Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
May 2024 5(5), 1551–1559

# FORECASTING BITCOIN PRICES WITH TIME SERIES ANALYSIS IN PYTHON AND EXCEL

JayKrishna Joshi[1], Anish Gharat[2], Snehee Chheda[3], Dr. R.P. Sharma[4], Dr. Aboo Bakar Khan[5]

[1]Mukesh Patel School of Technology Management & Engineering, SVKM's NMIMS Mumbai, India, Research Scholar Chhatrapati Shivaji Maharaj University Navi Mumbai, India
[2]Department of Data Science, Mukesh Patel School of Technology Management & Engineering, SVKM's NMIMS Mumbai, India
[3]Department of Data Science, Mukesh Patel School of Technology Management & Engineering, SVKM's NMIMS Mumbai, India
[4]Chhatrapati Shivaji Maharaj University, Navi Mumbai, India
[5]Department of Electrical and Electronics Engineering, Chhatrapati Shivaji Maharaj University, Navi Mumbai, India

**DOI**
10.29121/shodhkosh.v5.i5.2024.3457

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Copyright:** © 2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License.

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.

## ABSTRACT

Over the past few years the interest in trading with a decentralized or virtual form of currency has significantly increased. This has led to the rise of the cryptocurrency market in the early 2000's. Bitcoin, the pioneer in the field, has managed to dominate this volatile and cutthroat market till this day. Cryptocurrency is based on multiple technological frameworks and success stories of high returns in a short time frame have garnered the interest of young investors as well.
Similar to the traditional stock market, multiple machine learning, Artificial Intelligence, Time Series Analysis models have come up to help investors, understand trends, patterns, and derive a deeper understanding of the asset as well as the market they are investing in.
Our research aims to deal with this problem using time series methods such as Auto-Regressive (AR), Moving Average (MA) and Auto-Regressive Integrated Moving Average (ARIMA) models. In our analysis, we have implemented the models using both Excel method and Python. Our metric of evaluation is Mean Absolute Percentage Error (MAPE). In our work, data was taken from a website called 'Yahoo finance' [1] for Bitcoin cryptocurrency for a five-year time period i.e. from 1st June, 2017 to 31st May, 2024.

Our Python methodology has resulted in a MAPE of 0.16 for AR model, 0.14 for MA model and 0.11 for ARIMA model. The same has been verified using Excel and the score has been validated.

**Keywords**: Bitcoin Price Prediction, Time Series Analysis, Python, Excel. ARIMA, AR, MA

## 1. INTRODUCTION

Cryptocurrency has garnered the attention of multiple investors since its advent. High returns, and fundamentals based on decentralized and virtual currency has made it interesting for all its investors, including the younger age group. Other than popular cryptocurrencies being traded today such as Bitcoin, Ethereum, etc multiple other cryptocurrencies exist. These cryptocurrencies can be used for multiple purposes other than trading as well, such as making transactions online, buying virtual art or Non-Fungible Tokens (NFTs). At its core, it is a virtual currency. What was once viewed as an experimental financial instrument is now a multi-billion-dollar market, with Bitcoin dominating as the leading cryptocurrency.

Similar to traditional markets, cryptocurrencies are traded on a daily basis. Methodologies for traditional markets such as Machine Learning, Artificial Intelligence, Time Series Analysis, Algorithmic Trading, etc have investors make calculated decisions. These methods are not limited to traditional markets and can be expanded and applied to cryptocurrency as well. However, the cryptocurrency market is inherently more volatile and unpredictable than traditional financial markets. As a result, understanding price trends and making informed investment decisions requires advanced techniques beyond basic technical analysis.

## MODELS USED

1. Auto-Regression (AR): In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself. Thus, an autoregressive model of order '*p*' with $\varepsilon_t$ as white noise [2].

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

2. Moving Average (MA): While the AR model uses past values of the variable for future predictions, the MA model uses the past errors. An MA model with order 'q' and εt as white noise can be depicted by following equation:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2)$$

3. Auto-Regressive Integrated Moving Average (ARIMA): It is a combination of three components: AR; Integration (I), which is the differencing done to make the series stationary; and MA model. ARIMA(p,d,q) model consists of p being the order of the AR component, d being the degree of differencing applied and q being the order of the MA model. It is denoted as below with $y_t'$ as the differenced series. $y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$

The emergence of cryptocurrencies has completely changed the financial landscape and drawn interest from investors globally, particularly from younger demographics. In addition to being used for trading, cryptocurrencies like Bitcoin and Ethereum have become quite popular for other purposes like online transactions, buying virtual art, and Non-Fungible Tokens (NFTs). Fundamentally, a cryptocurrency is a virtual, decentralized money system based on blockchain technology. Once thought of as an experimental financial tool, this idea has grown into a multibillion-dollar industry, with Bitcoin emerging as the industry leader. This change demonstrates how essential cryptocurrencies have become to contemporary digital economies.

Compared to conventional monetary units like the US dollar or the euro, cryptocurrencies are essentially different. They are mostly free from interference from governments and are not subject to central authorities like national banks. One of the main draws for investors is its decentralized structure, which provides a degree of independence and autonomy not present in traditional financial systems. Cryptocurrencies are practically impervious to fraud or counterfeiting since they employ cryptographic techniques to secure transactions. Blockchain technology, which records every transaction on a public ledger available to all network users, strengthens this security.

Compared to conventional monetary units like the US dollar or the euro, cryptocurrencies are essentially different. They are mostly free from interference from governments and are not subject to central authorities like national banks. One of the main draws for investors is its decentralized structure, which provides a degree of independence and autonomy not present in traditional financial systems. Cryptocurrencies are practically impervious to fraud or counterfeiting since they employ cryptographic techniques to secure transactions. Blockchain technology, which records every transaction on a public ledger available to all network users, strengthens this security.

The cryptocurrency market is intrinsically more vulnerable to events such as abrupt shifts in regulatory rules, shifts in market sentiment, and world events than traditional markets, where price moves can frequently be predicted by fundamental and technical research. Therefore, sophisticated analytical approaches are needed to comprehend price trends and make informed investing decisions in cryptocurrencies. Moving averages and support and resistance levels are two examples of basic technical analysis methods that frequently fail to adequately capture the complexity of cryptocurrency markets. Sophisticated models from time series forecasting are utilized to handle these complications.

## METRIC USED

1. Mean Absolute Percentage Error (MAPE): Mean absolute percentage error is a relative error measure that uses absolute values to keep the positive and negative errors from cancelling one another out and uses relative errors to enable you to compare forecast accuracy between time-series models [5].

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

## 2. LITERATURE REVIEW

Multiple researches have been done using methods such as Long-Short Term Memory (LSTM), Neural Networks, Machine learning, etc. In a research article, the author uses methods such as Support Vector Machines (SVM), ARIMA and Bayesian models to predict the price of Bitcoins. It was concluded that SVM performed the best, but the other 2 methods performed fairly well as well [6].

Another paper focuses on the importance of having a significant amount of data to train the models and how different metrics could be used. Along with that, it uses LSTM and concluded that LSTM models due to their need for a high volume of data tend to overfit and are computationally heavy [7].

In another article, classical ML models are applied to 4 different kinds of cryptocurrencies namely Bitcoin, Ethereum, Dogecoin, and Bitcoin Cash. The authors emphasized the challenges in predictions including the volatility of the market, the lack of historical data, and the impact of external factors and concluded that traditional Linear Regression worked the best for their study [8].

A lot of studies have also used sentiment analysis from multiple social media sites to predict the prices of Bitcoins. SVM has proven to be a useful algorithm here as well and that sentiment analysis can help predict the trend of the currency. It helped understand market sentiment and price fluctuations. The limitation of this study included the inability to predict the magnitude of the price changes [9].

Neural Networks (NN) such as Recurrent Neural Networks (RNN) and random neural networks have been applied to the same task as well. NN can capture the patterns and non-linear relationships of the data and RNN has superior performance with respect to long term trends. The drawback of neural networks involved the challenges associated with hyperparameter tuning, optimization and model training needs such as high volume of data [10].

## 3. METHODOLOGY

The proposed methodology is a Python based code with Excel validation. We used pre-built models in the 'statsmodels' Python library. For Excel, the calculations have been done from scratch. The 5 years data taken is first resampled to monthly data to get a smoother curve while visualizing it. The general flow for all the models was the same for the Python code, which was to prepare the data, then fit the required model using the 'statsmodels' library, calculate the MAPE based on the predictions and visualize the results.

For the code, the AR and MA model follow the same workflow, whereas the ARIMA model follows a slightly different one, due to the functionality of the models. The AR and MA models are simpler and can be trained and predicted on the entire dataset, whereas in an ARIMA model, a train and test set creation are more optimal for prediction.

Fig. 1 shows the workflow of AR and MA models. Both of them follow a similar structure, with MAPE being the metric. Fig. 2 represents an in-depth view of how the AR and MA models are implemented. The working is divided into 6 main parts as indicated.

Fig. 3 and Fig. 4 similarly represent the workflow and pseudocode for the ARIMA model.

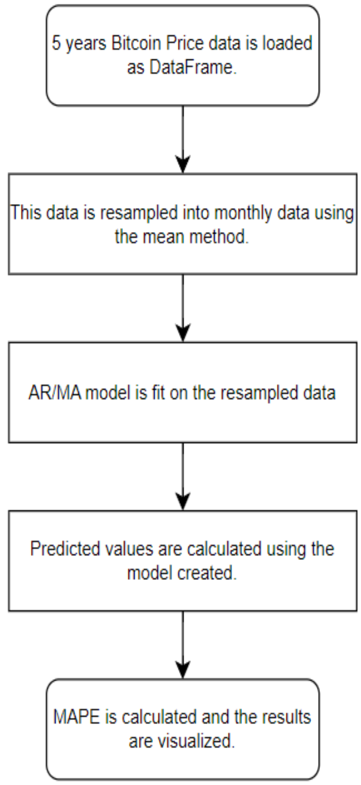


**Fig 1:** Flow of work for AR and MA models

```
Start

1. Load the Bitcoin price data from the CSV file into a DataFrame.
   a. Convert the date column to datetime format.
   b. Set the date column as the index.

2. Resample the data to monthly frequency:
   a. Group the data by month.
   b. Calculate the mean price for each month.

3. Fit the AR/MA model:
   a. Choose appropriate AR (p) and MA (q) parameters.
   b. Fit the ARIMA model to the resampled data.

4. Predict future values:
   a. Predict the Bitcoin prices for the next n months using the model.
   b. Save the predicted values.

5. Calculate MAPE:
   a. Define a function to calculate MAPE:
      - MAPE = (mean of absolute differences between actual and predicted values) / actual values * 100
   b. Apply this function to the actual and predicted values to get the MAPE score.

6. Visualize the results:
   a. Create a plot showing the actual Bitcoin prices over time.
   b. Plot the predicted Bitcoin prices on the same graph.
   c. Display the calculated MAPE on the plot.

End
```

**Fig 2:** Pseudocode for AR and MA

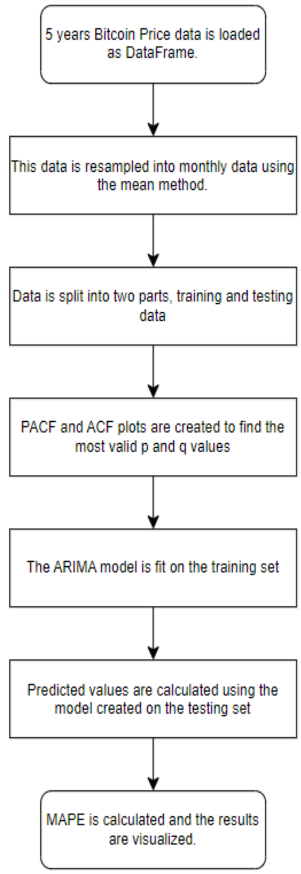


**Fig 3:** Flow of work for ARIMA model

```
Start

1. Load the Bitcoin price data from the CSV file into a DataFrame.
   a. Convert the date column to datetime format.
   b. Set the date column as the index.

2. Resample the data to monthly frequency:
   a. Group the data by month.
   b. Calculate the mean price for each month.

3. Split the data into training and testing sets:
   a. Define a ratio (e.g., 80% for training and 20% for testing).
   b. Split the monthly data accordingly.

4. Create PACF and ACF plots:
   a. Plot the PACF to determine the value of p (AR order).
   b. Plot the ACF to determine the value of q (MA order).

5. Fit the ARIMA model on the training data:
   a. Use the ARIMA model with the chosen p and q values from PACF and ACF.
   b. Fit the model to the training dataset.

6. Predict values on the testing data:
   a. Generate predictions using the model for the test period.
   b. Store the predicted values.

7. Calculate MAPE:
   a. Define a function to calculate MAPE:
      - MAPE = (mean of absolute differences between actual and predicted values) / actual values * 100
   b. Apply the function to the actual testing values and the predicted values.

8. Visualize the results:
   a. Create a plot showing the actual Bitcoin prices for both the training and testing sets.
   b. Plot the predicted Bitcoin prices for the testing period.
   c. Display the calculated MAPE on the plot.

End
```

**Fig 4** Pseudocode for ARIMA

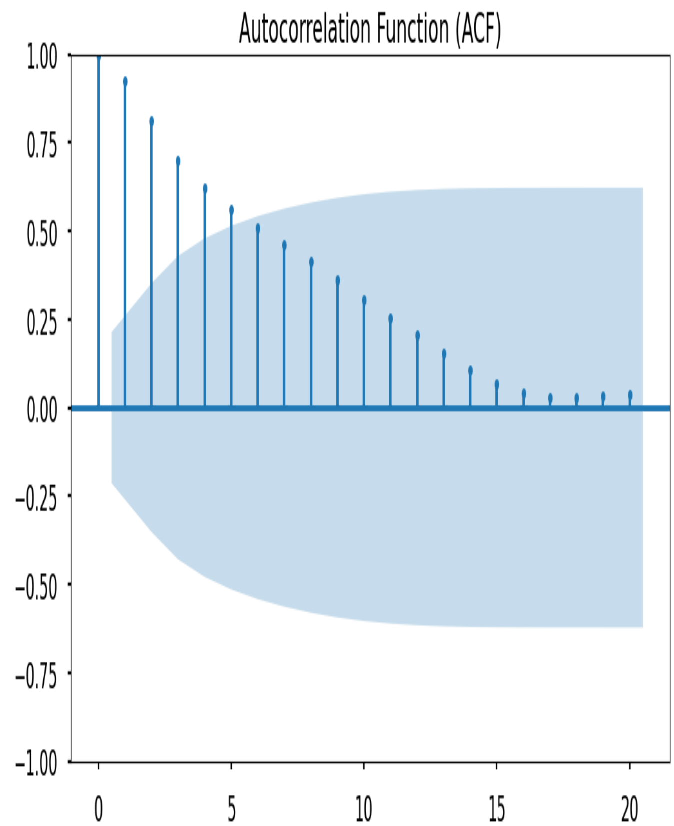


**Fig 5:** Partial Autocorrelation function for ARIMA model

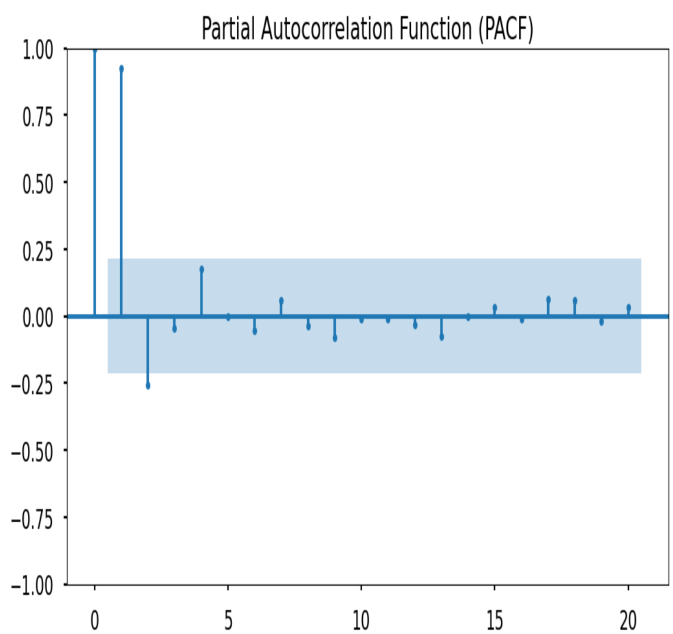


**Fig 6:** Autocorrelation function for ARIMA model

Fig. 5 and Fig. 6 show the PACF and ACF plots, that were used to find *p* (number of lagged observations) and *q* (number of lagged error) values respectively. The *d* value was set to 1 as differencing it once gave a stationary form of the data.

In the Excel calculations, for the AR model the *p* value is considered to be 1. The observations have been shifted by the considered *p* value. The slope and intercept are calculated considering the shifted/lagged values as the fixed variable and the original values as the target variable. This slope and intercept are used to predict the values and MAPE is calculated accordingly.

For the MA model the calculations were done using the average of the past *q* values, in this case 1, these moving averages calculated are the predicted values of the model.

The ARIMA model had a more detailed process than the other two models. First the data was made stationary by differencing the data once. The *phi* and *theta* values are taken from the Python code as these values are estimated using iterative processes and nonlinear equations which are not straightforward to calculate manually. Using these values the first value is predicted. The error is calculated using this predicted value, further values are predicted with the culmination of the current, lagged, error, *phi* and *theta* values.

## 4. RESULTS

The metric of evaluation is taken as MAPE as it's easy to interpret and is not affected by the scale of the values. Also, MAPE is a better metric when the data is positive in nature, and as we are using a price dataset, it plays the perfect role of a good metric of evaluation.

MAPE was calculated from the Python code and validated from the Excel findings. The MAPE calculated from the models are as follows:

**TABLE 1: MAPE calculated from using Python and Excel**

| Model | Python | Excel |
|---|---|---|
| AR | 0.16 | 0.16 |
| MA | 0.14 | 0.14 |
| ARIMA | 0.11 | 0.12 |

The TABLE1. depicts that the code gives a promising output and the values have been verified by the calculations done on Excel. The original and predicted values have also been visualized in Python to get a better understanding of the results. Fig. 7 represents the '*Monthly adj_close*', denotes the actual data, represented with the colour blue whereas the orange and green lines denote the AR and MA models respectively.

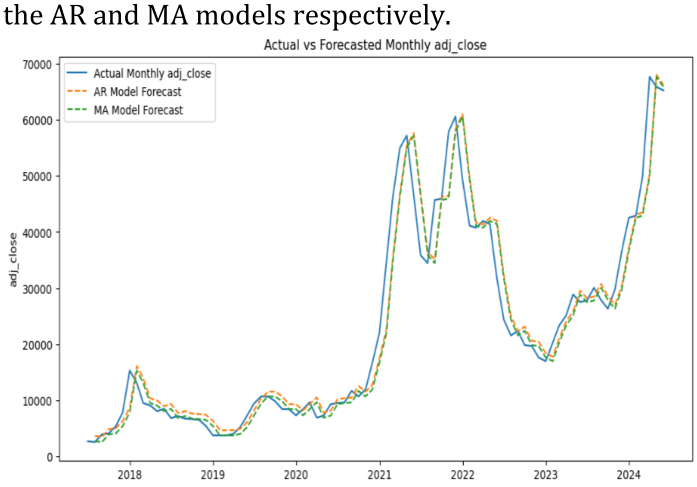


**Fig 7:** Original and forecasted values for AR and MA models

Fig. 8 shows the data for the ARIMA models. The '*BTC history*' represents the training set, '*BTC actual price*' represents the testing set and '*BTC predicted*' are the predicted values, represented with the green, red and blue line respectively.

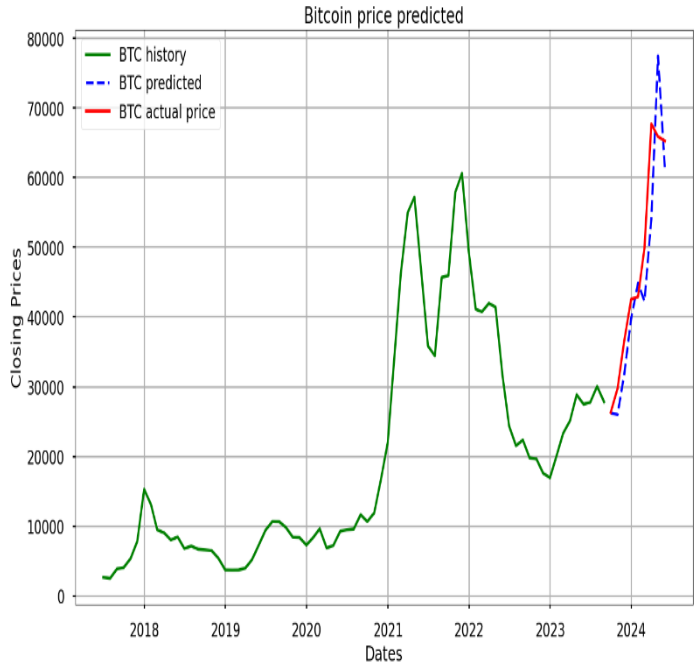


**Fig 8**: Original (Train and Test) and forecasted values for ARIMA models

**TABLE2: Forecast prices calculated from using Python and Excel**

| Date | Actual | Excel | Python |
|---|---|---|---|
| **30-09-2023** | 26306.14 | 26086.07 | 26237.28 |
| **31-10-2023** | 29755.9 | 26298.84 | 26022.07 |
| **30-11-2023** | 36596.16 | 31869.3 | 31788.25 |
| **31-12-2023** | 42546.89 | 39656.35 | 39752.09 |
| **31-01-2024** | 42919.61 | 44560.19 | 44854.1 |
| **29-02-2024** | 49875.17 | 42078.06 | 42285.19 |
| **31-03-2024** | 67702.44 | 54575.53 | 54069.22 |
| **30-04-2024** | 65882.38 | 76106.98 | 77462.81 |
| **31-05-2024** | 65266.32 | 60304.92 | 61205.62 |

TABLE 2. represents the values of the test set with respect to the ARIMA model. The monthly data for the last 9 months present in the dataset were used as a test case. The actual values and the values predicted using Excel and Python are represented. We can also observe the values predicted by Excel and Python are really close proving the correctness of the analysis.

## 5. CONCLUSION

To conclude models and algorithms such as Time Series Analysis, Machine Learning etc. can help investors gain a sense of the market. It must be noted that these are just predictions and the market is more volatile in nature. The predictions made are an aiding tool and not the sole factor to be considered while investing.

This analysis explains in detail how time series analysis is applicable and effective for predicting volatile markets like cryptocurrency. Even with this, Time series models have an assumption that the data needs to be stationary which might not always be true. The use of Python is effective and quick for any more data received in the future, the verification of the model using Excel also proves that analysis is correct and dependable to be used in the future.

## CONFLICT OF INTERESTS

None.

## ACKNOWLEDGMENTS

None.

## REFERENCES

"Yahoo is part of the Yahoo family of brands." https://finance.yahoo.com/quote/BTC-USD/history/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAACf8tC51NXB7TvNjDH3ij6imZ5Tw3Qg1PL6PLK5sNSUbw-grUCCVsINURkGcu2rnAzSywAf0cfNLTc8VeeEi4C426irNlbEESiVqsgKIuLTmj3glh5fiEe52DftkX7Maeuj_rs2UICbuST1CvZ0wruHNPzxMa904SRF0dHYQQlBJ

"8.3 Autoregressive models | Forecasting: Principles and Practice (2nd ed)." https://otexts.com/fpp2/AR.html

"8.4 Moving average models | Forecasting: Principles and Practice (2nd ed)." https://otexts.com/fpp2/MA.html

"8.5 Non-seasonal ARIMA models | Forecasting: Principles and Practice (2nd ed)." https://otexts.com/fpp2/non-seasonal-arima.html

"IPM Insights metrics include MAPE (Mean Absolute Percentage Error)," Oracle Help Center, Sep. 03, 2024. https://docs.oracle.com/en/cloud/saas/planning-budgeting-cloud/pfusu/insights_metrics_MAPE.html

M. Khedmati, F. Seifi, and M. J. Azizi, "Time Series forecasting of Bitcoin price based on autoregressive integrated moving average and machine learning approaches," International Journal of Engineering, vol. 33, no. 7, Jul. 2020, doi: 10.5829/ije.2020.33.07a.16.

Ijraset, "Cryptocurrency Price Prediction and Forecasting Market momentum Using Machine Learning Techniques," IJRASET. https://www.ijraset.com/research-paper/cryptocurrency-price-prediction-and-forecasting-market-momentum

"Cryptocurrency price prediction using supervised machine learning algorithms." https://revistas.usal.es/cinco/index.php/2255-2863/article/download/31490/30390?inline=1

G. Serafini et al., "Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches," 2020. https://www.semanticscholar.org/paper/Sentiment-Driven-Price-Prediction-of-the-Bitcoin-on-Serafini-Yi/ff388671b7038383e721d3d0e8b2f03766bb0ba7

K. Struga and O. Qirici, "Bitcoin Price Prediction with Neural Networks," 2018. https://www.semanticscholar.org/paper/Bitcoin-Price-Prediction-with-Neural-Networks-Struga-Qirici/78227a1267464c132236b0bf25a0db812788b864