The ISC Int'l Journal of Information Security

ISeCure

August 2019, Volume 11, Number 3 (pp. 139–144)

http://www.isecure-journal.org

SELECTED PAPER AT THE ICCMIT'19 IN VIENNA, AUSTRIA

# Using Machine Learning ARIMA to Predict the Price of Cryptocurrencies☆

Saad Ali Alahmari [1],*

[1] *Department of Computer Science, Shaqra University, Shaqraa, Saudi Arabia*

ARTICLE INFO.

*Keywords:*
ARIMA, Cryptocurrency, Machine Learning, Time-series analysis, Bitcoin.

**Abstract**

The increasing volatility in pricing and growing potential for profit in digital currency have made predicting the price of cryptocurrency a very attractive research topic. Several studies have already been conducted using various machine-learning models to predict crypto currency prices. This study presented in this paper applied a classic Autoregressive Integrated Moving Average (ARIMA) model to predict the prices of the three major cryptocurrencies âĂŤ Bitcoin, XRP and Ethereum âĂŤ using daily, weekly and monthly time series. The results demonstrated that ARIMA outperforms most other methods in predicting cryptocurrency prices on a daily time series basis in terms of mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE).



## 1 Introduction

Predicting cryptocurrency prices has become essential for many different fields of computational sciences because of the high volatility of cryptocurrency prices and the significant potential for investor profits. Cryptocurrencies are virtual currency that are anonymously processed over a decentralized network. There are over 2000 cryptocurrencies in use, most of which are traded anonymously, using blockchain technology. This study focused on the three major cryptocurrencies, Bitcoin, XRP and Ethereum, which represent the largest share of market capitalization. Bitcoin, the first cryptocurrency, was introduced by Satoshi Nakmoto in 2009. It uses blockchain technology as the platform for its transactions. XRP, the virtual currency of the Ripple payment system, was developed as an open source Internet protocol in 2013. Ethereum is a decentralized system that is fully autonomous and has substantial capabilities as a whole network. Unlike Bitcoin and Ethereum, XRP does not use blockchain technology because it has its own technology, called the Ripple Protocol Consensus Algorithm (RPCA).

Time series predictions are used in several domains, including finance, education, and marketing. Many of these predictions use the Autoregressive Integrated Moving Average (ARIMA) statistical model and, according to literature, a significant amount of that use is for the prediction of stationary datasets [1, 2]. The central challenge faced by researchers in predicting the price of cryptocurrencies, however, is the extreme volatility in prices [1–3]. In using the ARIMA model for this study, the researchers addressed this issue by adequately rendering the datasets stationary over

☆ The ICCMIT'19 program committee effort is highly acknowledged for reviewing this paper.

* Corresponding author.

Email address: s.alahmari@su.edu.sa (S. A. Alahmari)

ISSN: 2008-2045 

different time series dimensions.

This paper is structured as follows: Section 2 presents related work of previous research in price prediction of cryptocurrencies. Section 3 describes the theoretical background of ARIMA model. Section 4 addresses the experimental evaluation, presetting the data collections methods and the extracting of data features, discussing the stationary of data sets and defining the relevance ARIMA parameters. Section 5 discusses the experimental results. Section 6 addresses conclusions and observations, showing limitations to this approach and recommendations to future researches

## 2 Related Work

Several researchers have tried using variety of statistical and machine learning algorithms to predict the price of Bitcoin. The primary aim of those studies was to identify the most relevant features that could affect price predictions; however, by limiting the predictions to a single cryptocurrency (Bitcoin) they may not have been able to obtain sufficient affirmative scientific findings. Thus, this research elected to study the three cryptocurrencies (Bitcoin, XRP and Ethereum) which had the highest market capitalization value.

Reference [2] used ARIMA and SEQ2SEQ recurrent deep multi-layer neural network (SEQ2SEQ) to predict Bitcoin pricing, however, their models showed that recurrent neural networks (RNN) outperformed ARIMA in the long term only with additional input sources. Furthermore, the significant volatility in the Bitcoin datasets over the study period led to varied results. Reference [3] studied the short-term prediction of volatility for Bitcoin and U.S. dollars using a number of statistical tests (e.g. exponential weighted moving average (EMWA)) and machine learning models (e.g. random forests and Gaussian processes) using an hourly time series. They found that the extreme gradient boosting (XGT) and the elastic-net (ENET) achieved the best accuracy among other prediction models. Reference [4] applied a Bayesian optimized RNN and a long short-term memory (LSTM) network to predict the Bitcoin pricing process. Their results showed the LSTM network had a better classification accuracy of 52% and a lower regression error of 8%. They then compared the results with that of the ARIMA model without regressors.

Reference [5] used blockchain network features combined with Bitcoin basic features and included several regression and classification models (linear, logistic, SVM, and neural network). Although the best accuracy results achieved was 55% using the neural network classification, the authors noticed that using the network-based features did not have a significant effect on Bitcoin price predictions. Reference [6] focused on the Bitcoin process, considering blockchain and macroeconomic features, using the Bayesian neural networks (BNNs).

Reference [7] experiments several Regression and Deep Learning models using 1-minute interval Bitcoin trading data for six years period. These models are Theil-Sen Regression and Huber Regression, Long short-term memory (LSTM) and Gated Recurrent Unit (GRU). The GRU model gives the best results of MSE at 0.00002 and R2 at 0.992, followed by the results of the LSTM model. Reference [8] examines the estimation of volatility of three classic currencies against three cryptocurrencies by combing traditional Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and with the machine learning Support Vector Regression (SVR).

The results show that SVR-GARCH models outperforms all GARCH and its extensions models based the value of RMSE, MAE, the Diebold-Mariano test p-values and HansenâĂŹs Model Confidence Set. It should be noted that previous research neglected the importance of predicting different cryptocurrencies using the same model and then evaluating the model that was used to predict selected cryptocurrencies. This study considered three cryptocurrencies (Bitcoin, XRP and Ethereum) applying the ARIMA model to each and then training the dataset on daily, weekly, and monthly time series to predict the potential price and short-term direction.

## 3 ARIMA Method

The ARIMA method relies on previous values in the series for forecasting. It was invented by George Box and Gwilym Jenkins and is a widely used forecasting model [6]. The ARIMA model consists of the Auto-Regressive terms $AR$ and the Moving Average $MA$ terms.

Applying the lag operator $denotedL$, Autoregressive $AR$ terms are lagged values of the dependent variable and refers to it as lag order $p$ as the number of time lags. A non-seasonal $AR(p)$ can be formulated as follow:

$$AR(p): \emptyset(L) = 1 - \emptyset_1 L - \emptyset_2 L^2 - ... - \emptyset_p L^P \quad (1)$$

Moving Average $MA$ terms are lagged forecast errors in the predictions between past actual values and their predicted values and refers to it as the order of moving average $q$. A non- seasonal $MA(q)$ can be formulated as follows:

$$MA(q): \theta_(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q \quad (2)$$

Where $X_t$ is the time series for $t = 1, , n$, $Z_t$ is white noise, and the $d$ is the number of times that

**Table 1**. A Summary Statistics

| Statis. | Bitcoin | | | XRP | | | Ethereum | | |
|---|---|---|---|---|---|---|---|---|---|
| | D. | W. | M. | D. | W. | M. | D. | W. | M. |
| #Observ. | 2058 | 295 | 69 | 1963 | 282 | 65 | 1230 | 132 | 41 |
| Mean | 2297.791 | 2292.023 | 2278.808 | 0.174 | 0.165 | 0.166 | 219.885 | 210.4 | 209.84 |
| STD. | 3427.13 | 3417.74 | 3380.23 | 0.369 | 0.338 | 0.320 | 288.533 | 253.524 | 273.873 |
| Min. | 68.43 | 78.501 | 90.51 | 0.003 | 0.003 | 0.004 | 0.4829 | 0.522 | 0.66 |
| Max. | 19497.4 | 17667.07 | 15294.27 | 3.840 | 2.95 | 1.877 | 1432.880 | 1270.772 | 1103.64 |

the observations are differenced. The ARIMA model can be formulated as follows:

$$ARIMA(p, d, q) : (L) = 1 - \emptyset_1(L - 1)^d X_t = \theta_(L)Z_t \quad (3)$$

## 4 Experimental Evaluation

### 4.1 Data Collection

The cryptocurrency datasets used in this study included 200 digital cryptocurrencies and was collected from online data source [9] with updates extracted from the CoinMarketcap [10]. A subset for the three cryptocurrencies (Bitcoin, XRP, and Ethereum) was extracted as they were the focus of the research. The time series selected for the datasets varied among the three cryptocurrencies. Unlike the traditional monetary markets, there were dramatic changes in the price of the cryptocurrencies during these selected periods. Each dataset row represented daily market details for the cryptocurrencies worldwide and identified the following eight features: Date, Currency, Open, High, Low, Close, Volume, and Market Capital. The "Close" feature was chosen specifically as a dependent variable because it showed the closing price of the day and the open price for the following day. The "Volume" feature was designated as an independent variable because it represented the volume of transactions on the given day.

The original dataset comprised 84,080 observations. After extracting the sub-datasets related to the three cryptocurrencies (Bitcoin, XRP and Ethereum), the number of observations was reduced to 5,251 observations.

### 4.2 Feature Extraction

In comparing the mean values for cryptocurrency datasets, the researchers noted significant variances and considered minimizing the variances by using different time frequencies, such as hourly [11] and daily [12–14]. However, selecting a specific time series frequency (e.g. daily, weekly, monthly) observation is feature-engineering task that can lead to different insights. For this study the datasets were downsampled in three different levels to obtain greater insights. The model was trained with the "Close Price" feature

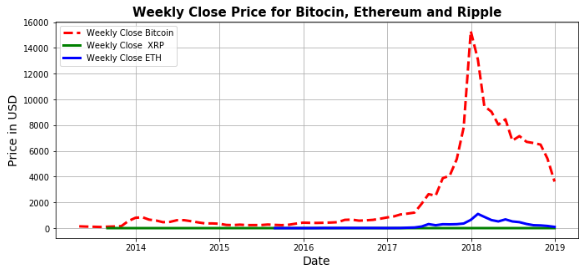


**Figure 1**. Historical datasets of Bitcoin, XRP and Ethereum

in three dimensions for each of the three cryptocurrencies (Bitcoin, XRP and Ethereum) as follows:

(1) Daily subdatasets: daily transactions for Bitcoin (2013-04-28 to 2018-12-15), XRP (2013-08-04 to 2018-12-18) and Ethereum (2015-08-07 to 2018-12-18).
(2) Weekly subdatasets: datasets were downsampled (mentioned 1) to weekly base by taking the average transaction value per week for Bitcoin, XRP and Ethereum.
(3) Monthly subdatasets: datasets were downsampled (mentioned 1) to monthly base by taking the average transaction value per month for Bitcoin, XRP and Ethereum.

The datasets plotted in Figure 1 represents the historical datasets for Bitcoin, XRP and Ethereum for the selected time series. The huge volatility in Bitcoin prices can be clearly seen, starting in the middle of 2017 and extending to the end of the year. This increase positively affected the prices for XRP and Ethereum, however, it is hard to see those increases because the changes are very small in comparison to the Bitcoin prices, infrequently reaching \$3.8 and \$1270, respectively.

Summary statistics for these datasets are listed in Table 1. It should be noted that the variances are minimized as the datasets were downsampled, particularly with regards to the large datasets, such as Bitcoin. Significant price increases in major cryptocurrencies in the third and fourth quarter of year 2017, however, have positively skewed the datasets.

The datasets were partitioned, based on the date feature; historical data was used for training and the most recent data was used for testing. The partitioned datasets were then divided into 80% training set and 20% testing sets. Three different measures were used for prediction accuracy: mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE). They are defined as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{4}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}e_t^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2} \tag{6}$$

Where $n$ denotes the total number of samples forecasted, $e$ denotes the actual value of the sample, and $t$ denotes the forecasting value of the sample.

### 4.3 Analysing Stationary of the Time Series

The DickeyâĂŞFuller test used to check the stationarity resulted in a p-value $> 0.05$ which indicated that the datasets were not stationary [15]. To make the datasets stationary, the seasonality and trends from the series were reduced by using the differencing technique as follows:

- For the daily dataset samples, the XRP datasets was stationary and the differencing technique was applied to the Bitcoin and Ethereum datasets at the 12-month base level.
- For the weekly downsampled datasets, the differencing technique was applied to the Bitcoin and Ethereum datasets at the 12 month-base level. For the XRP dataset, which had a higher seasonality, the differencing technique was applied at the 3 month-base level. Figure 2 shows that the randomness in the time-series datasets and the varied positive correlations with significant lags on the weekly dataset samples starts at approximately 50 lags in Bitcoin, 45 lags in XRP, and 30 lags in Ethereum.
- For the monthly downsampled datasets, the differencing technique was applied to the Ethereum datasets at the 12 month-base level. For the Bitcoin and XRP datasets the differencing technique was applied at the 3 month-base level. Although, the XRP dataset was stationary on the weekly base resampling, it required further differencing, which was probably due to lower number of observations with larger outliers.

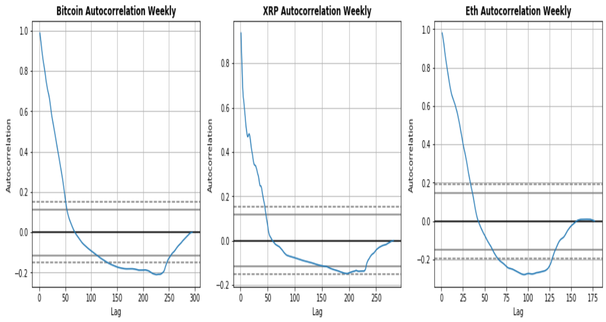


**Figure 2**. Randomness of Bitcoin, XRP and Ethereum time-series

### 4.4 Defining ARIMA Parameters

Once the datasets were stationary, the ARIMA parameters $p, d, andq$ were used to model the time series aspects. First, the best starting points for the $AR$ parameter $p$ were determined using the Auto-Correlation Function $ACF$. Then, the $MA$ parameter $q$ was determined using the Partial Auto-Correlation Function $PACF$. As significant values were found for both ACF and PACF it appears that the ARIMA model addressed the research problem.

The higher the downsampling is, the larger the range of values would be for the $ACF$ and $PACF$. The analysis showed that the best starting points for the $AR$ parameter $p$ of the daily base model were: Bitcoin ≈155 , XRP≈135 and Ethereum ≈95. For the weekly base model, shown in Figure 3 those values were: Bitcoin≈20 , XRP≈16 and Ethereum≈12, and for the monthly based model the values were: Bitcoin≈4, XRP≈3 and Ethereum≈2.

Consequently, there was a greater chance for spikes for $q$ for the datasets with greater downsampling. The analysis showed that there many spikes in the plots outside the insignificant zone (shaded) in the model. For the daily base model those were: Bitcoin $= \{60, 50, 36, 20, \dots 6, 2\}$, XRP $= \{35, 32, .., 7, 2\}$ and Ethereum $= \{105, 100, \dots 7, 6, 5\}$. For the weekly base model the values were: Bitcoin $= \{2\}$, XRP $= \{2, 5\}$ and Ethereum $= \{2, 4, 5\}$ ; Figure 4 shows the plot spikes outside the insignificant zone and the longest spikes for $q$ of the weekly model

## 5 Result and Discussions

This section describes the results of using the ARIMA model to forecast pricing for the Bitcoin, XRP and Ethereum cryptocurrencies. The experimental results are shown in Table 2. The experiment was repeated for daily, weekly, and monthly datasets by applying ARIMA with different parameters, depending on the AR and MA results for every dataset of or the Bitcoin, XRP and Ethereum.

**Table 2**. Experimental Results for MAE, MSE and RMSE Tests

| Test. | Bitcoin | | | XRP | | | Ethereum | | |
|---|---|---|---|---|---|---|---|---|---|
| | **D.** | **W.** | **M.** | **D.** | **W.** | **M.** | **D.** | **W.** | **M.** |
| **MAE.** | 313.894 | 764.24 | 1937.78 | 0.041 | 0.125 | 0.254 | 12.949 | 31.249 | 94.05 |
| **MSE** | 294560 | 1389560 | 10012856 | 0.0097 | 0.0681 | 0.1822 | 410.01 | 1712 | 14818 |
| **RMSE** | 542.735 | 1178.8 | 3164.31 | 0.096 | 0.261 | 0.427 | 20.25 | 41.380 | 121.72 |

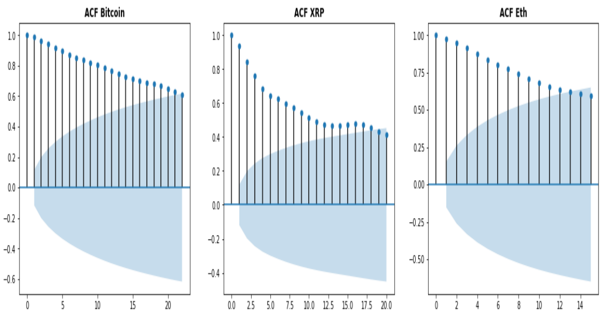


**Figure 3**. (ACF) Values

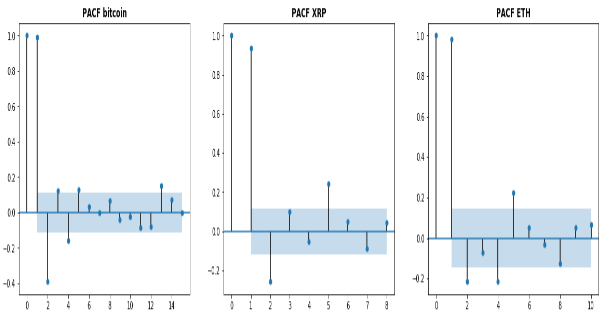


**Figure 4**. (PACF) Values

Table 2 summarizes the results of training errors and test errors. The researchers observed that using the ARIMA model on daily base re-sampling outperformed other models in terms of MAE, MSE and RMSE for predicting the price of Bitcoin, XRP and Ethereum. The ARIMA model performed better on weekly datasets than on monthly values for Bitcoin and XRP in terms of MAE, MSE and RMSE, whereas the error of the same model in Ethereum is worse due to the smaller dataset size. it is noted that the smallest time series re-smapling, is the better results on MAE, MSE and RMSE tests.

Figure 5 indicates the overall short-term direction for prices using the ARIMA model. The x-axis represents time scale by days, weeks, and months, respectively. The y-axis represents change in price, in US dollars; the green line is the predicted price. From the weekly and monthly ARIMA models the positive direction of both prices can be seen with a slight fluctuation in the short term. Interesting, the ARIMA model for the daily base trends in the negative direction, with many fluctuations, specifically for Bitcoin and XRP.

## 6 Conclusion

Predicting the price of cryptocurrency is a research area with significant profit potential, given the huge market capitalization. This study attempted to produce indications of price predictions for three major cryptocurrencies: Bitcoin, XRP and Ethereum. The researchers found that using the ARIMA model with weekly base re-sampling outperformed other models in terms of MAE, MSE and RMSE for predicting the price of Bitcoin, XRP and Ethereum. As result, the ARIMA model on weekly base showed a positive direction for prices in the short term for Bitcoin, XRP and Ethereum. Although the price increases, as well as the dataset sizes, were more dramatic for Bitcoin, in comparison with XRP and Ethereum, price indications for all three cryptocurrencies performed similarly. This leads the researchers to believe that the dataset variances could not affect the results as much as the algorithm used in the study.

In addition, other related data sources can be trained in the study, such as economic factors, number of transactions and number of financial institutions accepting cryptocurrencies. Applying other nonlinear algorithms is required to compare the overall accuracy of the ARIMA model

## References

[1] Noorhazlina Abu Bakar and Sofian Rosbi. Autoregressive integrated moving average ( arima ) model for forecasting cryptocurrency exchange rate in high volatility environment : A new insight of bitcoin transaction. 2017.

[2] Jonathan Rebane, Isak Karlsson, Panagiotis Papapetrou, and Stojan Denic. Seq2seq rnns and arima models for cryptocurrency prediction : A comparative study. In *Proceedings of SIGKDD Workshop on Fintech (SIGKDD FintechâĂŹ18) :*, 2018.

[3] T Guo and N Antulov-Fantulin. An experimental study of bitcoin fluctuation using machine learning methods. *arXiv preprint arXiv:1802.04065*, 2018.

[4] Sean McNally, Jason Roche, and Simon Caton. Predicting the price of bitcoin using machine

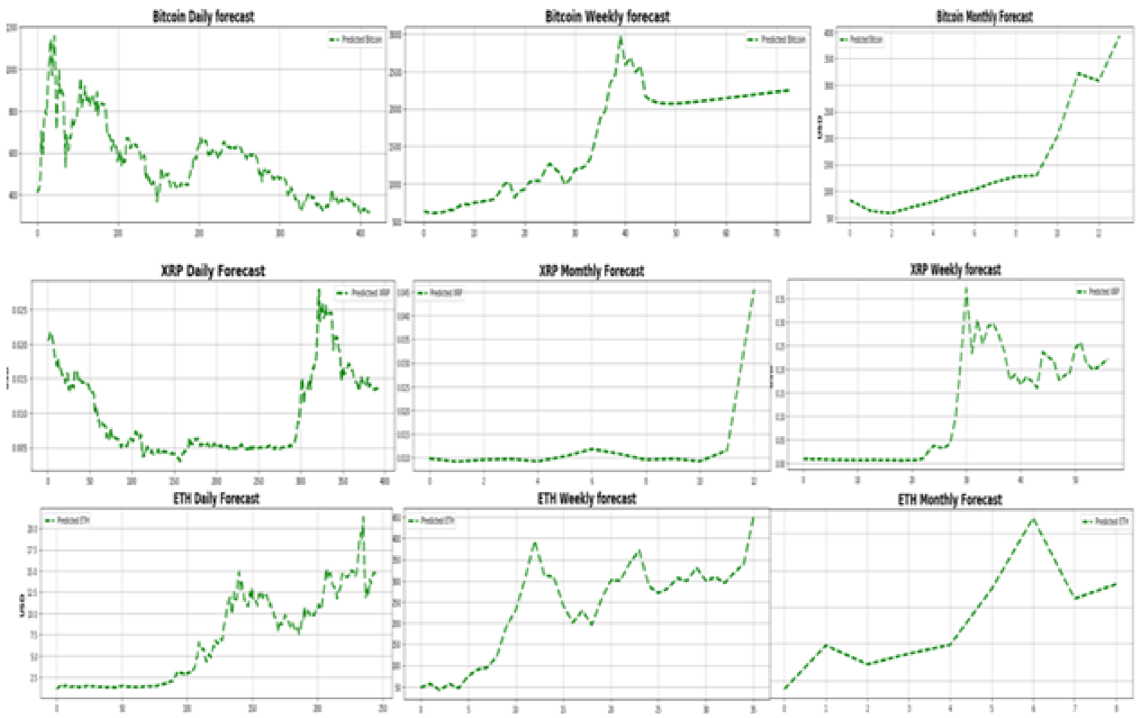


**Figure 5**. Overall ARIMA Model Price Directions

learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, pages 339–343. IEEE, 2018.

[5] Alex Greaves and Benjamin Au. Using the bitcoin transaction graph to predict the price of bitcoin. *No Data*, 2015.

[6] Huisu Jang and Jaewook Lee. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, 6:5427–5437, 2018.

[7] Thearasak Phaladisailoed and Thanisa Numnonda. Machine learning models comparison for bitcoin price prediction. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pages 506–511. IEEE, 2018.

[8] Yaohao Peng, Pedro Henrique Melo Albuquerque, Jader Martins Camboim de Sá, Ana Julia Akaishi Padula, and Mariana Rosa Montenegro. The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. *Expert Systems with Applications*, 97:177–192, 2018.

[9] Kaggle. http:// www.kaggle.com. accessed 15 June 2018.

[10] Coinmarktcap. http:// www. coinmarketcap.com. accessed 18 Dec 2018.

[11] Isaac Madan, Shaurya Saluja, and Aojia Zhao. Automated bitcoin trading via machine learning algorithms. 2014.

[12] Javier Contreras, Rosario Espinola, Francisco J Nogales, and Antonio J Conejo. Arima models to predict next-day electricity prices. *IEEE transactions on power systems*, 18(3):1014–1020, 2003.

[13] Dennys CA Mallqui and Ricardo AS Fernandes. Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*, 75:596–606, 2019.

[14] Siddhi Velankar, Sakshi Valecha, and Shreya Maji. Bitcoin price prediction using machine learning. In *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pages 144–147. IEEE, 2018.

[15] Christos Agiakloglou and Paul Newbold. Empirical evidence on dickey-fuller-type tests. *Journal of Time Series Analysis*, 13(6):471–483, 1992.

**Saad Ali Alahmari** Holds MSc in software engineering from University of Scranton USA, PhD in Computer Science from University of Southampton UK.He is currently a Assistant Prof. at the Computer Science department, Shaqra University, Saudi Araiba. His areas of research include Software Engineering, Machine Learning, Data Science, and Data Mining.