

Autoregressive Integrated Moving Average Model based Prediction of Bitcoin Close Price

Anupriya
Dept of CSE
BIT, Mesra
Ranchi, INDIA
apriya@nvidia.com

Shruti Garg, Assistant Professor
Dept of CSE
BIT, Mesra
Ranchi, INDIA
gshruti@bitmesra.ac.in

Abstract--- Analysis of any digital currency is performed for identifying and quantifying uncertainties, estimating their impact on results with real-time market value. Cryptocurrency, as an encrypted form of currencies which is used for shopping, investment, money transfer and in other purpose now days. Most popular use of Bitcoins are investment because its price was unexpectedly high in past few years (data shown in content of paper). In this paper prediction of Bitcoin close price by using the ARIMA model has been performed. The ARIMA model is found suitable for the prediction of bitcoin prices because this model is used for prediction of time series data. The forecast of future values is provided based on seasonality and trend present in the price data. In terms of visualizations, results are manifest by using R programming language. The obtained results are then compared with actual prices and percent mean error is calculated. The present mean error is found here less than 6% for most of the values.

Keywords--- Bitcoin; Time Series; Forecasting; Analysis; ARIMA; R

I. INTRODUCTION

Cryptocurrencies are the online digital or virtual currency with economic attributes generated to be used as a way of trade. Today it becomes a global circumstance familiar to most people. Whereas still some way geeky and not understand by most people, banks, governments and numerous companies are conscious of its significance. Cryptocurrency work with cryptography to secure and authenticate negotiation as well as to control the establishment of a new component of a certain cryptocurrency.

In early 2008, *Satoshi Nakamoto* introduced the first most familiar and common cryptocurrency known as Bitcoin which is the original cryptocurrency among all. It was launched in 2009 as a decentralized “electronic payment system”. It can be sent or received from a channel of transfer protocol i.e., client to client as well as peer-to-peer. The first bitcoin was used by a Florida programmer for purchasing a pizza in May 2010. Also the price of Bitcoin were above

\$10000 dollar in late 2017. That is why Bitcoin started commanding the trade of digital currencies, representing 56.83% of the total market capitalization (coinmarketcap.com)

Blockchain (BC), the innovation behind Bitcoin cryptographic money framework, is fundamental for shaping the spine for guaranteeing improved security and protection for different applications in numerous different spaces including the Internet of Things (IoT) biological system. Universal research is at present being led in both scholarly community and industry applying Blockchain in differed spaces. The Proof-of-Work (PoW) numerical test guarantees BC security by keeping up an advanced record of exchanges that are thought to be unalterable. Furthermore, BC utilizes an alterable Public Key (PK) to record the clients' character that gives an additional layer of protection. The effective appropriation of BC has been actualized in assorted non-money related frameworks, for example, the in internet voting, the decentralized informing cona veyed distributed storage frameworks, confirmation of-area, social insurance et cetera. Decentralization allows the blockchain technology to have increased capacity, better security, and faster settlement. [1]

Cryptocurrency implemented by the theory of solving encryption algorithms to create unique hashes that are finite in number. The opening of unbelievable exploratory cryptographic forms of money known as altcoins. [2]

There are many ways of earning money through cryptocurrency, one of the safe ways to earn money in cryptoland is “Buying and Holding”. Where buying a good cryptocurrency that includes a crucial utilize case and holds them until they get a reasonable market share. For example, cryptocurrencies like Bitcoin, Ethereum, Litecoin, Monero etc. So, in this paper,

we are performing the prediction of Bitcoin because at present a lot of media attention is there, and its total market value has reached 20 billion USD in March 2017. More importantly, several central banks began as of late to investigate the appropriation of cryptocurrency and blockchain innovations for retail and large-value installments. For example, the People's Bank of China aims to develop a worldwide digital currency based on blockchain technology. [3] Bitcoin is one of the most popular cryptocurrency since 2009, the market price of Bitcoin increases from an essential value of \$0 to a value \$6,386.07 at the time of this composing.

In this paper, Bitcoin close price was selected for prediction because most of the investor is interested in closing price. ARIMA model is used for forecasting because it is recommended by the literature that the Time Series data prediction using the ARIMA model could give superior results than the forecasting directly[4]. Autoregressive Integrated Moving Average (ARIMA) model was introduced in 1979 by Box and Jenkins and it is used to transform the non-stationary data to stationary data. Box and Jenkins referred a technique to composed a set of exercises for distinguishing, evaluating and diagnosing ARIMA models with time series data. Long-standing time esteem of a variable could be a straight mixture of past values and past errors which is shown in the ARIMA model as follows:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (1)$$

Where, Y_t is an actual value, ϵ_t is a random error at 't', ϕ_i and θ_j are coefficients, p and q are integers that are often referred to as autoregressive and moving average, respectively. (Ariyo, Adewumi, & Ayo, 2014) In the ARIMA model, assuming the upcoming value of the variables is depend upon autoregression, integration, and moving average respectively which is denoted as ARIMA(p,d,q). Deciding the value of p,d, and q is considering a vital role for constructing the operation of ARIMA model. It also regularly rehashed several times until didn't final the satisfactory model. [6]

R is one of the programming languages which was invented by Ross Ihaka and Robert Gentleman in the year 1993. It possesses a graphical and statistical method. For R language, R-Studio may be a free and effective coordinates development environment which permits the client to execute the ARIMA Model. R-

Studio permits the client to run R scripts in a convenient domain. [7]

In this paper, the prediction of Bitcoin close price has been performed for 534 days. ARIMA model is applied by using R language in R studio. Then the accuracy of forecasted price has been judged by comparing forecasted price with the actual price. For the comparison of data, it has been divided into training and test data by 60:40 ratio respectively. Then mean percentage error has been calculated. The next section2, describes previous work done in prediction of stock data. Section 3 tells about methodology and section 4 showing results found from experiment. The last section 5, is the conclusion of this work.

II. LITERATURE REVIEW

There are different types of analysis done on cryptocurrencies like statistical analysis, empirical analysis, SWOT(Strength, Weakness, Opportunity, and Threat) analysis, and time series analysis.

One of the researchers focuses on the statistical analysis of the top seven type of cryptocurrencies i.e., Bitcoin, Ripple, Litecoin, Monero, Dash, MaidSafeCoin, and Dogecoin. The investigation and results for the exchange rate of Bitcoin when contrasted with the U.S. Dollar by utilizing a noncausal autoregressive process with Cauchy mistakes. The results demonstrate that the delivered Bitcoin/USD conversion scale indicates nearby patterns which shows a time of theoretical conduct from web-based exchanging. [8]

The thesis is also done on a model of the market price of cryptocurrencies by using statistical analysis and they modeled the data by linear regression. It is a method to model a bond between a dependent variable and a lay of independent variables. In this analysis, their goal is to decrease the complexity of the model by removing insignificant variables and receive the best possible model. Their results reflect the price of cryptocurrencies which depends stiffly on the search traffic on the particular cryptocurrency designation on Google's search engine. [9]

The empirical data analysis of Cryptocurrencies was done in [10] and [11] exhibits long memory, leverage, stochastic volatility, and heavy-tailedness. Phillip et. al. did the analysis of cryptocurrency to find out investability, they found t-distribution suitable for 224 different cryptocurrencies. The vector autoregressive (VAR) model was applied in [11]. They use the Vector Error Correction (VEC) model which is based on the VAR model to examine the short-run properties of the cointegrated series. They select VEC because it takes into a report of any cointegrating relation among the

variables, In their paper, they consider Bitcoin more as a resource Bitcoin cost can be affected beneath the macroeconomic list and imperative resources cost record, in another way they said Bitcoin isn't as it was driven by its possess request and supply. In a credit money, the esteem can as it was being driven by its supply and request, from this point of see, Bitcoin is presently distant from getting to be a genuine money.

The SWOT analysis gives strengths, weaknesses, opportunities, and threats of bitcoins [12] [13]. As strengths one can say it is best performing cryptocurrency in last few years and gaining popularity very fast. It doesn't have many weaknesses but it is difficult to see it wallets as well as one block is shared to multiple user, thus it may lead to some attack. The biggest opportunity for bitcoin is it can act as gold in future. It is having very few hurdles among them biggest hurdle is user acceptance, although bitcoin became very popular these days, even it is not equally popular as physical currency among investor.

In time series regression analysis, use the dependent and the independent variables in their natural logarithm form. Their objective is to discover signs with respect to both short-run and long-run flow between the factors. In this analysis, they conducted market price of Bitcoin which is displayed by Google trends in the weekly analysis (163 observation) and Wikipedia views in the daily analysis (1139 observations). A time series is a covariance stationary when its expected distribution is secure over time, it has a constant mean, a constant variance, and the covariance of two dissimilar points in the sequence depend only on the step among them. [14] Time series is additionally valuable to imagine it as an item of amassing distinctive layers, subsequently, the method of decomposing each layer gives an appealing strategy to bring a coordinate person elucidation to the model. [15]

One of the researchers performed analysis on 'Stock price prediction using ARIMA model' and they have concluded their research that they capture the best result on a short-term basis to satisfy the stock price by demonstrating the potential of ARIMA model. [5] They also applied data mining techniques in R for prediction of stock market data and the conclusion of their research was ARIMA model gives very accurate prediction for short term data.

After going through literature, it has been found that the ARIMA model with the help of R code is suitable for forecasting of digital currency i.e., Bitcoin used in current paper. In this paper, we applied ARIMA

model for predicting the close price of Bitcoin by using previous data of Bitcoin price from 2015-2018.

III. METHODOLOGY

In this paper, the method used to evolve the ARIMA model for Bitcoin close price forecasting is described below in details.

R studio of version 3.5.1 is used to implement the R programming language. In terms of obtaining the solution for prediction of values historical as well as current data of Bitcoin price (coindesk.com) is collected as a primary step, which covers the interval from 1st January 2015 to 23rd September 2018 with the total number of observations i.e., 1362 shown in figure 1. Data contains four columns: open price, low price, high price, and close price in terms dollars respectively. The close price was selected for prediction because it depicts the enterprises of the index in a day of business.

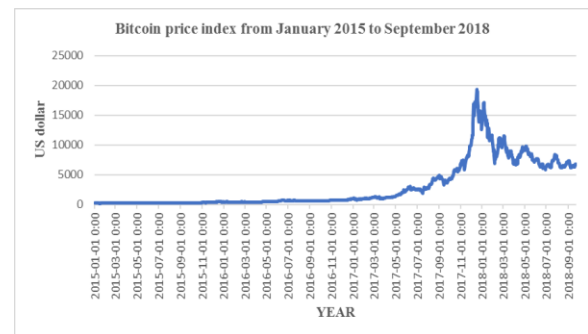


Fig. 1. Market close price value of Bitcoin from January 2015 to September 2018

The ARIMA (p,d,q) model is used to process the data. where-

- p = order of the autoregression
- d = degree of first difference
- q = order of the moving average

It is difficult to select suitable value of p, d and q, so "auto.arima()" function of R programming was used to find out these terms automatically.

- Auto Regression (AR) – Denoted with "p"
The future values which are based on the foregoing value are estimated by autoregression technique. Where AR(1) first-order process of an autoregression, means the successive value will depend on present value. In AR(2), the present value depends upon the foregoing two values.

- Moving Average (MA) – Denoted with “q”
Inside a data set, moving average technique is used to observe the general idea. It generally predicts the future value which depends upon the foregoing values.

The data of Bitcoin close price was divided into two parts, the first one is training data and the second one is testing data, which is taken as the ratio of 60:40. In which training data is used to calculate the log and differentiate the variables. Then Autocorrelation Function(ACF) and Partial Autocorrelation Function(PACF) was applied with the lag value 20 to find correlation in data. It also helps to check whether the data is stationary or not. Then the dicky fuller test was applied by using “adf.test()” function to convert the non-stationary time series into stationary.

The Dickey-Fuller test checks the null hypothesis i.e., a unit root which displays in an autoregressive model. The alternative hypothesis i.e., a variable which was created by a stationary handle. This paper did alternative hypothesis for the dicky fuller test which includes the lag value i.e., lag(9) and did differentiate the variables in the regression. If the value of ‘p’ is found as more than .5 then hypothesis will become null and series will declared as non-stationary. Here the p-value for instock and differentiate of instock is an alternative hypothesis which becomes stationary, is showed in the figure 2. below-

```
> adf.test(instock)

Augmented Dickey-Fuller Test

data: instock
Dickey-Fuller = -3.6128, Lag order = 9, p-value = 0.03145
alternative hypothesis: stationary

> adf.test(diffinstock)

Augmented Dickey-Fuller Test

data: diffinstock
Dickey-Fuller = -9.1986, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Fig. 2. Screenshot of the dicky fuller test.

The model of ARIMA(p,d,q) shows much values of p,d, and q respectively shown in table 1. In present case we have got values of p,d,q as (0,1,0), (2,0,0) and frequency 24 with a drift, which represent ARIMA model “Random walk with drift”, Obtained actual data are stored in temporary memory of ‘fitinstock’. The actual data is differentiated in R code for forecasting the new values.

TABLE 1: SPECIAL CASES OF ARIMA MODEL

SPECIAL CASES	ARIMA VALUES
White noise	ARIMA(0,0,0)
Random walk	ARIMA(2,1,2)
Random walk with drift	ARIMA(2,0,0)
Auto regression	ARIMA(p,0,0)
Moving average	ARIMA(0,0,q)

Now calculate the percentage error by using the actual price and forecasted prices (the real value which is taken for the calculation) and forecasted price (obtained value using the ARIMA model).

The formula to calculate the percentage error is given below:

$$\text{Percentage error} = \frac{\text{Forecasted Price} - \text{Actual Price}}{\text{Actual Price}} \times 100 \quad (2)$$

And lastly, the mean percentage error is calculated by using percentage error. The formula to calculate the mean percentage is given below:

$$\text{Mean percentage error} = \frac{1}{n} \text{percentage error} \quad (3)$$

The ‘n’ denotes the total number terms in series.

IV. RESULTS AND DISCUSSION

In the forecasting process, future data is predicted which is a bargain upon previous and current data. For the evaluation of future Bitcoin close price, miscellaneous prediction methods were used in past. There use time series prediction with ARIMA model which predicting the future data based on previous and current data by using ‘forecast’ package. Here is the graph shown in figure 3 shows the forecasted price for 545 days. Where the x-axis shows the number of days taken for forecasted the values and the y-axis shows the price in terms of the US dollar.

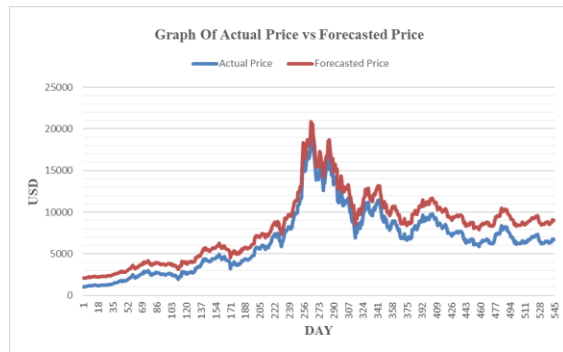


Fig. 3. Graph of Actual Price vs Forecasted Price

After getting the values of actual price and forecasted price, we calculated the percentage error by using formula given in equation no 2 and the mean percentage error is calculated by the formula given in equation 3, which is found less than 6% for most of the values. Here, figure 4. Shows the percentage error for all the values. Where x-axis shows the total number of days taken to forecasted and y-axis shows the error percentage.

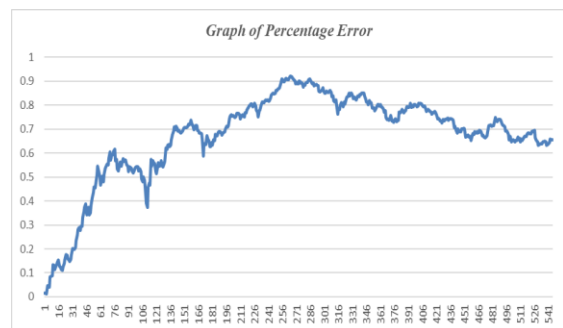


Fig. 4. Graphical representation for the mean Percentage error

REFERENCES

- [1] D. Kuo, C. Lee, and Y. Wang, "Cryptocurrency : A new investment opportunity ? Cryptocurrency: A New Investment Opportunity?," vol. 20, no. 3, pp. 16–40, 2018.
- [2] Y. Sovbetov, "M P RA Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litecoin, and Monero Factors Influencing" *J. Econ. Financ. Anal.*, vol. 2, no. 2, pp. 1–27, 2018.
- [3] J. Chiu and T. V. Koepl, "The Economics of Cryptocurrencies – Bitcoin and Beyond," *Ssrn*, pp. 1–40, 2017.
- [4] J. Kamalakannan, I. Sengupta, and S. Chaudhury, "Stock Market Prediction using Time Series Analysis," pp. 1–5.

Graph in figure 5. shows the forecasted price from 2018 to 2020 by using ARIMA(p,d,q) model. The suitable fitted value of p,d, and q i.e., ARIMA(2,1,2)(2,0,0) with drift and their frequency is taken as 24. In this graph, the x-axis shows the year and y-axis show the US dollar.

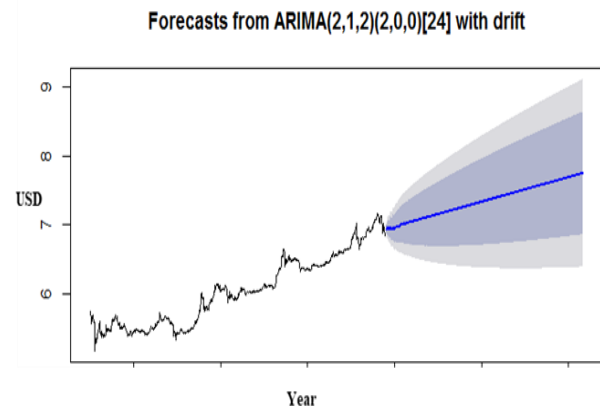


Fig. 5. Forecasted graph using ARIMA(p,d,q)

V. CONCLUSION

There are different researchers performed different kind of analysis for predicting the Bitcoin close price. The ARIMA model was used to predict closed price in this paper. For checking of the accuracy of ARIMA model the mean percentage error has been calculated after comparing the actual price and the forecasted price. The obtained result is less than 6% for most of the values. Then mean percentage error was calculated which which shows ARIMA gives accuracy around 60-70%. In future the same ARIMA prediction model can be used for prediction any other time series data which passes dicky fuller test i.e. the p-value must be lesser than .5.

- [5] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," *2014 UKSim-AMSS 16th Int. Conf. Comput. Model. Simul.*, pp. 106–112, 2014.
- [6] J. J. Wang, J. Z. Wang, Z. G. Zhang, and S. P. Guo, "Stock index forecasting based on a hybrid model," *Omega*, vol. 40, no. 6, pp. 758–766, 2012.
- [7] Angadi, M., Kulkarni, A.: Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R. International journal of advance research in computer science. 6(6), 1–5, (2015).
- [8] J. Osterrieder, S. Chan, J. Chu, and S. Nadarajah, "A Statistical Analysis of Cryptocurrencies," *Ssrn*, no. February, pp. 1–23, 2017.
- [9] E. Pärland and O. Rydén, "Explaining the market price of Bitcoin and other Cryptocurrencies with Statistical Analysis," 2015.

- [10] A. Phillip, J. Chan, and S. Peiris, "A new look at Cryptocurrencies," *Econ. Lett.*, vol. 163, no. November, pp. 6–9, 2018.
- [11] Y. Zhu, D. Dickinson, and J. Li, "Erratum to: Analysis on the influence factors of Bitcoin's price based on VEC model," *Financ. Innov.*, vol. 3, no. 1, p. 7, 2017.
- [12] P. D. Devries, "An Analysis of Cryptocurrency, Bitcoin, and the Future," *Int. J. Bus. Manag. Commer.*, vol. 1, no. 2, pp. 1–9, 2016.
- [13] T. E. D. J. May, "A Detailed SWOT Analysis Example (Applicable To All Industries)," 2018.
- [14] G. Kavvadias, "What Drives the Value of Cryptocurrencies ? A Time Series Analysis of Bitcoin," no. November, 2017.
- [15] Poyser & Perdiguero, "Exploring the determinants of Bitcoin's price: an application of Bayesian Structural Time Series, " June, 2017.