

Received March 19, 2022, accepted April 1, 2022, date of publication April 7, 2022, date of current version April 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3165621

Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model

NAILA ASLAM^{1,2}, FURQAN RUSTAM³, ERNESTO LEE⁴, PATRICK BERNARD WASHINGTON⁵, AND IMRAN ASHRAF⁶

¹School of Electronics and Information Engineering, Hebei University of Technology, Tianjin 300401, China

²Department of Computer Science, School of Systems and Technology, University of Management and Technology, Lahore 54000, Pakistan

³Department of Software Engineering, School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan

⁴Department of Computer Science, Broward College, Broward County, Fort Lauderdale, FL 33301, USA

⁵Division of Business Administration and Economics, Morehouse College, Atlanta, GA 30314, USA

⁶Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si 38544, South Korea

Corresponding authors: Imran Ashraf (ashrafimran@live.com) and Ernesto Lee (elee@broward.edu)

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 The cryptocurrency market has been developed at an unprecedented speed over the past few years. Cryptocurrency works similar to standard currency, however, virtual payments are made for goods and services without the intervention of any central authority. Although cryptocurrency ensures legitimate and unique transactions by utilizing cryptographic methods, this industry is still in its inception and serious concerns have been raised about its use. Analysis of the sentiments about cryptocurrency is highly desirable to provide a holistic view of peoples' perceptions. In this regard, this study performs both sentiment analysis and emotion detection using the tweets related to the cryptocurrency which are widely used for predicting the market prices of cryptocurrency. For increasing the efficacy of the analysis, a deep learning ensemble model LSTM-GRU is proposed that combines two recurrent neural networks applications including long short term memory (LSTM) and gated recurrent unit (GRU). LSTM and GRU are stacked where the GRU is trained on the features extracted by LSTM. Utilizing term frequency-inverse document frequency, word2vec, and bag of words (BoW) features, several machine learning and deep learning approaches and a proposed ensemble model are investigated. Furthermore, TextBlob and Text2Emotion are studied for emotion analysis with the selected models. Comparatively, a larger number of people feel happy with the use of cryptocurrency, followed by fear and surprise emotions. Results suggest that the performance of machine learning models is comparatively better when BoW features are used. The proposed LSTM-GRU ensemble shows an accuracy of 0.99 for sentiment analysis, and 0.92 for emotion prediction and outperforms both machine learning and state-of-the-art models.

INDEX TERMS Cryptocurrency, sentiment analysis, Text2Emotion, emotion analysis, machine learning.

I. INTRODUCTION

Cryptocurrency market has been developed at an exceptional pace since its emergence. Cryptocurrency is a digital currency however it is not controlled by any central authority to make online payments. It uses system ledger entries called 'tokens' to make online payments for goods and services. Elliptical curve encryption and public-private key pairs are used as

The associate editor coordinating the review of this manuscript and approving it for publication was Seifedine Kadry¹.

cryptographic algorithms. Similarly, hashing functions are utilized to protect online payments and ensure legitimate and unique transactions. Bitcoin was the first blockchain-based cryptocurrency introduced in 2009 and it remains important and leading the market today. In addition to Bitcoin, a large number of cryptocurrencies have been introduced over time, each with its opportunities and functions to provide different features and specifications. Such cryptocurrencies include Bitcoin clones, as well as, entirely new currencies with additional features.

Cryptocurrency investors expect both profit and loss due to ups and downs in the crypto market. For this purpose, many tools are available which can forecast the crypto market and occasionally investors invest based on such forecasts. The rise and fall in the demand for cryptocurrencies are also affected by general public opinion or Governmental policies. In this regard, peoples' sentiments and emotions can help in determining the up and down of cryptocurrency market value, especially, sentiment analysis is trendy nowadays for investment in cryptocurrency [1], [2]. Investors first perform an analysis of peoples' sentiment for a specific currency and then make investments according to the sentiments [3]. Because of that, sentiment analysis on cryptocurrency markets has become a task of great importance [4]. Studies show that tweets containing positive sentiments have a substantial impact on the demand for cryptocurrencies and vice versa [5], [6].

Despite the proposal of several sentiment analysis approaches, several challenges require further research efforts. For example, sentiment annotation is challenging when the sentence structure is complex. Often, simple sentences are needed to produce high-accuracy annotations. Similarly, a single approach cannot be generalized and applicable to all the corpus. An approach designed for sentiment analysis in one domain does not necessarily produce good results in another domain. In addition, the role of a specific feature extraction technique cannot be ignored fully. From this perspective, this study is specially designed for predicting people's sentiments and emotions on the cryptocurrency market using supervised machine learning models. Owing to the wide use of TwitterTM for expressing opinions and thoughts on specific topics, this study leverages a tweets dataset for this purpose. This study makes the following contributions

- An ensemble model is proposed to perform sentiment analysis with high accuracy. For this purpose, the advantages of long short-term memory (LSTM) and gated recurrent unit (GRU) are combined.
- Sentiment analysis and emotion analysis are performed. TextBlob is used for annotating the sentiments data while emotions are annotated using the Text2Emotion model. Positive, negative, and neutral sentiments are used while emotions are classified into happy, sad, surprise, angry, and fear.
- The suitability and performance of three feature engineering approaches are studied including term frequency-inverse document frequency (TF-IDF), bag of words (BoW), and Word2Vec. Experiments are performed using several well-known machine learning models such as support vector machine (SVM), logistic regression (LR), Gaussian Naive Bayes (GNB), extra tree classifier (ETC), decision tree (DT), and k nearest neighbor (KNN). Additionally, the performance of LSTM and GRU models is also analyzed.

The rest of the paper is structured as follows. Important research papers related to the current study are discussed in Section II. The proposed approach, the dataset used for

experiments, and machine learning algorithms are presented in Section III. Section IV provides the analysis and discussion of results. In the end, the conclusion is given in Section V.

II. RELATED WORK

Sentiment analysis has emerged as an important research area due to the wide use of social media platforms. As a result, a large body of literature can be found on sentiment and emotion analysis. For example, [7] proposes a machine learning approach for the automatic detection of emotions from the text posted on social networks. Emotions are detected by performing the text classification. The study investigates several problems including semantic complexity of text messages, casual style of micro-blogs, multiple sentiments in text, and different states of emotions. Binary classifiers are used to distinguish tweets with emotions and tweets without emotions. Two main tasks of the approach include offline training and online classification task. The developed emotion classification system Emotex can obtain a classification accuracy of 90% for text messages.

Similarly, emotion detection and emotion intensity degree is predicted in [8]. For this purpose, natural language processing (NLP) tools are used on sexist tweets which are categorized into indirect harassment, physical harassment, and sexual harassment. Additionally, emotions of anger, joy, sadness and fear are investigated containing low, medium, and high intensity. For multilabel classification SVM, Naive Bayes (NB), KNN, Multi-layer perceptron (MLP), LSTM, and convolutional neural network (CNN) are used with Word2Vec, global vector (Glove), and FastText for achieving high classification accuracy. In the same manner, the study investigates 3 categories of speech containing sexist remarks to find the intensity of each emotion. Results show that joy feeling and indirect harassment have direct relation and anger is associated with sexual harassment when the intensity is considered. Similarly, anger, joy, and sadness feelings are associated with physical harassment.

The study [11] conducts experiments to detect tweets' emotions using the AIT-2018 dataset. The authors propose a new model that leverages EmoSenticNet and WordNetAffect for detecting emotions. Results show that the performance is affected by the small dataset and language ambiguity problems. Accuracy is reduced by the text containing multiple emotions. The authors perform sentiment analysis on online social networks in [15] using machine learning and lexicon-based methods. A multilabel learning algorithm is introduced for this purpose. The proposed approach aims at multiple level emotion detection concerning user view and incorporates machine learning approaches to achieve a multilabel emotion detection system. The authors discover social correlation and temporal correlation, as well as, emotion label correlation. Despite its capability, the proposed approach is limited by the use of a small dataset and low accuracy. A deep learning-based nested LSTM model is presented for emotion detection in [12]. The authors aim at increasing the classification accuracy of Ekman's emotions. A large dataset

TABLE 1. A summary of the discussed research works.

Ref.	Year	Model	Objective	Dataset
[9]	2017	SVM	Emotion detection from text	-
[10]	2018	K-means, Naive Bayes and SVM	Emotion Detection and Recognition from Text	Tweet data
[7]	2019	SVM	Text stream message emotion detection	Tweet data
[8]	2019	SVM, NB, KNN ,MLP,CNN, LSTM	detects emotions on sexist Tweet	Tweet data
[11]	2019	Naive Bayes, DT, SVM, Proposed Model Unifying WordNetAffect and EmoSenticNet	Tweet emotion detection	Tweet data AIT-2018
[12]	2019	Nested LSTM, LSTM,SVM	Emotion detection on text	Tweet data, 980,549 training data, 144,160 testing data
[13]	2019	DT,CNN, Doc2Vector	Text emotion detection	Facebook and tweet
[14]	2019	Naive Bayes Classifier	Text Classification into four different emotional classes	Social Media
[15]	2020	ML-KNN,Proposed model, Multi-label Learning Approach for Emotion Detection	Online Social Networks emotion detection	Tweet data
[16]	2020	Emotion Extraction Model	detect emotions	Dataset ('International Survey on Emotion Detection Antecedents and Reactions' (ISEAR))
[17]	2020	LSTM	Bitcoin price prediction using sentiment analysis	Twitter™and Reddit
[18]	2021	LSTM	Cryptocurrency sentiment analysis for Chinese language data	Chinese social media platform Sina-Weibo dataset
[19]	2021	RF	Cryptocurrency sentiment analysis	Tweets dataset

containing 980,549 training samples and 144,160 testing samples is used for experiments. Nested LSTM has the best accuracy among other methods with a 99.16% accuracy score. The drawback that needs to be rectified is that the results are not significantly different from other models.

Besides the emotion analysis, several works perform cryptocurrency market prediction for which sentiment analysis of cryptocurrency-related tweets is utilized. Generally, positive comments containing positive sentiments are associated with higher market demand for cryptocurrency and vice versa. For example, study [18] proposes a sentiment analysis approach for cryptocurrency value prediction using a machine learning approach. Posts related to cryptocurrency are extracted from the Chinese social media platform Sina-Weibo for analysis. The proposed crypto-based sentiment dictionary and the LSTM model are used for prediction. The proposed approach outperforms the previous works by 18.5% for precision and 15.4% for recall. Similarly, the authors perform sentiment analysis in [19] on specific cryptocurrency coins using the tweets dataset and machine learning approach. Experiments are performed on NEO and manually labeled tweets dataset. Experiments performed using RF indicate 77% accuracy for sentiment classification.

The study [20] conducts experiments for predicting cryptocurrency prices using the sentiment analysis approach. The proposed approach predicts the cryptocurrency market price in real-time using the news and tweets data. The authors use the LSTM model for price prediction. Similarly, another study [17] proposes an approach for predicting bitcoin price using sentiment analysis and future forecasting techniques.

The study [17] finds a correlation between the bitcoin price and tweets sentiments. The study uses LSTM and ARIMA for the price prediction of bitcoin.

The above-discussed studies predict cryptocurrency prices using the sentiment analysis of cryptocurrency-related tweets. Results from the machine and deep learning models show low classification performance. For this reason, further research is required to enhance the classification performance of cryptocurrency-related tweets. Other than that the main objective of such studies is the cryptocurrency price prediction and not the sentiment analysis itself. So, the scope of such studies is limited. This study addresses these issues by proposing a high accuracy ensemble classifier for sentiment analysis on cryptocurrency-related tweets. The study [12] also uses an ensemble model, where two LSTM models are united in the nested way with an inner join. The current study, on the other hand, makes an ensemble of two different recurrent neural networks comprising LSTM and GRU. Secondly, these models are combined in a stacked manner where the output of one model serves as the input of the second model. Joining LSTM and GRU as a stacked model proves to be more accurate than LSTM with inner join.

III. MATERIALS AND METHODS

A. PROPOSED METHODOLOGY

This study performs experiments for sentiment analysis and emotion detection on cryptocurrency-related tweets. For this purpose, an ensemble model is proposed to obtain sentiment and emotion detection with improved classification

accuracy. Figure 1 shows the architecture of the proposed methodology. All experiments are carried out using an Intel Core i7 11th generation machine with Windows operating system. Machine and deep learning models are implemented in Python language using TensorFlow, Keras, and the scikit learning frameworks.

In the proposed approach, the first step is Twitter™ data collection using the Tweepy library. In this regard, a Twitter™ developer account is generated and Tweets are scrapped. Tweets are collected using specific tags such as “#cryptocurrency”, “cryptocurrency”, “#cryptomarket”, and “#BTC”. A total of 40,000 tweets are collected in this process. The data collection is carried out starting from July to August 2021. Sample text from the collected tweets is shown in Table 2.

TABLE 2. Sample text from the dataset.

User	Location	Text	Date
Aureum Victoria	Sachsen	Nice! My #Trading Bot 5 just sold \$NEO/ \$USDT with 4.73% profit on #binance! Binance -20% on the Fee:â€I https://t.co/hhJIKeZD1b	7/30/2021
tron_age	USA	Striatum is the World's First Unreinforced 3D-Printed Concrete Arched Bridge #blockchain #dlike #cryptocurrency https://t.co/10i2IUAbG4	7/30/2021

Tweets contain links, tags, usernames, numbers, and other characters which are not useful for the machine learning models' training. We remove these meaningless data from tweets using preprocessing techniques such as stemming, lemmatization, spell correction, stopwords removal, etc.

After data preprocessing, the dataset is annotated for both sentiment analysis and emotion detection. For sentiment annotation, TextBlob is used while for emotion annotation Text2Emotion libraries are utilized. TextBlob and Text2Emotion results on the sample preprocessed data are shown in Table 3.

TextBlob gives a 0.6 polarity score to the first tweet which means it is highly positive while the second tweet's polarity score is 0.2 which shows that tweet is positive but not highly. While for emotion detection Text2Emotion predicts the highest score of 0.5 for happy class on the first tweet and the second tweet's prediction is the highest score of 0.5 for surprise emotion. The ratio of negative, positive, and neutral tweets is shown in Figure 2a, whereas the ratio of each emotion in the dataset is given in Figure 2b.

Data splitting is the next step, where the data is divided into a ratio of 85 to 15. For training the models, 85% data is used while 15% is used for testing the models. This splitting ratio is more suitable for the models because 85% data is enough to train deep learning models. Table 4 shows the ratio of training and testing sets for both emotion detection and sentiment analysis datasets.

TABLE 3. Sentiment scoring using TextBlob and Text2Emotion.

Text	TextBlob		Text2Emotion	
	Polarity score	Sentiment	Highest score	Emotion
nice trad bot just sold neo usdt profit on binance binance fee	0.6	Positive	0.5	Happy
striatum world first unreinforced print concrete arch bridge blockchain cryptocurrency	0.2	Positive	0.5	Surprise

TABLE 4. Number of samples for training and test datasets.

Set	Samples
Training set	34,000
Testing set	6,000
Total data	40,000

Feature extraction is the performed which is required for machine learning models. Deep learning models, on the other hand, do not need a feature extraction process. The BoW, TF-IDF, and Word2Vec are used for feature extraction. After feature extraction, machine learning models are trained using the training data and in the end, the performance is evaluated in terms of accuracy, precision, recall, and F1 score.

B. PREPROCESSING

To increase the learning efficiency of machine learning models, preprocessing techniques are used to clean the data. The following steps are carried out for data preprocessing.

1) TOKENIZATION

It is the process of breaking down a text into smaller parts known as 'tokens'. A token can be a number, a phrase, or any symbol that includes all the relevant information about the data while maintaining its security.

2) PUNCTUATION REMOVAL

Natural language processing methods are used to remove punctuation from the tweets. Punctuation is a set of symbols often used in phrases and remarks for increasing the text's understandability for humans. However, it limits the learning capability of machine learning algorithms and must be eliminated to enhance their learning process. Common punctuation signs include the colon, question marks, comma, semicolon, full-stop/period, and so on [21].

3) NUMBER REMOVAL

This is another aspect of preprocessing that improves the performance of machine learning algorithms. Numbers in the text do not provide meaningful information for models' training and their removal reduces the feature space. Removing the numbers improves model efficiency while decreasing the complexity.

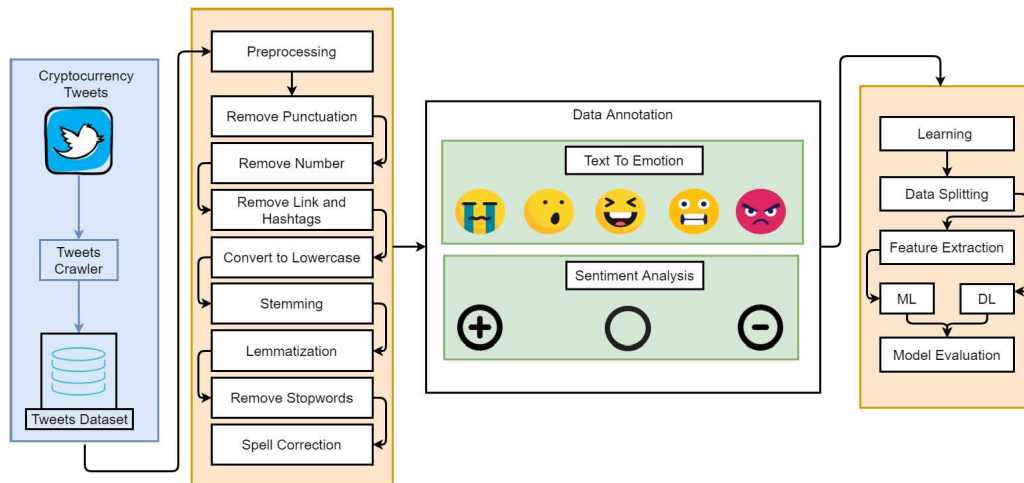


FIGURE 1. Architecture of the proposed methodology.

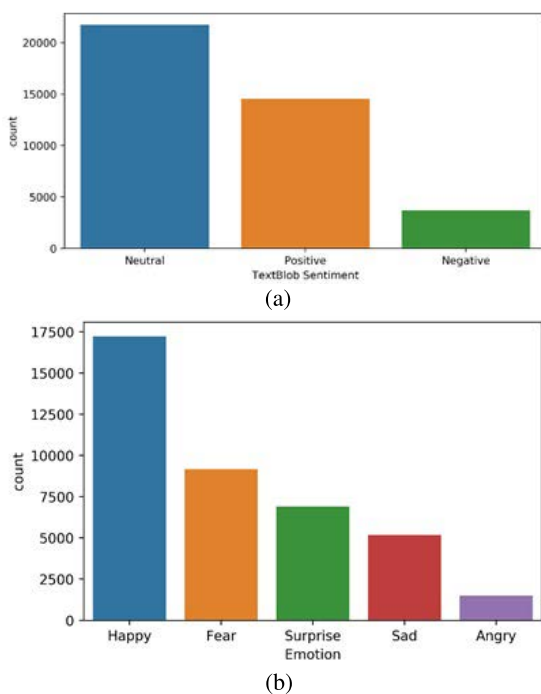


FIGURE 2. Number of samples, (a) Each class in the dataset, (b) Emotions found in the dataset.

4) STEMMING

Stemming is an important element of preprocessing because it improves efficiency by clarifying affixes from sentences/comments and returning the comments to their original form. The process of converting a word into its root form is known as stemming. For instance, several terms may have the same meaning, such as ‘goes,’ ‘going,’ and ‘gone,’ are all modified versions of ‘go.’ The Porter stemmer algorithms are used to implement stemming [22].

5) LEMMATIZATION

Lemmatization converts an extended word to its root form. Lemmatization can determine the intended part of speech

correctly, as well as, the sense of a word in a sentence. Stemming and lemmatization differ significantly where lemmatization evaluates the context first and then converts the word to its proper root form while stemming simply removes ‘s’ or ‘es’ at the end of a word. As a result, stemming often produces incorrect or incomplete words involving spelling mistakes. For example, lemmatization would appropriately identify the base form of ‘Studies’ to ‘Study’, whereas, stemming would cutoff the ‘es’ part and convert it to ‘Studi’.

‘Studies’ -> Lemmatization -> ‘Study’

‘Studies’ -> Stemming -> ‘Studi’

6) STOP WORDS REMOVAL

Stop words are English words that do not add any meaning to a sentence for machine learning models. As a result, stop words can be deleted without altering the meaning of a phrase. The elimination of stop-words improves the model’s performance while decreasing the complexity of input characteristics. We used pre-defined NLTK corpus stopwords [22].

7) SPELL CORRECTION

The practice of correcting misspelled words is known as spelling correction. During this step, the spelling checker is utilized to check for misspelled words and replace them with the right term. The Python package TextBlob is used for the experiments and offers the essential functionality for checking misspelled words [23]. Sample text from tweets is shown in Table 5 to illustrate the impact of preprocessing steps.

C. FEATURE EXTRACTION

For feature extraction three approaches are used including the BoW, TF-IDF, and Word2Vec whose short description is provided here for completeness.

TABLE 5. Sample text before and after the preprocessing.

Before preprocessing	After preprocessing
Nice! My #Trading Bot 5 just sold \$NEO/ \$USDT with 4.73% profit on #binance! Binance -20% on the Fee:hata€! https://t.co/hhJIKeZD1b	nice trad bot just sold neo usdt profit on binance binance fee
Striatu is the World's First Unreinforced 3D-Printed Concrete Arched Bridge #blockchain #dlike #cryptocurrency https://t.co/10i2IUAbG4	striatu world first unreinforc print concrete archi bridge blockchain cryptocurrency

1) BAG OF WORDS

The BoW technique is one of the simplest and most widely used feature extraction techniques for text analysis. Despite being easy to implement and understand, BoW often produces good results as compared to many complicated techniques. The BoW is commonly used for language modeling and text classification. This study uses the 'CountVectorizer' library to implement BoW. CountVectorizer counts the frequency of unique words and constructs a feature vector. Based on the simple occurrence of words, the feature space is often sparse [16].

2) TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY

In TF-IDF, IDF stands for inverse document frequency of the word, while TF stands for term frequency. The TF-IDF is a statistical method for determining the number of relevant words in a given text. The value rises with the number of times a word comes into the text, but it is adjusted by the word's frequency in the document [24]. TF is the number of times a term appears in a document's text. Because each document is different in size, a word will probably appear more frequently in longer documents compared to shorter ones. The term frequency is additionally divided by the length of the text to normalize it.

$$TF(t) = \frac{\text{No. of times } t \text{ appears in a document}}{\text{Total no. of terms in the document}} \quad (1)$$

The IDF is a measure of how often a word appears in a document. Based on the word's frequency, IDF shows the importance of a word. The IDF score for rare words is higher.

$$IDF(t) = \log_e \frac{\text{Total no. documents}}{\text{No. of documents that contain term } t} \quad (2)$$

TF-IDF is then calculated by multiplying TF and IDF.

3) Word2Vec

Word2Vec is a feature extraction technique used for text processing. Developed by Tomas Mikolov *et al.* at Google, it uses a neural network model as the base to learn the association between words from the text. The Word2Vec method extracts text features for specific words. When the model is trained on a text corpus, it can detect synonymous words or recommend additional words for partial sentences [25]. A list of numbers called 'vector' is used in Word2Vec to represent each different

word. The selection of the vectors is made such that the cosine similarity can be used to determine the semantic similarity between the vectors.

D. EMOTION DETECTION AND SENTIMENT ANALYSIS METHODS

1) TextBlob

Textblob is a famous Python library used for performing different tasks on text data. Several processes on text data can be carried out using TextBlob including noun phrase extraction, translation and sentiment analysis, etc. [26]. This study uses TextBlob to find the sentiments from cryptocurrency tweets. The TextBlob sentiment function is used to determine the text polarity and subjectivity. Subjectivity is used to check whether the text is objective or subjective and polarity is used to check whether the text gives a positive or negative sense.

2) Text2Emotion

Text2Emotion is a Python package created to identify appropriate emotions embedded in the text data [27]. Humans use emotions in the appropriate context while communicating and the words used to represent those emotions are properly aligned. Text2Emotion processes the textual data identifies the embedded emotion and provides the output in the form of a dictionary. Five main emotion types are well-suited for tweets including 'happy', 'angry', 'sad', 'surprise', and 'fear'.

E. MACHINE LEARNING MODELS

This study uses several machine learning and deep learning models for sentiment analysis and emotion detection. For example, RF, DT, KNN, SVM, GNB, and LR are used for experiments. In addition, ELM is also used for performance comparison. ELM does not use a gradient-based technique, so it tunes hyperparameters once and does not follow an iterative approach which improves its computation time than traditional machine learning models [28]–[30]. For the current study, ELM is deployed with the SoftMax layer and hot encoding function. The machine learning models are used with the best hyperparameter settings which are obtained using the grid search method. The hyperparameters setting of all machine learning models is shown in Table 6.

F. PROPOSED ENSEMBLE MODEL

This study proposes an ensemble model combining the LSTM and GRU for sentiment analysis and emotion detection [31], [32]. LSTM-GRU ensembles LSTM and GRU sequentially which are types of recurrent neural networks (RNN). GRU is a newer application of RNN as compared to LSTM but both are suited for text data and therefore combined to obtain high classification accuracy. In LSTM-GRU, the output of the LSTM will be the input for the GRU for the prediction. The architecture of the LSTM-GRU is shown in Figure 3. The first layer of the ensemble model is an embedding layer with a 5000 vocabulary size

TABLE 6. Hyper-parameters setting for machine learning models.

Model	Hyperparameters	Tuning range
Random Forest	n_estimators= 300; max_depth=300	n_estimators= 50 to 500; max_depth=50 to 500;
Decision Tree	max_depth=300	n_estimators= 50 to 500;
K Nearest Neighbour	n_neighbors= 3; weights=uniform	n_neighbors= 1 to 5; weights= {uniform, distance}
Support Vector Machine	kernel= linear; C=3.0	kernel= {linear, poly}; C=1.0 to 5.0
Gaussian Naive Bayes	var_smoothing= 1e-9	var_smoothing= 1e-9
Logistic Regression	solver= liblinear; C=3.0	solver= {liblinear, saga, sag}; C=1.0 to 5.0

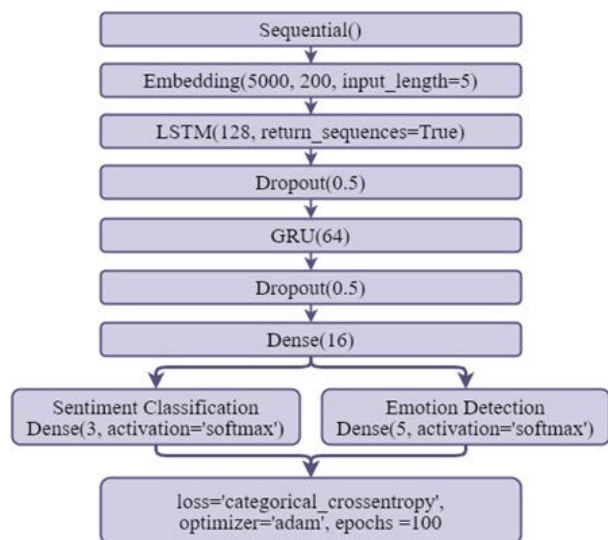


FIGURE 3. Structure of the proposed deep learning based ensemble model.

and 200 output dimensions. This embedding layer is followed by the LSTM layer with 128 units. LSTM layer processes the input data and produces a significant feature sequence for GRU. A dropout layer is added right after LSTM and GRU layer with a 0.5 dropout rate. This dropout layer helps to reduce the complexity of the ensemble model. GRU layer is used with 64 units and it processes the output by LSTM. Then a dense layer with 16 neurons is used to handle sparse output by the GRU. After that, the output layer with the different number of neurons for sentiment analysis and emotion detection is added based on the number of target classes. LSTM-GRU is compiled with the categorical_crossentropy function because of the multi-class problem and ‘adam optimizer’ is used for training [33]. LSTM-GRU is trained using 100 epochs.

IV. RESULTS AND DISCUSSIONS

This section presents the result of the proposed approach for sentiment analysis and emotion detection using machine learning and deep learning models. Results are presented with each feature extraction technique separately in terms of accuracy, precision, recall, F1 score, number of correct predictions (CP), number of wrong predictions (WP), and geometric mean (G mean) [34].

A. EMOTION DETECTION RESULTS USING MACHINE LEARNING MODELS

Experiments are performed using the machine learning models with BoW, TF-IDF, and Word2Vec features. Experimental results indicate that SVM and LR show more significant results as compared to other models with BoW features, each with a 0.90 accuracy score. This significant performance with both linear models is based on the large feature set. The 40,000 tweets generate a large feature set that is suitable for the SVM and LR. While ETC and DT achieve 0.85 and 0.84 accuracy scores, respectively, KNN and GNB could not perform well. Overall, SVM and LR show a superb performance with a large feature set and multiclass prediction. The results of machine learning models with BoW are shown in Table 7.

TABLE 7. Results using machine learning models with BoW features.

Model	Accuracy	Precision	Recall	F1 score	G mean
SVM	0.90	0.89	0.89	0.89	0.89
LR	0.90	0.90	0.88	0.89	0.89
GNB	0.38	0.41	0.44	0.35	0.42
ETC	0.85	0.88	0.77	0.81	0.82
DT	0.84	0.83	0.79	0.80	0.82
KNN	0.69	0.81	0.56	0.63	0.67

Results of machine learning models are also good with TF-IDF features. SVM and LR are the best performers with the TF-IDF features as well, each with a 0.90 accuracy score. The performance of models is slightly decreased when TF-IDF features are used. Although the performance of SVM and LR is not affected, other models’ performance is degraded, e.g., ETC, DT, KNN, and GNB. It indicates that the performance is affected with TF-IDF features due to complex features set as compared to simple BoW features.

Table 9 shows the performance of models using the Word2Vec features for emotion detection. The performance of models is not good with word2vec features as compared to TF-IDF and BoW features. LR achieves the highest 0.76 accuracy score for emotion detection using Word2Vec features. This poor performance of the model is attributed to Word2Vec’s inability to handle unknown or out-of-vocabulary (OOV) words. If a model has not encountered a word before, it can not interpret or build a vector for it. Consequently, the performance of machine learning models is substantially reduced when used with Word2Vec features.

Table 10 shows the performance comparison of machine learning models for emotion detection in terms of correct

TABLE 8. Results using machine learning models with TF-IDF features.

Model	Accuracy	Precision	Recall	F1 score	G mean
SVM	0.90	0.88	0.86	0.87	0.87
LR	0.90	0.90	0.84	0.87	0.87
GNB	0.37	0.40	0.44	0.35	0.42
ETC	0.84	0.89	0.75	0.80	0.82
DT	0.83	0.81	0.77	0.79	0.79
KNN	0.69	0.81	0.56	0.63	0.67

TABLE 9. Results using machine learning models with Word2Vec features.

Model	Accuracy	Precision	Recall	F1 score	G mean
SVM	0.75	0.74	0.67	0.70	0.70
LR	0.76	0.74	0.68	0.70	0.71
GNB	0.42	0.47	0.43	0.37	0.45
ETC	0.74	0.87	0.60	0.67	0.72
DT	0.63	0.55	0.55	0.55	0.55
KNN	0.72	0.74	0.61	0.66	0.67

TABLE 10. Results comparison using machine learning models.

Model	BoW		TF-IDF		Word2Vec	
	CP	WP	CP	WP	CP	WP
SVM	5,417	583	5,389	611	4,514	1,486
LR	5,406	594	5,375	625	4,539	1,461
GNB	2,251	3,749	2,243	3,757	2,526	3,474
ETC	5,102	898	5,054	946	4,435	1,565
DT	5,048	952	4,964	1,036	3,784	2,216
KNN	4,117	1,883	4,117	1,883	4,315	1,685

predictions (CP) and wrong predictions (WP). The highest number of correct predictions are obtained by the SVM using BoW features which is 5,417 out of 6,000 and gives the lowest wrong predictions of 583 as compared to all other models. LR is just behind the SVM with 5,406 correct predictions. GNB gives the highest number of wrong predictions for emotion detection using the BoW features with 3,749 wrong predictions out of 6000 total predictions. Figure 4 shows the comparison between feature extraction techniques concerning machine learning models for sentiment analysis.

B. SENTIMENT ANALYSIS USING MACHINE LEARNING MODELS

This section presents the results of machine learning models in terms of accuracy, precision, recall, F1 Score, WP, and CP for sentiment analysis. Sentiment analysis is based on three sentiments including negative, positive, and neutral as compared to emotion detection which involved five target classes. Consequently, better performance is expected for sentiment analysis with the machine learning models.

Table 11 contains the result of machine learning models for sentiment analysis using the BoW features and SVM and LR outperform all other models, each with an accuracy score of 0.98. In terms of F1 score and recall SVM performs significantly better than LR. SVM’s significant performance in both emotion detection and sentiment analysis shows that the large feature sets are more suitable for SVM. DT also achieved a 0.98 accuracy score while ETC is just behind with

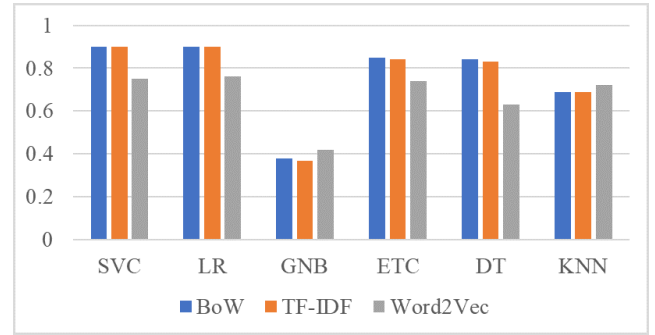


FIGURE 4. Comparison between feature extraction techniques with respect to machine learning models for emotion detection.

TABLE 11. Sentiment analysis results using BoW features.

Model	Accuracy	Precision	Recall	F1 score	G mean
SVM	0.98	0.98	0.97	0.98	0.97
LR	0.98	0.98	0.96	0.97	0.97
GNB	0.47	0.55	0.57	0.45	0.56
ETC	0.97	0.97	0.93	0.95	0.95
DT	0.98	0.97	0.97	0.97	0.97
KNN	0.78	0.86	0.67	0.72	0.76

a 0.97 accuracy score, however, their precision, recall, and F1 scores are marginally lower than SVM and LR. GNB still shows the worst performance with a 0.47 accuracy score.

The performance of models using TF-IDF features is given in Table 12. Results indicate that SVM and two tree-based models ETC and DT jointly achieve the highest accuracy score of 0.98. The performance of the ETC is improved with TF-IDF features. ETC is a tree-based ensemble model that uses majority voting criteria to predict the target class and tends to show better performance with few target classes as compared to a large number of target classes. ETC is followed by LR with a 0.97 accuracy score.

TABLE 12. Sentiment analysis results using TF-IDF features.

Model	Accuracy	Precision	Recall	F1 score	G mean
SVM	0.98	0.98	0.96	0.97	0.97
LR	0.97	0.98	0.94	0.96	0.95
GNB	0.45	0.58	0.57	0.44	0.57
ETC	0.98	0.97	0.95	0.96	0.96
DT	0.98	0.96	0.96	0.96	0.96
KNN	0.78	0.86	0.67	0.72	0.69

The performance of models using Word2Vec features is shown in Table 13 and results suggest that models’ performance is severely degraded when used with Word2Vec features. Although SVM and LR achieve the highest accuracy score of 0.89 each, the classification accuracy of other machine learning models is reduced on average. However, GNB achieves its highest accuracy score of the study with Word2Vec features in the sentiment analysis case as it gives a 0.61 accuracy score.

For sentiment analysis, the correct and wrong prediction ratio for each model is provided in Table 14. SVM gives the

TABLE 13. Sentiment analysis results using Word2Vec features.

Model	Accuracy	Precision	Recall	F1 score	G mean
SVM	0.89	0.88	0.82	0.85	0.85
LR	0.89	0.88	0.82	0.85	0.85
GNB	0.61	0.56	0.63	0.56	0.59
ETC	0.82	0.89	0.72	0.77	0.74
DT	0.74	0.67	0.68	0.68	0.68
KNN	0.82	0.89	0.75	0.78	0.76

TABLE 14. Sentiment analysis results comparison using machine learning models.

Model	BoW		TF-IDF		Word2Vec	
	CP	WP	CP	WP	CP	WP
SVM	5,896	104	5,895	105	5,322	678
LR	5,879	121	5,845	155	5,326	674
GNB	2,846	3,154	2,705	3295	3,657	2,343
ETC	5,813	187	5,860	140	4,942	1,058
DT	5,890	110	5,871	129	4,443	1,557
KNN	4,673	1,327	4,673	1,327	4,895	1,105

highest number of correct predictions for sentiment analysis using the BoW and TF-IDF features with 5,896 and 5,895 correct predictions, respectively, out of a total of 6,000 predictions. GNB is the worst performer with 3,294 wrong predictions using TF-IDF features for sentiment analysis. Figure 4 shows the comparison between feature extraction techniques concerning machine learning models for sentiment analysis. Figure 5 shows the comparison between feature extraction techniques concerning machine learning models for sentiment analysis.

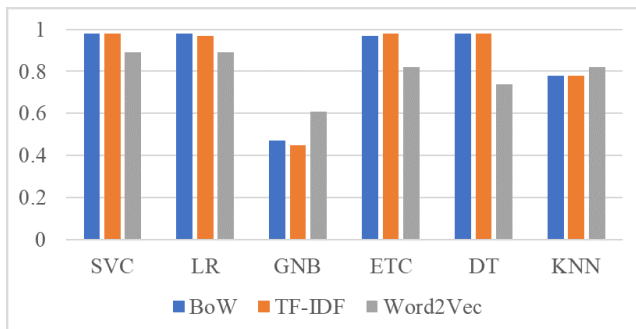


FIGURE 5. Comparison between feature extraction techniques with respect to machine learning models for sentiment analysis.

C. EMOTION DETECTION AND SENTIMENT ANALYSIS USING DEEP LEARNING MODELS

Besides using the machine learning models, two deep learning models are also used for experiments including LSTM and GRU, in addition to the proposed ensemble model. The performance of the proposed model is shown in Table 15. The ensemble recurrent structure model LSTM-GRU outperforms with significant performance in comparison with other models including both machine and deep learning models. The highest achieved accuracy by the models is highlighted in bold. In both cases, LSTM-GRU performs well in terms of

TABLE 15. Sentiment analysis and emotion detection results using deep learning models.

Approach	Model	Acc.	Prec.	Recall	F1	G mean
Emotion Detection	LSTM	0.90	0.89	0.89	0.89	0.89
	GRU	0.91	0.89	0.89	0.89	0.89
	LSTM-GRU	0.91	0.90	0.89	0.89	0.89
Sentiment Analysis	LSTM	0.99	0.98	0.97	0.97	0.97
	GRU	0.98	0.98	0.97	0.98	0.98
	LSTM-GRU	0.99	0.99	0.98	0.98	0.98

accuracy, precision, recall, number of correct predictions, and number of wrong predictions.

Table 15 shows that for emotion detection with five target classes GRU shows better performance as compared to LSTM with 5,435 correct predictions while LSTM gives 5,426 correct predictions. Conversely, when LSTM and GRU are ensembled, they show superior performance with 5,482 correct predictions as shown in Table 16. Similarly, in the sentiment analysis case, LSTM is better as compared to GRU which shows that their performance varies as the ratio of target classes changes. LSTM gives 5,913 correct predictions while GRU gives 5,902. However, when combined into an ensemble, the number of correct predictions is increased to 5,828. These results show that the proposed model LSTM-GRU can perform better because of its ensemble structure where LSTM and GRU are combined to get the better of both models. This versatility makes LSTM-GRU a significant model as compared to an individual model.

TABLE 16. Results comparison using deep learning models.

Approach	Model	CP	WP
Emotion Detection	LSTM	5,426	574
	GRU	5,435	565
	LSTM-GRU	5,482	518
Sentiment Analysis	LSTM	5,913	87
	GRU	5,902	98
	LSTM-GRU	5,928	72

D. RESULTS WITH RANDOM UNDERSAMPLING BALANCED DATASET

Since the dataset is balanced so the probability of the models' overfitting can not be ignored. For handling this problem, this study carries out the data balancing using the random undersampling (RUS) approach. Undersampling is preferred over oversampling, as data leakage may occur when oversampling is performed before the train-test split. RUS decreases the number of records from the dataset to achieve an almost similar data distribution for different classes. For this purpose, the number of majority class samples is randomly removed. In the dataset used in this study, the negative class has only 3,712 samples which is less than both the positive and neutral classes, so RUS is used to select an equal number of records for positive and neutral classes to make the dataset balanced. Similarly, for emotions, the angry class has a lower number of records, and the RUS approach is used to select the same number of records from each class. Table 17 shows the distribution of samples after applying RUS.

TABLE 17. Number of samples for different classes after random undersampling.

Dataset	Class	Original count	After sampling
Sentiment analysis	Positive	14,541	3,712
	Neutral	21,747	3,712
	Negative	3,712	3,712
Emotion detection	Happy	17,236	1,503
	Sad	5,175	1,503
	Fear	9,176	1,503
	Surprise	6,910	1,503
	Angry	1,503	1,503

Experiments are performed using the under-sampled dataset using the proposed LSTM-GRU model and results are given in Table 18. Results suggest that the performance of the model is better with the original dataset as compared to the balanced dataset using the RUS approach. RUS removes samples from the dataset which ultimately reduces the feature set size. Deep learning models are data-intensive and require a larger data size to obtain a good fit which is not possible with undersampled data. It degrades the performance of LSTM-GRU from a 0.99 accuracy score to 0.97 for the sentiment analysis and a 0.90 accuracy score to 0.83 for emotion detection. Despite the decrease in the classification accuracy, the models still perform better than machine learning models with good precision and F1 scores.

TABLE 18. Experimental results using LSTM-GRU on the balanced dataset.

Dataset	Accuracy	Precision	Recall	F1 score
Sentiment analysis	0.97	0.97	0.97	0.97
Emotion detection	0.83	0.83	0.82	0.82

E. COMPARISON FOR COMPUTATIONAL COMPLEXITY

We evaluate the computational time for all models with each approach and the results are shown in Tables 19 and 20. According to the results performance of the proposed model LSTM-GRU is significantly better in terms of computational time as the model takes only 1178 seconds for sentiment analysis and 1438 seconds for emotion detection. While from machine learning models, BoW and TF-IDF give good results but computational time for ETC and DT is very high as ETC takes 9351 seconds with the TF-IDF feature for sentiment analysis. On the other hand, GNB takes the lowest time but accuracy is very low by GNB. For the deep learning models, the computational time of the LSTM model is significantly low than both the GRU and the proposed model, however, its accuracy is also low.

F. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART APPROACHES

For analyzing the significance of the proposed LSTM-GRU model, this study carries out the performance analysis concerning other state-of-the-art approaches for sentiment analysis. Four recent studies have been selected in this regard

TABLE 19. Computational time (seconds) of machine learning models.

Model	Emotion Detection			Sentiment Analysis		
	BoW	TF-IDF	Word2Vec	BoW	TF-IDF	Word2Vec
SVC	333	348	241	175	195	270
LR	321	349	166	3	195	245
GNB	43	46	229	48	31	2
ETC	2704	2664	198	2345	9351	198
DT	1403	1521	149	1654	1677	149
KNN	3567	1977	122	2121	2447	117

TABLE 20. Deep learning models' computational time (seconds).

Approach	Model	Time (Sec.)
Sentiment Analysis	LSTM	127
	GRU	901
	LSTM-GRU	978
Emotion Detection	LSTM	201
	GRU	1015
	LSTM-GRU	964

TABLE 21. Performance comparison with state-of-the-art approaches.

Ref.	Year	Model	Task	Acc.
[35]	2020	LR-SGDC	Sentiment analysis	0.98
			Emotion detection	0.90
[31]	2021	CNN-LSTM	Sentiment analysis	0.93
			Emotion detection	0.83
[36]	2021	Stacked Bi-LSTM	Sentiment analysis	0.97
			Emotion detection	0.90
[37]	2021	ConBiLSTM	Sentiment analysis	0.95
			Emotion detection	0.89
Current study	2021	LSTM-GRU	Sentiment analysis	0.99
			Emotion detection	0.91

that uses deep learning models for emotion detection and sentiment analysis.

Study [35] perform sentiment analysis on users' tweets for the US airline companies using an ensemble model. The ensemble model uses the LR and SGDC under voting criteria. The study [31] conducts sentiment analysis using deep learning CNN and LSTM by stacking them for high performance. The ensemble mode is used to enhance the capability of sarcasm detection in tweets. Similarly, [36] carried out experiments for sentiment analysis on tweets related to deepfake technology. A stacked Bi-LSTM model is proposed to obtain improved results. Along the same directions, [37] proposes an ensemble model, convolutional Bi LSTM (ConBi-LSTM) by combining CNN and BiLSTM for sentiment analysis. Table 21 shows the comparative analysis of the above-discussed research works and the proposed LSTM-GRU model for sentiment analysis and emotion detection. Results suggest that the proposed model achieves results that are significantly better than state-of-the-art approaches.

We also deployed the ELM to analyze its performance against the proposed model and other machine learning models. ELM is faster as compared to traditional neural networks as it does not use the gradient-based technique. However, its performance is poor on the selected dataset for both sentiment analysis and emotion detection, as shown in Table 22. The best performance is observed when used with TF-IDF for

TABLE 22. ELM results using BoW, TF-IDF, and Word2Vec features.

Approach	BoW	TF-IDF	Word2Vec
Sentiment Analysis	0.64	0.68	0.61
Emotion Detection	0.60	0.65	0.60

sentiment analysis where it obtains a 0.68 accuracy score, followed by a 0.64 accuracy score using BoW. The performance of ELM is poor because of non-linear text features found in the current dataset.

V. CONCLUSION

This study performs sentiment analysis and emotion detection on tweets related to cryptocurrency. Sentiment analysis of cryptocurrency holds potential significance as it is widely used for predicting the market price of the cryptocurrency which necessitates sentiments classification with high accuracy. For experiments, tweets are extracted from TwitterTM, and the dataset is annotated using TextBlob and Text2Emotion for sentiments and emotions, respectively. Besides the use of several machine learning and deep learning models for classification, this study leverages recurrent neural networks LSTM and GRU to form an ensemble model to enhance classification performance. In addition, BoW, TF-IDF, and Word2Vec features are used as feature extraction techniques for the machine learning models. Results indicate that machine learning models perform well with BoW features compared with TF-IDF and Word2Vec. The proposed model achieves the highest performance for sentiment analysis with a 0.99 accuracy score and the highest precision and recall of 0.99 and 0.98, respectively. Similarly, LSTM-GRU outperforms all other models in terms of correct and wrong predictions for both sentiment analysis and emotion detection. Dataset balancing using the random undersampling suggests that LSTM-GRU performance is decreased due to fewer training data. This study considers the sentiment analysis for cryptocurrency-related tweets, we intend to perform cryptocurrency market price prediction based on the analyzed sentiments in the future.

ACKNOWLEDGMENT

(Naila Aslam and Furqan Rustam are co-first authors.)

REFERENCES

- [1] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," *SMU Data Sci. Rev.*, vol. 1, no. 3, p. 1, 2018.
- [2] S. Colianni, S. Rosales, and M. Signorotti, "Algorithmic trading of cryptocurrency based on Twitter sentiment analysis," CS229 Project, Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015, pp. 1–5.
- [3] A. Inamdar, A. Bhagatani, S. Bhatt, and P. M. Shetty, "Predicting cryptocurrency value using sentiment analysis," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 932–934.
- [4] D. L. K. Chuen, L. Guo, and Y. Wang, "Cryptocurrency: A new investment opportunity?" *J. Alternative Investments*, vol. 20, no. 3, pp. 16–40, 2017.
- [5] K. Wołk, "Advanced social media sentiment analysis for short-term cryptocurrency price prediction," *Expert Syst.*, vol. 37, no. 2, p. e12493, Apr. 2020.
- [6] C. Lamon, E. Nielsen, and E. Redondo, "Cryptocurrency price prediction using news and social media sentiment," *SMU Data Sci. Rev.*, vol. 1, no. 3, pp. 1–22, 2017.
- [7] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing Twitter data," *Int. J. Data Sci. Anal.*, vol. 7, no. 1, pp. 35–51, Feb. 2019.
- [8] S. Sharifirad, B. Jafarpour, and S. Matwin, "How is your mood when writing sexist tweets? Detecting the emotion type and intensity of emotion using natural language processing techniques," 2019, *arXiv:1902.03089*.
- [9] F. Calefato, F. Lanubile, and N. Novielli, "EmoTxt: A toolkit for emotion recognition from text," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos (ACIIW)*, Oct. 2017, pp. 79–80.
- [10] S. A. Salam and R. Gupta, "Emotion detection and recognition from text using machine learning," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 6, pp. 341–345, Jun. 2018.
- [11] F. M. Shah, A. S. Reyadh, A. I. Shaafi, S. Ahmed, and F. T. Sithil, "Emotion detection from tweets using AIT-2018 dataset," in *Proc. 5th Int. Conf. Adv. Electr. Eng. (ICAEE)*, Sep. 2019, pp. 575–580.
- [12] D. Haryadi and G. Putra, "Emotion detection in text using nested long short-term memory," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 1–7, 2019.
- [13] F. Ghanbari-Adivi and M. Mosleh, "Text emotion detection in social networks using a novel ensemble classifier based on Parzen tree estimator (TPE)," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8971–8983, Dec. 2019.
- [14] E. Junianto and R. Rachman, "Implementation of text mining model to emotions detection on social media comments using particle swarm optimization and naive Bayes classifier," in *Proc. 7th Int. Conf. Cyber IT Service Manage. (CITSM)*, Nov. 2019, pp. 1–6.
- [15] X. Zhang, W. Li, H. Ying, F. Li, S. Tang, and S. Lu, "Emotion detection in online social networks: A multilabel learning approach," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8133–8143, Sep. 2020.
- [16] D. Seal, U. K. Roy, and R. Basak, "Sentence-level emotion detection from text based on semantic rules," in *Information and Communication Technology for Sustainable Development*. Singapore: Springer, 2020, pp. 423–430.
- [17] S. M. Raju and A. M. Tarif, "Real-time prediction of BITCOIN price using machine learning techniques and public sentiment analysis," 2020, *arXiv:2006.14473*.
- [18] X. Huang, W. Zhang, X. Tang, M. Zhang, J. Surbiryala, V. Iosifidis, Z. Liu, and J. Zhang, "LSTM based sentiment analysis for cryptocurrency prediction," 2021, *arXiv:2103.14804*.
- [19] E. Şaşmaz and F. B. Tek, "Tweet sentiment analysis for cryptocurrencies," in *Proc. 6th Int. Conf. Comput. Sci. Eng. (UBMK)*, Sep. 2021, pp. 613–618.
- [20] T. Mehta, G. Kolase, V. Tekade, R. Sathe, and A. Dhawale, "Price prediction and analysis of financial markets based on news, social feed, and sentiment index using machine learning and market data," *Int. Res. J. Eng. Technol.*, vol. 7, no. 6, 2020. [Online]. Available: <https://www.irjet.net/archives/V7/i6/IRJET-V71688.pdf>
- [21] S. L. Cessie and J. C. V. Houwelingen, "Ridge estimators in logistic regression," *J. Roy. Stat. Soc., C, Appl. Statist.*, vol. 41, no. 1, pp. 191–201, 1992.
- [22] F. Pedregosa, S. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Dec. 2011.
- [23] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Amsterdam, The Netherlands: Elsevier, 2014.
- [24] 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, Feb. 2021, Art. no. e0245909.
- [25] S. Khomsah, "Sentiment analysis on YouTube comments using word2vec and random forest," *Telematika, Jurnal Informatika dan Teknologi Informatika*, vol. 18, no. 1, pp. 61–72, 2021.
- [26] 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, Feb. 2021, Art. no. e0245909.
- [27] S. Dhar and I. Bose, "Emotions in Twitter communication and stock prices of firms: The impact of COVID-19 pandemic," *Decision*, vol. 47, no. 4, pp. 385–399, Dec. 2020.
- [28] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, "Non-iterative and fast deep learning: Multilayer extreme learning machines," *J. Franklin Inst.*, vol. 357, no. 13, pp. 8925–8955, 2020.
- [29] J. Zhang, W. Xiao, Y. Li, S. Zhang, and Z. Zhang, "Multilayer probability extreme learning machine for device-free localization," *Neurocomputing*, vol. 396, pp. 383–393, Jul. 2020.

- [30] W. Xiao, J. Zhang, Y. Li, S. Zhang, and W. Yang, "Class-specific cost regulation extreme learning machine for imbalanced classification," *Neurocomputing*, vol. 261, pp. 70–82, Oct. 2017.
- [31] 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, Aug. 2021.
- [32] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah, A. A. Reshi, and I. Ashraf, "Sentiment analysis and topic modeling on tweets about online education during COVID-19," *Appl. Sci.*, vol. 11, no. 18, p. 8438, Sep. 2021.
- [33] Z. Zhang, "Improved Adam optimizer for deep neural networks," in *Proc. IEEE/ACM 26th Int. Symp. Quality Service (IWQoS)*, Jun. 2018, pp. 1–2.
- [34] R. P. Espindola and N. F. Ebecken, "On extending F-measure and G-mean metrics to multi-class problems," *WIT Trans. Inf. Commun. Technol.*, vol. 35, pp. 25–34, May 2005.
- [35] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. Choi, "Tweets classification on the base of sentiments for U.S. Airline companies," *Entropy*, vol. 21, no. 11, p. 1078, Nov. 2019.
- [36] V. Rupapara, F. Rustam, A. Amaar, P. B. Washington, E. Lee, and I. Ashraf, "Deepfake tweets classification using stacked bi-LSTM and words embedding," *PeerJ Comput. Sci.*, vol. 7, p. e745, Oct. 2021.
- [37] S. Tam, R. B. Said, and Ö. Ö. Tanrıöver, "A ConvBiLSTM deep learning model-based approach for Twitter sentiment classification," *IEEE Access*, vol. 9, pp. 41283–41293, 2021.



ERNESTO LEE is currently working as a Professor with the Department of Computer Science, Broward College, Broward County, FL, USA. His current research interests include block-chain, the IoT, data mining, mainly working machine learning & deep learning-based IoT and text mining tasks.



PATRICK BERNARD WASHINGTON is currently working as a Professor with the Division of Business Administration and Economics, Morehouse College, Atlanta, GA, USA. His current research interests include block-chain, the IoT, data mining, mainly working machine learning & deep learning-based IoT and text mining tasks.



NAILA ASLAM received the B.S. and M.S. degrees in computer science from The Islamia University of Bahawalpur, Pakistan. She is currently pursuing the Ph.D. degree from the Hebei University of Technology, Tianjin, China. Since, she has been working as a Lecturer with the Department of Computer Science, University of Management and Technology, Lahore, Pakistan. Her research interests include artificial intelligence, data mining, and machine learning.



FURQAN RUSTAM received the Master of Science degree in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. He worked as a Research Assistant with the Fareed Computing & Research Center, KFUEIT. He is currently working as a Lecturer with the Department of Software Engineering, University of Management and Technology, Lahore, Pakistan. His current research interests include data mining, machine learning, and artificial intelligence, mainly working on creative computing and supervised machine learning.



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 & communication engineering from Yeungnam University, Gyeongsan-si, South Korea, in 2018. He has worked as a Postdoctoral Fellow with Yeungnam University, where he is currently working as an Assistant Professor with the Information and Communication Engineering Department. His research interests include indoor positioning and localization, indoor location-based services in wireless communication, smart sensors (LIDAR) for smart cars, and data mining.

...