

# A Comparative Analysis based approach for Bitcoin Price Forecasting

Yash Wadalkar, Yellamraju V H Sai Tarun,  
Jaiesh Singhal  
UG Student  
Dept. of Electronics and Telecommunication  
Sardar Patel Institute of Technology  
Mumbai, India

Prof. Reena Sonkusare  
Head of Department  
Dept. of Electronics and Telecommunication  
Sardar Patel Institute of Technology  
Mumbai, India

**Abstract**— Bitcoin, one of the most famous and high-demand cryptocurrencies, is a type of digital asset that is extremely difficult to track and make predictions upon. In addition, Bitcoin price does not correlate with market-movements, therefore, predicting its price action and its locus is an ordeal. In this paper, we have followed a comparative analysis approach, wherein we are using four different models to predict the trend of BTC Time series data. The results justify that the models have achieved accurate forecasting trends. During the period of 16th to 31st December 2020, Bitcoin prices experienced considerably high swings, due to the increased demand for it. In quantitative terms, the prices experienced fluctuations to the tune of 8000 USD. Despite these enormous price changes, we were able to achieve a model, that helped us attain a Mean Absolute Error (MAE) of 153.55 USD and Mean Square Error (MSE) of 43231.80 USD. Conventional Bitcoin price predicting researches follow a single to two model approach. However, for a highly volatile asset like Bitcoin, making long-term predictions and generalizing them based on limited number of models results in low accuracy outputs. This gap has been bridged in our research, we have worked with different models, as well as fragmented the time intervals into smaller portions, post which the prediction was made for only 2 days. Using this approach, we attained results with least error rates. The results obtained clearly show that ARIMA is the best model for predicting the future trends for BTC time series data. It takes into account the different types of decompositions like Regular Trend, Sessional and Residual Trend making the model give the best results.

**Keywords**—Cryptocurrency; blockchain; bitcoin; Time-series Analysis; Facebook Prophet; LSTM; ARIMA; XGBoost

## I. INTRODUCTION

Technology has been disrupting the way humans live, work and even transact. Globally, economies and financial institutions have been going digital at an unprecedentedly fastpace[1]. The advent of Fin-Tech has left the conventional financial system archaic. One of the most recent developments in this space is the advent of Cryptocurrencies, more specifically, the most talked about, Bitcoin. Bitcoin, one of the original and first cryptocurrencies, currently has a market capitalization north of around 600 billion USD, which would only grow steeply in the forthcoming years.

This digital currency derives its popularity from two features: its decentralized nature [2] and the extraordinary volatility [3] that this asset exhibits. The returns that the

investors yield on this investment is significantly higher than that of traditional banking or market schemes. Moreover, since there is no government intervention, the price is solely controlled by the public, for the public.

Generalising, the prices of cryptocurrencies are not driven by the market therefore technical analysis becomes important in determining the price range of bitcoin over a period. Since technical analysis does not depend on external economic data, solely past price patterns are used to predict the price of Bitcoin [4].

In recent researches, LSTM and ARIMA models were used to forecast prices of Indian stocks for a period of 5 months. LSTM showed better results compared to ARIMA [5]. Another research where Gradient Boost Tree Model was used to capture Twitter data and public sentiments for a tenure of 3.5 weeks, analysed each tweet, for which over 50 percent accuracy was achieved [6]. A study carried out by Wint forecasted the closing price of the Myanmar Stock Price Index (MYANPIX) using ARIMA and Facebook Prophet. For all three periods (daily, weekly and monthly) prophet, having less error rate, has outperformed ARIMA [7].

Understanding the recent developments and with reference to researches published in this domain, we have carried out condensation of various approaches, in our research, which were followed to make predictions for equity markets and other cryptocurrencies. A comparative analysis of four models has been incorporated to furnish the best possible prediction model, wherein each model individually, contributes to the increased efficiency of the entire mechanism.

## II. METHODOLOGY

The below drawn flowgraph shows our approach towards building the models and comparing them based on various performance metrics.

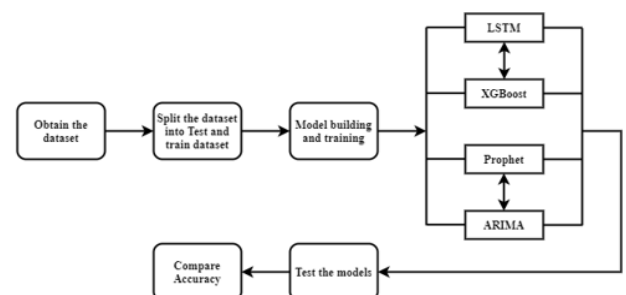


Fig 1. Block Diagram of the proposed approach

First, we choose the data set on which the different models are to be trained. The dataset used in our case is the Bitcoin Historical Dataset. It has the data between the period of January 1st 2012 to December 31st 2020, with the Open High Low Close (OHLC) prices of Bitcoin. Additionally, data of the volume of Bitcoin transacted, as well as the corresponding currency's volume was taken into consideration [8].

This is followed by filtering the time series data by data pre-processing, followed by dividing the dataset into Test and Train data. The four models: LSTM, XGBoost, Prophet and ARIMA are trained on the training data and then the comparison of the results is made on the test data.

The aforesaid models have been described in detail in following section, which justifies their suitability for our project.

### III. MACHINE LEARNING MODELS

#### A. Long Short Term Memory (LSTM):

LSTM is a type of Recurrent Neural Network (RNN) that is used in the field of deep learning. It's used in a variety of fields, including machine translation and speech recognition, and it's also good at classifying, processing, and forecasting time-series data [10]. For building the LSTM model, the first step contains the filtering of the information which is not required in the cell states and is thrown away. Then, information which is to be processed is stored in the cell states. This is followed by a Sigmoid layer which acts as a filter for what information is to be taken in input layer. And then, based on these filtered inputs going through the sigmoid layer and the series of tanh functions the final output is obtained.

#### B. XGBoost Classifier:

XGBoost is a gradient boosting algorithm that uses a decision tree based ensemble Machine Learning algorithm. It's made to be extremely powerful, adaptable, and portable. XGBoost is the best option for structured/tabular data ranging from small to medium sizes and also decision tree based problems because it offers parallel tree boosting. When adding new models, it uses the gradient descent algorithm to reduce the loss. Therefore, it is known as Gradient boosting. XGBoost basically aims to minimize the regularized objective function (L1 and L2). This is accomplished by integrating the uncertainty through the penalty term with the convex loss function dependent on the discrepancy between the expected and goal outputs. After each phase of the training process, new trees are introduced that project the residuals/errors of the prior trees. The final prediction is made by combining these.

#### C. Facebook Prophet:

Prophet is a model for forecasting time series data that was created by Facebook. Non-linear patterns align with annual, weekly, and regular seasonality in the Generalized Additive Model (GAM). It is extremely effective at predicting highly seasonal data for any kind of pattern, whether the data is highly seasonal or missing.

The Prophet employs a three part time series model with trend, seasonality, and holidays as its components. The equation representing the model is given by:

$$y(t) = g(t) + s(t) + h(t) + \epsilon t \quad (1)$$

In Equation (1),  $g(t)$  denotes logistic growth,  $s(t)$  denotes periodic shifts,  $h(t)$  denotes effect due to holiday, and  $\epsilon t$  takes into account any missed irregularities by the model [7][9].

#### D. Auto Regressive Integrated Moving Average (ARIMA):

This Model is one of the variants of the most used Autoregressive Moving Average (ARMA). This model is designed for time series data analysis, processing, and forecasting. Each ARIMA Model is denoted by the letters p, d, and q, where p denotes how many numbers of lagging observations does the model have, d denotes the number of times difference between the raw observations are observed, and finally q is the size of the moving average (MA) window. Whereas for Seasonal data, ARIMA is denoted with a similar expression of the non-seasonal data with 4 additional parameters P, D, Q and m: where m denotes the count of the periods each season has, P refers to the autoregressive part of the model, D represents the differencing term and Q represents the Moving average (MA) part of the ARIMA model [12][13].

### IV. RESULT ANALYSIS

Based on the different methods/models used, the models were trained on the Historical data of Bitcoin Prices from period from 16th December 2020 to 29th December 2020 and the results are compared on the prediction made by all the models from 29th December 2020 to 31st December 2020. This specific period experienced a steep increase of 8000 USD approx. in the Bitcoin Prices. So, the extreme volatility of Bitcoin Prices is also taken into consideration which helps in identification of the best model in the worst circumstances.

The two main performance metrics, namely Mean Average Error (MAE) and Mean Square Error (MSE), computed for each model [14]. The results are tabulated in Table 1.

A detailed explanation of these results has been provided below. On the basis of values obtained in Table 1, we have compared each of the model's performance with the other prediction models, alongside justifying the reason for choosing or rejecting the particular model for the fulfilment of our purpose.

#### A. Long Short-Term Memory (LSTM):

Figure 2 depicts the LSTM Model's forecast vs. Actual Prices from the time period of December 23rd to December 31st. The LSTM is trained via sigmoid activation function along with Adam optimizer for 100 epochs.

The Error between predictions and actual is compared via Mean Squared Error (MSE) and Mean Average Error (MAE). The MSE for LSTM is 114466.39 and MAE is 264.28 USD, which is not that high compared to other models but is not the best model for our use case.

**B. XGBoost:**

Figure 3 depicts the plot of XGBoost Forecast vs. Actual Prices from December 23rd to December 31st. The MSE for XGBoost is 190613.16 and MAE is

273.17, which is less accurate compared to LSTM Model. Thus, XGBoost is not the most accurate model for our use case.

Dec 23-31 2020 Forecast vs Actuals

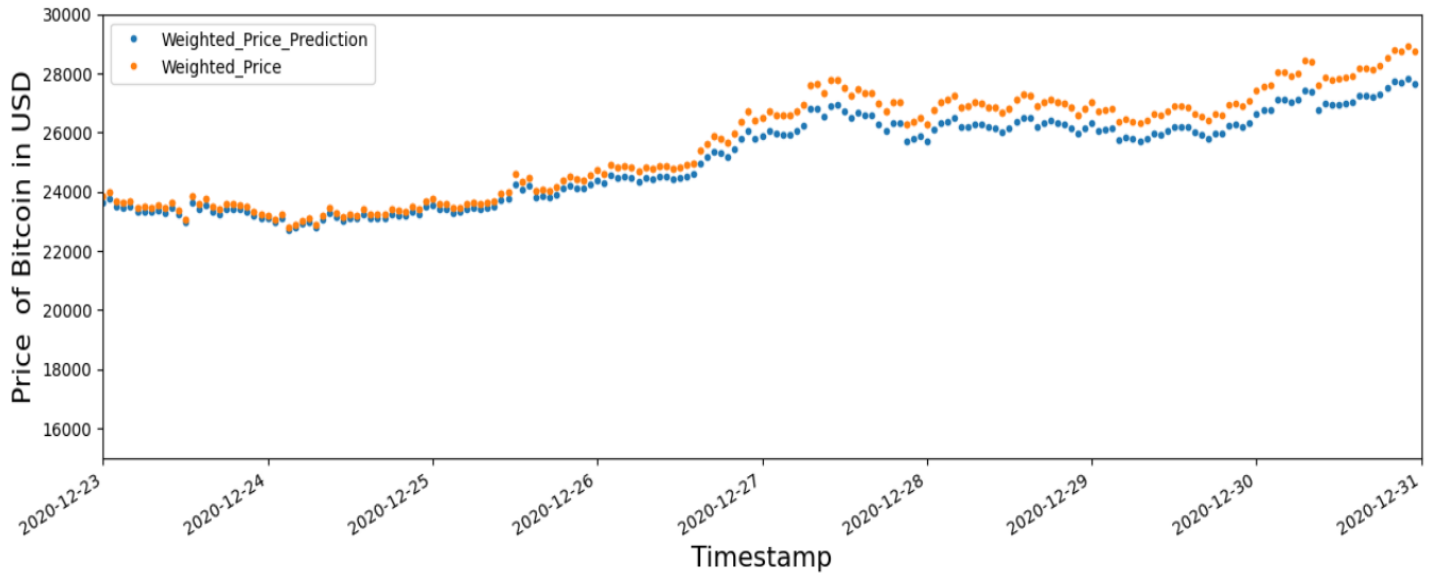


Fig 2. Forecast vs Actual for LSTM Model during Dec 23-31

Dec 23-31 2020 Forecast vs Actuals

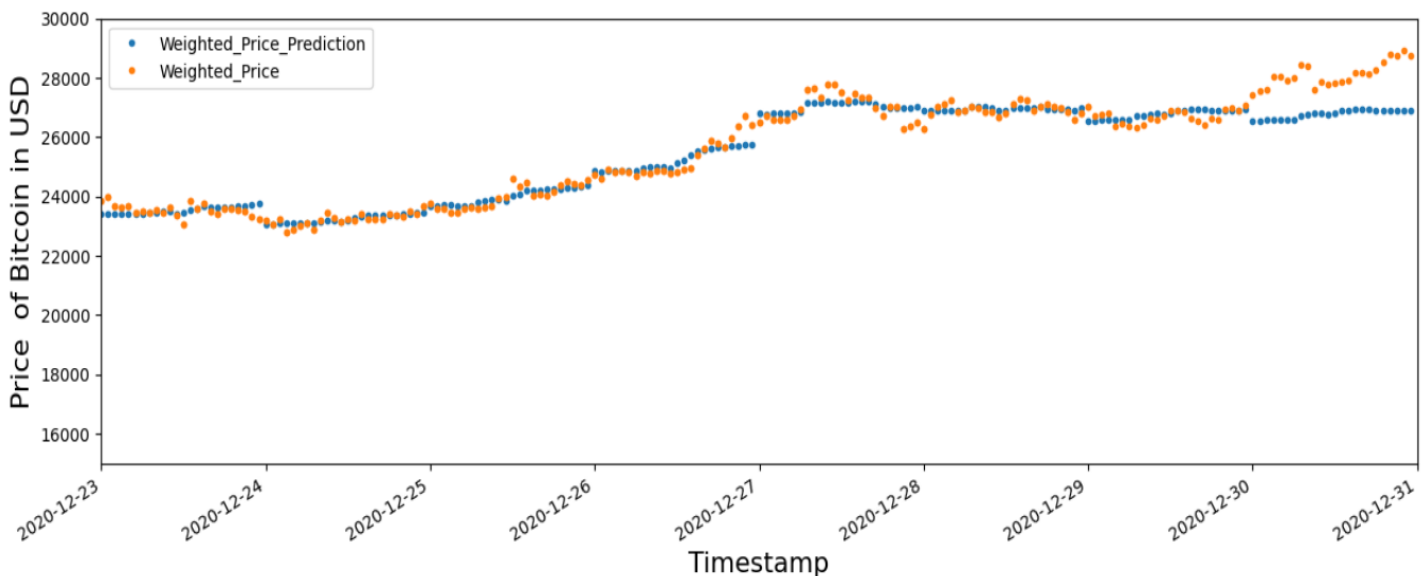


Fig 3. Forecast vs Actual for XGBoost Model during Dec 23-31

**C. FacebookProphet:**

Figure 4 depicts the plot of Facebook Prophet Forecast vs. Actual Prices from December 23rd to December 31st. The MSE for Facebook Prophet is 334612.50 and MAE is 419.39 USD, which is less accurate compared to both LSTM and XGBoost. This shows that Facebook Prophet is not as flexible for volatile data prediction. Therefore, Facebook Prophet is also not the best

model in our case.

**D. Auto Regressive Integrated Moving Average (ARIMA):**

The ARIMA Forecast vs Actual Prices for the period between December 23rd to December 31st is shown in Figure 5. The MSE for ARIMA is 43231.80 and MAE is 153.55 USD, which is the least compared to all of the other models. Therefore, ARIMA is the best model for this particular use case with the most accuracy.

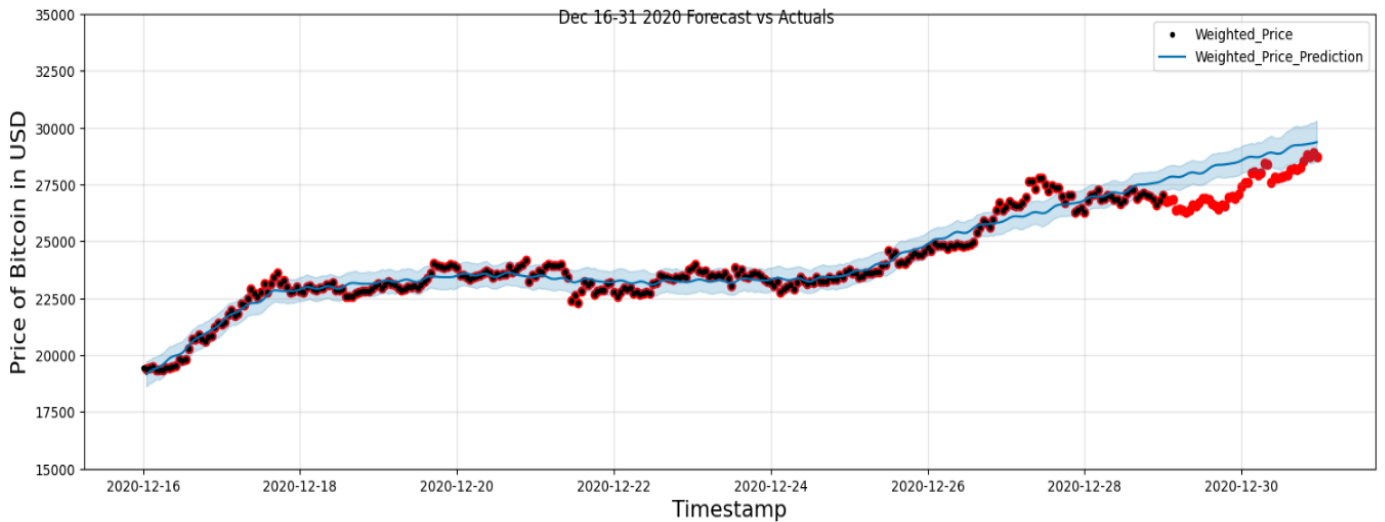


Fig 4. Forecast vs Actual for Facebook Prophet during Dec 16-31

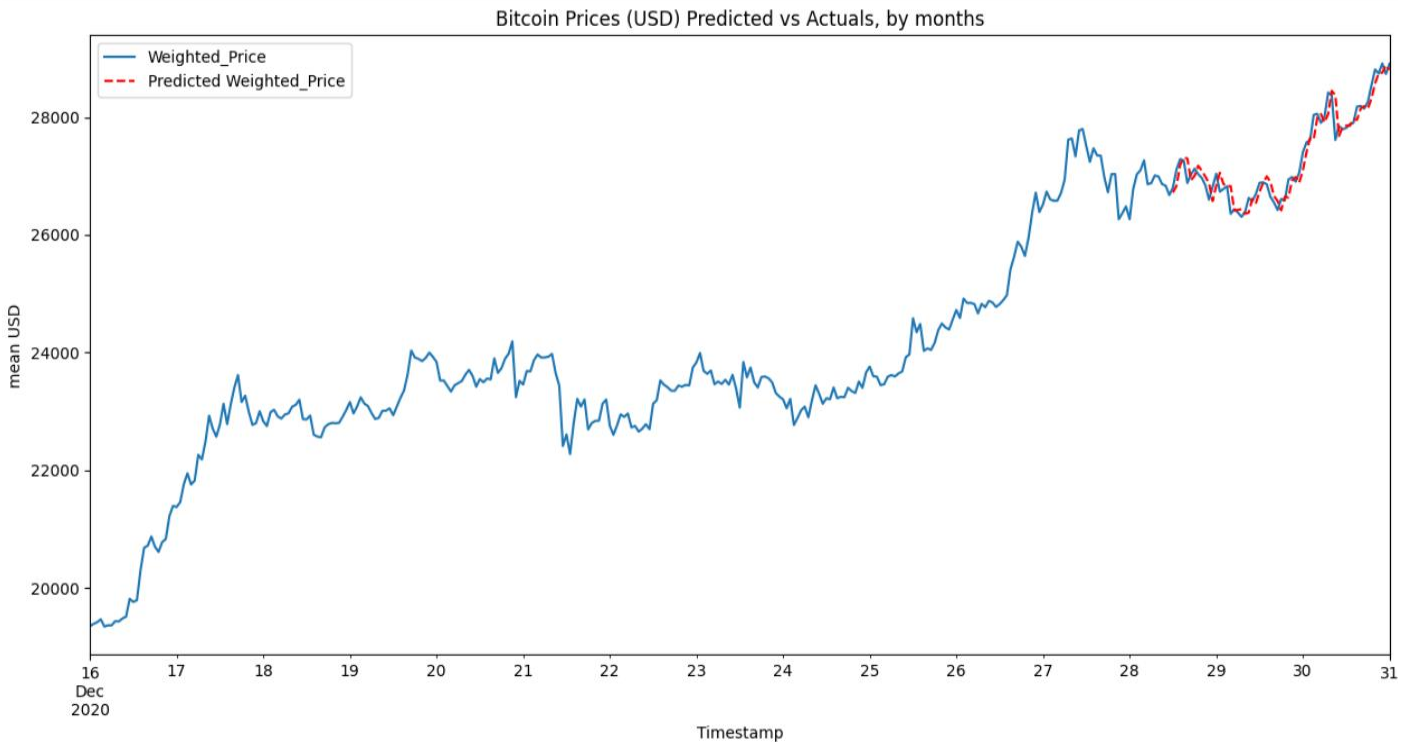


Fig 5. Forecast vs Actual for ARIMA Model during Dec 16-31

TABLE I. Comparison of MSEs and MAEs in USD of different Models

Model	MSE (in USD)	MAE (in USD)
LSTM	118736.68	281.84
XGBoost	190613.16	273.17
Prophet	334612.50	419.39
ARIMA	43231.80	153.55

### V. CONCLUSION

This paper presents an enhanced, inclusive solution for carrying out the time-series analysis for BTC dataset. The BTC dataset contains the OHLC (Open High Low Close) data between the period of Jan 1st 2012 and Dec 31st 2020. Due to the extreme volatility in the price value, we have chosen to predict the prices from the period of 29th Dec 2020 to 31st Dec 2020. Four different models, namely LSTM, ARIMA, XGBoost and Facebook Prophet were used for achieving the

aforementioned objective. The performance parameters taken into consideration for our purpose were Mean Absolute Error (MAE) and Mean Squared Error (MSE). The results thus obtained clearly indicate that ARIMA with MAE of 153.55 USD and MSE of 43231.80 USD emerges as the best model among the other three. [Table 1]. The better performance of ARIMA compared to others can be attributed to the fact that it takes into account the different types of decompositions like Regular Trend, Seasonal and Residual Trend. Thus, even when the deviation of prices is higher than normal the ARIMA Model is very accurate while making predictions for both short as well as long term prices.

In further advancements to this study, capturing governing factors would result in better accuracy of prediction. For example, obtaining data on the trending page of Twitter would provide us with insights about public

feelings regarding Bitcoin, which would directly impact the demand for it, in turn driving its price globally. Analyzing the posts on the Reddit platform would also help similarly, as it would help us to dive deeper and capture market sentiments of Bitcoin investors and the general public [6]. These insights can be reinforced into the training process of models, which could further increase efficiency and accuracy.

## REFERENCES

- [1] M. Mudassir, S. Bennbaia, D. Unal and M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach," *Neural Computing & Applications*, 2020
- [2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008
- [3] D. G. Baur and T. Dimpfl, "Realized bitcoin volatility," *SSRN Electronic Journal*, 2017
- [4] Y.S. Abu-Mostafa and A.F. Atiya, "Introduction to financial forecasting," *Applied Intelligence*, 1996, vol. 6, no. 3, pp. 205-213
- [5] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K.P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, India, 2017, pp. 1643- 1647
- [6] T.R. Li, A.S. Cham Rajnagar, X.R. Fong, N.R. Rizik and F. Fu, "Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model," *Frontiers in Physics*, 2019, vol. 7, no. 98, pp. 1-8
- [7] W.N. Chan, "Time Series Data Mining: Comparative Study of ARIMA and Prophet Methods for Forecasting Closing Prices of Myanmar Stock Exchange," *Journal of Computer Applications and Research*, 2020, vol. 1, no. 1, pp. 75-80
- [8] Bitcoin Historical Dataset, Bitstamp Exchange, Feb.2021. [Online]. Available: <https://bitcoincharts.com/charts/bitstampUSD>
- [9] S.J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, 2018, vol. 72, no. 1, pp. 37-45
- [10] W. Fang, P. Lan, W. Lin, H. Chang, H. Chang and Y. Wang, "Combine Facebook Prophet and LSTM with BPNN Forecasting financial markets: the Morgan Taiwan Index," *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Taipei, Taiwan, 2019, pp. 1-2
- [11] W. Zhang, P. Wang, X. Li and D. Shen, "Quantifying the cross-correlations between online searches and Bitcoin market," *Physica A: Statistical Mechanics and its Applications*, 2018, vol. 509, pp. 657-672
- [12] G.P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, 2003, vol. 50, pp. 159-175
- [13] L. Wang, H. Zou, J. Su, L. Li and S. Chaudhry, "An ARIMA-ANN Hybrid Model for Time Series Forecasting," *Systems Research and Behavioral Science*, 2013, vol. 30, pp. 244-259
- [14] C.J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model," *Climate Research*, 2005, vol. 30, no. 1, pp. 79-82
- [15] J. Rebane, I. Karlsson, P. Papapetrou, and S. Denic, "Seq2Seq RNNs and ARIMA models for Cryptocurrency Prediction: A Comparative Study," in *Proceedings of SIGKDD Workshop on Fintech (SIGKDD Fintech'18)*, 2018.
- [16] M. Nakano, A. Takahashi and S. Takahashi, "Bitcoin technical trading with artificial neural network," *Physica A: Statistical Mechanics and its Applications*, 2018, vol. 510, pp. 587-609
- [17] T. Shintate and L. Pichl, "Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning," *Journal of Risk and Financial Management*, Jan. 2019, vol.12, no.1, p.17