Predicting Cryptocurrency Prices Using Various Statistical Techniques

Pritish Sharma (19210582), Mitul Verma (19210961)

MSc in Computing (Data Analytics)

pritish.sharma3@mail.dcu.ie

mitul.vermam3@mail.dcu.ie

Abstract— **In this paper, we would be comparing various statistical techniques to predict the cryptocurrency prices. We employed Simple Moving Average, ARIMA and Long Short Term Memory (LSTM) for price prediction of three major cryptocurrencies. We aim to gain insights into the statistical methods that best describe the model for cryptocurrency price prediction. These insights could help crypto investors to refine their investing methods.**

***Keywords*— Cryptocurrency, Blockchain, Bitcoin, Ethereum, Ripple, LSTM, ARIMA.**

I. Introduction

Cryptocurrencies have captured the imagination of many around the globe, evident by their ever-increasing numbers, rising adoption, and jaw-dropping market capitalizations. However, the outlook towards the cryptocurrencies is not all that merry. There is considerable scepticism and many regard the current cryptocurrency trend as irrational and unsustainable. This mix of optimism and distrust in cryptocurrencies is visible in the wild price fluctuations. As investors seek to maximise their returns, the statistical analysis on the various factors impacting the crypto prices is being carried out extensively. We aim to compare and analyse the price prediction results obtained from Moving Average Technique, LSTM and ARIMA for three largest cryptocurrencies by market cap – Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP).

II. Related Work

It has been demonstrated that LSTM could yield accuracy as high as 27.2% when applied on stock data [1].

ARIMA is being increasingly used for predictions in many time series areas such as weather forecast and crop production [2][3].

Researchers have demonstrated that cryptocurrency price predictions could be improved by taking into account the sentiments expressed on social media [4].

III. Dataset and Exploratory Analysis

The dataset for three major cryptocurrencies – Bitcoin, Ethereum, and Ripple was acquired from Coin Market Cap [5]. The data was collected from January 1, 2017, to December 10, 2019.

A. Descriptive stats on data

Total of 1078 records are present for each of the currency. Below is the table describing the statistical properties of ‘Closing Price’ from Jan 1, 2017, to Dec 10, 2019.

TABLE I
STATISTICS ON CLOSING PRICES

Closing Price	BTC	ETH	XRP
Mean	6310.316	299.0179	0.396394
Std	3559.159	246.3541	0.37363
Min	777.76	8.17	0.005408
Max	19497.4	1396.42	3.38

B. Analysis of distribution of daily percent change in prices

Cryptocurrencies are known for their volatility. They have shown the tendency to swing in either direction by as much as 20% in one day. Enormous changes in daily price makes it hard to apply conventional machine learning models to make price prediction. Below are the boxplots for all three crypto currencies showing the distribution of the daily percent change in price.

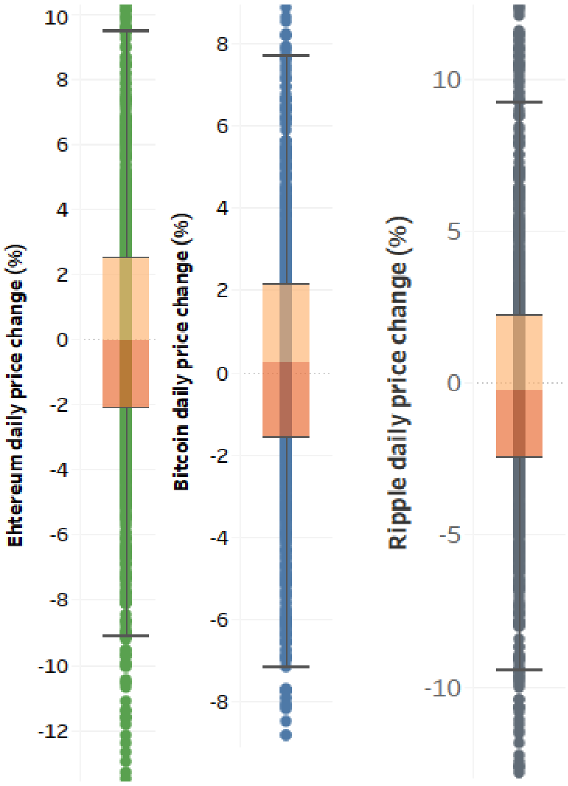

Fig. 1 Boxplot of daily percent change for ETH, BTC and XRP

It is evident from the boxplots that there are large number of outliers. These outliers in the historical data are a major concern while applying prediction algorithms.

IV. METHODS

A. *Moving Average*

For all three currencies, there was a major period of volatility between the second half of 2017 and first half of 2018. A Major proportion of the outliers in the Fig. 1 come from that specific time period. Additionally, cryptocurrencies often have short periods of wild swings which skew the net average.

Simple Moving Average (SMA) provides an elegant and simple solution to immunize against the short-term volatility.

We coded a 200-Day SMA and fed closing price data of 800 days for each currency to predict the future price. We used the closing price of next 265 days to calculate our Root Mean Square Error (RMSE)

Results of Moving Average

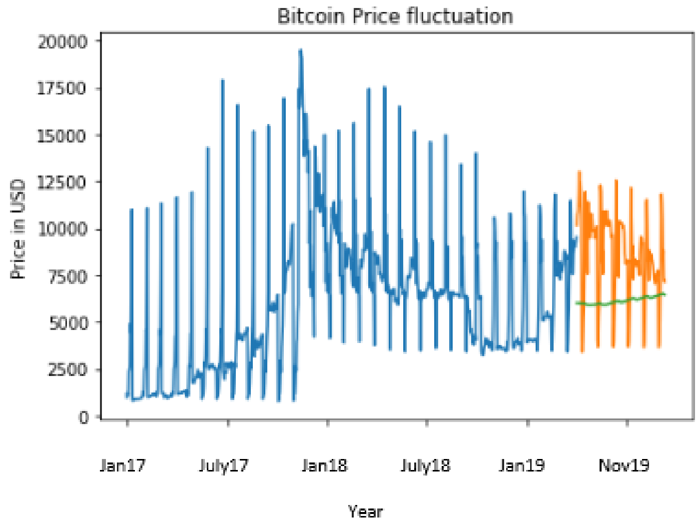

Fig. 2 Moving Average applied to Bitcoin closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

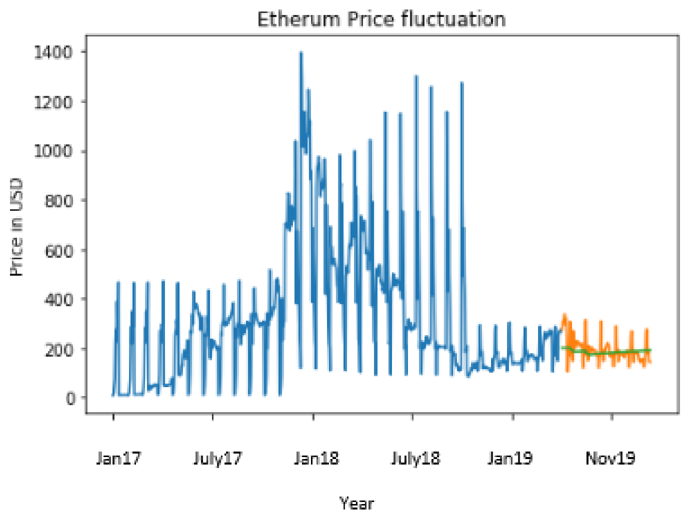

Fig. 3 Moving Average applied to Ethereum closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

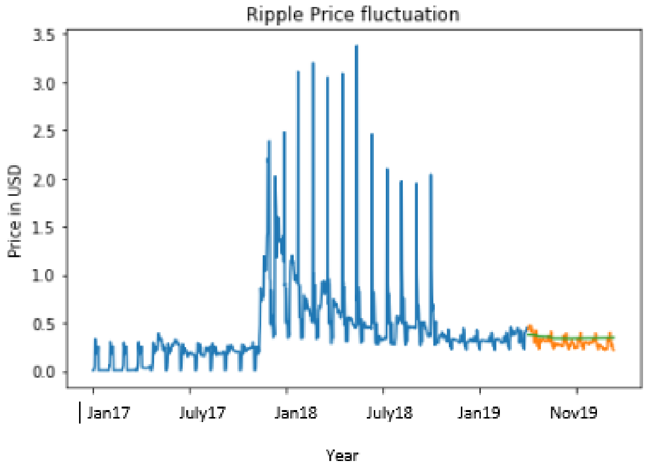

Fig. 4 Moving Average applied to Ripple closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

TABLE III
ROOT MEAN SQUARE ERROR FOR SIMPLE MOVING AVERAGE

	BTC	ETH	XRP
RMSE	3052.34	44.73	0.068

B. *ARIMA*

Since our data is univariate time-series (only taking into consideration the daily closing price of each crypto currency), Auto-Regressive Integrated Moving Averages (ARIMA) fits our objectives well.

By applying transformation to the closing price data to make it stationary, our data fulfils second condition necessary to apply ARIMA. For auto-tuning of ARIMA parameters and pre-processing of data to stationery state, we used Auto ARIMA.

We fed closing price data of 800 days for each currency to Python's implementation of Auto ARIMA to predict the future price. We used the closing price of next 265 days to calculate our Root Mean Square Error (RMSE).

Results of ARIMA

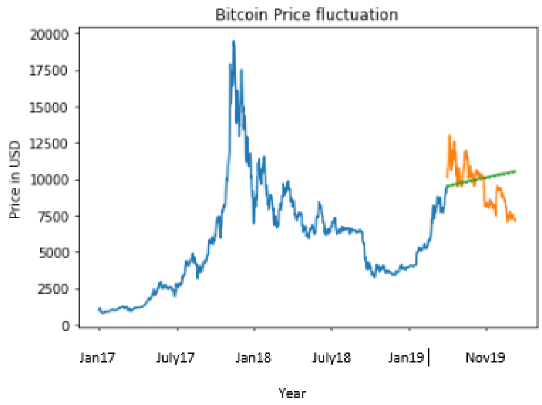

Fig. 5 ARIMA applied to Bitcoin closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

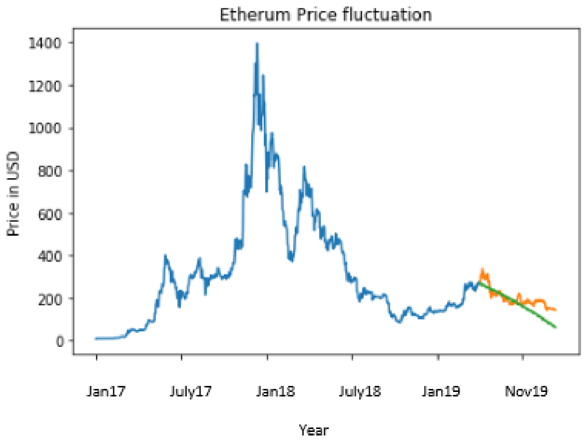

Fig. 6 ARIMA applied to Ethereum closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

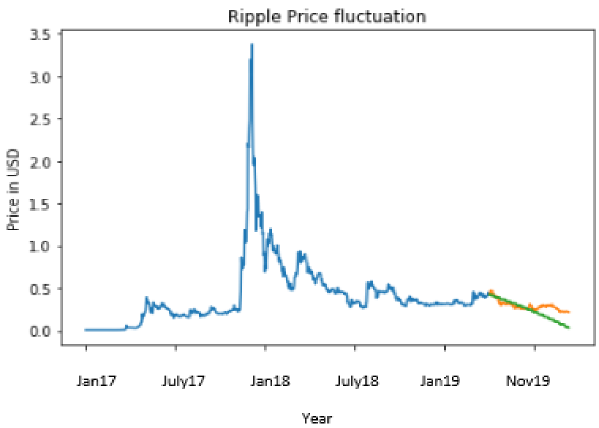

Fig. 7 ARIMA applied to Ripple closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

TABLE IIIII
ROOT MEAN SQUARE ERROR FOR ARIMA

	BTC	ETH	XRP
RMSE	1502.39	45.86	0.095

C. Long Short-Term Memory

The major challenge to predict the price of cryptocurrencies or any timer-series data for that matter is to develop a way to consider significant events of the past and neglect the insignificant ones. In SMA and ARIMA, this factor was not taken into consideration.

LSTM is a sublime solution to the problem. With backpropagation through time and layers, LSTMs, in extremely simplified terms, retain the memory of the significant events which allows for better prediction.

We fed closing price data of 800 days for each currency to Python's implementation of LSTM to predict the future price. We used the closing price of next 265 days to calculate our Root Mean Square Error (RMSE).

Results of LSTM

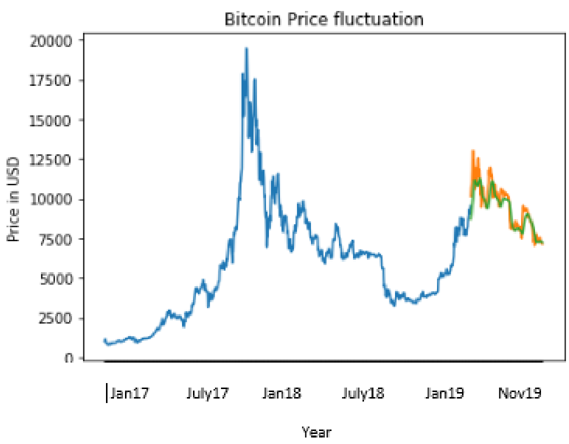

Fig. 8 LSTM applied to Bitcoin closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

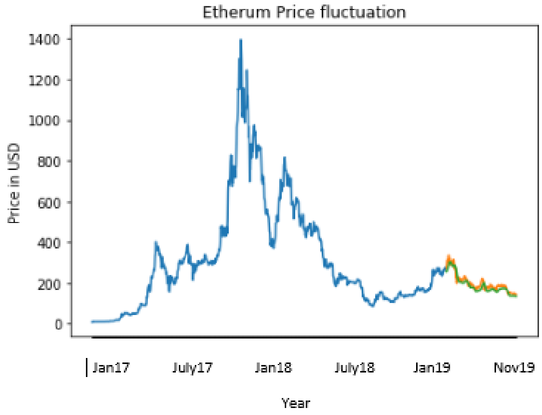

Fig. 9 LSTM applied to Ethereum closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

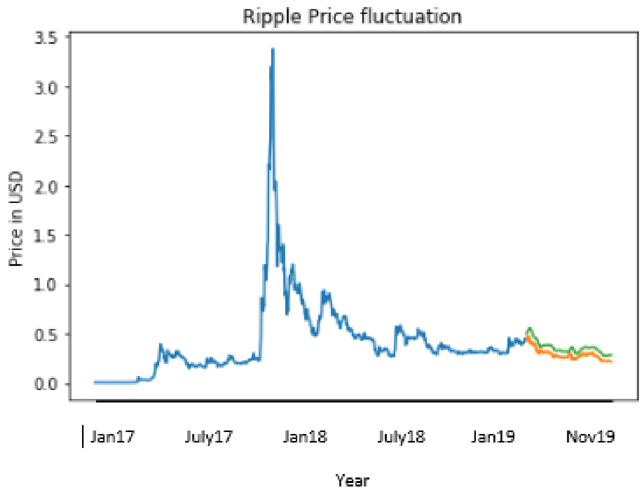

Fig. 10 LSTM applied to Ripple closing prices. Blue indicates the training data, orange indicates the test data and green indicates the prediction

TABLE IVV
ROOT MEAN SQUARE ERROR FOR LSTM

	BTC	ETH	XRP
RMSE	722.57	18.26	0.071

V. RESULTS AND FINDINGS

The following findings were made in the research paper.

1. LSTM vastly outperformed other techniques.

TABLE V

	BTC	ETH	XRP
RMSE for Simple Moving Average	3052.34	44.73	0.068
RMSE for ARIMA	1502.39	45.86	0.095
RMSE for LSTM	722.57	18.26	0.071

LSTM's superior prediction performance is explained by the incorporation of the memory of the significant events of the past. Both SMA and ARIMA had similar performance for Ethereum and Ripple, but ARIMA outperformed SMA significantly in the case of Bitcoin.

2. Bitcoin is significantly harder to predict than other two currencies. This could be explained by large standard deviation in Bitcoin's closing prices. This greatly affects the prediction capability of the algorithms.

3. Although LSTM significantly improved prediction, traditional algorithms SMA and ARIMA also performed well on Ripple and Ethereum. Both Ripple and Ethereum are significantly less volatile than bitcoin and therefore, with fewer outliers, prices of both Ethereum and Ripple could be predicted better.

VI. Conclusions

Due to rapidly evolving crypto environment, changing regulations, and prevailing uncertainty, it is hard to predict the prices. We showcased that using both traditional (SMA and ARIMA) and more modern (LSTM) techniques could predict the prices to a certain extent, with LSTM producing the best predictions. By taking into consideration sentiments expressed on the social media and news, and by incorporating other external factors such as stock and commodity prices into the prediction model, the results could be further improved. These techniques could help the investors in optimization of their crypto investments. The code used in this assignment is available on GitHub[6].

References

[1] K. Chen, Y. Zhou and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market - IEEE Conference Publication", *Ieeexplore.ieee.org*, 2015. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7364089.

[2] G. Jain and B. Mallick, "A Study of Time Series Models ARIMA and ETS", 2017. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2898968.

[3] B. BISWAS, L. DHALIWAL, S. Singh and S. SANDHU, "Forecasting wheat production using ARIMA model in Punjab", *https://www.researchgate.net/publication/304103256_Forecasting_wheat_production_using_ARIMA_model_in_Punjab*, 2014.

[4] S. Colianni, S. Rosales and M. Signorotti, "Algorithmic Trading of CryptocurrencyBased on Twitter Sentiment Analysis", Cs229.stanford.edu. [Online]. Available: http://cs229.stanford.edu/proj2015/029_report.pdf.

[5] "Cryptocurrency Market Capitalizations | CoinMarketCap", *CoinMarketCap*, 2019. [Online]. Available: https://coinmarketcap.com/.

Code Repository

[6] https://github.com/mitulverma14/CryptocurrencyStatisticalAnalysis