National College of Ireland
BSc in Computing
2014/2015

Sumit Tiwari
x15046885
X15046885@student.ncirl.ie

Bitcoin Modelling Using Data Mining Techniques

Technical Report

Table of Contents

Executive Summary....5

1 Introduction....6

1.1 Project Motivation and Background....6
1.2 Project Statement....7
1.3 Project Aims ,Objectives and Contributions....7
1.4 Technologies Used....8
2. Literature....9

2.1 Literature Review....9
3. Scientific Methodology Approach Used....15

3.1 Scientific Methodology Approach Used....15
4. Requirement Specifications and Process Flow Design....18

4.1. Requirements Specifications....18
4.1.1 Functional requirements....18
4.1.2 Use Case Diagram....18
4.1.3 Data requirements....20
4.1.4 Environmental requirements....20
4.1.5 Usability requirements....20
4.1.6 Performance/Response time requirement....21
4.1.7 Availability requirement....21
4.1.8 Recovery requirement....21
4.1.9 Robustness requirement....21
4.1.10 Security requirement....22
4.1.11 Reliability requirement....22
4.1.12 Maintainability requirement....22
4.1.13 Portability requirement....22
4.1.14 Extend-ability requirement....22
4.2 Project Process Flow Design, Dataset and Data Preparation....23
4.2.1 Project Process Flow Design....23
4.2.2 Data Visualization,....24
4.2.3 Data Preparation and Feature Section....27

5. Evaluation and Results of Bitcoin Models..... 30
5.1 Introduction.....30
5.2 Implementation,Evaluation and Results of Prediction Models.....30
5.2.1. ARIMA Model..... 30
5.2.1.1 Implementation..... 30
5.2.1.2 Evaluation and Results.....31
5.2.2 Time Series(Prophet) Model.....37
5.2.2.1 Implementation..... 37
5.2.1.2 Evaluation and Results.....38
5.3 Linear Regression Model.....41
5.3.1 Implementation.....41
5.3.2 Model Evaluation and Results..... 41
5.4 Multiple Regression..... 42
5.4.1 Implementation.....42
5.4.2 Model Evaluation and Results..... 42
5.5 Polynomial Regression.....42
5.5.1 Implementation.....42
5.5.2 Model Evaluation and Results..... 45
6. Conclusions and Recommended Future Work.....46
Acknowledgement..... 47
References.....48
6 Appendix.....52
6.1 Project Proposal..... 52
6.1.1 Introduction..... 52
6.1.2 Motivation and Background.....52
6.1.3 Project Objectives and Goals.....54
6.1.4 Technical Approach.....56
6.1.5 Technical Details.....57
6.1.6 Evaluation..... 58
6.2 Project Plan.....58
6.3 Monthly Journals..... 59
6.3.1 September Journal..... 59

6.3.2 October Journal.....60
6.3.3 November Journal..... 62
6.3.4 December Journal..... 64
6.3.5 January Journal.....66
6.3.6 February Journal.....67
6.3.7 March Journal.....68

Executive Summary

In recent years bitcoin has been attracted eminent attraction all around world in the area of cryptocurrency and its unique peer to peer transaction system. Analyzing and forecasting is most common task for data scientists that helps organizations to improve there business strategies.

This report is built around the fact that in bitcoin trading and analysis, the factors underlying are basically the traditional price predictions using the data mining technique.The project aim to gather data on bitcoin ,analyse and forecast in order to see the fall or rise in bitcoin price.The bitcoin Dataset used is bitcoinmarketcap. This report intend to help the reader and understand what approach and methodology used to complete different milestones.This report summarizing all the aspect used throughout this project.

This is done using different techniques and algorithm to predict and forecast the rise or fall of bitcoin with the use of Crisp-DM methodology.The techniques used in the projects are Time Series ARIMA ,Time Series(Facebook prophet API), Linear Regression and and Multiple Regression.

1 Introduction

Bitcoin is a new virtual currency also known as electronic payment system or cryptocurrency. It is the world's most valuable cryptocurrency, introduced by Satoshi Nakamoto in 2009.The two parties can make bitcoin transaction over internet with no middle man, by installing a bitcoin software called bitcoin client/wallet. Bitcoin transaction are posted in blocks to an open ledger called Block-chain. Individual users have a unique block-chain key. There is reasonable transaction cost compare to the other banks. Using bitcoin's has its own pros and cons. User identity and money is hidden. They do not have to disclose his/her personal identity. It has been related to its own conspiracies and has made people a little too skeptical towards investing in it.

1.1 Project Motivation and Background

Bitcoin is a peer-to-peer based virtual currency which is currently providing a digital alternative to the different currencies in the world. Its concept of having a single value throughout is what makes it popular among its owners and traders. Considering that bitcoin follows a decentralized model, it has often been linked to the dark web. Another reason for bit-coins being such a fascinating reason for research is that no researcher has yet been able to decode the social science behind people being open and accepting it as a currency despite not having any physical significance.

I have chosen this topic due to my interest in the concept of cryptocurrency and bitcoin, which has become the topic of interest for many investors (despite the cons) over the past few years. Some researchers have gone as far as to say that investing in this is as safe as investing in real estate but with bigger and quicker returns.

This research is based on the research that bitcoin prices have only gone up since their existence (almost doubling in value every year approximately on a minimum), and now, when they are gaining popularity

and are increasing in value, their transaction and its data has become a relevant part of the financial industry to understand the sudden boom of a currency which didn't even exist a few years ago.

1.2 Project Statement

- Project Statement : To Forecast the bitcoin price
- Project Statement: To evaluate the performance of forecasting models.
- Project Statement: Comparison of model developed .

1.3 Project Aims ,Objectives and Contributions

The main Objectives: The Literature review on bitcoin, customer involved with bitcoin transaction with the aim of identify whether there is a need of prediction models,challenges encountered by bitcoin customers and impact of bitcoin globally.

Contribution: The Results of the revised literature and identified gaps in the body of knowledge (What is missing).

The scope of this project is to develop prediction models to understand bitcoin prediction factor and improve management of bitcoin globally by analyzing and visualizing data and provide clear picture of the bitcoin data. The main datasets required will be supervised (kaggle) and unsupervised datasets which will be extracted from block-chain and twitter etc.

The main objectives are to develop prediction models (ARIMA, Time Series,linear and multiple regression). In order to develop this models first data pre-processing (data cleaning) will be conducted to remove (noise, null vales, duplicates values etc.). The main methodology approach to be used will be CRISP-Dm. Once the data processing been performed the next project goal will be to revisit and redesign the architectural and technical design in order to address any new information. Once this been completed the implementation of prediction models will start's and the computation of simulation model will be done. In addition,

evaluation and validation of the developed prediction models will be conducted and the results will be analyzed and visualized.

The main contributions of this project will be fully developed and evaluated prediction models. The results of the evaluation are another contribution.

1.4 Technologies Used

The main technologies used during implementation (chapter 5) are@

- R: R language will be used to clean and visualize data according to the project requirements. It is an integrated suit for data manipulation, calculation building models and visualization.
- Python: To support with the building models and visualizations
- Tableau: Tableau will be used to create dashboard and visualize .

The rest of the technical report is structure as follows. Chapter 2 presents the reviewed literature. Chapter 3 introduces the scientific methodology approach used and .Chapter 4 presents Requirement Specifications and Process Flow Design. Chapter 5 presents Evaluation and Results and Chapter 6 introduces Conclusion and Future work.

2. Literature

2.1 Literature Review

This research is mainly focused on predicting the Bitcoin prices so as to execute profitable trades, we also show that Bitcoin price data exhibit desirable properties such as constancy and mixing. The existing models which can predict the outcome of the Bitcoin predicting with various methods have been failing to predict the exact outcome of the trades such as ARIMA models and also lack a probabilistic interpretation **(M. Alrasheedi, 2012).** In spite of these limitations the research is mainly focusing on two important theoretical framework, they are predicting and trading the state Bitcoin charges, second, bringing high accuracy in predicting the trading to achieve high return on Bitcoin Investment.

At present modern era it is a common knowledge and fact that data flow of information is continuous as indexed by timestamps around the world from stock market data to social media, from search engine to webpage analytics (**Muhammad J Amjad, 2017**). This research mainly focuses on real-time accurate data forecasts using predicting algorithms, which can adapt to different structure that may be present in various domains. According to **Ciaian** (**2014**), the relationship between the price of Bitcoin and the demand and supply fundamentals of some global financial indicators, and Bitcoin's attractiveness for investors. The value of Bitcoin and its relationship to different financial data was examined by **van Wijk** (**2013**). The authors were able to conclude that the euro-dollar exchange rate have a significant impact on the price of Bitcoin in the short run but only has a significant impact on the value of Bitcoin in the long run. Also, the researchers concluded that other variables, like the dollar-yen exchange rate have no statistically significant effect on the formation of Bitcoin price.

The price returns and volatility changes in Bitcoin market were studied by **Bourie et al**. (**2016**). Furthermore, their analysis shows a negative relation between the US implied volatility index and Bitcoin realized volatility. According to **Kristoufek**

(**2015**), Fundamental factors like usage in trade, money supply and price level were found to be important in the determination of Bitcoin price in the long run.

Works on Financial Data Forecasting:

According to **I.M. Yassin et al.** (**2017**), to get an accurate trading data flow of information many companies have worked on several predicting models. Few of the models are summarized as below;

According to **Hafezi R, Shahrabi J, Hadavandi E.** (**2015**), to predict the quarterly price movements of the DAX German stock exchange index over the time period of 8 years used the framework of a four-layer multi-agent. This framework is also called as Bat Neural Network Multi Agent System (BNNMAS), which consists of several autonomous decision makers who are recognized as agents of the trade. This type of framework has been beneficial and a reliable tool for stock price predictions for long period.

According to **Marček D** (**2004**), the company Vahostav introduced a tool called Fuzzy Neural System (FNN). This structure consists of an input layer, fuzzy neurons and an output unit. This type of transfer function and formulation can affect the accuracy of the FNN. It also reflects marginal effect with 63.5 Root Mean Square Error [RMSE]. According to **Wah B W, Qian M L.** (**2006**), a Recurrent Neural Network (RNN) was used to predict the prices of several blue-chip stocks namely Citigroup, International Business Machine (IBM) and Exxon-Mobil. It was proven as an excellent model for Citigroup, IBM and Exxon-Mobil respectively over a 10-day prediction horizon.

According to **Shah V H.** (**2007**), introduced the machine learning techniques to predict the surge and depression of stock prices. The model which used both fundamental and technical analysis to forecast and made recommendations used called as hybrid Support Vector Machine (SVM). To enhance the training process and get more accurate result the system used Boosting Algorithm. By this the result suggest that the proposed SVM classifiers had managed to obtain MSE of 0.467 and 0.6 respectively. According to **Di X** (**2014**), in predicting stock trends, a tree-based algorithm was used to feature selection and technical indicators. The

forecasting was predicted using Support Vector Machine with Radial Basis Function (RBF). The reduction appears to help overcome data overfitting with 950 examples used for training and prediction was done on next 50 days to validate the accuracy of prediction. The results achieved 56% accuracy on predicting the next day price trend noting that accuracy can be improved to 70% by including additional sentiment data to the prediction. According to **Almeida J** (**2016**), an Artificial Neural Network (ANN)-based Bitcoin forecasting model was constructed to predicts its price over a one-day prediction horizon using its previous-day price and trading volume. According to **McNally S** (**2016**), explored several models (RNN, Auto-Regressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM)) to predict Bitcoin prices sourced from the Bitcoin Price Index (BPI). Training was performed using Graphics Processing Unit (GPU) due to the volume of computations required to train the models. The RMSE of the LSTM, RNN and ARIMA models were 6.87%, 5.45% and 53.47% respectively indicating that RNN was the best choice between the three models to perform Bitcoin forecasting. According to **Greaves A** (**2015**), a comparison between several regression models was performed for Bitcoin price prediction. The baseline, linear, SVM and ANN regression models were used to predict Bitcoin's future prices based on past data. The results indicated that the linear regression method was the most suitable for prediction.

According to Barber et al. (**2012**), for better understanding the rise and adoption of Bitcoin a detailed analysis of the major factors to the success Bitcoin have been summarized as follows;

No central point of trust: There are no particular approach or methods for trading and pricing. Bitcoin is completely distributed in nature and is able to operate entirely without any central authority, the blockchain and mining process allow for a majority vote for dispute resolution.

Incentives and economic system: The clever part behind the idea of Bitcoin is the incentive for users to become miners which allows anyone to participate in the transaction process and be rewarded for the participation with Bitcoin's.

Predictable supply: Bitcoin currency is only added to the available supply as a reward for mining process. The total supply and transactions can be defined mathematically and also can predict the patterns.

Divisibility and fungibility: The currency can be easily divisible up to the minimal unit of eight decimals. This allows users to transact nearly any amount and combine any amount.

Versatility and openness: The development of Bitcoin is completely open; the source code is developed using an open-source license and has hundreds of developers contributing to the software.

Scripting: The Bitcoin blockchain allows for the inclusion of simple scripts, which can be used to create complex transactions and financial services such as escrow and dispute mediation.

Transaction irreversibility: Transactions within the Bitcoin network are irreversible once seen by other peers on the network, and once added to the blockchain are practically impossible to reverse without a 51% attack.

Low fees: The Bitcoin network has extremely low transaction fees with no additional costs for international transfers, which in addition to transaction irreversibility is appealing to merchants.

Numerous implementations: There are numerous implementations of Bitcoin, supporting all major operating systems and devices, in addition to cloud-based services making access readily available. In this research we mainly focus on Time Series Analysis model to predict the trading and pricing of Bitcoins.

TIME SERIES ANALYSIS:

By acquiring the abundance knowledge of Bitcoin and its models of prediction, this section is going to mainly focus on time series analysis by using the Arima model. According to **V. S̨. Ediger** (**2007**), these models rely on linear assumptions regarding the data. Due to the highly non-linear nature of the Bitcoin price it was not expected to perform well. The data was changed to stationary to make it more consistent and meet the necessary assumptions. A time series can be considered in the same regard as a signal. The intention was to investigate

using ARIMA as an input to the neural network model. ARIMA analysis was used to analyze the correlation of variables in dataset on temporal scale.

Univariate ARIMA Model and Forecast:

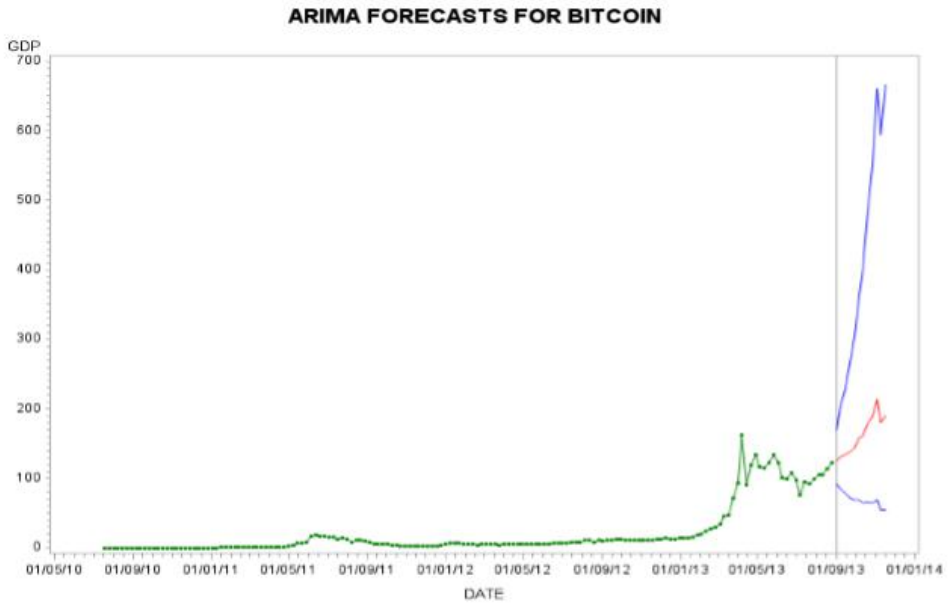
The below algorithm model is going to explain about final univariate model.

According to **Geraskin, P.** (**2010**), the below said algorithm and time series model explains the variables in dataset.

The final univariate ARIMA model is as follows:

Let wt = Δln(yt).

In order to obtain the above model, an ARIMA procedure in SAS and obtained an initial SBC .A repetition of the ARIMA procedure without an intercept resulted in a 10 higher SBC, so the intercept was retained in the model. Observations of the trend and correlation analysis showed likely correlation at a lag of 3 weeks, 10 weeks, 20 weeks, 30 weeks, and 40 weeks. Testing these above said combinations with the lags as both AR and MA shows that a model with AR term of lag 3 and MA term of lag 40 provides the lowest SBC. The below is an example of the prediction of the forecast. Below is the diagram of Forecast model of Bitcoin from the year 200 to 2014.



Conclusion:

Bitcoin values would be driven yy their worth as a medium of exchange. Bitcoin is a successful cryptocurrency, and it has been extensively studied in fields of economics and computer science. In this study, we analyze the time series of

Bitcoin price with ARIMA model in addition to macro-economic variables and address the recent highly volatile Bitcoin prices. Given the data of the entire time range, experimental results show that the ARIMA model learned with the selected features effectively describes processes of Bitcoin log price and log volatility. Adoption of rollover framework experimentally demonstrates the predictive performance of Arima is better than other models of prediction. Variability of Bitcoin must be modeled and predicted more appropriately. This can be achieved by adopting other extended computing language, such a contribution will help to enrich Bitcoin time series analysis in addition to opening various fields for new and advanced developments.

3. Scientific Methodology Approach Used

3.1 Scientific Methodology Approach Used

The methodology will be used in this projects will adhere to the path process of the CRISP-dm and following this path will allow for this project to be completed in full. Below is an image and a brief explanation of the CRISP-dm is:

Methodology:-

This research uses the 'Cross industry standard process for data mining' (CRISP-DM) methodology. Chapman et all (2000) stated that the CRISP-DM is vertical process which open up the six level break downs which we have used during this research. The CRISP-DM is been altered as per this research require to obtain the data into figures. The altered foam can be seen between two figure indicate by arrow signs. First one is data aggregating, data cleaning and data extraction. Second one is data modelling. Third one is execution of forecasting models. Forth one is assessment process of the forecasting model. Fifth one is expected value of forecasting model. Sixth and the last one is to achieve the research objective followed by the fifth level expected value of forecasting model.

As per (Shumway and Stoffer, 2006) The time series data can be processed by using time series R package (zero t series). this time series data is uni-variate as it carry on the single attribute that was used in forecasting model. After cleaning up the data, it will be ready to analyze into the forecasting model. The time series data has comprise in the foam of plots to analyze the trend, noise and seasonality. Data has been analyzed for linearity and non stationary using qqline plot, qqnorm plot and Auto-Correlation function (ACF) (Theiler et al , 1992).

The forecast package of R consist of all the forecasting models used for this research which is capable of processing the bitcoin data time series data as per (Brockwelll and Devis, 2016). To call a forecast model foam the package their specific function names are needed to pass.

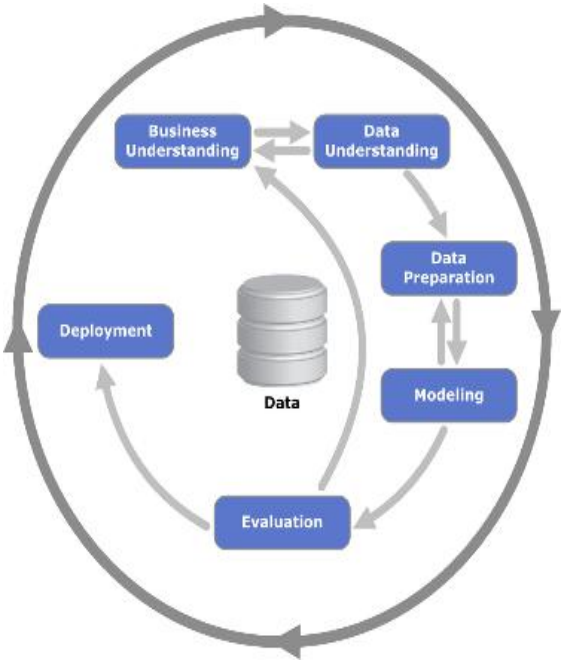


Figure 1: Crisp-dm Model

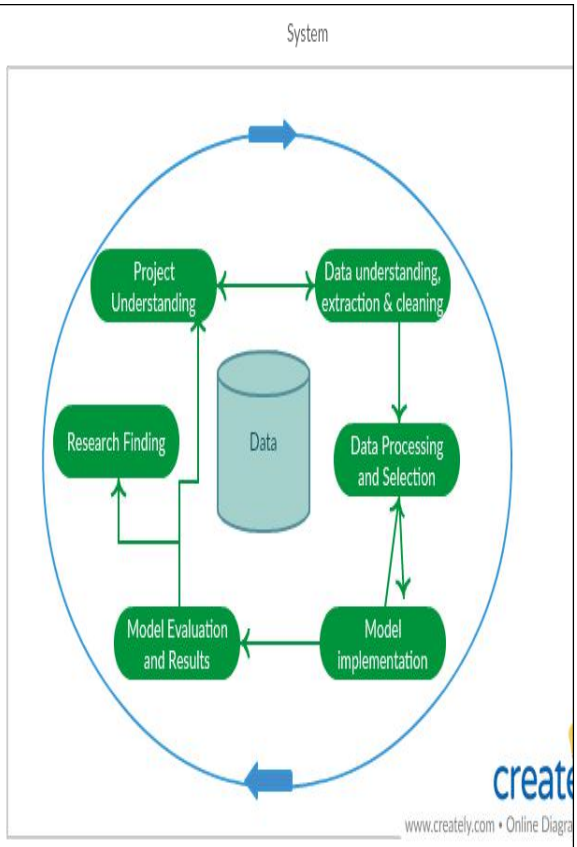


Figure 2: Modified Crisp-dm Model

Business Understanding: In the beginning of this phase focuses on understand the project aim's and scopes from a business point of view and then initialise this knowledge into the data mining problem.

Data Understanding: In this phase, initialise the data gathering from relevant sources and get familiar with datasets, data quality and problems with the data or any hidden information about the data. Data understanding is slightly differing then business understanding.

Data Preparation: This part of the phase is cover all the activities conduct to prepare the finalise data such as cleaning data, construct data, integrate data and data formation, removing duplicate values/record and transformation of data for modelling phase.

Modelling: In this phase, select the data modelling techniques and applied to the data. Generate test design model to test the model. There are different types of technique for same data mining problem.

Evaluation: In Evaluation phase ,the model been build to get the result from the data for the project and understand the outcomes of the data.

4. Requirement Specifications and Process Flow Design

4.1. Requirements Specifications

4.1.1 Functional requirements

Learnability: User should be able to learn prediction model after interacting with developed models ie ARMIA,Time Series ,Linear and Multiple Regression.

User-ability: The developed models results will be visualized using dashboard. The developed dashboard should be easy to use in terms of user ability. The developed models (prediction and simulation) should provide users with accurate prediction result.

4.1.2 Use Case Diagram

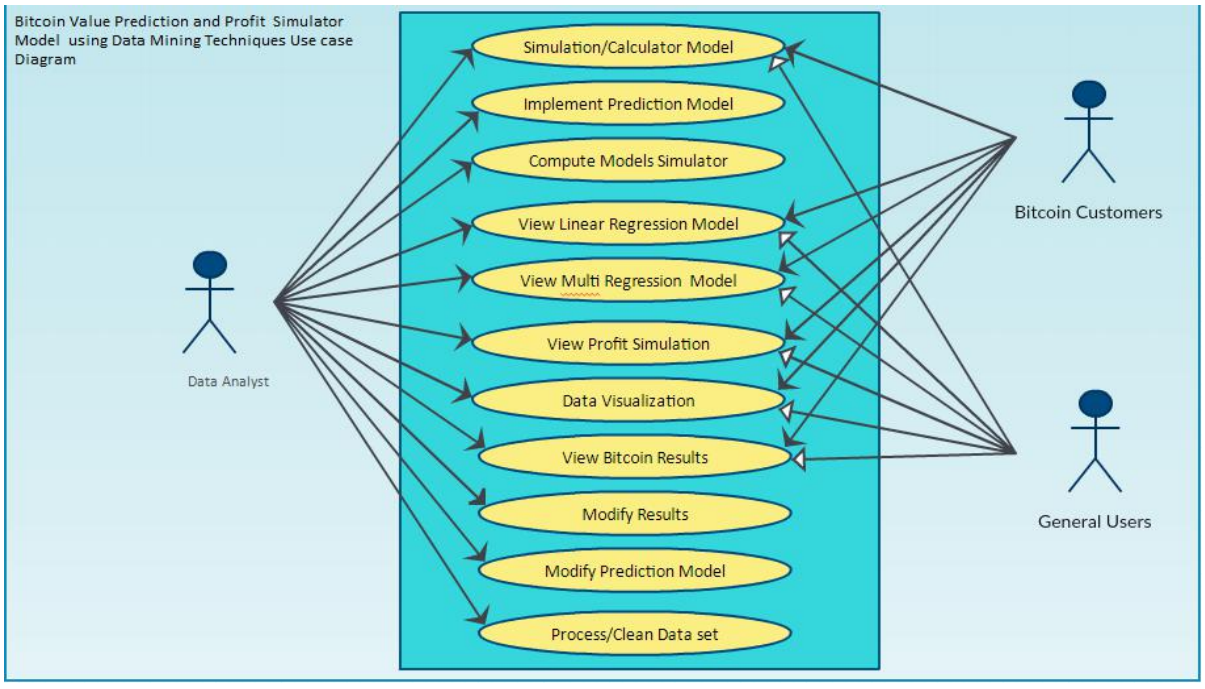


Figure 3 : Use case diagram

4.1.2Requirement 1 <User Input>

4.1.2.1 Description & Priority

The user input should be valid for the code to give a logical output. If integers are not used for values and characters are, the input is invalid.

4.1.2.2 Use Case

Each requirement should be uniquely identified with a sequence number or a meaningful tag of some kind.

Scope

The scope of this use case is to understand how the functional requirements of the proposed system will be handled.

Description

This use case describes the basic procedure followed to tackle all the requirements of the system.

Use Case Diagram

The use case diagram shows clearly the three groups of people who will interact with the system. The Data analyst, bitcoin customer and general users.

Flow Description

Precondition

The system is in initialization mode, where its practical state is at rest but in the back-end, the calculations for any updates market values are already setup and awaiting variable values such as amount and time period.

Activation

This use case starts when an <Data Analyst> makes a calculation based on a logic in the system. He can modify the logic based on market requirements and this is a liberty only the logic designers are granted. The business user and the normal user differ on the grounds that the normal user doesn't have an official bitcoin trading account as compared to the business user.

Main flow

- The system identifies the flow of events.
- The <Normal User> enters values and gets outputs based on those.
- The system can easily view the results, input values to get the output.
- The <Analyst> sets up the logic to help the user view their requirements.

Exceptional flow

E1: <The user enters invalid number>

- The system throws an exception saying “enter a valid amount”
- The <User> re-enters the correct value: integer
- The use case continues at position 3 of the main flow.

Termination

The system presents the next step where the users can see their output.

Post condition

The system goes into a wait state

4.1.3 Data requirements

The data requirement for this project, will be extracted from all the possible platforms, supervised Dataset will be retrieve from Kaggle, Block-chain and unsupervised from URLs and social Media APIs i.e. Twitter, Facebook etc.

4.1.4 Environmental requirements

To Execute this project, the Tableau Desktop suit will be used.Python 2.7 or 3.4 will be required with the necessary packages installed. R studio will be used to import and clean the supervised and unsupervised datasets.

4.1.5 Usability requirements

This section has been sanctioned to understanding the problem or the Pain Points which led to this research being an ideal solution to. This section holds a critical spot in the execution of the whole research as I am defining what problems are being focused on for this research and how the client placed them

in their requirements. I would like to specify that the client here has been said to be the people who might use this application model and the people who are already using a similar application for the same purpose or similar purposes.

New Users Perspective: For this category of users, the idea of a simulator is to investigate the prospects of their investment is new and they have various ideas of how this might be achievable. A client X may see this simulator and its visuals in the form of a growth chart where all he enters is an amount and a duration for investment and he sees how his figures hit the roof in terms of numbers and in terms of the peak of a graph.

4.1.6 Performance/Response time requirement

The response time is the key to high performance of the application in terms of user satisfaction. If the models take too much time to load the graphical representations or doesn't come up with at least a decent low graphic requirement image on the instant of a click so that the application gets a few Nano seconds in to load up the actual image, the users might perceive the application as slow.

4.1.7 Availability requirement

The application is an open source one and it serves a basic functionality, it is desired to have a high availability but going for paid high performance cloud services to get a high up time is not something that is needed.It has to available in case of hardware failure.

4.1.8 Recovery requirement

If the running system fails, the developed models should be easy to recover without losing any data.

4.1.9 Robustness requirement

This is limited to a small problem which I can fore see, a person entering invalid values in amount, such as alphabets and invalid dates. These are small problems

and the solutions to these can be hard coded in the form of output replied such as ‘Invalid Input’, ‘Please re-enter the Value’, etc.

4.1.10 Security requirement

Considering the open source nature and not much actual data being gathered from this, this is not much of an issue, though if the application is later sponsored by anyone and we get advertisements, a private cloud could be used to store user’s information which they can use to sign up on our application with their email, phone number, etc. because of data integrity and user’s authentication globally when interacting with developed dashboard.

4.1.11 Reliability requirement

Though the numbers are supposed to be as accurate as possible, the users would be aware of the fact that all investments are subjected to market risks and they will have to read any documents before investing.

4.1.12 Maintainability requirement

The application will have to be maintained based on how the results are different compared to the expected outcomes, this means very minor changes in the logic of the calculator and not much on the front-end.

4.1.13 Portability requirement

The application should not be limited to use on a single device, hence, it should have minimum network requirements to load on minimal service of network on browsers which are not very supportive of high end applications as well.

4.1.14 Extend-ability requirement

The code in the back-end will always have a scope for improvement and extended features, but no such basic requirements are needed.

4.2 Project Process Flow Design, Dataset and Data Preparation

4.2.1 Project Process Flow Design

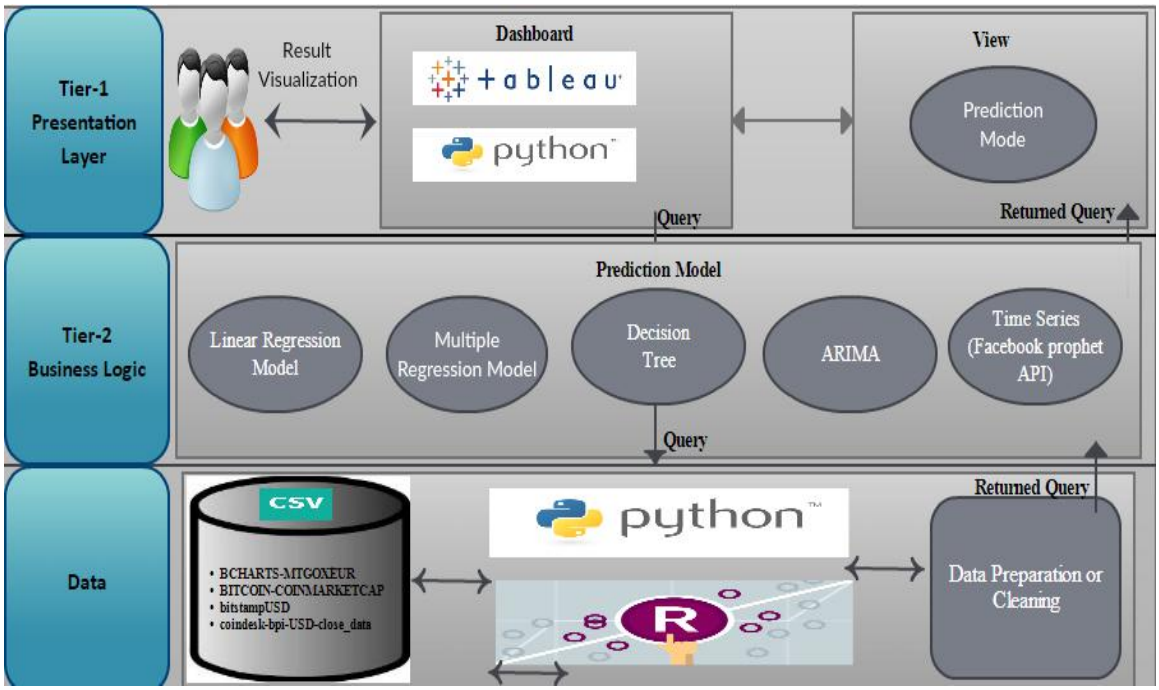


Figure 4: Project Flow diagram

Figure 4: Architecture diagram explains the process of how the system will work.

Tier 1: Also called as Presentation Layer. In this tier user will be able to communicate with developed models to get the result visualization through Dashboard which will be implement by using Tableau, Python etc. Through this dashboard user will be able to see the prediction and simulation Models. The reason to use this dashboard is to provide the clear picture of Prediction and Simulation to the users. Users will enter their query/requirement into Dashboard which is Tier1 then Tier will send the user query to Tier 2 according to user requirement.

Tier 2 or Application Layer: In this Tier Prediction and Simulation will be developed. In this we will use 2 types of regression called Linear and Multi. Both will help in determine the prediction of Bitcoin rise/ fall in price. Not only this, a

statistical model will be used to analyse the rise and fall of the bitcoin. It will use calculator and a stimulation to predict the profit or loss.

Data Access Layer: In this Layer the datasets will gathered, supervised dataset will be retrieve from Kaggle and unsupervised from twitter/Facebook. After gathering the Dataset, Data cleaning and Preparation will be done by using Python, R.

The flow of system goes as follow: Tier 1 and Tier 2.

4.2.2 Data Visualization,

Data analysis on the original datasets been done by Tableau.

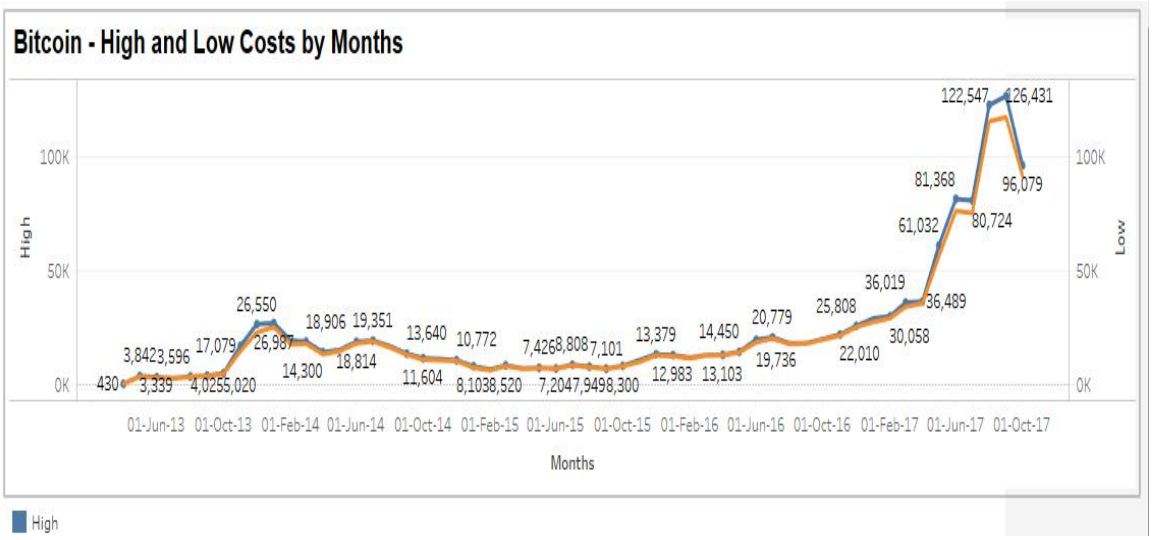


Figure 5: Bitcoin-High and Low Costs by Months

The datasets were analysed to visualize the high and low costs incurred by Months(Figure 5) which shows the high and low rate of the bitcoin by months.From the figure 5 we can see the rise in bitcoin prices by every month.

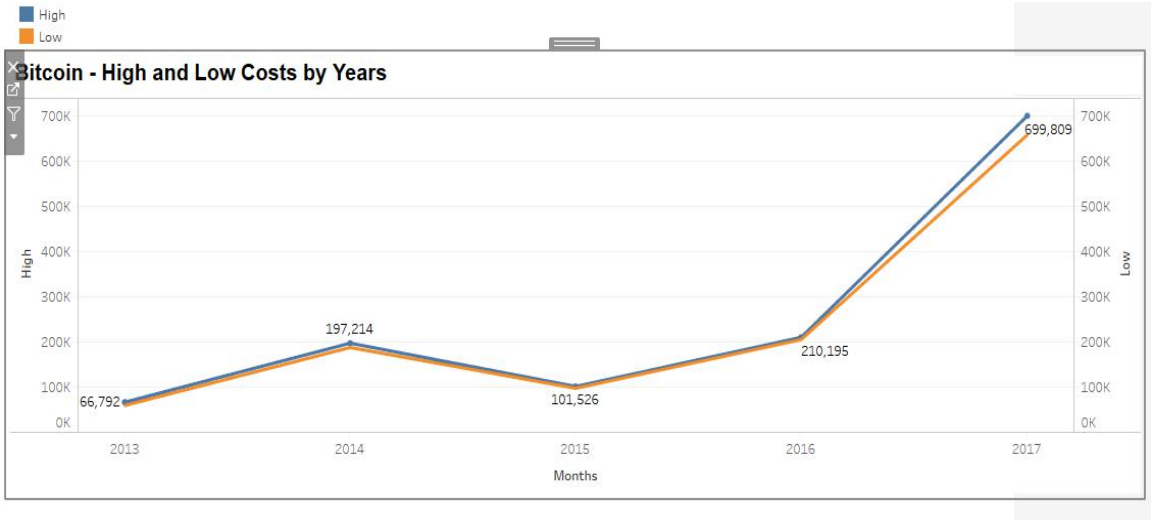


Figure 6: Visualization of Bitcoin High and Low Cost by Years

The datasets were analyzed to visualize the high and low costs incurred by Annum(Figure 6) which shows the high and low rate of the bitcoin by years.From the figure 6 we can see the how bitcoin prices effecting every year, reaching to the top in all crypto-currencies.

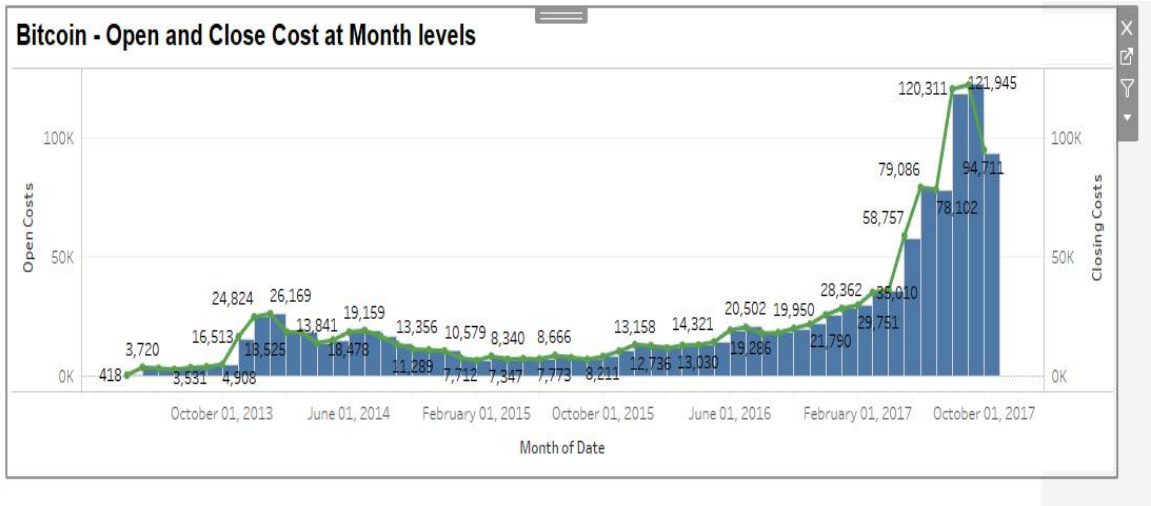


Figure 7 : Bitcoin-Open and Close cost monthly

The datasets were analyzed to visualize the Open and Close costs incurred monthly (Figure 7) which shows the open and close rate of the bitcoin at what rate it was open and closed on monthly basis.

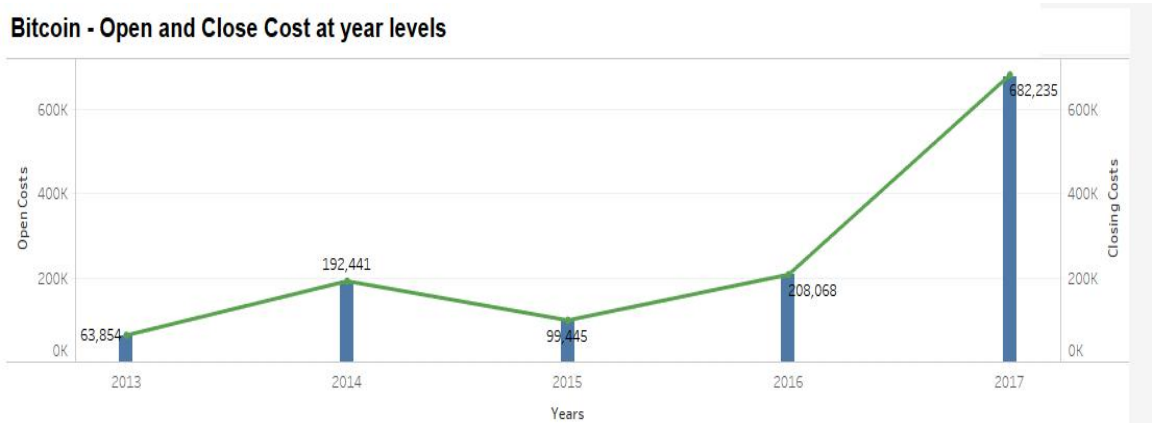


Figure 8: Bitcoin-Open and Close at Year levels

The datasets were analysed to visualize the Open and Close costs level incurred yearly(Figure 7) which shows the open and close rate of the bitcoin at what rate it was open and closed on yearly basis.

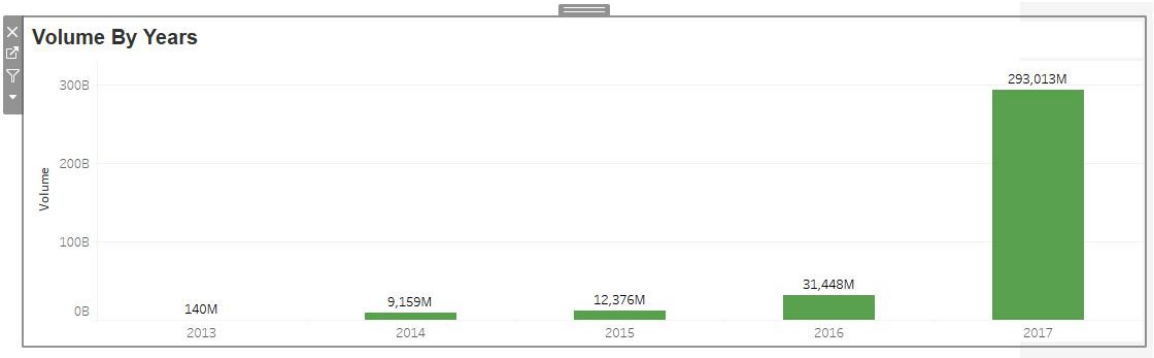


Figure 9: Volume By Years

The datasets were analyzed to visualize the volume costs incurred yearly (Figure 9) which shows volume rate of the bitcoin.From the figures we can see there is

sharp rises in the volume in 2013 it was around 140 m and just after 5 years times reached to the sky around 293013 m .

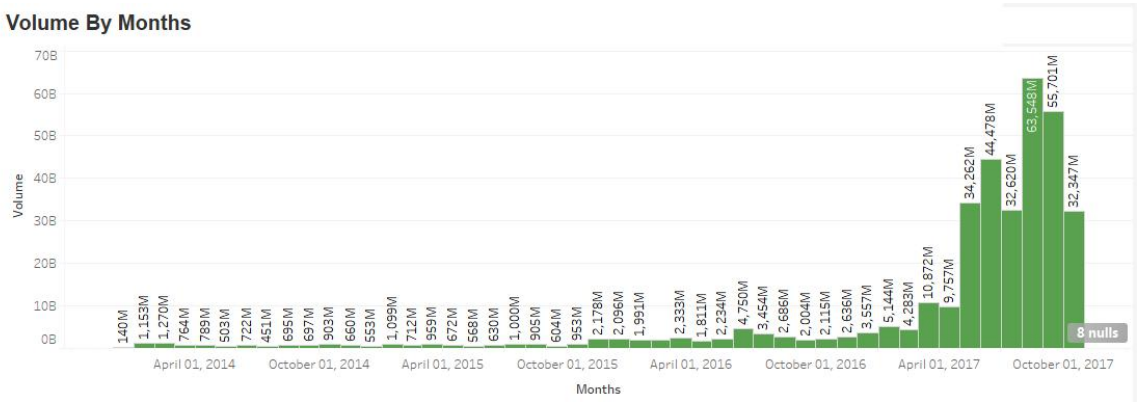


Figure 10: Volume by Month

The datasets were analysed to visualize the volume costs incurred monthly(Figure 9) which shows volume rate of the bitcoin.From the figures we can see from what month of the year bitcoin start booming and never stopped.

4.2.3 Data Preparation and Feature Section

In order to implement this project there is few different datasets have been used and gathered from different website such as Kaggle ,coin-basket.Firstly,the raw data was imported into the analyst application and visualization application for the cleaning and feature selection process. In the early stages some of the R and Python libraries were utilize to get interesting fact and Data visualization done in Tableau.For this project ,few different datasets were used.The cleaning process is removing un-valuable data ,changing attribute and format of the attribute as data was quite small there wasn't much cleaning needed.Once the cleaning process of Dataset is finished.Dataset will be transform, combine and ready for analysis for the feature selection depending on the requirement and the data will

be visualized in form of relevant graphs.In order to implement the ARIMA model R studio and its function been used in order to get the interesting result of the Dataset and to do prediction for the bitcoin.Following snippets of running function showing the importing the data and cleaning process

```
install.packages("zoo")
install.packages("xts")
install.packages("tseries")
install.packages("forecast")
install.packages("fpp2")
install.packages("ggplot2")
install.packages("lubridate")

library("zoo")
library("xts")
library("tseries")
library("forecast")
library("fpp2")
library("ggplot2")
library("lubridate")
library("dplyr")
```

Figure 11: Install packages

Figure 11 ,show that the packages and library installed in R studio.

```
setwd("C:/Users/SUMIT/Desktop/MainProject/data")

bitcoin= read.csv("BITCOIN-COINMARKETCAP.csv", header=TRUE, sep=",", na.strings="NA",

str(bitcoin)#structure of the dataset

summary(bitcoin)
```

Figure 12: Reading Dataset

In figure 12,show that path file was setup and read the data file in R-studio.

Str(bitcoin) give us the information about the data file and its attributes.

Summary(bitcoin) provides the summary of the attributes ie. Mean, Median, Min, and max of the data file.

```
bitcoin_close <- subset(bitcoin[c(1,5)])
str(bitcoin_close)
View(bitcoin_close)
bitcoin_close$Date <- mdy(bitcoin_close$Date)
bitcoin_close <- arrange(bitcoin_close,Date)
bitcoin_close <- bitcoin_close[-c(1,2,3),]
bitcoin_close <- bitcoin_close[-c(1618:1636),]
bitcoin_close <- bitcoin_close[-1]
```

Figure 13: Cleaning Process

Figure 13, show the cleaning and selection process.From the first code of the line removing the unwanted column.Afterward change the format of the date column and rearrange and removing unwanted data from the dataset.

5. Evaluation and Results of Bitcoin Models

5.1 Introduction

The Programming language used for this research is R and Python. Cleaned datasets have been loaded in R and Python using read.csv() function. All the execution part is done in R studio for (ARIMA and Time Series prophet) (IDE) using R language and Linear and Multiple Regression using Python. For processing the data set package like t-series, Zoo, ggplot2, fpp2,numpy ,panda ,glob ,prophet,dplyr and xlsxwriter have been installed by using the installation process. The these package are loaded into memory using library() function.

5.2 Implementation,Evaluation and Results of Prediction Models

5.2.1. ARIMA Model

5.2.1.1 Implementation

Bitcoin data for this project has been gathered from the Kaggle and been used for forecasting ,evaluation and comparison of models.Downloaded dataset has been cleaned using R and removed unwanted columns.For processing, packages and libraries like forecast,tseries,fpp2,ggplot2 etc has been installed and loaded using library() function. Cleaned data are then loaded in R using read.csv() function and implementation in done R studio.The data is then converted into Time Series using ts() function and plot the decompose model to check the trend using decompose() function for graphical representation of time series data.

ARIMA forecasting model has been implemented by using auto.arima() and then forecast() function is applied to forecast the 365 days of bitcoin value prediction.Then the Auto-correlation ACF() and Partial PACF() is applied function to check the data is noisy or not .

5.2.1.2 Evaluation and Results

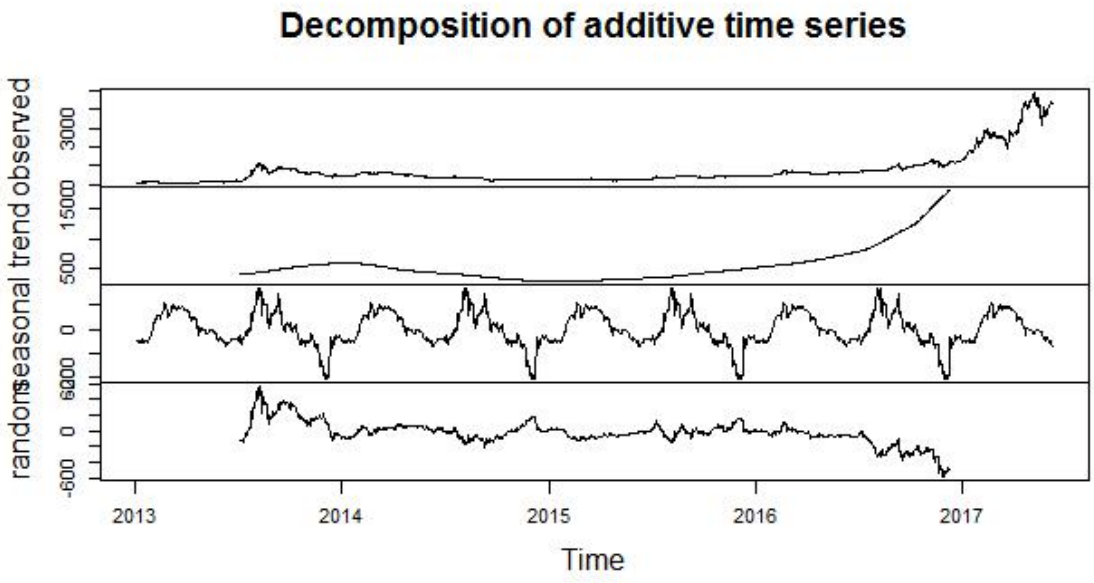


Figure 14: Random and Seasonal Trend

Figure14, shows that random in starting 2013 was steady and bit of flexuation end of 2013. in 2014 to 2015 it was steady. we seen a little fall in the end of 2016. seasonal in 2013 was a bit up and down ad shows the same performance until 2017. Trends in 2013 was average, priced went up in 2014, came to its lowest level in 2015, and has a rapidly growth in 2017. observed was on his lowest in 2013, seen a gradual growth in 2014, 2015 and 2016 was the same and had a dramatically rise in 2017.

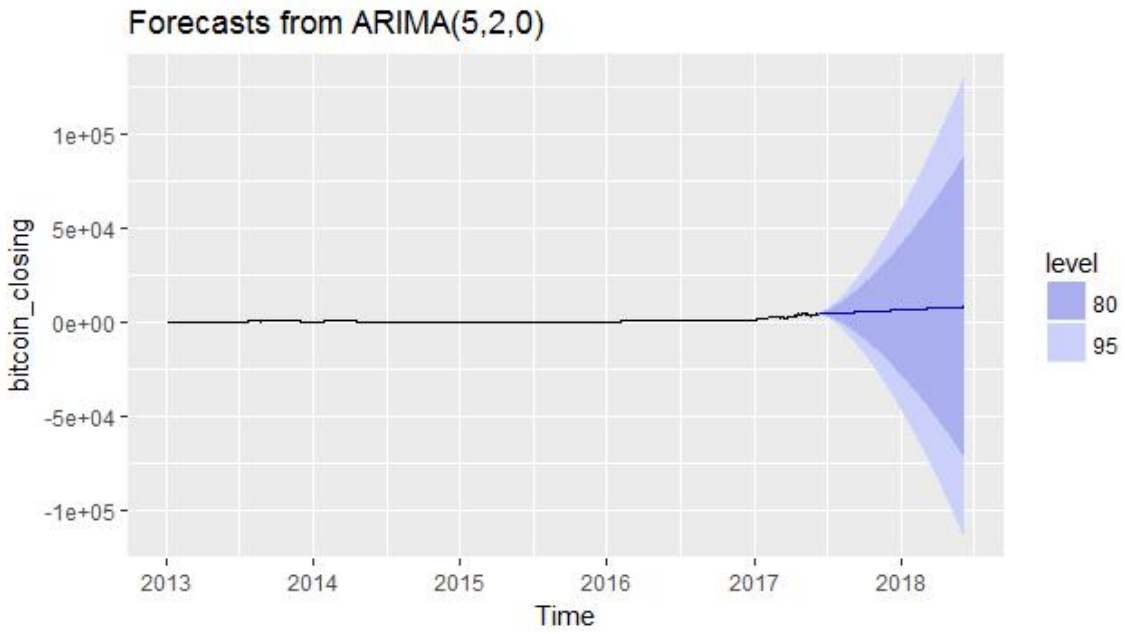


Figure 15:ARIMA Model

FIgure15 ,show the forecasts of the bitcoin data of 2013 to 2018. From 2013 to 2017 it was steady. In 2018 it was drastically growth after first half of the year 2017, as the data shows in the chart, the dark blue area cover the 80% and light blue area cover the 95% of the data.It also shows the trend between actual and predicted values.Dark blue line are the actual values and blue line is the predicted values.In blue shaded area, shows the prediction values.

```
> summary(arima_closing)
Series: bitcoin_closing
ARIMA(5,2,0)

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -0.8962  -0.7013  -0.5212  -0.4384  -0.1950
s.e.   0.0245   0.0313   0.0335   0.0314   0.0245

sigma^2 estimated as 3297:  log likelihood=-8831.06
AIC=17674.11   AICc=17674.16   BIC=17706.43

Training set error measures:
                    ME     RMSE      MAE         MPE     MAPE        MASE
Training set 0.04058749 57.29487 23.32579 0.01472822 2.858354 0.03847022
                    ACF1
Training set -0.03065137
>
```

```
> arima_closing
Series: bitcoin_closing
ARIMA(5,2,0)

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -0.8962  -0.7013  -0.5212  -0.4384  -0.1950
s.e.   0.0245   0.0313   0.0335   0.0314   0.0245

sigma^2 estimated as 3297:  log likelihood=-8831.06
AIC=17674.11   AICc=17674.16   BIC=17706.43
>
```

```
> checkresiduals(arima_closing)

        Ljung-Box test

data:  Residuals from ARIMA(5,2,0)
Q* = 661.81, df = 725.5, p-value = 0.956

Model df: 5.   Total lags used: 730.5
```

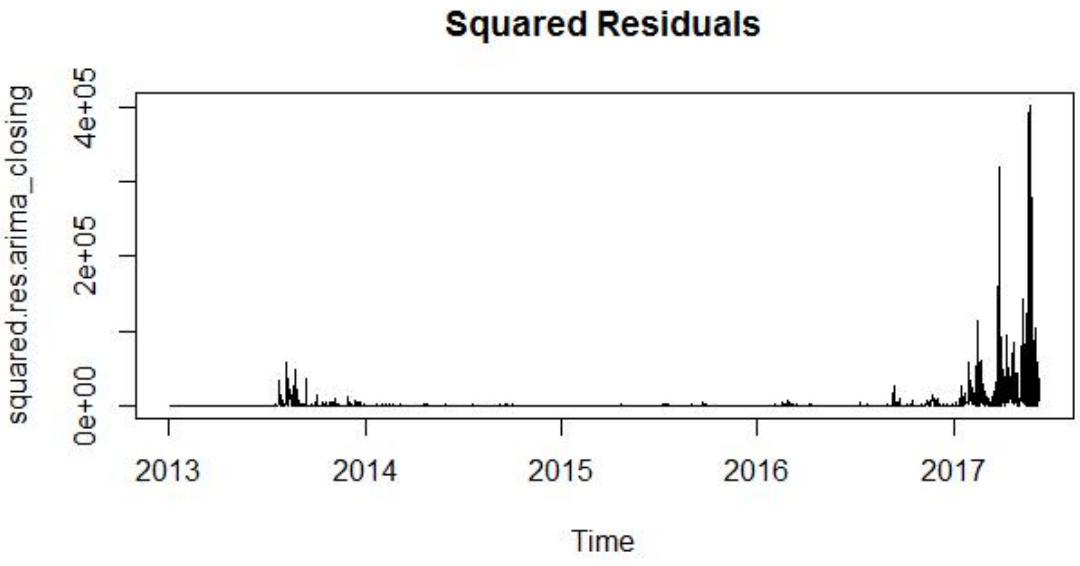
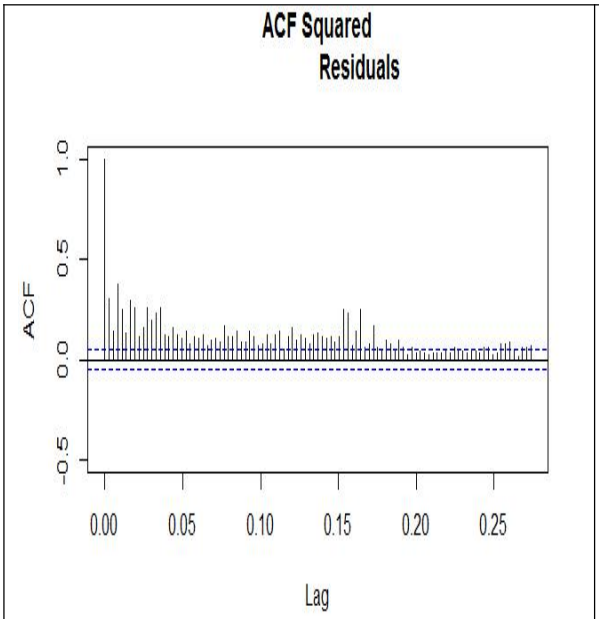


Figure 16:Residual ARIMA model

Figure16, show the squared residuals value in time frame between 2013 to 2017. End of the year 2013 the value has been rise a little bit and came back down to average in 2014. Between 2014 to mid 2017 value stayed steady,in 2nd half of 2017 and start making its impact in the market and reached to the highest recorded value ever as we can see in the graph.



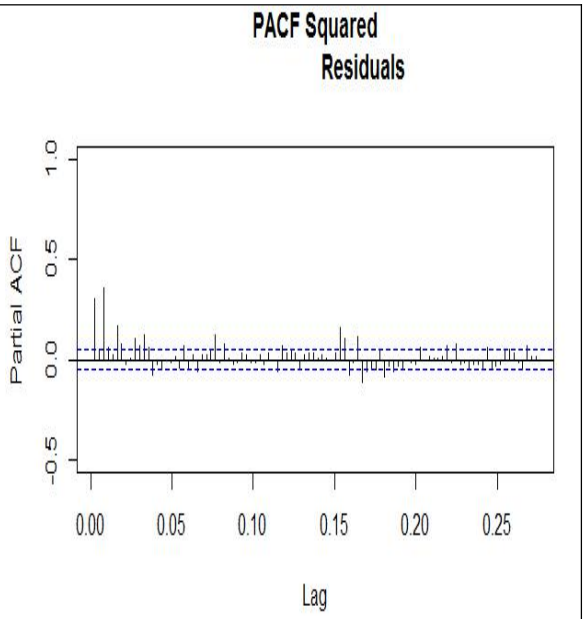


Figure 17: ACF and PACF Models

Figure 17, is showing the auto correlation and partial auto correlation of bitcoin data and it is clear from the graphs there is seasonality and relationship of past and current values.

```
garch(x = res.arima_closing, order = c(0, 8), trace = F)

Model:
GARCH(0,8)

Residuals:
      Min        1Q    Median        3Q       Max
-4.464621 -0.132343  0.003624  0.142311  3.902218

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 1.971e+03   1.079e+02   18.269  < 2e-16 ***
a1 1.364e-01   3.465e-02    3.936 8.28e-05 ***
a2 1.050e-01   3.725e-02    2.818  0.00483 **
a3 1.116e-01   2.741e-02    4.072 4.66e-05 ***
a4 4.824e-02   3.050e-02    1.582  0.11373
a5 1.187e-01   4.118e-02    2.883  0.00394 **
a6 1.111e-01   2.389e-02    4.650 3.32e-06 ***
a7 4.119e-02   4.775e-02    0.863  0.38838
a8 2.463e-11   4.170e-02    0.000  1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 12794, df = 2, p-value < 2.2e-16


        Box-Ljung test

data:  Squared.Residuals
X-squared = 39.385, df = 1, p-value = 3.479e-10
```

Figure above shows that after Jarque Bera and Box-Ljung Test have been performed to analyzed bitcoin data is not noisy as p-value is lower .05 which proves that data can be used to do forecasting.

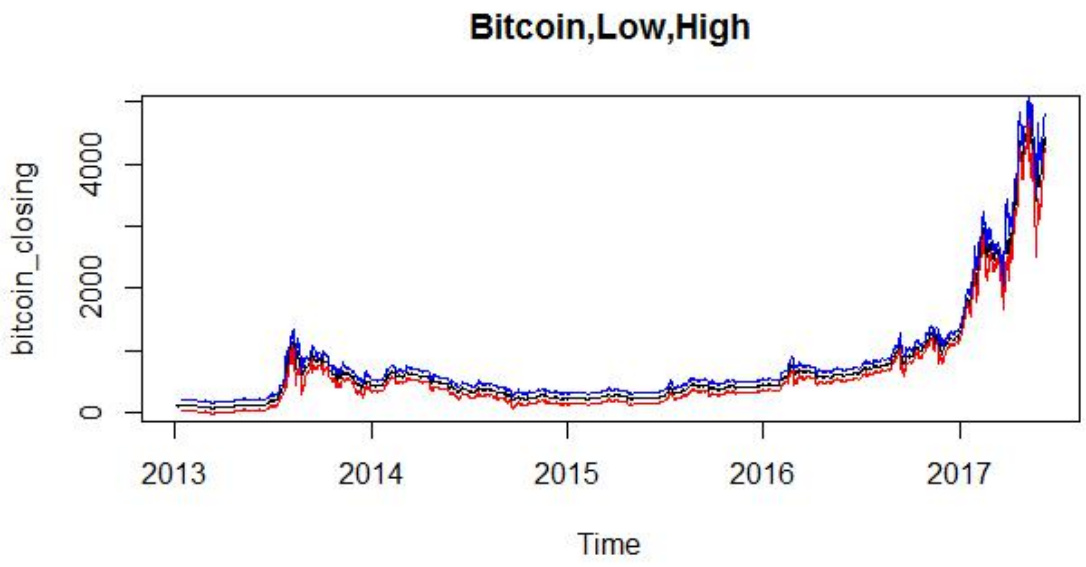
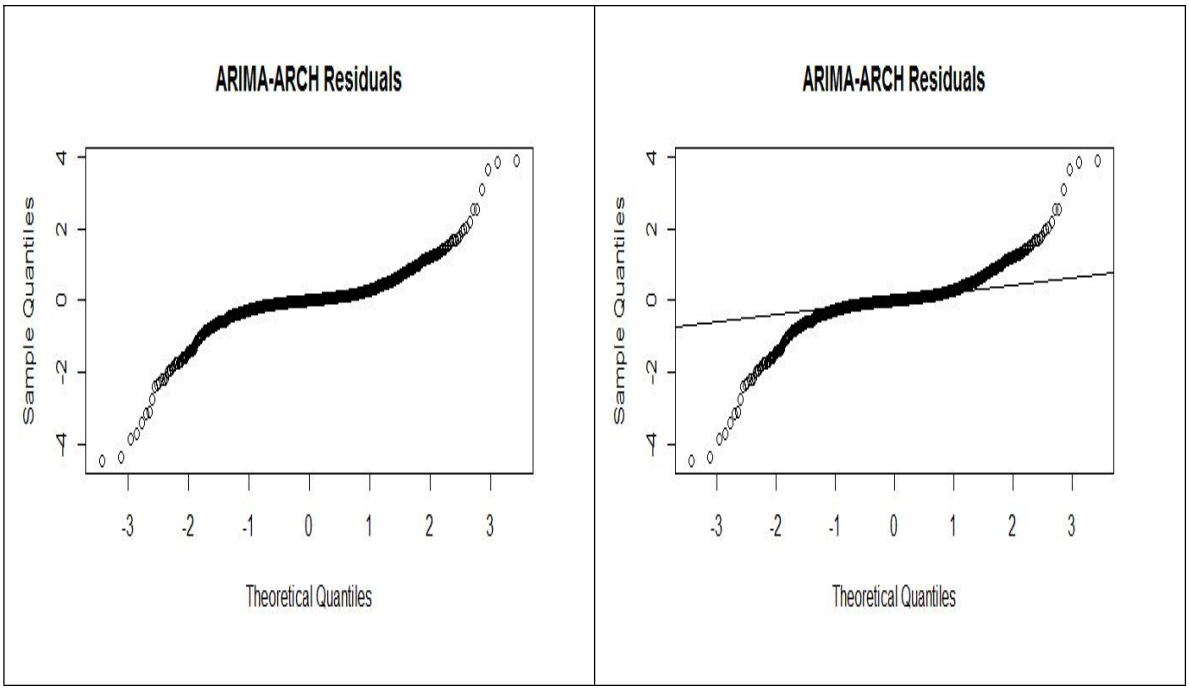


Figure 18: Low and High Bitcoin Value

Figure 17, shows the lowest and highest values by the bitcoin.As we can from the figure bitcoin start making its market around 2013 but after 2017 it made the huge impact in the crypto currency and got the eminent attraction from the organizations.In red line shows the lowest and in blue show the highest value reached during the 2013 and 2017 and forecasted value after 2017 to 2018.



5.2.2 Time Series(Prophet) Model

Time Series Prophet is a latest package model for forecasting at scale provided by Facebook's data science team. It is a procedure of forecasting the data in time series.It handles and robust of any missing data, drastic changes,outliers in the data.It forecast in very quick period of time and performs better than any other model in most cases.

5.2.2.1 Implementation

Time Series (Prophet Facebook API) model is applied on the bitcoin time series data using prophet() function.The prophet coding is done in prophet package which is also called prophet.For forecasting make_future_dataframe() function has been used and prediction is done by predict() function.Bitcoin price for over a year time has been forecasted and predicted by prophet.Then the visual representation of the prophet have been done by using simple plot() and prophet_plot_components() functions.

5.2.1.2 Evaluation and Results

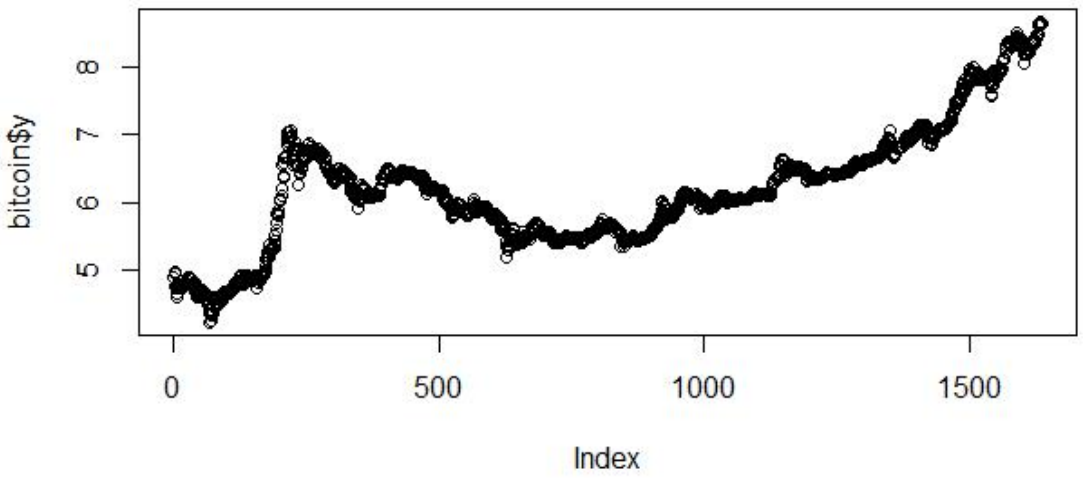


Figure 19: Prophet plot on Close

Figure 18,This is graph is taken from the time series prophet() model.It shows the data fluctuation over the time and it gradual increment over the period of time.As we seen before in ARIMA model, there is rise in price of bitcoin value over the time.

```
> m <- prophet(bitcoin)
Disabling daily seasonality. Run prophet with daily.seasonality=TRUE to override thi
s.
Initial log joint probability = -11.9449
Optimization terminated normally:
  Convergence detected: relative gradient magnitude is below tolerance
```

Figure 20: Joint Probability

The figure above show the joint probability of the data.

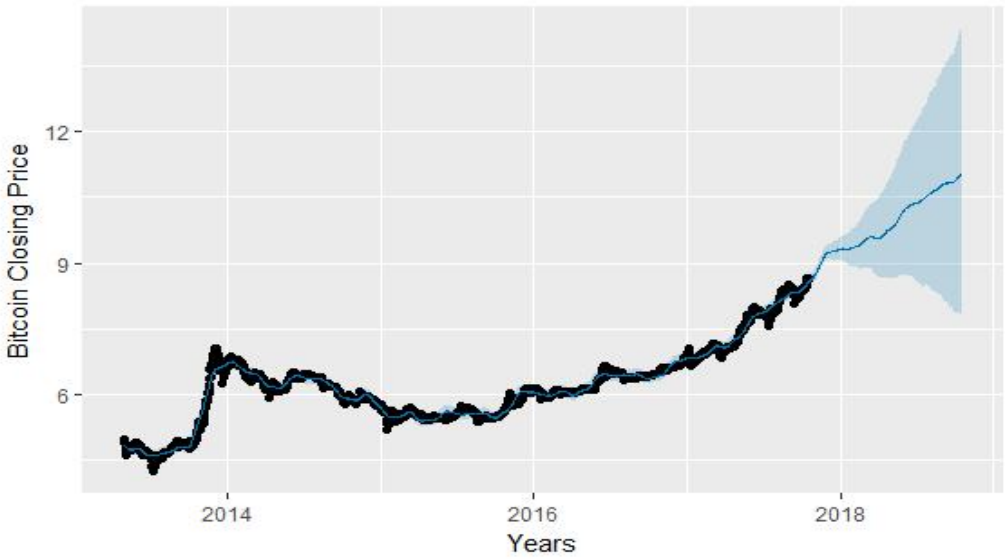


Figure 21: Actual Vs Predicted Values

Figure 20,show the trend between actual and predicted values.Dark black dotted are the actual values and blue line is the predicted values.In blue shaded area, shows the prediction values in yhat_lower and yhat_upper.We can there is clear trend throughout 2013 to 2017 and its continues during 2017 and 2018 it gives confidence intervals or prediction during 2017-2018.In early time of 2017,blue shaded area confidence interval is quite narrow as the time increase so the forecasted confidence interval increase.

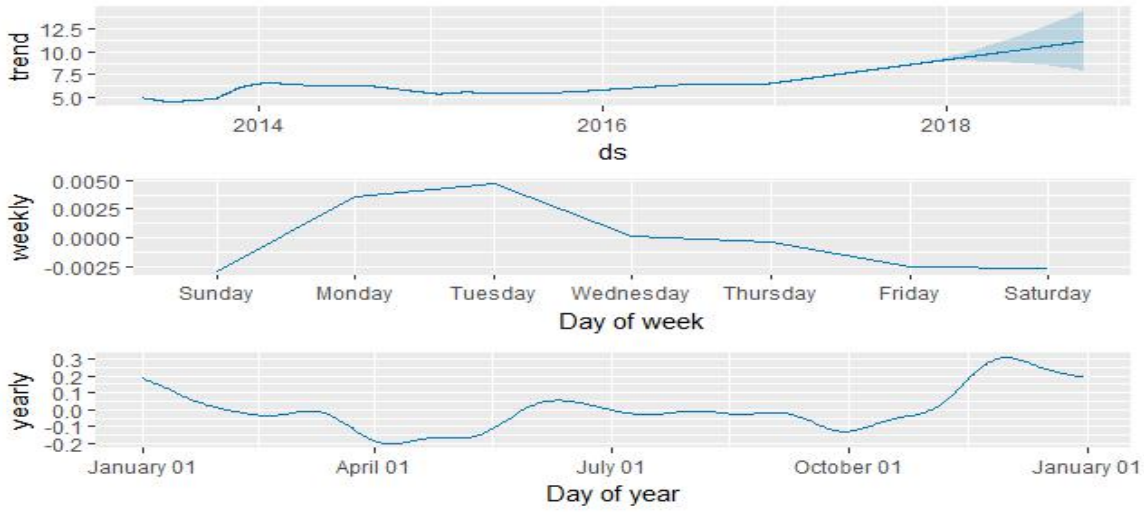


Figure 22 : Component Plot

In figure 21, we can see that the forecast as been broken down into trend component ,weekly seasonality and yearly seasonality.In trend component we can there is clear trend throughout 2013 to 2017 and its continues during 2017 and 2018 it gives confidence intervals or prediction during 2017-2018.In early time of 2017, confidence interval is quite narrow as the time increase so the forecasted confidence interval increase.

From the weekly graph, we can see that the values goes highest on Tuesday and lowest on sat and sun.We also can see that beginning of the week value rises while towards the end in drops down sharply.

From the yearly graph, we can see that it rises in start of the year and drops down after end of 1st quarter and rise again towards the end of the year.

5.3 Linear Regression Model

5.3.1 Implementation

Bitcoin data for this project has been gathered from the Kaggle(BITCOIN-COINMARKETCAP.csv) and been used for forecasting ,evaluation and comparison of models.Downloaded dataset has been cleaned in Python.For processing, packages and libraries like pandas for cleaning,merging and for data frame, numpy for calculation on data and matplotlib package to plot to graphical representation, etc has been installed and loaded and the data has been divide in test (40%) and train data (60%) randomly.After that LinearRegression() function have been used to do regressor and regressor.fit().regressor.predict() function is used to do prediction on the data.

5.3.2 Model Evaluation and Results

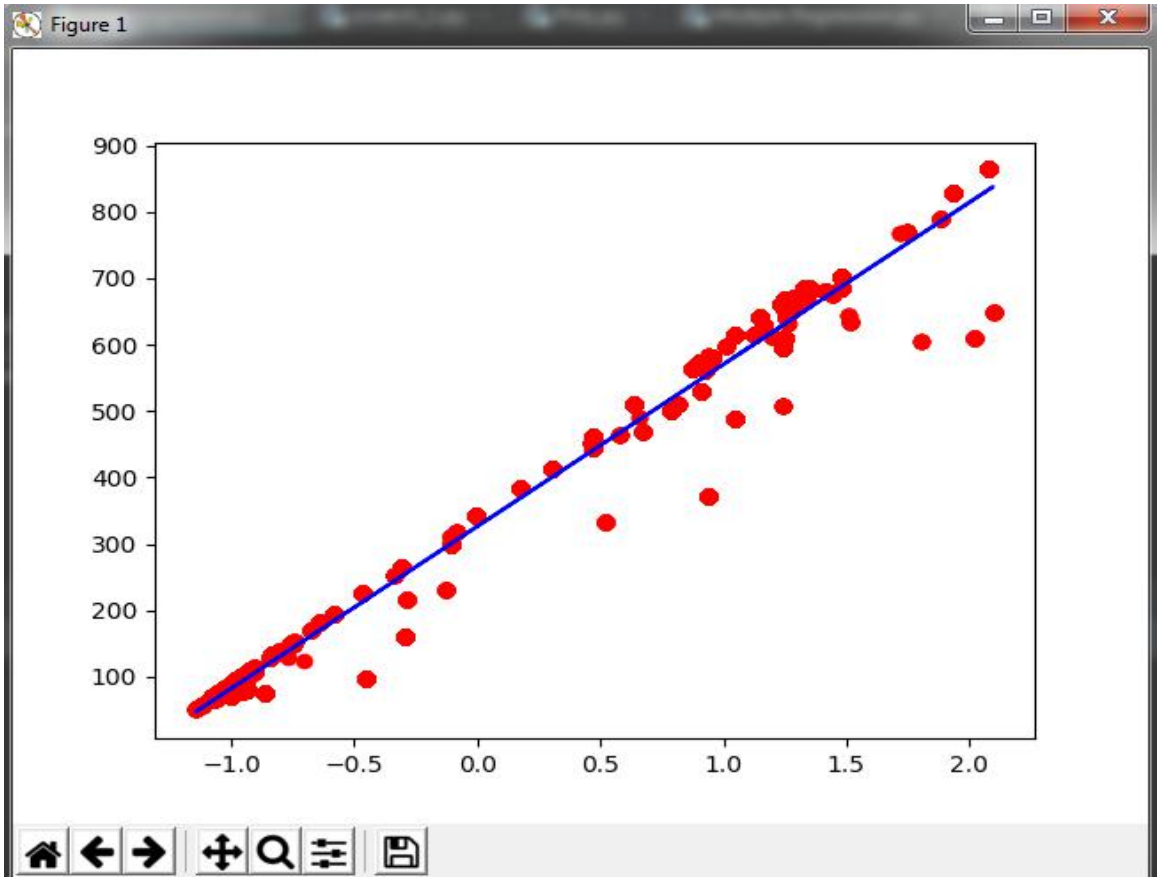


Figure 23: Linear Regression Model

Figure 23,show the graphical representation of linear regression model.As we can see that most of the data is on or close to the line which proves that the data is quite good for forecasting.

5.4 Multiple Regression

5.4.1 Implementation

For Multiple regression model I have used 4 different data file.Cleaned and merged them into one file through Python.For this model I have used glob for reading the file format ,pandas for cleaning,merging and for data frame, numpy for calculation on data and matplotlib package to plot to graphical representation.

5.4.2 Model Evaluation and Results

5.5 Polynomial Regression

5.5.1 Implementation

For Polynomial regression model I have used 4 different data file.Cleaned and merged them into one file through Python.For this model I have used glob for reading the file format ,pandas for cleaning,merging and for data frame, numpy for calculation on data and matplotlib package to plot to graphical representation.

```
pattern = '*.csv'
# Save all file matches: csv_files
csv_files_name = glob.glob(pattern)


# Print the file names
#print(csv_files_name)
# Load the second file into a DataFrame: csv2
df0 = pd.read_csv(csv_files_name[0])
df1 = pd.read_csv(csv_files_name[1])
df2 = pd.read_csv(csv_files_name[2])
df3 = pd.read_csv(csv_files_name[3])

################## Merging file
#df9 = df9.rename(columns={'Country1': 'Country'}) ## Renaming column name for merging

df0['Date'] = pd.to_datetime(df0['Date']).dt.strftime('%d/%m/%Y')
#print(df0.head(10))

df1['Date'] = pd.to_datetime(df1['Date']).dt.strftime('%d/%m/%Y')
#print(df2.head())


df2 = df2.rename(columns={'Timestamp': 'Date'}) ## Renaming column name for merging
df2 = df2.drop('Date', axis=1) ## Droping the column date
```

Figure above shows, loading the different data files into python and after that have done some cleaning process on the data frames.

```
############################## Merging of all the excel sheet
Merge1 = pd.merge(left=df0, right=df1, left_on='Date', right_on='Date')
#print(Merge1.head(10))

############################

Merge2 = pd.merge(left=Merge1, right=df2, left_on='High_x', right_on='High')
#print(Merge2.head(10))


################################################## separating the independent and dependent columns

x=Merge2.iloc[:,20:21].values




#print(x.head(5))

y=Merge2.iloc[:,2].values
```

From figure above we can see the merging the files into a one file.

```
from sklearn.cross_validation import train_test_split
x_train, x_test, y_train, y_test =train_test_split(x,y,test_size=0.2,random_state=0)

## Feature scaling
from sklearn.preprocessing import StandardScaler
sc_x=StandardScaler()
x_train=sc_x.fit_transform(x_train)
x_test =sc_x.fit_transform(x_test)

###### Polynomial  linear regression

from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
poly_reg=PolynomialFeatures(degree=2)
x_poly=poly_reg.fit_transform(x)

lin_reg_2=LinearRegression()
lin_reg_2.fit(x_poly,y)

## Visualing the Polynomial regression
plt.scatter(x,y,color='red')
plt.plot(x,lin_reg_2.predict(poly_reg.fit_transform(x)),color='blue')
plt.title('Predit')
plt.xlabel('Positive')
plt.ylabel('Negetive')
plt.show()
```

Figure above shows that ,the data has been split into test and train by using sklearn.

5.5.2 Model Evaluation and Results

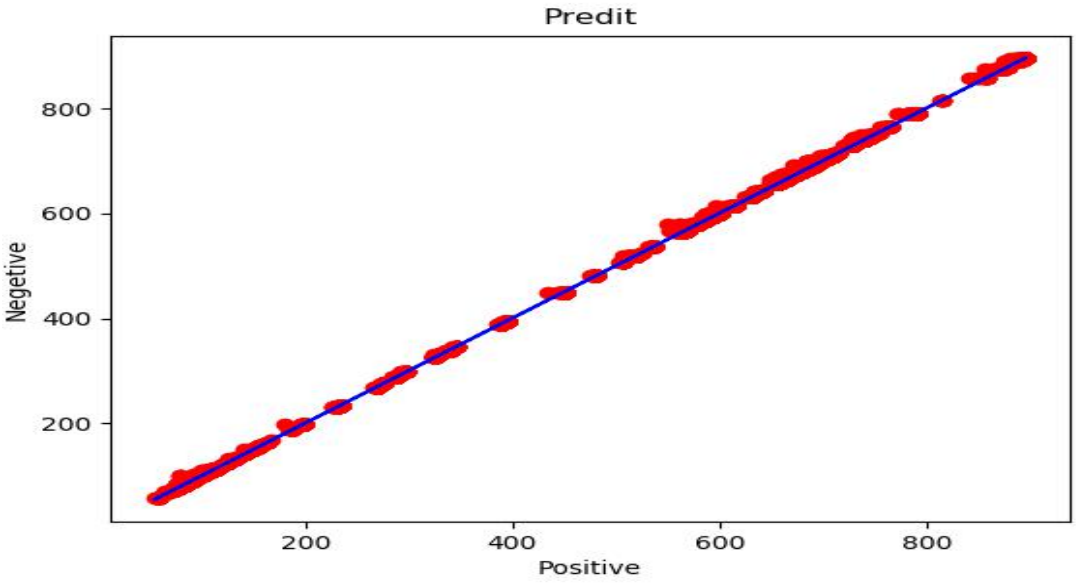


Figure 23,show the graphical representation of Polynomial regression model.As we can see that most of the data is appears to be on the line which shows that the data is good enough to do forecasting.

6. Conclusions and Recommended Future Work

In this project, I build two time series models ARIMA and Time series with Prophet and three regression models to check if the data is suitable or not for forecasting.They are quite similar in the way they work but some specs can be compared as slightly better than the other.They both works quite well in forecasting and analyzing. After completing my research and project I found out Time series Prophet model is quite easy to understand and build.It analyses faster than the ARIMA model.Prophet model performs well with reasonable number of missing data,large outliers and changes in historic trends.Prophet produce forecast accurately more often then the other models with less effort. Prophet model produces results in no times. This model comes handy if we are talking about speed and accuracy between two model.

Acknowledgement

It has been a great learning and experience for me throughout the whole semester so far specially the main project.I wasn’t sure if I could pull this off as it was more challenging than I was prepared for. However, with the support and guidance of all lecturers and special thanks to my supervisor(Catherine Mulwa).She is really patience and helpful person I wouldn't be able to finish my project without her guidance and support i feel it wouldn't have been possible. Classes were as productive as they could be. The whole module was set to equipped students with a great knowledge and experience when it comes to real world. Can deny it wasn’t hectic or I dint feel pressure but I was sure it was for my greater good.

As the project begun I was in touch with supervisor once or twice in every 2 weeks. Sometime often as I was facing difficulties at some stages of the module. Although It was my idea to work on that but sure needed a guidance how I can make it work to live up to their expectations. Whenever I needed to see my supervisor I was given time even on short notice. I was gaining momentum each time I had meeting. I cant thank my supervisor enough to help me though my project.

It surely is one of my great learning experience in my academic career.

References

http://proceedings.mlr.press/v55/amjad16.pdf : **Muhammad J Amjad, 2017.**

Research Journal of Business Management : 2012, M. Alrasheed.
Ciaian, P., Rajcaniova, M., and Kancs, A. (2014). The economics of bitcoin price formation. Technical report. http://arxiv.org/pdf/1405.4498.

van Wijk, D. (2013). What can be expected from the Bitcoin? PhD thesis, Erasmus Universiteit Rotterdam. https://thesis.eur.nl/pub/14100/Final-version-Thesis-Dennis-vanWijk.pdf.

Bourie, E., Azzi, G., and Dyhrberg, A. (2016). On the return-volatility relationship in the bitcoin market around the price crash of 2013. AEconomics: The OpenAccess, Open-Assessment E-Journal, October (2016), (2016-41). http://www.economicsejournal.org/economics/discussionpapers.

Kristoufek, L. (2015). What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis? PLoS ONE, (10(4)). http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0123923.

I.M. Yassin et al. ; file:///C:/Users/Harini%20Harish/Downloads/2842-6401-1-PB.pdf; 2017.

Hafezi R, Shahrabi J, Hadavandi E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. Applied Soft Computing, 2015.

Marček D. Some intelligent approaches to stock price modelling and forecasting.Journal of Information, Control and Management Systems.

Wah B W, Qian M L. Constrained formulations and algorithms for predicting stock prices by recurrent FIR neural networks. International Journal of Information Technology and Decision Making, 2006.

Shah V H. Machine learning techniques for stock prediction. 2007, https://pdfs.semanticscholar.org/5c6d/3b9959a8556a533a58e0ac4568dfc63d6ff4.pdf

Di X. Stock trend prediction with technical indicators using SVM. 2014, https://pdfs.semanticscholar.org/af39/96f74a477b649f4cf5c87645dae12b6232b9.pdf

Almeida J, Tata S, Moser A, Smit V. Bitcoin prediciton using ANN. 2015, http://s3.amazonaws.com/academia.edu.documents/39637630/Bitcoin_prediciton_using_ANN_Group_7.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1501566310&Signature=KVu8Z0%2FsHtadPadPmSzLvjs9kVs%3D&response-content-disposition=inline%3B%20filename%3DBitcoin_Stock_Prediction_Using_Artificia.pdf [20] McNally S. Predicting the price of Bitcoin using Machine Learning. Master thesis, Dublin: National College of Ireland, 2016.

McNally S. Predicting the price of Bitcoin using Machine Learning. Master thesis, Dublin: National College of Ireland, 2016.

Greaves A, Au B. Using the Bitcoin transaction graph to predict the price of Bitcoin. 2015.

codeburst. 2018. Introduction to Time Series forecasting with Prophet by Facebook. [ONLINE] Available at: https://codeburst.io/introduction-to-time-series-forecasting-with-prophet-by-facebook-ae0a54d09e88. [Accessed 02 May 2018].

http://proceedings.mlr.press/v55/amjad16.pdf : Muhammad J Amjad, 2017. Research Journal of Business Management : 2012, M. Alrasheed. Ciaian, P., Rajcaniova, M., and Kancs, A. (2014). The economics of bitcoin price formation. Technical report. http://arxiv.org/pdf/1405.4498.

van Wijk, D. (2013). What can be expected from the Bitcoin? PhD thesis, Erasmus Universiteit Rotterdam. https://thesis.eur.nl/pub/14100/Final-version-Thesis-Dennis-vanWijk.pdf.

Bourie, E., Azzi, G., and Dyhrberg, A. (2016). On the return-volatility relationship in the bitcoin market around the price crash of 2013. AEconomics: The OpenAccess, Open-Assessment E-Journal, October (2016), (2016-41). http://www.economicsejournal.org/economics/discussionpapers.

Kristoufek, L. (2015). What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis? PLoS ONE, (10(4)). http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0123923.

I.M. Yassin et al. ; file:///C:/Users/Harini%20Harish/Downloads/2842-6401-1-PB.pdf; 2017.

Hafezi R, Shahrabi J, Hadavandi E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. Applied Soft Computing, 2015.

Marček D. Some intelligent approaches to stock price modelling and forecasting.Journal of Information, Control and Management Systems.

Wah B W, Qian M L. Constrained formulations and algorithms for predicting stock prices by recurrent FIR neural networks. International Journal of Information Technology and Decision Making, 2006.

Shah V H. Machine learning techniques for stock prediction. 2007, https://pdfs.semanticscholar.org/5c6d/3b9959a8556a533a58e0ac4568dfc63d6ff4.pdf

Di X.Stock trend prediction with technical indicators using SVM. 2014, https://pdfs.semanticscholar.org/af39/96f74a477b649f4cf5c87645dae12b6232b9.pdf

Almeida J, Tata S, Moser A, Smit V. Bitcoin prediciton using ANN. 2015, http://s3.amazonaws.com/academia.edu.documents/39637630/Bitcoin_prediciton_using_ANN_Group_7.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1501566310&Signature=KVu8Z0%2FsHtadPadPmSzLvjs9kVs%3D&response-content-disposition=inline%3B%20filename%3DBitcoin_Stock_Prediction_Using_Artificia.pdf [20] McNally S. Predicting the price of Bitcoin using Machine Learning. Master thesis, Dublin: National College of Ireland, 2016.

McNally S. Predicting the price of Bitcoin using Machine Learning. Master thesis, Dublin: National College of Ireland, 2016.

Greaves A, Au B. Using the Bitcoin transaction graph to predict the price of Bitcoin. 2015.

V. S¸. Ediger and S. Akar, “Arima forecasting of primary energy demand by fuel in turkey,” 2007.

Geraskin, P., Fantazzini, D. 2010. Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask,European Journal of Finance. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1752115

6 Appendix

6.1 Project Proposal

6.1.1 Introduction

Bitcoin is a new virtual currency also known as electronic payment system or cryptocurrency. It is the world's most valuable cryptocurrency, introduced by Satoshi Nakamoto in 2009.The two parties can make bitcoin transaction over internet with no middle man, by installing a bitcoin software called bitcoin client/wallet. Bitcoin transaction are posted in blocks to an open ledger called Block-chain. Individual users have a unique block-chain key. There is reasonable transaction cost compare to the other banks. Using bitcoins has its own pros and cons. User identity and money is hidden. They do not have to disclose his/her personal identity. It has been related to its own conspiracies and has made people a little too skeptical towards investing in it.

6.1.2 Motivation and Background

Bitcoin is a peer-to-peer based virtual currency which is currently providing a digital alternative to the different currencies in the world. Its concept of having a single value throughout is what makes it popular among its owners and traders. Considering that bitcoin follows a decentralized model, it has often been linked to the dark web, which has been proven wrong by many researchers. Using bit-coins has its own pros and cons.

Pros

- A reasonable transaction cost, which is much less compared to third party merchants and card companies.
- High security as all the transactions are encrypted and implying that this currency is not linked to government rules.

- Its value is not hindered by factors such as inflation or deflation in prices

Cons:

- It is still not a legal tender in any company in the world.
- Its value is not determined by any government inflation but it still varies drastically (the biggest change was when in December 2013 the value was $1100 and it went down to $177 in January 2015)
- Been a large base for theft and fraud.

Another reason for bit-coins being such a fascinating reason for research is that no researcher has yet been able to decode the social science behind people being open and accepting it as a currency despite not having any physical significance.

I have chosen this topic due to my interest in the concept of cryptocurrency and bitcoin, which has become the topic of interest for many investors (despite the cons) over the past few years. Some researchers have gone as far as to say that investing in this is as safe as investing in real estate but with bigger and quicker returns.

This research is based on the research that bitcoin prices have only gone up since their existence (almost doubling in value every year approximately on a minimum), and now, when they are gaining popularity and are increasing in value, their transaction and its data has become a relevant part of the financial industry to understand the sudden boom of a currency which didn't even exist a few years ago.

The basic plan is to identify the various categories of data that can be generated from such financial transactions and use that to generate a report showing how

the number of transactions has led to the profits for people who invested in bit-coins initially. This report also aims at determining the best times to start a transaction in the bitcoin industry and what could be the best time for a safe withdrawal to assure that no losses are suffered. The analysis will also use data provided to provide a rough estimate of how much the cryptocurrency would cost at any given point of time using a logic based self-designed calculator which again will utilize stored data and statistics based on those data to calculate profit margins for people who might want to withdraw their iinvestments. The investors would be provided with a front-end calculator with analysis on profit margins and time based predictions for bitcoin rates. This will take into consideration various parameters such as exponential growth till date, leading parameters such as increase in user base which will be hampered with time, increase in popularity and published interests by major banks also impact the value of this currency, which will make the overall analysis more complex, yet accurate as the information used will be precise. Though the source of this data is not determined yet, financial websites such as coin metrics, blockchain.info and bitcoincharts.com will be used to take the input data for this research.

Having a technical knack towards visualizing and simplifying data, I chose this topic as it conveniently involves not only a topic which is trending in the market, but also would allow me to show the best of my capabilities in terms of user friendly tool building and understanding the concept in more details.

Cryptocurrency has been related to its own conspiracies and has made people a little too skeptical towards investing in it, now knowing how safe it is. To bring out such features of the Bitcoin, I have decided to work on this research and show them visual proof of how this is not only the future but also a clean investment.

6.1.3 Project Objectives and Goals

The proposed research offers to provide a deep understanding on how cryptocurrency works and provide aims to provide a visualization tool which uses logic's coming from scripts in SQL to utilize relevant information from the data. The idea is to implement a simple visualization tool such as Tableau or Microsoft

Visual Studio. The aim is not only restricted to providing analysis on bitcoin trends but also enable the data to be processed quickly while visualizing, to do that, my proposal is to implement stored functions for the execution of reports.

This will divide the overall research into various parts, first being where I determine what exactly the output should be with respect to the potential users of my research, second being where I generate a technical document which shows exactly what technology I want to use for this research and the final part which will implement and test the outcomes of this research.

To further simplify the aim of this research, we can consider a hypothetical scenario (which is the desired output of this research), where a user enters fields such as number of bit-coins, their current value, the time he or she wishes to invest in the Bitcoin, and gets an output which can show a rough amount of how much his investment will be worth at the time he might want to make a withdrawal.

This calculator won't just show the amount and result of investment in the certain duration, but will also generate graphs and suggestion models based on previous data such as what time periods during the previous years did the value of Bitcoin go down, or what could be the best times near the time when the users want to sell their Bit-coins to make maximum profit. All the logic of this report will be based upon data generated and the graphs and other similar visualizations will be generated on the universally adapted Tableau.

The aim is not just restricted to making an efficient and user friendly output but also to understand how the Bitcoin cryptocurrency works and how its trading has manipulated the market. One of the best biggest target audience for this type of tool is people who are future investors and those who are not yet very much aware of this currency, as it would show them a data oriented futuristic picture of the pros of investing in it.

This research will uncover how data inputs can be used to deliver the best quality outputs to people who want to invest in bitcoin using any third-party medium with a data based calculator and visually appealing outputs to basically lure people into seeing what they already know from the data in terms of their profit on the

investment and other such statistics which will be prepared along the duration of this project.

6.1.4 Technical Approach

This research will incorporate a very structural pattern in technical terms. The database used will be Amazon Web Services based Red-shift, to store and view the data. Further down, once the data is obtained via the websites, it will be arranged, and schema will be designed to assure that all the data is sorted/cleaned and that is when the analysis will be performed. The idea is to be able to perform analysis on that data using simple SQL statements. These SQL statements will then be used to generate visual reports for Tableau software. The basic idea is to generate an in depth analytical report for the data and to predict the future values for the Bit-coins based on various tables and designed columns in the schema as well.

Another critical aspect of this project is the calculator, the logic for which will be designed depending on the parameters which can be expected as an output for the project. This part of the project will be deeply enunciated once the Software Requirement Specification document is ready for the same, which shows exactly what the users of the tool and potential clients of the same could expect from me as a researcher and a developer of the tool.

The idea here is to perform a logic based calculation and building that logic would involve using simple core java based operations, the graphs generated would use Python for visualizing the results.

The research will include constructing a brief section which uses information extracted from similar previous works on cryptocurrency and some works specifically designed for Bitcoin, then a requirements elicitation process would be designed which could show what is needed from such a tool to make it more accurate as well as user friendly and more of a generic tool than a specialized and precise one. This will be followed by the implementation process, which would be done after the analysis and calculations logic's have been designed,

along with schema being sorted and all the research having been done, to ensure that the end user can see what this project plans to convey in all its entirety.

Special resources required

For the implementation of this project, I will require to download the following:

- Tableau v 10.0.
- AWS
- Access to Red-shift Database on AWS

MS Office (basically to check results of calculator using visualization tools vs normal graphs in Excel)

6.1.5 Technical Details

The languages used in this project would be:

1. Java: To design the calculator logic and add necessary fields as needed
2. Python: To support with the final visualizations using the calculator
3. SQL: To perform the required analysis on data which is taken from the websites.

Tools:

1. SQL Developer/ Microsoft SQL Server
2. Tableau
3. Eclipse

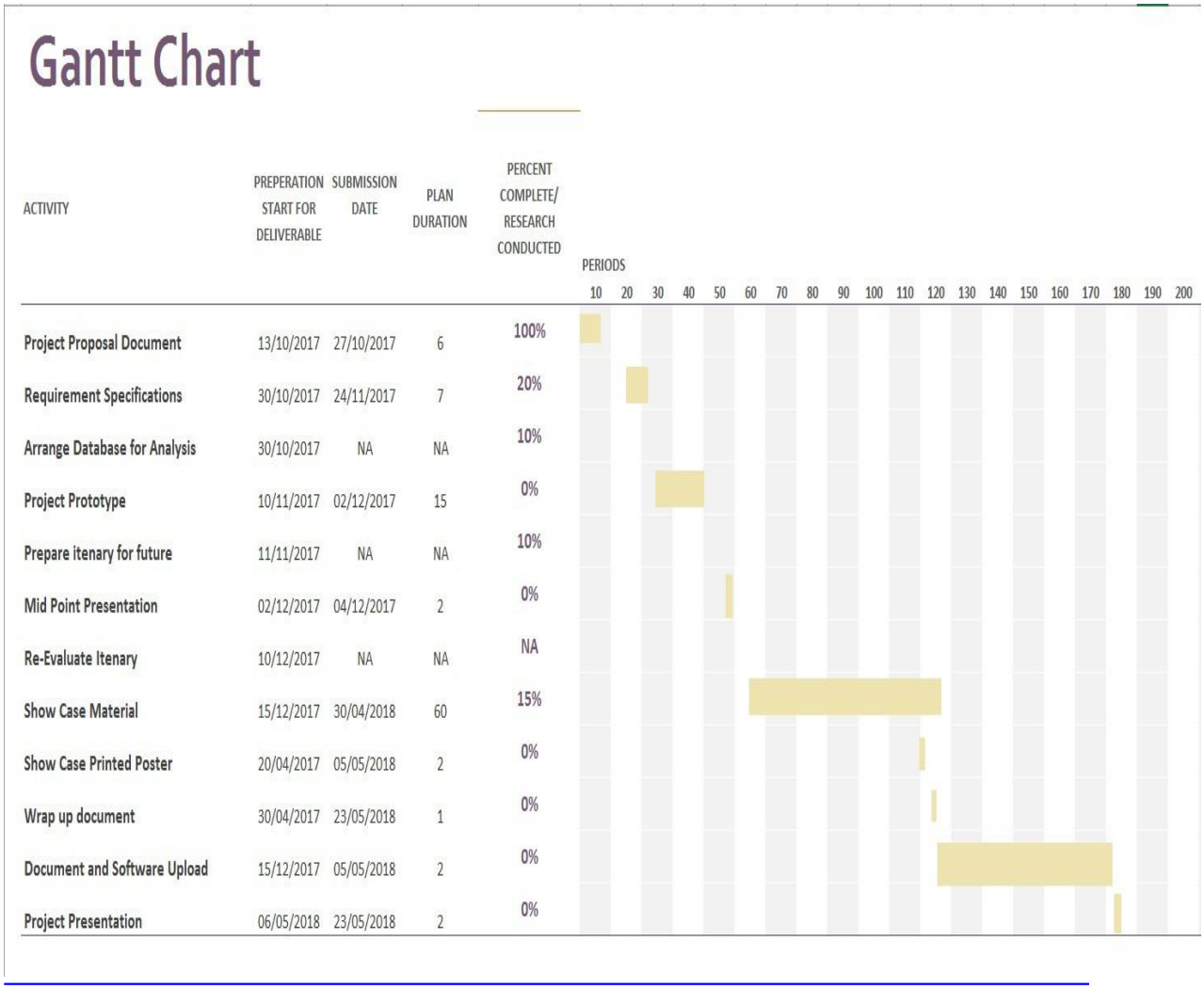
Libraries:

1. Plotly: Python based
2. Geoplotlib : This would be to narrow down the regions where cryptocurrency is popular to determine if that also plays a role in its rate (for example if Bitcoin is popular in a location with a high GDP, will it mean that its rate will be higher in comparison to a highly populated but economically backward region?)

6.1.6 Evaluation

The system will be tested based on the predictions which the calculator makes compared to the predictive patterns generated as the test results in previous years and seeing how much it compares to those. There have been many calculators which have predicted values for bitcoin prices for the upcoming months and this is unique in this case as I will not only use data input but also imply insights such as expected user base, a factor which has never been included in any of the calculations or predictions of Bitcoin prices ever, which is one of the strongest USP's of my research project.

6.2 Project Plan



6.3 Monthly Journals

6.3.1 September Journal

September Reflective Journal

Student name: Sumit

Programme : BSc in Computing in Data Analytic

Project: Bitcoin Prediction Modelling

My Achievements

Today is the first day of writing this journal, on 28th September I started to do some research for my final project. I am in final year of my course and have chosen Data Analytic as a specialization.

I did lots of research online while doing the research I visited many different types of the website which were related to data analytics and gained knowledge of application that can help me achieve my goals such as Kaggle, Tableau, Python etc. The findings of the research that I did today was all new to me, I did not have any previous experience of using Python or creating a project that require to do analysis, I would need focus more on this project in order to get high marks.

I have looked at many online pages and read blogs. Today I decided to do a project on predicting price of house in Ireland. I will be presenting this idea to three judges, at the moment I am not sure whether this idea will be accepted or not, I will try my best to describe it to judges and hope that it will be accepted so then I can make a quick start with this project.

I presented the idea of predicting house sale to three judges. Before going into the room, I was aware that it was all about telling your idea to the judges and getting their feedback. As I was very nervous, while presenting the idea I wasn't able tell exactly where I was going to get data from and what the whole concept was. Many question were asked during the presentation such as "what is novelty

of this project?", "where are you going to get the data from?", "what language are you going to use".

My Reflection

The question that were asked to me were very simple but as I was not prepared to answer them properly, I think I did not address the question correctly because my idea was not unique and there wasn't anything new that is not in the market therefore my idea was not accepted and I was told to do more research and make it bit complex.

I did more online research and tried to see if I could add more functionality or do something new with housing price. I also phoned few of my friends to discuss idea with them and see if more can be added to it. At this moment, I was getting help from where ever possible.

Intended Changes

I realized that idea of house sale was not good enough for final project and there wasn't much to do with it therefore I dropped the idea and started looking for a new fresh idea that have potential to achieve good grades. I continued to do my research.

Supervisor Meetings

Date of Meeting: Not Assigned Yet

Items discussed: Nothing to discussed

Action Items: Come up with new project idea

6.3.2 October Journal

October Reflective Journal

Student name: Sumit Tiwari

Programme : BSc in Computing in Data Analytic

Project: Bitcoin Prediction Modelling

My Achievements

This month I came up with the idea about what my main project will be. I did study over current Cryptocurrency and thought with more research and details I could make that work. Earlier this month I had meeting with supervisor and I proposed my idea to her. I was bit nervous at first as I was hopeful that I can make it work. During the meeting I was able to convince the supervisor about my idea and it was approved. Real work begins now.

Had to update the supervisor on project once every two weeks. Other labs and assignments also started about same time so time management was really important.

My Reflection

I wont lie, it was harder than I thought. It was a little hard to shape the project as I proposed it to the supervisor. As I started digging more and more It was coming all together gradually. Not focusing and make it work was not an option as I had another meeting later this month with supervisor on progress. I also found it hard to be on top other study work building up.

Intended Changes

I realized that I need to just give equal time to all on going workload in order to meet the deadlines.I presented more info on my project. It went well just wasn't happy with myself as I thought at that stage I would have more work done. Had a plan in place from this day hope to stick to that. All other Ca's and labs were getting hectic as well so it will be challenging.

Supervisor Meetings

Date of Meeting: 12/10/2017 and 23/09/2017

Items discussed: Progress on project, Project proposal report

Action Items: Research on Project,Project Proposal ,Detailed specification

6.3.3 November Journal

November Reflective Journal

Student name: Sumit Tiwari

Programme : BSc in Computing in Data Analytic

Project: Bitcoin Prediction Modelling

My Achievements

This month I figured project requirement specification in details as a result of my meeting with supervisor last month. I put an action plan in place as my project was getting in shape. Other module projects were going on and real work started building up. So I seriously needed some plan in to be on top of things.

Had another meeting with supervisor on progress. Work on specs requirement was key focused area.

This report was raw material of the project . I managed to put this in order and submitted my report by end of this month. It required a lot of focus and time management but I was glad to put in work because my project was going somewhere

My Reflection

I felt it did work well and supervisor feedback has been very productive for me. Sometimes the one specs doesn't work that you badly need thankfully there are alternatives. I see that as a good learning and exploring more about my project that wasn’t part of the plan,

However, I feel the pressure due ti time and workload stacking up from other modules.

Intended Changes

Not much changes there to apply just to stick with the plan so I don’t overdo the work on the project and fall behind on others.

Supervisor Meetings

Date of Meeting: 30/10/2017 ,09/11/2017 and 17/11/2017

Items discussed: Project Requirement specifications

Action Items: Work on specs that are not working properly

6.3.4 December Journal

December Reflective Journal

Student name: Sumit Tiwari

Programme : BSc in Computing in Data Analytic

Project: Bitcoin Prediction modelling

I this month I had to submit another report on tech. Also had a mid term presentation this week. A lot of work put in to meet the deadline. I still had a few tings to work on before I could present it. Project prototype was also on way this month. However I was confident as preparation was going well. Just the pressure of approaching dues dates of other projects was getting on my nerves.

My Achievements

This month,I was able to meet the deadlines and was very confident about presentation.I was hoping some glitch in tech department. Project prototype demo went well. Also was able to do good with other assignment submissions.With hectic schedule might effect the quality of the information put in the project. I was hard to manage but I was dedicated to have it completed with my best efforts.

My Reflection

I felt, it worked well and all hard work paid off so far. I would want to do more and brush up my skills on my project but splitting time between others assignments was challenging.

However, I was not successful in mid term presentation. Successful submitted the tech report. Project prototype demo went as I expected.

Intended Changes

I realized that I need to keep with the pace I am on at the moment. Months ahead are crucial all assignments about to meet their deadline. I am pretty confident with research and work ethics.

Supervisor Meetings

Date of Meeting: 04/12/2017 and 20/12/2017

Items discussed: Feedback on submitted Documents and intended changes need to be done in report.

Action Items: Start working on building models and Correcting report.

6.3.5 January Journal

January Reflective Journal

Student name: Sumit Tiwari

Programme : BSc in Computing in Data Analytic

Project: Bitcoin Prediction Model

By this month I had a lot of work done. My research had to go on I kept working on my prediction model. Faced some challenges regarding correct coding and correct language to make it user friendly. Was getting some success as well.

My Achievements

While main project work was going on I was able to prepare for mid term exams. I was a little relieved as I was in strong position with my prediction model. All the research in my past have made me very confident and I have come across new learning.

My Reflection

After having a good hold on my model I wanted to focus on my coming term exams. Dint want workload effect my grade. I had to put myself together and be prepared.

Intended Changes

Next month, I will try to focus more on exams as still have time to submit prediction model. Research for my program still goes on as I don't want to drop my work on it completely.

Supervisor Meetings

Date of Meeting:12/01/2018 and 25/01/2018

Items discussed: Mid-term presentation feedback,Wok on feature selection and models

Action Items:Build models,Work on document

6.3.6 February Journal

February Reflective Journal

Student name: Sumit Tiwari

Programme : BSc in Computing in Data Analytic

Project: Bitcoin Prediction Model

This month semester 2 starts. Have to review the whole model and look for flaws. That would have been very decisive period for the project. Was not able to face any sort of failure at this stage wit a lot of work building up.

My Achievements

This month, I was able to put implement plans in process as it will continue till final stage. I was hoping some flaws with coding but wasn't prepared for anything that can make me miss the deadline or anything unfortunate.

My Reflection

I felt, it was very busy period and head wrecking as was dealing with exams last month. I was working on both side restlessly. Was committed to continue as I have made it that far.

Intended Changes

Next month, as I been mentioning that time management and focus is everything. This month my focus will be shifting back on the model.I realized that I need to be prepared as more assignments awaiting at the start of this 2nd semester. Have labs each week as well.

Supervisor Meetings

Date of Meeting:06/02/2018 ,15/02/2018 and 26/02/2018

Items discussed:Report Progress,Progress on models,discuss solution and feedback

Action Items:Modify models,Pick the right models and continue work on report.

6.3.7 March Journal

March Reflective Journal

Student name: Sumit Tiwari

Programme : BSc in Computing in Data Analytic

Project: Bitcoin Prediction Model

This month other assignments and CA continues for other subjects. Implementation and progress on my model also continues. That took majority of my time, even more than I planned.

My Achievements

With all the work going on I managed to work on coding. There were some expected glitches in model. Initial idea did not work so had to do more research. Was getting constant feedback from supervisor. Some reality checks as well that made me work twice as hard. I went even deeper in coding and seek for solutions.

My Reflection

I felt this month was the most stressful of them all. Those irregularities made me nervous but was able to work on them. I was able to get over it up to some extant . Not being able to keep up with other CA and lab wasn't an option.

Intended Changes

Just had to plan my day better. That was the crucial time where I needed to set my priorities. I was getting real and all submission date are approaching. I feel, the amount of work and dedication I am investing I will be able to produce great results.

Supervisor Meetings

Date of Meeting: 05/03/2018,15/03/2018 and 28/03/2018

Items discussed: Literature Review,Review on models , Report progress , Report Architecture,Update on report

Action Items:Complete build models and Literature Review