International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Exploratory Case Study on COVID 19 Omicron using Twitter Analytics

Sivagurunathan S[1], Chandrakumar Thangavel *[2] , Vaishnavi V[3], Prem Nitin P[4]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract: The Covid-19 pandemic is the most disruptive event worldwide and it majorly affects public health. Social media plays a significant role in people's lives, and constantly gets bombarded with messages, tweets, memes, and posts about Covid-19 and Omicron. The Omicron is another variant of Covid-19 which is widely spread across the globe thereby increasing the percentage of people being affected. During this research, the tweets are collected using the Twitter API to perform Sentiment Analysis. An exploratory case study has been developed using Twitter analytics, relying upon pragmatic evidence stemming from the case study about Covid-19 and Omicron. This research aims at scrutinizing people's thoughts and opinions regarding omicron and covid by comparing the results. The tweepy library for accessing the Twitter API and the Valence Aware Dictionary for Sentiment Reasoning (Vader), a lexicon and rule-based sentiment analysis tool accessible in the Python programming language, are used.

Throughout this research 80,000 tweets were fetched with hashtags #covid, #covid19, #coronavirus, #corona, and other sets of 80,000 tweets were fetched with hashtags #Omicron, #Omicold, #OmicronVariant, #OmicronVirus around the globe. The tweets were obtained from 22/12/2021 to 30/04/2022.

Keywords: *sentiment analysis, corona, social media, NLTK vadar*

1. Introduction

On January 26th, 2020, roughly 2761 people were infected in China as a result of a viral epidemic that occurred in a small seafood market in Wuhan, Hubei province, and this virus was later recognized as SARS-Co V-2 or coronavirus [1,2]. Coronavirus outbreak across the globe was declared a pandemic by the World Health Organization on 30th January 2020 (WHO). Later, in late November 2021, a new form of SARS-CoV-2 was discovered in Gauteng Province, South Africa. The new strain was dubbed Omicron, and it increased the number of infected individuals [3]. The virus led to the outbreak epidemic around the world from 20th May 2020 and has infected around 290,959,019 people and caused 5,446,753 deaths reported to WHO. Many countries have imposed lockdowns as a result of the growth in the number of cases of infected persons by omicron around the world, and people are sharing their concerns on social media. As of 30th December 2021, 58.7% of the world population have taken at least one dose, around 49.6% are fully vaccinated and among them, 6.6% of the people have taken boosters.

Social media has become an integral part of daily routine. It builds a connection throughout the world. It provides a way to vitrine our lives, discretely, conveniently.

The widespread usage of social media facilitates the flow of information and the expression of people's views on current situations [4]. The data can be collected from twitter as it provides millions of tweets every day [5]. One of the trending topics on Twitter for the past two years is Covid-19 and it is discussed by people to date. It doesn't allow the people to get to know the seriousness of the outbreak of Omicron but acts as a platform to share people's opinions about it. It also allows for the analysis of people's sentiment, which is dynamic during the virus's outbreak, offering insights regarding prevalent sentiment and its effects. Recently published research focused on the automatic recognition of tweets concerning Covid -19. Studies to identify the sentiment of tweets on Covid-19 are recently conducted [6-9].

Erik Cambria, Dipankar Das, et.al. (2017) show in their research that sentiment analysis helps us to find people's opinions from their interactions through social media thereby analysing the different aspects of the events [10]. Among the most important uses of Natural Language Processing (NLP) is to assess people's sentiment on social media using comments, tweets, and posts, among other things. To extract the tweets and analyse them python programming language is used, and the analysis is

[1] Computer Appl, Gandhigram Rural Deemed university, Tamilnadu, India
ORCID ID : 0000-0002-8545-7303
[2]Dept of AMCS, Thiagarajar College of Engineering, Tamilnadu, India
ORCID ID : 0000-0002-1186-5988
[3] Dept of AMCS, Thiagarajar College of Engineering, Tamilnadu, India
ORCID ID : 0000-0002-4897-1733
[4] Dept of AMCS, Thiagarajar College of Engineering, Tamilnadu, India
ORCID ID : 0000-0003-1476-5090
** Corresponding Author Email: t.chandrakumar@gmail.com*

done using Natural Language Toolkit (NLTK) Vader to categorize the tweets of people according to their sentiments.

The motivation behind this study is the massive spread of Covid-19 thus making the world suffer and decreasing the global economy. To protect themselves from covid people took vaccinations and they were at ease for a course of time but soon omicron started to infect people around the globe. This study focuses on comparing the sentiments of people after covid vaccination and omicron.The research is carried on by the following structure which includes literature that deals with existing works related to COVID and sentimental analysis. The research methodology consists of a detailed explanation of the methods and algorithms used followed by the result and discussion section which describes the output obtained. The conclusion and future enhancement were also discussed.

2. Literature Review/ Background:

The study conducted by Ahmed, W., Bath, P. A. et.al. (2017) shows that Twitter data can be used to carry on research [11]. Nowadays, many researchers are working with Twitter to analyze the sentiment of the people. To analyze people's behavior during covid pandemic researchers used Twitter, in this section includes some of the essential papers. Dubey, A. D. (2020) analyzed the people's emotions during the covid pandemic were anger, anticipation, contempt, fear, grief, joy, surprise, and trust, according to the NRC emotion lexicon. Positive and negative sentiments were separated into two categories. The outcomes were presented in a variety of visualization [12]. Pokharel, B. P. (2020) used TextBlob, a naive Bayes probabilistic model, to do sentiment analysis on Twitter data about covid. This study examined people's sentiments in Nepal from May 21st to May 31st, 2020, and also displayed the frequently tweeted terms about covid each day using WordCloud. [13].

R. Medford, S. Saleh, A. Sumarsono et.al. [14] retrieved tweets using hashtags related to Covid-19 and performed sentiment analysis to find the emotional valence and the dominant emotions of the people with Latent Dirichlet Allocation (LDA). J. Samuel, G. Ali, et.al. (2020) analyzed tweets to determine feelings and thoughts about Covid-19. R programming, as well as sentiment analysis packages, is utilized. Geo-tagged Analytics is used to analyze public data in the United States [15]. R. Muthausami, A. Bharathi et.al. (2020) in their study performed sentiment analysis of Twitter data regarding Covid-19. The machine learning methodology is used to analyze the tweets to arrive at an accurate result. A naive Bayes Classifier is used to classify the data into different sentiments [16]. S. BoonItt and Y. Skunkan (2020) researched the public's view of the Covid-19 outbreak on Twitter using Sentiment Analysis and topic modeling. The NRC Lexicon is used here to categorize the tweets as different sentiments. To find the most popular themes in tweets, topic modeling uses latent Dirichlet allocation (LDA) [17].

Sattar, N. S., and Arifuzzaman, S. (2021) conducted research in the United States to find the sentiment of the people using the tweets gathered from Twitter and the predicted immunization population. Classification of tweets into different sentiments is done with the help of TextBlob and VADER. To predict the vaccinated population Time series forecasting is used [18]. Shamrat, M. F. M. J., et.al. (2021) obtained data from Twitter using multiple hashtags related to covid 19. The K- nearest neighbor is a classification method used to classify tweets into positive, negative, and neutral attitudes, and the model was trained using the NLP model (TextBlob). Text polarity and subjectivity scores were calculated and visualized for three different vaccines namely Moderna, Pfizer, and AstraZeneca. [19]. Pastor, C. K. (2020) researched the feelings of persons in Filipinos who are quarantined because of covid-19. Twitter was used as the primary data source, while RapidMiner was used to collect it. The acquired data was entered into a spreadsheet and processed. The sentiments of the tweets were determined using the AYLIEN Sentiment Analysis API [20].

In Garcia, K., & Berton, L. (2021) research, the tweets were collected in English and Portuguese language. Natural Language Processing is used in both languages. NLTK library is used for English texts and SpaCy for Portuguese texts. The polarity /emotions classification was done with emotion intensity regression (EI-reg) and the average was calculated [21]. Naseem, U., Razzak, and others (2021) demonstrated in their study that proactive decisions must be made to combat the development of negative emotions among individuals. Tweets are retrieved via the Twitter API, and machine learning and deep learning classifiers such as Support Vector Machine, Nave Bayes, Decision Tree, and Random Forest are used to categorize the tweets [22]. Lwin, Sheldenkar (2020) has done Twitter Sentiment Analysis using CrystalFeel algorithm. The data was collected from January 28th, 2020 to April 9th, 2020 using covid-related hashtags. The tweets are classified into fear, anger, sadness, and joy. To demonstrate the trend of the emotions across time Pearson correlation is used and the global sentiment is found [23].

3. Research Methodology

3.1Data Collection:

Data was gathered using hashtags associated with Covid-19 and Omicron, such as #covid19, #covid, #coronavirus, #corona, #Omicron, #OmicronVariant, and #OmicronVirus. The Python programming language is used to authenticate with the Twitter Application Programming Interface (API), which allows us to get tweets using the tweepy package. The hashtags that are mentioned above are used as search terms in the tweepy cursor and the tweets are fetched from 22nd Dec 2021 to 30th Dec 2021. As a whole 80000 tweets are collected for both Covid and Omicron of which after removing duplicates 28,077 and 23,149 tweets were available in the dataset. The Tweets, the date it was posted, and the location of the tweet were noted in the data frame.

3.2 Data Pre-Processing:

Python is used to clean up the raw tweets/data. The pre-processing is done as follows [24]:

1. The texts from the tweets are tokenized and then converted to lower case.
2. Using Regular Expression (regex) the hashtags, usernames, RT, colon, and hypertext (URLs) from the tweets extracted are removed since it doesn't contribute any useful information for the analysis of sentiment.

3. The punctuations and special characters are also removed [25].

4. Non-Ascii characters are removed from the raw tweets as the only point of focus is based on English.

Table 1. shows the raw and pre-processed tweets

Raw Tweets	Pre-processed Tweets
_Breaking: Omicron is now the dominant covid strain in the u.s., rising from 3% of cases to 73% in just one week	Breakingomicron is now the dominant covid strain in the u.s rising from 3% of cases to 73% in just one week
Well, that escalated quickly. Omicron now makes up 92% of sequenced cases in the New York and New Jersey region	well, that escalated quickly omicron now makes up 92% of sequenced cases in the new york and new jersey region

3.3 Data Analysis:

The preprocessed data is being used to investigate people's responses to covid vaccination and Omicron. To evaluate people's feelings Python's Natural Language Toolkit (NLTK) package is used for Natural Language Processing (NLP). Sanyal, S., &Barai, M. K. (2021) proposed that the accuracy of VADER's negative sentiment classification is more elevated compared to TextBlob [26]. So, VADER is used instead of TextBlob for Sentiment Analysis. The tweets are classified into three categories as follows [13]:

- Positive
- Negative
- Neutral

Tocategorize tweet in VADER, the "SentimentIntensityAnalyzer" model is employed; It returns the probability of tweets appearing in positive, negative, and neutral classifications, as well as a compound score. The tweets are classified into the above-mentioned groups using the compound score.

Table 2. depicts the classification of sentiment according to the polarity score

Polarity Score	Classification
Less than zero	Negative Sentiment
Equal to zero	Neutral Sentiment
Greater than zero	Positive Sentiment

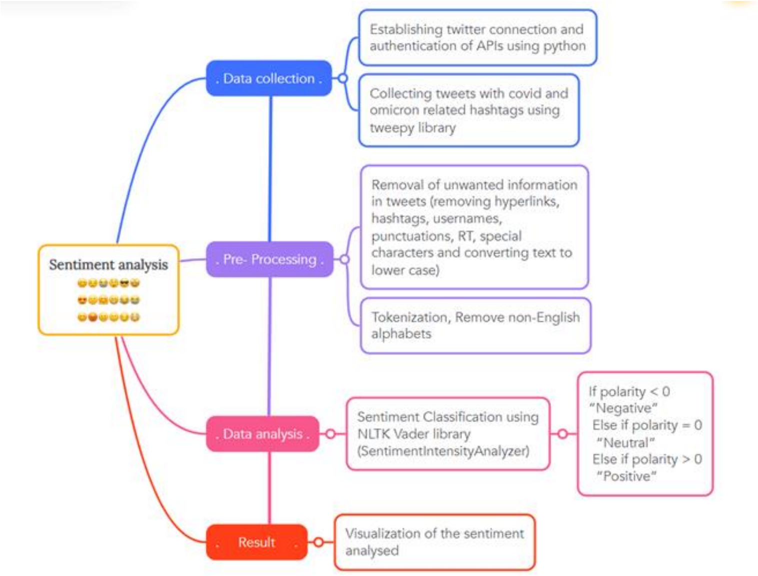
The tweets are categorized based on their polarity scores: if the polarity score is less than zero, it is defined as "Negative Sentiment," if the polarity score is equal to zero, it is labeled as "Neutral Sentiment," and if the polarity score is more than zero, it is classified as "Positive Sentiment."

```
Algorithm 1: Calculate Polarity Score
  Data: Tweet
Initialize: sid = SentimentIntensityAnalyzer();
Call sid.polarityscores(tweet);        /* returns dictionary with
  compound score */
return CompoundScore
```

4. Results and Discussion

```
Algorithm 2: Determine sentiment using polarity score
if $score < 0$ then
    return negative;
else if $score == 0$ then
    return neutral;
end
return positive;
```

The result of this study will be discussed in this section. The most frequently tweeted words which are related to Covid and Omicron are visualized with the help of WordCloud. The sentiments of the people are classified according to their polarity score and they are visualized using python they are shown below.



Sentiment analysis
. Data collection .
Establishing twitter connection and authentication of APIs using python
Collecting tweets with covid and omicron related hashtags using tweepy library
. Pre- Processing .
Removal of unwanted information in tweets (removing hyperlinks, hashtags, usernames, punctuations, RT, special characters and converting text to lower case)
Tokenization, Remove non-English alphabets
. Data analysis .
Sentiment Classification using NLTK Vader library (SentimentIntensityAnalyzer)
If polarity < 0 "Negative" Else if polarity = 0 "Neutral" Else if polarity > 0 "Positive"
. Result .
Visualization of the sentiment analysed

Fig. 1. Depiction of the flow of the research's methodology

4.1. Covid After Vaccination

Figures Fig.4. Represents the most tweeted words corona, covid, coronavirus, vaccine, hospital.The sentiments of people after Covid Vaccination are shown using a pie chart.People's positive sentiment reflects feelings such as happiness, surprise, delight, confidence, and optimism. Lockdowns were relaxed after vaccination, and fewer persons were infected with the virus, as well as the fatality rate.

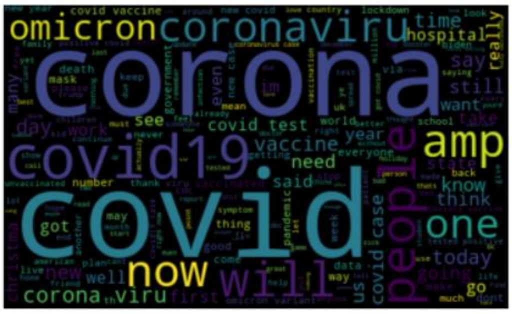


Fig 2: Most frequently used words regarding covid

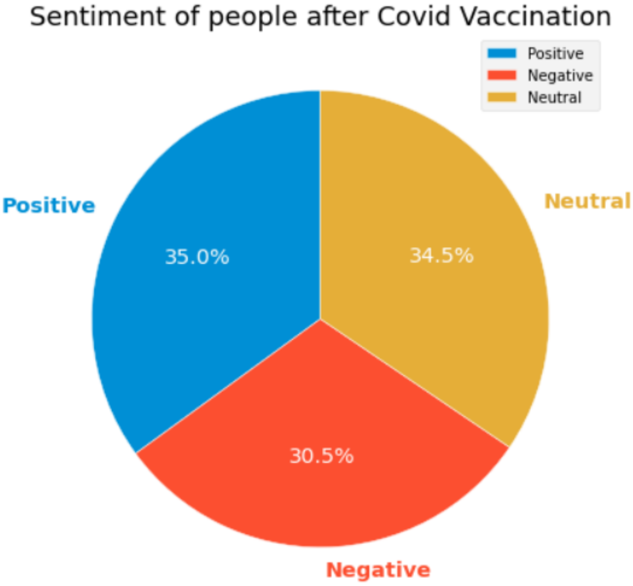


Fig. 1. Overall view of sentiments after covid vaccination

People were relieved to be able to resume their normal routines. The neutral sentiment of the people indicated emotions such as calm, relaxed, and no opinion. Even though people were vaccinated in large amounts there is no opinion regarding the large-scale vaccination. The negative sentiment of people indicates the emotions such as worry, depression, and fear. The lockdown was relaxed after vaccination, thereby making the classes offline. School and college students were worried since the examinations to be conducted were only offline. The work-from-home facility was also removed once after the relaxation of the lockdown. The focus of this study is to determine people's emotional and behavioral states following Covid vaccination. According to the above pie chart (Fig.5), around 35% of tweets have a good attitude, which is seen among persons after vaccination, followed by neutral sentiment and negative sentiment.

4.2. The emergence of Omicron:

The Fig.6. Represents the most tweeted words omicron, people, lockdown, vaccine, new, and omicron case. The sentiments of people after the emergence of the omicron are shown using a pie chart.

The positive sentiment of the public indicates feelings such as happiness, surprise, joy, confidence, hopeful. People aren't afraid of the outbreak of omicron as it doesn't have much mortality rate. The neutral sentiment of the people indicated emotions such as calm, relaxed, and no opinion. People don't have any awareness about omicron.

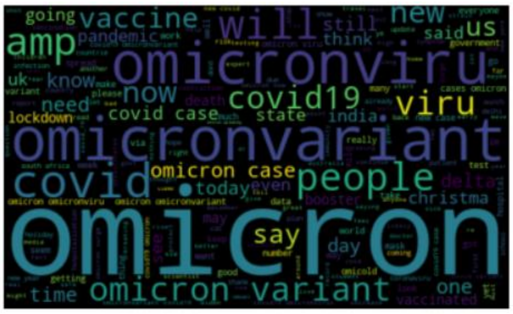


Fig. 2. Most frequently used words regarding omicron

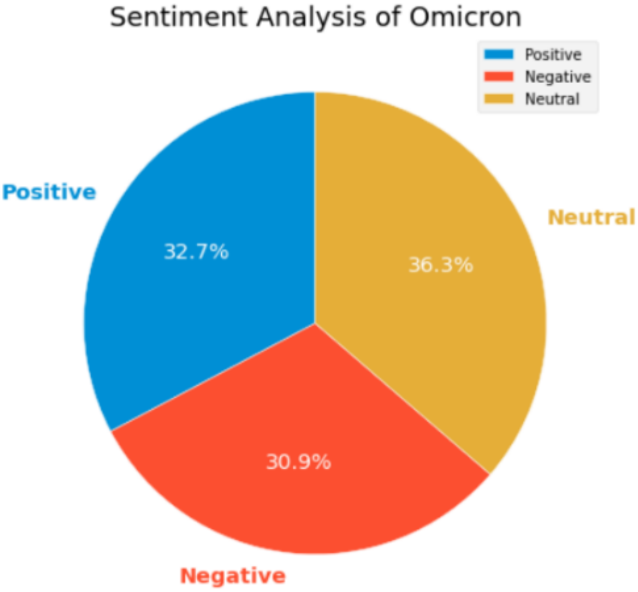


Fig. 3. Overall view of sentiments after the emergence of omicron

The negative sentiment of people indicates the emotions such as worry, de-pression, and fear. People don't wear masks or sanitizers adequately. Some people are worried about the outbreak of omicron. People fear that the number of people affected might increase. The goal of this study is to determine people's emotional and behavioral states of people after the emergence of the omicron variant. As it can be interpreted in the above pie chart (Fig.7,.), approximately 36.3% of tweets consist of neutral sentiment which is seen among the people after the emergence of the omicron variant followed by positive sentiment and negative sentiment.

5. Conclusion

A In a contemporary study, sentiment analysis was used on data from Twitter relating to international COVID-19 vaccination and omicron epidemics. From December 22nd to December 30th, 2021, tweets were collected following covid vaccination and an omicron outbreak. By comparing Fig.5. and Fig.7. the conclusion can be given as after the emergence of the omicron variant the positive sentiment is decreased while the negative sentiment is increased this depicts that the fear among the people is increased. One of the most preferred media to gather information during the pandemic or epidemic is Twitter and it is also highly effective. In the reference to the above study, it can be concluded that individuals' responses fluctuate from Coronavirus to omicron by posting their opinions via web-based media explicitly Twitter. Here NLTK VADER is used but there are plenty of tools available which can perform sentiment analysis and can give us varying results.

6. References

[1] Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. nature, 579(7798), 270-273.

[2] Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., ... & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269.

[3] Network for Genomic Surveillance in South Africa (NGS-SA). SARS-CoV-2

[4] Sequencing Update 26 November 2021 [Internet]. Network for Genomic

[5] Surveillance in South Africa (NGS-SA); 2021. Available from:

[6] https://www.nicd.ac.za/wp-content/uploads/2021/11/Update-of-SA-sequencingdata-from-GISAID-26-Nov_Final.pdf

[7] Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul;22(4):418-421 [FREE Full text] [CrossRef] [Medline]

[8] Kumar, S., Morstatter, F., & Liu, H. (2014). Twitter data analytics (pp. 1041-4347). New York: Springer.

[9] C. E. Lopez, M. Vasu, and C. Gallemore, "Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset," 2020,

[10] arXiv:2003.10359. [Online]. Available: http://arxiv.org/abs/2003.10359

[11] J. E. C. Saire and R. C. Navarro, "What is the people posting about symptoms related to coronavirus in Bogota, Colombia?" 2020, arXiv:2003.11159. [Online]. Available: http://arxiv.org/abs/2003.11159

[12] L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou, "'Go eat bat, chang!': An early look on the emergence of sinophobic behavior on Web communities in the face of COVID-19," 2020, arXiv:2004.04046. [Online]. Available: http://arxiv.org/abs/2004.04046

[13] J. Zhou, S. Yang, C. Xiao, and F. Chen, "Examination of community sentiment dynamics due to COVID-19 pandemic: A case study from Australia," 2020, arXiv:2006.12185. [Online]. Available: http://arxiv.org/abs/2006.12185

[14] Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, Antonio Feraco (Eds.), A Practical Guide To Sentiment Analysis, Springer International Publishing,Cham, Switzerland, 2017

[15] Ahmed, W., Bath, P. A., &Demartini, G. (2017). Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. The ethics of online research.

[16] Dubey, A. D. (2020). Twitter Sentiment Analysis during COVID-19 Outbreak. Available at SSRN 3572023.

[17] Pokharel, B. P. (2020). Twitter sentiment analysis during covid-19 outbreak in nepal. Available at SSRN 3624719.

[18] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. Lehmann, "An ' Infodemic ': Leveraging High -Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak," 2020.

[19] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification," SSRN Electron. J., no. May, pp. 1–21, 2020.

[20] R. Muthusami, A. Bharathi, and K. Saritha, "Covid-19 outbreak: Tweet based analysis and visualization towards the influence of coronavirus in the world," Gedragen Organ., vol. 33, no. 2, pp. 534–549, 2020.

[21] BoonItt, S., &Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. JMIR Public Health and Surveillance, 6(4), e21978.

[22] Gill, D. R. . (2022). A Study of Framework of Behavioural Driven Development: Methodologies, Advantages, and Challenges. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 09–12. https://doi.org/10.17762/ijfrcsce.v8i2.2068

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

[24] Shamrat, M. F. M. J., Chakraborty, S., Imran, M. M., Muna, J. N., Billah, M. M., Das, P., & Rahman, O. M. (2021). Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. Indonesian Journal of Electrical Engineering and Computer Science, 23(1), 463-470.

[25] Pastor, C. K. (2020). Sentiment analysis of Filipinos and effects of extreme community quarantine due to coronavirus (COVID-19) Pandemic. Available at SSRN 3574385.

[26] Pepsi M, B. B. ., V. . S, and A. . A. "Tree Based Boosting Algorithm to Tackle the Overfitting in Healthcare Data". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 5, May 2022, pp. 41-47, doi:10.17762/ijritcc.v10i5.5552.

[27] Garcia, K., & Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Applied Soft Computing, 101, 107057.

[28] Patil, V. N., & Ingle, D. R. (2022). A Novel Approach for ABO Blood Group Prediction using Fingerprint through Optimized Convolutional Neural Network. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 60–68. https://doi.org/10.18201/ijisae.2022.268

[29] Naseem, U., Razzak, I., Khushi, M., Eklund, P. W., & Kim, J. (2021). Covidsenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. IEEE Transactions on Computational Social Systems.

[30] Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR public health and surveillance, 6(2), e19447.

[31] 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 Sep 25;15(9):e0239441 [FREE Full text] [CrossRef] [Medline]

[32] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. New York, NY, USA: Springer; 2013.

[33] Sanyal, S., &Barai, M. K. (2021). Comparative Study on Lexicon-based sentiment analysers over Negative sentiment.