00.94	00.82	00.87	00.91
	0	00.89	00.97	00.93	
	1	00.93	00.91	00.92	
GBC	-1	00.95	00.71	00.81	00.83
	0	00.88	00.82	00.85	

(continued)

Table 3. (continued)

Algorithm	Sentiment	Precision	Recall	F1 measure	Accuracy
XGB	1	00.75	00.92	00.83	00.80
	-1	00.92	00.65	00.76	
	0	00.88	00.79	00.83	
SVM	1	00.71	00.91	00.80	00.91
	-1	00.95	00.83	00.89	
	0	00.93	00.93	00.93	
	1	00.89	00.95	00.91	

Table 4. Result obtained from Omicron_Dataset after applying different machine learning algorithms using BERT (-1 = Negative, 0 = Neutral, 1 = Positive sentiment).

Algorithm	Sentiment	Precision	Recall	F1 measure	Accuracy
NB	-1	00.83	00.99	00.90	00.85
	0	00.98	00.58	00.73	
	1	00.92	00.45	00.60	
RF	-1	00.90	01.00	00.95	00.91
	0	01.00	00.74	00.85	
	1	00.98	00.70	00.82	
GBC	-1	00.87	00.99	00.93	00.89
	0	00.98	00.66	00.79	
	1	00.91	00.62	00.74	
XGB	-1	00.85	00.99	00.92	00.87
	0	00.99	00.62	00.76	
	1	00.93	00.53	00.68	
SVM	-1	00.90	01.00	00.94	00.92
	0	01.00	00.72	00.84	
	1	00.98	00.69	00.81	

For the Omicron dataset, SVM gives the highest accuracies of 91% and 92% respectively for VADER and BERT. On the other hand, XGB gives the lowest accuracy of 80% for VADER, and for BERT, Naive Bayes performs worst with an accuracy of 85%. RF gives the second-best performances in both cases. Also, BERT improves accuracy for all the algorithms.

Figure 6 displays the word clouds of the most frequent words for VADER on the omicron dataset. From 6(a) “covid”, “omicron”, “break us”, “continue second” are significant negative tweet words that indicate people are talking about frustrations and the continuation of the second waves of this pandemic. “insights analytics”, “analytics team”, “data covid” implies positive sentiment towards more research and analysis on this issue as shown in Fig. 6(b). Notable neutral words include “white tailed”, “first time”, “tailed deer”, “staten island” which means people are talking about the discovery of omicron in a wild animal for the first time illustrated in Fig. 6(c). Figure 7 presented the word clouds for BERT. Important negative, positive, and neutral words are the same as covid dataset.

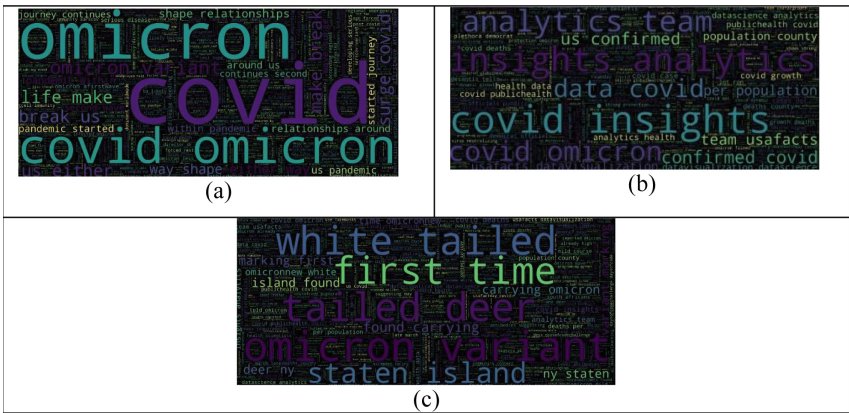


Fig. 6. Wordclouds for (a) Negative, (b) Positive and (c) Neutral sentiment using VADER library on omicron dataset

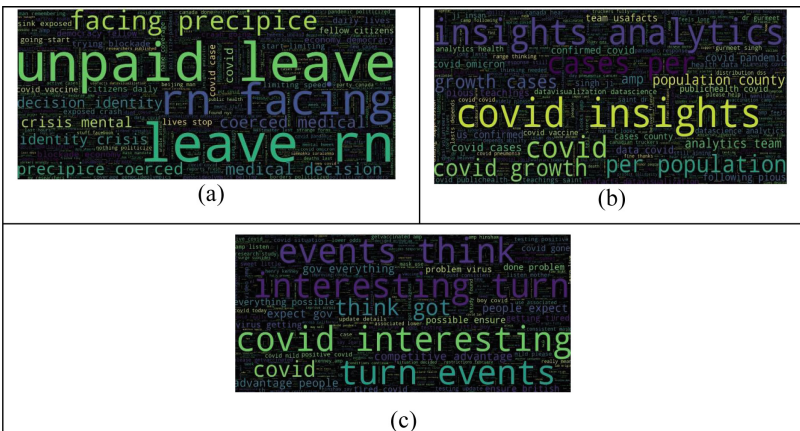


Fig. 7. Wordcloud for (a) Negative, (b) Positive and (c) Neutral sentiment using BERT model on omicron dataset

5 Conclusion and Future Works

The purpose of this study was to analyze public sentiment and emotions on covid-19 and omicron using VADER and BERT and then compare which technique performs well on two individual datasets. Data was collected from Twitter, as nowadays people like to share their concerns on social media. From our research, we can infer that most of the time VADER and BERT were both successful in identifying the negative tweets. In classifying sentiments, SVM performed best in all cases and most of the time XGB gave the lowest accuracy. RF algorithm also worked well with categorical data as evident from our study which gave the second-best result. After experimenting and analyzing the results, we can conclude that BERT outperforms VADER in all supervised algorithms (except in one case as shown in Table 1 and 2) because BERT is a transformer-based, deep left-to-right and right-to-left model. In the future, this study can be utilized in analyzing public sentiments on other covid variants, vaccinations, and on other social issues too. However, there are other algorithms and other combinations of algorithms that may improve the accuracy, which will be explored in the future.

References

1. Gligorić, K., Anderson, A., West, R.: Adoption of Twitter's new length limit: is 280 the new 140? arXiv preprint [arXiv:2009.07661](https://arxiv.org/abs/2009.07661) (2020)
2. Bao, Y., Quan, C., Wang, L., Ren, F.: The role of pre-processing in twitter sentiment analysis. In: Huang, D.-S., Jo, K.-H., Wang, L. (eds.) ICIC 2014. LNCS (LNAI), vol. 8589, pp. 615–624. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09339-0_62
3. Thu, P.P., New, N.: Implementation of emotional features on satire detection. In: 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 149–154. IEEE (2017)
4. Tonkin, E.L.: A day at work (with text): A brief introduction. In: Working with Text, pp. 23–60. Chandos Publishing (2016)
5. Rufai, S.R., Bunce, C.: World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *J. Public Health* **42**(3), 510–516 (2020)
6. Viana, R., et al.: Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 1–10 (2022)
7. Kreier, F.: Deltacron: the story of the variant that wasn't. *Nature* **602**(7895), 19 (2022)
8. Nair, A.J., Veena, G., Vinayak, A.: Comparative study of twitter sentiment on covid-19 tweets. In: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1773–1778. IEEE (2021)
9. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford I, no. 12 (2009)
10. Elbagir, S., Yang, J.: Twitter sentiment analysis using natural language toolkit and VADER sentiment. In: Proceedings of the International Multiconference of Engineers and Computer Scientists, vol. 122, p. 16 (2019)
11. Amin, A., Hossain, I., Akther, A., Alam, K.M.: Bengali vader: a sentiment analysis approach using modified vader. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–6. IEEE (2019)
12. Mao, J., Liu, W.: A BERT-based approach for automatic humor detection and scoring. In: IberLEF@ SEPLN, pp. 197–202 (2019)

13. Shamrat, F.M.J.M., et al.: Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. *Indones. J. Electr. Eng. Comput. Sci.* **23**(1) (2021)
14. Kaur, H., Ahsaan, S.U., Alankar, B., Chang, V.: A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets. *Inf. Syst. Front.* **23**(6), 1417–1429 (2021)
15. Mahbub, N.I., et al.: Sentiment analysis of microblogging dataset on coronavirus pandemic. In: 2021 5th International Conference on Electrical Information and Communication Technology (EICT), pp. 1–6 (2021)
16. Dubey, A.D.: Twitter sentiment analysis during COVID-19 outbreak. Available at SSRN 3572023 (2020)
17. Getting started—tweepy 4.6.0 documentation. https://docs.tweepy.org/en/stable/getting_started.html. Accessed 17 Mar 2022
18. Ridhwan, K.M., Hargreaves, C.A.: Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *Int. J. Inf. Manage. Data Insights* **1**(2) (2021)
19. Chandrasekaran, R., Mehta, V., Valkunde, T., Moustakas, E.: Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. *J. Med. Internet Res.* **22**(10), e22624 (2020)
20. Botchway, R.K., Jibril, A.B., Kwarteng, M.A., Chovancova, M., Oplatková, Z.K.: A review of social media posts from UniCredit bank in Europe: a sentiment analysis approach. In: Proceedings of the 3rd International Conference on Business and Information Management, pp. 74–79 (2019)
21. Tenney, I., Das, D., Pavlick, E.: BERT rediscovers the classical NLP pipeline. arXiv preprint [arXiv:1905.05950](https://arxiv.org/abs/1905.05950) (2019)
22. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) (2018)
23. Neogi, A.S., Garg, K.A., Mishra, R.K., Dwivedi, Y.K.: Sentiment analysis and classification of Indian farmers' protest using twitter data. *Int. J. Inf. Manage. Data Insights* **1**(2) (2021)
24. Jivani, A.G.: A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl* **2**(6), 1930–1938 (2011)