Tesla stock price prediction using Random Forest Machine Learning Algorithm

Project report submitted in partial fulfillment of the requirement for degree of

Bachelor of Technology

In

Computer Science and Engineering

By

(DEVENDRA MISHRA(171364)

Under the supervision of

(DR. GEETANJALI RATHEE)

To

Department of Computer Science & Engineering

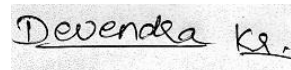**Jaypee University of Information TechnologyWaknaghat,Solan-173234**
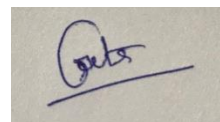
Himachal Pradesh

Certificate

I hereby declare that the work which presented in this report entitled **"Tesla Stock Price Prediction Using Random Forest Machine Learning Algorithm"** , in partial fulfillment of the requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** submitted in the Department Computer Science & Engineering and Information Technology, Jaypee University of Information Technology waknaghat, is an authentic record of our own work carried out over a period from July 2020 to may 2021 under the supervision of Dr. Geetanjali Rathee, Assistant professor, Computer Science & Engineering and Information Technology

The matter embodied in the report has not been submitted for award of any other degree or diploma.

Devendra Mishra(171364)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Dr. Geetanjali Rathee)

Assistant Professor

Computer Science & Engineering and Information Technology
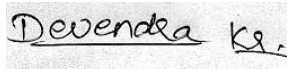
Dated:- 16th may 2021

No volume of words is enough to express our gratitude towards my guide, Dr. Geetanjali Rathee, Assistant Professor, Computer Science and Engineering Department, Jaypee University of Information Technology, Solan, who has been very concerned and has aided for all the materials essential for the preparation of this major project report. She has helped me to explore this vast field in an organized manner and provided us with all the ideas on how to work towards a research oriented venture.

I am also thankful to prof. (Dr.) Samir Dev Gupta, Head of Computer Science and Engineering Department and Mr. Ravi Raina, for the motivation and inspiration that triggered us for the major project work.

Most importantly, I would like to thank my parents, friends and the almighty for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope.

Devendra Mishra

9260980181

Table of Content

List of Abbreviations

ML:-Machine Learning

RFA:-Random Forest Algorithm

RMSE:-Root mean Squared Error

TSLA:-Tesla

Wrt:-with respect to

Vs:- versus

List of Figures

List of Graphs

List of Tables

Abstract

The project aims to provide retail investorsto navigate through the stock market.This is achieved through the use of Random forest machine learningalgorithm. A deeper understanding of Random Forest by taking OHLC (open-high-low-close) data for Tesla stock for 5year period. This will keep unrelated discussion at a minimum and try to focus on everything that is related to our understanding of Random Forest for this post. It would be helpful if you are familiar with concepts such as mean squared error and R-Squared statistics but it's not necessary to understand this Concept.

No prediction is 100% accurate. Therefore, the upper bound and lower bound of the stock prices will be displayed to illustrate the trading range the investors should be looking at.

Keywords: random forest, machine learning, mean squared error, R-squared statistics, Tesla stock

1.1 Introduction

The prediction of stock price is a task which tries to predict the future stock price value of a company stock and other factors which on an average will be traded financially.

If our system will tell the future price value of stock successfully then, we can gain a very significant profit.

The very first reason behind why we used this Random forest algorithm in our project is that it checks every condition based on decision tree which used to choose some of the important useful and the best attributes those affects most of the the problem. Hence we can predict the best possible output.

1.1.1 Introduction of Machine Learning

Machine learning (ML) is a way (and science) of making machines to learn old and new things which are not programmed by outside.

It involves mostly mathematics similarly it also involves computer science.

In it, the computer gets trained on a given dataset, and then to use the training data set in predicting the value of a new dataset.

For example, if we give animal data to train computer by giving it 5000 pictures of cats and 5000 other pictures those are other than a cat, and predict each time by computer that given picture is cat or other animal.

Now if we give the computer a new picture, then from the training dataset above, computer will be able to predict whether this new picture is cat or other.

This Procedure of training and testing that is predicting uses some of the specialized algorithms. We give the training dataset to the algorithm, and then the algorithm uses these training datasets to give future values means predictions of a new test data.

There are so many machine learning algorithms.

1.1.2 Introduction to Random Forest

Random forest algorithm is basically based on set of decision trees.

Random Forest is constructed by using a method known as Baggingthat work similar to the figure given below.

for every decision trees, we always choose a random sub set of the training dataset & make it to be fitted into the decision tree . repeat this as much time as there are numbers of tree given for any forest.

now, every model is up to some extent which is Predictive on a sub set of dataset. since this sub set has taken randomly , chances of mistake or error for every model must not be Correlated and these are some of the necessary things.

if i will take the average of all those different types of models, then i am actually merging the information very effectively which is contained in all of those models that is represented by the full dataset.

The main idea regarding the random forests are to create multiple decision trees from those random subset of data , since those subsets choosen are totally random so we cover the total dataset .

Those subsets are used for making multiple different predictions.

1.1.2Introduction to Linear Regression

linear regression is a approach that works linearly used for modelling the relationship between a scalar response(present there) and also some explanatory variables present there(these are also called as dependent and independent variables).

Linear regression was theregression which was of first type regression analysis, which wasthere to be studied rigorously, and was extensively used in many practical applications. The reason behind this was the models which depend linearly on the unknownvariables are easier and very good to fitthan those models which are related non-linearly to their variables and parameters and also one reason is that the statistical properties of the estimator resulting there are not difficult determine.

The linear regression model assumes the following relationship between dependent and independent variables-

Y=mX+c

where

dependent variable is Y,

independent variable is X,

m->slope of the line(change in y for one unit increase in x)

c-> Y intercept

1.2 Problem Statement

Telsa has been on the rising and rising since starting of 2013 which happened to be shortly after they launched their first car. There were various factors those were affecting the tesla.

There exists so many factors which have affected the price more than that of other factor. We must try to know these factors and the features and we should tell the stock price value for a particular day and also give comparision of it with the actual known value of output. The over fitting problem that ocured is also a cause, which has to be overcome.

Some of Those factors are open, close,adj close, low, high, date, etc.

The retail investor invests a many times to know the opportunity of investments. The investors who are wealthier can have Professional, financial and economical advisary service, But to simple and not wealthier retail investors, these costs are not able to be afforded.

hence, the retails investor has to know out that market by theemself and need to make to generate the Decision by their Own basis of information. That makes the investment process a lot stressfull in the deeveloping society. unfortunately, we homosapiens are not rational from the nature.

without Quantitative, and based on data driven model, the decision got swayyed by the cognitive biasness or the personal attachments, which results in the non important loss to them. even if the investor were enough cautious , maximum of them do'nt had required skill to processes a big amount of data required to take the desired decisions. The institutional investor trusts in good model which is supported by technology to stay away from the trap, but the Retail investor dont has access to these technology and mostly they found themself falling back and back in the market.

Therefore, our final year major project aims to serve them as the potential tool in finding unknown pattern in those trends of the stock price.

1.3 Objectives

The main aim is To design and develop a model to predict the price of stock of any company like tesla stock of a particular day and see if it actually similar to the output and analyse manually by what value of percentage its differs.

The ultimate goal of this major Project is to serve the retail investors those who use Machine Learning for helping them to get informed about the very fast changing trend of the Stock markets . the Project have aim for introducing and democratizing those recent Machine Learning technology for the Retail investor for their use.

We have experimented with some features in our dataset to predict the adjusted closing price for Tesla stock of a particular day.

These predictions weren't necessarily much good but these are only used for illustration purpose only , Since any prediction is not 100 percent accurate . therefore, we took the lower bound and the upper bound of Stock price and it would be display for illustrating the treding limit that investor would be taking care for

1.4 Methodology

It is a technique of machineries from that we could train the machines to understand by many datasets. On using different algorithm and walking through the series of Cost functions the logic of Machine Learning algorithm shall used for Predictive models while applyingg the relations among those factor . with use and advancing of Deep Learning on our model we were able to implement few intricate points for better accuracy and also better performance of this model . our model would use the recent science for an better and error free function.
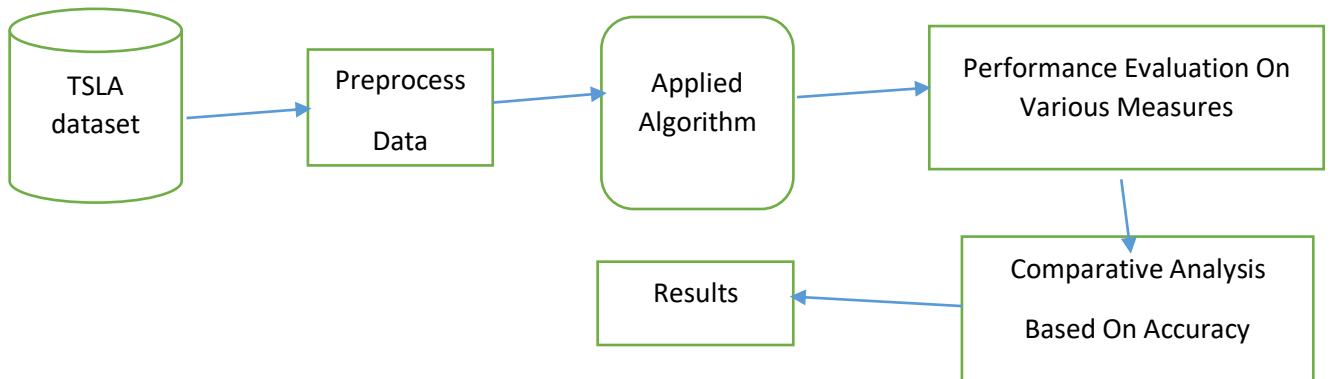

Fig. 1. Proposed Model Diagram

1.4   Organization

This model that we have implemented was decision tree and Random forest classifier and linear regression algorithm.

All type of coding was done in Jupyter Notebook platform.

Extensive use of python libraries like pandas, numpy, matplotlib and sklearn packages.

Our Dataset had 7 different features with 2416 instances .

Features was numerical in nature.

Finally we got the very best model which can classify unbalanced/imbalanced multiclass dataset that was obtained.

1.5 Tools to be used

Jupyter Notebook

Pandas

Numpy

Matplotlib

Scikit Learn

MSExcel

Chapter-2LITERATURE SURVEY

I have always wondered why some companies are growing at such rapid growth as Facebook and Google while other companies like yahoo and Tata Nano are not growing at that rate despite good market strategy.

After reading a few things one thing that stuck in my mind is that the success of any company depends largely on the price of the stock.

That is why I have decided to establish a system that can predict good releases and if I can do that then it can bring great success to many companies.

According to the paper of research titled "machine learning in stock price trend forecasting" published by y. dai no- y. zhang at stanford university, he uses the parameters such as pe ratios,px Volume, px EBITDA , ten day instability, fifty day average of moving , etc in predicting the next day Stock Price and long term Stock Price.

 Machine Learning algorithm which has been used in this study are logistic regression, gaussian discrimination analysis, quadratic discriminant analysis, and the SVM.

Accuracy measure can be defined   as the no. of total days in which the project has properly divided test dataset on the total no. of test day . with the small term model which predicts the price of the just next days , with the lowest accuracy, Quadratic Discriminant Analysis is the leading of all types, with a accuracy of 58.2% .

With a long-term model predicting future price , as long the time, the accuracy of the SVM will be better. with a 44-day time windows , the accuracy of the svm model has reached 79.3 percentage . Furthermore, this has been found that the accuracy will increase as the nimver of features increases.

when all the 16 features was used , the accuracy of the model gets 79 percent , whereas it decreased to 64 percentage when only 8 feature was used , and 55 percentage when just a single features were used.

our model would too investigates that how timing will effect the accuracies of Price estimates on various types.

Since, the model has to reach a specific level in order for it to be important for users to act as a reference, it is important that we use our model to determine what appropriate parameters and model structure is our goal of predicting prices.

The research paper "Predicting the movement of stock indexes and stocks using trend deterministic data preparation & Machine Learning strategies" ehich has been Written by j. patel ,s.shah, p. thakkar, noK. According to the international journal Expert Systems with Applications, the use of dataset to determine the Stock Price movements.

they took a study with the use of ten signals of Technical indicator such as input, and then by using forecasting models to predict whether stocks would increase or decrease on the next ten day, Technical Analysis indicator which includes sma, ema, momentum, stochastic sk, stochastic sk, macd, rsi, etc .

The forecasting model that they used including ann, svm, random forest, and naive bayesianmodel . These models emits 'up' or 'down' movements signal.

Studies has shown that the informal forests has achieved the greatest efficiency with an 83.56% accuracy of its results

Researchers have shown that macroevolution can automatically produce neural network models and hyperparameters with higher performance compared to high-end artificial model.

On a study conducted in year 2017 , a major emergence of the discovery of image separation in neural networks.

 It starts with the majority of people with 1 simple simulated models, and then gradually transforms the Population by the removal of the bad design and then creating an another design by changing the parameter of the best design on every iterations.

After the hundred of total hour of using the algorithms by great computation powers, many human model have achieve technical results in cifar data set.

For every iterations , just a Simple modification which change one parameters were used, allowing search in a large search space.

This paper has shown that it is possible to find good models using a lot of computing power in retrieving humans Machine Learning expert and to lay the foundations for democratic Learning using automl.

The project problem has been set in the prediction of the next day's stocks Price. The 'next day' has been chosen as the time frame like a small term Price movement often rely heavily in the trend dynamics and Price patterns, While the long term Price movement depends in the basics for stocks (example: corporate managements skill, Revenue Model , Market Demand, Macroeconomic features, etc).

RMSE and other schools are described as providing in-depth information on the performance of model prediction and financial-based comparisons between investor models

On the research part , the System ha been designed as to be as robusts as Possible to the simplify models testings . each models could be known by the pairs of the design option & the inputs option , which specify the configuration of the design & the input input.

It speeds up the processes of test of a different combination of planning and installation.

The two keys feature is selected as the inputs . 1st is the long consistent lists of the certain Raw historic datasets such as the stock prices & the daily % changes.

Selected specified size determines the historical size of the retrospective time from today When future's pricing is forecast.

By referring of the Principle of Technical Analysis , as soon the Stock Price reflect all the relevantinginformations, the technicals analysis will focuses on stock tradings patterns rather than the looking on the economicical and corporate fundamental.

So, By determining the historical pricing period such as the inclusion of a training design , this can be pieces of the informations that is useful to find the trading Pattern & therefore prediction of the future's Stock price trends.

From the setted observation time, it has been assumed that predictable Price Movement pattern will occur in the specified historical period.

The second input is the moving average of the figures. 1 of the most obvious ways to the market investor in the identification of market trends using the moving average . with a Robust System model, a different-moving average time can be use as an inclusion in stock price forecast model, for example, a set of 20, 30, 300, 600 day moving averages, which are widely used by investors.

At 1 day's estimate, there will be one outflow units that is the stock on the next days.

Predicted Stock Price would be the just another forecast, in predicting the Stock Price on 2nd day , the process will keep repeating until all the ten forecasts is made.

Methodology – Design

After obtaining the appropriate stock data from a third-party data provider, the backend Process Data & will build the model.

 after it,the prediction are made using appropriate model and algorithm similar to here using a linear regression and random forest algorithm that uses decision trees to predict.

Decision Trees

As the name suggests, decisions are made by dividing the data into groups that are similar within them but different from others.

 Separation is based on a specific process (points) that achieves the best classification of data in a way that prepares that process.

 We call this, the acquisition of information that basically tells us how much our conditions improve as we continue to make that distinction.

This is one of my favorite algorithms and I use it often. It is a type of supervised learning algorithm that is widely used for editing problems.

Surprisingly, it works on both variable and dependent variables. In this algorithm, we divide a population into two or more identical sets.

This is done based on the most important / variable variables that make the groups as different as possible.
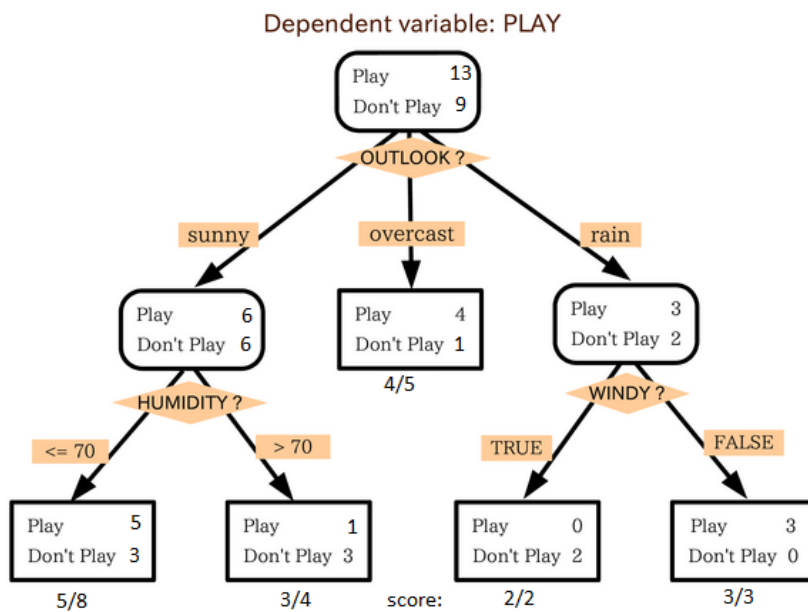
Fig 2 Decision Tree

Above the picture shows how the population is divided into four different groups according to many symbols to indicate 'they will play or not'.

A good way to understand how the decision tree works is to play Jezzball - an old game from Microsoft (pictured below). In fact, you have a room with moving walls and you need to build walls so that the surface can be removed without balls.
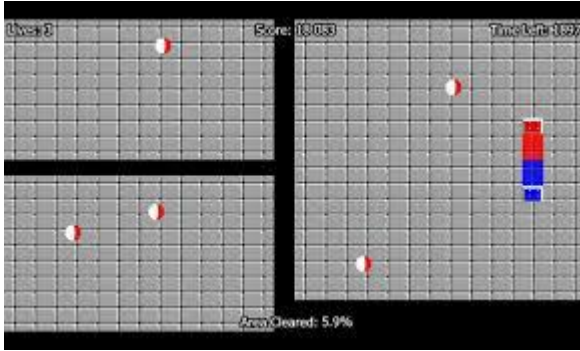
Fig 3 jezzball

So, every time you separate a room from a wall, you try to create 2 different people in one room. Decision trees work in a similar way by dividing the population into as many different groups as possible
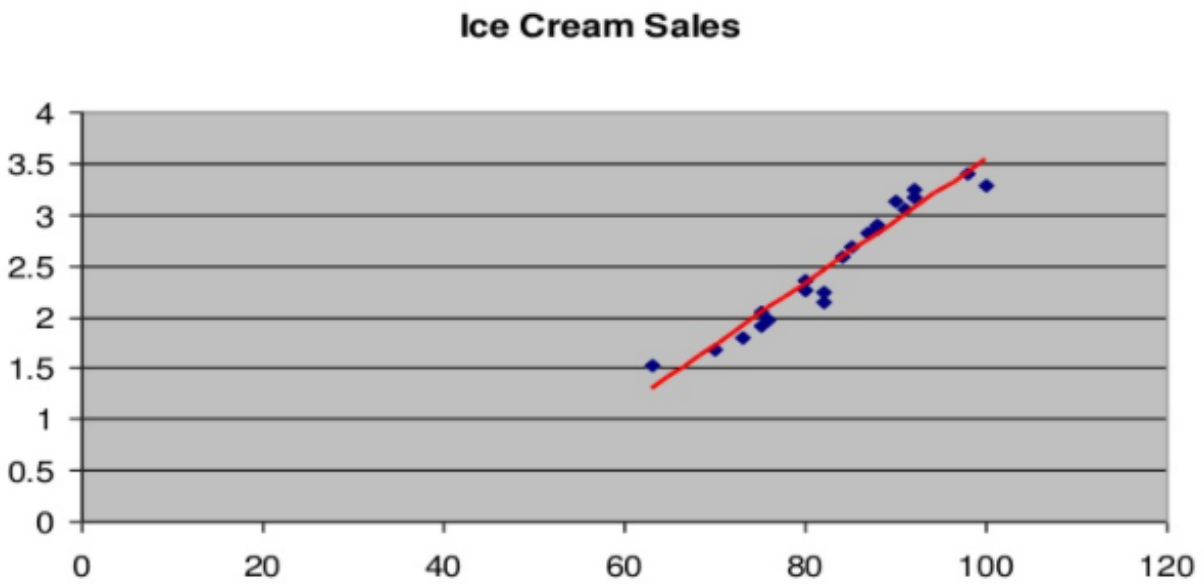
Linear regression

While using linear regression one takes into account the regression line which is of the form Y=mX+c. regression generates least square regression line.

Least square approach states that when we get predicted value of Y for each value of X in the data, then there is some deviation from the actual value. then to get more accurate result we square the differences in the Y value and add these squared differences across all the predictions. We get a number called residual or error sum or squares. Then we minimize this number using calculus to minimize the average error in the prediction.

Some examples are given below-

X Temperature	Y Sales
63	1.52
70	1.68
73	1.8
75	2.05
80	2.36
82	2.25
85	2.68
88	2.9
90	3.14
91	3.06
92	3.24
75	1.92
98	3.4
100	3.28
92	3.17
87	2.83
84	2.58
88	2.86
80	2.26
82	2.14
76	1.98

Ice Cream Sales

Regression line data representation

Chapter-3 SYSTEM DEVELOPMENT

In this Chapter, we have used the random forest machine learning algorithm and linear regression in our database to predict Tesla prices.

The main reason for using a random forest algorithm in this project is that it works based on a decision tree that selects some of the best features that affect the problem. It is therefore likely to predict a very positive outcome.
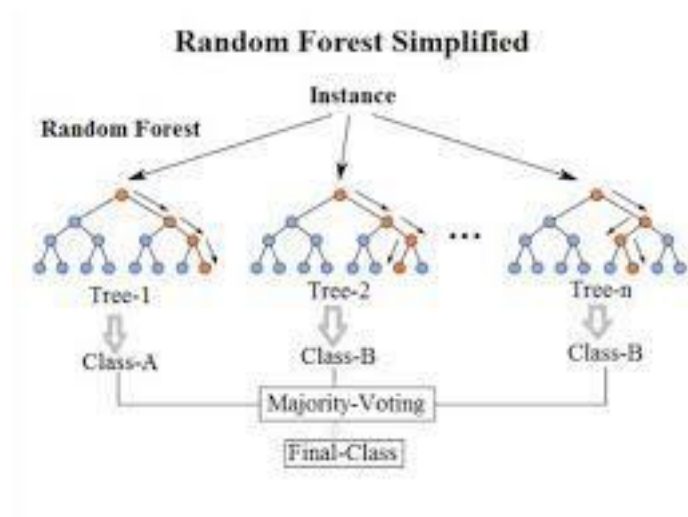
3.1 Random Forest Classifier

The random forest to differentiate Machine Learning algorithm, create many Decision Tree & eventually combine those to provide a final and last Prediction . Many Decision making tree is based on many Conditional statement that help with good accuracies. With Such a large no. of Random Decision - making tree , We could avoid being overly balanced.

A random forest is a way of learning together the division, retreating and other activities that work by building a pile of decision trees during training and classroom extraction which is a way of class (division) or predicting the meaning (return) of individual trees. Random decision-making forests correct the practice of over-pruning trees in their training set.

The first algorithm of random forest forests was developed by Tin Kam Ho using a random subspace method, which, in its formulation, is a method of applying the "racial discrimination" method according to the categories proposed by Eugene Kleinberg.

The algorithm extension was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark. The extension incorporates Breiman's concept of "integration" with random selection of elements, introduced first by Ho and later independently by Amit and Geman to create a set of decision-making trees with a controlled variety.

Figure 4-Classification in Random Forest

Linear regression:

Least square linear regression is used. And by the help of scatter plot and regression line , all the predictions are made. Scatter plot is basically a mathematical graphical representation of values for two variables for a set of data. It also explains easily the increase and decrease in then values present in ordered pair.
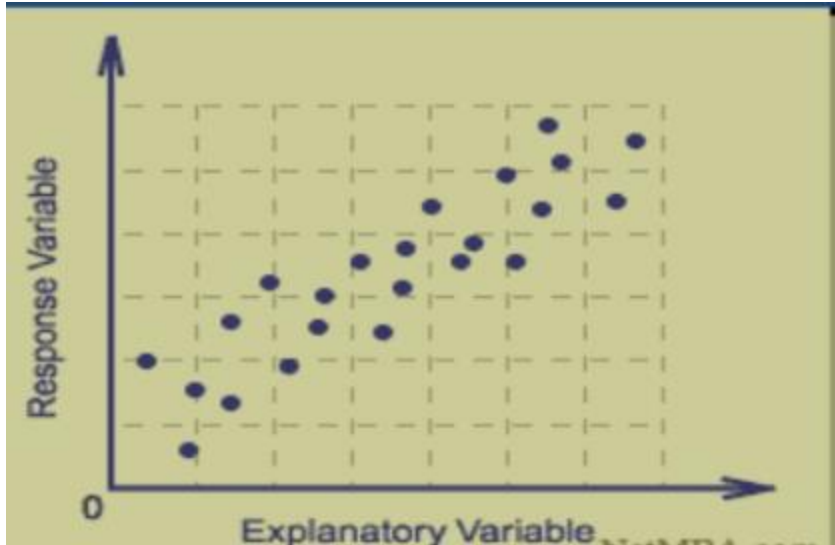
Scatter plot representation

3.2 Task Performed

3.2.1 Dataset Description

We collected tesla dataset from Kaggle and performed computer simulation on one dataset. Dataset is a Tesla Price Prediction dataset. Dataset contained 7 different features with 2416 instances. The features describe date, open , high, low, close , adj close, volume. The dataset contains features expressed on nominal, ordinal, or interval scales.

3.3 Random Forest

According to Breiman's method, each decision tree in the forest is formed by random selection first, in each place, a small group of input links (also called features or variables after this) to divide and, second, by calculating the best

variation in these elements in the training set. The tree is grown using the CART method to a large size, without pruning. This random action scheme is combined with a fundraiser to sample, by replacement, training data is set each time each new drug is developed.

3.4 Feature Importance

Out of all the 7 features Date, open, close, high, low, adj close and volume that is going in predicting the stock price . Those every feature will perform a necessary role to predict the price.

3.5 Test Plan

Test Plan of this stock price design prediction has been taken from the actual stock price data from kaggle. We have developed a variety of methods or techniques to create a systematic and harmonious way of using information in the purpose of the design . These Data's are specific to key Point needed in the study. In addition, testing program is aligned in the design & could assist in next development and development of the design and designs which is related to theses types of the research.

3.6. Data Set

For the purpose of this study, Tesla dataset is considered as it is the best dataset for the present study. It contains 2416 records. Every record has 7 attributes out of those one attribute is object variable. All 7 attributes contains only numeric data and are not null. Each record contains information about single day.

Date:

Open:

High:

Low:

Close:

Adj close:

Volume:

Outcome:

3.7 Data Preprocessing

The data processing are the most necessary aspects needed for training of the model . since every column and row may not be important to the design and set of available dataset are not in a way that could be use to the training of the machines. In those case dataset processing are a necessary factors determining the fine beginning of the design.

The data processing are a method which is use in converting the new dataset into a important format . This could be by filling in the null value , converting by phase to the Binary formats by the encoding.

Features	values
date	2416 non-null object

pen	2416 non-null float64
high	2416 non-null float64
low	2416 non-null float64
close	2416 non-null float64
Adj close	2416 non-null float64
volume	2416 non-null int64

Table 3.7 features and their values

3.8 Algorithm

X_train, X_test, y_train, y_test =train_test_split(X,y,test_size = 0.3)

Model=RandomForestRegressor(n_estimators=40,bootstrap=True,min_samples_leaf=25)

model.fit(X_train, y_train)

y_pred=model.predict(X_test)

conf = model.score(X_test, y_test)

print(conf)

model.fit(X,y)

predictions = model.predict(X_Check)

```
print(predictions)
```

As it was a very good experience to learn great. So, we went through some exciting observations in the project , By following a very unique approach to get a special output. Finally, We could go through the specific result and we could explore the correct outputs. All the machine learning algorithms which had been considered are proven as the important for collecting the right data for the validation purpose.

for the Machine Learning project that we create , we shall find different hyper parameters & check each for the validity of the model. Accuracy MSE and RMSE were the major measure for the performance. These can be calculated easily.

Some operations and data manipulations performed are

df.shape

output- (2416,7)

#plotting the data

plt.figure(figsize=(16,8))

plt.title('Close Price History')

plt.plot(df['Close'], color='red')

plt.xlabel('Date', fontsize=18)

plt.ylabel('Close Price USD', fontsize = 18)

plt.show()

output-

Close Price History

Graph showing close prices vs date

4.1 If test size =0.2

X_train.shape

Output-

(1884, 4)

#Calculate the confidence value by applying the model to testing data

conf = model.score(X_test, y_test)

print(conf)

output-

0.9316009542537967

#Plot the results

result.plot(figsize=(20,10), linewidth=1.5)
plt.legend(loc=2, prop={'size':20})
plt.xlabel('Date')
plt.ylabel('Price')

output-

Text(0, 0.5, 'Price')

Graph showing forecasted value of price with respect to date

4.2 If test size =0.3

X_train.shape

Output –(1670, 4)

#Calculate the confidence value by applying the model to testing data

```
conf = model.score(X_test, y_test)
print(conf)
```

output-

0.9042539650830713

#Plot the results

```
result.plot(figsize=(20,10), linewidth=1.5)
plt.legend(loc=2, prop={'size':20})
plt.xlabel('Date')
plt.ylabel('Price')
```

output-

Text(0, 0.5, 'Price')

Plot of forecasted price vs date

Error in random forest model-

We can see the values of mean absolute error ,mean squared error, and root mean square error for random forest algorithm.

```
[143]: print('Mean Absolute Error:', metrics.mean_absolute_error(y_test,predicted))
       print('Mean Squared Error:', metrics.mean_squared_error(y_test,predicted))
       print('Root Mean Squared Error:', math.sqrt(metrics.mean_squared_error(y_test,predicted)))

       Mean Absolute Error: 131.30530628248383
       Mean Squared Error: 27883.39569086618
       Root Mean Squared Error: 166.98321978829543
```

Analysis of linear regression implementation:

We have seen the results and errors in above implementation of random forest algorithm. now let us see the linear regression model implementation results and errors.
We have defined the model as linear regression.and calculated the values of regressor coefficient ,intercept and head of predicted dataframe.

```python
#define the model as linear regression
regressor = LinearRegression()
```

```python
regressor.fit(X_train, y_train)
```

```python
conf = model.score(X_test, y_test)
print(conf)
```

0.9867325642740683

```python
print(regressor.coef_)
```

[-2.67075363 -0.73758434 113.7745966   7.27303308]

```python
print(regressor.intercept_)
```

192.6326786181686

Figure showing linear regression model definition

```python
[9]: df2 = pd.DataFrame({'Actual': y_test.flatten(), 'Predicted' : predicted.flatten()})
```

```python
[2]: df2.head(25)
```

[2]:

	Actual	Predicted
0	331.100006	306.723031
1	218.259995	240.672450
2	212.279999	216.490885
3	37.509998	41.887407
4	34.740002	33.426461
5	331.660004	306.448244
6	246.949997	267.151604
7	225.009995	261.898209
8	102.400002	113.294676
9	268.260010	251.294555
10	32.540001	34.927703
11	513.489990	339.282572
12	207.460007	207.395432
13	199.630005	220.230705

rch

Figure showing Head of dataframe of predicted and acual values

```
]: print('Mean Absolute Error:', metrics.mean_absolute_error(y_test,predicted))
   print('Mean Squared Error:', metrics.mean_squared_error(y_test,predicted))
   print('Root Mean Squared Error:', math.sqrt(metrics.mean_squared_error(y_test,predicted)))

   Mean Absolute Error: 23.805900174686098
   Mean Squared Error: 1281.2836342261069
   Root Mean Squared Error: 35.79502247835733
```
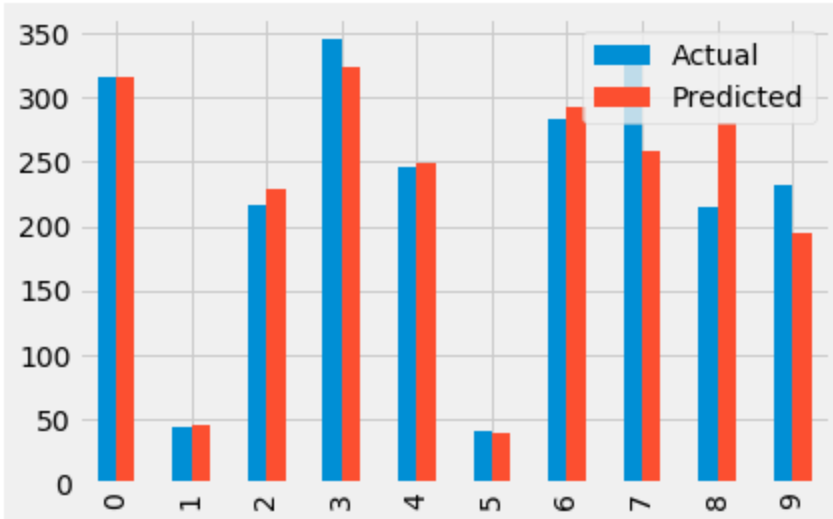
Figure showing the error in linear regression model

```
graph = df2.head(10)
```

```
graph.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x18a18e2d860>
```


Bargraph showing the actual and predicted values

4.2 Results

Above figures represents different performance values of our model algorithm calculated on various measures like confidence value and means square error and RMSE etcis having a great difference in both the cases. From above analysis it is analyzed that linear regression is best suited for this prediction showing the maximum accuracy.

As We used a pre-modified and neat and clean dataset that was already feature-engineered so we don't need to bother about any discussion around that. And

although, we selected and worked on training linear regression and random forest algorithm a pretty decent model, so we got better results .we can even do better by using more features and hyper parameters. decent dataset and more organized features will give best results.

Chapter-5CONCLUSIONS

5.1 Conclusion

One of the important real-world problems is the prediction of stock prices for any company like Tesla and using these results for the growth of company and also to be used by the retail investors in stock market. In this study, systematic efforts are made in designing a system which results in the prediction of tesla stock prices. During this work, two machine learning algorithms were studied and evaluated on various measures. Experiments are performed on TSLA Stock price Database. Experimental results determine the adequacy of the designed system with appreciable accuracy using linear regression algorithm. In future, the designed system with the used machine learning algorithm can be used to predict stock prices for other companies. The work can be extended and improved for the automation stock market and stock price analysis including some other machine learning algorithms.

5.2 Future Scopes

We target to take large numbers of feature in the future like companies future plans and checking their dependency on share capital. we do targets to grow sample size too for getting some more correct outputs to give more appropriate idea to the retail investors . As large dataset we will consider, the appropriate the sample to get Cross Validation in the criterias which will execute in our project.

It would be a most necessary insrument for the future that has to come & would be recognised as a greatest invent in order to Solve the Stock Market's Problem. The amount of person with stock market investment would increase gradually &therfore it would be a profit for the humanity.

Investors and prediction makers may used this in their Validations , in Consultancy etc . people who have any doubt or only who wants to has a simple investment may took it's Considerations in their accounts.
Therefore, with current technology and in the advancement of future's article, these problem will has a some good solution & timely using of the System will be beneficial.

Hence, by the use of Machine Learning algorithms these designs are able to find an answer under the desired period of time .however in Future's requirement on these department's there is a large observation which could be taken in the considerations.
Apart from that this model can be deployed as an web application or android application which can be available for free and can be tested by real users. These results can help us improving the model if we shift towards reinforcement learning. In today's time where ML is growing at higher rate, With proper guidance and using of Principal Component Analysis, This model accuracy can be reached to a mark of 90%.

REFERENCES

1.https://en.wikipedia.org/wiki/Random_forest

2.https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.html

3.https://github.com/prnvg/Stock-Price-Prediction

4.https://www.cse.ust.hk/~rossiter/fyp/103_RO4_Final_201819.pdf
5.https://www.slideshare.net/Tech_MX/linear-regression-14155467
6.https://www.google.com/search?q=kaggle+datasets&oq=kaggle&aqs=chrome.2.69i60j69i57j0i395i433j0i433j0j69i60l3.5155j1j7&sourceid=chrome&ie=UTF-8