Time-Series Prediction of Cryptocurrency Market using Machine Learning Techniques

Mahir Iqbal[1], Muhammad Shuaib Iqbal[2], Fawwad Hassan Jaskani[1,*], Khurum Iqbal[2] and Ali Hassan[2]

[1]Department of Computer Systems Engineering, Faculty of Engineering, Islamia University of Bahawalpur
[2]Department of Computer Systems Engineering, College of Electrical and Mechanical Engineering, National University of Science and Technology, Rawalpindi

Abstract

In the market of cryptocurrency the Bitcoins are the first currency which has gain the significant importance. To predict the market price and stability of Bitcoin in Crypto-market, a machine learning based time series analysis has been applied. Time-series analysis can predict the future ups and downs in the price of Bitcoin. For this purpose we have used ARIMA, FBProphet, XG Boosting for time series analysis as a machine learning techniques. The parameters on the basis of which we have evaluated these models are Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R^2. We conduct experiments on these three techniques but after conducting time series analysis, ARIMA considered as the best model for forecasting Bitcoin price in the crypto-market with RMSE score of 322.4 and MAE score of 227.3. Additionally, this research can be helpful for investors of crypto-market.

*Corresponding author. Email: favadhassanjaskani@gmail.com

1. Introduction

With regard to the analysis of volatility and predictions in cryptocurrency prices [1] divides the factors influencing cryptocurrency prices into domestic and external factors [2]–[4]. Three external influences exist: Crypto market: popularity (attraction), market trend, speculations, macro-financial: equity, exchange rates, gold price, interest rate, other policies; political: legalization (fitting), restrictions (ban), other markets, external markets; The main factors are: supply & demand, cost of transactions, compensation scheme, hash rate, circulation of coins and forks (rule changes) [5], [6]. Crypto market-related variables like beta, trade volume and uncertainty, both short- and long-running, and the strong volatility of bitcoins, seem to be significantly determining in all five cryptocurrencies (Bitcoin, Ethereum, Dash, Litecoin and Monero) [7].

Cryptocurrency relies on the Blockchain automated bookkeeping framework. By implementing an access management mechanism, Blockchain systems provide ways of ensuring the privacy and protection of user data [8], [9]. A Blockchain is a decentralized connected data structure characterized by its inherent data alteration resistance, but search query deficiencies are mostly due to the inferior data formatting [10]–[12]. In [13] author shown ChainSQL, the open-source Blockchain integration framework with the database, i.e. the Blockchain application platform, which has Block chain's decentralized, distributed and audible functionality and fast query processing and a well-designed database data structure [14], [15]. The currency is based on a decentralized peer-to-peer network that creates currencies and management of transactions without central authority [16], [17]. All Bitcoin transactions are posted in blocks to an open Blockchain directory, which is called Blockchain [18], [19]. This authentication is carried out in a non-trust

environment without the intermediary having to transfer the funds from the sender to the recipient [20]–[22]. In fact, the advent of cryptocurrency and actual trading is a time series issue [23]. It differs from conventional financial markets because of its unique nature and high volatility so it offers an important subject for price forecasting [24].

Studies [25], [26] show that the price pattern of cryptocurrency has many causes and is difficult to understand. For potential investors and government agencies it is necessary to establish a cryptocurrency price prediction mechanism.

For this purpose we have used different machine learning algorithms to forecast the bitcoin prices in the market. The main goal of this research is to predict the future trend of Bitcoins. Three models: Autoregressive integrated moving average (ARIMA), Facebook Prophet and XG Boosting algorithm has been used to predict the price of bitcoin in future according to the future trends.

2. Literature Review

This section shows the previous studies conducted on the prediction of cryptocurrency time series forecasting using machine learning methods. In this section we will also conclude the research studies.

In paper [1] author focused on the short-term prediction model for machine-learning cryptocurrencies. The updated Binary Auto Regressive Tree (BART) was adapted to series data and standard models. Study shows that BART is more accurate than the ARIMA. ARFIMA model in slow-growing and transitional dynamic times.

In paper [21] study uses advanced artificial intelligence frameworks for the fully-connected Artificial Neural Network (ANN) and Long Short Memory (LSTM) network to analyses the business dynamics of Bitcoin, Ethereum and Ripple. It was that ANN uses longer-term history while LSTM uses more fast dynamics, meaning that LSTM's usefulness is greater than ANN's effectiveness in using usable historical data.

In paper [15] author uses education technologies to estimate the prices of the three digital currencies most widely traded - Bitcoin, Digital Cash and Ripple. This is the first task to use the best of our cryptocurrency prediction experience. Although the LSTM model has a greater computing burden than nonlinear pattern brutality, profound learning was essentially highly efficient in predicting the volatile dynamics inherent in crypto-currency markets [27]–[29].

In order to forecast and assess the factors affecting cryptocurrency prices, researchers have used several machine-study and deep-learning algorithms in [30][31] like Gated Recurrent (GRU) [32], Neural Networks (NN) and Long-Term Speech (LSTM) units. This paper provides a prediction method for hybrid crypto-currency LSTM that focuses only on two crypto-currencies, namely lite coin and Montero. The results show that the proposed scheme quite reliably predicts the values, showing that the scheme can be used in different cryptocurrency forecasts. In paper [31] author identify Bitcoin prices through regular prices and high frequency prices to predict Bitcoin prices through machine learning techniques at different frequencies. In comparison with the usual price benchmark results, XGBoost and SDA has higher results with highest statistical accuracy and 66 and 65.3% machine learning algorithms respectively.

In paper [6] work is aimed at extracting and comparing the accuracy of the Bitcoin prediction with various machine learning algorithms. The results of the experiment are correlated with the decision tree and the regression model.

In paper [24] author develop an investment strategy to trade cryptocurrency products in exchange markets has been proposed. This paper shows all learning algorithms, whether they are resampling methods or not, overcoming Buy and Hold (B&H) strategies in the vast majority of the 100 surveyed markets. However, the unweighted average achieves the best overall performance from learning algorithms, namely up to 59.26% time sampling accuracy. However, both alternative sampling methods tested were found to offer much higher returns and a smaller probability of data being retrieved in a timely manner.

LSTM model structure used in the research [23]. Analysis shows that by comparing f1 values, the LSTM model exceeds the time series price range booster model, which is an overall master learning model considered to be of reasonably good predictive performance. Compared to the GB model with the LSTM, with improved efficiency of around 7 percent.

In [33] ensemble Learning assumes the best output of five Comparable Signal (assemblies 5) models with an annualized ratio of 80,17%, and 91,35% for Sharpe with an annualized return of 9,62% and 5,73%, respectively (around 0,5%). The positive findings support the argument that machine learning offers robust methods for the predictability of cryptocurrencies and the creation of effective trading strategies even under adverse conditions in these markets.

In paper [3] different feature selection methods have been evaluated for the best prediction attributes in the first step of the technique. For prediction of price trends, the behaviour of Artificial Neural Networks (ANN), SVM and Ensemble algorithms has been investigated (after recurrent neural networks and the clustering process for k-means). ANN and SVM are also used for returning Bitcoin maximum, minimum and closing rates. In addition, regression results have also been used as inputs to increase price route forecasts. The results showed that the selected attributes and the best model of machine learning were improved.

In paper [18] author use the learned method of machine learning when forecasting future Bitcoin prices. Many dynamic factors affect Bitcoin prices and precise projections provide a good foundation for decision-making on investment. The projected output was also compared to 11 regression algorithms. Lasso's regression

with the combination of generalised linear regression has shown itself to improve other regression algorithms by 9%.

In paper [22] two computer teaching techniques, namely linear regression (LR) and SVM, use a regular market closing series for ether cryptocurrency in order to forecast concentric pricing. In ether cryptocurrency price prediction, different window lengths of filters with various weight coefficients are used. The training method uses a cross-validation approach to build a high-performance model regardless of the data set. Two machine learning methods are used to incorporate the proposed model. The SVM method (96.06 percent) is more accurate than the LR method when the proposed model is employed (85.46 percent) Furthermore, by adding features to the SVM process, the accuracy of the proposed model can be increased to 99 percent.

Following table shows the comparative analysis of previous techniques:

Table 1. Comparative Analysis of Previous Researches

Ref	Techniques	Outcome	Evaluation
[32]	GRU	Bitcoin Price Prediction	RMSE: 435.645
[31]	Linear Regression and LDA	Bitcoin Price Prediction	Accuracy: 65% and 60% respectively
[34]	LSTM, ANFIS, SVM	Bitcoin Exchange Rate Prediction	RMSE 354.5. 430.8 546.9 Respectively
[2]	RNN and LSTM	Bitcoin Price Prediction using Tweets	Accuracy 90%, 91.3% respectively
[3]	SVM, ANN, RNN, KNN	Daily close price of Bitcoin Prediction	MEA 0.1, 0.2, 0.19. 0.18 Respectively

3. Methodology

This section shows the dataset and proposed machine learning algorithms in detail for time series analysis of bitcoin prices

A. Dataset

Bitcoin Historical data has been collected from Kaggle, an open source website which contains 8 features. Below table shows the complete description of features of dataset:

Table 2. Features Description

Feature	Description	Unit
Timestamp	The time at which each instance or entry has been collected from crypto-currency stock market	It can be in the form of dd/mm/yyyy
Open Price	Open price for each day according to each timestamp	Million USD
High Price	Highest price on that day in which data has been collected.	Million USD
Low Price	Lowest price on that day in which data has been collected.	Million USD
Close Price	Final price of bitcoin on that day in which data has been collected.	Million USD
Volume of BTC	Turnovers in the price of BTC	Million USD
Volume of Currency	Turnovers in the price of exchange rates of currencies	Million USD
Weighted Price	The average of the shared prices of all companies invested in BTC	Million USD

Figure 1 shows the exploratory analysis of each feature according to the time0serires in which the data has been collected
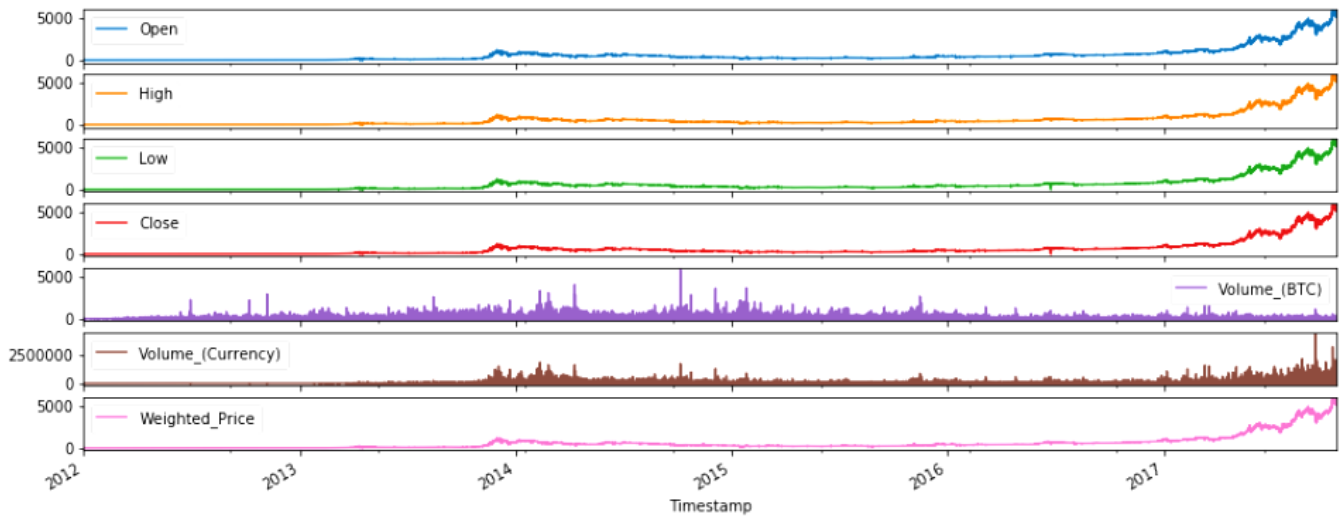
Figure 1. Exploratory Analysis of Each Feature

B. Preprocessing of Dataset

The removal of unwanted data from a dataset is called Preprocessing. For this purpose, we have normalized some features, removed outliers and significant attributed analysis has been also conducted.

C. Modelling

For the proposed model we have selected three Algorithms for Bitcoin Price Prediction:

- ARIMA
- XGBoost
- FBProphet

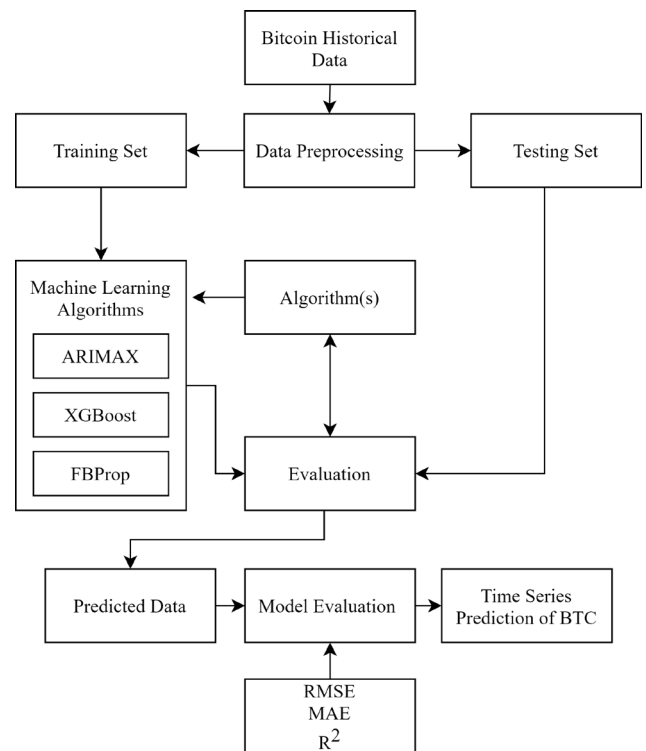Below figure 2 shows the basic methodology for this research:

Figure 2. Basic Methodology Flowchart

D. Machine Learning Algorithms

i. ARIMAX (Auto-regression Integrated Mean Average Algorithm)

ARIMAX is the machine learning model used for time series analysis. It takes data and make observations. On the basis of previous observations it takes mean average

and differentiate between consecutive two time-stamps in order to make time series stationary.
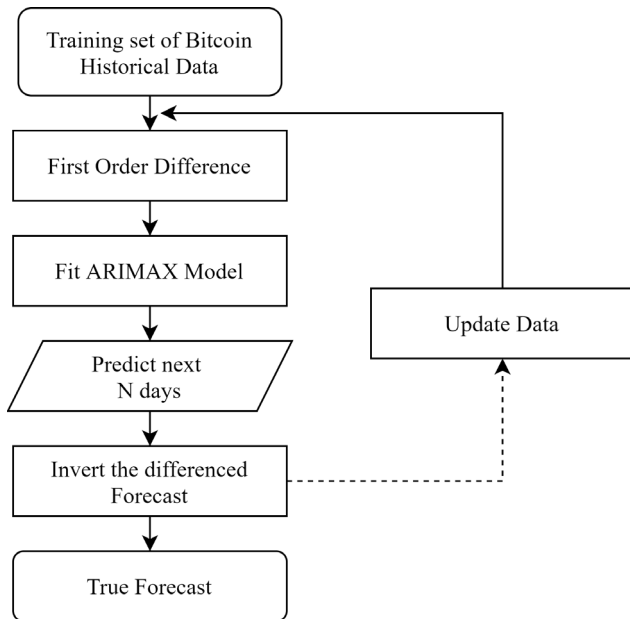

Figure 3. ARIMAX Model for BTC Prediction

ii. XGBoost Algorithm

This is the transformed model of Gradient Boosting algorithm which used transformed data in supervised learning for time-series analysis. Figure 4 shows the basic methodological steps involved in XGBoost time series analysis.
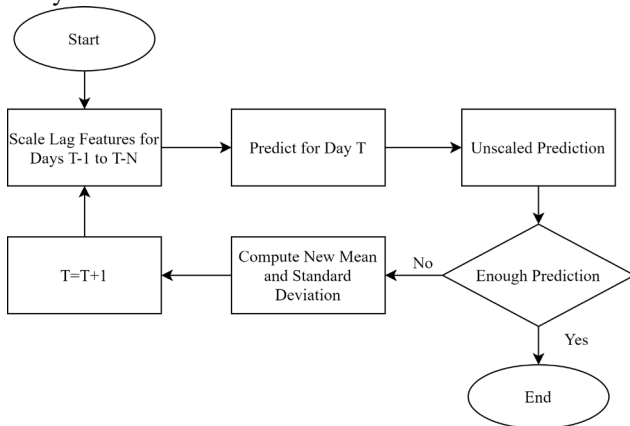

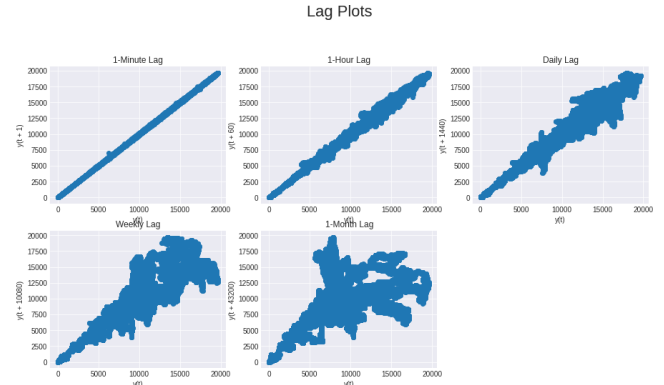**Figure 4.** XGBoost Model for Prediction

Lag Plots

Figure 5. Lag Plots

In figure 5, it can be seen that there is no correlation of timestamps with each other. So we have to resample our data for daily level.

iii. FBProp Algorithm

It is an algorithm using Forecasting time series data using an additive model where non-linear trends are fit with annual, weekly, and daily seasonality as well as holiday impacts.
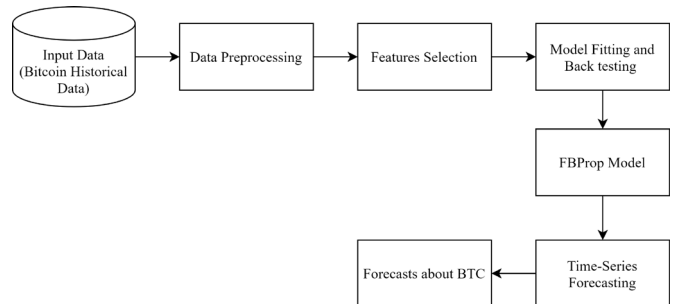

Figure 6. FBProp Algorithm for BTC

4. Results

Performance of proposed models has been measures by calculating RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and R^2 (R-Square). These parameters has widely been used in previous research studies. For conducting tests for these approached we have used 30% testing data.

a. Root Mean Square Error

RMSE is the standard deviation over the prediction errors in the time series analysis. RMSE shows how far the regression point of true data are from predicted data. It can be calculated as:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}} \qquad \dots (1)$$

b. Mean Absolute Error

It is a measure of errors between paired (True Data, Predicted Data) observations.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad \dots (2)$$

    c.   R-Square

It is a proportion of vatiance between actual and predicted data in the regression model.

$$R^2 = 1 - \frac{RSS}{TSS} \quad \dots (3)$$

A)   Performance of ARIMAX Model

This model has used some observations simulataneously with lagged observations to predict the weighted price. This model has used the differntiation between two consecutive timestamps in order to make the time series stationary. Below figure 6 shows the weighted obsered and predicted time forecast for the Bitcoin Stock Price.
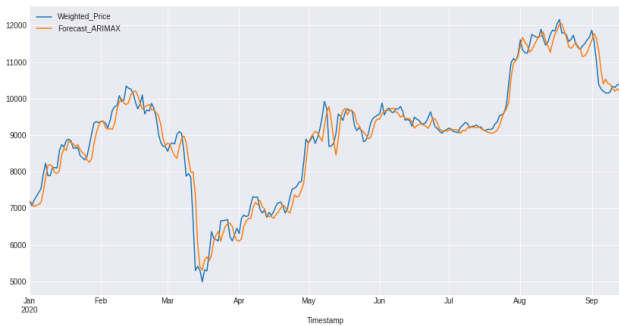

Figure 7. ARIMAX Performance

Model has been evaluated on the basis of RMSE, MAE and R^2. Table 3 shows the performance on the basis of performance parameters.

Table 3. Performance Analysis of ARIMAX

Model	Parameter	Score
ARIMAX	RMSE	322.4
ARIMAX	MAE	227.3
ARIMAX	R^2	189.6

B)   Performance of FBProp Model

This model works on Trends, Seasonality and Holidays.

    Y (t) = g (t) +s (t) +h (t) +e (t) … (4)

Where g (t) is the linear and logistic trends in the model, S (t) is the seasonality changes, h (t) is the effect of holidays and e (t) is the idiosyncratic errors.

Figure 8 below shows the weighted obsered and predicted time forecast for the Bitcoin Stock Price.
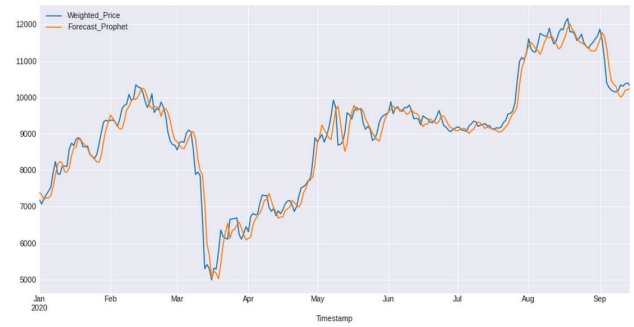

Figure 8. FBProp Performance

Model has been evaluated on the basis of RMSE, MAE and R^2. Table 4 shows the performance on the basis of performance parameters.

Table 4. Performance Analysis of FBProp

Model	Parameter	Score
FBProp	RMSE	229.5
FBProp	MAE	323.00
FBProp	R^2	205.4

C)   Performance of XGBoost Model

This is the transformed model of Gradient Boosting algorithm which used transformed data in supervised learning for time-series analysis. Below figure 9 shows the weighted obsered and predicted time forecast for the Bitcoin Stock Price

Figure 9. XGBoost Performance

Model has been evaluated on the basis of RMSE, MAE and R^2. Table 5 shows the performance on the basis of performance parameters.

Table 5. Performance Analysis of XGBoost

Model	Parameter	Score
XGBoost	RMSE	369
XGBoost	MAE	470.00

XGBoost	R^2	435.4

D) Comparative Analysis

All of the three models has been compared on the basis of their RMSE, MAE and R-square. Figure 10 shows the comparative analysis graph of these three models for time series forcasting.

Figure 10. Comparative Analysis of ARIMAX, FBProp and XGBoost

Table 5 shows the comparative perofrmacne according to RMSE, MAE and R square for each model of regression.

Table 6. Comparative Performance of Each Model

Model	Parameter	Score
ARIMAX	RMSE	322.4
ARIMAX	MAE	227.3
ARIMAX	R^2	189.6
FBProp	RMSE	229.5
FBProp	MAE	323.00
FBProp	R^2	205.4
XGBoost	RMSE	369
XGBoost	MAE	470.00
XGBoost	R^2	435.4

5. Conclusion

Our dataset contains the timestamps of yearly, monthly daily close, open, high, low and weighted price of bitcoins. We have pre-processed that data according to our requirement of normalization. Then we have applied three machine learning algorithm for time series forecasting of the bitcoin prices in the cryptocurrency market. We found that ARIMAX is the best algorithm to forecast the change in the bitcoin price in the market with RMSE of 322.4. While FBProp and XGBoost algorithms have gained the RMSE score of 229.5 and 369 respectively which is far less then ARIMAX algorithm. This study can further be improved by Hypertuning the

parameters of time-series analysis algorithms in order to improve RMSE value.

References

[1] V. Derbentsev, N. Datsenko, O. Stepanenko, and V. Bezkorovainyi, "Forecasting cryptocurrency prices time series using machine learning approach," vol. 02001, pp. 1–7, 2019.

[2] A. Mittal, V. Dhiman, A. Singh, and C. Prakash, "Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data," *2019 Twelfth Int. Conf. Contemp. Comput.*, pp. 1–6, 2019.

[3] D. C. A. Mallqui and R. A. S. Fernandes, "Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques," *Appl. Soft Comput. J.*, 2018, doi: 10.1016/j.asoc.2018.11.038.

[4] R. M. Parizi, A. Dehghantanha, K. K. R. Choo, and A. Singh, "Empirical vulnerability analysis of automated smart contracts security testing on blockchains," *arXiv*, 2018, doi: 10.5555/3291291.3291303.

[5] D. Gibert, C. Mateu, and J. Planes, "The rise of machine learning for detection and classification of malware: Research developments, trends and challenges," *J. Netw. Comput. Appl.*, vol. 153, no. March, p. 102526, 2020, doi: 10.1016/j.jnca.2019.102526.

[6] K. Rathan, S. V. Sai, and T. S. Manikanta, "Crypto-Currency price prediction using Decision Tree and Regression techniques," *2019 3rd Int. Conf. Trends Electron. Informatics*, no. Icoei, pp. 190–194, 2019.

[7] S. Xiaolei, L. Mingxi, and S. Zeqian, "A novel cryptocurrency price trend forecasting model based on LightGBM," no. November, 2018, doi: 10.1016/j.frl.2018.12.032.

[8] C. Gurdgiev and D. O. Loughlin, "Journal of Behavioral and Experimental Finance Herding and anchoring in cryptocurrency markets : Investor reaction to fear and uncertainty," *J. Behav. Exp. Financ.*, vol. 25, p. 100271, 2020, doi: 10.1016/j.jbef.2020.100271.

[9] T. Fang, Z. Su, and L. Yin, "International Review of Financial Analysis Economic fundamentals or investor perceptions ? The role of uncertainty in predicting long-term cryptocurrency volatility," *Int. Rev. Financ. Anal.*, vol. 71, no. 39, p. 101566, 2020, doi: 10.1016/j.irfa.2020.101566.

[10] S. Singh, T. K. Madan, J. Kumar, and A. K. Singh, "Stock Market Forecasting using Machine Learning: Today and Tomorrow," *2019 2nd Int. Conf. Intell. Comput. Instrum. Control Technol. ICICICT 2019*, pp. 738–745, 2019, doi: 10.1109/ICICICT46008.2019.8993160.

[11] S. Mehtab and J. Sen, "A Time Series Analysis-Based Stock Price Prediction Using Machine Learning and Deep Learning Models," *arXiv*, pp. 1–46, 2020, doi: 10.13140/RG.2.2.14022.22085/2.

[12] X. Li *et al.*, "Empirical analysis: stock market prediction via extreme learning machine," *Neural Comput. Appl.*, vol. 27, no. 1, pp. 67–78, 2016, doi: 10.1007/s00521-014-1550-z.

[13] M. Abraham, "Studying the patterns and long-run dynamics in cryptocurrency prices," no. October, pp. 1–16, 2019, doi: 10.1002/jcaf.22427.

[14] S. Mohapatra, N. Ahmed, and P. Alencar, "KryptoOracle : A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments," pp. 5544–5551, 2019.

[15] N. Science, C. Phenomena, S. Lahmiri, and S. Bekiros, "Chaos , Solitons and Fractals Cryptocurrency forecasting with deep learning chaotic neural networks," *Chaos, Solitons Fractals Interdiscip. J. Nonlinear Sci. Nonequilibrium Complex Phenom.*, vol. 118, pp. 35–40, 2019, doi: 10.1016/j.chaos.2018.11.014.

[16] C. Time-series, I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas, "Ensemble Deep Learning Models for Forecasting," pp. 1–21, 2020, doi: 10.3390/a13050121.

[17] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning," *Physica A*, vol. 551, p. 124569, 2020, doi: 10.1016/j.physa.2020.124569.

[18] I. Publishing, "Sensors & Transducers Cryptocurrency Price Prediction :," vol. 244, no. 5, pp. 44–47, 2020.

[19] D. Anghel, "A Reality Check on Trading Rule Performance in the Cryptocurrency Market: Machine Learning vs. Technical Analysis," *Financ. Res. Lett.*, p. 101655, 2020, doi: 10.1016/j.frl.2020.101655.

[20] C. Y. Chen and C. M. Hafner, "Sentiment-Induced Bubbles in the Cryptocurrency Market," no. 2017, 2019, doi: 10.3390/jrfm12020053.

[21] W. Yiying, "Cryptocurrency Price Analysis With Artificial Intelligence," *2019 5th Int. Conf. Inf. Manag.*, pp. 97–101, 2019.

[22] M. Poongodi *et al.*, "Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system ☆," *Comput. Electr. Eng.*, vol. 81, p. 106527, 2020, doi: 10.1016/j.compeleceng.2019.106527.

[23] D. Kwon, J. Kim, J. Heo, C. Kim, and Y. Han, "Time Series Classification of Cryptocurrency Price Trend Based on a Recurrent LSTM Neural Network," vol. 15, no. 3, pp. 694–706, 2019.

[24] T. A. Borges and R. F. Neves, "Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods," *Appl. Soft Comput. J.*, vol. 90, p. 106187, 2020, doi: 10.1016/j.asoc.2020.106187.

[25] T. E. Koker and D. Koutmos, "Cryptocurrency Trading Using Machine Learning," 2020, doi: 10.3390/jrfm13080178.

[26] A. Mikhaylov, "Cryptocurrency Market Analysis from the Open Innovation Perspective," 2020.

[27] Z. Hu, Y. Zhao, and M. Khushi, "A Survey of Forex and Stock Price Prediction Using Deep Learning," *Appl. Syst. Innov.*, vol. 4, no. 1, p. 9, 2021, doi: 10.3390/asi4010009.

[28] S. Mehtab, J. Sen, and A. Dutta, "Stock price prediction using machine learning and LSTM-based deep learning models," *arXiv*, 2020, doi: 10.1007/978-981-16-0419-5_8.

[29] J. Sen and E. Security, "Stock Price Prediction Using Machine Learning and Deep Learning Frameworks," no. December, 2018.

[30] M. Maheshkumar, S. Tanwar, R. Gupta, and N. Kumar, "Journal of Information Security and Applications A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions," *J. Inf. Secur. Appl.*, vol. 55, no. June, p. 102583, 2020, doi: 10.1016/j.jisa.2020.102583.

[31] Z. Chen, C. Li, and W. Sun, "Journal of Computational and Applied Bitcoin price prediction using machine learning : An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, p. 112395, 2020, doi: 10.1016/j.cam.2019.112395.

[32] A. Dutta, S. Kumar, and M. Basu, "A Gated Recurrent Unit Approach to Bitcoin Price Prediction," no. Poyser 2017, 2020.

[33] H. Sebastião and P. Godinho, "Forecasting and trading cryptocurrencies with machine learning under changing market conditions," *Financ. Innov.*, 2021, doi: 10.1186/s40854-020-00217-x.

[34] W. Chen, H. Xu, L. Jia, and Y. Gao, "Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants," *Int. J. Forecast.*, no. xxxx, 2020, doi: 10.1016/j.ijforecast.2020.02.008.