ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Evaluation of Online Opinions During the COVID-19 Pandemic in India

[1] Mrs. G. Haritha Rani, [2] Maddukuri Uday Kiran,

[1]Associate Professor,Dept.of MCA, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

[2]Students,Dept.of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

Abstract—Recently, sentiment analysis—a subfield of natural language processing—has become more important and popular than any other subfield. Emotions may be retrieved from text using sentiment analysis. Social media has been an essential tool for the spread of information during the COVID-19 epidemic. Conclusions drawn from such analyses are useful for decision-makers. Examining how Indians felt during the lockdown, this research makes use of NLP and ML models. Our data comes straight from Twitter using the Tweepy API. We utilize the VADER and TextBlob libraries to annotate it. Data preparation is done using the NLTK tool. The results show that the Ensemble model with unigram performs well in experiments. When asked about the government's decision to impose a lockdown, the vast majority of Indians expressed their approval.

Terms for Indexing—COVID-19, Natural Language Processing, Natural Language Knowledge Base, Machine Learning, and Lockdown.

INTRODUCTION

Rapid global expansion characterizes the COVID-19 pandemic, which started in Wuhan, China. The SARS-CoV-2 virus was responsible for the COVID-19 pandemic. The illness allegedly initially manifested itself at a Wuhan seafood market. Its direct impact on respiratory systems makes the disease very dangerous. One of the main symptoms of the illness is a diminished sense of smell and taste. Multiple failed attempts to control the virus before it spread were made. In an effort to stem the spread of the virus, all international flights were subject to screening after the first confirmed case in China. There has been a disturbing uptick in the COVID-19 pandemic transmission rate. This made disease management a major challenge. There was a subsequent exponential growth in the number of reported events. The rate of rise of COVID-19 cases is directly proportional to the current active case count. There was a lag between its inception and its global pandemic status. There has been no indication of any strain of the sickness so far. The COVID-19 virus has many known varieties, including alpha, gaama, delta, and BF. Omicron is another variety. Since certain vaccinations worked against one variation but not another, this made vaccine development more challenging. Since the human immune system is unable to combat delta, it was determined that this variation was the most lethal of the bunch. After each wave, the number of active cases drops dramatically until the next wave arrives. The only way to control this contagious illness was to isolate people from one another. The only successful strategy to manage the illness was social distance, such as lockdown and quarantine. Many were in favor of a lockdown, while others were against it. Lockdown advocates knew that quarantine was the only way to stop the spread of the virus and save lives. But there are other factors that significantly impact the imposition of lockdown, including as demographics, population density, and the economics of the nation. Finding people's opinions and perspectives is, hence, crucial. The government can make better judgments in the event of a future epidemic if it analyzes public mood. Key goals of the team Our goal is to compile all of the tweets sent by Indian individuals on the nationwide lockdown imposed to manage the coronavirus. Using a variety of machine learning models, we will classify user sentiment as either very positive, somewhat positive, neutral, moderately negative, or very negative. For the purpose of evaluating machine learning models. Here is the breakdown of the remaining sections of the paper: In Section II, we cover the relevant literature. Section explains the proposed strategy, while Section IV analyzes the experimental outcomes. The article is concluded in Section V.

RELATEDWORK

Since the majority of people express their opinions on this platform, sentiment research on social media may be quite beneficial. A number of studies are using sentiment analysis to find out what people think about different topics. We gathered tweets pertaining to COVID-19 and determined the prevalence of terms like racial bias, vaccination, and preventative measures in [1]. A total of 5,31,96 tweets from around 12,000 individuals were compiled. Tweets about COVID-19 have been on the upswing since January 21, 2020. Frightened people were the most common emotion conveyed in the messages, with 30% expressing astonishment. There was a clear correlation between the amount of racist postings and the number of current COVID-19 positives. Financial and political ramifications of the COVID-19 pandemic were among the most talked-about issues. In [2], the feelings and thoughts of people in the US and India are examined. The collection of tweets took place from April 1, 2020, to April 9,

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

2020. The NRC Emotion Lexicon was used to get opinions from the general public. An overwhelming majority of tweets about the Indian prime minister are favorable. About half of the people felt positively about the US Prime Minister.

Analyzed in[3] are the feelings of people from various nations about the COVID-19. The cleansed tweets' emotions and polarity were determined using an LSTM model. Complete quarantine was shown to have little supporters among the world's nations. In order to manually identify 10% of the data gathered from January to March 2020 for tweets connected to COVID-19 as positive or negative, the Naïve Bayes technique was used (as mentioned in reference 4). In recent years, academics have increasingly favoured using social media data to analyze user sentiment about an event, issue, or product [5] - [6]. In [7], remote supervision is used to do sentiment analysis on tweets. The training data, which served as noise, consisted of tweets accompanied by emoticons. The models were constructed using the Naive Bayes classifier, SupportVectorMachine (SVM), and maximumentropy (MaxEnt). Bigrams, unigrams, and POS were their characteristics. A support vector machine (SVM) model trained on unigram data outperforms alternatives. During the epidemic, false hypotheses and misleading reports have been widely circulated [8].

A number of recent research have shown that sentiment analysis on Twitter is quite effective. The various methods and instruments needed to conduct modulated analysis are detailed in [9]. Annotating data is made easier using TextblobandVader. To quantify user sentiment, VADER was used in the study [10]. Posts made by health groups in [11] [12] were used to forecast the future of pandemic situations. We were able to make accurate predictions because we quickly gathered data from many sources. Researchers looked at the use of social media during the COVID-19 outbreak in [13]. Credible people, such journalists and physicians, only accounted for 1% of the tweets. Approximately 84% of the tweets included untrustworthy medical information, and 16% were determined to be deceptive. In determining the central argument, sentiment analysis plays a vital role.

PROPOSEDAPPROACH

The COVID-19 pandemic shown in Figure 1 provides the foundation for the suggested technique to assess attitudes on Twitter data. During the Corona epidemic in India, we used this approach to do a sentiment analysis of the lockdown. From March 25, 2020, to April 14, 2020, the Indian government instituted the COVID-19 lockdown.

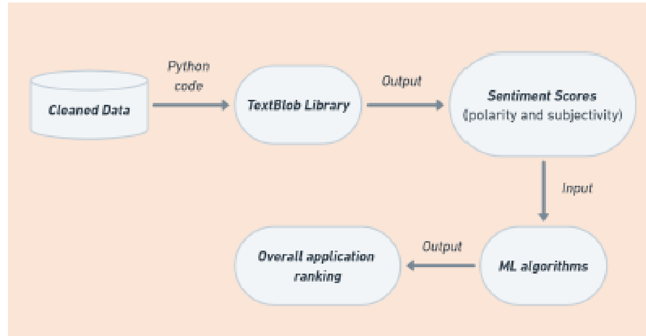

Fig. 1. Framework for Sentiment analysis

DataExtraction Sentiment analysis was conducted on this data. Consequently, we generated the dataset manually as objective-specific data is not accessible. The worsening situation has led to a dramatic increase of remarks on the pandemicon on social media from March 2020. From April 5, 2020, to April 17, 2020, we gathered over 10,000 tweets using the hashtag "#Indialockdown" using the Tweepy API.

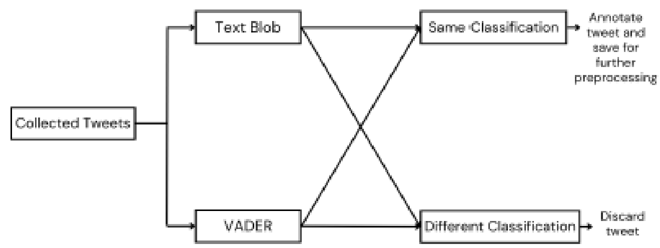

Fig. 2. Data annotation process

Annotation of Data

Once we have extracted data from Twitter and collected all of the tweets, we may sort them into five categories: very positive, somewhat positive, neutral, somewhat negative, and very negative. This method is shown in Figure 2. We do this by allocating all of these tweets to various polarities using the TextBlob and VADER libraries. The next step is to refine the polarity by combining the findings from TextBlob and VADER. Tweets with different polarity results are not suitable and will not be processed further.

Pre-processing Data

Links, Twitter-specific keywords (such as hashtags and tags beginning with @), one-letter words, digits, and punctuation marks are some examples of the accidental terms that might be in the data that we gather. During the testing and training of the four-classifier, these phrases might serve as noise and adversely impact the model's outputs. The classifier's efficiency may be enhanced by eliminating the noise in the labeled data set. In order to clean the data set of any contaminants, we use a pre-processing module that adheres to the procedures shown in Figure 3. In this stage, a module is

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

used to eliminate all of the aforementioned contaminants. We remove punctuation, stop words, tokenization, stemming, and lemmatization when data pre-processing is complete, and then we turn the data into a data frame.

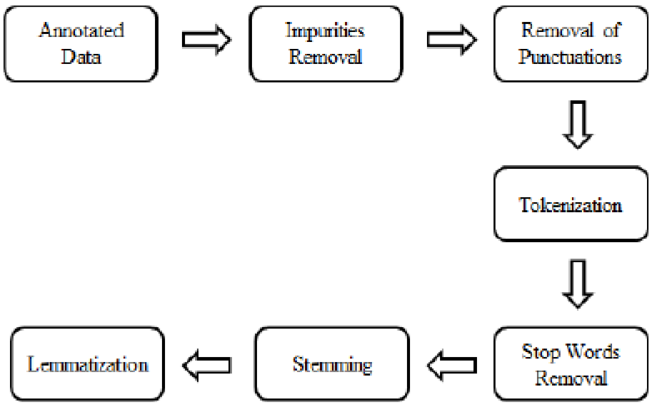

Fig. 3. Data preprocessing

The process of vectorization When training a model, machine learning models can only use numerical data. Hence, the text data must be transformed into numerical values. Using the extractor's CountVectorizer function, we determine word frequencies. The sparse matrix that CountVectorizer uses to determine how often each word appears in the text is shown in Table I. "His friend went out shopping with his brother and sister," Doc1 says, as an example. This text will be converted into a sparse matrix with an alphabetical index of the words using CountVectorizer, a package we use for vectorization:

TABLE I
EXAMPLE MATRIX FOR COUNTVECTORIZER

Index	0	1	2	3	4	5	6	7	8
Doc1	1	1	1	2	1	1	1	1	1

"He went out,":7, "shopped,":6, "was with":8, "brother":1, and "sister":5, all while his buddy was at 2:2. Classifier Training and Testing Machine learning models are fed the data after feature extraction. Logistic Regression, Bernoulli Naive Bayes, AdaBoost Classifier, Perceptron, MultinomialNaive-Bayes, LinearSVC, PassiveAggressiveClassifier, RidgeClassifier, and the Ensembled model are the eight machine learning classi-fies that we have used in our job. We trained the classifiers using 80% of the data and tested them using 20%. Uni-, bi-, and tri-gram analysis has been performed on the aforementioned classifiers.

EXPERIMENTALRESULTSANDDISCUSSION

We have covered the various characteristics, including as accuracy, precision, recall, and F1-Score, and how they are used to evaluate the results given by the classifiers. The k-fold cross-validation procedure (k=10) was used to validate the dataset using unigram, bigram, and trigram. The acronyms TP, TN, FP, and FN denote True Positive, True Negative, False Positive, and False Negative, respectively, in the confusion matrix. Below you can find the performance metrics that were used in this project:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

Precision: It is a parameter which tells the number of positive predictions that was made are correct:

$$Precision = TP/(TP + FP) \quad (2)$$

Recall: Recall is a parameter that tells us the number of positive cases correctly predicted by a classifier over all positive cases in data.

$$Recall = TP/(TP + FN) \quad (3)$$

F1-Score: Harmonic mean of both precision and recall is known as F1-Score.

$$F1 - score = 2(Precision Recall)/(Precision + Recall) \quad (4)$$

The percentage of accurate predictions relative to the total number of guesses is called accuracy. Multiple-Validation Results To compare and contrast the models, this section discusses and compares the cross-validationscores for unigram, bigram, and trigram.

TABLE II
CROSS VALIDATION SCORES OF THE MODELS FOR UNI, BI AND TRI-GRAMS

Models	Uni-Gram	Bi-Gram	Tri-Gram
Logistic Regression	0.65	0.56	0.54
Bernoulli Naive-Bayes	0.56	0.52	0.52
Adaboost Classifier	0.58	0.56	0.53
Perceptron	0.68	0.58	0.54
Multinomial Naive-Bayes	0.60	0.51	0.50
LinearSVC	0.66	0.56	0.54
Passive Aggressive	0.66	0.56	0.54
Ridge Classifier	0.66	0.57	0.54
Ensemble Model	0.66	0.56	0.54

According to Table II, the perceptron model has the greatest cross-validation score, which is around 0.68 in the case of the unigram. Ensembled model and Ridge Classifier, the two models after Perceptron, both have cross-validation scores of about 0.66. Additionally, Perceptroni only gives the greatest cross-validation results for bigrams. Logistic Regression, Perceptron, Ridge Classifier, LinearSVC, Passive Aggressive, and Ensemble model are the six models that provide trigrams with almost identical cross-validation scores (0.54).

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

TABLE III
ACCURACY, EXECUTION TIME AND RNK OF THE MODELS FOR UNIGRAM

Models	Accuracy(%)	Execution Time(sec)	Accuracy Rank
Logistic Regression	68.5	28.9305	6
Bernoulli Naive-Bayes	62	3.36085	9
Adaboost Classifier	63.5	80.6551	8
Perceptron	70.5	6.82662	3
Multinomial Naive-Bayes	66.0	2.35138	7
LinearSVC	70.5	1.68393	2
Passive Aggressive	69	16.0927	5
Ridge Classifier	70	4.23037	4
Ensemble Model	72	48.5128	1

Accuracy
Here we go over the details of each model that was examined for unigram, bigram, and trigram, including their accuracy ratings, execution time, and rank. Precision for Unigram: The precision, time to execution, and ranking of unigram models are shown in Table III. Unigram execution time for LinearSVC is at least 1.68 seconds. The Ensemble Model outperforms all other models with a 72% accuracy rate. Execution time determines their rankings, and both the Perceptron and Lin-ear SVC models have substantial accuracy.

TABLE IV
ACCURACY, EXECUTION TIME AND RANK OF THE MODELS FOR BIGRAM

Models	Accuracy(%)	Execution Time(sec)	Accuracy Rank
Logistic Regression	60.5	78.02	6
Bernoulli Naive-Bayes	55.5	8.815	8
Adaboost Classifier	60	431.51	7
Perceptron	61	7.74	3
Multinomial Naive-Bayes	54.5	2.44	9
LinearSVC	61	2.09	2
Passive Aggressive	61.5	24.699	1
Ridge Classifier	61	10.305	4
Ensemble Model	61	37.3381	5

TABLE V
ACCURACY, EXECUTION TIME AND RANK OF THE MODELS FOR TRIGRAM

Models	Accuracy(%)	Execution Time(sec)	Accuracy Rank
Logistic Regression	57	53.89	5
Bernoulli Naive-Bayes	55.5	18.15	8
Adaboost Classifier	58	266.14	1
Perceptron	56.5	8.982	7
Multinomial Naive-Bayes	52	2.89	9
LinearSVC	57	2.24	2
Passive Aggressive	57	24.88	4
Ridge Classifier	57	17.58	3
Ensemble Model	57	58.26	6

Bigram Precision: The acquired accuracy scores for each model for bigram are shown in Table IV.

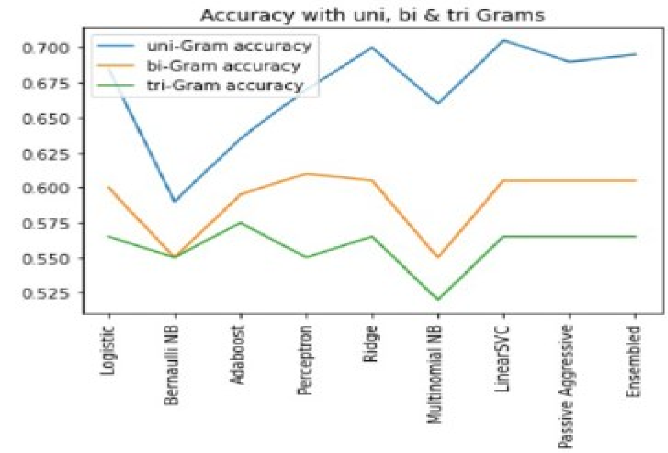

Fig. 4. Accuracy with uni, bi & tri-Grams

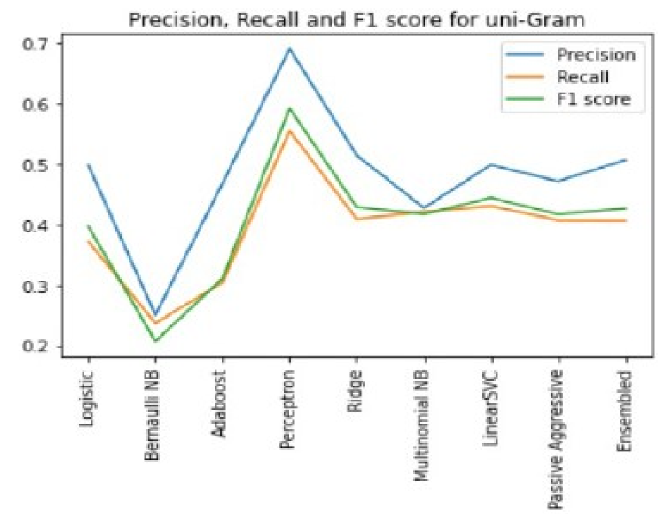

Fig. 5. Precision, Recall and F1 score for uni-Gram

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

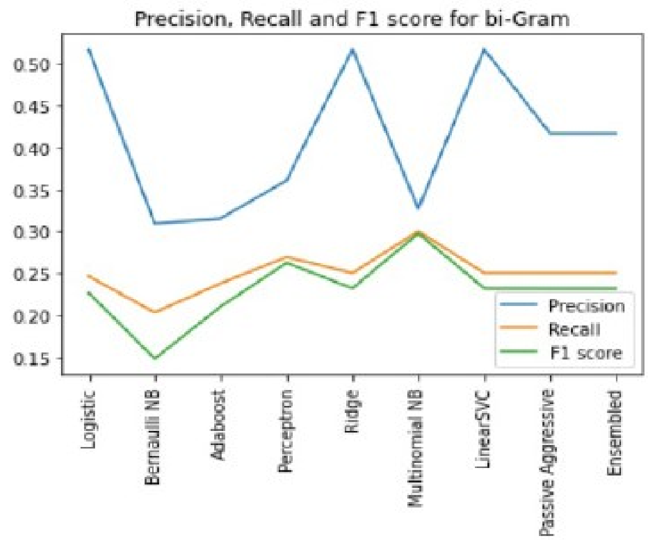

Fig. 6. Precision, Recall and F1 score for bi-Gram

With bigram, LinearSVC may be executed in as little as 2.09 seconds. Among all the models, Passive Aggressive has the highest accuracy rate at 61.5%. The order of execution duration determines the accuracies of the Perceptron, LinearSVC, Ridge Classifier, and Ensemble Models, all of which are almost identical at 61%. Precision for Trigram: Table V displays the accuracy, execution time, and rack for each trigram model. With a trigram accuracy of 58%, AdaboostClassifier is the most accurate among the models in Table IV. Execution time determines the order of the models, which include Logistic Regression, LinearSVC, Passive Aggressive, Ridge Classifier, and Ensemble Model, all of which achieve almost identical 57% accuracies.

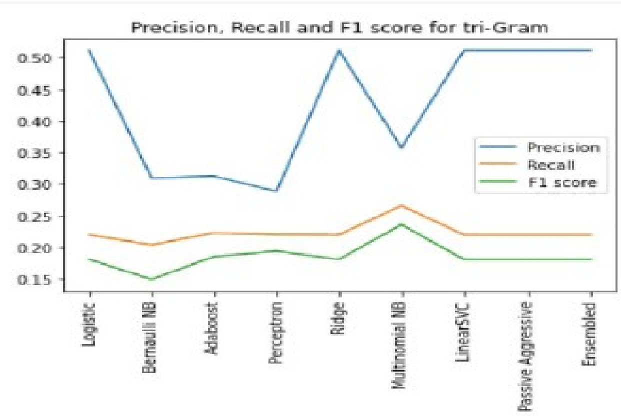

Fig. 7. Precision, Recall and F1 score for tri-Gram

shown in the figure are eight ML models that use N-grams as their metric. See Figures 4–7. When it comes to Unigram models, the Ensem-ble Model is the gold standard. The accuracy of the Perceptron and LinearSVC models is comparable. Unigram models consistently outperform bigram and trigram. Except for Bernoulli-NB and multinomial-NB, all models are almost as accurate for bigram. Adaboost is the most accurate model for Trigram.

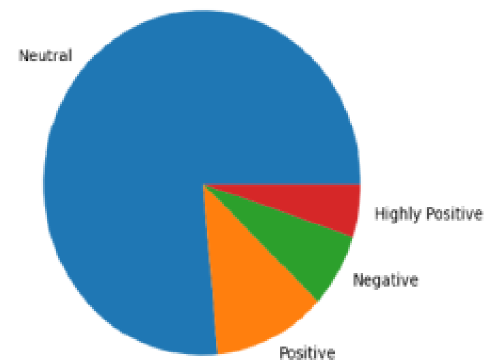

Fig. 8. Distribution of People Perception Towards LockDown

The Ensemble Model is the best option for analyzing sentiment during the COVID-19 lockdown because unigrams are more accurate than bigrams and trigrams. Precision, Recall, and F1-Score are metrics by which Perceptron stands out among Unigram models. The LogisticModel is the most accurate of the Bigram models. After the Logistic Model and Bernoulli-NB models, the Linear SVC and Ridge models are the second most and third least accurate, respectively. Multinomial-NBishaving the greatest value among other models in case of recall and F1-score as well. In terms of accuracy, Trigram, Logistic, Bernoulli NB, and Ensembled classifier models are on par with Perceptron, but the latter is noticeably less so. There is a very low value for all models with a trigram that follow the same recall and F1-Score values. The Multinomial-NB model outperforms all others in terms of Recall and F1-score. Discovering the outcomes A total of eight models using unigram, bigram, and trigram were included in the investigation. Unigram ensemble models outperform other models. Using an ensemble model and unigram feature extraction, Fig. 8 shows the distribution of people's perceptions regarding the lockdown during COVID-19 in India. It has been noted that 5.0% of the population feels very positively about the lockdown, 12.0% feel indifferent, and 6.5% feel negatively about it. All signs point to a lockdown being necessary. When it comes to the lockdown, the majority of people are on the fence.For many reasons, including financial concerns, rumors, etc., there are relatively few individuals who are opposed to the lockdown.

CONCLUSION

The most efficient and user-friendly method of reaching people all around the globe and getting their opinions on COVID-19 is via social media. In order to understand how Indians feel about shutdown, we have compiled the tweets. Emotion analysis utilizing several machine learning models has been completed. With an accuracy of 72%, the Ensemble model and Unigram were determined to be the best. Using this

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

combination, we were able to forecast the public's mood in tweets during the lockdown and discovered that the majority of people (76.5% to be exact) remain neutral towards the lockdown, indicating that people are still on the fence about it. Only 5% of the population strongly supports the lockdown, and 12% are in favor of it; these individuals are certain that the only way to stop the spread of the coronavirus is to implement the lockdown. For various reasons (such as money problems, food shortages, rumors, etc.), about 6.5% of the population is opposed to the lockdown. Only a small number of individuals have voiced their strong opposition to the lockdown. In order to assist local administration manage rumors and make appropriate choices during future pandemics, it may be beneficial to investigate using other deep learning models to improve the performance on the Lockdown data.

REFERENCES

[1] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. Lehmann, “An ‘infodemic’: Leveraging high-volume Twitter data to understand public sentiment for the COVID-19 outbreak,” medRxiv, Jan. 2020, doi: 10.1101/2020.04.03.20052936.

2. A. D. Dubey, "Decoding the Twitter sentiments towards the leadership in the times of COVID-19: A case of USA and India," SSRN Electron. J., Apr. 2009, doi: 10.2139/ssrn.3588623.

[3] "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets" (IEEE Access, 2020, vol. 8, pp. 181074-181090), by A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra.

"Philippine Twitter Sentiments during the COVID-19 Pandemic using Multinomial Naïve Bayes," the International Journal of Advanced Trends in Computer Science and Engineering, volume 9, issue 1, pages 408-412, 2020.

[5] "SemEval-2016 task 4: Sentiment analysis in Twitter," in Proceedings of the 8th International Workshop on Semantic Evaluation, 2014, pp. 1–18, by R. Sara, R. Alan, N. Preslav, and S. Veselin.

[6] "SemEval-2016 task 4: Sentiment analysis in Twitter," in Proceedings of the 10th International Workshop on Semantical Evaluation, June 2016, pp. 1-18, by P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov.

As a result of the severe community quarantine in the Philippines brought on by the COVID-19 pandemic, C. K. Pastor conducted a sentiment analysis in March 2020 for the Asian Journal of Multidisciplinary Studies, volume 3, issue 1, pages 1–6.

In their 2020 article titled "Covid-19 and the 5g conspiracy theory: social network analysis of twitter data," W. Ahmed, J. Vidal-Alaball, J. Downing, and F. L. Segu'ı report on their findings in the Journal of Medical Internet Research, volume 22, issue 5, page e19458.

Paper presented at the 2014 Eighth International AAAI Conference on Weblogs and Social Media by C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," (9).

[10] In "Analyzingbrexit's impact using sentiment analysis and topic modeling on Twitter discussion," published in 2020 in The 21st Annual International Conference on Digital Government Research, pp. 1-6K, S. H. W. Ilyas, Z. T. Soomro, A. Anwar, H. Shahzad, and U. Yaqub are cited. "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," Kurdistan Journal of Applied Research (KJAR), pp. 54-63, 2020, by H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin.

[11] "Rethinking Social Interaction: Empirical Model Development," journal article published in 2020 by J. Bjornestad, C. Moltu, M. Veseth, and T. Tjora, vol. 22, no. 4, 2020.

[12] Critical Impact of Social Networks Infodemic in Defeating Coronavirus COVID-19 Pandemic: A Twitter-Based Study and Research Directions, by A. Mourad, A. Srour, H. Harmanani, C. Jenainatiy, and M. Arafeh, published in 2020 in Computer Science.