# Comparative study of Twitter Sentiment On COVID - 19 Tweets

Anu J Nair
Department of Computer Science and Applications
Amrita Vishwa Vidyapeetham
Amritapuri, India
anu.j.nair022@gmail.com

Aadithya Vinayak
Department of Computer Science and Applications
Amrita Vishwa Vidyapeetham
Amritapuri, India
aadithyavinayak75@gmail.com

Veena G
Department of Computer Science and Applications
Amrita Vishwa Vidyapeetham
Amritapuri, India
veenag@am.amrita.edu

*Abstract:* **Recently, the number of tweets on COVID-19 are increasing at an unprecedented rate by including positive, negative and neutral tweets. This diversified nature of tweets has attracted the researchers to perform sentiment analysis and analyze the varied emotions of a large public towards COVID-19. The traditional sentiment analysis techniques will only find out the polarity and classify it as either positive, negative or neutral tweets. As an advanced step, the proposed research work attempts to find the sentiment of tweets using Logistic Regression sentiment analysis, VADER sentiment analysis and BERT sentiment analysis. The proposed analysis methods are more sensitive to sentiment expressions in social media contexts, while it can be generalized on the basis of the domain. Even though 3 different algorithms are implemented, all the preprocessing and further steps excluding the sentiment analysis algorithm will remain identical. The identical processing steps will help to compare the proposed three different sentiment analysis algorithms. Furthermore, there are many useful applications with this proposed analysis, as this work obtains a public opinion for the government officials or even for the health officials and help them to work on the basis of the obtained results.**

Keywords—*Natural Language Toolkit (NLTK), Twitter sentiment analysis , Valence Aware Dictionary and sentiment Reasoner (VADER), Aspect-based Sentiment Analysis, Bidirectional Encoder Representation from Transformers (BERT) Sentiment Analysis, Natural Language Processing (NLP)*

## I. INTRODUCTION

Sentiment analysis is nothing but making use of natural language processing (NLP) and text analysis tools to identify and extract the feelings of a writer from his/her piece of text present in the form of tweets. These feelings or sentiments will be either positive, negative or neutral. The basic idea of sentiment is the feeling of the author. These may vary from one person to other.

Social media technologies are the most vast and effective ways to express ones emotion towards one topic. There exist several different forms of social media technologies, such as blogs, business networks, photo sharing, forums, microblogs, enterprise social networks, video sharing networks, and social networks. As the number of social media technologies has increased, various online social networking services, such as Facebook, YouTube, and Twitter, have become popular because they allow people to express and share their thoughts and opinions about life events.

Different types of social media sites exist, wherein some of them are LinkedIn, Twitter, Facebook, Google, Instagram, WeChat, Whatsapp etc. As the number of social media users increases, the platform or the opportunity for the writer to express their feelings also increases but the best place for them is Twitter as they not only allow the users to express but also to share their thoughts and opinion. So if an analysis is carried out on tweets, the emotions or feelings of a large group of people towards this pandemic situation can be obtained.

These sentiments of writers can be found out from the words used and for that, sentiment analysis tools and lexicons, dictionary of words and the polarity within it are used. Polarity is nothing but whether the text is written with positive sentiment in the writer's mind or with a negative sentiment or neutral sentiment.

Sentiment analysis on twitter data can be done by extracting the tweets from the twitter. Then perform some sentiment analysis algorithms on these tweets. Then the basic idea of the sentiment of a tweet will be obtained further the best feature for determining the sentiment of that tweet will be analyzed. These features will be later used for finding the

polarity of the sentence. This final step is done because there may be some statements with multiple aspects.

These multiple aspects may be from positive and negative part i.e. Even though COVID is a pandemic situation to do work from home. So here it is observed that the term COVID is considered negative and work from home is having a good feeling. So this feature selection will help us in more efficient ways.
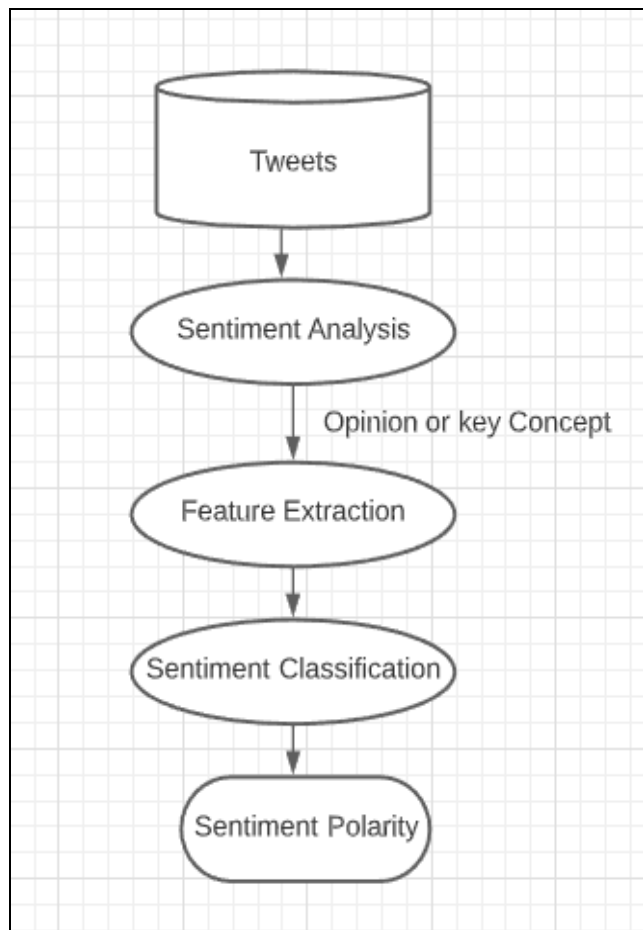


Fig 1: Sentiment Analysis Process

The pinnacle of our paper is to study the sentiment of users using sentiment analysis algorithm. For that, three different algorithms are utilized. They are sentiment analysis using logistic regression, VADER sentiment and BERT sentiment analysis. Then, the three algorithms are compared based on its accuracy for prediction.

## II. RELATED WORKS

Shihab Elbagir and Jing Yang .Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment [5].

Sentiment Analysis of Twitter data has been a topic of research recently as Twitter is considered one of the best social media sites through which the emotions of a large group of people towards a topic has been obtained. In this paper, they have used valence aware dictionary for sentiment reasoning and VADER Sentiment Analysis to predict the sentiment of a group of people towards the 2016 US election. In their work, it is further classified into multiclass. They have used a small amount of they got good accuracy in multiple classes. The NLTK and VADER were applied on the tweets on the 2016 election to data, lexicon with categorising specific data and the data was not trained too.

Jihang Mao, Wanli Liu. A BERT-based Approach for Automatic Humor Detection and Scoring [6]. In this paper, they have taken tweets from the corpus and predict if the tweet is a joke or not and also to give its score. For that purpose, they have made use of BERT sentiment analysis and pre-trained the Humor Analysis based on Human Annotation task, HAHA task. The tweet is passed to the output layer of classification. The score is found out using the mean square error of labels and predicted scores. This method is good for multilingual text classification tasks.

C.J. Hutto Eric Gilbert.VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text [7]. In their paper, they have made use of tweet, from Twitter directly and applied VADER on it. They also made a comparative study of Valence Aware Dictionary for sEntiment Reasoning, VADER with sentiment analysis algorithms like LIWC, ANEW, SentiWordNet and Support Vector Machine (SVM) algorithm. Using both quantitative and qualitative methods they had first validated the feature selected and combined it with five rules of grammar generally to get the sentiment intensity. VADER performs better than the rest of the algorithm used. But this VADER is not making use of any aspects in the tweets or sentences.

Heather Newman and David Joyner. Sentiment Analysis of Student Evaluationsof Teaching [8]. In this paper, they have made use of VADER to analyse student evaluation by different teachers of a single source. For these evaluation purposes that have made use of official evaluations, comments in the forum and through other ways too. They have compared frequently used comments and will find out the polarity of the comment as some questions may be asked common for all the teachers in all the universities. VADER provides a quick look feature on positive and negative comments. Correlation between the total sentiment score of the comment and the total sentiment of the class will be taken as a measurement.

Al Amin, Imran Hossain, Aysha Akther and Kazi Masudul AlamIn: Bengali VADER: A Sentiment Analysis Approach Using Modified VADER [9].In this paper, they have done research on Bengali NLP. For that, they have made a change in VADER so that it supports Bengali sentiment identification of polarity. They have improvised the functionality of English VADER so that they can classify the Bengali language

without any translator in it. From there experiment, it improved the sentiment analysis of Bengali text over the current model. They also improvised the Bengali lexicon polarity with the help of English lexicon polarity. They made use of stemming, listing out the boosting words of Bengali, apply bigram and trigram to give a better result. They have not used in sentences with both positive and negative meaning in it.

## III. PROPOSED WORK

This paper attempts to describe the proposed work in detail.

As discussed earlier, tweets related to COVID hashtags will be extracted and then sentiment analysis algorithms like Logistic Regression, BERT and VADER will be applied.

### A. *Sentiment Analysis Using Logistic Regression*

There are different types of classification models. They are discriminative and generative model. Logistic regression is a model of discriminative classification.

So in this model, first all the stopwords should be removed from the dataset. Then it is tokenized and vectorized for further process.
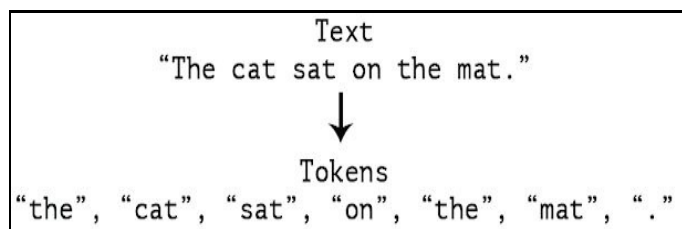


Fig 2: Tokenization in Sentences

For the vectorization purpose, IDs are assigned to each token. Then, the vectors are passed to the CountVectorization method to find the total number of times the word occurs in the dataset i.e. TF-IDF. Then it will be passed to the logistic regression model for sentiment analysis.

### B. *BERT*

BERT stands for Bidirectional Encoder Representation from Transformers. This algorithm is trained to consider the aspect or concept from both the direction that is from left to right or from right to left simultaneously. As a result, we will be able to get the result more accurately. BERT extract more features compared to other sentiment analysis algorithms.

They also make use of MLM methods in some cases. The purpose of MLM is to mask a random word in a sentence with a small probability. When the model masks a word it replaces the word with a token [MASK] [13]. BERT is generally using Recurrent Neural Network named Long Short -Term Memory (LSTM).to train right to left and left to right simultaneously and concatenate them later. BERT uses Transformers instead of LSTM to get the context of the words because the transformer is more attention based algorithm.

At first, we have to tokenize our data and then vectorize it to get the corresponding ids of the word used. While tokenizing we have to add [CLS] token to the beginning of the sentence and [SEP] token to the end of each sentence.
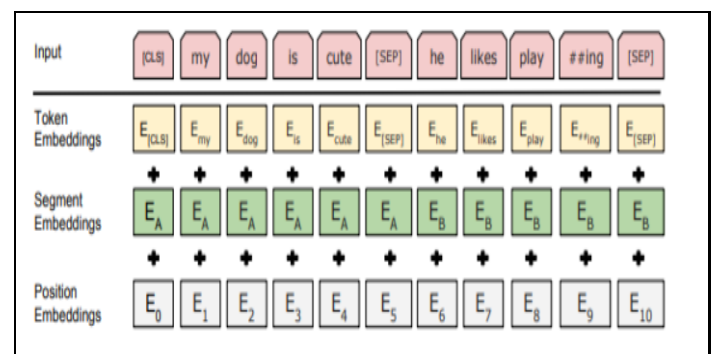


Fig 3: Tokenization And Vectorization in BERT

Then we have to give a small probability to each token by adding a small MASK value to it and finally perform an activation function on it to see if the score or the polarity value is exceeding the threshold set. If so then it is positive else it may be negative or neutral.

### C. *VADER*

VADER is Valence Aware Dictionary and sEntiment Reasoner. This VADER makes use of lexicons for getting valence i.e. strength of each word and polarity of each word. This VADER not only gives us whether a statement is positive, negative or neutral but also how much we will get it is positive, negative or neutral.So we will take the polarity of each word separately and apply average in it by normalizing it.

Normalized score = (score) / (sqrt ((score *score) + alpha))

The data is first taken for preprocessing and in this stage, we are removing all the stop words, punctuations and performing tokenization in it. This tokenization is done by taking each sentence and then splitting it into words and assigning the token id to all later on these ids will be used by the sentiment analysis algorithm to understand and increase the efficiency of the algorithm. These tokens are then taken for boosting by using bi-gram, tri-gram methods. These boosting is done in order to reduce the bias and variance to minimal length so that there will be less error while predicting. And finally, the valence score is calculated for each token separately.

As we get the set of tweets with the sentiment analysis applied to it we could go for the feature selection process in our work we will be making use of the key concept or the opinions of the users from there tweets to extract the features. After feature extraction, we will be making use of this data to classify the tweets into positive, negative or neutral and thus the sentiment polarity is derived. This takes for one single tweet. We will be collectively doing the same for all tweets in the dataset and then we will be finding out the accuracy and all for the above three algorithms.

## IV. EXPERIMENTS AND RESULTS

We have taken dataset related to tweets with different hashtags related to COVID -19. This may include #COVID-19, #WUHAN, #Corona Virus. We have taken a dataset with more than 25000 data in it.

Initially we have started our work by finding out the sentiment of a large group of people towards this pandemic situation.

We applied different algorithm in our dataset to see the accuracy of each algorithm with respect to same dataset. We have divided our data set into 75 and 25 per cent for training and testing purposes respectively. Which is 18,750 for training and 6,250 for testing. At first we applied sentiment analysis with logistic regression to get the polarity score of the dataset.

We are taking 75 and 25 percent because it is good to give more dataset for training as it will give more precise result while doing testing with the testing dataset.

TABLE 1: TO GET HOW MUCH FEATURE IS THERE IN EACH POLARITY

| Dataset | positive tweet | negative tweet | neutral tweet |
|---------|---------------|----------------|---------------|
| Covid-19 | 15340 | 9620 | 40 |

The above table has the number of tweets that is actually positive, negative and neutral .This is done manually as we have to get a reference to check is the predicted value is similar to that of the actual value.
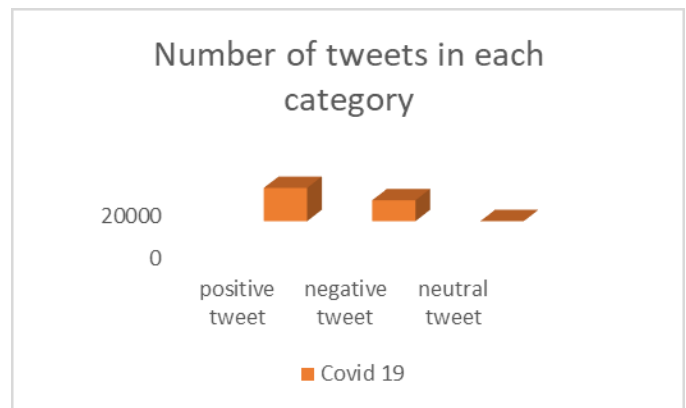


Fig 4: Number of tweets in each polarity

We can see that from 25,000 data 15,340 were having a positive, 9,620 having negative and 40 having neutral feelings or sentiment towards COVID time.

As we get the sentiment classification of each tweets from three different algorithms we could make a collective analysis to get the accuracy,precision,recall and f1-score on three different algorithms.And the datas are as follows:

TABLE 2: ACCURACY, PRECISION, RECALL AND F1-SCORE FOR 3 DIFFERENT ALGORITHMS

| Algorithm Used | Logistic Regression | BERT | VADER |
|----------------|--------------------|----|-------|
| Accuracy | 0.83 | 0.92 | 0.88 |
| Precision | 0.84 | 0.9 | 0.94 |
| Recall | 0.83 | 0.94 | 0.93 |
| F1 SCORE | 0.83 | 0.92 | 0.88 |

In order to get a clear idea on the matric score of the three different algorithm we will go for the following bar graph representation.
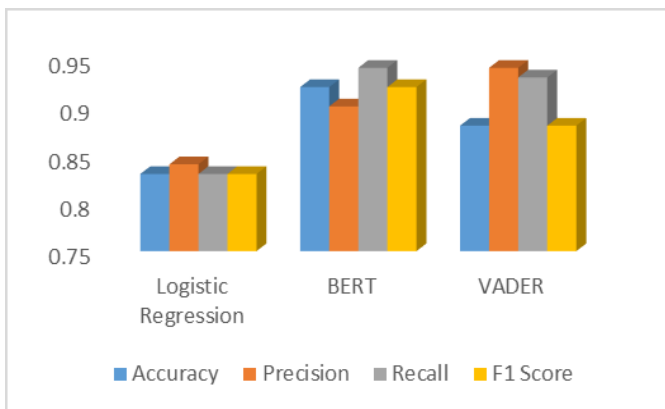
Fig 5:Accuracy,Precision,Recall and f1-score of three different algorithm.

From the figure 5 we could clearly see that applying Logistic regression on COVID tweets is having very low accuracy as compared to the other two algorithms so now let's concentrate on the remaining two algorithm mainly VADER and BERT.

Here, we can see that the accuracy of BERT is good accuracy (0.92) compared to the other two algorithms as they are doing the checking of context forward and backward. BERT algorithm on a large dataset consumes more time to give the output so it is better to use along with GPU.

In the case of precision VADER is having more value. As VADER is not taking care of all the aspects of a sentence we cannot consider it as the best algorithm for sentiment analysis. VADER is considered to be a rule-based algorithm and BERT to be an Aspect- based rule.

## Conclusion

In this paper we have applied three different algorithms say Logistic Regression, BERT, VADER sentiment analysis algorithm. We have normalized the score of the three algorithm with in a range of -1 and 1. This is done so that the comparison will be fair and easy to identify. On comparing we can see that BERT is more accurate (92%) than VADER and Logistic Regression. BERT is more accurate than other algorithms because they look for the aspect of the sentences. VADER will look for the valence and polarity score which will reduce the performance of the feature selection. Logistic regression is not looking for the polarity strength or aspect of the tweets. Moreover we have applied

this algorithms on large data and hence it take lot of time to process and get the output.

The limitation of our work is it is domain based and we have not looked towards the mood of the user. So we have kept the concept of mood based sentiment analysis for future work.

## *References*

[1] Rohith Gandhi," K Nearest Neighbours—Introduction to Machine Learning Algorithms",

[2] Kavitha K R, Ajith Gopinath, Midhun Gopi. "Applying Improved SVM Classifier for Leukemia Cancer Classification using FCBF." 978-1-5090-6367-3/17/$31.00 ©2017 IEEE

[3] Kavitha, K. R., G. SyamiliRajendran, and J. Varsha. "A correlation based SVM-recursive multiple feature elimination classifier for breast cancer disease using microarray." 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2016.

[4] Thushara, M. G., M. S. Krishnapriya, and Sangeetha S. Nair. "A model for auto-tagging of research papers based on keyphrase extraction methods." Advances in Computing, Communications, and Informatics (ICACCI), 2017 International Conference on. IEEE, 2017.

[5] Shihab Elbagir and Jing Yang, "Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment," Proceedings of the International MultiConference of Engineers and Computer Scientists, IMECS 2019, Hong Kong, March 13-15, 2019.

[6] Jihang Mao, Wanli Liu, "A BERT-based Approach for Automatic Humor Detection and Scoring," Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019).

[7] C.J. Hutto Eric Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.

[8] Heather Newman and David Joyner , " Sentiment Analysis of Student Evaluationsof Teaching,"©Springer International Publishing AG, part of Springer Nature 2018

[9] Al Amin, Imran Hossain, Aysha Akther and Kazi Masudul AlamIn, "Bengali VADER: A Sentiment Analysis Approach Using Modified VADER," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.

[10] K. A. Hasan, A. Mondal, and A. Saha, "A context free grammar and its predictive parser for bangla grammar recognition," in Computer and Information Technology (ICCIT), 2010 13th International Conference on. IEEE, 2010, pp. 87–91.

[11] I. San Vicente and X. Saralegi, "Polarity lexicon building: to what extent is the manual effort worth?" in LREC, 2016.

[12] P. J. Stone, D. C. Dunphy, and M. S. Smith, "The general inquirer: A computer approach to content analysis." 1966.

[13] Jacobo Rouces, Mickel Hoang and Oskar Alija Bihorac, "Aspect-Based Sentiment Analysis Using BERT," 8th international workshop on semantic evaluation (SemEval 2014)

[14] M. A. Smith et al., "Analyzing (social media) networks withvNodeXL," in Proceedings of the fourth international conference on Communities and technologies, 2009, pp. 255–264.

[15] D. L. Hansen, B. Shneiderman, and M. A. Smith, "Analyzing social media networks with NodeXL: Insights from a connected world" Morgan Kaufmann, 2010.

[16] S. Bird, E. Klein, and E. Loper, "Natural language processing with Python: analyzing text with the natural language toolkit." O'Reilly Media, Inc.," 2009.

[17] Sungheetha. Akey. and Rajesh Sharma. "Transcapsule model for sentiment classification." Journal of Artificial Intelligence 2, no. 03 (2020): 163-169.

[18] Mitra, Ayushi."Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 03 (2020): 145-152.

[19] Anu J Nair, Rizwana Rasheed, Maheeshma KM ,Aiswarya LS. "An Ensemble-Based Feature Selection and Classification of Gene Expression using Support Vector Machine, K-Nearest Neighbor, Decision Tree." 2019 International Conference on Communication and Electronics Systems (ICCES).

[20] B. Wagh, J. V Shinde, and P. A. Kale, "A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques," Int. J. Emerg. Res. Manag. Technol., vol. 6, no. 12, pp. 37–44, 2018.

[21] S. B. Mane, Y. Sawant, S. Kazi, and V. Shinde, "Real Time Sentiment Analysis of Twitter Data Using Hadoop," Int. J. Comput.Sci. Inf. Technol., vol. 5, no. 3, pp. 3098–3100, 2014.

[22] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," IEEE Access, vol. 3536, no. c,pp. 1–21, 2017.

[23] M. A. Smith et al., "Analyzing (social media) networks with NodeXL," in Proceedings of the fourth international conference on Communities and technologies, 2009, pp. 255–264.

[24] D. L. Hansen, B. Shneiderman, and M. A. Smith, Analyzing social media networks with NodeXL: Insights from a connected world.Morgan Kaufmann, 2010.

[25] S. Bird, E. Klein, and E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit. "O'Reilly Media, Inc.," 2009.

[26] Natural Language Toolkit http://www.nltk.org/ (Date Last Accessed, November 20, 2018).