**Bitcoin Price Prediction using ARIMA Model**

**KAVLEEN KAUR**

**DR. BEBONCHU ATEMS**

### **Clarkson University, 8 Clarkson Avenue, Potsdam, NY, 1369**

**A B S T R A C T**

The paper focuses on predicting the price of Bitcoin using the ARIMA and the Prophet model by Facebook. The paper demonstrates the implementation and the effectiveness of both the models stated above while predicting the price of Bitcoin. Bitcoin being a decentralized cryptocurrency, possed an altogether difficult challenge as the price of all cryptocurrencies is volatile. In addition, the paper also illustrates the market character of Bitcoin by also forecasting the volume of Bitcoin. Together, the predicted price and the volume can help us evaluate the predicted market character of Bitcoin. In the end, the paper would compare the two models and finally unveil the better model.

**1. INTRODUCTION**

Bitcoin is the first digital currency that uses peer-to-peer technology to facilitate financial transactions based on Blockchain technology. It is commonly abbreviated as “BTC” and was created in January of 2008 offering the promise of lower transaction than any traditional online payment mechanism. It has witnessed a wider adoption and attention than any other cryptocurrency out there. One reason for such a broad adoption was its promise for lower transaction cost and the idea of the currency being a decentralized currency, meaning that the currency is not government-issued and is not operated by any decentralized authorities including banks. Gradually, Bitcoin was started to being used to buy merchandise anonymously. It helped bigger industries to overcome international transaction costs as Bitcoin was not subjected to any regulations. Smaller industries and businessmen liked the currency for smaller reasons like avoiding credit card fees. Thus, a rapid increase in price especially in 2017, and also its limited availability has made Bitcoin a store of value for investors.

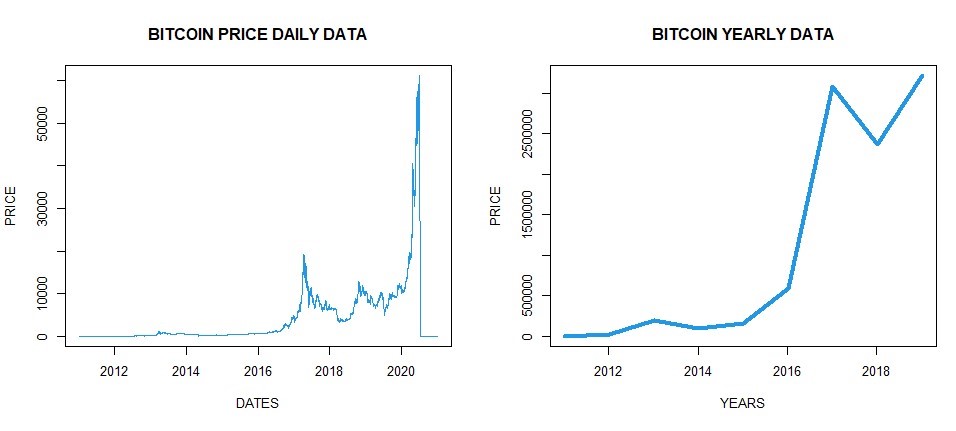
Bitcoin system is a collection of nodes or miners that run Bitcoin’s code and stores its blockchain. A blockchain is a collection of blocks which in turn are collections of transactions. Because all computers (nodes or miners) have the same transaction blocks, no one can defeat the system. Balances of Bitcoin are kept using public and private keys. The public keys are comparable to a bank account number while a private key is comparable to a secret ATM PIN. Bitcoin even has a market exchange called ”bitcoin exchange” that allows people to buy and sell different currencies. Coinbase is a leading exchange along with Bitstamp. Like any other currency, Bitcoin can also be transferred digitally using wallets. Bitcoins are not physical in nature. They are mined using computers where people solve complex math problems to mine bitcoin. Currently, a winner is rewarded with 12.5 billion roughly every 10 minutes.

Bitcoin is now proving to be a real revolution for the financial industry. JPMorgan Chase & Co is the latest Wall Street firm floating the idea of investors using Bitcoin as a way to diversify their portfolios. It has surged fivefold in the past as prominent investors are announcing huge investments responsible for purchasing Bitcoin. Thus, the fact that Bitcoin is surpassing the profits gained from the stock market for the last five years is not at all surprising. With the increase in the popularity of bitcoin, a theory has been arising in the market and that is whether investing in Bitcoin is a good idea or not. While it’s a common belief in the investors that bitcoin can be used as a hedging instrument, some economists even have raised their opinions stating that it can be used as a hedge against inflation too.

Thus, despite its rapidly changing nature, the price of Bitcoin has been an area of interest for all researchers and they have used all the available resources and knowledge to forecast the price of Bitcoin. An issue associated with forecasting the price of bitcoin is its price volatility. On November 20th, 2015, Bitcoin (BTC) got indexed as other stocks. From the very first day constantly growing and was worth 327.00 USD. It reached its highest peak on December the 15th, 2017. The BTC value was 19’650.01. The market and even professional investors went crazy. In December 2018, the Bitcoin price chart dropped significantly. Many people lost 85% or even more of their capital invested in Bitcoins. BTC was now worth a bit more than $3,000. In June 2019, the Bitcoin price chart raised to another mountain peak. The renewed rise has surprised all skeptics. But at $11,865 the zenith was reached again. Since then, the Bitcoin price had leveled around $7 to $8k. We can see it in Fig 1(a),(b). The statistics predicted the value for bitcoin to rise in 2021. BTC always regains its value after a significant drop as it is quite resilient even when the market faced high uncertainty duringCOVID pandemic.

To predict the price of the bitcoin, the data is collected from an API of the website quadln.com which mainly contains the closing price, the opening price, and the volume of BTC during a particular date. The data is collected daily from October of 2011 to March of 2021. The data also contains the high and the low price of the day. It also has data containing the volume of currencies combined which provided us with the weighted price of the day.

For the ARIMA model, the data collected for the price and volume is made stationary by taking the log differencing and is made seasonally adjusted by removing the seasonal element from it. After making it stationary and seasonally adjusted, several models were applied to find models with minimum AIC value. The models found were then used for forecasting future price and volume. For the Prophet model by Facebook, we do not make the variables stationary or make it seasonally adjusted because Prophet detects the following trend and seasonality from the data first, then combine them to get the forecasted values. Prophet also requires a fixed format of the data frame. The format that Prophet is a ‘ds’ column that contains the date field (in our case date) and a ‘y’ column that contains the value we are wanting to model/forecast.( closing price and volume of the bitcoin). The prophet is thoughgood at forecasting the data efficiently, it did not outperform the ARIMA model.



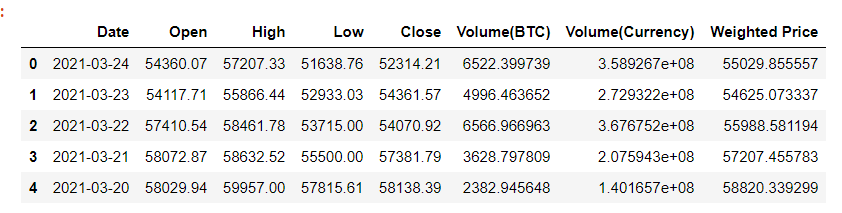
**Fig1(a) Fig1(b)**

**2. LITERATURE REVIEW**

The popularity of Bitcoin had reached its peak in 2017 and due to the steep and sudden rise of its demand, it caught the attention of many researchers that they could predict the price of bitcoin using various models. Our focus in this research is to predict the price of Bitcoin using the ARIMA and the Prophet model. Researchers have already researched a wide variety of models. ARIMA has been widely used in the literature to predict the price of the bitcoin along with many machine learning models such as RNN which also uses data on Twitter’s hashtag. These models outperform the ARIMA model. Thus, our paper uses another such model called the Prophet model by Facebook to see if it outperforms the ARIMAmodel. No one has covered the prophet model before in the paper. Many researchers have also stated that due to the high volatility of the price of bitcoin, the models tend to have larger MSE. Some researchers have also stated recently that bitcoin is resilient enough to bounce back after a significant drop in times of high uncertainity like in the COVID pandemic because of which is why we have data from the years 2021 and 2020 is even included in this study. Apart from this, studies also have been done on whether taking log differencing is beneficial or not. The results are ambiguous so, here we have taken the log differencing. No paper has yet studied the market character of Bitcoin which we will cover in this paper by also predicting the volume of Bitcoin.

**3. DATA ANALYSIS:**

The data is collected from an API of the website quadln.com which mainly contains the data on Bitcoin converted into every currency. This data is collected in USD rate conversion. It has eight features namely-Date, opening price, closing price, High and low price for the day, Volume of BTC, Volume of currency, and the Weighted Price of that particular date. The data is collected daily from October of 2011 to March of 2021. The following figure ( Fig 2) shows few rows of the dataset sorted by date in descending order:



**Fig 2**

Before we start implementing our proposed ARIMA model, we need to look at the trend of both the price and the volume of Bitcoin (fig3). More importantly, we also need to understand all the variables in the given table to understand using what field will we try to forecast the price and the volume.

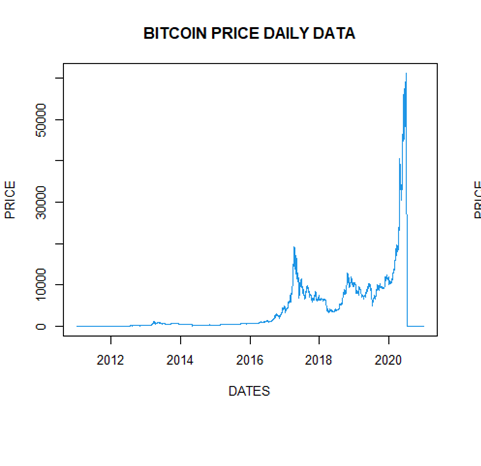


Fig 3

**TABLE 1:**

**VARIABLE**  **DEFINITION TYPE**

OPEN DEFINES THE OPENING PRICE ON THAT NUM

PARTICULAR DATE

CLOSE DEFINES THE CLOSING PRICE ON THAT NUM

PARTICULAR DATE

HIGH DEFINES THE HIGHEST PRICE ON THAT NUM

PARTICULAR DATE

LOW DEFINES THE LOWEST PRICE ON THAT NUM

PARTICULAR DATE

VOLUME\_BTC DEFINES THE VOLUME OF ONLY BITCOIN NUM

ON THAT PARTICULAR DATE

VOLUME\_CURRENCY DEFINES THE VOLUME OF EVERY CYPTO NUM

ON THAT PARTICULAR DATE

WEIGHTED \_PRICE DEFINES THE WEIGHTED PRICE ON THAT NUM

PARTICULAR DATE

**STATISTICAL ANALYSIS OF THE DATASET:**

The given figure illustrates some important parameters about the data. These parameters might be helpful when trying to analyze the dataset and understanding the dataset in debt.

Date Open High Low

Length:3470 Min. : 0.0 Min. : 0.0 Min. : 0.0

Class :character 1st Qu.: 232.1 1st Qu.: 236.3 1st Qu.: 227.4

Mode :character Median : 648.9 Median : 663.4 Median : 628.8

Mean : 4253.6 Mean : 4387.7 Mean : 4104.2

3rd Qu.: 6944.5 3rd Qu.: 7223.1 3rd Qu.: 6759.6

Max. :61177.0 Max. :61781.8 Max. :58959.6

Close Volume\_BTC Volume\_Currency Weighted\_Price

Min. : 0.0 Min. : 0 Min. :0.000e+00 Min. : 0.0

1st Qu.: 232.2 1st Qu.: 3656 1st Qu.:1.504e+06 1st Qu.: 232.7

Median : 649.6 Median : 7100 Median :6.768e+06 Median : 646.4

Mean : 4265.6 Mean : 9657 Mean :4.150e+07 Mean : 4250.1

3rd Qu.: 6963.4 3rd Qu.: 12336 3rd Qu.:4.781e+07 3rd Qu.: 6983.9

Max. :61165.2 Max. :137070 Max. :1.369e+09 Max. :60429.1

The data collected is daily data so the frequency used for this is 365.

**4. THEORY AND HYPOTHESES**

The market character of a Bitcoin is dependent on two variables namely- the price and the volume of the BTC. The price of the bitcoin illustrates the direction of the trend of the bitcoin while the volume of the bitcoin demonstrates the strength of the trend. Both together can only help us visualize the real market character. The market character is defined by two terms namely – the bullish character and the bearish character. The market is said to follow a bullish trend when both the price and the volume are showing a similar trajectory. But when the market is bearish, both the price and the volume will show a different trajectory.

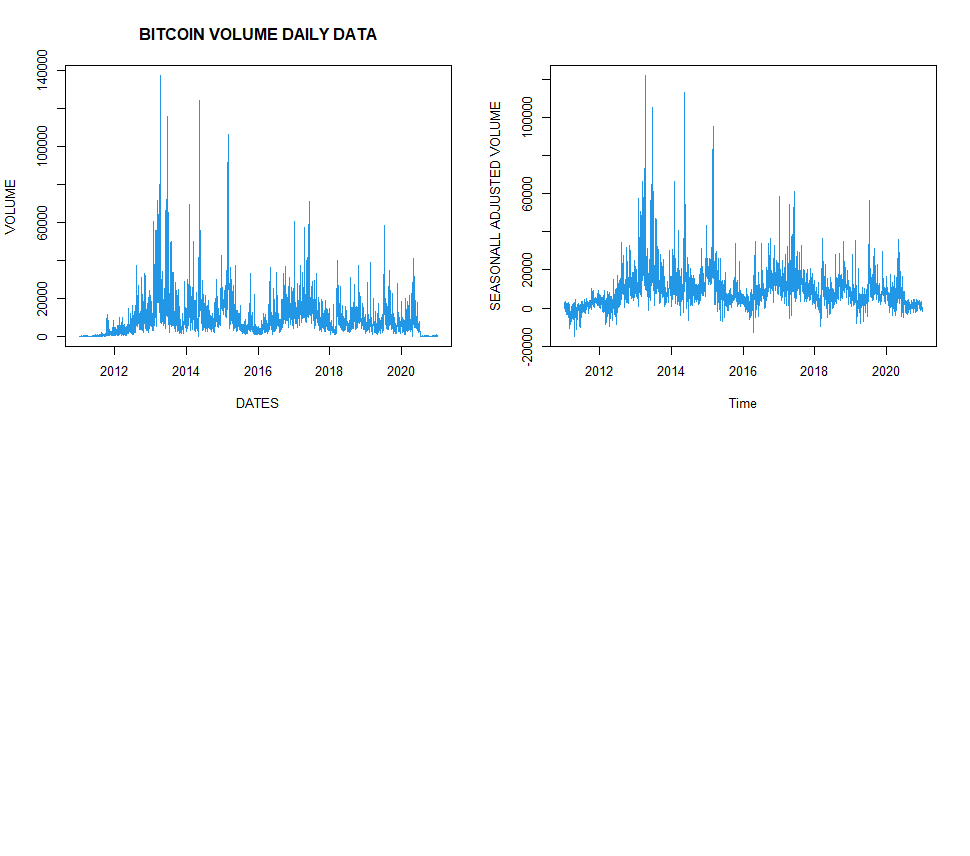
Predicting both the price and volume together does not pose a challenge while applying the ARIMA model. Due to a specific requirement of data-frame by Prophet model, it was required to separately analyze the variables.

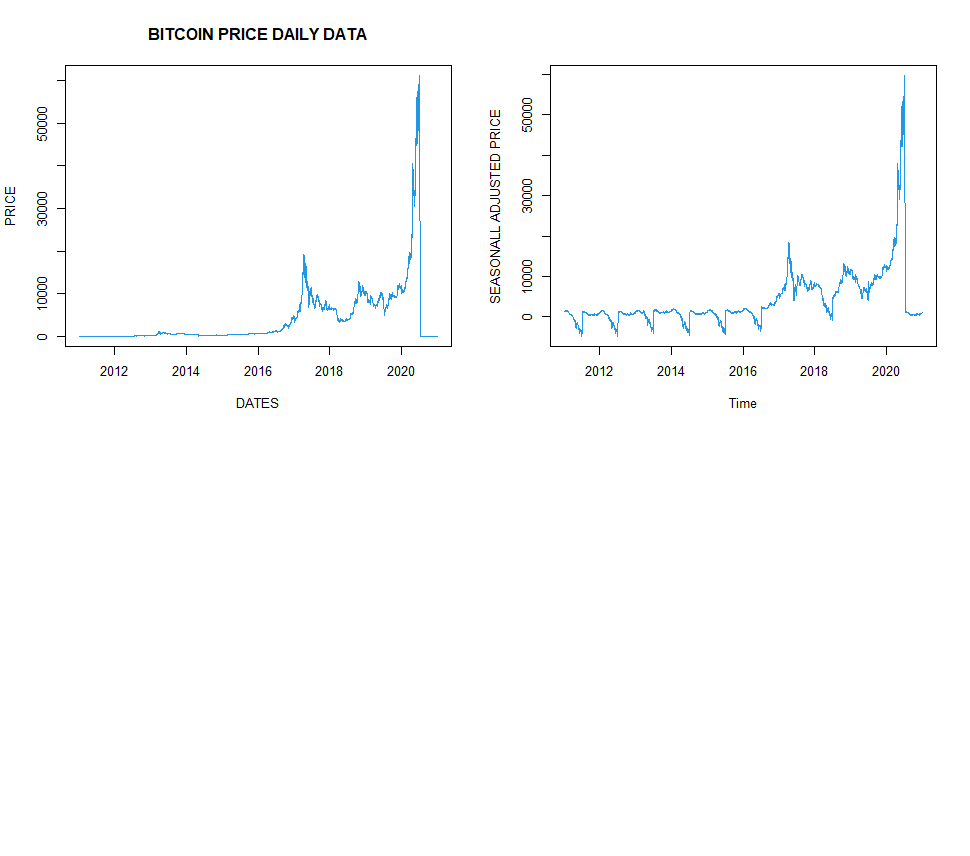
For, this particular project, we intend to predict the price and volume of BTC by using two different models which will help us to get two conclusions namely- the market character of BTC and which model of the two performs better.

**5. EMPIRICAL METHODOLOGY**

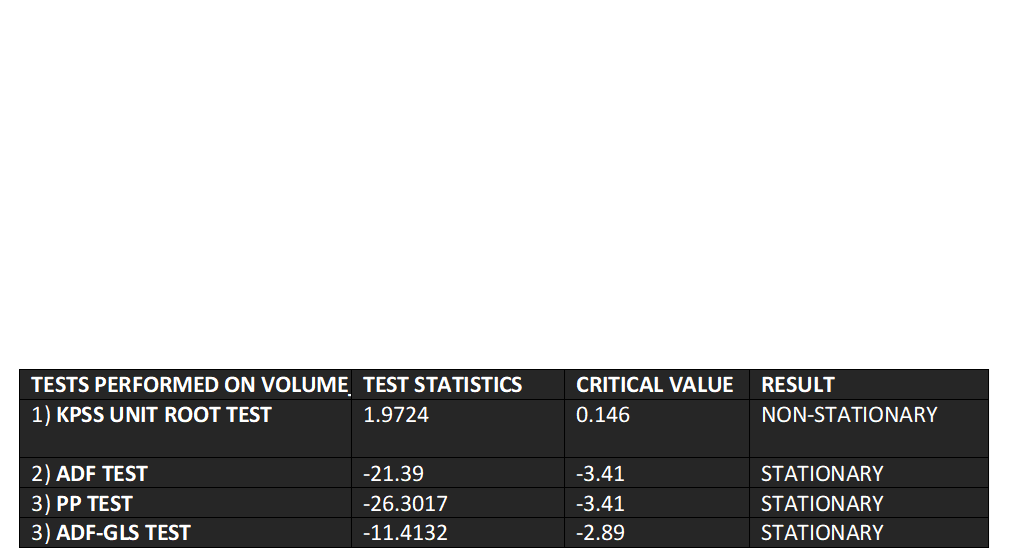
Seasonality is variations at specific timeframes which should be removed. The mean of the data is not around zero so it means that the data is not even stationary. So, after removing the seasonality of the data by removing the seasonal part as done for both the price and the volume as shown in Fig 4(a),(b), we would have to make the data stationary too.

Purely based on the figures, we cannot determine if it is stationary or not. To apply the ARIMA model, we need to make sure that the price and the volume of the dataset are stationary or not. We can make sure by performing stationarity tests like Augmented Dickey-Fuller, Kwiatkowski-Phillip-Schmidt-Shin, ADF-GLS, and Phillip-Perron test. These tests have been performed for both these variables to check the stationarity of these variables in Fig 5(a),(b).

 **Fig 4(a)**



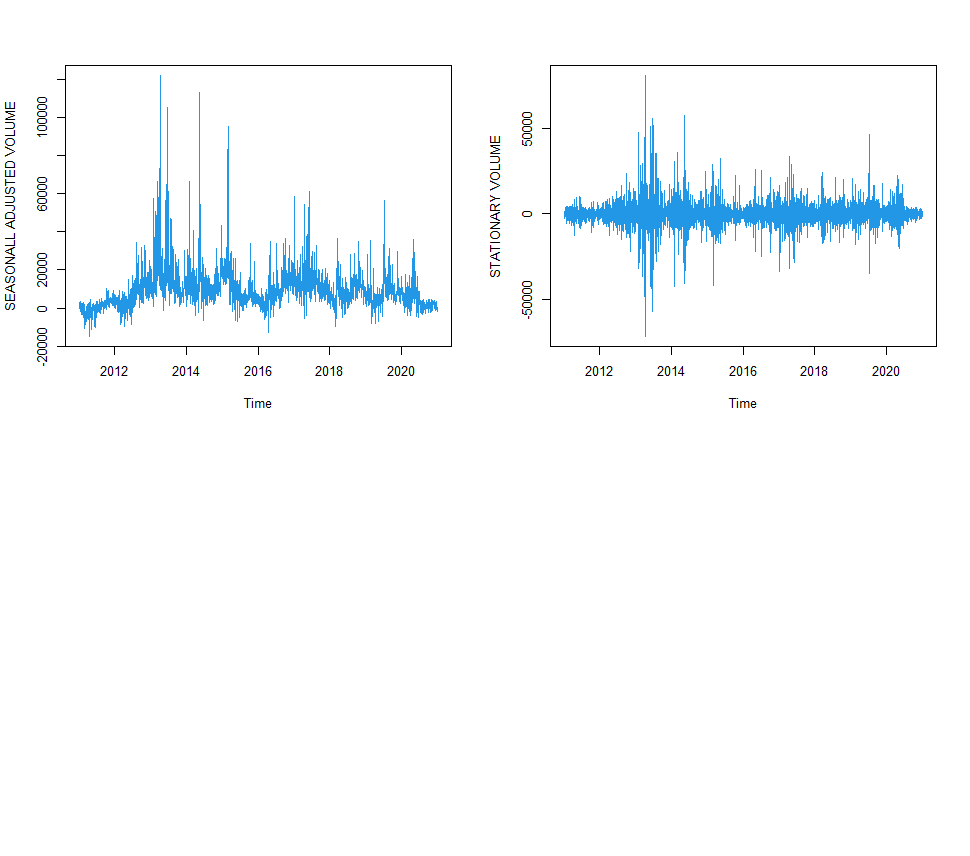
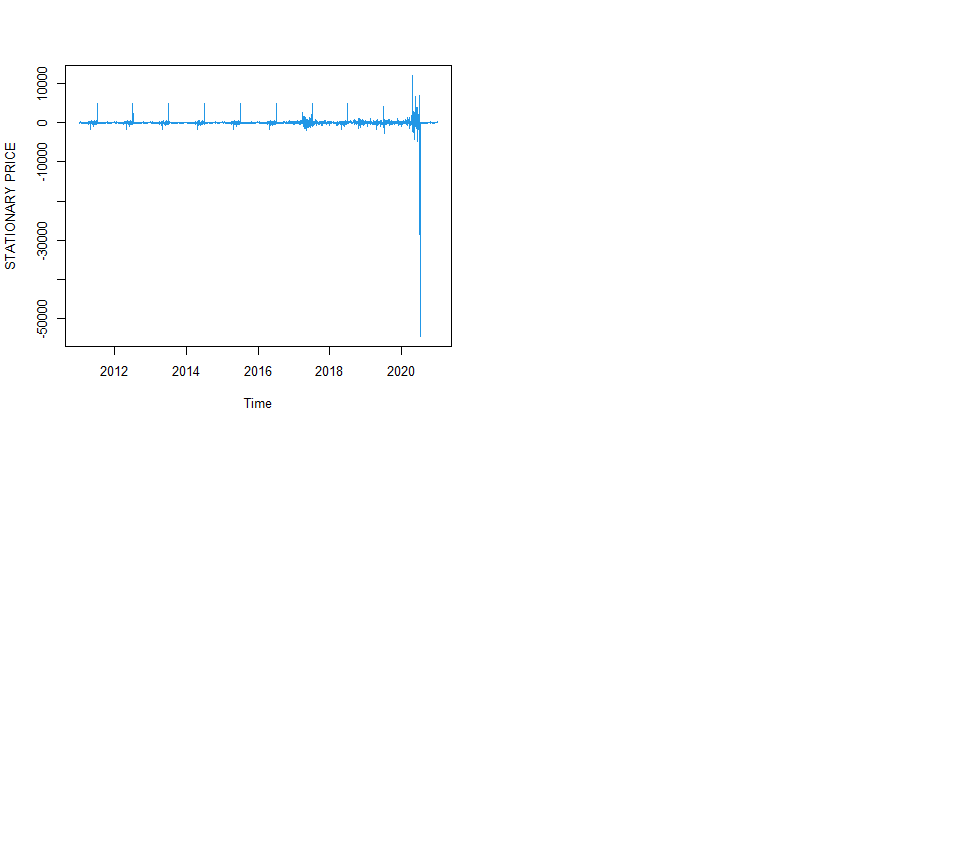
**Fig 4(b)**

**n Fig 5(a) (FOR VOLUME)**

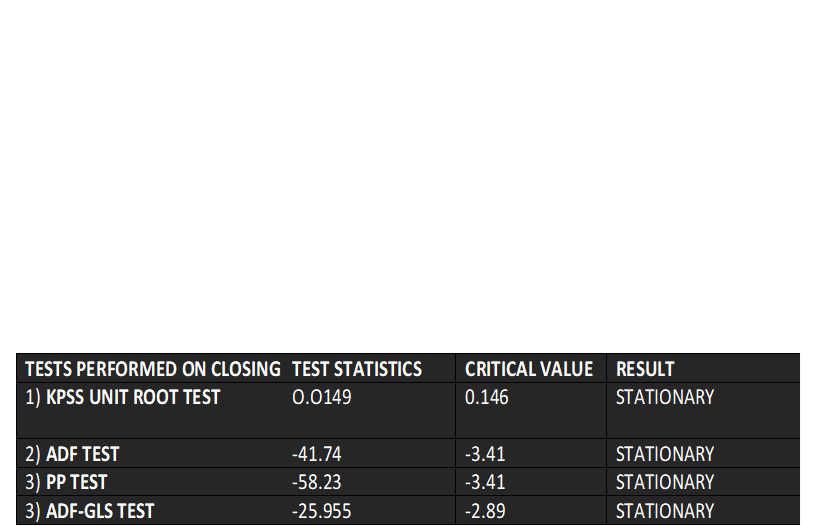


**Fig 5(b) (FOR PRICE)**

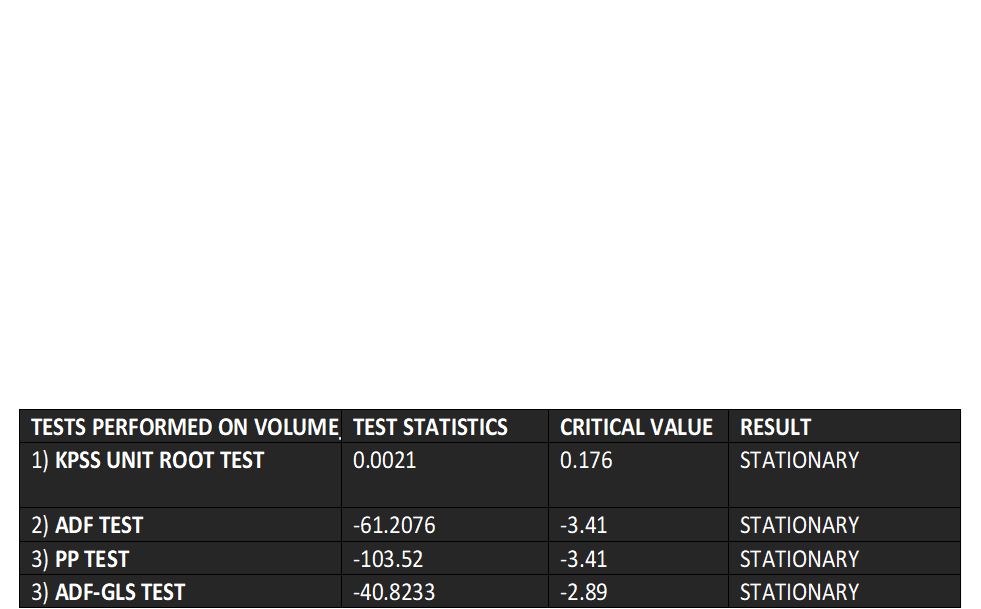
Fig 5(a)(b) illustrates that the variables are not stationary. So, we will take the log differencing of the variable to make the variables stationary. The plots and the stationarity test after performing log differencing are shown in fig 6 and fig 7.



**Fig 6**



**( FOR PRICE)**



**(FOR VOLUME)**

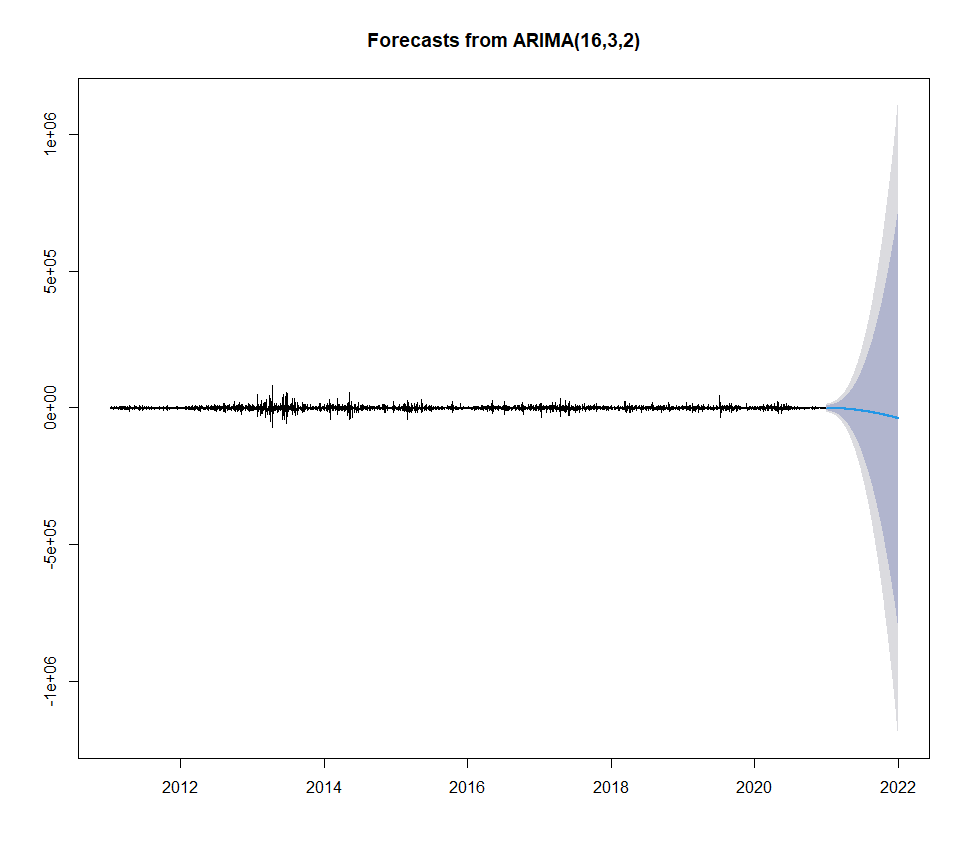
**Fig 7**

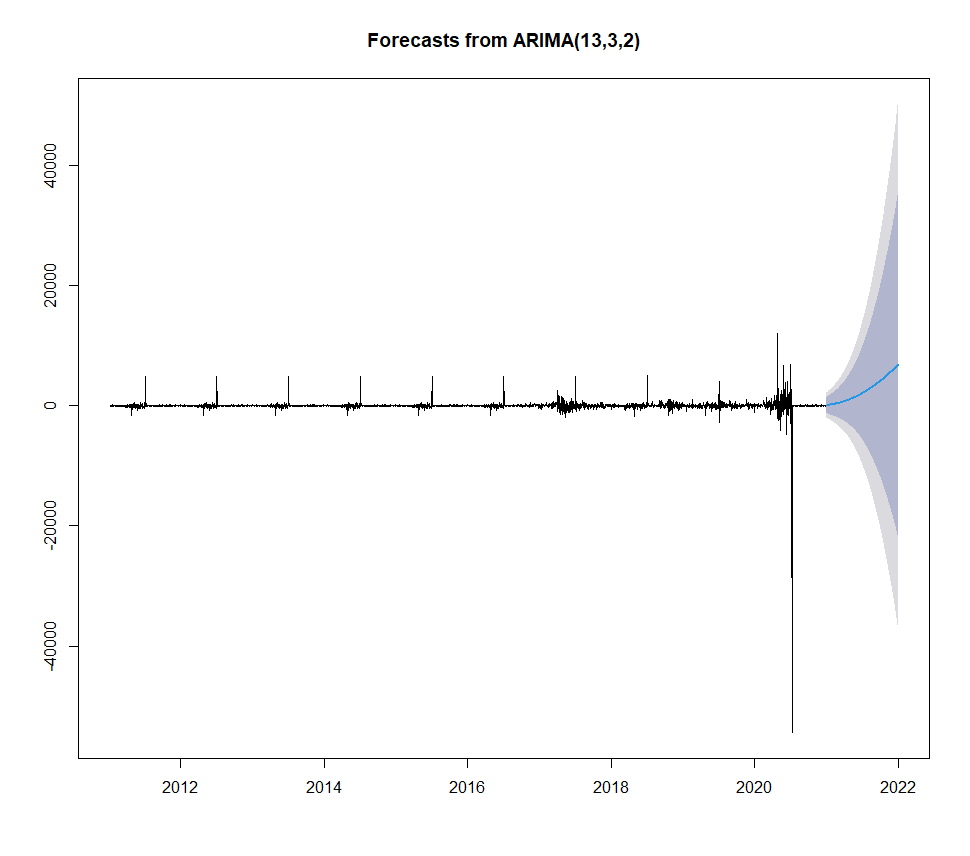
For the implementation of the ARIMA model, the dataset is stationary and is seasonally adjusted. For implementing the ARIMA model, we need p,d, and q values. ARIMA model is actually a type of model that explains a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. An ARIMA model is characterized by 3 terms: p, d, q where,p is the order of the AR term, d is the number of differencing required to make the time series stationary and q is the order of the MA term. To be precise ‘p’ is the number of lags that should be used as predictors. And ‘q’ is the order of the ‘Moving Average’ (MA) term which refers to the number of lagged forecast errors that should go into the ARIMA Model. After putting p, q, d values into the ARIMA model, we can find the model that best fits the data by finding the minimum AIC VALUE. In this particular case, the model that best fits the data is ARIMA(13,3,2). This model has the least AIC values among the rest for the price variable and the volume, the model that works best is ARIMA(16,3,2). For the price, lags that are used as predictors are p which is the order of AR term which is 13 and the order of Moving Average is 2. For the volume, lags that are used as predictors are p which is the order of AR term which is 16 and the order of Moving Average is 2. Now, we will fit the data into the models and then use it to forecast daily values for the next year and display the forecast with its Root Mean Square Error (RMSE) which is the evaluation metric for our models. The forecasted figure and their RMSE for both the price and the volume are given in figure 8(a)(b).

For implementing the Prophet model, we need to make a data frame with ‘ds’ and ‘y’ columns. The ‘ds’ column contains the dates of the entire dataset and the y column can contain only one field either the price or the volume of the BTC. This is the reason why we had to do it separately for each variable. One by one, we can add the variables into the y column and then would apply the prophet model to the data frame made. We have to keep in mind that the data is not seasonal or stationary. After the model is generated, we will try to predict future values and forecast 365 days ahead. We will use the metric library to find the mean error of the forecasted model and take its square root to get the RMSE of the model. The forecasted plots and their RMSE are given in figure 9(a)(b).

The metric used to evaluate the models is RMSE which stands for Root Mean Square Error. RMSE is the square root of Mean Error (ME). Mean Error stands for the difference between the actual and the predicted values of the model. The lesser is the difference, the better the model is. Thus, the RMSE value should be lower for a model to have performed better. In this case, the ARIMA model has lesser RMSE so it outperforms the prophet model.

**FORECASTING USING ARIMA MODEL ( FOR VOLUME)**

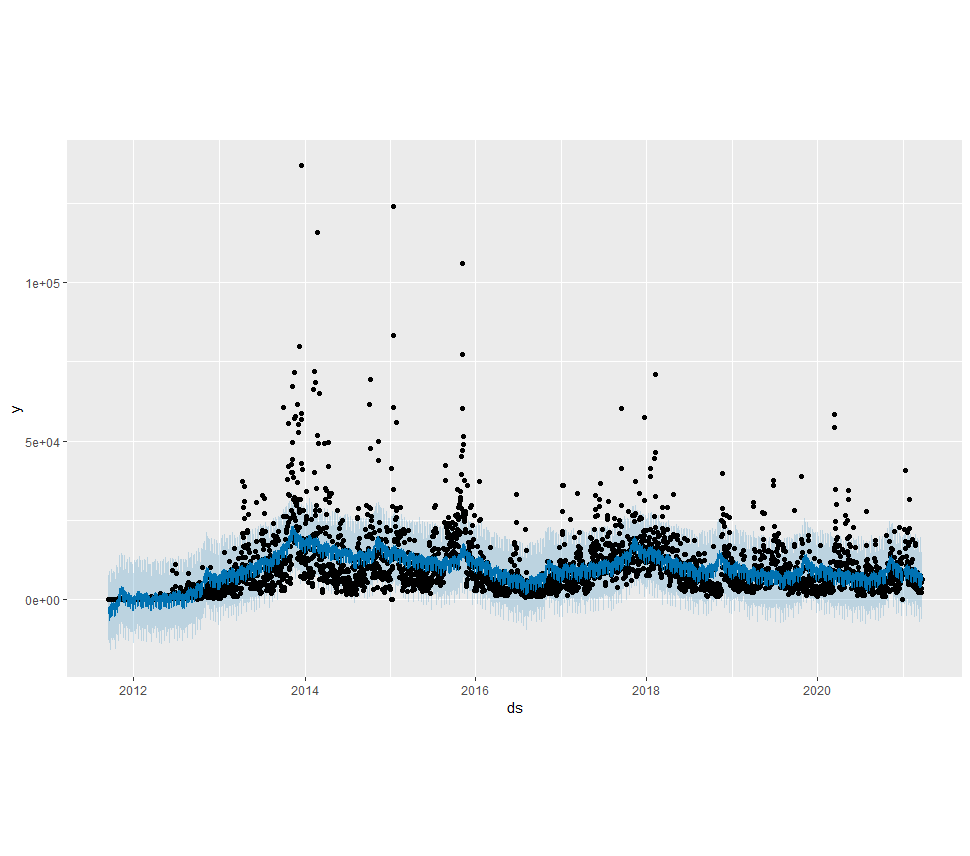
 **RMSE: 7404.244**

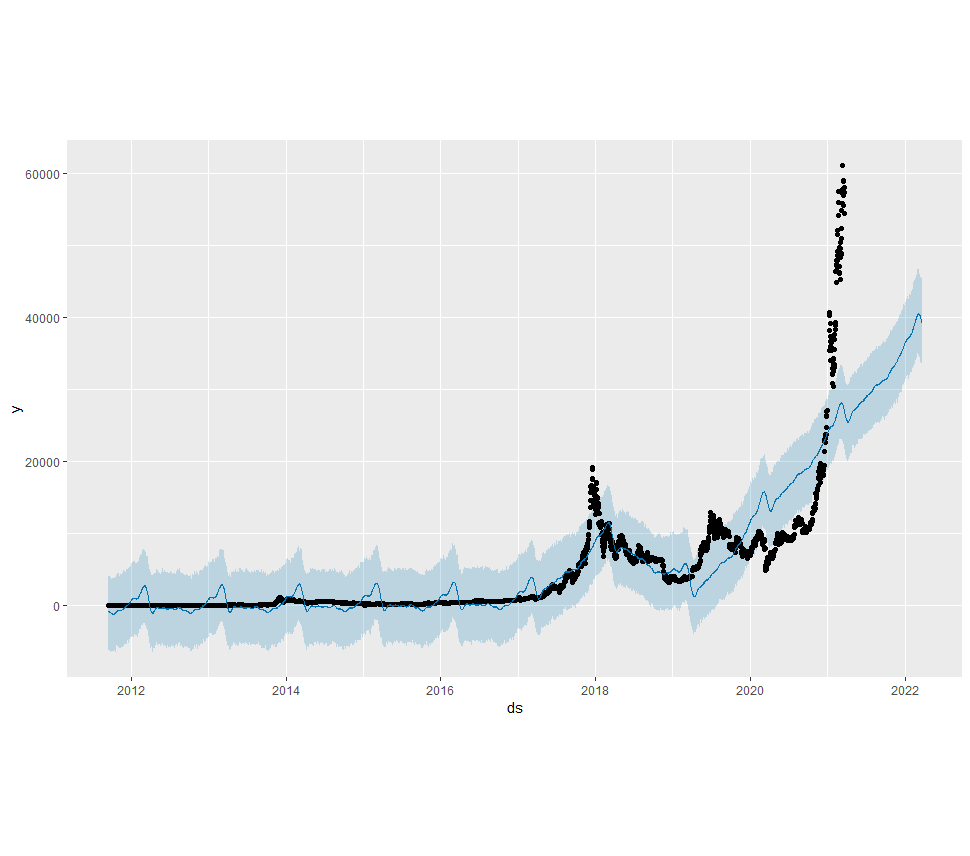
**FORECASTING USING ARIMA MODEL(FOR PRICE)**

**RMSE: 1067.921**

**FIG 8(a)(b)**

**Prediction using Prophet ( VOLUME) RMSE: 8495.901**



**PREDICTION USING PROPHET (FOR PRICE) RMSE: 3988.216** 

**FIG 9(a)(b)**

**6. RESULTS**

The study illustrates that according to the ARIMA and the Prophet models, the price of Bitcoin will go up but the volume of the bitcoin will show a downward trend. The market character of the BTC will witness a bearish trend meaning the overall market would see a downward trend. Forecasting these results, it would be wise to state that investing in Bitcoin would not be a suitable choice. We can say that because a bearish trend is seen when the price and the volume are showcasing a different trajectory curve in their graphs which apparently in this case they are.

Both the models produced have proven to be not the best for forecasting the price of Bitcoin and especially the volume of the Bitcoin. The RMSE values for both are very high though are higher for predicting the Volume of Bitcoin. But the Arima model has lower RMSE for both the variables so, we can say that the ARIMA model outperforms the Prophet model. RMSE, the Root Mean Square Error should be less as that demonstrates the square root of the difference between the actual and the predicted values.

The plots in fig 8(a)(b) show 80% and 95% confidence level. The area inside the dark grey region shows an 80% confidence level and the area in the light grey region shows a 95% confidence level. The area where the confidence interval enlarges cannot be much trusted but the area where the confidence interval is narrow can be trusted. As the confidence interval enlarges, the model losses its ability to forecast.

The plots in fig 9(a)(b) show both the actual and the predicted values. The black line shows the actual values while the lines in blue show the predicted values. Both, the plots show that the actual and the predicted values follow the same trend.

**7. CONCLUSION**

The paper is responsible for predicting the price and the volume of the bitcoin data to understand the market character of Bitcoin. It accomplishes the task by using two models namely the ARIMA and the Prophet model. The purpose of the paper was to analyze the market character and also find the better model among the two models presented. To conclude, we would say that the market shows a bearish character that means that the overall market of the bitcoin will go down even if the price increases because the volume of the bitcoin shows a downward trajectory. A comparison was made between the two models that showed the same results in general based on the error metric RMSE. The error metric clearly stated that the ARIMA model outperformed the Prophet model.

For future works, models such as the stock to flow model created by Twitter could be researched upon. This model is considered as most accurate and is based on bitcoin scarcity and the halving events of bitcoin. The model has fairly accurately tracked the price of bitcoin in the past. The stock to flow is a number that shows how many years at the current production rate are required to achieve the current stock. The higher the number, the higher the price so the stock to flow model based on scarcity and the production rate of new bitcoin. Apart from this, models like RNN, LSTM mainly machine learning, or deep learning models could do better as seen in many other research works. Research could also be done on finding better ARIMA models than the one presented in the paper.

**REFERENCES**

**1.** [Bitcoin Price Prediction Using Time Series Forecasting | by Ayushi Asthana | Towards Data Science](https://towardsdatascience.com/bitcoin-price-prediction-using-time-series-forecasting-9f468f7174d3)

2. [Bitcoin Definition (investopedia.com)](https://www.investopedia.com/terms/b/bitcoin.asp)

3. [What is bitcoin? - CNNMoney](https://money.cnn.com/infographic/technology/what-is-bitcoin/index.html)

4. [**Arthur Gervais**](https://www.researchgate.net/profile/Arthur-Gervais)**,** [**Ghassan Karame**](https://www.researchgate.net/profile/Ghassan-Karame)**,** [**Vedran Capkun**](https://www.researchgate.net/scientific-contributions/Vedran-Capkun-2087714579)**,** [**Srdjan Capkun**](https://www.researchgate.net/profile/Srdjan-Capkun)**,** Is Bitcoin a Decentralized Currency?, [IEEE Security and Privacy Magazine](https://www.researchgate.net/journal/IEEE-Security-and-Privacy-Magazine-1540-7993)

5. Bitcoin Price Prediction by Ahmer Sabah and Mahpari Ansari- investigated the effect of p, q, and d values on the achieved MSE in the prediction.

6. Seq2Seq RNN and ARIMA models for cryptocurrency prediction: A comparative study by J. Rebane, I.Karlsson, Sdenic, and P. Papatetrou –RNN outperforms the ARIMA model.

7. Shukla S, Dev S(2020)- BITCOINS BEAT CORONAVIRUS BLUES.

8. Time series forecasting of Bitcoin using high-dimensional features: a machine learning approach by Mohammed Mudassir, Shada Bennbaia, Devin Unal, and Mohammad Hammoudeh- machine learning model like ANN, LSTM, SVM, SANN outperforms the ARIMA model.

9. Munim ZH, Shakil MH, Alon I (2019) Next-day Bitcoin price forecast. J Risk Financ Manag 12(2):103. https://doi.org/10.3390/ jrfm12020103

10. Jang H, Lee J (2017) An empirical study on modeling and prediction of Bitcoin prices with bayesian neural networks based on blockchain information. IEEE Access 6:5427–5437

11. Shah D, Zhang K (2014) Bayesian regression and Bitcoin. In: 2014 52nd annual Allerton conference on communication, control, and computing (Allerton), IEEE, pp 409–414

12. Chen GH, Nikolov S, Shah D (2013) A latent source model for nonparametric time series classification. In: Advances in neural information processing systems, pp 1088–1096

13. Greaves A, Au B (2015) Using the Bitcoin transaction graph to predict the price of Bitcoin. https://doi.org/10.1109/CEC.2010.5586007

14.. McNally S, Roche J, Caton S (2018) Predicting the price of Bitcoin using machine learning. In: Proceedings—26th EuromiCRO international conference on parallel, distributed, and network-based processing, PDP 2018 pp 339–343. https://doi.org/10.1109/ PDP2018.2018.00060

15. M. Amjad and D. Shah, “Trading bitcoin and online time series prediction,” in NIPS 2016 Time Series Workshop, 2017, pp. 1–15.

16. S. McNally, J. Roche, and S. Caton, “Predicting the price of bitcoin using machine learning,” in IEEE 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), 2018, pp. 339–343.

17. H. Jang and J. Lee, “An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information,” IEEE Access, vol. 6, pp. 5427–5437, 2018.

18. P. J. Brockwell, R. A. Davis, and M. V. Calder, Introduction to time series and forecasting. Springer,2002, vol. 2.