*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 10 Issue 9 September 2022

A Machine Learning Approach To Identify The Best Cryptocurrency For Investment

Maanav Mehta [(a)]

Grade 12,
Jamnabai Narsee S
Maanavmehta0@gmail.com

Reetu Jain [(b)]

Mentor, On My Own Technology Pvt. Ltd.
Mumbai - 400049, Maharashtra, India.
reetu.jain@onmyowntechnology.com

Abstract

Cryptocurrencies (CTC) are the most common form of digital assets mostly used for the investment purpose. However due to low amount of oversight and uncertainty of the future, CTCs are the least favored as an investment option by the investors. Due to these reasons, there is a need for developing a model that can not only forecast the price of the CTCs but can also determine the most stable, high return with least risk CTCs. For achieving the aims and objectives an Artificial Neural Network (ANN) and a Support Vector Machine (SVM) models are developed which are trained by time series data scraped from data.cryptocompare.com. The website stores the data for 100 different CTCs but for analysis the top five CTCs with the highest average market capitalization are considered. The ANN model showed training and testing accuracy of 0.9876 and 0.9198 respectively whereas the SVM model showed training and testing accuracy of 0.796 and 0.7981 respectively. The ANN and SVM model predicted Ethereum (ETH) and Dogecoin (DGC) as the best investment option for CTCs based on the data of 1st August 2022

Keywords: Cryptocurrency, Artificial neural network, Support vector machine, Stability, and price prediction.

1. Introduction

In the present-day scenario, cryptocurrencies (CTC) are the most common form of digital currencies. The CTC are circulated in the digital medium through a system of online networks and their valuation depends on the demand and their availability. In recent times, a high growth is seen in the demand of CTC as it is not regulated by any government i.e., it is a decentralized currency [1]. The newness concept of CTC is appealing for some investors while it is perceived oppositional by maximum of the investors [2]. The two major disadvantages for investing in CTC are the low amount of oversight and uncertainty of the future [3]. As a result, CTC is not yet the is not the first choice of the investor due to the market's erratic behavior and price volatility. This is where price prediction for cryptocurrency comes in and the need to have a price forecasting model.

In recent times, with the development and advancement made in the field of artificial intelligence (AI) and machine learning (ML), it is often applied to predict the price of cryptocurrencies [4]. Building AI and ML models in predicting the price of CTC will enable the investors to get a vivid picture of the behavior of the CTC. This will not only help them to understand the pattern of price variation but will also answer the question which CTC is the most stable and is associated with low level of risk

1.1. Motivation and novelties

From the literatures scrutinized for putting down the present study, certain gaps that are identified which are to be tackled in this research paper are as follows:

a. Although there is much literature to forecast the price of CTCs, there is limited literature that uses AI concepts to determine the most stable CTC.
b. The existing literature also fails to identify the CTC which has the maximum return with the least risk associated with it.

The overall aim of the present study is to develop a robust, efficient as well as effective AI model that not only forecasts the price but also determines the most stable, high return with least risk CTCs.

The remainder of this paper is organized as follows. Section 2 summarizes the contemporary literature which is followed by section 3 that describes the methodology adopted to solve the problem described in section 3. Section 4 summarizes the result obtained from solving the problem by the adopted methodology and finally section 5 concludes the study.

2. Review of the contemporary literatures

Many specific studies have been carried out on the research domain of price forecasting of crypto currencies, some of the most recent and significant papers are reviewed for writing this research paper. In the reference [5], took the help of a regression model and decision tree to identify the price trend on day-by-day changes in the CTC price while giving knowledge about CTC price trends with the aim of deriving the accuracy of CTC prediction using different machine learning algorithms and comparing their accuracy. Time-series forecasting model such as autoregressive moving average (ARMA), generalized autoregressive conditional heteroskedasticity (GARCH) and k-nearest neighbor (KNN) are employed in [6] to measure market liquidity as the long rates of bid-ask spreads in a sample of three CTCs namely Bitcoin (BTC), Ethereum (ETH), and Ripple (XRP) and 16 major fiat currencies over 1 year. In [7], ML model is employed to answer the question what price indicators can be used to predict the closing price of BTC on a given day and found that the high price has the biggest influence. In [8], the global price movement trend of CTCs are investigated with respect to the social media communication data by analyzing the topical trends in the online communities and social media platforms to understand and extract insights that could be used to predict the price fluctuations in crypto-currencies. In the reference [9], proposed a deep-learning-based hybrid model that includes Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) to predict the price of Litecoin (LTE) and Zcash with inter-dependency of the parent coin. A sentiment analysis integrated LSTM model is developed in [10] to predict the volatile price movement of CTCs. Two DL models were developed in [11] to forecast the daily price of BTC. The first model is a conventional LSTM model, and the second model is LSTM integrated with the Autoregressive (AR) model and the outputs are statistically scrutinized to identify the robust model. In [12], leveraged the accurate forecast of BTC prices via the normalization of a particular dataset with the use of LSTM machine learning. In [13], it proved that GRU performed better in predicting three types of cryptocurrencies, namely BTC, LTC, and ETH, then the LSTM and bidirectional LSTM models. In [14], developed a multiple-input LSTM model for the prediction of cryptocurrency BTC, ETH, and XRP price and movement. A Markovian model is proposed in [15], to describe BTC historical movements and LSTM model to predict future movements. Another multiple input LSTM-based prediction model is proposed in [16], which is integrated with the Black-Scholes model to address the challenges in predicting the prices of CTCs. A comparison of the results obtained from ARIMA model and seq2seq recurrent deep multi-layer neural network is conducted to identify the model that could return the best result in predicting CTC prices [17]. ARIMA model was employed to forecast the price of BTC by analyzing the price time series in a 3-years-long period [18]. Also, ARIMA model is employed to create a highly accurate short-term prediction model, to predict the price of BTC several days ahead [19]. In [20] applied the ARIMA model of [19] to predict the price of CTCs using the data from April 28th, 2013 to July 31 st , 2017. In [21] ARIMA and ML model was further applied to predict the closing price of BTC the next day. In the paper, a comparison is made between the two models. In [22], the predicted closing price of CTC is compared for the different ML and DL models with the actual value. In

[23], the results obtained from ARIMA Time Series Model and the LSTM DL algorithm to estimate the future price of BTC.

3. Materials and Methods

In this study, two different predictive models namely Artificial Neural Network (ANN) and Support Vector Machine (SVM) were employed to obtain the aims and objectives of the present study. The two models are discussed in brief in this section of the paper. Starting the section by giving a brief background of the dataset used for the study.

3.1. Dataset

The dataset has been extracted from https://data.cryptocompare.com from 1st March 2021 till 25th July 2022. The website comprises price data for 100 CTCs, out of which the most common CTCs are considered to do the price prediction. After the extraction of the dataset, the next step involves the removal of the attributes that are not used in the prediction and simulation process. After dropping the unused attributes from the dataset, the timestamps are converted into the dd-mm-yyyy format

3.2. Artificial Neural Network (ANN)

ANN is a DL model that is inspired from the central nervous system of animals. ANN is applied to solve the problems that showed high levels of nonlinearity and uncertainty [24]. ANN comprises several layers namely input layer, hidden layer and output layer. Each layer comprises artificial neurons that exchange information between neurons of different layers. The number of neurons in the input layer is equal to the number of input parameters and the number of neurons in the output layer is equal to the number of outputs. However, the number of hidden layers and neurons in the hidden layer depends on two factors, that are accuracy and swift prediction. The performance of the output from the ANN is measured based on the error values computed between the target and actual output from the network and the execution time.

An artificial neuron comprises two parts: the summing factor and the transfer factor. The summing factor sums the weighted inputs to the neuron which is then passed through a transfer factor which is a linear or nonlinear function. A pictorial representation of a typical artificial neuron is shown in figure 1.

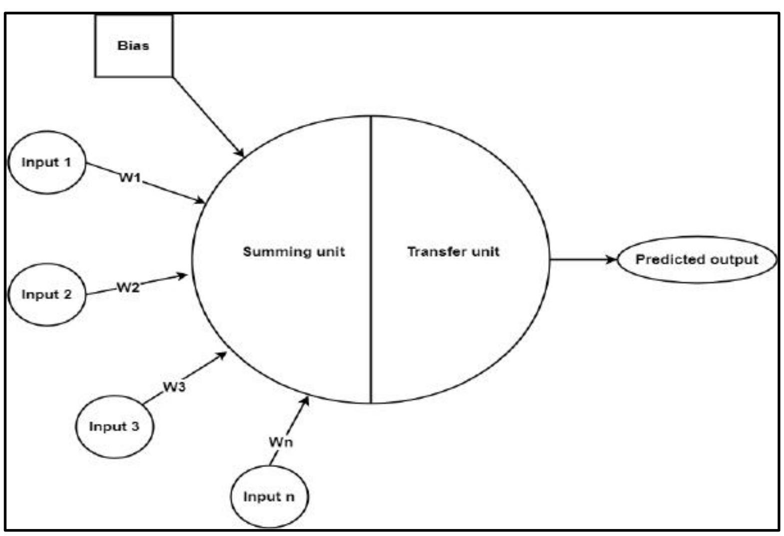


Figure 1: Pictorial representation of a typical artificial neuron

In the figure 1, the x are the input which is an output from a neuron of the predecessor layer having weight w and b is the bias value that guarantees even when all the inputs are zeros there will still be an activation in the neuron. A typical representation of the ANN network is shown in figure 2.

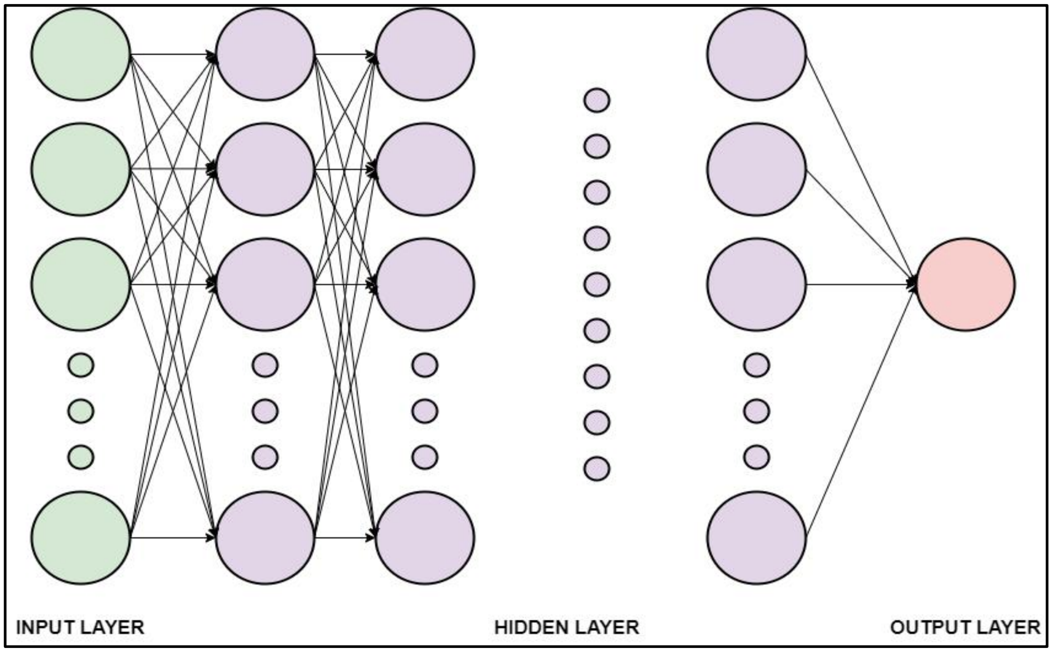


Figure 2: Pictorial representation of a typical ANN network

ANN has two phases (a) training and (b) learning phase. In the training phase the weights of the neurons are computed and in the learning phase the weights are updated. During the training phase, the output from the ANN is computed by moving in the forward direction and the performance is evaluated by the computation of the error. Then the error is propagated in the backward direction updating the interconnect weights of the neuron.

3.3. Support Vector Machines (SVM)

SVM is a member of the supervised ML family primarily used for learning the pattern for a non-linear dataset [26]. SVM which is a non-parametric technique is used for both regression and classifications [27]. The SVM model takes the help of some mathematical functions called kernels to transform the inputs into the desired form. The regression problems where SVM is employed uses linear function and those problems with non-linear regression map the input vector to n-dimensional space called a feature space. Considering the ML context for a multivariate training dataset (x_n) with with N number of observations having observed response of y_n. The mathematical equation for the linear function is given in Eq. (1).

$$f(x) = \beta.x + b \tag{1}$$

The objective of the SVM is to find the value of $f(x)$ and $\beta'\beta$ to fit the minimization function as per the Eq. (2).

$$g(\beta) = \frac{1}{2}\beta'\beta \tag{2}$$

with a special condition of the values of all residuals not more than ε, as in the Eq. (3).

$$\forall_n: |y_n - (\beta.x_n + b)| \tag{3}$$

The flowchart of the analysis of the research study is shown in figure 3.

4. Results and Discussions

As per the flowchart for the analysis shown in figure 3, the price prediction of the five CTCs with the highest average market capitalization shall be conducted. To find the CTCs with the highest average market capitalization is computed as follows:

$$mar_{ave} = \frac{\sum_{i}^{N} \ (mar\_cap)}{N} \tag{1}$$

The mar_cap in Eq. (1) denotes market capitalization and N is the total number of CTC units of a particular type traded from 1st March 2021 till 25th July 2022. The mar_{ave} is the average market capitalization for a particular type of CTC. The mar_{ave} value computed for each CTC kind is shown in the form of a bar graph in figure 4.

From the bar graph in figure 4, it is observed that the Bitcoin (BTC), Ethereum (ETH), Tether (TET), Litecoin (LTC) and Dogecoin (DGC). Therefore, these five CTCs are considered for price prediction using SVM and ANN. In the next step of the analysis, the number of the features are reduced using the Principal Component Analysis (PCA) technique. It is a statistical procedure that allows summarizing the information content in large data tables by means of a smaller set often termed as summary indices that can be more easily visualized and analyzed. With the help of the PCA algorithm, the number of features is reduced to 3.

In the next step of the analysis, the data are fed in the developed ANN and SVM model. The code for the ANN and SVM is done in Python 3.10 and runs on a Windows computer with 8 Gb RAM and intel i5 processor.

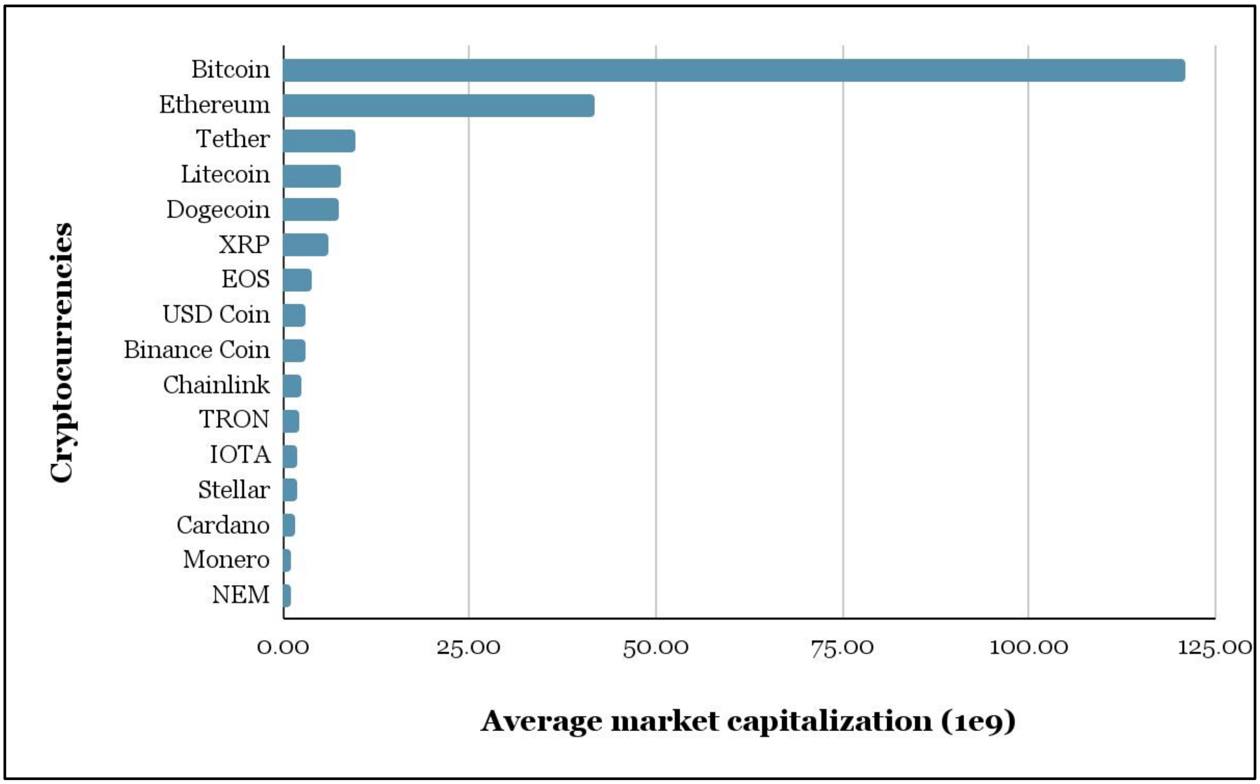


Figure 4: Bar graph showing the average market capitalization for the different CTCs

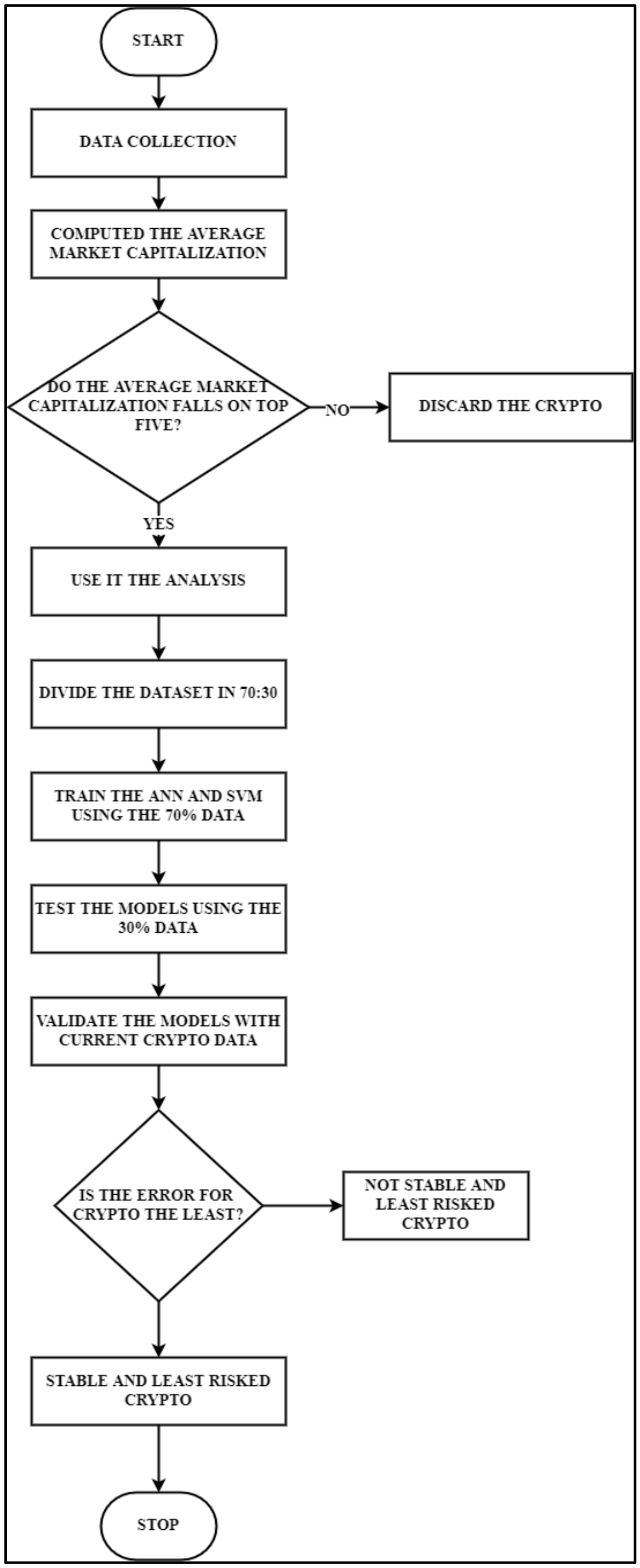
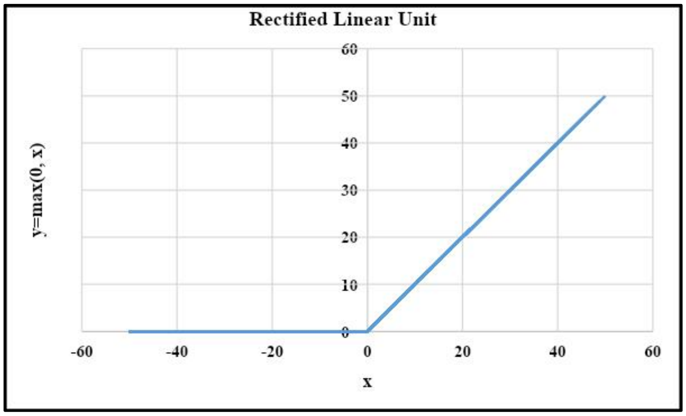


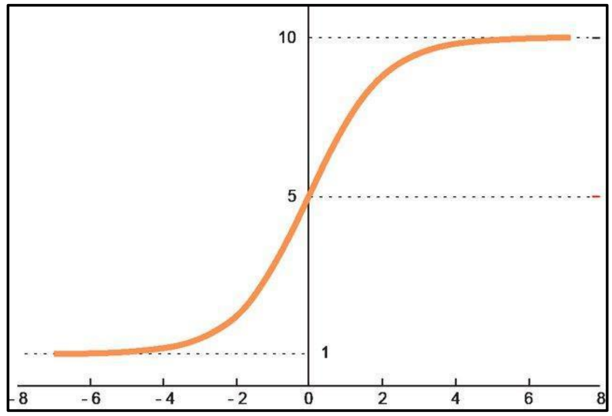
Figure 3: Flowchart of the analysis

.4.1. Result from ANN

The architecture of the ANN developed for predicting the price of the CTCs involves 2 hidden layers with 32 and 16 hidden neurons in the first and second hidden layer respectively. The rectified linear unit is employed as the transfer function for the input and the hidden layer whereas SOFTMAX transfer function is used for transferring the output layer. The figures for the rectified linear unit and SOFTMAX transfer function are shown in figure 5(i) and 5(ii).



(i)



(ii)

Figure 5: Pictorial representation of the (i) rectified linear unit and (ii) SOFTMAX transfer function

The ANN model is run for 50 epochs with 40 steps per epoch and 1 validation step. The loss and accuracy curve for the ANN model is shown in figure 6 and the validation loss and validation accuracy is shown in figure 7

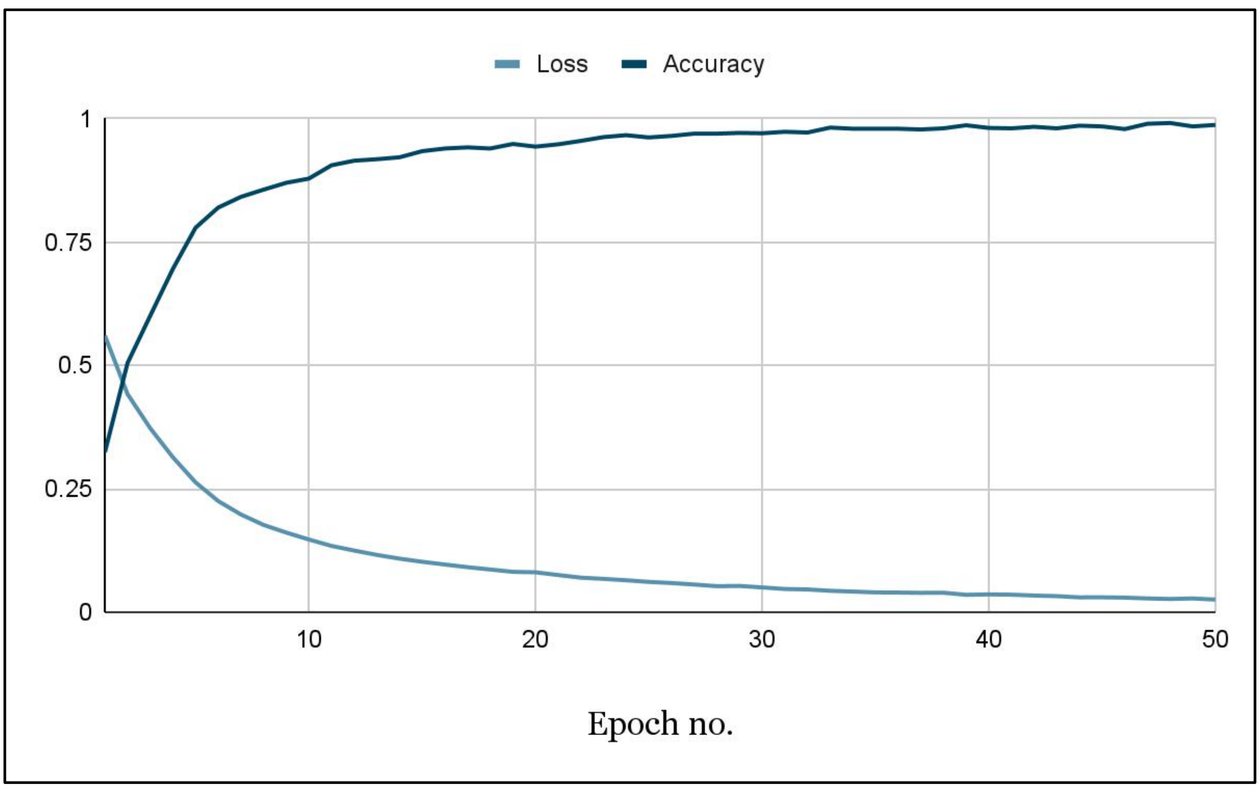


Figure 6: Loss and Accuracy curve for the ANN model

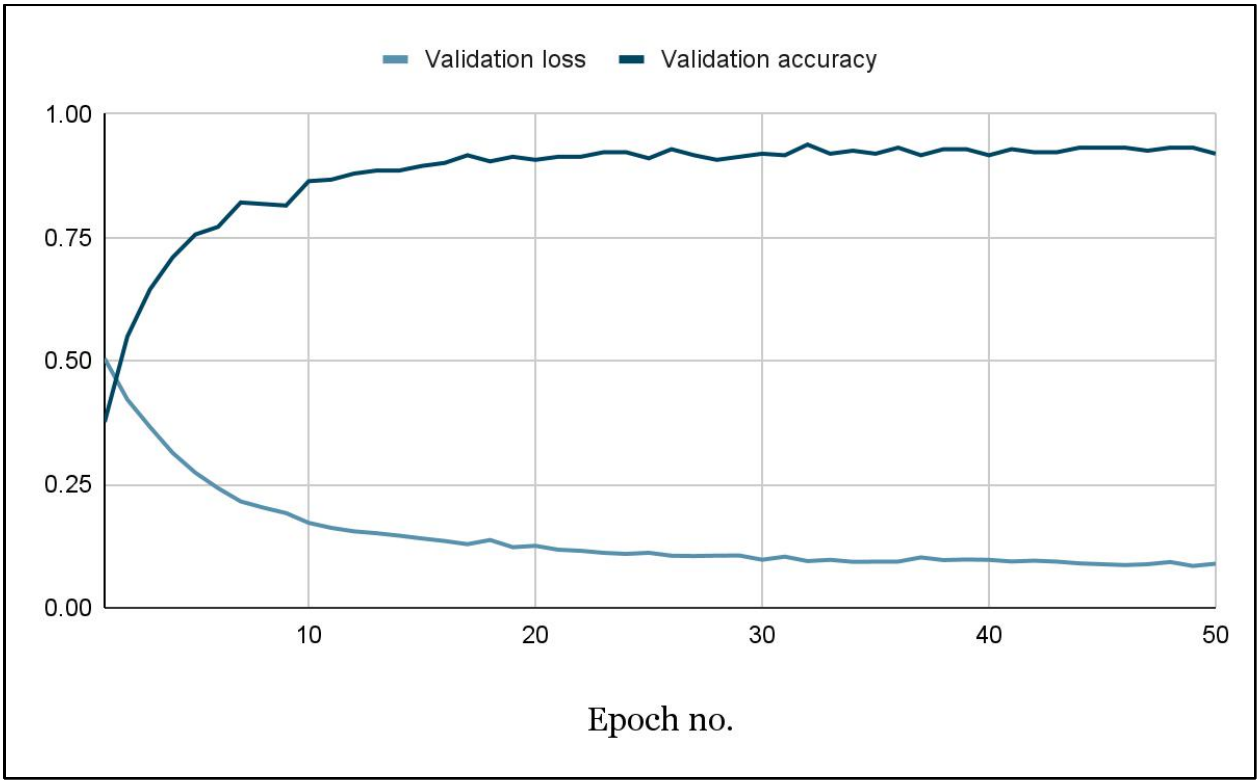


Figure 7: Validation loss and validation accuracy curve for the ANN model

The loss for the ANN model is computed by the Mean Square Error which is computed according to the Eq. 2.

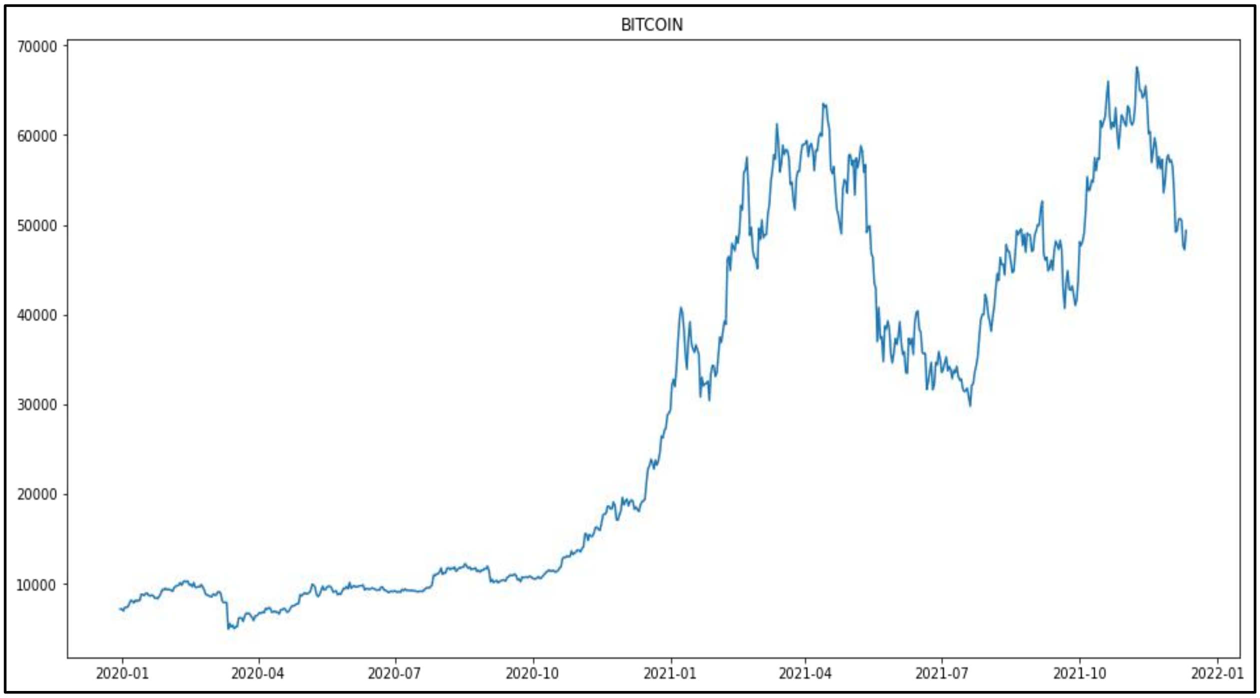
$$e = \frac{\sum_{i=1}^{n} (y - y_{pred})^2}{n} \quad (2)$$

Where e in Eq. (2) denotes the error computed by MSE, n is the total number of observations, y and y_{pred} are the actual and predicted output value respectively. The training accuracy computed for the ANN model is increased from 0.3243 in the 1st epoch to 0.9876 in the 50th epoch whereas the loss value decreased from 0.5605 1st epoch to 0.0254 50th epoch. On the other hand, the testing accuracy and loss value computed for the 1st epoch is 0.5048 and 0.3765 respectively which are improved to 0.0893 and 0.9198 respectively in the 50th epoch.

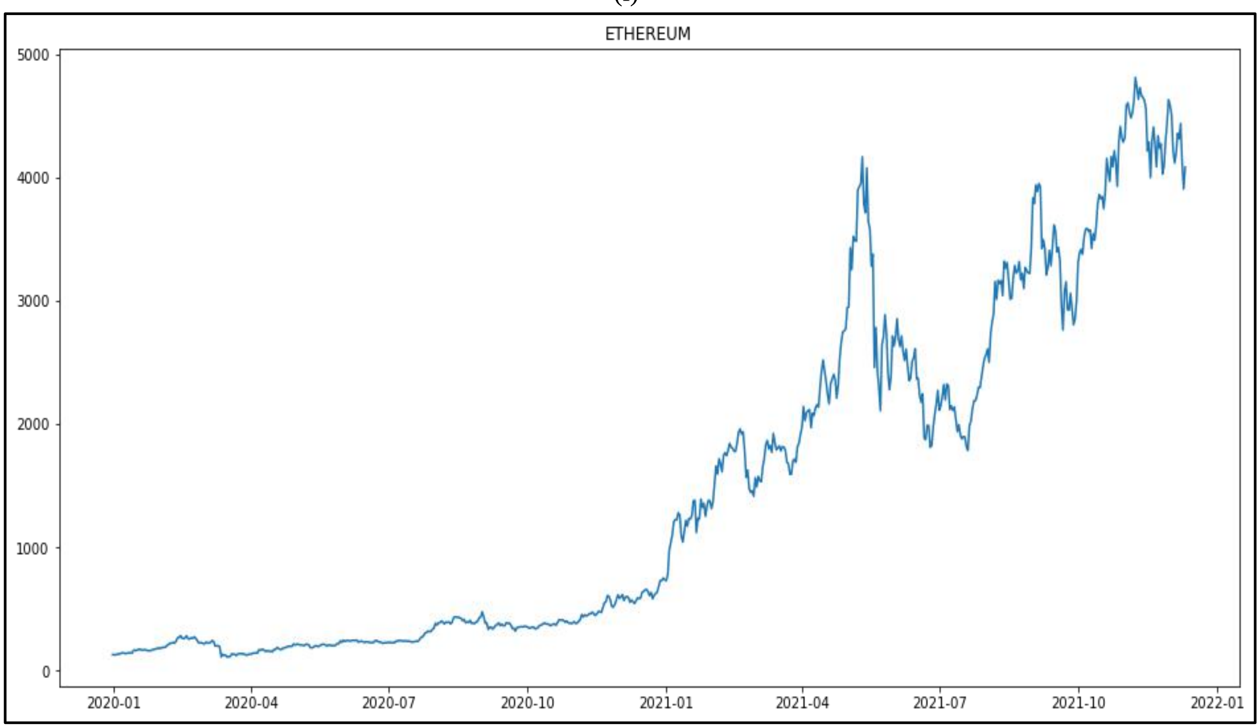
4.2. Validation of the ANN model

For validating the ANN model, the value of the CTCs is simulated for the data from January 2020 till January 2022 and the error value is computed according to the Eq. (2). The simulated graphs are shown in figure 8 and the error values are shown in table 1.

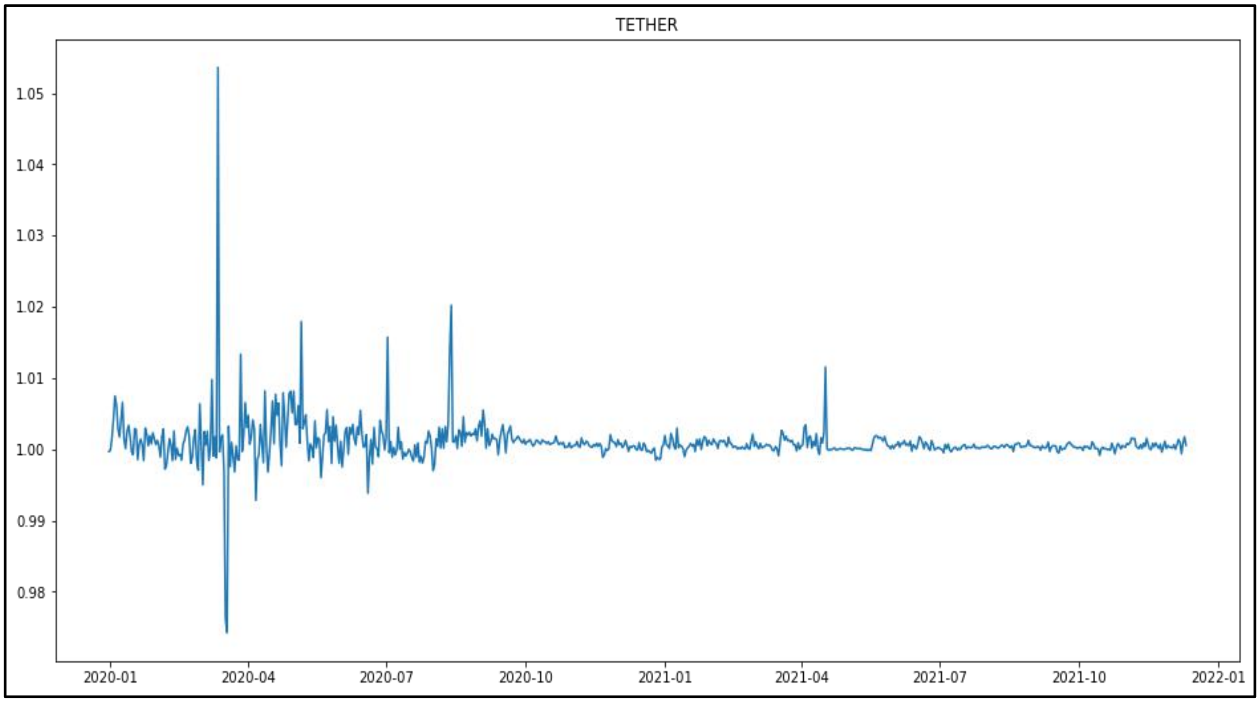
*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 10 Issue 9 September 2022



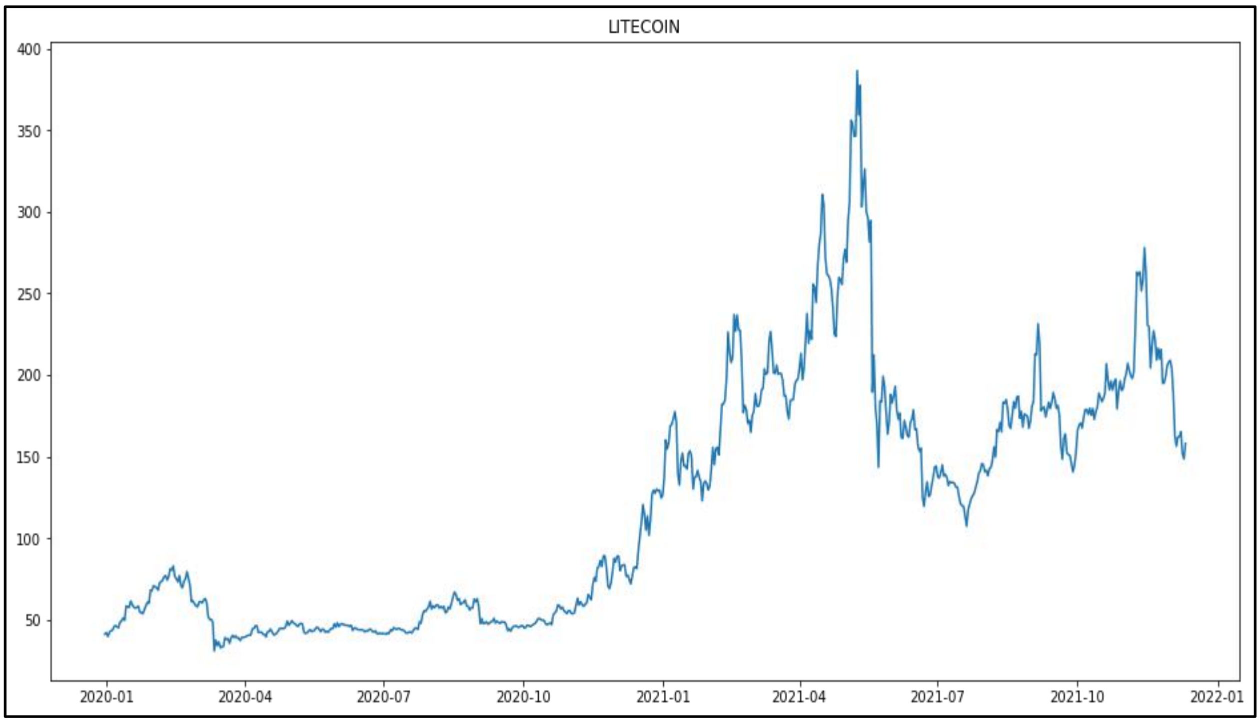
(i)



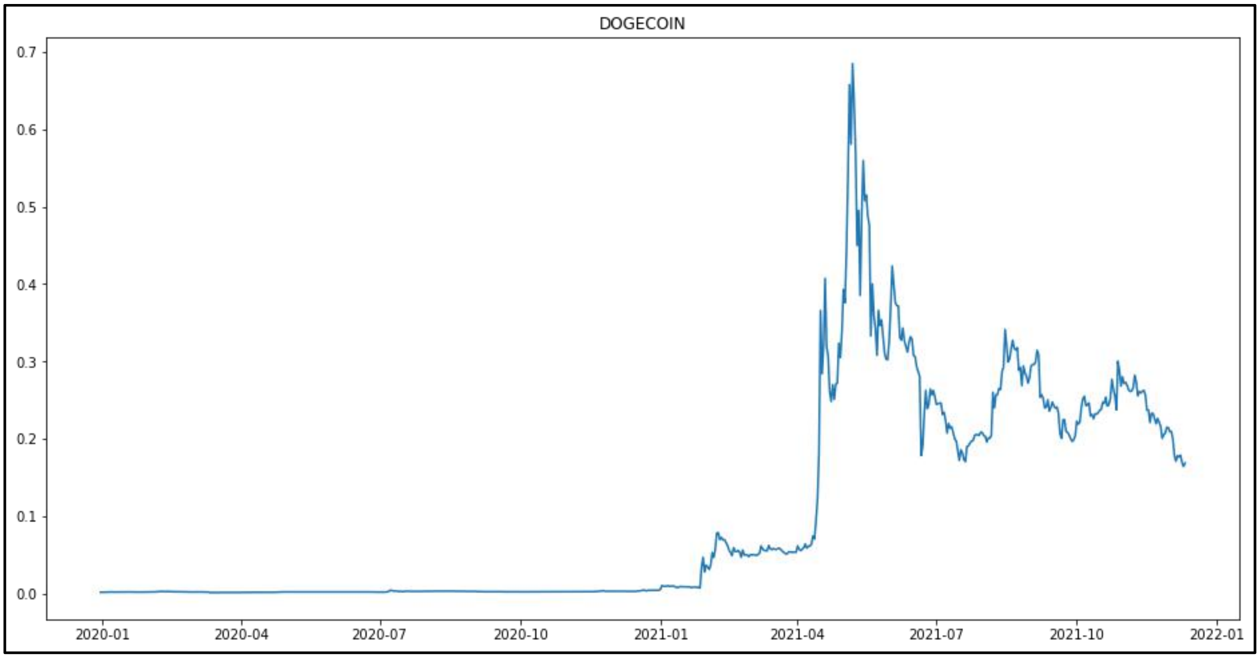
(ii)


TETHER
1.05
1.04
1.03
1.02
1.01
1.00
0.99
0.98
2020-01
2020-04
2020-07
2020-10
2021-01
2021-04
2021-07
2021-10
2022-01

(iii)


LITECOIN
400
350
300
250
200
150
100
50
2020-01
2020-04
2020-07
2020-10
2021-01
2021-04
2021-07
2021-10
2022-01

(iv)



(v)

Figure 8: Simulated graphs of (i) BTC, (ii) ETH, (iii) TET, (iv) LTC and (v) DGC

The MSE values computed for the simulated and the actual price of the CTCs are shown in table 1.

Table 1: MSE values for the CTCs

CTC	**BTC**	**ETH**	**TET**	**LTE**	**DGC**
MSE	1447.648	107.38	0.072	3.764	0.0045

4.3. Results from SVM

70% of the dataset are trained using the SVM algorithm to build the model. The remaining 30% of the dataset are used for testing the developed SVM model. The SVM model showed training and testing accuracy of 0.796 and 0.7981 respectively

4.4. Discussions

To predict the CTC that is the most stable, high return with least risk CTCs the data of 1st August 2022 is fed into the developed models. The result obtained SVM showed that DGC is best to invest, and ANN showed that ETH is the best. However, the trend showed that the investment in ETH has increased and since ANN is a better technique than SVM, ETH is the best investment option.

5. Conclusions

The overall intention of the present study is to develop a ML model that not only can predict the price of CTCs with accuracy but also determines the most stable, high return with least risk CTCs. In achieving the objective of the present study two models namely SVM and ANN were developed. The dataset used for the study is extracted from https://data.cryptocompare.com from 1st March 2021 till 25th July 2022. The website comprises price data for 100

CTCs but the five CTCs with the highest average market capitalization were considered for the study. Analysis showed that BTC, ETH, TET, LTE and DGC have the highest average market capitalization. The dataset is divided into 70:30 where 70% data are used to train the models and 30% data are used for testing it. The ANN model developed showed training and testing accuracy of 0.9876 and 0.9198 respectively. The ANN model computed training and testing loss of 0.5605 and 0.0893 respectively. The developed SVM model has training and testing accuracy of 0.796 and 0.7981 respectively. To achieve the final objective of the present research, the data of 1st August 2022 is fed into the developed models. The result obtained SVM showed that DGC is best to invest, and ANN showed that ETH is the best. Since the ANN is a better predicting technique than the SVM and above that the accuracy is far better for ANN than SVM, so the result predicted by ANN is considered. This concepts in the present study can be further extended to understand the stability and return from CTCs using different AI and ML models.

References

1. Laboure, M., H.-P. Müller, M., Heinz, G., Singh, S., & Köhling, S. (2021). Cryptocurrencies and cbdc: The route ahead. Global Policy, 12(5), 663-676.

2. Kinney, A. B. (2021). Embedding into an Emerging Money System: The Case of Bitcoin. Sociological Focus, 54(1), 77-92.

3. Rice, M. (2019). Cryptocurrency: History, Advantages, Disadvantages, and the Future.

4. Yiying, W., & Yeze, Z. (2019, March). Cryptocurrency price analysis with artificial intelligence. In 2019 5th International Conference on Information Management (ICIM) (pp. 97-101). IEEE.

5. Rathan, K., Sai, S. V., & Manikanta, T. S. (2019, April). Crypto-currency price prediction using decision tree and regression techniques. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 190-194). IEEE.

6. Cortez, K., Rodríguez-García, M., & Mongrut, S. (2020). Exchange Market Liquidity Prediction with the K-Nearest Neighbor Approach: Crypto vs. Fiat Currencies. Mathematics 2021, 9, 56.

7. Chouhan, S. S., Mukhija, M. K., & Dangi, P. (2021). Design Implementation of Machine Learning Based Crypto Currency Prediction System. EFFLATOUNARIA-Multidisciplinary Journal, 5(2), 1639-1650.

8. Poongodi, M., Nguyen, T. N., Hamdi, M., & Cengiz, K. (2021). Global cryptocurrency trend prediction using social media. Information Processing & Management, 58(6), 102708.

9. Tanwar, S., Patel, N. P., Patel, S. N., Patel, J. R., Sharma, G., & Davidson, I. E. (2021). Deep learning-based cryptocurrency price prediction scheme with inter-dependent relations. IEEE Access, 9, 138633-138646.

10. Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., ... & Zhang, J. (2021, April). Lstm based sentiment analysis for cryptocurrency prediction. In International Conference on Database Systems for Advanced Applications (pp. 617-621). Springer, Cham.

11. Wu, C. H., Lu, C. C., Ma, Y. F., & Lu, R. S. (2018, November). A new forecasting framework for bitcoin price with LSTM. In 2018 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 168-175). IEEE.

12. Andi, H. K. (2021). An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model. Journal of Soft Computing Paradigm, 3(3), 205-217.

13. Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. AI, 2(4), 477-496.

14. Livieris, I. E., Kiriakidou, N., Stavroyiannis, S., & Pintelas, P. (2021). An advanced CNN-LSTM model for cryptocurrency forecasting. Electronics, 10(3), 287.

15. Hashish, I. A., Forni, F., Andreotti, G., Facchinetti, T., & Darjani, S. (2019, September). A hybrid model for bitcoin prices prediction using hidden Markov models and optimized LSTM networks. In 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 721-728). IEEE.

16. Li, L., Arab, A., Liu, J., Liu, J., & Han, Z. (2019, July). Bitcoin options pricing using LSTM-based prediction model and blockchain statistics. In 2019 IEEE international conference on Blockchain (Blockchain) (pp. 67-74). IEEE.

17. Rebane, J., Karlsson, I., Papapetrou, P., & Denic, S. (2018). Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: A comparative study. In SIGKDD Fintech'18, London, UK, August 19-23, 2018.

18. Azari, A. (2019). Bitcoin price prediction: An ARIMA approach. arXiv preprint arXiv:1904.05315.

19. Wirawan, I. M., Widiyaningtyas, T., & Hasan, M. M. (2019, September). Short term prediction on bitcoin price using ARIMA method. In 2019 International Seminar on Application for Technology of Information and Communication (iSemantic) (pp. 260-265). IEEE.

20. Poongodi, M., Vijayakumar, V., & Chilamkurti, N. (2020). Bitcoin price prediction using ARIMA model. International Journal of Internet Technology and Secured Transactions, 10(4), 396-406.

21. Nguyen, D. T., & Le, H. V. (2019, November). Predicting the price of bitcoin using hybrid ARIMA and machine learning. In International Conference on Future Data and Security Engineering (pp. 696-704). Springer, Cham.

22. Hua, Y. (2020). Bitcoin price prediction using ARIMA and LSTM. In E3S Web of Conferences (Vol. 218, p. 01050). EDP Sciences.

23. Karakoyun, E. S., & Cibikdiken, A. O. (2018, May). Comparison of arima time series model and lstm deep learning algorithm for bitcoin price forecasting. In The 13th multidisciplinary academic conference in Prague (Vol. 2018, pp. 171-180).

24. Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. Applied Soft Computing, 38, 788-804.

25. Medhi, T., Hussain, S. A. I., Roy, B. S., & Saha, S. C. (2021). An intelligent multi-objective framework for optimizing friction-stir welding process parameters. Applied Soft Computing, 104, 107190.

26. Han, L. Y., Zheng, C. J., Xie, B., Jia, J., Ma, X. H., Zhu, F., ... & Chen, Y. Z. (2007). Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. Drug discovery today, 12(7-8), 304-313.

27. X. F. Du, S. C. Leung, J. L. Zhang, and K. K. Lai, "Demand forecasting of perishable farm products using support vector machine," International journal of systems Science, vol. 44, no. 3, pp. 556–567, 2013.