Bitcoin Price Prediction Using Machine Learning and Deep Learning Algorithms

Pratham Agarwala
4 PUBLICATIONS 47 CITATIONS
SEE PROFILE

Chittaranjan Chandwani
Maharaja Surajmal Institute Of Technology
2 PUBLICATIONS 9 CITATIONS
SEE PROFILE

Suman Mann
Galgotias University
75 PUBLICATIONS 431 CITATIONS
SEE PROFILE

2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Oct 13-14, 2022

2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 978-1-6654-7433-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICRITO56286.2022.9964677

Bitcoin Price Prediction Using Machine Learning and Deep Learning Algorithms

Bhavay Malhotra,
Student
Department of Information Technology
Maharaja Surajmal Institute of Technology
Guru Gobind Singh Indraprastha University,
New Delhi, India
malhotrabhavay2001@gmail.com

Chittaranjan Chandwani,
Student
Department of Information Technology
Maharaja Surajmal Institute of Technology
Guru Gobind Singh Indraprastha University,
New Delhi, India
chittaranjanchandwani@gmail.com

Pratham Agarwala,
Student
Department of Information Technology
Maharaja Surajmal Institute of Technology
Guru Gobind Singh Indraprastha University,
New Delhi, India
agarprath21@gmail.com

Suman Mann,
Associate Professor
Department of Information Technology
Maharaja Surajmal Institute of Technology
Guru Gobind Singh Indraprastha University,
New Delhi, India
sumanmann@msit.in

***Abstract*–A cryptocurrency or 'coins' is a virtual asset which can be used as an alternative to physical currency via a computer network that is not reliant on the government or any bank, to uphold or maintain it. This has led to the creation of numerous brands of cryptocurrencies. It was only after the boom in their price in 2011 that they began to be regarded as an investment asset. Since these coins are highly volatile in their pricing, there is a need for a good prediction of their closing price on which investment decisions can be made. To address this requirement, this paper studies the relative performances of different machine learning algorithms for a well-known cryptocurrency – the 'Bitcoin'. The performance measures of different machine learning models were undertaken to get the accuracy of the models for Bitcoin and results were obtained. The results show that the Auto Regressive Integrated Moving Average (ARIMA) is better than the other models and has the least mean absolute error. It is observed that the quality of training data and amount of the dataset used plays an important role for a successful prediction. When comparing the predicted value of Bitcoin through ARIMA with its actual value, the results obtained are found to be comparable for the entire four months of analysis.**

Keywords - Bitcoin; Deep Learning; Forecasting; Machine Learning;

I. Introduction

Cryptocurrency is a form of digital currency that exists as a string of encrypted data. To make it unique, this digital data is monitored and organised through a peer to peer network (p2p) called 'Blockchain'. This p2p network provides a secure and safe ledger for purchase, sale, and transfer transactions. These currencies, although new, are becoming increasingly pertinent in the financial industry and can be considered growing markets. Unlike physical money, cryptocurrencies are decentralized. Which means, it is not issued by the government or other financial institutions. Some popular cryptographic examples are Bitcoin, Ether, Litecoin and Monero.

Bitcoin is based on p2p transactions and is without government surveillance, meaning it is completely decentralized. Transactions and liquidity in the network are based on encryption instead. Due to the uniqueness and increasing adoption of its payment protocol, the Bitcoin ecosystem has received a lot of interest from investors, businesses, as well as consumers. This increasing demand and the use of Bitcoin has made it possible to trade on the stock exchange.

In recent years, the emergence and availability of a combination of trading hardware with high frequency and low latency and robust machine learning algorithms has allowed the stock market to maximize economic profits. Therefore, it makes sense to duplicate this prediction method in the Bitcoin world as networks gain liquidity and more people are interested in investing in systems for profit. With this intent, this study aims to use machine learning to predict the price of Bitcoin.

II. Literature Survey

There has been an enormous boom in the data sector of information technology that has allowed the immense potential of data to be recognized by various industries. This has encouraged companies to collect huge amounts of data and use it to gain insights for improvement and to predict the future of their businesses. On the same lines cryptocurrency too has seen a huge rise in attention due to its high valuation and profits, which makes it a popular investment asset despite its high volatility [1], particularly with Bitcoin which is one of the most popular and the oldest cryptocurrency. When investing in a volatile asset like cryptocurrency, one must be aware of historic prices in order to predict and invest correctly. It is here that machine learning meets cryptocurrency [2]. Using the huge historic continuous time dependent datasets of cryptocurrency, one can forecast future trends and invest more knowledgeably. This has led to a number of studies that compare different machine learning and deep learning algorithms using different kinds of time-series datasets

TABLE I. LITERATURE REVIEW

S.No.	Investigator	Methodology	Output	Findings
1.	Azizah Hitam, N.; Ritahani Ismail, A[3], Cao, L.; Tay, F.E.[4-5], Huang, W.; Nakamori, Y.; Wang, S.Y[6]	Neural networks, support vector machines, Deep learning models	SVM classifier performed best with 95% performance accuracy	While SVM worked best for Ethereum, Artificial Neural Networks(ANNs) gave the best result for Bitcoin
2.	Lipo Wang; Edwin Sim. [7]	Genetic Algorithm based Selective Neural Network Ensemble (GASEN)	The proposed approach GASEN generated 85% returns.	It outperformed the "previous day trend following" trading strategy which generated 38% returns. Also a single, best MLP model was used which generated 53% returns.
3.	Memon, R.A.; Li, J.P.; Ahmed, J[13], Hölbl, M.; Kompara, M.; Kamišali´c, A.; Nemec Zlatolas, L.[14]	A queuing theory-based model is proposed for understanding the working and theoretical aspects of the blockchain	71.428 transactions could be fit in one block	Analysis is done by simulating Bitcoin and Ethereum transactions for two months
4.	Samuel Asante Gyamerah[15],	Generalized linear model via penalized maximum likelihood, random forest, support vector regression with linear kernel, and stacking ensemble	Stacking ensemble model with two base learner (random forest and generalized linear model via penalized maximum likelihood) performed best a R-squared value of 0.9967	The Generalized linear model gave a value of 0.9966 , Random Forest gave 0.9835 and SVR(linear) gave 0.9960 R-squared value all on testing datasets.
5.	S.M. Raju; Ali Mohammad[17]	Long short-term memory(LSTM), Auto-Regressive Integrated Moving Average (ARIMA)	RMSE value of LSTM is 197.515 while for ARIMA it is 209.263	The RMSE value of LSTM on single feature is 198.448 and RMSE value on multiple feature is 197.515
6.	Feng Mai; Zhe Shan; Qing Bai; Xin Wang; Roger H.L Chiang[18],	Vector Error Correction Models (VECMs) are used to study the relationship between Bitcoin and Social Media	Social media sentiment is an important predictor in determining bitcoin's valuation, but not all social media messages are of equal impact.	Many researchers have used the volume of bitcoin traded over a period of time and data from different social platforms (like Google Trends) to gather information for model training.
7.	Skipper Seabold; Joseph Perktold[19]	Moving time series data is made stationary by the process of differencing (subtracting past values from present values) in the ARIMA model which then uses Linear Regression Model using its lags as predictors.	The algorithm is successfully implemented on moving time series data by making it stationary and applying regression with lags and predictors in python's statsmodel library.	Benchmarking tests were done to check the accuracy and suitability performance of the model for its intended purpose.

acquired from various sources like medical instruments, factory power usage, economics etc. Recent works in this field have been stated in **Table 1**.

III. METHODOLOGY USED AND THE DATASET

A. Regression Analysis

Regression analysis is one of the most primitive machine learning models that aims to understand the relation between two or more variables. In this one of the variables is dependent while the rest are independent. Accordingly, the functions chosen are fitted onto the data as per the error function to predict the outcome.

Estimate a function $f_\beta(.)$ (parameterized with β) given data points (x_i, y_i) □i ϵ {0, 1, …. , n-1} under a loss function $\Sigma_i\ l(f(x_i), y_i)$.

1) Linear Regression

In linear regression the aim is to generate a two dimensional line of data points. The mean square error parameter is minimized to achieve the desired linear regression.

Mathematically, linear regression solves the following problem:

Given P number of data points (x_i, y_i) where $x_i, y_i\ \epsilon\ R\ \forall i\ \epsilon$ {0, 1, ... , P-1} , fit a linear function

$$\hat{y} = f_\beta(x) = \beta_0 + \beta_1 x$$

by minimizing

$$min_{\beta} \sum_{p} \quad ||y^{p} - f(x^{p})||^{2}$$

2) Ridge Regression

The ridge regression technique is used to analyze multiple regression data that have multiple collinearity. When such multiple collinearity occurs, the least squares estimation is considered unbiased, however due to the large variance this can be incorrect. Ridge regression reduces this error by adding a bias to the estimate to provide a much more reliable estimate.

3) Least Absolute Shrinkage Selector Operator Regression (LASSO)

LASSO regression and ridge regression, both are used as normalizations for overfitting of training data points. However, LASSO offers additional benefits because it enhances the economics of learned weights. While the ridge regression forces the learned weights to be small but non-zero so as to reduce the overall norm. LASSO attempts to bring most of the weights closer to zero. This makes the weight matrix sparse which is efficient in implementation while providing a similar accuracy in fitting. Mathematically, by modifying the loss function LASSO regression solves a problem as shown.

Given P number of data points (x_i, y_i) where $x_i, y_i \in R \; \forall i \in$ {0,1,......, P-1}, fit a function f_β (parameterized with β) by minimizing

$$min_{\beta} \sum_{p} \quad ||y_i - f(x_i \quad)||^2 + \alpha||\beta||_1$$

where $\alpha \in R$ is a scaling factor and $||\beta||_1 = \Sigma_k|\beta_k|$.

LASSO and ridge regression differ in the sense that LASSO uses the weight L1 norm in place of the L2 norm. When using the L1 norm, loss function tends to increase sparsity of the learned weights.

B. Long-Short Term Memory (LSTM)

The LSTM is a Recurrent Neural Network-style model which is considered an acceptable and a measurable approach in cases where data is sequential in nature. It can be used to understand long-term dependency of the words in a sentence. It uses functional gates to supervise the flow of information between cells. The gates modify the information that travels through them and provides a filtered version of the inputs.

The following equations formulate the forward training processes in LSTM.

$$f_t = \sigma(W_f \,.\, [h_t - 1, x_t] + b_f)$$

$$i_t = \sigma(W_f \,.\, [h_t - 1, x_t] + b_i)c_t = f_t * c_t - 1 + i_t * tanh(W_c \,.\, [h_t - 1, x_t] + b_c)$$

$$o_t = \sigma(W_o \,.\, [h_t - 1, x_t] + b_c)$$

$$h_t = o_t * tanh(c_t)$$

Where,

it is the activation of input gate

ot is the activation of output gate

ft is the activation of the forget gate

Ct is the activation vector for the cell block

ht is the activation vector for the memory block

W is the weight matrix

b is the bias vector, respectively.

σ(◦) is the denotes the sigmoid function.

C. Gated Recurrent Unit (GRU)

Another type of recurrent neural network (RNN) is the Gated RNN which has shown feasibility and robustness in varying tasks involving sequential data. The functions showing the working of GRU are:

$$z_t = \sigma(W_{zxt} + V_{zht} - 1 + b_z)$$

$$r_t = \sigma(W_{rxt} + V_{rht} - 1 + b_r)$$

$$h_t = tanh(W_{cxt} + V_c(r_t . h_t - 1))$$

$$h_t = (1 - z_t). h_t - 1 + z_t . h_t$$

Where,

WeRdXd, VeRdXd, and beRdXd are the model parameters that are shared by time steps and learned in the training stage,

(.) represents the element-wise product, and k represents the dimensionality of the hidden vectors.

D. Auto-Regressive Integrated Moving Average (ARIMA)

ARIMA, which is a statistical analysis model, uses time-series data to either predict future trends or to better understand the dataset. For example, an ARIMA model might seek to predict the future price of some stocks based on their past performance or forecast the earnings of a company based on past periods.

The model is used on a stationary time series. The prediction obtained is given as a linear regression that includes features like time differences and moving averages. The implementation uses the 'Statsmodels' package [18]. In ARIMA, the data is differenced by transforming the price features to the difference between prices. The ARIMA equations are

$$(1 - \sum_{k=1}^{p} \quad \alpha_k L^k)(1 - L)^d X_t) = (1 - \sum_{k=1}^{q} \quad \beta_k L^k)\epsilon_t$$

where L is considered as the lag operator and p, d, q are hyper-parameters over which the lag operator is optimized.

At each time t, we use the price history to train our model and predict the price.

E. The Dataset

Time series data is any set of information that contains many disparate measurements that update continually over time as seen in the financial performance logs, healthcare records, and industrial or supply chain process reports. The stewards of such data want to forecast future performance more accurately but they too need help in understanding the evolution of the data. To address this requirement, the paper enables such forecasts in real-time.

The dataset used for analysis is obtained from CryptoCompare via API (https://min-api.cryptocompare.com/). The price information of Bitcoin has been taken for a period of 4 months (at the time of training) in order to reduce computational efforts. The data is available in the format as shown in **Table 2** and **Fig. 1**. It has 5 features:

- Close Price: Market closing price of the day.
- High Price: Particular day's highest price.
- Low Price: Particular day's lowest price.
- Open Price: Particular day's market opening price..
- Volume: The volume of currency that is being traded for that particular day.

F. Data Splitting

Following standard methods, the complete dataset is divided into 80% and 20% parts. The 80% is used for training the model and 20% for testing. The split is done by time so that the first data point in the test set is in the future with respect to the last data point in the training set. Hyperparameters are adjusted by training and testing in the training set. The final result is obtained through training and testing in the test set.

TABLE II. THE DATA IN TABULAR FORM

time	high	low	open	volumefrom	volumeto	close
2020-12-03	25261.84	24499.62	24928.35	107.98	2696330.33	25114.66
2020-12-04	25221.08	23928.52	25114.66	112.41	2750699.79	24002.57
2020-12-05	24689.49	23829.92	24002.57	45.59	1108897.32	24634.57
2020-12-06	24906.15	24253.18	24634.57	34.22	842652.10	24862.59
2020-12-07	24967.51	24304.60	24862.59	77.25	1904163.36	24626.17

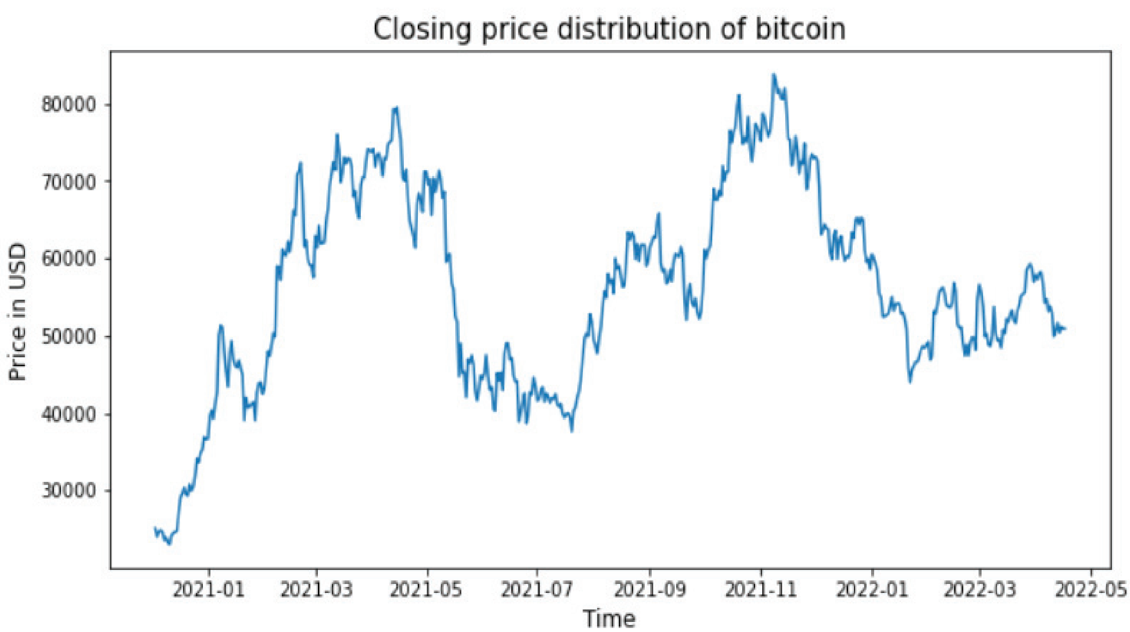

Fig. 1. Data in Graphical Form (Source: Authors)

IV. RESULT ANALYSIS

Several different models as discussed in the preceding section were trained with the same set of dataset.

The linear regression for the available data is as seen in **Fig. 2**.

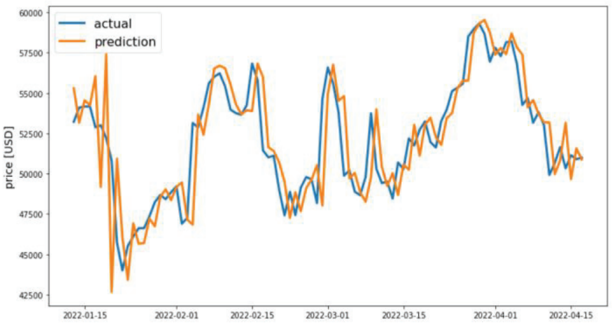

Fig. 2. Linear Regression (Source: Authors)

The ridge regression results for the present analysis is shown in **Fig 3.**

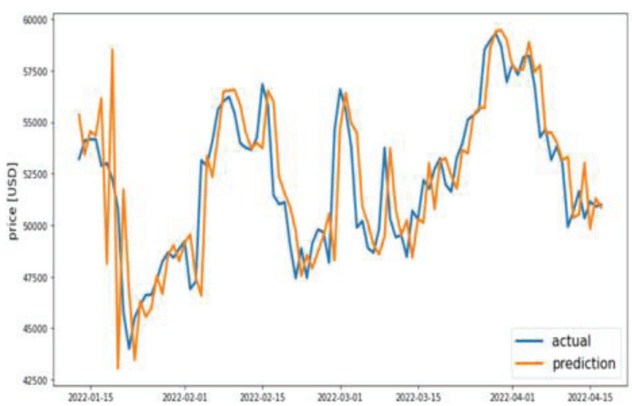

Fig. 3. Ridge Regression (Source: Authors)

Fig 4 shows the LASSO regression of the current analysis.

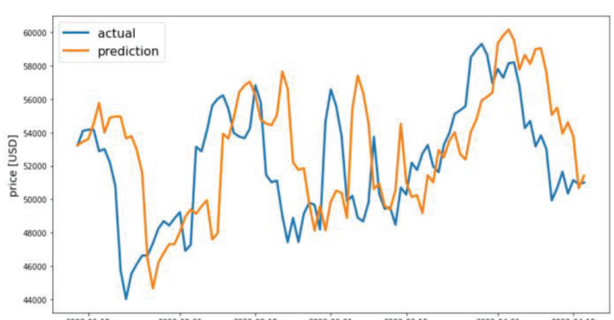

Fig. 4. Lasso Regression (Source: Authors)

The LSTM graph for the study is seen in **Fig. 5**.

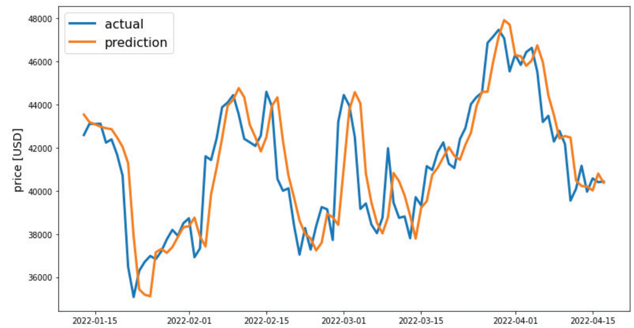

Fig. 5. LSTM Graph (Source: Authors)

The resulting output for the current analysis is as seen in **Fig. 6**.

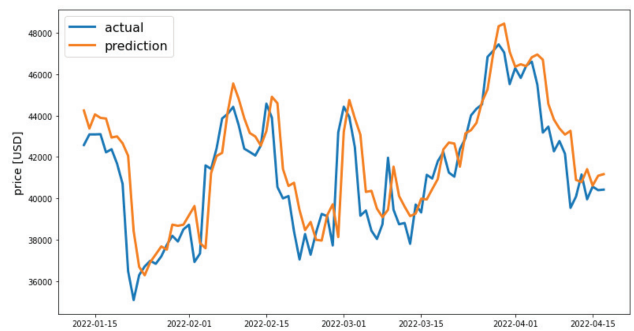

Fig. 6. GRU Graph (Source: Authors)

The changes in price acts as a sign of predictions to arrive at the ARIMA graph as seen in **Fig. 7**.

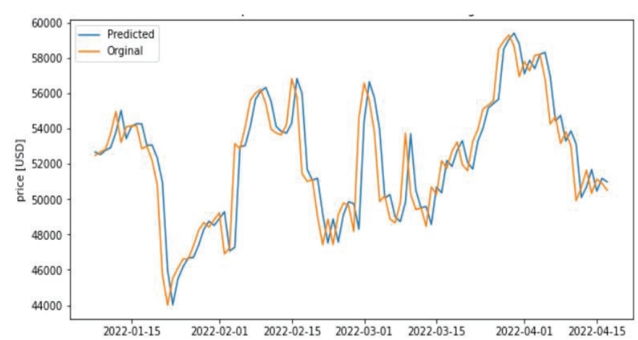

Fig. 7. ARIMA Graph (Source: Authors)

The models were then evaluated using the Mean Absolute Error. The results as obtained from this analysis are seen in **Table 3**.

TABLE III. MEAN ABSOLUTE ERROR (MAE) FOR EACH MODEL

SNo.	Algorithm	Mean Absolute Error (MAE)
1	Linear Regression	0.02897
2	Lasso Regression	0.05942
3	Ridge Regression	0.02967
4	LSTM	0.02821
5	GRU	0.02915
6	ARIMA	0.02294

The MAE values for the models with the best performance on the test data have been testified. Among all the classifiers, the ARIMA model provides the best results in comparison to the other models with a Mean Absolute Error of 0.02294. It is imperative to mention that the other models used are not designed specifically to work with time-series data, unlike the ARIMA-based model. It is thus natural that they underperform when compared to ARIMA. Apart from ARIMA, LSTM, GRU, Linear and Ridge Regression models give similar performances having almost equal Mean Absolute Error while the LASSO regression turns out to be the most unfit model for Bitcoin price prediction.

V. CONCLUSION

The present study has shown that the prediction of the nature of price fluctuations for the growing cryptocurrency market is possible with the help of machine learning practices. In doing so the performance of six prediction models, namely, Linear, Ridge, LASSO regression, LSTM, GRU and ARIMA have been evaluated and compared for Bitcoin using available market data for validation of the model. In addition, the study has demonstrated that cryptocurrency markets have enormous potential for financial time-series studies and research due to the high data accessibility and availability. However, there are a plethora of opportunities to explore in this area.

While all methods achieved less than 6% error, the best performance was achieved by the Auto Regressive Integrated Moving Average (ARIMA) model, due to its features and desirability towards time-series data. Other models fell short as they are not designed specifically to the time-series data used in the study. Since the results obtained are conclusive and encouraging, further studies will be undertaken to refine and study the dynamics of cryptocurrency markets.

VI. FUTURE SCOPE

Although the tested algorithms provided quite accurate predictions, there is still some scope of improvement. It is opined that ARIMA can be further optimized by fine tuning its parameters. Since cryptocurrency data is highly volatile, it is dependent on numerous factors. By considering more price varying factors, increasing the number of features in the dataset and increasing the data size would help in improving the performance of the model as well as provide better results. New modern technologies which have better computing powers to handle larger datasets and algorithms like Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) which are able to capture the volatility in data can be applied to obtain even better results thereby improving decision making capabilities for future investments. The versatility and usability of our study can be extended to other forms of cryptocurrencies such as Ethereum, Ripple and Litecoin as well.

REFERENCES

[1] Peter D. DeVries,"An Analysis of Cryptocurrency, Bitcoin, and the Future", International Journal of Business Management and Commerce ,2016, pp. 1-5.

[2] S Mann, T Jain, A Vyas, "The Blockchain Revolution: Paradigm Shifts in Traditional Voting Practices", International Journal of Computer Applications, 2020

[3] Azizah Hitam, N.; Ritahani Ismail, A.; "Comparative Performance of Machine Learning Algorithms for Cryptocurrency Forecasting" 2018

[4] Cao, L.; Tay, F.E. "Financial forecasting using support vector machines. Neural Comput." Appl. 2001, 10, 184–192.

[5] Cao, L.; Tay, F.E. "Application of support vector machines in financial time series forecasting". Omega 2001, 29, 309–317.

[6] Huang, W.; Nakamori, Y.; Wang, S.Y. "Forecasting stock market movement direction with support vector machine". Comput. Oper. Res. 2005, 32, 2513–2522.

[7] Lipo Wang; Edwin Sim, "Bitcoin Price Prediction Using Ensembles of Neural Networks",International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, 2017

[8] Ferreira, M.; Rodrigues, S.; Reis, C.I.; Maximiano, M., 'Blockchain: "A Tale of Two Applications"'. Appl. Sci. 2018, 8, 1506.

[9] MR Baker, DL Padmaja, R Puviarasi, S Mann, "Implementing Critical Machine Learning (ML) Approaches for Generating Robust Discriminative Neuroimaging Representations Using Structural Equation Model (SEM)", - Computational and Mathematical Methods in Medicine, 2022

[10] Trabelsi, N. Are "There Any Volatility Spill-Over Effects among Cryptocurrencies and Widely Traded Asset Classes?" J. Risk Financ. Manag. 2018, 11, 66.

[11] Corelli, "A. Cryptocurrencies and Exchange Rates: A Relationship and Causality Analysis. Risks" 2018, 6, 111.

[12] Cocco, L.; Tonelli, R.; Marchesi, M. "An Agent Based Model to Analyze the Bitcoin Mining Activity and a Comparison with the Gold Mining Industry". Future Internet 2019, 11, 8.

[13] Memon, R.A.; Li, J.P.; Ahmed, J. "Simulation Model for Blockchain Systems Using Queuing Theory". Electronics 2019, 8, 234.

[14] Hölbl, M.; Kompara, M.; Kamišali´c, A.; Nemec Zlatolas, L. "A Systematic Review of the Use of Blockchain in Healthcare". Symmetry 2018, 10, 470.

[15] Samuel Asante Gyamerah, "Are Bitcoins price predictable? Evidence from machine learning techniques using technical indicators",2019

[16] V Negi, S Mann, V Chauhan, "Devanagari character recognition using artificial neural network"- International Journal of Engineering and Technology, 2009

[17] S M Raju, Ali Mohammad Tarif, "Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis", 2020

[18] Mai F., Jay Shan, Qing Bai, Shane Wang, "How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis", Journal of Management Information Systems, 2018

[19] Skipper Seabold; Joseph Perktold. "Statsmodels: Econometric and Statistical Modeling with Python" ,January 2010