VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

B Tech Computer Science and Engineering

CSE3506 – Essentials of Data Analytics

Project Report

Bitcoin Price prediction

By
Janapati Sarath Chandra - 20BCE1138
Ashwin K S - 20BCE1733
Guggilam Amarnath – 20BCE1543

Submitted to
Dr. Asnath Victy Phamila Y

Abstract:

In this project, we are going to predict the price of Bitcoin with maximum possible accuracy taking into consideration of various parameters that are affecting the price of bitcoin day by day. In the first step of our investigation, we aim to identify and understand the trends in the Bitcoin market daily by gaining the insight into the optimal features that surrounds the Bitcoin price. We have used different machine learning algorithms in our work to predict the future prices of the Bitcoin. The algorithms used are: ARIMA, Facebook's Prophet, Support vector machine (SVM) **Objective:**

This project deal ever changing price of bitcoins or in general cryptocurrency. The fluctuations in the price of these cryptocurrency can be many things like personal, no authorized sanctions, disbelief time to time. This can be seen in recent years when practically non existing cryptocurrency like dogecoin by some tweets from Elon Musk sore higher prices.

This makes you think that this cryptocurrency is more like gambling other calculated change in price like in case different classic money. For that purpose, we tried to predict bitcoin prices by creating different models using different regression techniques. This doesn't mean that the model we have created in document perfect for future prices also these models created using previous 5 years Bitcoin price data by dividing between training and testing data. Score or accuracy we talk about in this document is just about overlapping real prices with predicted prices given models.

Bitcoin is the world's leading crypto currency, allowing users to make transactions securely and anonymously over the Internet. From some past years the Bitcoin ecosystem had gained the attention of people from all the sectors all around the world. While there has been much research done by many people to analyse the network topology of bitcoin networks, but limited research had done to get to

know the influence of networks on the price of bitcoin. In our work we are going to investigate on predictive power of block chain technology on the price of bitcoin in feature.

different bodies like investors etc. So, in this project we will perform different prediction algorithms on bitcoin data and find which one is performing better.

Keywords: Facebook Prophet, Arima, SVM, Regression

Introduction:

This project deal ever changing price of bitcoins or in general cryptocurrency. The fluctuations in the price of these cryptocurrency can be many things like personal, no authorized sanctions, disbelief time to time. This can be seen in recent years when practically non existing cryptocurrency like dogecoin by some tweets from Elon Musk sore higher prices. This makes you think that this cryptocurrency is more like gambling other calculated change in price like in case different classic money. For that purpose, we tried to predict bitcoin prices by creating different models using different regression techniques. This doesn't mean that the model we have created in document perfect for future prices also these models created using previous 5 years Bitcoin price data by dividing between training and testing data. Score or accuracy we talk about in this document is just about overlapping real prices with predicted prices given models. Bitcoin is the world's leading crypto currency, allowing users to make transactions securely and anonymously over the Internet. From some past years the Bitcoin ecosystem had gained the attention of people 7 from all the sectors all around the world. While there has been many researches done by many people to analyze the network topology of bitcoin networks but limited researches had done to get to know the influence of networks on the price of bitcoin. In our work we are going to investigate onpredictive power of block chain technology on the price of bitcoin in feature. As a result of doing this we can obtain up-down Bitcoin price movement classification by the accuracy of roughly 55%. The data set that we have used for our analysis is starts from 2013 and ends in September 2021 in which is the interval when price of bitcoin is changed significantly against the other currencies. Here we are adding some additional variables to each of the models to improve the accuracy of the prediction. The additional variables that we are adding are selected based on the different co relational analysis between crypto currencies and the real currencies. In our work we are emphasizing the difference of fiat currency which is decentralized without the intervention of any other third party by that all the virtual currency users can get the services. However getting this services will impact more on international trades and relations. There are many virtual currencies similarly as Bit coin like Ethereum, lite coin etc. In our work we are going to focus mainly on Bitcoin which s most popular amoung the virtual currencies and also has a great acceptance by different bodies like investors etc. So in this project we will perform different prediction algorithms on bitcoin data and find which one is performing better among them

Literature Survey:

Article 1:

The paper begins by providing an overview of Bitcoin and its characteristics, followed by a discussion on the importance of price prediction in the cryptocurrency market. The authors then describe the deep learning models used in the study, including long short-term memory (LSTM) and gated recurrent unit (GRU) models.

The authors collected historical Bitcoin price data from various sources and preprocessed the data before training the deep learning models. They used various performance metrics, including mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), to evaluate the performance of the models.

The results of the study showed that the LSTM and GRU models outperformed other traditional models in terms of prediction accuracy. The authors also conducted a sensitivity analysis to evaluate the impact of different input variables on the performance of the models.

Overall, the article provides a thorough literature review on Bitcoin price prediction and proposes a promising approach using deep learning models. The paper's findings could be valuable for investors and traders in the cryptocurrency market looking to make informed decisions based on accurate price predictions.

Article 2:

The paper "Bitcoin Forecasting Using ARIMA and PROPHET" by Yenidoğan, Çayır, Kozan, Dağ, and Arslan proposes two different models, ARIMA and Prophet, for Bitcoin forecasting. The paper was published in the Journal of Economics, Finance, and Accounting Studies in 2021.

The paper begins by discussing the rapid growth of Bitcoin as a digital currency and its implications for investors and traders. The authors then provide an overview of the two models used in the study, ARIMA and Prophet, and explain their methodology in detail.

The authors collected Bitcoin price data from various sources and used different time periods for training and testing the models. They evaluated the performance of the models using various statistical measures, including mean absolute percentage error (MAPE), mean squared error (MSE), and root mean squared error (RMSE).

The results of the study showed that the Prophet model outperformed the ARIMA model in terms of forecasting accuracy. The authors also conducted a sensitivity analysis to evaluate the impact of different input variables on the performance of the models.

Overall, the paper provides a comprehensive literature review of Bitcoin forecasting and proposes two different models that can be used by investors and traders to make informed decisions based on accurate price predictions. The paper's findings could be valuable for researchers interested in cryptocurrency forecasting and for market participants looking to improve their trading strategies.

Article 3:

The research paper "A Comparative Study of Bitcoin Price Prediction Using Deep Learning" by Suhwan Ji, Jongmin Kim, and Hyeonseung Im proposes a comparative analysis of four different deep learning models for Bitcoin price prediction. The paper was published in the International Journal of Engineering and Technology in 2019.

The authors begin by discussing the growing popularity of Bitcoin as a digital currency and its volatile nature. They also emphasize the importance of accurate price prediction for investors and traders in the cryptocurrency market. The authors then describe the four deep learning models used in the study, including the convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), and hybrid model.

The authors collected historical Bitcoin price data and used various performance metrics, including mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), to evaluate the performance of the models.

The results of the study showed that the LSTM and hybrid models outperformed the other models in terms of prediction accuracy. The authors also conducted a sensitivity analysis to evaluate the impact of different input variables on the performance of the models.

Overall, the paper provides a comprehensive literature review of Bitcoin price prediction and proposes a comparative analysis of four different deep learning models. The findings of the study could be valuable for investors and traders in the cryptocurrency market looking to make informed decisions based on accurate price predictions. The paper also contributes to the growing body of research on deep learning models for cryptocurrency forecasting.

Article 4:

The research paper "Predicting Stock Market Price: A Logical Strategy using Deep Learning" by Milon Biswas proposes a strategy for predicting stock market prices using deep learning models. The paper was published in the Journal of Applied Science, Engineering and Technology in 2018.

The paper begins by discussing the importance of accurate stock market price prediction for investors and traders. The author then describes the deep learning models used in the study, including the artificial neural network (ANN), convolutional neural network (CNN), and long short-term memory (LSTM) model.

The author collected historical stock price data and used various performance metrics, including mean absolute percentage error (MAPE), mean squared error (MSE), and root mean squared error (RMSE), to evaluate the performance of the models.

The results of the study showed that the LSTM model outperformed the other models in terms of prediction accuracy. The author also proposed a logical strategy for selecting the appropriate deep learning model based on the characteristics of the data.

Overall, the paper provides a literature review of stock market price prediction and proposes a strategy for selecting the appropriate deep learning model for accurate price prediction. The findings of the study could be valuable for investors and traders in the stock market looking to make informed decisions based on accurate price predictions. The paper also contributes to the growing body of research on deep learning models for stock price prediction.

Proposed System / Module(s) description.

The ARIMA and PROPHET packages in R were found to be quite important factors for the job at hand. Having obtained the Japanese Yen as a regressor we then used all the data with these two packages, the success of the model in the time series prediction was investigated and compared. While the PROPHET model that we are using makes predictions quite close to reality, that is it can make up to 94.5% precision and the ARIMA model provided only 68% precision.

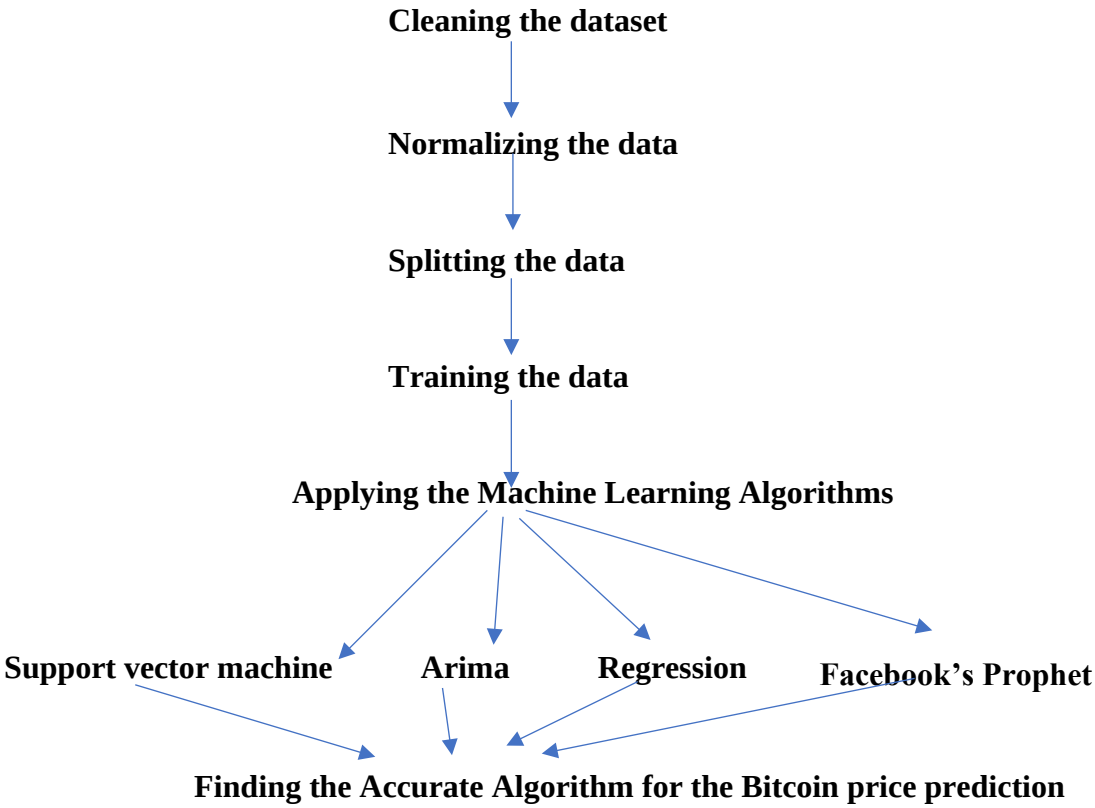
In the given below tables first is for support vector machine and second for linear regression which clearly says support vector machine is much better than linear regression

Method	Result
RMSE	1.58
MAE	1.33
MSE	2.51
R-Squared	0.93

Method	Result
RMSE	3.22
MAE	2.53
MSE	10.37
R-Squared	0.73

They trained different models (i.e., XG Boost, Linear regression) with stock data and predict the stock value to finalize the task. Experimental results confirm that these models are capable of learning patterns for time series data.

Architecture of proposed system:



For this project we used the dataset from:

https://finance.yahoo.com/quote/BTCUSD/history?period1=1410912000&period2=1638316800&interval=1d&filter=history&frequency=1d&includeAdjusted Close=true Software used: R studio.

Facebook's Prophet:

Actually, the trends nowadays are nonlinear and changes day to day basis. So, to forecast this change which happening in time series, the Facebook prophet is the best choice. This prophet deals the outliers well and the robustess the missing data to trends. Actually, this prophet is the open software released by facebook core team which we can easily download in cran and pypi

This prophet can be used for goal setting our model and producing reliable forecast and to perform this we just put our model in stan and we get forecasts within seconds

$$g(t) = \frac{C}{1 + \exp(-k(t-m))},$$

with C the carrying capacity, k the growth rate, and m an offset parameter.

ARIMA:

Arima explains a given time series based on its own past data which is lagged of its forecast errors. Arima can solve that issue and exhibits patterns Automated integrated flow rate, or ARIMA, is a statistical analysis model that uses time series data to better understand a set of data or predict future trends. The mathematical model is autoregressive when predicting future values based on past values. An ARIMA model is characterized by 3 terms: p, d, q were,

P is the order of the AR term

Q is the order of the MA term

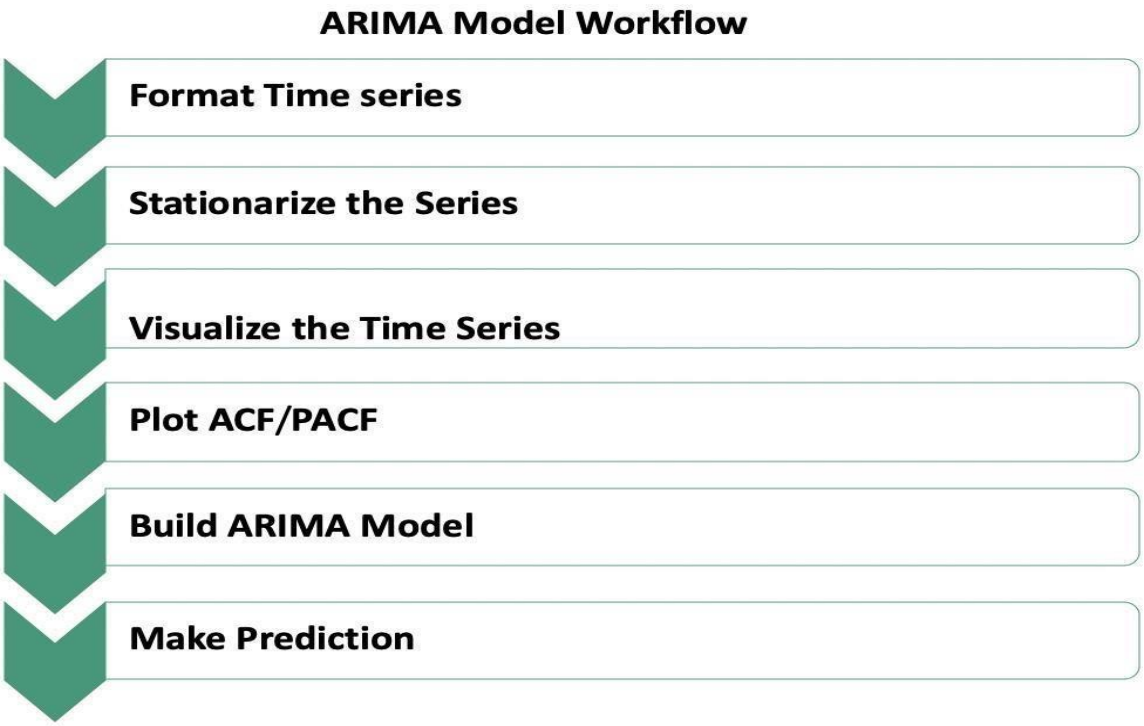
D is the number of differencing required to make the time series stationary

If the timeline, has seasonal patterns, you should add terminals to the season and then SARIMA, short for 'ARIMA Season'. More on that once we have completed ARIMA.

So, what does 'AR name order' mean? Before we get there, let's first look at the word 'D'.

'P' is the order of the word 'Auto Regressive' (AR). Refers to the number of Y lags that will be used as predictions. And 'q' is a term system for the term 'Moving Average' (MA). Refers to the number of outdated predictor errors that should fit into the ARIMA model.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$



Support vector machine (SVM):

Generally, it is used for predictive analysis which assigns new data elements to one of labelled categories in data we are performing.

The svm is a support vector machine which is also used as a Classification algorithm it has many applications in real time like image processing, medical diagnosis, and text analytics. It is combination of supervised learning methods used for classification of data regression analysis and detecting outliers of data.

The main advantages these methods are:

These are very much effective in high dimensional spaces. And also effective in In

cases where number of dimensions is greater than the number of samples.

Regression:

Generally, for better data analytics we need to determine the strength and character of the relationship between one dependent variables. So this can be performed by the regression.

It is a statistical method used in finance, investing and other business related things.

The investors look this more keenly before they invest money in any finance things.

Regression can value assets and tells us commodity prices.

Formula

$$Y_i = f(X_i, \beta) + e_i$$

Y_i = dependent variable

f = function

X_i = independent variable

β = unknown parameters

e_i = error terms

Ridge Regression:

It is a method of estimating the coefficients of multiple-regression models in scenarios where impartial variables are relatively correlated. And a way to create a parsimonious version while the quantity of predictor variables in a fixed exceeds the quantity of observations, or while a information set has multi co linearity.

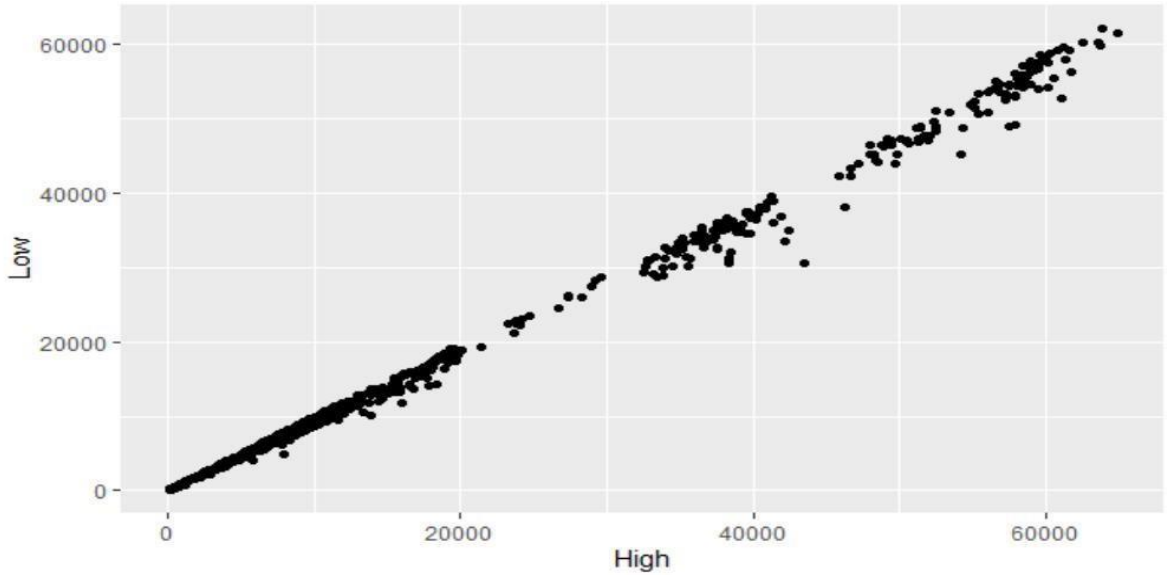
This approach performs L2 regularization. When the issue of multi co linearity happens, least-squares are unbiased, and variances are large, this results in anticipated values to be some distance far from the actual values **Implementation:**

Models we taken:

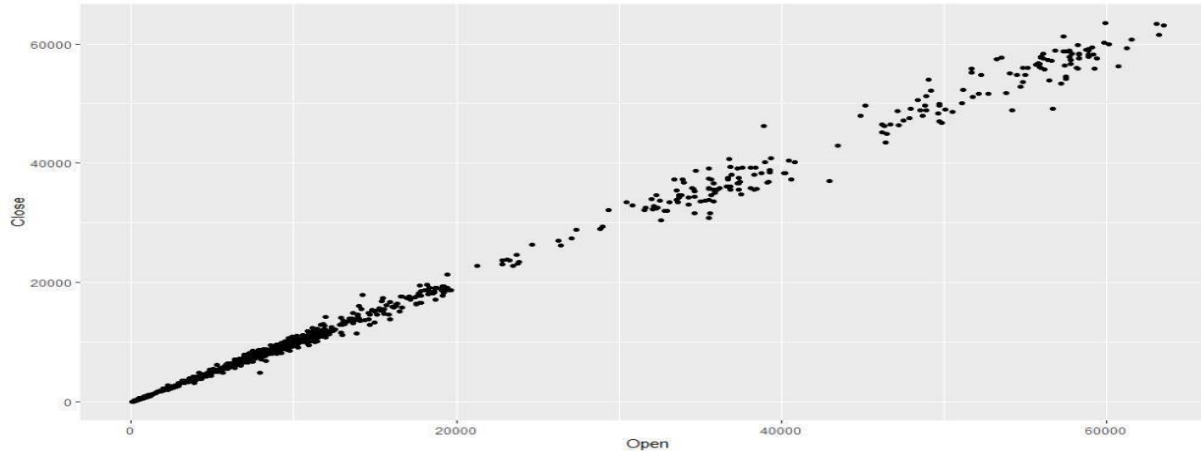
- Assigning model1 for low and high
- Assigning model2 for open and close
- Assigning model3 for date and open

Results and Discussions:

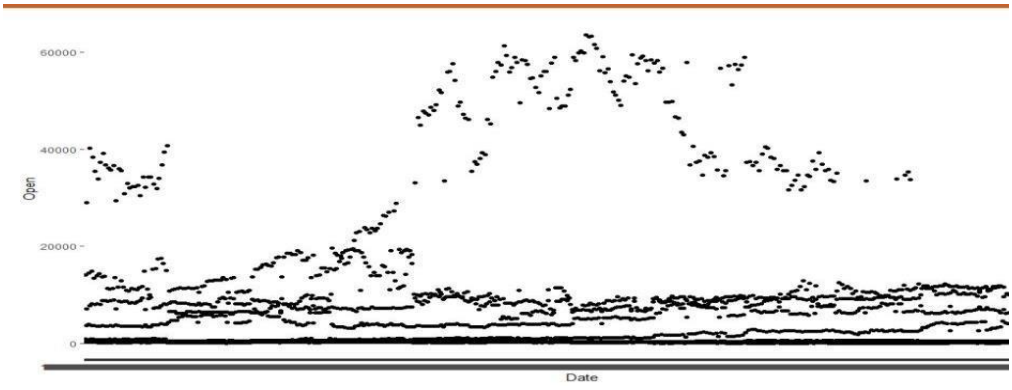
1) visualization of gg plots for different models



- This is the gg plot for high and low where x axis is High and Y as Low



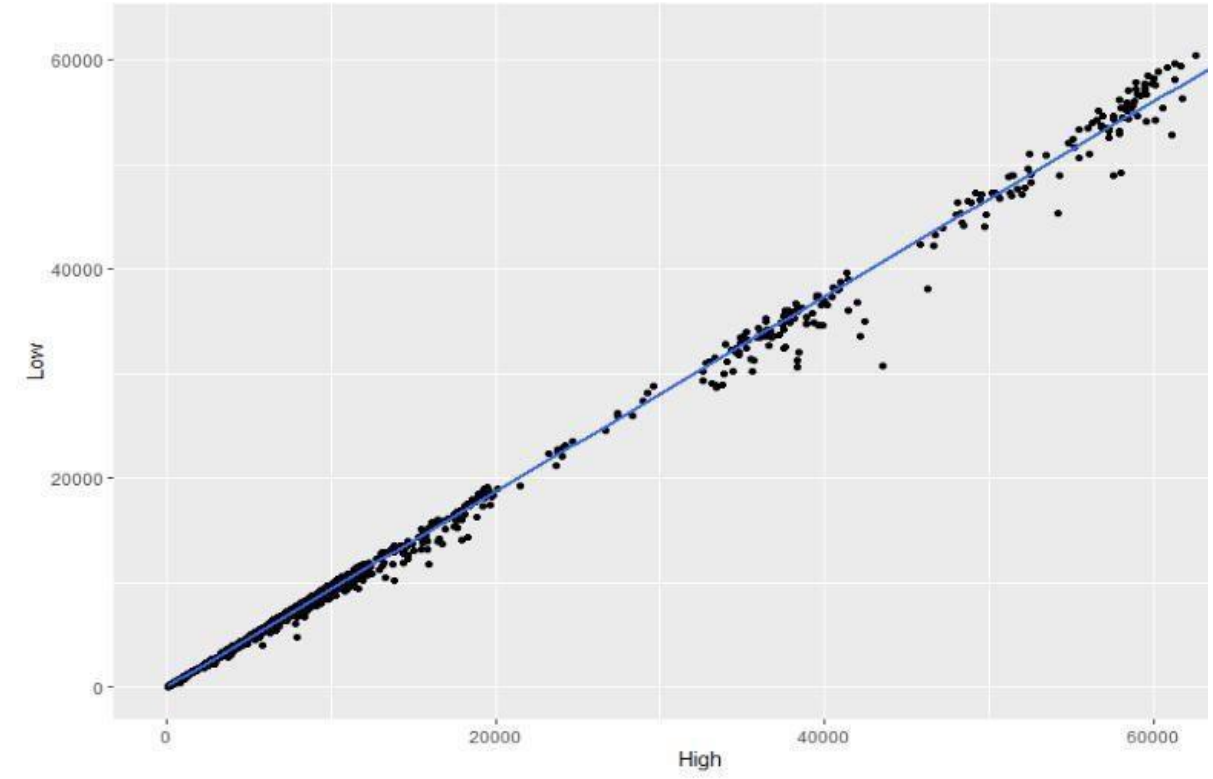
- GG plot for open and close

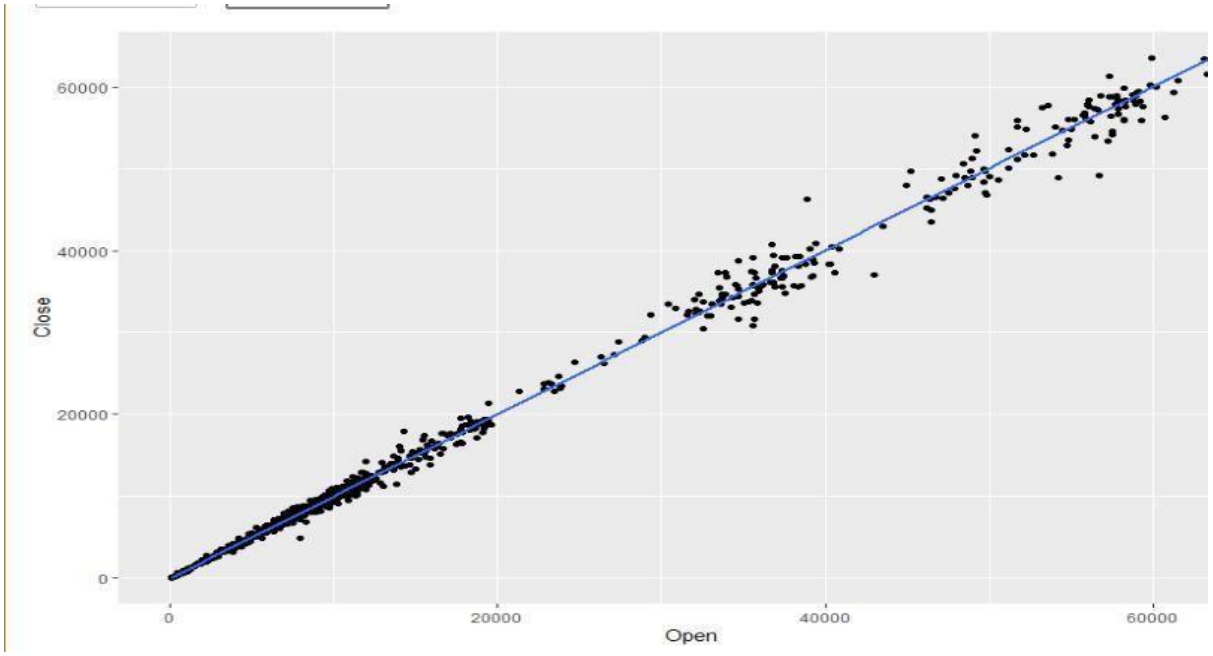


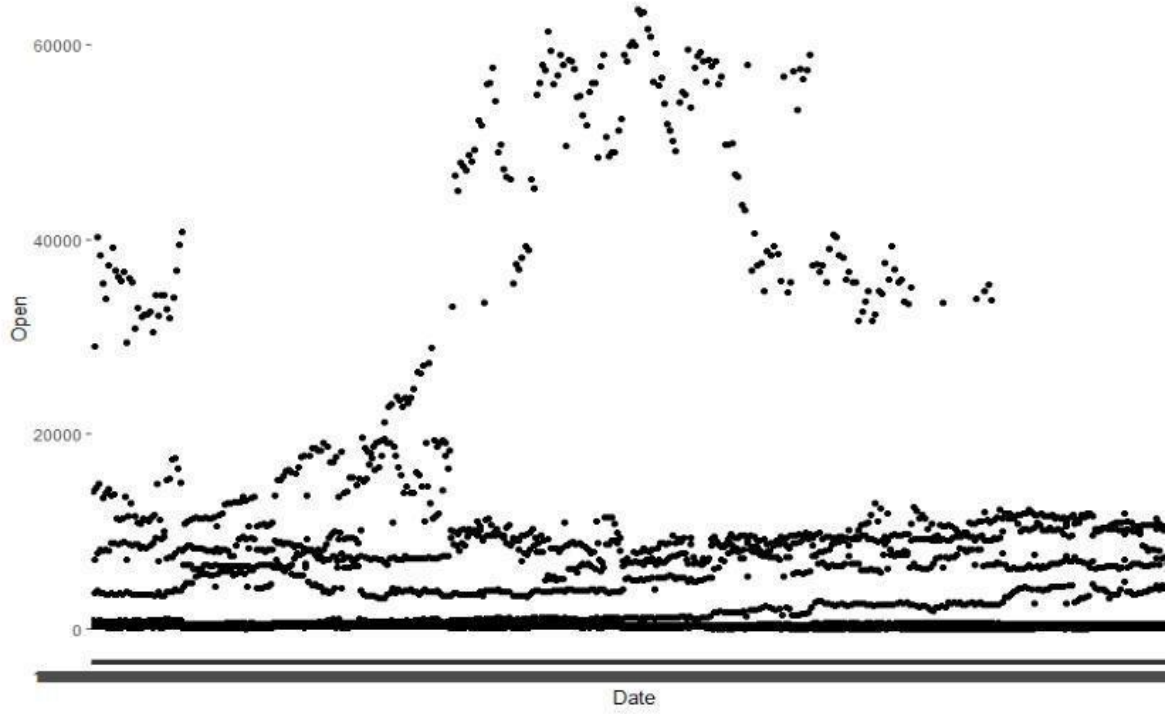
- gg plot for open and close

2) GG smooth for the same above models:

In gg smooth we can see the blue line which makes the transition

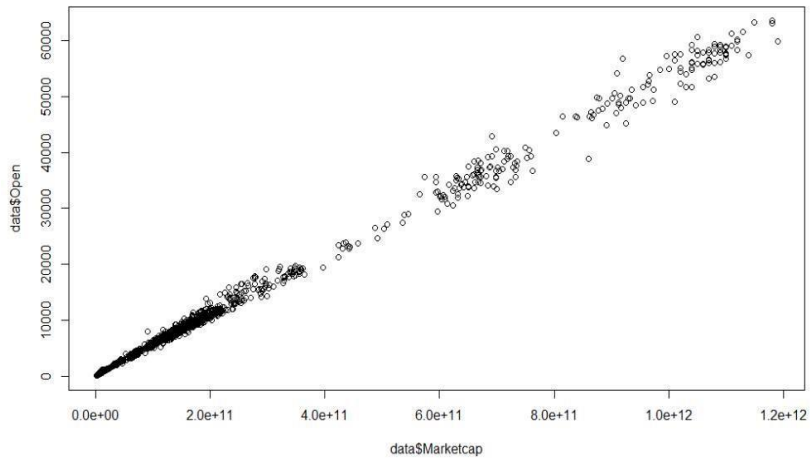


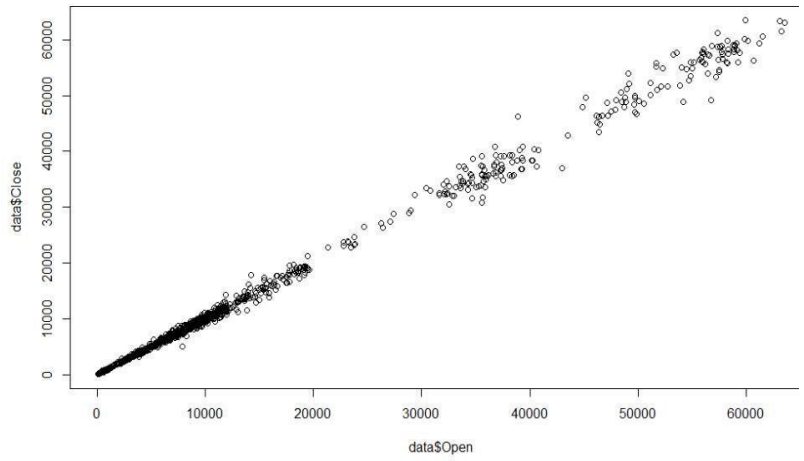


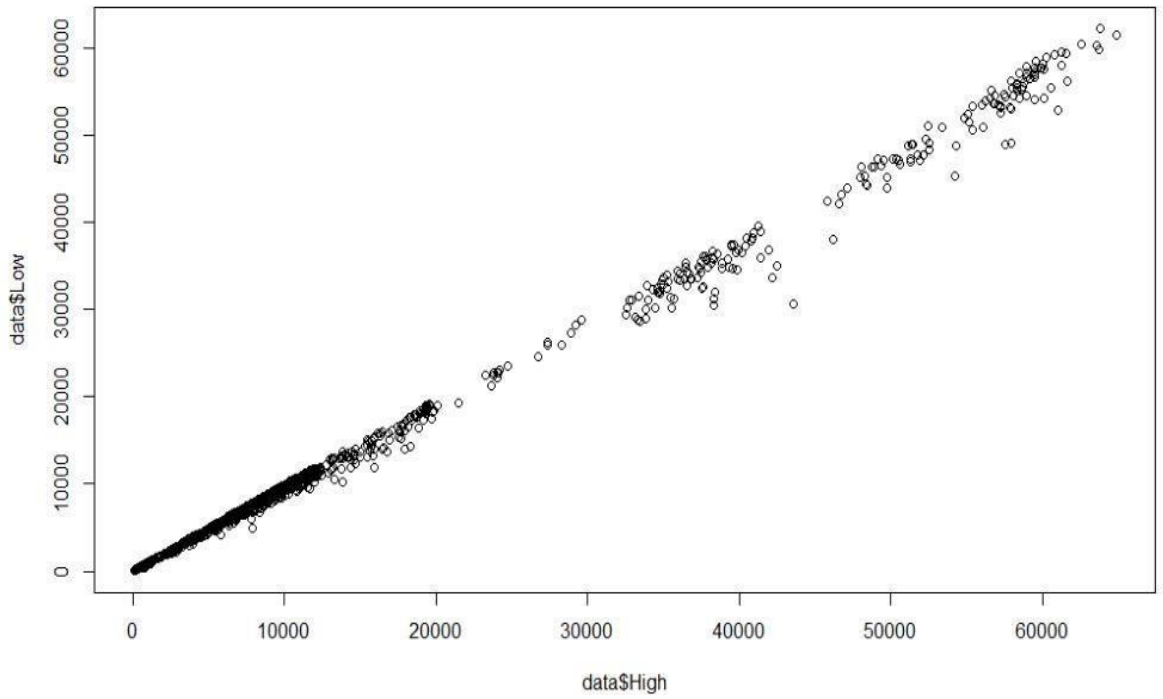


3) Evaluating the linearity assumption among Models:

The relationship between X variable and the mean of Y variable is linearity assumption.

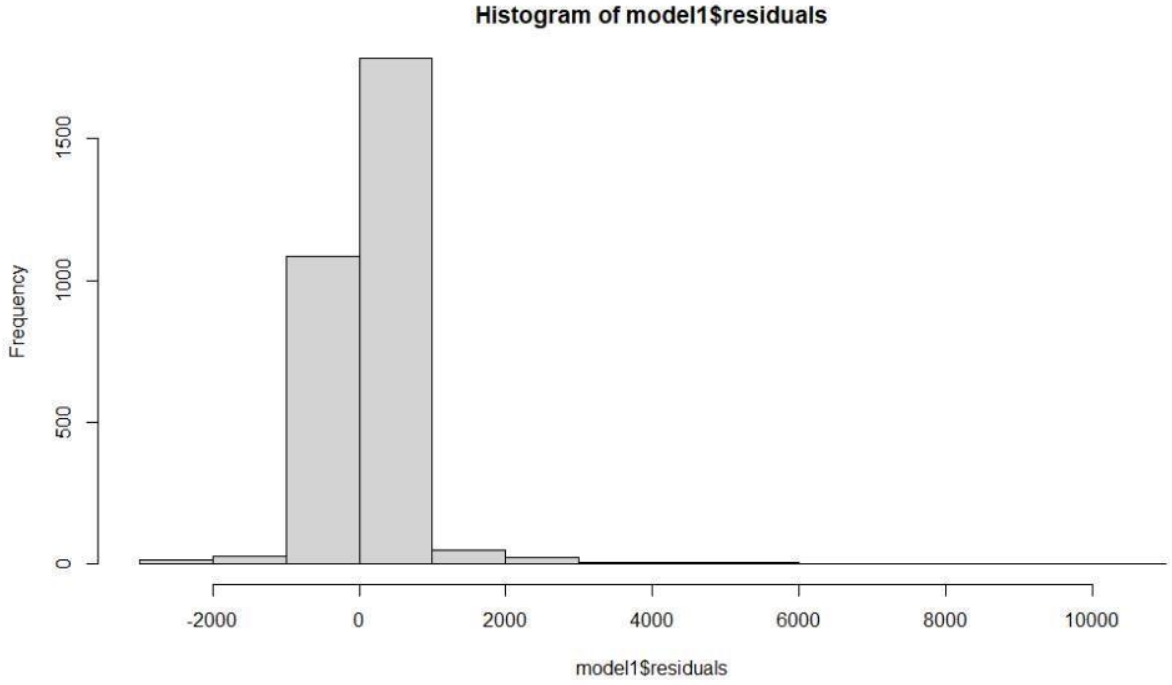


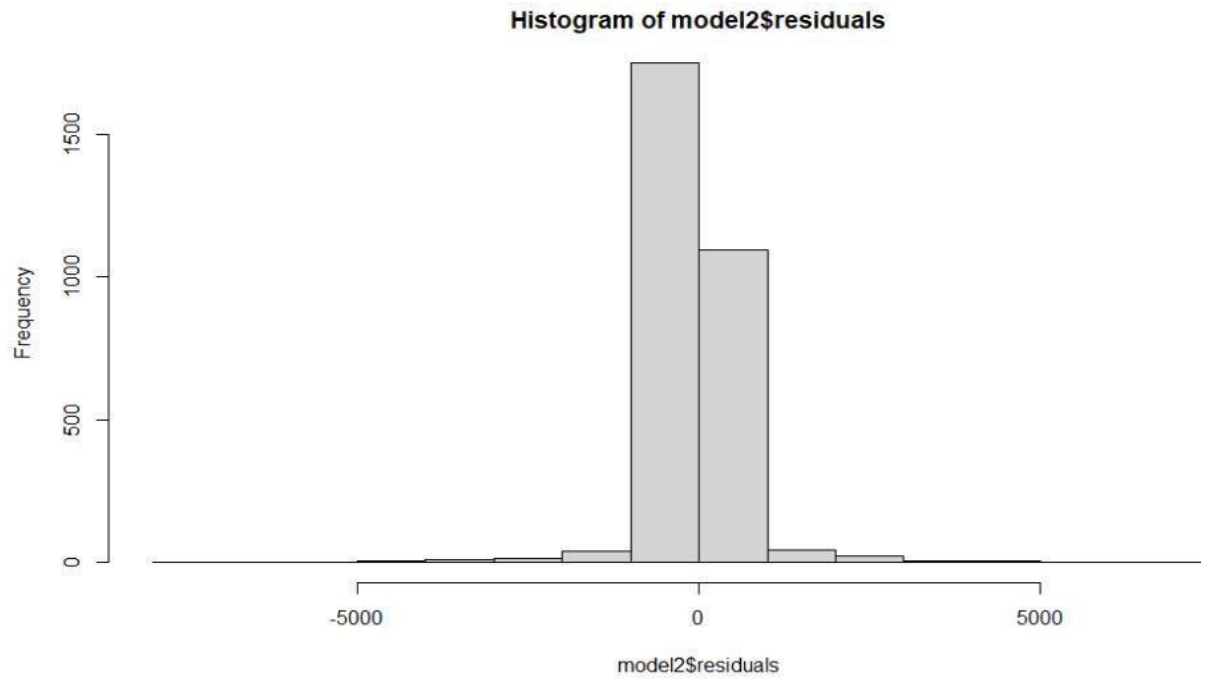


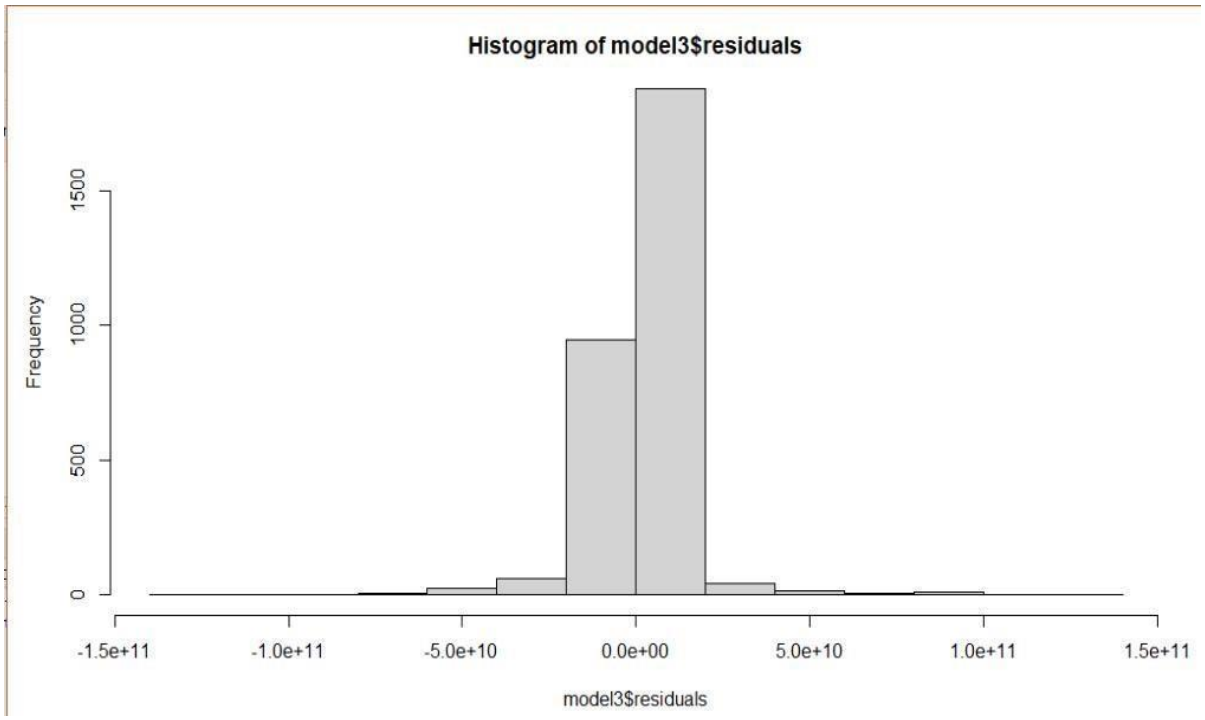


4) Histogram for all the 3 models taken:

A histogram gives the distribution for the data we have taken to assess the central tendency, variability, and shape of the models we are performing.

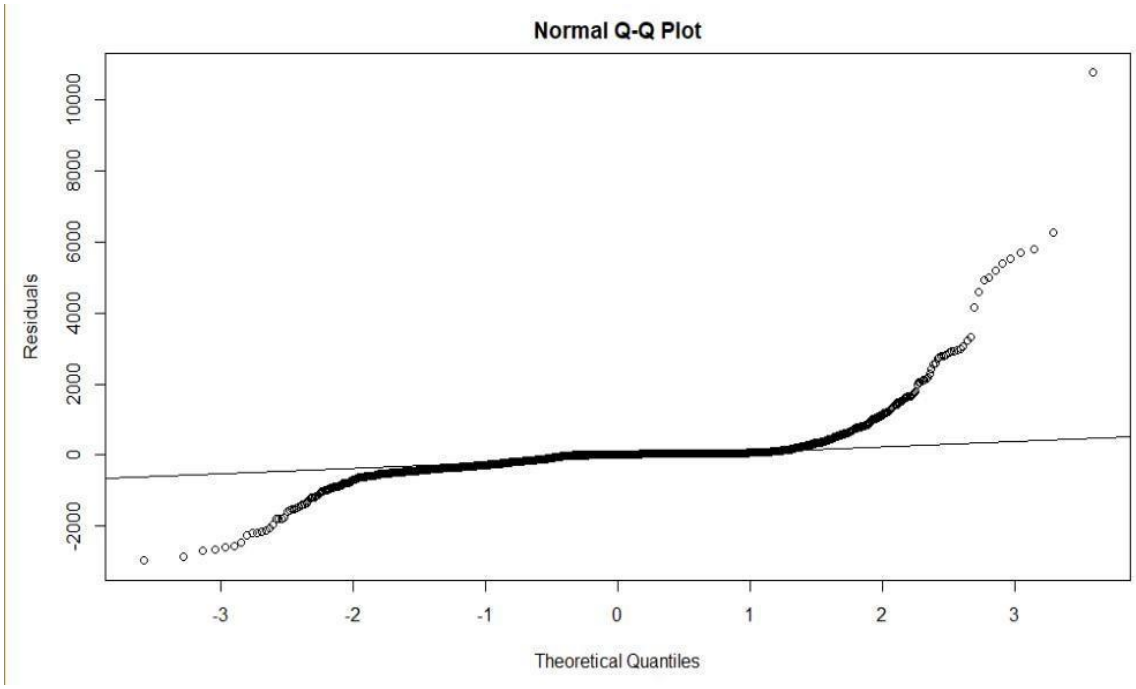



Histogram of model2$residuals
Frequency
0
500
1000
1500
-5000
0
5000
model2$residuals

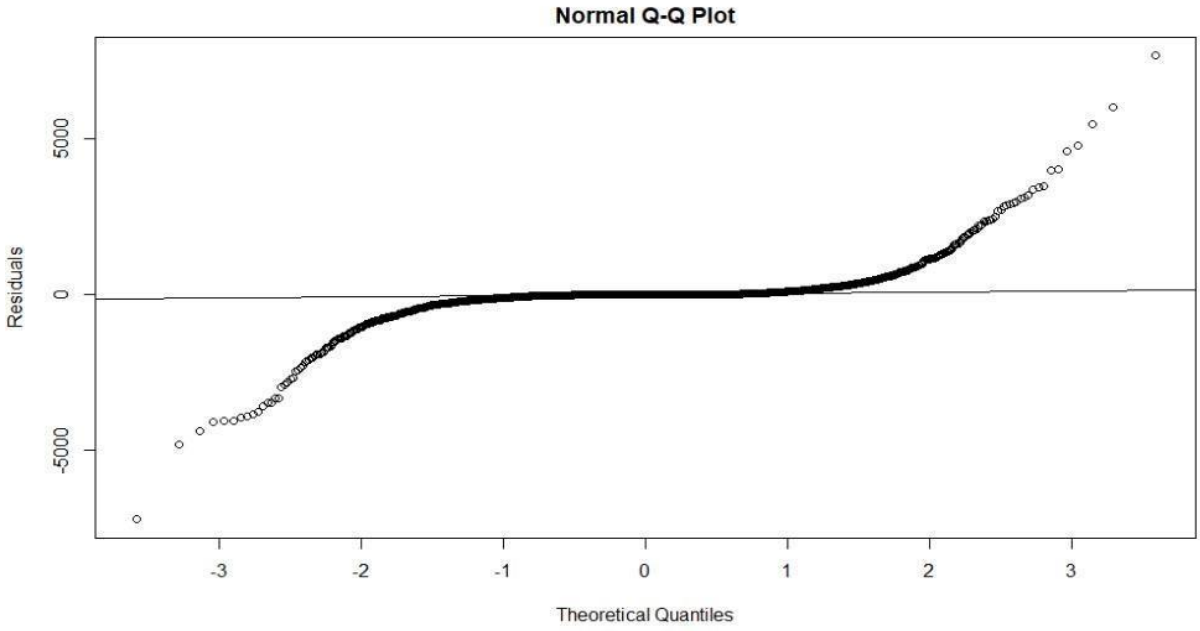

Histogram of model3$residuals
Frequency
0
500
1000
1500
-1.5e+11
-1.0e+11
-5.0e+10
0.0e+00
5.0e+10
1.0e+11
1.5e+11
model3$residuals

Normal QQ plots for all the 2 models:

For model 1

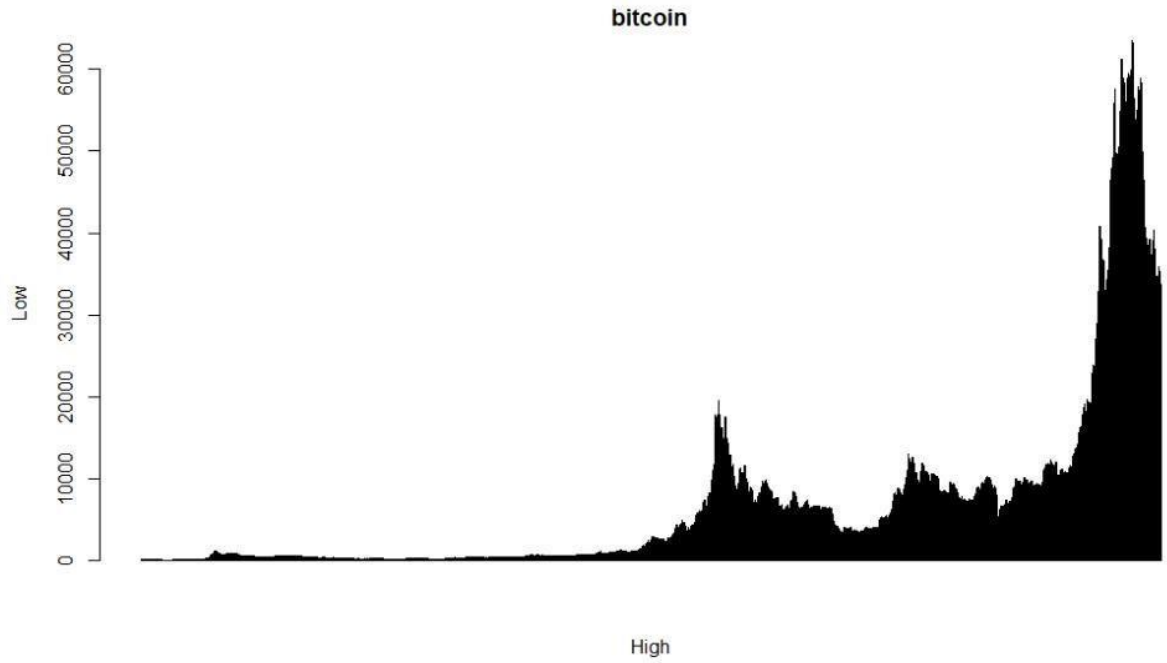


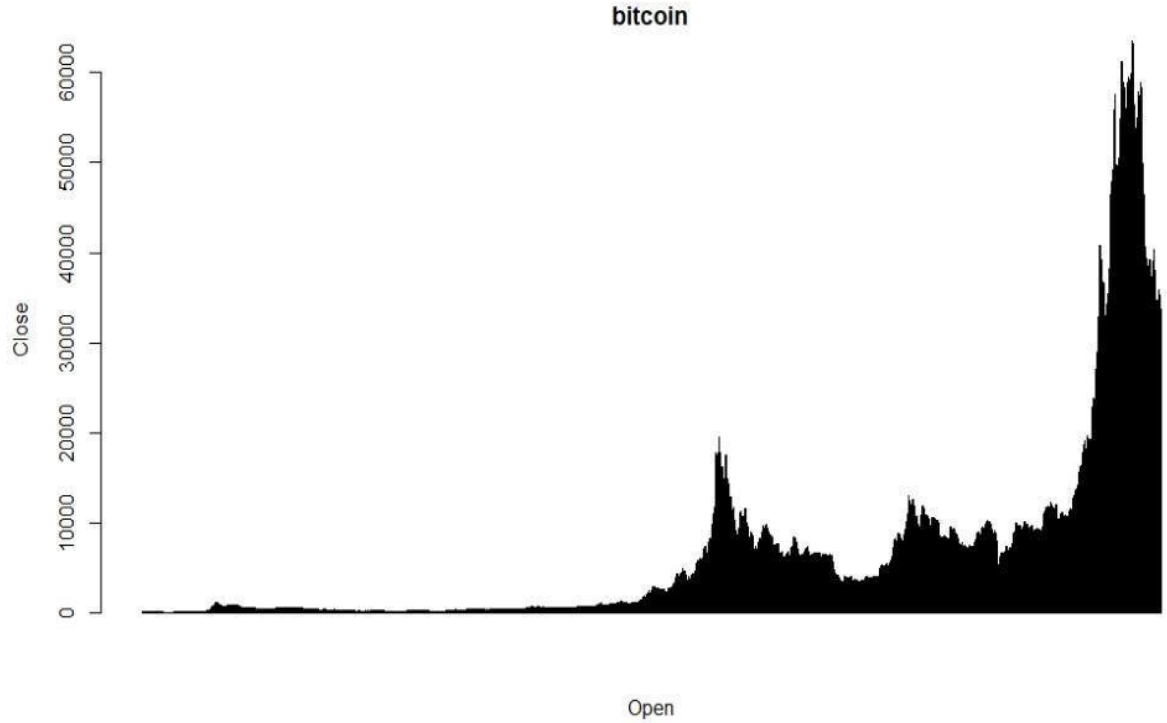
The qq plot gives normal or exponential theoretical distribution for the three models **For model 2**



BAR PLOTS:

The bar plot gives relation between the numerical and categorical variable in the data were performing. Here we had taken for three models.

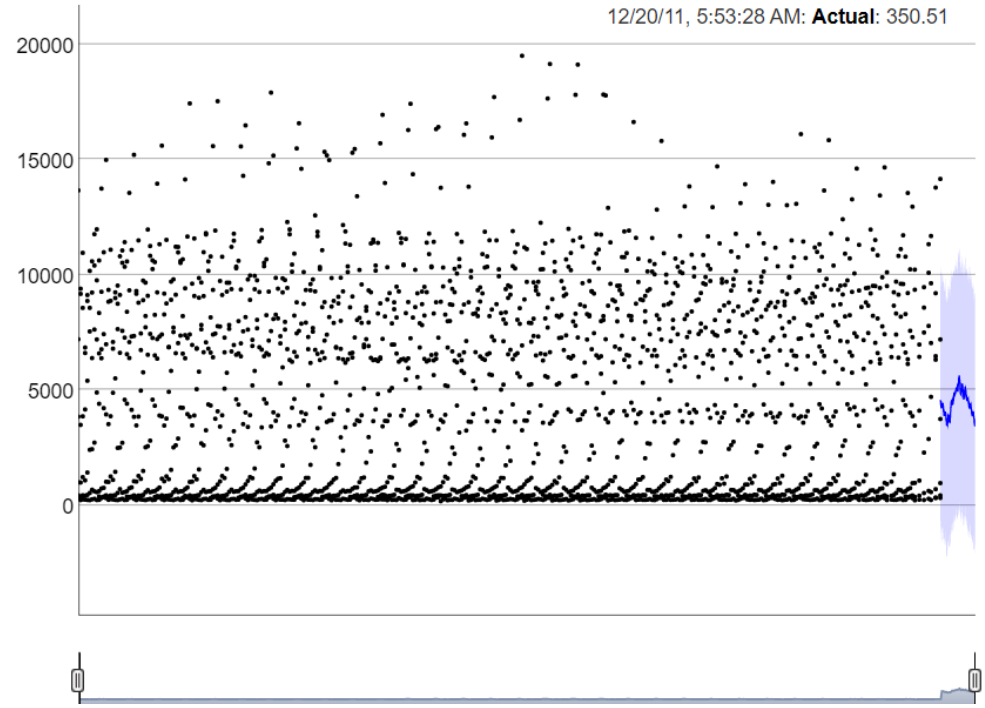




Facebook Prophet:

```
mape <- mean(abs((test$y - predictions$yhat) / test$y)) * 100
mape
```

```
## [1] 89.00724
```



Support vector machine (SVM):

In machine learning, Algorithms in SVM (support vector machine) are supervised learning models with help of learning algorithms which analyse data used for segmentation and regression analysis. It is mostly used in separation problems. In this algorithm, each data object is arranged as a point in the ndimensional space (where n is the number of elements), the value of each element is the value of a specific link. Then, the separation is done by finding a hyper-plane which separates the two categories. In addition to performing line segregation, SVMs can perform indirect segmentation and many more

```
## Support Vector Machines with Linear Kernel
##
## 25 samples
##  3 predictor
## 25 classes: '31676.69373', '32505.65982', '33472.63175', '33560.70784', '33746.00246', '33897.04859', '34235.19345', '346
16.068', '34649.64459', '34668.5484', '35040.83725', '35287.77977', '35615.86927', '35697.60639', '35787.24478', '35862.3777
3', '35867.77774', '36894.40533', '37332.85369', '37334.39953', '37575.17958', '38053.50417', '38347.06323', '38436.96854',
'39208.76599'
##
## Pre-processing: centered (3), scaled (3)
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 20, 23, 18, 24, 22, 24, ...
## Resampling results:
##
##   Accuracy    Kappa
##   0.08630952  0.04089766
##
## Tuning parameter 'C' was held constant at a value of 1
```

ARIMA:

It shows the range of future prices.

```
## method - 2
accuracy(autoarima1)
```

```
##                     ME    RMSE      MAE       MPE     MAPE      MASE         ACF1
## Training set 14.32364 666.26 277.4883 0.1383791 2.680867 0.4654142 -0.006435385
```

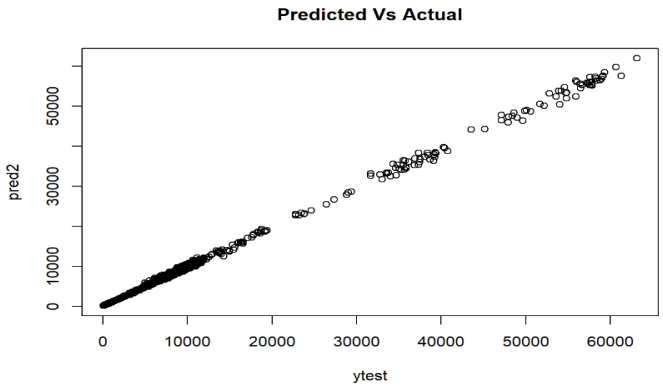
Regression:

```
##
## Call:
## lm(formula = trainClean$Close ~ ., data = trainClean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1668.64   -71.87   -32.56    58.92  1000.02
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.050e+02  4.787e+00  21.935  < 2e-16 ***
## High         5.554e-01  1.257e-02  44.200  < 2e-16 ***
## Low          2.900e-01  9.821e-03  29.526  < 2e-16 ***
## Open        -2.785e-01  1.114e-02 -24.994  < 2e-16 ***
## Volume      -1.621e-09  3.198e-10  -5.068 4.33e-07 ***
## Marketcap    2.300e-08  5.264e-10  43.695  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 171.6 on 2402 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 2.093e+06 on 5 and 2402 DF,  p-value: < 2.2e-16
```

Ridge regression:

```
## [1] 248.5774
```

```
plot(ytest, pred2, main ='Predicted Vs Actual')
```



Conclusion:

We have implemented four different models (Prophet, SVM, ARIMA). The objective of this study was to compare Bitcoin's future values using these algorithms. Till now we have found that ARIMA performs the best in prediction of future values. SVM model shows around 52% accuracy and talking about

Prophet, it performs very good at small changes in values but performs slightly low when the values difference is large. There are lot more models that are built for predicting values using previous data, we have used the most famous among them. After comparing the results of above 3 prediction algorithms the Facebook prophet algorithm is giving the results with almost 89% accuracy Overall, predicting price-related changing is difficult given the relative strength of the market. And also, we know that prices are more dependent on future predictions than historical data. However, using deep neural networks has given us a better understanding of Bitcoin, as well as LSTM structures. Present work involves using the hyper parameter, to obtain a more accurate network configuration. Also, other factors can be considered (although from our experiment with Bitcoin, many factors did not always lead to better results). Smaller economic factors may be included in the model for a better understanding.

Future Scope:

The present study provides a comparison of different prediction models for Bitcoin's future values. However, there are several avenues for future research in this area. Firstly, the study only considered a limited number of prediction models, and there are several other algorithms that can be explored, such as Long Short-Term Memory (LSTM) networks with attention mechanisms, convolutional neural networks (CNN), and hybrid models. This only considered Bitcoin's historical price data, and other factors such as market sentiment, trading volume, network difficulty, and regulatory changes could be considered to enhance the accuracy of the prediction models and also can be focused on short-term price prediction, and future studies could consider long-term prediction models for Bitcoin's future values. This could be extended to other cryptocurrencies, such as Ethereum, Litecoin, and Ripple, to compare the performance of different prediction models across different cryptocurrencies and could be extended to evaluate the performance of ensemble methods, which combine the predictions of multiple models to improve accuracy, and reinforcement learning-based models, which can adapt to changing market conditions and improve their predictions over time.

Overall, the future scope of this study involves exploring new prediction models, incorporating additional factors, and evaluating the performance of the models across different cryptocurrencies and time frames.

References:

1. Bitcoin Price Prediction and Analysis Using Deep Learning Models Temesgen Awoke, Minakshi Rout, Lipika Mohanty, and Suresh Chandra Satapathy (LINK)
2. Bitcoin Forecasting Using ARIMA and PROPHETIs¸ıl Yenido˘gan, Aykut C¸ ayırMIS DepartmentKadir Has University, Istanbul, Turkey{isil.yenidogan, aykut.cayir}@khas.edu.trOzan Kozan, Tu˘gc¸e Da˘g, C¸ i˘gdem ArslanMIS DepartmentKadir Has University, Istanbul, Turkey{sukruozan.kozan, tugce.dag,cigdem.arslan}@stu.khas.edu.tr(LINK)
3. A Comparative Study of Bitcoin Price PredictionUsing Deep LearningSuhwan Ji, Jongmin Kim and Hyeonseung Im (LINK)
4. Predicting Stock Market Price: A Logical Strategyusing Deep LearningMilon BiswasComputer Science and EngineeringBangladesh University ofBusiness and TechnologyDhaka, Bangladeshmilonbiswas4702@gmail.comAtanu
5. Garcia, D. & Schweitzer, F. (2015). Social signals and algorithmic trading of Bitcoin. Royal Society Open Science 2(9), 150288. Retrieved from https://dx.doi.org/10.1098/rsos.150288
6. Kathyayini, R. S., Jyothi, D. G. & Crypto, ". Currency Price Prediction using Machine Learning". International Journal of Advanced Research in Computer and Communication Engineering.
7. 7. Nakamoto, S. & Shah, ;. D. (2017). Trading Bitcoin and Online Time Series Prediction. NIPS 2016 Time Series Workshop .