TABLE OF CONTENTS

S. No	Chapter	Page Number
1	Abstract	3
2	Introduction	3-4
3	Literature Survey / Requirements	5
4	Proposed System / Module(s) description	6-9
5	Results and Discussion	10-21
6	Conclusion	22
7	Reference	23

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)

Keywords: models, data, price, predict, bitcoin

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

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 on predictive 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 / Requirements

- 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.

Proposed System / Module(s) description

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&includeAdjustedClose=true

Software used: R studio.

Facebook's Prophet:

Actually the trends nowadays are non linear 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

The prophet Algorithm is used for many applications across the face book for producing many reliable forecasts and for planning the goals etc. we have found that this prophet is performing better than any other algorithms in many cases. We are going to fit the models in Stan so that we can get the outputs for forecasts in just a few seconds of time

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

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

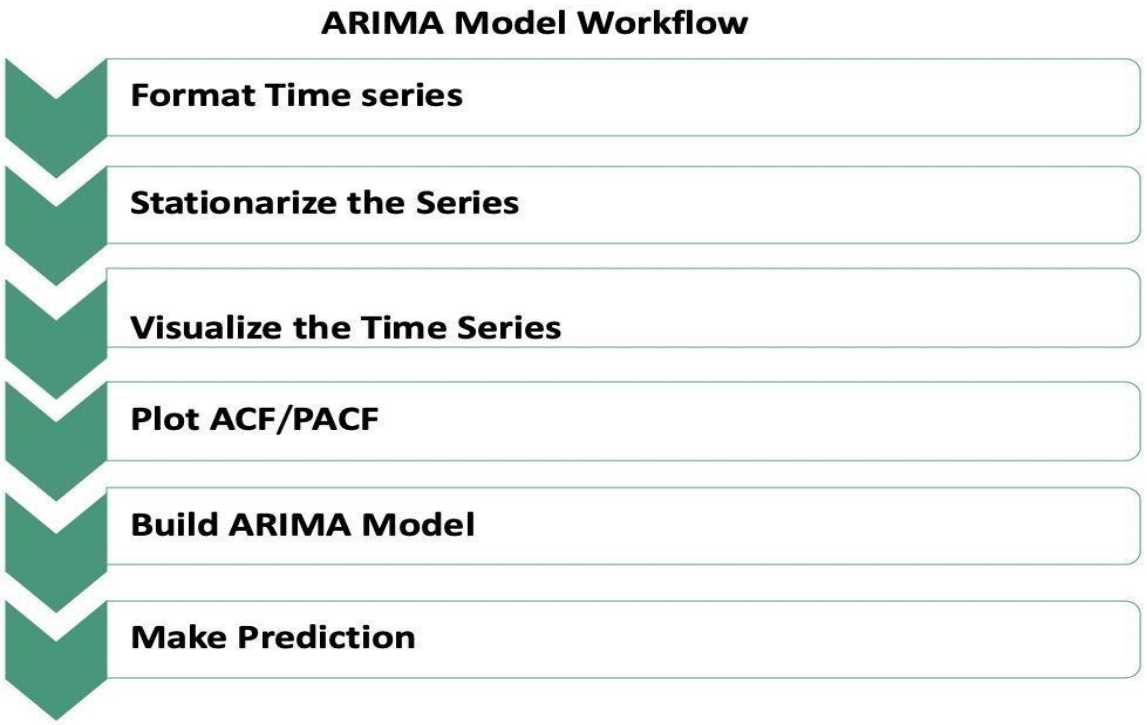
D is the number of differencing required to make the time series stationary

If the time line, 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 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.

$$\lambda\|\mathbf{w}\|^2 + \left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0, 1 - y_i(\mathbf{w}^T\mathbf{x}_i - b)\right)\right],$$

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 methods 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

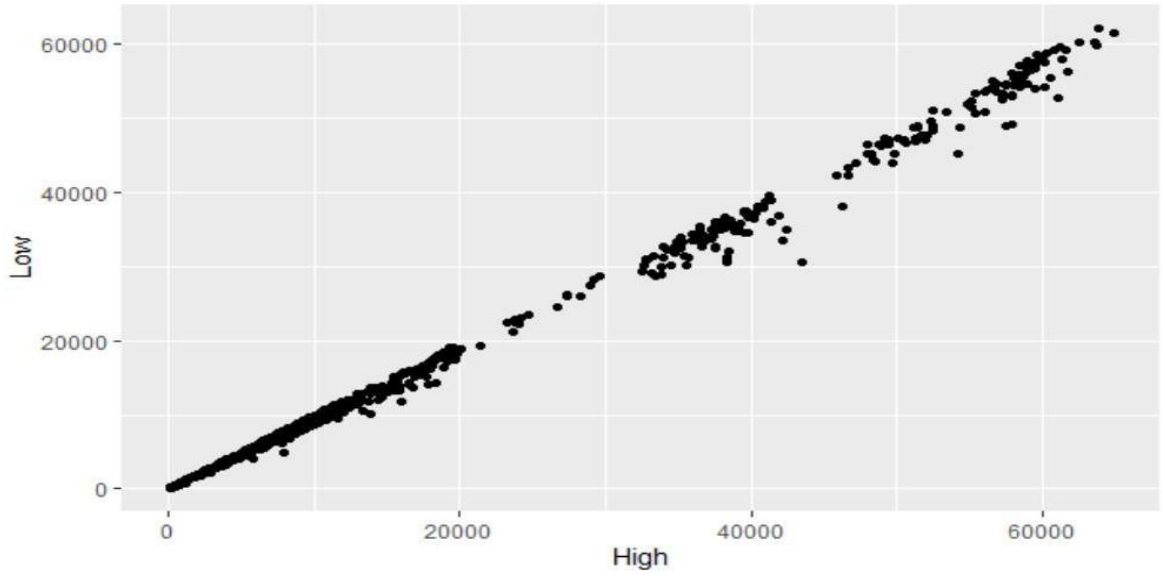
Results and Discussion

Models we taken:

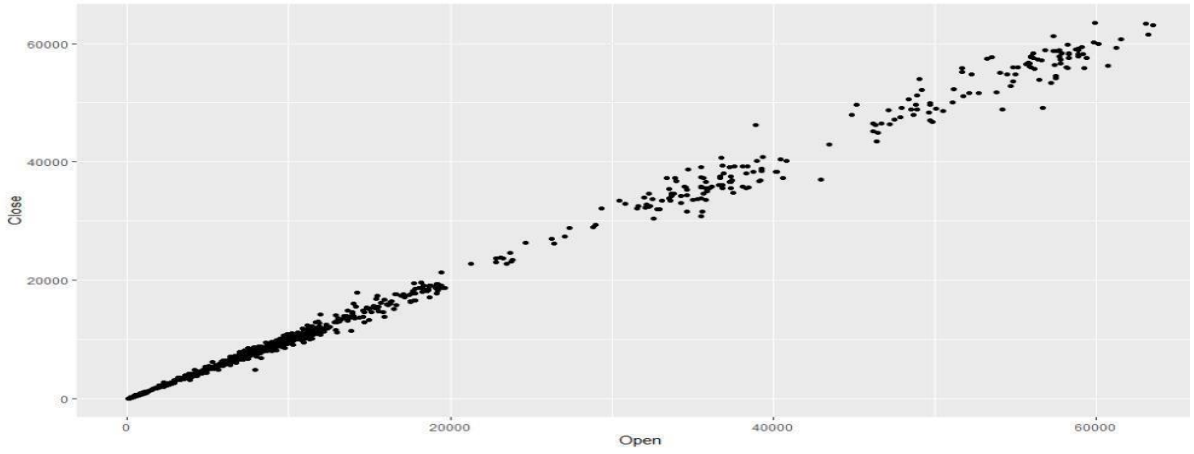
➔ Assigning model1 for low and high
➔ Assigning model2 for open and close
➔ Assigning model3 for date and open

PLOTS AND GRAPHS:

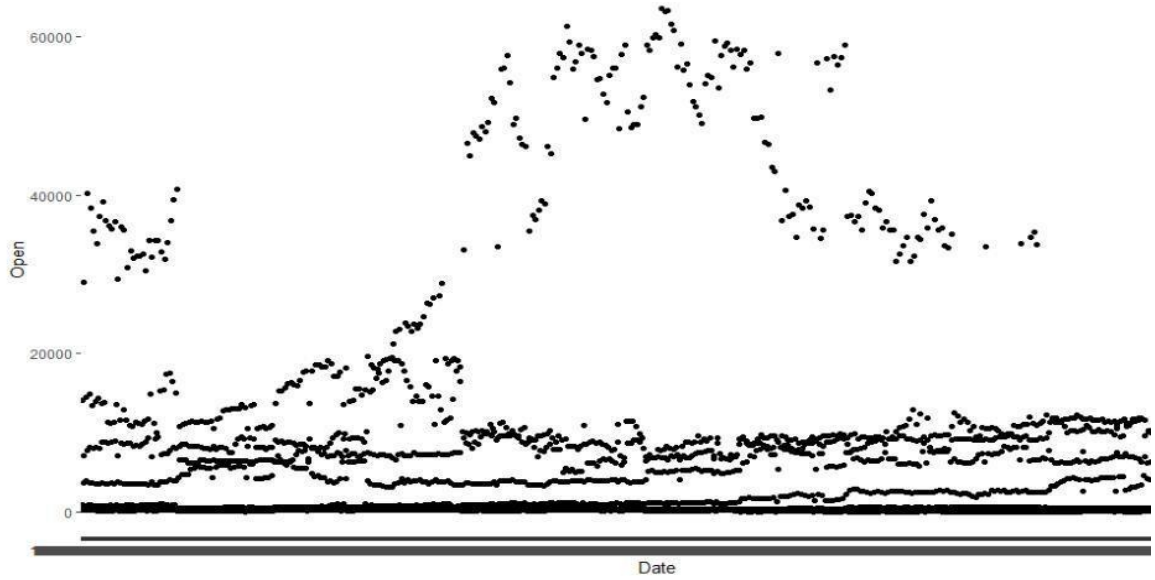
➔ **visualization of gg plots for different models**



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



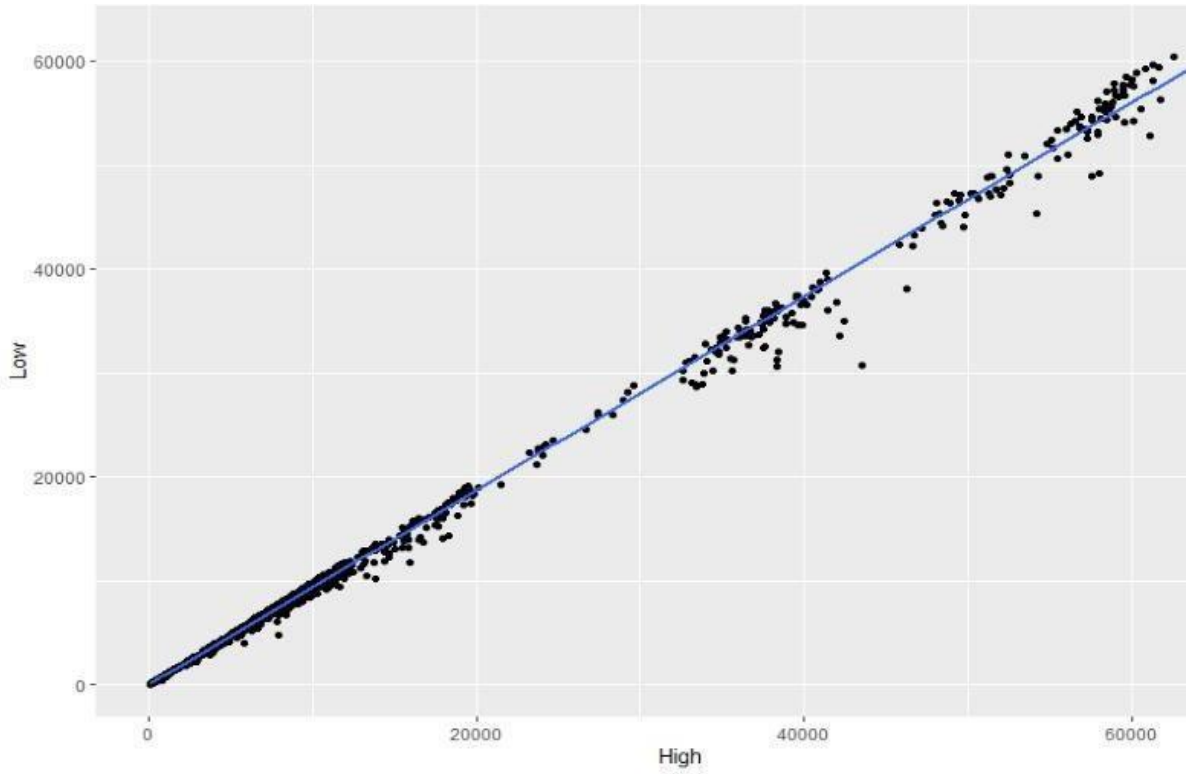
GG plot for open and close

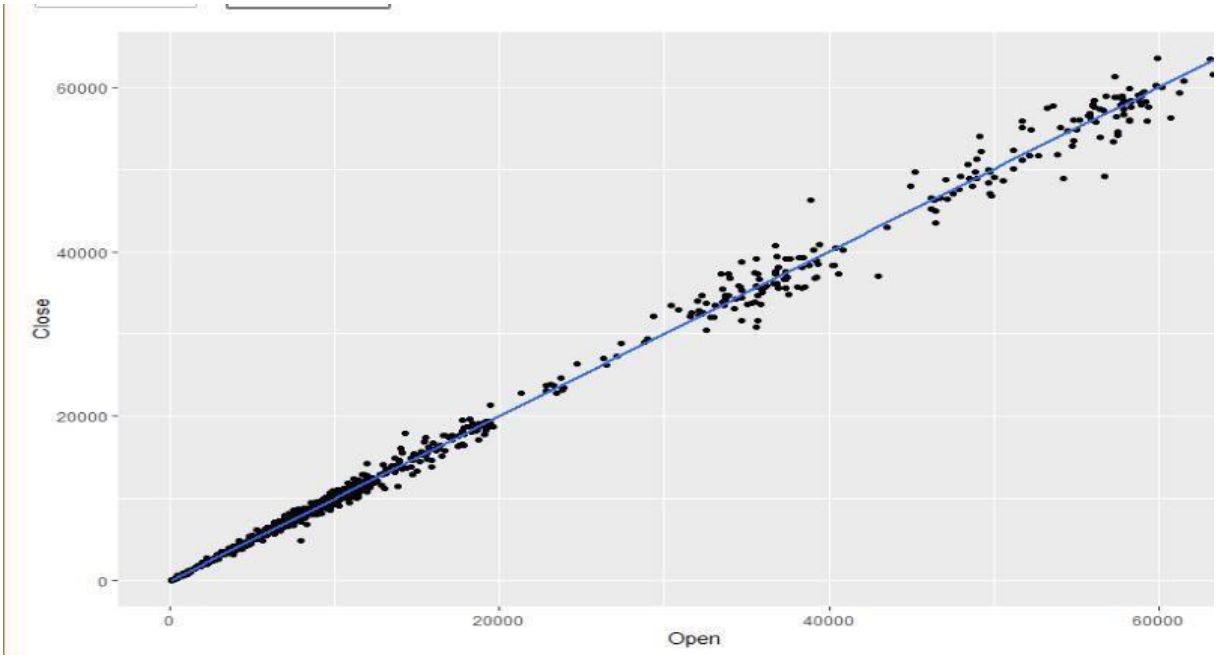


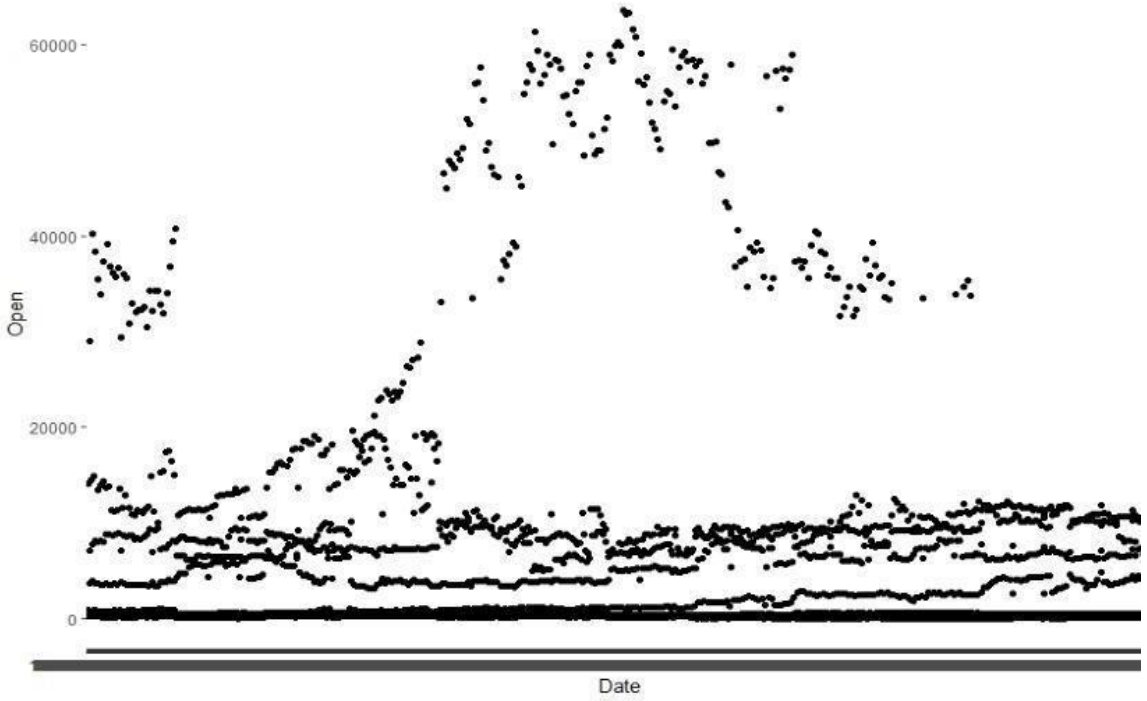
gg plot for open and close

➔ **GG smooth for the same above models:**

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

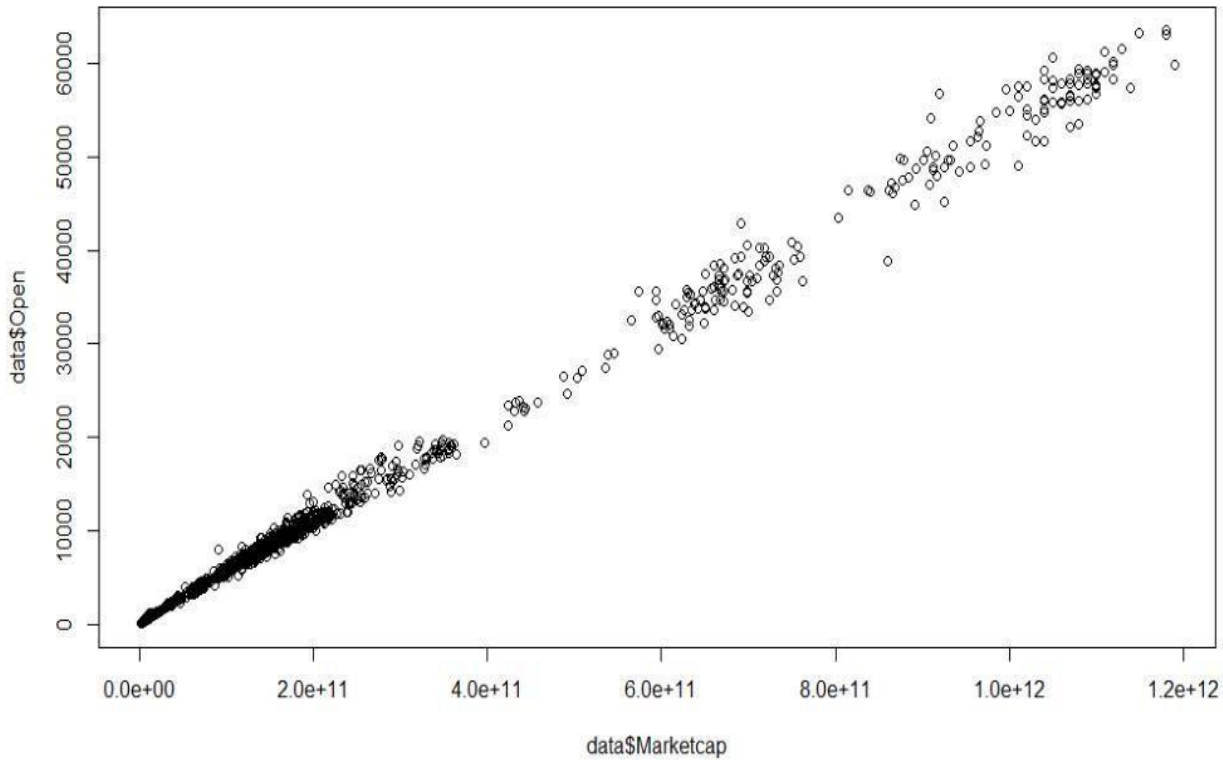


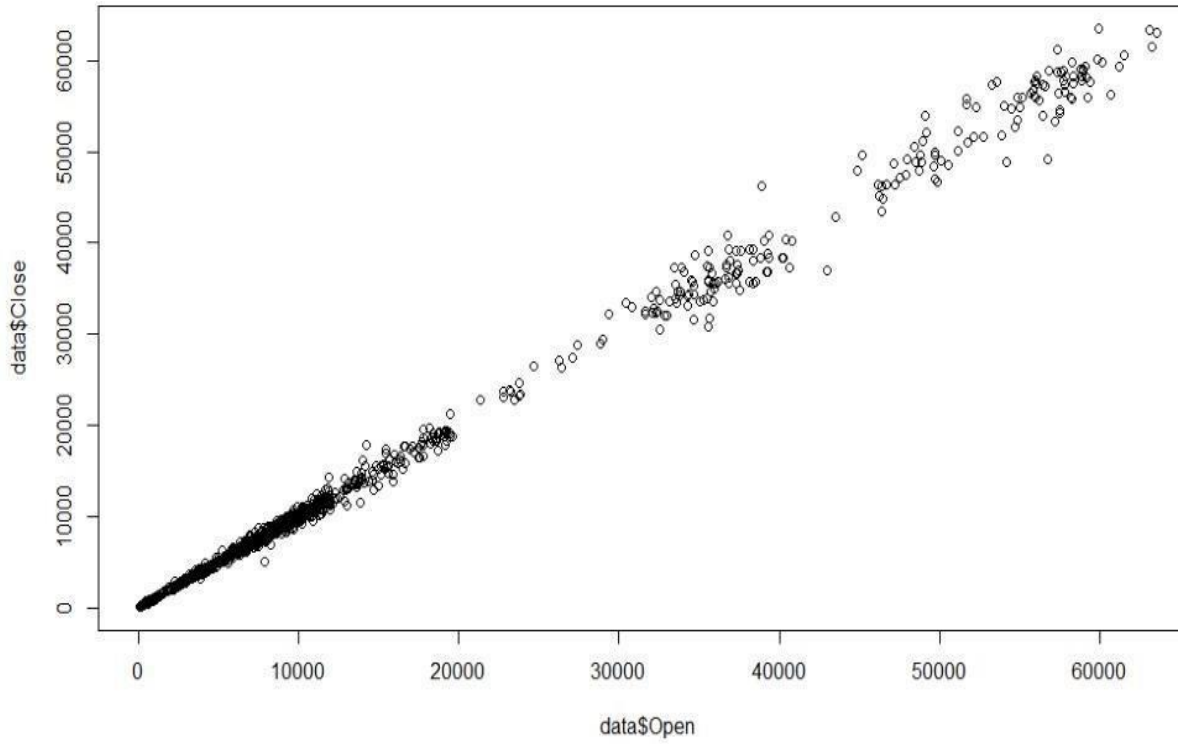

60000
40000
20000
0
Close
0
20000
40000
60000
Open

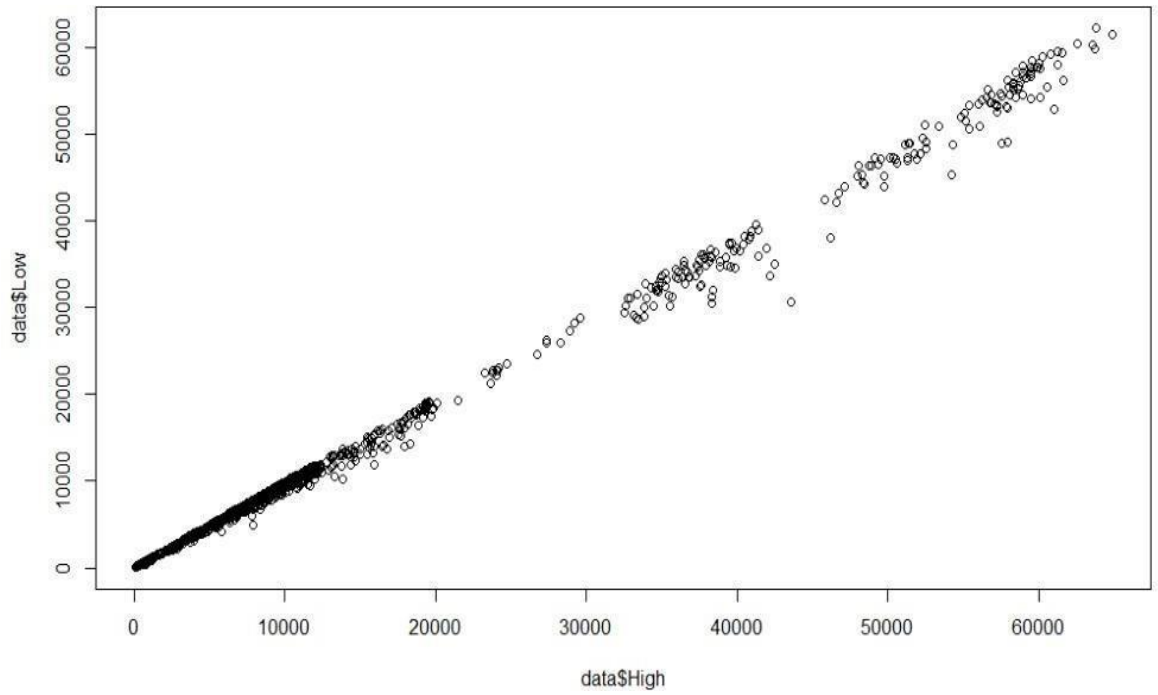

60000
40000
20000
0
Open
Date

➔ **Evaluating the linearity assumption among Models:**

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

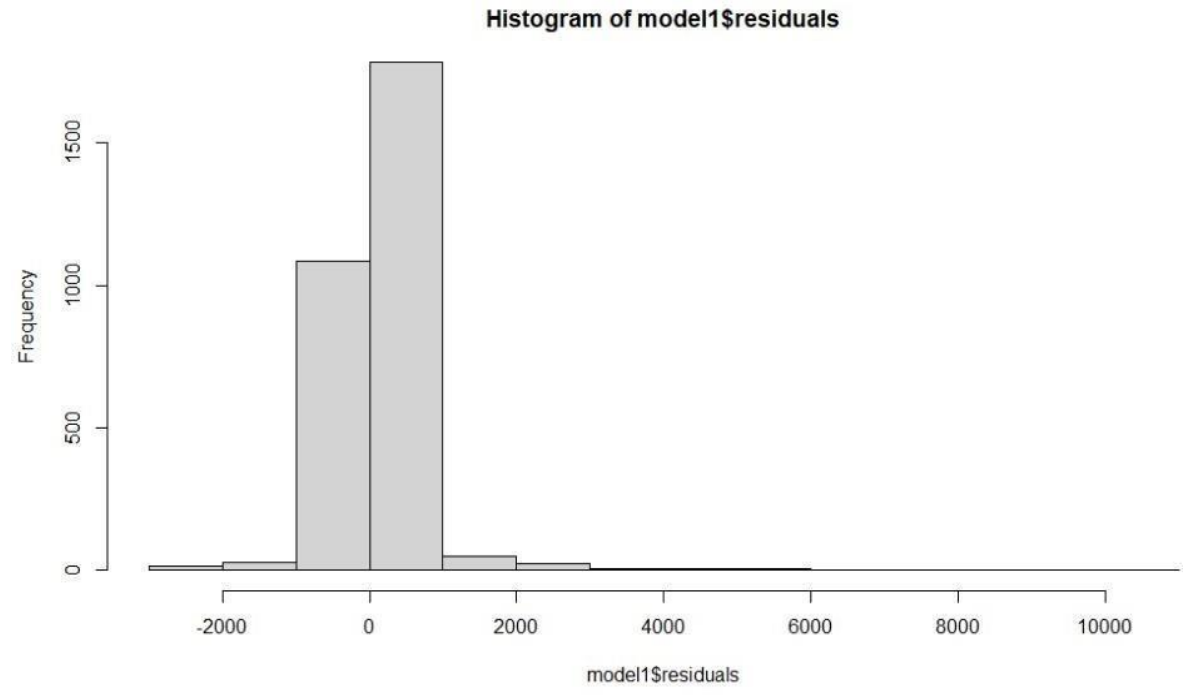


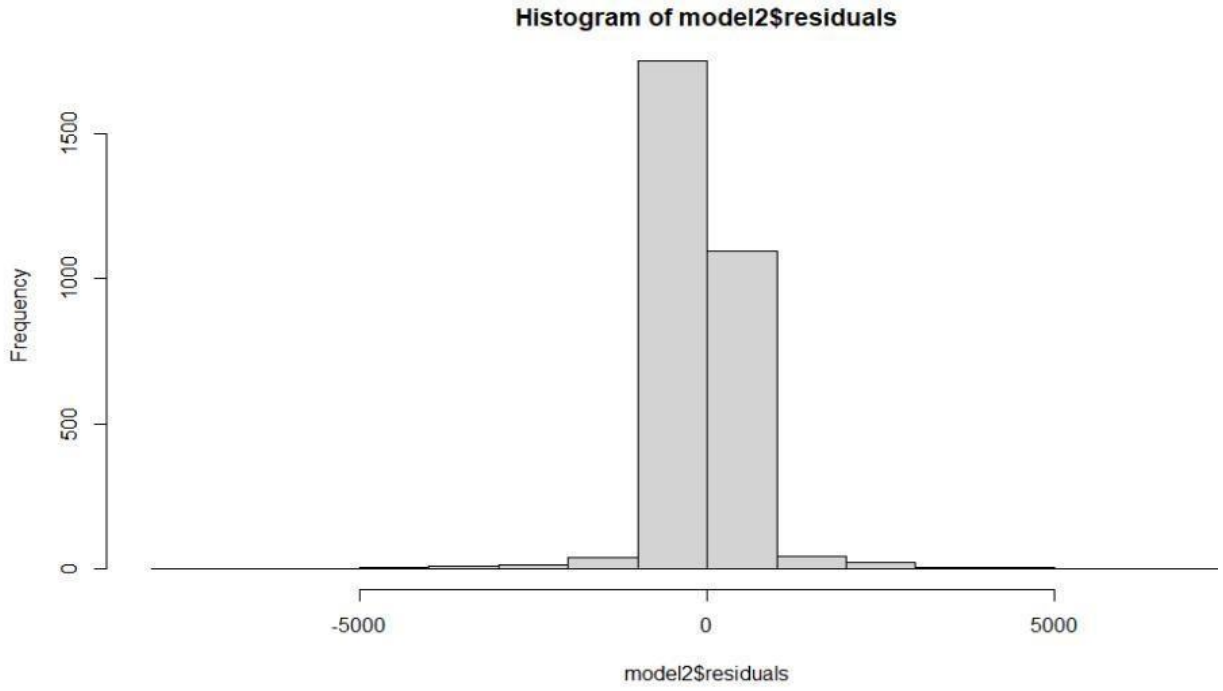


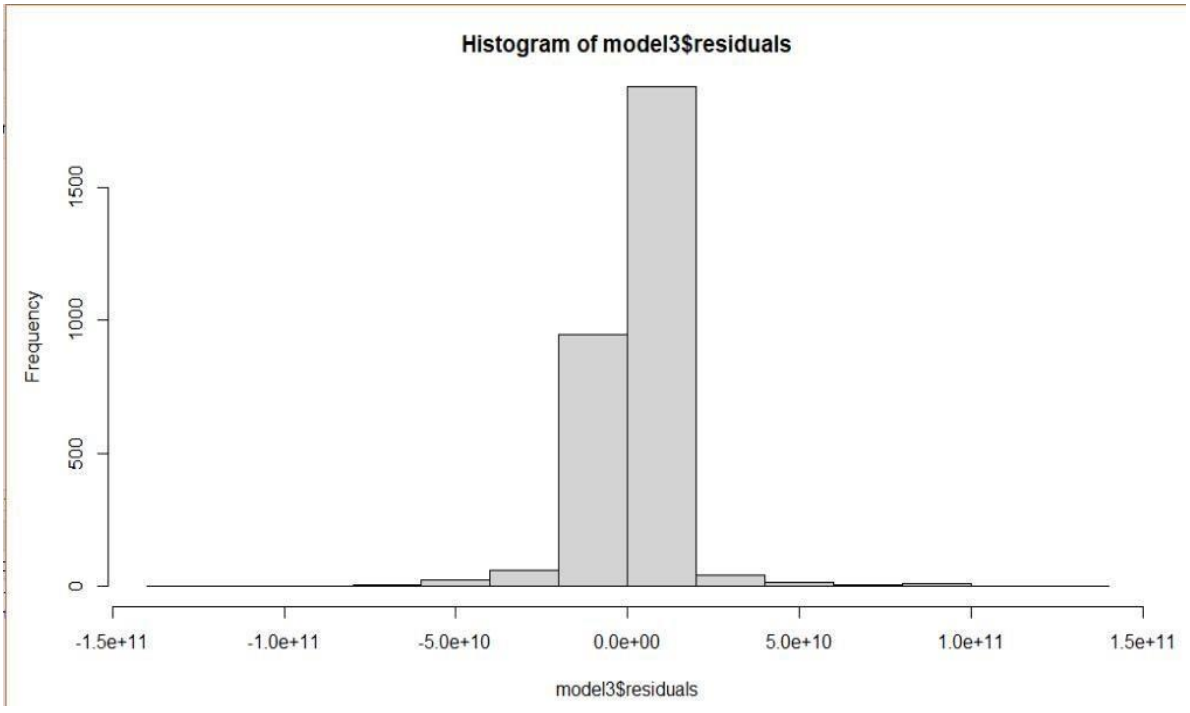


➔ **Histogram for all the 3 models taken:**

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

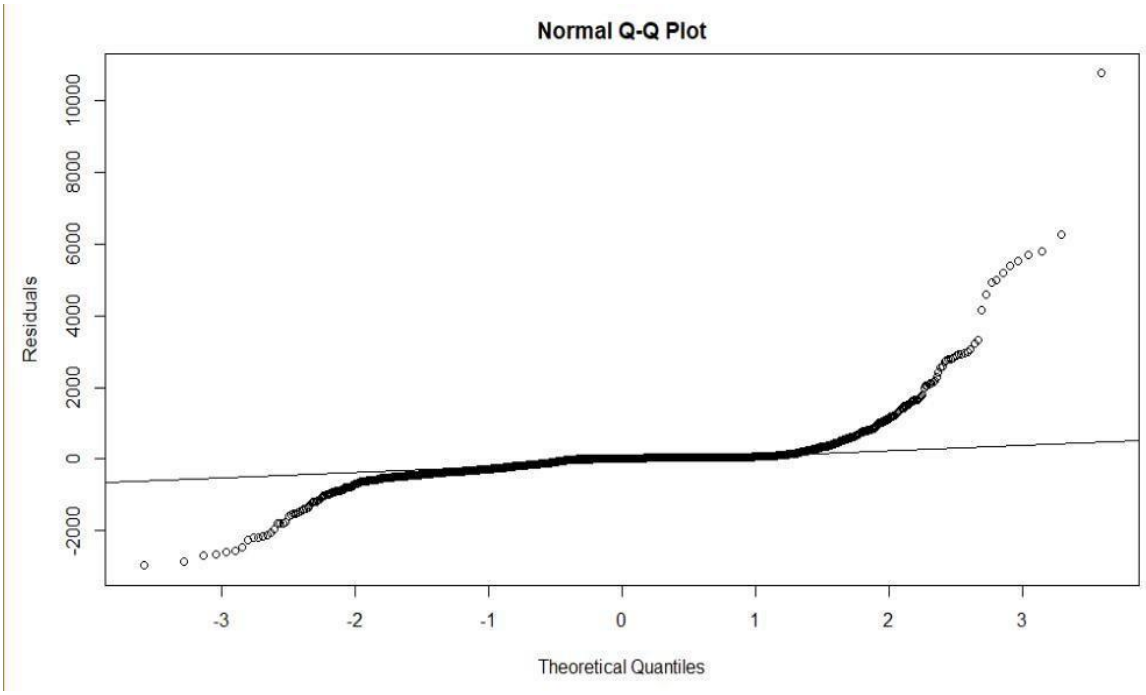



Histogram of model2$residuals
Frequency
0
500
1000
1500
-5000
0
5000
model2$residuals


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

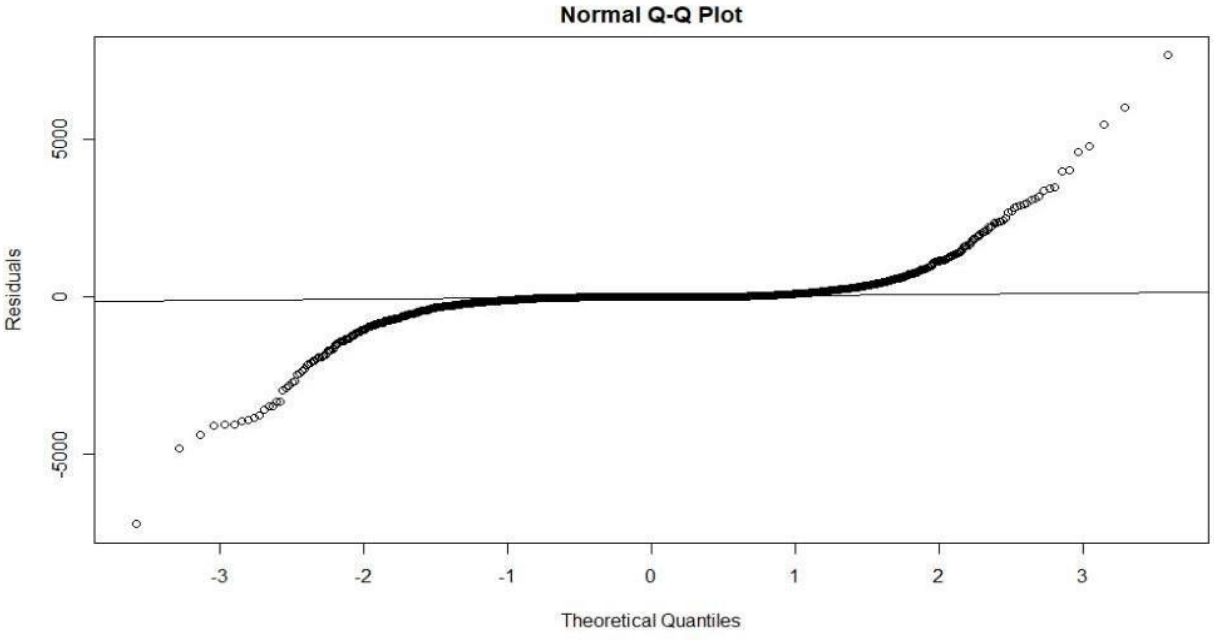
#normal QQ plots for all the 3 models:

For model 1



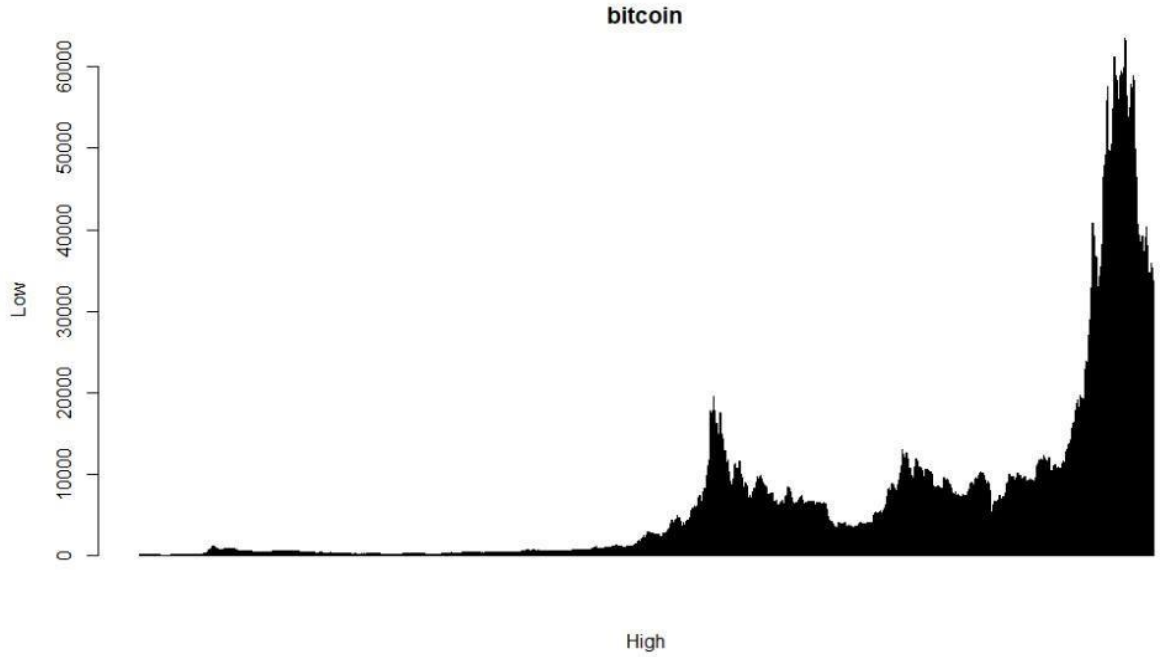
The qq plot gives normal or exponential theoretical distribution for the three models

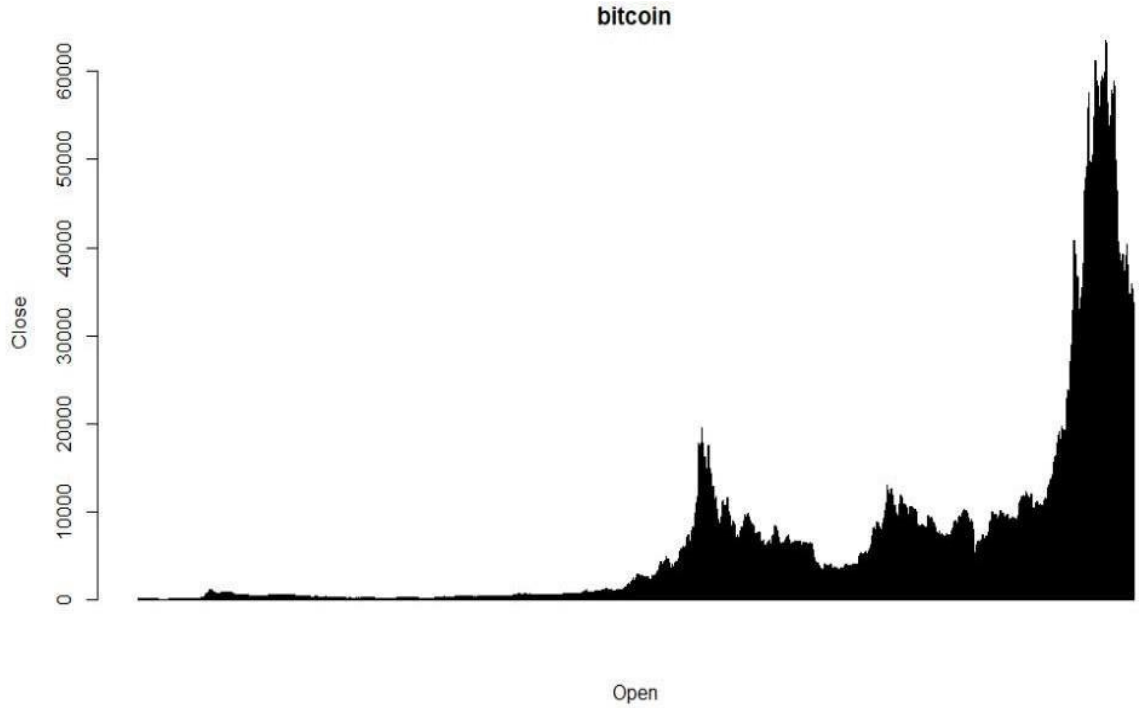
For model 2



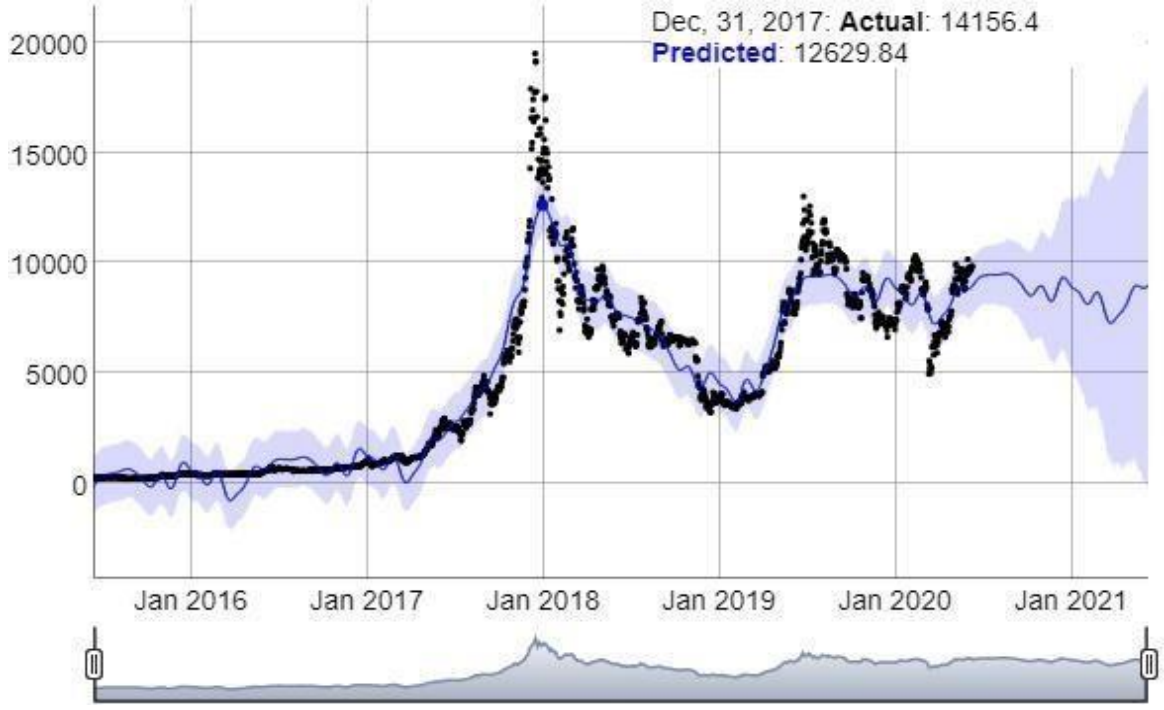
➔ **BAR PLOTS:**

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





Facebook's Prophet:



888

In the above graph blue line is predicted values and black dotted lines are actual value. So that you can see the Actual and Predicted values for each day.

Support vector machine (SVM):

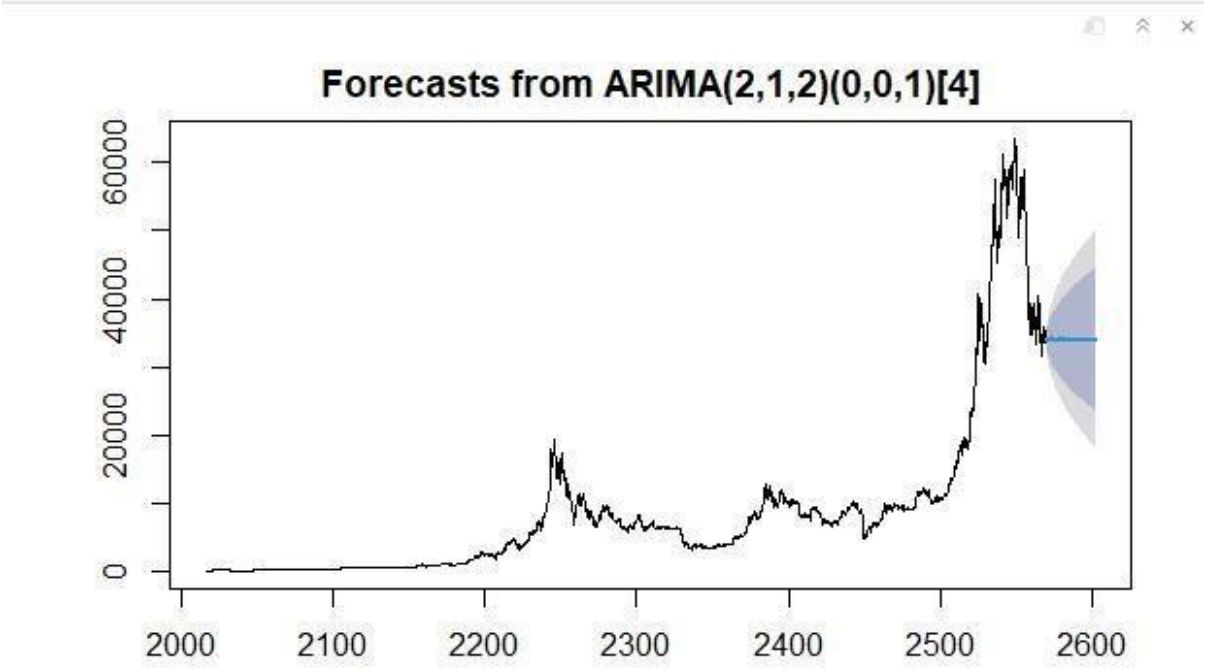
In machine learning, Algorithms in SVM (support vector machine) are supervised learning models with help of learning algorithms which analyze data used for segmentation and regression analysis. It is mostly used in separation problems. In this algorithm, each data object is arranged as a point in the n-dimensional 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

29 samples
 3 predictor
29 classes: '31637.78006', '32186.27767', '32505.65982', '33472.63175',
'33560.70784', '33572.11765', '33723.02898', '33746.00246', '34662.43589',
'34668.5484', '35040.83725', '35287.77977', '35551.95873', '35552.51715',
'35615.86927', '35678.1292', '35697.60639', '35787.24478', '35867.77774',
'36684.92452', '36702.59937', '37332.85369', '37334.39953', '37575.17958',
'38053.50417', '38436.96854', '39097.8609', '39208.76599', '40218.47786'

Pre-processing: centered (3), scaled (3)
Resampling: Cross-Validated (10 fold, repeated 3 times)
Summary of sample sizes: 22, 26, 25, 24, 28, 27, ...
Resampling results:

  Accuracy   Kappa
  0.1780612  0.07502403

Tuning parameter 'C' was held constant at a value of 1
```

ARIMA:

It shows the range of future prices



The table below shows the mean squared error

```
mean_squared_error(df['Open'][2:len(df)],x)
```

```
423537.3748746528
```

Regression:

```
      12       13       15       20       25       28
203.7242 200.4482 202.4182 207.3700 209.3324 216.0503
[1] "RMSE"
[1] NaN
[1] "R-Squared"
[1] 0.9996439
[1] "MSE"
[1] 45197.47
[1] "MAE"
[1] 107.4341
[1] "Adjusted R squared"
[1] 0.99977
```

Ridge regression:

```
(Intercept) 6.711254e+03 6.711243e+03 6.711228e+03 6.711207e+03
(Intercept) .            .            .            .
High        1.095542e-06 1.448235e-06 1.914478e-06 2.530821e-06
Low         1.173388e-06 1.551144e-06 2.050517e-06 2.710655e-06
Open        1.129101e-06 1.492600e-06 1.973126e-06 2.608349e-06
Volume      5.455924e-13 7.212392e-13 9.534336e-13 1.260379e-12
Marketcap   6.049238e-14 7.996723e-14 1.057117e-13 1.397443e-13

(Intercept) 6.711181e+03 6.711145e+03 6.711098e+03 6.711037e+03
(Intercept) .            .            .            .
High        3.345586e-06 4.422650e-06 5.846451e-06 7.728607e-06
Low         3.583315e-06 4.736910e-06 6.261880e-06 8.277772e-06
Open        3.448072e-06 4.558127e-06 6.025538e-06 7.965341e-06
Volume      1.666141e-12 2.202528e-12 2.911592e-12 3.848918e-12
Marketcap   1.847330e-13 2.442049e-13 3.228223e-13 4.267482e-13

(Intercept) 6.710955e+03 6.710847e+03 6.710704e+03 6.710515e+03
(Intercept) .            .            .            .
High        1.021667e-05 1.350566e-05 1.785338e-05 2.360060e-05
Low         1.094262e-05 1.446530e-05 1.912193e-05 2.527746e-05
Open        1.052960e-05 1.391931e-05 1.840015e-05 2.432331e-05
Volume      5.087981e-12 6.725903e-12 8.891059e-12 1.175313e-11
Marketcap   5.641293e-13 7.457340e-13 9.857959e-13 1.303128e-12

(Intercept) 6.710266e+03 6.709936e+03 6.709500e+03 6.708923e+03
(Intercept) .            .            .            .
High        3.119768e-05 4.123989e-05 5.451390e-05 7.205926e-05
Low         3.341427e-05 4.416987e-05 5.838677e-05 7.717829e-05
Open        3.215291e-05 4.250238e-05 5.618238e-05 7.426409e-05
Volume      1.553636e-11 2.053714e-11 2.714713e-11 3.588382e-11
Marketcap   1.722596e-12 2.277061e-12 3.009949e-12 3.978641e-12
```

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 perform 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 95% 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.

References

1. Bitcoin Price Prediction and Analysis Using Deep Learning Models Temesgen Awoke, Minakshi Rout, Lipika Mohanty, and Suresh ChandraSatapathy

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

3. A Comparative Study of Bitcoin Price PredictionUsing Deep LearningSuhwan Ji, Jongmin Kim and Hyeonseung Im

4. Predicting Stock Market Price: A Logical Strategyusing Deep LearningMilon BiswasComputer Science and Engineering,Bangladesh University ofBusiness and Technology Dhaka.

5. Garcia, D. & Schweitzer, F. (2015). Social signals and algorithmic trading of Bitcoin. Royal Society Open Science 2(9), 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. Nakamoto, S. & Shah, ;. D. (2017). Trading Bitcoin and Online Time Series Prediction. NIPS 2016 Time Series Workshop .

8. Santhosh, &., S. G. & ”. A Relative Study on Bitcoin Mining. Relative Study on Bitcoin Mining”, Imperial Journal of Interdisciplinary Research.

9. Awoke, Temesgen, et al. "Bitcoin price prediction and analysis using deep learning models." Communication software and networks. Springer, Singapore, 2021. 631-640.

10. Wirawan, I. Made, Triyanna Widiyaningtyas, and Muchammad Maulana Hasan. "Short term prediction on bitcoin price using ARIMA method." 2019 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE, 2019.