

Received January 3, 2022, accepted January 16, 2022, date of publication January 19, 2022, date of current version January 28, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3144659

Inquest of Current Situation in Afghanistan Under Taliban Rule Using Sentiment Analysis and Volume Analysis

ERNESTO LEE¹, FURQAN RUSTAM², IMRAN ASHRAF³,
PATRICK BERNARD WASHINGTON⁴, MANIDEEP NARRA⁵, AND RAHMAN SHAFIQUE²

¹Department of Computer Science, Broward College, Broward County, Fort Lauderdale, FL 33301, USA

²Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan

³Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea

⁴Division of Business Administration and Economics, Morehouse College, Atlanta, GA 30314, USA

⁵Indiana Institute of Technology, Fort Wayne, IN 46803, USA

Corresponding authors: Imran Ashraf (ashrafimran@live.com) and Furqan Rustam (furqan.rustam1@gmail.com)

This work was supported by the Florida Center for Advanced Analytics and Data Science funded by Ernesto.Net (under the Algorithms for Good Grant).

ABSTRACT Microblogging websites and social media platforms serve as a potential source for mining public opinions and sentiments on a variety of subjects including the prevailing situations in war-afflicted countries. In particular, Twitter has a large number of geotagged tweets that make the analysis of sentiments across time and space possible. This study performs volume analysis and sentiment analysis using LDA (Latent Dirichlet Allocation) and text mining over two datasets collected for different periods. To increase the adequacy and efficacy of the sentiment analysis, a hybrid feature engineering approach is proposed that elevates the performance of machine learning models. Geotagged tweets are used for volume analysis indicating that the highest number of tweets is originated from India, the US, the UK, Pakistan, and Afghanistan. Analysis of positive and negative tweets reveals that negative tweets are mostly originated from India and the US. On the contrary, positive tweets belong to Pakistan and Afghanistan. LDA is used for topic modeling on two datasets containing tweets about the current situation after the Taliban take control of Afghanistan. Topics extracted through LDA suggest that majority of the Afghanistan people seem satisfied with the Taliban's takeover while the topics from negative tweets reveal that issues discussed in negative tweets are related to the US concerns in Afghanistan. Sentiment analysis over two different datasets indicates that the trend of the sentiments has been shifted positively over three weeks.

INDEX TERMS Topic modeling, sentiment analysis, Taliban regime, latent Dirichlet allocation, machine learning.

I. INTRODUCTION

The current situation in Afghanistan is crucial after President Joe Biden announced the end of the war in Afghanistan and the withdrawal of allied military forces [1]. Taliban, which continued to capture and contest territory across the country despite ongoing peace talks with the Afghan government, ramped up attacks on ANDSF (Afghan National Defense and Security Forces) bases and outposts, began to rapidly seize more territory [2]. The U.S withdraw its army in August 2021 and the Taliban take control of 90% of Afghanistan [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang¹.

On 15th August Taliban entered and captured Kabul, the capital of Afghanistan, and take its full control. After this victory, they announced to make a new government comprising representatives from all Taliban groups. Although sooner or later, Taliban takeover was inevitable, the world is surprised at the pace ANDSF lost control of major cities. Several speculations and theories have been discussed over the news and social media over the past several days [3], [4]. In addition, different conjectures and contemplations are made by human rights activists and organizations about the change in social norms especially regarding the status of women in Taliban-controlled Afghanistan. Different opinions, reviews, and impressions are posted online, both by common people,

as well as, governments, politicians, and religious leaders about the current situation in Afghanistan. Such reviews represent significant potential to analyze the thought of people about and draw current situation in Afghanistan. Giving a mixed attitude with positive and negative comments, opinions of Afghanistan people are also very important, both living in Afghanistan and abroad.

Social networks have become an inevitable and integrated part of daily lives and hold an inexorable place in daily activities for many people [5]. People feel urged to freely express their opinions and feelings on social media platforms without restrictions of race, color, ethnicity, and location. Especially, Twitter has become one of the most widely used social media platforms where people express their thoughts, opinions on different topics, ideas, and personalities [6]. Such opinions and thoughts are posted in the form of tweets which are further shared on Facebook and other similar platforms. These tweets are useful to extract people's sentiments on Afghanistan's current situation. Besides personal opinions, such tweets indicate several problems related to the social and cultural situation. Sentiment analysis possesses the ability to analyze these tweets to determine the sentiments of people. In addition, topic modeling can be used to extract the most discussed topics and issues from tweets. Once the sentiments are categorized as positive and negative, the issues and opportunities can be obtained from negative and positive sentiment containing tweets, respectively. Topic modeling is an attractive approach that helps in extracting important topics from the given document (s). It helps in determining the importance of a topic by the frequency of its occurrence and its relation to other topics. This study investigates one of the challenging topics in current international affairs, Taliban rule in Afghanistan after US withdrawal, through the opinions of the public in tweets. In this regard, this study proposes an approach for sentiment analysis using machine learning techniques on Twitter data containing tweets related to Afghanistan. Following contributions are made in the current study

- Two datasets are collected containing tweets related to Afghanistan's current situation which are used for sentiment analysis. Datasets are collected for two time spans to perform a prospective analysis.
- For initial sentiment analysis, HFE (Hybrid Feature Engineering) approach is devised to increase the accuracy of sentiment analysis. To avoid the labor of data labeling, TextBlob is used for automatic annotation. Several machine learning models are used with TF-IDF (Term Frequency-Inverse Document Frequency), BoW (Bag of Words), and HFE for sentiment analysis.
- Topic extraction is performed using LDA (Latent Dirichlet Allocation). The analysis is performed for change in the sentiments over time considering the collected datasets. Extracted topics are analyzed in terms of their importance and sentiments are discussed with respect to geolocation.

The rest of the paper is structured as follows. Section II discusses important research works related to the current study. The proposed research methodology and its related contents are presented in Section III. Section IV provides the discussion of results. In the end, the conclusion and future work are given in Section V and Section VI, respectively.

II. RELATED WORK

During the last few years sentiment analysis has received a wide interest due to the popularity of social media platforms like Twitter, Facebook, etc. Additionally, the availability of a large amount of data in the form of tweets, reviews, and comments accelerated its pace. Consequently, a large body of work exists for sentiment analysis and topic modeling. Study [7] uses statistical approaches to analyze the effectiveness of fake review systems in the sentimental analysis domain. Different machine learning models are evaluated to investigate their accuracy for the task at hand. Fake news is detected based on sentiments like positive and negative reviews via the proposed system. NB (Naive Bayes), KNN (K Nearest Neighbors), K-means Clustering, SVM (Support Vector Machines), and DT-J48 algorithms are evaluated in the study. Results indicate that SVM performs best among the investigated models.

Due to the potential of topic modeling and sentiment analysis to extract important aspects of specific topics, several works use topic modeling approaches such as LDA, LSA (Latent Semantic Analysis), and PLSA (Probabilistic Latent Semantic Analysis), etc. along with the sentiments analysis. For example, [8] uses LDA with VADER (Valence Aware Dictionary and sEntiment Reasoner) to analyze important topics from global climate change tweets. Diverse topics on global climate changes are extracted and analyzed across time and space. The influence of change in policy towards climate change is investigated using LDA and sentiment analysis. Similarly, the authors perform topic modeling and sentiment analysis in [9] using big data analysis to identify the needs of customers that travel using Asian airlines. Owing to the increasing market size of Asian airlines, such analysis aims at increasing customers' satisfaction through meeting their in-flight needs. Online reviews by customers are utilized to analyze several important topics such as in-flight meals, entertainment, seat class, seat comfort, and staff service, etc. Feature extraction and feature selection techniques are also an important part of sentiment analysis. Feature selection plays a prominent role to enhance text classification accuracy and many research works investigate the efficiency of various feature extraction approaches. For example, study [10] proposed a summarization-based approach for feature reduction to improve the model's performance. The selected features are used to train the k nearest neighbor (KNN) classifier. Another study [11] analyzed feature extraction, feature reduction, and feature selection to improve the model's performance for text classification. They also used a KNN classifier which performs well with information gain and mutual information feature selection techniques.

TABLE 1. Sample tweets from the collected dataset.

Username	Location	Text	Date
maccaburbsvoice	Melbourne	The #Taliban are extremely well organised. https://t.co/VG9ulBooE6	8/13/2021
YoussefDisla	Minneapolis, MN	Let the #Taliban have #Afghanistan	8/13/2021
Asitseemstome		Taliban surge across Afghanistan continues as four more cities fall #Afghanistan #Taliban https://t.co/F4JcANimCB	8/13/2021



FIGURE 1. Location distribution of collected tweets.

Topic modeling and sentiment analysis are important, both from the perspective of individual customers, as well as, analyzing the public opinion about ongoing political conflicts at local and international levels. For example, [12] conducts sentiment analysis of tweets during primary debates during 2016-16 to uncover people’s emotions about Donald Trump. Of the tweets which are labeled as positive or negative, joy is found to be the dominant emotion from the tweets that supported Trump. Unlike, sentiment analysis for local governments where the scope is narrow, analysis of tweets about international conflicts and events holds large scope and depicts the thoughts of people worldwide. For example, the scope of tweets about, the US banning Huawei, the Iran-US nuclear deal, and peace talks between North Korea and the US, is much broader than the sentiment analysis of products and services. Several research works can be found in the literature that focuses on such international conflicts and events. For example, Wael *et al.*, [13] use sentiment analysis to carry out bias detection of Palestinian-Israeli conflict in Western media. This process involves finding misleading terms, vocabularies, and expressions used to shape the opinion when disseminating news about the Israel-Palestine war. Similarly, sentiment analysis is performed on tweets about post-conflict in Columbia using a crawled dataset from Twitter in [14]. The analysis is performed using tweets from Colombia residing Colombians and foreigners. Tweets from foreigners are considered to study public perception. Results of the research indicate that foreigners tend to show a more positive attitude and feelings compared to local people.

The authors investigate public opinions about Iran-US nuclear deal in [15]. Through the deal, Iran is restricted to stop its conflicting program, and the US, European countries, and China agreed to lift economic sanctions in exchange. The lexicon-based approach is utilized using machine learning models to analyze the opinions about the deal from people residing in different countries. The majority of the sentiments found in the collected tweets are found to be positive. Similarly, [16] investigates public sentiments about the US banning Huawei through tweets posted during the banned period. Analysis using lexicon-based approach indicates the majority of the people remain neutral about the US-China trade war, followed by those who oppose it. Besides, US and China being the countries involved in the trade war, large volumes of tweets are generated from Canada, the United Kingdom, India, Pakistan, and South Africa. Another work analyzing sentiments about China-US trade war is [17] which presents a real-time monitoring and reporting framework for public opinions. The data is obtained from 52 economic leaders and the China-US trade war is used as a case study to study the performance of the proposed framework. Results indicate that CNN (Convolutional Neural Network) tends to produce better results for real-time monitoring.

Topic modeling and sentiment analysis can play a significantly important role to find the prevalent issues and challenges triggered from local and international events and help the authorities to take corrective actions to mitigate their influence. For example, [18] discusses the problem of Syrian refugees by analyzing the tweets in Turkish and

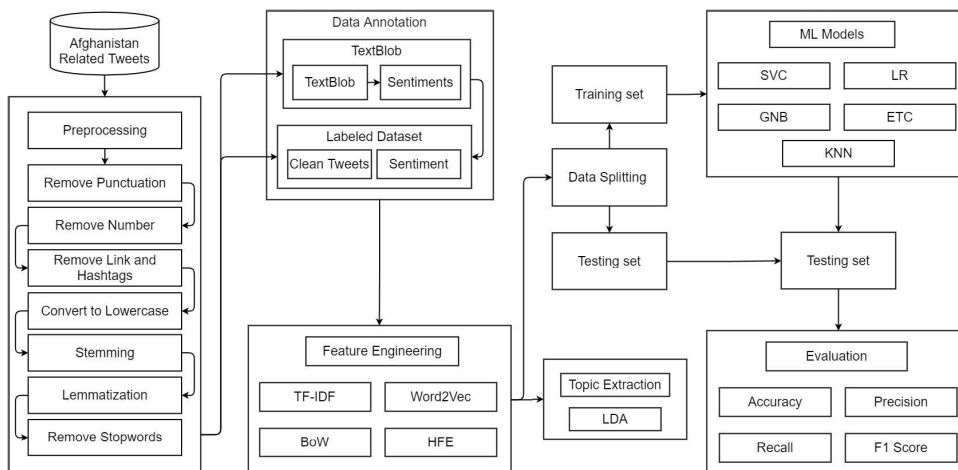


FIGURE 3. Architecture of proposed methodology for sentiment analysis.

are removed from the dataset to reduce the complexity. Tags and links are removed using regular expressions.

- Punctuation and number removal:** Punctuation is a necessary part of the sentence to make it more intelligible and meaningful for human readers. However, for machine learning models, punctuation creates extra complication and increase the feature vector size. Keeping in view the fact that punctuation does not contribute to the training process, they are removed from tweets.
- Convert to lowercase:** Machine learning models are case sensitive so the difference in the case cause complexity in models training as the ‘War’, ‘war’, and ‘WAR’ are considered as different words. It increases the feature space size as well as, the processing time. Conversion to lower case converts each character to its lower case and Python built-in function is used for that purpose.
- Stemming and lemmatization:** Stemming converts the words to their root forms such as ‘goes’, ‘going’, ‘gone’ are variations of the ‘go’. If not handled properly, machine learning models take them as different words, so stemming is performed to convert them all to their base form ‘go’ [25]. Although Lemmatization is similar to stemming, it is often more effective. Stemming removes a few characters which may lead to erroneous terms. On the contrary, lemmatization considers the context of the sentence and transforms the words into their proper base form. This study carries out stemming and lemmatization using Porter’s stemmer and WordNetLemmatizer.
- Stopwords removal:** Stopwords are an important part of the sentence to increase readability and meaningfulness. These are short and meaningless words such as the, in, he, it, an, etc., that do not contribute to the training of the machine learning models. Consequently, stopwords are removed to reduce the complexity of feature space.

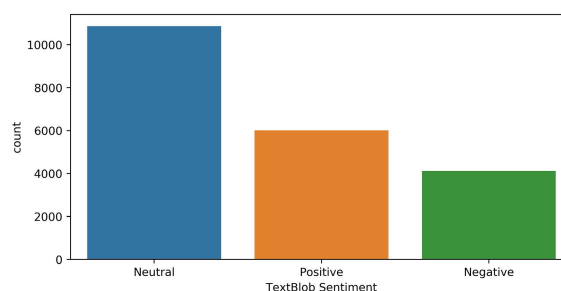


FIGURE 4. Distribution of positive, negative and neutral tweets.

Results of preprocessing steps applied on sample tweets from the collected dataset are shown in Table 2.

1) TEXTBLOB

After preprocessing of data, the dataset is annotated with the negative, positive and neutral sentiments using the Textblob library [26]. Textblob library finds the polarity score for each tweet to be annotated. Table 3 shows the score values that are used to assign sentiments for tweets.

Textblob is one of the most widely used lexicon-based techniques for sentiment scoring [26]. It finds the polarity score from the given text that ranges between -1 to 1. This study uses the Textblob technique to annotate the tweet dataset into negative, positive, and neutral tweets. These sentiments are used in this study as a target class to train the machine learning models. Textblob is used because it gives more correlated sentiments corresponding to text features. Using the TextBlob, the count of tweets with each score is given in Figure 4.

After data annotation, the dataset is split into training and testing datasets for models’ training and testing, respectively. Train-test split is 85:15 where 85% of the data are used for training while 15% is used for testing. The number of tweets after data split is given in Table 4.

TABLE 2. Results of using preprocessing steps.

Before Preprocessing	After Preprocessing
The #Taliban are extremely well organized. https://t.co/VG9ulBooE6	taliban extremely well organize
Taliban surge across Afghanistan continues as four more cities fall #Afghanistan #Taliban https://t.co/F4JcANimCB	taliban surge across afghanistan continue four more city fall afghanistan taliban

TABLE 3. Sentiment assignment using TextBlob.

Score	Sentiment
Polarity score >0	Positive
Polarity score =0	Neutral
Polarity score <0	Negative

TABLE 4. Number of tweets in training and testing sets.

Sentiment	Training	Testing	Total
Positive	5,086	925	6,011
Negative	3,518	605	4,123
Neutral	9,246	1,620	10,866
Total	17,850	3,150	21,000

2) FEATURES ENGINEERING

Once the data annotation is done, feature extraction approaches are applied. Feature extraction implies that we need to transform the text data into a numeric form that can be used to train machine learning models.

This study uses three well-known feature extraction techniques as TF-IDF, BoW, and Word2Vec. Additionally, Chi-square (Chi2) is used as a feature selection approach.

Bag of Words: BoW is one of the simplest feature extraction techniques used to extract features from text data [24]. It is an easy-to-implement and easy to interpret approach. Despite being simple, it often produces competitive results as compared to complex feature extraction approaches. It counts the frequency of each unique term in the corpus and makes a numerical feature vector, thereby reducing the complexity of the model’s training. The Sci-kit library is used to implement the BoW.

Term Frequency - Inverse Document Frequency: Originally introduced by Salton, TF-IDF is one of the most widely used feature extraction techniques for text analysis [27]. It gives weighted features as compared to simple count features by BoW. It computes the weight of each term in the corpus and these weighted features are more efficient in the training of models. TF-IDF comprises TF and IDF where the former counts the occurrence of each unique term while the latter assigns weights to each term based on its appearance in different documents. TF can be computed using (1) [28]

$$TF = \frac{t_D}{D_t} \tag{1}$$

where t_D is the number of times the term t appears in a document D and D_t is the number of terms in a document D .

IDF can be computed as (2) [28]

$$IDF = \log\left(\frac{N}{N_{D,t}}\right) \tag{2}$$

where N is the number of documents and $N_{D,t}$ is the number of documents containing the term t . So TF-IDF can be computed as

$$TF - IDF = TF * IDF \tag{3}$$

Word2Vec: Word2Vec is a feature extraction approach developed by the Google [29]. It is different from BoW and TF-IDF as they use the repetition of the term in a corpus while word2vec utilizes the similarity between words. It uses cosine similarity to find the similarity between the words. It is found to work better on both large and small datasets. Word2Vec is implemented using the Genism library.

3) HYBRID FEATURE ENGINEERING

HFE is a combination of feature extraction techniques T-IDF, BoW, and Word2Vec using feature selection technique Chi2. Chi2 is a feature selection technique that is used in this study to select the best features [23], once TF-IDF, BoW, and Word2Vec are applied to the preprocessed data. Chi2 selects the best features by finding the independence between the features’ variables [30]. The Chi-Square statistics are adjusted by the degree of freedom which varies with the number of levels of variables and the number of levels of the class variable. We used the Chi2 technique in our proposed hybrid feature engineering technique to select the best features.

Algorithm 1 Hybrid Feature Engineering.

Input: n, m .

- 1: $TF - IDF_F (n \times m) = TF - IDF(corpus)$
- 2: $BoW_F (n \times m) = BoW(corpus)$
- 3: $Word2Vec_F (n \times m) = Word2Vec(corpus)$
- 4: $TF - IDF_{BF(n \times p)} = Chi2(TF - IDF_F(n \times m))$
- 5: $BoW_{BF(n \times p)} = Chi2(BoW_F(n \times m))$
- 6: $Word2Vec_{BF(n \times p)} = Chi2(n \times m)$
- 7: $HEF_{(n \times q)} = Concat(TF - IDF_{BF(n \times p)}, BoW_{BF(n \times p)}, Word2Vec_{BF(n \times p)})$

In HFE, first, features are extracted from the text data using each feature extraction technique followed by the selection of 2,000 features from each extracted feature set. These 2000 features from each technique are concatenated to make the new hybrid feature set which has more diversity and less

complexity to boost the performance of the machine learning models. Algorithm 1 shows the steps followed in HEF

where n is the number of tweets and m is the number of features extracted using each technique, i.e., TF-IDF, BoW, and Word2Vec. $TF - IDF_F$, BoW_F , and $Word2Vec_F$ show TF-IDF features, BoW features, and Word2Vec features from the tweet dataset, respectively.

Chi2 performs feature selection using the following procedure. The p is the number of selected features by the Chi2 and BF is the best features. We combine these best features to propose a new feature set. $HEF_{(n \times q)}$ are the hybrid features used in the proposed approach to obtain higher accuracy. The HFE feature set is small in size as compared to TF-IDF, BoW and Word2Vec features sets and reduces the feature space complexity. The number of features with each technique is shown in Table 5 and the illustration of HFE is shown in Figure 5.

TABLE 5. Number of features using different techniques.

Technique	Features
BoW	23,993
TF-IDF	23,993
Word2Vec	23,993
HFE	6000

4) MACHINE LEARNING MODELS

This study used five machine learning models to perform the sentiment analysis aiming to select the one with the best performance. The selected models are optimized by fine-tuning several hyperparameters. Value range is used to analyze the performance of a model with a different set of parameters. From the literature review, different text classification models are selected along with their best-performing hyperparameters. For performance optimization, value ranges found in the literature are used on the collected dataset to analyze models' performance. For example, in the case of the LR model, the solver parameter is used with 'saga', 'sag'. Although LR has a third parameter for the solver, i.e., 'liblinear', however, 'liblinear' is not appropriate for multi-class problems. Similarly, for other models, value ranges are selected where possible to find the set of the best parameters to obtain high classification accuracy. For the ETC model, we found that if a feature set is large, the max depth of each decision tree has to be increased to obtain better results, so, we define the range of the 'max_depth' parameter between 50 to 500 to analyze its performance. The selected hyperparameters setting and

TABLE 6. Hyperparameter settings for machine learning models.

Models	Hyperparameters
LR	solver= "saga", C=2.0, multi_class = "multinomial"
ETC	n_estimators=300, max_depth= 300
SVC	Kernel= "linear", C= 2.0, random_state= 27
GNB	Default setting
KNN	n_neighbour= 5

TABLE 7. Hyperparameter values tuning range.

Models	Hyperparameters
LR	solver= {saga, sag, liblinear}, C={1.0 to 5.0}, multi_class = {multinomial, ovr}
ETC	n_estimators={50 to 500}, max_depth= {50 to 500}
SVC	Kernel= {linear, sigmoid, poly}, C={1.0 to 5.0}, random_state= {1 to 50}
GNB	Default setting
KNN	n_neighbour= {2 to 10}

their corresponding values are given in Table 6 and the tuning range for each hyperparameter is shown in Table 7.

A comprehensive description of the machine learning models is provided in Table 8.

IV. RESULTS AND DISCUSSIONS

Performance evaluation of the trained models is realized using accuracy, precision, recall, F1 score, and confusion matrix. Results of models are discussed with respect to each feature extraction technique such as BoW, TF-IDF, Word2Vec, and HFE. The performance comparison between machine learning and deep learning models is conducted as well. Similarly, the performance of the proposed approach is compared with state-of-the-art approaches.

A. MODELS' PERFORMANCE USING BoW FEATURES

Initially, the experiments are performed using BoW features and Table 9 shows the experimental results. Results indicate that tree-based models and linear models perform significantly better in terms of precision, recall, and F1 score. SVM outperforms all models by achieving the highest accuracy, precision, recall, and F1 score of 0.95, 0.94, 0.93, and 0.94, respectively. Performance of LR and RF is marginally lower with 0.93 and 0.92 accuracy scores, respectively. Owing to the fact that the dataset has a large feature set, the performance of the linear models is considerably better than other models. On the other hand, GNB and KNN did not perform well. GNB assumes that data follows a Gaussian distribution and the accuracy is affected when it is not the case with the data distribution under different labels.

The confusion matrix in Figure 6 shows the number of correct and wrong predictions by each model using the BoW features. SVM gives 2,981 correct predictions and 169 wrong predictions out of 3,150 total predictions. LR gives 2,922 correct and 228 wrong predictions while RF gives 2,902 correct predictions and 248 wrong predictions. The worst performer GNB gives only 1,423 correct predictions and 1,727 wrong predictions. 6.

B. RESULTS OF MACHINE LEARNING MODELS USING TF-IDF FEATURES

Experimental results for machine learning models using TF-IDF features are provided in Table 10. Results suggest that SVM performs the best even when used with TF-IDF features and obtains scores of 0.93, 0.93, 0.90, and 0.91 for accuracy, precision, recall, and F1 score, followed by ETC

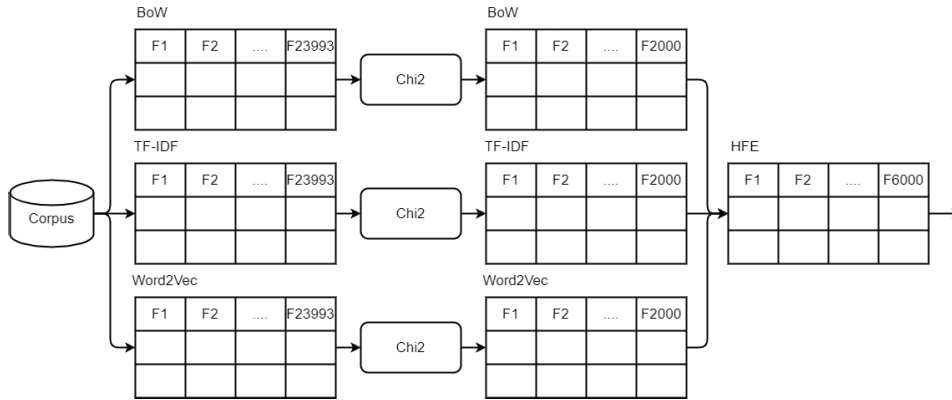


FIGURE 5. Methodology adopted for hybrid feature engineering.

TABLE 8. Description of machine learning models used in the study.

Model	Description
LR	LR is a linear model widely used for classification and makes use of the logistic function to separate the data and is named logistic regression [31]. This logistic function is also known as the Sigmoid function. It is an S-shaped function and can map real values between 0 and 1 space. The logistic function can be defined as: $\frac{1}{(1+e^{-a})}$ where e is the Euler's number and a is the actual numerical value that is to be transformed. LR equation can be defined as $y = \frac{e^{(y_0+y_1*x)}}{(1+e^{(y_0+y_1*x)})}$ where y is the predicted output, y_0 is the bias or intercept term and y_1 is the coefficient for the single input value (x). LR can be a good for sentiment analysis and performs better with large feature space.
ETC	ETC is a tree-based ensemble model used for text classification. Similar to the random forest, it predicts using the majority voting criteria. It generates the number of decision trees on the original dataset and performs voting between decision trees predictions to make the final prediction [25]. Mathematically ETC can be defined as $ETC = \text{mod } \sum_{i=1}^N t_i$ where t_i belongs to t_1, t_2, \dots, t_n which are trees in ETC, and M is the number of decision trees. ETC can be a good choice for the current approach because it can perform better even when the dataset is not linearly separable.
SVM	SVM is a linear model widely used for data classification. It utilizes hyperplanes to classify data. In feature space, SVM draws the hyperplanes with a good margin from each class to give significant results [32]. It performs well on both multi-class and binary class classification. SVM can be used with different kernels, and this study utilizes the linear kernel which makes it more efficient to perform sentiment analysis. SVM tends to show better performance for text classification tasks when the data has a large feature space.
GNB	GNB is a type of Naïve Bayes model that follows Gaussian normal distribution and performs well on the continuous data [33]. Naive Bayes models are based on the Bayes theorem used for the classification of data. GNB can be a good choice when working with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution.
KNN	KNN also known as a lazy learner is used for classification and regression [34]. In KNN, the K indicates the number of nearest neighbors considered to make a decision. It is very simple and easy to interpret as compared to other machine learning models and often shows competitive results. During training, it finds the similarity between features based on their distance with neighbors. During prediction, it compares the test data with training data and predicts the target class using a distance measure such as Euclidean, Manhattan, etc.

TABLE 9. Performance of models using BoW features.

Model	Accuracy	Precision	Recall	F1 score
SVM	0.95	0.94	0.93	0.94
LR	0.93	0.94	0.90	0.92
GNB	0.45	0.53	0.52	0.45
ETC	0.92	0.93	0.89	0.91
KNN	0.57	0.77	0.42	0.40

TABLE 10. Performance of models using TF-IDF features.

Model	Accuracy	Precision	Recall	F1 score
SVM	0.93	0.93	0.90	0.91
LR	0.90	0.92	0.87	0.89
GNB	0.42	0.46	0.47	0.42
ETC	0.92	0.93	0.88	0.90
KNN	0.66	0.68	0.58	0.60

with a 0.92 accuracy. GNB shows poor performance with TF-IDF features as well. Furthermore, the performance of the machine learning models is degraded when used with TF-IDF features. These results conform with [35] which shows that preprocessed data using BoW yields better classification accuracy for sentiments.

Figure 7 shows the confusion matrix for machine learning models using the TF-IDF features. SVM gives 2,915 correct predication and 235 wrong predictions using the TF-IDF

features. ETC model is the second-best performer and predicts 2,884 correctly while 266 predictions are wrong. These statistics show that the model's performance on BoW features is more significant as compared to the model's performance with TF-IDF features. Primarily, the features using the BoW technique are simple as compared to features using TF-IDF, and have a high probability of existence in the tweets. On the other hand, TF-IDF assigns higher weights to rare terms, and such terms are not common in tweets which makes the feature

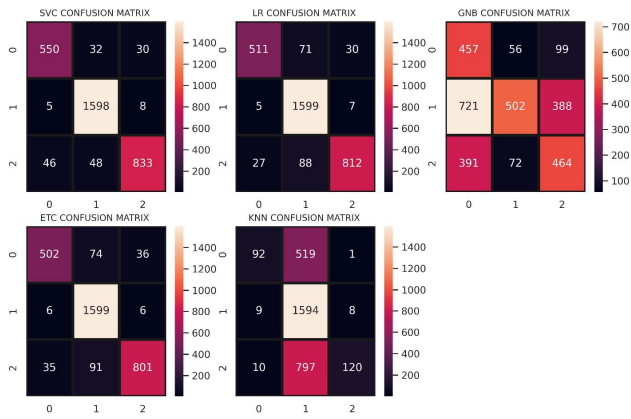


FIGURE 6. Confusion matrix for classifiers using BoW features.

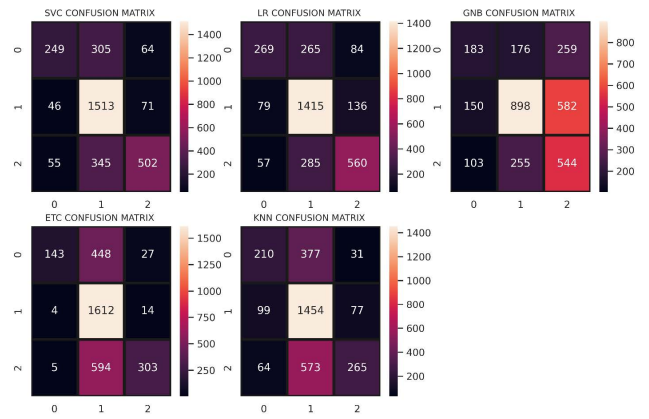


FIGURE 8. Confusion matrix for classifiers using Word2Vec features.

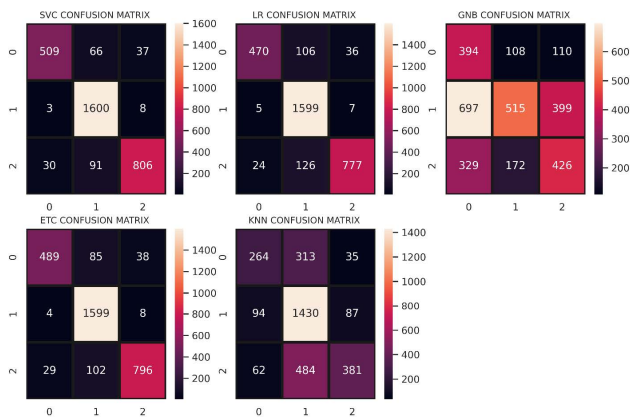


FIGURE 7. Confusion matrix for classifiers using TF-IDF features.

set complex. So models get less complexity on BoW features and perform better.

C. MODELS' PERFORMANCE USING WORD2VEC FEATURES

The performance of models using the word2vec features is shown in Table 11. Results reveal that the classification performance is substantially degraded when models are trained on Word2Vec features. For example, the accuracy of the best performing model SVM is 0.72 which was 0.93 with BoW and 0.95 with TF-IDF. Similarly, the accuracy of LR has been reduced to 0.71 when used with Word2Vec from 0.90 with TF-IDF. On the other hand, the performance of GNB has been elevated when used with Word2Vec features and it reaches to 0.52 accuracy score which was 0.45 and 0.42 with BoW

TABLE 11. Performance of models using Word2Vec features.

Model	Accuracy	Precision	Recall	F1 score
SVM	0.72	0.73	0.63	0.65
LR	0.71	0.70	0.64	0.66
GNB	0.52	0.50	0.48	0.48
ETC	0.65	0.81	0.52	0.54
KNN	0.61	0.63	0.51	0.62

TABLE 12. Performance of models using Hybrid features.

Model	Accuracy	Precision	Recall	F1 score
SVM	0.97	0.96	0.96	0.96
LR	0.96	0.96	0.95	0.95
GNB	0.87	0.86	0.81	0.83
ETC	0.93	0.93	0.91	0.92
KNN	0.63	0.79	0.49	0.50

and TF-IDF, respectively. Word2Vec maps word embedding into feature space which becomes complex as the number of interrelated used terms increases. This feature complexity often tends to reduce the performance of machine learning models.

The confusion matrix for the machine learning models using the word2vec features is shown in Figure 9. The performance of models is low using the word2vec features as compared to TF-IDF and BoW and that is the reason the wrong prediction ratio is higher. SVM gives 2,264 correct predictions and 886 wrong predictions while LR, ETC give 2,244 and 2,058 correction predictions, respectively.

D. PERFORMANCE OF MODELS USING HYBRID FEATURES

The performance of models using the proposed HFE features is shown in Table 12. The performance of the models is significantly better with the proposed features as compared to BoW, TF-IDF, and Word2Vec.

SVM outperforms all models with the highest 0.97 accuracy while the precision, recall, and F scores for SVM are 0.96 each. Followed by SVM, the performance of LR is marginally low with a 0.96 accuracy score. The values for F1 scores and accuracy are in proximity indicating the good fit of these models. For the most part, the performance of the machine learning models has been better when used with the HFE features. Results suggest that GNB, the poor performer with BoW, TF-IDF, and Word2Vec, shows far better performance with HFE and reaches the accuracy score of 0.87 with good values for other performance evaluation metrics. Selection of important features through Chi2 and joining them to make hybrid features leads to a better fit of the models which produces better results in return

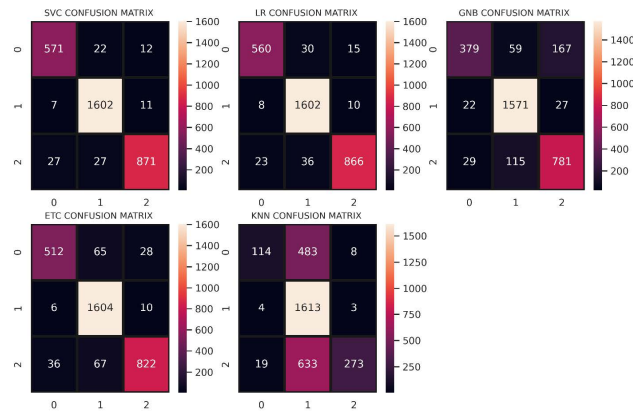


FIGURE 9. Confusion matrix for classifiers using hybrid features.



FIGURE 10. Accuracy comparison of machine learning models using BoW, TF-IDF, Word2Vec and hybrid features.

Confusion matrices in Figure 9 show the significance of HFE features as models achieved their best results when used with HFE features. According to the confusion matrix, SVM gives the highest correct predictions with 3,044 correct predictions and gives only 105 wrong predictions. LR also performs better with 3,028 correct predictions and 122 wrong predictions. GNB achieves its best accuracy with HFE features and gives 2,731 correct predictions and 419 wrong predictions. A comparison of models’ performance using each feature extraction technique is given in Figure 10.

E. RESULTS OF DEEP LEARNING APPROACHES

Together with machine learning models, deep learning models are also employed for the task at hand. For this purpose,

TABLE 13. Architecture and parameters used for deep learning models.

Layers	CNN-LSTM	LSTM	GRU
Layer 1	Embedding Layer (input_dim=5000, output_dim=100)	Embedding Layer (input_dim=5000, output_dim=100)	Embedding Layer (input_dim=5000, output_dim=100)
Layer 2	Conv1D (128, 5, activation = 'RELU')	Dropout (0.5)	Dropout (0.5)
Layer 3	MaxPooling1D (pool_size=4)	LSTM (100)	GRU (256)
Layer 4	LSTM(100)	Dense (3, activation = 'softmax')	SimpleRNN (128)
Layer 5	Dense (3, activation='softmax')		Dense(3, activation='softmax')
Compile		loss='categorical_crossentropy', optimizer='adam'	
Fit		Epochs=100, batch size=32, verbose =2	

three state of the arts models LSTM, CNN-LSTM [36], and GRU [37] are utilized in this study. These models are reported to show better performance for text classification in previous studies. A list of parameters and their associated values are provided in Table 13, as well as, the architecture of each layer of these models. Feature engineering is not required for deep learning models.

Table 14 shows the performance of deep learning models used in this study. Results suggest that both LSTM and CNN-LSTM obtain an accuracy of 0.92 each for sentiment classification while the classification accuracy of GRU is 0.87. It indicates that deep learning models are low-performing as compared to machine learning models. Predominantly, deep learning models perform better when trained with a large dataset. Using a small dataset, deep learning models tend to overfit and their accuracy is declined steadily [38], [39]. Conversely, the proposed HFE provides a set of important features to the machine learning models and produces better results.

F. DISCUSSION ON MACHINE LEARNING AND DEEP LEARNING MODELS

This study deploys state-of-the-art machine learning models including both linear and tree-based models with the hybrid feature engineering approach. Linear models such as SVM and LR perform well as compared to other models because of the large feature set. A combination of TF-IDF, BoW, and Word2Vec generates a large feature set that can be good for linear models. Consequently, SVM obtains the highest accuracy score as compared to other models. Although the performance of ETC is also good but compared with SVM or LR, it achieves lower accuracy. KNN is the worst performer in this study because of its lazy learning characteristics as it works well often when the feature set is small. In addition to the machine learning models, deep learning models are used as well which show comparatively lower performance. LSTM, GRU, and CNN-LSTM models have low performance on account of a smaller dataset. Deep learning models require a large dataset, as well as, large feature vector to obtain a good fit which is not possible with the current dataset. A critical analysis of machine and deep learning models is provided in Table 15 regarding the advantages and disadvantages in the context of this study. It includes the influence of the selected feature extraction approach, values of the parameter(s) and the structure, etc. on the overall performance.

TABLE 14. Performance of deep learning models used in the study.

Model	Accuracy	Precision	Recall	F1 score
LSTM	0.92	0.96	0.96	0.96
CNN-LSTM	0.92	0.96	0.95	0.96
GRU	0.87	0.87	0.77	0.88

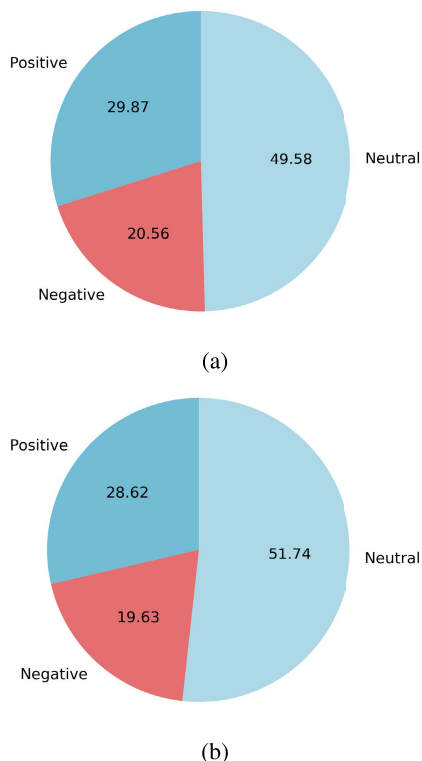


FIGURE 11. Comparing number of words and sentences in positive, negative and neutral tweets, (a) Words ratio and (b) Sentences ratio.

G. VOLUME ANALYSIS

Volume analysis aims at finding the information related to the volume of text/tweets which may be the number of total tweets, negative tweets, etc. Often it is used with the geolocation information to show the distribution of tweets concerning the location of the posted tweets.

For the collected dataset containing tweets about Afghanistan’s current situation, tweets are annotated with respect to the containing sentiments. For this purpose, the top five countries with the highest number of tweets are considered only which include India, the US, Pakistan, UK, and Afghanistan. The prime objective is to find the reaction of the people towards the current situation in Afghanistan under Taliban rule. It helps to determine the overall sentiment level of people from a specific country.

Figure 11 presents the comparison between the number of words and the number of sentences found in positive labeled tweets. The objective of this representation is to show the ratio of negative, positive, or neutral words in positive labeled tweets. The rationale for presenting this ratio is that even the positive tweets may contain negative and neutral words as the label is assigned based on the sentiment found in the majority

of words in tweets. So, even a positively labeled tweet may contain several negative and neutral words. Figure 11a shows that the majority of words used in positive labeled tweets is neutral. However, the number of positive words found in positive tweets is high as compared to negative words which are 29.87% and 20.56%, respectively. Similarly, Figure 11b shows the ratio of positive, negative, and neutral sentences in positive labeled tweets. Contrary to Figure 11a where the sentiment is determined for words, sentiment is given to each sentence for the data shown in Figure 11b. Then the ratio of sentences is determined, as shown in Figure 11b where neutral, positive, and negative sentences are 51.74%, 28.62%, and 19.63%, respectively.

Figures 12a and 12b show the distribution of negative and positive tweets, respectively for the top five countries with the highest number of tweets. The objective is to analyze the ratio of positive and negative tweets from different countries. It shows that 49% of all the negative tweets originate from India, followed by the US and the UK comprising 20% and 17%, respectively of the total negative tweets. The lowest number of negative tweets are originated from Pakistan and Afghanistan, respectively containing 5% and 9% of the total negative tweets. Similarly, for the positive tweets, Pakistan has 26% while Afghanistan has 23% of the total positive tweets. India contributes the lowest 15% of the positive tweets, followed by the UK with 17% of the total positive tweets.

Figures 11, 12 and 13 represent different aspects and characteristics of tweets collected for this study. Figure 11 shows the ratio for positive, negative, and neutral sentiments for words and sentences in the dataset exhibiting that even the positive annotated tweets contain several negative and neutral words and vice versa. Figure 12 indicates the ratio of positive and negative tweets for the top five counties from the collected dataset. The highest ratio of positive and the lowest ratio of negative tweets, is by Pakistan which indicates peoples’ support from Pakistan towards Afghanistan. On the other hand, the ratio for negative tweets is higher from India which highlights the trends of political interest. Figure 13 shows the ratio of tweets for the top five counties which show that a higher number of tweets are posted from India and Pakistan. Both countries have their political interest in the neighbor Afghanistan and the public also expresses their opinions regarding that.

Presenting only the ratio of positive and negative tweets to the total number of positive and negative tweets may not be enough to understand the trend of individual countries regarding positive and negative tweets. In addition, the distribution of positive, negative, and neutral tweets for each country is more meaningful as it can portray the overall sentiments of a specific country concerning the current situation in Afghanistan. Figure 13 shows the distribution of tweets sentiments for the top five countries with the highest number of tweets including India, the US, the UK, Pakistan, and Afghanistan. It shows that for the most part, the sentiments in the tweets are neutral with the highest ratio of

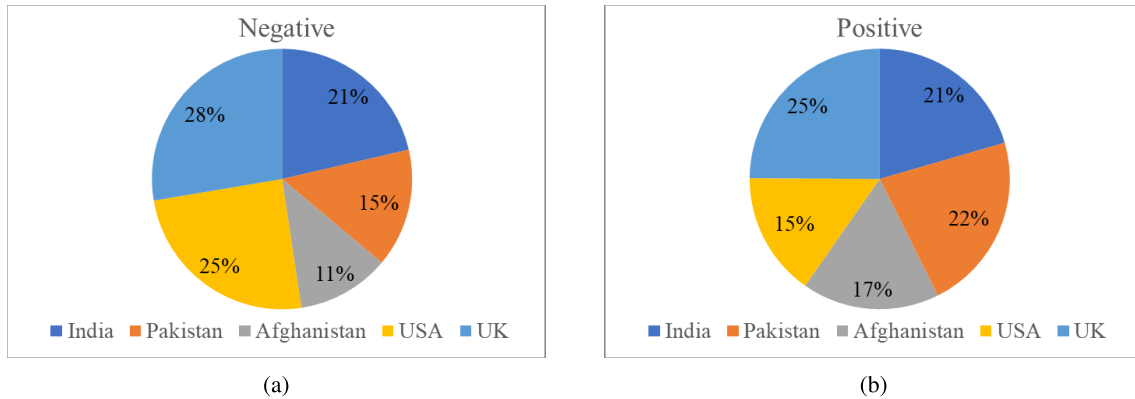


FIGURE 12. Top five countries in dataset in terms of number of tweets, (a) Negative tweets ratio, (b) Positive tweets ratio.

TABLE 15. Critical overview of machine learning models regarding feature and parameter selection.

Model	Feature	Parameter	Advantage	Disadvantage
SVM	HFE, TF-IDF, BoW, Word2Vec	Kernel=linear	Performs better with text data as it generates a large feature set which is suitable SVM for obtaining good accuracy	Higher computational time as compared to other machine learning models.
LR	HFE, TF-IDF, BoW, Word2Vec	Solver =saga, multi_class = multinomial	LR performs well when the number of features are high.	Multi-class classification performance is poor.
ETC	HFE, TF-IDF, BoW, Word2Vec	N_estimators=300, Max_depth=300	More suitable for small size datasets.	Poor performance for linear data and sparse feature set, as compared to SVM.
GNB	HFE, TF-IDF, BoW, Word2Vec	Default setting	GNB is good with simple features such as BoW.	Large feature set affects performance of GNB.
KNN	HFE, TF-IDF, BoW, Word2Vec	n_neighbour= 5	Simple to implement and interpret, low computational cost.	Shows poor performance on large feature set.
LSTMS	Embedding vector	LSTM layer, dropout layer, dense layer	Good for text data due to its recurrent architecture.	Requires a large dataset for a good fit.
GRU	Embedding vector	GRU layer, dropout layer, RNN layer, dense layer	Shows good results with text data due to its advanced recurrent architecture and is faster than LSTM.	Requires a large dataset for a good fit.
CNN-LSTMS	Embedding vector	LSTM layer, flatter layer, Maxpooling layer, dropout layer dense layer	CNN combined with LSTM increases important feature extraction process and improves the performance of LSTM.	CNN requires a large feature set to find significant features for good fit.

64% found in the tweets originating from Afghanistan. The neutral attitude from Afghanistan people shows a diplomatic mindset concerning the uncertainty of the situation. The lowest ratio of neutral tweets originates from the UK with 37% of the total tweets originating from the UK. For Pakistan, India, and the US, approximately half of the posted tweets has neutral sentiments. On the other hand, the ratio of negative sentiments is the highest in the UK, followed by the US and India, respectively. The lowest ratio of negative tweets is found in Afghanistan and Pakistan with respect to total tweets originating from these countries. Considering the total tweets from the top five countries only, the highest number of tweets are generated from India, followed by Pakistan, Afghanistan, the US, and the UK which is normal as the immediate stakeholders of the current situation in Afghanistan are its neighbors Pakistan and India.

H. ANALYSIS OF AFGHANISTAN'S SITUATION AFTER 14 DAYS OF INITIAL DATASET EXTRACTION

For analyzing if there is any shift in the sentiments, after the initial data collection, a new dataset is collected from 03 September 2021 to 08 September 2021. Sentiment analysis is performed on the tweets grouped by the day and the results are displayed in Figure 14. Results indicate that despite the neutral sentiments being in the highest number of tweets, an increasing trend is observed in the ratio of positive sentiments while negative sentiments seem to decline. It may be due to the fact that the situation in Afghanistan is on the route to normalization with the formation of a new Government.

I. TOPIC EXTRACTION USING LATENT DIRICHLET ALLOCATION

The sentiment analysis on Afghanistan-related tweets will be more meaningful with topic extraction so this study performs

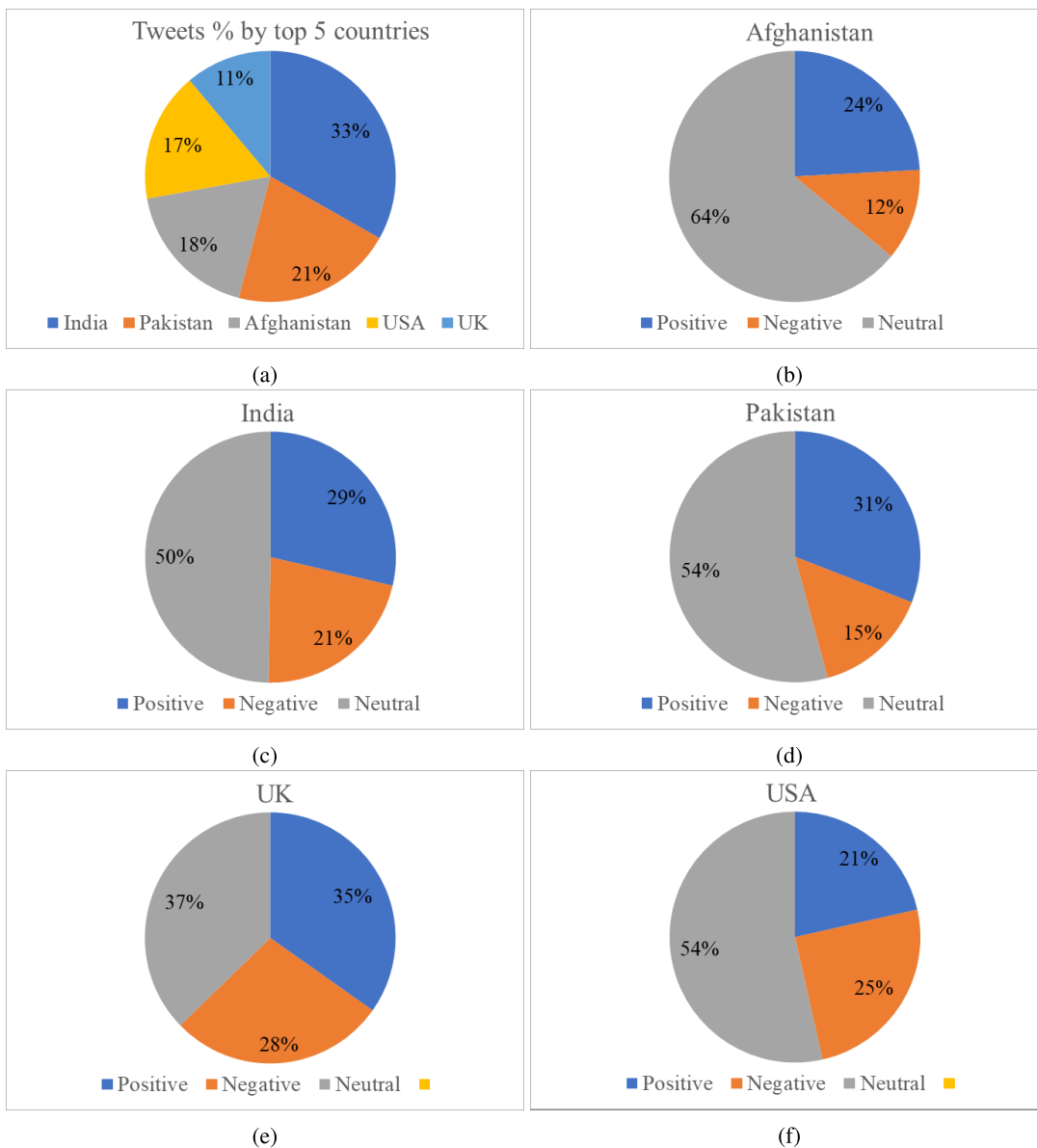


FIGURE 13. Tweets distribution for top five countries with highest number of tweets, (a) Total tweets, (b) Afghanistan tweets, (c) India tweets, (d) Pakistan tweets, (e) UK tweets, and (f) US tweets.

topic extraction using LDA. The underlying objective is to list down the current issues as presented in the tweets. LDA is a statistical model which discovers the topics from given documents [40]. LDA is fit on the data once the preprocessing and features extraction has been carried out. For this purpose, different hyperparameters are used for LDA. For example, n_components is set to 5 for extracting 5 most discussed topics. A list of other hyperparameters is provided in Table 16.

The top 5 most discussed topics are extracted for both positive and negative tweets as shown in Table 17. Based on the topics from positive tweets, it can be inferred that the people of Afghanistan seem satisfied with the Taliban takeover in Afghanistan as can be seen from Topic 1 and Topic 3 in positive tweets. On the contrary, topic extraction

TABLE 16. Hyperparameters used for topic modeling.

Hyperparameter	Value
n_components	5
max_iter	5
learning_method	online
learning_offset	50
random_state	0

from negative tweets shows that topics found in the negative tweets are mostly related to the US concerns in Afghanistan as shown in Topic 1, 3, and 4 of negative tweets.

Topic modeling results show that positive sentiments come from those tweets which are praising the Taliban’s bravery in fighting the US forces. Similarly, positive sentiments

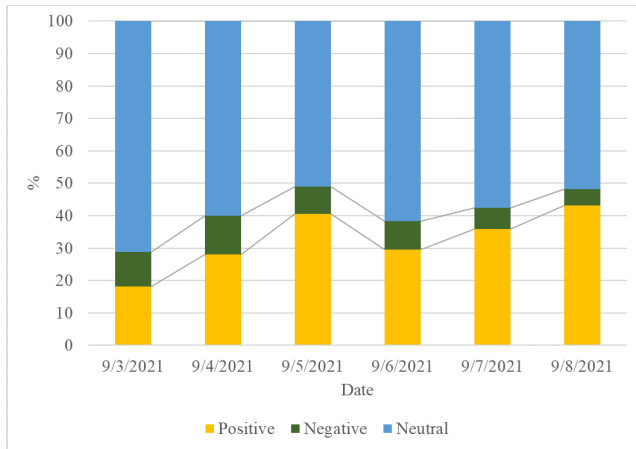


FIGURE 14. Data analysis after 14 days of gaps from initial data.

TABLE 17. Topic extracted from positive and negative tweets.

Topic	Positive
Topic 1	taliban right like usa great better afghanistan
Topic 2	people taliban said time says afghanistan good country right
Topic 3	pakistan media safe live afghanistan
Topic 4	afghan taliban forces special afghanistan kabul govt government
Topic 5	kabul taliban afghanistan good airport e
Topic	Negative
Topic 1	taliban killed trump usa fighters
Topic 2	taliban military time weapons rule afghan
Topic 3	taliban biden women afghans like media worse men long phase
Topic 4	taliban worse rule usa
Topic 5	world forces usa know went worse

TABLE 18. Mostly used words in tweets impact on negative and positive sentiments.

Count	Negative	Positive
1	killed	victory
2	destroy	better
3	warcrime	freedom
4	pathetic	conquer
5	terror	expertise
6	horrifying	easily
7	angry	free
8	failed	trust
9	defeated	winning
10	slaver	right

appeared in tweets favoring the local Government and people’s autonomy. Predominantly, these tweets are originated from the Middle East and Afghanistan neighboring countries. Many tweets contain negative sentiments for the Taliban due to several issues. First, many humanists and social workers show their reservations concerning the violation of human rights in general such as freedom of speech and action. Secondly, voices are raised regarding the ban on women’s education. Similarly, concerns have been shared regarding the forced implementation of various policies by the Taliban. Table 18 shows the words most commonly used in tweets containing negative and positive sentiments.

In addition to the most common words from the overall tweet datasets, further experiments are carried out on the

TABLE 19. Mostly used words in tweets impact on negative and positive sentiments.

Country	Positive	Negative
India	Peace, loving, success, good, brilliant	Worst, painful, terrorism, terror,killed
Pakistan	Victory, safe, success, love, best	Bad, kill, wrong,evil, hypocrisy
UK	Power, innocent, charities, right, congratulates	Destroyed, death, horrifying, Hell, war
US	Free, peace, powerful, inspired, love	Failed, shamelessly, dangerous, Terrorism,dangerous
Afghanistan	Winning, good, safe, best, peace	Worse, terrorist, fears, horrible, defeated

country level to find the positive and negative words. Table 19 shows the list of the top 5 countries with the highest number of tweets.

V. CONCLUSION

This study performs sentiment analysis concerning the current situation in Afghanistan under Taliban control. The study accomplishes two objectives: sentiment analysis of tweets and volume analysis. For the former, a proposed hybrid feature engineering approach is used with several well-known machine learning models, as well as, three deep learning models including CNN, CNN-LSTM, and GRU. The latter involves the analysis of sentiments for the top five countries from where the highest number of tweets are originated. Geotagged tweets are utilized for this purpose to present the positive, negative, and neutral attitudes of people towards the current situation in Afghanistan. Furthermore, topic modeling is performed to extract the most discussed topics from negative and positive tweets using LDA. Of the three feature engineering approaches including BoW, TF-IDF, and Word2Vec, for the most part, the performance of machine learning models is better with BoW. Among the used machine learning models, SVM tends to show the highest performance with an accuracy score of 0.97 when used with the proposed hybrid features. The hybrid feature engineering approach makes use of Chi2 to select the most appropriate features from BoW, TF-IDF, and Word2Vec and combines them to elevate the performance of models. Predominantly high performance of deep learning models is bounded with a large volume of data and apparently, the comparatively smaller size of the data used in this study is not suitable to achieve the significant results with deep learning models. Volume analysis indicates that the majority of the sentiments found in the tweets considering the top five countries comprise neutral sentiments. However, considering the negative sentiments, a large portion of the negative tweets is originated from India, followed by the US and UK. On the other hand, the majority of the positive tweets are initiated from Pakistan and Afghanistan regarding the current situation in Afghanistan. Topic modeling of the positive tweets reveals that a major part of Afghani people seems satisfied with the Taliban’s takeover. On the contrary, topics from the negative tweets are mostly related to the US

concerns in Afghanistan. Gap analysis reveals that on average there is a positive shift in the sentiments over the past two weeks.

VI. FUTURE WORK

This research provides insights into the sentiment of people from different countries regarding the current situation in Afghanistan under the Taliban regime. Keeping in view the topics found using LDA, Governments can grasp the gravity of the situation and can take actions accordingly. Social and news media agencies can further explore the sentiments of people, especially the people from Afghanistan, regarding the Taliban regime and highlight the probable future challenges. For the future, the authors aim to conduct a study for management implications and technical improvement providing guidelines on dealing with different current challenges.

ACKNOWLEDGMENT

(Ernesto Lee and Furqan Rustam are co-first authors.)

REFERENCES

- [1] Council Foreign Relations. *War in Afghanistan*. Accessed: Aug. 31, 2021. [Online]. Available: <https://www.cfr.org/global-conflict-tracker/conflict/war-afghanistan>
- [2] J. Schroden, "Afghanistan's security forces versus the taliban: A net assessment," *CTC Sentinel*, vol. 14, p. 1, Oct. 2021.
- [3] J. Krauss. *Taliban Take Over Afghanistan: What we Know and What's Next*. Accessed: Aug. 31, 2021. [Online]. Available: <https://apnews.com/article/taliban-takeover-afghanistan-what-to-know-1a74c9cd866866f196c478aba21b60b6>
- [4] J. Hollingsworth. *Who are the Taliban and How Did They Take Control of Afghanistan so Swiftly*. Accessed: Aug. 31, 2021. [Online]. Available: <https://edition.cnn.com/2021/08/16/middleeast/taliban-control-afghanistan-explained-intl-hnk/index.html>
- [5] N. Lyndon, A. Salman, and K. C. Hoe, "The enlivenment of public opinion in the new era: Exploring the power of social media on political consciousness in Saudi Arabia," *Humanities*, vol. 9, no. 2, pp. 209–219, 2021.
- [6] H. Vanam, "Analysis of Twitter data through big data based sentiment analysis approaches," *Mater. Today, Proc.*, Jan. 2021, doi: 10.1016/j.matpr.2020.11.486.
- [7] E. Elmurghi and A. Gherbi, "Detecting fake reviews through sentiment analysis using machine learning techniques," in *Proc. IARIA/Data Anal.*, 2017, pp. 65–72.
- [8] B. Dahal, S. A. Kumar, and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Social Netw. Anal. Mining*, vol. 9, no. 1, pp. 1–20, 2019.
- [9] H.-J. Kwon, H.-J. Ban, J.-K. Jun, and H.-S. Kim, "Topic modeling and sentiment analysis of online review for airlines," *Information*, vol. 12, no. 2, p. 78, 2021.
- [10] S. R. Basha, J. K. Rani, and J. P. Yadav, "A novel summarization-based approach for feature reduction enhancing text classification accuracy," *Eng., Technol. Appl. Sci. Res.*, vol. 9, no. 6, pp. 5001–5005, 2019.
- [11] S. Rahamat Basha, "Impact of feature selection techniques in text classification: An experimental study," *J. Mech. Math. Sci.*, vol. 1, no. 3, pp. 39–51, Sep. 2019.
- [12] M. Abdullah and M. H. Hadzikadic, "Sentiment analysis of Twitter data: Emotions revealed regarding donald Trump during the 2015-16 primary debates," in *Proc. IEEE 29th Int. Conf. Tools with Artif. Intell. (ICTAI)*, Nov. 2017, pp. 760–764.
- [13] W. F. Al-Sarraj and H. M. Lubbad, "Bias detection of Palestinian/Israeli conflict in western media: A sentiment analysis experimental study," in *Proc. Int. Conf. Promising Electron. Technol. (ICPET)*, Oct. 2018, pp. 98–103.
- [14] V. D. G. Vera, L. M. M. Suárez, and I. C. P. Lopera, "Sentiment analysis on post conflict in colombia: A text mining approach," *Asian J. Appl. Sci.*, vol. 6, no. 2, pp. 1–7, 2018.
- [15] E. Abedin, H. Jafarzadeh, and S. Akhlaghpour, "Opinion mining on Twitter: A sentiment analysis of the Iran deal," in *Proc. Pacific Asia Conf. Inf. Syst. (PACIS)*, Yokohama, Japan. Atlanta, GA, USA: Association for Information Systems, Jun. 2018, p. 12.
- [16] N. H. Khun and H. A. Thant, "Visualization of Twitter sentiment during the period of U.S. banned huawei," in *Proc. Int. Conf. Adv. Inf. Technol. (ICAIT)*, Nov. 2019, pp. 274–279.
- [17] M. AlKhatib, M. El Barachi, A. Aleahmad, F. Oroumchian, and K. Shaalan, "A sentiment reporting framework for major city events: Case study on the China-United States trade war," *J. Cleaner Prod.*, vol. 264, Aug. 2020, Art. no. 121426.
- [18] N. Öztürk and S. Ayvaz, "Sentiment analysis on Twitter: A text mining approach to the syrian refugee crisis," *Telematics Inf.*, vol. 35, no. 1, pp. 136–147, 2018.
- [19] P. Ghasiya and K. Okamura, "Understanding the Middle East through the eyes of Japan's Newspapers: A topic modelling and sentiment analysis approach," *Digit. Scholarship Hum.*, vol. 36, no. 4, pp. 871–885, Mar. 2021.
- [20] E. Cunliffe and L. Curini, "Isis and heritage destruction: A sentiment analysis," *Antiquity*, vol. 92, no. 364, pp. 1094–1111, 2018.
- [21] S. Mansour, "Social media analysis of user's responses to terrorism using sentiment analysis and text mining," *Proc. Comput. Sci.*, vol. 140, pp. 95–103, Oct. 2018.
- [22] J. Roesslein. (May 2009). *Tweeepy Documentation*. [Online]. Available: <http://tweeepy.readthedocs.io/en/v3>
- [23] F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan, and G. S. Choi, "Classification of shopify app user reviews using novel multi text features," *IEEE Access*, vol. 8, pp. 30234–30244, 2020.
- [24] R. Khan, F. Rustam, K. Kanwal, A. Mehmood, and G. S. Choi, "U.S. based COVID-19 tweets sentiment analysis using TextBlob and supervised machine learning algorithms," in *Proc. Int. Conf. Artif. Intell. (ICAI)*, Apr. 2021, pp. 1–8.
- [25] 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.
- [26] E. Saad, S. Din, R. Jamil, F. Rustam, A. Mehmood, I. Ashraf, and G. S. Choi, "Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums," *IEEE Access*, vol. 9, pp. 85721–85737, 2021.
- [27] G. Salton and C. T. Yu, "On the construction of effective vocabularies for information retrieval," *ACM SIGPLAN Notices*, vol. 10, no. 1, pp. 48–60, 1975.
- [28] Y.-T. Zhang, L. Gong, and Y.-C. Wang, "An improved TF-IDF approach for text classification," *J. Zhejiang Univ. Sci.*, vol. 6, no. 1, pp. 49–55, 2005.
- [29] J. Lilleberg, Y. Zhu, and Y. Zhang, "Support vector machines and Word2vec for text classification with semantic features," in *Proc. IEEE 14th Int. Conf. Cognit. Informat. Cognit. Comput.*, Jul. 2015, pp. 136–140.
- [30] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," in *Proc. 7th IEEE Int. Conf. Tools with Artif. Intell.*, 1995, pp. 388–391.
- [31] R. E. Wright, "Logistic regression," in *Reading and Understanding Multivariate Statistics*, L. G. Grimm and P. R. Yarnold, Eds. Washington, DC, USA: American Psychological Association, 1995, pp. 217–244.
- [32] I. Mazlin, I. M. Rawi, and Z. Zakaria, "Hadith arabic text classification using convolutional neural network and support vector machine," in *Proc. ICCST*, Pattaya, Thailand, vol. 724, Aug. 2020, p. 371.
- [33] R. L. Mustafa and B. Prasetyo, "Sentiment analysis using lexicon-based method with naive Bayes classifier algorithm on #newnormal hashtag in Twitter," *J. Phys., Conf. Ser.*, vol. 1918, no. 4, Jun. 2021, Art. no. 042155.
- [34] Z. Ge, "A fusion algorithm model based on KNN-SVM to classify and recognize spam," *J. Phys., Conf. Ser.*, vol. 1982, no. 1, Jul. 2021, Art. no. 012069.
- [35] A. P. Pimpalkar and R. J. Retna Raj, "Influence of pre-processing strategies on the performance of ML classifiers exploiting TF-IDF and BOW features," *Adv. Distrib. Comput. Artif. Intell. J.*, vol. 9, no. 2, pp. 49–68, Jun. 2020.
- [36] R. Jamil, I. Ashraf, F. Rustam, E. Saad, A. Mehmood, and G. S. Choi, "Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model," *PeerJ Comput. Sci.*, vol. 7, p. e645, Oct. 2021.
- [37] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Aamir, "An enhanced gated recurrent unit with auto-encoder for solving text classification problems," *Arabian J. Sci. Eng.*, vol. 4, pp. 1–15, May 2021.

- [38] D. Soekhoe, P. V. D. Putten, and A. Plaat, "On the impact of data set size in transfer learning using deep neural networks," in *Proc. Int. Symp. Intell. Data Anal.* Cham, Switzerland: Springer, 2016, pp. 50–60.
- [39] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Aug. 2018.
- [40] R. Xie, S. K. W. Chu, D. K. W. Chiu, and Y. Wang, "Exploring public response to covid-19 on weibo with lda topic modeling and sentiment analysis," *Data Inf. Manage.*, vol. 5, no. 1, pp. 86–99, 2021.



His research interests include bias in machine learning, applied deep learning and machine learning, and blockchain.

ERNESTO LEE received the bachelor's degree in physics from Old Dominion University, the Master of Science degree in systems engineering with a concentration in software engineering from the Virginia Tech, and the Ph.D. degree in business administration with a concentration in data science and analytics from the Baker College. Since 2020, he has been a Computer Science Professor at the Broward College and the Lead Researcher at the Florida Center for Advanced Analytics and Data



His research interests include data mining, machine learning, and artificial intelligence, mainly working on creative computing, and supervised machine learning.

FURQAN RUSTAM received the M.C.S. degree from the Department of Computer Science, Islamia University of Bahawalpur, Pakistan, in October 2017. He is currently pursuing the Master of Computer Science degree with the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. He is also working as a Research Assistant at the Fareed Computing & Research Center, KFUEIT.



His research interests include indoor positioning and localization, advanced location-based services in wireless communications, smart sensors (LIDAR) for smart cars, and data mining.

IMRAN ASHRAF received the M.S. degree in computer science from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2010, and the Ph.D. degree in information and communication engineering from Yeungnam University, Gyeongsan, South Korea, in 2018. He worked as a Postdoctoral Fellow at Yeungnam University. He is currently working as an Assistant Professor at the Information and Communication Engineering Department, Yeungnam University. His



As a leading authority on data science in finance, he has integrated deep and machine learning into his research while assisting companies in their implementation of artificial intelligence.

PATRICK BERNARD WASHINGTON is currently a Philosopher leading the world conversation at the intersection of finance, technology, and public policy. He is also a Distinguished Author and a Public Speaker providing innovative solutions, investing in transformative ideas, and changing people's financial trajectory. He is also an Associate Professor of finance at the Morehouse College. His research interests include intersections between investments, asset pricing, technology,



His research interests include machine learning, artificial intelligence, and data mining.

MANIDEEP NARRA received the master's degree in engineering management with information technology concentration from the Indiana Institute of Technology, Fort Wayne, IN, USA. He worked on different domain, including marketing services and healthcare. His accomplishments included designing, high quality, and deliverability with security. His research interests include machine



His research interests include data mining, machine learning, and artificial intelligence, mainly working on creative computing, and supervised machine learning.

RAHMAN SHAFIQUE received the master's degree in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. He is currently working as a Research Assistant with the Fareed Computing and Research Center, KFUEIT. His research interests include

...