International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Hybrid Approach for Sentimental Analysis of COVID-19 Twitter Data

Vysyaraju Venkata Rama Raju

Student, Rajam, Vijayanagaram, 532127, India.

ABSTRACT

Data science is the field of applying advanced techniques and scientific principles to extract valuable information from data. User-generated multi-media content, such as images, text, videos, and speech, has recently become prominent on social media platforms like Twitter during the COVID-19 period. The suggested emotion from the text can be successfully mined for many situations using sentiment analysis. The proposed system is a hybrid framework that combines a lexicon-based technique for analysis of sentiments in tweets and labeling with supervised Machine Learning techniques for tweet classification. The long short-term memory (LSTM) has been identified as the preferred ML technique in the framework because to its high accuracy and recall score. The evaluated and final results indicate that our hybrid framework has the potential to automatically classify large tweet volumes, like the tweets on COVID-19, according to the sentiments in society.

Keywords:Data Science, Machine Learning (ML), Twitter, COVID-19, Sentiment Analysis, Tweet Classification, Long Short-Term Memory (LSTM).

1. Introduction

The unique virus disease known as COVID-19, or coronavirus disease, first appeared in the year of 2019. The ubiquitous outbreak inflated social media updates in the form of tweets, messages, and posts. People adopted widely used social media and microblogging channels such as Facebook and Twitter, to share their ideas, thoughts, and reactions. Twitter, one of the most popular social networking sites, is a quickly expanding and intrusive online platform where users can do a variety of tasks, including writing, publishing, updating, and reading tweets, which are brief text messages. The user tweets that reflect our society's attitudes and thoughts throughout the pandemic make the examination of tweets regarding COVID-19 extremely insightful. Fear, anxiety, and other unpleasant symptoms were created by the shifts in sentiment that occurred throughout epidemic periods.

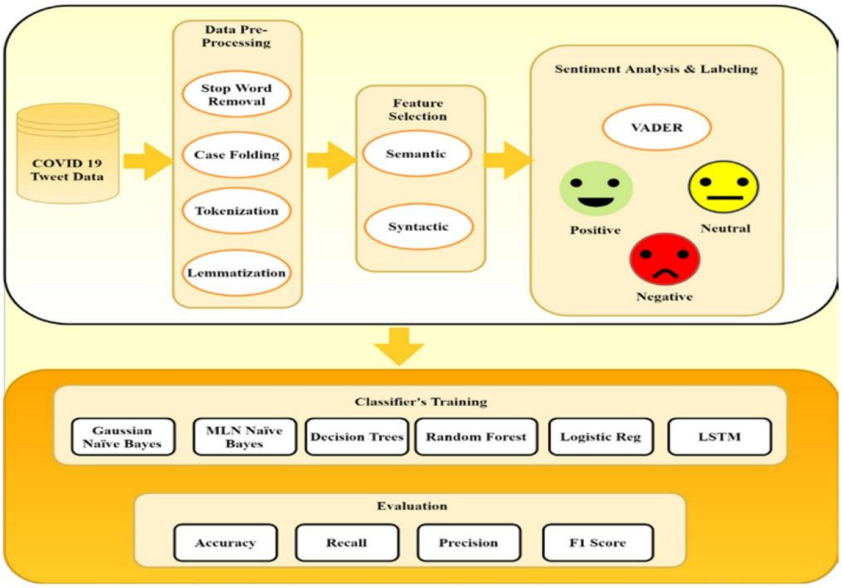

Fig-1: -Hybrid classification framework pipeline: automated lexicon-based labeling using the vader is combined with supervised ml techniques for classifier training.

Therefore, it is important to develop automated methodologies for analyzing and classifying tweets that represent the sentiments in society. On the other side, sentiment analysis (SA) is the method of recognizing and classifying a specific text's polarity in terms of document, sentence, and phrase. This work is a hybrid framework that combines a lexicon-based technique for tweet sentiment analysis and labeling with supervised ML techniques for tweet classification. VADER lexicon-based technique has been used to extract the sentiments that are utilized to label the tweets and the long short-term memory (LSTM) neural network has been selected as the preferred ML technique in the framework due to its high accuracy and high recall score.

Fig-1 shows an overview of the whole process. This work employs the VADER sentiment analysis which is a lexicon-based technique to extract the sentiments that are utilized to label the tweets. These labelled tweets feed into a supervised ML technique, such as Gaussian Naïve Bayes (GNB), multinomial Naïve Bayes (MLNB), logistic regression (LR), decision tree (DT), random forest (RF), and long short-term memory networks (LSTM), to predict the sentiments for novel unlabelled test datasets. To achieve the objective of automatic classification of people's sentiments, this novel hybrid approach combines a natural language processing (NLP) lexicon-based technique with a supervised ML technique. This work also involves a DL-based LSTM neural network to enhance the scope of our work. With the implementation of this framework in day-to-day lives people can automatically classify large tweet volumes, such as the tweets on COVID-19, according to the sentiments in the society.

2. Literature Survey

In paper [1] P. Gupta, S. Kumar, R. R. Suman, and V. Kumar Proposed a method using NLP and machine learning classifiers. Eight distinct classifiers (Multinomial Naive Bayes, Bernoulli Naive Bayes, Logit Regression, Linear Support Vector Classifier, AdaBoost Classifier, Ridge Classifier, Passive Aggressive Classifier, and Perceptron) are employed to effectively identify the data. The combined output of the Text Blob and VADER libraries is used to increase accuracy. On the basis of accuracy, precision, recall, F1-Score, and receiver operating characteristic (ROC) curves with various gramme sizes, the results of all classifiers have been analysed. Additionally, they used the k-fold cross-validation method to validate the chosen data for each model using unigrams, bigrams, and trigrams. The Linear SVC classifier and unigram exhibit the best performance, according to their observations.

In paper [2] WHA, S analyzed tweets by Indian citizens during the COVID-19 pandemic situation.The Bidirectional Encoder Representations from Transformers (BERT) model, a new deep-learning model for text analysis, was used to analyse the data. In order to compare the model performance with BERT, performance was also tested with three more models, including logistic regression (LR), support vector machines (SVM), and long-short term memory (LSTM). In this study, we focused on four Indian internet users' feelings—fear, sadness, anger, and joy—that are connected to both positive and negative emotional categories.

In paper [3] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund and, J. Kim introduced a latest large-scale sentimental data set COVIDSENTI. It consists of 90, 000 COVID-19-related tweets collected. The data set that consists of tweets has been collected by using Tweepy, an official Python Twitter API library.For evaluation and generalization reasons, it is further separated into three identically sized subsets called COVIDSENTI-A, COVIDSENTI-B, and COVIDSENTI-C in each sentiment class.The tweets were divided into three emotion categories: Positive, negative, and neutral. The emotional sentiment was divided into positive, negative, and neutral categories using the Text Blob tool3. We investigate the topic distributions utilizing LDA in order to statistically examine the themes in our data set. The text is constructed from a variety of themes using the topic modelling technique linear discriminant analysis. For vectorization, the phrase frequency-inverse document frequency (TF-IDF) has been utilised. Traditional approaches like TF-IDF, word embeddings, and hybrid approaches that integrate non-contextual word representation methods perform worse than BERT and its variations.

In paper [4] Zhang, X., Saleh, H., Younis, E. M., Sahal, R., & Ali, proposed system for real-time sentiment prediction. It consists of two primary parts: the creation of an offline sentiment analysis model and an online pipeline for sentiment prediction. Using Twitter streaming data, this work creates a real-time method to forecast public opinion towards the coronavirus pandemic. It gathers tweet data regarding the coronavirus using the hashtags #COVID-19 and #Coronavirus before retraining the tweet data with Text Blob. Afterward, uses the TF-IDF feature selection approach with various n-gram sizes. After comparing five machine learning classifiers, the best model for real-time coronavirus sentiment prediction is found.Spark Streaming API receives tweets from Kafka's topic for the analysis step and carries out the preprocessing operations, such as noise removal, tokenization, normalization, and stemming. The best model is then selected after features are extracted and sent to it in a vector structure. In order to classify each tweet's attitude concerning the coronavirus into three distinct classes—positive, negative, and neutral—Spark employs the best prediction model discovered during the offline phase.

In paper [5] Kandasamy, V., Trojovský, P., Machot, F. A., Kyamakya, K., Bacanin, N., Askar, S., &Abouhawwash, M. designed pandemic prediction using deep learning architecture in social media.For implementation details, they used the tweets' dataset collected and filtered by #COVID-19, #Coronavirus, and #COVID'19 hashtags.They have explored the deep learning concept in the real-time system for predicting COVID-19, and it has been developed into two phases. First, they perform Sentiment analysis with Latent semantic analysis pre-diction in offline mode. Secondly, they Exhibit a model in the online mode.In this work, they are using the most standard classifiers: DT, KNN, RF, and SVM. The proposed N-gram stack autoencoder incorporated inside an ensemble system getting-to-know scheme has been developed, validated, and benchmarked with competing schemes from the maximum current and applicable literature.Here, the streaming API of Twitter, apache-spark, and Kafka were involved.

3. Data Collection

The work has utilized the publicly available tweet dataset from Kaggle.com that consists of the COVID-19-related tweets along with other associated information, such as location, retweets, followers, friends, the number of tweets, and hashtags. The analyzed tweet data cannot be attributed to individuals. Therefore, ethics review and approval are not applicable. The utilized dataset has 179 108 tweets collected for 36 days over July and August 2020. This dataset follows the code of CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, which implies free usage of data without any credits. Since the Twitter API restricts daily data collection, the data have been collected for a limited period, i.e., between July and August 2020. During this period, COVID-19 cases were at their peak in many countries, prompting people to express their sentiments on social media platforms. Applying ML techniques to textual data is computationally demanding, requiring significant time and cost. We utilized the Google Cloud Platform (GCP) to implement our classification framework pipeline. Our configured compute engine consisted of 2 vCPUs, 1 tesla K80 GPU unit, 13-GB RAM, and 100-GB SD.

4. Automated Hybrid Tweet Classification

We used an unsupervised technique for sentiment analysis to get around the traditional requirement of a labelled sentiment dataset. We use a lexicon-based method, which links words with emotions like positive, negative, and neutral (all possible worlds of the English dictionary). For classifying words according to a polarity score lexicon-based technique is straightforward and effective. After the dataset has been labelled, we applied and compared a number of supervised machine learning approaches to identify the one that produces the highest accuracy and recall. We tuned the parameters using fivefold cross-validation to get the optimum performance out of the ML algorithms.We then go on to discuss the pipeline for our proposed hybrid categorization framework as shown in fig1.

4.1. Data Pre-processing:

Starting with the unprocessed tweet text data, we used NLP to clean the text and remove redundant information and anomalies to make it more informative.

1) Stop Word Removal: Using the Python-based NLP toolset, we have eliminated stop terms such "over," "under," "again," "further," "then," "once," and "there." Using regular expressions, we further deleted URLs, user mentions, special characters (*, $,!), and hashtags (#).

2) Case Folding: A word's polarity can be significantly impacted by its upper- and lowercase usage (positive, negative, and neutral). Depending on the nature of the corpus and the intended use of the corpus, words in the text may be changed from uppercase to lowercase or vice versa. Our dataset consists of tweets with a small word count that are frequently written in lowercase and in the typical communication style. Because there are few capital words in tweets, we have created case-folding.

3) Tokenization: We converted each twitter sentence into tokens (words) using tokenization, which enables machine learning (ML) algorithms to understand the meaning of each word.

4) Lemmatization: In NLP, lemmatization is the process of locating the vocabulary's base words in order to group all a word's variations into a single category (i.e., lemma). As an illustration, the fundamental word "stop" includes several variations, including "stopped," "stopping," and "stops." All these forms derive from the stop base class. To make the ML effective, we conducted the lemmatization.

4.2. Feature Selection:

Semantic and syntactic properties are extracted through text classification. While the syntactic feature consists of unigrams, bigrams, and n-grams (i.e., groups of associated words), the semantic feature consists of the sentiment associated with a word. A tweet's single words, such as "corona," "virus," and "death," are referred to as unigrams. As in "corona virus," "covid vaccination," and "positive cases," bigrams are the combining of two words. All n-grams, which are frequently used to classify texts, include Unigrams, Bigrams, Trigrams, and Quadgrams, which are all word chains.

4.3. Sentimental Analysis:

Twitter has a limited amount of textual data, and analysis of sentiments is the process of separating out a piece of information, like a tweet, that contains positive, neutral, and negative attitudes. When referring to sentiment analysis, the term "polarity-based analysis" is frequently used because polarity indicates a range of emotions. A popular sentiment analysis method is VADER. Because it can assess the sensitivity of messages, the rule-based VADER model was created especially for the sentiment analysis of social media posts. Although there are alternative methods, including the General Inquirer and linguistic inquiry and word count (LIWC), VADER beats them. As a result, VADER was utilised in our system for sentiment analysis. The compound scores that VADER provides can be used as a threshold value to distribute the tweets across the three sentiment classes, together with the polarity scores of the positive, negative, and neutral classes.

4.4. Data Labelling:

Before using supervised ML algorithms, tweet labelling is a crucial step. Based on the sentiment scores of the raw tweets, we have assigned labels. Following the sentiment analysis, we have produced a new dataset with all of the tweets' associated sentiment values (labels). By utilising Python's for loop and if-else conditional statements, the entire process of data tagging is automated. We have given a score of 0 to all tweets with a positive sentiment, 1, for tweets with a neutral sentiment, and 2, for tweets with a negative feeling.

4.5. Classifiers:

Since we employed supervised ML and DL methods to categorise the tweets automatically based on feelings, these methods are known as classifiers. As a classifier, we have specifically used GNB, MLNB, LR, DT, RF, and LSTM.

4.6. Evaluation:

By the support of performance metrics accuracy, precision, recall, and F1-score classification, we assessed the effectiveness of the used ML approaches. Formally, let TP stand in for the total number of true positives, or the total number of positively identified samples. TN stands for true negatives. The term "false positive" (FP) refers to the number of negative samples that were mistakenly labelled as positive. The term "false negatives" (FN) refers to the quantity of positive samples that were mistakenly categorised as negatives. Accuracy gauges total effectiveness. The accuracy of the results generated is measured. Recall counts the number of tweets that were properly categorised, and the F1-Score is a weighted average of recall and precision ratings. Formally,

$$\text{Precision} = TP/(TP + FP)$$

$$\text{Recall} = TP/(TP + FN)$$

$$\text{F1 score} = (2(\text{precision} \times \text{recall}))/(\text{precision} + \text{recall})$$

Accuracy = (TP + TN)/(TP + FP + FN + TN)

5. Methodology

This paper consists of lexicon-based automated tweet sentiment analysis and labeling in conjunction with state-of-the-art DL techniques for tweet classification.VADER as a sentiment analysis tool would likely not suffice to process these enormous text volumes. Therefore, in order to advance the scalability of text processing for sentiment analysis, we propose a hybrid of VADER with ML techniques. VADER provides the data labels according to a polarity score, i.e., score for positive, negative, and neutral sentiments and the labeled data form the basis for the training of the ML techniques. To obtain the best performance from the ML techniques, we have used fivefold cross-validation to tune the parameters. keeping in mind an unbalanced dataset scenario, we have also generated other evaluation measures namely, recall, precision, and F1 score. The LSTM technique outperformed all other ML techniques, achieving a high-test accuracy.

Gaussian Naïve Bayes:

Naive Bayes is a probabilistic ML algorithm which will be utilized in several classification tasks. Typical applications of Naive Bayes are classification of documents, filtering spam, prediction and then on. This algorithm relies on the discoveries of mathematician and hence its name.

The name “Naïve” is employed because the algorithm incorporates features in its model that are independent of every other. Any modifications within the value of 1 feature do not directly impact the worth of the other feature of the algorithm. the most advantage of the Naïve Bayes algorithm is that it is an easy yet powerful algorithm.

Multinomial Naïve Bayes:

Multinomial Naive Bayes is one amongst the variations of the Naive Bayes algorithm in machine learning which is incredibly useful to use on a dataset that is distributed multinomially. When there are multiple classes to classify, this algorithm may be used because to predict the label of the text it calculates the probability of every label for the input text and so generates the label with the best probability as output.

The multinomial Naive Bayes algorithm may be a probabilistic learning method that is mostly employed in linguistic communication Processing (NLP). The algorithm relies on the Bayes theorem and predicts the tag of a text like a bit of email or news article. It calculates the probability of every tag for a given sample so gives the tag with the very best probability as output.

RandomForest:

Random Forest is a widespread machine learning algorithm that belongs to the supervised learning technique. Classification and Regression problems in ML be done by using it. it is supported the concept of ensemble learning, which may be a process of joining multiple classifiers to unravel an intricate problem and to extend the performance of the model.

Random Forest is a classifier that encompasses several decision trees on various subsets of the given dataset and takes the common to improve the predictive accuracy of dataset. The larger number of trees within the forest ends up in higher accuracy and avoids the matter of overfitting.

Decision Tree:

Decision Tree is a supervised ML technique that is used for both classification and regression problems, but frequently it is chosen for solving classification problems. It is called Decision tree, because just like a tree, it starts with a root node, which expands on further branches and constructs a tree-like structure.

By calculating what quantity accuracy each split will cost us, employing a function. The split that costs least are chosen, this algorithm is recursive in nature because the groups formed will be sub-divided using same strategy. because of this process, this algorithm is additionally called the Greedy Algorithm, as we've an more desire of lowering the price. This makes the foundation node as finest predictor/classifier.

Logistic Regression:

For binary classification issues, the suitable supervised ML algorithm, logistic regression is applied. A logistic function is used to model the probabilities that indicate the probable outcomes of a single test. In essence, logistic regression classifies the data and reduces outliers using a logistic function.

$$Logistic\ function = \frac{1}{1+e^{-x}}$$

Multinomial logistic regression is capable of modelling in conditions like with two or more than two discrete potential outcomes. A model cannot fit new data or forecast observations that have not yet occurred if the training data are learned too carefully. Overfitting is the term for this situation. The overfitting will be reduced using the logistic function.

Long Short-term memory:

Long Transient Memory (LSTM) networks are a kind of Intermittent Brain Organization that can learn request reliance. The result of the past step is utilized as contribution to the ongoing move toward RNN. Hochreiter and Schmidhuber made the LSTM. It resolved the issue of RNN long haul reliance, in which the RNN cannot foresee words put away in long haul memory however can make more precise expectations considering current information. RNN does not give a productive presentation as the hole length rises. The LSTM might save data for quite a while naturally. It is utilized for time-series information handling, expectation, and arrangement.

6. Results & Discussions

Most of the examined tweets were very near the historic 140-character limit for tweets. This shows that COVID-19-related tweets are relatively lengthier than the overall average tweet length of around 90 characters. The relatively high character count of the COVID-19-related tweets could also be because of the gravity and complexity of the COVID-19 pandemic and the resulting inherent desire of individuals to precise their feelings in relatively rich detail in long tweets. Importantly, these long tweets increase the text volume that needs processing, intensifying the requirement for automated sentiment analysis with ML. At the identical time, the long tweets are rich in content and will convey nuanced sentiments that are necessary for gaining insights into societal sentiments. The VADER sentiment analysis gave the subsequent distribution of the user sentiments: 38.5% of tweets belong to the positive sentiment class, 34.7% belong to the neutral sentiment class, and 26.6% of tweets belong to the negative sentiment class. Thus, most sentiments were positive or neutral, indicating an optimistic opinion toward the pandemic. Table I illustrates some example tweets from all three sentiment classes, together with their polarity scores, whereby the contextual meaning of tweets indicates the user's sentiments.

Table I: - example tweets of all sentiment classes with polarity scores

Sentiments	**Example tweets**	**Scores**
Positive	"@brookbanktv the one gift #COVID19 has given me is an appreciation for the simple things that were always around me."	{'neg': 0.0, 'neu': 0.754, 'pos': 0.246, 'compound':0.7351}
Negative	"#coronavirus #covid19 deaths continue to rise. It is almost as bad as it ever was."	{'neg': 0.152, 'neu': 0.848, 'pos':0.0, 'compound': -0.4976}
Neutral	"Holy water in times of #COVID19"	{'neg': 0.0, 'neu': 1.0, 'pos':0.0, 'compound': -0.4976}

We compare all the mentioned techniques in to identify the best and required model for the automatic classification of societal sentiments. Below Table II shows the percentage values of the different evaluation measures for all examined ML techniques.

Table II: -evaluation measure scores for model performances

Models	**Test Accuracy**	**Recall**	**Precision**	**F1 Score**
Gaussian NB	68%	68%	70%	68%
Multinomial NB	70%	70%	70%	70%
Decision Tree	64%	64%	72%	63%
Logistic Regression	79%	79%	80%	78%
Random Forest	77%	77%	78%	76%
LSTM	83%	83%	82%	82%

We observe from Table II that DT and Gaussian NB perform relatively poorly with performance metrics generally under 70%, while multinomial NB consistently achieved 70% across all metrics. The RF technique achieves somewhat improved performance with metrics between 76% and 78%, while LR performed 2% better than RF for every metric. However, only the LSTM technique achieved consistent performance within the 82%–83% range for all performance metrics. The LSTM technique thus outperformed all other ML techniques, achieving a test accuracy of 83% and a recall score of 83%.

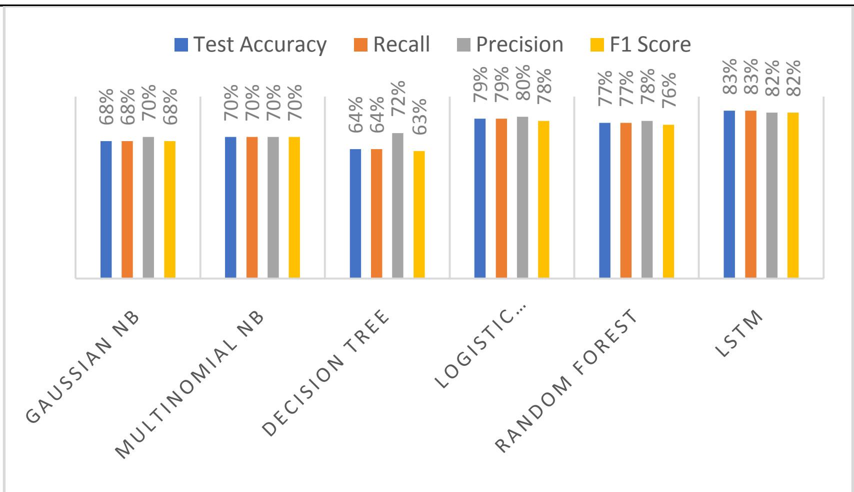

Fig-1: Graphical representation evaluation measure scores

7. Conclusion

For the goal of sentiment analysis in the COVID-19 topic area, we created a novel hybrid framework that combines a lexicon method for twitter sentiment analysis and labelling with a DL technique for tweet categorization. We specifically extracted the positive, negative, and neutral emotions by using the VADER lexicon technique to identify the COVID-19-related tweets based on their associated feelings in order to automatically classify the social sentiments on Twitter. We employed many ML and DL methods for the classification challenge. LSTM surpassed all other methods with a classification test accuracy of 83%. In comparison to the VADER method, the trained ML classifier was able to handle data up to an order of magnitude faster.

We should also consider that the diagnostic analysis of society sentiments was the main emphasis of this study. Analysing the purposeful manipulation of society sentiments is the most important aspect of related social media research. Examining how sentiment analysis and sentiment manipulation interact, for example, to discover intentional attempts to push public sentiments in a particular direction, is a significant area for future research.

REFERENCES

[1] P. Vyas, M. Reisslein, B. P. Rimal, G. Vyas, G. P. Basyal and P. Muzumdar, "Automated Classification of Societal Sentiments on Twitter With Machine Learning," in IEEE Transactions on Technology and Society, vol. 3, no. 2, pp. 100-110, June 2022, doi: 10.1109/TTS.2021.3108963.

[2] P. Gupta, S. Kumar, R. R. Suman and V. Kumar, "Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter," in IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 992-1002, Aug. 2021, doi: 10.1109/TCSS.2020.3042446.

[3] Singh, C., Imam, T., Wibowo, S., &Grandhi, S. (2022). A Deep Learning Approach for Sentiment Analysis of COVID-19 Reviews. *Applied Sciences*, *12*(8), 3709, doi: 10.3390/idr13020032.

[4] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund and J. Kim, "COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis," in IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 1003-1015, Aug. 2021, doi: 10.1109/TCSS.2021.3051189.

[5] Zhang, X., Saleh, H., Younis, E. M., Sahal, R., & Ali, A. A. (2020). Predicting coronavirus pandemic in real-time using machine learning and big data streaming system. *Complexity*, *2020*.

[6] Kandasamy, V., Trojovský, P., Machot, F. A., Kyamakya, K., Bacanin, N., Askar, S., &Abouhawwash, M. (2021). Sentimental Analysis of COVID-19 Related Messages in Social Networks by Involving an N-Gram Stacked Autoencoder Integrated in an Ensemble Learning Scheme. *Sensors*, *21*(22), 7582.

[7] Khan, R., Shrivastava, P., Kapoor, A., Tiwari, A., & Mittal, A. (2020). Social media analysis with AI: sentiment analysis techniques for the analysis of twitter covid-19 data. *Critical Rev*, *7*(9), 2761-2774.

[8] Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020). Word frequency and sentiment analysis of twitter messages during coronavirus pandemic. *arXiv preprint arXiv:2004.03925*.

[9] Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. *Ieee Access*, *8*, 181074-181090.

[10] Dubey, A. D. (2020). Twitter sentiment analysis during COVID-19 outbreak. *Available at SSRN 3572023*.

[11] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., &Hassanien, A. E. (2020). 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*, *97*, 106754.

[12] Naw, N. (2018). Twitter sentiment analysis using support vector machine and K-NN classifiers. *IJSRP*, *8*, 407-411.

[13] Bhat, M., Qadri, M., Noor-ul-Asrar Beg, M. K., Ahanger, N., & Agarwal, B. (2020). Sentiment analysis of social media response on the Covid19 outbreak. *Brain, Behavior, and Immunity*, *87*, 136.

[14] Wang, Y., Wang, M., & Xu, W. (2018). A sentiment-enhanced hybrid recommender system for movie recommendation: a big data analytics framework. *Wireless Communications and Mobile Computing*, *2018*.

[15] Jnoub, N., Al Machot, F., &Klas, W. (2020). A domain-independent classification model for sentiment analysis using neural models. *Applied Sciences*, *10*(18), 6221.

[16] Gupta, I., & Joshi, N. (2020). Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic. *Journal of intelligent systems*, *29*(1), 1611-1625.

[17] Geetha, R., Karthika, S., &Mohanavalli, S. (2020). Tweet classification using deep learning approach to predict sensitive personal data. In *Advances in Electrical and Computer Technologies* (pp. 171-180). Springer, Singapore.

[18] Barkur, G., & Vibha, G. B. K. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian journal of psychiatry*, *51*, 102089.

[19] Saleena, N. (2018). An ensemble classification system for twitter sentiment analysis. *Procedia computer science*, *132*, 937-946.

[20] Samuel, J., Ali, G. G., Rahman, M., Esawi, E., & Samuel, Y. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, *11*(6), 314.