Twitter Sentiment Analysis On Coronavirus Outbreak Using Machine Learning Algorithms

[1]Dr K B Priya Iyer, [2]Dr Sakthi Kumaresh,
[1]Associate Professor, Department of Computer Science
[2]Associate Professor, Department of Computer Science
M.O.P. Vaishnav College for Women (Autonomous), Chennai, India

ABSTRACT

Social media is a source that produces massive amount of data on an unprecedented scale. It serves as a platform for every person to share their perspectives, opinions and experiences apart from just being a platform that gives information to the public who search for information on the disease. As unexpected as the occurrence of coronavirus disease 2019 (COVID-19) was, it has been radically affecting people all over the world, there is a need to analyse the opinion of people on the pandemic COVID-19. This paper focuses on the sentiment analysis of COVID-19 using twitter data. The analyses are based on the machine learning algorithms. This article provides an analysis on how people react to a pandemic outbreak, how much they are aware of the disease and its symptoms, what precautionary measures they are taking and whether or not people are following government's guidelines etc. Understanding the posts on social media pages during a pandemic outbreak allows health agencies and volunteers to better assess and understand the public's insolences, sentiments and needs in order to deliver appropriate and effective information.

KEYWORDS: Twitter, Corona Virus, Machine Language

1. INTRODUCTION

In the end of 2019, the COVID-19, on-going coronavirus disease originated in Wuhan, China. The novel virus is believed to have created from an animal-to-human spill over event linked to seafood and live-animal markets like butcher shops. The virus has spread and communicated locally in Wuhan and other places in China, despite strict intervention measures and efforts implemented in the region. It is affecting 203 countries and territories around the world as on 2 April 2020. Coronavirus affected 936,725 people, claimed more than 47260 lives as of 2 April 2020. According to WHO, the fatality rate is around 2% as discussed in the press conference that was held on the 29th of January,2020. The World Health Organization declared the coronavirus pandemic outbreak as a Global Public Health Emergency.The on-going outbreak of coronavirus disease, has taken 58 lives, along with 2032 confirmed cases in India, as of 2nd April 2020. Due to COVID-19, several people had to lose their lives and it is surprising to know that the number of deaths associated with COVID-19 surpassed the other coronaviruses SARS-CoV, and MERS-CoV, which stood a highest threat to the world's public health.

All countries are taking various steps to control the pandemic such as Janata curfew, nation lockdown, cancelling transport facilities, impose social distancing restrictions etc. Twitter is one of the fastest information sharing platform among all online social networking media. Messages or tweets on twitter range from personal information to global news or events.

Analysing this continuously generated data is very interesting and informative enabling users or organisations to acquire knowledge. This helps the government or organisations to know how far the public is aware of the disease outbreak, its symptoms and precautionary measures.

Sentiment analysis is well studied using Twitter data in recent days to predict and/or monitor health related issues. Twitter contains huge number of meaningless messages and unwanted or polluted content, which negatively affects the perception analysis performance. The traditional techniques are not well suited because of the short length of tweets, spelling and grammatical errors, and the frequent use of informal languages. In this effort, information about illnesses and diseases is extracted from Twitter with spatio-temporal restraints during a given disease outbreak period. Sentiment analysis is used to understand the perception of the people about coronavirus disease and also to know what extent people's livelihood is affected. The research results will facilitate faster response to and preparation for epidemics and also be very useful for both public and governments to make more informed decisions.

Table 1:
Coronavirus affected countries details as on 2 April 2020

Location	**Confirmed**	**Cases per 1M people**	**Recovered**	**Deaths**
United States	124,697	381.24	3,231	2,227
Italy	92,472	1,463.97	12,384	10,023
China	81,439	59.3	75,448	3,300
Spain	73,235	1,477.99	12,285	5,982
France	37,611	556.65	5,700	2,314
Iran	35,408	437.18	11,679	2,517
United Kingdom	17,136	264.06	140	1,028
Indonesia	1,155	4.43	59	102
Philippines	1,075	10.31	35	68
Greece	1,061	98.44	52	32
India	979	0.77	87	25

In this research paper, the main aim is to obtain a better understanding of the social opinions and perspectives on COVID-19 and how it has changed people's thinking over the past few months. Social mediasuch as Twitter is mainly beneficial to extract information related to the user's sentiments, opinions and insights on a numerous number of topics. Twitter is considered to be a mini blogging social media platform and it has a huge and growing number of users every day. Twitter has reached 330 million active users every month and 145 million users every day. Most of the twitter users are between the age group of 35 and 65. In twitter, user post short messages or blogs of 140 or fewer characters to "tweet" about their opinions on coronavirus, to share information, and to have talks with their 'followers'. Hence tweets collected from twitter data for sentiment analysis of people on coronavirus using deep learning algorithms will help to study user's sentiments into three categories as positive, neutral and negativeduring the disease outbreak.

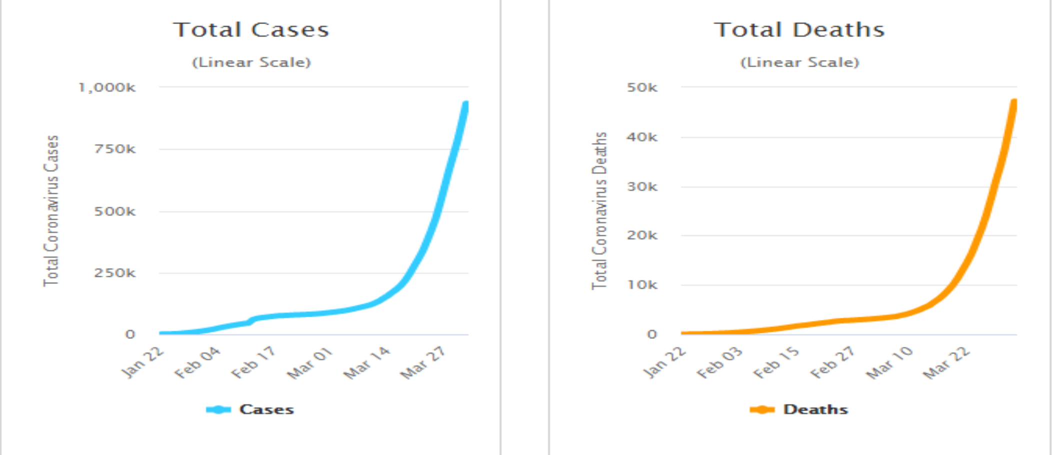


Figure 1: Total Coronavirus Cases VsDeaths data as on 2 Apr 2020

2. LITERATURE REVIEW

Two most common feature in NLP TF-IDF and N-Grams [1] was used on the SS-Tweet dataset. Six different algorithms are used for classification and they found that TF-IDF feature is giving better results (3-4%) as related to N-Gram features. Categorization of people views into positive, negative and neutral was done on 393,869 static twitter data that was taken from world website. [2] authors came out the with the result that maximum and supreme peoples' view is neutral. A hybrid approach [3] is developed for sentiment analysis that makes use of machine learning algorithms like Naïve Bayes and support vector machines (SVM). Comparison study on political views was made using sentiment analysis. [4] The paper discovers thenumerous sentiments applied to Twitter data and their outcomes. Varied techniques for Twitter sentiment analysis methods have been discussed, which includes machine learning, ensemble approaches and lexicon(dictionary) based approaches. Twitter sentiment analysis based on ensemble method and hybrid sentiment analysis techniques were looked into. [5] compare the sentiment with balloting data to see how ample correlation is shared.

Lexicon and Naive Bayes Machine Learning Algorithm was used to analyse the sentiment and label or tag tweets based on hashtag content, [7] The research has been carried by classifying an opinion/view/belief in the form of commentaries into two classes, which is positive and negative with the level of accurateness that is prejudiced by the training procedure. Public sentiment data to the tourist lures comprised in the positive sentiment. [8] This work accustomed to the sentence-level approaches to run on Android OS and measures their performance in terms of memory usage, CPU usage, and battery consumption. Their findings reveal sentence-level approaches that need almost no editions and run comparatively fast.

[10] examine the result of sentiment analysis features in locating ADR mentions. Methods. Results show that sentiment analysis features slightly advance ADR credentials in tweets and well-being related forum posts. This study shows that adding sentiment analysis features can slightly advance the performance of even a state-of-the-art ADR identification method. [11] paper suggests a technique of sentiment dictionary implanting which signifies sentiment

word's semantic relations. The outline was that combined encoding morphemes and their POS tags, and working out only significant lexical morphemes in the embedding space. As a result, the revised embedding approach enhanced the performance of sentiment classification.

[12] proved that AD-related tweets used to perpetuate public stigma, which impacted negative expectations of individuals with the disease. [13] analyses text sentiment in social media using lexical-opinion method. [14] analyses Twitter datasets in NLTK Corpora using a feature extraction technique. Various machine learning classifiers such as MultinomialNB, BernoulliNB, LogisticRegression are discussed. Experimental results demonstrate that BernoulliNB, LogisticRegression, and SGD classifier reached accuracy as high as 75%.

[15] designed a rule classifier with a voting-based ensemble of supervised classifiers. A set of rules based on the occurrences of emoticons and sentiment-bearing words are framed. Experimental results demonstrate the effectiveness of the method. [16] focussed on Natural Language Toolkit techniques for processing data from Twitter.Feature selection is extracted by Chi Square test and Naïve Bayes classifier is used for training.[17] focuses on sentiment analysis at entity level for Twitter. A lexicon based approach is adopted to perform entity-level sentiment analysis. Recall is improved by inclusion of likely tweets. The classifier helps to train and assign polarities to the entities in the new tweets.

3. MACHINE LEARNING APPROACH

Machine learning methods are trained on datasets and a model is created for evaluation. Based on the accuracy of the model, the machine learning method is acceptable. The three methods in machine learning algorithms are supervised learning, unsupervised learning and reinforcement learning. In supervised learning, the model is trained using labelled data which contains both input and results. The phases of processing are training phase and testing phase. Unsupervised learning methods do not use training data or labelled data. It finds the hidden structures or patterns from unlabelled data.

Supervised Learning

Supervised learning requires a well-labelled dataset to train. Supervised learning is of two types namely regression and classification. Classification techniques help to find the appropriate class labels which can predict the positive, negative and neutral sentiments. A machine learning model is developed which uses the labelled data to train, classify the tweets and predict the sentiments of the tweets. Decision Tree, Random Forest, Bayesian belief network, Naive Bayes and KNN classifiers are some of the algorithms that are used in this method.

Unsupervised Learning

Unsupervised methods are based on machine learning or lexicon. The requirement of the labelled datasets is not required in unsupervised learning. Sentiment analysis when done using unsupervised learning; it is generally based on a Sentiment Lexicon. Text classification helps to extract phrases which contain adjectives or adverbs to estimate a phrase’s semantic orientation. Semantic orientation is then used to classify the sentiments.

Naïve Bayes Classifier Definition

Sentiment analysis refers to field of study of extracting subjective emotions, thoughts, feelings and views from text.Usually the text reflection of positivity or negativity is figured out using sentiment analysis. Naive Bayes algorithm is the probability of antext belonging to a particular group based on the presence or absence of a particular charecter.The Naive Bayes algorithm assumes that occurrence of each point if independent of other. Bayes' Theorem is stated as the following equation:

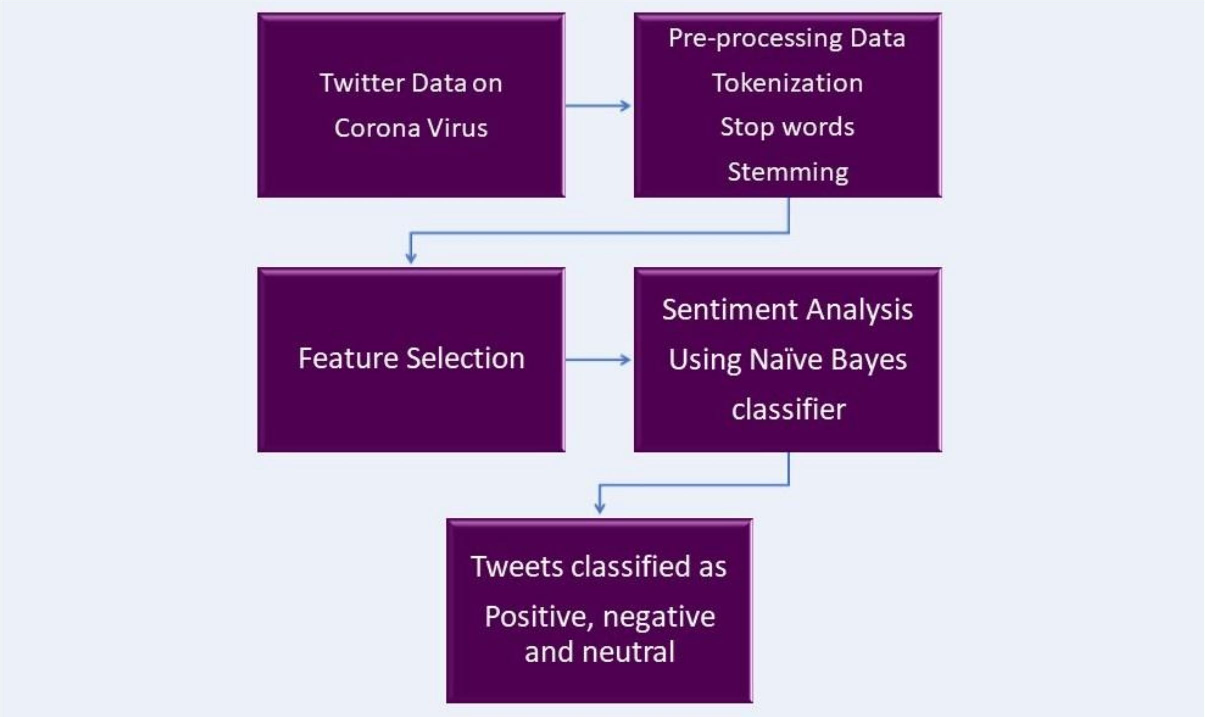
$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

$P(C_k|x)$: Probability (conditional probability) of occurrence of event C_k given the event x is true. $P(C_k)$ and P(x): Probabilities of the occurrence of event C_k and x respectively. $P(x|C_k)$: Probability of the occurrence of event x given the event C_k is true.

Proposed Model

The dataset is collected from twitter data using online API tweets. Data that contain IDs and sentiment scores of the tweets related to the COVID-19 pandemic and is processed through a set of five phases. The phases are Tweet collection& pre-processing, Tweets cleaning, Feature selection, Modelling and Evaluation.

Table 2 :
Corona Outbreak - Sentiment Analysis Model


Twitter Data on Corona Virus
Pre-processing Data
Tokenization
Stop words
Stemming
Feature Selection
Sentiment Analysis Using Naïve Bayes classifier
Tweets classified as Positive, negative and neutral

Phase I:Tweets Collection and Pre-processing

Process of transforming unstructured data into structured data is pre-processing stage. The data is downloaded from twitter directly through twitter API. Hashtags, whitespaces, hyperlinks, urls, usernames, stop words etc are removed from tweets.

Table 3:
Pre-processing stage

Remove hashtags	Remove whitespaces	Remove hyperlinks	
Remove URL address	Remove HTML special entities	Remove usernames	
Removal of stop words	Remove tickers	Remove Unicode strings from the tweets	

Phase II: Tweets cleaning

Tweets in its original form cannot be processed for analysis. Tokenization refers to splitting strings into smaller words called tokens. Tokenization consists of identifying nouns, verbs, adverbs and adjectives etc. Grouping of words with same meaning is one of the process under NLP. For example, ran, runs and running will be treated as one word instead of different words. Stemming and lemmatizationaretwo popular techniques of normalization in NLP. Stemming will remove unwanted suffixes from beginning or end of the word.Lemmatization analyses the word with respect to vocabulary. The cleaned tweets are now subjected to text processing where tokenization and stemming are done to the tweets.

Phase III: Building tweet dictionary & determining word density

After the text processing, the word dictionary is built. All words are categorised into positive, negative and neutral across all dataset. Later the density of each word is calculated as the count of occurrences of everyunique word across all the training dataset. The sentiments are usually not affected by stop words. Hence stop words are filtered and words like not and do not say about the polarity of the tweet. This dictionary helps in evaluating the testing set.

Table 4 :
Tweets collection from Twitter

corona	coronavirus	Covid	covid19
quarantine	Pandemic	sars cov2	social distancing
work from home	chinese virus	Vaccine	wuhan virus
Stayhomestaysafe	wash ur hands	hand sanitizer	Lockdown
wear a mask	corona vaccines	face shield	health worker

PhaseIV: Feature Selection

Feature selection is based on word density. Word density is determined by Term Frequency-Inverse Document Frequency (TF-IDF)method. TF-IDF identifies the frequently occurring words in the given tweet and the words that are not appearing frequently in the remaining training data set. Word density helps to know about the polarity of the tweet.

Polarity of the tweet is calculated through a term weight using TF-IDF. The positivity and negativity of the term is calculated based on number of times the term occurs in given tweet dataset.The TF-IDF is executed on all terms in the dataset to find the rank of each word. A high rank in TF-IDF shows the word is relevant in given tweet and can contribute much to the polarity of the tweet.

The overall approach works as follows:

Given a Data Set D, a term t, and an individual tweet (dt), dt$\in D$, we calculate:

$$\textit{Adsjusted TF-IDF} = f_{t,\, dt} * \log (|D| / f_{t+s,\, D}) \text{ -------- (1)}$$

Where $f_{t,\, ds}$ equals the no. of times term tappears in ds_1

$|D|$ is the size of the Data set, and $f_{t+s,\, D}$equals the number of tweets in which the term t and its corresponding synonym word appears in D.

For each term in the term data, the corresponding synonyms are fetched from word dictionary. Synonym are considered to be the words equivalent to the original term and hence taken for the calculation of $f_{t,\, D.}$Set of terms that are extracted using Adjusted TF-IDF are used to judge the polarity of the tweet.

Phase V: Model Building& Evaluation

A model is a way of describing rules and equations within a system. Each tweet in the dataset is labelled positive or negative. The dataset has twp ;parts where first part is used to build the model and second part is used to test the accuracy of the model. Naive Bayes classifier is used to build the model. The experimental analysis reflects the sentiments of people towards coronavirus.

When Naïve Bayes model is used, with feature values t, (Adjusted TF-IDF), P [YES] is computed, where Y is the number of positive instances and N is the number of negative instances. The Laplace estimator is used to avoid zero probabilities. This simply replaces Y and N by Y+1 and N+1.

$$\mathrm{P[YES]} = \frac{\mathrm{Y}}{\mathrm{Y+N}}\ \mathrm{P}\ \mathit{Adjusted}\ \mathrm{TF} - \mathrm{IDF[t|Yes]} \text{ ------ (2)}$$

Similar expression for p [No] is calculated. The overall probability that the given tweet is positive or negative can then be calculated as follows:

$$P = \frac{P[YES]}{P[YES] + P[NO]} \quad \text{------------ (3)}$$

Algorithm:Twitter_CoronavirusSentimentAnalyzer()

```
1. Create a twitter account
2. Get consumer key, consumer secret, access key, access secret from twitter login
3. Initialize twitter API
4. Tweets<- twitterAPI() //tweets downloaded from twitter live data on coronavirus
5. Tweet<-tweet_preprocessing(Tweets) // removes hashtags, usernames, urlsetc
6. Cltweets<-tweets_cleaning(Tweet) // performs tokenization, stemming etc
7. Pos_tweets<-sentiment.positive(Cltweets)
8. Neu_tweets<-sentiment.neutral(Cltweets)
9. Neg_tweets<-sentiment.negative(Cltweets)
10. Coronadataset=Merge(Pos_tweets, Neg_tweets,Neu_tweets)
11. NaiveBayesclassifier<-train(coronadataset)
12. Performance<-test(coronadataset)
```

In the above algorithm, the live tweets are downloaded from twitter using functions twitterAPI(). The hashtags, usernames, urls are removed using the functions tweet_preprocessing(). The output of the function is sent to tweets_cleaning() for tokenising and stemming. The tweets are categorised as positive, negative and neutral tweets using functions sentiment_positive(), sentiment_negative and sentiment_neutral(). The labelled tweets as merged into Corona dataset using Merge() function. The Naïve Bayes classification is applied on training dataset and performance is evaluated using test dataset.

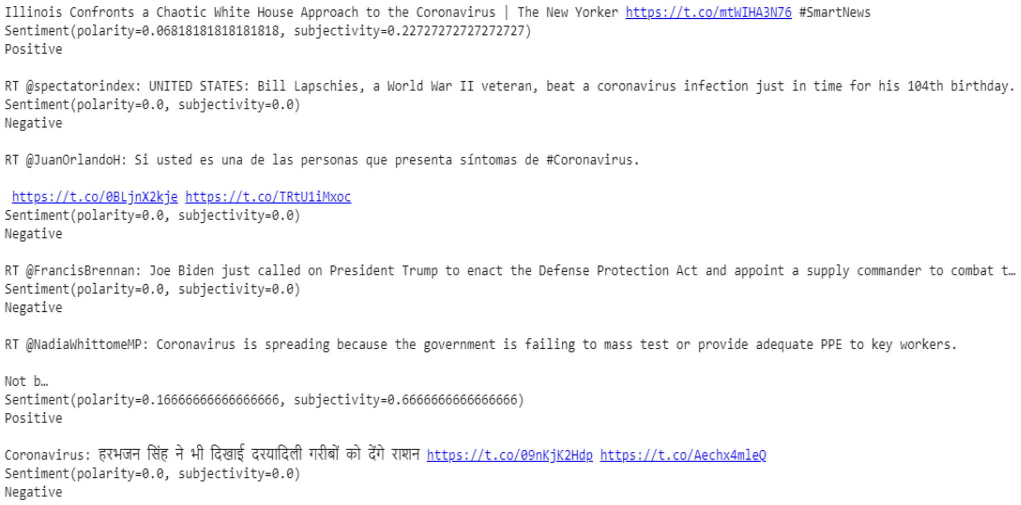
4. EXPERIEMENTAL ANALYSIS

The tweets are collected for 5 days from 1 Apr 2020 to 5 Apr 2020 on Corona Virus. Python is used for implementing the Naïve Bayes classification algorithm. The twitter data is collected by creating an application in twitter that interacts with the Twitter API and tweets are downloaded using words that matches the search key words. Oauth protocol is used for authentication while particular tweets are searched on Corona virus.From the twitter API, confidential keys such as consumer key, consumer secret, an access token and an access token secret is generated. These keys help to provide user authentication towards accessing twitter API. The default permissions from twitter account are read-only. Using these keys, the respective key words such as Corona Virus from the twitter is filtered in the particular location and languages.

The results of the corona virus outbreak using sentiment analysis and machine learning of the collected tweets are shown in Figure 3. Positive, neutral, and negative are the categories of tweet. The positive tweets showpeople support the coronavirus outbreak symptoms and precautionary measures etc. Figure 4 shows the visual representation of words used in tweets. Some of the positive and negative words extracted from the tweets are shown below

Positive: Recovery, Vaccine, protective, hope, trust, safety

Negative: pandemic, virus, sadness, worry, symptoms, crisis



```
Illinois Confronts a Chaotic White House Approach to the Coronavirus | The New Yorker https://t.co/mtWIHA3N76 #SmartNews
Sentiment(polarity=0.06818181818181818, subjectivity=0.22727272727272727)
Positive

RT @spectatorindex: UNITED STATES: Bill Lapschies, a World War II veteran, beat a coronavirus infection just in time for his 104th birthday.
Sentiment(polarity=0.0, subjectivity=0.0)
Negative

RT @JuanOrlandoH: Si usted es una de las personas que presenta síntomas de #Coronavirus.

 https://t.co/0BLjnX2kje https://t.co/TRtU1iMxoc
Sentiment(polarity=0.0, subjectivity=0.0)
Negative

RT @FrancisBrennan: Joe Biden just called on President Trump to enact the Defense Protection Act and appoint a supply commander to combat t…
Sentiment(polarity=0.0, subjectivity=0.0)
Negative

RT @NadiaWhittomeMP: Coronavirus is spreading because the government is failing to mass test or provide adequate PPE to key workers.

Not b…
Sentiment(polarity=0.16666666666666666, subjectivity=0.6666666666666666)
Positive

Coronavirus: हरभजन सिंह ने भी दिखाई दरयादिली गरीबों को देंगे राशन https://t.co/09nKjK2Hdp https://t.co/Aechx4mleQ
Sentiment(polarity=0.0, subjectivity=0.0)
Negative
```

Figure 3: Tweets polarity, subjectivity and sentiment

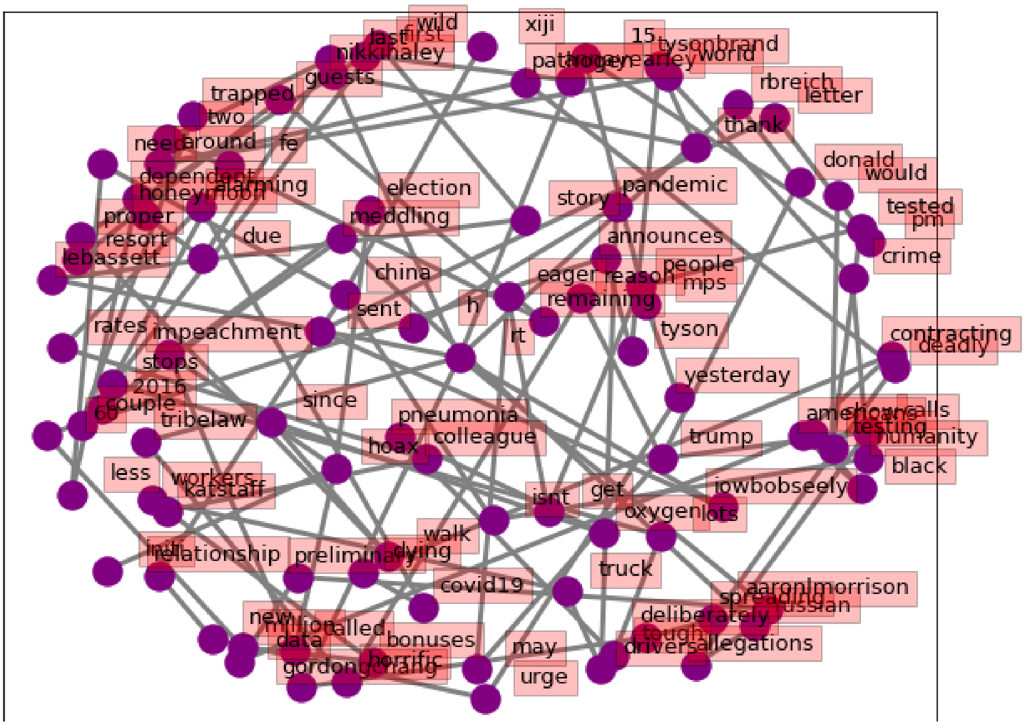


Figure 4: Visualization of Coronavirus outbreak

The polarity of tweets is shown in Figure 5 and Figure 6. From Figure 5, it is shown that most of the tweets are at 0.0 polarities. . The most common words and their frequency count are calculated online directly from tweets. These form the wordlist for processing the test data set. These are depicted in Figure 7 and Figure8.

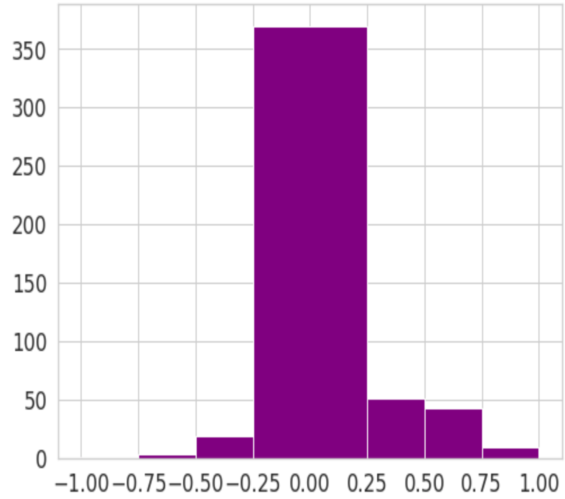
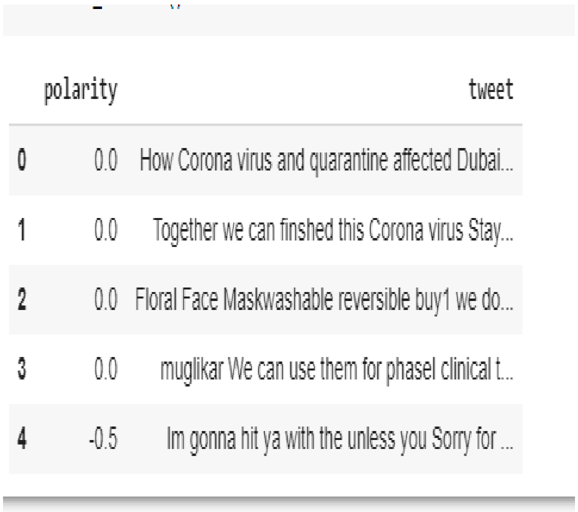
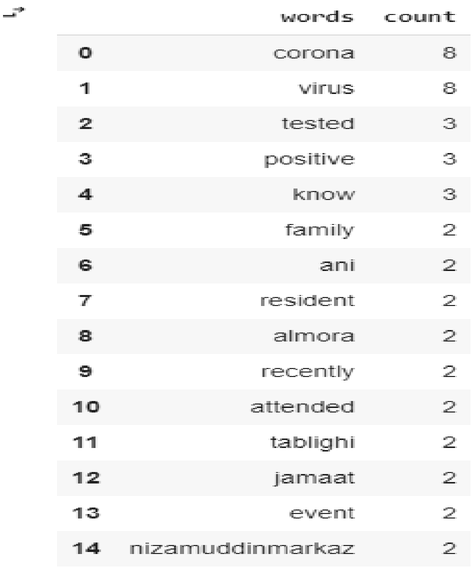


Figure 5: Sentiments from tweets on Corona Virus



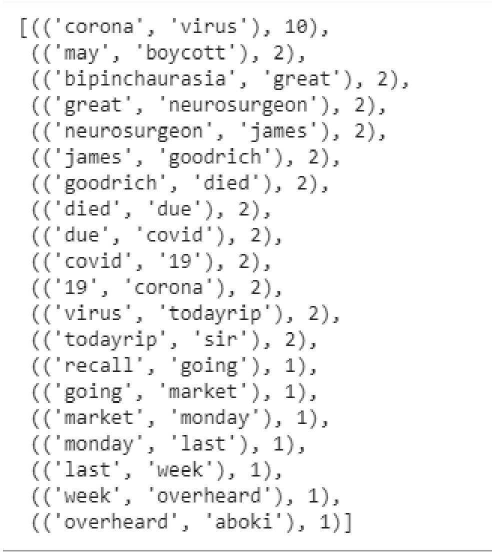
	polarity	tweet
0	0.0	How Corona virus and quarantine affected Dubai...
1	0.0	Together we can finshed this Corona virus Stay...
2	0.0	Floral Face Maskwashable reversible buy1 we do...
3	0.0	muglikar We can use them for phaseI clinical t...
4	-0.5	Im gonna hit ya with the unless you Sorry for ...

Figure 6: Polarity of each tweet after cleaning



	words	count
0	corona	8
1	virus	8
2	tested	3
3	positive	3
4	know	3
5	family	2
6	ani	2
7	resident	2
8	almora	2
9	recently	2
10	attended	2
11	tablighi	2
12	jamaat	2
13	event	2
14	nizamuddinmarkaz	2

Figure 7: words frequency



```
[(('corona', 'virus'), 10),
 (('may', 'boycott'), 2),
 (('bipinchaurasia', 'great'), 2),
 (('great', 'neurosurgeon'), 2),
 (('neurosurgeon', 'james'), 2),
 (('james', 'goodrich'), 2),
 (('goodrich', 'died'), 2),
 (('died', 'due'), 2),
 (('due', 'covid'), 2),
 (('covid', '19'), 2),
 (('19', 'corona'), 2),
 (('virus', 'todayrip'), 2),
 (('todayrip', 'sir'), 2),
 (('recall', 'going'), 1),
 (('going', 'market'), 1),
 (('market', 'monday'), 1),
 (('monday', 'last'), 1),
 (('last', 'week'), 1),
 (('week', 'overheard'), 1),
 (('overheard', 'aboki'), 1)]
```

Figure 8 : bigram and count of tweets

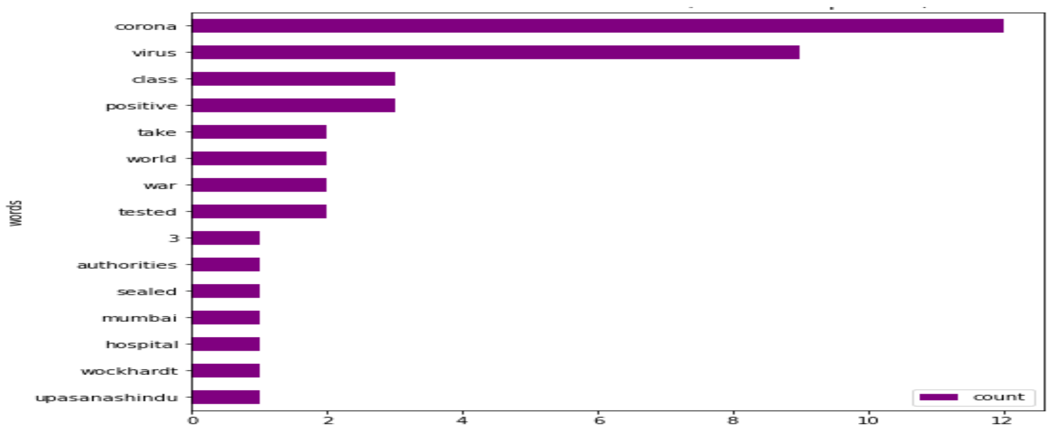
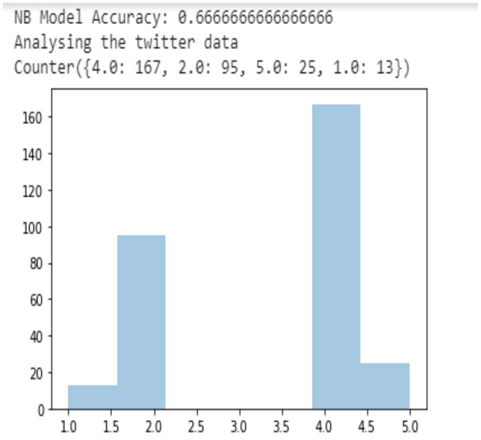


Figure 9: Common words found in tweets without stop words



```
NB Model Accuracy: 0.6666666666666666
Analysing the twitter data
Counter({4.0: 167, 2.0: 95, 5.0: 25, 1.0: 13})
```

Figure 10: NB Model Accuracy

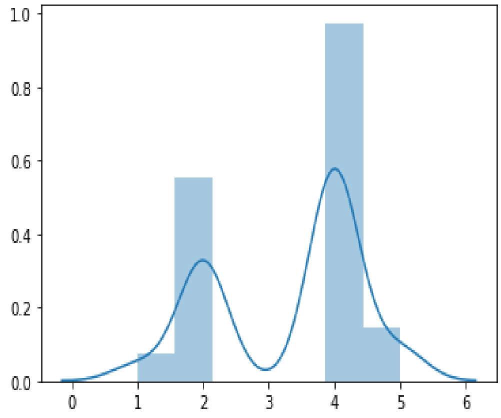
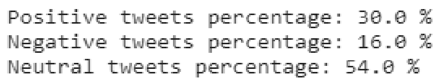


Figure 11: Tweet Analysis



```
Positive tweets percentage: 30.0 %
Negative tweets percentage: 16.0 %
Neutral tweets percentage: 54.0 %
```

Table 2: Percentage of Tweets sentiments

Figure 10 and Figure 11 shows the Naïve Bayes classification performance. The accuracy of the model is nearly 70%. The percentage of positive tweets are 30%, negative tweets are 16%. Most of the tweets nearly 56% are neutral on corona virus outbreak as shown in Table 2.

5. CONCLUSION

This work was carried on corona virus outbreak using twitter data from 1st April 2020 to 5th April 2020 where the virus spread across several countries and the outbreak became pandemic. This work helps to understand the people's perception about coronavirus and its impact on the public. The sentiments during the period were downloaded and the public's reaction towards the outbreak was analysed. Machine learning algorithm is applied for data analysis and accuracy of the model is nearly 70%. The people well understood the government policies, safety measures, its symptoms and precautionary measures to be taken during this period. They well followed and maintained the social distancing and sanitizing methods. This study helpsthe organisations to understand the opinion of people during theCorona Virus outbreak. As the virus is spreading vigorously, the study needs to be carried out on a weekly basis to have a better understanding on the sentiments of the people.

6. REFERENCES

[1] Ahuja, Ravinder& Chug, Aakarsha&Kohli, Shruti& Gupta, Shaurya&Ahuja, Pratyush. (2019). The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science. 152. 341-348. 10.1016/j.procs.2019.05.008.

[2] Ajay Bandi and Aziz Fellah. Socio-Analyzer: A Sentiment Analysis Using Social Media Data. Volume 64, 2019, Pages 61–67 Proceedings of 28th International Conference on Software Engineering and Data Engineering.

[3] Ali Hasan, Sana Moin, Ahmad Karim and ShahaboddinShamshirband "Machine learning-based sentiment analysis for twitter accounts", MDPI, 2018.

[4] Alsaeedi, Abdullah & Khan, Mohammad. (2019). A Study on Sentiment Analysis Techniques of Twitter Data. International Journal of Advanced Computer Science and Applications. 10. 361-374. 10.14569/IJACSA.2019.0100248.

[5] Brandon Joyce, Jing Deng. "Sentiment Analysis of Tweets for the 2016 US Presidential Election", in IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA: IEEE (2017)

[6] D T Hermanto et al Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms Naive Bayes Classifier. 2018 J. Phys.: Conf. Ser. 1140 012037

[7] JohnnatanMessias, Joao P. Diniz, Elias Soares, Miller Ferreira,MatheusAraujo, Lucas Bastos, Manoel Miranda, FabricioBenevenuto, Towards Sentiment Analysis for Mobile Devices . 2016

[8] KavyaSuppala and Narasingarao: "Sentiment analysis using Naïve Bayes classifiers", international journal of innovative technology and exploring engineering June 2019.

[9] I.Korkontzelos,A.Nikfarjam,M.Shardlow,A.Sarker,S. Ananiadou, and G. H. Gonzalez, "Analysis of the effect of sentiment analysis on extracting adverse drug

reactions from tweets and forum posts," Journal of Biomedical Informatics,vol. 62, pp. 148–158, 2016.

[10] Minchae Song, Hyunjung Park, Kyung-shik Shin"Attention-Based Long Short-Term Memory Network Using Sentiment Lexicon Embedding for Aspect-Level Sentiment Analysis in Korean." Information Processing & Management, 56 (3) (2019), pp. 637-653

[11] Nels Oscar, Pamela A. Fox, Racheal Croucher, Riana Wernick, Jessica Keune, and Karen Hooker . Machine Learning, Sentiment Analysis, and Tweets: An Examination of Alzheimer's Disease Stigma on Twitter . J Gerontol B PsycholSciSocSci, 2017, Vol. 72, No. 5, 742–751

[12] Rahman, S. A. El, F. A. AlOtaibi, and W. A. AlShehri. (2019, 3-4 April 2019). "Sentiment Analysis of Twitter Data", in t*he 2019 International Conference on Computer and Information Sciences (ICCIS)*.

[13] ShihabElbagir, Jing Yang. Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn.

[14] U. A. Siddiqua, T. Ahsan, and A. N. Chy, "Combining a rule-based classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog," in 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, pp. 304– 309.

[15] M. Vadivukarassi, N. Puviarasan and P. Aruna. Sentimental Analysis of Tweets Using Naive Bayes Algorithm. World Applied Sciences Journal 35 (1): 54-59, 2017

[16] Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2015). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE), 89, 1–8.