Forecasting Cryptocurrency Prices using Machine Learning

Data Analytics

Ashwini Chaudhari
Student ID: x18129676

School of Computing
National College of Ireland

Supervisor: Dr. Muhammad Iqbal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ashwini Chaudhari
Student ID:	x18129676
Programme:	MSc Data Analytics
Year:	2019-2020
Module:	Research Project
Supervisor:	Dr. Muhammad Iqbal
Submission Due Date:	12/12/19
Project Title:	Forecasting Cryptocurrency Prices using Machine Learning
Word Count:	4897
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	30th January 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Forecasting Cryptocurrency Prices using Machine Learning

Ashwini Chaudhari
x18129676

Abstract

Blockchain and cryptocurrencies have risen to popularity in the recent years to a great extent due to its increasing trading volumes and huge capitalization in the market. These cryptocurrencies are being used not only for trading but are being accepted for monetary transactions as well these days. As the prices fluctuate and return on investment increases investors, traders and general public are showing increased interest towards bitcoin and altcoins. This research focuses on implementing forecasting models that will return accurate price predictions for cryptocurrencies. Prices for Bitcoin, Ethereum and Litecoin are predicted using the traditional forecasting model for timeseries ARIMA, the Prophet Model and deep learning algorithm LSTM. The results of the three models were evaluated and the LSTM Model was found to outperform the Prophet as well as the ARIMA model.

1 Introduction

Bitcoin was launched in the year 2009 and since then the admiration and acceptance of both Blockchain as well as cryptocurrencies has been going in an upward direction only. There has been a tremendous growth in the applications of Blockchain in varied fields ranging from financial markets and healthcare to internet of things and other technologies. The unique properties of Blockchain such as security and transparency, lower costs and decentralization has led to its usage for solving various real-world issues Mallqui and Fernandes (2019).

However, initially this technology was developed to be used for virtual currencies and since the launch of Bitcoin there has been no going back as the number of cryptocurrencies in the market are ever increasing. There is no slowing down in this domain due to huge investments in this technology globally Wu et al. (2018). There has been a huge disruption in the financial industry due to this "distributed ledger technology" Brühl (2017). Along with investors and traders, common people have also started showing interest in cryptocurrencies owing to the huge return on its investments. The usage of cryptocurrencies for regular transactions in finance has also begun and hence this domain is only going to expand further leading to increased curiosity in understanding their behaviour Mallqui and Fernandes (2019). Various researches and studies have been carried out for price prediction of these virtual currencies and this research follows a similar approach but uses ARIMA, LSTM and Prophet for forecasting prices of Bitcoin, Ethereum and Litecoin using up-to date historical prices.

The main focus of this research is analysis of Bitcoin and two more altcoins using above mentioned three models and determining which model is the most efficient. Section 2 summarizes the popular prediction techniques for cryptocurrency prices that have been used so far. Section 3 discusses the design specifications and methodology adopted for this research and Section 4 elaborates on the techniques and models that are employed for prediction. Section 5 discusses the results for the experiments carried out using the models followed by Conclusion and Future work in Section 6.

Research Question

"How efficiently can machine learning algorithms help predict cryptocurrency prices?"

Keywords: ***machine learning, cryptocurrency, ARIMA, LSTM, Prophet***

2 Literature Review

The prevalence and impact of financial markets on multiple domains such as businesses, education, jobs, technology is ever increasing thus affecting various other sectors as well as the economy. This is one of the primary reasons for investors, researchers and traders to have developed a keen interest in the price movements of stocks and stock-market analysis. This has led to the development of various behaviour and price prediction models using different techniques which can be broadly categorized as Statistical, Pattern Recognition, Machine Learning and Sentiment Analysis . Dev et al. (2019).

Cryptocurrencies have been a revolutionary trend in the financial markets over the past few years and hence been under limelight due to its profitability. As its nature is very similar to stocks, popular cryptocurrency price prediction techniques can also be categorized as mentioned above. Figure 1 represents the taxonomy of the three most popular prediction techniques that have yielded favorable results in the recent past.

Saxena and Sukumar (2018) and Karakoyun and Çıbıkdiken (2018) both in their researches have used statistical model ARIMA and LSTM deep learning algorithm and the results obtained in the first research showed an RMSE of 700.69 for ARIMA and 456.78

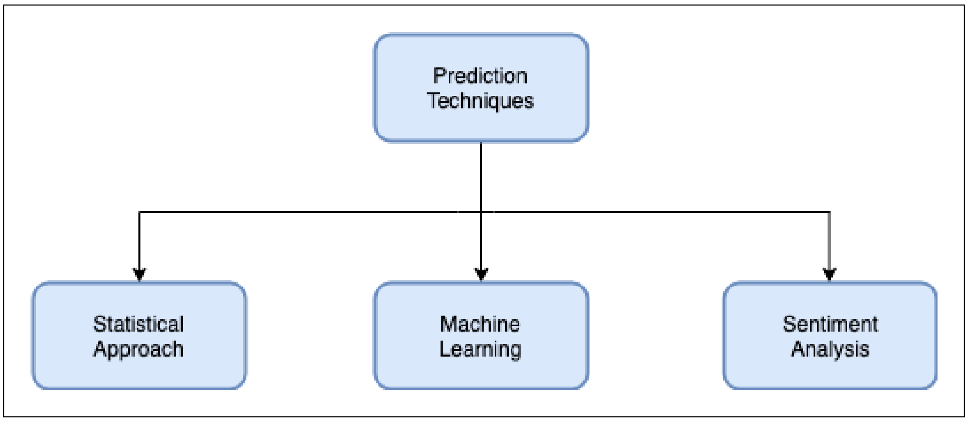

Figure 1: Popular techniques for predicting cryptocurrency prices

for LSTM whereas RMEE of 1146.07 for ARIMA and 93.27 for LSTM was observed in the research by by Karakoyun and Çıbıkdiken (2018). However, both the studies did conclude that LSTM helped to achieve more accuracy than ARIMA with an error of just 1.4%.

GARCH models were adopted by Yu et al. (2019) and Peng et al. (2018), where, in the first research price volatility of Bitcoin was analyzed for identifying the effect of volume on volatility from the perspective of asymmetry along with analyzing growth of user interest using Google Trends and Wikipedia page views while a combination of GARCH and ML approach SVR was used in the second research. The findings of the latter suggested that SVR-GARCH model outperformed all benchmarks of the GARCH models as RMSE and MAE values were lower for the model developed using machine learning approach .

Kristjanpoller and Minutolo (2018) in their study proposed the ANN-GARCH model accompanied by preprocessing for forecasting volatility in Bitcoin prices and compared it with volatility of EUR-USD. After obtaining MSE and MAPE for various horizons MCS test was applied where it was observed that for 5% significance there was no superiority of hybrid models as compared to the GARCH models.

LSTM is used as the base model for all three researches carried out by Lahmiri and Bekiros (2019) , Wu et al. (2018) and Alessandretti et al. (2018) but the first study uses only LSTM and showed that LSTM outperformed GRNN where as the other two studies combined LSTM with other models as well. Wu et al. (2018) compared the results of traditional LSTM to LSTM with AR(2) and highlighted that prices that were forecasted by both models are close to real values but MSE, MAE, MAPE and RMSE values are low for the LSTM-AR(2) model. Alessandretti et al. (2018) conducted an analysis on 1681 cryptocurrencies and results obtained suggested that XGBoost trees worked well for shorter duration's of 5 to 10 days but LSTM-RNN model predicted more accurate prices for longer duration's. Pant et al. (2018) also used RNN to predict Bitcoin price for the next day using historical prices. This research combined a sentiment analyser that can differentiate between positive and negative tweets with the RNN model and achieved an accuracy of 77.62%.

Mallqui and Fernandes (2019) employed ANN, SVM and Ensemble algorithms for prediction of direction of prices of Bitcoin. The results of regression were used as inputs for increasing prediction of price direction. The results presented an increased accuracy of 10% on the basis of chosen attributes. MAPE values of 1% - 2% were obtained with respect to the closing prices, maximum and minimum prices of Bitcoin thus proving the model more efficient as compared to others and suggested using technical and economic indicators to further improve the predictions.

Four conventional ML methods were applied by Singh and Agarwal (2018) for prediction of prices of Bitcoin where all features were implemented individually for Linear and Polynomial Regression and KNN, SVR and Polynomial Regression were used to perform hyper-tuning using grid-search logic. KNN Regression method outperformed all other models obtaining MSE of 0.00021. Jim et al. (2019) also used KNN, SVM and LASSO along with ElasticNet, SGD, Bayesian Regression, Decision Tree, AdaBoost, GTB, MLP to predict cryptocurrencies 30-day returns. Almost all algorithms were found to generate an accuracy between 50% - 60% but LASSO dominated as accuracy obtained for this model was 61%. The authors concluded with research by stating that huge cryptocurrencies that exhibit low volatility could be more accurately predicted as compared to the new cryptocurrencies in the market that are highly volatile.

Various researchers such as Rahman et al. (2018), Bibi et al. (2019), Wimalagunaratne and Poravi (2018), Steinert and Herff (2018) have used sentiment analysis for prediction of cryptocurrency prices by collecting tweets over time. Rahman et al. (2018) made use of various regression and classification algorithms out of which Naive Bayes showcased maximum accuracy of 89.65% and proposed that sentiment analysis could help predict change in price of cryptocurrencies where as Bibi et al. (2019) used Topic modeling in combination with opinion mining. The authors investigated users concerns accompanied by their sentiments at the top locations that were identified as places where cryptocurrency usage was the more prominent.

Wimalagunaratne and Poravi (2018) assessed the public perception, dependency amongst Bitcoin and altcoins and historical and trading data and applied Machine learning and sentiment analysis to increase the efficiency of price prediction. Steinert and Herff (2018) on the other hand focused on predicting returns of altcoins using sentiment analysis by applying linear regression on the basis of data collected for 45 days. Results obtained showed that prediction of short term returns from sentiments and Twitter activity is possible. The authors also and suggested using advanced approaches of prediction like Neural Networks may further improve the prediction accuracy. Yenidogan et al. (2018) have used ARIMA and Prophet Model for cryptocurrency price prediction using bitcoin data form 2016 to 2018 and the results show that with an R^2 of 0.94 for Prophet performed way better than ARIMA which could obtain an R^2 of only 0.68.

After studying and reviewing various papers it was proposed that this research will be using ARIMA Forecasting Model, deep-learning algorithm LSTM and the Prophet model developed by the Data Scientists at Facebook for forecasting Taylor and Letham (2018).

3 Methodology

The methodology used for implementing this research is CRISP - DM which is an industry-proven method for data mining projects and research. The Cross Industry Standard Process specifically for Data Mining (CRISP- DM) is very flexible as it provides an option to modify the methodology to suit specific project requirements and needs while offering a road-map for the entire project development stage. Figure 2 shows all phases of this methodology and the same are described in an elaborate manner below.

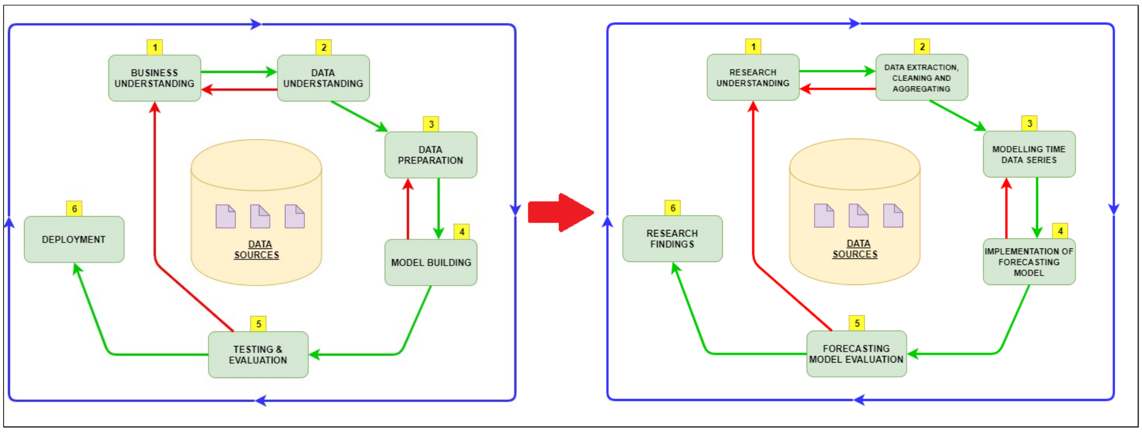

Figure 2: Modified CRISP-DM Process

3.1 Business Understanding

There have been a lot of breakthroughs and advancements in the technology sector in the past decade. But one particular technology that emerged out of nowhere and is now being employed by most sectors like finance, healthcare, IOT, etc. is Blockchain. This technology was first garnered attention when Bitcoin was launched and has been emerging ever since.

The past few years have been a massive boon for Bitcoin investors due to its great bull run and huge returns on the cryptocurrency investment. This has caused a lot of interest in cryptocurrency investments and hence the need for forecasting.

Along with constantly increasing market capitalization the interest is also increasing due to the various altcoins that are now available in the market like Ethereum, Litecoin, EOS, Dash, Ripple, etc. The total number of altcoins in the market currently is even more than 1500 different types of altcoins. Jim et al. (2019) We have over 170 hedgefunds catering specifically to cryptocurrencies and transactions in this domain daily exceed over 15 billion dollars. Alessandretti et al. (2018) The highly volatile market, huge market capitalization and the fluctuating asset value provides great scope for research and development in this domain and hence in this research we will be focusing the accuracy for predicting prices of cryptocurrencies.

3.2 Data Understanding

Data for the experiments is collected from the websites *Quandl* (2019) and *Poloniex* (2019) for top 3 cryptocurrencies which lets users download some cryptocurrency datasets for free. Data for top 3 cryptocurrencies namely, Bitcoin, Ethereum and Litecoin is used for analysis. The data extracted has various attributes similar to the attributes of stock prices and includes the opening and closing prices for the currencies, high as well as low price, the volume and the weighted price of the currency. Generally, the weighted price is considered for financial assets as it is the ratio of value to volume traded over a time horizon and so we will be using this price for forecasting.

3.3 Data Preparation

The required bitcoin pricing data was extracted using the free API for Bitcoin from the Quandl Website. The data for historical prices of Bitcoin from the Kraken exchange was collected. It is observed that supply and demand determines the Bitcoin prices and hence the prices vary for different exchanges. So the data that we extracted from Kraken alone cannot be relied for actual prices.

For overcoming this issue along with the observed down-spikes in Kraken pricing data from 3 more exchanges was also taken into consideration. The data for Bitcoin prices from Coinbase, Bitstamp and Itbit was extracted and was combined with the Kraken data to calculate the average prices of Bitcoin and all zero values were removed from the dataframe.

Following the Bitcoin data, data for altcoins was extracted using API from the Poloniex Website. The buying of altcoins works in a manner that most of the altcoins cannot be purchased using regular currencies. How the purchase works is that usually one buys the Bitcoins and these are then exchanged for altcoins. So we extracted bitcoin exchange rate for altcoins and made use of our previously collected Bitcoin data to get the value of this in USD by conversion. Data for each of the cryptocurrencies was then examined for any null values. Trend and seasonality of data was removed to make the data Stationary as required for time series data.

3.4 Modeling

The three models used for time-series forecasting of the price of cryptocurrencies in this research are:

- ARIMA:

  ARIMA is an acronym for "AutoRegressive Integrated Moving Average". This model for forecasting is used for examining the historical pricing and then forecasting the future vales based on this data. It is a traditional model for the forecasting of time-series in which the AutoRegressive (AR) and the Moving Average (MA) models when combined work to form the model ARIMA Karakoyun and Çıbıkdiken (2018).

Finite differencing is used to make the data stationary for a non-stationary time series in this model. This model usually comprises model identification, parameter estimation and diagnosis.

- LSTM: Long Short-Term Memory networks, generally referred to as LSTM's are RNN of a special that have the ability to learn about dependencies that are long-term. This means that they are very well capable of remembering information for an extended duration as compared to RNN. This deep learning model allows training of large architectures and are better in resistance to errors Saxena and Sukumar (2018).

  The LSTM model developed by Hochreiter and Schmidhuber (1997) has a network that comprises of multiple layers which include an input and an output layer along with either one or even multiple hidden layers. It keeps on learning over multiple steps of time and consists of remember and forget gates that help determine thee information to be passed based on its importance and strength Karakoyun and Çıbıkdiken (2018).

- Prophet:

  The Prophet Model developed by Facebook's Data Science Team for forecasting is based on a model that is additive in nature where the daily, weekly as well as yearly non-linear trends are fit. The model employs a modular regression approach that permit selection of components pertaining to a forecasting problem and make adjustments accordingly while also working well with parameters that are default.

  It also comprises of a forecast tracking and measuring system including forecast flagging that allows analysts to make adjustments and improve the forecasts by making incremental enhancements. Basically, Prophet is an adaptable and adjustable model that helps analyze different time series and provide a scalable performance Taylor and Letham (2018).

3.5 Evaluation

After implementing the three models their efficiency and performance are evaluated. The evaluation metrics used are:MSE, $RMSE$, MAE and $MAPE$ and R^2. Smaller error values indicate better performance for the model whereas a high R^2 value, specifically between 0.7 and 1 means that there is a high correlation between the observed and predicted values. The results obtained for all three models and elaborately discussed in Section 5.

4 Implementation

The implementation process for the entire research is illustrated in Figure 3. Data is extracted using *Quandl* (2019) and *Poloniex* (2019) API's and data processing is done including data cleaning, aggregation and calculation of altcoin prices. Bitcoin and altcoin prices are then merged into a single dataframe and data is made suitable for time-series.

Forecasting algorithms are then applied and the models are implemented. The output is then evaluated on the basis of evaluation metrics for all three models followed by comparison of results for determining the best forecasting algorithm.

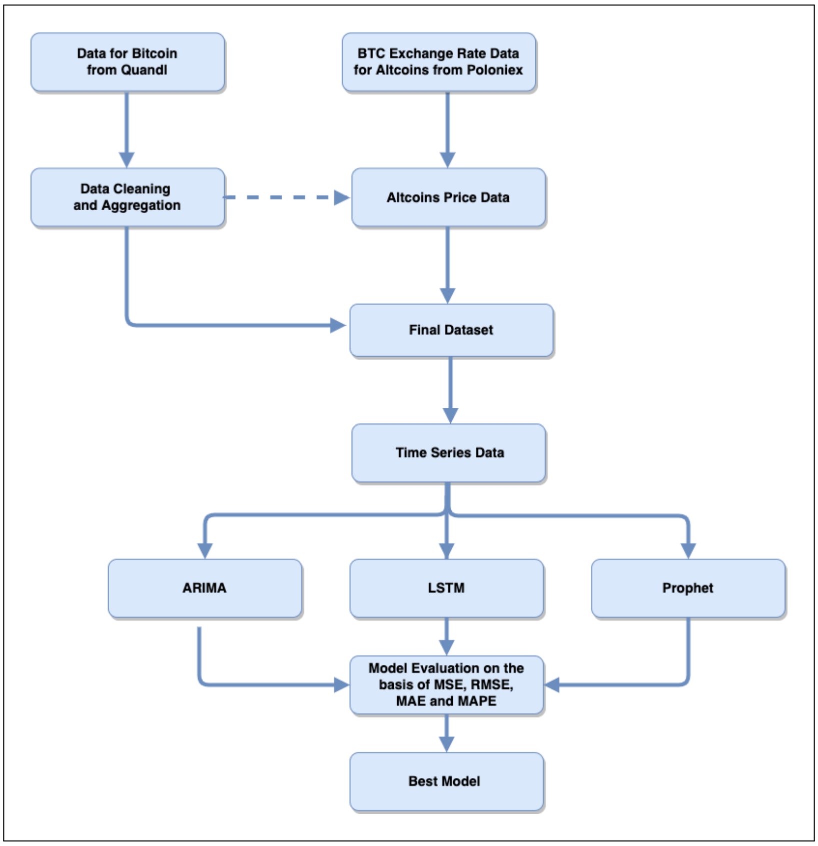

Figure 3: Process Flow of the Implementation method

Various datasets for bitcoin and other cryptocurrencies are available from multiple sources. So for this research data for Bitcoin was extracted from the Quandl Website using the Quandl API. The pickle library is used for saving as well as serializing the data which provides us with a dataframe consisting of all the extracted data. Bitcoin exchange data from Kraken was retrieved and visualized to check for any inconsistencies that might be present. It was observed that a few zero values were present.

It was noted that correct Bitcoin prices were not very easy to find as the way for determining prices is by taking into consideration the demand and supply. Bitstamp, Coinbase and Itbit are three major exchanges for Bitcoin. Price data from these three was extracted and then all four pricing values were combined into a single dataframe on the basis of their "Weighted Price". Again the data was visualized, a few inconsistencies were observed but mostly the prices were close to each other. The data was cleaned to get rid of any existing zero values and an average of all four prices was calculated to get the final price which will be used for the analysis. Figure 4 shows the process of data cleaning and aggregation of the bitcoin price data in an elaborate manner where the graph on the top left represents the prices extracted for Kraken and the graph at top right showcases the prices from all the four exchanges. The graph on the bottom left is the one obtained after cleaning and aggregation of data where as the graph on the bottom left is the one showing average bitcoin prices over the years.

To get data for our other cryptocurrencies the Poloniex API was used and Bitcoin exchange rate for each altcoin was extracted. Then using this data and our previous Bitcoin price data the altcoin prices were obtained in USD by multiplying weighted price of altcoins and average Bitcoin price. Lastly, a dataframe consisting of all prices was created for the analysis.

Figure 4: Cleaning and Aggregation of the Bitcoin Pricing Data

Once the data for bitcoin was cleaned and aggregated, the price data for Ethereum and Litecoin was collected and examined for any null values. This data was in the form of the exchange rate for Bitcoin and so the exchange rates were then multiplied by with our Bitcoin average prices to get the prices in USD for our altcoins Ethereum and Litecoin.

After obtaining data for the required cryptocurrencies, exploratory data analysis was performed to get a better understanding for the analysis. ARIMA and SARIMAX libraries were imported from the statsmodels module in python. The price data was analysed to check for any trends and seasonalities that might be present. It was observed that the data for Bitcoin exhibited changes over a period of time in variance and mean. A good timeseries dataset needs to be stationary in nature and hence presence of broad trends need to be taken care of. Seasonal Decomposition of the data was carried out and Dickey-Fuller test results confirmed that the data does not show presence of stationarity as it can be seen in Figure 5.

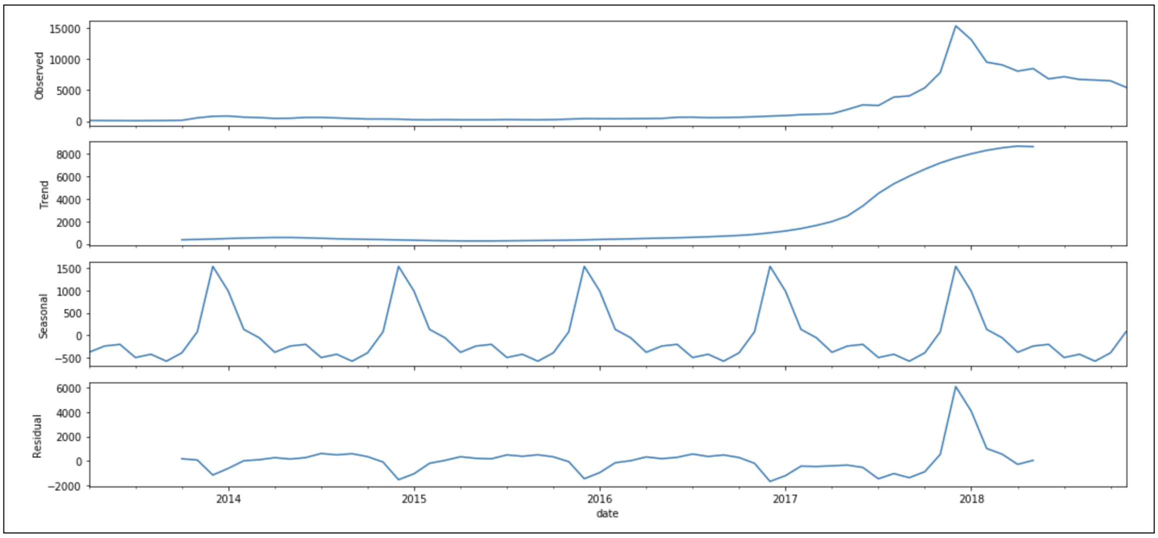

Figure 5: Seasonal Decomposition of Bitcoin Prices

Box-Cox transformations were then carried out for suppressing the variance followed by seasonal differentiation and regular differentiation. The strength of a particular observed datapoint with a previously observed datapoint is shown by the autocorrelation and so autocorrelation and partial autocorrelation factors are created for checking patterns in the data of the said timeseries. Positive correlation can be observed in the first three-four lags as shown in Figure 6 but most values are found to be in the blue shaded region which is the insignificant zone so in the residuals there is a possibility for presence of components that are seasonal. ARIMA model is applied by combinations of multiple parameters quality assessment of ARIMA model is fitted with SARIMAX() for each combination.

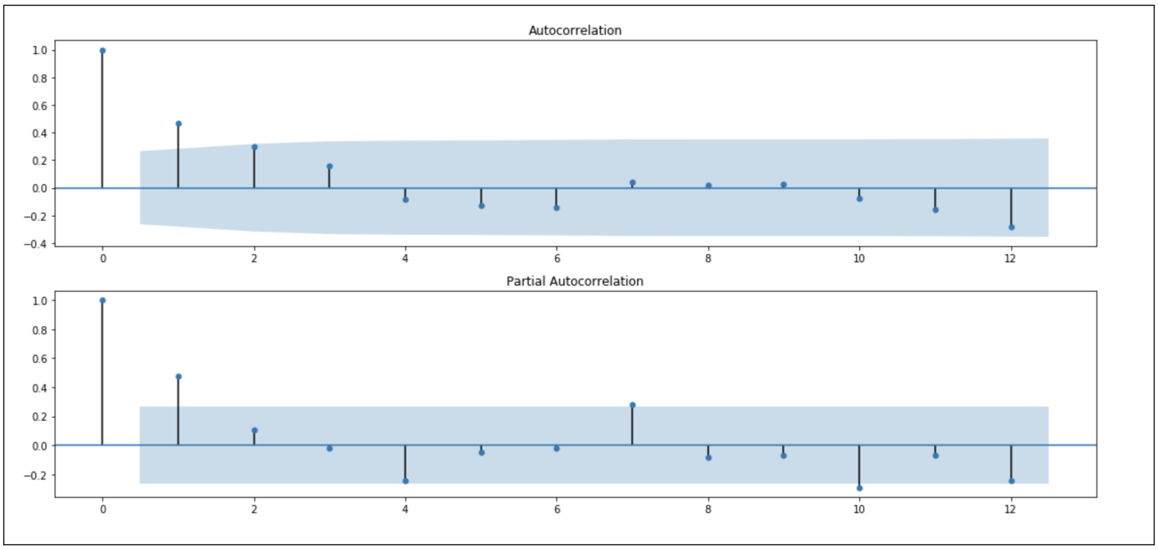

Figure 6: Cryptocurrency Prices in USD

Parameters are initially approximated followed by selection of model and best model summary diagnostics are examined. Once it was confirmed that normal distribution of residuals was present and there was absence of seasonality the prediction and forecasting was done using invboxcox() and price for Bitcoin was forecasted. The forecasted values were evaluated and are discussed in the Evaluation Section. The same procedure was followed to check the stationarity for Ethereum and Litecoin price data and results were evaluated.

For the LSTM Model, keras was used to import the LSTM library and other required dependencies. After loading the data, date was converted to datetime using pandas and model was bulid using a split_sequence function. Number of timesteps were assigned and data was split into training and testing sets. A bidirectional model was build while making use of the adam optimizer for LSTM and data was then trained. The bidirectional model was then used for the prediction and the results were evaluated for all the three cryptocurrencies separately.

The Prophet model was implemented using the fbprophet library. Data was loaded and the date and currency price columns were renamed to 'ds' and 'y' as it is a must for the column names of the dataframe to be in that manner for using fbprophet. The prices were converted to float and prophet model was applied to the dataframe for forecasting by considering the changepoint_prior_scale and daily_seasonality parameters. This is followed by creation of dates in future for which forecasting is to be done and then prices are predicted. This was done separately for all three cryptocurrencies and the forecasting results were then visualized and evaluated.

5 Evaluations

All the three models ARIMA, LSTM and Prophet were implemented successfully and thereafter were evaluated on the basis of their $Mean\ Square\ Error$ (MSE), $Root\ Mean\ Square\ Error$ ($RMSE$), $Mean\ Absolute\ Percentage$ (MAE), $Mean\ Absolute\ Percentage\ Error$ ($MAPE$) and R^2 values.

5.1 ARIMA Evaluations

There was a huge amount of processing carried out on the data before using it as a timeseries for the ARIMA Model as it was essential to remove trends and seasonalities present in the extracted data. After carrying out various statistical tests the stationary time series was used for the SARIMAX model and the results obtained were quite satisfactory in terms of the MAPE values. The R^2 obtained for Bitcoin and Ethereum indicate that the model was a good fit for the timeseries data of these two currencies, however, the model did not perform as well on the timeseries data for Litecoin.

Table 1 presents the results obtained using ARIMA for the three cryptocurrencies.

Crypto -currency	MSE	RMSE	MAE	MAPE	R^2
Bitcoin	2519603.08	1587.32	920.45	0.20	0.82
Ethereum	15552.79	124.71	66.81	0.31	0.73
Litecoin	1085.50	32.94	15.72	0.21	0.66

Table 1: ARIMA Results

Figure 7 shows the prices predicted by the ARIMA Model and we can clearly notice that there is a lag between the observed and forecasted values of all three cryptocurrencies. When the three plots are looked at individually, it is evident the the model is not as good a fit for prediction of Litecoin data as it is for the other two currencies and this is justified by the obtained R^2 values as well.

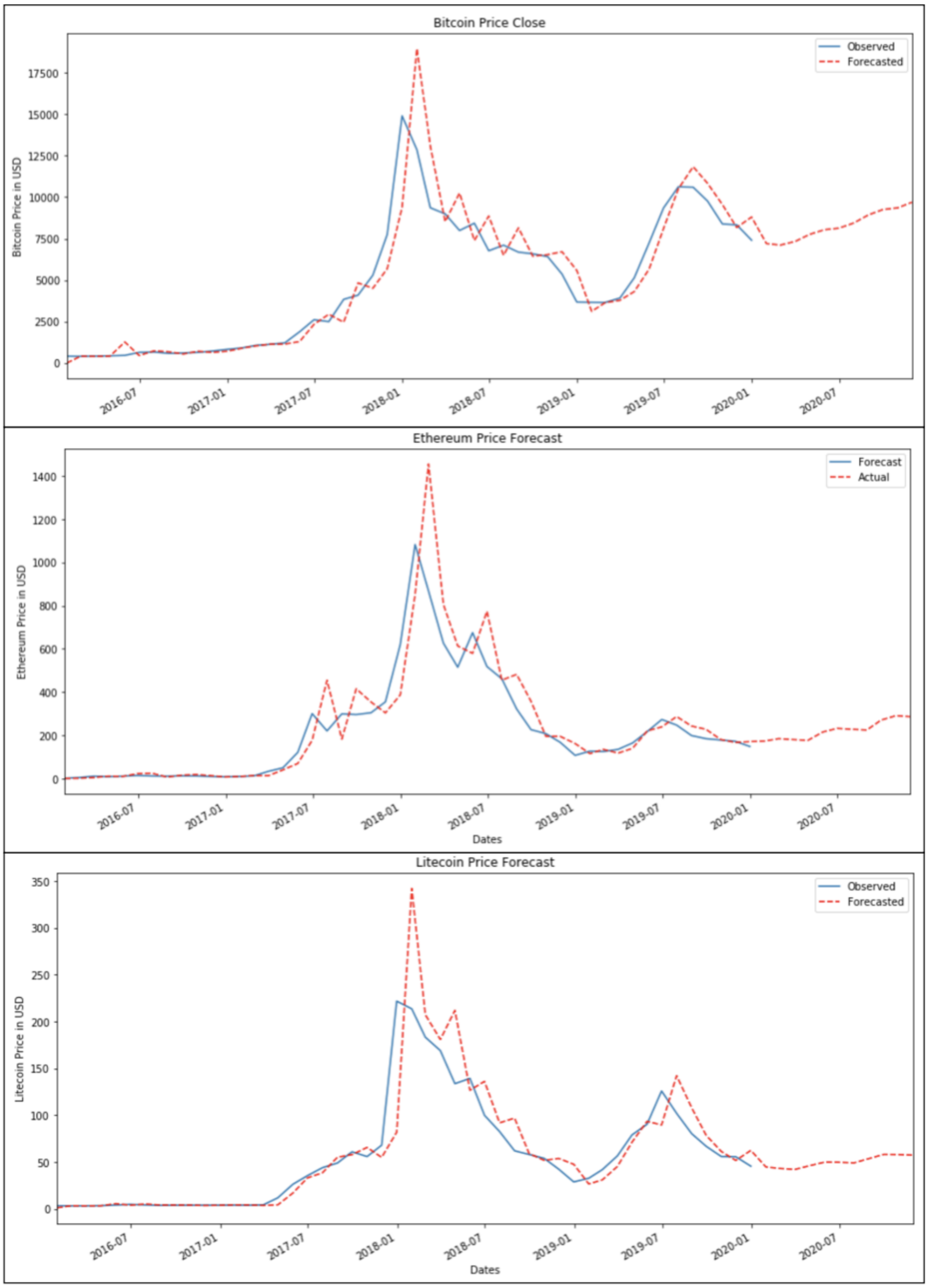

Figure 7: Plots using ARIMA for Forecasting

5.2 LSTM Evaluations

Table 2 presents the results obtained using LSTM for the three cryptocurrencies. The results obtained by the LSTM Model look very promising as the R^2 as well as MAPE values are very close to ideal. R^2 values as high as 0.98 and 0.99 convey that the model might be one of the best for forecasting prices. All other evaluation metrics for Ethereum

Crypto -currency	MSE	RMSE	MAE	MAPE	R^2
Bitcoin	67493.99	259.79	161.01	0.020	0.99
Ethereum	55.07	7.42	5.26	0.028	0.98
Litecoin	11.18	13.34	2.19	0.032	0.99

Table 2: LSTM

and Litecoin also look well and suggest that the LSTM model is overall a good fit. The MSE, RMSE and MAE values for Bitcoin look considerably high as compared to the other two cryptocurrencies but this could also be a result of the Bitcoin price "bullrun" in the recent past which might have resulted into these values.

Figures 8 and 9 show the prices predicted by the LSTM Model and we can still observe some lag in the actual prices and predicted prices of the three currencies. As it clearly noticeable, this model does showcase better performance as compared to ARIMA.

Figure 8: Bitcoin and Ethereum Plots using LSTM for Forecasting

Figure 9: Litecoin Plot using LSTM for Forecasting

5.3 Prophet Evaluations

The Prophet model required the least amount of processing as the only requirement was that the names of the columns should be specifically 'ds' and 'y'. The results obtained for the three cryptocurrencies are presented in Table 3. The R^2 values obtained for the Prophet model also look good as values for all three cryptocurrencies is above 0.90 but the MAPE values obtained are extremely high. This could be the case here because for low forecasts it is not possible for percentage error to exceed 1 i.e. 100 % but for really high forecasts no upper limit exists. This is one of the reasons that MAPE values are not used individually but always along with other error metrics. We can observe an opposite trend in the MAPE values as compared to the other error metrics in this case.

Crypto -currency	MSE	RMSE	MAE	MAPE	R^2
Bitcoin	589867.35	768.02	522.60	0.59	0.96
Ethereum	2329.30	48.26	31.91	1.07	0.96
Litecoin	260.93	16.15	10.20	1.48	0.93

Table 3: Prophet Results

Figure 10 shows the forecast plots by the Prophet Model. In the prophet plots, the black dots are the observed or actual prices and the blue line shows the forecasted price values. The plot looks interesting as it also shows the lower and upper bound for the forecasted prices. We can observe that the forecasted values are in sync with the actual price for most of the plot but does go out of range at various datapoints. Especially in the initial period, when the prices were considerably stable the forecast shows fluctuations and also some more inconsistencies such as just after a downfall predictions of small hikes in price which never actually happened.

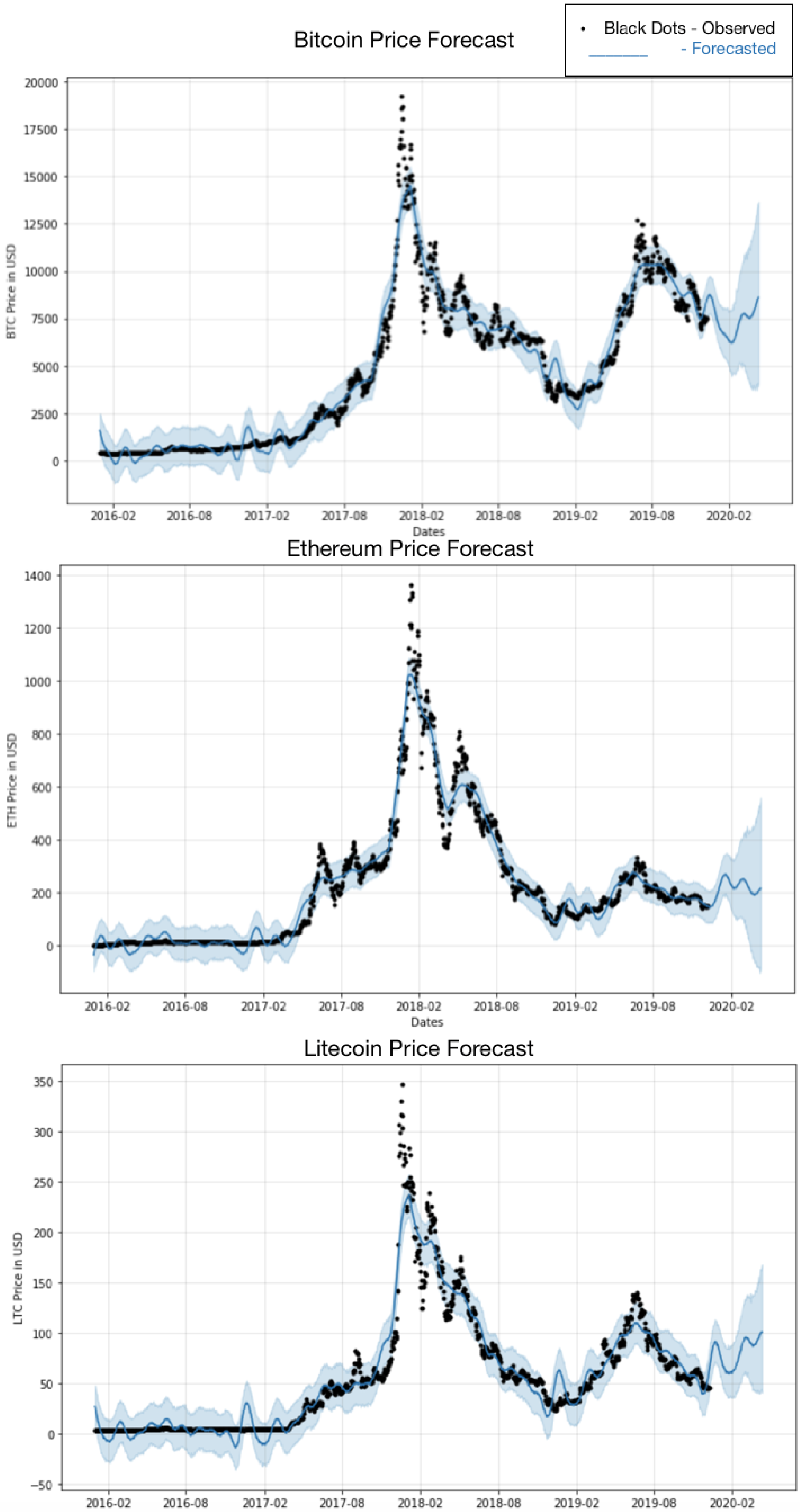

Figure 10: Plots using Prophet for Forecasting

5.4 Discussion

The above experiments presented the results for price forecasted for three different cryptocurrencies using three different algorithms from which the first one, ARIMA which is a statistical model, second being LSTM, a deep learning algorithm and the Prophet timeseries forecasting model. The nature of all these three algorithms is quite different and techniques each of them use to predict the prices is also varied. The results obtained are summarized below in Table 4 on the basis of R^2 values and it can be said that LSTM outperforms both ARIMA and Prophet in terms of correlation between the observed and forecasted prices. However, it needs to be noted that the MAPE values for Prophet showcased strange results and so there is scope for improving or optimizing the MAPE values.

R^2	ARIMA	LSTM	Prophet
Bitcoin	0.82	0.99	0.96
Ethereum	0.73	0.98	0.96
Litecoin	0.66	0.99	0.93

Table 4: R^2 values for the models

The results obtained are average and not extraordinary when comparing to state of art papers and this could be due to two possible reasons. The first limitation being, in this research only historical prices of the cryptocurrencies are considered for forecasting and not other factors that might be responsible for influencing the price of cryptocurrencies. So it might be a good idea to incorporate one more or multiple other features that may be responsible for fluctuation of the cryptocurrency prices. It is always a good idea to include other factors and forecast prices for small chunks of time or even daily if the model forecasts are to be used for trading rather than long periods as this gives scope to models for learning and reduces the error considerably.

Also, the comparatively high error values could be due to the rise and fall of cryptocurrency prices in the last two years as most of the research reviewed and referenced has used prices until 2018 and lacks the second major hike in the bitcoin historical price which occurred during the year 2019. One more factor to be considered is the huge difference in trading volumes and prices of the cryptocurrencies in the top tier and other altcoins. The three cryptocurrencies used in this research more or less show a similar trajectory in the past few years and hence the results obtained do not show any major signs of contradiction in the efficiency of the models. Hence, if altcoins from a different band are selected, which are either maybe extremely volatile or very stable the results obtained might vary.

6 Conclusion and Future Work

This research work implemented forecasting of Bitcoin, Ethereum and Litecoin prices based on historical price using three different machine learning models and achieved substantial results. The values obtained and the plots convey that LSTM outperformed ARIMA as well as the Prophet model in forecasting the price of cryptocurrencies. These models however are not limited only to price prediction but are capable of performing well for various other forecasts and predictions as well and could be used in sectors other that the stocks or financial markets.

To conclude, all three models did showcase a good performance and it could be said that neural network worked best in this case but two key points to be taken into consideration. Firstly, as mentioned earlier all three cryptocurrencies used have had a similar past, it will give a lot of scope for improvement and experimentation if various other altcoins are considered for the analysis. And secondly, note that these models work based on past data i.e. the historical prices and so it is very risky to use the forecasts for trading of cryptocurrencies as there are multiple other factors affecting the prices and by taking these factors into consideration and incorporating them for the prediction may yield even better results.

Acknowledgements

I would like to thank my supervisor Dr. Muhammad Iqbal for his guidance, continuous encouragement and constant support along with his various comments and suggestions that helped in development of the project and improving the writing and presentation of this paper.

References

Alessandretti, L., Elbahrawy, A., Baronchelli, A. and Aiello, L. (2018). Anticipating cryptocurrency prices using machine learning., *Complexity* **2018**.

Bibi, S., Hussain, S. and Faisal, M. I. (2019). Public perception based recommendation system for cryptocurrency., *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Applied Sciences and Technology (IBCAST), 2019 16th International Bhurban Conference on* p. 661.

Brühl, V. (2017). Virtual currencies, distributed ledgers and the future of financial services., *Intereconomics* **52**(6): 370 – 378.

Dev, S., Haruna, I. and Farhana, Z. (2019). Stock market analysis: A review and taxonomy of prediction techniques., *International Journal of Financial Studies* **2**: 26.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory., *Neural Computation* **9**(8): 1735.

Jim, L., Richard Ziyuan, L., Tamás, B. and Avinash, S. (2019). Cryptocurrency investing examined., *The Journal of The British Blockchain Association* .

Karakoyun, E. and Çıbıkdiken, A. O. (2018). Comparison of arima time series model and lstm deep learning algorithm for bitcoin price forecasting., *Proceedings of the Multidisciplinary Academic Conference* pp. 171 – 179.

Kristjanpoller, W. and Minutolo, M. C. (2018). A hybrid volatility forecasting framework integrating garch, artificial neural network, technical analysis and principal components analysis., *Expert Systems With Applications* **109**: 1 – 11.

Lahmiri, S. and Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks., *Chaos, Solitons and Fractals: the interdisciplinary journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena* **118**: 35 – 40.

Mallqui, D. C. and Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques., *Applied Soft Computing Journal* **75**: 596 – 606.

Pant, D., Neupane, P., Poudel, A., Pokhrel, A. and Lama, B. (2018). Recurrent neural network based bitcoin price prediction by twitter sentiment analysis., *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), Computing, Communication and Security (ICCCS), 2018 IEEE 3rd International Conference on* p. 128.

Peng, Y., Albuquerque, P. H. M., Camboim de Sá, J. M., Padula, A. J. A. and Montenegro, M. R. (2018). 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.

Poloniex (2019). `https://www.poloniex.com/`. [Online].

Quandl (2019). `https://www.quandl.com/`. [Online].

Rahman, S., Hemel, J. N., Anta, S., Muhee, H. A. and Uddin, J. (2018). Sentiment analysis using r: An approach to correlate cryptocurrency price fluctuations with change in user sentiment using machine learning., *2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR), Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR), 2018 Joint 7th International Conference on* p. 492.

Saxena, A. and Sukumar, T. R. (2018). Predicting bitcoin price using lstm and compare its predictability with arima model, *International Journal of Pure and Applied Mathematics* **119**(17): 2591–2600.

Singh, H. and Agarwal, P. (2018). Empirical analysis of bitcoin market volatility using supervised learning approach., *2018 Eleventh International Conference on Contemporary Computing (IC3), Contemporary Computing (IC3), 2018 Eleventh International Conference on* p. 1.

Steinert, L. and Herff, C. (2018). Predicting altcoin returns using social media., *PLoS ONE* **13**(12): 1 – 12.

Taylor, S. J. and Letham, B. (2018). Forecasting at scale, *The American Statistician* **72**(1): 37–45.

Wimalagunaratne, M. and Poravi, G. (2018). A predictive model for the global cryptocurrency market: A holistic approach to predicting cryptocurrency prices., *2018 8th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Intelligent Systems, Modelling and Simulation (ISMS), 2018 8th International Conference on, ISMS* p. 78.

Wu, C.-H., Ma, Y.-F., Lu, C.-C. and Lu, R.-S. (2018). A new forecasting framework for bitcoin price with lstm., *2018 IEEE International Conference on Data Mining Workshops (ICDMW), Data Mining Workshops (ICDMW), 2018 IEEE International Conference on, ICDMW* p. 168.

Yenidogan, I., Cayir, A., Kozan, O., Dag, T. and Arslan, C. (2018). Bitcoin Forecasting Using ARIMA and PROPHET, *UBMK 2018 - 3rd International Conference on Computer Science and Engineering* (September 2018): 621–624.

Yu, J. H., Kang, J. and Park, S. (2019). Information availability and return volatility in the bitcoin market: Analyzing differences of user opinion and interest., *Information Processing and Management* **56**(3): 721 – 732.