Automated Classification of Societal Sentiments on Twitter With Machine Learning

Piyush Vyas, Martin Reisslein, *Fellow, IEEE*, Bhaskar Prasad Rimal, *Senior Member, IEEE*, Gitika Vyas, Ganga Prasad Basyal, *Graduate Student Member, IEEE*, and Prathamesh Muzumdar

***Abstract*—The rapid growth in information sharing on social media has defined a new information age in our society. Microblogging sites, such as Twitter, gained immense popularity during the COVID-19 pandemic. We developed an automated framework to extract the positive, negative, and neutral sentiments from tweets, and to further classify the tweets through machine-learning (ML) techniques. The developed framework can help to understand the sentiments in our society during profound events, such as the COVID-19 pandemic. Our framework is novel in that it is a hybrid framework that combines a lexicon-based technique for tweet sentiment analysis and labeling with supervised ML techniques for tweet classification. We have evaluated the hybrid framework with a range of measures, such as precision, accuracy, recall, and F1 score. Our results indicate that the majority of the sentiments are positive (38.5%) or neutral (34.7%). Furthermore, with an accuracy of 83%, the long short-term memory (LSTM) neural network has been selected as the preferred ML technique in the framework. The evaluation results indicate that our hybrid framework has the potential to automatically classify large tweet volumes, such as the tweets on COVID-19, according to the sentiments in the society.**

***Index Terms*—COVID-19, Coronavirus tweets, hybrid framework, sentiment analysis, text classification, tweet classification, Twitter.**

I. Introduction

A. Motivation

THE WIDESPREAD COVID-19 pandemic has caused panic and despair among people as reflected in their social media updates [1]. The ubiquitous outbreak inflated social media updates in the form of tweets, messages, and posts [2]. Importantly, user-generated data on social media can be a prominent source of information at the time of crises [3]. People adopted widely used social media and microblogging channels [4], [5], such as Facebook and Twitter, to share their ideas, thoughts, and reactions [6]. Among all social networking platforms, Twitter is the third-largest online social networking platform [7]. The analysis of tweets about COVID-19 is highly informative because the user tweets reflect the mindsets and sentiments of our society during the pandemic [8].

The worldwide spread of the Coronavirus caused a range of thoughts and emotions. The COVID-19 pandemic has by its very nature, rapidly caused confusion and fear within society [9]. People in different nations responded with distinct reactions on social networking websites [6]. The variation in sentiments during pandemic times caused mental disturbances in the form of fear, anxiety, and many other dreadful symptoms [10]; the COVID-19 pandemic has contributed to exposing the vulnerabilities of urban dwellers [11] and poses a significant public health issue [12]. Phrases, such as "updates about confirmed cases," "COVID-19-related death," "early signs of the outbreak," "economic impact," and "preventive measures" in tweets indicate sentiments of anxiety and terror on microblogging sites [13]. Furthermore, public views on COVID-19-related news on microblogging sites have the potential to proliferate different sentiments [14].

The availability of enormous social media data opens the door for analysis of people's sentiments [3]. The analysis of such a large amount of information is very tedious due to the unstructured and noisy form of the data [3]. Therefore, it is important to develop automated methodologies for analyzing and classifying tweets that represent the sentiments in society. Machine-learning (ML) techniques can be used to automate the sentiment analysis. The study [6] relied on single deep-learning (DL)-based techniques for the classification of tweets; whereas we incorporate various ML techniques in our framework. Our study thus enhances the understanding of which ML techniques perform well and which do not for tweet classification. Furthermore, prior work, such as [1]–[3], focused purely on the sentiment analysis task; whereas we examine a wider scope of the sentiment analysis chain by automating the classification of COVID-19 tweets through a hybrid framework that encompasses lexicon-based tweet sentiment analysis and labeling in conjunction with ML techniques for tweet classification.

The popular supervised ML algorithms require external assistance for model building. These techniques need an extensive labeled dataset for the training to predict the desired outcomes [15]. Lexicon-based techniques work upon the semantics of the words in text data. Lexicon-based techniques can be used for calculating polarity, e.g., positive,

Manuscript received February 4, 2021; revised June 14, 2021; accepted August 24, 2021. Date of publication August 30, 2021; date of current version June 21, 2022. This work was supported by the State of South Dakota Board of Regents' Competitive Research Grant. *(Corresponding author: Martin Reisslein.)*

Piyush Vyas and Ganga Prasad Basyal are with the College of Business and Information System, Dakota State University, Madison, SD 57042 USA.

Martin Reisslein is with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: reisslein@asu.edu).

Bhaskar Prasad Rimal is with the Beacom College of Computer and Cyber Sciences, Dakota State University, Madison, SD 57042 USA (e-mail: bhaskar.rimal@dsu.edu).

Gitika Vyas was with the Department of Renewable Energy, Rajasthan Technical University, Kota 324009, India.

Prathamesh Muzumdar is with the Muma College of Business, University of South Florida, Tampa, FL 33613 USA.

Digital Object Identifier 10.1109/TTS.2021.3108963

negative, and neutral sentiments in the text. We incorporated both lexicon-based and supervised learning in our hybrid framework for the automatic classification of COVID-19 tweet sentiments.

We employ the valence-aware dictionary and sentiment reasoner (VADER) lexicon-based technique to extract the sentiments that are utilized to label the tweets. These labeled 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 (LSTM), to predict the sentiments for novel unlabeled test datasets. To achieve our objective of automatic classification of people's sentiments, our novel hybrid approach combines a natural language processing (NLP) lexicon-based technique with a supervised ML technique. We have also used a DL-based LSTM neural network to enhance the scope of our work.

B. Related Work

Recently, several studies, such as [10], [14], and [16]–[18], have focused on applying sentiment analysis on microblogging sites. Twitter has emerged as the hub of people's sentimental datasets [17], [19]. According to [20], identifying informative posts on social media has a vital role in a crisis and in learning what is happening during a pandemic. Twitter has an oceanic volume of information to exhume for knowledge extraction, especially during critical events, such as the proliferation of COVID-19. For example, the studies [21]–[25] examined social media data to infer public attitudes toward the use of masks, electronic cigarettes, tourism, and vaccination during the COVID-19 pandemic. These sentiment analysis studies can be treated as exemplary use cases of Twitter data analysis.

Sentiments can be broadly characterized as positive, negative, and neutral. Studies [17], [26] have shown that core sentiment classes are apt for short textual data, such as user tweets. The analysis of user emotions is another subcategory that comprises classes of distinct people's emotions, such as fear, anxiety, joy, happiness, and disgust. A few studies, such as [27]–[29], have analyzed the emotional expressions on social media. The term "infodemic" has been coined based on the proliferation of pandemic-related information on the Web, hence infodemiological analysis could be used to identify the perceptual biases in a society [30]. Moreover, topic modeling is a mechanism used by [31]–[34] to extract the diversified pandemic-related knowledge on Twitter; whereby, [31] utilized the demographics and tweets from Brazil and USA, while [32] studied COVID-19 associated topics for the North America region.

Furthermore, the study [35] extracted information regarding the pandemic emergency, COVID-19 control, various symptoms, and the spread of COVID-19. More specifically, the study [10] analyzed tweets by around 183 000 tweeters over five months and their results reveal that COVID-19 altered the sentimental behavior from being positive toward the negative. The study has shown that the variation in sentiments during the pandemic has caused mental disturbances in the form of fear, anxiety, and many other dreadful symptoms. The studies [14] and [16] used the data of the microblogging website named Weibo and collected the news of COVID-19 from the public views and further extracted the sentiments. The news collection segregates the main news from a wide set of news data and then applies K-means clustering on the headlines, followed by the observation of the closest news to the headlines to be the key news of the day. English texts are analyzed well by the ML methods rather than other methods; thus, ML methods are preferred for quality output [26].

Studies, such as [5] and [36], manually labeled tweets which can introduce human biases into the tweet labels. In contrast, our proposed hybrid framework employs an automatic tweet labeling approach that enables unbiased tweet labeling. We note that the studies [10], [14] were particularly focused on the sentiment analysis of specific demographics, namely, Australia and China, respectively. Similarly, studies [25], [37], [38] performed sentiment analysis for the United Arab Emirates, India, and the United Kingdom, respectively. In contrast, we utilized a dataset that contains COVID-19-related tweets from the entire world. Further, studies, such as [10], [37], [39], and [40] performed the sentiment analysis to examine community dynamics of different events, government policies, preparedness, management, and distinct programs; whereas we concentrate specifically on the COVID-19 pandemic.

The study [41] considered a binary sentiment extraction and classification by ML techniques; whereas we have performed a multiclass sentiment analysis and classification with ML and DL techniques. Similar to our approach, the study [10] used VADER to extract positive and negative sentiments; however, we extracted neutral sentiments as well because we believe that a large group of people would prefer to tweet neutrally. The study [12] concentrated on analyzing tweets composed of lockdown reopening information; whereas our dataset comprises a wide range of COVID-19-related information that includes the public outrage, sentiments, COVID vaccine, precautionary measures, government policies, and many more (see Fig. 3). Some recent studies, such as [28], [42], and [43], have used CristalFeel, Textblob, and recurrent neural network techniques as a tool to analyze the extracted core emotions from the tweets, whereas we have employed the dictionary (lexicon)-based sentiment reasoning in VADER to analyze the varied sentiments on Twitter.

DL and ML techniques have been utilized by [44]–[46] to analyze the sentiments of the COVID-19 tweets; whereas in addition to these approaches, we have utilized the amalgamation of the lexicon-based VADER technique and ML techniques to extract and classify the sentiments. Following the general principles from [19] and [20], we have considered user-generated content on Twitter for our sentiment analysis. Also, da Silva *et al.* [19] and Ghafarian and Yazdi [20] highlighted the importance and usability of supervised ML techniques which we have utilized in our framework. To the best of our knowledge, this study is the first to develop and evaluate a hybrid framework that automatically analyzes and labels the sentiments of COVID-19 tweets with a lexicon-based technique, such as VADER, and further classifies the tweets with state-of-the-art ML techniques.

C. Contribution

This article contributes to the existing knowledge base related to COVID-19 tweet classification by providing a novel hybrid classification pipeline that enhances the understanding of how to classify people's sentiments that are proliferated on microblogging sites. Our key contributions are as follows.

1) From the architectural perspective, we have developed a novel hybrid framework for sentiment analysis and classification through ML. The introduced novel hybrid framework consists of lexicon-based automated tweet sentiment analysis and labeling in conjunction with state-of-the-art DL techniques for tweet classification. Our hybrid framework demonstrated that the automatic lexicon-based labeling process overcomes the obstacle of manual tweet annotation.
2) From the performance perspective, to validate our hybrid framework design rationales, we have employed various evaluation measures, such as accuracy, precision, recall, and F1 score. We demonstrate a methodology to evaluate the best state-of-the-art ML techniques based on the different evaluation measures. Our study of various ML techniques and their performances related to COVID-19 sentiment classification contributes to the growing understanding of useful AI applications [49].

The automated sentiment classification helps to understand how people feel about the imposed government policies, regulations, and guidelines related to COVID-19. During pandemics, there is a necessity to act fast to control the pandemic through government policies and at the same time to mitigate any negative societal sentiments, which could contribute to noncompliance and undermine the pandemic control efforts. Caregivers, healthcare providers, and government officials can monitor people's sentiments expressed through the enormous volume of textual content available on social media by utilizing automated hybrid sentiment classification systems. For example, in May 2020, #antivax was trending on Twitter and turned into a worldwide protest. These provocative tweets directly correlate with the people's sentiments and lead to enormous volumes of text expressions of sentiments on social media.

Incremental efforts to scale up and to improve VADER as the sole sentiment analysis tool would likely not suffice to process these enormous text volumes. Therefore, in order to fundamentally advance the scalability of the text processing for sentiment analysis, we propose a hybrid (amalgamation) of VADER with ML techniques. In our hybrid design, VADER and ML complement each other in that VADER provides the data labels and the labeled data form the basis for the training of the ML techniques. The proposed hybrid system can help inform the preparation of policies and guidelines as well as publicity campaigns to inform people in a timely manner with the scientific facts.

II. Automated Hybrid Tweet Classification

To overcome the conventional requirement of a labeled sentiment dataset, we employ an unsupervised technique for sentiment analysis. In particular, we employ a lexicon-based approach that associates sentiments, such as positive, negative, and neutral, with words (all possible worlds of the English dictionary). Lexicon-based techniques are simple, yet powerful for labeling the words according to a polarity score, i.e., score for positive, negative, and neutral sentiments. After the labeling of the dataset, we applied and compared various supervised ML techniques to determine the best-supervised ML technique that achieves high accuracy and recall. To obtain the best performance from the ML techniques, we have used fivefold cross-validation to tune the parameters [50]. We proceed to describe our proposed hybrid classification framework pipeline, which is illustrated in Fig. 1, in the following sections.

A. Data Preprocessing

Starting from the raw textual tweet data, we performed NLP to clean the text for making it more informative by eliminating redundancies and anomalies.

1) Stop Word Removal: We have removed stop words, such as "over," "under," "again," "further," "then," "once," "here," and "there" with the Python-based NLP toolkit (NLTK). Further, we eliminated special characters (*, $, !), URLs, user mentions (@), and hashtags (#) by using regular expressions (RegEx).

2) Case Folding: Uppercases and lowercases of a word can have a significant effect on the word's polarity (positive, negative, and neutral). The conversion of words in the text from uppercase to lowercase and vice versa depends upon the nature of the corpus and the purpose of the corpus utilization. Our dataset comprises tweets that have a limited number of words and people often write them in the usual communication language, i.e., mostly in lowercases. Therefore, we have performed case-folding by lowering the limited uppercase words in all tweets.

3) Tokenization: We have performed tokenization to convert each tweet sentence into tokens (words), which helps ML techniques to learn about the meaning of each word.

4) Lemmatization: In NLP, lemmatization is a process of identifying the base words from the vocabulary to aggregate all different forms of a word into one category (i.e., lemma). For example, the base word "stop" has different forms, such as "stopped," "stopping," and "stops." All these forms belong to the same base class, i.e., stop. We performed the lemmatization to make the ML efficient.

B. Feature Selection

The text classification extracts the semantic and syntactic features. The semantic feature comprises the sentiment associated with a word; whereas the syntactic feature consists of the unigrams, bigrams, and n-grams (i.e., group of associated words) [4], [48]. Unigrams refer to single words in a tweet, such as "corona," "virus," and "death." Bigrams refer to the combination of two words, such as "corona virus," "covid vaccine," and "positive cases." All Unigrams, Bigrams, Trigrams, and Quadgrams are chains of words that are categorized under n-grams, which are widely used for text classification.

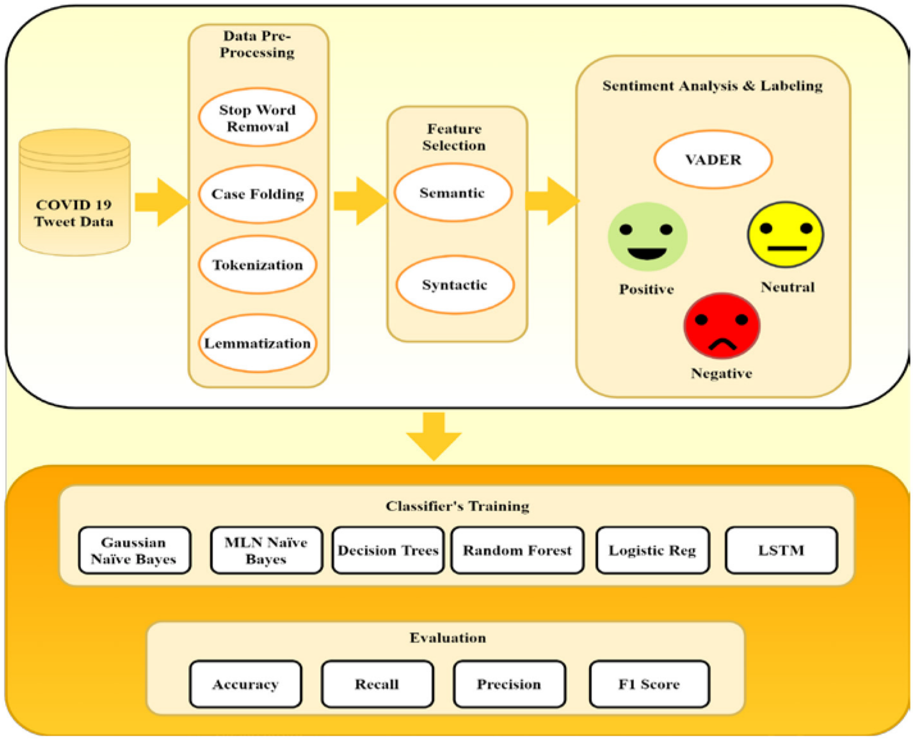


Fig. 1. Hybrid classification framework pipeline: automated lexicon-based labeling using the VADER is combined with supervised ML techniques for classifier training.

C. Sentiment Analysis

Twitter possesses small textual data and sentiment analysis is the process of extracting positive, neutral, and negative sentiments from a piece of information, such as tweets [51]. Sentiment analysis is often referred to as polarity-based analysis, whereby polarity implies different sentiments. VADER [52] is a widely adopted sentiment analysis technique. VADER is a rule-based model that has been specifically developed for the sentiment analysis of social media messages because of its ability to analyze the sensitiveness of messages [18]. Although there are other techniques available, such as linguistic inquiry and word count (LIWC) and the General Inquirer, VADER outperforms these other techniques [52]. Therefore, we used VADER for the sentiment analysis in our framework. VADER provides the polarity scores of positive, negative, and neutral classes along with a compound score, whereby the compound score can be used as a threshold value to distribute the tweets among all three sentiment classes.

D. Data Labeling

The labeling of tweets is an essential task before applying supervised ML techniques. We have labeled raw tweets based on their sentiment values. After the sentiment analysis, we have created a new dataset that consists of all tweets with their associated sentiment values (labels). The entire process of data labeling is automatic by applying the conditional statements (i.e., for loop and if-else) of Python programming. We have assigned 0 to all positive sentiment tweets, 1 to all neutral sentiment tweets, and 2 to all negative sentiment tweets.

E. Classifiers

Since we have used supervised ML and DL techniques to automatically classify the tweets based on the sentiments, these techniques are termed as a classifier. Specifically, we have used GNB, MLNB, LR, DT, RF, and LSTM as a classifier. A DT is a representation of nodes and branches, whereby nodes consist of the attributes for classification, and a branch holds the values of those attributes [15]. The DT has many versions of classifiers, we have used the modern classification and regression tree (CART) version. The RF is an amalgamation of various DTs and gives an output, that is, commonly elected by all trees. The RF technique is fast, scalable, noise-reduced, and does not overfit [53]. Although the name LR seems to be a regression technique, it is a statistical classification technique that generates the classification probability of the desired output (i.e., dependent variable) [53]. The Naïve Bayes mainly targets the text classification tasks. The architecture of Naïve Bayes depends on conditional probabilities. It creates trees based on their probability of occurring. These trees are also known as the Bayesian Network [15]. We have employed the Gaussian and multinominal Naïve Bayes techniques. The LSTM is a modified version of a recurrent neural network that consists of memory units to hold the old outcomes to learn efficiently to classify the inputs in different classes [54].

VADER as a text sentiment extraction technique is potent to deal with the polarity classification (i.e., positive, negative, and neutral) of tweet contents. According to [52], VADER achieved a superior 0.96 accuracy for extracting the sentiments from microblogging sites. Hence, the reliance on a dictionary

TABLE I
HYPERPARAMETER DESCRIPTION OF MODELS

Model	Type	Tuned Parameters	Description
Decision Tree	Classification	Criterion = "entropy", max_depth = 8	"Criterion" is a function that takes the value to calculate either gini index or information gain for entropies. "max_depth" is an integer type of variable that takes the value to stop the expansion of the tree.
Gaussian Naïve Bayes	Classification	var_smoothing = 0.4	"var_smoothing" is a float parameter that generates the value to stabilize the variance between all data features to gain maximum model performance.
Multinominal Naïve Bayes	Classification	alpha = 0.7	"alpha" is a float parameter for selecting additive (Laplace/Lidstone) smoothing for improving performance.
Logistic Regression	Classification	C = 0.01	C is a positive float regularization parameter, the smaller value of C, the stronger the stopping of the model from overfitting.
Random Forest	Classification	n_estimators = 100, Criterion = "gini", max_depth = 8	"Criterion" and "max_depth" are similar to Decision tree, whereas "n_estimators" is an integer type of parameter for selecting number of trees.
LSTM	Classification	Batch_size = 64, Dropout_rate = 0.5, Activation function "ReLu"	"Batch_size" is an integer parameter that defines the number of samples used in one iteration. "Dropout rate" is a float parameter between 0 and 1, used to restrain the neural network from overfitting. "Activation" is a parameter for selecting an activation function, such as rectified linear unit (ReLu), Sigmoid (logistic), or hyperbolic tangent activation function (tanh), for activating the neurons of the network by initializing the learning rate.

of lexicons and their associated sentiments makes VADER sufficiently accurate for extracting the sentiments from the tweets. Importantly, VADER can extract the sentiments based on the contextual meaning of a phrase, for example, "diminishing the COVID cases" would be labeled as positive sentiment.

The positive, negative, and neutral sentiments are the core classes for their subcategories (i.e., highly positive, positive, neutral, negative, and highly negative). Since the utilized data consists of short text snippets in the form of tweets, the limited-class sentiment extraction, i.e., two or three main classes, is appropriate, as demonstrated by prior studies, such as [17], [19], [26], and [55]. By utilizing a 0.05 threshold value of the compound score, we distributed the tweets among the positive, negative, and neutral classes by conditional computation. VADER computes the compound score by summing the valence scores of each word in the lexicon, adjusted according to VADER-internal compound score policies, and then normalized to be between -1 (most extreme negative) and $+1$ (most extreme positive). If the compound score is greater than or equal to the 0.05 threshold value, then the tweet belongs to the positive class; if the compound score is less than or equal to -0.05 then the tweet belongs to the negative class; and, if the compound score is between +0.05 and -0.05, then the tweet belongs to the neutral class.

III. PERFORMANCE EVALUATION

A. Dataset and Computational Infrastructure

We have utilized the publicly available tweet dataset from Kaggle.com [56] 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 collections, the data have been collected for a limited period, i.e., between July and August 2020. During this period, the 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.

B. Performance Metrics

We have evaluated the utilized ML techniques with the accuracy, precision, recall, and F1-score classification performance metrics. Formally, let TP represent the number of true positives, i.e., the number of positive samples that were correctly classified. TN denotes the number of true negatives, i.e., the number of negative samples that were correctly classified. FP is the number of false positive, i.e., the number of negative samples that were incorrectly classified as positive. FN is the number of false negatives, i.e., the number of positive samples that were incorrectly classified as negatives. Accuracy measures the overall performance. Precision measures the relevancy of the generated results. Recall measures the correctly classified tweets and the F1-Score is a weighted combination of recall and precision scores [54]. Formally,

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{1}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{2}$$

$$\text{F1 score} = (2(\text{precision} \times \text{recall}))/(\text{precision} + \text{recall}) \tag{3}$$

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN}). \tag{4}$$

C. Hyperparameter Tuning

Every ML technique has a critical component known as hyperparameters that are often responsible for profound performance changes. Table I shows the hyperparameters for the adopted ML techniques that we have implemented on our Twitter dataset.

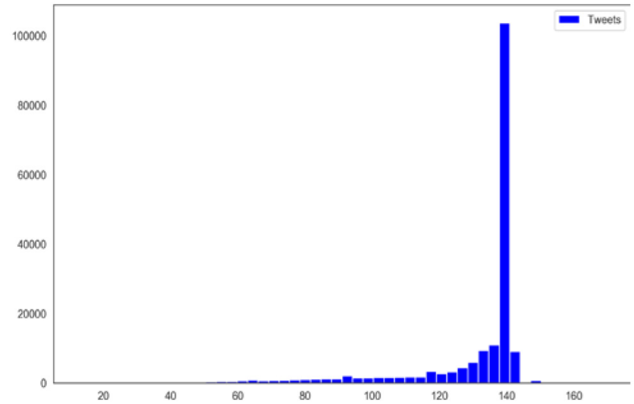


Fig. 2. Tweet length distribution: number of characters in tweets.

The hyperparameters can be adjusted through cross-validation, whereby the entire dataset or the training dataset is used to get the best performing values of the hyperparameters. In particular, the K-fold cross-validation method is often adopted, whereby K is the number of folds; depending on the dataset size, 5-fold or 10-fold cross-validation is often used. These folds indicate the division of datasets into different parts, such as 5-fold implies the division in five parts. While performing the K-fold cross-validation, all parts serve interchangeably as a training and testing set. Specifically, we consider accuracy as the performance metric in the cross-validation, and thus obtain the set of hyperparameter values that achieves the highest accuracy. We then adopt these hyperparameter values for the implementation of the corresponding ML technique.

D. Results

1) Sentiment Labeling: Fig. 2 shows the tweet length distribution, which indicates that the vast majority of the examined tweets were very close to the historic 140 character limit for tweets. This distribution shows that the COVID-19-related tweets are relatively more lengthy than the general average tweet length of around 90 characters [57]. The relatively high character count of the COVID-19-related tweets may be due to the gravity and complexity of the COVID-19 pandemic and the resulting inherent desire of people to express the sentiments in relatively rich detail in long tweets. Importantly, these long tweets add to the text volume that requires processing, intensifying the need for automated sentiment analysis with ML. At the same time, the long tweets are rich in content and may convey nuanced sentiments that are valuable for gaining insights into the societal sentiments.

The VADER sentiment analysis gave the following 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 II illustrates some example tweets from all three sentiment classes, along with their polarity scores, whereby the contextual meaning of tweets indicates the users' sentiments.

TABLE II
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's 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': 0.0}

Fig. 3 shows the word clouds of all three sentiment distributions that indicate the high-frequency words used in all tweets. Fig. 3(a) indicates that the positive sentiments are mainly related to the words "help," "mask," "need," "people," "support," and "thank" which were utilized in a context expressing a positive sentiment in the tweets. Interestingly, Fig. 3(b) indicates that these same words dominated in the neutral tweets. In contrast, Fig. 3(c) indicates that aside from the work "people," words related to death ("death," "death toll," and "died"), and words related to the former U.S. President Donald Trump ("realdonaldrump" and "trump") played an important role in the tweets expressing negative sentiments. The Donald Trump-related words may indicate that there is a political dimension to the negative sentiments, which may relate to the role that the COVID-19 pandemic played during the last year of the presidency of Donald Trump. It is also interesting to note that the word "amp," which is commonly interpreted to mean "ain't my problem" plays an important role in all three sentiment classes. Apparently, the "amp" word had been flexibly utilized to express a wide range of sentiments.

In our worldwide dataset, 24.3% (i.e., 43 673) of the tweets were generated by the USA, followed by other top COVID-19-related tweet producers, such as India, Canada, Australia, and South Africa. Since the USA generated the most tweets and led the number of global COVID-19 cases and deaths at the time, we show the USA heatmap as an example of the state-by-state (i.e., regional) tweet activity in Fig. 4. The right-hand side scale on the heatmap shows the correspondence between the total number of tweets per state and the associated heatmap color. A state with more than 4000 tweets is represented by a blue color shade and a state with close to zero tweets is represented by a red color shade. Within the USA, the states of California (most populous U.S. state) and New York (4th most populous U.S. state) dominated the tweet activity. The next most tweet-active state was Texas (2nd most populous U.S. state), followed by Florida (3rd most populous U.S. state). The next most tweet-active states were Georgia (9th most populous) and Pennsylvania (6th most populous) on the East Coast, as well as the Midwestern state of Illinois (5th most populous), while on the West Coast, the state of Washington (13th most populous) produced a large total number of tweets associated with COVID-19-related content. An interesting trend emerges from Fig. 4 in that the tweet activity (i.e., the number of tweets per state) follows generally the rank order of the

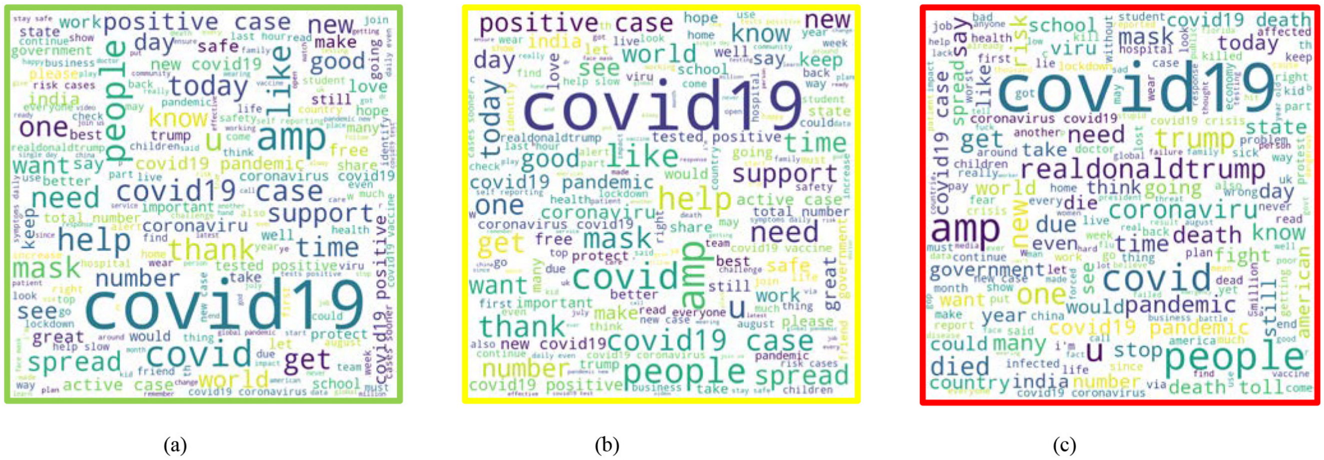


Fig. 3. Word clouds for sentiments: (a) positive sentiment word cloud, (b) neutral sentiment word cloud, and (c) negative sentiment word cloud.

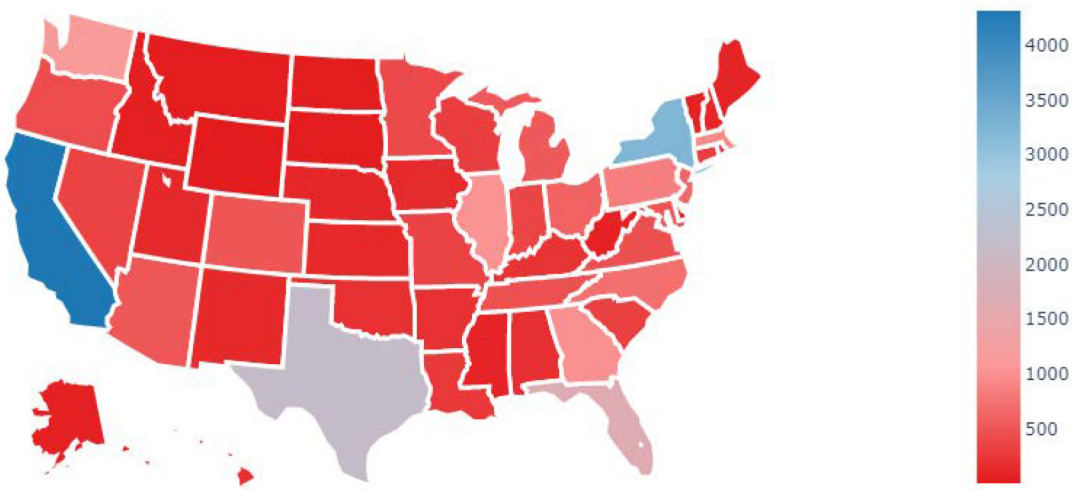


Fig. 4. USA state-wise heatmap of total tweets per state.

states in terms of population. However, there are some notable exceptions, such as the 4th most populous state New York is the second most tweet-active state, and Washington is very tweet-active for its relatively small population.

In order to further investigate the state-by-state tweet characteristics, we computed for each state an aggregate compound tweet sentiment score by summing up the compound VADER sentiment scores for all the tweets in a state. Fig. 5 shows the corresponding heatmap of the aggregate compound tweet sentiment score, which has a dark shade of red for a negative sentiment, a light shade of red for a neutral sentiment, and a dark shade of blue for a positive sentiment. We observe from Fig. 5 that at the extreme ends of the scale, Georgia and California have the most negative and most positive aggregate compound tweet sentiment score, respectively. Aside from California, the state of New York is the only other state to achieve a blue shade, indicting a relatively positive aggregate compound tweet sentiment score. The next most positive sentiments existed in Florida, Texas, Illinois, Washington, and Pennsylvania, while all the remaining states trended toward neutral or slightly negative sentiments.

Interestingly, there is an overall correspondence emerging from the aggregate compound tweet sentiment scores in Fig. 5 with the typical U.S. electoral map, e.g., for presidential elections: California and New York typically vote overwhelmingly for the Democratic Party (which is traditionally represented with the blue color) while most rural Midwestern states typically vote overwhelmingly for the Republican party (traditionally represented with the red color). Florida and Pennsylvania are traditionally "swing states" that may either vote Democratic or Republican. This emerging correspondence between aggregate compound tweet sentiment score and typical voting behavior provides further indication for the political dimension of the pandemic sentiment, which was noted above for Fig. 3(c).

Importantly, the heatmap results in Fig. 5 indicate that the sentiments in a population typically vary by region. Therefore, allocations of resources for public health education efforts and publicity campaigns should be responsive to the regionally prevailing sentiments. For instance, the results in Fig. 5 suggest to allocate resources to Georgia and other states with a relatively dark shade of red. The prevailing negative sentiments

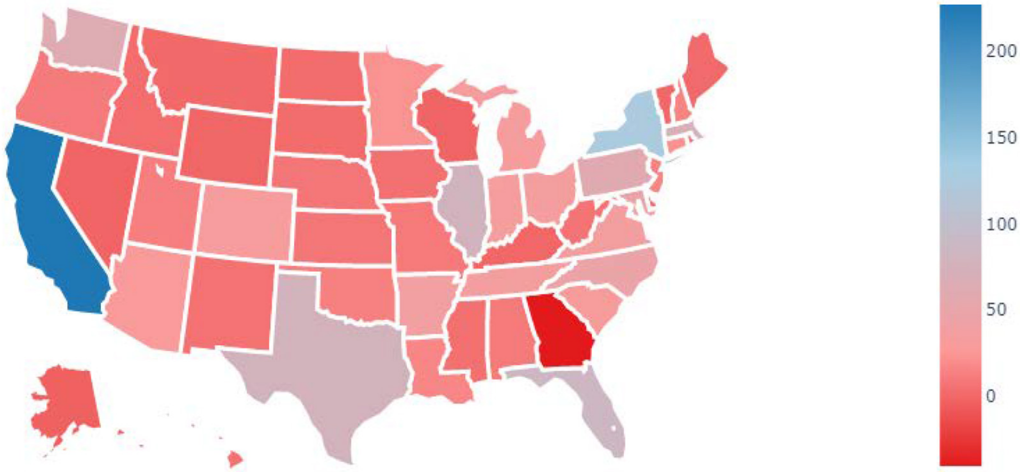


Fig. 5. U.S. state-wise heatmap of aggregate compound tweet sentiment score.

in these states might promote incompliance with public health measures and thus undermine the pandemic control efforts. High-speed automated sentiment analysis of enormous text volumes is a prerequisite for such a regionally responsive public health policy. We, therefore, evaluate next the accuracy and processing capabilities of the ML-based sentiment analysis relative to the VADER sentiment classification.

2) ML Performance: For the training of the adopted classifiers, we split the dataset into training and testing sets. We have used the train_test_split() method of the scikit learn python library (sklearn) for the split of the entire dataset into 70% as a training set and 30% as a testing set. More specifically, we labeled all data through VADER and then we split that data 70–30 into training set and testing set. We then trained the ML models on the training set (70% of the data with labels). Subsequently, we tested the ML models on the testing set (30% of the data, whereby the labels are not provided to the ML techniques, the VADER labels are then compared with the ML generated labels to evaluate the accuracy of the ML techniques). Thus, the test dataset was unknown and not a part of the model's training, i.e., we evaluated all models on unknown test data. As summarized in Table I, we have tuned all the selected classifier techniques through fivefold grid search cross-validation on the training dataset by utilizing scikit learn, a Python-based library.

The test dataset comprised approximately equal class distribution (i.e., ∼33% of positive sentiment, 33% of negative sentiment, and the rest for the neutral sentiment class). Thus, the test accuracy of models is a good measure to judge the ML performance. In general, a dataset is not always balanced, and it is not a good practice to rely on only a single evaluation measure, such as model accuracy. Therefore, keeping in mind an unbalanced dataset scenario, we have also generated other evaluation measures, namely, recall, precision, and F1 score.

With the typical contemporary computational infrastructure specified in Section III-A, VADER required 37.2 s for the analysis of the complete set of 179 108 tweets, while the ML technique required 13.3 s for the training from the training set and 1.14 s for the sentiment analysis of the testing set. Thus,

TABLE III
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%

VADER had a throughput of approximately $4.8\ 10^3$ tweets per second, whereas the trained ML classifier had a throughput of approximately $0.3 \times 179\,108$ tweets/1.14 s $= 47.1\ 10^3$ tweets per second, i.e., achieved a processing speed-up of roughly one order of magnitude.

We compare all techniques in Table III to identify the best model for the automatic classification of the societal sentiments. Table III shows the percentage values of the different evaluation measures for all examined ML techniques. We observe from Table III 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 each metric. However, only the LSTM technique achieved consistent performance in 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%.

IV. CONCLUSION AND FUTURE WORK

We developed a novel hybrid framework consisting of a lexicon approach for tweet sentiment analysis and labeling in combination with a DL approach for tweet classification for the purpose of sentiment analysis in the COVID-19 topic area. In particular, to automatically classify the societal sentiments on Twitter, we extracted the positive, negative, and neutral sentiments by utilizing the VADER lexicon technique to label the

COVID-19-related tweets based on their associated sentiments. For the classification task, we used different ML and DL techniques. With a classification test accuracy of 83%, LSTM outperformed all other techniques. The trained ML classifier achieved a processing speed-up of about one order of magnitude compared to the VADER technique. Hence, our results showed the potential for high-speed automatic classification of the societal sentiments on Twitter related to COVID-19, which may inform public health publicity campaigns.

One interesting direction for future research is to revisit the hyperparameter tuning by adding stratified sampling before cross-validation to further enhance and validate the high model accuracy level. Future research directions could consider the emotion lexicons of Canada's National Research Council (NRC) that consist of a wide variety of sentiments, such as Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, and Trust for the sentiment analysis and classification on microblogging sites. Furthermore, detection of misinformation regarding the COVID-19 pandemic is required to restrain the dissemination. Finally, pretrained transfer learning (TL) models, such as bidirectional encoder representations from transformers (BERT) and a robustly optimized BERT pretraining approach (RoBERTa) can be utilized to classify the COVID-19 tweets.

We also note that this study has focused on the diagnostic analysis of societal sentiments. Related social media studies have sought to analyze the deliberate manipulation of societal sentiments, e.g., [58]–[63]. An important future research direction is to examine the interplay between sentiment analysis and sentiment manipulation, e.g., to uncover deliberate efforts to steer societal sentiments into prescribed directions.

REFERENCES

[1] R. Khan, R. Khan, P. Shrivastava, A. Kapoor, A. Tiwari, and A. Mittal, "Social media analysis with AI: Sentiment analysis techniques for the analysis of Twitter COVID-19 data," *J. Crit. Rev.*, vol. 7, no. 9, pp. 2761–2774, 2020.

[2] N. K. Rajput, B. A. Grover, and V. K. Rathi, "Word frequency and sentiment analysis of Twitter messages during Coronavirus pandemic," Apr. 2020. [Online]. Available: arXiv:2004.03925.

[3] A. Kruspe, M. Häberle, I. Kuhn, and X. X. Zhu, "Cross-language sentiment analysis of European Twitter messages duringthe COVID-19 pandemic," Aug. 2020. [Online]. Available: http://arxiv.org/abs/2008.12172.

[4] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the OMG!," in *Proc. Int. AAAI Conf.*, 2011, pp. 538–541.

[5] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of Twitter data," in *Proc. Workshop Lang. Social Media*, 2011, pp. 30–38.

[6] A. S. Imran, S. M. Doudpota, Z. Kastrati, and R. Bhatra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning—A case study on COVID-19," *IEEE Access*, vol. 8, pp. 181074–181090, 2020, doi: 10.1109/ACCESS.2020.3027350.

[7] D. Antonakaki, P. Fragopoulou, and S. Ioannidis, "A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 114006. [Online]. Available: https://doi.org/10.1016/j.eswa.2020.114006

[8] J. Xue *et al.*, "Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach," *J. Med. Internet Res.*, vol. 22, no. 11, Nov. 2020, Art. no. e20550, doi: 10.2196/20550.

[9] R. Abbas and K. Michael, "COVID-19 contact trace app deployments: Learnings from Australia and Singapore," *IEEE Consum. Electron. Mag.*, vol. 9, no. 5, pp. 65–70, Sep. 2020, doi: 10.1109/MCE.2020.3002490.

[10] J. Zhou, S. Yang, C. Xiao, and F. Chen, "Examination of community sentiment dynamics due to COVID-19 pandemic: A case study from Australia," Jun. 2020. [Online]. Available: http://arxiv.org/abs/2006.12185.

[11] L. J. Robertson, A. Munoz, and K. Michael, "Managing technological vulnerability of urban dwellers: Analysis, trends, and solutions," *IEEE Trans. Technol. Soc.*, vol. 1, no. 1, pp. 48–59, Mar. 2020, doi: 10.1109/tts.2020.2975806.

[12] C. Burr, J. Morley, M. Taddeo, and L. Floridi, "Digital psychiatry: Risks and opportunities for public health and wellbeing," *IEEE Trans. Technol. Soc.*, vol. 1, no. 1, pp. 21–33, Mar. 2020, doi: 10.1109/tts.2020.2977059.

[13] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, "Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter," *PLoS One*, vol. 15, no. 9, Sep. 2020, Art. no. e0239441, doi: 10.1371/journal.pone.0239441.

[14] X. Yu, C. Zhong, D. Li, and W. Xu, "Sentiment analysis for news and social media in COVID-19," in *Proc. 6th ACM SIGSPATIAL Int. Workshop Emerg. Manage. GIS*, Nov. 2020, pp. 1–4, doi: 10.1145/3423333.3431794.

[15] A. Dey, "Machine learning algorithms: A review," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016.

[16] 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, 2020, doi: 10.3390/ijerph17062032.

[17] M. Bilgin and İ. F. Şentürk, "Sentiment analysis on Twitter data with semi-supervised Doc2Vec," in *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, 2017, pp. 661–666, doi: 10.1109/UBMK.2017.8093492.

[18] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and VADER sentiment," in *Proc. IMECS*, 2019, pp. 1–5.

[19] N. F. F. da Silva, E. R. Hruschka, and E. R. Hruschka, "Tweet sentiment analysis with classifier ensembles," *Decis. Support Syst.*, vol. 66, pp. 170–179, Oct. 2014. [Online]. Available: https://doi.org/10.1016/j.dss.2014.07.003

[20] S. H. Ghafarian and H. S. Yazdi, "Identifying crisis-related informative tweets using learning on distributions," *Inf. Process. Manag.*, vol. 57, no. 2, 2020, Art. no. 102145. [Online]. Available: https://doi.org/10.1016/j.ipm.2019.102145

[21] A. C. Sanders *et al.* (Sep. 2020). *Unmasking the Conversation on Masks: Natural Language Processing for Topical Sentiment Analysis of COVID-19 Twitter Discourse*. [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.08.28.20183863v3

[22] A. Depoux, S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith, and H. Larson, "The pandemic of social media panic travels faster than the COVID-19 outbreak," *J. Travel Med.*, vol. 27, no. 3, 2020, Art. no. taaa031, doi: 10.1093/jtm/taaa031.

[23] Y. Gao, Z. Xie, and D. Li, "Electronic cigarette users' perspective on the COVID-19 pandemic: Observational study using Twitter data," *JMIR Public Health Surveillance*, vol. 7, no. 1, Jan. 2021, Art. no. e24859, doi: 10.2196/24859.

[24] Y. Lu and Q. Zheng, "Twitter public sentiment dynamics on cruise tourism during the COVID-19 pandemic," *Current Issues Tourism*, vol. 24, no. 7, pp. 892–898, 2021, doi: 10.1080/13683500.2020.1843607.

[25] A. Hussain *et al.*, "Artificial intelligence–Enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: Observational study," *J. Med. Internet Res.*, vol. 23, no. 4, 2021, Art. no. e26627, doi: 10.2196/26627.

[26] K. Lagutina, V. Larionov, V. Petryakov, N. Lagutina, I. Paramonov, and I. Shchitov, "Sentiment classification into three classes applying multinomial Bayes algorithm, N-grams, and thesaurus," in *Proc. 24th Conf. Open Innovat. Assoc. (FRUCT)*, 2019, pp. 214–219.

[27] P. Singh, S. Singh, M. Sohal, Y. K. Dwivedi, K. S. Kahlon, and R. S. Sawhney, "Psychological fear and anxiety caused by COVID-19: Insights from Twitter analytics," *Asian J. Psychiatr.*, vol. 54, Dec. 2020, Art. no. 102280, doi: 10.1016/j.ajp.2020.102280.

[28] M. O. Lwin *et al.*, "Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends," *JMIR Public Health Surveillance*, vol. 6, no. 2, 2020, Art. no. e19447, doi: 10.2196/19447.

[29] A. Kumar, S. U. Khan, and A. Kalra, "COVID-19 pandemic: A sentiment analysis: A short review of the emotional effects produced by social media posts during this global crisis," *Eur. Heart J.*, vol. 41, no. 39, pp. 3782–3783, 2020, doi: 10.1093/eurheartj/ehaa597.

[30] Z. Hu, Z. Yang, Q. Li, A. Zhang, and Y. Huang, "Infodemiological study on COVID-19 epidemic and COVID-19 infodemic," to be published, doi: 10.21203/rs.3.rs-18591/v1.

[31] K. Garcia and L. Berton, "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107057, doi: 10.1016/j.asoc.2020.107057.

[32] H. Jang, E. Rempel, D. Roth, G. Carenini, and N. Z. Janjua, "Tracking COVID-19 discourse on Twitter in North America: Infodemiology study using topic modeling and aspect-based sentiment analysis.," *J. Med. Internet Res.*, vol. 23, no. 2, 2021, Art. no. e25431, doi: 10.2196/25431.

[33] H. Yin, S. Yang, and J. Li, "Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media," Jul. 2020. [Online]. Available: http://arxiv.org/abs/2007.02304.

[34] 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, 2020, Art. no. e19016, doi: 10.2196/19016.

[35] S. Boon-Itt and Y. Skunkan, "Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study," *JMIR Public Health Surveillance*, vol. 6, no. 4, 2020, Art. no. e21978, doi: 10.2196/21978.

[36] D. Whang and S. Vosoughi, "Dartmouth CS at WNUT-2020 task 2: Informative COVID-19 tweet classification Using BERT," in *Proc. 6th W-NUT*, Dec. 2020, pp. 480–484, doi: 10.18653/v1/2020.wnut-1.72.

[37] M. Alhajji, A. Al Khalifah, M. Aljubran, and M. Alkhalifah, "Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19," to be published, doi: 10.20944/preprints202004.0031.v1.

[38] G. Barkur, Vibha, and G. B. Kamath, "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India," *Asian J. Psychiatr.*, vol. 51, Jun. 2020, Art. no. 102089, doi: 10.1016/j.ajp.2020.102089.

[39] R. M. Merchant and N. Lurie, "Social media and emergency preparedness in response to novel coronavirus," *JAMA*, vol. 323, no. 20, pp. 2011–2012, 2020, doi: 10.1001/jama.2020.4469.

[40] S. Behl, A. Rao, S. Aggarwal, S. Chadha, and H. S. Pannu, "Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises," *Int. J. Disaster Risk Reduct.*, vol. 55, Mar. 2021, Art. no. 102101. [Online]. Available: https://doi.org/10.1016/j.ijdrr.2021.102101

[41] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 public sentiment insights and machine learning for tweets classification," *Information*, vol. 11, no. 6, p. 314, Jun. 2020, Art. no. 314, doi: 10.3390/info11060314.

[42] K. H. Manguri, R. N. Ramadhan, and P. R. M. Amin, "Twitter sentiment analysis on worldwide COVID-19 outbreaks," *Kurdistan J. Appl. Res.*, vol. 5, no. 3, pp. 54–65, 2020.

[43] L. Nemes and A. Kiss, "Social media sentiment analysis based on COVID-19," *J. Inf. Telecommun.*, vol. 5, no. 1, pp. 1–15, Jul. 2020, doi: 10.1080/24751839.2020.1790793.

[44] H. Kaur, S. U. Ahsaan, B. Alankar, and V. Chang, "A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets," *Inf. Syst. Front.*, to be published.

[45] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," *PLoS One*, vol. 16, no. 2, 2021, Art. no. e0245909.

[46] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106754. [Online]. Available: https://doi.org/10.1016/j.asoc.2020.106754

[47] G. A. Ruz, P. A. Henríquez, and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," *Future Gener. Comput. Syst.*, vol. 106, pp. 92–104, May 2020, doi: 10.1016/j.future.2020.01.005.

[48] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Comput. Surveys*, vol. 49, no. 2, pp. 1–41, Nov. 2016, doi: 10.1145/2938640.

[49] B. Shneiderman, "Design lessons from AI's two grand goals: Human emulation and useful applications," *IEEE Trans. Technol. Soc.*, vol. 1, no. 2, pp. 73–82, Jun. 2020, doi: 10.1145/2938640.

[50] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surveys*, vol. 4, pp. 40–79, Mar. 2010, doi: 10.1214/09-SS054.

[51] J. Carvalho and A. Plastino, "On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis," *Artif. Intell. Rev.*, vol. 54, no. 21, pp. 1887–1936, 2021.

[52] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 8, 2014, pp. 216–225. [Online]. Available: http://sentic.net/

[53] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 1310–1315.

[54] P. Vyas and O. F. El-Gayar, "Credibility analysis of news on Twitter using LSTM: An exploratory study," in *Proc. AMCIS*, 2020, pp. 1-5.

[55] J. Nowak, A. Taspinar, and R. Scherer, "LSTM recurrent neural networks for short text and sentiment classification," in *Proc. Int. Conf. Artif. Intell. Soft Comput.*, 2017, pp. 553–562.

[56] G. Preda. (2020). *COVID19*. Accessed: May 16, 2021. [Online]. Available: https://www.kaggle.com/gpreda/covid19-tweets

[57] A. B. Boot, E. T. K. Sang, K. Dijkstra, and R. A. Zwaan, "How character limit affects language usage in tweets," *Palgrave Commun.*, vol. 5, no. 1, pp. 1–13, 2019.

[58] N. Johnson, B. Turnbull, T. Maher, and M. Reisslein, "Semantically modeling cyber influence campaigns (CICs): Ontology model and case studies," *IEEE Access*, vol. 9, pp. 9365–9382, 2021, doi: 10.1109/ACCESS.2020.3048269.

[59] S. Karnouskos, "Artificial intelligence in digital media: The era of deepfakes," *IEEE Trans. Technol. Soc.*, vol. 1, no. 3, pp. 138–147, Sep. 2020, doi: 10.1109/tts.2020.3001312.

[60] S. Sharma and A. Jain, "Cyber social media analytics and issues: A pragmatic approach for Twitter sentiment analysis," in *Advances in Intelligent Systems and Computing*, vol. 924. Singapore: Springer, 2019, pp. 473–484, doi: 10.1007/978-981-13-6861-5_41.

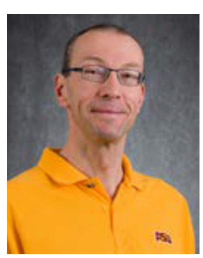
[61] B. Srivastava, F. Rossi, S. Usmani, and M. Bernagozzi, "Personalized chatbot trustworthiness ratings," *IEEE Trans. Technol. Soc.*, vol. 1, no. 4, pp. 184–192, Dec. 2020, doi: 10.1109/tts.2020.3023919.

[62] I. J. Strudwicke and W. J. Grant, "#JunkScience: Investigating pseudoscience disinformation in the Russian Internet Research Agency tweets," *Public Understand. Sci.*, vol. 29, no. 5, pp. 459–472, Jul. 2020, doi: 10.1177/0963662520935071.

[63] G. P. Zachary, "Digital manipulation and the future of electoral democracy in the U.S.," *IEEE Trans. Technol. Soc.*, vol. 1, no. 2, pp. 104–112, Jun. 2020, doi: 10.1109/tts.2020.2992666.

Piyush Vyas received the master's degree in information systems from Dakota State University, Madison, SD, USA, in 2020, where he is currently pursuing the Ph.D. degree in information systems with the College of Business and Information Systems.

His current research interests include text mining, association rule mining, machine learning, transfer learning, deep learning, and data mining.



Martin Reisslein (Fellow, IEEE) received the Ph.D. degree in systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 1998.

He is currently a Professor with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA.

Prof. Reisslein received the IEEE Communications Society Best Tutorial Paper Award in 2008, the Friedrich Wilhelm Bessel Research Award from the Alexander von Humboldt Foundation in 2015, and the DRESDEN Senior Fellowship in 2016 and 2019. He serves as an Associate Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, IEEE TRANSACTIONS ON EDUCATION, IEEE ACCESS, and *Computer Networks*. He is a Co-Editor-in-Chief of *Optical Switching and Networking*. He chaired the Steering Committee of the IEEE TRANSACTIONS ON MULTIMEDIA from 2017 to 2019 and was an Associate Editor-in-Chief of the *IEEE Communications Surveys & Tutorials* from 2007 to 2020.

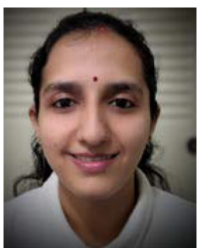
Bhaskar Prasad Rimal (Senior Member, IEEE) received the Ph.D. degree in telecommunications engineering from the University of Quebec, Quebec City, QC, Canada, in 2017.

He is an Assistant Professor of Computer and Cyber Sciences with Dakota State University, Madison, SD, USA. He was a Visiting Assistant Professor and an Academic Specialist with the Department of Computer Science, Tennessee Tech University, Cookeville, TN, USA. He was a Visiting Scholar with the Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. He was a Postdoctoral Fellow with the Electrical and Computer Engineering Department, University of New Mexico, Albuquerque, NM, USA.

Dr. Rimal was a recipient of the Doctoral Research Scholarship from the Quebec Merit Scholarship Program for Foreign Students of Fonds de Recherche du Quebec-Nature et Technologies, the Korean Government Information Technology Fellowship, the Kookmin University IT Scholarship, and the Kookmin Excellence Award as an Excellent Role Model Fellow. He has been serving as an Associate Editor for IEEE ACCESS since 2017. Further, he served as an Associate Technical Editor for the *IEEE Communications Magazine*, an Associate Editor for the *EURASIP Journal on Wireless Communications and Networking*, and an Editor for the *Internet Technology Letters*.

Ganga Prasad Basyal (Graduate Student Member, IEEE) received the Bachelor of Commerce degree in 2004, and the master's degree in business administration (marketing) from DAVV University, Indore, India, in 2009, and the master's degree in information systems from Dakota State University, Madison, SD, USA, in 2018, where he is currently pursuing the Ph.D. degree in information systems major in decision support and analytic.

His research areas include deep learning, transfer learning, machine learning, statistical analysis, and knowledge management for healthcare analytic using natural language processing techniques.



Gitika Vyas received the bachelor's degree in electronic instrumentation and control in 2014, and the master's degree in renewable energy from Rajasthan Technical University, Kota, India, in 2018.

Her research areas include data analysis in social networking, renewable energy, and solar energy.

Prathamesh Muzumdar is currently pursuing the Ph.D. degree in information systems with the Muma College of Business, University of South Florida, Tampa, USA.

His research interests mainly focus on behavioral information systems and design science methods.