Designing forecasting assistant of the Bitcoin price based on deep learning using the market sentiment analysis and multiple feature extraction

Sina Fakharchian (Sina.cbar@gmail.com)
Islamic Azad University Islamshahr Branch: Islamic Azad University Eslamshahr Branch

Research Article

Keywords: Artificial intelligence, Price prediction assistant, Deep learning, Feature selection, Sentiment analysis

Posted Date: August 30th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1341589/v1

Designing forecasting assistant of the Bitcoin price based on deep learning using the market sentiment analysis and multiple feature extraction

Sina Fakharchian[1]*

[1] Islamic Azad University Islamshahr Branch, Department of Computer Engineering, IRAN

[*]Corresponding Author Email Address: Sina.cbar@gmail.com

Abstract

Nowadays, the issue of fluctuations in the price of digital Bitcoin currency has a striking impact on the profit or loss of people, international relations, and trade. Accordingly, designing a model that can take into account the various significant factors for predicting the Bitcoin price with the highest accuracy is essential. Hence, the current paper presents several Bitcoin price prediction models based on Convolutional Neural Network (CNN) and Long-Short-Term Memory (LSTM) using market sentiment and multiple feature extraction. In the proposed models, several parameters, including Twitter data, news headlines, news content, Google Trends, Bitcoin-based stock, and finance, are employed based on deep learning to make a more accurate prediction. Besides, the proposed model analyzes the Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiments to examine the latest news of the market and cryptocurrencies. According to the various inputs and analyses of this study, several effective feature selection methods, including mutual information regression, Linear Regression, correlation-based, and a combination of the feature selection models, are exploited to predict the price of Bitcoin. Finally, a careful comparison is made between the proposed models in terms of some performance criteria like Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Median Absolute Error (MedAE), and coefficient of determination (R^2). The obtained results indicate that the proposed hybrid model based on sentiments analysis and combined feature selection with MSE value of 0.001 and R^2 value of 0.98 provides better estimations with more minor errors regarding Bitcoin price. This proposed model can also be employed as an individual assistant for more informed trading decisions associated with Bitcoin.

Keywords: Artificial intelligence; Price prediction assistant; Deep learning; Feature selection; Sentiment analysis

1. Introduction

The last decades have witnessed a remarkable growth in the use of digital currencies by people and organizations. Nowadays, the issue of cryptocurrencies has received much attention and is being widely examined in the literature (Chaudhari and Crane, 2020; Dai et al., 2021; ElRahman and Alluhaidan, 2021; Li et al., 2021; Zuiderwijk et al., 2021). In the modern world, cryptocurrency has been introduced as a novel and emerging topic which is governed by the cryptographic protocol using Blockchain (Chohan, 2017). Considering the concept of this cryptocurrency, the way people think about money has been revolutionized (Pant et al., 2018). Also, the value of the cryptocurrency has been significantly raised due to the continuous rise in adoption and widespread usage of that in the real world. According to the striking value of cryptocurrencies, some people consider them as equal with real currencies or Fiat currencies. In comparison, others regard them as a good opportunity to invest. On January 9, 2017, the value of a Bitcoin has increased from \$ 863 by 2000% and reached its highest price level, i.e., \$ 17,550, on December 11, 2017. Eight weeks later, on February 5, 2018, the price of Bitcoin became less than half of the price mentioned earlier, i.e., about \$ 7900. Nevertheless, the promising technology behind cryptocurrencies, namely the Chinese blockchain, is about to raise the use of cryptocurrencies. Kristoufek stated that Bitcoin is a unique asset, and the price of Bitcoin cryptocurrencies acts like a standard financial asset (Kristoufek, 2015). Bitcoin is regarded as the first decentralized digital currency in which the transactions are conducted directly between the users and no intermediary (Matta et al., 2015; Naimy and Hayek, 2018). This type of currency is fundamentally different from what is typically employed in a prevalent monetary system. Based on mining, the cryptocurrency is created, which has led to considerable variations in the online economic activities of users worldwide (Jain et al., 2018). Due to the fact that the price of cryptocurrencies does not behave as in the past, it is significantly difficult to predict the price of cryptocurrencies. Additionally, the large fluctuations in the price of cryptocurrency, random effects in the market, and the influence of various factors on the price of Bitcoin, have become a globally novel challenge. Hence, the issue of predicting the variations in the price of cryptocurrency Bitcoin is of great importance. On the other hand, there are many opportunities for better understanding the drivers of the Bitcoin price (Karalevicius et al., 2018).

Moreover, since no central governing authority controls the digital currency and is mainly affected by the general public, Bitcoin is regarded as a volatile currency that changes based on socially constructed ideas. Therefore, the issue of sentiment analysis in the prediction of Bitcoin is of great importance, and many authors have studied it in this regard. The idea of some economists such as Daniel Kahneman and Amos Tversky has proved that the decisions made in this field are influenced by sentiments (KAI-INEMAN and Tversky, 1979). The study of R. J. Dolan regarding "Emotions, Cognition, and Behavior" further confirms that decision-making is extremely affected by sentiments (Dolan, 2002). Actually, the sentiment analysis indicates that demand for a good product, and consequently price, maybe influenced more than its economic basics. In recent years, researchers have specifically found that purchasing decisions are made by people and are under the effect of online data collection (Mittal et al., 2019). Galen Thomas Panger stated that Twitter sentiments are related to people's overall sentimental state. In addition, it was revealed that social media such as Twitter has a calming effect rather than reinforcing the user's sentimental state (Panger, 2017). Based on a textual analysis conducted on a social context with the aim of investors called "Search Alpha”, Chen et al. stated that the comments outlined in the submitted articles of "Seeking Alpha" were highly effective and even could predict the astonishments of profitability (Chen et al., 2013). In a similar study, Tetlock demonstrated that high levels of media pessimism in the stock market directly affect trading volume (Tetlock, 2007). Finally, in another study, Gartner pointed out that most users use social media to make their final decisions for purchasing (Pettey, 2010).

Over time, extensive literature has developed on the effectiveness of tweet sentiments. Kouloumpis et al. showed that standard methods of natural language processing like sentence scoring were ineffective due to the short nature of tweets and the uniqueness of this writing style

(Kouloumpis et al., 2011). Pak and Patrick divided the individual tweets into positive, negative, or neutral categories that could better understand sentiments by the computer (Pak and Paroubek, 2010). O'Connor et al. indicated that the sentiments in tweets reflect the public opinion on various topics in public opinion surveys (O'Connor et al., 2010). This study identified sentiment analysis as a more cost-effective option versus public opinion surveys. Nevertheless, according to this concept, the sentiments generated by tweets more accurately reflect the sentiments of the majority of people on the topic. Hence, it can be considered for predicting demand and the results of variations in the products' price. In another study, the researchers found that employment-related searches were related to the unemployment rate (Ettredge et al., 2005). A relationship between the volume of inquiries and the volume of stock trading on NASDAQ was observed in the study of Bordino et al. (Bordino et al., 2012). Choi and Varian have also conducted specific studies on Google Trends and presented remarkable results (Choi and Varian, 2012). According to the results of this study, it can be concluded that simple seasonal models of trend data are considered input data that outperform models that did not use Google Trends. Also, Asur et al. found that the extent of how much a keyword is a trend in the newly released films accurately predicted their revenue in the box office (Asur and Huberman, 2010).

Overall, the data of sentimental can be used to predict variation in macroeconomic statistics, and many studies have been performed in this field. Several researchers, including Choi and Varian (Choi and Varian, 2012) and Ettredge et al. (Ettredge et al., 2005), have claimed that web-based search data, which is the same as Google Trends data, can be particularly utilized to predict the price of Bitcoin. Dennis and Yuan collected capacity scores in tweets associated with 500 S&P companies and realized a correlation between them and stock prices (Sul et al., 2014). De Jong et al. analyzed minute-by-minute stock prices and tweet data of 30 stocks at the Dow Jones industrial average (de Jong et al., 2017). Accordingly, it was revealed that 87% of stock returns were under the effect of such tweets; however, the authors also sought the vice versa in that stock. As a result, the prices affected tweets. Bollen et al. used a self-organizing fuzzy neural network to predict price changes in the DOW Jones Industrial Average and obtained 86.7% accuracy by Twitter sentiments (Bollen et al., 2011). Evita Stenqvist and Jacob L¨onn¨o presented a study, "Predicting Bitcoin price fluctuation by analyzing Twitter sentiments," and obtained striking results (Stenqvist and Lönnö, 2017). The authors collected and processed the tweets regarding Bitcoin and Bitcoin prices from May 11 to June 11. Then, the unrelated or unaffected tweets were eliminated from the analysis. After that, the authors used the VADER (Valence Aware Dictionary and Sentiment Reasoner) method to analyze the tweets' text. Besides, the authors categorized the sentiments of each tweet and labeled them negatively, neutrally, or positively. Connor et al. employed the sentiment of news headlines and tweets to predict price changes in Bitcoin, Light Coin, and Atrium (Lamon et al., 2017). The results of this study represent the remarkable performance of the logistic regression for classifying these tweets. The authors also accurately predicted the 43.9% price increase and 61.9% price reduction. Colianni et al. collected tweets from November 15, 2015, to December 3, 2015, and used Naive Bayes and Support Vector Machines to classify tweets, and reached higher accuracy for predicting price (Colianni et al., 2015). Finally, Shah et al. successfully presented a strategy using historical prices and Bayesian regression analysis (Shah and Zhang, 2014).

Traditional time series prediction techniques like Holt-Winters exponential smoothing models are fundamentally related to linear assumptions and need data with the capability of breaking down into trend, seasonal, and noise to be effective (Chatfield and Yar, 1988). Since the Bitcoin market mainly lacks seasonality and high volatility, the traditional methods are not useful. To tackle this drawback, deep learning (DL) technology has been introduced as a novel technique that reduces the costs and complexity of the calculations (McNally et al., 2018). Unlike the traditional linear statistical models, the artificial intelligence (AI) method is able to consider the nonlinear property. Notably, artificial neural networks (ANNs) with deep learning (DL) algorithms are regarded as the most thriving methods due to their remarkable predictive capabilities (Nakano et al., 2018). In the cutting-edge paper of 2017, A. Radityo et al. employed ANN to forecast the next-day price of Bitcoin (Radityo et al., 2017). Four types of ANN algorithms have been considered in this study, namely, Neuro Evolution of Augmenting Topologies (NEAT), Genetic Algorithm Neural Network (GANN), Genetic Algorithm Backpropagation Neural Network (GABPNN), and Backpropagation Neural Network (BPNN). Considering machine learning algorithms such as generalized linear models and random forests, Bitcoin price prediction was modeled by Madan et al. binomial classification problem (Madan et al., 2015). In 2008, Zhu et al. used the volume of the stock transactions as a neural network input to improve the forecasting performance in the medium and long term and presented acceptable results (Zhu et al., 2008). A modular neural network was employed by Kimoto et al. for predicting the best shopping point (Kimoto et al., 1990). Guresen et al. compared the performance of different neural networks in stock market prediction and proved that a multilayer perceptron (MLP) neural network outperforms the others (Guresen et al., 2011). In contrast, S. McNally stated that the capabilities of the recurrent neural network (RNN) and the long short term memory (LSTM) outweigh the benefits of MLP due to the temporal nature of Bitcoin data (McNally et al., 2018). Similarly, in 2019 S. Tandon et al. attempted to present the price prediction model to forecast the Bitcoin price using RNN and LSTM with 10-fold cross-validation. A careful comparison was made between the proposed model and other available models, including RNN with LSTM, Linear Regression, and Random Forest. The benefits of the proposed model were proved, and remarkable results were presented. In a major advance of 2020, Dutta et al. has used a gated recurrent unit method to forecast the Bitcoin price and obtained acceptable results (Dutta et al., 2020). In 2021, Ramadhan et al. also used (LSTM-RNN) for predicting the Bitcoin price (Ramadhan et al., 2021). A hybrid Bitcoin price prediction method based on ANN and using Bi-LSTM and Bi-RNN was also presented by Das et al. in 2021, and the benefits of the proposed method were revealed (Das et al., 2021). Despite this interest, no one as far as we know has studied the issue of Bitcoin price prediction considering Twitter data, news headlines, news content, Google Trends, Bitcoin-based stock, and finance using CNN and LSTM.

Considering CNN and LSTM, the current paper aims to propose a model for forecasting the variations in the price of cryptocurrency Bitcoin. For this purpose, a variety of methods of textual sentiment analysis such as news headlines, news, and tweets are considered. Such methods consist of using the Twitter API, a Python library, namely 'Tweepy,' extracting text and news content from the Telegram channel, the reference site regarding cryptocurrencies, namely Kevin Telegraph, receiving and extracting Google Trends data. In the beginning, using the tweets in which the Bitcoin is mentioned, the data are collected from the storage. Then, the tweets are analyzed to calculate the sentiments score and compare it to other days. After that, that day's price is examined to determine if there is a relationship between tweets and variations. As a result, variations in the price of cryptocurrencies can be determined using the sentiment. A careful comparison is made between the proposed models in terms of some performance criteria like Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Median Absolute Error (MedAE), and coefficient of determination (R^2). The major contributions of this paper are summarized as follows:

- Presenting Bitcoin price prediction models based on Convolutional Neural Network (CNN) and Long-Short-Term Memory (LSTM) using market sentiment and multiple feature extraction.
- Analyzing the VADER sentiments to examine the latest news of the market and cryptocurrencies.
- Pricing Bitcoin price considering feature selection methods, including mutual information regression, Linear Regression, correlation-based, and a combination of the feature selection models.

The remaining of this paper is organized as follows: The major fundamental concepts regarding the topic of the present study, including an overview of cryptocurrencies, Twitter, sentiment analysis, and Google Trends, are explained in the second section. The issue of data collection is also illustrated in the third section. The research method and how to select model inputs are examined in the fourth section. Finally, a summary regarding the present study and the main conclusions as well as suggestions for future studies are presented in the fifth section.

2. Preliminaries

The analysis presented in this paper needs an understanding of why and how cryptocurrencies are different from valid currencies or stocks in the companies of the traditional stock market. This section provides more information regarding such reasons and clarifies why these cryptocurrencies are used. Since cryptocurrencies are part of the more extensive technology (China Blockchain), Twitter activities can be considered very effective. It should be noted that Google Trends data and the volume of tweets represent an overall tendency to have cryptocurrencies. Hence, the basic concepts concerning cryptocurrencies, Twitter, sentiment analysis, and Google Trends are described here.

2.1 Blockchain and cryptocurrencies

The data of the first cryptocurrency in the world are analyzed in this paper. Bitcoin is the largest cryptocurrency in terms of market size, followed by Atrium. Bitcoin was the first cryptocurrency to be created. Creating Bitcoin is mysterious since it was created by a person or group of people using the name "Satoshi Nakamoto" in 2009. At the same time as launching Bitcoin, Satoshi Nakamoto presented a paper entitled "Bitcoin: A peer-to-peer Payment Method" (Nakamoto, 2008). In contrast to cash, this system outlines a peer-to-peer payment method using an electronic system. The cryptocurrency can be sent directly from one party to another without the use of a third party to verify the transaction between them. This innovation is presented employing the "blockchain," which is like a common ledger in the whole transaction. This is a peer-to-peer network in which the network verifies the whole transactions to prevent forging them.

Since the applications of blockchain technology go far beyond peer-to-peer payments, this technology provides security, privacy, and decentralization. A decentralized office exploits the blockchain for IoT applications, isolated storage systems, healthcare, and more (Xu and Croft, 1998). The range of blockchain applications has led to creating more blockchains and cryptocurrencies. Furthermore, using the blockchain increases the usage of cryptocurrencies and gives them intrinsic value whose amount depends on many factors. The main reason is this it is a new technological debate. Notably, the information regarding the type of currency and how it stores its new value is useful to improve understanding of what can lead to price changes.

2.2 Twitter

Twitter was created in July 2006 as an application that consists of other applications, websites (such as Instagram, Facebook, LinkedIn, etc.), and microblogging. A microblog is a medium that allows smaller and more frequent updates compared to blogging to be performed. Twitter allows users to send messages publicly (called "tweets") up to 140 characters long, which was doubled on November 6, 2017, to 280 characters per tweet. Users can add a "hashtag" to the tweet, denoted by the symbol of "#." This symbol follows a sequence of characters employed to identify the subject of a tweet and search for that. Hashtags are considered later when collecting tweets in the data section.

It is noteworthy that Twitter has received much popularity rapidly since its launch in 2006. Evidence that shows how much Twitter is essential dates back to January 15, 2009, when an Airways plane crashed in the Hudson River in the United States. An image that was posted on Twitter regarding that incident broke the record of the views' number. Because 83% of the world's leaders have Twitter accounts, Twitter earns nearly $ 330 million a month with 1.3 billion users. Due to such considerable statistics, it should be noted that the Twitter database can be significantly rich and efficient. It is considered a great source of information showing how people almost feel about anything you want. Also, you can observe how these feelings change over time since it has the capability to inform you when a tweet has been sent. Hence, Twitter is regarded as a remarkable resource for collecting textual data on a topic such as cryptocurrencies to explore possible relationships between them and their prices.

2.3 Sentiment analysis

It can be estimated that 90% of the global data has been generated in the last two years. Most of this data is in the form of textual data without structure. This data can also be in the form of tweets, articles posted on the Internet, text messages, emails, or others which create such a wide amount of unstructured data. "Natural language processing" (NLP) is considered a novel discussion that is being studied or developed. There is a set of methods for computers to analyze and understand the text. In this paper, a set of natural language processing tools called "emotion analysis" is employed. Sentiment analysis is conducted for extracting and measuring the sentiments or mental opinions outlined in the text. There are several methods to do this, but the "VADER" (Valence Aware) method is selected in this study (Manning and Schutze, 1999). The aim here is to use sentiment analysis in the collected tweets for determining what tweets have positive or negative comments regarding cryptocurrencies.

2.4 Google Trends

In many parts of the world, almost the whole aspect of daily life includes the Internet. Browsing the Internet is conducted through search engines and Google. Nowadays, the most popular search engine in the world, with 74.52% of searches in Google. Therefore, Google search data can provide credible insights into what the world is interested in and the extent of this interest in anything. Google makes this data available through Google Trends. The data provides information concerning the popularity of the searched words compared to other words. There is a variation in the ranking of Google Trends data at different times in cryptocurrencies, which can be related to increasing and reduction of the public profit and the price of cryptocurrencies.

2.5 Headline and the main text of the day news

Due to the fact that the price of cryptocurrencies significantly depends on positive and negative news and the cryptocurrency market follows more fundamental analysis, we decided to extract news from the most globally reputable news site in the field of cryptocurrency, i.e., Kevin Telegraph, for increasing the accuracy. In the period from "2021/02/05" to "2021/09/10", the extraction and analysis of sentiments based on Twitter data have also been conducted based on the news to see how the news is effective for determining the Bitcoin price.

3. Methodology

The main information regarding the proposed method of this study to predict the Bitcoin price is given in this section. Also, the main method and neural networks that are used to reach the final results and predict the Bitcoin price accurately are outlined in this section.

3.1 The proposed model

In this section, a price forecasting assistant or a predicter model based on CNN and long-short-term memory (LSTM) is analyzed using market sentiment and multiple feature extraction. The proposed model consists of different parts, and each part has information and details, which are described separately in the following section. Besides, the flowchart of the proposed method is demonstrated in Figure 1 for better understanding. Additionally, in the proposed models, VADER sentiment analysis is exploited to examine the latest market news of cryptocurrencies. In the proposed models, the Twitter data analysis, the news headlines, news content, Google Trends, Bitcoin stocks, and financials based on deep learning are employed to forecast the Bitcoin price better and more accurately

Moreover, due to the high extraction feature of different input data, the selection methods of the mutual information regression, Linear Regression, and correlation-based selection method are exploited. A combination of three feature selection methods is considered in a separated model to benefit from their advantages. According to the various input data in this section, nine different models are developed based on CNN and LSTM to forecast Bitcoin prices. In each of these proposed models, different layers and separated input data are considered to examine the effect of each input data on the Bitcoin price prediction. Finally, the various proposed models are compared with each other in terms of criteria such as Mean Square Error (MSE), Root Mean Square Error

(RMSE), Mean Absolute Error (MAE), Median Absolute Error (MedAE), and coefficient of determination (R^2).

According to the presented flowchart shown in Figure 1, this section consists of several main sub-sections, including data collection and data set, text preprocessing and text feature extraction, data normalization, VADER -based sentiment analysis, feature selection, proposed models based on deep learning, evaluating the performance criteria. In the following, these criteria and the main points and methods are illustrated in each subsection.

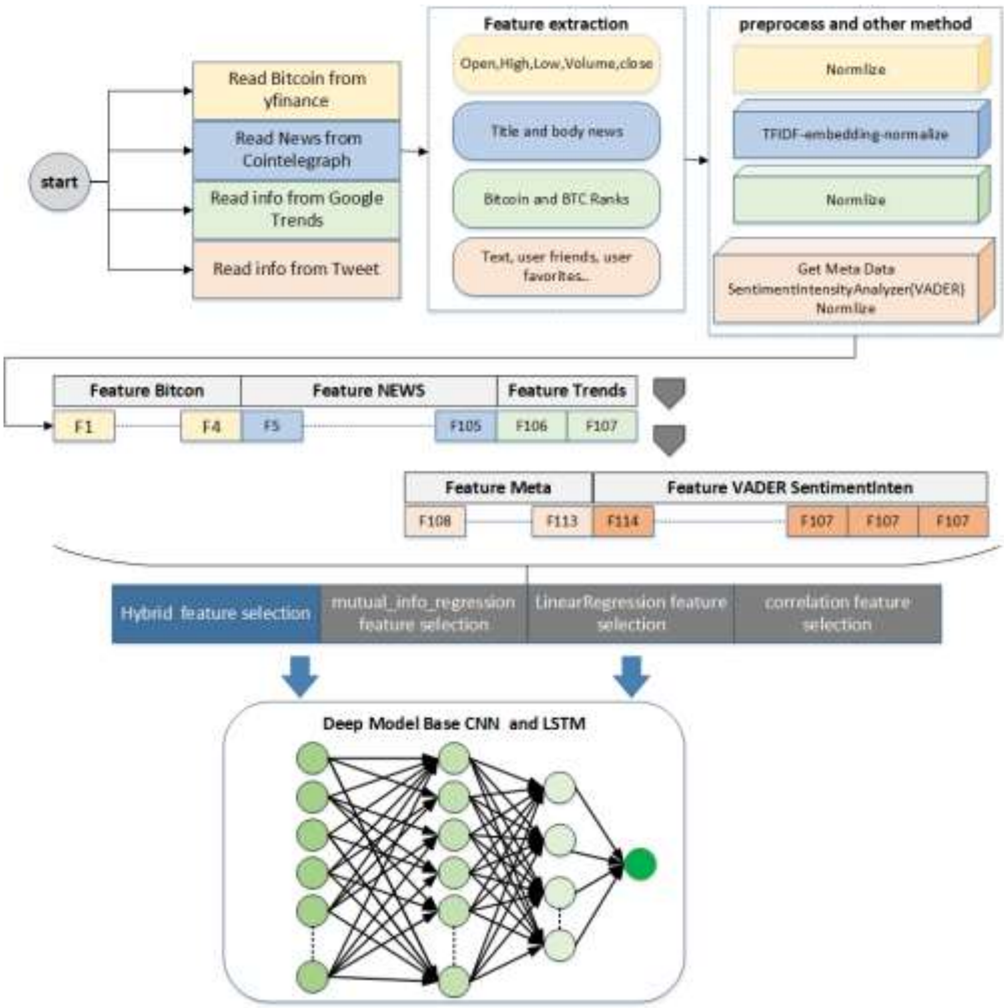

Figure 1. Flowchart of the proposed method based on multiple feature extraction and deep learning

3.2 Deep neural networks

A deep neural network (DNN), an artificial neural network (ANN) with multiple layers between the input and output layers, contains various types of neural networks but the same components: neurons, synapses, weights, biases, and functions always exist in this network. DNNs have been widely used in related work due to their remarkable application (P and M, 2021; Soni et al., 2021). In short, the literature pertaining to DNN (Awoke et al., 2021; Liu et al., 2021) strongly suggests that this technology is highly beneficial to developing the Bitcoin price-prediction models.

3.2.1 LSTM Networks

LSTM networks, abbreviated as "Long Short Term Memory," are a special type of recurrent neural network that has the ability to learn long-term dependencies. Hochreiter and Schmid Huber proposed these networks in 1997 for the first time. Notably, many researchers were involved in improving these networks, which are mentioned in the original text.

In fact, the major aim of designing LSTM networks was to deal with the problem of long-term dependency. It is noteworthy that memorizing information for long periods of time is the default and normal behavior of LSTM networks, and their structure is such that they can learn very distant information well, which is a striking characteristic of their structure.

The whole recurrent neural networks are in the form of iterative sequences (chains) of modules (units) of neural networks. In standard recurrent neural networks, these iterative modules have a simple structure. For instance, it has only one layer of hyperbolic tangent (tanh). Iterative modules have only one layer in standard recurrent neural networks.

LSTM networks have a similar sequence or chain structure, but the iterative module has various structures. They contain four layers rather than one layer of neural network that interact with each other according to a special structure. In LSTMs, iterative modules have four layers that interact with each other.

3.2.2 CNN neural networks

The convolutional neural network is similar to other neural networks (e.g., the MLP neural network) and is composed of neural layers with bias and weights and the ability to learn. The following items occur in each neuron:

- The neuron receives a set of inputs.
- Internal multiplication is conducted between the weights of the neurons and the inputs.
- The result is added to bias.
- Finally, a nonlinear function (the same as the activation function) is passed.

The above process is conducted layer by layer and reaches the output layer, creating the network forecast.

3.3 Feature selection methods

Feature selection is known as the process of specifying the least possible number of features in a data set that can describe this set and the main features (Alweshah et al., 2021). A feature selection aims to eliminate unnecessary features and select an important feature according to the data set and its class (Şahin et al., 2021). In the proposed model, three different selection models, including mutual information regression feature, Linear Regression, and correlation-based selection, are exploited. The proposed method is considered the correlation-based model since it can accurately identify the correlation between Bitcoin value and features. Therefore, the correlation-based model is useful to identify an important feature based on the correlation between the feature and the value of the class or Bitcoin. The mutual information feature selection method is one of the effective feature selection methods that is used in the proposed model (Kraskov et al., 2004). Mutual information shows a vital criterion of interdependence between features that are widely used in feature selection (Vergara and Estévez, 2014). In the third model, feature-importances of a Linear Regression model are exploited to have the features based on a regression model. This feature selection model is beneficial for having important features according to a Linear Regression model. The proposed deep model is considered for the feature selection methods, and the combination of these three feature models in Sections 4-6 are designed by Model-1, Model-2, and Model-3 models, respectively. Finally, the results that prove the superiority of these feature selection methods are highlighted in Section 6.

3.3.1 Correlation-based feature selection method

In this feature selection method, a subset of features is called a good subset whose features, on the one hand, have a high correlation with the "classification" or target feature, and on the other hand, are uncorrelated with each other. The extent of "merit" or goodness of a subset of features is calculated by the following equation:

$$Merit_{S_K} = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \tag{1}$$

3.3.2 Feature selection with mutual information

The features provide a lot of information from the output to the model, and the model estimates the amount of output in classification and regression projects based on this information. The mutual information method has a completely different approach to the previous methods and examines the relationship between a feature and the output instead of analyzing the mean and variance. Also, based on the amount of mutual information that a feature gives the output, it is scored. The approach of this method is significantly interesting and important, and it can accurately determine how much a feature is appropriate for estimating the output.

3.4 Statistical Analysis

This section considers the various criteria for examining the proposed method based on multiple feature extraction and deep learning. The majority of authors typically employ MSE error to make a comparison between the different models. In this paper, several main prediction criteria such as mean-square error (MSE), root-mean-square-error (RMSE), mean absolute error (MAE), median absolute error (MedAE), and coefficient of determination (R^2) are considered. In the following, the formulas of these criteria and their explanations are presented (Bui et al., 2018; Chou and Bui, 2014; Chou et al., 2016). Also, Table 1 summarizes the evaluation criteria of the proposed method.

Mean Square Error (MSE): This criterion calculates the mean square error of the distance between the predicted values of the proposed and actual Bitcoin method. The smaller the MSE values, the more accurate the Bitcoin prediction result of the proposed method. Through Equation 2, this criterion is calculated.

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

Where n denotes the number of samples, y_i is the experimental or actual Bitcoin values, and $\hat{y}_i$ represents the predicted Bitcoin values of the proposed method.

Root Mean Square Error (RMSE): If the square root of MSE is calculated, this criterion is called RMSE. In fact, the comparison between MSE and MAE is not correct due to the variation in the scale of the error value in MSE. Hence, it is necessary to define the RMSE criterion. This criterion is represented in Equation 3.

$$RMSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{3}$$

Where n shows the number of samples, the experimental or actual Bitcoin values are captured by y_i, and $\hat{y}_i$ indicates the predicted values of the proposed method. Notably, if the variance in individual errors is greater, the gap between the MAE and RMSE criteria becomes larger.

Mean Absolute Error (MAE): This criterion calculates the mean absolute difference between the predicted values of the proposed and actual Bitcoin method. The smaller the MAE values, the more accurate the prediction result of the proposed method. Through Equation 4, this criterion is calculated.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

According to Equation 4, n shows the number of samples, $y_i$is the experimental or actual values of Bitcoin, and $\hat{y}_i$ is the predicted values of the proposed method.

Median Absolute Error (MedAE): This median criterion is considered to calculate the absolute difference between the predicted values of the proposed and actual Bitcoin method. This criterion is shown in Equation 5.

$$MedAE = meadian(\,|y_1 - \hat{y}_2|, \ldots, |y_n - \hat{y}_n|) \tag{5}$$

According to Equation 5, n shows the number of samples, $y_i$is the experimental or actual Bitcoin values, and $\hat{y}_i$ is the predicted values of the proposed method.

Determination coefficient or detection coefficient (R^2): This criterion calculates how much the predicted values of the proposed method have a good agreement with the actual values of Bitcoin. In contrast to other criteria, the better it is to one. This criterion is calculated through Equation 6.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{6}$$

Accordingly, $\overline{y}_i$ represents the variables mean, n denotes the number of samples, the experimental or actual Bitcoin values are shown by y_i, and $\hat{y}_i$ demonstrates the predicted values of the proposed method.

Table 1. Summary of the criteria for evaluating the proposed method

Formula Number	Variables		Measure Full Name
(2)	Sample size The actual value of the ith Bitcoin The predicted value of the ith Bitcoin.	n y_i $\hat{y}_i$	Mean Squared Error (MSE)
(3)	Sample size The actual value of the ith Bitcoin The arithmetic mean of the Y	n y $\hat{y}$	Mean Absolute Error (MAE)
(4)	Sample size The actual value of the ith Bitcoin The predicted value of the ith Bitcoin.	n y_i $\hat{y}_i$	Root Mean Squared Error (RMSE)
(5)	Sample size The actual value of the ith Bitcoin The predicted value of the ith Bitcoin.	n y $\hat{y}$	Maximum Absolute Error(MedAE)
(6)	Sample size The actual value of the ith Bitcoin The predicted value of the ith Bitcoin. mean of response variables	n y_i $\hat{y}_i$ $\overline{y}_i$	The Coefficient of Determination (R^2)

3.5 VADER-based sentiment analysis

It is a sentiment analysis that aims to identify and extract users' opinions (Cambria et al., 2018). The primary aim of this section of the proposed method is to analyze the feelings of users' tweets. A variety of methods have been proposed for sentiment analysis, among which the VADER method is one of the successful methods in the field of sentiment analysis. As a matter of fact, VADER is a tool or library based on words and roll that can extract sentiments from text, emoticons, emojis, abbreviations, and terms accurately (Hota et al., 2021; Hutto and Gilbert, 2014). This tool has a better speed due to its vocabulary and roll, and its output is a 4-dimensional vector in which positive, negative, neutral, and compound values are generated for each input text. It should be noted that the positive, negative, and neutral values are normally considered between zero and one. Therefore, in the proposed method, the tweet text's positive, negative, neutral, and compound values are extracted in Table 2.

Table 2. The way of analyzing VADER-based sentiments in the proposed method

Tweet id	negative	neutral	positive	compound polarities
1	0.10	0.71	0.19	0.88
2	0.04	0.830	0.130	1
3	0.3	0.1	0.6	0.99
.	-	-	-	-
.	-	-	-	-
N-1	0.6	0.3	0.1	1
N	0.1	0.5	0.4	0.99

VADER-based sentiment analysis is performed in the proposed method according to Table 2, and finally, each of the positive, negative, neutral, and compound values is selected as a final feature. These features are examined in the feature selection step by feature selection methods in terms of effectiveness. If they were important, they would be considered in the final list of the feature selection.

4. Data collection

This section provides necessary information regarding the data used for analyzing the problem and the proposed models.

4.1 data set

In this part of the proposed method, four types of data, including information associated with news, tweets, Google Trends, and Bitcoin stocks, have been collected by different methods and through API, which is examined and presented in the following. The whole information was daily obtained from "05/02/2021" to "10/09/2021" for each one.

Bitcoin Information: The yfinance library is exploited to extract Bitcoin stock features, including open, close, high, low, volume and price. Bitcoin stock information was extracted daily from "05/02/2021" to "10/09/2021". Therefore, four Bitcoin features and a close feature are considered as real values for forecasting at this phase. Table 1 highlights an overview of this feature.

Tweet information: The Twitter API was employed to collect tweets extracted daily from "05/02/2021" to "10/09/2021". The collected data includes 1.2 million tweets related to the word BTC and Bitcoin. Finally, tweet information is grouped daily. In addition to the tweets' texts, the meta feature of the users is also collected at this step. Meta tweet information includes total followers, average followers, and so on, whose exact information is illustrated in Table 3.

Table 3. Different extraction features in the proposed method

No.	Group	Feature name	Method and final feature Count	Description
1	Bitcoin	Open	Direct➔1Feature extract	This feature is related to Bitcoin and has been extracted directly from the yfinance library.
2		High	Direct➔1 Feature extract	
3		Low	Direct➔1 Feature extract	
4		Volume	Direct➔1 Feature extract	
5	tweet	Text_tweet	VADER➔3 Feature extract	Tweet features are divided into textual and meta categories. Textual information is extracted from VADER based on sentiment analysis, negative, positive, and neutral features.
6		user_followers_sum	Sum➔1 Feature extract	
7		user_followers_mean	Mean➔1 Feature extract	
8		user_friend_sum	Sum ➔1 Feature extract	
9		user_friend_mean	Mean ➔1 Feature extract	
10		user_favourites_mean	Mean➔1 Feature extract	
11		user_verified_most	Most ➔1 Feature extract	
12		user_verified_mean	Mean➔1 Feature extract	
13	Google Trends	Bitcoin_rank	Count➔1 Feature extract	The amount of ranking is based on the two words "Bitcoin" and "BTC".
14		BTC_rank	Count➔1 Feature extract	
15	NEWS	Title news	TFIDF➔50:N Feature extract	Features are extracted from the headline and the news content based on the TFIDF method. In this feature extraction model,
16		Body news	TFIDF➔50:N Feature extract	

	at least 50 effective words are considered for the headline and the news content.

Google Trends Information: The ranking of the two words "Bitcoin" and "BTC" were extracted using the pay trends library at this step. The information of this step was also extracted daily from "05/02/2021" to "10/09/2021". More detailed information regarding these features is given in Table 3.

News headline and text information: In this step, the text and news related to Bitcoin were extracted from reputable sites like Coin telegraph using the Beautiful Soup and urllib libraries. Besides, each of the news headlines and text was extracted separately. In the next step, they were preprocessed, and then the TFIDF method was used to extract the effective features or words.

4.2 Text preprocessing and textual feature extraction

At this step of the proposed method, a series of preprocessing operations, including data clean, tokenization, stop word removal, and steaming, is applied to any tweet and textual news data. In natural language processing, algorithms do not have any understanding regarding the text; thus, the first and most important step is to identify or separate the words (signs and words), which is the task of the tokenization step (Jurafsky, 2000; Manning et al., 2014). The next step is to eliminate the stop words that are actually the repetitive words in the text without any information and are only used to connect the words in the sentence (Rani and Lobiyal, 2018). Stemming is the last step that needs to be performed in the preprocessing phase. In fact, the stem refers to the main meaning and concept of the word. Thus a limited number of stems are formed in natural language, and the rest of the words are extracted from these stems (Porter, 1980; Xu and Croft, 1998). Stem's major aim is to extract the stem and remove the affixer attached to the word (Manning and Schutze, 1999; Porter, 2001). Thus stemming is one of the main steps in natural language processing that must take place. Therefore, the steps of the word processing are given step by step below:

Data Cleaning Step: In this step of the proposed method, the blank textual data, numerical data, link address, and so on are eliminated from the textual news and tweets to prepare the text for the next steps of text processing.

Tokening step: In this step of the proposed method, unifying or tokenizing the sentences in each film is conducted.

Stop Word Removal step: In this step of the proposed method, stop word removal is conducted using nltk library and English Porter Stemmer.

Tokenization step: In this step of the proposed method, word stemming has been conducted using the nltk library and English Porter Stemmer.

After preparation, the Tweet data is sent to VADER for examining the sentiment analysis. Nevertheless, the textual news data in this paper is characterized by the TFIDF extraction method. F-IDF is known as a method to convert text to numerical values based on the importance of the words. This type of weighting is based on the belief that the words that distinguish a document from other headlines and news content are important words and thus have more weight (Salton and Buckley, 1988). According to Equation 7, in this type of weighting, the importance of words is measured based on the number of repetitions in the headlines and the news content and the whole documents in the content (data set).

$$TFIDF(t_i, d_j) = TF(t_i, d_j) \times IDF(t_i) \tag{7}$$

Where $TFIDF(t_i, d_j)$ examines the significance of the word based on the headline and the news content, and $IDF(t_i)$ calculates the significance of the word based on the headline and the news content including that word.

4.3 Data normalization

One of the crucial steps in preprocessing or preparing data sets in machine learning and deep learning algorithms is normalization and standardization methods. Normalization is conducted to scale the data values in a specific range of values. Most machine learning algorithms and deep data normalization more accurately predict the prices. The Min-Max normalization method is one of the scaling methods that are significantly popular and causes the data to be in the range between [0,1], which can be defined as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{8}$$

According to Equation 8, X_{min} represents the lowest value in a feature of the Bitcoin data, the value of each feature is denoted by X in Bitcoin data, and X_{max} indicates the maximum value in each feature of the Bitcoin data.

5. Proposed models based on deep learning

The first part of the proposed model revealed that various data, including meta-data tweets, sentiment tweet data (VS-data), news title-data, news content-data, Bitcoin data, and Google Trends data, have what type of features as shown in Table 3. Due to the various types of data collected in this paper. Nine different deep learning models have been designed according to the input data type in this part of the proposed method. Each model may have a different layer depending on the type of input data, such as the type of text. In addition, several models are designed based on the whole features and selecting the important features. This selection is based on the different features, including mutual-info-regression, Linear Regression, and correlation, and finally, a model is designed based on the combination of mutual-info-regression, Linear Regression, and correlation. Most of the models in this section are designed to indicate the impact of each data separately on the Bitcoin prediction. Through the combination of the whole existing features in the whole data, it is possible to specify how much these features are effective. Notably, in Section 6, the comparison results of the different criteria imply which models with which features have managed to predict Bitcoin more accurately.

5.1 The proposed Model-1 with Bitcoin data

In this model, a deep network based on convolutional layers and LSTM layers is designed with Bitcoin data input, as shown in Figure 1, which is called Model-1 in this paper. In this model, the Bitcoin stock data, including open, close, high, low, volume and price, is only considered to predict the Bitcoin price. The first model is composed of different layers, including three conv1-d layers, two max-pooling layers, a flatten layer, a dense layer, and an LSTM layer.

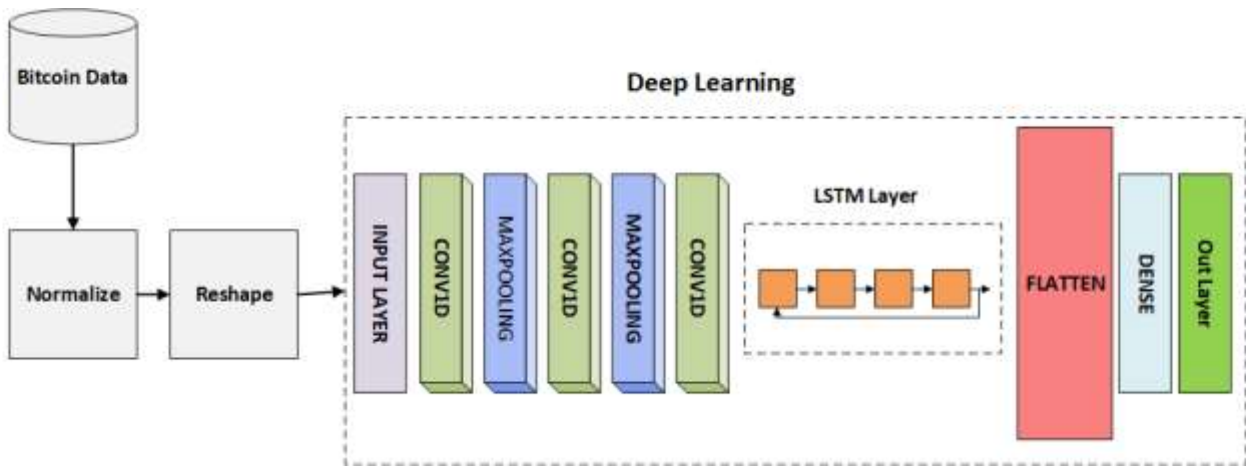

Figure 1. The proposed Model-1 architecture based on convolutional layer and LSTM layer with Bitcoin data

The proposed Model-1 architecture based on convolutional layers and Bitcoin data in Figure 2 shows that the convolutional layer is used to extract the better feature. Accordingly, the LSTM layer is utilized to maintain the temporal state of the data. Conv1-d with 500, 200, and 100 filters in this model, two max-pooling layers with two kernels, one LSTM layer with 32 units, and one dense layer with 20 units are set. Besides, the activation function is set with Relu except for the last layer, and the last layer is set according to the data type of the Sigmoid activation function. Notably, details of the loss function and the number of epochs and other hyper-parameters of Model-1 are given in Section 5.

5.2 The proposed Model-2 with Metadata

In this model, a deep network based on convolutional layers and Dense layers is presented with data input of metadata tweet. According to Figure 2, this model is called Model-2 in this paper. In this model, the metadata tweets include user-followers-sum, user-followers-mean, user-friend-sum, user-friend-mean, user-favorites-mean, user-verified-most, and user-verified-mean are considered to predict Bitcoin prices. The second model consists of different layers, including three conv1-d layers, two max-pooling layers, a flatten layer, a dense layer, and a dropout layer.

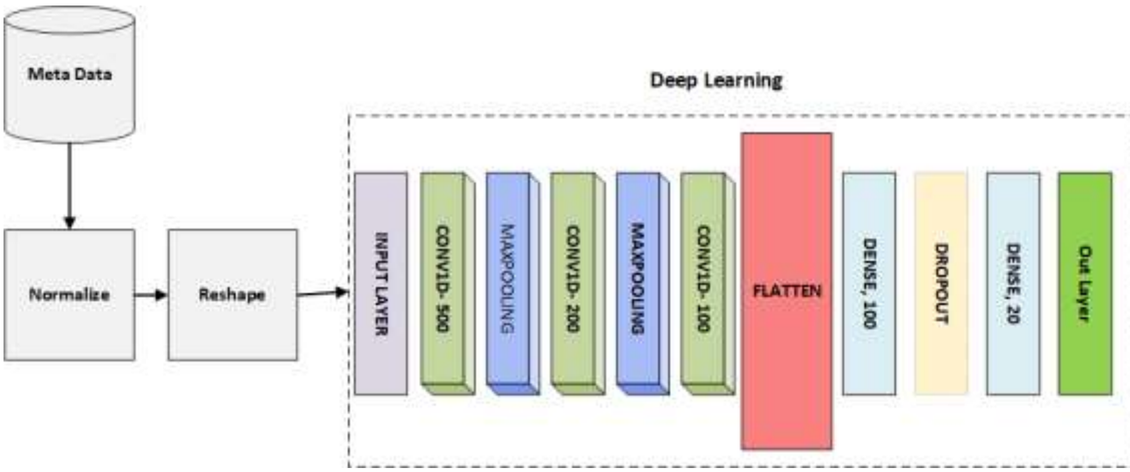

Figure 2. The proposed Model-2 architecture based on convolutional layers with Metadata

The proposed Model-2 architecture based on convolutional and metadata layers in Figure 2 shows that this model uses a convolutional layer to extract better features. Also, Dense and Dropout layer overfit problems are considered for better network training. Conv1-d with 500, 200, and 100 filters in this model, two max-pooling layers with two kernels, a dropout layer with 0.1%, and two Dense layers with 100 and 20 units are set. Also, in this model, the activation function is set with the Relu except for the last layer, and the last layer is set according to the data type of the Sigmoid activation function. In addition, more details regarding the loss function and the number of epochs and other hyper-parameters of Model-2 are presented in Section 6.

5.3 The proposed Model-3 with VADER data

In this model, a deep network is designed based on convolutional and Dense layers with data input of sentiment analysis of tweet text with Bitcoin data. As shown in Figure 3, this model is called Model-3 in this paper in which the data of the sentiment analysis of tweet text, including positive, negative, neutral, and compound values obtained from the VADER tool, are only used to predict Bitcoin prices. The third model consists of different layers, including three conv1-d layers, two max-pooling layers, a flatten layer, and a dense layer.

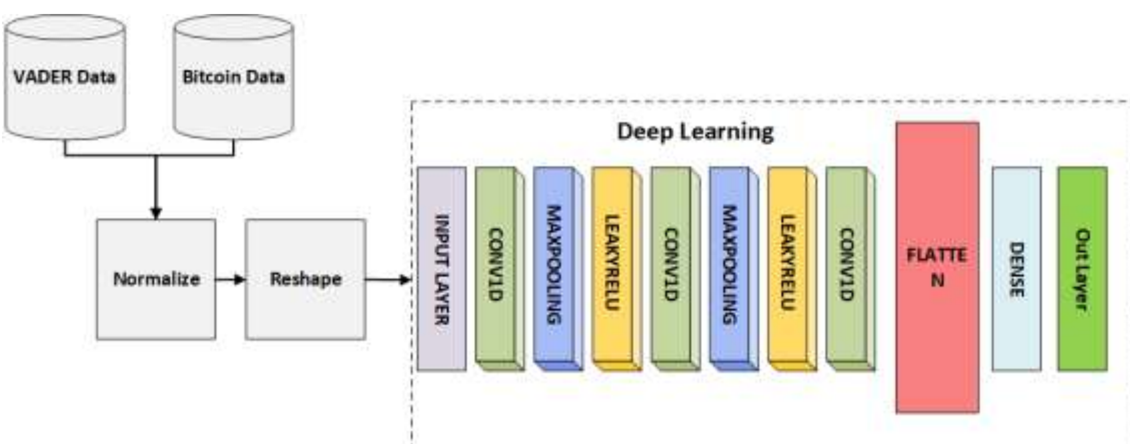

Figure 3. Proposed Model-3 architecture based on convolutional layers with sentiment analysis data of tweet text

The proposed Model-3 architecture based on convolutional layers and sentiment analysis data of tweet text and Bitcoin data in Figure 3 demonstrates that this model uses a convolutional layer to extract better features and uses the Dense layer for the linear state. Conv1-d with 500, 200, and 100 filters in this model, two max-pooling layers with two kernels, and a Dense layer with 100 units are set. Also, the last two-layer activation function is set with Relu, and the last layer is set according to the data type of the Sigmoid activation function. In this model, two leakyrelu activation functions are used after max-pooling layers. Notably, more details concerning the loss function and the number of epochs and other hyper-parameters of the Model-3 are presented in Section 6.

5.4 The proposed Model-4 with Meta + Bitcoin Data

This model presents a deep two-channel dense full-layer network with Bitcoin data input and metadata tweets. As shown in Figure 4, this model is named Model-4 in this paper. In contrast to other models, the proposed model has two input channels in which metadata tweets are used in the first channel, including user-followers-sum, user-followers-mean, user-friend-sum, user-friend-mean, user-favorites-mean, user-verified-most, user-verified-mean as input. Notably, the stock data, including open, close, up, down, volume, and price, are used in the second channel to predict the Bitcoin price. Finally, the two channels are combined by the concatenate layer.

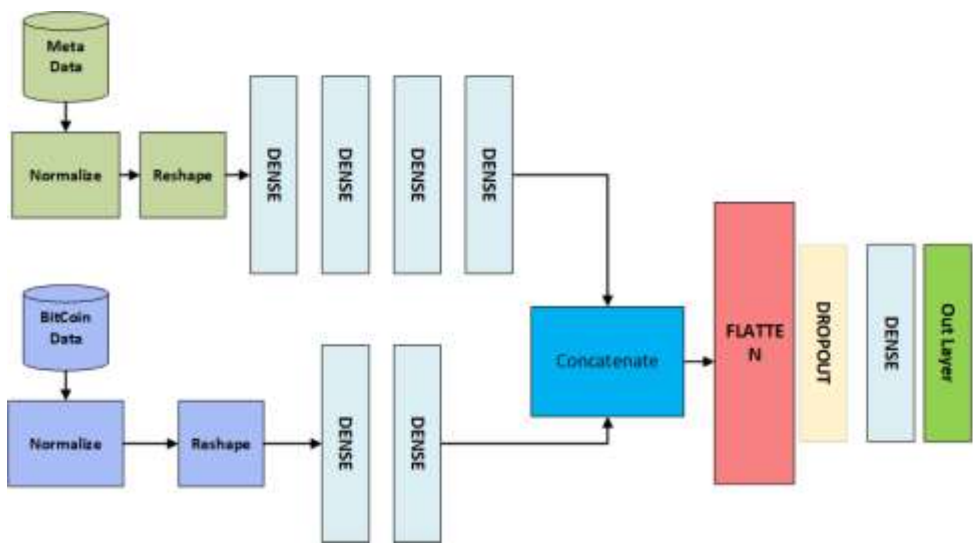

Figure 4. The proposed architecture of deep two-channel Model-4 with Meta + Bitcoin Data

The architecture of the proposed model-4 is based on a deep two-channel model with Bitcoin data and metadata tweets. Figure 4 indicates that this model has used the Dense layer and two-channel state to predict the Bitcoin price better. In this model, Dense layers with 500, 300, 200, and 100 units are set in the first channel. Dense layers with 200 and 100 units and a concatenate layer are set in the second channel. In addition, a concatenate layer is set from a dropout layer with 0.1%, and a Dense layer with 20 is used. Also, the activation function is set with Relu except for the last layer, and the last layer is set according to the data type of the Sigmoid activation function. Besides, more details about the loss function and the number of epochs and other hyper-parameters of the Model-4 are given in Section 6.

5.5 The proposed Model-5 with textual tweet data and Embedding layer

As regards Figure 5, in this model, a deep network based on convolutional layers with tweet textual data and an embedding layer is considered. This model is named Model-5 in this paper. In this model, the textual tweet data are directly used, and then the preprocessing operation with the Embedding input layer is considered. The fifth model consists of an embedding input layer from other layers, including three conv1-d layers, two max-pooling layers, a flatten layer, two dense layers, and a dropout layer. The major aim of this model is to use words and sentences directly in the text of the tweet to predict the Bitcoin price. In contrast to the third model, sentiment analysis is not considered.

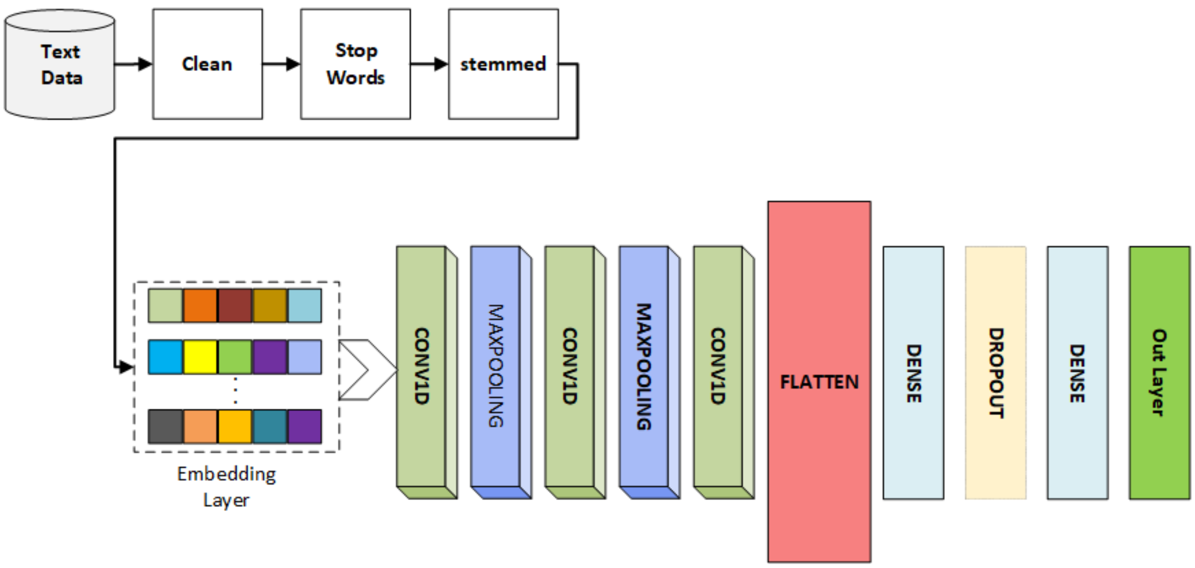

Figure 5. The model architecture proposed Model-5 with text tweet data with an Embedding layer

The proposed architecture of Model-5 based on convolutional layers with tweet textual data and Embedding layer is demonstrated in Figure 6, in which the convolutional layer and the Embedding layer are used to extract better features. Also, the Dense and Dropout layers are used to tackle the overfit problem and improve network learning. The embedding layer in this model with dimensions of 2000 * 500 is considered. After the Embedding layer, conv1-d with 500, 200, and 100 filters, two max-pooling layers with two kernels, one dropout layer with 0.1%, and two Dense layers with 100 and 20 units are set in this model. Furthermore, the activation function is set with Relu except for the last layer, and the last layer is set according to the data type of the Sigmoid activation function. Also, more details regarding the loss function and the number of epochs and other hyper-parameters of Model-2 are illustrated in Section 6.

5.6 The proposed Model-6, Model-7, and Model-8 with the whole data and three feature selection models

This section presents the three deep network models based on convolutional layers with the whole data and three feature selection models. As shown in Figure 6, these models are based on the mutual-info-regression, Linear Regression, and correlation feature selection methods, which are called Model-6, Model-7, and Model-8, respectively. Some important features are extracted from the whole data and given to the model based on learning data in these models. According to Table 1, the whole features are combined if the minimum number of text features and headlines is 100 features. These models contain at least 115 features, among which only 20% of the important features are selected by the feature selection method and given to the final model.

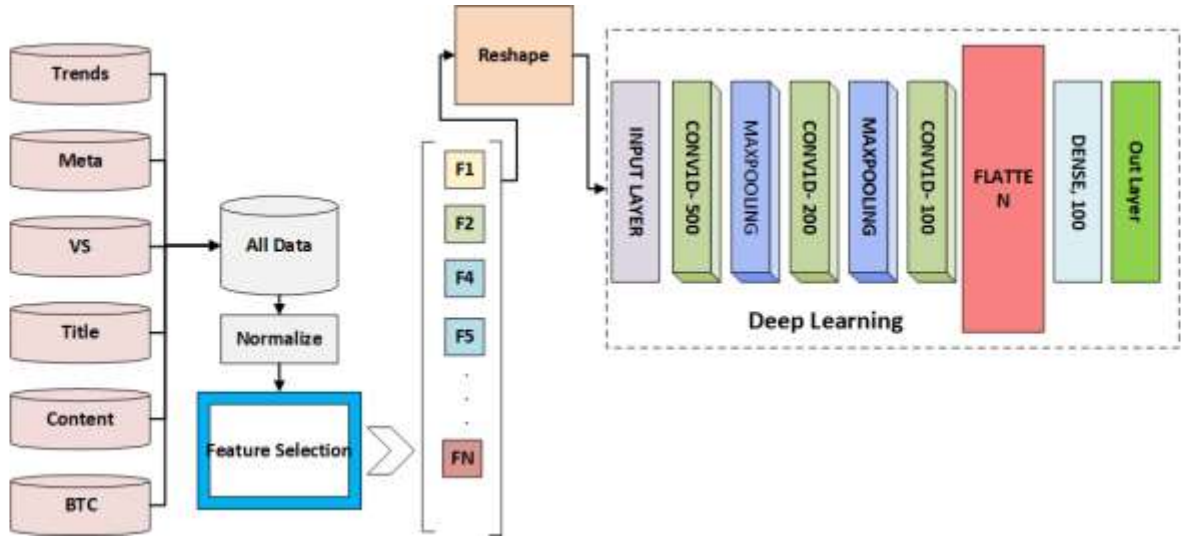

Figure 6. The proposed architecture of Model-6, Model-7, and Model-8 with the various data and three feature selection models

The proposed architecture of Model-6, Model-7, and Model-8 with different data and three feature selection models are considered in this section. With respect to Figure 6, the whole features are for better prediction, and also, a feature selection method is used in each of the proposed models to select important features and eliminate the extra features. Conv1-d with 500, 200, and 100 filters in this model, two max-pooling layers with two kernels, and a Dense layer with 100 units are set. Besides, the activation function is set with Relu except for the last layer, and the last layer is set according to the data type of the Sigmoid activation function. Also, the necessary details regarding the loss function and the number of epochs and other hyper-parameters of the Model-3 are illustrated in Section 6.

5.7 The proposed Model-9 with the various data and a combination of three feature selection models

A deep network model based on convolutional layers with the whole data and a combination of the three feature selection models are created in this model. As shown in Figure 7, this model is called Model-9 in this paper. Unlike models 6,7,8, the selected features are based on the combination of mutual-info-regression, Linear Regression, and correlation feature selection methods in this model. According to the learning data, each feature selection method selects 15% of the features. Then, the whole features of these three models are combined with each other and include about 45% of the total features. Nevertheless, since the repetitive feature may exist in the total selection feature, about 15 to 20% of the repetitive features are removed, and finally, about 20 to 25% of the important feature are selected by the combined feature selection method.

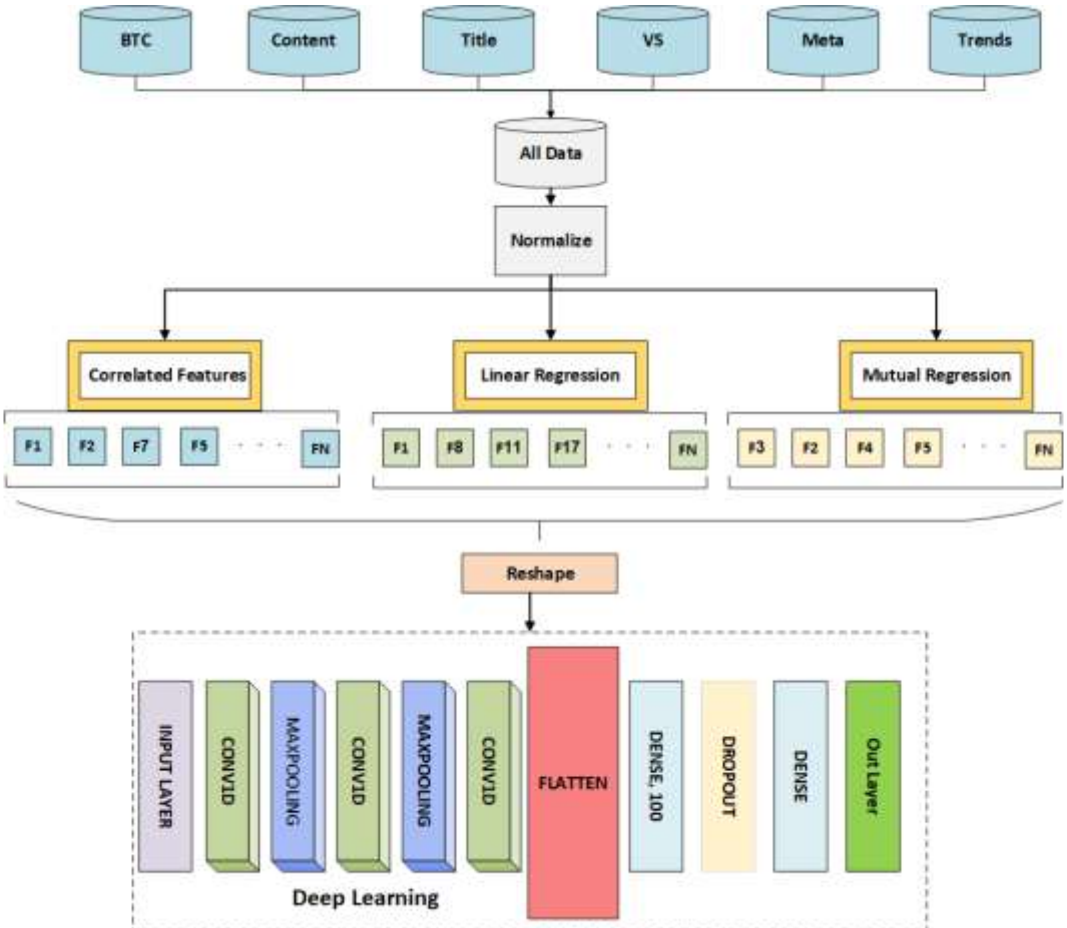

Figure 7. The proposed model architecture of Model-9 with the various data and a combination of three feature selection models

The proposed architecture of Model-9 with the various data and combinations of three feature selection models is presented here. Accordingly, the whole three feature selection methods are used in this model to predict all data more accurately and combine feature selection features. In this model, the advantages of three feature selection models are exploited to select important features and remove additional ones. Conv1-d with 500, 200, and 100 filters in this model, two max-pooling layers with two kernels, and a dense layer with 100 units are set. Notably, the activation function is set with Relu except the last layer, and the last layer is set according to the data type of the Sigmoid activation function. Moreover, the necessary details regarding the loss function and the number of epochs and other hyper-parameters of the Model-3 are presented in Section 6.

6. Evaluation and validation

In this section, the nine proposed models based on the convolutional neural network learning and LSTM are examined for predicting Bitcoin prices. The proposed model has been implemented and developed in the Google Columbine environment with 12 GB RAM and TensorFlow and keras libraries. TensorFlow library is one of the most widely used and popular neural network learning libraries in Python programming language that researchers and companies also exploit to create a variety of neural network architectures. In the whole experiment, the price of Bitcoin with a windows length of 1 was predicted due to the availability of the whole inputs for 78 days. Also, 80% of the data was considered for learning and 20% for the experiment. Some of the parameters of the proposed models were introduced in Section 3. Table 4 indicates the hyperparameters of each proposed model for implementation.

Table 4. Validation of hyperparameters of proposed models

Proposed model Name	hyperparameter
Model-1	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-2	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-3	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-4	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-5	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-6	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-7	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-8	epochs=100, batch_size=10, optimizer=Adam,loss=MSE
Model-9	epochs=100, batch_size=10, optimizer=Adam,loss=MSE

According to Table 4, in order to make a fair comparison between these values, the whole models are set with the same optimizer and loss. In this section, nine proposed models based on the convolutional neural networks learning and LSTM are compared for predicting Bitcoin price in terms of various criteria, including mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), median absolute error (MedAE) and determination coefficient (R^2). The first experiment for the loss function of the whole nine proposed models is based on the learning and test data shown in Figure 8 and Figure 9.

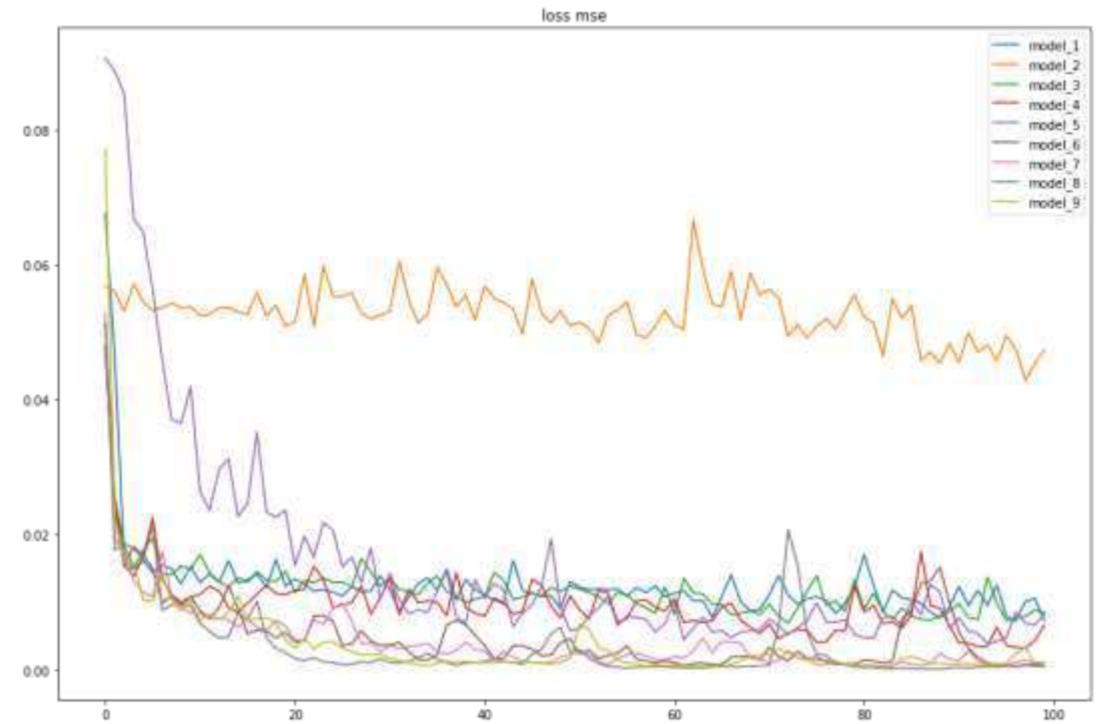

Figure 8. Comparison of proposed models in terms of loss function for the learning data

According to Figure 8, the comparison results of the proposed models in terms of loss function for the learning data show that the eighth and ninth models have better results than the other models, and also the second and third models have the worst performance in terms of the loss function. Notably, some models, such as the second model, are mainly considered to show the direct effect of words on the Bitcoin price. It should be noted that most algorithms can be optimized well on the learning data. Concerning the experimental data, the point is which model can have the best performance. In the following, the proposed models are compared in terms of loss function on the experimental data for more evaluation.

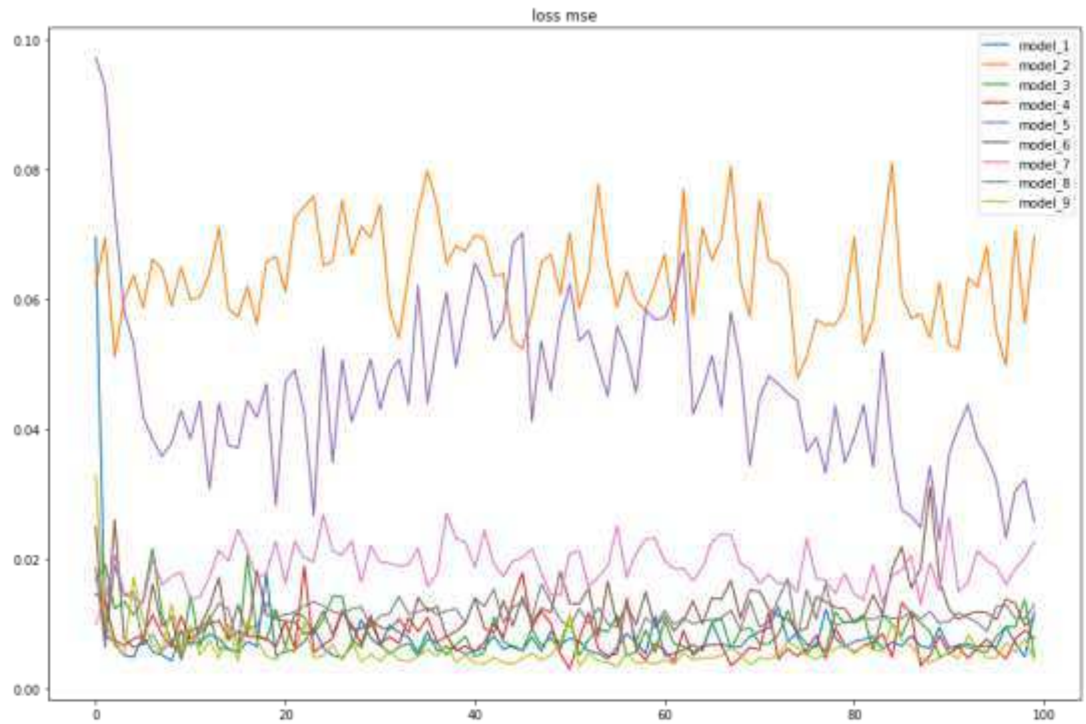

Figure 9. Comparison of proposed models in terms of loss function on the experimental data

As shown in Figure 9, the comparison results of the proposed models in terms of loss function on the experimental data highlight the fact that the eighth and ninth models performed better than the other models in the experimental data. However, some models like the fifth model were not able to have a better performance in the learning data since they only used the sentiment analysis. Besides, Models 2, 5, and 7 have acceptable performance in the experimental data.

It is worth mentioning that there is the direct interference of metadata tweets in predicting the Bitcoin price in Model 9 and the lack of interference of other features. Then, the second experiment is compared to the proposed models in terms of various criteria, including mean square error (MSE), root mean square error (RMSE), median absolute error (MedAE), and determination coefficient (R^2), as shown in Table 5.

Table 5. Comparison of different proposed models in terms of the various criteria

Model	MSE	RMSE	MAE	MedAE	R^2
Model-1	0.01029	0.10143	0.06812	0.05259	0.87864
Model-2	0.04925	0.22193	0.17161	0.12521	0.4190
Model-3	0.00698	0.08357	0.06006	0.04579	0.91762
Model-4	0.00362	0.06013	0.04170	0.03244	0.95735
Model-5	0.03420	0.185	0.1425	0.1145	0.5962
Model-6	0.00258	0.05082	0.02769	0.01285	0.96954
Model-7	0.01035	0.10176	0.07705	0.05846	0.87786
Model-8	0.00251	0.05013	0.03383	0.02759	0.9703
Model-9	**0.00151**	**0.0388**	**0.02519**	**0.01747**	**0.98219**

Table 5 gives the necessary information regarding the comparison made between the different proposed models in terms of different criteria. The ninth model with the value of 0.001 has obtained the best result in terms of MSE. Also, this model has a better performance compared to other models in the MAE and R^2 criteria with values of 0.02 and 0.98, respectively. Also, in terms of MSE criteria, the second model with the value of 0.04 has obtained the worst result. Compared to other models, this model has performed worse in terms of MAE and R^2 with values of 0.17 and 0.4190, respectively. Notably, based on the sixth, seventh and eighth models in these feature selection methods, it can be concluded that the two Model-6 and Model-8 have shown much better performance. Thus, the value of MSE for Model-6 is equal to 0.00258, and MSE value for Model-8 is equal to 0.00251. These results imply the fact that the feature selection of the two methods of mutual-info-regression and Linear Regression identified the important features correctly. Additionally, the fourth model in which VADER sentiment analysis is used has improved the whole criteria compared to the first model without using VADER sentiment analysis. Compared to the first model, the R^2 criterion has improved by 8%, and the results have been positive for the effectiveness of VADER sentiment analysis in predicting the Bitcoin price in the fourth model. In the third experiment, the values of the prediction diagram in each proposed model are graphically displayed for the whole Bitcoin data in Figures 10-18.

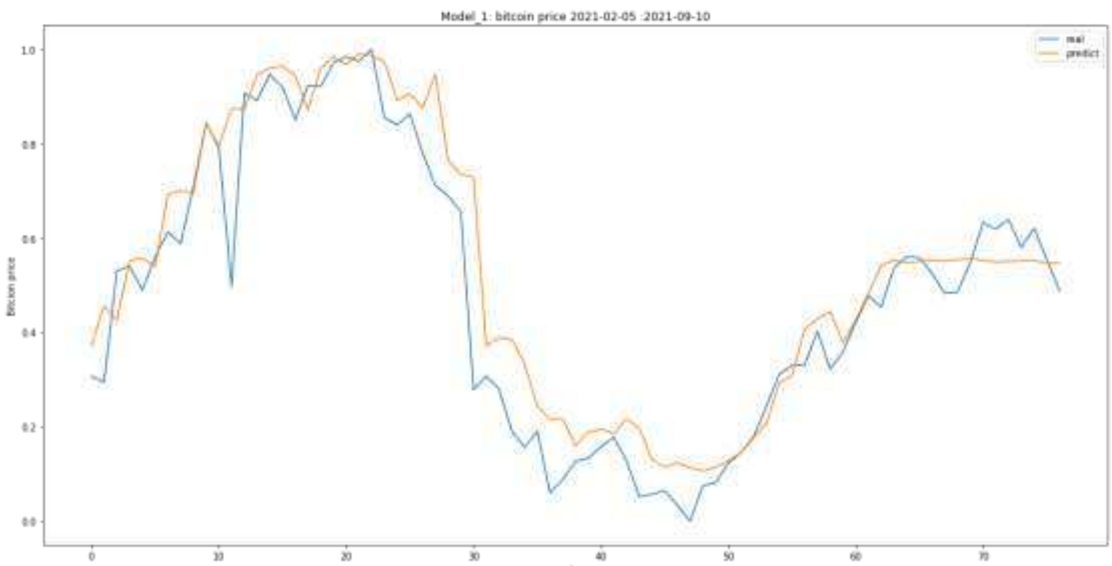

Figure 10. Bitcoin price prediction based on the first proposed model

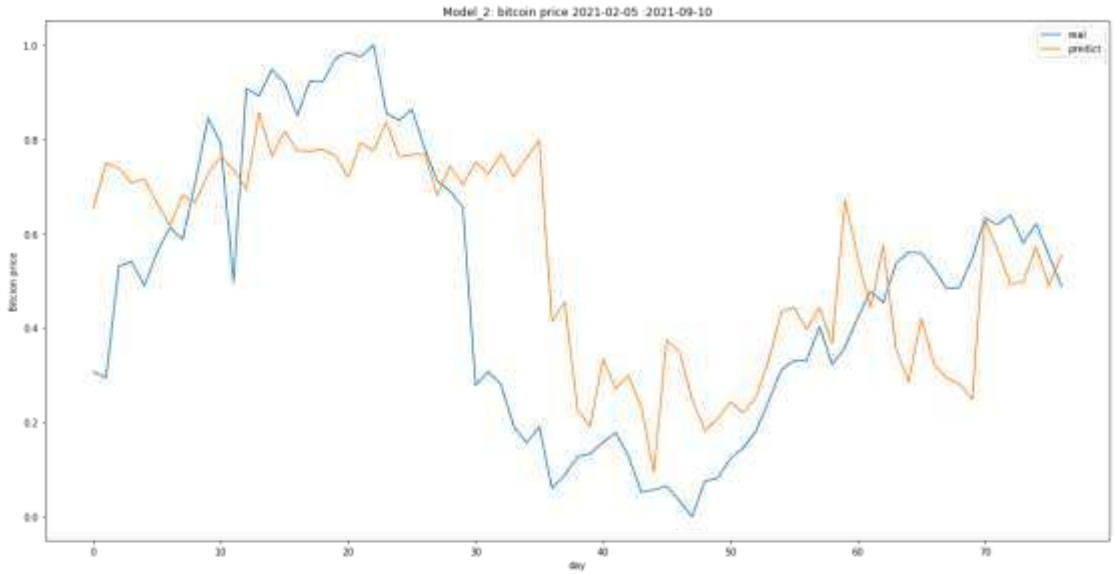

Figure 11. Bitcoin price prediction based on the second proposed model

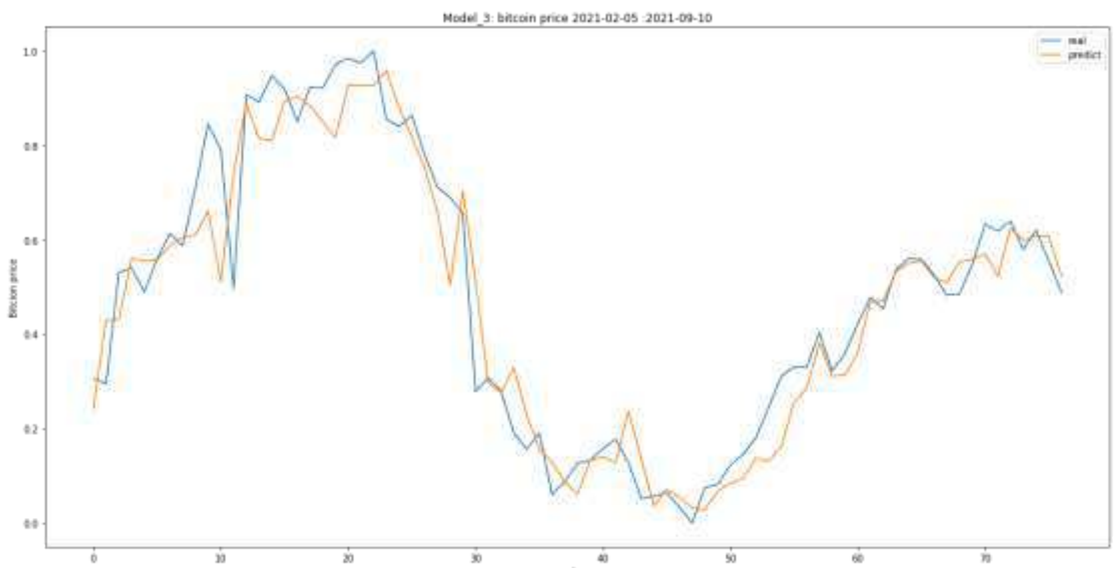

Figure 12. Bitcoin price prediction based on the third proposed model

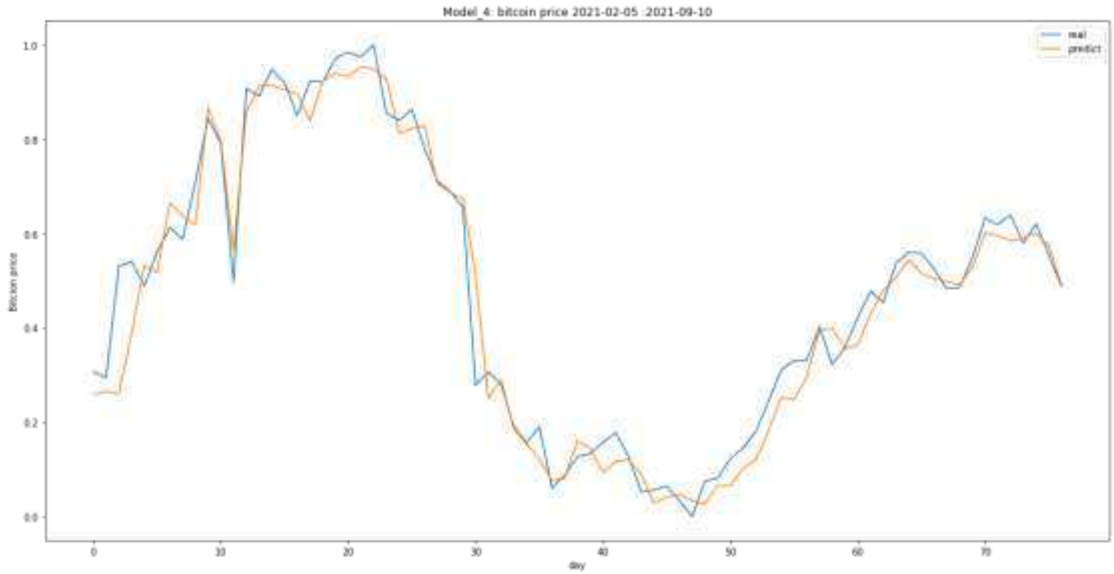

Figure 13. Bitcoin price prediction based on the fourth proposed model

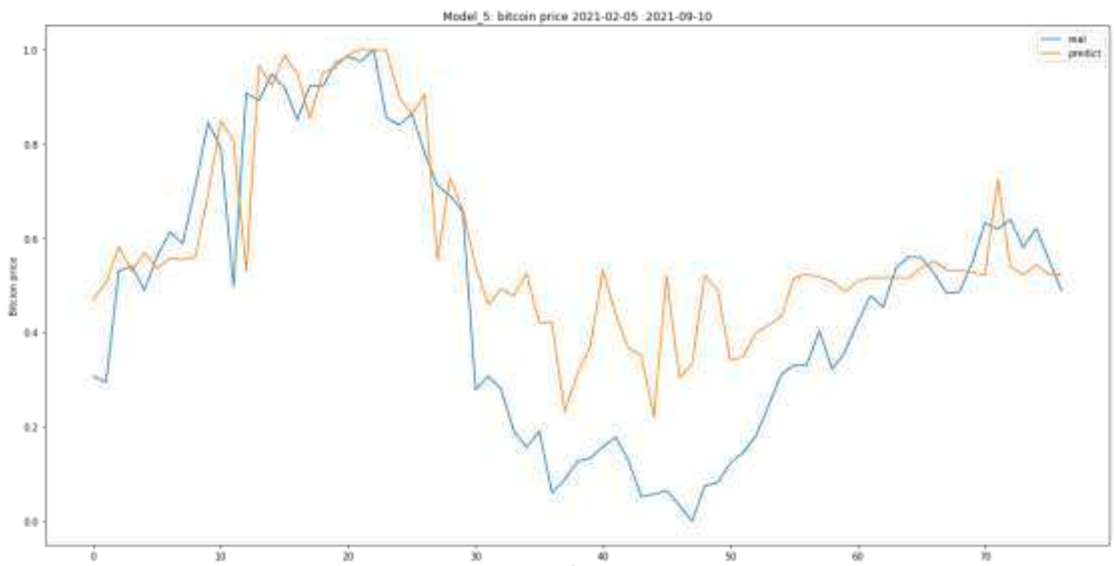

Figure 14. Bitcoin price prediction based on the fifth proposed model

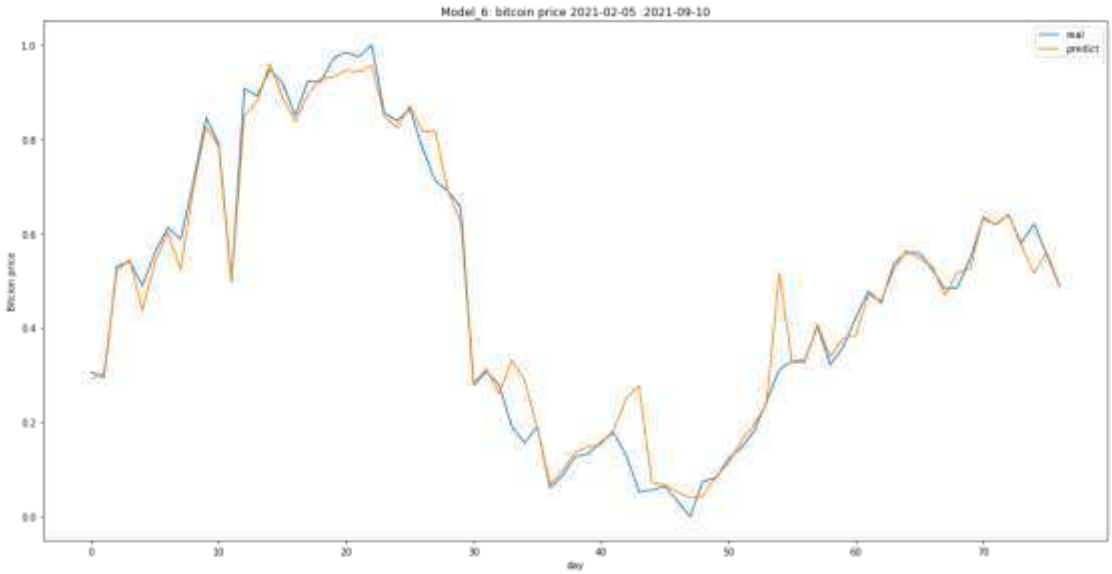

Figure 15. Bitcoin price prediction based on the sixth proposed model

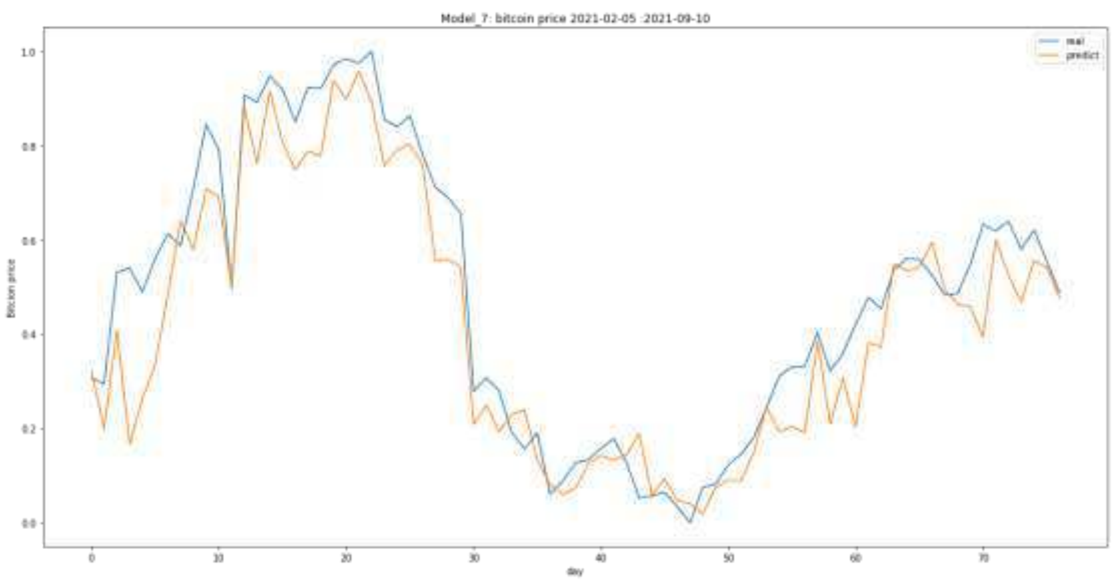

Figure 16. Bitcoin price prediction based on the proposed seventh model

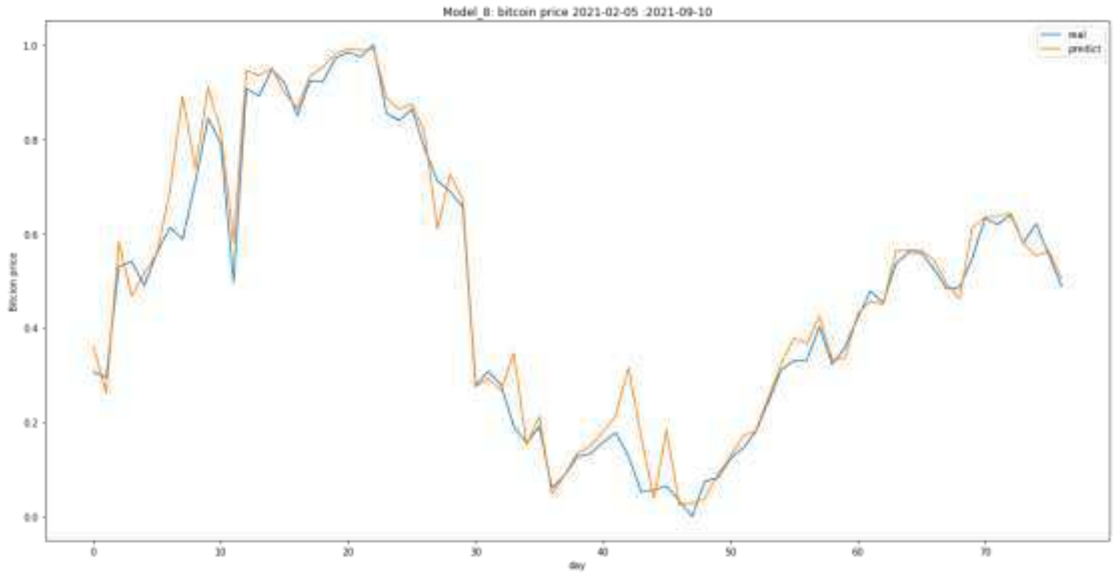

Figure 17. Bitcoin price prediction based on the proposed eighth model

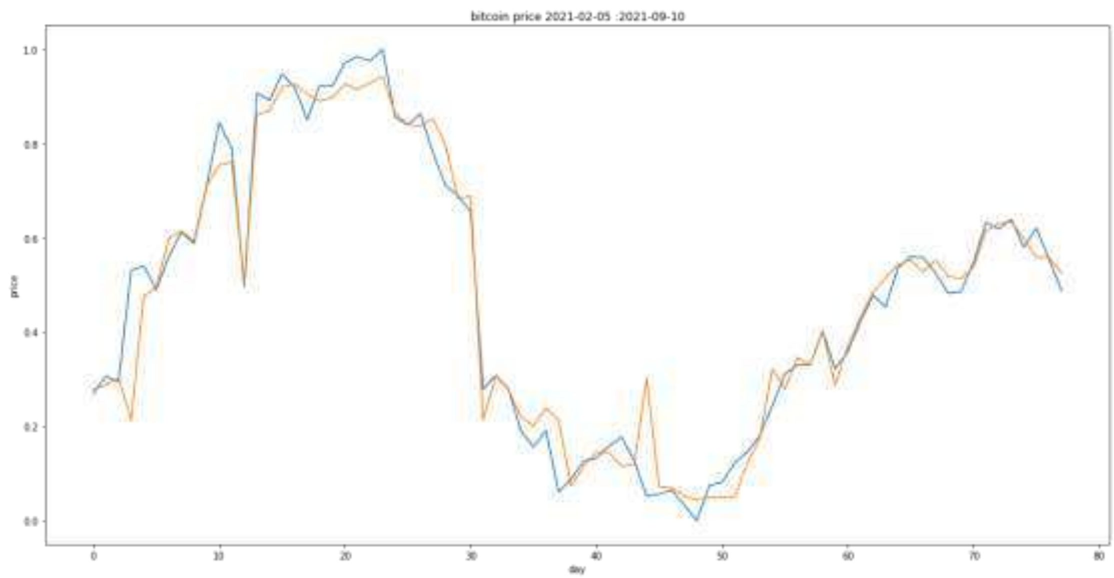

Figure 18. Bitcoin price prediction based on the proposed ninth model

The prediction diagram of the proposed models is presented on the whole Bitcoin data. This forecast was based on the loss function, and with respect to comparison results of the criteria in Table 5, it can be concluded that the ninth model has a better performance than other models in terms of accuracy of the price prediction. Notably, after that, the sixth and eighth models have a better performance in predicting the Bitcoin price.

7. Conclusion

In summary, the issue of Bitcoin price prediction using Deep Learning (DL) methods was considered in this research. This method is an advanced form of neural network algorithms that allows the extraction of low-level and high-level features from Bitcoin time data. Besides, deep learning methods can better consider the non-predictable nature of price. Several Bitcoin price prediction models based on CNN and LSTM were designed. Additionally, the sentiment analysis using the VADER tool and feature extraction of the Bitcoin news was employed in the proposed models. Twitter data analysis, news headlines, news content, Google Trends, Bitcoin stocks, and financials based on deep learning were considered in the proposed model to better and more accurately predict the Bitcoin price. Notably, due to the high extraction features of different input data, three methods of mutual information regression, Linear Regression, and correlation-based feature selection were exploited in this study. A combination of three feature selection methods was presented in a separated model to take advantage of such feature selection methods. Finally, the whole proposed models were compared with each other in terms of the performance criteria such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), median absolute error (MedAE), and coefficient of determination (R^2). The results of various implementations and experiments proved the remarkable performance of the proposed hybrid model based on sentiment analysis and combined feature selection with MSE value of 0.02 and R^2 value of 0.1 in obtaining better results and less error in predicting the Bitcoin price. Due to the input features, each model can be used as an individual assistant for more informed Bitcoin trading decisions depending on the input features. In future work, investigating the use of more data samples to experiment with the various proposed models might prove crucial. Further, combining deep learning models with robust machine learning algorithms can be considered an interesting topic for future study.

Acknowledgment

The authors appreciate the unknown referee's valuable and profound comments.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

ALWESHAH, M., ALKHALAILEH, S., ALBASHISH, D., MAFARJA, M., BSOUL, Q. and DORGHAM, O., 2021. A hybrid mine blast algorithm for feature selection problems. *Soft Computing*, **25,** 517-534.

ASUR, S. and HUBERMAN, B.A., 2010. Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 492-499.

AWOKE, T., ROUT, M., MOHANTY, L. and SATAPATHY, S.C., 2021. Bitcoin Price Prediction and Analysis Using Deep Learning Models, 631-640. Singapore.

BOLLEN, J., MAO, H. and ZENG, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, **2,** 1-8.

BORDINO, I., BATTISTON, S., CALDARELLI, G., CRISTELLI, M., UKKONEN, A. and WEBER, I., 2012. Web search queries can predict stock market volumes. *PloS one*, **7,** e40014.

BUI, D.-K., NGUYEN, T., CHOU, J.-S., NGUYEN-XUAN, H. and NGO, T.D., 2018. A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete. *Construction and Building Materials*, **180,** 320-333.

CAMBRIA, E., PORIA, S., HAZARIKA, D. and KWOK, K., 2018. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. *Proceedings of the AAAI conference on artificial intelligence*.

CHATFIELD, C. and YAR, M., 1988. Holt-Winters forecasting: some practical issues. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **37,** 129-140.

CHAUDHARI, H. and CRANE, M., 2020. Cross-correlation dynamics and community structures of cryptocurrencies. *Journal of Computational Science*, **44,** 101130.

CHEN, H., DE, P., HU, Y.J. and HWANG, B.-H., 2013. Customers as advisors: The role of social media in financial markets. In.

CHOHAN, U.W., 2017. Cryptocurrencies: A brief thematic review. *Available at SSRN 3024330*.

CHOI, H. and VARIAN, H., 2012. Predicting the present with Google Trends. *Economic record*, **88,** 2-9.

CHOU, J.-S. and BUI, D.-K., 2014. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings*, **82,** 437-446.

CHOU, J.-S., CHONG, W.K. and BUI, D.-K., 2016. Nature-inspired metaheuristic regression system: programming and implementation for civil engineering applications. *Journal of Computing in Civil Engineering*, **30,** 04016007.

COLIANNI, S., ROSALES, S. and SIGNOROTTI, M., 2015. Algorithmic trading of cryptocurrency based on Twitter sentiment analysis. *CS229 Project***,** 1-5.

DAI, B., JIANG, S., LI, C., ZHU, M. and WANG, S., 2021. A multi-hop cross-blockchain transaction model based on improved hash-locking. *International Journal of Computational Science and Engineering*, **24,** 610-620.

DAS, S., BILLAH, M. and MUMU, S.A., 2021. A Hybrid Approach for Predicting Bitcoin Price Using Bi-LSTM and Bi-RNN Based Neural Network, 223-233. Cham.

DE JONG, P., ELFAYOUMY, S. and SCHNUSENBERG, O., 2017. From returns to tweets and back: an investigation of the stocks in the Dow Jones Industrial Average. *Journal of Behavioral Finance*, **18,** 54-64.

DOLAN, R.J., 2002. Emotion, cognition, and behavior. *science*, **298,** 1191-1194.

DUTTA, A., KUMAR, S. and BASU, M., 2020. A Gated Recurrent Unit Approach to Bitcoin Price Prediction. *Journal of Risk and Financial Management*, **13,** 23.

ELRAHMAN, S.A. and ALLUHAIDAN, A.S., 2021. Blockchain technology and IoT-edge framework for sharing healthcare services. *Soft Computing*, **25,** 13753-13777.

ETTREDGE, M., GERDES, J. and KARUGA, G., 2005. Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, **48,** 87-92.

GURESEN, E., KAYAKUTLU, G. and DAIM, T.U., 2011. Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, **38,** 10389-10397.

HOTA, H.S., SHARMA, D.K. and VERMA, N., 2021. 14 - Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad. In: KOSE, U., GUPTA, D., DE ALBUQUERQUE, V.H.C. and KHANNA, A.s (Eds.): Data Science for COVID-19. 275-295.

HUTTO, C. and GILBERT, E., 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*.

JAIN, A., TRIPATHI, S., DWIVEDI, H.D. and SAXENA, P., 2018. Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis. *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 1-7.

JURAFSKY, D., 2000. Speech & language processing. Pearson Education India.

KAI-INEMAN, D. and TVERSKY, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, **47,** 363-391.

KARALEVICIUS, V., DEGRANDE, N. and DE WEERDT, J., 2018. Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance*.

KIMOTO, T., ASAKAWA, K., YODA, M. and TAKEOKA, M., 1990. Stock market prediction system with modular neural networks. *1990 IJCNN International Joint Conference on Neural Networks*, 1-6 vol.1.

KOULOUMPIS, E., WILSON, T. and MOORE, J., 2011. Twitter sentiment analysis: The good the bad and the omg! *Fifth International AAAI conference on weblogs and social media*.
KRASKOV, A., STöGBAUER, H. and GRASSBERGER, P., 2004. Estimating mutual information. *Physical review E*, **69,** 066138.
KRISTOUFEK, L., 2015. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one*, **10,** e0123923.
LAMON, C., NIELSEN, E. and REDONDO, E., 2017. Cryptocurrency price prediction using news and social media sentiment. *SMU Data Sci. Rev*, **1,** 1-22.
LI, D., HAN, D., WENG, T.-H., ZHENG, Z., LI, H., LIU, H., CASTIGLIONE, A. and LI, K.-C., 2021. Blockchain for federated learning toward secure distributed machine learning systems: a systemic survey. *Soft Computing*.
LIU, M., LI, G., LI, J., ZHU, X. and YAO, Y., 2021. Forecasting the price of Bitcoin using deep learning. *Finance Research Letters*, **40,** 101755.
MADAN, I., SALUJA, S. and ZHAO, A., 2015. Automated bitcoin trading via machine learning algorithms. *URL: http://cs229. stanford. edu/proj2014/Isaac% 20Madan*, **20**.
MANNING, C. and SCHUTZE, H., 1999. Foundations of statistical natural language processing. MIT press.
MANNING, C.D., SURDEANU, M., BAUER, J., FINKEL, J.R., BETHARD, S. and MCCLOSKY, D., 2014. The Stanford CoreNLP natural language processing toolkit. *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 55-60.
MATTA, M., LUNESU, I. and MARCHESI, M., 2015. Bitcoin Spread Prediction Using Social and Web Search Media. *UMAP workshops*, 1-10.
MCNALLY, S., ROCHE, J. and CATON, S., 2018. Predicting the Price of Bitcoin Using Machine Learning. *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, 339-343.
MITTAL, A., DHIMAN, V., SINGH, A. and PRAKASH, C., 2019. Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data. *2019 Twelfth International Conference on Contemporary Computing (IC3)*, 1-6.
NAIMY, V.Y. and HAYEK, M.R., 2018. Modelling and predicting the Bitcoin volatility using GARCH models. *International Journal of Mathematical Modelling and Numerical Optimisation*, **8,** 197-215.
NAKAMOTO, S., 2008. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review***,** 21260.
NAKANO, M., TAKAHASHI, A. and TAKAHASHI, S., 2018. Bitcoin technical trading with artificial neural network. *Physica A: Statistical Mechanics and its Applications*, **510,** 587-609.
O'CONNOR, B., BALASUBRAMANYAN, R., ROUTLEDGE, B.R. and SMITH, N.A., 2010. From tweets to polls: Linking text sentiment to public opinion time series. *Fourth international AAAI conference on weblogs and social media*.
P, S. and M, P.B., 2021. Diagnosis of lung cancer using hybrid deep neural network with adaptive sine cosine crow search algorithm. *Journal of Computational Science*, **53,** 101374.
PAK, A. and PAROUBEK, P., 2010. Twitter as a corpus for sentiment analysis and opinion mining. *LREc*, 1320-1326.
PANGER, G.T., 2017. Emotion in social media. University of California, Berkeley.
PANT, D.R., NEUPANE, P., POUDEL, A., POKHREL, A.K. and LAMA, B.K., 2018. Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, 128-132.
PETTEY, C., 2010. Gartner Says Majority of Consumers Rely on Social Networks to Guide Purchase Decisions. *Online im Internet: URL:< http://www. gartner. com/it/page. jsp*.
PORTER, M.F., 1980. An algorithm for suffix stripping. *Program*.
—, 2001. Snowball: A language for stemming algorithms. In.

RADITYO, A., MUNAJAT, Q. and BUDI, I., 2017. Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods. *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 433-438.

RAMADHAN, N.G., TANJUNG, N.A.F. and ADHINATA, F.D., 2021. Implementation of LSTM-RNN for Bitcoin Prediction. *Indonesia Journal on Computing (Indo-JC)*, **6,** 17-24.

RANI, R. and LOBIYAL, D.K., 2018. Automatic Construction of Generic Stop Words List for Hindi Text. *Procedia Computer Science*, **132,** 362-370.

ŞAHIN, D.Ö., KURAL, O.E., AKLEYLEK, S. and KıLıC, E., 2021. A novel permission-based Android malware detection system using feature selection based on linear regression. *Neural Computing and Applications*.

SALTON, G. and BUCKLEY, C., 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, **24,** 513-523.

SHAH, D. and ZHANG, K., 2014. Bayesian regression and Bitcoin. *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 409-414.

SONI, N., SHARMA, E.K. and KAPOOR, A., 2021. Hybrid meta-heuristic algorithm based deep neural network for face recognition. *Journal of Computational Science*, **51,** 101352.

STENQVIST, E. and LöNNö, J., 2017. Predicting Bitcoin price fluctuation with Twitter sentiment analysis. In.

SUL, H., DENNIS, A.R. and YUAN, L.I., 2014. Trading on Twitter: The Financial Information Content of Emotion in Social Media. *2014 47th Hawaii International Conference on System Sciences*, 806-815.

TETLOCK, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, **62,** 1139-1168.

VERGARA, J.R. and ESTEVEZ, P.A., 2014. A review of feature selection methods based on mutual information. *Neural Computing and Applications*, **24,** 175-186.

XU, J. and CROFT, W.B., 1998. Corpus-based stemming using cooccurrence of word variants. *ACM Transactions on Information Systems (TOIS)*, **16,** 61-81.

ZHU, X., WANG, H., XU, L. and LI, H., 2008. Predicting stock index increments by neural networks: The role of trading volume under different horizons. *Expert Systems with Applications*, **34,** 3043-3054.

ZUIDERWIJK, A., CHEN, Y.-C. and SALEM, F., 2021. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, **38,** 101577.