Valuation and Forecasting of Cryptocurrency: Analysis of Bitcoin, Ethereum and Dogecoin

Yang Shen[1, †], and Haoyuan Wang [2, *, †]

[1]Department of Computing, The Hong Kong Polytechnic University, Hong Kong

[2]Department of Decision Sciences and Managerial Economics, the Chinese University of Hong Kong, Hong Kong SAR, China

*Corresponding author: 1155184987@link.cuhk.edu.hk

†These authors contributed equally.

Abstract. Currently, cryptocurrency has become a research and investment topic of great concern and attracted considerable attentions in a wide range of fields. Stocks are well known as a form of investment, and there are countless studies on how its price trends. Cryptocurrencies, however, are not easily predicted due to their extremely high volatility. This paper implements the forecasting results by taking three major cryptocurrencies (i.e., Bitcoin, Ethereum, Dogecoin) as examples for the period starting from January 1st, 2020 to May 31st 2022. Four popular prediction models (XGBoost, LightGBM, GARCH, ARIMA) are applied in Python by training models and testing the prediction results. According to the analysis, XGBoost and LightGBM can forecast the future prices for all three cryptocurrencies exactly apart from the turning points when prices rise and drop suddenly. Although the predicting trends of GARCH, ARIMA have differences from the real price, they can forecast well except the unexpected situations such as COVID-19. Overall, relatively reliable and accurate prediction models can be provided for investors to apply and make wise decisions in investments based on the results of this study.

Keywords: cryptocurrency; bitcoin; Ethereum; dogecoin; valuation and forecast.

1. Introduction

It is observed that today's electronic payment system is completely dependent on the credit endorsement of third-party financial institutions to operate. What we need is an electronic cash system that does not need to trust a third party, but only through encrypted proofs, allowing both parties to make irreversible transfers [1]. Nowadays, there is evidence that digital currency has become a hot spot in the investment world and large quantities of capital flows into the market and its market value reached US $2.23 trillion on January 5, 2022. In view of the sharp rise in the market value of cryptocurrencies, they have received great attention from retail and institutional investors, regulators, and policymakers [2]. Recently, considerable research has grown up around the volatility of risk poses a series of problems for the investors. Inspired by bitcoin, other currencies (e.g., Ethereum and ripple) are also very popular in the encryption market. In the past two years, a currency has become popular because of its association with the memetic material "dog face", which is the face of Dogecoin [3]. Nowadays, the number of existing digital currencies has increased to nearly 2,000. Among them, Bitcoin and Ethereum have higher market value and these two occupy most of the digital currency market share. Dogecoin, on the other hand, has relatively low prices, but high trading volume.

A challenging problem which arises in this domain is how could we evaluate Bitcoin (BTC), Ethereum (ETH) and Dogecoin (DOGE). Based on the conclusion of evaluation, this article hopes to provide people who are planning to invest in digital currencies a more objective trading strategy. Plenty of studies have been conducted to forecast the future trend of Bitcoin (BTC), Ethereum (ETH) and Dogecoin (DOGE). Kristoufek made use of wavelet coherence to analyze the correlations between series and evaluation in time and scales, which emphasized the influence of time and frequency to bitcoin price [4]. Nematallah et al. used deep learning mechanisms like Recurrent Neural Network (RNN) and Long Short-Term memory (LSTM) to predict the bitcoin price trend. Mean Absolute percentage Error (MAPE) and Root Mean Squared Error (RMSE) are applied to estimate whether models fit digital currency data [5]. According to Pintelas [6], one can regard the forecast of digital currency price as a simple time series problem, such as ARIMA which is already used in price trend prediction. Apergis analyzed the role of the COVID-19 crisis in determining and predicting the traditional volatility return of eight cryptocurrencies through the asymmetric GARCH model in his article [7]. In this preliminary study, Zhang and Mani explored the volatility asymmetry and correlation between three popular crypto assets (bitcoin, Ethereum and dogecoin) and gold [8]. Several generalized autoregressive conditional heteroscedasticity (GARCH) models are analyzed.

As the most promising currency, digital currency has lower transaction cost than traditional financial transaction, faster transaction speed, and anonymity, which is conducive to the protection of privacy. With such advantages, digital currency has been widely discussed and more people started to trade them online currently. Besides, the price of digital currency in the market always fluctuates irregularly. To benefit more in such trading process, the most basic and straightforward strategy is to buy at a relative low price and sell at a higher price like stocks. Thus, doing research on the valuation of digital currency and developing a model to predict the future price are required.

On this basis, we decide to apply possible factors to test a variety of models, XGBoost, LightGBM, GARCH, ARIMA and choose several relatively fitting ones. Then, each model will be estimated in terms of Mean Square Error (MSE), Generalization Error (GE), and coefficient of determination (R^2). After that, the future price of Bitcoin (BTC), Ethereum (ETH) and Dogecoin (DOGE) can be predicted exactly. Besides, we will also suggest some digital currency transaction strategies to traders. The rest part of the paper is organized as follows. The Sec. 2 will talk about the data and methodologies we used in predicting future price of these three kinds of cryptocurrency, including the data retrieved from official website, variables, factors and four models. The Sec. 3 will present the forecasting results by running the data in Python, make a comparison to results in different models, explain limitations in this research, and briefly discuss our future outlooks. The Sec. 4 will summarize the full text and reach to the conclusion.

2. Data & Method

2.1 Data

This paper takes Bitcoin (BTC), Ethereum (ETH) and Dogecoin (DOGE) as the research objects and selects the daily data on Yahoo Finance as the sample. Because Dogecoin (DOGE) has started a large number of transactions in recent two years, the time period is from January 1, 2020, to May 31, 2022. This paper will use machine learning model and time series model to fit relevant data and factors at the same time and take the data from June to July 2022 as the measurement standard of model prediction accuracy.

Table. 1 Factor classification

Serial number	Type	Factor
1	Trends Indicators	SMA
2		MACD
3	Interval breakthrough Indicators	BBands
4		ATR
5	Energy Indicators	CCI
6	Other indicators	MFI

2.2 Variable description and Factor construction

GARCH model measures volatility through variables such as opening price, closing price, daily maximum price and minimum price, while ARIMA model requires that a stable curve be obtained through the time series measurement of the original data of the above currencies to predict future data. The key of multi factor model and machine learning model is to construct characteristic factors. Due to the highly autonomous nature of virtual currency, it is impossible to analyze its fundamental information, and only factors can be mined from transaction data. This paper constructs six single factors from trend, interval breakthrough, energy, etc. as a factor pool. On the one hand, these six factors have practical economic significance, on the other hand, the dimension complementarity between factors reduces the information deviation of single factors. Table. 1 lists the factor classification. As can be seen from the Figs 1-3, the correlation coefficient between the selected factors and the closing prices of the three currencies is very high.

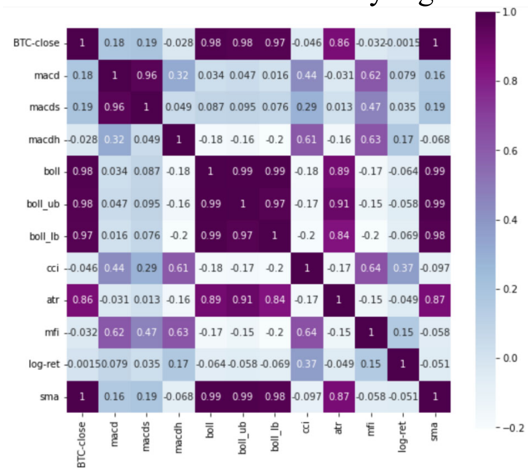

Fig. 1 BTC-Indicators Heatmap.

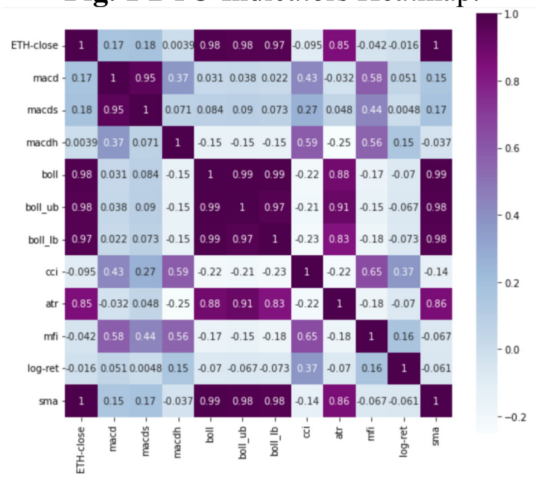

Fig. 2 ETH-Indicators Heatmap.

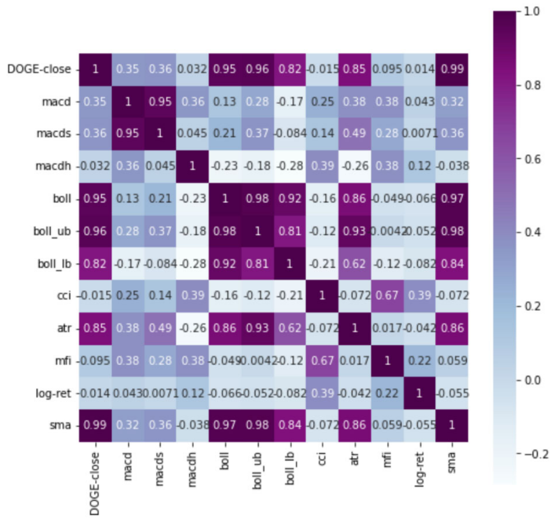

Fig. 3 DOGE-Indicators Heatmap.

2.3 Models

LightGBM (Light Gradient Boosting Machine) is a model to implement GBDT (Gradient Boosting Decision Tree) algorithm. The main idea for it is to utilize decision tree, a weak classifier to process iterative training to get an optimal model. It supports efficient parallel training, and has the advantages of fast training speed, low memory consumption, good accuracy, distributed support, difficult to overfitting and rapid processing of massive data.

One of the most famous GBDT tool for analyzing was XGBoost, a decision tree algorithm based on a pre-sorting method, before the introduction of LightGBM. There are some wonderful advantages for using XGBoost model, such as high accuracy and flexibility, regularization, shrinkage, column subsampling, missing value handling, parallelized operations. XGBoost is an additive expression consisting of ***k*** base models as:

$$\overline{y_i} = \sum_{i=1}^{k} f_i(x_i) \tag{1}$$

The loss function can be expressed by the prediction value $\overline{y_i}$ and the true value y_i, where ***n*** is the number of samples:

$$L = \sum_{i=1}^{n} l(y_i, \overline{y_i}) \tag{2}$$

The prediction accuracy of the model is jointly determined by the deviation and variance. The loss function represents the deviation of the model. With the purpose of getting a small variance, a simple model is required like:

$$Obj = \sum_{i=1}^{n} l(y_i, \overline{y_i}) + \sum_{i=1}^{k} \Omega(f_i) \tag{3}$$

Here, Ω is the regular term of the model.

ARIMA (AutoRegressive Integrated Moving Average) is a time series model for forecasting and analyzing data. It is widely used in the prediction of stationary data [9, 10]. ARIMA (p, d, q) consists of three parts:

- AR(p): AR (AutoRegressive) means the current value and regressions of past value are equivalent. Since it doesn't rely on other explanatory variables other than its own historical value, it is called autoregression. If rely on the recent ***p*** historical value in the past, the order is ***p*** and model is denoted as ***AR(p)***.
- I(d): I (Integrated) refers that the ARIMA model makes difference to the time series. For time series analysis, unstable series should be converted to be stable to maintain stationarity. ***D*** represents the order of difference. The difference between two values in adjacent moments is a new time series, which is named first-order difference series. The first-order sequence of the first-order difference sequence is called the second-order difference sequence, and so forth. Besides, there is also a seasonal difference ***S***, which reflects the seasonal difference sequence in a period of ***T***.
- MA(q): MA (Moving Average) indicates that the current value is equivalent to the regression of the prediction error in the past. Prediction error is equal to the difference between the model prediction value and the true value. When it depends on ***q*** prediction error, the order is ***q*** and the model is denoted as ***MA(q)*** [11].

ARCH model (autoregressive conditional heteroskedasticity model) is called "autoregressive conditional heteroskedasticity model", which solves the problem caused by the second assumption (constant variance) of time series variables in traditional econometrics. GARCH model is called generalized arch model, which is an extension of arch model and developed by Bollerslev. ARCH model can accurately simulate the change of the volatility of time series variables. It is widely used in the empirical research of financial engineering, so that people can more accurately grasp the risk (volatility), especially in the value at risk.

2.4 Metrics

To measure the model, five kinds of popular metrics (MAPE, MAE, R^2, MSE, RMSE) are applied. MAPE (Mean Absolute Percentage Error) is a statistical measurement to predict the accuracy of a forecasting method. MAE (Mean Absolute Error) is a measure of error between the actual price and predicted price. R^2 (Coefficient of Determination) is used to measure how better the model fits. The closer it is to 1, the better the model fits. MSE (Mean Squared Error) and RMSE (Root Mean Squared Error) have similar meanings as MAE. The formulae are given as follows.

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{actual_i - predict_i}{actual_i}\right|}{n} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n}|actual_i - predict_i|}{n} \tag{5}$$

$$R^2 = 1 - \frac{(predict_i - actual_i)^2}{\left(\overline{actual} - actual_i\right)^2} \tag{6}$$

$$MSE = \frac{\sum_{i=1}^{n}(predict_i - actual_i)}{n} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(predict_i - actual_i)}{n}} \tag{8}$$

3. Results & Discussion

3.1 Xgboost

In XGBoost, the prediction results of Bitcoin, Ethereum and Dogecoin measured by MAPE, MAE, R^2, MSE and RMSE are shown in Table. 2. R^2 of three cryptocurrencies are all close to 1 and MAPE are close to 0, which shows that the XGBoost model fits them well and can be used to forecast their future prices accurately.

Table. 2 Prediction Results Measured by Metrics through XGBoost.

	MAPE	MAE	R^2	MSE	RMSE
Bitcoin	0.0151	472.2422	0.9984	579978.0780	761.5629
Ethereum	0.0231	1.5931	0.9984	2717.6282	52.1309
Dogecoin	0.0450	0.0048	0.9861	0.0002	0.0146

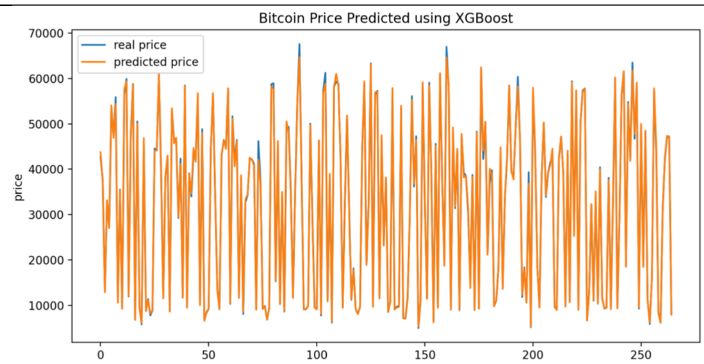

Fig. 4 Bitcoin Price Predicted using XGBoost.

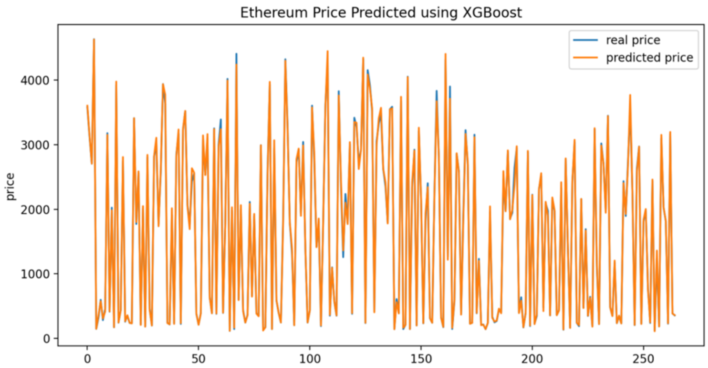

Fig. 5 Ethereum Price Predicted using XGBoost.

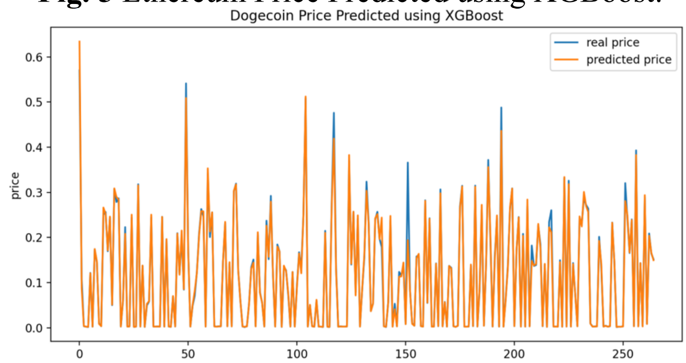

Fig. 6 Dogecoin Price Predicted using XGBoost.

Seen from th Figs. 4-6, the blue line refers to the real price while the yellow line represents the predicted price. The trends of those two lines are similar, which means it fit well for bitcoin, Ethereum and dogecoin. Thus, XGBoost is a reliable model to predict the future price of these three digital currencies.

3.2 Page Numbers

In LightGBM, the prediction results of Bitcoin, Ethereum and Dogecoin measured by MAPE, MAE, R^2, MSE and RMSE are shown in Table. 3. R^2 of three cryptocurrencies are all close to 1 and MAPE are close to 0, which shows that the LightGBM model fits them well and can be used to forecast their future prices accurately. As shown in Figs. 7-9, the blue line refers to the real price while the yellow line represents the predicted price. The trends of those two lines are similar, which means it fit well for bitcoin, Ethereum and dogecoin. Thus, LightGBM is a reliable model to predict the future price of these three digital currencies.

Table. 3 Prediction Results Measured by Metrics through LightGBM.

	MAPE	MAE	R^2	MSE	RMSE
Bitcoin	0.0238	641.4682	0.9972	1008460.7821	1004.2215
Ethereum	0.0333	40.5865	0.9979	3643.8484	60.3643
Dogecoin	0.0416	0.0072	0.9604	0.0006	0.0246

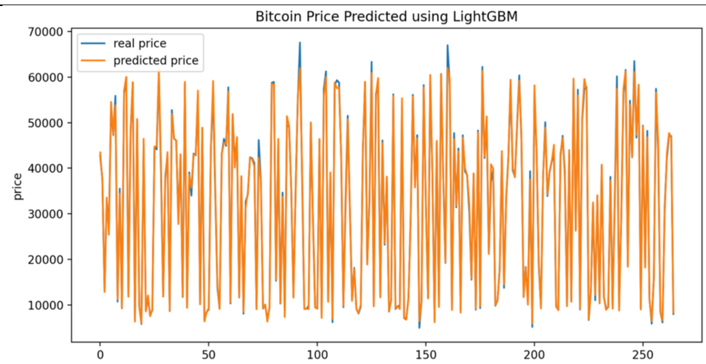

Fig. 7 Bitcoin Price Predicted using LightGBM.

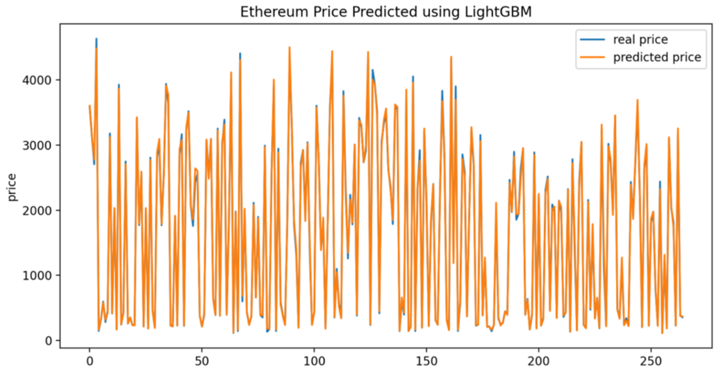

Fig. 8 Ethereum Price Predicted using LightGBM.

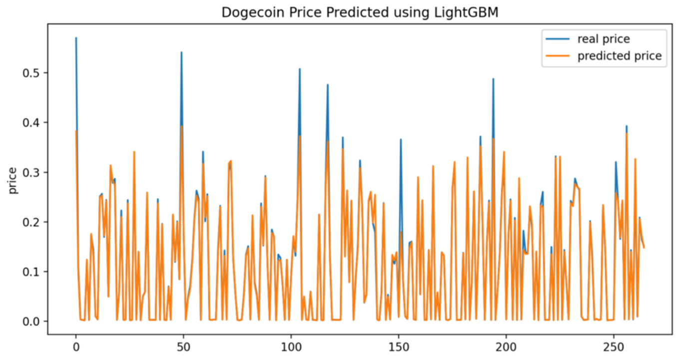

Fig. 9 Dogecoin Price Predicted using LightGBM.

3.3 Arima & GARCH

GARCH is a model for mapping volatility, and only raw data can be used. The mapping results are found to be not fit, but they are consistent with the volatility in early 2020. Therefore, it is speculated that the increasingly serious global spread of the epidemic has affected the trend of virtual currency. Similarly, Arima is an autoregressive model. We trained ARIMA (6, 1, 7) model, but the fitting result is still not ideal. Therefore, we can reasonably speculate that the current model is not fit enough because of the epidemic situation. In conclusion, under the influence of the global epidemic situation, the multi factor model rather than the simple time series model can better fit the virtual currency price prediction, it is more convenient for us to find suitable models and give investors suggestions.

3.4 Comparison

According to the analysis by adding some relevant factors to establish a machine learning model, the fitting and prediction results are relatively accurate, while the conventional time series model is difficult to establish a model with good prediction effect mainly due to the high volatility of virtual currency. Thus, in the future research, we can build more machine learning models to predict the virtual currency.

3.5 Limitations & Prospect

In fact, this study has some shortcomings and defects. Primarily, the number of factors may be insufficient and the number of virtual currencies used is too small. Besides, the situation under the influence of the epidemic situation is not considered. Therefore, in the future research, we will try to add some factors under the influence of the epidemic to the machine learning model, and at the same time use more data of virtual currency types to establish multiple models, so as to better compare the differences between virtual currencies so that investors can better choose suitable virtual currencies for investment.

4. Conclusion

In summary, this paper investigates the valuation of cryptocurrency based on the case study of Bitcoin, Ethereum and Dogecoin. Specifically, four prevailing prediction models (XGBoost, LightGBM, ARIMA and GARCH) are adopted in Python to forecast the future price values of three cryptocurrencies. According to the analysis, prediction results of cryptocurrencies through XGBoost and LightGBM are the most accurate, except the turning points. These two models can be applied in any situations for forecasting. Besides, since the breakout of pandemic, there are much uncertainty occurred while forecasting by utilizing ARIMA and GARCH. In addition, the number of factors may be insufficient. Moreover, the number of virtual currencies used is too small. Nevertheless, the situation under the influence of the epidemic situation is not considered. In the future, more factors under the influence of the epidemic ought to be taken into consideration in machine learning model, and at the same time use more data of virtual currency types to establish multiple models. Overall, these results offer a guideline for investors to better choose suitable virtual currencies for investment.

References

[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Decentralized Business Review, 2008: 21260.

[2] Akhtaruzzaman M, Boubaker S, Nguyen D K, et al. Systemic risk-sharing framework of cryptocurrencies in the COVID–19 crisis. Finance research letters, 2022: 102787.

[3] Agarwal B, Harjule P, Chouhan L, et al. Prediction of dogecoin price using deep learning and social media trends. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 2021, 8(29): e2-e2.

[4] Kristoufek L. What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. Plos one, vol. 10, pp: 1-15, 2015.

[5] Nematallah H F, Sedky A A H, Mahar K M. Bitcoin Price Trend Prediction Using Deep Neural Network. Webology (ISSN: 1735-188X), 2022, 19(3).

[6] Emmanuel G. Pintelas. Deep Neural Networks for Bitcoin Price Prediction, Reseach gate preprint, 2020.

[7] Apergis N. COVID-19 and cryptocurrency volatility: Evidence from asymmetric modelling[J]. Finance Research Letters, 2022, 47: 102659.

[8] Stephen Zhang and Ganesh Mani. Popular cryptoassets (Bitcoin, Ethereum, and Dogecoin), Gold, and their relationships: volatility and correlation modeling. Data Science and Management, 2022, vol. 4, pp: 30-39.

[9] Azari A. Bitcoin price prediction: An ARIMA approach. arXiv preprint arXiv:1904.05315, 2019.

[10] Hoerl A E, Kennard R W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970, 12(1): 55-67.

[11] Poongodi, M., Vijayakumar, V., & Chilamkurti, N. Bitcoin price prediction using ARIMA model. International Journal of Internet Technology and Secured Transactions, 2020, vol. 10, pp: 396.