IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 November 2023, accepted 7 December 2023, date of publication 13 December 2023, date of current version 20 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3342688

RESEARCH ARTICLE

Cryptocurrency Price Prediction Model Based on Sentiment Analysis and Social Influence

FATEMEH FEIZIAN AND BABAK AMIRI
School of Industrial Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran
Corresponding author: Babak Amiri (babakamiri@iust.ac.ir)

ABSTRACT Cryptocurrency as an alternative method of payment that acts both as a type of currency and as a virtual accounting system has always been of interest to investors. Since the public sentiment of a society about cryptocurrencies can affect the cryptocurrencies' prices, a machine learning model based on sentiment analysis has been proposed to forecast the future prices of cryptocurrencies such as Bitcoin, Ethereum, EOS, Cardano, and Ripple using machine learning models that are suitable for time series data analysis to reduce the risk of investing in this market. It was shown that by applying weights to the sentiment scores of tweets according to the influence factor of the individuals, the accuracy of the prediction will increase and a significant difference between the accuracy scores was observed using the LSTM model according to the MAPE indicator ($P = 0.045$). Also, a hybrid model is proposed based on the combination of features extracted from the texts by one of the dictionary-based text analysis models and the feature of weighted sentiment scores. It was shown that our proposed hybrid model outperformed the other models in predicting the prices of Ethereum, EOS, and Cardano according to the MSE indicator. Also, our proposed model based on weighted sentiment scores according to the influence factor of the Twitterers outperformed the other models in the prediction of the future prices of Bitcoin and Ripple, which indicates that the increase in the number of features will not always lead to an increase in the accuracy of our prediction models.

INDEX TERMS Cryptocurrency, machine learning, regression prediction model, sentiment analysis.

ACRONYMS

LSTM	Long Short-Term Memory.
ARIMA	Autoregressive Integrated Moving Average.
MLP	Multilayer Perceptron.
WLPS	Weightless Polarity Scores.
IBWPS	Influence-Based Polarity Score.
DBSA	Dictionary-based Sentiment Analysis.
HMSA	Hybrid Model Sentiment Analysis.
LIWC	Linguistic Inquiry Word Count.
MSE	Mean Suare Error.
RMSE	Root Mean Square Error.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.

I. INTRODUCTION

Cryptocurrency has been designed as a medium of exchange and can be considered a type of digital currency that is distinguished from traditional currencies because it is based on the principle of decentralized control, and the legal payment status of cryptocurrencies has been authorized by many countries [1]. Cryptocurrencies are being used all over the world, and many people are investing in this area [2].

It has been shown that Twitter, as a source for reflecting social emotions, can be used to predict the price fluctuations of cryptocurrencies [3]. Through sentiment analysis of the tweets, valuable insight can be provided to predict the future prices of cryptocurrencies [4]. Also, the volume of transactions for one of the cryptocurrencies, along with the sentiment of the tweets, can be used to predict the future prices of cryptocurrencies. It was proven that an increase in the average polarity scores of the tweets related to cryptocurrencies would result in an increase in the exchange volume and prices of Bitcoin [5]. In recent studies, techniques such as machine learning, natural language processing, and time series data analysis have been used to identify the patterns of cryptocurrencies' price fluctuations [6]. Traders can get the most financial benefit by buying and selling cryptocurrencies

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

on time. To make the right decision regarding the purchase or sale of cryptocurrencies, it is necessary to create a strong predictive system that can predict future prices with acceptable accuracy through the use of machine learning methods and by analyzing the opinions of individuals on social media, considering the historical prices of cryptocurrencies. Twitter, as a powerful inclusive social network in terms of the number of active users and a place for the wide reflection of their opinions, was selected in recent studies as a source of public opinions in the community. Through sentiment analysis of the tweets, it is possible to identify the negative or positive feelings associated with the published contents. In previous studies, it was shown that the volume of messages published by users on social networks can also be useful in predicting the prices of cryptocurrencies [7]. Time series data analysis methods have been used to analyze a sequence of data points collected over time intervals. This modeling approach can be used if there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables or if little knowledge is available about the underlying data-generating process [8]. The approaches to time series forecasting are traditional statistical models, including moving average, exponential smoothing, and autoregressive integrated moving average (ARIMA), which are all linear models in which the predictions of future values are constrained to be linear functions of past observations. ARIMA combines three different processes: an auto-regressive (AR) function regressed on past values of the process, an integrated (I) part to make the data series stationary by differencing, and a moving average (MA) function regressed on a purely random process. According to these hybrids, while the neural network model deals with nonlinearity, the ARIMA model deals with the non-stationary linear component [9]. To overcome the linear limitation of time series models, some nonlinear models have been proposed, such as multilayer perceptrons (MLPs), which are a subset of artificial neural networks [10]. Multilayer perceptrons can discriminate data that is not linearly separable and use a supervised learning technique called backpropagation for training. MLP, as a deep learning method, has a distributed memory and has the ability to work with insufficient knowledge [11]. Recurrent neural networks (RNNs), which have been widely adopted in all research areas, are not able to learn the relevant information from input data while the input gap is large. The long short-term memory (LSTM) model, because it has some gate functions, can handle the problem of long-term dependencies, so the LSTM model has become the focus of deep learning [12]. According to the previous research, with the availability of enough training data, LSTM will usually outperform, while ARIMA works better for smaller data sets [13]. In this study, three deep learning models that, according to previous studies, had acceptable results in predicting cryptocurrencies' prices have been selected. These models are ARIMA, LSTM, and MLP. The purpose of choosing these three models is to predict the prices of each of the cryptocurrencies as accurately as possible. Since any learning model that is efficient for the price prediction of one of the digital currencies may not be appropriate for another, it has been tried to use different learning models. Meanwhile, in this study, four distinct methods of sentiment analysis have been proposed, and the performance of each learning model for each of these methods has been measured according to the evaluation metrics. According to previous studies, sentiment analysis can be used to predict the prices of digital currencies [14], [15], [16], but using this method alone is associated with shortcomings.

According to the study conducted by Abraham et al., it was shown that tweet volumes about Bitcoin and Google Trends search queries were found to be associated with Bitcoin price fluctuations, while tweet volumes showed a slightly higher positive correlation with the price fluctuations than Google Trends [7]. It has been proven that deep learning models such as convolutional neural networks and different types of recurrent neural networks, including the long short term memory network, the stacked long short term memory network, the bidirectional long short term memory network, and the gated recurrent unit network, can be utilized to predict the cryptocurrency closing prices in real time with promising accuracy [17]. In other study, it was concluded that some factors, such as Ethereum-specific blockchain information, macro-economy factors, and the blockchain information of other cryptocurrencies, play roles in predicting Ethereum prices [18].

Kraaijeveld and De Smedt, using lexicon-based sentiment analysis, explored the predictive power of Twitter sentiment and found that the sentiment of the tweets had strong predictive power for predicting the prices of cryptocurrencies such as Bitcoin and Litecoin [19]. In the study conducted by Rezaei et al., a hybrid algorithm was proposed with the capability of extracting deep features and time sequences [20]. The results showed that the combination of CNN alongside LSTM and Complete Ensemble Empirical Mode Decomposition (CEEMD) or Empirical Mode Decomposition (EMD) will enhance the analytical power of the prediction model. In a study by Bhatt et al., various models were trained over historical data on cryptocurrencies. It was found that adding sentiment features to the prediction model resulted in better performance [21].

In the study conducted by Raj Pant in 2018, it was assumed that some tweets were more important than others, and by giving weight to their sentiment, more accurate results were obtained. In Raj Pant's study, a predefined list of important people, organizations, and countries was used to identify the important tweets to give double weight to their sentiment [22], while in our study, all the tweets got weights based on the normalized number of followers of the tweet's publishers. The innovative aspects of our research include the use of machine learning algorithms along with a new approach to sentiment analysis that provides more accurate results in the prediction of future prices of cryptocurrencies by introducing a new feature for our prediction model obtained from weighted polarity scores based on users' influence. Knowing that the individuals in social networks have different powers of influence over each other, the application of weights to the sentiment scores of the tweets can be done based on the number of followers of the tweet's publishers. The difference between the study conducted by Raj Pant and our study was the way we calculated the sentiment scores. In our proposed model, we tried to give more importance to tweets published by famous people, and it would be investigated whether including the influence factor of the publisher based on their followers could enhance the prediction model. Then, by comparing the accuracy score for the model in which the polarity scores are weighted based on the normalized number of followers with the model in which they are unweighted, it was indicated that the model works better when this factor is included, so our proposed model will make predictions based on different methods of sentiment analysis. In the study conducted by Kim et al., a novel approach was used to predict the cryptocurrencies' prices that used multi-variate on-chain time-series data. Various on-chain variables are selected, grouped according to their inherent characteristics, and used as input variables for price prediction [23].

According to the study of Critien et al., to discover the optimal time interval in which the sentiment expressed becomes a reliable indicator of price change, the relation between future price at different temporal granularities and the sentiment of texts was explored and evaluated by two different neural network models, one based on recurrent networks and one based on convolutional networks. It was shown that not only can price direction predictions be made, but the magnitude of price changes can also be predicted with relative accuracy [24].

In the study of Park et al., to predict the closing prices of cryptocurrencies, the performance evaluation of a genetic algorithm tuned for deep learning (DL) and boosted tree-based techniques was investigated [25]. By comparing the results of Convolutional Neural Networks (CNN), Deep Forward Neural Networks, and Gated Recurrent Units, it was found that the CNN model had the least mean average percentage error of 0.08 and produced a consistent and highest explained variance score of 0.96 (on average) compared to other models [26]. In another study, long short-term memory (LSTM) and gated recurrent unit (GRU) were used to predict the future prices of cryptocurrencies, and according to the Twitter sentiment analysis, an action recommendation model was proposed to recommend actions to investors for maximizing profits, such as "sell", "buy", and "wait". This study showed that the proposed method had better performance compared to the conventional methods, and it was statistically validated. Similarly, the LSTM model outperformed other models in terms of Bitcoin, Ethereum, and Litecoin cryptocurrencies and was found to be efficient for cryptocurrency price prediction when compared to the other models with 67.43% accuracy [27]. In accordance with the mentioned study, the evaluation of convolutional LSTM neural networks was conducted in forecasting the future price of cryptocurrencies, which generally outperformed the other models, while CNN neural networks were also found to be able to provide good results, especially in the prediction of prices for Bitcoin, Ether, and Litecoin cryptocurrencies [28]. In another study, a method for adjusting the result of the action recommendation model based on Twitter sentiment analysis was recommended. A support vector machine (SVM) was used to forecast the movement of cryptocurrency prices based on the above multi-source data by classifying the tweets into sentiment categories with the Valence Aware Dictionary and Entiment Reasoner (VADER) and constructing the corresponding sentiment indicators. It was shown that their proposed multi-source data can effectively predict the movement of cryptocurrency prices [15]. In that study, the Twitter-Robustly Optimized BERT Pretraining Approach (roBERTa) and models were used to analyze the sentiments expressed on social media.

According to the study of Zhang et al., to predict the daily close price and the fluctuation of cryptocurrencies, a model was proposed by comparing the daily closing prices of six popular and valuable cryptocurrencies: Bitcoin (BTC), Bitcoin Cash (BCH), Litecoin (LTC), Ethereum (ETH), Electro-Optical System (EOS), and Ripple (XRP) based on a weighted and attentive memory channel model. The results showed that their proposed model achieves state-of-the-art performance and outperforms the baseline models in prediction error, accuracy, and profitability [15].

It can be seen from previous studies that the LSTM method has an acceptable performance in predicting the prices of cryptocurrencies and is considered an efficient method. By using LSTM, future prices can be predicted according to historical and time-series data. This model has been extensively used in previous studies and shown to have acceptable accuracy in price forecasting, which helps investors make better decisions on any action in the cryptocurrency market. It has been proven that LSTM is suitable for forecasting time series data with respect to unknown time lags. Its relative insensitivity to gap length is considered an advantage. LSTM requires a relatively large amount of time and memory for training [29]. Again, in the study conducted in 2023, the prices of bitcoin and Ethereum were predicted using the bidirectional LSTM model and a feature selection and weighting approach based on the mean decrease impurity (MDI) feature. This research represents the finest performance scores with the MDI approach using the LSTM model, and it was revealed that the fluctuations of the prices can be predictable as well as predicting the close prices of cryptocurrencies. It was recommended to focus on several kinds of cryptocurrencies due to the increase in their popularity [30].

In most of the research on the prediction of cryptocurrency prices, the characteristics of the users who published the tweets, such as the number of followers and the number of friends, have not been included, and a comprehensive study has not been conducted in which all the factors affecting the price of cryptocurrencies and the degree of effectiveness of each factor have been identified. Investigation of the impact of each factor, such as the level of influence of users in social media based on the number of followers, the number of friends, the number of positive and negative reactions, as well as the opinions of the audiences regarding the published content, can be studied separately, but designing a prediction model considering the influential power of each person who is publishing any post in social media and the people who are affected by each tweet will be a complex problem. Since social media's influencers can be ten times more influential than an average person, this study has been conducted based on the assumption that cryptocurrency prices depend on the behavior and reactions of influential people and the type of content that is published by those people [31]. Therefore, in this study, a model is presented that includes factors such as the influence power of Twitterers that have been less considered in previous studies.

The aim of this study was to compare the performance of different methods of sentiment analysis in predicting the prices of five different cryptocurrencies with the LSTM, ARIMA, and MLP models. As the number of followers indicates the level of influence that anybody has on the social network, the polarity scores of the tweets published by Twitterers were weighted based on the publishers' influence factor, which can be identified by the number of followers.

As mentioned, there are many factors that have an impact on the prices of cryptocurrencies that need to be identified and used as input features for the learning models, and each of them has a different level of importance in having an impact on the future prices of cryptocurrencies. The various combinations of those features, considering the importance of each feature in the prediction model, could be tested and evaluated by the evaluation indicators. In this study, the sentiment factor of the published tweets was selected, and new methods of sentiment were proposed based on social influence. This research has been conducted on five kinds of cryptocurrencies, and different evaluation metrics were selected to evaluate the precision of the performance of each proposed method of considering sentiment scores. In this research, different models were used based on different ways of considering sentiment analysis by three machine learning models.

The remainder of this paper is organized as follows: Cryptocurrencies historical prices and Twitter data collection in Section II, Data preparation in Section III, Proposed model in Section IV, Results and discussions in Section V.

II. DATASET

There are two kinds of datasets that are used to feed the model. One of them is a collection of historical prices, and the other is a collection of tweets to be used in our sentiment analyzer model. In this section, the methods of collecting the data from Twitter and also collecting the historical prices of cryptocurrencies have been described.

A. DATA COLLECTION FROM SOCIAL MEDIA

Twitter, as a rich source of users' comments and the largest active community, was selected to collect the tweets containing the hashtags of our investigated cryptocurrencies. Although there are restrictions on collecting tweets from Twitter through Twitter's application programming interface (API), it is possible to extract tweets with libraries that can provide access to Twitter's data, such as Snscrape, which can partially overcome the limitations. In this research, a dataset containing 1.2 million extracted tweets with hashtags of various kinds of cryptocurrencies in nearly seven months of 2021 was extracted from Kaggel. The tweets of each cryptocurrency will be grouped according to the hashtags used in their texts. There were some tweets that could be used for more than one kind of cryptocurrency. For example, if a tweet contained the hashtags of four kinds of cryptocurrencies, that would be used to predict the prices of all four cryptocurrencies. Due to the lack of tweets in this dataset for some days, on days when the number of tweets wasn't sufficient, the tweets were randomly gathered from Twitter by snscrape, and the number of extracted tweets for each cryptocurrency was increased to 1500 tweets per day. This module from the SNSCRAPE library provides functions to interact with scraped tweets with the Twitter API, so the number of tweets analyzed in a period of 198 days reached a total of 300,000 tweets for each cryptocurrency in the investigated time interval.

B. DATA COLLECTION OF HISTORICAL CRYPTOCURRENCIES' PRICES

The finance division of Yahoo's website, by offering access to a wide range of financial cryptocurrencies' data, such as closing, opening, lowest, and highest prices and dates, was used to discover the historical prices of BTC, ETH, EOS, Cardano (ADA), and XRP. We selected the historical prices of cryptocurrencies in 198 days, from February 5, 2021, until August 21, 2021.

III. DATA PREPRATION

The stage of data preprocessing includes the preprocessing the texts of the tweets which has been described in section A, and the price normalization of cryptocurrencies which has been described in section B.

A. TWITTER DATA PREPROCESSING

Preprocessing of the texts of tweets includes removing abbreviations, hashtags, links, emojis, and stop words. The library of Natural Language Toolkit (NLTK) was used for this purpose. Duplicate tweets are also deleted. To completely preprocess the tweets, another set of stop words, including 485 words, was added to the predefined array of stop words in NLTK to be removed from the texts of the tweets.

B. CRYPTOCURRENCIES PRICE NORMALIZATION

To compare the prediction accuracy of the learning model for different cryptocurrencies, it is necessary that the prices of all cryptocurrencies be placed in a similar range, or, in other words, that the prices of cryptocurrencies be normalized. For this purpose, the prices of all cryptocurrencies took on values between 0 and 1. To normalize the price of cryptocurrencies, assuming that P_{actual} is the actual price of the cryptocurrency in a day, the normalized value of p ($P_{normalized}$) is calculated using the formula (1), where Max is the highest price of the cryptocurrency in the investigated time period and Min is the lowest price of the cryptocurrency in this interval.

$$p_{normalized} = \frac{p_{actual} - min}{Max - min} \tag{1}$$

IV. PROPOSED MODEL

This section includes the sentiment analyzer model and the learning model. The proposed sentiment analyzer model is being described in Part A, and the learning model will be described in Part B.

A. SENTIMENT ANALYZER MODEL

The next stage after data preprocessing is sentiment analysis. As mentioned earlier, sentiment analysis is an approach that determines the emotional tone behind the text. According to the model proposed by Pant et al., tweets are pre-processed before being fed to the sentiment analysis model [22].

In this study, cryptocurrencies' prices are also normalized before being fed to the learning model. To get insights from the texts, polarity scores were obtained using the TextBlob library. According to the obtained polarity scores, we can figure out how much of the tweet is positive or negative. The values of the polarity scores are between −1 and 1. If the polarity score of a tweet is between −1 and 0, this tweet will be considered to have negative emotions, and if the score is between 0 and 1, it can be understood that the tweet contains positive emotions.

Some of the limitations that we encountered were the number of tweets that we had to analyze. The greater the number of tweets we had, the more accurate the results we obtained. Also, there were limitations in distinguishing between the tweets published by bots and the tweets published by Twitter users. In this paper, it was assumed that the tweets published by bots could also affect the users' behaviors. In the sentiment analysis procedure, some of the words can have a negative or positive tone, depending on the context in which they exist. In the IBWPS method, the polarity scores will be identified by a lexicon-based sentiment analysis model without considering the context. By proposing the DBSA model that the LIWC library is being used, it is tried to overcome this limitation. This tool will identify the positivity and negativity of the tones in the tweets, considering the context in which the words appear.

In this study, the linguistic inquiry word count (LIWC) will be used as a tool to extract the type of positive or negative emotion more accurately. The output of the first phase of the research is a table containing features such as daily historical prices, sentiment scores of the tweets, and dates. The second phase of the research is teaching the learning model based on the mentioned features. After calculating the average daily polarity score for all cryptocurrencies, about 87% of these numerical scores were between 0.5 and −0.5, and only 13% of tweets had a polarity score of −0.5 to −1 or 0.5 to 1. This means that the days when the average polarity score of their tweets is greater than 0.5 or smaller than −0.5, or, in other words, the more this polarity score is closer to 1 or −1, the learning model would be more sensitive to the increase or decrease in the price of cryptocurrencies.

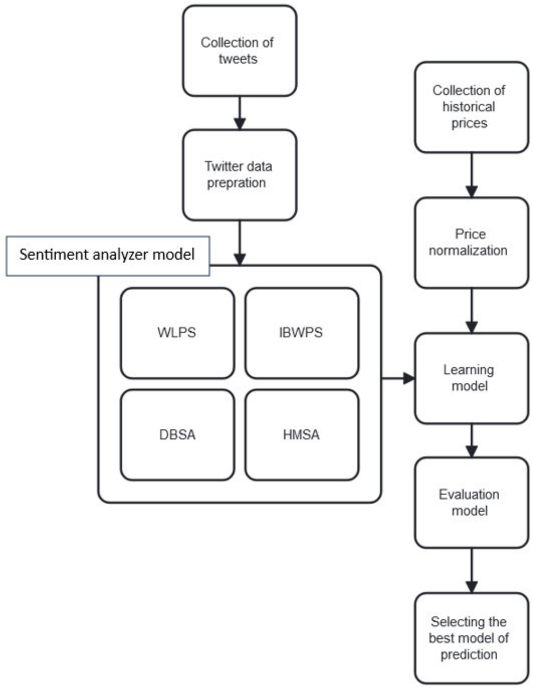

FIGURE 1. The proposed price prediction model.

Figure 1 illustrates our proposed model for predicting the future prices of cryptocurrencies.

1) PREDICTION BASED ON UNWEIGHTED POLARITY SCORES

At this stage, the average of the polarity scores on a daily basis was calculated. Along with the calculated daily polarity scores, the historical prices of Bitcoin, Ethereum, EOS, Cardano, and Ripple in the investigated time interval were gathered from the Yahoo website. Considering that keeping the tweets whose polarity score is 0 causes this calculation of the average polarity score to become close to 0, the tweets whose polarity score is 0 were eliminated.

Historical prices and the sentiment of the tweets were used as the features on which predictions were made, and the values of the cryptocurrencies in the last 20 percent of our investigated time period will be predicted based on these features. In this way, by specifying the normalized historical prices and unweighted polarity scores, the operation of prediction will be performed to predict the future prices of the mentioned cryptocurrencies using machine learning algorithms, which will be explained in Section C. Since this method is based on unweighted polarity scores, it is named the weightless polarity scores (WLPS) model.

2) PREDICTION BASED ON INFLUENCED BASED WEIGHTED POLARITY SCORES

Considering that the greater the number of followers a person has in a social network, the more influential that person is, at this stage, the obtained polarity scores have been weighted based on the number of followers. By studying the topology of the social network, we find that one of the indicators for evaluating the importance of a node is the degree of centrality, which calculates the degree of importance of a node based on the number of links connected to that node. The greater the number of links connected to that node, the more important that node is in the network and can have a greater impact on the others [32]. It can be concluded that the number of followers of a person is the number of edges connected to a node, and the more followers a person has, the more influence that person will have in the social network. Therefore, since one of the factors used to identify the level of influence of a person on social networks is the number of followers, the polarity score of each tweet will be weighted based on the normalized value of the number of followers. To weight the polarity scores of tweets, it is necessary to multiply the normalized number of followers of the Twitterer by the polarity score of that tweet. For this purpose, the normalized number of followers is in the range between zero and one. If "f" represents the number of followers of a user, the normalized value of "f" ($f_{normalized}$) is calculated from the actual value of "f" (f_{actual}) using the following formula:

$$f_{normalized} = \frac{f_{actual} - min}{Max - min} \tag{2}$$

where Max represents the largest number of followers that the Twitterers in our dataset have, and the value of min is the smallest number of followers. For each tweet, the polarity score of each tweet will be multiplied by the normalized number of followers of the related publisher, and then the weighted average polarity score for each day is calculated. In this section, historical data and weighted average polarity scores will be used as features to predict the prices of cryptocurrencies in the last 41 days. Since this method is based on influenced-based weighted polarity scores, it is named IBWPS.

3) PREDICTION BASED ON LINGUISTIC INQUIRY WORD COUNT DICTIONARY-BASED SENTIMENT ANALYSIS

In this research, we seek to have a more accurate division of positive and negative emotions. One of the tools used in the process of text mining is the LIWC tool, which can identify more than a hundred dimensions from the texts and find the hidden features of the text. This tool can provide a rich insight into the publishers' psychological states, including emotions, thinking styles, and social concerns. In brief, decades of empirical research, especially research that uses LIWC as a scientific tool, provide specialized ways to understand, explain, and quantify psychological, social, and behavioral phenomena [69-71]. In this analysis method, for the prediction of cryptocurrencies' prices, the positive and negative tones of the tweets will be extracted by this tool. Since the higher the number of words in a text, the more certain the tool will be to classify the words in that text, all the tweets published on each day for each investigated cryptocurrency were supposed to be used as input. In this method, the percentage of the words that have a positive or negative tone will be calculated. These extracted features will be used to feed the learning model to predict the future prices of cryptocurrencies. Since this method is based on the percentage of positivity or negativity of the tones of the tweets and the sentiment analysis was conducted with the LIWC tool, this method is named DBSA, which stands for dictionary-based sentiment analysis.

4) PREDICTION BASED ON HYBRID MODEL OF SENTIMENT ANALYSIS

A combination of two methods that can identify the types of emotions published in texts using a tool like LIWC that has more than a hundred dictionaries can provide an accurate categorization of the positivity and negativity of the tones in texts alongside weighted sentiment scores. In this method, the combination of features from the two methods mentioned in sections II and III is considered the input to our model.

B. LEARNING MODEL

A machine learning model is created based on regression machine learning models, which are suitable for time series data analysis. The learning models that have been selected to make predictions in this study are LSTM, ARIMA, and MLP. For this purpose, 80% of the data are selected as training data, and 20% of them are selected as test data. The tweets published in a time interval of 198 days in 2021 for the studied cryptocurrencies are given to the learning models according to the methods mentioned in Section IV.

The prediction model works using different types of features extracted by using the different methods of sentiment analysis, which can be weighted or unweighted. The machine learning models that are useful for time series data analysis have been selected. The data from the first 158 days was used as training data, and the data from the last 40 days was used for testing. Since the prices of the investigated cryptocurrencies were in different ranges and we wanted to compare them with each other, we turned the prices of each cryptocurrency in the investigated time intervals into the range of 0 and 1 in the process of normalization. The predicted prices of each cryptocurrency on day "n", depend on the values of the prediction model for the n-1 days before. The prediction model is based on the one-step-ahead prediction method. The difference between the WLPS, IBWPS, DBSA, and HMSA methods is the way they consider the sentiment scores of the tweets and the number of features needed to make predictions. For example, in the WLPS method, the feature of the average of unweighted polarity scores was used alongside the historical prices of cryptocurrencies, while in the IBWPS method, the normalized number of followers for a tweet's publisher is multiplied by the polarity score of that tweet. The average of weighted polarity scores has been calculated on a daily basis for each cryptocurrency to be used as a feature to predict its price. In WLPS and HMSA, the predictions are made by unweighted sentiment scores, while in IBWPS and HMSA methods, the predictions are made by weighted scores. The HMSA method uses the weighted sentiment scores beside the percentage of daily positive and negative tones of the tweets, which is a hybrid model consisting of the features in the IBWPS and DBSA methods. Then, the prediction accuracy of the prediction models will be measured by the evaluation indicators.

We used a nonseasonal ARIMA model that has three parameters: p as the number of autoregressive terms, q as the number of lagged forecast errors in the prediction equation, and d as the number of nonseasonal differences needed for stationary, which we set in the range of 0 and 2:

$$p = d = q = \text{range}\ (0, 2) \tag{3}$$

The number of input layers for MLP has been defined as the same as the number of features in each of the proposed methods. For example, if the number of features in a method is three, then the number of input layers will be set to 3. The batch size for the learning models is defined as the number of samples to work on before the internal parameters of the model are updated, and we defined it with a value of 32.

C. EVALUATION MODEL

The prediction model is proposed based on features such as cryptocurrencies' historical prices and the average daily polarity scores of tweets. After applying weights to the polarity scores of each tweet, the prediction model will predict the prices of cryptocurrencies for each of the methods proposed in Section IV, and the accuracy of the prediction models will be compared with each other. The results of the investigation will indicate how much the accuracy of the model will increase by applying weights to the polarity scores, which are determined by the evaluation indices. Four indicators—mean square error (MSE), root mean square error (RMSE), mean absolute magnitude of error (MAE), and mean absolute percentage of error (MAPE) will be used to evaluate the accuracy of the model.

1) MSE

The mean squared error, or MSE, is a method for the evaluation of forecasting methods that regulates large forecasting errors. In this evaluation indicator, each of the errors is squared. According to the formula (3), y_i is the actual price of a cryptocurrency on day i and $\hat{y}_i$ is the value that has been predicted.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 \tag{4}$$

2) RMSE

The values of the RMSE indicator can be obtained by taking the root of the values obtained by the MSE indicator.

$$RMSE = \sqrt{MSE} = sqrt(\frac{1}{N} \sum_{i=1}^{N} \left(\mathrm{y_i} - \hat{y}_i\right)^2) \tag{5}$$

3) MAE

MAE is an indicator for measuring the errors between paired observations. In formula (5), y_i is the actual or expected price of a cryptocurrency on day i, and $\hat{y}_i$ is the predicted value. The average of all differences between these two values is calculated on a daily basis.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{6}$$

4) MAPE

The mean absolute percentage error, or MAPE, is calculated using the absolute errors of each period divided by the actual values [33]. The value obtained by this indicator is the value obtained by MAE that has been multiplied by 100.

$$MAPE = \frac{100}{n} \sum_{i=1}^{N} \frac{y_i - \hat{y}_i}{y_i} \tag{7}$$

V. RESULTS AND DISCUSSION

In this section, the result of the prediction of the cryptocurrencies' prices according to the proposed model is presented. In this model, the prices of five selected cryptocurrencies were predicted by the application of machine learning models using weighted and unweighted polarity scores. The evaluation metrics that have been used to evaluate the accuracy of the proposed prediction model are MSE, MAE, RMSE, and MAPE, and the results of the prediction based on the mentioned indicators were compared with each other. Each of sections A, B, C, D, and E describes the results of each investigated cryptocurrency.

A. PRICE PREDICTION OF BITCOIN

The results of predicting the future prices of bitcoin using three different machine learning models are shown in FIGURE 2 to FIGURE 13. By analyzing the results of Bitcoin price prediction in four different methods according to the results shown in TABLE 1, it can be figured out that the accuracy of the prediction model in the IBWPS, DBSA, and HMSA methods is higher than the WLPS method, in which the daily average of unweighted polarity scores was used in the model. The accuracy score obtained in the IBWPS method, in which the polarity scores are weighted based on the impact of tweets' publishers according to the number of followers, indicates that the highest accuracy for the prediction of bitcoin was obtained by the use of the IBWPS method. The accuracy of prediction in all three machine learning models for the HMSA method, which combined features of the IBWPS and DBSA methods for the prediction of bitcoin, was slightly lower than the IBWPS method (0.0003 according to

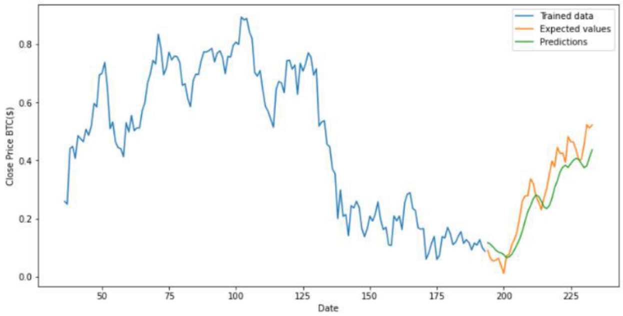

FIGURE 2. **The price prediction of Bitcoin using MLP model for WLPS method.**

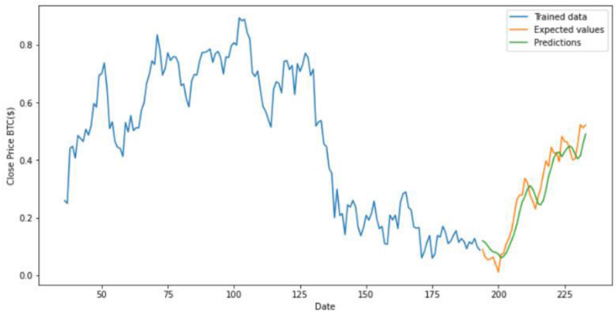

FIGURE 3. **The price prediction of Bitcoin using MLP model for IBWPS method.**

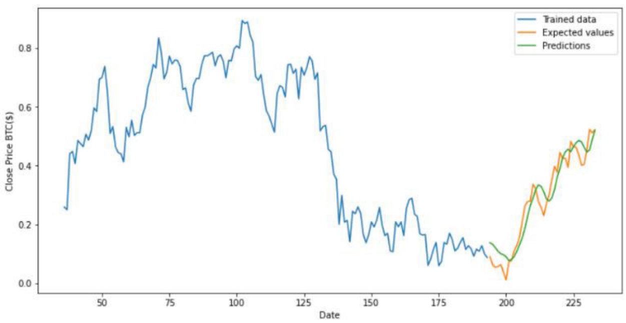

FIGURE 4. **The price prediction of Bitcoin using MLP model for DBSA method.**

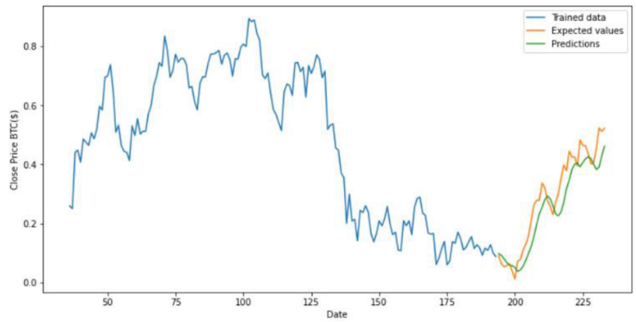

FIGURE 5. **The price prediction of Bitcoin using MLP model for HMSA method.**

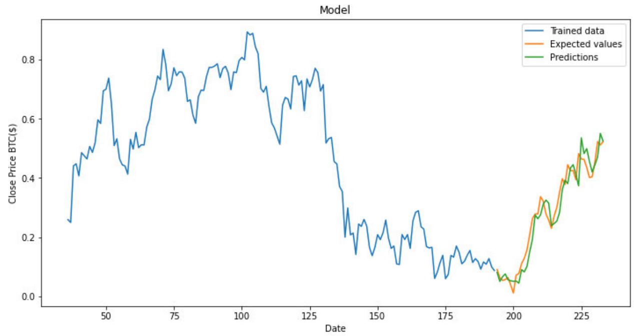

FIGURE 6. **The price prediction of Bitcoin using ARIMA model for WLPS method.**

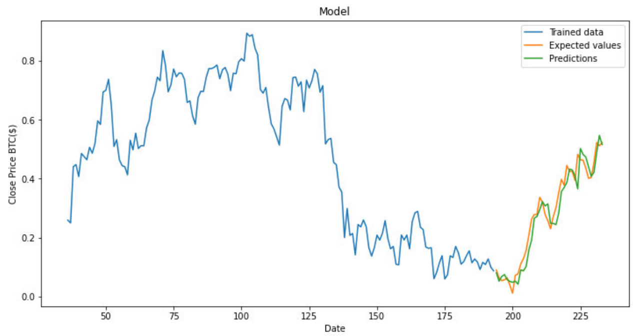

FIGURE 7. **The price prediction of Bitcoin using ARIMA model for IBWPS method.**

the MSE indicator). In all methods, the accuracy of prediction by ARIMA is the highest. According to the MSE indicator, the obtained scores in the ARIMA model for IBWPS, DBSA, and HMSA were 15.91%, 9.66%, and 11.93% lower than WLPS, respectively. For the LSTM model, the obtained scores for IBWPS, DBSA, and HMSA were 38.38%, 17.3%, and 30.27% better than WLPS, and the obtained MSE scores in the MLP model for IBWPS, DBSA, and HMSA methods were 48.89%, 15.93%, and 46.02% lower than WLPS, respectively.

B. PRICE PREDICTION OF ETHERIUM

According to TABLE 2, in which the results of prediction for Ethereum have been shown, LSTM had the most accurate prediction in the HMSA method, and then the LSTM model had the most accurate prediction in the IBWPS method. According to the MSE indicator, the scores obtained by this model for IBWPS, DBSA, and HMSA were 52.53%, 50.51% and 62.47% better than WLPS, respectively. The prediction models for the DBSA method, in which the input of the prediction model is the result of feature extraction by the LIWC tool, had a weaker performance than the IBWPS method, in which the polarity scores of the tweets were multiplied based on the normalized values of the number of followers of that tweet's publisher. The lowest accuracy was achieved by the WLPS method. In the WLPS and IBWPS methods, the highest accuracy of prediction was obtained by the ARIMA model. In the DBSA and HMSA methods, LSTM performed better than ARIMA, with

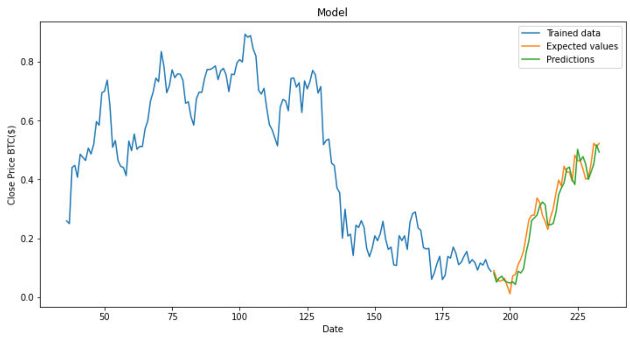

FIGURE 8. **The price prediction of Bitcoin using ARIMA model for DBSA method.**

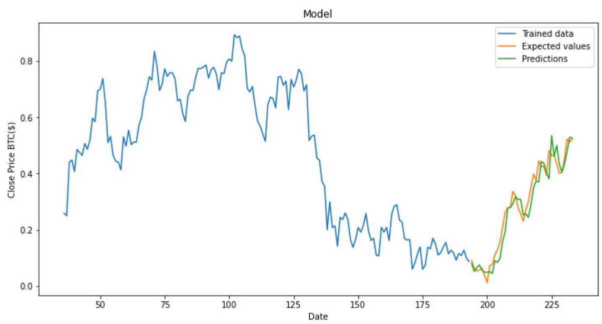

FIGURE 9. **The price prediction of Bitcoin using ARIMA model for HMSA method.**

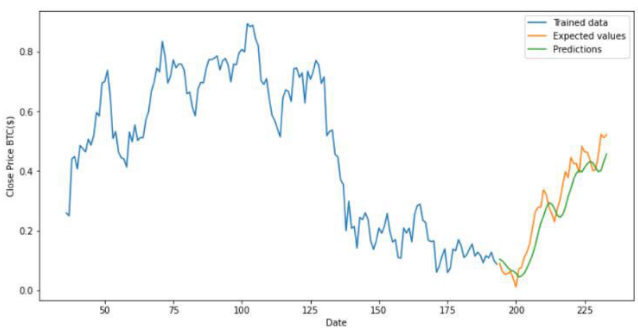

FIGURE 10. **The price prediction of Bitcoin using LSTM model for WLPS method.**

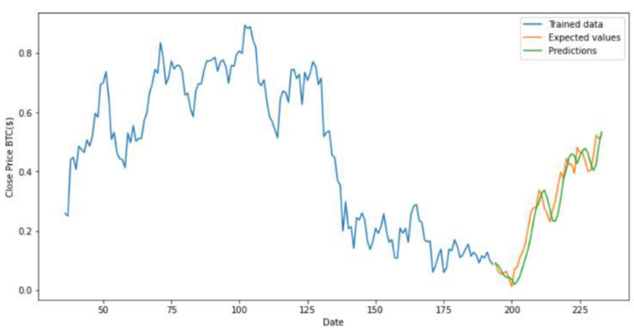

FIGURE 11. **The price prediction of Bitcoin using LSTM model for IBWPS method.**

a small difference (0.000002 according to the MSE indicator). FIGURE 14 to FIGURE 25 show the results of the prediction based on MLP, ARIMA, and LSTM models in four methods.

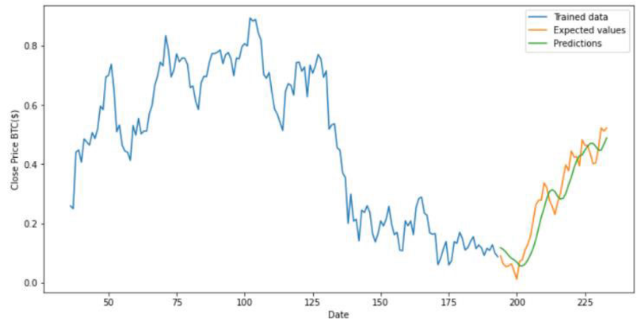

FIGURE 12. **The price prediction of Bitcoin using LSTM model for DBSA method.**

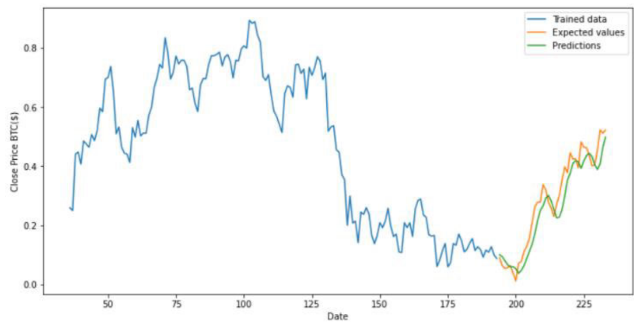

FIGURE 13. **The price prediction of Bitcoin using LSTM model for HMSA method.**

TABLE 1. **Bitcoin price prediction accuracy score using LSTM, ARIMA, and MLP algorithms by MSE, RMSE, MAE, and MAPE indicators for WLPS, IBWPS, DBSA, and HMSA methods.**

Method	Algorithm	Performance Indicators			
		MSE	RMSE	MAE	MAPE
WLPS	MLP	0.00452	0.06723	0.05685	5.68592
	ARIMA	0.00176	0.04205	0.03407	3.40747
	LSTM	0.00370	0.06085	0.05150	5.15083
IBWPS	MLP	0.00231	0.04813	0.04031	4.03161
	ARIMA	0.00148	0.03855	0.03052	3.05205
	LSTM	0.00228	0.04780	0.03972	3.97294
DBSA	MLP	0.00380	0.06172	0.05127	5.12748
	ARIMA	0.00159	0.03989	0.03215	3.21512
	LSTM	0.00306	0.05535	0.04826	4.82646
HMSA	MLP	0.00244	0.04947	0.04274	4.27499
	ARIMA	0.00155	0.03945	0.03036	3.03698
	LSTM	0.00258	0.05084	0.04191	4.19193

C. PRICE PREDICTION OF EOS

According to TABLE 3 related to the price prediction of EOS, all three LSTM, ARIMA, and MLP models have the highest accuracy in the HMSA method. The accuracy of EOS price prediction in the IBWPS method was higher in predicting the price of EOS compared to the DBSA method, according to

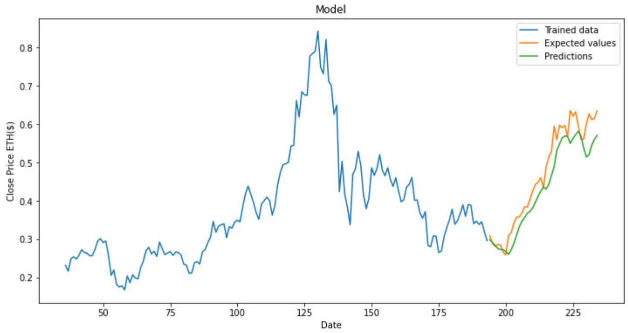

FIGURE 14. Price prediction result for Ethereum using LSTM for WLPS method.

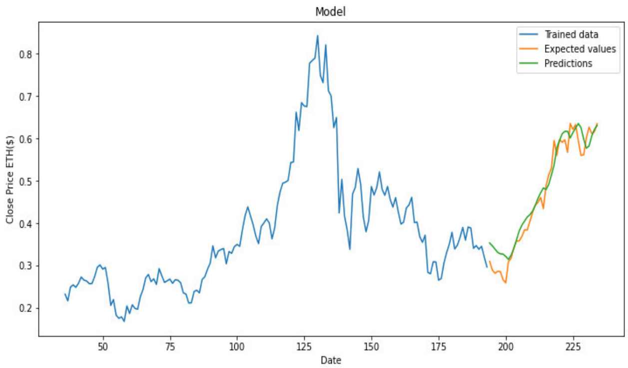

FIGURE 15. Price prediction result for Ethereum using LSTM for IBWPS method.

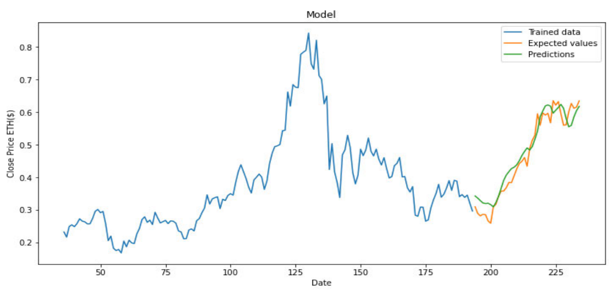

FIGURE 16. Price prediction result chart for Ethereum using LSTM for DBSA method.

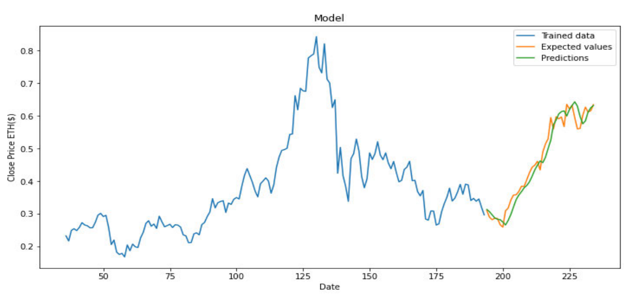

FIGURE 17. Price prediction result for Ethereum using LSTM for HMSA method.

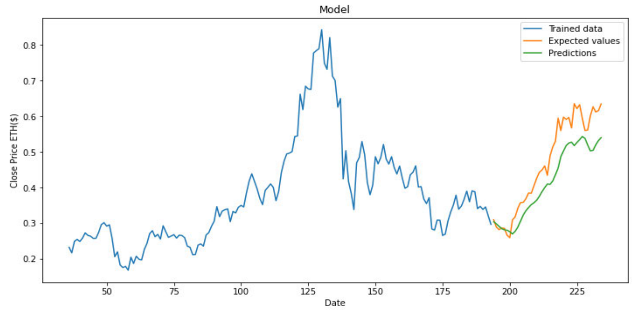

FIGURE 18. Price prediction result for Ethereum using MLP for WLPS method.

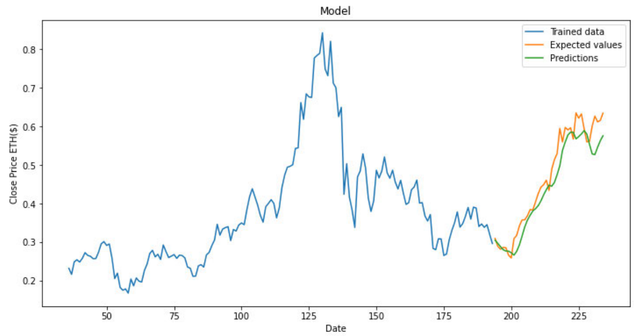

FIGURE 19. Price prediction result for Ethereum using MLP for IBWPS method.

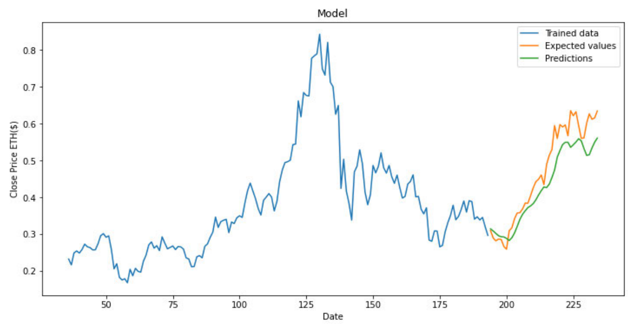

FIGURE 20. Price prediction result for Ethereum using MLP for DBSA method.

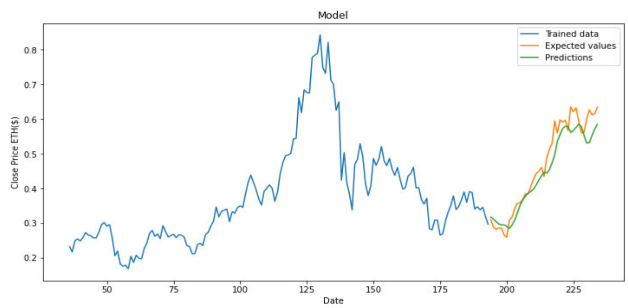

FIGURE 21. Price prediction result for Ethereum using MLP for HMSA method.

the LSTM and ARIMA models while MLP has performed more accurately in predicting the prices of EOS in DBSA than the IBWPS method. The lowest accuracy for all three models was obtained in the WLPS method, while the most accurate prediction among all other models for EOS was obtained in the HMSA method using the ARIMA model. FIGURE 26 to

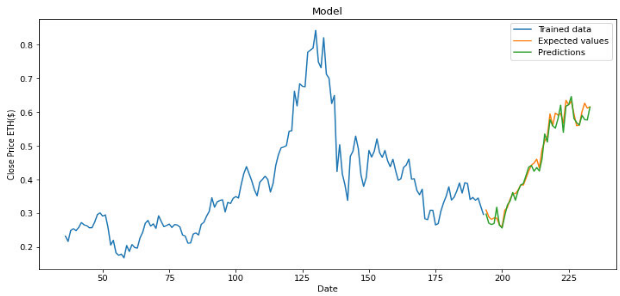

FIGURE 22. Price prediction result chart for Ethereum using ARIMA for WLPS method.

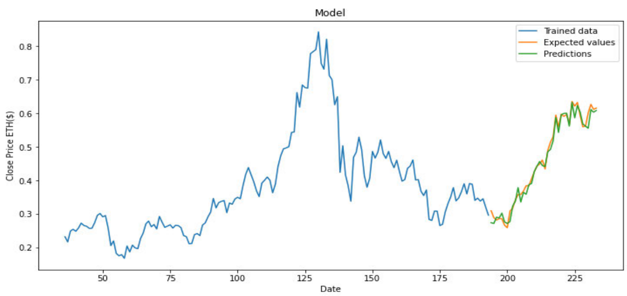

FIGURE 23. The price prediction for Ethereum using ARIMA for IBWPS method.

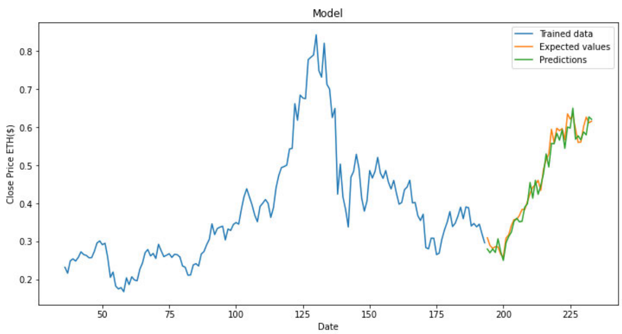

FIGURE 24. The price prediction for Ethereum using ARIMA for DBSA method.

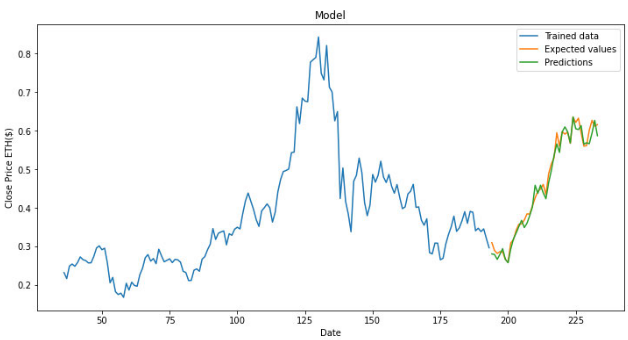

FIGURE 25. The price prediction for Ethereum using ARIMA for HMSA method.

FIGURE 37 show the results of predicting the prices of EOS for MLP, LSTM, and ARIMA models using four different methods have been shown.

TABLE 2. Ethereum price prediction accuracy score using LSTM, ARIMA, and MLP algorithms by MSE, RMSE, MAE, and MAPE indicators for WLPS, IBWPS, DBSA, and HMSA methods.

Method	Algorithm	Performance Indicators			
		MSE	RMSE	MAE	MAPE
WLPS	MLP	0.004636	0.068092	0.05828	5.82898
	ARIMA	0.001207	0.034747	0.02700	2.70017
	LSTM	0.002273	0.047679	0.03879	3.87954
IBWPS	MLP	0.001681	0.041011	0.03253	3.25341
	ARIMA	0.001067	0.032677	0.02649	2.64970
	LSTM	0.001079	0.032852	0.02600	2.60070
DBSA	MLP	0.002929	0.054123	0.04486	4.48675
	ARIMA	0.001127	0.033574	0.02585	2.58581
	LSTM	0.001125	0.033555	0.02863	2.86341
HMSA	MLP	0.001588	0.039856	0.03188	3.18867
	ARIMA	0.000993	0.031517	0.02506	2.50661
	LSTM	0.000853	0.029212	0.02323	2.32336

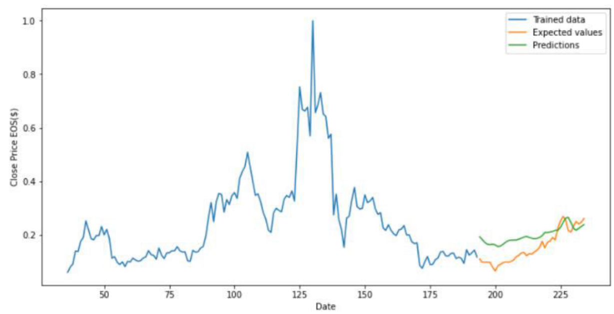

FIGURE 26. The price prediction for EOS using LSTM for WLPS method.

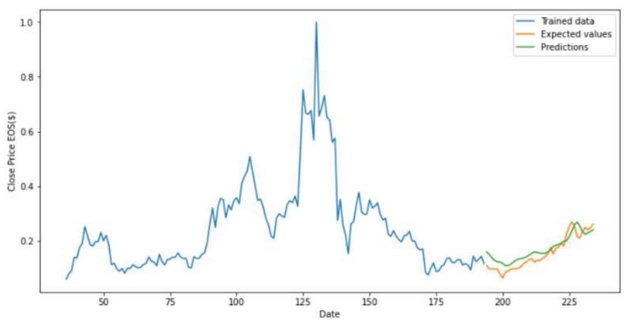

FIGURE 27. The price prediction for EOS using LSTM for IBWPS method.

D. PRICE PREDICTION OF CARDANO

According to the TABLE 4 related to the results of Cardano price prediction, it has been concluded that the HMSA method had the highest accuracy than the others in all three learning models, and then the highest accuracy was achieved in the IBWPS method. The accuracy of prediction using the WLPS method, where the average of unweighted polarity scores was calculated, was the lowest among all learning models. MLP had the lowest accuracy in predicting the prices of Cardano among all machine learning models. The highest accuracy was obtained in predicting the prices of Cardano

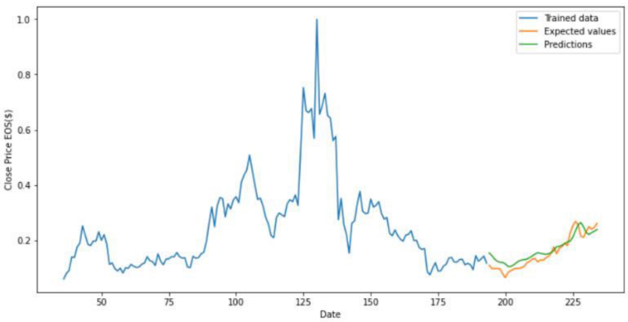

FIGURE 28. The price prediction for EOS using LSTM for DBSA method.

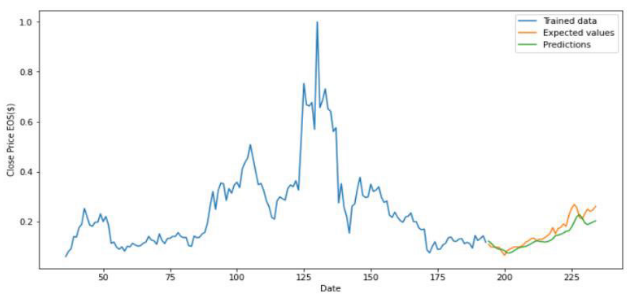

FIGURE 29. The price prediction for EOS using LSTM for HMSA method.

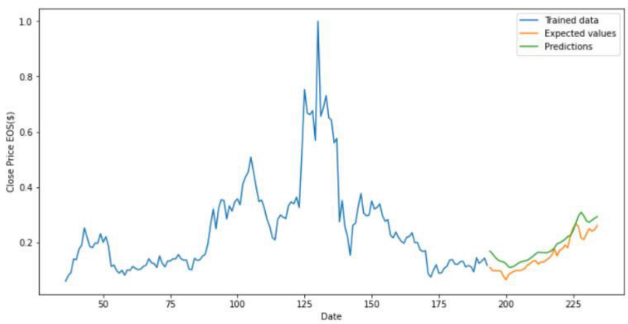

FIGURE 30. The price prediction for EOS using MLP for WLPS method.

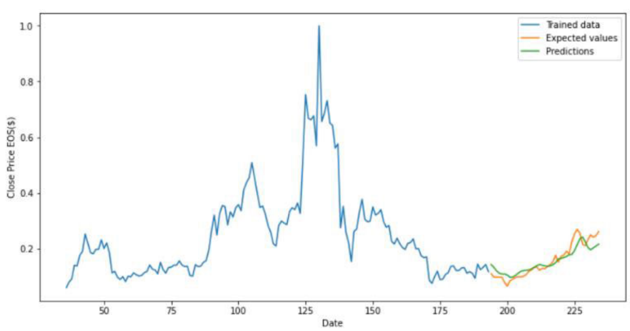

FIGURE 31. The price prediction for EOS using MLP for IBWPS method.

using the HMSA method with the ARIMA model according to the MSE indicator. FIGURE 38 to FIGURE 49 show results of predicting the prices of Cardano for MLP, ARIMA, and LSTM models.

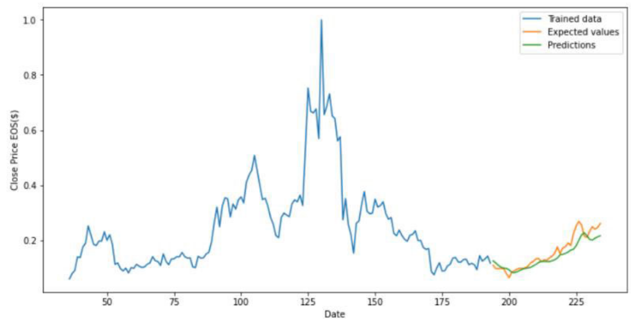

FIGURE 32. The price prediction for EOS using MLP for DBSA method.

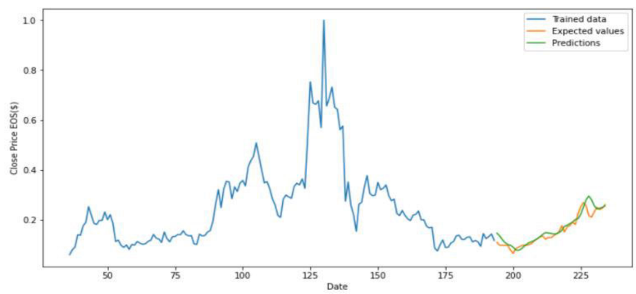

FIGURE 33. The price prediction for EOS using MLP for HMSA method.

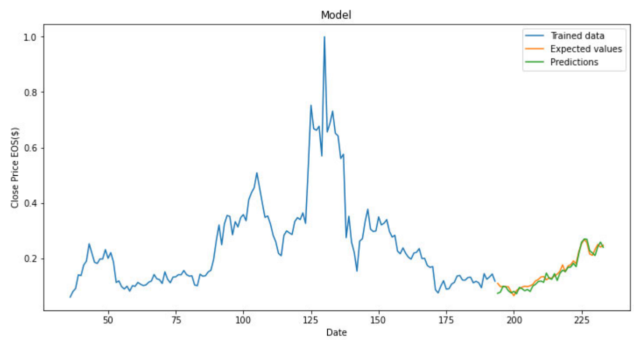

FIGURE 34. The price prediction for EOS prediction using ARIMA for WLPS method.

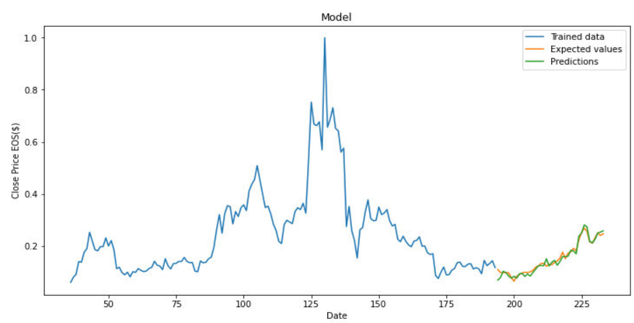

FIGURE 35. Price prediction result chart for EOS using ARIMA for WLPS method.

E. PRICE PREDICTION OF RIPPLE

According to TABLE 5, all three machine learning models in the IBWPS method had the highest accuracy in predicting

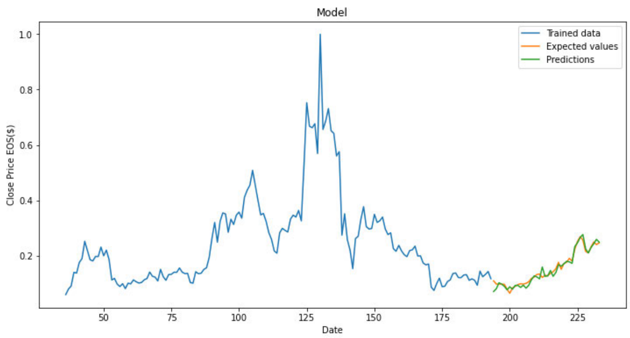

FIGURE 36. The price prediction for EOS using ARIMA for DBSA method.

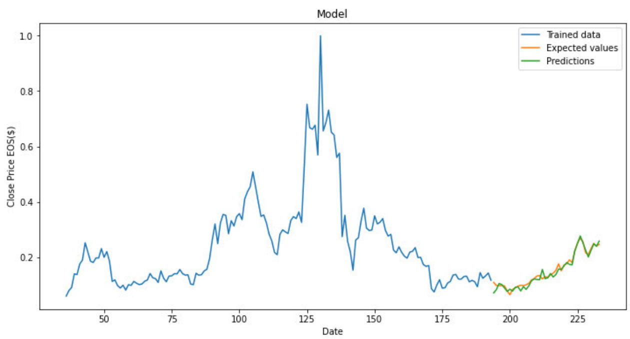

FIGURE 37. The price prediction for EOS using ARIMA for HMSA method.

TABLE 3. EOS price prediction accuracy score using LSTM, ARIMA, and MLP algorithms by MSE, RMSE, MAE, and MAPE indicators for WLPS, IBWPS, DBSA, and HMSA methods.

Method	Algorithm	Performance Indicators			
		MSE	RMSE	MAE	MAPE
WLPS	MLP	0.001023	0.031984	0.027503	2.75038
	ARIMA	0.000423	0.020569	0.015917	1.59170
	LSTM	0.003203	0.056600	0.050845	5.08456
IBWPS	MLP	0.000747	0.027336	0.020831	2.08315
	ARIMA	0.000369	0.019223	0.014287	1.42875
	LSTM	0.000824	0.028707	0.025085	2.50851
DBSA	MLP	0.000684	0.026163	0.020920	2.09206
	ARIMA	0.000380	0.019507	0.014836	1.48369
	LSTM	0.000962	0.031028	0.023856	2.38569
HMSA	MLP	0.000537	0.023176	0.016193	1.61932
	ARIMA	0.000349	0.018705	0.014976	1.49765
	LSTM	0.000676	0.026011	0.022409	2.24091

the price of XRP. The accuracy of predicting the price of XRP by LSTM was generally higher than ARIMA. Similar to the price prediction of other cryptocurrencies, the lowest accuracy was obtained with the WLPS method. MLP had the lowest accuracy in the WLPS and DBSA methods, while the lowest accuracy in the IBWPS and HMSA models was obtained by ARIMA. According to the MSE indicator, the accuracy of the prediction of LSTM compared to MLP

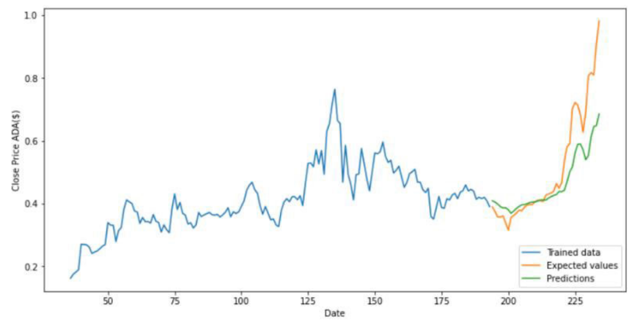

FIGURE 38. Price prediction result for Cardano using MLP for WLPS method.

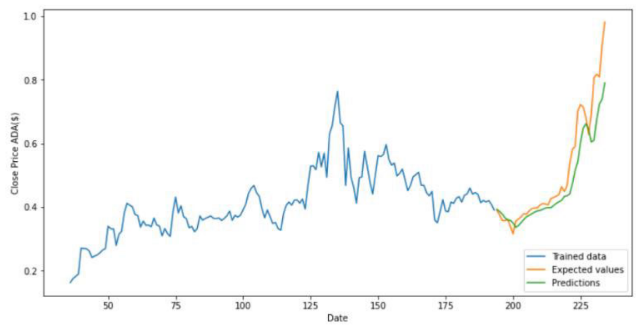

FIGURE 39. Price prediction result for Cardano using MLP for IBWPS method.

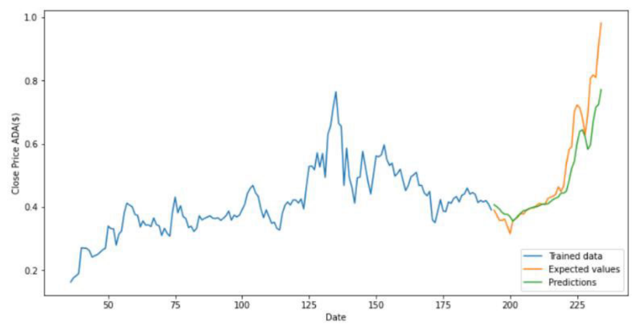

FIGURE 40. Price prediction result for Cardano using MLP for DBSA method.

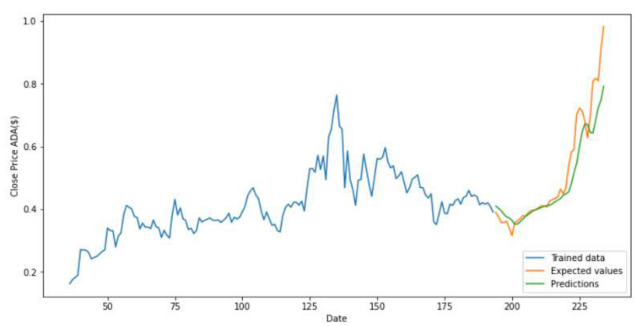

FIGURE 41. Price prediction result for Cardano using MLP for HMSA method.

was almost twice higher than MLP in all four methods. FIGURE 50 to FIGURE 61 are the results of predicting the prices for Cardano for MLP, ARIMA, and LSTM models.

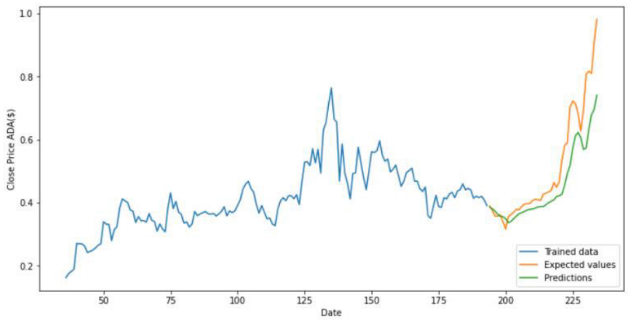

FIGURE 42. Price prediction result for Cardano using LSTM for WLPS method.

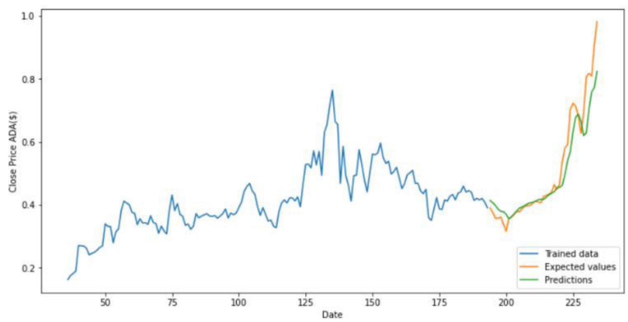

FIGURE 43. Price prediction result for Cardano using LSTM for IBWPS method.

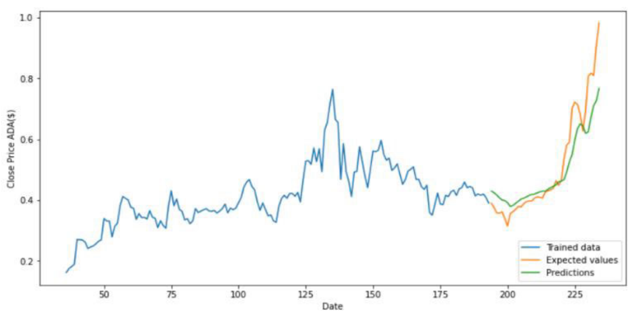

FIGURE 44. Price prediction result for Cardano using LSTM for DBSA method.

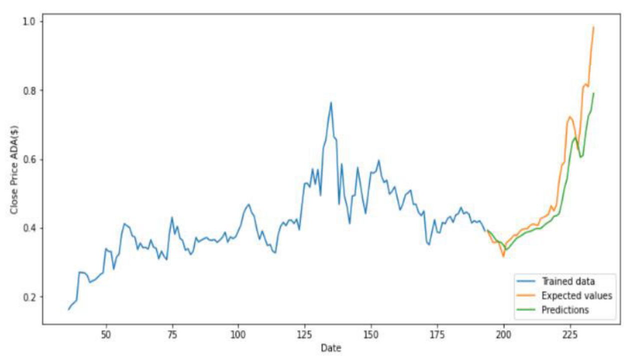

FIGURE 45. Price prediction result for Cardano using LSTM for HMSA method.

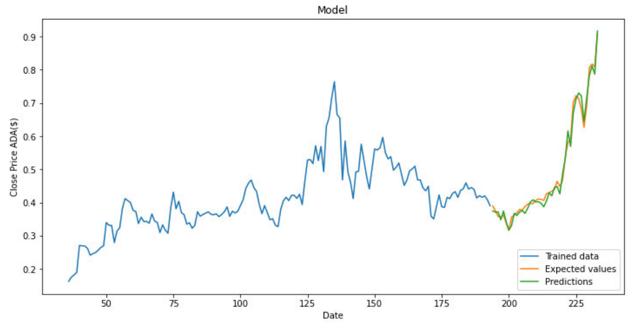

FIGURE 46. Result of price prediction for Cardano using ARIMA for WLPS method.

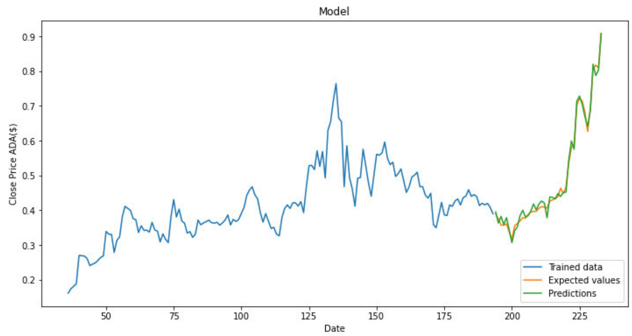

FIGURE 47. Price prediction result for Cardano using ARIMA for IBWPS method.

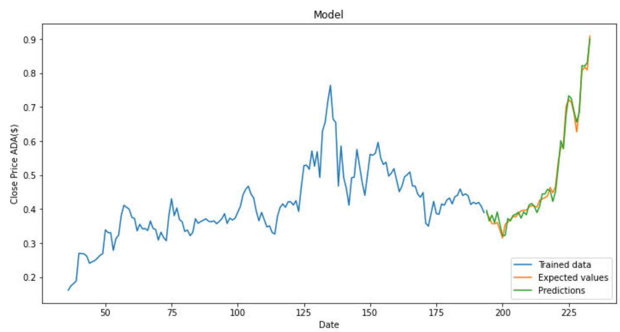

FIGURE 48. Price prediction result for Cardano using ARIMA for DBSA method.

The average MSE score for all five cryptocurrencies for each of the proposed methods is shown in Table 6. The results indicate that the proposed methods, including IBWPS, DBSA, and HMSA using all three deep learning models have better results compared to the WLPS method.

In the IBWPS method, the obtained MSE score is 51.9818% less than the MSE score obtained in WLPS while predicting based on the MLP model. Also, in the DBSA method, the MSE score is 35.8175% less than the MSE score obtained by WLPS. Furthermore, in the HMSA method, the obtained MSE score is 53.9017% less than the WLPS.

By having predictions based on the ARIMA model, the obtained results showed that there was an improvement of 11.4856% while using IBWPS compared to the WLPS method, and the DBSA model performed 5.43329% less than

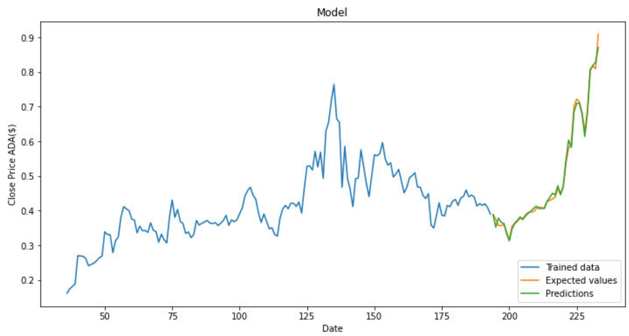

FIGURE 49. Price prediction result for Cardano using ARIMA for HMSA method.

TABLE 4. Cardano price prediction accuracy score using LSTM, ARIMA, and MLP algorithms by MSE, RMSE, MAE, and MAPE indicators for WLPS, IBWPS, DBSA, and HMSA methods.

Method	Algorithm	Performance Indicators			
		MSE	RMSE	MAE	MAPE
WLPS	MLP	0.01082	0.10406	0.06693	6.69305
	ARIMA	0.00194	0.04409	0.03078	3.07824
	LSTM	0.00860	0.09275	0.06371	6.37107
IBWPS	MLP	0.00539	0.07348	0.04916	4.91676
	ARIMA	0.00177	0.04213	0.02828	2.82852
	LSTM	0.00366	0.06055	0.038507	3.85070
DBSA	MLP	0.00597	0.07729	0.04816	4.81658
	ARIMA	0.00188	0.04345	0.03033	3.03337
	LSTM	0.00580	0.07615	0.05347	5.34779
HMSA	MLP	0.00475	0.06896	0.044307	4.43071
	ARIMA	0.00174	0.04172	0.02645	2.64566
	LSTM	0.00317	0.056387	0.04545	4.54521

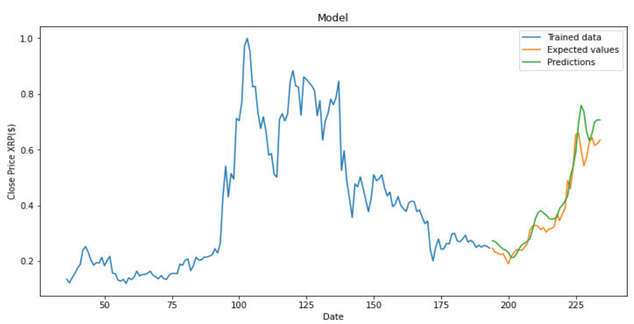

FIGURE 50. Price prediction result for Ripple using MLP for WLPS method.

WLPS. Also, the HMSA method had 10.4539% better results compared to WLPS.

The results obtained by LSTM indicate that IBWPS performed 56.3293%, DBSA performed 37.5356%, and HMSA performed 57.8048% better than WLPS.

Based on the MAPE indicator, by getting the average accuracy score of the five investigated cryptocurrencies in both WLPS and IBWPS methods, it was concluded that there is a significant difference between the accuracy obtained from WLPS and IBWPS using the LSTM model ($p = 0.045$), considering that the average accuracy score obtained

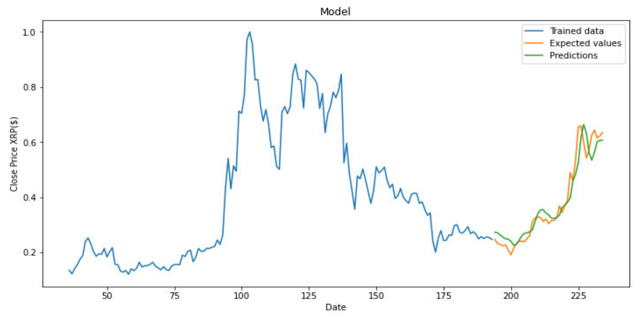

FIGURE 51. Price prediction result for Ripple using MLP for IBWPS method.

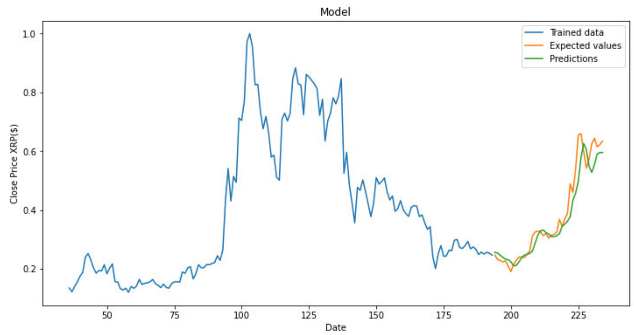

FIGURE 52. Price prediction result for Ripple using MLP for DBSA method.

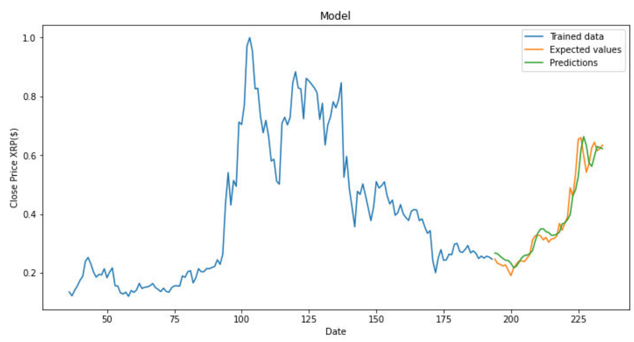

FIGURE 53. Price prediction result for Ripple using MLP for HMSA method.

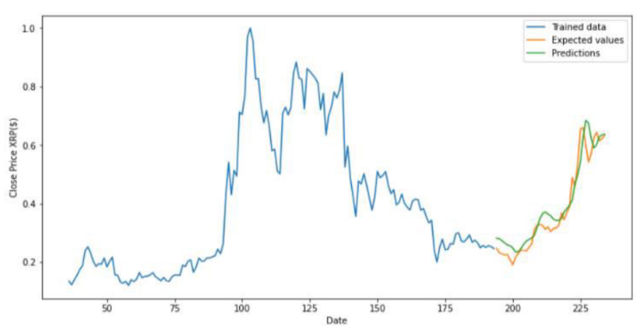

FIGURE 54. Price prediction result for Ripple using LSTM for WLPS method.

using the LSTM model by the WLPS method according to the MAPE indicator was 4.675474 and for IBWPS was 2.970972.

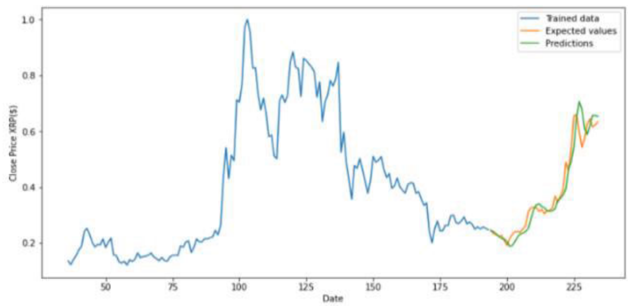

FIGURE 55. **Price prediction result for Ripple using LSTM for IBWPS method.**

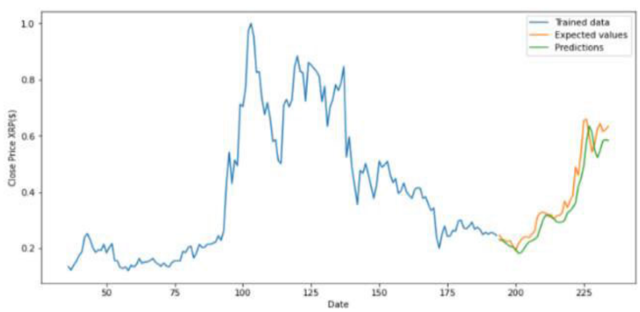

FIGURE 56. **Price prediction result for Ripple using LSTM for DBSA method.**

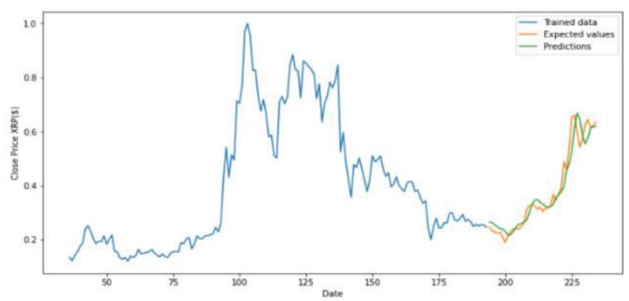

FIGURE 57. **Price prediction result for Ripple using LSTM for HMSA method.**

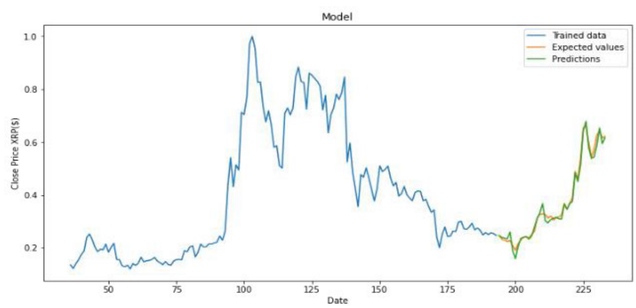

FIGURE 58. **Price prediction result for Ripple using ARIMA for WLPS method.**

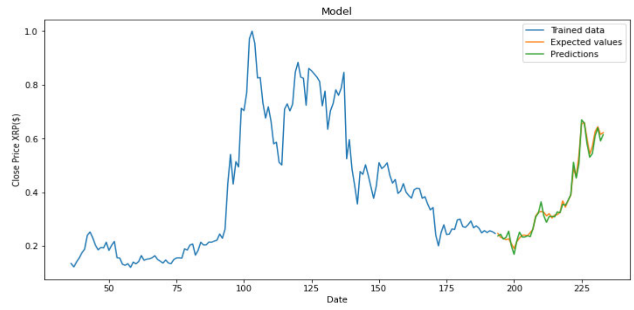

FIGURE 59. **Price prediction result for Ripple using ARIMA for IBWPS method.**

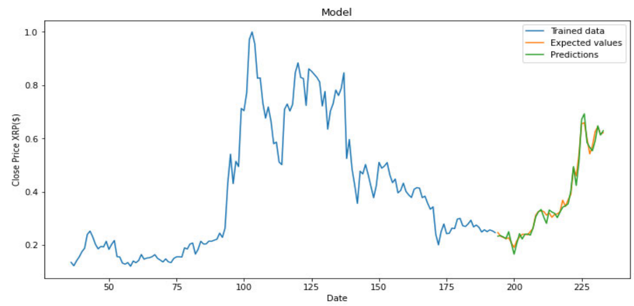

FIGURE 60. **Price prediction result for Ripple using ARIMA for DBSA method.**

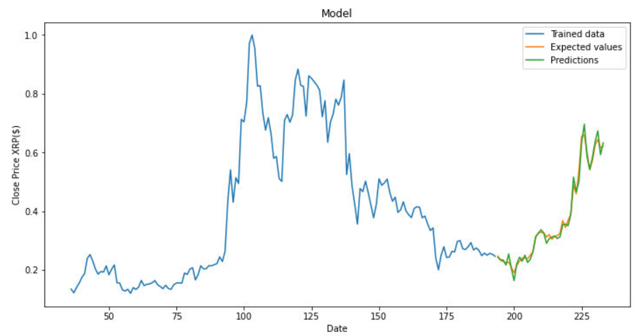

FIGURE 61. **Price prediction result for Ripple using ARIMA for HMSA method.**

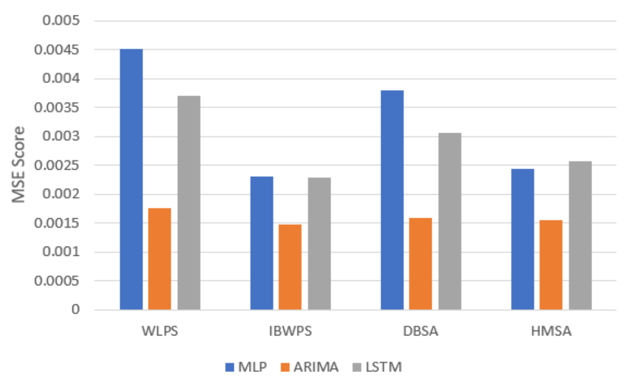

FIGURE 62. **The accuracy scores of Bitcoin according to MSE indicator per each method.**

According to the results obtained in all proposed models for predicting the price of Bitcoin and Ripple, it was shown that the accuracy in predicting the prices is mainly dependent on the content published by people who have higher influence in social networks. By including the number of followers in the process of giving weight to the calculated polarity scores of each tweet, the most accurate results in

TABLE 5. Ripple price prediction accuracy score using LSTM, ARIMA, and MLP algorithms by MSE, RMSE, MAE, and MAPE indicators for WLPS, IBWPS, DBSA, and HMSA methods.

Method	Algorithm	Performance Indicators			
		MSE	RMSE	MAE	MAPE
WLPS	MLP	0.00322	0.05680	0.04158	4.15884
	ARIMA	0.00194	0.04412	0.03099	3.09901
	LSTM	0.00154	0.03927	0.02891	2.89137
IBWPS	MLP	0.00150	0.03883	0.02773	2.77389
	ARIMA	0.00175	0.04184	0.02901	2.90164
	LSTM	0.00059	0.02441	0.01922	1.92201
DBSA	MLP	0.00216	0.04656	0.03165	3.16553
	ARIMA	0.00190	0.04361	0.02959	2.95932
	LSTM	0.00112	0.03357	0.02631	2.63123
HMSA	MLP	0.00185	0.04310	0.03192	3.19287
	ARIMA	0.00188	0.04343	0.02969	2.96996
	LSTM	0.00087	0.02958	0.02385	2.38514

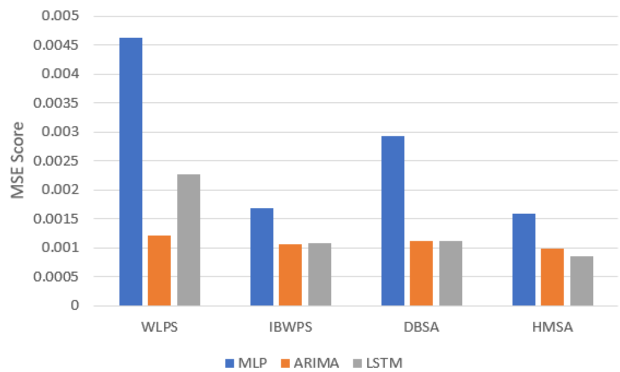

FIGURE 63. The accuracy scores of Ethereum according to MSE indicator per each method.

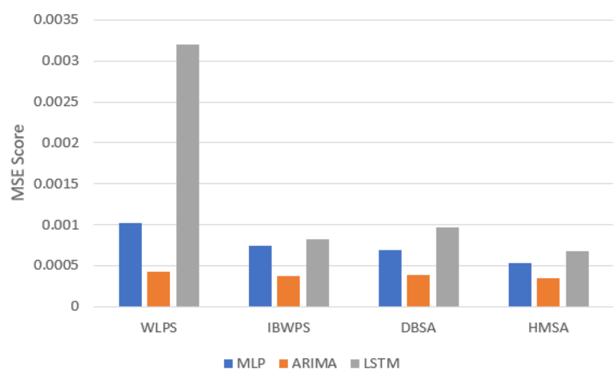

FIGURE 64. The accuracy scores of EOS according to MSE indicator per each method.

predicting the price of Bitcoin and Ripple can be obtained. As shown in FIGURES 62 to FIGURE 66, the MSE scores of each proposed model for Bitcoin, Ethereum, Cardano, EOS, and Ripple were calculated. These charts illustrate that all methods using the WLPS method had the lowest accuracy.

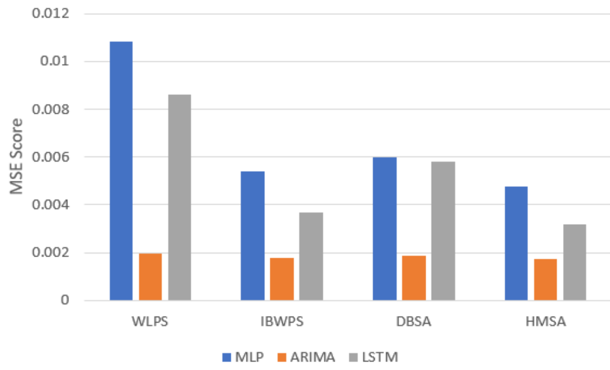

FIGURE 65. The accuracy score of Cardano based on MSE indicator per each method.

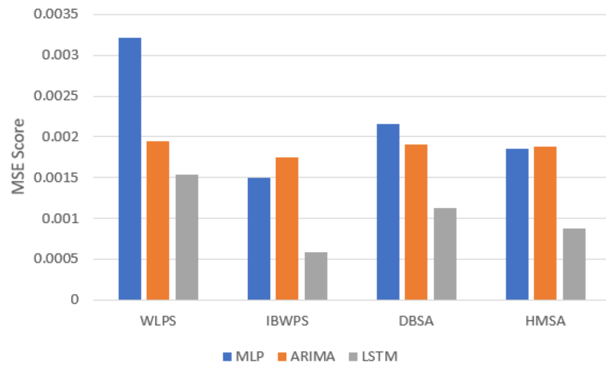

FIGURE 66. The accuracy score of Ripple based on MSE indicator per each method.

TABLE 6. The average MSE scores obtained for all five investigated cryptocurrencies.

	Method			
Algorithm	WLPS	IBWPS	DBSA	HMSA
MLP	0.004844	0.002326	0.003109	0.002233
ARIMA	0.001454	0.001287	0.001375	0.001302
LSTM	0.003863	0.001687	0.002413	0.00163

In the research conducted by Abraham et al. in 2018, the dependency between the prices of two investigated cryptocurrencies, Bitcoin and Ethereum, and the average of daily polarity scores was proven [7]. The difference between the current research and the research conducted by Abraham et al., was applying weight to the polarity scores based on the number of followers of the tweet's publishers. It was shown that by applying a coefficient to the polarity scores based on the influence of the publishers and identifying the percentage of positivity and negativity of tones in the tweets of each cryptocurrency on a daily basis, more accurate results can be achieved in predicting the prices compared to the case where the polarity scores are unweighted. In Serafini et al.'s research, the prediction method was proposed based on various combinations of features such as volume of trades, sentiments of tweets, and the volume of published tweets to predict the price of Bitcoin. According to Serafini et al.'s study, by comparing the accuracy of prediction based on the MSE indicator based on all different combinations of the mentioned features, it was concluded that if the prediction model is implemented based on only the sentiment scores of tweets and the ignorance of other features, the results will be more accurate [34]. So, in our research, only the feature of sentiment scores has been used in our proposed model, and other features like the volume of trades and the volume of publications have been ignored in our price prediction model. Usually, the prices of cryptocurrencies are affected by the sentiment of a society in the recent days, and according to the previous studies, researchers will use the sentiment of the tweets published today for predicting the prices of cryptocurrencies in the next day [35], and that wouldn't be possible to predict the prices in the distant future unless being informed about the world's events that may occur in the future.

VI. CONCLUSION

This paper proposed new approaches for predicting the prices of cryptocurrencies using the various methods of considering the sentiment scores of the tweets alongside the historical prices of cryptocurrencies to propose a prediction model that has the most accurate results in predicting the prices of cryptocurrencies, considering the specifications of the users on social media. The results have shown that the inclusion of the influence factor of twitterers results in more accurate predictions compared to the models in which the influence level of tweets' publishers has not been included. In this research, a novel approach, including the influence factor of the Twitterers, was proposed to be used as input data for machine learning models. MLP, ARIMA, and LSTM were selected to make predictions due to their suitability for time-series data analysis. To be able to compare the results of the proposed prediction models for predicting the prices of cryptocurrencies such as Bitcoin, Ethereum, EOS, Cardano, and Ripple, the prices of all investigated cryptocurrencies were normalized to be in the range between 0 and 1. Meanwhile, after preprocessing the tweets by removing all non-English characters, punctuation, and stop words, the polarity scores of each tweet were calculated, and the process of applying weights to the polarity scores was conducted based on the normalized number of followers of each tweet's publishers. In this study, four methods of considering the polarity scores, named WLPS, IBWPS, DBSA, and HMSA, have been proposed to be used in predicting the prices of cryptocurrencies. In the WLPS method, the same as in traditional models, the daily average of unweighted polarity scores was used in the prediction model, and the lowest accuracy was obtained by the use of the WLPS method. In the second proposed method, where the influence factor of the twitterers was included in the prediction model, named IBWPS, the polarity scores were weighted based on the number of followers of the twitterers, and the most accurate results among all other methods were obtained for predicting the prices of Bitcoin and Ripple using the IBWPS method. In the third proposed method (DBSA), the percentage of positive and negative tones in the daily tweets has been separately used in predicting the prices of the investigated cryptocurrencies by one of the dictionary-based tools (LIWC). According to the results of predicting the cryptocurrencies' prices using DBSA compared to WLPS, it was shown that DBSA outperformed WLPS, although in both of the WLPS and DBSA methods, the influence factor of the twitterers was not included in the prediction model. The fourth proposed method for predicting the prices of cryptocurrencies was a hybrid model (HMSA), where the daily average of weighted polarity scores along with the percentage of positive and negative tones were used to predict the prices of cryptocurrencies. The results showed that the most accurate results were obtained in predicting the prices of Cardano, Ethereum, and EOS using the HMSA method with all three machine learning models.

In this study, only the influence factor of twitterers on social media was investigated, and the effectiveness of including this factor in the prediction model was proven. Knowing that the identification of all effective factors that could be extracted by the analysis of the network's topology and the measurement of the effectiveness of each item for predicting the future prices of cryptocurrencies, although encountering complexities, could enhance the performance of prediction models, it is recommended to conduct deep investigations on the network's topology to extract the effective features to be included in the prediction models to make it possible to select a model that could have the most accurate results in predicting the cryptocurrency prices and outperformed the other models.

REFERENCES

[1] Z. Zhou, Z. Song, H. Xiao, and T. Ren, "Multi-source data driven cryptocurrency price movement prediction and portfolio optimization," *Expert Syst. Appl.*, vol. 219, Jun. 2023, Art. no. 119600.

[2] G. Hileman and M. Rauchs, "2017 global cryptocurrency benchmarking study," *J. SSRN*, 2017.

[3] T. R. Li, A. S. Chamrajnagar, X. R. Fong, N. R. Rizik, and F. Fu, "Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model," *Frontiers Phys.*, vol. 7, p. 98, Jul. 2019.

[4] K. Rathan, S. V. Sai, and T. S. Manikanta, "Crypto-currency price prediction using decision tree and regression techniques," in *Proc. 3rd Int. Conf. Trends Electron. Inform. (ICOEI)*, Apr. 2019, pp. 190–194.

[5] D. Garcia and F. Schweitzer, "Social signals and algorithmic trading of Bitcoin," *Roy. Soc. Open Sci.*, vol. 2, no. 9, Sep. 2015, Art. no. 150288.

[6] M. Ortu, S. Vacca, G. Destefanis, and C. Conversano, "Cryptocurrency ecosystems and social media environments: An empirical analysis through Hawkes' models and natural language processing," *Mach. Learn. Appl.*, vol. 7, Mar. 2022, Art. no. 100229.

[7] 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.

[8] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *Int. J. Forecasting*, vol. 14, pp. 35–62, Mar. 1998.

[9] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003.

[10] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2664–2675, Mar. 2011.
[11] A. Pinkus, "Approximation theory of the MLP model in neural networks," *Acta Numerica*, vol. 8, pp. 143–195, Jan. 1999.
[12] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.
[13] D. S. D. O. S. Júnior, J. F. L. de Oliveira, and P. S. G. de Mattos Neto, "An intelligent hybridization of ARIMA with machine learning models for time series forecasting," *Knowl.-Based Syst.*, vol. 175, pp. 72–86, Jul. 2019.
[14] K. Wołk, "Advanced social media sentiment analysis for short-term cryptocurrency price prediction," *Expert Syst.*, vol. 37, no. 2, Apr. 2020, Art. no. e12493.
[15] Z. Zhang, H.-N. Dai, J. Zhou, S. K. Mondal, M. M. García, and H. Wang, "Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115378.
[16] H. A. Dolatsara, E. Kibis, M. Caglar, S. Simsek, A. Dag, G. A. Dolatsara, and D. Delen, "An interpretable decision-support systems for daily cryptocurrency trading," *Expert Syst. Appl.*, vol. 203, Oct. 2022, Art. no. 117409.
[17] T. Zoumpekas, E. Houstis, and M. Vavalis, "ETH analysis and predictions utilizing deep learning," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113866.
[18] H.-M. Kim, G.-W. Bock, and G. Lee, "Predicting Ethereum prices with machine learning based on blockchain information," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115480.
[19] O. Kraaijeveld and J. De Smedt, "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices," *J. Int. Financial Markets, Institutions Money*, vol. 65, Mar. 2020, Art. no. 101188.
[20] H. Rezaei, H. Faaljou, and G. Mansourfar, "Stock price prediction using deep learning and frequency decomposition," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114332.
[21] S. Bhatt, M. Ghazanfar, and M. Amirhosseini, "Machine learning based cryptocurrency price prediction using historical data and social media sentiment," *Comput. Sci. Inf. Technol.*, vol. 13, no. 10, pp. 1–11, 2023.
[22] D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, and B. K. Lama, "Recurrent neural network based Bitcoin price prediction by Twitter sentiment analysis," in *Proc. IEEE 3rd Int. Conf. Comput., Commun. Secur. (ICCCS)*, Oct. 2018, pp. 128–132.
[23] G. Kim, D.-H. Shin, J. G. Choi, and S. Lim, "A deep learning-based cryptocurrency price prediction model that uses on-chain data," *IEEE Access*, vol. 10, pp. 56232–56248, 2022.
[24] J. V. Critien, A. Gatt, and J. Ellul, "Bitcoin price change and trend prediction through Twitter sentiment and data volume," *Financial Innov.*, vol. 8, no. 1, pp. 1–20, Dec. 2022.
[25] A. A. Oyedele, A. O. Ajayi, L. O. Oyedele, S. A. Bello, and K. O. Jimoh, "Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119233.
[26] J. Park and Y.-S. Seo, "Twitter sentiment analysis-based adjustment of cryptocurrency action recommendation model for profit maximization," *IEEE Access*, vol. 11, pp. 44828–44841, 2023.
[27] J. B. Awotunde, R. O. Ogundokun, R. G. Jimoh, S. Misra, and T. O. Aro, "Machine learning algorithm for cryptocurrencies price prediction," in *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*. Cham, Switzerland: Springer, 2021, pp. 421–447.
[28] S. Alonso-Monsalve, A. L. Suárez-Cetrulo, A. Cervantes, and D. Quintana, "Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113250.
[29] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
[30] M. Rafi, Q. A. K. Mirza, M. I. Sohail, M. Aliasghar, A. Aziz, and S. Hameed, "Enhancing cryptocurrency price forecasting accuracy: A feature selection and weighting approach with bi-directional LSTM and trend-preserving model bias correction," *IEEE Access*, vol. 11, pp. 65700–65710, 2023.
[31] A.-D. Vo, "Sentiment analysis of news for effective cryptocurrency price prediction," *Int. J. Knowl. Eng.*, vol. 5, no. 2, pp. 47–52, 2019.
[32] J. Golbeck, *Introduction to Social Media Investigation: A Hands-On Approach*, Syngress, 2015.
[33] S. Prayudani, A. Hizriadi, Y. Lase, and Y. Fatmi, "Analysis accuracy of forecasting measurement technique on random K-nearest neighbor (RKNN) using MAPE and MSE," *J. Phys., Conf. Ser.*, vol. 1361, Nov. 2019, Art. no. 012089.
[34] G. Serafini, P. Yi, Q. Zhang, M. Brambilla, J. Wang, Y. Hu, and B. Li, "Sentiment-driven price prediction of the Bitcoin based on statistical and deep learning approaches," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.
[35] R. Parekh, N. P. Patel, N. Thakkar, R. Gupta, S. Tanwar, G. Sharma, I. E. Davidson, and R. Sharma, "DL-GuesS: Deep learning and sentiment analysis-based cryptocurrency price prediction," *IEEE Access*, vol. 10, pp. 35398–35409, 2022.

FATEMEH FEIZIAN received the B.S. degree in computer engineering from Alzahra University, Tehran, Iran, and the M.S. degree in information technology from the Iran University of Science and Technology. She was a Data Analyst in different companies. Her research interests include data mining and machine learning.

BABAK AMIRI received the Ph.D. degree in information technology from The University of Sydney, in 2014. He is currently an Assistant Professor with the Iran University of Science and Technology. He has a track record of publishing quality research articles in respected outlets within the fields of complex systems, machine learning, and data science. Notably, the academic community has well-received research on topics, such as social media analysis, complex networks, customer analytics, and text analysis.