Using a Hybrid-Classification Method to Analyze Twitter Data During Critical Events

SAADAT M. ALHASHMI, AHMED M. KHEDR, IFRA ARIF, AND MAGDI EL BANNANY

1Department of Information Systems, University of Sharjah, Sharjah 27272, United Arab Emirates
2Department of Computer Science, University of Sharjah, Sharjah 27272, United Arab Emirates
3Mathematics Department, Zagazig University, Zagazig 44519, Egypt
4College of Business Administration, Umm Al Quwain University, Umm Al Quwain, United Arab Emirates
5Faculty of Business, Ain Shams University, Cairo 11566, Egypt

Corresponding author: Ahmed M. Khedr (akhedr@sharjah.ac.ae)

ABSTRACT In this paper, sentiment analysis of two critical events is presented using machine learning (ML) techniques. COVID-19 has put immense pressure across the globe and sentiment analysis of data from Twitter using ML techniques has become a hot topic. We extract the COVID-19 and Expo2020 data from twitter. First, we evaluate the Twitter data of these two significant events for sentiment analysis and then use the classification algorithm to find out the usefulness of the proposed methodology. A hybrid approach that uses supervised learning model Support Vector Machine (SVM) combined with Bayes Factor Tree Augmented Naive Bayes (BFTAN) technique is proposed to accurately classify the input tweet while keeping in mind the different challenges of sentiment analysis. Our study has four main contributions: a) hybrid classification techniques are thoroughly explored for sentiment analysis, b) a novel hybrid classification approach is proposed for sentiment analysis, c) a new Twitter dataset related to COVID-19 that can be used for future research, d) empirical study to show that the hybrid-classification approach can achieve comparable performance in improving accuracy, identifying the polarity of comparative sentences, distinguishing the intensity of opinion words, considering negative words, and handling sarcasm as well. The experimental results show that the proposed approach is robust in producing correct classification results with the tradeoff of poor time efficiency. Also, the accuracy of the proposed model is comparable to other classifiers, which is encouraging. Class distribution of each dataset demonstrates that more than 60% of tweets are negative.

INDEX TERMS Sentiment analysis, Twitter data, COVID-19, Expo2020, support vector machine (SVM), Bayes factor tree augmented naive (BFTAN).

I. INTRODUCTION

For the past few decades, social media has been an essential platform for its users to share their thoughts, views, and feelings and help create a virtual bond between users. People contact each other and build relationships through messages, comments, posts, and likes. Social media platforms are used for sharing information, advertising, social events (crisis) and political purposes. Twitter is among the online social media platforms where users can read, comment and update short messages known as tweets. Twitter data is increasingly being used for numerous purposes, such as forecasting financial exchange prices, box office revenue for movies, and identifying clients with negative sentiments [1]. Twitter platform is used by journalists in critical events to gather information posted by users and re-tweet valuable messages. The output of textual documents on social media increases exponentially.

Scientific research on the semantic content of such tweets is known as sentiment analysis [2]. A survey was conducted in 2018 to estimate the number of active Twitter users. There were 326 million active users on Twitter, and they Tweet over 500 million tweets per day [3]. Analysis of sentiment helps determine the polarity of a given text. It is used for extracting people’s opinions, feelings and thoughts using Natural Language Processing (NLP) [4]. The aim is to identify the positive and negative value according to the standard categorization. Sentiment analysis is often referred to in a document classification task as opinion mining. The objective of sentiment analysis is to analyze the attitude of various individuals to a particular subject [5]. It is a challenging job to extract the sentiment information due to the volume and variety of data [5], [6]. Recently, sentiment analysis using
machine learning has become a significant research field. It is considered a classification problem (machine learning) in which the user can determine the algorithm for classification [7]. In a moment of crisis or critical incidents (health crises or natural disasters) where communities are disturbed, online social platforms including Twitter become a significant platform for information sharing.

Sentiment analysis of data from Twitter during critical events is considered a complicated task. Earlier, Lexicon-based approaches (unsupervised method) were used to anticipate the polarity of the given text, where text data is organized into a predefined set of sentiment classes and the sentiment lexicons are used to determine the sentiment score of the text [8]. However, lexicon-based methods are not practical approaches since the polarity of text data can often vary, as shown in the lexicon. So instead of lexicon-based methods, many scholars tend to use context-based lexicons to avoid issues with term polarity. In context-based lexicons, word orders and the syntactic relationship between words are considered.

A critical case is a situation characterized by the violation of a certain threshold by the entire country or by a single individual, which would result in a new and uncertain condition in response to which mixed and ambiguous emotional reactions are elicited [9]. The recent coronavirus pandemic (COVID-19) has put much strain on many countries worldwide after its severity was officially acknowledged in early January 2020 [10]. The majority of the inbound and outbound market sources (China, USA, Italy, France, Spain) are severely affected and significantly impact the country’s economy. Many countries have strict restrictions on leaving houses, airlines operations and hotel bookings. Most of the mega-events including Expo-2020, initially expected to open in October 2020, have already been delayed because of the safety precautions. Discussions about the impact and expected dates are on social media. Given this, the motivation behind this research is twofold. The first is to evaluate the Twitter data of these two significant events for sentiment analysis and then use the classification algorithm to find out the usefulness of the proposed methodology. Neutral emotions were also analyzed in [10] that are difficult to detect.

Critical events can be categorized into three relevant phases: the early phase of the event, the critical situation itself and the post-crisis recovery [11]. In the early phase, events such as earthquakes, climate changes are detected and are notified promptly by monitoring the Twitter platform. During the core of the crisis, using Twitter, we can obtain information about people’s opinions on the critical event’s nature that can help improve the calamity recovery system. In the reconstruction phase, Twitter communication can help improve critical events and disasters. Looking at the emotions, more people may work and donate for critical events or help develop mental health to confront the psychological impacts of the crisis. Part of the government’s and local officials’ attempts to manage the crisis rely on social media, and upon posting a social media request, people may expect a quick response [12].

This paper provides sentiment analysis of Twitter data using a hybrid classification (SVM and BFTAN) methodology. The relationship among words are identified using this classifier. We analyze the performance of this hybrid classifier using two Twitter datasets: the COVID-19 dataset and the Expo2020 dataset. Data extracted from Twitter are preprocessed, and then classified to compute accuracy, recall, and precision. Our hybrid-based approach aims to address the following challenges: improving accuracy, identifying the polarity of comparative sentences, distinguishing the intensity of opinion words, considering negative comments, and handling Sarcasm. Results demonstrate the efficacy of the suggested approach based on the accuracy and class distribution of each dataset. Briefly, the work presented here has four main contributions: a) hybrid classification techniques are thoroughly explored for sentiment analysis, b) a novel hybrid classification approach based on BFTAN is proposed for sentiment analysis, c) a new Twitter dataset related to the recent event (COVID-19) that can be used further in future research, d) it is empirically shown that the hybrid-classification approach can achieve comparable performance in improving accuracy, identifying the polarity of comparative sentences, distinguishing the intensity of opinion words, considering negative words, and handling Sarcasm as well. It also demonstrates that more than 60% of tweets are negative. The remaining part of the paper is: A literature review on the use of classifiers for sentiment analysis is given in Section II. Section III presents the various challenges of sentiment analysis. Section IV outlines the methodology used, data collection and data analysis. Section V describes the study findings and results interpretation, and some limitations and future directions are elaborated in Section VI. Finally, the paper is concluded in section VII.

II. LITERATURE REVIEW

This section provides a brief overview of sentiment analysis, followed by previous approaches used for sentiment analysis of data from Twitter at important events. Sentiment analysis or opinion mining helps assess the polarity of the review (text) as either positive or negative. Different critical classification levels are utilized in sentiment analysis and various sentiment analysis approaches that use classifiers for data from Twitter are proposed. A sentiment analysis method stated in [3] investigates the public sentiments at the Syrian refugee crisis. They gathered English and Turkish tweets using the Twitter package. A total of 2,381,297 tweets were collected for analysis. The sentiment analysis of Turkish tweets showed that the tweets classified into the positive category were higher than that of negative and neutral tweets. In English tweets, the number of neutral categories showed more numbers relative to other categories, with a ratio of 35% of all tweets to 12%. In [13], public opinions extracted from Twitter related to the most challenging international affair - The Iran Deal that took place in July 2015 between Iran and other countries
(China, the United Kingdom, the United States, Russia, and the European Union). After analysis, 41% of tweets were classified as positive, the remaining 35% as negative and 24% as neutral. Then, they applied five classification algorithms; Neural Networks, SVM, Decision Trees (DT), Naive Bayes (NB) and K-NN on the tweets. SVM showed more accurate prediction with an accuracy rate of 80.95%.

In [14], a novel approach for Twitter sentiment classification is presented with a non-iterative Deep Random Vectorial Functional Link (D-RVFL) that consists of four hidden layers having the same number of neurons and activation functions at each layer. The effectiveness was tested using Precision, Recall, Accuracy and F1 score and compared with Random Vector Functional Link (RVFL), SVM and Random Forest (RF). It showed that the proposed approach surpasses other methods based on F1 score. An automatic text classification system proposed in [15] classifies tweets according to various general requirements in times of crisis such as food, shelter, water, electricity, medical emergency, collapsed structure and trapped people. The best features were chosen utilizing term frequency and Chi-square. Finally, data classification is performed using SVM and NB. The results showed that SVM outperforms other NB algorithms. Disaster Risk Management Priorities can be determined in which public feedback can be used to make local communities better in times of disaster.

In [16], a Bidirectional Recurrent Neural Network (BRNN) analyzes the collected textual data and sequential data. They divided the data corpus in which 85% of the training data and 15% of the testing data were used to construct a BRNN model that achieved an 81.67% accuracy rate. Machine learning-based approach for sentiment analysis combined with features of character n-grams language model is used in [17] for SemEval-2013 Task 2 (Task B) for “Message Polarity Classification”. This system combined the SVM classifier results with the n-gram language model to make a final prediction. This addressed high lexical variation in Twitter data and achieved a good performance.

In [18], a deep learning sentiment analysis was used in conjunction with ensemble methods to enhance the algorithm performance. Seven public datasets collected from microblogging and movies reviews domain were used to test the performance of this approach where it outperformed other baseline approaches based on F1 score. In this work, performance was enhanced by taking advantage of the existing ensemble classifiers and combining them with sentiment-trained word embedded manually developed features. The strategy suggested in [19] to encode the subject information through word embedding was used for Twitter sentiment classification. In this approach, the topic information of tweets was generated using Latent Dirichlet Allocation (LDA). Then a topic-enhanced recursive autoencoder used to encode the topic information and used SVM classifier to evaluate word representation. Finally, a topic-enhanced word embedding model was integrated with the traditional model for greater accuracy. By using topic-enhanced word embedding as a feature, this method gave a 78.57% macro-F measure to predict the positive/negative polarity of the tweets. The integrated model gave an improved 81.02% macro-F-measurement.

A novel sentiment analysis approach was presented in [20] to detect seven sentiment classes on micro- posts such as tweets at the time of situational awareness. Different machine learning algorithms were compared and various features used to detect micro-posts were analyzed. The results of the classifiers were analyzed to detect highly-crisis-relevant information. A sentiment value of tweets in times of disaster was accurately detected in [21]. They used several methods to test the flexibility of the proposed system. Four different methods such as SentiWordNet, Emoticons, AFNN and Bayesian Networks were compared. Bayesian Networks with SentiWordNet provided the best precision and recall and improved sentiment detection. An algorithm was proposed in [22] to monitor real-time tweets and detect target events, such as earth- quakes, where each Twitter user is regarded as a sensor. They first applied a semantic analysis to extract tweets on the target event. Then the NB algorithm was applied to the training data to obtain a probabilistic model that can detect the target event.

In [23], a hybrid approach was used in which a lexicon/supervised learning approach, and two supervised machine-learning approaches were evaluated in multi-way sentiment analysis. Information from the lexicon was paired with the information from the NB Classifier with a set of features. Supervised learning methods used in this approach are NB classifier with feature selection and MCST classifier based on hierarchical SVMs considering label similarity. The results showed that the mix of lexicon and corpus-based information is superior to other state-of-the-art systems. A work presented in [7] used Bayesian Network Classifiers for sentiment analysis on two Spanish datasets: the Chilean earthquake (2010) and the Catalan independence referendum (2017). Five different classifiers are presented with one being the variant of Tree Augmented Naive Bayes (TAN) model. SVM achieved 80% accuracy rate on both datasets with respect to the accuracy of the models. Both these models find it hard to perceive the relationship between features at the time of classification and the drawback is dealt using the Bayesian Network Classifier. A downside of using the suggested method is that it relies heavily on event and time.

In [24], a hybrid model is proposed by using Gated Recurrent Unit (GRU) and Symbiotic Organisms Search (SOS), named Symbiotic Gated Recurrent Unit (SGRU). Text mining and Natural language processing techniques are used to pre-process the construction site accidents data. This proposed approach is evaluated using other classification techniques based on average weighted F1-score. A hybrid model is presented in [25] to detect epileptic seizure using GA and Particle Swarm Optimization (PSO) to optimize parameters for SVM. Experimental results shows that the performance of the PSO-SVM and GA-SVM are comparable and they both outperform simple SVM. In [26], a deep transfer learning model is presented to detect people who are not wearing
face masks with the help of surveillance cameras to control the transmission of COVID-19. A deep transfer learning model known as Resnet50 combined with machine learning techniques for classification using decision trees, SVM and ensemble methods. Comparison techniques are performed to find the most suitable approach that will result in highest accuracy and less time.

In [27], US Presidential Elections-2012 and Karnataka State Elections 2013 data set were analyzed using SVM with PCA. PCA helps in reducing dimensions and achieving better accuracy. However, it is not too much efficient and does not provide consistent output. It showed an accuracy of 88%. In [28], sentiment analysis of microblogging services using Naive Bayes, Maximum Entropy, and SVM were analyzed. It achieves higher accuracy by doing preprocessing steps. But emoticons and other language tweets are not considered in this approach. The method showed an accuracy of 80%.

In [29], Sina microblog data is analyzed using SVM. The method saves more time and shows good performance. In [30], manually labeled data of 300 positive, negative and neutral tweets were analyzed using SVM. Prediction accuracy is good as compared to keyword-based approach. But it does not handle dialects or domain specific issues. The method showed an accuracy of 86.89%. In [31], the model analyzed 9,195 tweets and 2,181 hashtags collected in one week period using SVM. The result showed an accuracy of 84.13%. However, the classification of hashtags are not employed. In [32], NB, RF and SVM approaches are used to monitor consumer opinion about their products. The work, however, didn’t consider neutral tweets. Internet movie database is considered in [33]. They used NB and obtained an improved accuracy rate of 81.42%. At the same time, French movie reviews were analyzed in [34]. They used SVM and showed an improved classification performance with an accuracy of 93.25%. Three different movie review datasets were analyzed in [35] using NB and SVM, and achieved impressive accuracy levels.

Chinese sentiment corpus with a size of 1021 documents were analyzed in [36] using centroid classifier, K-nearest neighbor, winnow classifier, Naive Bayes and SVM. SVM exhibits the best performance for sentiment classification and reported an information gain of 0.90 for SVM. In [37], microblogging data were analyzed using Naive Bayes, RF and SVM. Naive Bayes showed consistently accurate results. 2000 movie reviews were analyzed in [38] using NB and SVM. The method is language independent, however computationally expensive. The results showed an accuracy of 86.35%.

Table 1 represents the comparison of most common methods in machine learning for sentiment analysis.

### III. CHALLENGES IN SENTIMENT ANALYSIS

After conducting a literature review in the previous section, it can be concluded that sentiment analysis is a difficult task and many challenges can be considered when doing sentiment analysis. However, a few of the challenges of sentiment analysis are highlighted below:

- Applying sentiment analysis to non-English languages is a challenging task as most of the available online sentiment dictionaries are in English, so translation problems may occur and impact the system’s output.
- Identifying the context and sense of the word is very challenging. For example, the term ‘small’ refers to the size, is sometimes considered a negative adjective for cars and a positive adjective for computer devices.
- Identifying the polarity of comparative sentences is very difficult to calculate. For example, “i7 core is faster than i5 core”, which indicates that the word faster is associated with i7 core.
- Mostly negation words are not appropriately handled while doing sentiment analysis that may be the reason for incorrect results. For example, “the weather is not good today” contains positive polarity for the word “good”. Still, due to the negative word “not”, the sentence’s meaning is changed completely.
- Distinguishing the intensity and strength of the opinion word (positive vs extremely positive) is also a challenging task.
- Developing algorithms and techniques that help improve the system’s accuracy is also a challenging task as there is no clear way to identify such algorithms.
- Real-time opinion mining and real-time data collection are required as colossal growth can be seen in social networking websites such as Facebook, Twitter etc. it is crucial to have an automated system.
- Identifying Sarcasm is also a challenging task. Some writers include ironic comments to form a positive meaning of the sentence. In contrast, the intended meaning is negative or vice versa.

Different techniques addressed different challenges of sentiment analysis. Some techniques focused on increasing the accuracy of the system using hybrid approach but didn’t address other challenges of intensity, comparative words, negation and Sarcasm etc. A limited work is done to handle Sarcasm in text. A new approach is required that can solve all these challenges in real time. The literature on hybrid approach is very limited and is mostly developed to solve the limitations of other previous approaches. In this paper, we use two supervised machine learning techniques to create a hybrid classification framework in a novel manner for twitter sentiment analysis. The proposed hybrid approach used available resources, a set of rules, a supervised learning model SVM combined with Bayes Factor Tree Augmented Naive Bayes technique to accurately classify the input tweet while keeping in mind these different challenges of sentiment analysis. Although, we also investigated traditional methods for designing a better approach for twitter sentiment analysis, yet these techniques are popular among research community and also have been investigated.
TABLE 1. Comparison of common sentiment analysis techniques.

| Ref. | Dataset | Technique used | Advantages | Limitations | Reported accuracy |
|------|---------|----------------|------------|-------------|------------------|
| [24] | Construction Accident Narrative Historical Data | Hybrid SORU | It represents the best algorithm in terms of classification performance. | Does not address imbalance class issues. | Highest precision among others |
| [25] | electroencephalogram (EEG) signal data | GA and PSO | High classification accuracy. | Hybrid classifiers require more time for execution. | 99.38% |
| [26] | face masked datasets | Decision trees, SVM and ensemble algorithm | More efficient. | More time is consumed | 99.76% for decision tree & 100% for SVM |
| [27] | US Presidential Elections 2012 and Karnataka State Elections 2013 | SVM with PCA | PCA helps in reducing dimensions and offers better accuracy. | Does not provide consistent output, not too much efficient and fast | 88% |
| [28] | microblogging services | Naive Bayes, Maximum Entropy, & SVM | Achieves higher accuracy by doing pre-processing steps | Emotions and other language tweets are not considered in this approach. | 80% |
| [29] | Sina microblog data | SVM | Good performance, saves time | Not mentioned in the paper | 62.9% |
| [30] | Manually labeled data of 300 positive, negative & neutral tweets | SVM | Good prediction accuracy | Does not handle dialects or domain specific issues. | 86.69% |
| [31] | 29,193 tweets & 2,181 hashtags (one week time) | SVM | Competitive performance is achieved. | Classification of hashtags are not employed. | 84.13% |
| [32] | Stanford Sentiment!140 corpus & HCR | NB, RF and SVM | Monitor consumer opinion about their products. | Neutral tweets are not considered in this approach. | 75.81% |
| [33] | Internet movie database | NB | Improved accuracy rate. | Burden of labeling many example sets in target domain. | 81.42% |
| [34] | French movie reviews | SVM | Improved classification performance. | Not mentioned in the paper. | 93.25% |
| [35] | 3 movie review dataset | NB, SVM | Achieved impressive accuracy levels. | Computing PMI values is a very time consuming process. | 81.14% |
| [36] | Chinese sentiment corpus with a size of 1021 documents | centroid, k-nearest neighbor, winnow, Naive Bayes & SVM | SVM shows best performance for sentiment classification. | Sentiment classifiers are severely dependent on domain or topics. | SVM: Information gain is 0.9 |
| [37] | microblogging data | Naive Bayes, RF and SVM | Naive Bayes shows consistently accurate results | Skewness in dataset that impacts recall | 72.70% |
| [38] | 2000 movie reviews | NB and SVM | Language independent | Computationally expensive approach | 86.35% |

S. M. Alhashmi et al.: Using a Hybrid-Classification Method to Analyze Twitter Data During Critical Events

IV. PROPOSED SVM-BFTAN METHOD

A. DATA COLLECTION

The efficiency of the proposed approach is measured using two sets of twitter data- The COVID-19 and the Expo2020 datasets. The tweets are collected to measure the user positivity and negativity ratio/response on these two events. The dataset is collected by implementing the following steps:

- Twitter Search Strategy: All tweets were extracted and collected using twitter search strategy.
- Hashtags selection: The hashtags that are related to the chosen events are listed in Table 2.
- Tweets collection: Tweets are collected using the listed hashtags. The collected tweets are directly made available for further processing. The dataset structure is shown in Table 2.

B. PRE-PROCESSING

A tweet can have different views expressed by people in a variety of ways. The input data obtained from twitter is redundant and inconsistent, so pre-processing is the fundamental task. This step is implemented over the collected tweets to remove redundant data, reshape them and make it available...
for the feature extraction step. The main objective of this step is to remove unwanted parts, URLS, hashtags, punctuations, numbers, stop words and to highlight only significant parts that will make the further processing easy and accurate. In our approach, pre-processing is done in two phases and are implemented sequentially i.e., the output of each step is taken as an input of another step.

**Phase 1:** This phase is used to eliminate the unwanted noise/unwanted elements from twitter dataset such as:
- Eliminate all URLS, email addresses etc. using regular expression matching
- Remove all hastags (#) and the word that follows it.
- Remove twitter terms that start with (@) symbol that is used to tag some entity.
- Remove all symbols, parenthesis, backward slashes, forward slashes, numbers, punctuations and dashes from tweets using regular expression matching.
- Substitute a single white space for multiple white spaces.
- Remove all non-English letters using regular expression matching.
- Detection and separation of emoticons

**Phase 2:** Two dictionaries, stop words and acronym are used to enhance the accuracy and precision of the twitter dataset processed in phase 1. The following steps are involved in this phase:
- All tweets are changed to lower case for data uniformity.
- Common stop words like a, the, is, etc. are omitted by comparing with the stop word dictionary.
- Duplicate tweets are removed from the same user ID.
- Negation words are detected and separated.

As an example, each step of pre-processing is illustrated in Table 3.

**C. FEATURE EXTRACTION**
From the pre-processed information, the proposed framework recognizes and separates the feature sets. The feature set is portrayed as a part or attribute of an object for which the planned assessment is to be performed. These feature sets are utilized to characterize and classify the data. The involved features address practically every aspect of the tweets, to address the objective targets of negative words, sarcasm detection and improving accuracy. Our proposed system make use of two feature extraction techniques- word2vec and sentiment score, that are used to get better results and reduce the effect of real neutrals [39]. A set of parameters

---

**TABLE 3. Preprocessing explained using input example.**

| Phase | Steps | Output |
|-------|-------|--------|
| 1     | URL removal | This had me dying. Make us corona-free @Vakkstar123 #go2quad #Sidemen |
|       | Hashtags removal | This had me dying. Make us corona-free @Vakkstar123 |
|       | Hash tags removal | This had me dying. Make us corona-free @Vakkstar123 |
|       | Twitter term removal | This had me dying. Make us corona-free |
|       | Punctuation removal | This had me dying Make us corona free |
|       | Multiple white space removal | This had me dying Make us corona free |
|       | Non-English removal | This had me dying Make us corona free |
|       | Emoticons detector | This had me dying Make us corona free |
| 2     | Convert to lower case | this had me dying make us corona free |
|       | Stop removal words | dying make us corona free |
|       | Remove duplicate tweets | dying make us corona free |
|       | Negation detector | dying make us corona free |
TABLE 4. Features considered in Tweet analysis.

| No  | Features                      | Description                                                                 | Values          | Effect |
|-----|-------------------------------|-----------------------------------------------------------------------------|-----------------|--------|
| 1   | Identifying polarity          | Polarity of the words that are presented in the tweet.                      | Numerical [-1]  | Linear |
| 2   | Distinguish intensity         | Intensity of the words (intensive and exclamation detector)                 | Categorical    | Associative |
| 3   | Handle negative words         | Handling negative words that are present in the tweet                      | Categorical    | Associative |
| 4   | Handle Sarcasm                | Handle irony and sarcasm comments in the tweets                             | Categorical    | Associative |

TABLE 5. Different polarity categories and its scores.

| Category     | Sentiment Score | Value |
|--------------|-----------------|-------|
| Very Positive| 0.1 or higher   | 3.2   |
| Positive     | Higher than 0   | 1     |
| Neutral      | 0               | 0     |
| Negative     | Less than 0     | -1    |
| Very Negative| Less than 0.1   | -2,-3 |

that are considered while constructing the model is identified and described in the Table 4.

1) FEATURE EXTRACTION WITH THE HELP OF SENTIMENT SCORE

We took a list of positive and negative English opinion words or sentiment words and each tweets were labeled with negative or positive, based on the opinion and the sentiment words. First, we calculate the sentiment polarity [40] of all tweets using (1),

\[
\text{SentimentScore} = \frac{\text{Positive} - \text{Negative}}{\text{Positive} + \text{Negative} + 2}
\]

This equation is used to get the direction and strength of the sentiment [41], where positive represents positive word count and negative represents negative word count in a tweet. Sentiment class C is represented by two-discrete values i.e.,

\[
C \in [-1, 1].
\]

This variable C is used in capturing sentiment values and the distances between them. Sometimes, we cannot decide the emotionality degree of the tweet because the polarity measure fails to do so (Sentiment Score=0) or the positives and negatives cancel each other out. In this case, we cannot decide whether the tweet is positive, negative or neutral. Hence, we use the following definition [7]:

\[
C(w) = \begin{cases} 1 \text{ (Positive tweet),} & \text{if } \text{SentimentScore} \geq 0.1 \\ -1 \text{ (Negative tweet),} & \text{if } \text{SentimentScore} < 0.1 \end{cases}
\]

After the polarity of the tweet is calculated using (3) and according to the values of the involved features, the tweet is classified into one of the following categories and their respective scores are presented in Table 5:

- If polarity is greater than or equal to 0.1, then the tweet is classified as Very Positive.
- If polarity is less than 0.1 and greater than 0.0, then the tweet is classified as Positive.
- If polarity is equal to 0.0, then the tweet is classified as Neutral.
- If polarity is less than 0.0 and greater than −0.1, then the tweet is classified as Negative.
- If polarity is less than or equal to −0.1, then the tweet is classified as Very Negative.

2) FEATURE EXTRACTION WITH THE HELP OF Word2Vec

Word2vec is an unsupervised learning algorithm. The main idea of Word2Vec is to supply the word sequence with weights of each word, making the embedding/ input layer a vector representation of the texts [42]. Word2Vec is one of the most popular technique to learn word embeddings using shallow neural network. Word embedding is one of the most popular representation of document vocabulary. It is capable of capturing context of a word in a document, semantic and syntactic similarity, relation with other words, etc. We have used the CBOW (continuous bag of words), where the model predicts the word under consideration given context words within specific window. The number of dimensions in which the current word must be expressed at the output layer is defined in the hidden layer. By traversing the dataset, Word2Vec vectors are created for each examination in the data set. We obtain the word embedding vectors for each word in the review by simply applying the model to that word. We will apply an average over all the vectors of words in a sentence to represent a sentence from our dataset. These Word2Vec vectors are then used for classification in the next phase. We used a simple equation to allocate polarity to the words in the vocabulary, with a single positive seed word and a single negative seed word as in [43], as follows:

\[
C(w) = \text{sim}(w, P) - \text{sim}(w, N)
\]

where P is the positive seed word for the domain represented by its corresponding word vector w and N is the negative seed word for its word vector. sim represents cosine distance between word vectors. The polarity is positive if the value is greater or equal to zero, and negative otherwise.

D. PROPOSED HYBRID CLASSIFICATION APPROACH

In our proposed hybrid approach, we use different classifiers that are used to deal with different problems so that the accuracy of the system can be improved. The proposed approach is based on hybrid-based approach of SVM and Bayes Factor Tree Augmented Naive Bayes (SVM & BFTAN). Support Vector Machine is a supervised learning model that finds the hyper-plane with the [44]. SVM is also referred to as
constrained optimization problem, where the class \(c\) \([1, -1]\) represents positive and negative training samples of document \(d\) and the solution represented by the given vector \(w\),

\[
w = Xj \alpha j cj dj, \alpha j \geq 0
\]

where \(j\)'s can be found by solving a dual optimization problem, the documents \(dj\) that exceed zero are support vectors, and are the sole vectors that contribute to the vector \(w\). Classification of these instances consists of finding the side of a hyper-plane on which they fall. It performs well with a small dataset and provides a non-linear solution with a kernel function to map the input variables into high dimensional space. The aim of SVM is to split the datasets into classes in order to find the maximum marginal plane (MMH). It works well with high dimensional space and offers good accuracy while not working well with overlapping classes. Linear splines kernel is used to deal with data vectors. Input data is normalized so that features are on same scale and compatible. \([0, 1, 1000]\) and \(\gamma [0.001, 1]\).

Bayesian networks graphical model takes the joint probability distribution of a set of discrete random variables and encode it \([45]\). Given an input data point, Bayesian network and the Bayesian model selection factor, the considered datasets is depicted in Table 2.

A sentiment analysis was performed using the proposed Gaussian Naive Bayes. The steps for sarcasm extraction are

1. Choosing the classifier with best accuracy.
2. Training the model with best classifier.
3. Testing the model with best classifier.
4. Real time testing of tweets.

\section*{E. SARCASTIC DETECTION}

People do not always express their opinions in the same way. Opinion of each person or individual is diverse in light of the perspective. A few individuals express it in sarcastic manner that is by all accounts positive at the point when perused yet are really not. Comments made by them may mean something contrary to what they have said. They are generally made to offend someone or might be to criticize something in a comical manner. Example: My flight is delayed...that’s amazing...! \([35]\). The detection of these kinds of words are very difficult to predict so we used a set of classifiers and the best one if used for testing and functioning of the model. The sarcasm detection in the proposed model is performed using classifiers such as Decision Tree, Random Forest, Gradient Boosting, Adaptive Boosting, Logistic Regression, and Gaussian Naive Bayes. The steps for sarcasm extraction are as follows:

\begin{itemize}
\item Classification of the prepared data using five classifiers.
\item Choosing the classifier with best accuracy.
\item Training the model with best classifier.
\item Testing the model with best classifier.
\item Real time testing of tweets.
\end{itemize}

\section*{V. EXPERIMENTAL RESULTS}

A sentiment analysis was performed using the proposed approach, using two Twitter datasets. A brief description of the considered datasets is depicted in Table 2.

1) COVID-19 dataset The dataset from the Twitter is connected between Twitter API and Rapidminer. Then we use the “Search Twitter” operator to search recent tweets on the topic. 120 K tweets are collected between...
11-05-2020 to 03-10-2020 and #coronavirus, #Covid-19, #corona, #virus are used as the keywords. The raw data is pre-processed to indicate positive and negative tweets based on sentiment and opinion words present in the tweet.

2) Expo-2020 dataset The dataset from the Twitter is connected between Twitter API and Rapidminer. We use “search twitter” operator to search tweets specific to the “United Arab Emirates”. 5000 tweets are collected between 14-05-2020 to 16-05-2020 using #Expo2020, #delayExpo, #postponedexpo2020 and others as keywords. The raw data is then pre-processed to indicate positive and negative tweets based on sentiment and opinion words present in the tweet.

A. SIMULATION SETUP

We use Python 3.8.3 with Anaconda distribution environment to implement our proposed classifier and then we use the “execute python” operator in Rapidminer to embed it. We train our both datasets using the proposed classifier and compare them by computing their True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). After training the datasets, the following performance measures were computed: (i) Accuracy, (2) Precision, and (3) Recall.

The feature vectors gathered from the prior step is given as an input to the proposed approach (SVM-BFTAN) i.e., a classification module that uses SVM and BFTAN to classify the data. We randomly sampled our previously described datasets into train and test data. We use 70% of the dataset for generating the training set and 30% for the test set. We first train our classifier on training dataset and then the confusion matrix is computed using the test set such as:

- **True Positives (TP):** Number of positive tweets that are predicted correctly.
- **True Negatives (TN):** Number of negative tweets that are predicted correctly.
- **False Positive (FP):** Number of positive tweets that are predicted incorrectly.
- **False Negative (FN):** Number of negative tweets that are predicted incorrectly.

Subsequently, the following performance measures are computed:

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)
\]
\[
\text{Recall} = \frac{TP}{TP + FN} \quad (12)
\]
\[
\text{Precision} = \frac{TP}{TP + FP} \quad (13)
\]

The splitting procedure is executed 7 to 10 times for both datasets; 70% of the training samples and 30% of the test samples. For each set, the classification performance is measured on test set and then the average of each measurement is reported. Class imbalance problem occurs when the class sizes are different. Mostly, any one of the classifiers will be biased towards the majority class. Class distribution of the datasets shows that the 60% of the tweets are negative.

Performances of five classification algorithms- BFTAN, TAN, NB, SVM and RF are compared. The results shown in Table 6 indicate the comparative results of the suggested approach and the existing approaches for the above three parameters with sentiment score feature extraction using dataset 1. The results of our empirical study suggest that the accuracy of the proposed approach is comparable to other classifiers. Table 7 displays the comparative results of the suggested approach and the existing approaches for the above three parameters with sentiment score feature extraction using dataset 2. The results demonstrate that SVM-BFTAN has shown 90.84%, 91.22% and 90.08% accuracy, precision and recall values. The recall value of SVM-BFTAN is the highest compared to the other classifiers while RF showed the highest accuracy values followed by SVM-BFTAN on dataset 2. These results suggest that improved performance can be achieved by combining different techniques in a systematic manner.

The results shown in Table 8 indicate the comparative results of the proposed approach and the existing approaches

### TABLE 6. Comparison of the proposed method with the existing methods in terms of accuracy, precision and recall using dataset 1.

| Algorithm | Covid-19 Dataset | Expo-2020 Dataset |
|-----------|------------------|-------------------|
|           | Accuracy | Precision | Recall |
| 1: SVMBFTAN | 82.8% | 86.8% | 82.6% |
| 2: BFTAN | 72.5% | 73.9% | 73.5% |
| 3: TAN | 69.8% | 72.1% | 70.9% |
| 4: NB | 72.13% | 72.08% | 73.56% |
| 5: SVM | 69.62% | 70.89% | 70.43% |
| 6: RF | 70.50% | 71.70% | 70.02% |

### TABLE 7. Comparison of the proposed method with the existing methods in terms of accuracy, precision and recall using dataset 2.

| Algorithm | Covid-19 Dataset | Expo-2020 Dataset |
|-----------|------------------|-------------------|
|           | Accuracy | Precision | Recall |
| 1: SVMBFTAN | 90.84% | 91.22% | 90.08% |
| 2: BFTAN | 88.50% | 89.95% | 89.04% |
| 3: TAN | 87.62% | 89.91% | 87.14% |
| 4: NB | 87.30% | 89.80% | 88.56% |
| 5: SVM | 90.45% | 91.75% | 86.01% |
| 6: RF | 91.50% | 90.61% | 90.02% |

### TABLE 8. Performance evaluation of the algorithms with feature extraction approach (Word2Vec) using both datasets.

| Algorithm | Covid-19 Dataset | Expo-2020 Dataset |
|-----------|------------------|-------------------|
|           | Accuracy | Precision | Recall |
| 1: SVMBFTAN | 0.81 | 0.75 | 0.79 | 0.90 | 0.87 |
| 2: BFTAN | 0.74 | 0.77 | 0.80 | 0.86 | 0.89 |
| 3: TAN | 0.76 | 0.79 | 0.71 | 0.90 | 0.88 |
| 4: NB | 0.70 | 0.75 | 0.79 | 0.89 | 0.83 |
| 5: SVM | 0.77 | 0.73 | 0.85 | 0.86 | 0.91 |
| 6: RF | 0.75 | 0.77 | 0.81 | 0.82 | 0.89 |
for the above three parameters with Word2Vec feature extraction using dataset 1 and dataset 2. The results demonstrate that the proposed classifier has the highest accuracy, precision and recall values. It shows that the comparable performance can be achieved by combining two supervised classification techniques. SVM outperforms other techniques in terms of time efficiency. As expected, our hybrid classification approach is very time consuming (mean time across all features $\approx 980$ sec) as it combines two individual techniques. Hence the total time elapsed by each technique adds up in hybrid classification.

We compare five classification algorithms on two different events with Ruz et al. [7] results. We use the same approach of sentiment analysis [7] with the same size of data but on two different events to see the performance of the algorithm. Comparison results are shown in Figure 2. The results show that our values are less accurate compared to the benchmark paper on dataset 1. Figures 3 and 4 show the precision and recall values of the five classifiers on dataset 1. Our values are slightly less compared to the benchmark paper on COVID-19 dataset.

Figure 5 shows the accuracy values of the five classifiers on dataset 2. The results show that our values are more accurate compared to the benchmark paper on dataset 2. Figures 6 and 7 show the precision and recall values of five classifiers on dataset 2. Our results are better compared to the benchmark mark paper on the Expo2020 dataset.

By looking at the results, we can see that the performance of the five classifiers on dataset 1 is not the best compared to the Ruz et al. approach but the performance is better on dataset 2, so we use an ensemble of Bayesian Boosting (BB) on the classifiers to improve our results on Covid-19 dataset. Table 9 shows the improved version of the five classifiers after adding the ensemble of BB, in terms of accuracy,
VI. DISCUSSIONS AND FUTURE DIRECTION

In this section, we present some interesting observations related to the strengths and limitations of the proposed approach arose during this study which are worth discussing. Our results suggest that SVMBFTAN is the only classifier that shows an above 80% accuracy values on both the datasets. Our proposed hybrid technique offers better quality classification than the other traditional techniques. In dataset 1, there is enough data to support SVMBFTAN model therefore, the accuracy values of SVMBFTAN are better compared to the BFTAN and the other Naive Bayes classifiers, while in dataset 2 there were not many examples to promote SVMBFTAN model. In this case, the generalization performance of SVMBFTAN is competitive with RF. RF offers better time efficiency. If output quality and time efficiency are addressed together, this raises the question as how to tradeoff between these two. If one approach gives better quality, then its time efficiency is compromised so it’s very challenging which should be preferred more. As a suggestion, output quality should be considered more as compared to its time consumption because at the end it’s only the quality that matters the most and our approach gives better quality results. However, it might be interesting to see how the performance of our proposed approach could change in terms of solution quality and time by using many other techniques in different manner or using high-performance computing (HPC) devices for better results. The comparison of our approach to the
traditional methods in social sciences (in-depth interviews, focus groups, questionnaires) and the big data analysis of social media including Twitter, shows us a broader and a more dynamic image of critical events. However, the difficulties in coding and classifying the twitter data may remain in the future as well. During critical events, the number of tweets and re-tweets are increasing day by day, so it is challenging to identify relevant and meaningful communication patterns. In some cases when the twitter information does not match the ground information obtained through technical devices from natural locations, it may lead to wrong decisions by the authorities. Several challenges need to be resolved in the future, such as approaches to minimize error while collecting information on critical events and introducing new techniques that are time efficient and improving geo-location techniques. As our analysis shows, the classification of Twitter data has improved our communication dynamics in critical events. A statistical debate on the representativeness and relevance of Twitter data has been an important research subject over the past few years. In general, there are two questions: if these Tweeted reflect the real society or if these Twitter samples portray Twitter communications. Our focus in this study is on the second question. Twitter data is accessed in the following two ways. The most common of them is the API and the Streaming API that is partitioned into two sources: Filter API (in which we can search using parameters, i.e., keywords, hashtags, user accounts and geographic locations) and Sample API (that provides one percentage of all tweets with no parameters). The next source is the Representational State Transfer or REST API that gives timelines of specific users that is confined to 3200 recent tweets. Twitter do not share information on its sampling techniques, so it is challenging to determine if the one percentage of Twitter data that is public is a true reflection of twitter communication or not.

This paper focusses to use a hybrid classifier for sentiment analysis during critical events. The performance of this methodology is largely dependent on the quality of the training examples. In some cases, there may be a chance of biased or noisy data. In this case, the sentiment polarity of hashtags is used as a feature in the classification process [46]. The number of positive and negative hashtags is taken as the input feature in this case.

For the future work, we explore hashtag classification at initial level (hashtag-level sentiment classification) for COVID-19 dataset such as the recent conflict in October 2019 [41], [46], or in the case of a natural disaster like an earthquake. Hashtags related to positive feelings as well as negative impacts, could be analyzed. In the following stage, tweet sentiment analysis may be done. More approaches to integrating hashtag-level sentiment classification with tweet-level sentiment classification to achieve more reliable and robust findings can be discussed. The suggested approach to sentiment analysis is time-consuming and case dependent, which means that the relevance and predictive capability are limited to the brief term or short time period of the event. For example, if we train our classifier for a potential case of the same nature, new hashtags will arise that can create complications, when people need to use additional terms to express their thoughts and emotions. The classifier not trained according to the new event makes its use limited. However, the design has a network architecture that is interlinked with words and how they classify the event, ensuring the conceptual validity of the model.

If the predictive model changes unexpectedly, either slowly or rapidly for events in the future in a setting which was initially modeled, an online learning approach is adopted. Alternatively, an unsupervised learning approach can be used on new event and words. We can subsequently couple this unsupervised network with a fine-tuned version of the original Bayesian network to obtain the words that appear in both the events. Techniques to handle missing data and its flexibility makes machine learning an up-and-coming area for sentiment analysis in critical situations. In future, other qualitative approaches such as grounded theory, can be combined with the proposed classifier to obtain better results. This may inspire researchers and scientists to prioritize machine learning techniques in their studies and to recognize Twitter as a good platform for reflecting emotions and social facts.

VII. CONCLUSION
This paper introduced a novel hybrid classification approach to analyze the feelings of tweets using the SVM and BFTAN methods. The method was tested on two Twitter datasets, the COVID-19 and the Expo2020 datasets. It classifies the input tweets into four phases: (i) collecting data, (ii) preprocessing of the tweets, (iii) feature extraction, (iv) proposed hybrid classification approach. Our hybrid-based approach is proposed to address the following challenges: improving accuracy, identifying the polarity of comparative sentences, distinguishing the intensity of opinion words, considering negative comments, and handling Sarcasm. Results demonstrate the efficacy of the suggested approach based on the accuracy and class distribution of each dataset. The approach was compared with other classifiers - BFTAN, TAN, NB, SVM and RF. Accuracy, precision and recall were computed for all the considered datasets. When enough data is available to support training examples, Bayesian classifier shows effective performance. Dataset 1 provides sufficient data to support training samples, so in this case, SVMBFTAN shows more accurate results compared to others. In contrast, in dataset 2, RF shows competitive accuracy values. Our method demonstrates higher accuracy values compared to previous methods, although an improvement in time is desired. However it might be interesting to see how the performance of our proposed approach could change for both in terms of solution quality and time by using many other techniques in different manner or using high-performance computing (HPC) devices for better results. Future work will therefore explore more accurate and flexible techniques by introducing some feature selection methods, preprocessing techniques for data and dealing with the hashtag-level sentiment analysis for the classification of tweets.
REFERENCES

[1] S. Valenzuela, S. Puente, and P. M. Flores, “Comparing disaster news on Twitter and television: An intermediary agenda setting perspective,” J. Broadcast. Electron. Media, vol. 61, no. 4, pp. 615–637, Oct. 2017.

[2] D. Diamantini, A. Mircoli, D. Potena, and E. Storti, “Social information discovery enhanced by sentiment analysis techniques,” Future Gener. Comput. Syst., vol. 95, pp. 816–828, Jan. 2019.

[3] N. Oztürk and S. Ayvaz, “Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis,” Telematik Informat., vol. 35, no. 1, pp. 136–147, Apr. 2018.

[4] D. Danneman and R. Heimann, Social Media Mining With R, U.K.: Packt Publishing Ltd, 2014.

[5] A. C. Pandey, D. S. Rajpoot, and M. Sarawat, “Twitter sentiment analysis using hybrid cuckoo search method,” Inf. Process. Manage., vol. 53, no. 4, pp. 764–779, 2017.

[6] E. S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, O. S. Siordia, and E. A. Villaseñor, “A case study of Spanish text transformations for Twitter sentiment analysis,” Expert Syst. Appl., vol. 81, pp. 457–471, Sep. 2017.

[7] G. A. Ruiz, P. A. Henriquez, 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.

[8] P. Chen, Z. Sun, L. Bing, and W. Yang, “Recent attention network on memory for aspect sentiment analysis,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2017, pp. 452–461.

[9] K. E. Weick and K. M. Sutcliffe, Managing the Unexpected: Sustained performance in a Complex World, Hoboken, NJ, USA: Wiley, 2015.

[10] C. Caragea, A. C. Squicciariini, S. Stehle, K. Nepalli, and A. H. Tapia, “Mapping moods: Geo-mapped sentiment analysis during hurricane sandy,” in Proc. ISCRAM, 2014, pp. 132–138.

[11] T. Sakaki, M. Okazaki, and Y. Matsuou, “Tweet analysis for real-time event detection and earthquake reporting system development,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2012.

[12] N. Pourerebriah, S. Sultana, J. Edwards, A. Gochanour, and S. Mohanty, “Understanding communication dynamics on Twitter during natural disasters: A case study of hurricane Sandy,” Int. J. Disaster Risk Reduction, vol. 37, Jul. 2019, Art. no. 101176.

[13] E. Abedin, H. Jafarzadeh, and S. Akhlaghpour, “Opinion mining on Twitter: A sentiment analysis of the Iran deal,” in Proc. PACIS, 2018, pp. 220.

[14] P. A. Henriquez and G. A. Ruiz, “Twitter sentiment classification based on deep random vector functional link,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–6.

[15] J. R. Ragni and P. M. R. Anand. “An empirical analysis and classification of crisis related tweets,” in Proc. IEEE Int. Conf. Comput. Intell. Res. (ICCIR), Dec. 2016, pp. 1–4.

[16] M. J. Canon, A. Satuito, and C. Sy, “Determining disaster risk management priorities through a neural network-based text classifier,” in Proc. Int. Symp. Comput., Consum. Control (IS3C), 2009.

[17] M. Al-Ayyoub, S. B. Essa, and I. Alsmadi, “Lexicon-based sentiment analysis of Arabic tweets,” Int. J. Social Netw. Mining, vol. 2, no. 2, pp. 101–114, 2015.

[18] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, “Topic sentiment analysis in Twitter: A graph-based hashtag sentiment classification approach,” in Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2011, pp. 1031–1040.

[19] Ankit and N. Saleena, “An ensemble classification system for Twitter sentiment analysis,” Proc. Comput. Sci., vol. 132, pp. 937–946, 2018.

[20] P. Mcfiville, W. Gryc, and R. Lawrence, “Sentiment analysis of blogs by combining lexical knowledge with text classification,” in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2009, pp. 1275–1284.

[21] H. Ghorbel and D. Jacot, “Further experiments in sentiment analysis of French movie reviews,” in Advances in Intelligent Web Mastering—3, vol. No. E. Mugellini, P. S. Szczepaniak, M. C. Pettenati, and M. Sokho, Eds. Berlin, Germany: Springer; 2011, pp. 19–28, doi: 10.1007/978-3-642-18029-3_3.

[22] V. K. Singh, R. Piryani, A. Uddin, P. Waila, and Marisha, “Sentiment analysis of textual reviews; evaluating machine learning, unsupervised and SemiWordNet approaches,” in Proc. 5th Int. Conf. Knowl. Smart Technol. (KST), Jan. 2013, pp. 122–127.

[23] S. Tan and J. Zhang, “An empirical study of sentiment analysis for Chinese documents,” Expert Syst. Appl., vol. 34, no. 4, pp. 2622–2629, 2008.

[24] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera, “Opinion mining and sentiment analysis on a Twitter data stream,” in Proc. Int. Conf. Adv. ICT Emerg. Regions (ICTer), Dec. 2012, pp. 182–188.

[25] E. Boiy and M.-F. Moens, “A machine learning approach to sentiment analysis in multimedia web texts,” Int. Retr., vol. 12, no. 5, pp. 526–558, 2009.

[26] J. R. Alharbi and W. S. Alhalabi, “Hybrid approach for sentiment analysis of Twitter posts using a dictionary-based approach and fuzzy logic methods,” Int. J. Semantic Web Inf. Syst., vol. 16, no. 1, pp. 116–145, Jan. 2020.

[27] M. Grčar, D. Cherepnalkoski, I. Mozetič, and P. K. Novak, “Stance and influence of Twitter users regarding the brexit referendum,” Comput. Social Netw., vol. 4, no. 1, pp. 1–25, Dec. 2017.

[28] S. V. Wavre and S. N. Deshmukh, “Sentiment classification using machine learning techniques,” Int. J. Sci. Res., vol. 5, no. 4, pp. 819–821, 2016.

[29] A. Feizollah, S. Ainin, N. B. Anuar, N. A. B. Abdullah, and M. Hazim, “Halal products on Twitter: Data extraction and sentiment analysis using stack of deep learning algorithms,” IEEE Access, vol. 7, pp. 83354–83362, 2019.

[30] A. G. Pablos, M. Cuadros, and G. Rigau, “A comparison of domain-based word polarity estimation using different word embeddings,” in Proc. 16th Int. Conf. Lang. Resour. (ICLR), Dec. 2012, pp. 182–188.

[31] M. Bordoloi and S. K. Biswas, “Machine learning based sentiment analysis using graph based approach,” in Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCTN), Jul. 2019, pp. 1–5.

[32] H. E. Kyburg and J. Pearl, “Probabilistic reasoning in intelligent systems: Networks of plausible inference,” J. Philosoph. vol. 88, no. 8, p. 434, Aug. 1991.

[33] R. I. Ogie, R. J. Clarke, H. Forehead, and P. Perez, “Crowdsourced social media data for disaster management: Lessons from the PetaJakarta.Org project,” Comput., Environ. Urban Syst., vol. 73, pp. 108–117, Jan. 2019.