Prediction of Litecoin Prices using ARIMA and LSTM

MSc Research Project
Msc in FinTech

Ranjani Chandrasekaran
Student ID:X18108423

School of Computing
National College of Ireland

Supervisor: Victor Del Rosal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Ranjani Chandrasekaran
Student ID:	X18108423
Programme:	MSc in FinTech
Year	2018-2019
Module:	Research Project
Supervisor:	Victor Del Rosal
Submission Due Date	
Project Title:	Prediction of Litecoin prices using ARIMA and LSTM
Word Count	Count 6201 Pages 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: Ranjani Chandrasekaran

Date: 12th August 2019

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):	

Prediction of Litecoin prices using ARIMA and LSTM

Ranjani Chandrasekaran
X18108423

Abstract

Since its inception in 2009, cryptocurrency has attracted attention from investors and the academia as well. The significant growth in market capitalization, types and the volatility of cryptocurrencies has spurred considerable interest and studies by researchers. However, most studies are on the hugely popular bitcoin. The latest study shows that Litecoin though occupying the fourth place at 2.74% market cap as on 1st August 2019, has attracted institutional investors. This is a breakthrough, wherein for the first time, cryptocurrency was seen in equal terms with fiat currency. This significant development is attributed to Litecoin's unique features of low cost of mining, real-time speed of transactions, backed by strong blockchain based cost-effective architecture which are capable of handling higher volumes than bitcoin. This research is aimed at the price prediction of Litecoin using machine models. Two models Auto-Regressive Integrated Moving Average (ARIMA) and Long short-term Memory (LSTM) are used for data analysis to understand the best possible model for price prediction. Data collected over a five-and-a-half-year period from 2014 to 2019 is analysed and both the models are evaluated with MAPE, ME, MAE, and RMSE performance parameters. The best results for Litecoin price prediction are achieved when LSTM model is used with a MAPE of 5.759%.

1 Introduction

The past two decades have witnessed revolutionary changes in our life the internet and internet of things (IOT) have literally brought about the world to our fingertips. The nature of changes is like the one that was witnessed in the sphere of transportation, brought about by the advent of diesel engines – which transformed travel - from horse carriages to luxury cars. As jets and ships ferried people and goods across continents and seas, trade and economies changed – from that of isolation and localized markets to a boundary-less global market. The present changes brought about by internet and Internet of Things are in the field of knowledge and communication. Trade and financial transactions today happen though the web. As monopolies in trade are phased out a new era has begun with the advent of cryptocurrencies which pose a direct challenge to fiat currencies.

What started as a single (Bitcoin) and individual attempt has grown to more than 2000+ different currencies now. Cryptocurrencies are characterized by key parameters viz., Price, Market capitalization, market share, volume, rank and age. Each of the newer cryptocurrencies like Ethereum, Litecoin build on the strengths of the forerunner while ensuring weaknesses are eliminated. For example, Litecoin has a much-reduced mining time and mining infrastructure as compared to Bitcoin. Further Litecoin operates at real time transaction speeds and consequently become as strong as fiat currencies. These key aspects have ensured that institutional investors are keen to transact in Litecoin. The interest shown by institutional investors have vested cryptocurrencies with a new legitimacy. This increased acceptance has motivated established and mainstream companies like Facebook to actively consider entering the cryptocurrency arena.
Apart from legitimacy, the key concern for cryptocurrency investors is the volatility.

Bitcoin for example, saw price fluctuation of more than 600% within a five-month span ranging from US$ 2735 in Aug 2017 to US$ 19343 in Dec 2017 (Independent.ie, 2019). Likewise, Litecoin too fell in value during August 2019 to 64.33 US dollars per coin. Litecoin hit the highest price of 237.57 US dollars per coin in December 2017 (Statista, 2019).
Litecoin became strong as against dollar in the first half of 2019 and the sudden drop in prices makes it volatile and hard for speculators to judge and invest (Statista, 2019).
Every cryptocurrency goes through price fluctuations though not as steep as Bitcoin. Since 2013, Cryptocurrencies have shown stability in long term properties. This important parameter has ensured acceptance in the market, and investors are gaining confidence with each passing year. With increased confidence investors treat the cryptocurrency market like any other market. and allocate funds based on the market sample studies.
The cryptocurrency market is growing at a steady rate and as on 1st August 2019 is 275 Billion USD (CoinMarketCap,2019).

The key factors which determine the market cap of a cryptocurrency are as follows:
1. The cost of mining
2. The complexity of mining
3. Speed of operations &
4. Coins in circulation

In terms of market share, Litecoin is the fourth largest cryptocurrency ranked behind Bitcoin, Ethereum and Bitcoin Cash and has a market cap of 2.34% of the overall cryptocurrency market as on 1st August 2019 (CoinMarketCap,2019). Yet in terms of complexity of mining infrastructure requirement, it is the lowest and mining time is the shortest. Further by introduction of high-speed transaction it has ensured that speed of operations is real time and at par with fiat currency.

Although in the cryptocurrency market Litecoin has only a small presence of the market share now, it is also unique at it attracts institutional investors. This is due to several important factors like less speed of transactions, better security, lesser transaction costs and the algorithm used is Scrypt. Further studies have also established that Litecoin has a definite and causal relationship with fiat currency like Chinese Yuan, Thai Baht, Taiwan Dollar (Corelli A. 2018)

Therefore, Litecoin has potential for enormous growth and the possibility of being one of the first cryptocurrency to be adopted for standard exchange rates of fiat currency.
Cryptocurrency market has two unique features. First is the development and continued growth of a self-organized market of virtual currency. Secondly the assets and the value thereof generated by social consensus (CoinMarketCap,2019). Both these factors have attracted the scientific community towards research of cryptocurrency.

There are many researchers' here are many researches on the price prediction and volatility of cryptocurrency with respect to various aspects of markets using various statistical measures (Alessandretti et al., 2018). However, the studies are either generic for cryptocurrencies or are focused on the more popular one viz., Bitcoin. The research aims to fill in the gap of study and forecasting price fluctuations of Litecoin.

This research aims to answer the question: "What is the impact of ARIMA and LSTM in the level of accuracy for prediction of Litecoin prices?"

In this paper, we have used two of time series analysis methods – ARIMA & LSTM in order to find out the pattern of Litecoin price movement and forecasting the closing price for one month as well as analyzing the comparative performance of the time series models. ARIMA is chosen as it is best suited for time-series analysis wherein prediction is done for one variable at different time points.

LSTM is an effective method in sequence predictions since the method factors in values of previous prices of stock to accurately predict future prices.

The study is conducted for a period of five and a half years from 2014 to 2019. Data collection is done on everyday basis to accurately predict Litecoin prices over a span of 31 days (Table 1 and Table 3). This paper is organized as follows: Section 2 highlights the related work of earlier research papers in the field of cryptocurrency; section 3 describes the research methodology carried out; section 4 details the design specification required to carry out research; section 5 explains the implementation process;

section 6 is the evaluation between two models and section 7 is the conclusion of the research and opportunities for future work exploring the same subject in further detail.

2 Related Work

As a result of increasing investor interest, research on cryptocurrency too has increased in the past few years. Recent advancement in machine learning has facilitated the research process. Literature on Litecoin is limited and this provides the opportunity for investigation.

(Greaves and Au, 2015) analyze and predict the price of Bitcoin using well established statistical models and transaction graphs. The authors have used Logistic Regression, Neural Network and SVM methods for the analysis of transactions. The comparative results obtained in their research are using Linear Regression with MSE at 1.94, with SVM 1.98 and baseline at 2.02. In the classification models the highest accuracy obtained is with neural network at 55.1%, with baseline model at 53.4%, with Logistic regression at 53.4% and with SVM at 53.7%. (Madan, Saluja and Zhao, 2015) have researched on data collected over ten-minute interval spans to predict price at varying levels of noisiness. The data is sourced from okcoin.com. This has over 25 features and data has been captured daily over a period of five years. Applying binomial classification and generalized linear models for data, the researchers were able to predict Bitcoin price with an accuracy of 98.7%. (KARASU et al.,2018) analyses and predict bitcoin prices. For the research data is collected daily and the authors have applied established models such as Linear Regression and support vector machine (SVM) with kernel functions. Having computed various combinations, the authors finalized the model with least error after evaluation using measurement parameters such as MAE, MSE, RMSE and Pearson correlation. The authors have concluded that polynomial support vector machine with weighted moving average filter of two days gave the most accurate results. (Mcnally, Roche and Caton,2018) The authors have analyzed data collected for prediction of bitcoin prices. They have used Bayesian optimized recurrent neural network, LSTM and ARIMA models for price prediction. The authors factored in the concept that small margins can make a big difference in financial markets and therefore collected data for a period of three years. The researchers concluded the study with an accuracy levels of 52% for LSTM and RMSE of 8%. (Shah and Zhang,2018) The authors have sourced data from okcoin.com and analyzed in detail to conclude on return of investments for bitcoin. The research has effectively used Bayesian regression and efficacy for prediction of bitcoin price movements over a period. The authors have researched on possible strategy for doubling of investment within a period of sixty days. (Shah and Zhang,2018) have arrived at following conclusions: Firstly, there is an inverse relation between number of trades and threshold. Second, there is a direct relation between threshold and holding time. Thirdly a direct correlation between threshold and average profit made, with a sharp ratio of 4.10 the return of investments in bitcoin is about 89% over a period of fifty days. (Phaldisailoed and Numnonda, 2018) The research aims to understand the best way of predicting price of bitcoin.

The authors have strived to understand which model delivers the best results - both in terms of effectiveness and accuracy. The authors have used data collected over a period of six years daily with one-hour intervals.

Deep learning regression models viz., LSTM and gated recurrent units (GRU) and among the regression models Theil-Sen regression and Huber regression models are used and results are compared. The authors conclude that deep learning models took more time for processing data but were more accurate than the regression models. The accuracy measured in R-square was 99.2% and the mean square error at 0.00002 in the GRU. (Sin and Wang 2017) The authors research focused on the correlation between the day end prices of bitcoin on one day to the changes of price on the following day. The authors used the data collected and analyzed using following deep learning methods; Artificial neural network and genetic algorithm based selective neural network (GASEN) models. The research concluded the profitable returns stood at 85% an accuracy of the models ranging from 58% to 63%. (Mallqui and Fernandes,2018). The authors research focused on two results of bitcoin. The first objective aimed to predict all three intra-day parameters – the max, min and closing prices.

Another objective aimed to predict the direction of the movement of the price. For the first objective of the day's max, min and closing prices, the authors made use of regression models like ANN, RNN and SVM. Whereas for meeting the second objective of obtaining the price movement, the authors have used Support vector machine, Artificial neural network and ensemble algorithms – particularly the recurrent neural network and K-means clustering tree method. The authors concluded that the ANN method had the highest accuracy at 62.91% while the accuracy of SVM method was at 52.73%. (Roy, Nanjiba and Chakrabarty, 2018) The research aimed at forecasting bitcoin prices using time series analysis models. The data collected was for a period of four years. Three models viz., Auto regressive moving average ARIMA, auto regression (AR) and moving average (MA) used for forecast. The authors concluded that Auto regressive integrated moving average model gave the best results with an accuracy of 90% whereas the other models had accuracy levels of 89% and 87.58%. The study predicted price data for a period of ten consecutive days. (Saad and Mohaisen,2018) The authors research focused on concluding the most accurate model for prediction of bitcoin prices. The authors researched data evaluated over a twenty-month period. Using regression and deep learning models – Linear Regression (LR), gradient descent and random forest. Different attributes of cryptocurrency and estimation of features have been analyzed. On comparison the researchers concluded the linear regression models gave the best desired results with an accuracy rate of 99.44%, RMSE rate of 0.0113 and MAE of 0.0060 (Wu et al,2018) The authors of the research paper aimed at concluding the best way for forecasting bitcoin prices daily. The authors used two different models to map and analyze the data which was collected for a seven-month period cumulating to two hundred and eight data sets. LSTM conventional and LSTM with auto regression were the two models chosen for the study. Both the models were gauged based on RMSE, MSE, MAPE & MAE. The conclusion arrived at by the authors is that the results of LSTM with autoregression is more accurate than the LSTM conventional model. The authors (Abraham et al,2018) researched the price predictions of bitcoin using data collected from social media. The authors have explained the impact of social media like Twitter and Google trends in the decision making. Higher volumes of messages on social media like Twitter and other news medium influences peoples thought process and consequently purchasing decisions. The authors have used only the volumes of transactions. Though the sentiments of the messages whether positive or negative were also mapped and observed to be positive, the authors did not use the sentiment-based data and used only the volume-based transactions. After the data was filtered for noise linear regression model with correlation matrices were used for analysis. The authors concluded that among the cryptocurrencies Bitcoin and Ethereum had correlation to the volume of social media outputs. (J R and Das, 2018) The study focused on price prediction for six leading cryptocurrencies viz., Dash, Ethereum, Monera, Ripple, Litecoin and Bitcoin. The authors used one model LSTM model for the analysis. The study finalized that as compared to other cryptocurrencies, Bitcoin prediction was accurate with MAE of 0.038. This research paper does not take any other parameter or model for analysis and therefore could not conclude whether LSTM gives the best results. (Garg,2018) The research has focused on prediction of price for bitcoin using auto regression integrated moving average ARIMA model. (Sun, Zhou and Lin, 2019) the authors have used random forest algorithm to predict prices of cryptocurrencies. The research paper concluded that certain factors are significant in predicting price movements.

The other conclusion is with data of longer duration a model can make more accurate predictions. (Matta, Lunesu and Marchesi, 2019) The research is on prediction of bitcoin price prediction using Automated sentiment analysis. The authors have cross correlated data of Google media trend and volume of tweets to predict the movement of prices.

(S. Vassiliadis, 2017) In this research paper of combination of GARCH model with the SVR is used to evaluate performances of popular cryptocurrencies like Bitcoin, Ethereum and Dash with fiat currencies like Euro, Pound & Japanese Yen. Daily data classified as low and hourly data classified as high, were used to predict the volatilities. The research concluded that the errors, RMSE and MAE, high frequency data are much lower than those calculated for low frequency data. (Y.Peng, 2018) The paper concluded that Bitcoin volatility had a definite relation with quantum of transactions done and also with stock market indices.

(M. Balcilar,2017) has researched on Bitcoin price volatility and trading volume. The researcher concluded that while volume had a causal relationship with returns but did not show the same results with volatility. Data is collected from coinmarketcap.com data base for a period three years and nine months starting from Jan 2015. Bitcoin opening, closing, high and low prices for the day was mapped. The study achieved accuracy level of 60 to 70% with error of less than 6%. (Rane and Dhage 2019) The authors have emphasized the fact that due to increased volume of transactions of cryptocurrencies large volume of data is available. Completing the with this large data the authors have used a variety of models to conclude on the best possible way for accurate price prediction. The authors have used following models: support vector machine model, binomial generated linear model, non-linear auto regressive model, auto regressive moving average model, long short-term memory model, multilayer perception neural network model and regression model for the comparative study. They have concluded that among all the models the best was the ARIMA model.

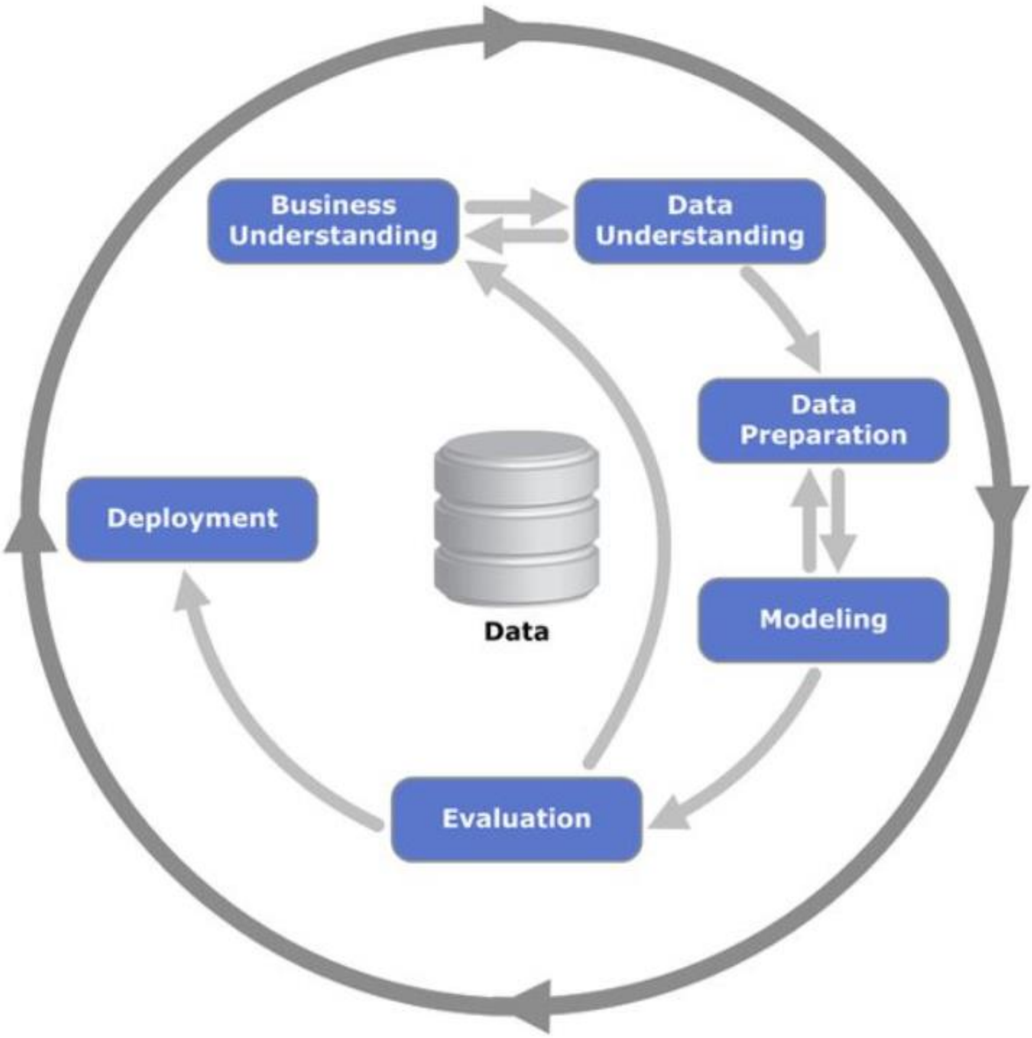
3 Research Methodology

This research paper uses the Cross industry standard process that is CRISP-DM for data mining. The CRISP-DM data mining process was developed by a consortium of Daimler Chrysler, SPSS and NCR in the years 1999- 2000. The process envisages a clear link between business understanding and data mining. As Wlliam Vorhies, one of the creators of CRISP-DM confirms that any data mining activity must begin with a clarity in the business process. Business understanding and clarity will not only help to collect data but also to retain relevant data by cleaning unwanted aspects. In this way CRISP-DM gives a firm guidance and roadmap for successful data mining.

The CRISP-DM Process Lifecycle:

The CRISP-DM model maps the whole process of data mining project in six phases (Fig 1) The phases do not have specific and rigid sequence. Back and forth movement between different phases is required to achieve desired results. The result of one phase will help to choose which is the next phase or even which specific task of the next phase that needs to be performed. The directions indicated in the figures gives the dependencies and key result-determining phases. The learning process is continuous and is therefore symbolically shown in the outer circle. The process does not end once the solution is created and deployed. The results obtained from a process is not an end by itself and is often observed to start fresh and pointed business related queries. In fact, the latter data mining processes will get the advantages of the earlier results (Smart Vision - Europe, 2019).

Cross industry standard process for data mining



CRISP-DM process diagram (Wikimedia Commons, 2012)

Figure:1

3.1 Business Understanding:

The business objective of this research paper is to predict the prices of Litecoin to enable considered investment by institutional investors and individual investors. There are a lot of factors that influence the cryptocurrency market which make the coins highly volatile. However, a huge volume of data is available. With this enormous data, a detailed study using appropriate deep learning algorithm models will enable accurate prediction of prices (Smart Vision - Europe, 2019). Among the several factors, this research paper will focus on price and consider that as a variable for prediction. Well established machine learning algorithms such as ARIMA and LSTM are used to achieve this objective.

3.2 Data Understanding:

The data collected must be absolutely in alignment with the business or research needs. Otherwise there is the risk that the data collected and analysed is not in line with the requirements. Hence right at the outset, it is essential that there needs to be total clarity regarding the data to be collected. A step by step process is utilised to meet data understanding needs.

First, the data collected must be ensured to meet business plans and objectives; at times data collection will throw up the need to revisit business plans. Second is to ensure that data collected is explored in detail for usage of appropriate formats, samples and timelines. Third is to ensure the verification of the quality of data. With these steps the process of data understanding will be appropriate, and the analysis will yield useful outcomes for the business objectives (Smart Vision - Europe, 2019).

3.3 Data Preparation:

Data preparation is a very crucial aspect which determines the accuracy of the implementation. This is because data collected is raw and is prone to many anomalies. These anomalies must be identified and treated appropriately before the data is taken for analysis. The clean-up and transformation of data includes identification and appropriate treatment of missing entries, elimination of duplicate entries, standardizing and integration of similar entries to reduce (Smart Vision - Europe, 2019). Data thus verified and enriched is then uploaded as csv file for application of the deep learning techniques.

3.4 Modelling:

Applying the most appropriate techniques to fit the model in the data is key to achieving desired results. In this research two deep learning modelling techniques are used to predict price of Litecoin. The models used are: ARIMA-Autoregressive Integrated Moving average and Long short-term memory. The key aspects of these modelling techniques are as explained below.

3.4.1 ARIMA:

The ARIMA model is most effective for analysis of time series data and to forecast future results. This technique uses a combination of auto-regression and moving averages. As a first step of the process the data needs to be made stationary that is to ensure there is constant variance, mean and covariance. The methods that are used to make the data stationary are differencing for linear trends, applying log in case of non-linear trends and applying log seasonal difference in case of seasonality (Medium, 2019). An easier and effective way to do this is to by installation of Auto ARIMA in the library. The algorithm in the auto ARIMA itself performs the stationarity and plots the ACF and PACF.

3.4.2 LSTM:

Long short-term memory models are an RNN architecture which are very effective in sequence prediction. LSTM cell has an input gate and two output gates viz., the forget and remember gates (Brownlee, 2019). The cell evaluates the information and can pass or block the information obtained based on the strengths of the signals. LSTM is programmed to evaluate and comprehend the time period of holding earlier information while in parallel it considers what to remember and what to forget (Brownlee, 2019). The data that is retained is also connected with the new inputs which enables it to carry forward meaningful long-term dependencies. The weights are also propagated back and forward through layers unlike other modelling techniques. This singular characteristic ensures that the weightage is retained is for entire processing (Brownlee, 2019). This is important in our research since the previous price of cryptocurrency is crucial in predicting the future price.

3.5 Evaluation:

The final and key aspect of research is to understand the accuracy of the forecast that is done.
Since two models are used in the research, to conclude on the performance, both the models are compared.

4 Design Specification

This paper has used R Studios to build the project and the code is written in R language which is widely used programming language. Among the languages, the R language is well suited for statistical computing and graphics since it provides the ideal integrated open source environment. The other key advantage of R language is that it integrates seamlessly with comma separated values in which data is collected and prepared for the research. This research uses RStudio cloud for neural network which is compatible with tensorflow. This research paper is completed using neural networks of Keras libraries.

5 Implementation

While cryptocurrencies are the sought-after investment options preferred by investors, it also poses higher risks due to inherent volatility which are consequence of its framework and structure. The solution provided by this research is to establish accurate models that will help both individual and institutional investors to make informed decisions on investing. The copious data available is leveraged and analyzed using established techniques like ARIMA and LSTM to conclude on the most accurate method. The data used in the research is collected from coinmarketcap.com. Although Litecoin was incepted since 2011, the data available in the site is from the year 2014 onwards. Data is collected for a period of five and a half years which is from 14th January 2014 to 7th July 2019. Each detail of the data is keenly understood. The data set has a total of 1991 attributes with six key variables within a day. In these, five variables relate to price which is highest price, lowest price, opening price, closing price and weighted price while the sixth variable is volume of trade. For the implementation and prediction, closing prices are considered as a predictor. The data is read in the comma-separated format (csv). All seven variables mentioned above are imported into R studio. The data is univariate time which means prediction of one variable will occur at different time points.

The objective of this research paper is to predict prices of Litecoin in an accurate manner. In this section research approach and the methods deployed are explained. Two machine learning models, ARIMA and LSTM are used on the data collected.

Implementation of Proposed Methodology in Price predicting of Litecoin:

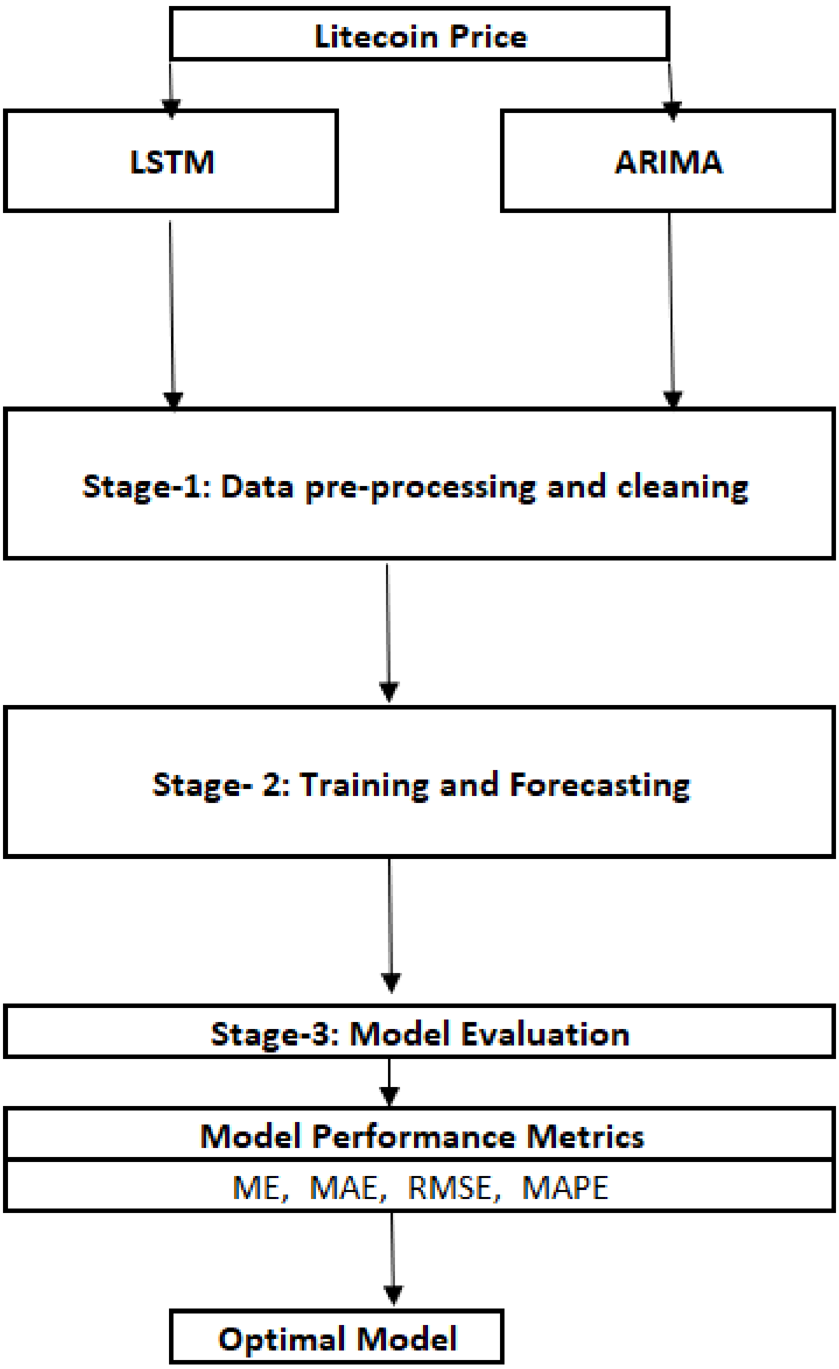


Figure:2

5.1 Implementation Of ARIMA:

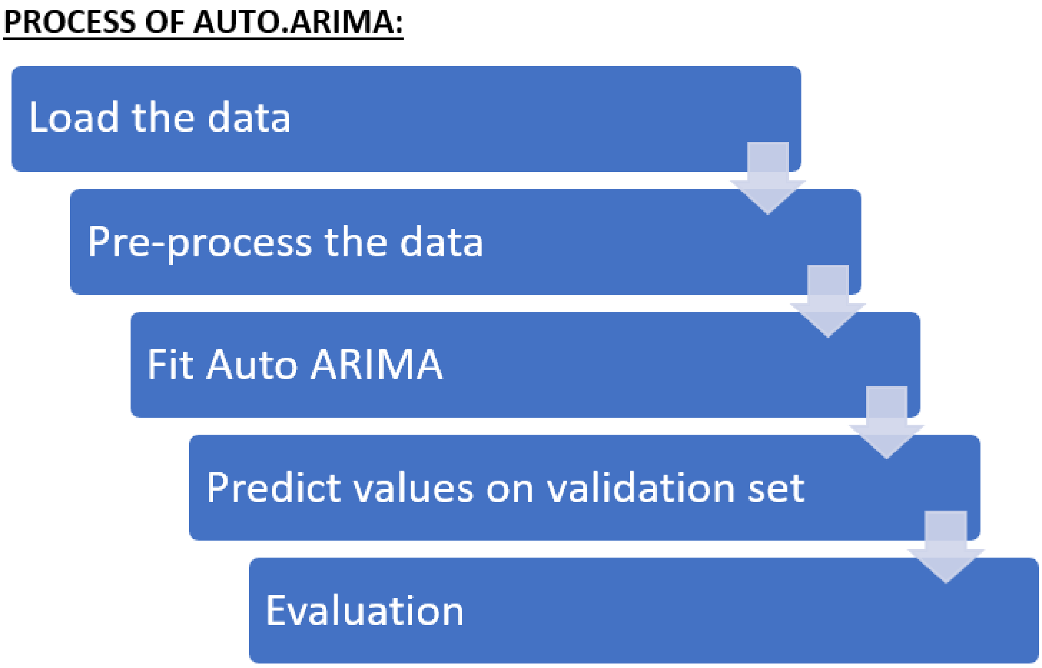


Figure 3

Implementation of ARIMA is in a series of stages. First it starts with importing the data that is collected. After the data is imported the data must be cleaned to handle missing values and treatment of outliers so that results are accurate. In the dataset of Litecoin prices there were no missing values and no outliers to be treated. The data after pre-processing is made ready to fit model. Before fitting model for prediction of values few packages like CaTools, Lubridate (RStudio,2019) are needed to be installed as library in RStudio. The CaTools has several basic functions like moving (running, rolling) and other windows statistic files in addition to read, write, GIF file, roundoff, cumsum and other features in addition to having the predict algorithm. The Lubridate is a convenient and efficient package for date and time format, especially useful for time series studies (RStudio,2019). Few other libraries such as fpp2, seasonal, fma, grid, forecast, dplyr, Mlmetrics, stargazer is installed (RStudio,2019). The data needs to be converted into time series data and to format the date in the data set. The closing price is retained as that is a variable for prediction. The entire data is split into test and train modules, with 60% of data for training and remaining 40% for testing. Initially a plotting is made on close price daily to observe and deal with the data. Seasonality is taken into consideration wherein the seasonality component is looked deeper by season plot and seasonal subseries plot whereby this helps us to understand the arrangement and identifies any major shifts in seasonality. Plotting of Auto correlation function proves that if pairs of data show correlation, in other words measures extent of linear relationship between lagged values of a variable. After the ACF has been plotted, the time plot shows clear trend and seasonality where in the exponential decaying of ACF shows trend and anything above the blue line is statistically significant.

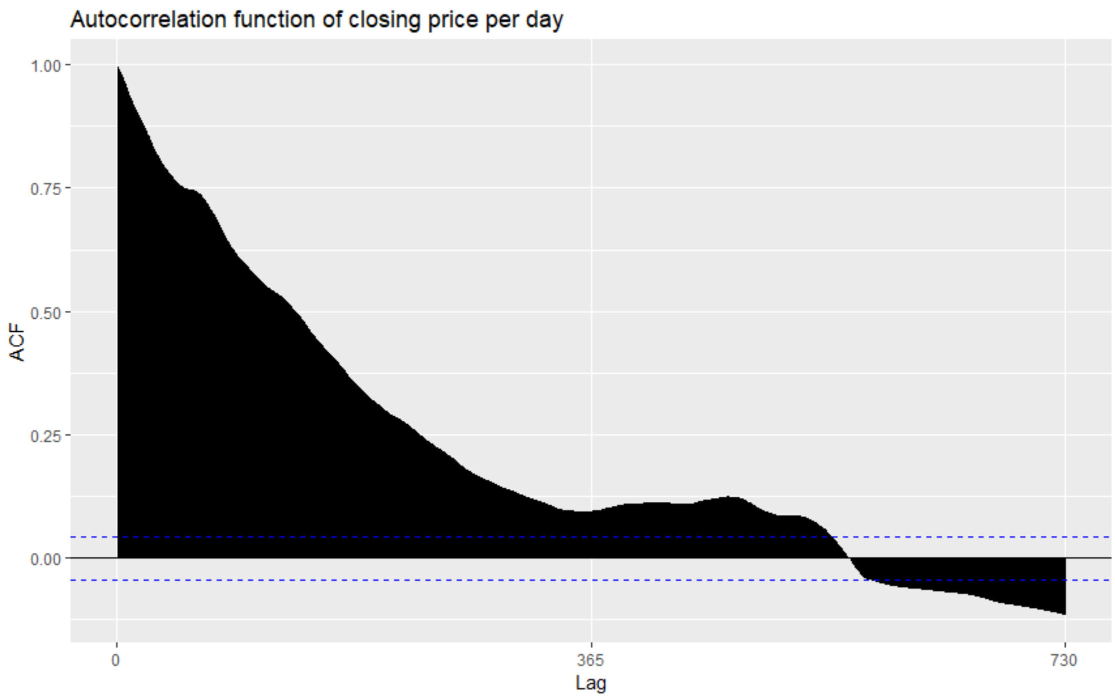


Figure 4

The data is forecasted to arrive at the mean, naïve, seasonal naïve and random walk drifted. Finally, graph for multiplicative decomposition is drawn below in figure.

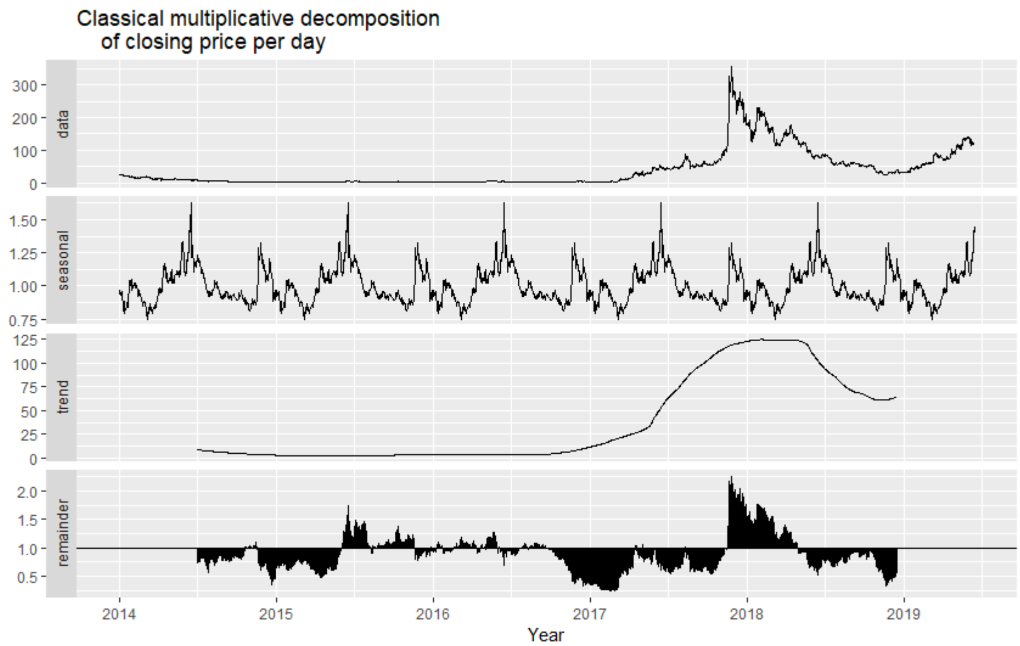


Figure: 5

Fitting the model of auto arima the results obtained are up to third order of lag, differencing is first order and third order of error, that is p=3, d=1, q=3. The arima (3,1,3) is fitted to the model in training data. The inference of this observation is that the correlation can be made for 3 previous day's prices wherein the sigma square 27.57, aic in 12072, bic is 12111.06 and aicc is 12072.06. After obtaining the order, the model is tested on the remaining 40% test data where the estimated /predicted value is obtained.

5.2 Implementation of LSTM:

Implementation of Long Short-Term Memory is done in the following stages. First stage is the importing of following packages and libraries to ensure effective and speedy processing. Library readr version 1.2.0 (RStudio,2019) to provide fast and user-friendly reading of csv files. Library tseries for time series analysis and computational finance. Tidyverse which is a system for data manipulation, exploration and visualisation required for workflow coverage (RStudio,2019). Keras for core data structure and layering. Lubridate for effective handling of date and time. caTools for several utility functions like moving (running, rolling) and other windows statistic files in addition to read, write, GIF file, roundoff, cumsum and other features in addition to having the predict algorithm. With the all the required library packages is completed we proceed to the next stage, which is importing the data set, formatting the date. Data is processed in two sub-sample sets. 60% for testing and 40% for training to validate the results obtained. and plotting for examining the dataset. The data is tested for its stationary by KPSS test where in the p value is 0.01 which is less than alpha = 0.05, so here the null hypothesis of stationarity is rejected, and alternative hypothesis of non-stationarity is accepted.

```
	KPSS Test for Level Stationarity

data:  data$close
KPSS Level = 9.5589, Truncation lag parameter = 8, p-value = 0.01
```

The data for analysis in LSTM model is to be done on supervised mode. This means that the data has a target variable Y and a predictor X. Here the target variable is the Litecoin closing prices which is used for prediction. Transformation of Litecoin closing prices is done by using log and looking back by 1 is taken into consideration. The data is normalized by using feature range parameter with a default value (0,1). The next step is to use the default activation function which for LSTM is the sigmoid function with the range (-1 to 1). A function is defined for invert scaling. The network is looped which examines every window till all training data of 1194 observations are completed.
The LSTM model is fit by having numeric class, batch size 1 and units =1. In this the mean square error is used to reduce errors. The model summary is as follows:

```
Model: "sequential"
________________________________________________________________________________
Layer (type)                        Output Shape                       Param #
================================================================================
lstm (LSTM)                         (1, 1)                             12
________________________________________________________________________________
dense (Dense)                       (1, 1)                             2
================================================================================
Total params: 14
Trainable params: 14
Non-trainable params: 0
________________________________________________________________________________
```

The epochs which is 25 iterations is on 1194 observations which is the training data.

```
1194/1194 [==============================] - 4s 3ms/sample - loss: 0.0049 - acc: 8.3752e
-04
1194/1194 [==============================] - 4s 3ms/sample - loss: 0.0051 - acc: 8.3752e
-04
1194/1194 [==============================] - 4s 3ms/sample - loss: 0.0049 - acc: 8.3752e
-04
1194/1194 [==============================] - 4s 3ms/sample - loss: 0.0049 - acc: 8.3752e
-04
1194/1194 [==============================] - 4s 3ms/sample - loss: 0.0049 - acc: 8.3752e
-04
1194/1194 [==============================] - 4s 3ms/sample - loss: 0.0051 - acc: 8.3752e
-04
```

Once the iterations is done, modeling is done on the test (797 observations) data.
Plotting the predictions:

Prediction close price for 797 observations

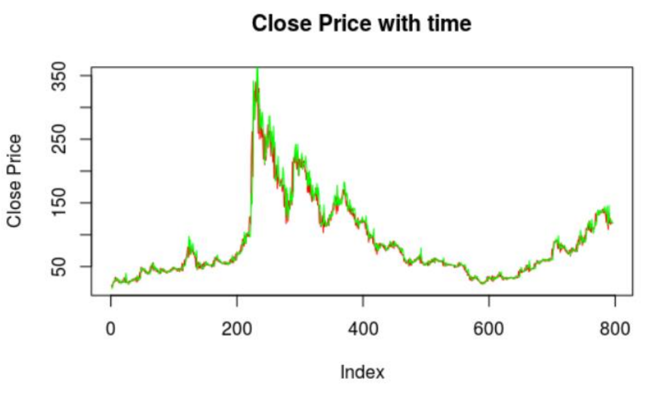


Figure: 6

Prediction for 1194 observations

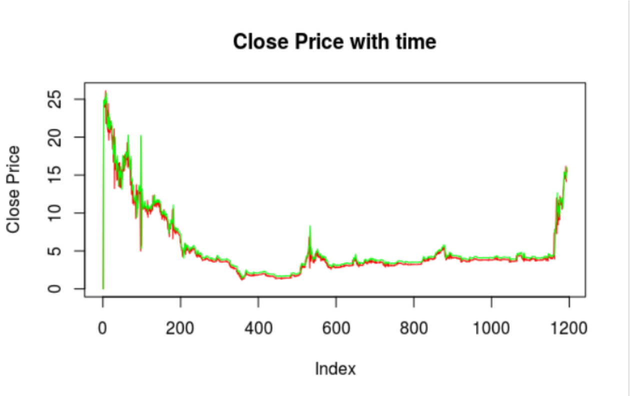


Figure: 7

Prediction close price for 1991 observations:

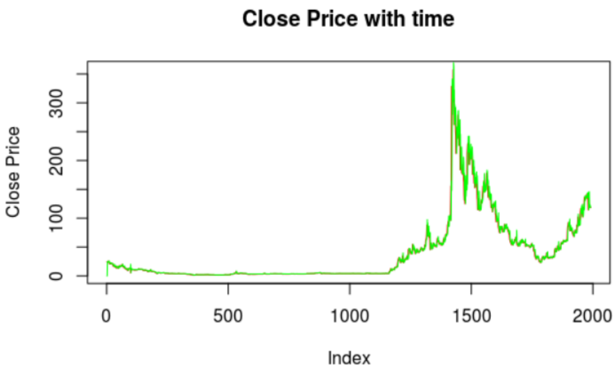


Figure: 8

6 Evaluation

Since the data is analyzed using two different models, it is essential to evaluate the models with appropriate metrics to conclude on the best fit. In the research, the observed value and forecasted value are recorded. The difference between two recorded values is the error. Using the error values ARIMA and LSTM are evaluated for the following parameters: Mean error, Mean Absolute error, Root Mean Squared error, and Mean Absolute percentage error.

6.1 Evaluation of ARIMA

By de-composing and fitting the model in train data, it is observed that the result matches to ARIMA 3,1,3 model. The ARIMA (3,1,3) thus obtained is run in the test data and the estimated value for 31days is recorded. The 31-day data is compared with the actual data and the absolute values are defined. For these absolute values Mean error, mean absolute error, Root mean squares error and MAPE -mean absolute percentage error is calculated and tabulated below in Table 1. The absolute value is calculated with each day as the difference between the true value and estimated value. Subsequently the mean error, absolute error, root mean square error and mean absolute percentage error are calculated and placed in table 2.

ARIMA (Values in dollars ($) per litecoin)			
Date	True_value	Estimated_value	Absolut_valu
07/06/2019	118.51	117.754517	0.7555
08/06/2019	114.87	117.0456026	2.1756
09/06/2019	129.83	118.0432328	11.7868
10/06/2019	136.08	118.0581625	18.0218
11/06/2019	136.16	117.5359661	18.6240
12/06/2019	130.86	118.5375884	12.3224
13/06/2019	132.71	118.0578934	14.6521
14/06/2019	138.35	117.9139382	20.4361
15/06/2019	136.95	118.7277526	18.2222
16/06/2019	134.19	117.973186	16.2168
17/06/2019	135.13	118.2424109	16.8876
18/06/2019	136.83	118.7113691	18.1186
19/06/2019	135.78	117.9233369	17.8567
20/06/2019	139.07	118.5131325	20.5569
21/06/2019	141.77	118.5680223	23.2020
22/06/2019	136.83	117.9596748	18.8703
23/06/2019	135.4	118.6966779	16.7033
24/06/2019	135.51	118.3726502	17.1373
25/06/2019	130.52	118.0842537	12.4357
26/06/2019	114.24	118.7712942	4.5313
27/06/2019	119.46	118.1921098	1.2679
28/06/2019	133.44	118.2660885	15.1739
29/06/2019	122.16	118.7367311	3.4233
30/06/2019	122.67	118.0762432	4.5938
01/07/2019	118.68	118.4579139	0.2221
02/07/2019	121.97	118.6164259	3.3536
03/07/2019	119.67	118.0504131	1.6196
04/07/2019	118.53	118.6124168	0.0824
05/07/2019	118.31	118.4512094	0.1412
06/07/2019	118.33	118.113209	0.2168
07/07/2019	118.51	117.754517	0.7555

Table: 1

ARIMA	
Performances Metrics	**Values**
Mean Error	10.2097
Mean Absolute Error	10.6568
Root Mean Square Error	13.2853
Mean Absolute Percentage Error	0.07926

Table: 2

The auto.arima () function installed from forecast package and has arrived at the ARIMA (3,1,3) which is the best fit model for predicting prices of Litecoin. The selection of best fit of model is done in the background by R (Box et al.; 2015).

The result generated by the ARIMA model is shown below

method	character [1]	'ARIMA(3,1,3)'
model	list [18] (S3: ARIMA, forecast_ARI	List of length 18
level	double [2]	80 95
mean	double [30] (S3: ts)	118 117 118 118 118 119 ...
lower	double [30 x 2] (S3: mts, ts, matri:	111.0 107.2 105.9 103.9 101.5 100.9 107.5 101.9 99.5 96.4 92.9 91.5 ...
upper	double [30 x 2] (S3: mts, ts, matri:	124 127 130 132 134 136 128 132 137 140 142 146 ...
x	double [1961] (S3: ts)	24.5 24.5 24.0 24.0 24.9 24.9 ...
series	character [1]	'v1_d_train'
fitted	double [1961] (S3: ts)	24.5 24.5 24.5 24.0 24.0 24.9 ...
residuals	double [1961] (S3: ts)	2.45e-02 -1.05e-06 -4.98e-01 1.18e-02 9.02e-01 -3.72e-02 ...

Figure: 9

6.2 Evaluation of LSTM

The dataset for LSTM model is split into 60% train and 40% test with a batch size of 1 and units 1. The LSTM model is run with 25 epochs. The details of results for 31 days are listed below in dollars per Litecoin.

LSTM (Values in $ per litecoin)			
date	true_value	estimate_value	absol_est
06/07/2019	117.08	104.1328	12.9472
06/08/2019	118.51	111.6928	6.8172
06/09/2019	114.87	117.3628	2.4928
06/10/2019	129.83	119.0006	10.8294
06/11/2019	136.08	115.1528	20.9272
06/12/2019	136.16	130.1128	6.0472
6/13/2019	130.86	136.3628	5.5028
6/14/2019	132.71	137.6477	4.9377
6/15/2019	138.35	131.1428	7.2072
6/16/2019	136.95	132.9928	3.9572
6/17/2019	134.19	138.6406	4.4506
6/18/2019	135.13	137.2622	2.1322
6/19/2019	136.83	134.4728	2.3572
6/20/2019	135.78	135.4128	0.3672
6/21/2019	139.07	137.1139	1.9561
6/22/2019	141.77	136.0628	5.7072
6/23/2019	136.83	139.3528	2.5228
6/24/2019	135.4	142.9885	7.5885
6/25/2019	135.51	137.2072	1.6972
6/26/2019	130.52	135.6828	5.1628
6/27/2019	114.24	137.5524	23.3124
6/28/2019	119.46	141.1800	21.7200
6/29/2019	133.44	114.5228	18.9172
6/30/2019	122.16	119.7428	2.4172
07/01/2019	122.67	141.6939	19.0239
07/02/2019	118.68	122.4428	3.7628
07/03/2019	121.97	124.7918	2.8218
07/04/2019	119.67	118.9628	0.7072
07/05/2019	118.53	122.5728	4.0428
07/06/2019	118.31	119.9968	1.6868
07/07/2019	118.33	118.8128	0.4828

Table: 3

The error values is used to calculate the performance metrics below:

LSTM	
Performances Metrics	**Values**
Mean Error	-0.7701
Mean Absolute Error	7.2707
Root Mean Square Error	10.2684
Mean Absolute Percentage Error	0.05759

Table: 4

Here in the below chart, date in in x axis and prices in y axis, the estimated (predicted) value is in blue and true/actual value is in red. The prediction shown below is the actual and predicted values for predicted 31 days (07/06/2019 to 07/07/2019).

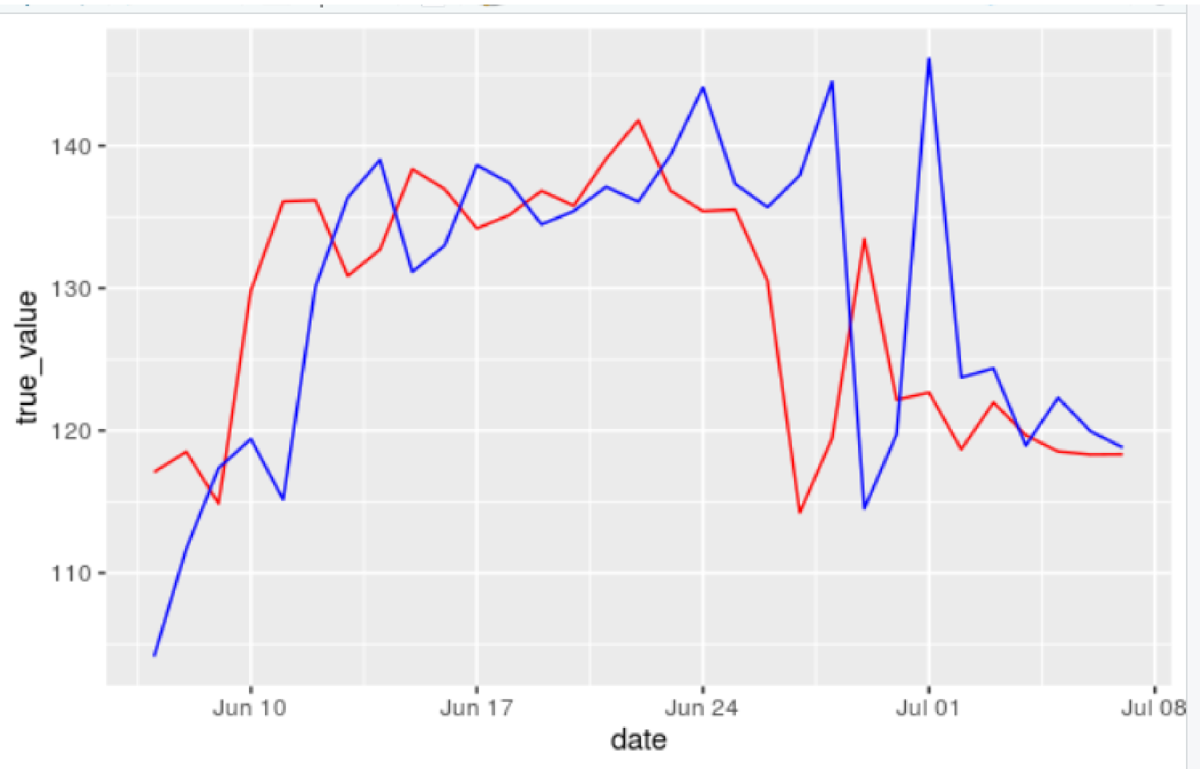


Figure: 9

6.3 Discussion

The objective of this research paper is to understand the impact of machine learning on the price prediction of Litecoin. To meet the objective, two models ARIMA and LSTM are chosen. While ARIMA is known to be robust and efficient in time series forecasting, LSTM model is powerful in sequence prediction problems. This is because of important feature of using a “remember and forget” gated architecture and carry forward relevant past information for accurate long-term dependencies.
In the design specification template, a compatible hardware of core i5+ 8th Gen CPU, 4 GB RAM is used. To effectively run the models, intelligent and user-friendly software combination of Rstudio, Rstudio cloud, Lubridate, caTools, fpp2, forecast, Mlmetrics, dplyr, stargazer, seasonal installed. Data is collected for a period of five and half years on daily basis providing an effective dataset for analysis. The pre-processed and cleaned data is run with a split between testing and validation. For ARIMA (3,1,3) the data set after first order differencing has become stationary and is correlated to 3 previous day’s closing prices. The results of the models are compared by different error calculations of MSE, MAE, MAPE and RMSE to evaluate the accuracies of the results to understand the impact of machine learning models on accurate price predictions. Additionally, the best fit model with the least error is also ascertained.

7 Conclusion and Future Work

7.1 Conclusion

It is well known that Machine learning algorithms are useful in solving real world problems. In this research, the main goal is to ascertain the impact of machine learning models ARIMA and LSTM is terms of accuracy for prediction of Litecoin prices. Both ARIMA and LSTM are time series models well suited for the research. Despite having different architectures, it is observed that with available dataset, an appropriate design specification and well implemented process, both the models predict Litecoin prices accurately.

The dataset is collected for a period over five and half years from 14th January 2014 to 7th July 2019 for modelling and prediction of price for a period of 31 days. The data collected is pre-processed to normalize and eliminate outliers. The data is then split into 60% for training and 40% to validate by testing. The actual prices and estimated prices are recorded. Assessment of both the models are made by using same performance metrics: the mean error, mean absolute error, root mean square error and mean absolute percentage error. Using a dataset of 1991 observations and predicting prices for 31 days with 25 epochs, the results of the LSTM model are with mean error of negative 0.7701. The negative here indicates that the predicted value is less than actual values; The mean absolute error is 7.2707, mean absolute percentage error is 5.759% and the root mean square error is 10.2684. The LSTM model values are lesser than ARIMA model values which has a mean error of 10.209, mean absolute error of 10.6568, mean absolute percentage error of 7.926% and the root mean square error of 13.2853.

From the above comparison of metrics for the two models used, the LSTM predicts the price of Litecoin with greater accuracy than ARIMA. Since both individual and institutional investors prefer investing in large amounts, even marginal variations in price prediction will result in huge impact in their portfolios. Hence, we must consider the model which gives us the least error which is LSTM as concluded in this research paper.

7.2 Future Work

In this research paper the prediction model of Litecoin is developed with reference to only one variable that is by considering the closing price for the day. Therefore, the research is limited in certain aspects. Improving on these aspects will provide the opportunity for future research work which will enable the predictions to be more accurate. Firstly, the premise of the prediction model which considered only one factor of closing price. In the real world, prices are impacted by a variety of factors like market sentiments and governmental policies. Secondly since data is live streamed from the markets, price fluctuations within the day can be mapped on hourly basis.
Study on price prediction considering opening, intra-day high, intra-day low and closing prices can form the basis of future research work.

Further this research paper is focused on two models ARIMA and LSTM Price prediction can be explored using other machine learning models. Deep learning models require large quantity of data for better results. Hence a future study can also be done with datasets exceeding the 1991 observations which is used in this research paper.

References

Greaves, A. and Au, B. (2015). Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin.

Madan, I.,Saluja, S. and Zhao, A. (2015). Automated Bitcoin Trading via Machine Learning Algorithms.

Shah, D. and Zhang, K. (2014). Bayesian regression and Bitcoin. In: 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton).

Mallqui, D. and Fernandes, R. (2018). Predicting the Direction, Maximum, Minimum and Closing Prices of Daily Bitcoin Exchange Rate Using Machine Learning Techniques. [online] Available at: https://doi.org/10.1016/j.asoc.2018.11.038

McNally, S., Roche, J. and Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. In: 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing.

Phaladisailoed, T. and Numnonda, T. (2018). Machine Learning Models Comparison for Bitcoin Price Prediction. In: 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE).

Alessandretti, L., ElBahrawy, A., Aiello, L. and Baronchelli, A. (2018). Anticipating Cryptocurrency Prices Using Machine Learning.

KARASU, S., ALTAN, A., SARAÇ, Z. and HACIOĞLU, R. (2018). Prediction of Bitcoin Prices with Machine Learning Methods using Time Series Data.

Corelli, A. (2018). Cryptocurrencies and Exchange Rates: A Relationship and Causality Analysis. Risks, 6(4), p.111.

Sin, E. and Wang, L. (2017). Bitcoin Price Prediction Using Ensembles of Neural Networks. In: 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017.

Roy, S., Nanjiba, S. and Chakrabarty, A. (2018). Bitcoin Price Forecasting Using Time Series Analysis. In: 2018 21st International Conference of Computer and Information Technology (ICCIT).

Abraham, J., Higdon, D., Nelson, J. and Ibarra, J. (2018). Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis.

Saad, M. and Mohaisen, A. (2018). Towards Characterizing Blockchain-based Cryptocurrencies for Highly Accurate Predictions. In: Hot Topics in P2. pervasive Mobile and Online Social Networking.

J R, P. and S Das, V. (2018). Cryptocurrency Price Prediction Using Long-Short Term Memory Model. In: International Journal of Research and Scientific Innovation (IJRSI) | Volume V, Issue VII, July 2018 | ISSN 2321–2705.

Wu, C., Ma, Y., Lu,, C. and Lu, R. (2018). A New Forecasting Framework for Bitcoin Price with LSTM. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW). [online] Singapore: 2018 IEEE International Conference on Data Mining Workshops (ICDMW

Tamuly, B. (2019). Litecoin [LTC] Price Analysis: Volatility awaited as bears take control of silver coin. [online] AMBCrypto.

Catania, L., Grassi, S. and Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. International Journal of Forecasting, 35(2), pp.485-501.

Garg, S. (2018). Autoregressive Integrated Moving Average Model based Prediction of Bitcoin Close Price. IEEE.

Rane, P. and Dhage, S. (2019). Systematic Erudition of Bitcoin Price Prediction using Machine Learning Techniques. In: 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE.

Independent.ie. (2019). Bitcoin stumps analysts as currency falls to more than one-year low - Independent.ie.

CoinMarketCap. (2019). Litecoin (LTC) price, charts, market cap, and other metrics | CoinMarketCap.

Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015) Time Series Analysis: Forecasting and Control. John Wiley & Sons, Hoboken.

Sun, J., Zhou, Y. and Lin, J. (2019). Using machine learning for cryptocurrency trading.

Matta, M., Lunesu, I. and Marchesi, M. (2019). Bitcoin Spread Prediction Using Social And Web Search Media.

S. Vassiliadis, P. Papadopoulos, M. Rangoussi, T. Konieczny, J. Gralewski, Bitcoin Value Analysis Based On Cross-Correlations, Journal of Internet Banking and Commerce 22 (S7) (2017) 1.

Y. Peng, P. H. M. Albuquerque, J. M. C. de S´a, A. J. A. Padula, M. R. Montenegro, The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression, Expert Systems with Applications 97 (1) (2018) 177–192

M. Balcilar, E. Bouri, R. Gupta, D. Roubaud, Can volume predict bitcoin returns and volatility? A quantiles-based approach, Economic Modelling 64 (2017) 74–81

RStudio. (2019). R Packages - RStudio. [online] Available at: https://www.rstudio.com/products/rpackages/ [Accessed 10 Aug. 2019].

Medium. (2019). Time series Forecasting — ARIMA models. [online] Available at: https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06 [Accessed 10 Aug. 2019].

Smart Vision - Europe. (2019). What is the CRISP-DM methodology [online] Available at: https://www.sv-europe.com/crisp-dm-methodology/ [Accessed 10 Aug. 2019]?

Brownlee, J. (2019). How to Develop LSTM Models for Time Series Forecasting. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting.

Statista. (2019). Litecoin cash price monthly 2016-2019 | Statista. [online] Available at: https://www.statista.com/statistics/807160/litecoin-price-monthly/