CRYPTOCURRENCY PRICE PREDICTION USING ARIMA MODEL

Amritha Subburayan
Engineering and Applied Sciences
University at Buffalo, SUNY
Buffalo, New York
asubbura@buffalo.edu

Kamalnath Sathyamurthy
Engineering and Applied Sciences
University at Buffalo, SUNY
Buffalo, New York
kamalnath@buffalo.edu

Sanjay Aravind Loganathan Ravichandran
Engineering and Applied Sciences
University at Buffalo, SUNY
Buffalo, New York
sanjayar@buffalo.edu

Yaswanth Kumar Reddy Kaduru
Engineering and Applied Sciences
University at Buffalo, SUNY
Buffalo, New York
ykaduru@buffalo.edu

***Abstract*— Cryptocurrency is a new form of asset which is used for paying or investing digitally. The upsurge in the various cryptocurrency prices over the past 10 years has increased the curiosity of researchers and investors to analyze and forecast its prices in the future. The primary focus of this research is to develop an Autoregressive Integrated Moving Average (ARIMA) model, which is a time-series statistical model, for forecasting the cryptocurrency prices. This research focuses on six different cryptocurrencies – Bitcoin, DogeCoin, Ethereum, Binance Coin, XRP and Cardano due to their popularity, and discusses the price movements and the stability of these cryptocurrencies using various exploratory data analysis techniques and visualization techniques. The dataset for these six coins has been collected from Kaggle datasets and has been merged into a single dataset. In this research, the dataset has a record of the prices of these six cryptocurrencies dated from 2018 to 2021. Finally, the paper discusses the results of the time-series model, broader impacts of this research and the future enhancements that can be done.**

Keywords—time-series forecasting, ARIMA, cryptocurrencies, price prediction, blockchain.

1. INTRODUCTION

Cryptocurrency is a form of digital asset based on a network that is distributed across a large number of computers by the means of blockchain technology. This decentralized structure allows them to exist outside the control of governments and central authorities which offers anonymity for all the users. All the transactions and exchange information regarding cryptocurrencies are stored in a general ledger which requires owners' authorization to perform any operation on cryptocurrency. Due to tremendous increase in online transactions, it led to an increase in demand for cryptocurrencies. With the growth of cryptocurrency transactions, the price of all cryptocurrencies is increasing tremendously. First ever cryptocurrency was implemented in 2009 ('Bitcoin'). At the time of implementation, the value of bitcoin was worth only 0.0008 to 0.008 US dollars. But as of today, the price of each bitcoin is almost 63000 US dollars. This increase led to new investment opportunities. Predicting the future stake of cryptocurrency helps common people to invest better. Our focus in this research paper is to use the application of Auto Regressive Integrated Moving Average (ARIMA) in time series forecasting to predict the future stake of six Cryptocurrencies – Bitcoin, DogeCoin, Ethereum, Binance Coin, XRP and Cardano. Even though LSTM (Long-Short Term Memory) is effective in the time series forecasting model, it is not exemplary when it comes to data which are not stationary. ARIMA model uses the approach of differencing in order to make data stationary and predicts the future trend of cryptocurrency more accurately than other Recursive Neural Networks or LSTM. Our research differs in the idea of more accurately forecasting the trend of multiple cryptocurrencies. We combined individual dataset of each cryptocurrency into a single dataset and formulated to get the opening and closing price of each cryptocurrency for every day from multiple time frame dataset. With this dataset, we aimed at fully utilizing the effectiveness of ARIMA by using both p (order of the auto-regressive (AR)), d (degree of differencing) and q (order of the moving average (MA)) values for regular forecasting and P, D and Q which is for seasonal forecasting of same cryptocurrencies which helps in forecasting more accurately.

2. MOTIVATION

Bitcoin, the first cryptocurrency, was invented in 2008 by an anonymous group or person named Satoshi Nakamoto. It was first publicly released as open-source software in 2009 and got first traded in 2010 with as low as $0.0008 per coin. The highest, at that time, was just $0.08 per coin. The first significant surge in price occurred only two years after the first introduction when the Electronic Frontier Foundation (EFF) accepted Bitcoin for donations. Despite the foundation revoking that decision, Bitcoin's price touched 1 US dollar. The price went up to 100 times the year's starting price. But by the end of the year 2011, the price plummeted to just under $5 per coin. In the same year (2011), XRP, one of the cryptocurrencies, was introduced based on the protocol developed in 2004 by Ryan Fugger. The coin was also called Ripple Coin.

In 2013, Bitcoin saw a rise from $13 to $1,100, and the investors' gains were almost 6,600%. DogeCoin was first introduced to the market on December 6, 2013. It was created by Billy Markus and Jackson Palmer, who took inspiration from an online meme and started the oin as a joke.

In 2015, Ethereum, which was based on a paper authored by Vitalik Buterin, was introduced to the general public.
In 2017, Bitcoin experienced a second significant surge, rising from $1,100 to $20,000. This rise of 20x times in less than 12 months was partly because the public got aware of the cryptocurrency. In the same year, Cardano and Binance Coin, other cryptocurrencies were also introduced to the public.

In 2021, more than ten years after Bitcoin introduction, it was the most dominant in the cryptocurrency market with one coin priced at around $60,000. The highest price ever recorded for Bitcoin was on November 5, 2021: $68,521 per coin. Ethereum is the second according to the market capitalization and one coin is priced at around $4,309.55. The highest price ever recorded for Ethereum was in November 2021: $4,800 per coin. Binance Coin stands third in the market with the current price per coin around $578.65. The highest price ever recorded for Binance Coin was $690.93. While Cardano, XRP & DogeCoin, stands at sixth, eighth & eleventh positions, with the current price around $1.39, $0.81 & $0.8206 respectively.

Cryptocurrency	Date of introduction	Market Capitalization	Volume	Circulating supply	Current price
BitCoin	January 2009	$953,439,434,119	686,385 BTC	18,894,312 BTC	$50,275.75
XRP	May 2011	$38,422,177,457	3,441,394,521 XRP	7,247,295,769 XRP	$0.8132
DogeCoin	December 2013	$23,498,884,769	6,805,158,473 DOGE	132,426,963,311 DOGE	$0.1777
Ethereum	July 2015	$507,852,465,184	5,472,233 ETH	118,645,865 ETH	$4,277.08
Cardano	September 2017	$46,351,813,922	1,696,781,923 ADA	33,367,268,119 ADA	$1.39
Binance Coin	July 2017	$96,322,966,986	4,038,251 BNB	166,801,148 BNB	$576.64

Note: Market Capitalization: Current Price * Circulating Supply. Where, Circulating Supply is the number of coins that are in public hands. Volume is how much of a cryptocurrency was traded in the last 24 hours. No one could possibly predict what's going to happen in the near-by future. Either cryptocurrencies can thrive to become a commonly used digital currency and people could even use it to buy a cup of coffee. Or the government regulations can hinder cryptocurrencies' further growth and could collapse the entire industry. Considering the rise in cryptocurrency popularity and usage, it has become one of the investors' favorite choices to make lucrative profits. While the future remains an oblivion, the current technology advancement with regard to machine learning can be useful to make fair predictions. This enlightens investors about when's the right time to invest, when to withhold their investments, and when to exit the market to avoid loss. Thus, enabling investors to understand the risks and make better decisions.

3. METHODOLOGY

3.1. DATA COLLECTION

Source: Kaggle Crypto-Currency dataset.

Crypto-Currency dataset by Kaggle consists of 244855 observations of six popular cryptocurrency details with 6 independent and 1 dependent variables.

```
    Date              Open               High               Low                Close              Volume
 Length:244855      Min.   :    0.00   Min.   :    0.00   Min.   :    0.00   Min.   :    0.00   Min.   :       0
 Class :character   1st Qu.:    0.27   1st Qu.:    0.27   1st Qu.:    0.27   1st Qu.:    0.27   1st Qu.:       0
 Mode  :character   Median :   23.62   Median :   23.67   Median :   23.53   Median :   23.62   Median :       6
                    Mean   : 3482.03   Mean   : 3496.36   Mean   : 3466.72   Mean   : 3481.92   Mean   :   13060
                    3rd Qu.: 1646.12   3rd Qu.: 1654.73   3rd Qu.: 1636.91   3rd Qu.: 1646.16   3rd Qu.:    1287
                    Max.   :64593.00   Max.   :64755.00   Max.   :64366.00   Max.   :64623.00   Max.   :20084416
 CryptoCurrency_Name
 Length:244855
 Class :character
 Mode  :character
```

Figure 1: Summary of the dataset

Below is the detailed description on variables of the dataset.

Date: Crypto-currency price on a given day.
Open: Indicates the starting price before the trading.
High: Indicates how high the price went on given a timestamp.
Low: Indicates how low the price went on a given timestamp.
Close: Final price after trading on a given day.
Volume: Gives Volume of transactions on a given day.

Here, Close is the dependent variable, which needs to be predicted with the support of other independent variables.

3.2. PRE-PROCESSING

All the six crypto-currencies datasets are merged to one dataset where a column indicating cryptocurrency name is created for clear understanding of the trends.

	Date	Open	High	Low	Close	Volume	CryptoCurrency_Name
1	2019-06-19 09:30:00	0.09097	0.09105	0.09063	0.09063	38996.0	CARDANO
2	2019-06-19 10:00:00	0.09063	0.09080	0.09050	0.09080	11653.5	CARDANO
3	2019-06-19 10:30:00	0.09080	0.09080	0.09055	0.09055	2117.9	CARDANO
4	2019-06-19 11:00:00	0.09055	0.09056	0.09034	0.09042	15849.2	CARDANO
5	2019-06-19 11:30:00	0.09042	0.09055	0.09037	0.09055	1810.3	CARDANO
6	2019-06-19 12:00:00	0.09055	0.09063	0.09040	0.09040	3101.8	CARDANO
7	2019-06-19 12:30:00	0.09040	0.09040	0.09000	0.09023	64563.3	CARDANO
8	2019-06-19 13:00:00	0.09023	0.09042	0.09000	0.09042	16159.8	CARDANO
9	2019-06-19 13:30:00	0.09042	0.09042	0.08987	0.08988	4235.4	CARDANO
10	2019-06-19 14:00:00	0.08988	0.08988	0.08950	0.08955	5383.1	CARDANO
11	2019-06-19 14:30:00	0.08928	0.08928	0.08888	0.08888	3523.7	CARDANO
12	2019-06-19 15:00:00	0.08888	0.08888	0.08888	0.08888	283.6	CARDANO

Showing 1 to 13 of 244,855 entries, 7 total columns

Figure 2: List of variables and its values

Outliers are identified in the dataset which affects the mean of the data by skewing the data to either left or right thus making the predictions biased to one solution. Hence, few outliers are removed from the dataset with the help of boxplot.

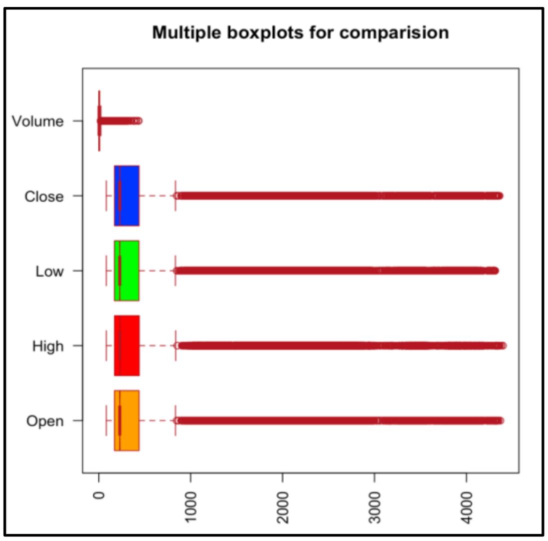

Figure 3: Outliers in cryptocurrency dataset

Above boxplot, shows the presence of outliers in a given dataset.

Feature scaling is an important step in pre-processing the dataset. Feature scaling needs to be done when features have high variability in magnitude between the values with other features. Feature scaling is applied for volume variables since the maximum value is far away from other independent variables. This can normalize the features to a finite range in the dataset.

```
> summary(data8$Volume)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
       0        0        6    13060     1287 20084416
> summary(data8$Close)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
    0.00     0.27    23.62  3481.92  1646.16 64623.00
```

Figure 4: Before Feature Scaling

```
> summary(data8)
      Date                 Open.V1              High.V1              Low.V1
 Min.   :2018-07-10   Min.   :-0.370339   Min.   :-0.370057   Min.   :-0.370676
 1st Qu.:2019-08-22   1st Qu.:-0.370311   1st Qu.:-0.370029   1st Qu.:-0.370648
 Median :2020-05-22   Median :-0.367827   Median :-0.367552   Median :-0.368160
 Mean   :2020-04-12   Mean   : 0.000000   Mean   : 0.000000   Mean   : 0.000000
 3rd Qu.:2020-12-29   3rd Qu.:-0.195263   3rd Qu.:-0.194919   3rd Qu.:-0.195651
 Max.   :2021-07-29   Max.   : 6.499608   Max.   : 6.483671   Max.   : 6.511619
      Close.V1             Volume.V1         CryptoCurrency_Name
 Min.   :-0.370356   Min.   : -0.07640   Length:244855
 1st Qu.:-0.370328   1st Qu.: -0.07640   Class :character
 Median :-0.367844   Median : -0.07637   Mode  :character
 Mean   : 0.000000   Mean   :  0.00000
 3rd Qu.:-0.195261   3rd Qu.: -0.06887
 Max.   : 6.503301   Max.   :117.42392
```

Figure 5: After Feature Scaling

There are two types of Feature Scaling techniques,
1) Normalization (values are scaled between 0 and 1)
2) Standardization (Values are transformed such that the mean value is 0 and the standard deviation is 1)

3.3. EXPLORATORY DATA ANALYSIS

Closing price for the past three years has been analyzed for six cryptocurrencies. Even though the prices are not consistent, Bitcoin value is found to be higher compared to other cryptocurrencies. With this plot, it is clearly stated that Bitcoin is popular among all other cryptocurrencies.

Figure 6: Relationship between Date and Closing Price

The figure 6 explains the historical price movements of the 6 cryptocurrencies and helps us to analyze the trends in the variations.

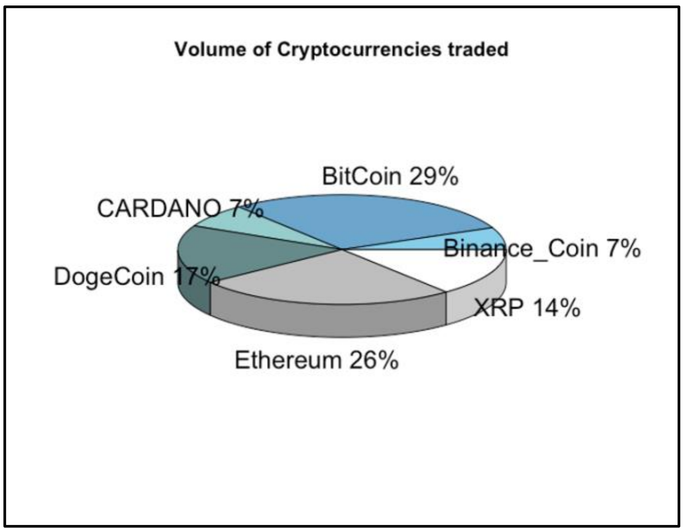

Figure 7: Volume of Cryptocurrencies traded

This pie chart represents which cryptocurrency has been branded highly in the past. We can see that the Bitcoin has accounted for 29% of trading which is the highest and Cardano, Binance Coin, DogeCoin account for the lowest percentage of trading.

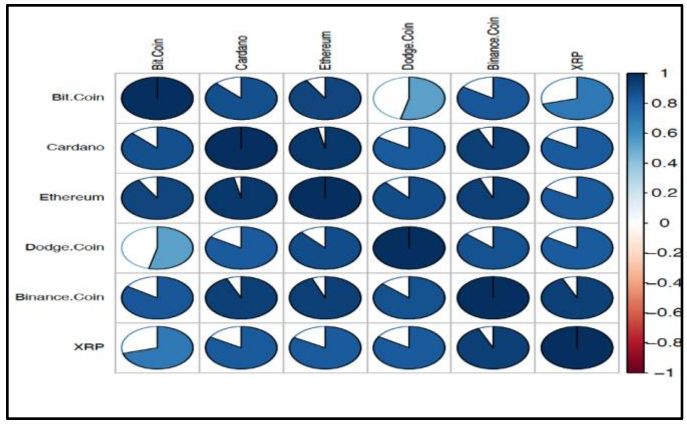

Figure 8: Correlation between each cryptocurrency

This correlation matrix (Figure 8) represents the relationship between different cryptocurrencies. For example, the correlation between Cardano and Ethereum are high, which indicates that the prices of Ethereum have increased whenever prices of Cardano have increased and this will be a helpful analysis for the investors.

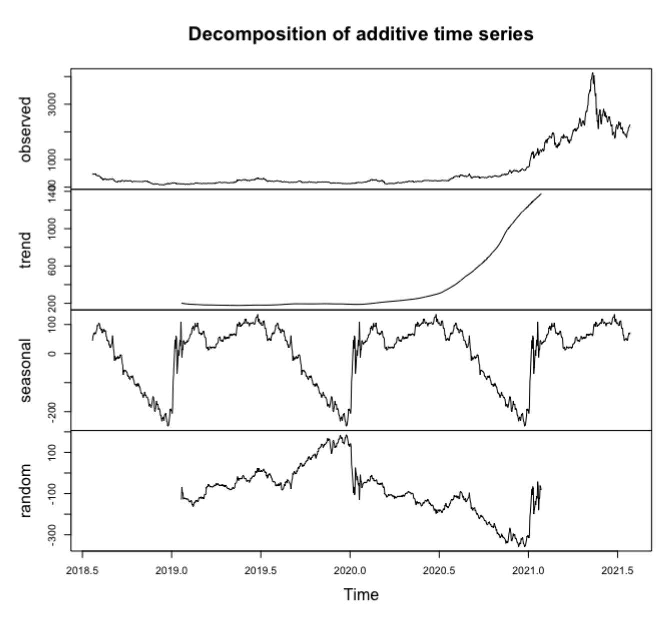

Figure 9: Additive time series decomposition for Bitcoin

Additive time series decomposition graph (Figure 9) explains about the trend, seasonal and random data in the dataset.

3.4. MACHINE LEARNING MODEL

The proposed method involves analyzing and building different machine learning models like supervised learning methods to predict the prices of various crypto currencies for the given period. Here, a Time-Series forecasting model will be initiated which helps to capture data at every interval.

This project involves using a very popular time series forecasting model – Autoregressive Integrated Moving Average (ARIMA) model. ARIMA model is the statistical model used for forecasting the time series. The ARIMA model is characterized by below = 3 terms,

- p is the number of autoregressive terms,
- d is the number of nonseasonal differences for stationary series
- q is number of moving average terms

p refers to AR (Auto Regressive), which gives the number of lags of Y to be used as predictors. q refers to MA (Moving Average), which gives a number of lagged forecast errors. d gives a minimum number of differencing to make the series stationary. When the series is already stationary then we initialize it as 0. The model needs the data to follow a stationary process, where covariance, mean and autocorrelation of the series should not change over the period.

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

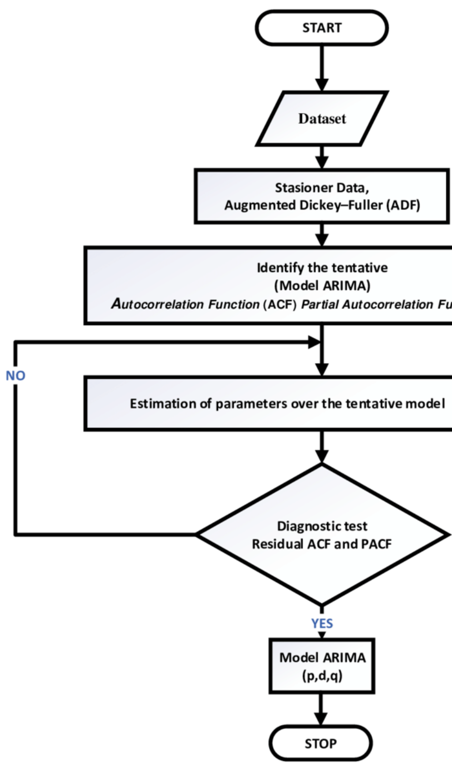

Figure 10: Arima Model Algorithm

The values for p and q can be determined using the sample autocorrelation function (ACF), partial autocorrelation function (PACF), and/or Extended autocorrelation function (EACF).

Autocorrelation function (ACF)

It is a complete auto-correlation function which gives the value of autocorrelation of any series with its lagged values. The below figure represents the ACF plot for the dataset used for this research.

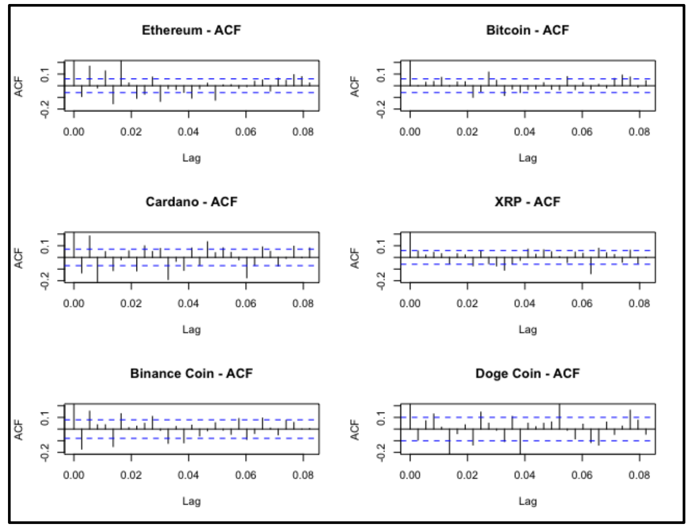

Figure 11: ACF plot

Partial Autocorrelation Function (PACF):

PACF is a statistical measure that measures the correlation between two variables after controlling the effects of the variables. This helps to calculate direct correlation between the time series and a lagged version of itself.

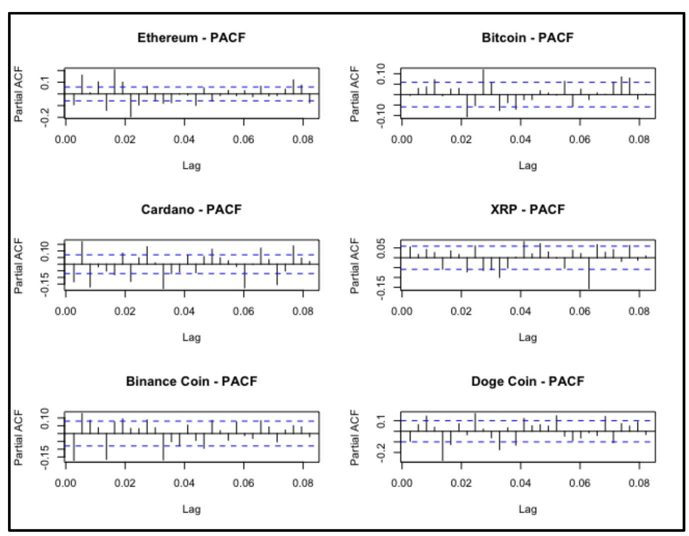

Figure 12: PACF Plot

Stationarity:

For any time-series forecasting model, stationarity is the most important concept an algorithm needs before modelling.
A Time-Series is considered as stationary if below conditions are met,

1) A time Series says St has constant mean.
2) A time Series says St has constant Standard Deviation.
3) If there is no seasonality in St. If St has a repeating pattern within a year, then it has seasonality.

Stationarity can be checked with the help of the hypothesis test.
Augmented Dickey-Fuller Test (ADF):
ADF is a statistical test used to test whether a time series is stationary or not. This can be achieved with the help of null and alternative hypotheses.

a) Null Hypothesis: Time Series is stationary. This gives a time-dependent trend.
b) Alternate Hypothesis: Time-Series is non-stationary. This doesn't depend on time.
c) If ADF < critical values (5%): Accept the null hypothesis. Hence the time series is stationary.
d) If ADF > critical values (5%): Failed to reject null hypothesis. Hence the time series is non-stationary.

After having a stationary time series, seasonality check is essential to the accuracy of the forecast.

Seasonality

Seasonality is not present in this time series hence application of Seasonal ARIMA model is not required for this time-series.

4. RESULTS

R Studio has an in-built function called auto.arima() that automatically finds the p, d, and q parameters for the ARIMA model. This function builds the model and can be used to forecast the future prices.

```
> (close_arima_btc <- auto.arima(close_ts_btc,D=1))
Series: close_ts_btc
ARIMA(0,1,0)(0,1,0)[365]

sigma^2 estimated as 1219014:  log likelihood=-6297.48
AIC=12596.96   AICc=12596.96   BIC=12601.57
```

Figure 13: ARIMA Model and its AIC summary

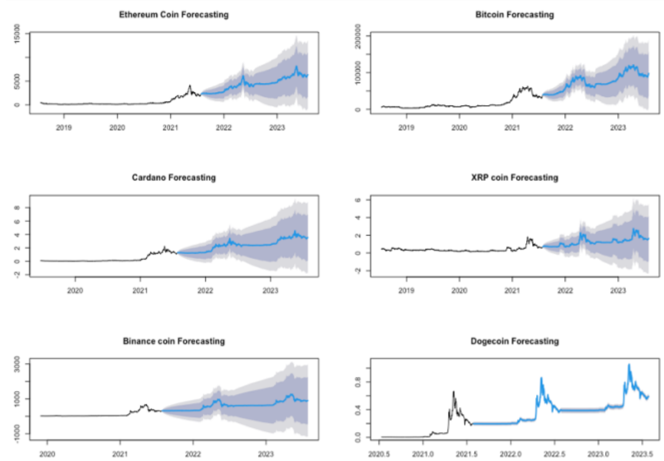

Figure 14: Trends in forecasted prices

The figure below represents the price movements of the six cryptocurrencies. The blue line indicates the trends in future prices.

Model Performance Evaluation:
In this paper, the other useful criterion is Akaike Information Criterion (AIC), which is used to determine the order of a non-seasonal arima model. It is given by:

$$\mathbf{AIC = -2Log(L) + 2 * (p + q + k)}$$

Where L is the likelihood of the data and k is the intercept of the ARIMA model.
The objective is to minimize this AIC value to obtain a better model. Lower the AIC value, better the model.

5. CONCLUSION AND FUTURE WORK

Considering the rise in cryptocurrency popularity and usage, it has become one of investors' favorite choices to make lucrative profits. The only drawback is that cryptocurrencies are fiat-currencies, and their market is highly volatile. Investors need to know when's the right time to invest, when to withhold their investments, and when to exit the market to avoid loss. By increasing the accuracy of the prediction model of the cryptocurrency, it helps common people and investors make better strategies and to invest better. In the prediction of cryptocurrencies prices, we are able to forecast prices for the only six popular cryptocurrencies for next two years using the traditional time series forecasting model **(ARIMA).**

The future work involves finding the prices for all the cryptocurrencies and analyzing the trends in variations. Also, implementing the gradient descent model, XGBoost model, can be used to predict prices for longer periods of time series for the complex multivariate time series models.

6. REFERENCE

[1] *Xue Tan, Rasha Kashef, "Predicting the Closing Price of CryptoCurrencies: A Comparative Study,", Data 2019, Dubai (references)*

[2] *Yan Pang, GaneshKumar Sundararaj, Jiewen Ren, "Cryptocurrency Price Prediction using Time Series and Social Sentiment Data" BDCAt'19, December 2-5. 2019, Auckland, New Zealand. (references)*

[3] *AbedAlqader Sweden, Ahmad N. Khuffash, Othman Othman M.M., Ahmed Awad, "Detection and Prevention of Malicious Cryptocurrency Mining on Internet-Connected Devices", ICFNDS'18, June 26–27, 2018, Amman, Jordan (references)*

[4] *Wang Zhengyang, Li Xingzhou, Ruan Jinjin, Kou Jiaqing, "Prediction of Cryptocurrency Price Dynamics with Multiple Machine Learning Techniques", (references)*

[5] *Shenjia Ji, Hongyan Yu, Yinan Guo, Zongrun Zhang: "Research on Sales Forecasting Based on ARIMA and BP Neural Network Combined Model"*

[6] *Peter T Yamak, Li Yujian, Pius K Gadosey: "A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting", ACAI'19, December 2019, Sanya, China*

[7] *Yangyang Xie, Yuansheng Lou, "Hydrological Time Series Prediction*

by ARIMA-SVR Combined Model based on Wavelet Transform"

[8] *Vinay B Gavirangaswamy, Gagan Gupta, Ajay Gupta, Rajeev Agrawal, "Assessment of ARIMA-based Prediction Techniques for Road-Traffic Volume"*

[9] *Zhixian Yan, "Traj-ARIMA: A Spatial-Time Series Model for Network-Constrained Trajectory"*

[10] *"List_of_cryptocurrencies." Wikipedia, Wikimedia Foundation, [Online]., Available: en.wikipedia.org/wiki/List_of_cryptocurrencies. [Retrieved 7 Dec 2021.]*

[11] *"Bitcoin." Wikipedia, Wikimedia Foundation, [Online]., Available: en.wikipedia.org/wiki/Bitcoin. [Retrieved 7 Dec 2021.]*

[12] *"Dogecoin." Wikipedia, Wikimedia Foundation, [Online]., Available: en.wikipedia.org/wiki/Dogecoin. [Retrieved 7 Dec 2021.]*

[13] *"Cardano_(blockchain_platform)." Wikipedia, Wikimedia Foundation, [Online]., Available: en.wikipedia.org/wiki/Cardano_(blockchain_platform). [Retrieved 7 Dec 2021.]*

[14] *"Ethereum." Wikipedia, Wikimedia Foundation, [Online]., Available: en.wikipedia.org/wiki/Ethereum. [Retrieved 7 Dec 2021.]*

[15] *"Ripple." Wikipedia, Wikimedia Foundation, [Online]., Available: en.wikipedia.org/wiki/Ripple. [Retrieved 7 Dec 2021.]*

[16] *"CoinMarketCap: Cryptocurrency Prices, Charts and Market Capitalizations" CoinMarketCap Foundation, [Online]., Available: https://coinmarketcap.com/. [Retrieved 7 Dec 2021.]*

[17] *"The 4 Possible Scenarios for Bitcoin In The Long Term." Medium, A Medium Corporation, Nov 22, 2019, [Online]., Available: https://medium.com/swlh/the-4-possible-scenarios-for-bitcoin-in-the-long-term-b1e93298b347. [Retrieved 7 Dec 2021.]*

[18] *"Choosing the best q and p from ACF and PACF plots in ARMA-type modeling" [Online]., Available: https://www.baeldung.com/cs/acf-pacf-plots-arma-modeling/. [Retrieved 7 Dec 2021.]*