Sentiment Analysis On Twitter Data Of Vaccines For COVID-19

Amrita Mishra[#1], Mohd. Saif Wajid[*2], Upasana Dugal[#3]

[#1]*Department of Computer Science & Engineering, School of Engineering, Babu Banarasi Das University, Uttar Pradesh, India*

[*2]*Department of Computer Science & Engineering, School of Engineering, Babu Banarasi Das University, Uttar Pradesh, India*

[#3]*Department of Computer Science & Engineering, School of Engineering, Babu Banarasi Das University, Uttar Pradesh, India*

***Abstract*— The unprecedented outbreak of the 2019 novel corona virus, termed as COVID-19 by the World Health Organization (WHO) on 11 March 2020, has placed numerous governments around the world in a precarious position. The government with the help of municipal authorities took several measures like distribution of PPE kit, sanitizers, medical masks, etc to suppress the harmful effects. Research and drug institutes like Russia based Gamaleya Research Institute (GMI), India based Indian Council of Medical Research (ICMR) and National Institute of Virology (NIV) have developed vaccines to combat COVID-19.This research paper is a thorough effort to perform Sentiment Analysis of the vaccines developed for emergency and preventive use for coronavirus disease. In this paper, we propose a framework for retrieving social media Twitter data for Sputnik V, Moderna and Covaxin vaccines. Furthermore the positive, negative and neutral sentiments that are obtained from Lexicon based approaches are validated using Naïve Bayes Algorithm with admissible accuracy upto 79%, 70%, 78% for Sputnik V, Moderna and Covaxin respectively which is overlooked in other research studies. The Visual analysis using pie charts have been done to analyse about the sentiments of different vaccines.**

***Keywords* — Sentiment Analysis, Visual Analysis, Twitter, Lexicon, Word Tokenization.**

I. INTRODUCTION

Twitter is one of the world's most popular social media platforms with over 335 million users. Twitter represents one of the largest and most dynamic datasets of user generated content approximately 200 million users post 400 million tweets per day [14]. Twitter provides a microblogging service in which users post status messages, called "Tweets" mentioned under user's handle, with no more than 140 characters. Twitter platform is beneficial in reaching a wide audience and connecting with customers such as in Digital Marketing of products. Twitter information is also helpful in disseminating health information. Recognizing the fact the government agencies like World Health Organization WHO and Center for Disease Control and Prevention (CDC) have adopted use of twitter in resolving public and health issues. Relating to this fact, in the first 12 weeks of the Zika virus outbreak in late 2015, the WHO Twitter account was retweeted over 20,000 times. The tweets posted by users may convey a lot more than mere set of words [1]. They also serve as data grounds for opinion mining. Based on locations of Twitter User, information about masses or groups undergoing same kind of problematic situation or joyful experience are resolved. The data analysis for similar patterns may facilitate identical research problems in future.

COVID-19 is an infectious disease caused by recently found virus known as SARS-CoV-2(Severe Acute Respiratory Syndrome). Its outbreak is beyond the previous observations of this virus and is thus considered pandemic by World Health Organization .Some of the major vaccines developed for preventive and emergency use in COVID-19 are Sputnik V, Pfizer BioNTech, Moderna and Covaxin.

Sputnik V is developed by Russia on 12 August, 2020 by Gamaleya Research Institute (GMI) in collaboration with Russian Defense Ministry is the first registered vaccine. U.S also developed Pfizer BioNTech and Moderna on 11 December, 2020 and 17 December, 2020 respectively. On December 11, the Food and Drug Administration (FDA) issued an Emergency Use Authorization (EUA) for emergency use of Pfizer-BioNTech for prevention of coronavirus disease 2019 (COVID-19) for individuals. Moderna vaccine is used for active immunization to prevent COVID-19 caused by severe acute respiratory syndrome coronavirus 2(SARS-CoV-2) in individuals 18 years of age or above. Covaxin, India's indigenous COVID-19 vaccine by Bharat Biotech is developed in collaboration with the Indian Council of Medical Research (ICMR) and National Institute of Virology (NIV).Covaxin has been granted approval for emergency restricted use in India on January 3, 2021.

The main objective of this research study is to analyze the positive, negative and neutral sentiments of the vaccines among Twitter users in Pandemic and to validate the results using Machine Learning Algorithm. Furthermore, study of Text Classification techniques and fetching bi grams and trigrams of vaccines is also performed. Consequently, we propose a

framework to collect and process tweets relating to vaccines and apply Machine Learning Based Algorithm to calculate the correctness of opinions flooding on Twitter. The results of this process are visualized to show the trends and dynamics of polarity of individual vaccines.

The rest of the paper is arranged as follows: Section II is literature review of the studies conducted by Researchers on Twitter Data Sentiment. Section III describes the lexicon based and machine learning methodologies on Twitter data. It also provides a detailed explanation of the proposed framework and the implementation process. Section IV contains the architecture. Section V consists of the results obtained from the implementation stage Section VI consists of conclusion and future scope.

II. LITERATURE REVIEW

Sentiment Analysis has been of avid interest to researchers lately. A lot of work has been put into it and there is a vast domain of its applications. Gaurav Bhatt et al.,2016,[21] have performed Sentiment Analysis over Educational institutions Using Twitter Dataset of IIT,NIT and AIIMS Colleges in India with SVM, Naïve Bayes and ANN algorithms and accuracy of 89.6%.The area of Neural Networks has been investigated for performing sentiment analysis on benchmark dataset consisting of online product reviews. Bespalov, Bi,Qi and Shokoufandeh , 2015, [20] carried out binary classification on Amazon and TripAdvisor dataset using Perceptron classifier and obtained one of the lowest error rates among their experiments of 7.59 and 7.37 on the two datasets respectively. Researchers have also been working upon prediction of accuracy of tested datasets using Machine Learning Algorithms .Kanakraj and Guddeti , 2015, [3] used Natural Language Processing techniques for Sentiment Analysis and compared Machine Learning Methods and Ensemble Methods to improve on the accuracy of classification. Shahheidari, Dong and Bin Daud, 2013, [4] used a Naïve Bayes Classifier for classification and tested it for news, finance, job, movies and sportstaking into consideration Data Mining on basis of two emoticons (☺ and ☹). Prediction Of Election Results is another domain in which massive population expresses opinion over Social Networks. Rincy Jose and Varghese S Chooralil, 2015 [7] have used Twitter Data with Classifier Ensemble Approaches with accuracy of 71.48% in predicting election results. Rincy Jose, et. al, 2015 [9] have also predicted election results with Word Sense Disambiguition with accuracy of 78.6% .

Mohd. Saif Wajid and et al. 2017, [8] have used Sentiment Analysis Based on A.I. Over Big Data. They have introduced the methodology for creating user recommended data group (Big data) by elaborating a matrix for user recommended data group for big data which is then reduced by dimension reduction technique. Neethu M.S and Rajasree R, 2013 [5] used twitter post on electronic products, compared the accuracy between different Machine Learning Algorithmn and further improved accuracy by replacing repeated character with two occurrences, including a slang dictionary and taking emoticons into consideration. Jotheeswaran and Koteeswaran, 2015 [6] performed binary classification on the IMDB dataset by employing a Multi-Layer Perceptron Neural Network and using Decision Tree -Based Feature Ranking for feature extraction and a hybrid algorithmn (based on Differential Evolution and Genetic Algorithm)for weight training, thereby obtaining a maximum classification accuracy of 83.25%.

Laszlo and Attila, 2020 [31] have used fresh scraped data collections over the Recurrent Neural Networks to determine what emotional manifestations occurred in given time interval in COVID-19 .The Sentiment Analysis helps in monitoring area based upon the opinion raised in different territories.

III. METHODOLOGY

The proposed framework technique follows four main processes; Extraction of COVID-19 vaccines datasets using Twitter developer account, preprocessing, computing sentiment analysis score for tweets Using Lexicon Method which is worked upon Natural Language Processing for Tweets of Sputnik V, Moderna and Covaxin and finally validating the outcome of results using Naïve Bayes Algorithm of Machine Learning under performance evaluation. The detailed explanation including research designing, research procedure and is as follows.

A. COVID-19 VACCINES DATASET

Tweets have been extracted after creating a developer account on Twitter. Using the Tweepy library in Python Anaconda environment and developer account details a set of 2000 tweets are extracted. Unique tweets referring to query Sputnik V, Moderna and Covaxin in month of March 2021 are gathered in json format. The json format is converted into data frame and csv format. Examples of the tweet dataset obtained can be seen in Table.1.

Table 1:Examples of the tweet dataset obtained

S.No.	Tweets
1	RT @ amitsurg : “Clinical trial mode” for Covaxin.
2	RT @htpune: Covishield or Covaxin? Choice is yours.
3	RT @sidhant: All in All, India will send Iran 500,000 doses of Indian Vaccine COVAXIN.
4	RT @cnnphilippines: Indian Ambassador to PH: Covaxin was deployed immediately when India started its vaccination drive in January.
5	RT @karmashxo: Iran received 150,000 doses of indigenous Bharat Biotech’s #Covaxin today.
6	RT @snehamordani: Centre writes to states on DCGI clearance of COVAXIN for Emergency Use Authorisation .

B. PRE-PROCESSING

Twitter dataset provide information about tweet id, text, retweet, timestamp, date, username. In order to segregate the user opinion from user information, preprocessing was performed on the tweets. Removal of stopwords, punctuation, duplicate tweets, extra spaces and usernames is performed using concept of Regular Expression. As a result 902 processed tweets are gathered for Sputnik V, 998 for Moderna and 1214 for Covaxin. These words plays an important role in English language but are not essential in sentiment analysis. Examples of the preprocessed tweet dataset obtained can be seen in Table.2.

Table 2: Raw and Pre-Processes Tweets

Raw tweets	Preprocessing
RT @ amitsurg : “Clinical trial mode” for Covaxin.	clinical trial mode for covaxin
RT @htpune: Covishield or Covaxin? Choice is yours.	covishield or covaxin choice is yours
RT @sidhant: All in All, India will send Iran 500,000 doses of Indian Vaccine COVAXIN.	all in all india will send iran 500000 doses of Indian vaccine covaxin
RT @cnnphilippines: Indian Ambassador to PH: Covaxin was deployed immediately when India started its vaccination drive in January.	indian ambassador to ph covaxin was deployed immediately when india started its vaccination drive in January.
RT @karmashxo: Iran received 150,000 doses of indigenous Bharat Biotech’s #Covaxin today.	Iran received 150000 doses of indigenous bharat biotechs covaxin today.

C. PROPOSED SOLUTION

LEXICON BASED SENTIMENT ANALYSIS USING TEXTBLOB

Following the preprocessing and data cleaning phase the Final Tweets are collected for Sputnik V, Moderna and Covaxin. The Dictionary Based approach is implemented using the TextBlob for Sentiment Analysis. A magnitude of 0 score was considered as neutral, greater than 0 as positive and lesser than 0 as negative. Based on this the resultant data is classified as positive, negative and neutral and sentiment score is given as shown in Table 3. The text and sentiment score columns of cleaned dataset procured after classification is split into 80 percent training data and used in Machine Learning Multinomial Naïve Bayes Classifiers while rest of data is used as testing data model.

Table 3: Tweet and sentiment score

Tweets	Sentiment score
clinical trial mode for covaxin	1
all in all india will send iran 500000 doses of Indian vaccine covaxin	1

Covishield or Covaxin? Choice is yours.	0
indian ambassador to ph Covaxin was deployed immediately when india started its vaccination drive in January.	1
covaxin out of clinical trial mode, granted restricted emergency use authorisation	-1
centre writes to states on dcgi clearance of covaxin for emergency use authorisation .	-1

USING NATURAL LANGUAGE PROCESSING AND COUNT VECTOR

The Twitter dataset obtained after preprocessing is also utilized in implementing Natural Language Text Processing methods that includes Tokenization, Stemming, Count Vectorization, bi-grams and tri-grams for sentiment analysis.

Tokenization: Tokenization is a way of separating a piece of text into smaller units called tokens. Tokenization can be broadly classified into 3 types – word, character, and subword (n-gram characters) tokenization. Example of Tokenization in the dataset is, "new Covaxin developed by Hyderabad based .." as [new, co, developed, Hyderabad, based..].

Stemming: Applying Porter Stemmer in python provides the root of words. Elimination of words coming from same roots is performed as these words are considered having same root and meaning. For Example: connect, connection, connected, connections, connect comes from "connect". Example of Stemming in the dataset is, "new Covaxin developed by Hyderabad based ." as [new, co, develop, Hyderabad, based..].

Count Vectorization: The count vectorization is performed as it provides the capability for generating the vector representation of text also making it a highly flexible feature representation module for text. After count vectorizer, it is possible to analyze the words with two or three or more in the text. Table 4 shows words and vector numbers.

Table 4: Words and Count

Dataset	Word	Count
Covaxin	Covaxin	1433
Covaxin	vaccine	634
Moderna	de	403
Sputnik V	vaccin	110
Sputnik V	sputnik	601

Converting to N-grams : Building n gram model helps to predict most probably words that might follow this sequence. Some N2 grams i. e word classification in 2 words are :
Bharat Biotech, 324
Efficacy 81, 193
Clinical trials, 154
Some N3 grams i.e word classification in 3 words are:
('milh es doses', 148),
('urgente governo bolsonaro', 143)

Implementing Multinomial Naïve Bayes Algorithm for Sentiment Analysis

Naive Bayes is based on Bayes' theorem, where the adjective Naïve says that features in the dataset are mutually independent. Naive Bayes is a probabilistic classifier, meaning that for a document d, out of all classes c ∈ C the classifier returns the class ˆc which has the maximum posterior ˆ probability given the document. In Eq.1 we use the hat notation ˆ to mean "our estimate of t he correct class".

$$\hat{c} = \arg\max_{c \in C} P(c|d) \quad ..(1)$$

D. PERFORMANCE EVALUATION

Precision: Precision quantifies the number of positive class predictions that actually belong to the positive class.
Precision = true positive / (true positive +false positive)

Recall: Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.
Recall = true positive / (true positive + false negative)

F1-score: F-Measure provides a single score that balances both the concerns of precision and recall in one number.
F1-score = (2* Precision *Recall) / (Precision + Recall)

Accuracy: Classification accuracy is the total number of correct predictions divided by the total number of predictions made for a dataset.

Accuracy = (True positive + True Negative) / (True Positive + True Negative + False Positive + False Negative)

IV. ARCHITECTURE

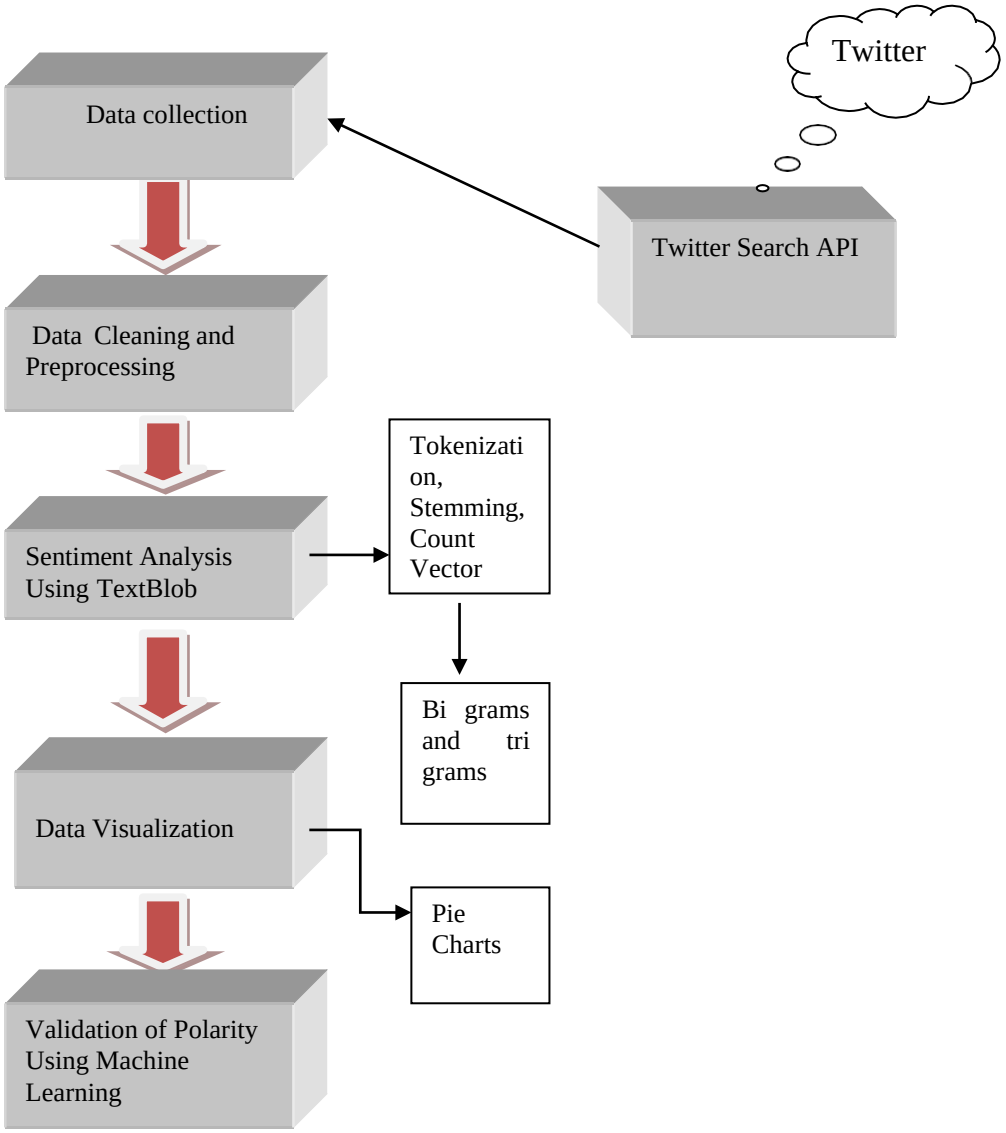

Figure 1: The Architecture of the Proposed System

V. RESULTS AND DISCUSSION

This section gives an insight into the results obtained from the above experiment. Figure 2(a) shows the pie chart percentage distribution of tweets for Sputnik V. The percentage of Positive tweets is 22.3%, Negative tweets is 9.7% and 68.0% for neutral tweets. Figure 2(b) shows the pie chart percentage distribution of Moderna. The Positive tweets are 10.6%, Negative tweets are 12.1% and neutral tweets are 77.3%.Figure2(c) shows the pie chart percentage distribution of tweets for Covaxin. The percentage of Positive tweets is 35.0%, Negative tweets is 17.7% and 47.2% for neutral tweets.

In case of positive tweets, the highest have been found for Covaxin vaccine .The Moderna vaccine has more neutral percentage distribution 77.3% than Sputnik V that is 68.0%. But as positive tweets are more for Sputnik V than Moderna, it is evident that more people are favoring Sputnik V than Moderna over social media.

From Table 5, a total of 446 tweets are regarded as positive, 194 as negative and 1360 as neutral for Sputnik V. 212 tweets are regarded as positive, 241 as negative and 1547 as neutral for Moderna. 701 tweets are regarded as positive, 354 as negative and 945as neutral for Covaxin.

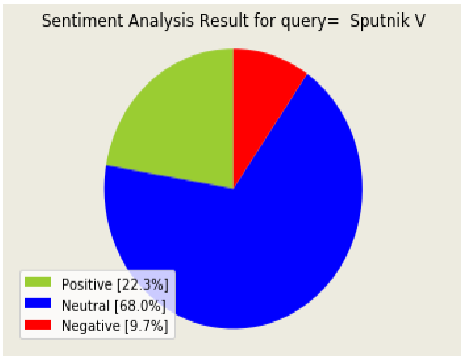

Figure 2(a)

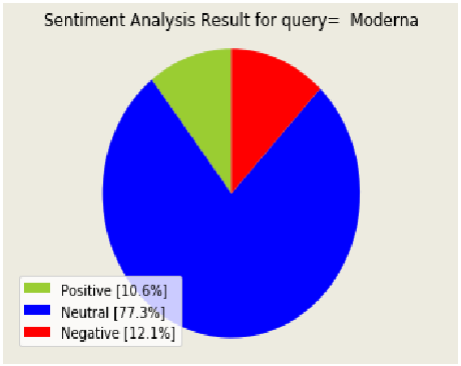

Figure 2(b)

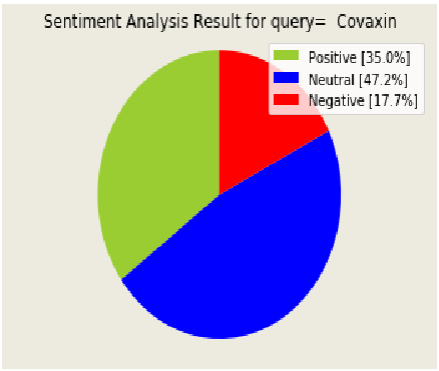

Figure 2(c)

Figure 2(a) Pie chart distribution for Sputnik V using TextBlob, 2(b) Pie Chart distribution for Moderna using TextBlob, 2(c) Pie Chart distribution for Covaxin.

Table 5: Number of distribution of tweets

Vaccine	Positive Tweets	Negative Tweets	Neutral Tweets
Sputnik V	446	194	1360
Moderna	212	241	1547
Covaxin	701	354	945

Table 6: Percentage of distribution of tweets

Vaccine	Positive (%)	Negative (%)	Neutral(%)
SputnikV	22.3	9.7	68.0
Moderna	10.6	12.1	77.3
Covaxin	35.0	17.7	47.2

From Table 6, Covaxin has highest percentage of positive tweet distribution that is 35.0%. Sputnik V positive tweets distribution of 22.3% which is higher than Moderna positive percentage of 27.3%.This translates to the observation that among three vaccines the positive tweets about for Covaxin are more positive in the magnitude of their sentiment and also indicates that it is most positively talked than other vaccines. Among the Russia based Sputnik V and India's manufactured Covaxin more people are favoring Covaxin as its positive percentage distribution is more. The highest negative percentage distribution for Covaxin which is 17.7 % signifies that out of the total 2000 tweets, 354 negative tweets are found negative. Hence Covaxin is also most negatively discussed vaccine among twitter.

Table 7: Classification report of Multinomial Naïve Bayes

Vaccine	Score	Precision	Recall	F1-score	Accuracy
Sputnik V	-1	0.92	0.12	0.21	79%
	0	0.79	1.00	0.88	
	1	1.00	0.07	0.12	
Moderna	-1	0.90	0.25	0.39	70%
	0	0.68	0.99	0.81	
	1	1.00	0.08	0.14	
Covaxin	-1	0.85	0.57	0.68	78%
	0	0.84	0.85	0.84	
	1	0.67	0.78	0.72	

Table 7 represents classification report for the accuracy, precision, recall and F1 score for different vaccine. The results obtained by Lexicon approaches (Table 5 and Table 6) are validated using Multinomial Naïve Bayes Algorithm. It is seen that Sputnik V has maximum accuracy of 79% followed by Covaxin accuracy 78% and Moderna accuracy 70%.The precision, Recall and F1 score are also computed.

VI. CONCLUSION & FUTURE SCOPE

Through this paper a deep insight to the approaches of Sentiment Analysis is dealt with. Using Twitter Data of vaccines it is observed that India based Covaxin is most favored vaccines among twitter users. Among the U.S based Moderna and Russia's Sputnik V, Sputnik V has more positive sentiments on Twitter. The results have been validated using Machine learning Naïve Bayes Algorithm upto an accuracy of 79%, 70% and 78%. The precision, Recall and F1-score are some machine learning metrices computed in our research.

The future scope includes sentiment analysis using recurrent neural networks, Logistic Classifiers and SVM that can help to improve accuracy of prediction. Also, sentiment in different languages could be worked out using machine learning approaches. More comprehensive research requires tweets to be taken on successive dates to identify the fluctuating trends over social media. Tweets may also be collected on basis of location to study about the opinion about vaccines for specific geographical locations**.**

REFERENCES

[1] Simranpreet Kaur, Pallavi Kaul, Pooya M.Z, "Monitoring the Dynamics of Emotions during COVID-19 Using Twitter Data", 10th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare, pp 423-430, 2020 .

[2] King R.A, Racherla P. and Bush V.D, "What We Know And Don't Know About Online Word-of-Mouth. A Review And Synthesis of the Literature", Journal of Interactive Marketing, Vol. 28, issue 3, pp.167-183, 2014.

[3] Kanakaraj M., Guddeti R M.R, "Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques", 9th IEEE International Conference on Semantic Computing, pp. 169-170, Anaheim, California, 2015.

[4] Shahheidari S., Dong H., Bin Daud M.N.R "Twitter sentiment mining: A multidomain analysis",7th IEEE International Conference on Complex, Intelligent and Software Intensive Systems,pp.144-149, Taichung, Taiwan. 2013.

[5] Neethu M. S. and Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", 4th IEEE International Conference on Computing, Communications and Networking Technologies, pp. 1-5, Tiruchengode, India,2013.

[6] Jotheeswaran J. and Koteeswaran S, "Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis", Journal of Engineering and Applied Sciences, Vol. 10, issue 14, pp. 5883-5894,2015.

[7] Rincy Jose and Varghese S Chooralil, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Word Sense Disambiguition", International conference on Control Communication and Computing In India(ICCC),2015.

[8] Mohd. Saif Wajid , Shivam Maurya , Ramesh Vaishya , "Sentence Similarity based Text Summarization using Clusters", International Journal of Scientific & Engineering Research , Vol. 4,Issue 5, ISSN 2229-5518,2013.

[9] Rincy Jose and Varghese S Chooralil, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach.", International Conference on Control Communication and Computing In India(ICCC),2015.

[10] S.Rajalakshmi, S. Asha and N. Pazhaniraja, "A Comprehensive survey on sentiment analysis", Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, India, 2017, pp. 1-5, doi: 10.1109/ICSCN.2017.8085673.

[11] Vijayaraj, R. Saravanan, P.VicterPaul, R.Raju, "Comprehensive Survey on Big Data Analytics Tools", IEEE Third International, Conference on Innovatives in Information, Embedded And Communication Systems(ICIIECS'16), Coimbatore, Volume 2, pp 32-40,2016,ISSN:978-1-4673-8203-3.

[12] Avinash C.Pandey, S.R Seth and M.Varshney, " Sarcasm Detection of Amazon Alexa Sample Set", Springer Nature Pte Ltd. Advances in Signal Processing and Communications,2019.

[13] U.Ravi Babu, " Sentiment Analysis Of Reviews for E-Shopping Websites", IJECS Vol. 6 Issue 1, pp No.19965-19968,2017.

[14] Medha Khurana, Anurag Gulati, Saurabh Singh, " Sentiment Analysis Framework of Twitter Data Using Classification", 5th IEEE International Conferenceon Parallel, Distributed and Grid Computing, India, 2018.

[15] Navonil Majumder, Soujanya Poria, Haiyun Peng, Niyati Chhaya, Erik Cambria, "Sentiment and Sarcasm Classification With Multitask Learning",IEEE Intelligent Systems, pp.1541-1672, 2019.

[16] D. Ciregan, U. Meier and J. Schmidhuber, " Multi-column deep neural networks for image classification", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 16-21 June, pp. 3642-3649, doi: 10.1109/CVPR.2012.6248110, 2012.

[17] Farhan Hassan Khan, Saba Bashir, Usman Qamar, " TOM: Twitter opinion mining framework using hybrid classification scheme Decision Support Systems", 57 245257, 2013 Elsevier B.V.

[18] S. Porwal , G. Ostwal , A. Phadtare, M. Pandey and M. V. Marathe , " Sarcasm Detection Using Recurrent Neural Network" ,Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 746-748, doi: 10.1109/ICCONS.2018.8663147, 2018.

[19] M. Bouazizi and T. Otsuk Ohtsuki , " A Pattern - Based Approach for Sarcasm Detection on Twitter", IEEE Access,Vol.4,pp.54775488,2016,doi:10.1109/ACCESS.2016.2594194 .

[20] Bespalov D., Bai B., Qi Y., and Shokoufandeh A ,"Sentiment classification based on supervised latent n-gram analysis", 20th ACM international conference on Information and knowledge management, pp. 375-382, USA, 2011.

[21] N.Mamgain, E. Mehta, A. Mittal and G. Bhatt, " Sentiment analysis of top colleges in India using Twitter data", International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), 2016, pp. 525-530,2016.

[22] P. Nambisan, Z. Luo, A. Kapoor,T. B. Patrick and R. A. Cisler. "Social Media, Big Data, and Public Health Informatics: Ruminating Behavior of Depression Revealed through Twitter"48th Hawaii International Conference on System Sciences, Kauai, HI, USA, 2015, pp. 2906-2913, doi: 10.1109/HICSS.2015.351 ,2015.

[23] Tran, T., Lee, K, " Understanding citizen reactions and Ebola- related information propagation on social media", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),IEEE. pp. 106–111, 2016.

[24] Song P, Karako T, " COVID-19: Real-time dissemination of scientific information to fight a public health emergency of international concern. 2020 Biosci Trends" ,March 16;14(1):1-2.doi: 10.5582/bst.2020.01056. Epub 20 PMID: 32092748, 2020.

[25] Saroj Kumar, Singh A.K., Priya Singh, Khan A.M., Vibhor Agrawal, Wajid M.S, "Sentiment Analysis Based on A.I. Over Big Data", In: Satapathy S., Bhateja V., Joshi A. (eds) Proceedings of the International Conference on Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing , vol 469. Springer, Singapore . https://doi.org/10.1007/978-981-10-1678-3_6, 2017.

[26] Koyel Chakraborty, Surbhi Bhatia, Siddhartha Bhattacharyya, Jan Platos, Rajib Bag, Aboul Ella Hassanien. " Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media", Applied Soft Computing, Vol.97,PartA106754, ISSN1568- 4946,2020 .

[27] M.M. Truşcă. Efficiency of SVM classifier with Word2Vec and Doc2Vec models, "Proceedings of the International Conference on Applied Statistics,", Vol. 1(1), pp. 496-503, Sciendo,2019.

[28] M. Bilgin, İ.F. Şentürk , "Sentiment analysis on twitter data with semi supervised doc2vec. "International Conference on Computer Science and Engineering (UBMK)" IEEE, pp. 661–666,2017.

[29] Koyel Chakraborty, Siddhartha Bhattacharyya, Rajib Bag, Aboul Alla Hassanien, " Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Techniques", Social Network Analytics,pp127-147, 2020.

[30] Siddique Latif, Muhammad Usman, Sanaullah Manzoor, Waleed Iqbal, Junaid Qadir, Gareth Tyson, et al., "Leveraging data science to combat COVID-19: A comprehensive review", TechRxiv., 2020.

[31] Laszlo and Attila, " Social Media Sentiment Analysis Based on COVID-19 ", Journal of Information and Telecommunications, Informations, DOI: 10.1080/24751839.2020.1790793, 2020.