Time-Series forecasting of bitcoin prices using machine learning and deep learning techniques

Tarang gupta

Dissertation Report

­

August 2021

**ABSTRACT**

This research consists of bitcoin price predictions using different AI techniques to maximise the profits. Cryptocurrency investments are increasing rapidly as people are getting aware more about its sudden fluctuations and potential of giving the profits. Although Bitcoin is not the only cryptocurrency today which is attracting the attention of everyone but this research is limited to the prediction of bitcoin price only but the same techniques applied in this research can be applied to any other cryptocurrency price prediction as well. In this research a good correlation of about 86% has been found between the price of bitcoin and Ethereum. This research is done with sole purpose of maximizing the profits as cryptocurrencies tend to give much larger amount of profits within short span of time compared to any other investment strategies like stock market investment, mutual funds etc. To achieve this task, Statistical approaches like Simple Moving Averages are used to understand the predictability and followed by Additive Regression models using Fbprophet and also Deep Learning techniques such as LSTM has been implemented. Among all of the applied techniques we found Recurrent Neural Network like LSTM works very well and achieved the root mean squared error of 784.

LIST OF TABLES

1. Table 1: Performance (MSE) of different ML algorithms by Maleki ...…........….20
2. Table 2: Performance of different ML algorithms by Reddy and Sriramya ……..20
3. Table 3: Performance Metrics of different DL techniques by Hashish ………….24
4. Table 4: Performance Metrics of different techniques applied in this research…..43

LIST OF FIGURES

1. Figure 1: First Blockchain Block……………………………………………........8
2. Figure 2: Data Features from different data sources……….………………….....17
3. Figure 3: Data Processing workflow………………….………………………….18
4. Figure 4: Prediction trend using ARIMA …...……………………………….......19
5. Figure 5: Word Embedding …………………………………………………........21
6. Figure 6: Stacked Artificial Neural Network……………………………………..23
7. Figure 7: Proposed method using HMM and LSTM……………………………..23
8. Figure 8: Forecasting Process using Fbprophet …….……………………………29
9. Figure 9: Long Short Term Memory (LSTM) ……………………………………31
10. Figure 10: Correlation Heatmap of collected dataset ……………………………32
11. Figure 11: Trends and Patterns in collected dataset………………………………33
12. Figure 12: Highest value of BTC price yearly ……………………………………34
13. Figure 13: Highest value of BTC price monthly for each year …………………..35
14. Figure 14: Correlation Heatmap of Altcoins and Bitcoin Price ………………….36
15. Figure 15: Moving Average for each month ……………………………………..36
16. Figure 16: Simple Moving Average when N=10 …………………………………37
17. Figure 17: Simple Moving Average when N=30 …………………………………37
18. Figure 18: FBProphet Forecast …………………………………………………..38
19. Figure 19: Retrained FBProphet Forecast ………………………………………..39
20. Figure 20: Forecast Included Unseen 60 Future Dates …………………………..39
21. Figure 21: Prepared Data for LSTM ……………………………………………..40
22. Figure 22: Summary of LSTM Neural Network ………………………………....40
23. Figure 23: Summary of LSTM Neural Network 2 ……………………………….41
24. Figure 24: Evaluation Metrics for Regression ……………………………………42

TABLE OF CONTENTS

Chapter 1: Introduction 7

* 1. Background of the study 10
  2. Aims & Objectives 11
  3. Significance of the study 12
  4. Scope of the study 12
  5. Structure of the study 12

Chapter 2: Literature Review ……………………………………………………………………....14

2.1 Introduction ………………………………………………………………….14

2.2 Data used for Bitcoin Price Predictions ……………………………………..15

2.3 Modelling Techniques for Bitcoin Price Predictions ….…………………….18

2.4 Summary.…..…………………………………………………………….......24

Chapter 3: Research Methodology ………………………………………………………......25

3.1 Introduction………………………………………………………………………...25

3.2 Data Collection and Understanding………………………………………….25

3.3 Data Pre-processing…………………………………………………………..26

3.4 Exploratory Data Analysis …………………………………………………..26

3.5 Data Selection for Modelling ………………………………………………..27

3.6 Modelling Techniques ………………………………………………………27

3.7 Summary ……………………………………………………………………31

Chapter 4: Experimentation & Analysis………………………………………………..........32

4.1 Introduction…………………………………………………………………..32

4.2 Data Analysis……………………..………………………………………….32

4.3 Experimentations.…………………………………………………………….36

4.4 Summary …………………………………………………………………….41

Chapter 5: Results & Discussions.……………………………………………………..........42

5.1 Introduction………………………………………………………………….42

5.2 Results Obtained ……………………………………………………………42

5.3 Summary…………………………………………………………………….43

Chapter 6: Conclusion……………………………………………………………................44

6.1 Conclusion…………………………………………………….…………….44

6.2 Recommendations ……..……………………………………………………44

REFRENCES

APPENDIX A: Research Plan

APPENDIX B: Research Proposal

CHAPTER 1

INTRODUCTION

Bitcoin is not a coin like our traditional coins which can be seen and touched rather it is a digital coin which resides on the internet only. Bitcoin is the first decentralized cryptocurrency invented and implemented in 2009 by Mr. Satoshi Nakamoto. The idea to develop the Bitcoins is to give the Financial Independence to Individual So he can send money from one place to another without any Bank’s or any other Government authority’s intervention. For which it uses the wonderful technology named Blockchain which is now being used other than developing the Cryptocurrencies such as in Hospital Management Systems, Airlines Management Systems etc. With Blockchain the Security in Finance became more reliable. To understand in brief, you can make fool of one person or may be 5-10 people but you won’t be able to make fool a group of thousands of people. Yes, that’s how Blockchain technology works at its core. It is a chain of Blocks and each block is associated with a hash key of the previous blocks which is distributed among the nodes in the Network of Blockchain. Bitcoin is the first application which follows the principles of Blockchain technology. In 2009 the price of Bitcoin was almost zero. There were no exchanges, there was no market. Only the fans of this new cryptocurrency were sending the bitcoins just for fun reward to each other. In 2017 Bitcoin had major jump in its price. Which increased to $9925 and made the record and in 2020-2021 Bitcoin broke all the limits and crossed $50,000. Some of the Countries have banned transactions in Bitcoin or any other cryptocurrencies such as Ethereum, Litecoin, Bitcoin Cash, Litecoin, Ripple and the current trending one Dogecoin. And some of the countries have imposed strict regulations for Cryptocurrency world as it affects the country’s financial stability by blocking the cash flow mechanism and promotes cyber-crimes as what they say. Yes, there is no doubt that in cyber world for all the unethical stuff happen there would be Crypto transaction involved in it either on Dark web, Deep web or usual Ransomware attacks ex. Wnnacry Ransomware in 2017. (Demir et al., 2019) says Bitcoin is the first cryptocurrency which uses Block chain technology at its core and used worldwide for digitalx payment and for investment purposes. Bitcoinx is decentralized which means there is no control of government or bank or any other authority on Bitcoin. Bitcoin is independent and its price depends upon its demand and supply. According to (Gartner, 2021) Bitcoin is a cryptocurrency, based on a 2008 white paper by the pseudonymous Satoshi Nakomoto. Bitcoin is a peer to peer version of electronic cash that allows online payments to be sent in real time from party to counterparty without a financial institution as intermediary and without a central bank as a clearing house. The system of bitcoin payments leverages a cryptographic algorithm known as ‘proof of work’, and distributed ledger technology. Transactionss made by Bitcoins are easy and not limited to any specific country or region which is the case in traditional banking system. To send out some amount of money overseas using our traditional system it would take more than 2-3 days or may be a week. During this time the bank or some higher authority actually validate the transaction. This process has completely abandoned and replaced by fast and secure implementation of Block chain when it comes to Cryptocurrency transactions and has capability of getting the overseas transactions within seconds. Investment can be done through various marketplacesx known as bitcoin exchanges such as Binance, Coinbase, CoinDCX, WazirX etc. These Bitcoin Exchange wallets allow the investor to sell/buy Bitcoins using different currencies like INR, USD etc. The most popular and largest Bitcoin exchange wallet is Mt Gox which was launched in July 2010 and by 2014 it was handling over 70% of all bitcoin transactions worldwide. Bitcoins are stored in these exchange digital wallets just like a virtual bank account. The data of all the transactions at each timestamp is stored in the Block chain Network which can be viewed on (Blockchain, 2019). Each record in a blockchain is known as block. Each block contains address of its own block and address of previous block known as hash. The data of transactions in blockchain network is encrypted and During transactions the name of user is not revealed which gives the user the flexibility of anonymity but only their wallet ID is made public. Below is the information of first ever block of blockchain network extracted from (*Block: 1 | Blockchain Explorer*, no date)

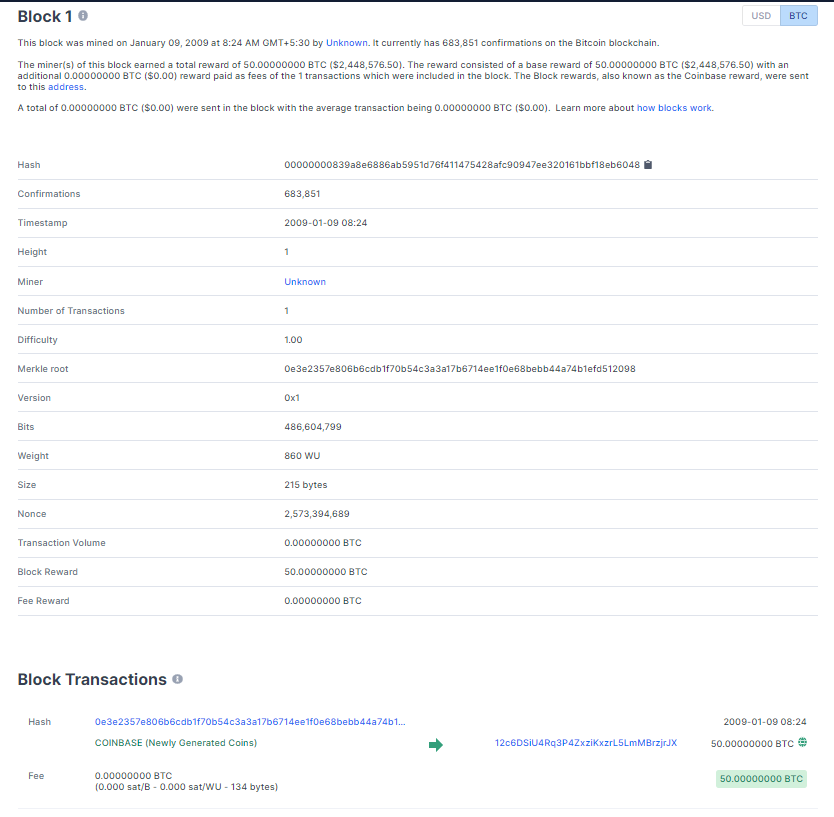


Figure 1: First Blockchain Block

According to (Raju and Mohammad, 2020) says Bitcoin is an online payment system which can be used globally in highly stabilized and secured manner. It is Peerx - to - Peer value transfer protocol. Bitcoin transactions gets verified by the nodes in the block chain network and gets published on a public ledger. The Smallest unit of BTC known as Satoshi represents 0.00000001 bitcoin which is One hundred -millionths of a xbitcoin. So we got the good understanding about what Bitcoin (BTC) is but how do we earn them? (Gupta, 2017) says The earning methods of Bitcoin includes: Bitcoin can be purchased just by paying the current amount of Bitcoinx and get that in your crypto exchange wallet. If One doesn’t want to invest much then there is no need of buying a whole Bitcoin One can buy the fraction of Bitcoin which is known as Satoshi. The Bitcoin will be safe in your Crypto Exchange Wallet (You can use Coinbase wallet or Binance wallet) and when the value of Bitcoin again fluctuates analyse when you feel yourself in Profit you can sell that Bitcoin. Another way of earning the BTC is through mining: - The way of creating the Bitcoin or it’s fraction Satoshi is known as Bitcoin Mining. There is a limitation in number of Bitcoins in the real world i.e. 21 million which implies that the number of Bitcoins can never be more than 21 million in the Market. Due to this rule there will be continuous increase we will see in BTC price. Just Like in Normal Transactions when a sender sendss money to Receiver there is always an involvement of Banks and some other government authorities to verify or to complete the Transactions which may take days to happen but in Bitcoin Transactions there is no one but a miner will be there. Every transaction first gets distributed in the Blockchain Network in the encryptedx form where several distributed computing machines or (nodes of network) tries to validate and verify that transaction and one of them at the end finishes the verification which acts as a Bitcoin miner. Each Transaction contains some Math Problemss and these math problems are actually the decryption algorithms of the encrypted transaction. Now to verify or to decrypt the algorithm of the Transactionn the miner needs powerful CPUs and GPUs or ASIC (Application Specific Integrated Circuits) to solve the Math Problems. Resultant the miner will be awarded with some Bitcoins or the fraction of Bitcoin i.e. Satoshi to verify the transaction. That’s how a bitcoin is created and new bitcoins make its existence in the real world. This entire process of creating new bitcoins is called Bitcoin Mining. Supply and Demand is not the only factors which affects the price of Bitcoins there are few more factors like rewards issues to bitcoin miners, the number of completing cryptocurrency, the exchange platform where it trades on etc.

**1.1 Background of the study**

Investing in stocks in order to maximize the profit is not new to us. People have been investing in stock market to get the proportional profit of their invested amount in a particular stock. Same is the case for Cryptocurrencies as well which may give more than 500% of return in just few days. Take example of Dogecoin which went from $0.006 to $0.4498 in just about 2 months. What if one already knew about the future predictions? Imagine the possibilities of knowing what will happen in the future! Imagine what it would be if you had invested in Dogecoin in Jan 2021 at the price of $0.006. Many people regret for not investing in Bitcoins back in 2016 or 2017 but how would they even know about when the Bitcoin price going to rise or for any other cryptocurrency. Anyways not to regret any more we can capture the patterns or trends of rising prices of Bitcoins using Machine Learning. Machine Learning (ARIMA) or Deep Learning (LSTMs/RNNs) will not tell us the future but it might help in capturing the trend or pattern of the Bitcoin price movement. However, the parameters which actually affect the prices of Bitcoin is not just the previous trend. Bitcoin price fluctuates on the basis of **demand** and **supply** but a rare occurrence contributes the most in fluctuation like getting the temporary ban of Bitcoin Trading in any country or temporary ban in Bitcoin mining, Since Bitcoin Mining is highly proportional to the Electricity consumption as mining process takes too much of time and computational power to validate the transaction and make new bitcoin or satoshi (fraction of bitcoin) existence in the real world. Which leads to significant drop in the Price of Bitcoin. In the same way, Acceptance of Bitcoin as the payment option in buying Tesla made a positive impact to the Bitcoin prices, Ban Removals from country also tend to make the positive impacts and so the Bitcoin Prices jumps to its peak. Recent advancements in computer GPUs and their affordability, awareness of Crypto world to newbies also made a significant impact on Bitcoin prices. A single tweet by Elon Musk made Dogecoin popular on the internet. Thus more demand and trust in Dogecoin and so the prices of Dogecoin rose. There has been recent research to predict the Bitcoin price using the thoughts of individuals which they post on Tweeter, Reddit, LinkedIn and also the user search insights via Google Trends. According to (Sin and Wang, 2017) As the markets is moderately new, existing works identified with determining in this market is genuinely restricted. One examination showed that Google Trends information and volume of tweets identified with Bitcoin on Twitter have positive relationship with Bitcoin's cost and subsequently might have the option to anticipate the vacillations in cost of Bitcoin. In another investigation, Bayesian Regression, xBinary classification algorithm was utilized to anticipate value variety in Bitcoin and the expectation gave practically 200%s returns in under 60 days when utilized with a trading technique. The examination reasoned that there might be 'information' in Bitcoin's chronicled data that can help anticipate future value varieties but reading the thoughts won’t help much in predictions of Bitcoin as People tend to discuss more than actually taking the action on it. Indeed, when positive thoughts are posted by some popular personality like Elon Musk or famous Influencer which tend to change the mind of thousands or lacks of people, definitely there will be increase in Demand and so the Bitcoin price will rise. According to (Mudassir *et al.*, 2020) At the point when Bitcoin started to stand out enough to be noticed at end of 2013, it saw a critical vacillation in its cost. A strands of literature has inspected the consistency of Bitcoin returns through different boundaries, for example, socially based media consideration and Bitcoin-related authentic specialized pointers. One gathering considered the period from September 4, 2014, to August 31, 2018, by catching the occasions the term Bitcoin has been tweeted. The outcomes showed that the quantity of tweets on Twitter can impact Bitcoin exchanging volume for the next day. Besides, examined the impact of clients remarks in online stages on value variances and number of exchange of digital currencies and found that Bitcoin is especially corresponded with the quantity of positive remarks via web-based media. They detailed the accuracy of 79% alongside Granger causality test, which infers that user conclusions are valuable to anticipate the value changes. With regards to time series forecasting, there are three distinct sorts of statistical model based methodologies for time series forecasting. The primary methodology, pure models, which just uses the chronicled information on the variable to be predicted. Instances of pure time series forecasting models are Autoregressive Integrated Moving Average (ARIMA) and Generalized AutoRegressive Conditional Heteroskedasticity (GARCH).

**1.2 Aim and Objective**

The Primaryx aim and objective of this research is to develop a trustworthy and interpretable predictive time series model which can capture the past trends or patterns in the Price of Bitcoin So one can make better decisions for future investment in Cryptocurrency (Bitcoin) to maximize the ROI.

The research objectives are formed dependent on the point of this examination which are as per the following:

1. Understanding the time series data of Bitcoin price.
2. Finding relation of Bitcoin price with Altcoin’s price such as Ether, Ripple etc.
3. Developing a robust machine learning /deep learning model for the successful prediction of bitcoin future price that could yield significant profit.
4. Evaluating the predictive models to identify most profitable model with least errors.
5. Comparative Analysis on Machine Learning and Deep Learning Techniques.

**1.3 Significance of the Study**

Everyone wants to have the passive source of income. Either it is stock market or Affiliate Marketing or Rental Income. We have been seeing these examples but the most underrated example of one of the passive sources of income is trading in Cryptocurrencies. This is something which is very well known by youngsters (not limited to). It becomes vital to have good understanding of Crypto World to get started in Bitcoin investment just like in Stocks. Predictions of Bitcoin price will help individual to gain maximum profit and save them in losing the big amount of money. This research will add to decipher or clarify the forecast produced by different prescient AI or profound learning models to help the choice in crypto speculation.

**1.4 Scope of the Study**

* The data for research is taken directly from Kaggle.
* The Research is limited for the people who look for passive source of income by investing in Cryptocurrency.
* The end model in this research is suitable for Value Investors and not for Traders as we have date collected on daily basis and not hourly basis.
* The Techniques/ Algorithms going to be used for the Research will be Time Series analysis and Deep Learning but we will try other traditional machine learning algorithms as well such as SVM, Random Forest Repressors as in recent work they have performed well.

**1.5 Structure of the Study**

In this research, the structure of the study is formed in multiple section. First Section has more of the research problem and aim and objective of the research, Second Section is about the recent studies which are relevant to solve our business problem. Where the results obtained by researchers will be demonstrated through which the understanding of relevant techniques can be filtered out. Third section will be formed basis on the succession of the second section where the most relevant and other experimentation will be finalised to get implemented in the fourth section. Which is followed by fifth chapter that includes the results obtained with the experimentations done in fourth chapter and later concluded in sixth chapter.

CHAPTER 2

LITERATURE REVIEW

**2.1 Introduction**

Cryptocurrency is still a new technology and Bitcoin is the world’s most expensive cryptocurrency as of now in 2021. Several researches have been done to understand ups and downs in the price of Cryptocurrencies. Capturing price and understanding its patterns makes it Time Series problem and So the Bitcoin price prediction problem is, Researchers have applied Time Series Regression techniques such as ARIMA, SARIMA etc. Researchers have also gathered data of Cryptocurrency’s user’s thoughts from Social Media such as Tweeter, Reddit, Pinterest etc and performed Sentiment Analysis on that. Deep Learning techniques never stay behind. Deep Learning has very good predictable power when it comes to capturing the Sequential data like Time Series with the help of Recurrent Neural Networks. RNNs itself has its variants which resolved each other’s problems like Gradient Vanishing or huge amount of parameters issues. Researchers have used LSTM (Long Term-Short Term Memory) Recurrent Neural Networks and predicted the pricing of Bitcoin with very minimal errors. (Felizardo *et al.*, 2019) compares several popular machine learning algorithms adopted in bitcoin price prediction tasks such as Back propagation Neural Network (BPNN), Autoregressive Integrated Moving Average mode (ARIMA), Random Forest (RF), Support Vector Machine (SVM), Long Short-Term Memory (LSTM) and WaveNets. Few Researchers have gathered the bitcoin price real time data with the Interval of one minute which can be very useful for Daily Traders to maximise their profits. (Hashish *et al.*, 2019) implemented a hybrid model using Hidden Markov Models to deal with the volatility of cryptocurrencies and predict the future fluctuations with LSTM. Other researches also proposed hybrid methods between traditional machine learning and ARIMA to predict the bitcoin price. (McNally, Roche and Caton, 2018) says Research on predicting the price of Bitcoin using machinex learning algorithms specifically is lacking. Implementation of a latent source model to predict the price of Bitcoin noting 89% return in 50 days with a Sharpe ratio of 4.1. There has also been work using text data from social media platformss and other sources to predict Bitcoin prices. Researchers investigated the strong relationship between the price of Bitcoin and Social Media posts, tweets and the views for bitcoin on Google Trendss and implemented a methodology to predict not the price of bitcoin but the trading volume. Traditional machine learning techniques such as SVM (support vector machines) have also been used to predict the bitcoin price along with ANNs (Artificial Neural Networks) which reported price direction accuracy of 55%. Other than SVMs Researchers have also implemented models like Random Forest, Binomial GLM (Generalised Linear Models) However, these models were limiting the generalisability of the results. (McNally, Roche and Caton, 2018) says Traditional time series prediction methods such as Holt Winters exponential smoothing models heavily rely on linear assumptions and requires data that can be broken down into seasonality, trends and noise. This kind of methodology is much suitable for the tasks like sales forecasting where seasonality effects actually exist in the data. Due to the high volatility in the Bitcoin market data and lack of seasonality these methods are not very effective for this task. Given the complexity of the task, deep learning makes for an interesting technological solution based on its performance in similar areas. According to (Li and Wu, 2019) Identifying the determinants of bitcoin price has always been a key issue for predicting or forecasting the bitcoin price. Most of the studies by researchers found a very good correlation between Bitcoin and queries searched on Google Trends and Wikipedia. Indeed, there is a good correlation but it is hard to say if there is causation exist too. Other than analysing correlation of Wikipedia and Google search queries, Researchers have also evaluated the influence of public opinions and social media posts. According to (Balfagih and Keselj, 2019) After exploring the relationship between tweets related to bitcoin and the prices of bitcoin utilizing different-2 language modelling techniques such as word embedding of tweeter data and N-Gram modelling, discovered that the bitcoin price’s fluctuations are primarily driven by the popularity of Bitcoin Thus, there is a positive causation between the people thoughts or opinions and the price of bitcoin. (Mittal *et al.*, 2019) argues that Indeed, There is a very good relationship between Google Trends insights, Tweet volume data with the price of Bitcoin while no significant correlation with the sentiments of tweets is discovered. Few researchers also found that the sentiment ratio of twitter is positively correlated with bitcoin prices, while the value of Bitcoin is negatively affected by the exchange rate between the USD and the euro has negative relation with bitcoin price.

**2.2 Data used for Bitcoin Price Predictions**

There are multiple sources to collect historical data for Cryptocurrency prices. (Demir *et al.*, 2019) collected data from databases named Quandl and CoinmarketCap which is captured on daily basis for last five years with different time intervals available. CoinMarketCap keeps a track of all the cryptocurrencies available in the market. They keep a record of all the transactions by recording the amount of coins in circulation and the volume of coins traded in the last 24-hours. Quandl holds databases related to financial, economic, and social background from over 500 publishers. Data available on Quandl can be used on different platforms such as Python, MATLAB, Maple and Strata. The researcher was able to procure datasets for Bitcoin for up to 5 years of timestamp data with specifications such as –Data high, Data low, Open, Close, volume of transaction, weighted price. Before applying any machine learning technique, it is recommended to normalise the data. The author normalised the bitcoin data with multiple features using multiple techniques like listed below.

* Log Transformation: In this method, the range of our data gets compressed and makes the value near to zero.

(2.1)

* MATLAB predefined method ‘normc’: normc() function is heavily used in MATLAB to normalize database columns by compressing the range to the best possible extent as compared to other methods.
* Standard Deviation Normalisation: Here, we take into consideration the difference of every value with respect to the mean value. The advantage of this technique is that we get the negative values as well due to proper compression of the Y axis. The formula is given below. This method is also known as Z-score normalisation.

(2.2)

(Raju and Mohammad, 2020) has collected data in two phases. In first phase researcher collected the real time bitcoin price data through an API and in second phase they collected the data from tweeter or reddit to perform sentiment analysis. These datasets are collected in real time with the help of APIs. The author has collected the data from multiple sources like Coinmarketcap, Bitstamp, Coinbase, and Blockchain Info. These data sources provide publicly available API to collect the data in real time. The dataset from these data sources is gathered an interval length of every minute. Figure 2 is the image showing data features collected by author.

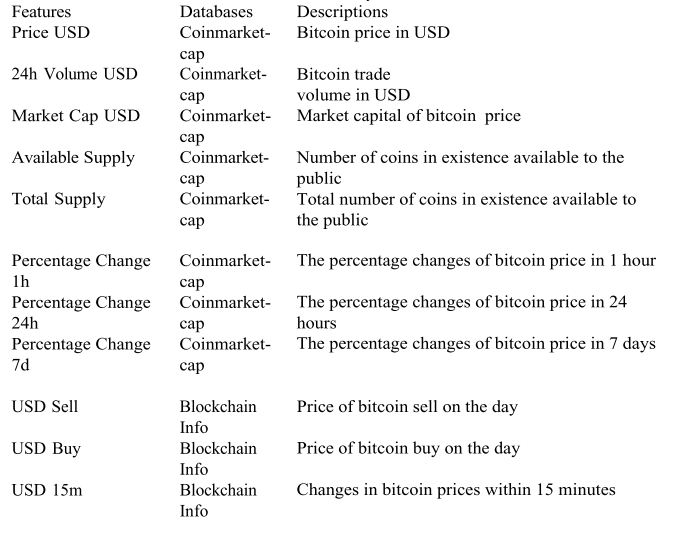


Figure 2: Data Features from different data sources

Researchers have performed Sentiment Analysis on Social Media data of People opinions and actions shared on internet. For this the author collected data from tweeter using API provided by Tweeter named tweepy. After scraping the tweets posted by users, pre-processing the textual data is a really big challenge as text/tweets consist of emotions, acronyms etc. To pre-process the tweets, the author has applied techniques like Tokenization, stop-words removal, Regex matching for removing special characters etc. Post that to identify the sentiment of the tweet, pre-trained models are being used which are available to use directly in a form of python library such as textblob, and classified the tweets as Positive, Negative or Neutral. Later the author has combined the data collected in both the phases and created a new dataframe which contains only the closing price, time features and the sentiment of each tweet. The flow chart of entire process followed by (Raju and Mohammad, 2020) is shown in Figure 3.

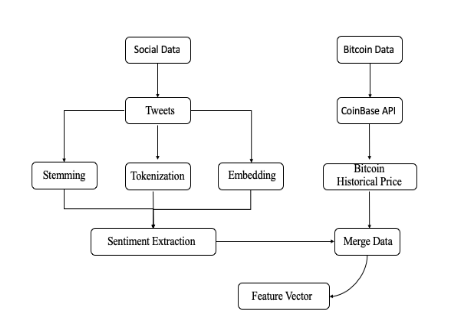


Figure 3: Data Processing workflow

**2.3 Modelling Techniques for Bitcoin Price Predictions**

1. Traditional Statistical Methods:

There are various traditional techniques to solve any Time Series problem such as Autoregressors (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA) which combines or integrates the features of AR and MA. (Maleki *et al.*, no date) Autoregressive is utilized to characterize time-fluctuating systems in nature, monetary viewpoints, and so forth In AI, an AR model increases from a progression of planned advances and acknowledges assessments from past exercises as inputs for a regression model to anticipate whenever step's worth. The forecasts for autoregressive models are regressed on past observations in time series and these type of models use weighted sums of past values to predict future values. Moving Average (MA) is a univariate time series methodology which determines that the upshot variable depends directly on the present and different past values of a stochastic (imperfectly forecastable) term. ARIMA model is integration of an autoregressive and moving average models. Both of the models are fitted to time series data either to all the more likely comprehend the information or to predict future data points in the series. The AR part of ARIMA determines that the interest's developing variable is regressed on its own slacked values. The MA part denotes that the regression fault is a linear compound of fault terms whose values coincided at various times in the past. The "I" shows that the data values have been altered with the diversity between their values and the olden values. Author has applied ARIMA to multiple cryptocurrencies price to predict bitcoin price. These cryptocurrencies which are treated as feature for Bitcoin price prediction includes Ethereum, Litecoin, Zcash. After trying different parameter values experimentation of ARIMA model the best parameter values the author found as p=1 and q=1 with one difference. The same experimentation has been applied to each cryptocurrency and best parameter for each are ARIMA(2, 1, 1), ARIMA(2, 1, 1), and

ARIMA(1, 1, 1) for Ethereum, Litecoin, and Zcash, respectively. (Raju and Mohammad, 2020) implemented ARIMA and LSTM and found that ARIMA model’s RMSE is 209.263 and also found best parameters as (10,1,0) and prediction trend can be seen in below figure.

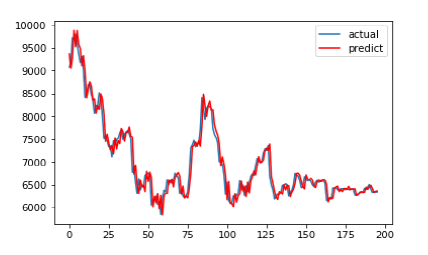


Figure 4: Prediction trend using ARIMA

1. Machine Learning Methods:

Machine Learning algorithms like Linear Regression, Logistic, Support Vector Machines or any tree based technique are not capable of capturing the trend or pattern hidden in sequential data but they are well versed in capturing the relationships or distribution of the data and make wonderful predictions. (Maleki *et al.*, no date) utilized machine learning techniques to analyse which cryptocurrency or Altcoins affects the price of bitcoin the most. The author has implemented multiple regression models on the dataset of Ethereum, Zcash, Litecoin which he used as independent features and the price of bitcoin as dependent variable. These regression models include Linear Regression, Gradient Boosting Regressor, Support Vector Regressor, Lasso Regressor, Random Forest Regressors etc and compared the MSE for each model. Below is the table of MSE obtained for each machine learning model applied by the author and the best results the author has obtained is with Gradient Boosting Algorithm having minimum mean square error (MSE).

Table 1: Performance (MSE) of different ML algorithms by Maleki

|  |  |
| --- | --- |
| **Machine Learning Algorithm** | **MSE** |
| Linear Regression | 421034.06468415906 |
| Support Vector Regressor | 516804.797833455 |
| Random Forest Regressor | 23365.60049452004 |
| Gradient Boosting Regressor | 18359.190867996942 |

(Reddy and Sriramya, 2020) gathered the bitcoin price data from CoinMarketCap and applied pre-processing techniques and different normalisation techniques like Log Transformation, Standardization etc and implemented machine learning techniques such as k-Nearest Neighbors(KNN), Ridge Regression, Polynomial Regression, Linear Regression, Random Forest Regressor and concluded that Linear Regression is the one which gave him most accurate results efficiently. Below is the table of results the author has achieved.

Table 2: Performance of different ML algorithms by Reddy and Sriramya

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model Test | k-Nearest Neighbors(KNN) | Ridge Regression | Polynomial Regression | Linear Regression | Random Forest |
| Training RSS(Residual Sum of Squares) | 92.99% | 75.7% | 84.2% | 94.6% | 89.78% |
| Test(Residual Sum of Squares) | 90.5% | 12.4% | 79.99% | 96.99% | 71.56% |

Similarly, (Rathan, Sai and Manikanta, 2019) extracted the data from Quandl and applied algorithms like decision tree, linear regression etc and found that Linear Regression outperforms Decision tree regressors with 97% accuracy whereas Decision tree achieve accuracy of 95%. (Felizardo *et al.*, 2019) implemented ARIMA, Machine learning techniques like Random Forest Regressor, Support Vectors and also deep learning techniques like LSTM, Wavenets and compared all of them with different-2 hyper parameters combinations and used 30 cross validation for each model. The author suggests to play more with the hyper parameters of neural network.

1. Sentiment Analysis:

Many researchers have performed sentiment analysis on people’s opinions and thoughts posted on social media and found good co-relation with the price of bitcoin. (Raju and Mohammad, 2020) collected tweeter’s tweet data using Twitter’s streaming API and tweepy. Tweepy allows filtering the data using hashtags or key words, pre-processed the tweets by applying Tokenization, Stop words removal, Regex matching special characters removal and then used textblob pre-trained python library to tag each tweet either positive, negative or neutral. (Balfagih and Keselj, 2019) applied sentiment analysis on tweets by collecting the tweeter data and prepared the textual data using popular language modelling techniques like Tweet Embedding, N-Gram modelling and formed 30 new datasets to be tested in the data mining process then passed all these to traditional machine learning classification methods such as Decision trees, Random Forest, Naïve Bayes, K-Nearest Neighbour and deep learning methods such as Weightless Neural Network, Radial Basis Function Network to figure out the best form of modelled features. Below is the graphical representation of word embedding which the author has used as part of process of preparing textual data of tweets.

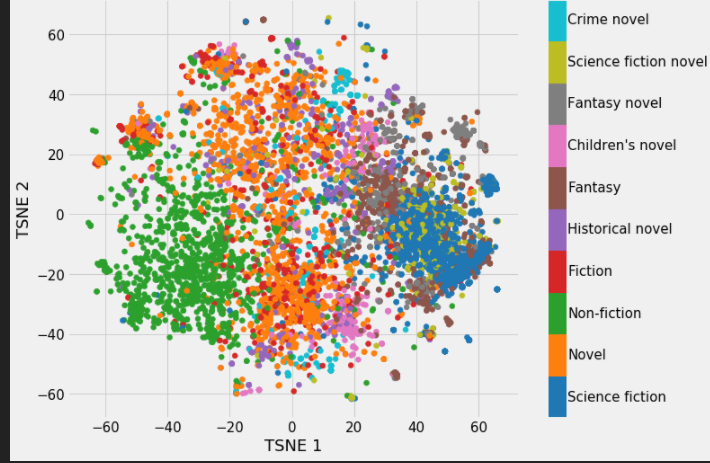


Figure 5: Word Embedding

Many studies found very good relationship of tweeter data, reddit posts, google search data with the price of bitcoin. In another study it is discovered that the bitcoin price’s fluctuations are primarily driven by the popularity of Bitcoin Thus, there is a positive causation between the people thoughts or opinions and the price of bitcoin. According to (Li *et al.*, 2019) very large number of researches have been conducted on sentiment analysis of tweets or other social media platform’s posts for the predictions of price. The author says unlike the traditional stock markets, the Bitcoin market lacks a mature financial regulatory regime and supervisory framework this is why it is easily affected by online market sentiment and which many researchers have already investigated and found very good correlation between the sentiments of social media and the price of bitcoin.

1. Deep Learning Techniques:

The abundant data and deep learning techniques is proven to be the best combination to solve many AI related problems either it be a NLP problem or Computer Vision problem and this is extended by one of the most popular and complex deep learning technique known as Recurrent Neural Networks (RNNs) to solve any Sequential problem. RNNs has its own variants namely Long Short Term Memory (LSTMs) and Gated Recurrent Units (GRUs), Researchers have solved many NLP and CV related sequential problems using RNNs and have also applied it to predict the stock market prices, and can be applied to predict the cryptocurrencies price as well. (Raju and Mohammad, 2020) achieved 198.448 RMSE using LSTMs with the learning rate of 0.01 which outperforms ARIMA model with 209.263 RMSE. (McNally, Roche and Caton, 2018) also implemented LSTM with RMSprop optimizer and found that the model training is converged between 50 to 100 epochs with early stopping, and compared it with Traditional Simple RNNs and ARIMA model, found that no model could effectively learn from the data but recommends RNN and LSTM are more effective for Bitcoin price predictions with LSTM more capable for recognising longer dependencies. (Mudassir *et al.*, 2020) implemented Artificial Neural Network (ANN) and found satisfactory results with stochastic gradient based optimizer Adam, learning rate as 0.08 and with the Relu activation function then stacked multiple such ANNs. The stacked ANNs trained using the training split of five cross validations and each model trained on separate fold. Figure 6 illustrates the stacked ANN technique implemented by the author.

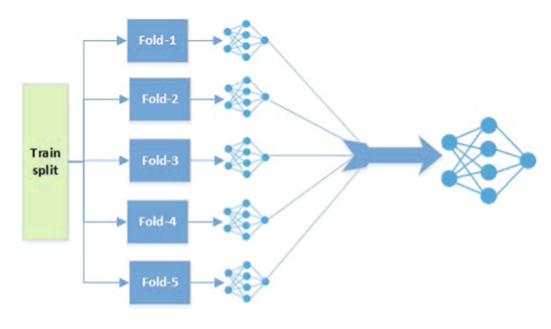


Figure 6: Stacked Artificial Neural Network

(Hashish *et al.*, 2019) collected 2-minutes frequency bitcoin data from Coinbase exchange wallet and proposed a hybrid model of Hidden Markov Model (HMM) and Long Short Term Memory (LSTM). Furthermore, the researcher has exploited Genetic Algorithms (GA) to optimize and fine tune the parameters of Neural Network. Hidden Markov Models are generative stochastic models which can be used to model time series data and have extensively used in the implementation of applications such as Speech Recognition, Molecular Biology, Computer Vision etc. Genetic algorithms are optimisation techniques which the author has utilised to fine tune and optimise the LSTM Network. Figure 7 illustrates the entire proposed method by the author.

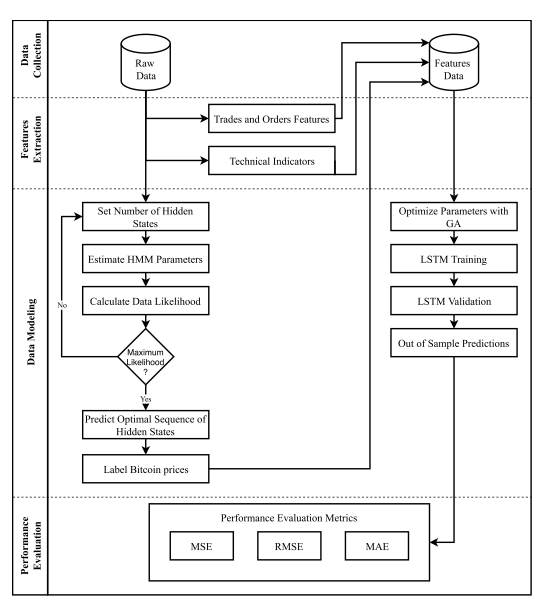


Figure 7: Proposed method using HMM and LSTM

The optimized parameters of LSTM before the training includes number of epochs, batch size, number of layers, number of neurons, dropout rate, optimizer, loss function and evaluation metric. The researcher has utilised HMM to create a new feature by clustering the prices with respect to a pre-defined number of hidden states. The author has compared optimized LSTM with ARIMA and traditional LSTM, results are shown in Table 3.

Table 3: Performance Metrics of different DL techniques by Hashish

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Name** | **MSE** | **RMSE** | **MAE** |
| ARIMA | 20153.722 | 141.964 | 112.060 |
| LSTM | 49.089 | 7.006 | 2.652 |
| Optimized LSTM | 33.888 | 5.821 | 2.510 |

**2.4 Summary**

Thera are several researches being done to understand and predict the causation of ups and downs of bitcoin price using artificial intelligence techniques. Which includes machine learning techniques, deep learning techniques on bitcoin price data. In most of the studies traditional time series forecasting doesn’t fit well to give descent results whereas due to advancements of computational power available today deep learning techniques like LSTM, GRU worked very well at least outperforms the traditional statistical models like ARIMA. Many Studies have been done to understand the people thoughts via social media analysis by extracting data from tweeter, reddit and google search insights. Researchers have found there is very good correlation of social media on bitcoin price. Many Researchers claims that more popular the bitcoin will be among people more the price will go high as per demand and supply principles which we can predict through social media analytics by understanding people’s though on the topic of cryptocurrency or bitcoin in specific.

CHAPTER 3

RESEARCH METHODOLOGY

**3.1 Introduction**

In this research, we are majorly focusing on predicting the price of bitcoin to have a good predictability and interpretability for future investments and to gain maximum of profits thus minimizing the losses. For this in this research we are planning to use naive statistical approach like Simple Moving Average Predictions and Additive Regression technique using FBProphet introduced by Facebook. Time Series machine learning algorithms such as ARIMA is widely used algorithm which we will be using as well for making the predictions of Bitcoin price of very next day by making my model learn the sequential history of Bitcoin price data which is taken from Kaggle. Similarly, we will try deep learning technique which is mostly used for sequential data known as Recurrent Neural Networks (RNNs) and specifically we will try out different flavours of RNN i.e. LSTMs (Long Short Term Memory) or GRUs (Gated Recurrent Units).

**3.2 Data Collection and Understanding**

The data for our research is collected directly from Kaggle Data Source : (SRK, 2021). Cryptocurrency Historical Prices which includes price history in Dollars not only of Bitcoins but also of Altcoins such as Dogecoin, Ethereum, Litecoin, Ripple, BinanceCoin etc. These records are captured from 29-04-2013 to 27-02-2021 with below features information.

1. Date: Date of the day the data is captured along with time i.e. 23:59 for each datapoint.
2. High: Highest value of the crypto for the day.
3. Low: Lowest value of the crypto for the day.
4. Open: Starting/Open Price of Bitcoin for the day.
5. Close: Closing Price of Bitcoin for the day.
6. Volume: Volume of Bitcoin for each day.
7. Marketcap: Market Capitalisation of Bitcoin for each day.

The dataset is structured day wise and captured at 23:59 time. Thus the end model is suitable for value investors and not for the Traders. The highest value of the Bitcoin price for that day is given in column High and similarly the lowest value is given in column Low and similarly Open and closed value of Bitcoin price is also given in the dataset along with the Market Capitalization of Bitcoin for each day. However, the information given in features are highly correlated as one affects the other. This is why we can move forward with any of one feature along with the Timestamps and which we are choosing as column ‘High’. The reason behind that is one would be more interested in the price which has been highest for the day and not the open or closing price for any day.

**3.3 Data Pre-processing**

After collecting the data, we observed that dates in our data are actually of type string/object. Date column has not only the dates but also has the timings. Although the time in each record is same i.e. 23:59 as that’s the time when the data is being captured. We have removed the time from the date column as it doesn’t have any useful information or variance resultant it will not contribute to any of our analysis or modelling, then we have changed the format of dates in python understandable datetime format. Before doing so we have already extracted the Year, Month and Day from the dates and made individual columns which can help us in further analysis.

**3.4 Exploratory Data Analysis**

We will perform exploratory analysis on this time series data to find out whether the data is stationary or not and if not we will make the data stationary by using differencing methods as that is the prerequisite for any Time Series machine learning algorithm such as ARIMA or SARIMA and will target the least RMSE as evaluation metric in ARIMA model. We observed the fluctuations in the price of Bitcoin in Dollars from 2013 to 2021 where we saw that from 2013 to 2017 the fluctuations are minimal almost Flat but post that in 2018 we see a significant spike in the price. We all remember this time when Bitcoin was too hyped in the marker and newbies were introduced with the term Bitcoin. This is the reason we can see a very slight spike in the volume at the same time because people started trusting on bitcoin and started trading/investing in bitcoin. This was the time when bitcoin price broke all the records till that time by crossing 19000 USD / 14 Lacks INR. After that the price of bitcoin again fell down and started increasing again in 2020 and hit more than 19000 USD and kept on breaking its own record. In 2021 Bitcoin price hit 60000 USD which is the latest record in the price history of bitcoin. The reasons for that can be the popularity of bitcoin which could be seen in the graph of Volume. Volume has a significant increase in 2021 which implies that trust among more people has been made i.e. more people have started trading or investing in Bitcoins, more newbies came to know about the technological advancements in Finance Sector which tend to increase in the demand of bitcoin thus the more verification work for Miners which is also one of the major factors of bitcoin price’s fluctuations. The reason of this popularity of bitcoin has a lot more to do with the tweets of Influencers like Elon Musk. Significant fluctuation has been seen just by his single tweet either the acceptance of bitcoin as payment option for Tesla or any random meme shared by Musk related to bitcoin which is why the past trend or pattern is not enough for bitcoin price predictions. There are many factors which may affect its price but we will be limiting this research to the Time Series Analysis. The overall returns or profits has a huge spike from 2013 to 2021 which is near about 5000% increment. That implies money invested in 2013 would have become 5000x in 2021 i.e. if One would have invested 1 USD in 2013 that would become 5000 USD by 2021. To find out if altcoin’s price has any correlation with the price of bitcoins, we have taken price history data of altcoins like Dogecoin, Ethereum, Ripple from 2015-8-8 and filtered the bitcoin price history data from the same date and plotted a heatmap and observed that there is a strong and positive correlation between the price of bitcoin and ethereum. There is strong relationship of 86% between the price of bitcoin and ethereum. Similarly, we can see the correlation between bitcoin and dogecoin is 76%. Well that’s not the case with Ripple, the correlation between bitcoin and ripple’s price is 42% which is not as strong as other altcoins but still there is a positive correlation. Probably there would be other altcoins too which may have strong correlation with bitcoin price. Which implies that prices of other cryptocurrencies may help in predicting the price of Bitcoin.

**3.5 Data Selection for Modelling**

The data set includes records from April - 2013 to Feb – 2021. When making predictions for the next day in 2021, data from very past may not include much relevant information eventually it may lead to become noise for predictions resultant give inaccurate results. The data set has been filtered and being used from 2018 for modelling purpose.

**3.6 Modelling Techniques**

1. Simple Moving Averages:

Simple Moving Average is very naive approach of predicting the future. Financial experts mostly take averages for guessing the future price as Stock data or Crypto data or most of time series data tend to follow the past trend or in other words History repeat itself when it comes to Time Series Data So taking averages of a particular interval can give us good idea of future price. According to (Hansun, 2013) A Simple Moving Average (SMA)s is a common average of the previous N data points in time series data. Each point in the time series data is equally weighted. When window size is taken as 10 i.e. the next day price would be the mean of just past 10 day’s price, our model is over-fitted as my prediction line is overlapping the actual price line and may not be able to predict for the future records. But when we consider more days and find the moving average that would become more informative and generalised.

1. Additive Regression model using FBProphet:

Fbprophet is package developed by Facebook especially for Time Series Analysis. According to (Taylor and Letham, 2017) “Prophet was initially developed for the purpose of creating high quality business forecasts as Analysts who can produce high quality forecasts are quite rare because forecasting is a specialized data science skill requiring substantial experience. Prophet is completely automatic forecasting library which uses Additive Regression technique at its core. The library has designed in such a way that it’s super easy for experts and non-experts to make high quality forecasts that keep up with demand. This library tries to address the following difficulties common to many business time series:

* Seasonal effects caused by human behaviour: weekly, monthly and yearly cycles, dips and peaks on public holidays.
* Changes in trend due to new products and market events.
* A reasonable number of missing observations or large outliers.

The authors claim that, even with the default settings, in many cases, their library produces predictions as accurate as those delivered by experienced analysts. Moreover, Prophet has a number of intuitive and easily interpretable customizations that allow gradually improving the quality of the forecasting model. What is especiallys important, these parameters are quite comprehensible even for non-experts in time series analysis, which is an area of data science requiring certain skills and experience. This research claims that by the combination of automatic forecasting and analyst in the loop forecasts for special cases, the possibility of solving wide variety of business forecasting problems. The illustration of forecasting process to work at scale is shown in Figure 8.

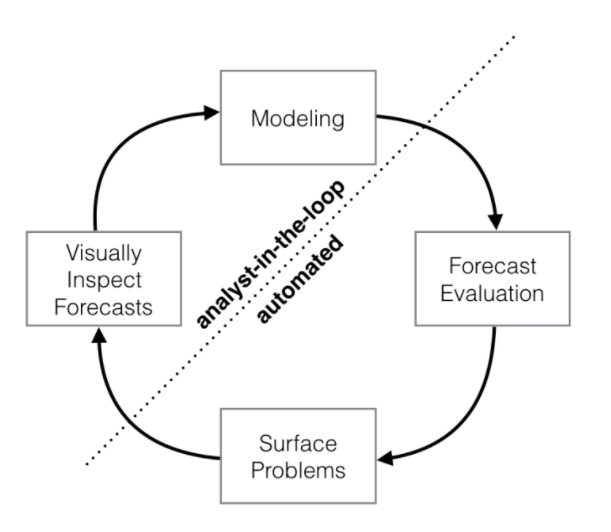


Figure 8: Forecasting Process using Fbprophet

According to (Indulkar, 2021) Fbprophet is an auto AI bundle for Time Series Analysis, that assists with foreseeing the data utilizing irregularity which can be founded day by day, week by week, month to month, and yearly. The irregularity of seasonality in data can be something that the data eyewitnesses repeating after a specific time, the model that is best suited for the irregularity of seasonality is the changes of Ice-cream sales during the winters and the summers. This is significant for fitting the information at its best. Fbprophet can be utilized to fit just as anticipate the information dependent on any irregularity and later used to get the segments for such prediction, that can be changed in the pattern of data dependent on week after week change, month to month change, and day by day change. The Equation for Additive Regression Model used in Fbprophet can be understood as:

The following components of Fbprophet for fitting of dates are follows:

* y(t) = Additive Regressive Model
* g(t) = Trend Factor
* h(t) = Holiday Component
* s(t) = Seasonality Component
* et = Error Term

1. Deep Learning forecasting using LSTM:

Similarly, For Deep Learning approach we will use traditional RNNs or LSTMs which are widely used for sequential data modelling like Time Series or Textual data for NLP tasks but we have vanishing gradient issue in traditional RNNs in case of deep network which is solved by Long Short Term Memory (LSTM) or Gated Recurrent Units (GRU). According to (Colah’s Blog, 2015) The Problem of Long-Term Dependencies Generally, Even Humans look at the very recent information to take action on the future work. For example, a NLP model which is created to predict the next word in the sentence by looking at the previous word. In case we are interested in predicting the last word in ‘The kite is flying in the …’ Sentence. It is obvious that the next word will be sky. In cases like these, where the distance between the relevant knowledge and the place that it’s needed is small, RNNs are capable of learning past information to make predictions but there are some cases where we need to have more context. Consider trying to predict the last word in the text “I grew up in India… I speak Hindi fluently.” Recent information suggests that the next word would be the name of a language but if we want to filter on which language, we need the context of India. It’s entirely possible for the gap between the relevant information and the point where it is needed to become very large.

Unfortunately, as this distance becomes huge, RNNs becomes incapable in learning due to Gradient Vanishing.

Problems with RNN:

1. The problem of vanishing and exploding gradients.
2. Vanishing Gradient don’t allow learning to take place in the network.
3. Vanilla RNNs are only able to learn short – term dependencies in practice.

Since the date we will be dealing with is a Time Series data where we don’t know that how many days of records to use as previous time stamp to predict next record which is basically a hyper parameter which we will have to tune according to the outcome. Though to deal with the issues in Vanilla RNN, many attempts/research happened so Vanilla RNNs can also learn the deep/long dependencies and the most popular and successful tweek in RNNs is known as Long Short Term Memory (LSTMs). According to (Colah’s Blog, 2015) Long Short Term Memory networks usually called LSTMs are a special kind of RNNs which are capable of learning long term dependencies. They were introduced by [Hochreiter & Schmidhuber (1997)](http://www.bioinf.jku.at/publications/older/2604.pdf), and were refined and popularized by many people in following work. They work pretty well on a large variety of problems, and are now widely used. (Raju and Mohammad, 2020) says LSTMs enables the neural network to learn more about many time steps by maintaining a more steady error which enables the network to learn long term dependency. LSTM cell contains forget and remember gates which allows the cell to take decision about which information to discard and which information to be transmit based on the strength and importance of the information. Figure 9 depicts the working of LSTMs.

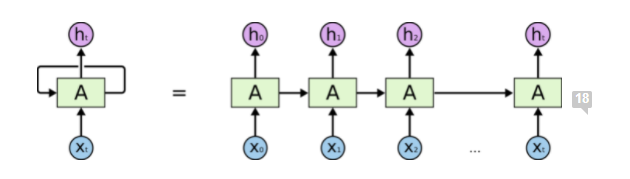


Figure 9: Long Short Term Memory (LSTM)

**3.7 Summary**

This Research’s motivation is to apply multiple machine learning techniques to predict the price of Bitcoin and come up with robust and reliable model. In this research we will try to understand the predictable power of simple moving averages then will compare the results of each and every model we built which includes Additive Regression Models using FBprophet, Sequential Modelling experts like ARIMA and RNNs or LSTMs.

CHAPTER 4

EXPERIMENTATION AND ANALYSIS

**4.1 Introduction**

There are multiple experiments being done in this study to come to an optimal solution and few observations has been made which is being discussed in this chapter. This chapter includes all the data analysis and experimentations including hyper parameter tuning. which is being done with the multiple time series predictive techniques discussed in Chapter 3 Research Methodology.

**4.2 Data Analysis**

As the data collected from Kaggle includes highest, lowest, opening and closing value, volume and Market capitalisation of bitcoin. Which is highly correlated to each other as can be seen in Figure 10 which is why one feature to predict the price is more than enough along with the timestamps.

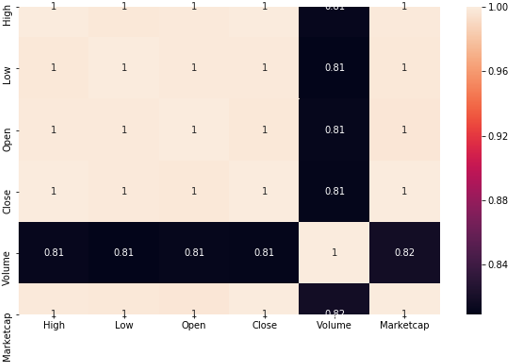


Figure 10: Correlation Heatmap of collected dataset

The trends or patterns are also very similar to each other and easily familiarise with the fluctuations of the price of bitcoin in the past years as can be seen in Figure 11. Where we also observed that volume in 2021 has very significant spike which indicates that people started trusting the bitcoins thus bought and invested in bitcoin. As the demand and supply kept on increasing the price also increased and can be seen in open, close, high and low price graphs. This causation is rare in stock market as volume doesn’t actually makes any impact on the price of share but in crypto world it does. This research is focused to predict the continuity of this trend.

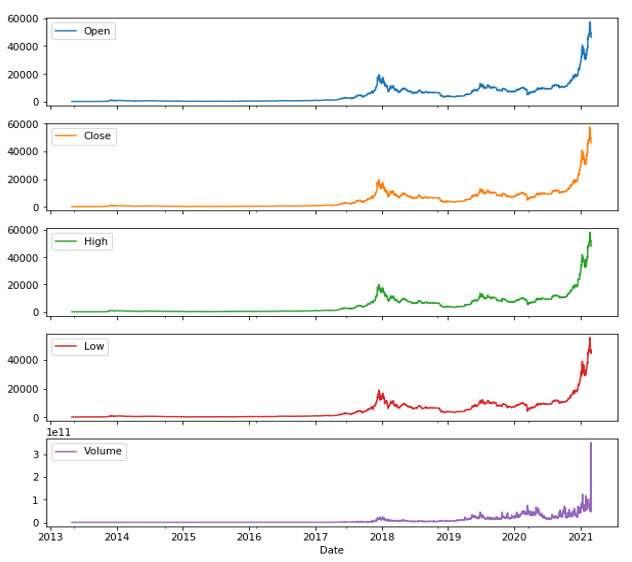


Figure 11: Trends and Patterns in collected dataset

To have a better idea of the bitcoin price history line chart in Figure 12 indicates the highest value of bitcoin price in each year from 2013 to 2021 (Please note the data for 2021 is only till 27th Feb) and Figure 13 indicates the maximum price of bitcoin month wise for each year.

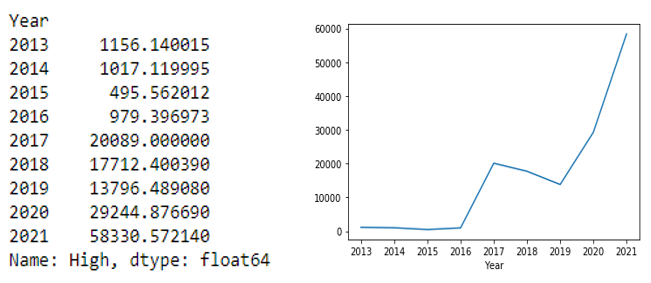
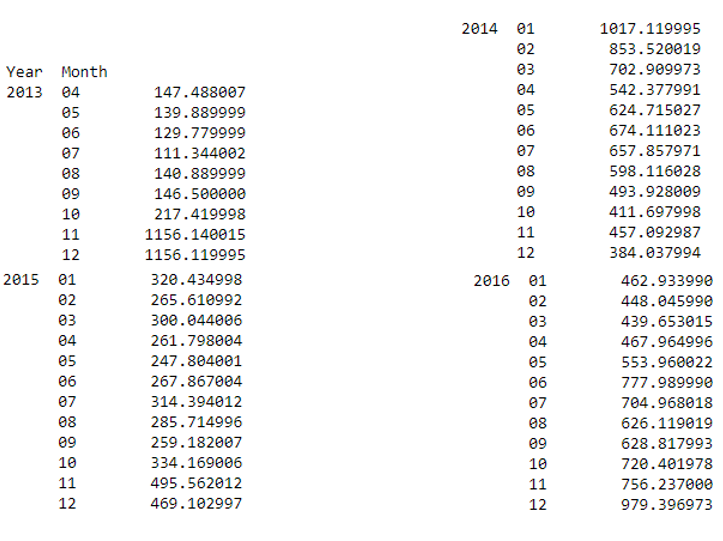


Figure 12: Highest value of BTC price yearly



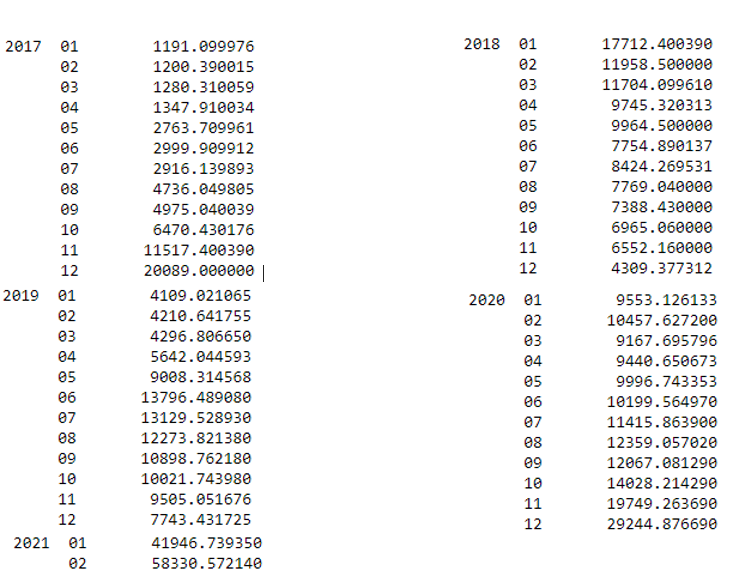


Figure 13: Highest value of BTC price monthly for each year

The data of altcoins such as Dogecoin, Ethereum and Ripple has also been taken from the same data source and filtered from 8th Aug 2015 along with Bitcoin data to analyse the relationship or correlation among them. For this purpose, Figure 14 of correlation heatmap is plotted and found that Ethereum has the strongest correlation with the bitcoin which can be used as a feature in further studies. This is followed by the Dogecoin which has 76% of correlation with the price of bitcoin. Although Ripple doesn’t seem to have a good relationship with bitcoin price as it has only 42% correlation with bitcoin.

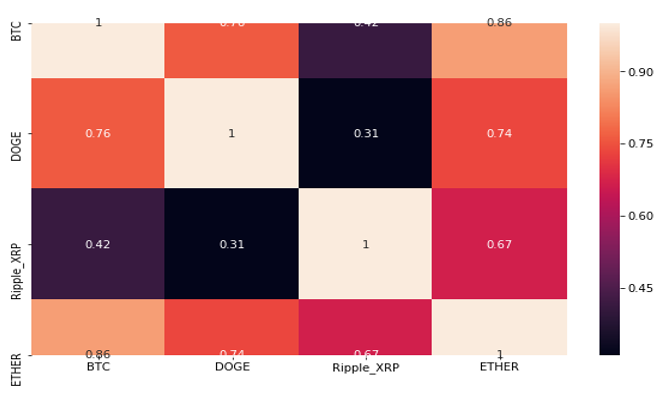


Figure 14: Correlation Heatmap of Altcoins and Bitcoin Price

**4.3 Experimentations**

* Simple Moving Average:

Simple Moving Average is the very simple approach of predicting the future trend. It is basically the average of previous N data points where N is hyper-parameter. A graphical representation of bitcoin’s highest price’s monthly moving average can be seen in Figure 15.

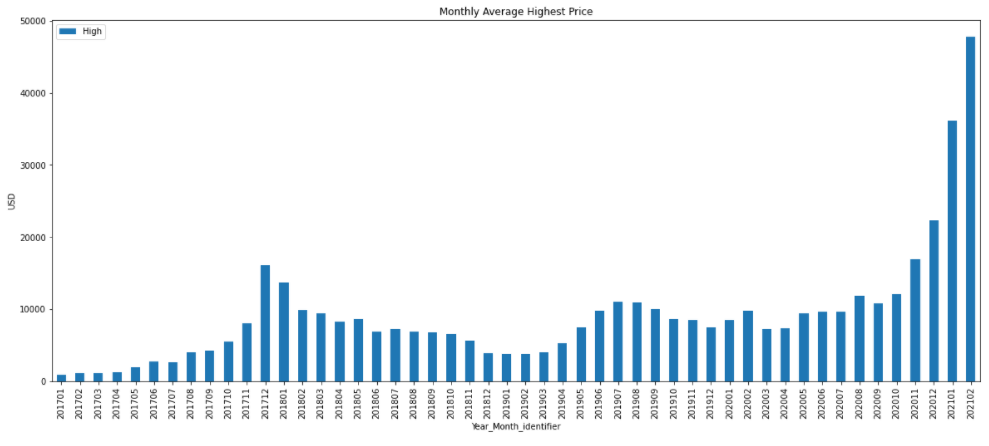


Figure 15: Moving Average for each month

Moving Average is basically the average of past N data points, when N is taken as 10 the prediction line is overlapping with the actual data points thus has overfitting involved as shown in Figure 16. The RMSE came out to be 1033 with N =10. As N is the hyper-parameter so when N is taken as 30 i.e. a month’s average the RMSE came out to be 2141 and prediction line very nicely generalised the actual data points as shown in Figure 17.

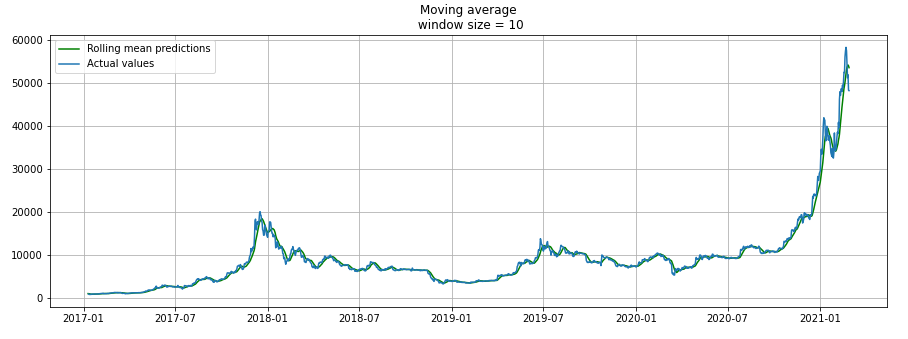


Figure 16: Simple Moving Average when N=10

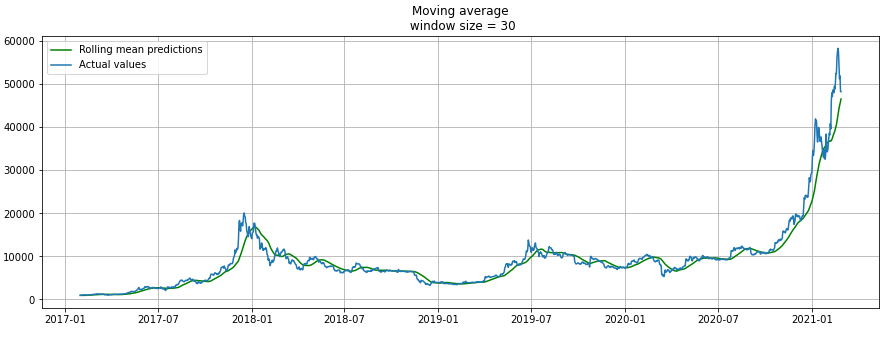


Figure 17: Simple Moving Average when N=30

* Additive Regression:

Fbprophet introduced by Facebook for Time Series Analysis is very easy to use. In this study we utilised Fbprophet library to implement Additive Regression model on Bitcoin price data. Prophet requires the data columns as ‘ds’ for dates columns and ‘y’ for the price data column on which the model has to be fitted or in other words the target column has to be renamed to ‘y’ for prophet to work. These are kind of prerequisites for the prophet library and rest it can take care by its own. As we are building the model to predict the highest price for the day so the column named High is being renamed to ‘y’. The entire data from 2017 to 2021 has been splitted into train and test at first where data from 2017 to 2020 has been considered as train data and data of 2 months of 2021 as test data. After applying or fitting the prophet on train dataset, we have created the future data frame with the prediction size of 58. This number is basically the number of records in our test dataset and the added 58 dates in this future data frame are the dates of our test dataset. Post predicting the highest price of bitcoin on future data frame, the root mean squared error (RMSE) is came out to be 4476. When plotted we observed that the model was able to predict well on the trained data and was not able to generalise on test data or the 2 months’ data of 2021 because there is a sudden huge peak in the actual price of bitcoin price which model has no idea. This sudden peak in price can be seen in Figure 18.

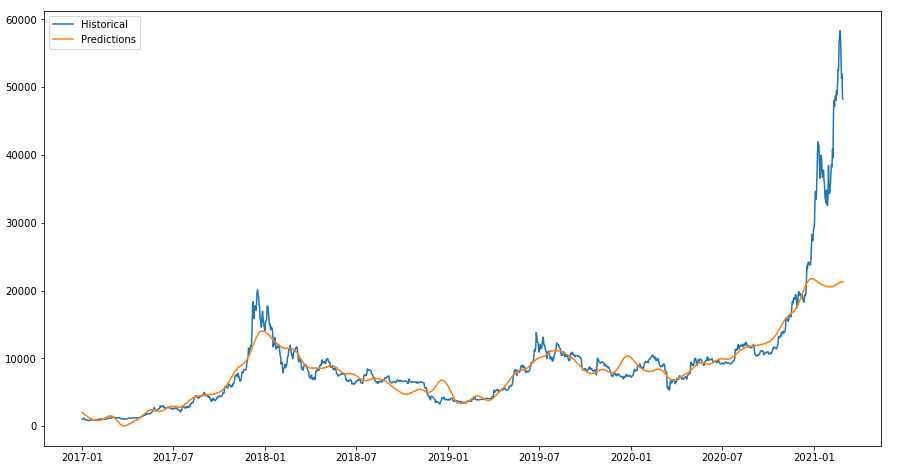


Figure 18: FBProphet Forecast

Due to this sudden peak and model failure, we trained the model again but this time on complete dataset i.e. from 2017-01-01 to 2021-02-27 so the model learns the sudden price jump patterns too and the results are better than before as the RMSE came out to be as 1644 and when plotted the model is able to generalise the data points as can be seen in Figure 19. But this time the model is completely predicting on what it is trained on. So we created new future data frame with completely unseen 60 future dates added and predicted and plotted the forecast to see the trend of future and unseen dates to make the investment decision. This can be seen in Figure 20.

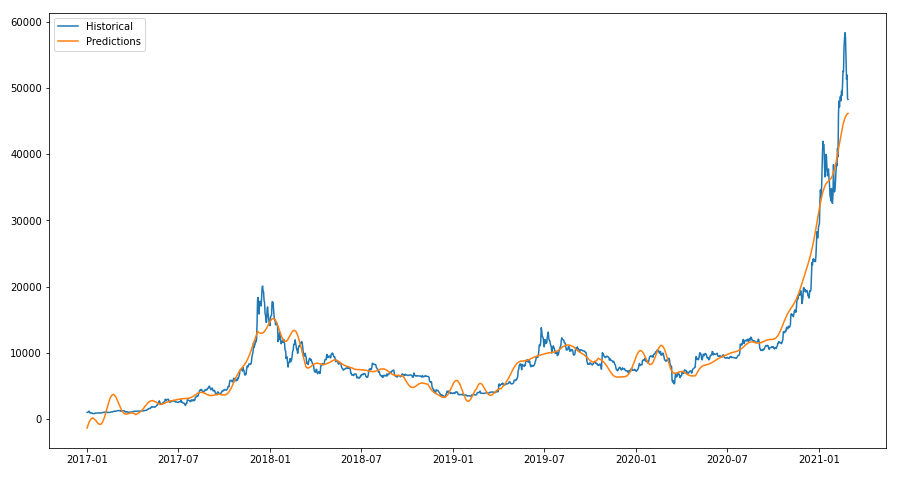


Figure 19: Retrained FBProphet Forecast

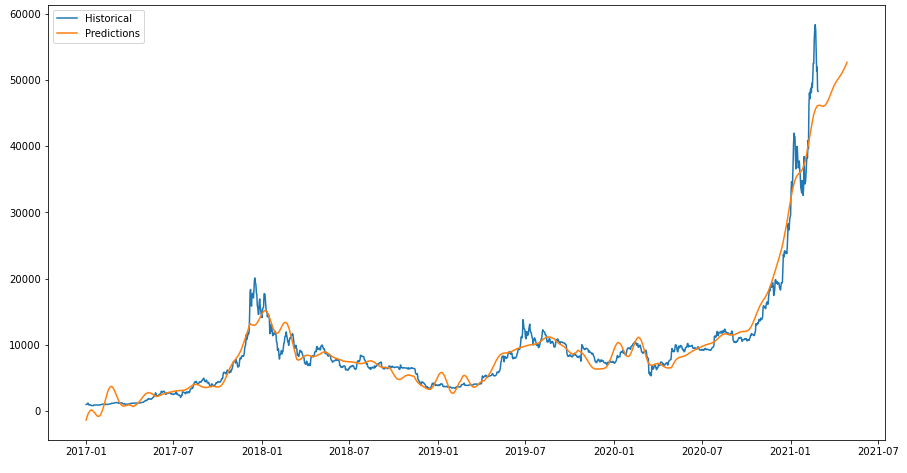


Figure 20: Forecast Included Unseen 60 Future Dates

The forecast of next 60 days i.e. from 2021-02-28 to 2021-04-28 indicates that the price of bitcoin will increase which happen to be the case when validated online.

* Deep Learning Technique:

Recurrent Neural Networks are widely used when it comes to sequential data and so we utilised the Long Short Term Memory Network using open source API provided by Google named Tensorflow and used Keras as backend for tensorflow. We took the entire data from 2017-01-01 to 2021-02-27 along with the columns of highest price as we did for Prophet and Simple Moving Average. We removed the date column from the data as there is no longer need of timestamps when it comes to training LSTMs for time series because LSTM understands only the numerical values and tries to learn based on the previous history of data. For that purpose, our data is stored in feature\_X using which our model will learn and also same list of price is stored in variable target\_y which we are going to use to prepare our data to make it understandable for LSTM. We are using the Tensorflow API to prepare the data in such a way that the model learns the past N days’ data to predict the price of very next day. Where N is the window length which is a hyper parameter. The tensorflow’s Time series generator make the data available in the batches given the batch size we want and prepares the data in the manner where X is the data of past N days and y is the price of very next day and then moves its pointer to the very next record and takes another set of N day’s price to predict the very next day price given the stride as 1. Stride is again a hyper parameter which is to be tuned but in our case we are taking stride as 1. After preparing the data using tensorflow’s Time series generator, the data for LSTM looks like shown in Figure 21 when batch size taken as 1 and window size as 3.

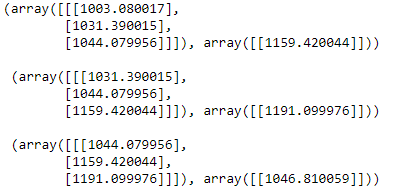


Figure 21: Prepared Data for LSTM

We then fed this prepared data with window size of 3 and batch size of 32 to a Neural Network of 3 hidden LSTM layers with activation function as relu, optimizer as Adam, loss as mean\_squared\_error and metric as RMSE and ran it for 15 epochs. The summary of the Neural Network Architecture can be seen in Figure 22. We got the best RMSE as 981 in 4th epoch using this Network.

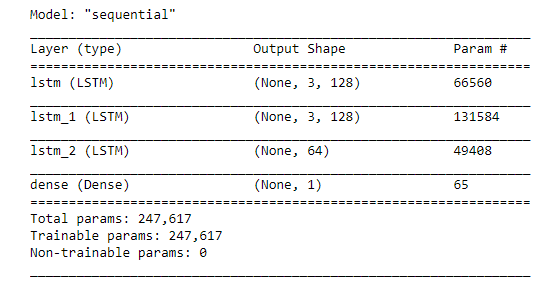


Figure 22: Summary of LSTM Neural Network

We then added the dropouts of 20% after each LSTM layer and increased the window length to 6 and this network doesn’t seem to be working better than the previous one. With this network we got the best score of RMSE as 1504 in 19th epoch. Then we added one Dense layer of 128 neurons before the output layer and increased the window size to 8 which again doesn’t outperform out first ever build LSTM Network. After multiple experimentation we observed that increasing the length of window is actually making the model worst. This is why we again took window size as 3 and added one more LSTM layer of 64 neurons with no dropouts and the summary of the Network Architecture can be seen in Figure 23.

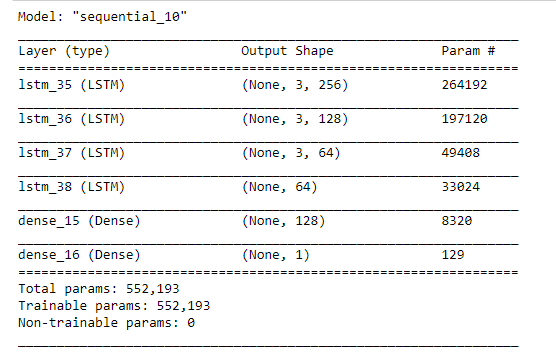


Figure 23: Summary of LSTM Neural Network 2

We continuously reduced the learning rate to update the weights in Neural Network using the callback function ReduceLROnPlateau available in tensorflow’s keras. We got the best value of RMSE as 784 in 11th epoch which is way better than any other Network Architecture.

**4.4 Summary**

After all the analysis and experimentation with different kind of machine learning and deep learning techniques and tuning their hyper parameters. We found that Long Short Term Memory (LSTM) Neural Network works best when it comes to reducing the loss and learns the training data so well. Which happen to be the case with Chapter 2 Literature Review as well. Researchers had found LSTMs to be the best for Time Series Analysis and so the case with us too.

CHAPTER 5

RESULTS AND DISCUSSIONS

**5.1 Introduction**

In the previous chapters, we discussed about data collections, selection and pre-processing on it and also different-2 techniques to build a robust generalizable machine learning or deep learning model. In this chapter, we will discuss about the results obtained from the applied techniques and also will determine the best technique among all of those.

**5.2 Results Obtained**

The evaluation metrics for time series prediction are very similar to the traditional regression problems as can be seen in Figure 24.

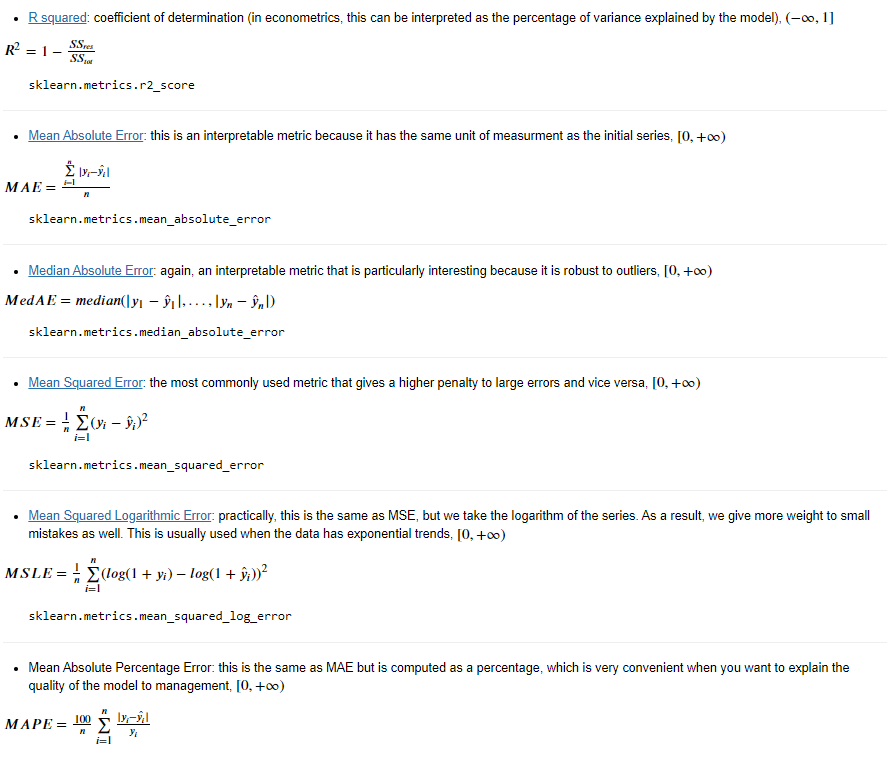


Figure 24: Evaluation Metrics for Regression

In this research, we have chosen the RMSE or Root Mean Squared Error as the evaluation metric for all the applied techniques. With Simple Moving Average, we found that when N is taken as 10 the RMSE is reasonably okay but when plotted the prediction line was overlapping the true data points which indicates the model has become overfit. To avoid the problem of overfitting we tuned the value of N and when N is taken as 30 the plot looked fine and RMSE came out to be as 2041. Similarly, Additive Regression Model using Facebook’s prophet library, the RMSE we achieved is 1644 but this number is reduced significantly when tried Recurrent Neural Network like LSTM and RMSE we achieved is 784 with the window size of 3 and decaying learning rate with the fraction of 0.5. The results of all the applied techniques can be seen in Table 4.

Table 4: Performance Metrics of different techniques applied in this research

|  |  |  |
| --- | --- | --- |
| **Techniques** | **Parameters** | **Results (RMSE)** |
| **Simple Moving Average** | N = 10 | 1033 |
| N = 30 | 2141 |
| **FbProphet** |  | 1644 |
| **LSTM** | window = 3 | 981 |
| window = 6 | 1504 |
| window = 8 | 1761 |
| window = 3 with reducing learning rate | 784 |

**5.3 Summary**

Being a time series problem Bitcoin’s price prediction seems be to very similar to any other time series problem but the sudden ups and downs like in bitcoin price are very rare to see in other time series problems which is also have to be treated as outliers in those cases but that’s not the case with bitcoin price data as this is very common behaviour in cryptocurrencies to have sudden peaks in upward or downward directions. Which is why capturing those high peaks becomes difficult sometimes as they are never seen before. The predictions with the build models are not relevant for long term predictions and can help only in making investment decisions to see few days’ trends for future.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

**6.1 Conclusion**

The sole purpose of this research is to have robust interpretable artificially intelligent model which can help in making intelligent decisions for maximising the profits for cryptocurrency investments and bitcoin in particular. Although the same techniques can be applied to any other cryptocurrencies as well. In this research, we have observed that other cryptocurrency’s data also can help in the prediction of bitcoin price as there is very good co-relation among them. Ethereum and Dogecoin has very good co-relation with the price of bitcoin and can be used as features for predicting the price of Bitcoin in future work.

**6.2 Recommendations**

Time Series analysis is very common problem in industry as most of the financial and sales data is sequential in nature which becomes very critical to have understand the seasonality or any trend in the data. This is why there are many auto-ml libraries available today and are becoming so popular which includes time series analysis as one of the functionalities or feature. Which can be tried out for future work. Few of the libraries which can be tried out includes PyCaret by Moez Ali and Greykite by Linkedin.

**3.7 REFERENCES**

Balfagih, A. M. and Keselj, V. (2019) ‘Evaluating Sentiment C1assifiers for Bitcoin Tweets in Price Prediction Task’, *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, pp. 5499–5506. doi: 10.1109/BigData47090.2019.9006140.

*Block: 1 | Blockchain Explorer* (no date). Available at: https://www.blockchain.com/btc/block/00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048 (Accessed: 13 June 2021).

Blockchain (2019) *Blockchain Explorer - Search the Blockchain | BTC | ETH | BCH*, *Politics, Gender, and Concepts*. Available at: https://www.blockchain.com/explorer?utm\_campaign=dcomnav\_explorer%0Ahttps://www.blockchain.com/btc/tx/4410c8d14ff9f87ceeed1d65cb58e7c7b2422b2d7529afc675208ce2ce09ed7d (Accessed: 13 June 2021).

Colah’s Blog (2015) *Understanding LSTM Networks -- colah’s blog*. Available at: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (Accessed: 28 April 2021).

Demir, A. *et al.* (2019) ‘Bitcoin Price Prediction Using Machine Learning Methods’, *1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings*, pp. 144–147. doi: 10.1109/UBMYK48245.2019.8965445.

Felizardo, L. *et al.* (2019) ‘Comparative study of Bitcoin price prediction using WaveNets, Recurrent Neural Networks and other Machine Learning Methods’, *BESC 2019 - 6th International Conference on Behavioral, Economic and Socio-Cultural Computing, Proceedings*. doi: 10.1109/BESC48373.2019.8963009.

Gartner (2021) ‘What CFOs Are Saying About Bitcoin’, (March), pp. 1–10.

Gupta, T. (2017) *Bitcoins | What are Bitcoins? | How to earn Bitcoin?*, *Google Blogger*. Available at: https://tambitcoins.blogspot.com/2017/06/what-are-bitcoins-how-to-earn-bitcoin.html (Accessed: 25 April 2021).

Hansun, S. (2013) ‘A new approach of moving average method in time series analysis’, *2013 International Conference on New Media Studies, CoNMedia 2013*. doi: 10.1109/conmedia.2013.6708545.

Hashish, I. A. *et al.* (2019) ‘A Hybrid Model for Bitcoin Prices Prediction using Hidden Markov Models and Optimized LSTM Networks’, *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2019-Septe, pp. 721–728. doi: 10.1109/ETFA.2019.8869094.

Indulkar, Y. (2021) ‘Time series analysis of cryptocurrencies using deep learning fbprophet’, *2021 International Conference on Emerging Smart Computing and Informatics, ESCI 2021*, pp. 306–311. doi: 10.1109/ESCI50559.2021.9397004.

Li, L. *et al.* (2019) ‘Bitcoin options pricing using LSTM-Based prediction model and blockchain statistics’, *Proceedings - 2019 2nd IEEE International Conference on Blockchain, Blockchain 2019*, pp. 67–74. doi: 10.1109/Blockchain.2019.00018.

Li, X. and Wu, W. (no date) ‘A Blockchain Transaction Graph based Machine Learning Method for Bitcoin Price Prediction’, pp. 1–10.

Maleki, N. *et al.* (no date) *Bitcoin Price Prediction Based on Other Cryptocurrencies Using Machine Learning and Time Series Analysis*.

McNally, S., Roche, J. and Caton, S. (2018) ‘Predicting the Price of Bitcoin Using Machine Learning’, *Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018*, pp. 339–343. doi: 10.1109/PDP2018.2018.00060.

Mittal, A. *et al.* (2019) ‘Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data’, *2019 12th International Conference on Contemporary Computing, IC3 2019*, pp. 4–9. doi: 10.1109/IC3.2019.8844899.

Mudassir, M. *et al.* (2020) ‘Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach’, *Neural Computing and Applications*, 6. doi: 10.1007/s00521-020-05129-6.

Raju, S. M. and Mohammad, A. (2020) ‘Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis’.

Rathan, K., Sai, S. V. and Manikanta, T. S. (2019) ‘Crypto-currency price prediction using decision tree and regression techniques’, *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, 2019-April(Icoei), pp. 190–194. doi: 10.1109/icoei.2019.8862585.

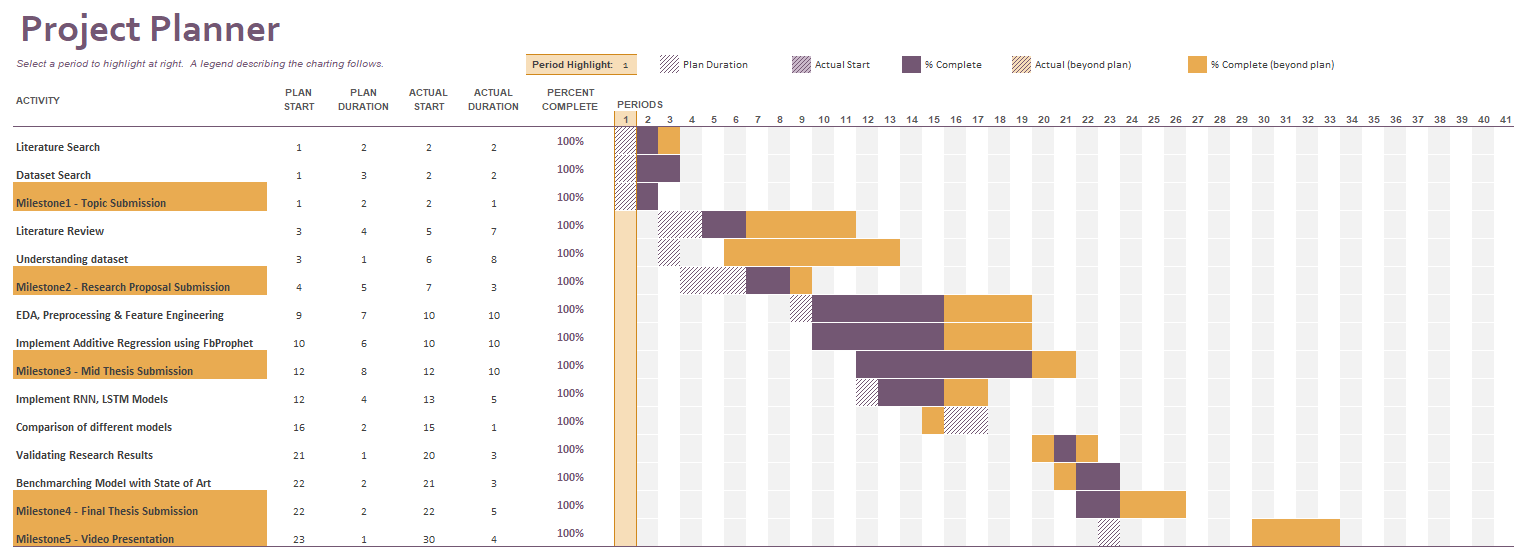
Reddy, L. S. and Sriramya, P. (2020) ‘A research on bitcoin price prediction using machine learning algorithms’, *International Journal of Scientific and Technology Research*, 9(4), pp. 1600–1604.

Sin, E. and Wang, L. (2017) ‘Bitcoin Price Prediction Using Ensembles of Neural Networks’, *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 666–671.

SRK (2021) *Cryptocurrency Historical Prices | Kaggle*, *Kaggle*. Available at: https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory (Accessed: 13 June 2021).

Taylor, S. and Letham, B. (2017) *Prophet: forecasting at scale - Facebook Research*, *Facebook*. Available at: https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/ (Accessed: 13 June 2021).

APPENDIX -A: RESEARCH PLAN



APPENDIX -B: RESEARCH PROPOSAL

Time-Series forecasting of bitcoin prices using machine learning and deep learning techniques

Tarang gupta

Research Proposal

December 2020

Table of Contents

[Introduction 3](#_Toc72827567)

[Background and Related work 5](#_Toc72827568)

[Aim and Objective 6](#_Toc72827569)

[Significance of the Study 7](#_Toc72827570)

[Scope of the Study 7](#_Toc72827571)

[Research Methodology 7](#_Toc72827572)

[Expected Outcome 9](#_Toc72827573)

[Required Resources 9](#_Toc72827574)

[Research Plan 9](#_Toc72827575)

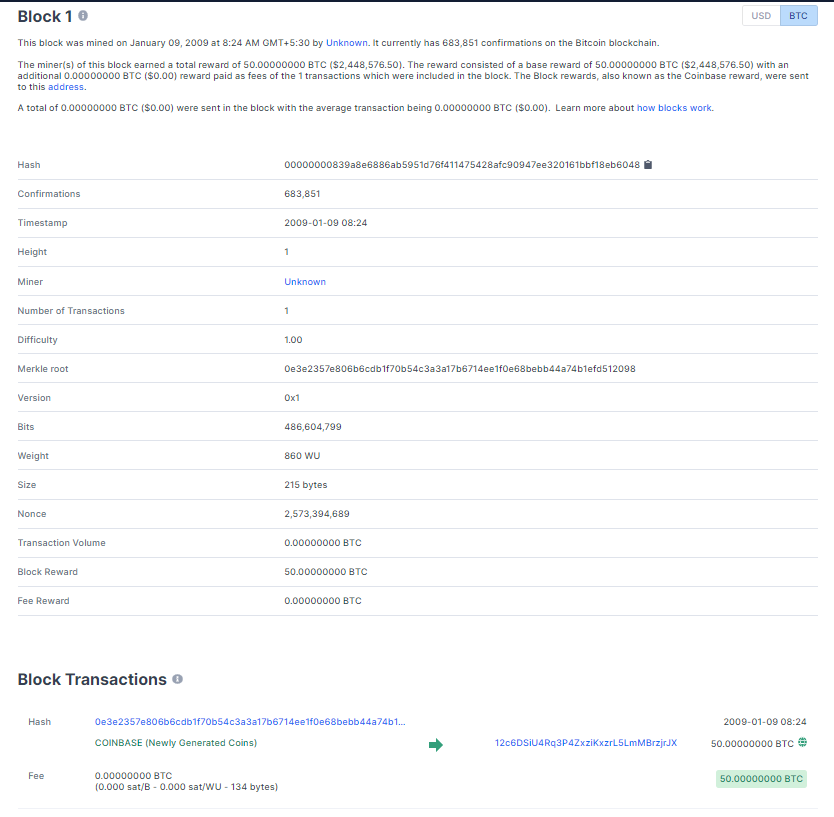
[References 10](#_Toc72827576)

# Introduction

Bitcoin is not a coin like our traditional coins which can be seen and touched rather it is a digital coin which resides on the internet only. Bitcoin is the first decentralized cryptocurrency invented and implemented in 2009 by Mr. Satoshi Nakamoto. The idea to develop the Bitcoins is to give the Financial Independence to Individual So he can send money from one place to another without any Bank’s or any other Government authority’s intervention. For which it uses the wonderful technology named “Blockchain” which is now being used other than developing the Cryptocurrencies such as in Hospital Management Systems, Airlines Management Systems etc. With Blockchain the Security in Finance became more reliable. To understand in brief, “You can make fool of one person or may be 5-10 people but you won’t be able to make fool a group of thousands of people” Yes that’s how Blockchain technology works at its core. It is a chain of Blocks and each block is associated with a hash key of the previous blocks which is distributed among the nodes in the Network of Blockchain. Bitcoin is the first application which follows the principles of Blockchain technology. In 2009 the price of Bitcoin was almost zero. There were no exchanges, there was no market. Only the fans of this new cryptocurrency were sending the bitcoins just for fun reward to each other. In 2017 Bitcoin had major jump in its price. Which increased to $9925 and made the record and in 2020-2021 Bitcoin broke all the limits and crossed $50,000. Some of the Countries have banned transactions in Bitcoin or any other cryptocurrencies such as Ethereum, Litecoin, Bitcoin Cash, Litecoin, Ripple and the current trending one Dogecoin. And some of the countries has imposed strict regulations for Cryptocurrency world as it affects the country’s financial stability and promotes cyber-crimes as what they say. Yes, there is no doubt that in cyber world for all the unethical stuff happen there would be Crypto transaction involved in it either on Dark web, Deep web or usual Ransomware attacks ex. Wnnacry Ransomware in 2017.

“Bitcoin is the first cryptocurrency which uses Block chain technology at its core and xused worldwide for digitalx payment and for investment purposes. Bitcoinx is decentralized which means there is no control of government or bank or any other authority on Bitcoin. Bitcoin is independent and its price depends upon its demand and supply. Transactionss made by Bitcoins are easy and not limited to any specific country or region which is the case in traditional banking system. To send out some amount of money overseas using our traditional system it would take more than 2-3 days or may be a week. During this time the bank or some higher authority actually validate the transaction. This process has completely abandoned and replaced by fast and secure implementation of Block chain when it comes to Cryptocurrency transactions and has capability of getting the overseas transactions within seconds. Investment can be done through various marketplacesx known as bitcoin exchanges such as Binance, Coinbase, CoinDCX, WazirX etc. These Bitcoin Exchange wallets allow the investor to sell/buy Bitcoins using different currencies like INR, USD etc. The most popular and largest Bitcoin exchange wallet is Mt Gox which was launched in July 2010 and by 2014 it was handling over 70% of all bitcoin transactions worldwide. Bitcoins are stored in these exchange digital wallets just like a virtual bank account. The data of all the transactions at each timestamp is stored in the Block chain Network which can be viewed on <https://www.blockchain.com/explorer?view=btc> . Each record in a blockchain is known as block. Each block contains address of its own block and address of previous block known as hash. The data of transactions in blockchain network is encrypted and During transactions the name of user is not revealed which gives the user the flexibility of anonymity but only their wallet ID is made public”, (Demir *et al.*, 2019)

Below is the information of first ever block of blockchain network extracted from <https://www.blockchain.com/btc/block/00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048>



According to (Raju and Mohammad, 2020) “Bitcoin is an online payment system which can be used globally in highly stabilized and secured manner. It is Peerx - to - Peer value transfer protocol. Bitcoin transactions gets verified by the nodes in the block chain network and gets published on a public ledger. The Smallest unit of BTC known as Satoshi represents 0.00000001 bitcoin which is One hundred -millionths of a xbitcoin.” So we got the good understanding about what Bitcoin (BTC) is but how do we earn them? According to (Gupta, 2017) “The earning methods of Bitcoin includes: Bitcoin can be purchased just by paying the current amount of Bitcoinx and get that in your crypto exchange wallet. If One doesn’t want to invest much then there is no need of buying a whole Bitcoin One can buy the fraction of Bitcoin which is known as Satoshi. The Bitcoin will be safe in your Crypto Exchange Wallet (You can use Coinbase wallet or Binance wallet) and when the value of Bitcoin again fluctuates analyse when you feel yourself in Profit you can sell that Bitcoin. Another way of earning the BTC is through mining: - The way of creating the Bitcoin or it’s fraction Satoshi is known as Bitcoin Mining. There is a limitation in number of Bitcoins in the real world i.e. 21 million which implies that the number of Bitcoins can never be more than 21 million in the Market. Due to this rule there will be continuous increase we will see in BTC price. Just Like in Normal Transactions when a sender sendss money to Receiver there is always an involvement of Banks and some other government authorities to verify or to complete the Transactions which may take days to happen but in Bitcoin Transactions there is no one but a miner will be there. Every transaction first gets distributed in the Blockchain Network in the encryptedx form where several distributed computing machines or (nodes of network) tries to validate and verify that transaction and one of them at the end finishes the verification which acts as a Bitcoin miner. Each Transaction contains some Math Problemss and these math problems are actually the decryption algorithms of the encrypted transaction. Now to verify or to decrypt the algorithm of the Transactionn the miner needs powerful CPUs and GPUs or ASIC (Application Specific Integrated Circuits) to solve the Math Problems. Resultant the miner will be awarded with some Bitcoins or the fraction of Bitcoin i.e. Satoshi to verify the transaction. That’s how a bitcoin is created and new bitcoins make its existence in the real world. This entire process of creating new bitcoins is called Bitcoin Mining.”

# Background and Related work

Investing in stocks in order to maximize the profit is not new to us. People have been investing in stock market to get the proportional profit of their invested amount in a particular stock. Same is the case for Cryptocurrencies as well which may give more than 500% of return in just few days. Take example of Dogecoin which went from $0.006 to $0.4498 in just about 2 months. What if one already knew about the future predictions? Imagine the possibilities of knowing what will happen in the future! Imagine what it would be if you had invested in Dogecoin in Jan 2021 at the price of $0.006. Many people regret for not investing in Bitcoins back in 2016 or 2017 but how they even know about when the Bitcoin price going to rise or for any other cryptocurrency. Anyways not to regret anymore we can capture the patterns or trends of rising prices of Bitcoins using Machine Learning. Machine Learning (ARIMA) or Deep Learning (LSTMs/RNNs) will not tell us the future but it might help in capturing the trend or pattern of the Bitcoin price movement. However, the parameters which actually affect the prices of Bitcoin is not just the previous trend. Bitcoin price fluctuates on the basis of **demand** and **supply** but a rare occurrence contributes the most in fluctuation like getting the temporary ban of Bitcoin Trading in any country or temporary ban in Bitcoin mining, Since Bitcoin Mining is highly proportional to the Electricity consumption as mining process takes too much of time and computational power to validate the transaction and make new bitcoin or satoshi’s (fraction of bitcoin) existence in the real world. Which leads to significant drop in the Price of Bitcoin. In the same way, Acceptance of Bitcoin as the payment option in buying Tesla made a positive impact to the Bitcoin prices, Ban Removals from country also tend to make the positive impacts and so the Bitcoin Prices jumps to its peak. Recent advancements in computer GPUs and their affordability, awareness of Crypto world to newbies also made a significant impact on Bitcoin prices. A single tweet by Elon Musk made Dogecoin popular on the internet. Thus more demand and trust in Dogecoin and so the prices of Dogecoin rose. There has been recent research to predict the Bitcoin price using the thoughts of individuals which they post on Tweeter, Reddit, LinkedIn, Google Trends. According to (Sin and Wang, 2017) “As the markets is moderately new, existing works identified with determining in this markets is genuinely restricted. One examination showed that Google Trends information and volume of tweets identified with Bitcoin on Twitter have positive relationship with Bitcoin's cost and subsequently might have the option to anticipate the vacillations in cost of Bitcoin. In another investigation, Bayesian Regression, xBinary classification algorithm, was utilized to anticipate value variety in Bitcoin and the expectation gave practically 200%s returns in under 60 days when utilized with a trading technique. The examination reasoned that there might be 'information' in Bitcoin's chronicled data that can help anticipate future value varieties.” but reading the thoughts won’t help much in predictions of Bitcoin as People tend to discuss more than actually taking the action on it. Indeed, when positive thoughts are posted by some popular personality like Elon Musk or famous Influencer which tend to change the mind of thousands or lacks of people, definitely there will be increase in Demand and so the Bitcoin price will rise. According to (Mudassir *et al.*, 2020) “At the point when Bitcoin started to stand out enough to be noticed at end of 2013, it saw a critical vacillation in its cost. A strands of literature has inspected the consistency of Bitcoin returns through different boundaries, for example, socially based media consideration and Bitcoin-related authentic specialized pointers. One gathering considered the period from September 4, 2014, to August 31, 2018, by catching the occasions the term Bitcoin has been tweeted. The outcomes showed that the quantity of tweets on Twitter can impact Bitcoin exchanging volume for the next day. Besides, examined the impact of clients remarks in online stages on value variances and number of exchange of digital currencies and found that Bitcoin is especially corresponded with the quantity of positive remarks via web-based media. They detailed the accuracy of 79% alongside Granger causality test, which infers that user conclusions are valuable to anticipate the value changes. With regards to time series forecasting, There are three distinct sorts of statistical model based methodologies for time series forecasting. The primary methodology, pure models, which just uses the chronicled information on the variable to be predicted. Instances of pure time series forecasting models are Autoregressive Integrated Moving Average (ARIMA) and Generalized AutoRegressive Conditional Heteroskedasticity (GARCH). ARIMA based time arrangement gauge at following day Bitcoin costs.”

# Aim and Objective

The Primaryx aim and objective of this research is to develop a trustworthy and interpretable predictive time series model which can capture the past trends or patterns in the Price of Bitcoin So one can make better decisions for future investment in Cryptocurrency (Bitcoin) to maximize the ROI.

The research objectives are formed dependent on the point of this examination which are as per the following:

1. Understanding the time series data of Bitcoin price.
2. Finding relation of Bitcoin price with Altcoin’s price such as Ether, Ripple etc.
3. Developing a robust machine learning /deep learning model for the successful prediction of bitcoin future price that could yield significant profit.
4. Evaluating the predictive models to identify most profitable model with least errors.
5. Comparative Analysis on Machine Learning and Deep Learning Techniques.

# Significance of the Study

Everyone wants to have the passive source of income. Either it is stock market or Affiliate Marketing or Rental Income. We have been seeing these examples but the most underrated example of one of the passive sources of income is trading in Cryptocurrencies. This is something which is very well known by youngsters (not limited to). It becomes vital to have good understanding of Crypto World to get started in Bitcoin investment just like in Stocks. Predictions of Bitcoin price will help individual to gain maximum profit and save them in losing the big amount of money. This research will add to decipher or clarify the forecast produced by different prescient AI or profound learning models to help the choice in crypto speculation.

# Scope of the Study

* The data for research is taken directly from Kaggle.
* The Research is limited for the people who look for passive source of income by investing in Cryptocurrency.
* The Techniques/ Algorithms going to be used for the Research will be Time Series analysis and Deep Learning but we will try other traditional machine learning algorithms as well such as SVM, Random Forest Repressors as in recent work they have performed well.

# Research Methodology

In this research, we are majorly focusing on predicting the price of bitcoin to have a good predictability and interpretability for future investments and to gain maximum of profits thus minimizing the losses. For this we are planning to use traditional regression techniques such as Linear Regression, Support Vector Machine Regressors, Random Forest Regressors. Although these machine learning algorithms are not capable of capturing the information hidden in the sequential data/ Time Series data but they are well versed in learning the distributions of data and make wonderful predictions.

Time Series machine learning algorithms such as ARIMA is widely used algorithm which we will be using as well for making the predictions of BTC price of very next day by making my model learn the sequential history of BTC price data which is taken from Kaggle.

Similarly, we will try deep learning technique which is mostly used for sequential data known as Recurrent Neural Networks (RNNs) and specifically we will try out different flavours of RNN i.e. LSTMs (Long Short Term Memory) or GRUs (Gated Recurrent Units).

Before getting started with the model building strategy. We will have a look at our dataset which contains records of BTC price from 2013 to 2021 with features like mentioned below -

**Date, High, Low, Open, Close, Volume, Marketcap**

The dateset is structured day wise and captured at 23:59 time. Where the highest value of the BTC price for that day is given in column “High” and similarly the lowest value is given in column “low” and similarly Open and closed value of BTC price is also given in the dataset along with the Market Capitalization of Bitcoin for each day. We will perform exploratory analysis on this time series data to find out whether the data is stationary or not and if not we will make the data stationary by using differencing methods as that is the prerequisite for any Time Series machine learning algorithm such as ARIMA or SARIMA and will target the least RMSE as evaluation metric in ARIMA model. Similarly, For Deep Learning approach we will use traditional RNNs or LSTMs which are widely used for sequential data modelling like Time Series or Textual data for NLP tasks but we have vanishing gradient issue in traditional RNNs in case of deep network which is solved by Long Short Term Memory (LSTM) or Gated Recurrent Units (GRU). According to (Colah’s Blog, 2015) “The Problem of Long-Term Dependencies: Generally, Even Humans look at the very recent information to take action on the future work. For example, a NLP model which is created to predict the next word in the sentence by looking at the previous word. In case we are interested in predicting the last word in ‘The kite is flying in the …’ Sentence. It is obvious that the next word will be sky. In cases like these, where the distance between the relevant knowledge and the place that it’s needed is small, RNNs are capable of learning past information to make predictions Butx there are some cases where we need to have more context. Consider trying to predict the last word in the text “I grew up in India… I speak Hindi fluently.” Recent information suggests that the next word would be the name of a language but if we want to filter on which language, we need the context of India. It’s entirely possible for the gap between the relevant information and the point where it is needed to become very large.

Unfortunately, as this distance becomes huge, RNNs becomes incapable in learning due to Gradient Vanishing.

Problems with RNN:

1. The problem of vanishing and exploding gradients.
2. Vanishing Gradient don’t allow learning to take place in the network.
3. Vanilla RNNs are only able to learn short – term dependencies in practice.

Since the date we will be dealing with is a Time Series data where we don’t know that how many days of records to use as previous time stamp to predict next record which is basically a hyper parameter which we will have to tune according to the outcome. Though to deal with the issues in Vanilla RNN, many attempts/research happened so Vanilla RNNs can also learn the deep/long dependencies and the most popular and successful tweek in RNNs is known as Long Short Term Memory (LSTMs).

According to (Colah’s Blog, 2015) “Long Short Term Memory networks usually called LSTMs are a special kind of RNNs which are capable of learning long term dependencies. They were introduced by [Hochreiter & Schmidhuber (1997)](http://www.bioinf.jku.at/publications/older/2604.pdf), and were refined and popularized by many people in following work. They work pretty well on a large variety of problems, and are now widely used.

At the end we will compare the results of each and every model we built which includes Traditional Regression, SVM Regressors, Random Forest Regressors, Sequential Modeling experts like ARIMA and RNNs and LSTMs.

# Expected Outcome

This research is focused on getting a robust predictive time series or deep learning model with a better accuracy of forecasting, which can help in predicting the price of Bitcoin so one can make better decisions for future investment and plan accordingly. Best results are expected for the closing price forecasting of the very next day. The research will also help in finding the best technique out of Machine Learning or Deep Learning approaches and compare different models like basic regression, SVM and Random Forest Regressors, Auto Regressive Integrated Moving Average (ARIMA) and Recurrent Neural Networks like LSTM.

# Required Resources

The Research implementation will be done using Python.

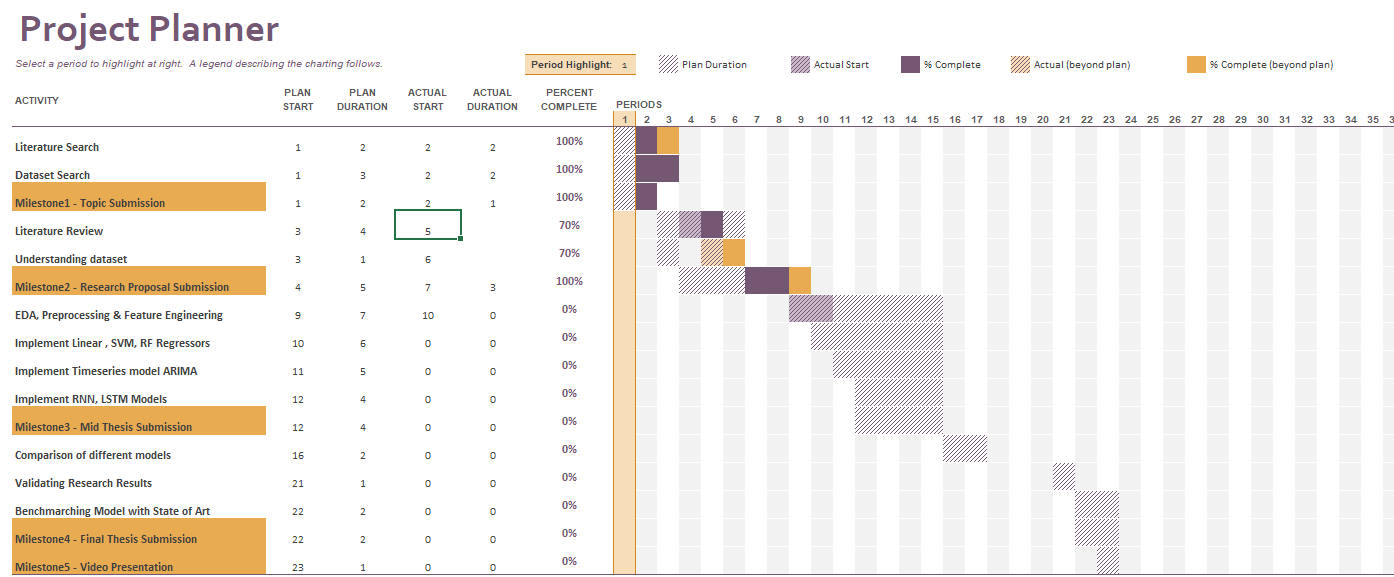
Software/tools requirements:

1. Conda Environment with Jupyter Notebook
2. Data Analysis and preprocessing Packages like Pandas, Matplotlib, Seaborn, scikit-learn.
3. Time Series Analysis Libraries like statsmodels, Facebook Prophet, Pycaret.
4. Deep Learning libraries like Keras with tensorflow as backend.

Hardware requirements:

1. Minimum of 4 GB RAM.
2. Processing power of Intel(R) Core(TM) i5-7020U CPU @2.30GHz
3. Google Colab in case of requirement of GPUs or TPUs

# Research Plan



# References

Balfagih, A. M. and Keselj, V. (2019) ‘Evaluating Sentiment C1assifiers for Bitcoin Tweets in Price Prediction Task’, *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, pp. 5499–5506. doi: 10.1109/BigData47090.2019.9006140.

*Block: 1 | Blockchain Explorer* (no date). Available at: https://www.blockchain.com/btc/block/00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048 (Accessed: 13 June 2021).

Blockchain (2019) *Blockchain Explorer - Search the Blockchain | BTC | ETH | BCH*, *Politics, Gender, and Concepts*. Available at: https://www.blockchain.com/explorer?utm\_campaign=dcomnav\_explorer%0Ahttps://www.blockchain.com/btc/tx/4410c8d14ff9f87ceeed1d65cb58e7c7b2422b2d7529afc675208ce2ce09ed7d (Accessed: 13 June 2021).

Colah’s Blog (2015) *Understanding LSTM Networks -- colah’s blog*. Available at: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (Accessed: 28 April 2021).

Demir, A. *et al.* (2019) ‘Bitcoin Price Prediction Using Machine Learning Methods’, *1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings*, pp. 144–147. doi: 10.1109/UBMYK48245.2019.8965445.

Felizardo, L. *et al.* (2019) ‘Comparative study of Bitcoin price prediction using WaveNets, Recurrent Neural Networks and other Machine Learning Methods’, *BESC 2019 - 6th International Conference on Behavioral, Economic and Socio-Cultural Computing, Proceedings*. doi: 10.1109/BESC48373.2019.8963009.

Gartner (2021) ‘What CFOs Are Saying About Bitcoin’, (March), pp. 1–10.

Gupta, T. (2017) *Bitcoins | What are Bitcoins? | How to earn Bitcoin?*, *Google Blogger*. Available at: https://tambitcoins.blogspot.com/2017/06/what-are-bitcoins-how-to-earn-bitcoin.html (Accessed: 25 April 2021).

Hansun, S. (2013) ‘A new approach of moving average method in time series analysis’, *2013 International Conference on New Media Studies, CoNMedia 2013*. doi: 10.1109/conmedia.2013.6708545.

Hashish, I. A. *et al.* (2019) ‘A Hybrid Model for Bitcoin Prices Prediction using Hidden Markov Models and Optimized LSTM Networks’, *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2019-Septe, pp. 721–728. doi: 10.1109/ETFA.2019.8869094.

Indulkar, Y. (2021) ‘Time series analysis of cryptocurrencies using deep learning fbprophet’, *2021 International Conference on Emerging Smart Computing and Informatics, ESCI 2021*, pp. 306–311. doi: 10.1109/ESCI50559.2021.9397004.

Li, L. *et al.* (2019) ‘Bitcoin options pricing using LSTM-Based prediction model and blockchain statistics’, *Proceedings - 2019 2nd IEEE International Conference on Blockchain, Blockchain 2019*, pp. 67–74. doi: 10.1109/Blockchain.2019.00018.

Li, X. and Wu, W. (no date) ‘A Blockchain Transaction Graph based Machine Learning Method for Bitcoin Price Prediction’, pp. 1–10.

Maleki, N. *et al.* (no date) *Bitcoin Price Prediction Based on Other Cryptocurrencies Using Machine Learning and Time Series Analysis*.

McNally, S., Roche, J. and Caton, S. (2018) ‘Predicting the Price of Bitcoin Using Machine Learning’, *Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018*, pp. 339–343. doi: 10.1109/PDP2018.2018.00060.

Mittal, A. *et al.* (2019) ‘Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data’, *2019 12th International Conference on Contemporary Computing, IC3 2019*, pp. 4–9. doi: 10.1109/IC3.2019.8844899.

Mudassir, M. *et al.* (2020) ‘Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach’, *Neural Computing and Applications*, 6. doi: 10.1007/s00521-020-05129-6.

Raju, S. M. and Mohammad, A. (2020) ‘Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis’.

Rathan, K., Sai, S. V. and Manikanta, T. S. (2019) ‘Crypto-currency price prediction using decision tree and regression techniques’, *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, 2019-April(Icoei), pp. 190–194. doi: 10.1109/icoei.2019.8862585.

Reddy, L. S. and Sriramya, P. (2020) ‘A research on bitcoin price prediction using machine learning algorithms’, *International Journal of Scientific and Technology Research*, 9(4), pp. 1600–1604.

Sin, E. and Wang, L. (2017) ‘Bitcoin Price Prediction Using Ensembles of Neural Networks’, *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 666–671.

SRK (2021) *Cryptocurrency Historical Prices | Kaggle*, *Kaggle*. Available at: https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory (Accessed: 13 June 2021).

Taylor, S. and Letham, B. (2017) *Prophet: forecasting at scale - Facebook Research*, *Facebook*. Available at: https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/ (Accessed: 13 June 2021).