

Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter

Prasoon Gupta^{ID}, Sanjay Kumar, R. R. Suman, and Vinay Kumar^{ID}

Abstract—With the rapid increase in the use of the Internet, sentiment analysis has become one of the most popular fields of natural language processing (NLP). Using sentiment analysis, the implied emotion in the text can be mined effectively for different occasions. People are using social media to receive and communicate different types of information on a massive scale during COVID-19 outburst. Mining such content to evaluate people's sentiments can play a critical role in making decisions to keep the situation under control. The objective of this study is to mine the sentiments of Indian citizens regarding the nationwide lockdown enforced by the Indian government to reduce the rate of spreading of Coronavirus. In this work, the sentiment analysis of tweets posted by Indian citizens has been performed using NLP and machine learning classifiers. From April 5, 2020 to April 17, 2020, a total of 12741 tweets having the keywords "Indialockdown" are extracted. Data have been extracted from Twitter using Tweepy API, annotated using TextBlob and VADER lexicons, and preprocessed using the natural language tool kit provided by the Python. Eight different classifiers have been used to classify the data. The experiment achieved the highest accuracy of 84.4% with LinearSVC classifier and unigrams. This study concludes that the majority of Indian citizens are supporting the decision of the lockdown implemented by the Indian government during corona outburst.

Index Terms—Coronavirus, COVID-19, lockdown, machine learning, natural language processing (NLP), opinion mining, sentiment analysis, social media.

I. INTRODUCTION

INITIATED in Wuhan, China, the exponential spread of Coronavirus disease (COVID-19) has caused public health crisis regionally and internationally as well [1]. The Centers for Disease Control and Prevention (CDC) actuated its Emergency Operations Center (EOC), and the World Health Organization (WHO) published its first report regarding the situation about Coronavirus disease 2019 (COVID-19) on January 20, 2020 [2]. The WHO recognized and gave the name "2019-nCoV" to novel Coronavirus. The seriousness of the COVID-19 pandemic had been underrated until the National Health Commission (NHC) categorized it as a B-type infectious disease officially and took measures to fight against this pandemic on January 20, 2020 [3]. WHO later announced it as a health emergency internationally on January 30, 2020.

Manuscript received June 12, 2020; revised August 28, 2020 and October 16, 2020; accepted December 1, 2020. Date of publication December 21, 2020; date of current version August 2, 2021. (Corresponding author: Vinay Kumar.)

The authors are with the National Institute of Technology Jamshedpur, Jamshedpur 831014, India (e-mail: 2018pgscs15@nitjsr.ac.in; 2017rscs001@nitjsr.ac.in; rrsuman.cse@nitjsr.ac.in; vkumar.cse@nitjsr.ac.in).

Digital Object Identifier 10.1109/TCSS.2020.3042446

One and a half months later, on March 11, 2020 COVID-19 is classified as pandemic [4]. Coronavirus disease (COVID-19) is an infective pandemic caused by a newly discovered virus named Corona. Most of the formidable diseases are generated from unhygienic habits. Hygiene measures and sanitation, such as hand washing, could play an important and cost-effective role in reducing the spread of pandemics, such as the COVID-19 [5]. The disease causes respiratory illness (like the flu) with symptoms such as cough, fever, and in more severe cases, difficulty breathing. Mostly, the symptoms in the person infected with COVID-19 are low to medium respiratory illness and recover without any specific treatment. Aged persons and those already having some medical issues like chronic respiratory disease, diabetes, cancer, and cardiovascular disease are more vulnerable to acute infection from COVID-19. We can protect ourselves by avoiding touching our faces, washing our hands frequently, and avoiding close contact (1 m or 3 feet) with the unwell people. The most likely ways of spreading COVID-19 are the following: direct contact with an infected person; contact with the droplets of a virus carrier; or direct touch of contaminated objects or surfaces with COVID-19 virus, and then rubbing the nose or touching the mouth [6].

At this time, there are no specific vaccines or treatments for COVID-19, and this new pandemic creates fear for the world because of its estimated mortality rate of 2%–5% [7].

The most cost-effective ways to become safe are precaution and social distancing in the absence of vaccination for this infectious disease. Social distancing is a crucial way to limit the spreading of the COVID-19 pandemic, where different physical distancing restrictions are applied to fight against COVID-19 [6]. People are advised to stay at their homes and to ensure social distancing almost all the countries enforced lockdown to ensure social distancing at public places. In the history of human civilization, these times were never experienced where the whole world has been residing under lockdown [8]. Lockdown is an exigency protocol that forbids people from moving freely in public places. A total lockdown means people must stay wherever they are and not even go outside their building. We can understand lockdown as the curfew with some relaxation to essential services. All non-essential services are closed for the full lockdown period. More than 200 countries, areas, and territories are rapidly impacted by the COVID-19 pandemic [9].

Being the most affected country in the month of March 20, Italy enforced a nationwide lockdown on March 9, 2020.

After that, almost all the affected countries imposed lockdown to prevent the close contact of people and hence the spreading of COVID-19. Looking at the numbers of COVID-19 infected, recovered, and death cases of China, the USA, Britain, Italy, and other countries, the Indian government was well aware that extreme steps were required in India to reduce the exponentially increasing numbers [10]. The government of India initiated this step with a one-day lockdown on March 22, 2020, and then a 21-day complete lockdown from March 25, 2020 to April 14, 2020. Observing the positive results from lockdown 1.0, India extends it till May 3, 2020, as the second phase. To reduce the spread of the Coronavirus, almost every country around the world is facing enormous challenges in the implementation of preventive steps [11]. Though lockdown is the only way to reduce the rate of spreading of COVID-19 until the production of its vaccination, it created some severe issues in every nation like downfall in the economy and thus in GDP. Due to the closure of industries, daily wage workers are struggling to make both ends meet. This causes an increase in frustration and, hence, in depression due to unemployment, freeness, isolation, and unavailability of other medical facilities, etc. Generally, every person is supporting this decision, but due to these issues, different people have different opinions about imposing the nationwide lockdown. The exponential spread of Coronavirus has generated a tremendous need for contriving rapid analytical approaches for interpreting the information flow and the evolution of mass opinion in different pandemic scenarios [12]. This creates the scope to analyze the sentiments of the common public on this decision, which is also the motivation of this work.

Sentiment analysis is the process of diagnosing the emotions behind the words. It can generally be depicted as the assigning of opinion categories and scores, according to keyword and phrase, which are matched with sentiment score lexicons and customized dictionaries [12]. It is a subsection of text mining that includes the processing of natural language to automate the process of extracting and classifying the emotions in the written text. This process helps in specifying the writer's attitude toward any entity, topic, product, etc. [13]. The sentiment analysis of the common public regarding lockdown is important because it affects the success of this step. Due to any reason, if the people are not happy with this decision, it is almost impossible for the administration to achieve the objective of the lockdown. India is one of the largest markets for digital consumers, with 565 million Internet users [14]. So, the best possible way to get genuine texts written by common people is social media. Twitter is the best site to carry sentiment analysis on written texts due to limitations in the length of posts. Even social media platforms are being used by organizations, individuals, and governments to communicate with each other on various important occasions and health emergencies like COVID-19. Thus, analyzing such content could be very crucial for the administration in deciding the policies appropriately, and it could also help health care organizations in assessing the needs of their stakeholders [15]. During this outburst of COVID-19, people are frequently using Twitter to share their opinions and to acquire needful information on the steps taken by the administration [16].

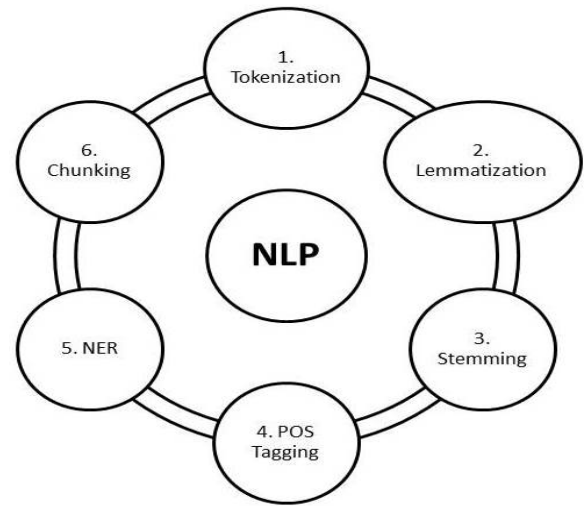


Fig. 1. Steps of NLP.

Also, there is limited knowledge of the general public's sentiment about the main topics which are in discussion over time [2].

The sentiment categorization can be implemented by using various approaches. We can majorly classify these approaches in the following three types: 1) lexicon-based approach; 2) machine learning/deep learning approach, and 3) hybrid approach. We have used a machine learning approach to classify the opinion of India's common public in this work. To use machine learning for sentiment analysis, data pre-processing on raw data is a prerequisite since the efficiency of the algorithm used is directly proportional to the quality of training and testing dataset. In sentiment analysis, the preprocessing of text is known as natural language processing (NLP). It includes the following six steps [17] depicted in Fig. 1: tokenization; lemmatization; stemming; part of speech tagging; name entity recognition; and chunking.

The rest of this article is organized as follows. Section II presents a literature review related to sentiment analysis and COVID-19. The methodology used to execute the experiment and the experimental setup used in this work are discussed in Section III. Results and analysis have been explained in Section IV. Section V contains the most frequent challenges faced and possible limitations of this work.

The conclusion of the research with future work is mentioned in Section VI.

II. LITERATURE REVIEW

Sentiment analysis from social media data is one of the highly emerging research fields. It could play a critical role in case of medical emergencies like the COVID-19 pandemic, and hence it is more crucial. Though a lot of research from various angles on sentiment classification and NLP is still in progress, some of the completed works are as follows.

Wu *et al.* [1] used data from December 31, 2019, to January 28, 2020, on the count of infectious persons exported from Wuhan to deduct the number of cases in Wuhan from December 1, 2019, to January 25, 2020. Cases exported intra-country were then predicted. They predicted the COVID

cases across the country by using the flight booking data and COVID positive persons travelled through flight and thus prognosticated the national and international spread of COVID-19 after calculating the impact of the metropolitan-wide quarantine of Wuhan and surrounding cities which was started in China from Jan 23–24.

Medford *et al.* [2] fetched tweets related to COVID-19 and calculated the frequency of keywords subject to infection prevention practices, inoculation, and racial partiality. They executed a sentiment analysis to observe sentimental valence and dominant emotions. They performed topic modeling to extract and explore hot discussion topics over time. They extracted 126049 tweets posted by 53196 different users. The frequency of COVID-19-related tweets suddenly increased from January 21, 2020 onward. The sentiments are classified from approximately half (49.5%) of all posts expressed fear to about 30% expressed surprise. The number of racial posts closely matched the frequency of new cases of COVID-19 positives. The financial and political effects of the COVID-19 were the most common topics in discussion.

Li *et al.* [3] collected and analyzed the Weibo posts from 17865 active Weibo users using online ecological recognition (OER) based on some machine learning prognostic models. They evaluated word frequency, sentiment indicator scores (e.g., depression, anxiety, indignation, and happiness), and cognitive indicators (e.g., social risk assessment, and life gratification) from the extracted posts. To evaluate the differences in the same group they have executed the opinion mining and the paired sample t-test before and after the confirmation of COVID-19 on January 20, 2020. The results evaluated by them reflect that the sensitivity to social risks and the negative sentiments increased while the scores of positive sentiments and life gratification have been decreased.

Pandey *et al.* [5] addressed the gap between the information and risk of misinformation by developing a life-long learning model that provides genuine information in Hindi, the most commonly used local language in India. They matched the sources of authentic and genuine info, such as the news provided by WHO by using machine learning and NLP. They observed a Cohen's Kappa of 0.54 with the best performing combination and was deployed in their application.

Kayes *et al.* [6] collected 100000 tweets with the keyword #coronavirus within Australia. Among these 1 lakh tweets, 3076 contain the keyword "social distancing" and #socialdistancing. They use 8000 tweets for training & validation and 2000 tweets for testing the model. They achieved an accuracy of 83.70% and an F1-Score of 81.62% on the test data. They applied the trained model on the 3076 tweets that contain the keyword "social distancing." They observed that 80% plus tweets discussed "social distancing" have a positive opinion, as illustrated. They concluded that people in Australia supported social distancing as well as accepted it.

Pastor *et al.* [7] performed a study to get the opinion of the students regarding the online mode of delivery of instructions because of extreme community quarantine during the COVID-19 pandemic. They performed this study by taking students' opinions in the College of Business and Public Administration (CBPA) of Pangasinan State University, Lingayen Campus. Firstly, they invited all the students to

answer some questions regarding the issues they may face during the online study. They found that most of the students feel that they might face some issues, and many of them were worried about Internet connectivity in the area. They concluded that maximum students are not prepared for online delivery of instructions, and thus they suggested that an alternative way of instructional delivery should be provided by the institutes so that educational excellence could be maintained.

Dubey *et al.* [8] performed a study to compare the opinions and emotions present in Indian and US national tweets while they mentioned Narendra Modi and Donald Trump, respectively. The tweets extracted for the opinion mining were posted from April 1 to April 9, 2020. NRC Emotion Lexicon has been used to analyze emotions and sentiments in these tweets. They concluded that 64.53% of tweets mentioning Narendra Modi are containing positive sentiments, while for Donald Trump, it was 48.71%.

Chen *et al.* [9] have demonstrated a study of the topic regarding the mention of controversial and non-controversial words related to COVID-19 on Twitter during the pandemic. They used LDA to collect topics from the controversial and non-controversial tweets extracted from Twitter and then compared them through both sets of tweets qualitatively. They found that topics in the controversial tweets are mostly associated with China, even after removing the keywords connected to the "Chinese virus" before the study, whereas discussions present in the non-controversial tweets are about facing and fighting with COVID-19 in the USA.

Barkur *et al.* [10] dealt with Indian citizens' sentiments after the Indian government announced the lockdown. For analysis, they used the social media platform Twitter. They examined the tweets to extract the sentiments of the Indians regarding lockdown. They extracted the tweets using the two frequently used hashtags: #IndiaLockdown and #Indiafightscoronavirus from March 25, 2020 to March 28, 2020. They examined 24000 tweets for the study using software R and generated a word cloud that evaluates the sentiments of the tweets. They found that there were sadness, fear, negativity, and disgust about the lockdown, still the positive sentiments were present prominently in the tweets. They concluded that Indians were determined that they had to reduce the spreading rate of COVID-19 and were devoted to it.

Alhajji *et al.* [11] analyzed a total of 53127 tweets from the Saudi citizens regarding COVID-19 and found that the numbers of positive tweets are greater than negative tweets for almost all the measures. They found that the most positive sentiments were present in the religious practices-related steps. They concluded that Saudi Twitter users have positive sentiments and support toward the infection control steps in fighting with COVID-19, and this supportive attitude of Saudi citizens results in the overarching confidence of the Saudi government. According to them, at times of pandemic, religious beliefs may also play an important role in preparing believers. They collected separate tweets on various steps taken by the Saudi government. They extracted 9924 tweets after the announcement of the Grand Mosque closure and found that 76.72% of the tweets were positive. Similarly, they collected tweets for Qatif lockdown, closure of schools and universities, shopping malls, parks, and restaurant closure

measures, sports competition suspension hashtags, for the congregational and weekly Friday prayers suspension measure, and finally for nationwide curfew measure.

Samuel *et al.* [12] demonstrated some knowledge about the progression of fear sentiments over time as COVID-19 approached to apex in the USA, using powerful textual analytics aided by essential textual data visualizations. They have analyzed problems regarding public sentiments ruminating deep concerns about virus and COVID-19, steering to the identification of an increase in fear and negative sentiment. They also presented the use of exploratory and extended textual analytics and textual data visualization approaches to find initial insights. Finally, they provided a comparative study of textual categorization mechanisms used in AI applications and manifested their importance for tweets of different lengths.

Abd-Alrazaq *et al.* [15] included approximately 2.8 million tweets for their study. Out of them, 167 073 tweets from 160 829 different users met the inclusion criteria. They analyzed the tweets on 12 topics and then grouped them into four main themes: sources of the virus; its origin; its effect on population, countries, and the economy; and methods of extenuating the danger of infection. They found the positive mean sentiment for all except two topics: one is causalities caused by COVID-19, and second is increased racism. They observed the minimum mean of 2722 tweets for increased racism and a maximum of 13 413 for economic losses. They also found the highest mean for likes of 15.4 for economic losses and the lowest for travel bans and warnings of 3.94.

Burnap *et al.* [16] constructed models to estimate the information flow size and survival using data retrieved from the famous microblogging site Twitter by following the terrorist event in Woolwich, London in 2013. They explained data flow as the propagation over time of info posted to Twitter via retweets. They used zero-truncated negative binomial and Cox proportional hazards regression methods to calculate the estimated value of social, content, and temporal elements of the tweet.

Naiknaware *et al.* [17] used the sentiment analysis score method to predict the popularity of different schemes offered by India's government. They used the below seven steps process to find the results: 1) extracting relevant tweets using Twitter API; 2) tweets preprocessing; 3) storing the processed tweets in CSV File format; 4) apply score. The sentiment () method; 5) generate sentence score; 6) analyzing the sentiment polarity of each tweet; and 7) prepare results. Based on the sentence score found in the fifth step, they divide the sentence's polarity into positive, negative, and neutral and, thus, stated their prediction.

Wu *et al.* [18] developed a novel decision support system using sentiment analysis, support vector machine, and generalized autoregressive conditional heteroscedasticity (GARCH) modeling. This model has been widely used to forecast time series containing features of autocorrelation and heteroscedasticity. They apply GARCH modeling and employ the results into the SVM model to accommodate complicated nonlinear and asymmetric relations engrafted in fickle forecasting. First, they manually label polarities for the postings of the data set. Then they use sentiment analysis with a manually annotated

data set to extract features from the text written on the stock forum and thus to predict the polarity of other posts automatically. After that, they integrate the postings of each stock daily. Then they used the resultant GARCH-SVM model to predict randomness in future stock prices. They also compare this model accuracy, which was 81.82% with the lexicon approach, which generates an accuracy of 75.58%.

Ding *et al.* [19] constructed an entity-level opinion analysis tool named SentiSW for issue comments consisting of sentiment classification and entity recognition. The objective of the developed tool is to classify issued comments into three polarities: positive, negative, and neutral, and to recognize the entity of comments. They build data set manually by annotating 3000 comments selected from 231 732 issued comments taken from different GitHub projects. They used a tenfold cross-validation technique to evaluate the SentiSW tool and received 63.98% average recall, 68.71% average precision, and 77.19% accuracy. They manually label 660 comments by "Person" and "Project" entity to evaluate entity recognition and achieved an overall 88.73% recall, 76.58% precision, and 75.15% accuracy.

Pota *et al.* [20] applied the sentiment analysis on political tweets using a neural-network-based approach. They used the SemEval-2017 data set, which contains 20 633 tweets for training and 12 284 tweets for testing the trained model. Their method represents the text by dense vectors, including information of subwords to observe similarities between words by using morphology and semantics both. Then, they used the convolutional neural network technique and trained the model based on the labeled data set. They applied that model on a collection of tweets gathered during the tenure of some days before U.K. General Elections. According to observed results, they concluded that the CNN approach is better as compared to lexicon-based approaches for classification of sentences into positive and negative polarities.

Tomar *et al.* [21] performed opinion mining on GST using data from Twitter to find out the public opinion about GST. They construct the data set by manually annotating 1000 tweets plus 2000 reviews from the Internet Movie Database (IMDB). They train and test the classifier by applying the data set in three different ways. Model-1 used the IMDB data set for training and testing with the K-fold validation process using $K = 10$, Model-2 used IMDB data set for training and manually annotated data set for testing. Lastly, Model-3 used a full data set for training and testing with the K-fold validation process using $K = 10$. They achieved an accuracy of 74.75%, 67.1%, and 72.3% for Model-1, Model-2, and Model-3, respectively. Observing the results, they stated that if a training data set is constructed using different sources instead of single-source, better accuracy can be achieved.

Das *et al.* [22] constructed the data set with the gathered 20 000 tweets using Twitter API and NodeXL software about GST during its implementation in India. The data set contains 2006 positive words, 4782 negative words, and words to identify misspelled words that frequently appear in social media. They also proffered some idea about how the presence of any sentimental word in a sentence can change its polarity. After implementing, they concluded that the Naïve Bayes

classifier is the most popular because it is relatively easier to implement, yet it is relatively complicated and beats many other complicated algorithms in performance.

Li *et al.* [23] answered the questions like how to figure out and classify the situation based information on social media and what are the different predictability of highlights of the spreading scale of various types of situation based info using COVID-19-related conversations on Sina Weibo, the widely used microblogging site in China (like Twitter). They categorized the data into seven situations: caution & advice, notifications, donations, emotional support, help-seeking, criticizing, and counter rumors. They concluded that the chosen features for various kinds of situational data could also assist the administration in organizing their COVID-19-related info to increase or reduce the reposting of their posts.

Luo [24] used a susceptible-infected-recovered (SIR) model to predict the life cycle of COVID-19 in many countries and the world. To predict the COVID-19 life cycle, he took daily updated data about COVID-19 from “Our World in Data” website and retrogress the mathematical SIR model using publically available codes from the Milan Batista website. He ran regression for countries individually and modified it daily with the newer data. The model then formed is used to predict the life cycle of a full pandemic and then construct the life cycle curve. He fitted the data to plot an initial segment of the curve, and the remaining segment is estimated. He predicted 97%, 99%, and completely ending dates of individual countries and the whole world. They concluded that this pandemic might end by the end of November 2020 from the world.

Rao *et al.* [25] proposed an approach to conduct the screening of people efficiently. According to him, gather the travel history with some more general manifestations using a mobile-based online survey. The data thus collected can help in the primary screening and early recognition of COVID-19-positive individuals. Data points can be gathered and refined through the artificial intelligence (AI) model, which can finally evaluate persons who may be positive and classify them into four classes: no risk, minimal risk, moderate risk, and high risk of being septic with the Coronavirus. The recognition of the high-risk cases can then be quarantined on priority, thus reducing the probability of spread.

Dutta *et al.* [26] performed a study to determine whether machine learning could be used to evaluate how much predictions about confirmed, negative, released, and death cases are close to real values. They used deep learning neural networks, long short-term memory (LSTM), and gated recurrent unit (GRU) for training the data set. Prediction results are then cross-checked by real data. They concluded that the combined LSTM-GRU model gave comparatively better results in predicting confirmed, negative, released, and death cases.

From the above discussion, it is clear that sentiment analysis is one of the prominent decision-making methods. In such a pandemic situation, sentiment analysis may play a critical role in taking steps toward controlling and managing the pandemic by the Indian government. Very few studies have been reported on sentiment analysis of Indian people over

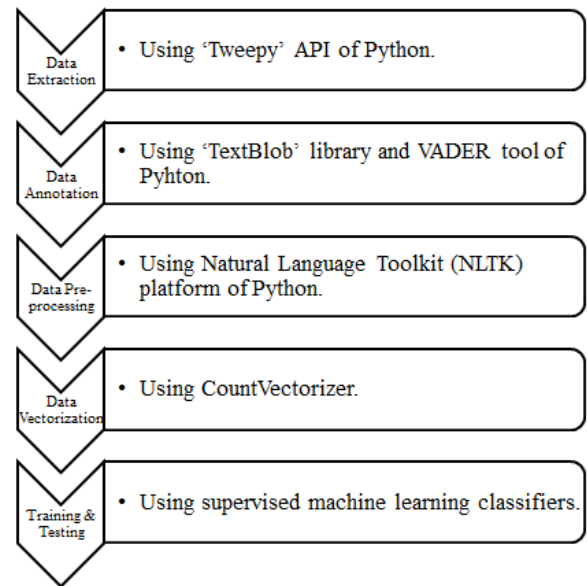


Fig. 2. Methodology used.

COVID-19. Therefore, the main objective of this research is to get the ratio of people in favor of lockdown in India and people against this. Such steps taken by the government will only be successful if the people are supporting it. So, we suggested a methodology to get the views of common people quickly, or we can say in real-time, which can assist the government in decision-making.

The key objectives of this article are as follows.

- 1) To mine the sentiments of Indian citizens regarding the nationwide lockdown enforced by the government to minimize the rate of spreading of Coronavirus.
- 2) The sentiment analysis of tweets posted by Indian citizens has been performed using NLP and machine learning classifiers.
- 3) To classify the data accurately, eight different classifiers are used.
- 4) To improve accuracy, a consolidated result of TextBlob and VADER libraries is used.

III. PROPOSED APPROACH

In this section, we have proposed a framework for sentiment analysis during the COVID-19 pandemic. The framework is shown in Fig. 2. The various phases of the framework to perform the sentiment analysis of lockdown during corona outbursts are as follows.

A. Data Extraction

Sentiment analysis is performed on the lockdown imposed by the Indian government from March 25, 2020 to April 14, 2020, during COVID-19. Thus, the objective specific data are not available, and hence we have prepared the data set manually. In light of the deteriorating situation, discussions of the pandemic on social media have drastically increased since March 2020 [9]. We have extracted 12 741 tweets having the keyword “Indialockdown” from April 5, 2020 to April 17, 2020 using Tweepy.

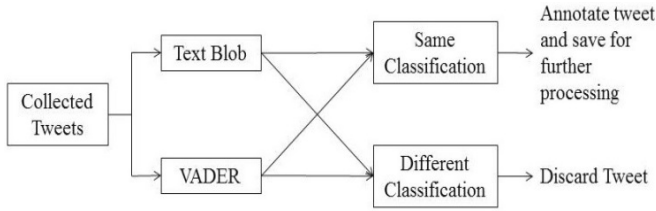


Fig. 3. Data annotation process.

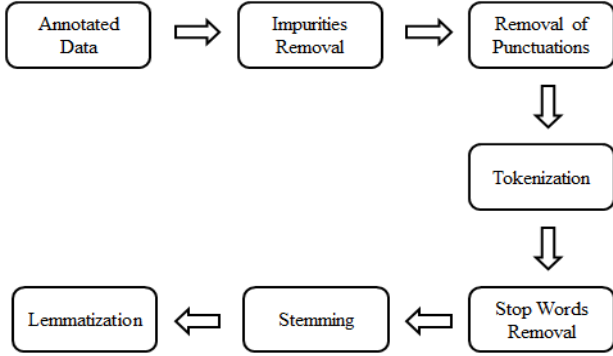


Fig. 4. Data preprocessing.

B. Data Labeling

After tweets collection, we have used the following approach shown in Fig. 3 to label the tweets as positive, neutral, and negative. We have generated each tweet's polarity using the TextBlob library and VADER (Valence Aware Dictionary for sEntiment Reasoning) tool of the Python. Next, we have taken the intersection of TextBlob and VADER results to consolidate the polarities. After this step, we are left with 7284 tweets having 3545 with positive polarity, 2097 with neutral polarity, and 1642 with negative polarity.

C. Data Preprocessing

The data we have collected may hold some unsought and sentiment fewer words like links, Twitter-specific words such as hashtags (starts with #) and tags (starts with @), single letter words, numbers, etc. These types of words can play the role of noise in our classifier training and testing. To amend classifier efficiency, it is necessary to remove noise from the labeled data set before feeding the classifier.

Our pre-processing module separates noise from the labeled data set [21]. The steps of pre-processing are shown in Fig. 4. In this step, we implemented a module to remove the above-specified impurities, converted the data set into a data frame, and then executed removal of string punctuations, tokenization, and removal of English stop words, stemming, and lemmatization.

D. Vectorization

The machine learning classifiers cannot take the input written in any language except numbers. Thus, before using the text data for predictive modeling, it is required to convert it into features. We have used the CountVectorizer feature extractor to calculate word frequencies. CountVectorizer counts the frequency of each word present in the document and creates a sparse matrix, as shown in Table I. For example, Doc1:

TABLE I
SAMPLE MATRIX BY COUNTVECTORIZER

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

“She was young the way an actual young person is young.” CountVectorizer will convert this text into the following sparse matrix with an index of the words in alphabetical order as follows: {“she”: 4, “was”: 6, “young”: 8, “the”: 5, “way”: 7, “an”: 1, “actual”: 0, “person”: 3, “is”: 2}.

This matrix is not sparse because we are converting the only single document. In the case of multiple documents, it is frequent that a word present in one document can be missing from some other documents, and hence the corresponding cells are filled up with zero, and the resultant matrix will become sparse.

E. Training and Testing the Classifiers

After feature extraction of the preprocessed data set, we have passed the data to machine learning classifiers. We have used eight classifiers (Multinomial Naive-Bayes, Bernoulli NaiveBayes, LogisticRegression, LinearSVC, AdaBoostClassifier, RidgeClassifier, PassiveAggressiveClassifier, and Perceptron) for this purpose. We have used 80% data for training and 20% data for testing the classifiers. We have extracted the performance of the classifiers mentioned above using 1-g, 2-g, and 3-g.

F. Experimental Setup

All the experiments are carried out using core i3, 2.4 GHz, 3 MB L3 cache processor with 4 GB of RAM in Python 3.0 platform. Anaconda IDE is used to code the experiment in Python language. We have used “Tweepy” API of Twitter to extract tweets; natural language toolkit (NLTK) library of Python for data pre-processing. Matplotlib, pandas, numpy, and sklearn libraries are also used in this work.

IV. RESULTS AND ANALYSIS

In this section, the result analysis of all the classifiers based on accuracy, precision, recall, F1-Score, and receiver operating characteristics (ROC) curves with different grams has been discussed. Along with that, we have validated our data set for each model with unigrams, bigrams, and trigrams using k-fold cross-validation technique with $k = 10$. Accuracy can be calculated using (1) and (2) which is the one out of many metric units for evaluating classifiers. It is defined as the number of correct predictions over the total number of predictions.

Accuracy

$$= \text{number of correct predictions} / \text{total number of predictions.} \quad (1)$$

Technically, we can understand accuracy in terms of positives and negatives as

$$\text{Accuracy} = (\text{True Positives} + \text{True Negatives}) / \text{total number of predictions} \quad (2)$$

Predicted Class		Actual Class	
		Positive	Negative
	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Fig. 5. Pictorial representation of positives and negatives.

TABLE II
ACCURACIES OF CLASSIFIERS WITH UNIGRAM

Classifier	Accuracy (%)	Execution Time (in sec)	Accuracy Rank
MultinomialNB()	67.5	5.46	8
BernoulliNB()	73.7	5.02	6
LogisticRegression()	82.2	5.56	3
LinearSVC()	84.4	7.34	1
AdaBoostClassifier()	68.6	13.15	7
RidgeClassifier()	78.3	19.88	5
PassiveAggressiveClassifier()	84.0	4.11	2
Perceptron()	82.0	4.13	4

TABLE III
ACCURACIES OF CLASSIFIERS WITH BIGRAMS

Classifier	Accuracy (%)	Execution Time (in sec)	Accuracy Rank
MultinomialNB()	62.2	11.00	7
BernoulliNB()	55.2	8.5	8
LogisticRegression()	80.2	13.46	4
LinearSVC()	80.4	24.19	3
AdaBoostClassifier()	68.6	34.5	6
RidgeClassifier()	78.4	46.42	5
PassiveAggressiveClassifier()	80.4	8.87	2
Perceptron()	82.0	9.36	1

where the total number of predictions is the sum of true positives, true negatives, false positives, and false negatives. True positives are the number of correct predictions of the positive class. Similarly, the number of correct predictions of the negative class is known as true negatives, the number of incorrect predictions of the positive class is known as false positives, and the number of incorrect predictions of the negative class is known as false negatives. Fig. 5 depicts the pictorial representation of positives and negatives.

With unigram, we have observed the highest accuracy of 84.4% from the LinearSVC classifier; the perceptron classifier gives the best accuracy of 82% with bigrams and 81.2% with trigrams. We have used the execution time as the tie-breaker parameter. The accuracy and execution time of all the classifiers are shown in Table II (unigrams), Table III (bigrams), and Table IV (trigrams).

Fig. 6 summarizes the accuracy comparison of all the classifiers with unigram, bigrams, and trigrams.

K-fold cross-validation is the best technique to test the effectivity of the machine learning model. The re-sampling approach of the k-fold cross-validation technique is much useful in measuring the efficiency of any machine learning

TABLE IV
ACCURACIES OF CLASSIFIERS WITH TRIGRAMS

Classifier	Accuracy (%)	Execution Time (in sec)	Accuracy Rank
MultinomialNB()	60.6	13.21	7
BernoulliNB()	52.0	13.74	8
LogisticRegression()	78.4	21.73	2
LinearSVC()	78.3	41.33	3
AdaBoostClassifier()	68.6	55.26	6
RidgeClassifier()	76.7	82.67	5
PassiveAggressiveClassifier()	77.8	13.94	4
Perceptron()	81.2	14.08	1

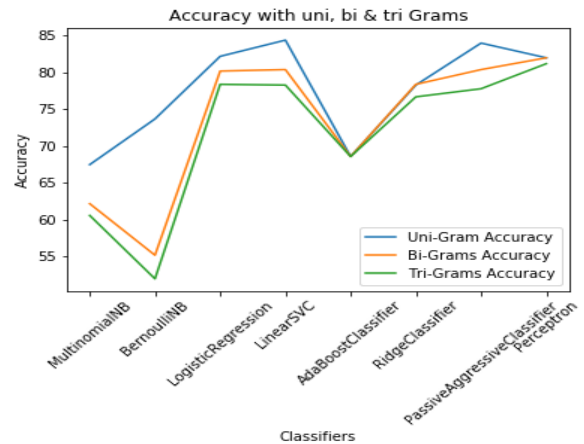


Fig. 6. Comparison of accuracies with uni-, bi-, and trigrams.

model with a limited amount of input data. In this technique, the data set is divided into k equal parts or folds and then any one fold is used as a testing set and rest $k-1$ folds are used for training the model and the cross-validation score for this particular permutation is recorded. This process gets repeated for k times having a new fold as a testing set and rest as a training set. Then the mean of the scores of all the permutations is calculated which is the final cross-validation score of a model. We have calculated the cross-validation score of each model with unigrams, bigrams, and trigrams. We have found the highest cross-validation score of 0.85 with unigram, 0.82 with bigrams, and 0.8 with trigrams using LinearSVC classifier. The comparison between cross-validation scores of all the classifiers with unigram, bigrams, and trigrams is shown in Fig. 7. Further cross-validation scores for all the classifiers are summarized in Table V (unigrams), Table VI (bigrams), and Table VII (trigrams).

Precision, recall, and F1-Score are other metrics to evaluate the models. The proportion of positive identifications that actually belong to the positive class is known as precision, which is calculated using (3). Recall can be calculated using (4) which shows the number of positive predictions that are identified correctly out of all positive examples. F1-Score given by (5) is the weighted harmonic mean of precision and recall

Precision

$$= \text{True Positives} / (\text{True Positives} + \text{False Positives}) \quad (3)$$

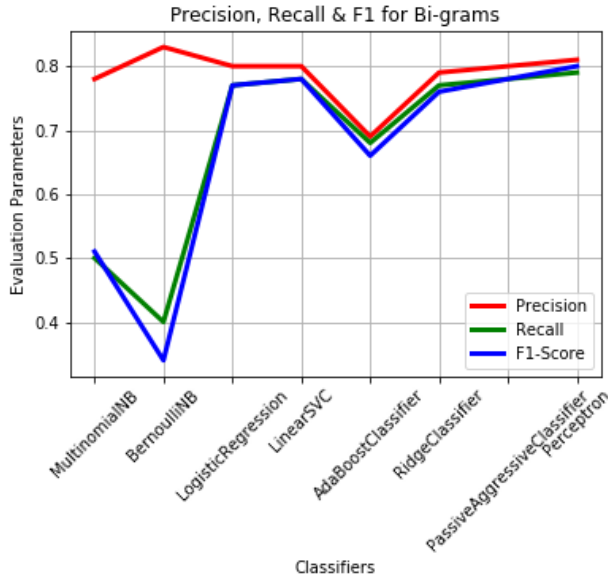


Fig. 7. Tenfold cross-validation score comparison between unigram, bigrams, and trigrams.

TABLE V
TENFOLD MEAN CROSS-VALIDATION SCORE OF CLASSIFIERS WITH UNIGRAMS

Classifier	Mean Cross-Validation Score
MultinomialNB()	0. 672235
BernoulliNB()	0. 740828
LogisticRegression()	0. 838668
LinearSVC()	0. 849275
AdaBoostClassifier()	0. 515899
RidgeClassifier()	0. 793021
PassiveAggressiveClassifier()	0. 845918
Perceptron()	0. 791419

TABLE VI
TENFOLD MEAN CROSS-VALIDATION SCORE OF CLASSIFIERS WITH BIGRAMS

Classifier	Mean Cross-Validation Score
MultinomialNB()	0. 613528
BernoulliNB()	0. 523814
LogisticRegression()	0. 812219
LinearSVC()	0. 821920
AdaBoostClassifier()	0. 529850
RidgeClassifier()	0. 797411
PassiveAggressiveClassifier()	0. 813449
Perceptron()	0. 772373

Recall

$$= \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \quad (4)$$

F1 – Score

$$= 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}). \quad (5)$$

We have compared the precision, recall, and thus F1-Score of all the classifiers with unigrams, bigrams, and trigrams.

TABLE VII
TENFOLD MEAN CROSS-VALIDATION SCORE OF CLASSIFIERS WITH TRIGRAMS

Classifier	Mean Cross-Validation Score
MultinomialNB()	0. 596960
BernoulliNB()	0. 509350
LogisticRegression()	0. 795119
LinearSVC()	0. 797058
AdaBoostClassifier()	0. 530027
RidgeClassifier()	0. 774320
PassiveAggressiveClassifier()	0. 791597
Perceptron()	0. 768503

TABLE VIII
N-GRAMS PERFORMANCE WITH EVALUATION METRICS

Grams & Metrics	Precision	Recall	F1-Score
Unigram	83.5% (PassiveAggressiveClassifier)	82.4% (LinearSVC)	82.2 (LinearSVC)
Bi-grams	82.7 (BernoulliNB)	78.9 (Perceptron)	79.8 (Perceptron)
Tri-grams	80.5 (Perceptron)	78.1 (Perceptron)	79 (Perceptron)

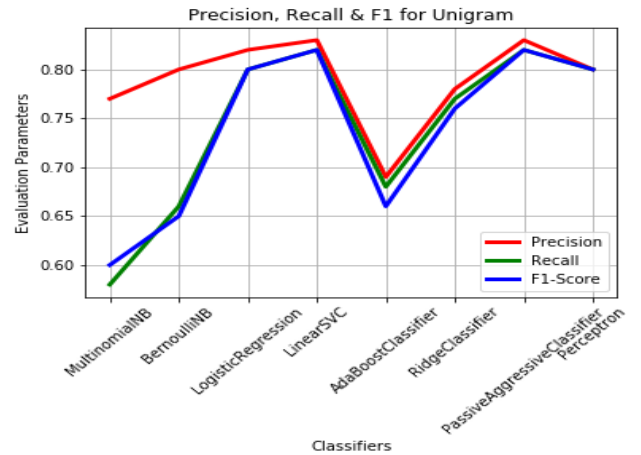


Fig. 8. Graphical representation of precision, recall, and F1-Score (Unigram).

We observe the highest precision of 83.5% with passive aggressive classifier and highest recall and F1-Score of 82.4% and 82.2% respectively with LinearSVC classifier in case of unigram. In bigrams, we got the best precision of 82.7 with BernoulliNB classifier, highest recall, and F1-Score of 78.9% and 79.8%, respectively, with perceptron classifier. Perceptron classifier attains the highest precision of 80.5%, recall of 78.1%, and F1-Score of 79% when used with trigrams. Table VIII summarizes the best results observed in precision, recall, and F1-Score with n-grams.

Below figures sum up the comparison between precision, recall, and F1-Score of all eight classifiers used for the experiment with unigram shown in Fig. 8, bigrams shown in Fig. 9, and trigrams shown in Fig. 10.

Though perceptron classifier performs better in many cases yet LinearSVC classifier achieved the highest accuracy, recall,

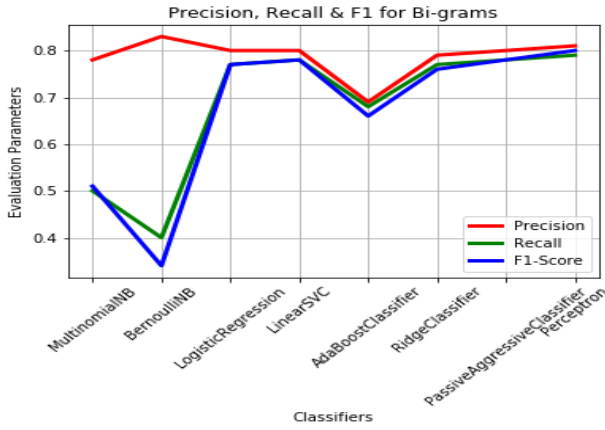


Fig. 9. Graphical representation of precision, recall, and F1-Score (Bigrams).

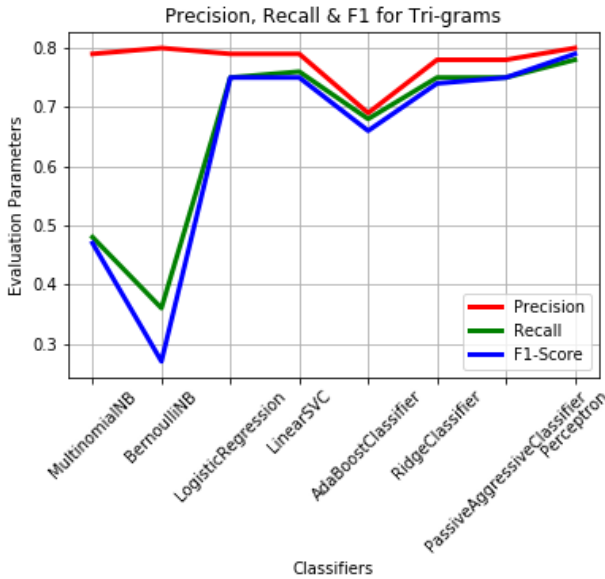


Fig. 10. Graphical representation of precision, recall, and F1-Score (Trigram).

and F1-Score with unigrams, which are also higher than the results of perceptron with bigrams and trigrams; thus, we choose LinearSVC classifier with unigrams as our prediction model.

Effective Analysis: After selecting the LinearSVC classifier for our prediction, we have evaluated our model using the confusion matrix and AUC-ROC curve. A confusion matrix is a table that is used to examine the performance of a classifier with the help of data whose true values are known. It is also used to construct the ROC curve. ROC curve is the graphical way to observe the relation between true-positive rate and false-positive rate. It shows the classifier's performance on every verge of classifications. The area under ROC (AUROC) is used to examine the performance of the aggregate measure or to quantify the classifier's performance. ROC curve is generally used to examine binary classifiers. To evaluate the multiclass models, we have to generate n curves for the n -classification model (one for each class versus rest all classes). We got the confusion matrix after applying our data on LinearSVC with unigram, which is depicted in Table IX.

Since this experiment includes three classes, we have observed three ROC curves for negative class, neutral,

TABLE IX
CONFUSION MATRIX OF LINEARSVC WITH UNIGRAM

Classes	-1	0	1
-1	211	55	58
0	9	359	13
1	34	52	624

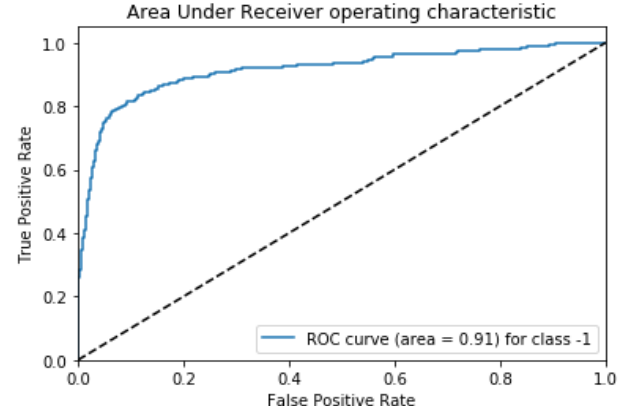


Fig. 11. AUROC curve for negative class.

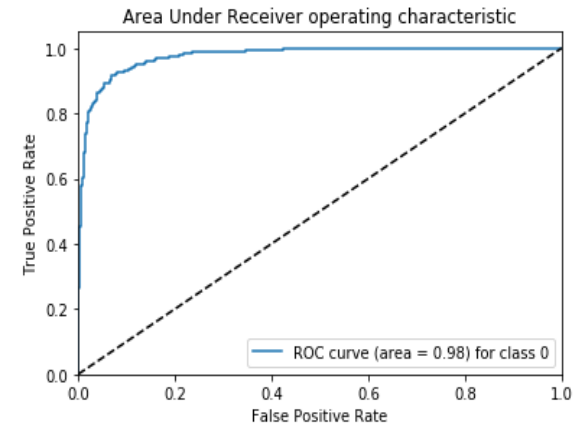


Fig. 12. AUROC curve for neutral class.

and positive ones in Figs. 11–13, respectively, with respectable AUCs. We achieved areas of 0.91, 0.98, and 0.95 for each classification in our experiment.

A. Results

After observing the accuracies, precision, recall, F1-Score, and k -fold cross-validation scores of eight different classifiers with unigram, bigrams, and trigrams, we have got the best results with LinearSVC and unigram. After that, we have verified the same by observing the confusion matrix and AUROC curves for our model. So, we have predicted a total of 7071 tweets using this combination and we have found that 48.69% of people are talking positive about the lockdown, 29.81% are neutral, and 21.5% of the people are feeling negative due to some reason as depicted in Fig. 14.

The results are according to our expectations as maximum people are supporting it because lockdown could be the best precautionary measure in this critical pandemic. Also, people need to share information because any critical situation comes

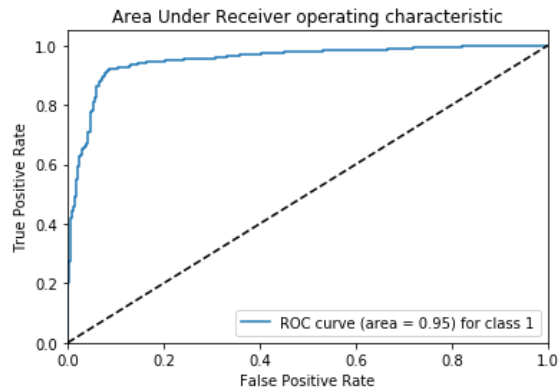


Fig. 13. AUROC curve for positive class.

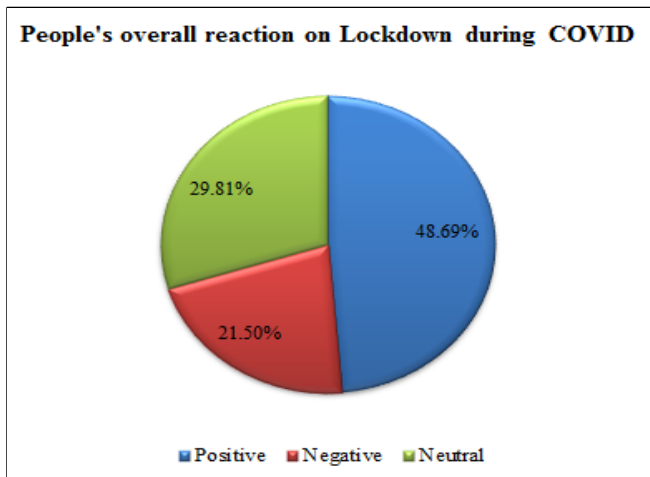


Fig. 14. People reaction on lockdown in India.

with lots of rumors, which is the main reason behind a specific amount of neutral tweets, and we have already mentioned various reasons for negative sentiments; thus, some of the sentiments are negative also.

V. LIMITATIONS AND CHALLENGES

The most common challenge about the sentiment analysis of the written text is that we cannot neglect the importance of the processing of natural languages. The accuracy and performance of the experiment are directly proportional to the granularity of the data set, which is constructed after NLP in the case of sentiment analysis. We have to tackle many irregularities, diversity, and subjectivity in the data while dealing with the natural language. The major limitations of this work are that we have taken the tweets of a specific phase during the lockdown, but with the change of phases, the surrounding conditions might get changed; thus, the sentiments of the public could also be changed. We have not considered the sentiments in emoticons and hashtags of the tweets, as we believe it could hamper classifiers' efficiency. Users can post negative sentiment with positive hashtags and vice-versa, and they can also use wrong emoticons in sarcasm. In this work, we have used two lexicons to annotate our data set, using more could make the data set more granular.

VI. CONCLUSION

Social media is witnessing a massive increase in the number of users per day. People prefer to share their honest opinions on social media instead of sharing with someone in person. Using the posts from Twitter, we examined the common public's aggregate reaction toward the implementation of lockdown by the Indian government during the spread of COVID-19. Motivated by the mixed reactions coming after the announcement of lockdown in India, we collected tweets during phase 2 of lockdown in India. We have applied collected data to eight supervised machine learning techniques with different grams of text after annotation and preprocessing. We have observed the best performance with the LinearSVC classifier and unigram. The combination gives us an accuracy of 84.4%, which is best in all the combinations which we have executed on our data set. We have consolidated the performance by calculating precision, recall, F1-Score, and tenfold cross-validation for all the combinations, and we got the best results with LinearSVC and unigram. So, we executed the sentiment analysis of tweets by the public during lockdown using this combination and found that almost half of the population (48.69%) is talking positive about the lockdown, 29.81% are neutral, and 21.5% of the people are feeling negative due to some reason. We have further evaluated our model by observing the confusion matrix and AUROC curves. Overall, it can be observed that Indians have positively taken the fight against the pandemic, and most of the population supports and agrees with the government for deciding to enforce the lockdown to reduce the spread rate. The positive response has been noticed regarding lockdown and implies that India has prevented the exponential spread of Coronavirus up to a great extent [10]. In health emergencies like COVID-19, these types of works could assist the policymakers and health care departments as well.

Sentiment analysis of natural languages itself contained a vast scope to work on, and due to health emergencies, this work is also manifested with a wide range of future scopes. Future studies can consider the tweets before the start of the first lockdown and after the end of the last and can demonstrate the changes in sentiments of the people in both cases and their consequences. The factors which can affect mental stability during pandemics can also be studied, and the study of the impact of fake news on the public can also play an important role in assisting the administration and policymakers in controlling the situation [10]. From the technical point of view, future studies can look to improve the accuracy of the model and can experiment on a large corpus.

REFERENCES

- [1] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modeling study," *Obstetrical Gynecology Surv.*, vol. 75, no. 7, pp. 399–400, Jul. 2020.
- [2] 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](https://doi.org/10.1101/2020.04.03.20052936).
- [3] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, "The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users," *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, p. 2032, Mar. 2020.

- [4] WHO Statement Regarding Cluster of Pneumonia Cases, WHO, Wuhan, China, 2020.
- [5] R. Pandey *et al.*, "A machine learning application for raising WASH awareness in the times of COVID-19 pandemic," 2020, *arXiv:2003.07074*. [Online]. Available: <http://arxiv.org/abs/2003.07074>
- [6] A. S. M. Kayes, M. S. Islam, P. A. Watters, A. Ng, and H. Kayesh, "Automated measurement of attitudes towards social distancing using social media: A COVID-19 case study," Tech. Rep., Oct. 2020.
- [7] C. K. Pastor, "Sentiment analysis on synchronous online delivery of instruction due to extreme community quarantine in the Philippines caused by Covid-19 pandemic," *Asian J. Multidisciplinary Stud.*, vol. 3, no. 1, pp. 1–6, Mar. 2020.
- [8] 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](https://doi.org/10.2139/ssrn.3588623).
- [9] L. Chen, H. Lyu, T. Yang, Y. Wang, and J. Luo, "In the eyes of the beholder: Analyzing social media use of neutral and controversial terms for COVID-19," 2020, *arXiv:2004.10225*. [Online]. Available: <http://arxiv.org/abs/2004.10225>
- [10] G. Barkur, Vibha, and G. B. Kamath, "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India," *Asian J. Psychiatry*, vol. 51, Jun. 2020, Art. no. 102089.
- [11] M. Alhajji, K. A. Al, M. Aljbran, and M. Alkhalifah, "Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19," Dept. Social Behav. Sci., College Public Health, Temple Univ., Philadelphia, PA, USA, Tech. Rep., doi: [10.20944/preprints202004.0031.v1](https://doi.org/10.20944/preprints202004.0031.v1).
- [12] J. Samuel, A. GG, M. Rahman, E. Esawi, and Y. Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification. Nawaz and Rahman, Md. Mokhesur and Esawi, Ek and Samuel, Yana," *Information*, vol. 11, no. 6, pp. 1–22, Apr. 2020, doi: [10.3390/info11060314](https://doi.org/10.3390/info11060314).
- [13] R. Liu, Y. Shi, C. Jia, and M. Jia, "A survey of sentiment analysis based on transfer learning," *IEEE Access*, vol. 7, pp. 85401–85412, 2019.
- [14] N. Kaka *et al.*, "Digital India: Technology to transform a connected nation," McKinsey Global Inst., India, Tech. Rep., Mar. 2019. [Online]. Available: <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20India%20Technology%20to%20transform%20a%20connected%20nation/MGI-Digital-India-Report-April-2019.pdf>
- [15] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah, "Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study," *J. Med. Internet Res.*, vol. 22, no. 4, Apr. 2020, Art. no. e19016.
- [16] P. Burnap *et al.*, "Tweeting the terror: Modelling the social media reaction to the woolwich terrorist attack," *Social Netw. Anal. Mining*, vol. 4, no. 1, p. 206, Dec. 2014.
- [17] B. R. Naiknaware and S. S. Kawathekar, "Prediction of 2019 Indian election using sentiment analysis," in *Proc. 2nd Int. Conf.*, Aug. 2018, pp. 660–665.
- [18] D. D. Wu, L. Zheng, and D. L. Olson, "A decision support approach for online stock forum sentiment analysis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 8, pp. 1077–1087, Aug. 2014.
- [19] J. Ding, H. Sun, X. Wang, and X. Liu, "Entity-level sentiment analysis of issue comments," in *Proc. 3rd Int. Workshop Emotion Awareness Softw. Eng.*, Jun. 2018, pp. 7–13.
- [20] M. Pota, M. Esposito, M. A. Palomino, and G. L. Masala, "A subword-based deep learning approach for sentiment analysis of political tweets," in *Proc. 32nd Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, May 2018, pp. 651–656.
- [21] N. Tomar, R. Srivastava, and J. K. Verma, "Analysing public sentiment on GST implementation in India," in *Proc. Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Sep. 2018, pp. 1101–1106.
- [22] S. Das and A. K. Kolya Sense GST, "Text mining & sentiment analysis of GST tweets by naive Bayes algorithm," in *Proc. 3rd Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN)*, Nov. 2017, pp. 239–244.
- [23] L. Li *et al.*, "Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 2, pp. 556–562, Mar. 2020.
- [24] J. Luo. (2020). *When Will COVID-19 End Data-Driven Prediction*. [Online]. Available: <https://ddi.sutd.edu.sg>
- [25] R. AS and V. JA, "Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine," *Infection Control Hospital Epidemiol.*, vol. 1, pp. 1–5, Jan. 2020.
- [26] B. SK and S. Dutta, "Machine learning approach for confirmation of COVID-19 cases: Positive, negative, death and release," *medRxiv*, vol. 2, pp. 172–177, Jan. 2020, doi: [10.5281/zenodo.3822623](https://doi.org/10.5281/zenodo.3822623).



Prasoon Gupta received the B.Tech. degree in computer science and engineering from Abdul Kalam Technical University, India. He is currently pursuing the M.Tech. degree with the Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jamshedpur, India.

His research interests include sentiment analysis, data analysis, and machine learning.



Sanjay Kumar received the M.Tech. degree in computer science and engineering from the Birla Institute of Technology, Ranchi, India. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jamshedpur, India.

His research interests are information security, digital watermarking, WSN, and mathematical modeling.



R. R. Suman received the M.Tech. degree in computer science and engineering from IIT Kharagpur, Kharagpur, India.

He is currently an Associate Professor with the Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jamshedpur, India. His research interests are dependability analysis, software engineering, image encryption and data analysis.



Vinay Kumar received the Ph.D. degree in safety analysis of computer-based systems from IIT (BHU), Varanasi, India.

He is currently an Assistant Professor with the Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jamshedpur, India. His research interests are reliability, safety, and mathematical modeling.