INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY



RESEARCH ARTICLE



GOPEN ACCESS

Received: 20.05.2021 **Accepted:** 15.07.2021 **Published:** 11.08.2021

Citation: Ho A, Vatambeti R, Ravichandran SK (2021) Bitcoin Price Prediction Using Machine Learning and Artificial Neural Network Model. Indian Journal of Science and Technology 14(27): 2300-2308. https://doi.org/10.17485/IJST/v14i27.878

*Corresponding author.

v2ramesh634@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2021 Ho et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN

Print: 0974-6846 Electronic: 0974-5645

Bitcoin Price Prediction Using Machine Learning and Artificial Neural Network Model

Alvin Ho¹, Ramesh Vatambeti^{1*}, Sathish Kumar Ravichandran¹

1 Computer Science and Engineering, CHRIST (Deemed to be University), Bengaluru, India

Abstract

Objective: This paper explains the working of the linear regression and Long Short-Term Memory model in predicting the value of a Bitcoin. Due to its raising popularity, Bitcoin has become like an investment and works on the Block chain technology which also gave raise to other crypto currency. This makes it very difficult to predict its value and hence with the help of Machine Learning Algorithm and Artificial Neural Network Model this predictor is tested. Methodology: In this study, we have used data sets for Bitcoin for testing and training the ML and AI model. With the help of python libraries, the data filtration process was done. Python has provided with a best feature for data analysis and visualization. After the understanding of the data, we trim the data and use the features or attributes best suited for the model. Implementation of the model is done and the result is recorded. Finding: It was discovered that the linear regression model's accuracy rate is very high when compared to other Machine Learning models from related works; it was found to be 99.87 percent accurate. The LSTM model, on the other hand, shows a mini error rate of 0.08 percent. This, in turn, demonstrates that the neural network model is more optimized than the machine learning model. Novelty: In this work, a small GUI has been created using the tkinter library that will allow the user to input the High, Low, and Open features values and then predict the next value for the coin. This paper compares the prediction outcomes of a machine learning model and an artificial neural network model. Because linear regression provided the highest accuracy compared to the other machine learning models, we used it to compare it to the LSTM model.

Keywords: Bitcoin; Block chain; Crypto currency; Machine Learning; Artificial Neural Network

1 Introduction

Bitcoin is a digital crypto currency that operates on an online decentralized network; it can be traded using an online peer-to-peer Bitcoin network that is not reliant on a central bank or a single administrator⁽¹⁾. Because it is accepted in over 40 countries worldwide (including Germany, Canada, and Croatia), the emergence of new alternative coins has resulted from its growing popularity. Bitcoin is also used to

exchange other crypto currencies, products, and services (2,3). Since the introduction of this crypto currency in the year 2009, no hacker has been able to infiltrate it due to block chain technology, where each electronic coin is encrypted with a unique digital signature which makes it easier to track and can be trusted. Each owner signs a digital hash from the previous transaction and adding the public key of the next owner before passing it on (4,5).

The price of Bitcoin in January 2017 was 1,000USD and by the end of December 2017, its value went up to 16000 USD and its value as on July 2021 is 32818 USD ^(6,7). We can say that the crypto market is very volatile, and among all the crypto currencies in the market, Bitcoin is experienced by most of the investors due to its anonymity and transparency in the system ^(8,9). This research aims to work on the prediction system for Bitcoin using various Machine learning algorithms and deep learning models to predict the price. There are various factors affecting the price of Bitcoin, in this project we will focus on open, close, high, and low factors.

The data set contains day transactions from 29th August 2017 to 9th August 2020. The data is first tested out with certain regression techniques and then a few deep learning models are implemented as deep learning tends to provide better accuracy compared to machine learning concepts when there is a high or more number of data sets. This paper also consists of 6 libraries.

- Pandas: provides an environment to python for creating creative and practical statistical computing for financial data analysis applications (10).
- Seaborn: effective visualization for better understanding the graphs and charts (11).
- Scikit-Learn: implementation of various algorithms. It contains a larege variety of supervise and unsupervised learning algorithms (12).
- Tkinter: To create a faster and quicker GUI application, with cooler features including the implementation of CSS support (13).
- Pickle: serialization and de-serialization of python object structure to store it in a file/database, maintaining program state, and transfer of data over a network (14,15).

In the rest of the research paper, Section 2, provides the work of other authors similar to this project, Section 3, includes the proposed idea and methodology, followed by Section 4 where we see a brief description of the data sets and the results, and finally, Section 5 will show the conclusion and future work.

2 Related Work

The prediction of crypto coins using the SVM and SVM-PSO method is suggested, where they used the day trading method to predict the values of ETH, BTC, XEM, XRP, XLM, LTC. SVM-PSO shows the optimized results. Performance accuracy of different Classifiers differs from coin to coin. However, this paper works only with a machine learning algorithm, and hence the data can be further improved by implementing the Deep Learning concept⁽¹⁶⁾.

The prediction of Bitcoin price using a transaction graph is proposed. The experiment consists of the Baseline, Logistic Regression, SVM, and Neural Network model with an accuracy of 53.4%, 54.3%, 53.7 and 55.1%. The feature selection in this paper is based on the Bitcoin block chain network which tends to be the least informative feature for the prediction of the Bitcoin price (17).

Predicting the crypto currency prices using sentimental analysis and Machine learning concepts like SVM and Random Forest on ETH, BTC, and XRP with BTC being the highest accuracy of 0.72. This accuracy rate is very low since machine learning algorithms were applied and it can be improved by testing with deep learning models (18).

A prediction model was proposed using four major algorithms, Gradient Boosted Tree, Neural Network, Ensemble Learning Method (with the best accuracy of 92.4%), and KNN model (With the least accuracy due to noisy random features and extreme volatility) (19).

The prediction system using Log regression, SVM, ANN, and random forest was proposed and shows that SVM has the best accuracy regarding a time-scale activity consisting of daily, 15-, 30- and 60- minutes return. Although SVM does tend to show better results out of the 4 algorithms, the prediction system can still show better results when Deep learning concepts are applied (20).

A linear regression model was used to predict the various cryptocurrency price using the open, low, and high cost. The experiment shows an accuracy of 99.3%. This paper does consist of a high accuracy rate but the data set used is comparatively small for a model to work on a real-time chart⁽³⁾.

3 Proposed Work

3.1 Problem Statement

Everybody wants to grow their money by investing in the stock market, but with the growing technology and the introduction of e-money what can be a better way to make your money grow by investing in crypto currency? (21) Bitcoin or any other crypto currency is not under the influence of any country or government, for this reason, it can be invested by anyone around the globe without the fear of being imposed of taxes from other countries. (7) The success of Bitcoin is measured by its huge capitalism growth and price, it leads to the emerging of various other crypto currencies which differ from Bitcoin in just a few parameters. (22)

One of the primary reasons for people to dive into the crypto market is that it's very easy and simple to buy and sell assets via trading platforms such as WazirX, Binance, etc. These platforms are very easy to use and it does not take much time to create an account and start trading. It's as simple as creating a Gmail account or a Facebook account, you just need to fill in your details and provide legal proof of yourself like an adhaar card or a PAN card. While stock market investment includes interacting with brokers, legal representatives, or agents, which then adds up to your expense, crypto trading is done via a peer-to-peer network which removes the problem of a third party.

Buying a fraction of crypto assets is what makes crypto trading flexible. You need not spend all your money to buy just a single asset. You can always purchase a fraction of it and invest the rest of your expense on other coins. Unlike stock market trading, the Crypto market functions 24/7 which allows businessmen and other retail investors who are usually busy working during the daytime to have an equal footing compared to all the other traders.

Compared to traditional exchanges, crypto exchanges are way faster as assets can be bought and sold in a minute wherein traditional exchange method we require a day or two for the process to take place. This is where blockchain technology comes into the picture. Traditional exchanges use convoluted technology which requires checking and rechecking, but with the introduction of blockchain, settling can be done instantly. This also helps you to keep a track of your order and you can see the depth of your trading of each asset conveniently.

3.2 Limitation

Although crypto trading has become a new trend, the increase in the number of digital coins and the adaptation of block chain technology causes the biggest concern i.e., scalability. It is still dwarfed by the number of transactions that, VISA, processes each day. Additional to that is the speed of transaction which the crypto market cannot compete with the players like VISA and MasterCard until the infrastructure delivering these technologies is massively scaled.

The crypto market is very volatile and can never be predicted at 100 percent accuracy. The market depends on human sentiment too; you may never know when a person owning at least 100 Bitcoin can suddenly sell his entire asset and create a big dip in the crypto market. We can never predict a human emotion even with the advanced technology we have in hand.

The analysis of any technical chart composes of mainly 3 major topics, the trend and momentum which indicate the direction and strength of direction, support, and resistance which indicates the potential stopping points of those directions, and the pattern in general, which indicates the information about the market psychology. Cryptocurrencies have not been around for long enough to provide sufficient information regarding the resistance and key support compared to the stock market, currencies, and commodities. This makes it difficult to predict and practice.

3.3 Methodology

Data collection: Data Collection is the first step we take in order to start any project. It is defined as the procedure of collecting, measuring, and analyzing accurate insights for research using standard validation techniques. An analyst would then be able to assess their theory dependent on gathered information. By and large, information assortment is the essential and most significant advance for research, independent of the field of examination. The methodology of information assortment is diverse for various fields of study, contingent upon the necessary data. The most important objective of data collection is ensuring that the gathered information is rich in content and reliable for statistical analysis so that data-driven decisions can be made efficiently and effectively. The data set contains day transactions from 29th August 2017 to 9th August 2020. The data is first tested out with certain regression techniques and then a deep learning model is implemented to provide better accuracy compared to machine learning concepts when there are high or more data sets.

		Date	Symbol	Op	en	High	Low	Close	Volume BTC	Volume USDT
0	2020-08-05 0	3-РМ	BTCUSDT	11617.	56 1169	93.94 1	1593.01	11678.72	2108.69	24576947.18
1	2020-08-05 0	2-PM	BTCUSDT	11609	99 116	44.65 1	1466.00	11617.56	5554.28	64363597.97
2	2020-08-05 0	1-PM	BTCUSDT	11562	86 116	20.00 1	1542.32	11609.99	4847.99	56151259.42
3	2020-08-05 12	2-PM	BTCUSDT	11438	06 115	84.60 1	1391.59	11562.86	5612.11	64457434.95
4	2020-08-05 1	I-AM	BTCUSDT	11393	24 1145	50.00 1	1382.21	11438.06	3069.55	35035195.65
,,,,	.tail()	D	ate Syr	nbol	Open	Hig	ph L	ow Clo	se Volume BTC	Volume USD1
260	23 2017-08-	17 08-	AM BTCU	SOT	4349.99	4377.0	35 4333	3.32 4360.0	69 0.9495	9 4139.70
260	24 2017-08-	17 07-	AM BTCU	SDT	4324.35	4349.1	99 4287	4349.5	99 4.4400	19241.06
			AM BTCU	SOT	4315.32	4345.4	45 4309	37 4324.3	35 7.2300	31282.31
260	26 2017-08-	17 00	Jem BICO	Dept.						
260 260					4308.83	4328.6				

Fig 1. Display of the Data collected

Feature Selection: Now that we have the required data for the project, we need to start the next procedure called data segregation or feature selection. This is a process where we trim out the unwanted data or we remove the unnecessary data from the data set. This step is necessary as we require only those features which can contribute to our prediction as unnecessary data can cause noise in our final output. To put it in simple words, we segregate data so that we can have a better model which provides us with an optimized result, reduce the property of over-fitting or redundancy and reduce the training time so that the system can generate output faster and with higher accuracy. In this project, I have implemented a few predefined python libraries which help in data visualization and can help you understand the important features which are required by the system. Data visualization is a technique where data or information is represented in a diagrammatic format for better understanding. Data visualization helps us to communicate with the relationships of data using the help of images. These images are in form of patterns that can be understood very easily. This is one of the main reasons how machine learning helps in analyzing data. Whether you work in the finance department or marketing or technical or design, you need to visualize data to understand it. This makes data visualization an important factor in today's world.

Features	Definition
Open	Opening value of trade at that time stamp
Close	Closing value of trade at that time stamp
High	Highest trade value in the time stamp
Low	Lowest trade value in the time stamp
Volume BTC	Total trade volume in BTC in the given timestamp
Volume USDT	Total trade volume in USDT in the given timestamp
Date	The given date and time of each bid
Symbol	Symbolic representation of coin

Fig 2. The features represented in the data

With the help of data visualization libraries, we can see the correlation between features and pinpoint the ones which we require. A sample image is shown below to show the correlation graph between the features in the given data set. You can notice in the given image Figure 3 that the relation between "Volume of USDT" and "Volume of BTC" with the other features is not clear, due to which they are segregated and removed from the training data set.

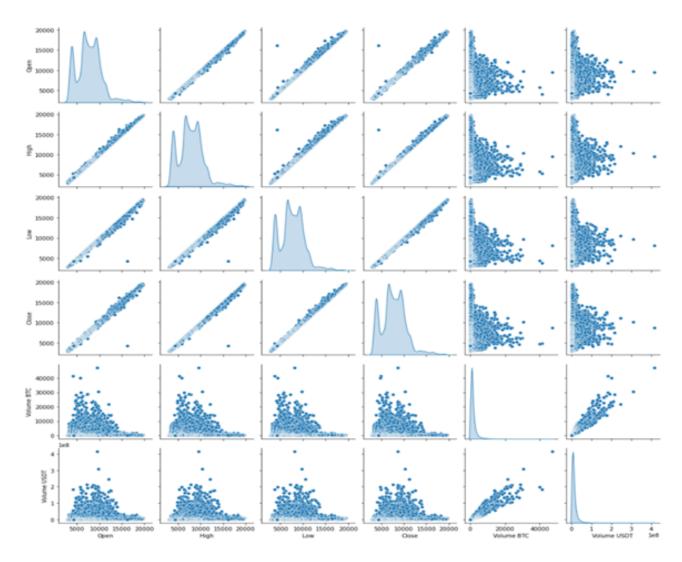


Fig 3. Correlation graph between the features.

Data Preparation: When variables that are measured in different scales it does not contribute equally in model fitting which will lead to model learned function to create a bias. Thus, standardization or normalization of data is very much essential for better accuracy and result. When working with a Machine Learning model or Deep Learning models where we require back propagation to be more stable and even faster (9), proper scaling of data is necessary.

$$x_{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

3.4 Algorithm Implemented

Linear Regression:

This technique is used to identify the relationship between dependent and independent variables and is leveraged to predict future outcomes. When we use only one dependent and one independent variable then it is called the simple linear regression. As the number of independent and dependent variable increase, it is then referred to as multi-linear regression. The graph is

plotted using a straight line across the graph which seeks to be the best fit by calculating the method of least square.

$$y = mx + C$$
 $C = y$ intercept
 $m = slope$
 x, y are the points on the graph
(2)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2$$

$$MSE = \text{mean squared error}$$

$$n = \text{number of data points}$$

$$Y_i = \text{predicted values}$$

$$\widehat{Y}_i = \text{predicted values}$$
(3)

Long Short-Term Memory (LSTM):

It is a deep learning concept or particularly a Recurrent Neural Network concept that avoids the vanishing gradient problem. The main reason for using this algorithm is that it avoids the back propagation error from vanishing or exploding, instead, these errors can flow backward through an unlimited number of virtual layers unfolded in space. LSTM mainly works on time series graphs with data sets that consist of events that occur thousands or millions of discrete-time steps earlier. It works with given long delays between significant events and can also handle signals with a mixture of low and high-frequency components. Over a lot of researchers have used LSTM to predict time series related data sets for stock prediction and have achieved greater or higher accuracy compared to other algorithms. LSTM is capable of recognizing context-sensitive language unlike any other previous models based on Hidden Markov Models (HMM) and other similar concepts.

The main formulation of the result in LSTM is based on Mean Absolute Error, the equation

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

$$y = prediction$$

$$x = true \ value$$

$$n = total \ number \ of \ data \ points$$
(4)

4 Result and Discussion

After the data analysis process we find that the only four features were well suited for the testing of this project. The data was trimmed and only the selected features were left as shown in Figure 4.

	Open	High	Low	Close
0	11617.56	11693.94	11593.01	11678.72
1	11609.99	11644.65	11466.00	11617.56
2	11562.86	11620.00	11542.32	11609.99
3	11438.06	11584.60	11391.59	11562.86
4	11393.24	11450.00	11382.21	11438.06

Fig 4. Attribute/Features selected are Open, High, Low, and Close

We can see the output of two models, one which is the Machine Learning model i.e. Linear regression, and the other one is the Recurrent Neural Network model i.e. Long Short-Term Model which shows us the two different outcomes. Linear

regression tends to work based on the Mean Squared Equation which tells us the accuracy of the linear graph with respect to the continuous-time frame data set. We see that the accuracy of the training data is approximately 99.97% and the accuracy of the testing data is tending to be approximately 99.97% as shown in Figure 5. Meanwhile, the LSTM model tends to find the accuracy with respect to the Mean Absolute Error which shows the error rate approximately to be 0.08% as shown in Figure 6.

```
In [38]: model.score(x_train, y_train)
Out[38]: 0.9997158887216999
In [39]: pred = model.predict(x_test)
In [40]: model.score(x_test, y_test)
Out[40]: 0.9997966090479169
```

Fig 5. Accuracy obtain from the training and testing data set using Linear regression model

S.No	Open	High	Low	Close	Expected Result
1	11617.56	11693.94	11593.01	11678.72	11669.05
2	11609.99	11644.65	11466.00	11617.56	11530.74
3	11562.86	11620.00	11542.32	11609.99	11603.13
4	11438.06	11584.60	11391.59	11562.86	11525.64
5	11393.24	11450.00	11382.21	11438.06	1140.47

Fig 6. Testing of Linear regression model

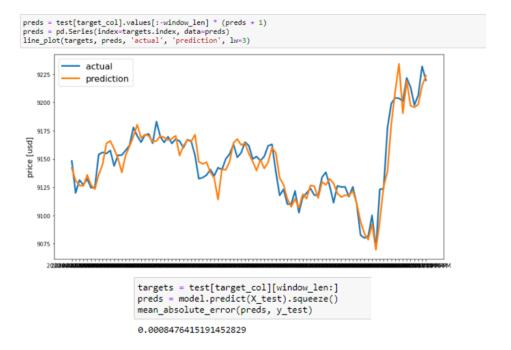


Fig 7. Final Resultant graph of LSTM and the Mean Absolute Error rate (0.08%)

Discussion: The Data visualization shows the correlation between all the features and only the four selected features have a sharp correlation. Data is then fitted into the model using the predefined commands accessible to python. These data models were trained and tested out with a limited number of data sets and provided the result. With the growing technology and the raise in the data sets we can still work on the model with various other alternative crypto currencies. The model shows a better prediction rate for LSTM but with a very slight difference compared to the linear regression model.

5 LIMITATIONS

Although crypto trading has become the new trend, the increase in number of digital coins and the adaptation of the block chain technology, causes the biggest concern i.e., scalability. It is still dwarfed by the number of transactions that, VISA, processes each day. Additional to that is the speed of transaction which the crypto market cannot compete with the players like VISA and MasterCard until the infrastructure delivering these technologies is massively scaled. The analysis of any technical chart composes of mainly 3 major topics, the trend and momentum which indicate the direction and strength of direction, support and resistance which indicates the potential stopping points of those directions and the pattern in general, which indicates the information about the market psychology. Cryptocurrencies have not been around for long enough to provide sufficient information regarding the resistance and key support compared to stock market, currencies and commodities. This makes it difficult to predict and practice.

6 CONCLUSION

The study reveals that the best accuracy rate is shown in Long Short-Term Memory than Linear Regression. This study is used to compare the features: open, close, high, and low only, hence the result may differ if we tend to take various other features into considerations. Because the crypto market is volatile and influenced by social media and other external factors, data sets cannot be the only reason for forecasting. As technology advances, new data can be collected, analyzed, and practiced, resulting in better results for this experiment.

Future Scope

- To work on a better User Interface so that people can access these data easily and effortlessly.
- Implementing IOT model for smart automatic analysis.
- Implementing more algorithms to find out the best method for predicting the crypto currency

References

- 1) Sin E, Wang L. Bitcoin price prediction using ensembles of neural networks. In: 2017 13th International conference on natural computation, fuzzy systems and knowledge discovery. IEEE. 2017;p. 666–671. doi:10.1109/FSKD.2017.8393351.
- 2) Shankhdhar A, Singh AK, Naugraiya S, Saini PK. Bitcoin Price Alert and Prediction System using various Models. *IOP Conference Series: Materials Science and Engineering*. 2021;1131(1):012009. Available from: https://dx.doi.org/10.1088/1757-899x/1131/1/012009.
- 3) Mittal R, Arora S, Bhatia MP. Automated cryptocurrencies prices prediction using machine learning. *ICTACT Journal on Soft Computing*. 2018;8(4):1758–1761. Available from: http://ictactjournals.in/paper/IJSC_Vol_8_Iss_4_Paper_8_1758_1761.pdf.
- 4) Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2019. Available from: https://git.dhimmel.com/bitcoin-whitepaper/.
- 5) Sebastião H, Godinho P. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation*. 2021;7(1):1–30. Available from: https://dx.doi.org/10.1186/s40854-020-00217-x.
- 6) Phaladisailoed T, Numnonda T. Machine learning models comparison for bitcoin price prediction. 10th International Conference on Information Technology and Electrical Engineering. 2018;p. 506–511. Available from: 10.1109/ICITEED.2018.8534911.
- 7) Jaquart P, Dann D, Weinhardt C. Short-term bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*. 2021;7:45–66. Available from: 10.1016/j.jfds.2021.03.001.
- 8) Rane PV, Dhage SN. Systematic erudition of bitcoin price prediction using machine learning techniques. 5th International Conference on Advanced Computing & Communication Systems (ICACCS). 2019;p. 594–598. Available from: 10.1109/ICACCS.2019.8728424.
- 9) Roy R, Roy S, Hossain MN, Alam MZ, Nazmul N. Study on nonlinear partial differential equation by implementing MSE method. *Global Scientific Journals*. 2020;8(1):1651–1665.
- Mckinney W. Pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing. 2011;14(9):1–9. Available from: https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf.
- 11) Waskom M. Seaborn: statistical data visualization. *Journal of Open Source Software*. 2021;6(60):3021. Available from: https://dx.doi.org/10.21105/joss.03021.
- 12) Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics. 2014;8:14. Available from: https://dx.doi.org/10.3389/fninf.2014.00014.
- 13) Polo G. PyGTK, PyQT, Tkinter and wxPython comparison. . *The Python Papers*. 2008;3:26–37. Available from: https://www.mclibre.org/descargar/docs/revistas/the-python-papers-3-1-en-200804.pdf.

- 14) Awoke T, Rout M, Mohanty L, Satapathy SC. Bitcoin price prediction and analysis using deep learning models. In: Communication Software and Networks 2021. Springer.; p. 631–640. doi:10.1007/978-981-15-5397-4_63.
- 15) Livieris IE, Stavroyiannis S, Pintelas E, Kotsilieris T, Pintelas P. A dropout weight-constrained recurrent neural network model for forecasting the price of major cryptocurrencies and CCi30 index. *Evolving Systems*. 2021;3:1–6. Available from: 10.1007/s12530-020-09361-2.
- 16) An optimized support vector machine (SVM) based on particle swarm optimization (PSO) for cryptocurrency forecasting. *Procedia Computer Science*. 2019;163:427–433. Available from: https://doi.org/10.1016/j.procs.2019.12.125.
- 17) Greaves A, Au B. Using the bitcoin transaction graph to predict the price of bitcoin. 2015. Available from: http://snap.stanford.edu/class/cs224w-2015/projects_2015/Using_the_Bitcoin_Transaction_Graph_to_Predict_the_Price_of_Bitcoin.pdf.
- 18) Valencia F, Gómez-Espinosa A, Valdés-Aguirre B. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. Entropy. 2019;21(6):589. Available from: https://dx.doi.org/10.3390/e21060589.
- 19) Chowdhury R, Rahman MA, Rahman MS, Mahdy MRC. An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. *Physica A: Statistical Mechanics and its Applications*. 2020;551(124569). Available from: https://dx.doi.org/10.1016/j.physa.2020.124569.
- 20) Akyildirim E, Goncu A, Sensoy A. Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research.* 2021;297(1-2):3–36. Available from: https://dx.doi.org/10.1007/s10479-020-03575-y.
- 21) Huang X, Zhang W, Huang Y, Tang X, Zhang M, Surbiryala J, et al.. 2021.
- 22) Muniye TA, Mohanty MR, Satapathy S. Bitcoin Price Prediction and Analysis Using Deep Learning Models. In: Communication Software and Networks. 2020. Available from: 10.1007/978-981-15-5397-4_63.