How to Cite:
Bahekar, K. B., Gautam, P., & Sharma, S. (2022). Global sentiment analysis over third wave COVID19 tweets. *International Journal of Health Sciences*, *6*(S10), 225–240. https://doi.org/10.53730/ijhs.v6nS10.13411

Global sentiment analysis over third wave COVID19 tweets

Kirti Bala Bahekar
Research Scholar, Department of Computer Science Engineering, RNTU University Bhopal, India
*Corresponding author email: kkirti06@gmail.com

Dr. Pratima Gautam
Department of Computer Science and IT, RNTU University Bhopal, India
Email: pratimaguatam@aisectuniversity.ac.in

Dr. Shailja Sharma
Department of Computer Science Engineering, RNTU University Bhopal, India
Email: shailja.sharma@aisectuniversity.ac.in

***Abstract*---**Internet users are increasing rapidly during the last decade, and after the Covid-19 outbreak, social media platforms became the favorite source to express public responses. They are using Twitter, a free microblogging site, to express their thoughts, joys, and sorrows spontaneously. Researchers take great interest in analyzing public sentiments with the help of Data science techniques like natural language processing and machine learning methods, to predict public suggestions on topics of social concerns. In the proposed research article, we have collected public tweets during the third wave of Covid19 from 21st to 31st January 2022, and public sentiments are observed with 12 popular Machine Learning algorithms and commonly used words are represented as n-grams and here three n-grams (Unigram, Bigram, and Trigram) are collected and prediction is also observed on these data. It is observed that in all the cases LinearSVC presents the highest classification accuracy of approx. 96% for covid datasets. It has also worked well on all the three n-gram datasets with accuracies of 96.39% for unigram, 95.24 for bigram, and 87.70% for trigram. During the third wave, the public has an abundance of faith in government policies and supported all the methods of controlling the pandemic including participation in a large number of vaccination drives. The conclusion of the study can be stated as citizens are overcoming their anxiety about the disease.

***Keywords*---**data science, natural language processing, machine learning, prediction, COVID19, N-grams.

Introduction

Covid19 outburst in March 2020, affected the lives of almost all people around the world, everybody witnessed thousands of death during the first wave encounter in March 2020 due to the SARS-CoV-2 variant of coronavirus, and the second wave due to the Delta variant of the virus is more destructive then its earlier form, this wave was active during April-May 2021(Kausar et al., 2021). During this medical emergency, almost all the scientific communities were engaged in developing an effective vaccine against this disease. As a result of their desperate efforts, many pharm companies come up with positive results in the vaccine. Within no time all the countries planned and executed several vaccination drives to protect their citizens from the deadly virus.

At the end of the year 2021, a new variant of coronavirus named Omicron came into the picture which has a high infection rate and it spreads speedily in almost every urban population on the globe. Apart from vaccination drives public suffered from the deadly various, which caused various up and downs in the public sentiment, at the same time they had lots of queries regarding the effectiveness of newly invented vaccines on upcoming variants of the virus. Nowadays, Twitter has become the usual channel to share thoughts, views, suggestions, and put queries. Analyzing public sentiment turns out to be possible through natural language processing methods of Data sciences(Aqlan et al., 2019). In the proposed experiment, we are exploring public opinions by using the machine learning method in two phases. The first phase involves an examination of views with Covid19-related tweets, in the second phase widespread words are collected from the tweets to generate n-grams datasets, here three n-grams are considered collection of single words known as 1-gram or unigram, a combination of two well-known words are 2-grams or bigrams, and three words phrases are 3-gram or trigram. We have analyzed the performance of 12 classifiers in both phases in terms of accuracies and time needed to construct the model by the methods.

Literature Review

During the ongoing pandemic, everybody adopted social media as a communication channel; hence sentiment analysis based on public reactions on social media is in trend and such researches are also helpful for the policymakers in such a crucial time of the pandemic. Various researchers are conducted opinion classification and using Natural language processing is in trends and some works among them are discussed as follows: To determine the prevalence of terms connected to prevention, racial bias, and infection prevention methods, (Medford et al., 2020) retrieved tweets about COVID-19. To determine sentimental valence and predominate feelings, they conducted a sentiment analysis. To identify and investigate popular conversation topics over time, they used topic modeling. They took out 126049 tweets sent by 53196 unique individuals. Tweets about COVID-19 began to appear more often starting on January 21, 2020. Nearly half (49.5%) of all posts were posted with feelings of dread, and about 30% of them were made with feelings of surprise. The rate of new COVID-19 positive cases and the number of racial posts were very closely related. The COVID-19's financial and political ramifications were frequently brought up in conversation.

Following the announcement of the lockdown by the Indian government, (Barkur et al., 2020) addressed the reactions of Indian residents. They made use of Twitter's social media platform for analysis. To determine the Indians' opinions on lockdown, they looked through the tweets. From March 25, 2020, to March 28, 2020, they retrieved tweets containing the two hashtags #IndiaLockdown and #Indiafight- corona. Using the R software, they analyzed 24000 tweets for the study and produced a search string that assesses the sentiment of the tweets. They discovered that despite the lockdown generating despair, anxiety, negativity, and disgust, positive sentiments predominated in the tweets. They concluded that Indians were committed to reducing the rate at which COVID-19 spread and were determined to reduce the loss.

The one lakh tweets are collected from Australia with the hashtag #coronavirus by (A.S.M Kayes, M.S Islamm, P.A watter , A. Ng, 2020). 3076 of these tweets contain the hashtags #socialdistancing and the term "social distancing." For training and validation, they utilized 8000 tweets, and for model testing, they use 2000 tweets. On the test data, they attained correctness of 83.70 percent and an F1-Score of 81.62 percent. Mostly on 3076 tweets that include the keyword "social distancing," they used the trained model. They noted that more than 80% of tweets that referenced "social distancing" expressed support for it, as seen in the illustration. They adopted that Australians accepted and encouraged the social distance method of prevention of Covid19. When Donald Trump and Narendra Modi were referenced in national tweets from India and the US, respectively, (Dubey, 2020) carried out a comparison of the opinions and feelings expressed in those tweets. Tweets were published between April 1 and April 9, 2020, and were collected for such opinion mining. The NRC Emotion Lexicon was used to examine the emotions and sentiments expressed in these tweets. They found that, compared to tweets about Donald Trump, 64.53 percent of those for Narendra Modi had positive feelings.

Tweets from April 5, 2020, to April 17, 2020, are collected by (Gupta et al., 2021) to observe the public sentiments on Lockdown in India, a country with a large and dense population and Covid19 widespread may cause great life loss. The Indian government has imposed the lockdown in several phases to control the widespread of the disease; the sentiment analysis of this situation supported the authorities to plan new policies. Natural language processing (NLP), and ML (Machine Learning) based eight algorithms were implemented for prediction, and it is observed that the LinearSVC method presents an accuracy of 84.4%. Popular N-grams and prediction of n-grams are also represented in the research work. Tweets collected on the social behavior of the public during a pandemic are analyzed by (Singh et al., 2021) they have collected tweets and observed the changes in public opinion during a pandemic for two datasets collected geographically, from the world and India. They used the BERT (Bi-directional Encoding representation) classifier model for prediction which gives an accuracy of approx. 94percent. During this phase the trends in society changed a lot, these changes are evaluated by M. (Mansoor et al., 2020) as tweets on corona, work from home, online studies, and some datasets present in data repositories like Kaggle, they had implemented Artificial neural network (ANN) and Long-short term memory algorithm (LSTM) for predictions with results 76% and 84.5%.

People are in favor of working from home and supporting online studies during this crucial time.

Materials and Methods

In this block of paper, we have proposed a methodology for sentiment analysis during the third wave of the pandemic. The structure is shown in figure 1. The steps followed for analyzing public views are as follows:

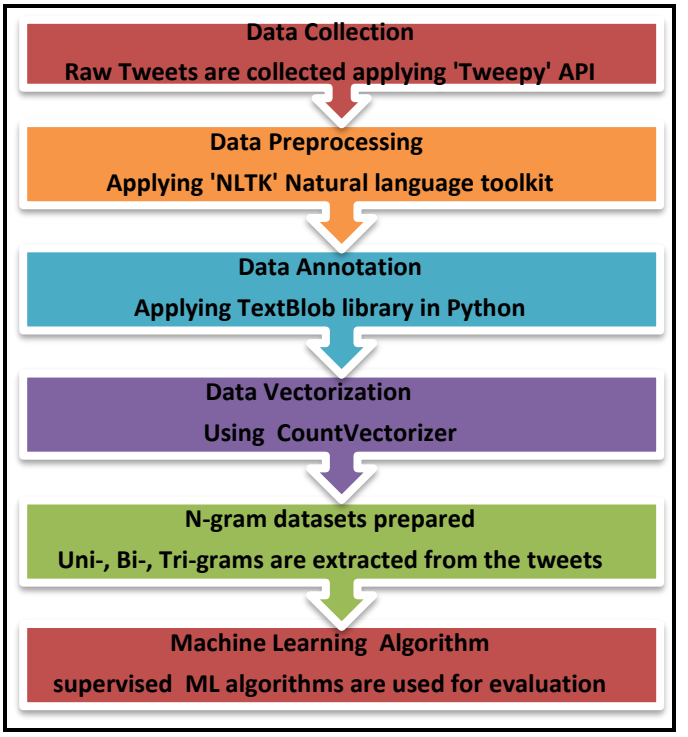


Figure 1. NLP methodology

Data Collection

In December 2021, a new variant of coronavirus came into existence named Omicron, the infection rate of this variant is very high and it affected almost 80% of the public who came in contact with an infected person, during this phase, people have many doubts on the usefulness of the Covid-19 vaccines available in the market. So, here is the study to evaluate the public outlook during the period. Tweets are collected during the third wave of the pandemic(Hung et al., 2020; Kwan, 2020) i.e. from 21 January 2022 to 31 January 2022 applying the Tweepy Python library. A covid19-related trending hashtag is observed to design the dataset using "Covid", "Covid19", and "coronavirus". 8910 tweets are collected to form a dataset; further common n-grams are also extracted from the data as Unigram, Bigram, and Trigram datasets.

Data Preprocessing

The raw tweets are collected from Twitter which holds spontaneous and slight words(Ikonomakis et al., n.d.) like links, hashtags (includes #), and tags (includes @). Single letter words, numbers, etc. These type of words acts as impurities for the training and test process, to apply the classifier efficiently these noise word should be removed.Therefore, Data preprocessing is required to separate noisy words and phrases from the dataset. The steps mentioned in figure 2 are applied to preprocess data like removing URLs and hashtags, Cleaning data by removing meaning-less words, removing stopwords, deleting punctuations, stemming and applying vectorization to the text, and data set in converted into a data frame for further processing.

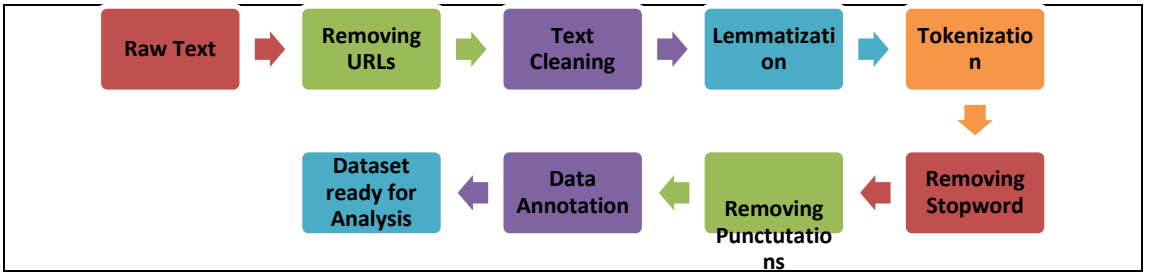


Figure 2. Steps applied in Data Preprocessing

Data Annotation

After implementing data preprocessing data annotation is applied to label the tweets TextBlob library of python (Prakash, 2020) is used to assign Positive, Negative, and Neutral classes to the tweets. After performing this step, among the total 8910 tweets, 3395 positive, 1416 negative, and 4099 had neutral sentiments polarity.

Table I
Distribution of total Covid19 Tweets

Covid19 Sentiments	
Neutral	4099
Positive	3395
Negative	1416
Total Tweets	8910

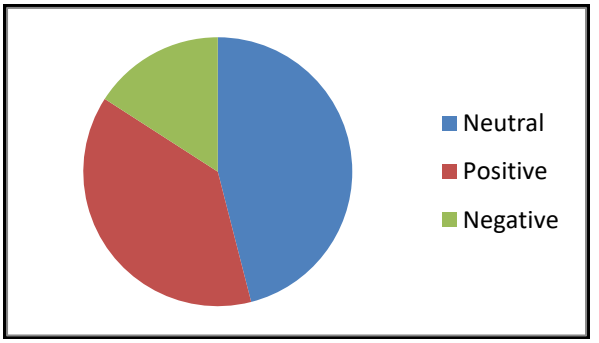


Figure 3. Distribution of Covid19-related Tweets

Data Vectorization

The machine learning classifiers accept input only in form of numbers excluding words in any language, therefore before the prediction process all the text is featured with polarity, TextBlob NLTK library (Shelar & Huang, 2018) is used to assign polarities to the tweets, and according to the polarities text is categorized, if the polarity is less than zero it is assigned as negative sentiment, for greater than zero value it is positive otherwise, it is neutral sentiment. These datasets can be further used for classification and prediction methods.

N-gram Extraction

New datasets are collected by the generation of common and repeated n-grams in the tweets, Unigram of one word, Bigrams combination of two words and Trigrams are three popular word arrangements (Aljameel et al., 2021), those are collected from the text of the databases, in this n-gram are assigned the polarity using TextBlob library and classified into three categories positive, negative, and neutral.

Supervised ML classifier

Data preprocessing, Vectorization and annotation are followed by passing datasets into various machine learning algorithms for prediction. In all we have considered 12 supervised machine learning algorithms(Chintalapudi et al., 2021):

- MultinomialNB(): Multinomial NaiveBayes algorithm.
- BernoulliNB():Bernoulli NaiveBayes algorithm
- LogisticRegression(): Logistic Regression algorithm
- LinearSVC(): A Linear- Support Vector Classifier algorithm
- AdaBoostClassifier():An AdaBoost classifier.
- RidgeClassifier: Classifier using Ridge regression.
- PassiveAggressiveClassifier():Passive Aggressive Classifier
- Perceptron():Linear perceptron classifier
- KNeighborsClassifier():Classifier implementing the k-nearest neighbors vote
- DecisionTreeClassifier():A decision tree classifier.
- LinearDiscriminantAnalysis(): A Linear Discriminant analysis
- MLPClassifier: Multi-layer Perceptron classifier

Experimental Setup

The experiment is performed using a Laptop with an i3 processor with 4GB ram, Data collection, and the pre-processing task is designed in an Anaconda Jupyter notebook in python. Data is collected using the "Tweepy" API of Twitter to extract required tweets, NLTK tool kit libraries are used for pre-processing of data, Matplotlib, pandas, and NumPy is also used, and dataset are categorized as Negative, Positive, and Neutral statements using Textblob, and Sklearn libraries are used for classification purposes.

Results and Discussions

The most effective method for assessing the performance of a machine learning algorithm is K-fold cross-validation. The k-fold cross-validation technique's re-sampling strategy is very helpful for assessing the effectiveness of any machine learning approach. In this method, the dataset is split into k equal folds parts and used to train the model. One fold is then deployed as a test dataset for the cross-validation score, and the remaining k folds are employed to record the results for this specific permutation. A fresh fold is used as a testing set and the rest is used as a training set as this method is repeated k times. The final cross-validation score of a system is then derived as the mean of the scores of all the permutations.

In the first phase of the research, we collected tweets on Covid using various related popular hashtags i.e. "#Covid", "#Covid19", and "#coronavirus".After performing all necessary data preprocessing, data are trained and tested by using 10-fold cross-validation methods. Then the results of all the classifiers are compared in the measure of accuracy, mean cross-validation score, and execution time of the classifier on the Covid19 data set. Along with that, n-gram datasets i.e. unigram, bigram, and trigram datasets are also evaluated using 10 fold cross validation method. Accuracy (Raza et al., 2019)is the major criterion to compare the performances of the different algorithms, which can be calculated by the formulas given below:

Accuracy= No. of correct predictions/No. of total predictions (1)

In simplified terms it can also be expressed as:

Accuracy= (TP+TN)/Total predictions (2)

Where the total number of predictions are total of true positives (TP) is correctly predicted positive values, true negative (TN) is correctly calculated negative values, false positive (FP) is an incorrect prediction of positive values, and false negative (FN) is incorrectly predicted negative values. Figure 4 depicts the representation of positive and negative classification.

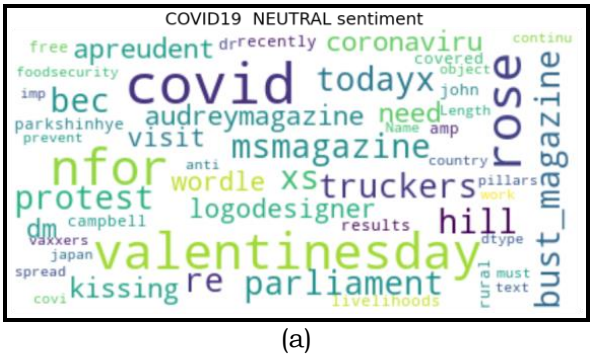
		Actual Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figure 4. Pictorial representation of Confusion matrix

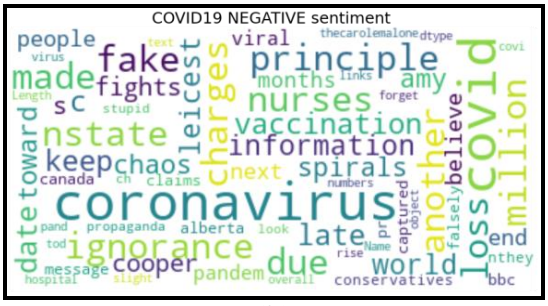
The most commonly used words in the tweets can be represented by the word cloud, figure 5 gives the display of frequently used words for representing various types of sentiments in the tweets.

Table II
Accuracy of the classifiers on the Covid19 dataset

Accuracies of Classifiers on 10-Fold method COVID COMBINE Tweets			
Classifier	Accuracy	Mean Cross validation Score	Rank
MultinomialNB()	74.28	0.74285	11
BernoulliNB()	76.12	0.76122	10
LogisticRegression()	85.17	0.85172	2
LinearSVC()	85.97	0.85971	1
AdaBoostClassifier()	78.87	0.78872	9
RidgeClassifier	83.46	0.83460	5
PassiveAggressiveClassifier()	83.31	0.83306	6
Perceptron ()	83.74	0.83740	4
KNeighborsClassifier()	59.27	0.59273	12
DecisionTreeClassifier()	84.62	0.84624	3
LinearDiscriminantAnalysis()	79.55	0.79546	8
MLPClassifier	81.78	0.81776	7



(a)



(b)

(c)

Figure 5. Words used in representing (a) Neutral (b) Negative and (c) Positive sentiments

Table II represents the experimental results of all the classifiers on the Covid19 dataset, the table contains the accuracy score, mean cross-validation score, and time taken by the algorithms to build a model. The accuracy of 85.1715% and mean cross-validation score of 0.85971 of the LinearSVC classifier are higher among all the other classifiers considered. The performance of all the predictor methods on the Covid19 dataset is shown in Figure 6. The time required by the classifiers is given by execution time; it is displayed in table III and fig.7 below. The classifier Logistic Regression requires less time to build the model for classification on the training data.

Table III
Execution Time

Classifier	Time (in a sec)
MultinomialNB()	11.4388
BernoulliNB()	13.9662
LogisticRegression()	8.15251
LinearSVC()	31.4771
AdaBoostClassifier()	398.513
RidgeClassifier	31.8507
PassiveAggressiveClassifier()	109.385
Perceptron()	36.8449
KNeighborsClassifier()	698.258
DecisionTreeClassifier()	577.772
LinearDiscriminantAnalysis()	461.975
MLPClassifier	1187.61

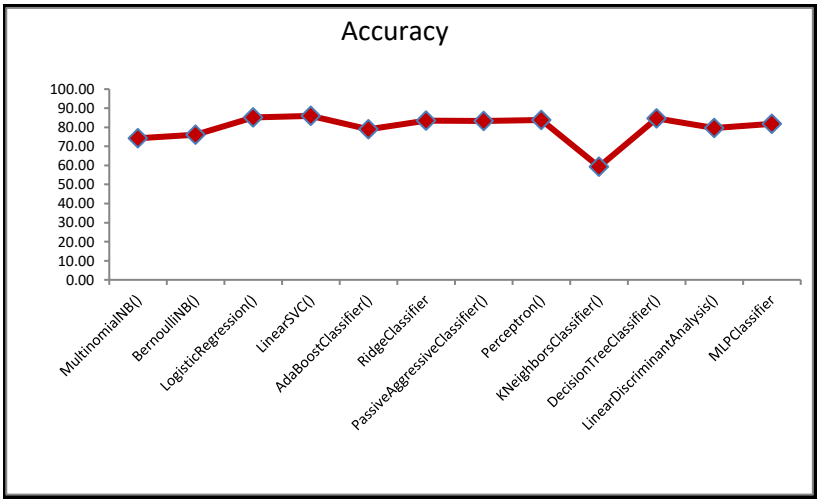


Figure 6. Accuracy of different classifiers with Covid19 data

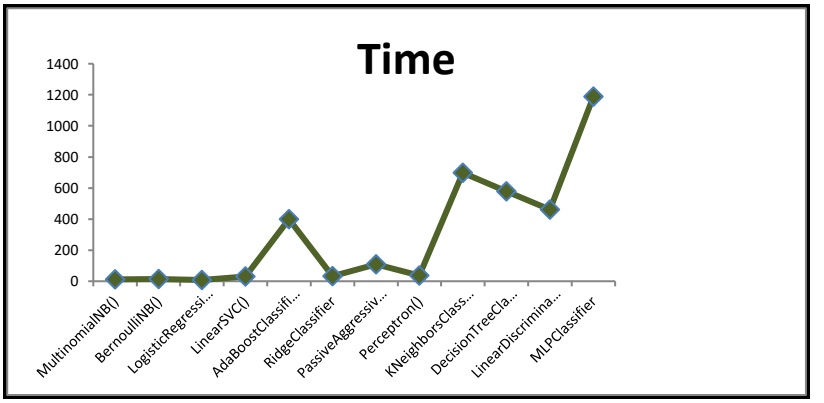


Figure 7. Time required by the classifiers to build model.

Efficient Observation: After evaluating the results of the first phase of the experiment, we have observed that the LinearSVC algorithm of the support vector machine best-fit algorithm performed well. The model is future used for the evaluation of the Confusion matrix and AUC-ROC curve. A table applied to examine the performance of the classifiers is represented as a Confusion matrix, where data is represented whose true values are known, here table IV displays the frequency distribution by the LinearSVC model of the test dataset. ROC (Receiver Operating Characteristics) curve plots true positive rate and false positive rate. Since the ROC curve is used to examine binary models, here for multiclass problems, n curves are generated for the n-classification model, considering one class at a time. The experiment has three classes, therefore three ROC curves are drawn in figure 8-10 with neutral, negative, and positive ones, having AUCs (Area Under the ROC Curve) 0.951, 0.923, and 0.940 respectively.

Table IV
Confusion Matrix by LinearSVC

Class	Negative	Neutral	Positive
Negative	178	42	29
Neutral	24	775	35
Positive	24	58	617

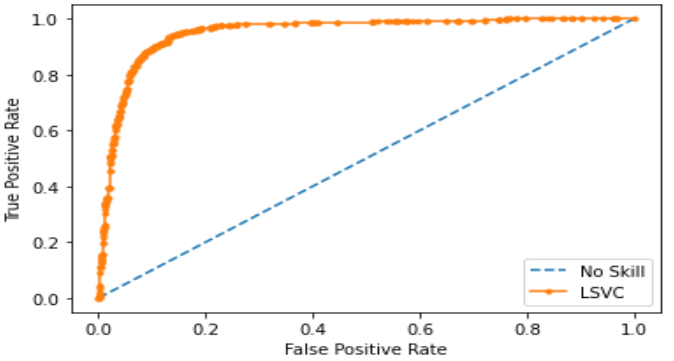


Figure 8. AUROC Curve of LinearSVC classifier for Neutral class

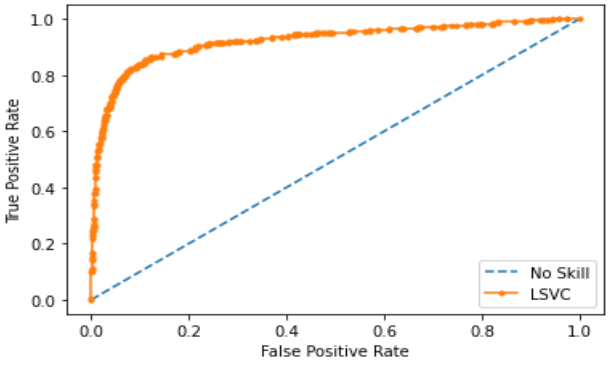


Figure 9. AUROC Curve of LinearSVC classifier for Negative class

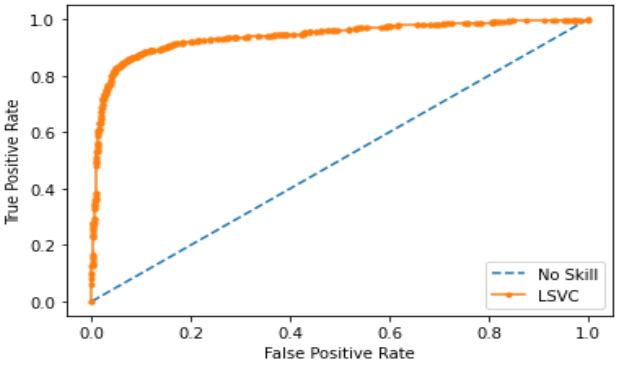


Figure 10. AUROC Curve of LinearSVC classifier for Positive class

Now, in the second phase of the research experiment, the commonly used combination of words are observed as n-grams where a single word is called Unigrams, and repeated combinations of two and three words are called bigrams and trigrams respectively. Table V represents the size of unigram, bigram, and trigram datasets, and their sentiment classification according to the polarity value if the polarity is zero then they are categorized as neutral, for less than zero it is categorized as negative; in other cases it is positive. We have also calculated each classifier's cross-validation score in the case of the unigrams, bigrams, and trigrams Table V shows the opinion representation of popular words in 1-, 2-, and 3- gram datasets. Table VI shows accuracies results with different classifiers with n-gram datasets. It is observed that in the case of unigrams or 1-grams most of the algorithms present the same results, whereas In the case of bigram and trigram LinearSVC algorithms (Gupta et al., 2021) give satisfactory results. In table VII time required by the classifier to build the model is shown, it is observed that the Logistic regression algorithm also produces considerable accuracies in less time as compared to all other algorithms in the case of bigram and trigram.

Table V
Sentiment distribution in n-gram datasets

N-Gram Sentiment Distribution				
n-gram	Neutral	Negative	Positive	Total
Unigram	17146	302	332.00	17780
Bigram	47051	2939	4947.00	54937
Trigram	45680	4327	8034.00	58041

Table VI
Accuracies of Classifiers with Uni, Bi, and Trigrams

Classifier	Unigram Accuracy	Bigram Accuracy	Trigram Accuracy
MultinomialNB()	96.39	95.16	87.69
BernoulliNB()	96.39	95.17	87.69
LogisticRegression()	96.39	95.16	87.70
LinearSVC()	96.39	95.24	87.70
AdaBoostClassifier()	96.23	92.49	87.67
RidgeClassifier	96.39	95.22	87.67
PassiveAggressiveClassifier()	96.26	95.14	87.36
Perceptron()	96.39	95.04	80.85
KNeighborsClassifier()	96.39	92.18	85.30
DecisionTreeClassifier()	95.63	95.18	87.20
LinearDiscriminantAnalysis()	96.39	95.26	87.67
MLPClassifier	96.34	95.09	87.44

Table VII
Time taken by Classifiers with Uni, Bi, and Trigrams

Classifier	Unigram Time	Bigram Time	Trigram Time
MultinomialNB()	0.92	20.26	5.57
BernoulliNB()	0.95	23.08	6.35
LogisticRegression()	1.42	9.51	4.74
LinearSVC()	310.27	14.68	17.13
AdaBoostClassifier()	20.91	663.25	157.63
RidgeClassifier	0.95	25.02	5.04
PassiveAggressiveClassifier()	0.70	41.60	10.01
Perceptron()	0.70	25.11	6.36
KNeighborsClassifier()	13.10	3211.81	1179.54
DecisionTreeClassifier()	1.36	2732.42	226.73
LinearDiscriminantAnalysis()	1.13	185.08	25.83
MLPClassifier	472.13	2425.57	860.35

Figure 11 shows the frequency distribution of some popular words used in tweets. Covid, cases, coronavirus, new, deaths, and vaccine are some commonly used words in the tweets. Based on these words datasets are designed with three n-grams considered and various classifiers are applied to them Figure12 shows that a graphical representation of the accuracies and time required to build a model is represented in the graph as Figure 13, the time required by KNN Classifier (Babu & Katta, 2015)is very high in almost all the cases whereas in bigram and trigram Logistic regression (LR) algorithm(Raza et al., 2019)requires lesser time.

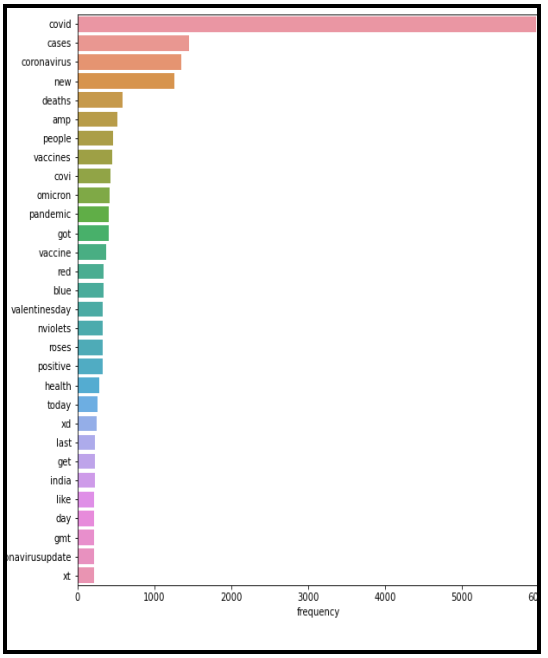


Figure 11. Frequency distribution of popular words in tweets

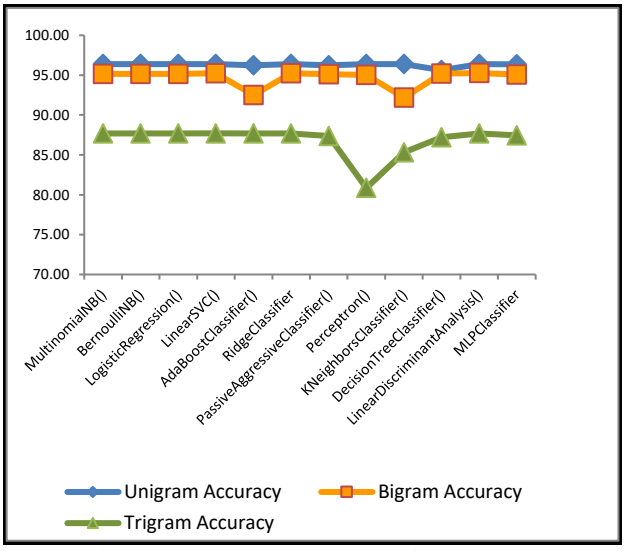


Figure 12. Accuracies with n-grams

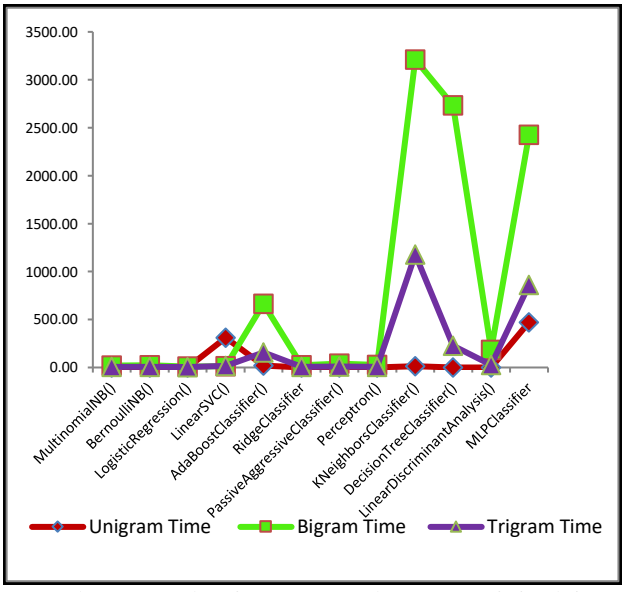


Figure 13. Time required constructing a model with n-grams

Conclusions

In the present study, we have analyzed the public sentiment during Covid19 third wave, though they have a lot of queries. Still, more than half of the tweets hold neutral polarity, which helps in controlling the panic situation during a tough time. Evaluation on Covid19 data shows LinearSVC algorithm indicates satisfactory results of accuracy of 85.9% in the limited amount of time. In the investigation n-grams (1-, 2-, and 3-grams) datasets are designed for classification accuracies, it is observed that in the case of unigram most of the datasets present the same results of 96.39%, and in the case of bigram Linear

Discriminant Analysis method and LinearSVC algorithms produce results of 95.26% and 95.24%, and in trigram dataset Logistic Regression and LinearSVC method produces same results with 87.70%. Overall, we can state that after people had been satisfied to a great extent with the vaccination drives and they are almost relaxed with the government policies to control health emergencies. They have noticed that the newly emerged Omicron variant of coronavirus is also not disastrous as its previous Delta version. These kinds of projects could help the Medical health authorities and officials to control situations where there are health emergencies, like Covid19. Because of health crises, the field of emotion analysis with machine learning methods, and NLP, has expanded significantly in the years after it was first developed. Studying the elements that may influence public mental health during crucial health emergencies and the effects of false information on the populace can both help the government and policymakers keep the situation under control(Maynard & Funk, 2011). Future research can investigate a huge corpus and look to increase the model's accuracy from a technical standpoint.

References

A.S.M Kayes, M.S Islamm, P.A watter , A. Ng, and H. kayes. (n.d.). View of Automated measurement of attitudes towards social distancing using social media_ A COVID-19 case study _ First Monday. https://doi.org/https://doi.org/10.5210/fm.v25i11.10599

Aljameel, S. S., Alabbad, D. A., Alzahrani, N. A., Alqarni, S. M., Alamoudi, F. A., Babili, L. M., Aljaafary, S. K., & Alshamrani, F. M. (2021). A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent covid-19 outbreaks in Saudi Arabia. International Journal of Environmental Research and Public Health, 18(1), 1–12. https://doi.org/10.3390/ijerph18010218

Aqlan, A. A. Q., Manjula, B., & Lakshman Naik, R. (2019). A study of sentiment analysis: Concepts, techniques, and challenges. In Lecture Notes on Data Engineering and Communications Technologies (Vol. 28). Springer Singapore. https://doi.org/10.1007/978-981-13-6459-4_16

Babu, M. S. P., & Katta, S. (2015). Artificial immune recognition systems in medical diagnosis. Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, 2015-Novem, 1082–1087. https://doi.org/10.1109/ICSESS.2015.7339240

Barkur, G., Vibha, & Kamath, G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian Journal of Psychiatry, 51(January). https://doi.org/10.1016/j.ajp.2020.102089

Chintalapudi, N., Battineni, G., & Amenta, F. (2021). Sentimental analysis of COVID-19 tweets using deep learning models. Infectious Disease Reports, 13(2), 329–339. https://doi.org/10.3390/IDR13020032

Dubey, A. D. (2020). Decoding the Twitter Sentiments towards the Leadership in the times of COVID-19: A Case of USA and India. SSRN Electronic Journal, April, 19–22. https://doi.org/10.2139/ssrn.3588623

Gupta, P., Kumar, S., Suman, R. R., & Kumar, V. (2021). Sentiment Analysis of Lockdown in India during COVID-19: A Case Study on Twitter. IEEE Transactions on Computational Social Systems, 8(4), 939–949. https://doi.org/10.1109/TCSS.2020.3042446

Hung, M., Lauren, E., Hon, E. S., Birmingham, W. C., Xu, J., Su, S., Hon, S. D., Park, J., Dang, P., & Lipsky, M. S. (2020). Social network analysis of COVID-19 sentiments: Application of artificial intelligence. Journal of Medical Internet Research, 22(8), 1–13. https://doi.org/10.2196/22590

Ikonomakis, E. K., Kotsiantis, S., Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (n.d.). Text Classification Using Machine Learning Techniques EVOLUTIONARY INTELLIGENCE (SPRINGER): Special Issue on “Intelligent and Fuzzy Systems in Data Science and Big Data” View project Data preprocessing View project Text Classification Using Machine Learning Techniques. https://www.researchgate.net/publication/228084521

Kausar, M. A., Soosaimanickam, A., & Nasar, M. (2021). Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak. International Journal of Advanced Computer Science and Applications, 12(2), 415–422. https://doi.org/10.14569/IJACSA.2021.0120252

Kwan, J. S. (2020). Understanding Public Sentiments , Opinions and Topics about COVID-19 using Twitter. June, 623–626. https://doi.org/10.1109/ASONAM49781.2020.9381384

Mansoor, M., Gurumurthy, K., U, A. R., & Prasad, V. R. B. (2020). Global Sentiment Analysis Of COVID-19 Tweets Over Time. October 2020. http://arxiv.org/abs/2010.14234

Maynard, D., & Funk, A. (2011). Automatic detection of political opinions in tweets. CEUR Workshop Proceedings, 718, 81–92.

Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. (2020). An “ Infodemic ”: Leveraging High -Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak.

Prakash, bishow. (2020). Twitter Sentiment analysis during COVID-19 Outbreak in Nepal Bishwo Prakash Pokharel M . Phil in ICT 3 rd semester Nepal Open University , Nepal Abstract Keywords. May 2020, 26.

Raza, H., Faizan, M., Hamza, A., Mushtaq, A., & Akhtar, N. (2019). Scientific Text Sentiment Analysis using Machine Learning Techniques. In IJACSA) International Journal of Advanced Computer Science and Applications (Vol. 10, Issue 12). www.ijacsa.thesai.org

Shelar, A., & Huang, C. Y. (2018). Sentiment analysis of twitter data. Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018, 1301–1302. https://doi.org/10.1109/CSCI46756.2018.00252

Singh, M., Jakhar, A. K., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. Social Network Analysis and Mining, 11(1), 1–11. https://doi.org/10.1007/s13278-021-00737-z