International Journal of Computer Sciences and Engineering
Vol.**11**, Issue.**10**, pp.**45-50**, October **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Paper

Aanalysis on Impact of Social Media on Human Behaviour Due to Covid Vaccine Tweets

Jatin Panjavani[1]

[1]Department of Computer Science, LJMU, Liverpool, UK

Author's Mail Id: jatin2707@gmail.com

Received: 03/Sept/2023; **Accepted**: 06/Oct/2023; **Published**: 31/Oct/2023. **DOI**: https://doi.org/10.26438/ijcse/v11i10.4550

Abstract: The Covid-19 outbreak has created a challenge for the whole of mankind and impacted everyday life worldwide. The pandemic had a devastating effect on many people. It also created anxiety, fear, and depression among people. To eradicate the disease, many scientists at major pharmaceutical companies and institutes working together to develop vaccine. Social media is one of the best platforms to discuss the latest trending topics and express your view about them. The Covid-19 Vaccine promotion across the world has created lots of discussion in Twitter, social media platforms where user love to express their feeling and opinion. However, due to lack of knowledge and understanding about Covid – 19 vaccines, it has created negative sentiments towards vaccine among few. Also, there has not been much research work done on in-depth analysis of people's opinion or sentiment towards various vaccine and its brands. In this study will be using publicly available Covid-19 vaccine tweets to understand public opinion or feeling about various covid-19 vaccine brands. This research will be using publicly available Covid-19 vaccine tweets to understand public opinion or feeling about various Covid-19 vaccine brands. The study was conducted using publicly available datasets from online resource, Kaggle. The dataset contains 228207 tweets from Kaggle about the opinion for Covid-19 vaccines during 12 December 2020 to 24 November 2021. After the extraction, vaccine sentiments identified across all brands. After identification of sentiments, accuracy be evaluated based on prediction using various metrics. This research will find the difference in public opinion on Covid-19 vaccines. Understanding the sentiments and public opinions towards vaccine will help health agencies to increase positive awareness about covid-19 vaccine across world.

Keywords: Covid-19 Vaccine, Corona Vaccine, Sentiment Analysis, Twitter Sentiments, Machine Learning, Deep Learning

1. Introduction

There are many social networking sites like Facebook, Instagram, Twitter, WhatsApp, and Facebook, have many data with opinions and feelings. These platforms can provide lots of knowledge or information about public thoughts, service, product using various social media platform [1]. To both suppliers and customers, this data is highly valuable. During any online shopping, first thing consumers usually prefer to check public opinions about the service and product. Based on the customer's opinion or sentiment, the company or manufacturer can learn about its service or product benefits as well as drawbacks. Although both individuals and business organizations can benefit from these opinions. For researchers, it is a very interesting area to study and understand the opinions conveyed in this text content. This modern area of study is also known as Opinion Mining or Sentiment Analysis [2].

In the current world of machine learning and AI, machines are given the chance to analyze and resolve the many challenges by understanding the trend various dataset by themselves. Understanding of hidden patterns and trends predicts and avoids future potential problems. A machine-learning model analyze a different kind of data to solve more questions using the trend and patterns unseen within data. There are many companies or organizations with large volumes of data are using machine learning to solve their business problems. In current modern world, many cloud based cost-effective computational processing and data-storage options have allowed the design and implementation of algorithms to examine huge amounts of data. To gain maximum value from big volume of data, businesses need to understand which model can fit the accurate statistical algorithm [3].

1.1 Research Goal and Method

The research is performed to understand the applications and its solutions of sentiment analysis for the analyzing and classification of opinions using multiple Machine Learning (ML) techniques. These approaches need to search various solutions to the problems found. It also requires certain classifications and methods to evaluate the solutions to certain limitations. Finally, the review ends with the stability and significance of the proof. We focused on the following research questions:

- RQ1: What are sentimental analysis applications using Machine learning techniques?
- RQ2: What are sentiment analysis applications in Covid Vaccine tweets used Machine Learning and Deep Learning techniques.
- RQ3: How much activity was carried out recently?
- RQ4: What are the different classifications, methodology and techniques that were used?
- RQ5: How do the different techniques prove themselves as best?

1.2 Need of Sentiment Analysis

The current situation in world consider data is new oil. As lots of new people started using social media platforms – Twitter, Facebook Instagram. The scope as well as opportunities also have been opened for organizations to understand customed opinion or feedback or behaviors. Tweets on social media are a meaningful resource for examining sentiment and opinion for organizations, governments or at individual's level. (Events, people, products, and trends). Social media platforms like Twitter create a huge volume of tweets or text every day. The study performed in 2015 shows around twenty-one million comments or tweets per hour. It is required to automate the sentiment analysis method to promote the tasks of examining the opinion or views of the people opinion without manually reading those million comments or tweets [4].

2. Related Work

Sentiment classification is part of textual analytics area which is evolving rapidly. The model processing of unstructured text, opinions, feelings, expression, and emotion is highlighted in this research. The research of recognizing or understanding comments or opinion analysis of text comments highlighting public's feeling, expression, or emotion utilizing social media platform has increased during last few years. Multiple related study on textual; sentiment classifications are highlighted below.

Chakraborty [5] presented study on failure of the public tweets associated to the accounts linked to WHO and Covid-19 in guiding public during this epidemic crisis. The study considered public tweets associated and collected which are related epidemics. The use case with approximately more than 23,000 comments or tweets were collected during the period 1st January'19 – 23rd March'20. The research highlighted that most of public in their 40-population commented positive tweets related to Covid-19. The remaining stayed occupied with negative tweets. The Word Cloud, and graphs related to the most frequent word from tweets shows not much significant words.

Praveen [6] used ML techniques to understand is there any shift in general people opinion during times of epidemic emergency and to gather opinions of general people emotion for contact tracing. This study highlights the people's very crucial questions regarding numeral sickness supervision.

Chen [7] projected a Twitter sentiment classification technique with a plenty of emoji importance. The study shows two ways inserting underneath negatively or positively Twitter comments, then, the study classifies or highlights the sentiments with the 2-sense emoji importance using a Deep Learning technique, LSTM model. The study presented that the two ways integrating model is effective at detecting emojis' expression-mindful inserting and outpacing cutting edge favored.

Reddy and Reddy classified tweets or comments into negative or positive opinions, however, the study used detached phrases and words to classify comments or tweets instead of traditional methods of preprocessing twitter comments. The study used LSTM, CNN, and ANN. CNN and LSTM are used to picture words from text however ANN was used to spread sentence appearance. The study in addition suggested the best model techniques or algorithm to create sentiment lexicons for sentiment classification using latest methodologies.

Hasan [9]. suggested a fusion method relating to machine learning and sentiment classification. After that, the study also compared sentiment analysis methods to analyze political viewpoints through various supervised ML based methods like SVM and Naïve Bayes.

Xue [10] highlighted how users from Twitter discussed on Covid-19 situation and their reaction related to same. The study used ML methods to assess Tweets or comments about Covid-19. The study identified around 11 relevant or significant subjects which later categorized them in ten different subcategory texts. The Sentiment Classification outputs from study specify that anxiety about the Covid-19 doubtful nature leads all concerns. The study also discussed the research's shortcomings and significance.

Sanders [11] examined more than 1 million comments from Twitter collected from March'20 to July'20 to predict common behavior or viewpoints towards use of mask at the time of Covid-19 pandemics. The study used segmentation, NLP, and Sentiment classification techniques to organize comments from Twitter on mask use under high-level forms, later study used automated text mixture to express stories for each text category. The study examined topic grouping based on mask associated Twitter comments, it gives insights on people attitudes towards Covid-19 and precautionary strategy. The study also examined huge increase in the volume or strength of mask associated comments. the highlighted insights can be referred by the medical or other government community to perform qualitative evaluations of people emotion towards health involvement approaches in real time.

Gupta [12] highlighted analyzing sentiments analysis of Indian towards statewide lockdown enforced by the Indian leadership to reduce the spread of Covid-19 spread. Sentiment Classification of tweets presented by national citizens of India was performed in study using machine learning and NLP models. Tweets data or comments were collected from Twitter using the API with the help of Valence

Aware Dictionary and TextBlob for Sentiment classification using different NLP techniques. The highlighted tweets were categorized using different classifiers techniques. The study used the unigram and Linear and SVC classifier to attain the highest precision of approximate 84 percent. As per outcomes of the study, the most of Indian citizen agree with decision taken by government of India to enforce a lockdown during the Covid-19 pandemics.

Das and Kolya [13] suggested a different approach for achieving sentiment analysis evaluation with accuracy relating on Covid-19 comments from Twitter and potential increase in case prediction using a deep learning method. The study collected a huge tweet from Covid-19 tweets. The study also separated the tweets comments among train and test dataset. After that, the study examined predictive analytics and polarity in parallel. The study also showed apprised prediction in increase for Covid-19 cases in coming time.

Gambino [14] presented various tweets or thoughts from Twitter users to examine Sentiment Classification responsibilities like Emotion Mining and Sentiment Polarity. Most of tweet's data are focused based on researcher' opinion that, the study published by a researcher is examined to determine the expressed opinion on that. The study showed tweet data focused on individual' opinions. The created dataset consists of stories and news from three different papers and the six stated expressions faced by social media platform readers on the stories or editorials. The data was used to examine how different six reactions are expressed by social media platform users after studying the news articles or reports. The study also highlighted a few observations from a ML technique which was used for predicting the distribution of feelings in earlier unspecified news articles.

3. Research Methodology

The study will analyze the feeling and opinions of the tweets received from Twitter on Covid-19 Vaccine to predict the sentiments. The study has used Twitter dataset of 228207 tweets extracted from Kaggle on the Covid-19 vaccines during period of 12th December'20 to 24th November'21. To perform detailed analysis, the study will explore EDA, most frequently used word, word cloud and predictive classification model. Figure 1 [5] highlights the summarized framework to analyze public opinions and predict the sentiments from tweets. There is multiple analysis module in framework. First, data is collected from twitter then performed exploratory data analysis to understand tweets. Data visualization is used to understand relationship between different features.

TextBlob libraries used to identify the sentiments from tweets. Predictive model is applied for classifying the sentiments using traditional ML and DL models. This study will explore traditional Machine Learning models like LR, DT, RF and GB models and Deep Learning models like Long Short-Term Memory (LSTM) model and word embedding models.

4. Proposed Scheme

Methodology for this study consists of four stages.

- Data Collection on Covid-19 Vaccine tweets from Kaggle. The data will be unstructured with lots of noise.
- Data Pre-processing like Removing Special Characters, Hyperlinks. Data will further be analyzed using tokenization, Lemmatization and Normalization [15]
- Calculating Polarity and Subjective Score from cleaned data
- Sentiment classification using Classic Machine Learning Model like Logistic Regression, Decision Tree, Random Forest Classifier and Deep Learning model like BiLSTM and Word Embedding
- Visualizing the output of model to generate insight for Government and other agencies.

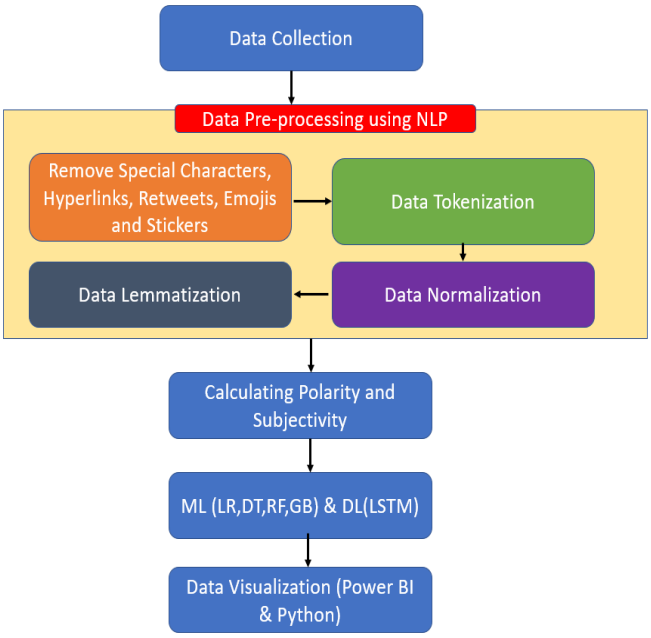

Figure 1 Processing steps for Model

5. Results and Discussion

Overall, the COVID-19 vaccination brands impact on public and at gender level as seen in social media texts is assessed using our suggested data analytics-based approach. This is the first strategy that applies data analytics techniques like descriptive, diagnostic, predictive, and prescriptive analysis to analyze the sentiments of tweets to ascertain public perceptions about various COVID-19 vaccine brands. The suggested strategy works very well in helping specialists, politicians, and members of the government offer useful recommendations for administering the proper immunizations. The following highlights a handful of our method's important observations.

In suggested framework, the study explores the necessary elements which completed the dataset's descriptive analysis. Here, the study has noticed that there are more male candidates than female candidates. the percentage of female

candidates and most people express positive sentiments rather than unfavorable ones in tweets about vaccines, By investigating, the diagnostic analysis yields significant discoveries. the words that appear frequently and frequently in the tweets for both sexes, as depicted in Figure 2. Positive tweets, according to what we've observed, frequently include great or free slots or second dose for vaccinations. Public believe that by getting the right vaccination, the effects of the pandemic can be reduced as soon as possible, and regular life can be resumed. In negative tweets, people typically express their worries about getting immunized, such as stay at home, side effects, etc. According to the diagnostic analysis depicted in Figure 4. The study has also discovered a relationship between the various sentiment polarities regarding vaccine brands from the perspective of gender level, and we have concluded that Pfizer or Moderna received more positive feedback females while Moderna and Biotech received more positive comments from males based on their expressed sentiments on tweets. The predictive analysis aids in identifying the polarities of the sentiments in tweets about vaccines. The predictive analysis is an outcome of the experiment. demonstrate that the Convolution and Random Forest outperforms other basic machine and deep learning algorithms. The learning models with the highest 97% accuracy. Followed Predictive analysis aids the specialist's prescriptive analysis.

The findings from the various data analytics methodologies show that the predominately favorable attitudes toward particular vaccine brands and gender may be useful for professionals in suggesting an optimal immunization approach. This research does have some restrictions, though. The current study investigation is limited to assessing five well-known vaccination brands launched during data collected duration. While examining the tweets about vaccine brands, the study was unable to consider the location and age of the individuals. The research employs a machine-dependent method to identify a user's gender based on their username, while manually labelled datasets with sex or gender may yield better results. By evaluating social media texts on numerous occasions where human current interests are involved, our proposed data analytics-based framework can also be taken into consideration in other domains of application, such as education, agriculture, cyber-security, etc.

Social media contains more textual information every day. People discuss the COVID-19 pandemic outbreak on social media sites like Twitter, expressing their wants, feelings, and opinions. Thus, by utilizing the benefits of data analytics and sentiment analysis, we propose a systematic framework in this research effort that may be helpful to provide valuable recommendations for appropriate vaccination by medical professionals, administrative authorities, and legislators. The visualization of pandemic information connected to vaccines, the evaluation of the efficacy of various vaccine brands, the prevention of adverse effects, and the provision of both emotional and physical enjoyment for both sexes are all possible uses for our suggested technique. The study was able to carry out descriptive, diagnostic, predictive, and prescriptive studies by putting a data analytics strategy into place. The study concluded by offering specialists practical advice on how to proceed with the right distribution of various vaccine brands among males and females using sophisticated data analysis. In the future, the study will consider a person's age, location, and other contextual factors to help them decide exactly what action to take beforehand.

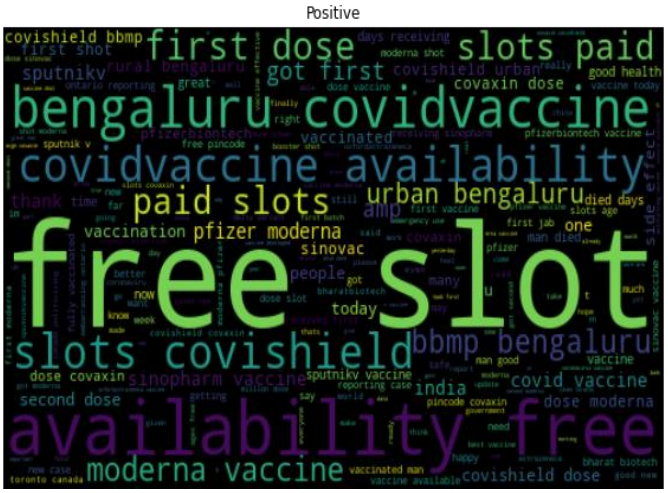

Figure 2. Word Cloud - Positive

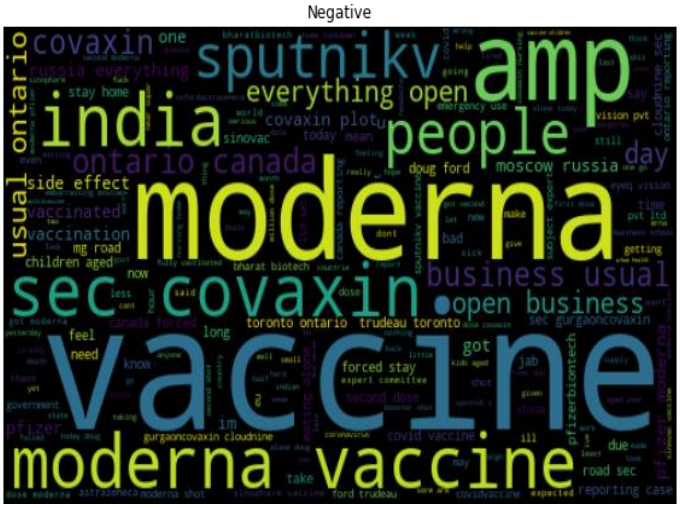

Figure 3. Word Cloud - Negative

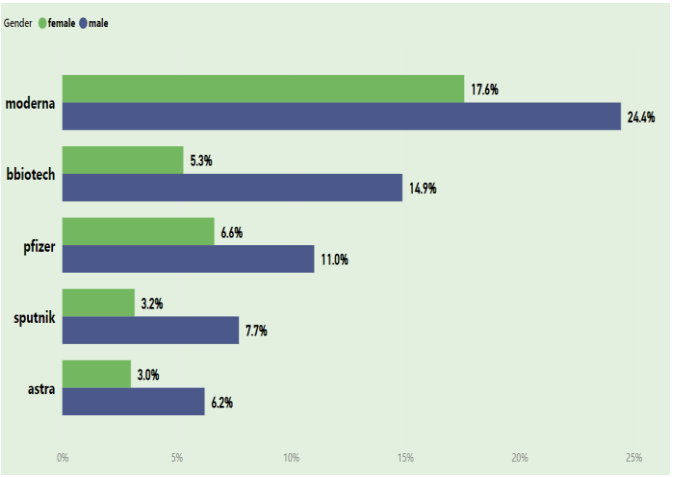

Figure 4. Gender Distribution for Vaccine Brands

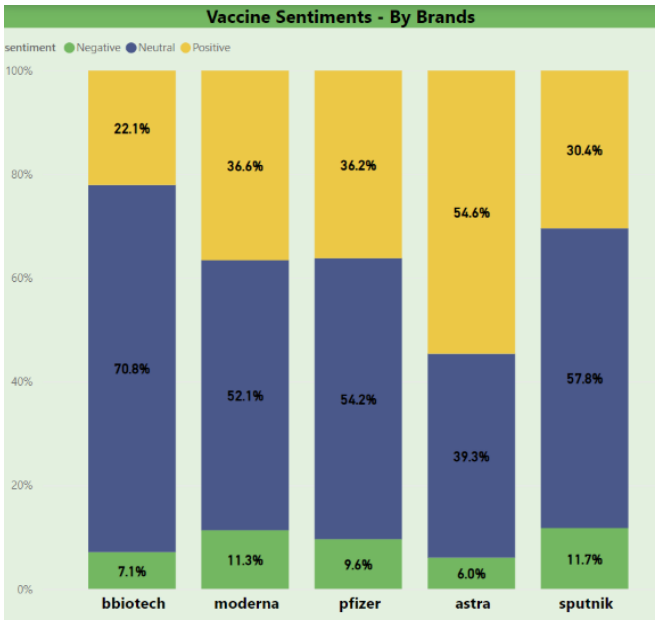

Figure 5. Vaccine Brands Sentiment Ratio

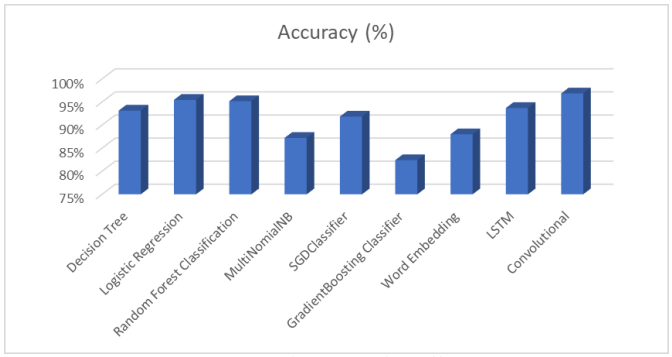

Figure 6. Comparative study of accuracy for different tradition and Depp learning model

The study train various deep learning and machine learning models on dataset and evaluates their performance against that of the Convolution model. The study can see from Figure 5 that Convolution, Logistic and Random Forest models perform effectively and produce results with an accuracy level of greater than 93%. The comparison's findings demonstrate that the Convolutional model outperforms other models and achieves a maximum accuracy of roughly 96%. The results of Table 1 demonstrate that the Convolution model outperforms other deep learning models in terms of accuracy, recall, and f1-measure for weighted averages of positive reviews.

Table 1

Model	Precision	Recall	F1-Measure
Decision Tree	0.95	0.96	0.96
Logistic Regression	0.95	0.99	0.97
Random Forest Classification	0.95	0.99	0.97
MultiNomialNB	0.92	0.92	0.92
SGDClassifier	0.9	1	0.94
GradientBoosting Classifier	0.77	1	0.87
Word Embedding	0.97	0.87	0.91
LSTM	0.94	0.87	0.91
Convolutional	0.91	0.89	0.9

6. Conclusion and Future Scope

Overall, the COVID-19 vaccination brands impact on public and at gender level as seen in social media texts is assessed using our suggested data analytics-based approach. This is the first strategy that applies data analytics techniques like descriptive, diagnostic, predictive, and prescriptive analysis to analyze the sentiments of tweets to ascertain public perceptions about various COVID-19 vaccine brands. The suggested strategy works very well in helping specialists, politicians, and members of the government offer useful recommendations for administering the proper immunizations. The following highlights a handful of our method's important observations.

In the suggested framework, the study explores the necessary elements which completed the dataset's descriptive analysis. Here, the study has noticed that there are more male candidates than female candidates. By investigating, the diagnostic analysis yields significant discoveries. the words that appear frequently and frequently in the tweets for both sexes, as depicted in word cloud. Positive tweets, according to what we've observed, frequently include great or free slots or second dose for vaccinations. Public believe that by getting the right vaccination, the effects of the pandemic can be reduced as soon as possible, and regular life can be resumed. In negative tweets, people typically express their worries about getting immunized, such as staying at home, side effects, etc. According to the diagnostic analysis depicted in Figure 4, the study has also discovered a relationship between the various sentiment polarities regarding vaccine brands from the perspective of gender level, and we have concluded that females have provided more positive feedback for Pfizer or Moderna while Moderna and Biotech received more positive feedback from males based on their expressed sentiments on tweets. The predictive analysis aids in identifying the polarities of the sentiments in tweets about vaccines. The predictive analysis is an outcome of the experiment. demonstrate that the Convolution and Random Forest outperforms other basic machine and deep learning algorithms. The learning models with the highest 97% accuracy. Followed Predictive analysis aids the specialist's prescriptive analysis. To make helpful recommendations by classifying vaccination brands into men and women as seen in Figure 4.

Authors' Contributions

The primary aim of the researcher in this study is to check sentiments of public on Covid-19 Vaccine launched across world. So far, most study in this area focus on analyzing the sentiments of tweets from twitter using one model, either traditional or deep learning

In this study, the researcher has tried to shift focus from one model to comparing the multiple models to understand best model to predict the sentiments of public from twitter on Covid-19 vaccine. This is one of the first studies to combine both tradition machine learning and deep learning algorithm to analyze the data. The researcher used python software to build the predictive model that gives the sentiments of public providing the comments on twitter for Covid-19 Vaccine.

References

[1] Fuyong Zhang, Wang Yi, Liu Shigang, Wang Hua. Decision-based evasion attacks on tree ensemble classifiers. World Wide Web; 23(5):2957–77, **2020.**

[2] Vohra, S., & Teraiya, J. (2013). A Comparative Study of Sentiment Analysis Techniques. International Journal of Information, Knowledge and Research in Computer Engineering, 2(2), pp.**313-317, 2013.**

[3] Machine Learning & its Applications Outsource to India, **2020.**

[4] Jain, A. P., & Dandannavar, P. (2016). Application of machine learning techniques to sentiment analysis. Second International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT), 1(1). pp.**628–632, 2016.** DOI: https://doi.org/10.1109/ICATCCT.2016.7912076

[5] Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. Appl Soft Comp. 2020;97:106754. https://doi.org/10.1016/j.asoc.2020.106 754 PMid:33013254 PMCid:PMC7521435

[6] Praveen SV, Ittamalla R. General public's attitude toward governments implementing digital contact tracing to curb COVID-19–a study based on natural language processing. Int J Per Comp Comm. **2020**. https://doi.org/ 10.1108/IJPCC-09-2020-0121

[7] Chen Y, Yuan J, You Q, Luo J. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. In Proceedings of the 26th ACM international conference on Multimedia, pp.**117-125, 2018.** https://doi.org/ 10.1145/3240508.3240533

[8] Reddy DM, Reddy DN. Twitter Sentiment Analysis using Distributed Word and Sentence Representation. arXiv preprint arXiv:1904.12580. **2019**

[9] Hasan A, Moin S, Karim A, Shamshirband S. Machine learning-based sentiment analysis for twitter accounts. Math Comp App. **2018**;23(1):11. https://doi.org/10.3390/ mca23010011

[10] Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PloS one. 2020;15(9):e0239441. https://doi.org/10.1371/journal.pone .0239441 PMid:32976519 PMCid:PMC7518625

[11] Sanders AC, White RC, Severson LS, Ma R, McQueen R, Paulo HCA, Bennett KP. Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of COVID-19 Twitter discourse. medRxiv, 2020-08. 2021. https://doi.org/10.1101/2020.08.28.20183863

[12] Gupta P, Kumar S, Suman RR, Kumar V. Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter. IEEE Trans Comp Soc Sys. 2020. https://doi.org/10.1109/TCSS.2020.3042446

[13] Das S, Kolya AK. Predicting the pandemic: sentiment evaluation and predictive analysis from large-scale tweets on Covid-19 by deep convolutional neural network. Evol Intell. 2021:1-22. https://doi.org/10.1007/s12065-021-00598-7 PMid:33815622 PMCid:PMC8007226

[14] Gambino OJ, Calvo H, García-Mendoza CV. Distribution of emotional reactions to news articles in twitter. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). **2018.**

[15] Shahriar, K.T., Islam, M.N., Anwar, M.M. and Sarker, I.H., (2022) COVID-19 analytics: Towards the effect of vaccine brands through analyzing public sentiment of tweets. Informatics in Medicine Unlocked, [online] 31February, p.100969, **2022.** Available at: https://doi.org/10.1016/j.imu.2022.100969

AUTHORS PROFILE

Jatin Panjavani Earned his B.E (C.E), MBA and M.S (ML &AI) in 2006, 2009, and 2022, respectively. He is currently working as Senior Data Analyst in IT Company. He has more than 13 years of experience in the field of analytics and data visualization. His main research work focus on Natural Language Processing, Cybersecurity and Data Analytics.

