YES BANK STOCK CLOSING PRICE PREDICTION

VINIT LADSE, PRATIKSHA KHARODE, GAURAV BHAKTE

CAPSTONE PROJECT-II

ALMABETTER,BANGLORE

ABSTRACT:

- Accurate prediction of stock market returns is a very challenging task due to the volatile and non-linear nature of the financial stock markets. With the introduction of machine learning, time series forecasting and increased computational capabilities, programmed methodsof prediction have proved to be more efficient in predicting stock prices.
- In this work, regression and time series forecasting techniques have been utilized for predicting the next day closing price for one company belonging to the Finance sector of operation.
- The financial data: Open, High, Low and Close prices of stock are used for creating new variables which are used as inputs to the model.
- The models are evaluated using standard strategic indicators: RMSE and MAPE.
- The low values of these two indicators show that the models are efficient in predicting stock closing prices.

OBJECTIVE:

Yes Bank is a well-known bank in the Indian financial domain. Since 2018, it has been in the news because of the fraud case involving Rana Kapoor. Owing to this fact, it was interesting to see how that impacted the stock prices of the company and whether Time series models or any other predictive models can do justice to such situations. This dataset has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month. The main objective is to predict the stock's closing price of the month.

PROBLEM STATEMENT:

- Perform regression analysis using multiple models to predict the closing price of the stock and compare the evaluation metrics for all of them to find the best model.
- Prediction of Yes Bank stock closing price.
- Getting accuracy score of several machine learning model.
- The Objective of the Project is to come up with the time series model or predictive model to predict the stock closing price of the month using the Yes Bank dataset of monthly stock prices of the month.

DATA SUMMARY:

The dataset of YES BANK has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month of around 185 observations.

It contains the following features:

- **Date:** It denotes date of investment done (in our case we have month and year).
- **Open:** Open means the price at which a stock started trading when the opening bell rang.
- **High:** High refer to the maximum prices in a given time period.
- **Low:** Low refer to the minimum prices in a given time period.
- **Close:** Close refers to the price of anindividualstockwhenthe stock exchange closedfor the day.

INTRODUCTION:

- Stock market is characterized as dynamic, unpredictable and non-linear in nature. Predicting stock prices is a challenging task as it depends on various factors including but not limited to political conditions, global economy, company's financial reports and performance etc. Thus, to maximize the profit and minimize the losses, techniques to predict values of the stock in advance by analyzing the trend over the last few years, could prove to be highly useful for making stock market movements. Traditionally, two main approaches have been proposed for predicting the stock price of an organization.

- Technical analysis method uses historical prices of stocks like closing and opening price, volume traded, adjacent close values etc. of the stock for predicting the future price of the stock.

- The second type of analysis is qualitative, which is performed on the basis of external factors like company profile, market situation, political and economic factors, textual information in the form of financial news articles, social media and even blogs by economic analysts.

- Nowadays, advanced intelligent techniques based on either technical or fundamental analysis are used for predicting stock prices. Particularly, for stock market analysis, the data size is huge and also non-linear.

- To deal with this variety of data an efficient model is needed that can identify the hidden patterns and complex relations in this large data set. Machine learning techniques in this area have proved to improve efficiencies by 60-80 percent as compared the past methodology.

STEPS INVOLVED:

- Collection Of Data

  Before building any machine learning model, it is vital to understand what the data is, and what are we trying to achieve. Data exploration reveals the hidden trends and insights and data preprocessing makes the data ready for use by ML algorithms. So, let's begin. . . To proceed with the problem dealing first we will load our dataset that is given to us in a csv file into a data frame. Mount the drive and load the csv file into a data frame

- Discusing Problem Statement

  After analyzing the datasets we discussed with every single problem to overcome it. We all decided to divide our task and initialized with our own problem statement. The problem statement were based on target variable we took for analysis.

- Data cleaning

  The next task was data cleaning which was easy with this dataset. As mentioned in above points the data were float64 dtype , int64 dtype, object dtype,datetime64.

- Exploratory Data Analysis

  After data cleaning it was sure to target some important columns for Exploratory Data Analysis. Matching the data with correct suitable problem by python libraries to result some insightful visualization was great task. These also gives us a more information and graphs.

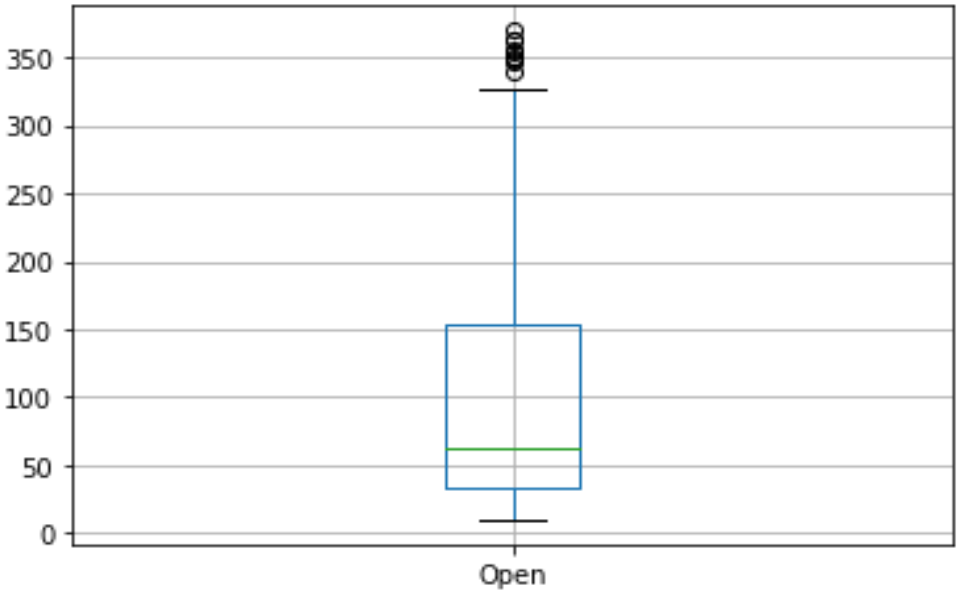
- Visualization of Analysis :

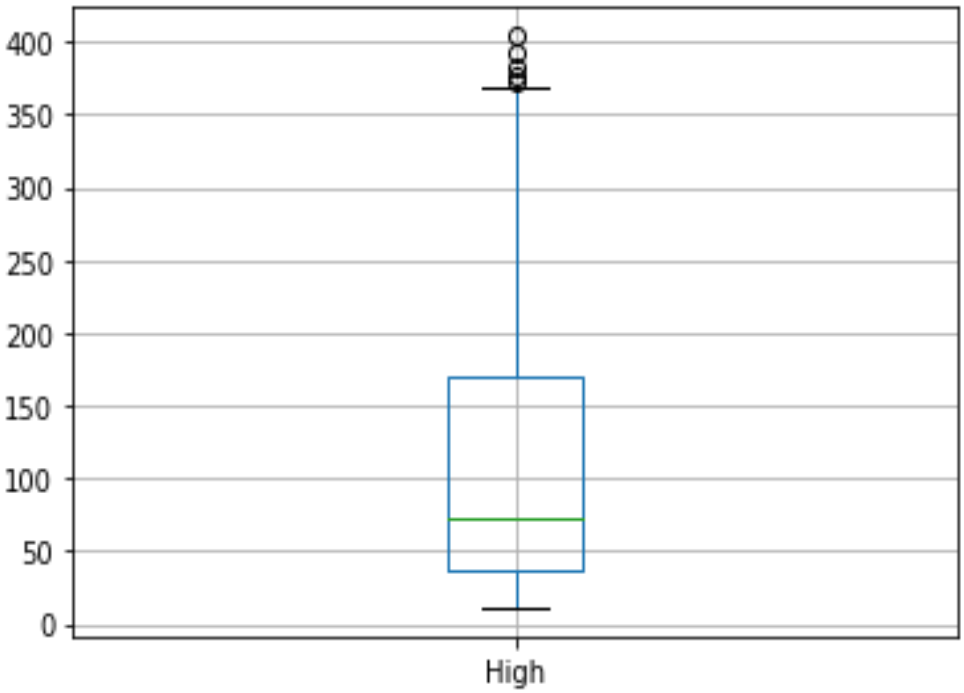
  The EDA parts make more clear about data in a picture and graphical form. Mainly we perform matplotlib and seaborn libraries of python for the data analysis. The libraries helps a lot with graphs.

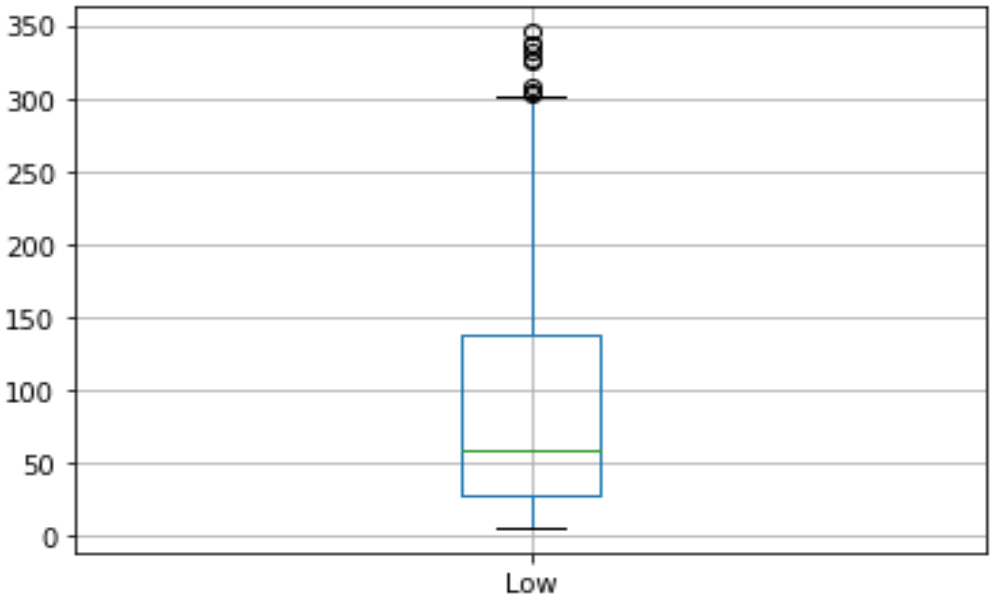
1)Missing values:

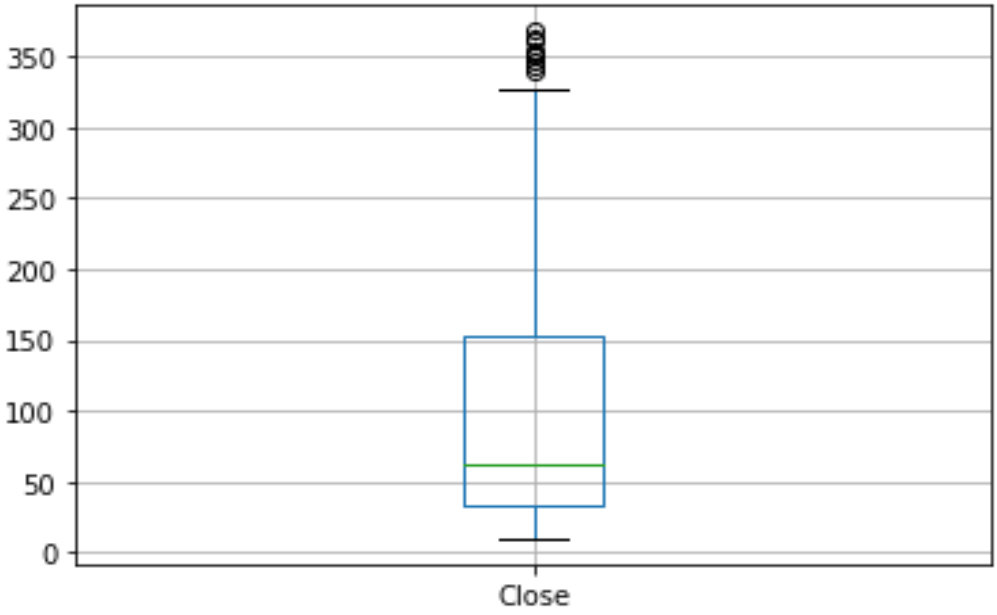
No missing values in dataset.

2)Checking Outliers:





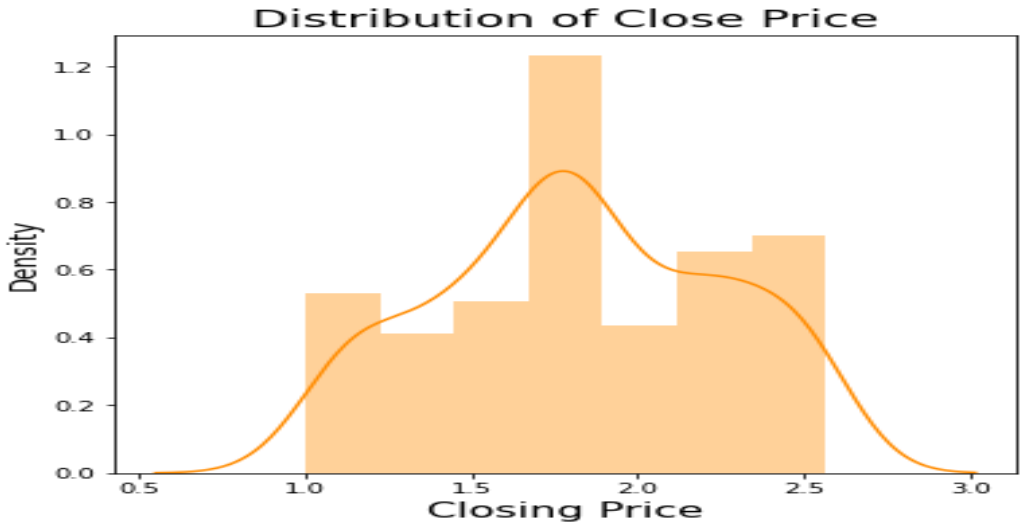




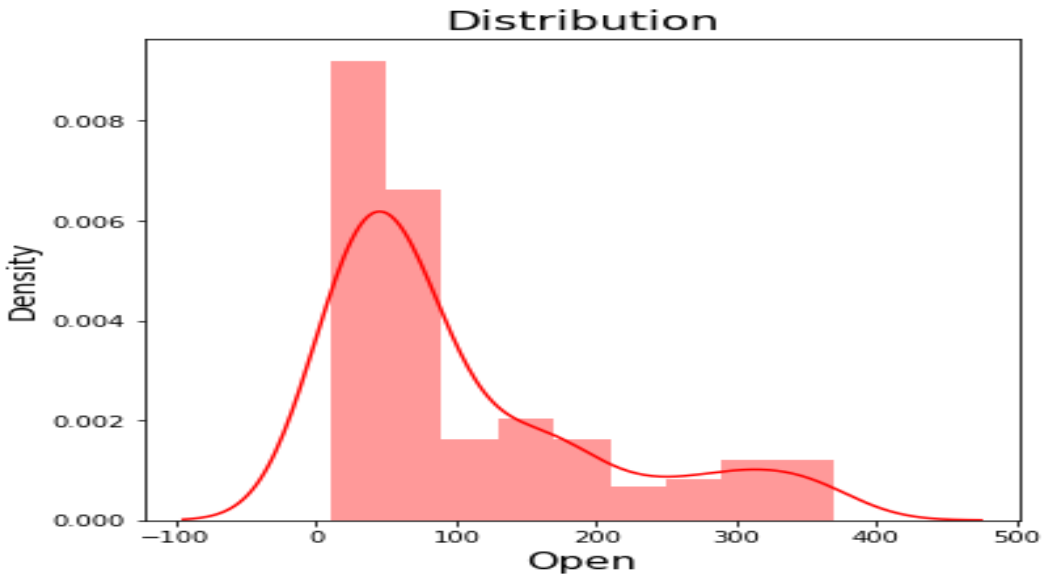
EXPLORATORY DATA ANALYSIS/ DATA PREPROCESSING:

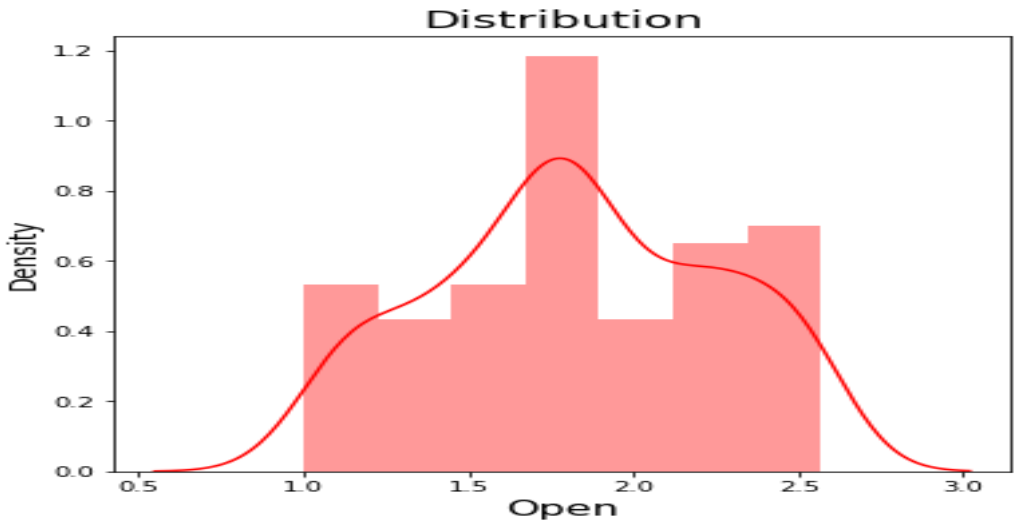
The primary goal of EDA is to support the analysis of data prior to making any conclusions. Exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization method.

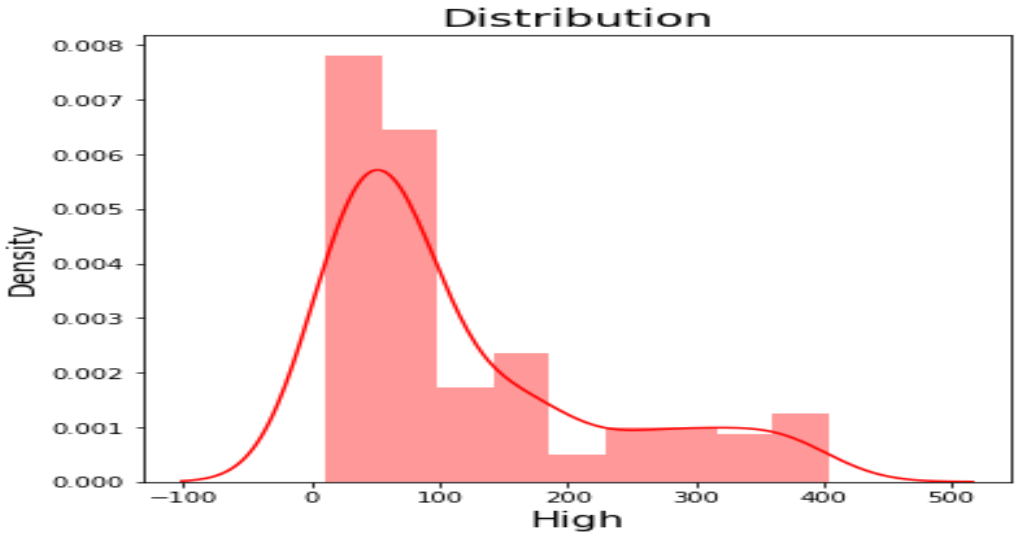
Dependant Variable of Close Price Of stock:

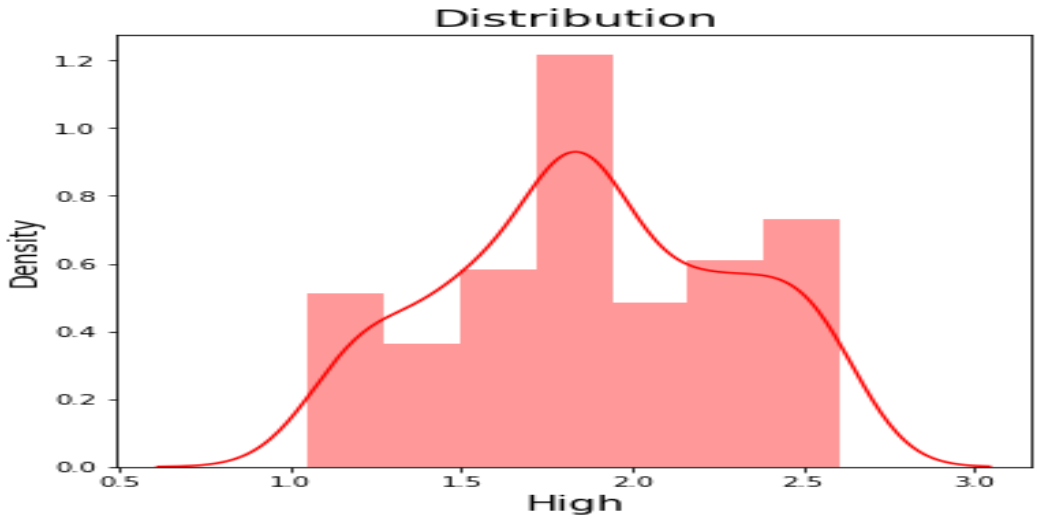


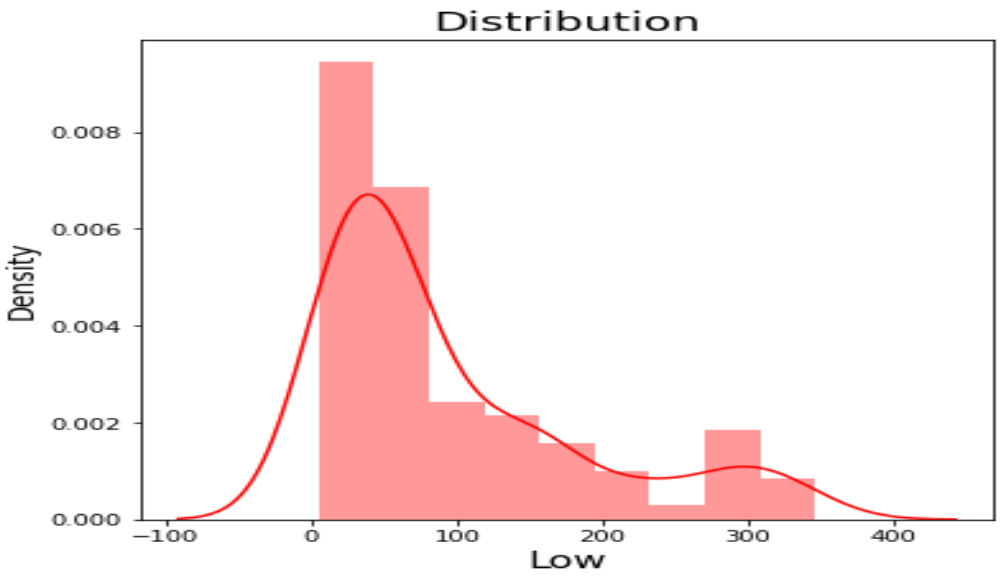
Independant Variable of Open ,High And Low Price Of stock:

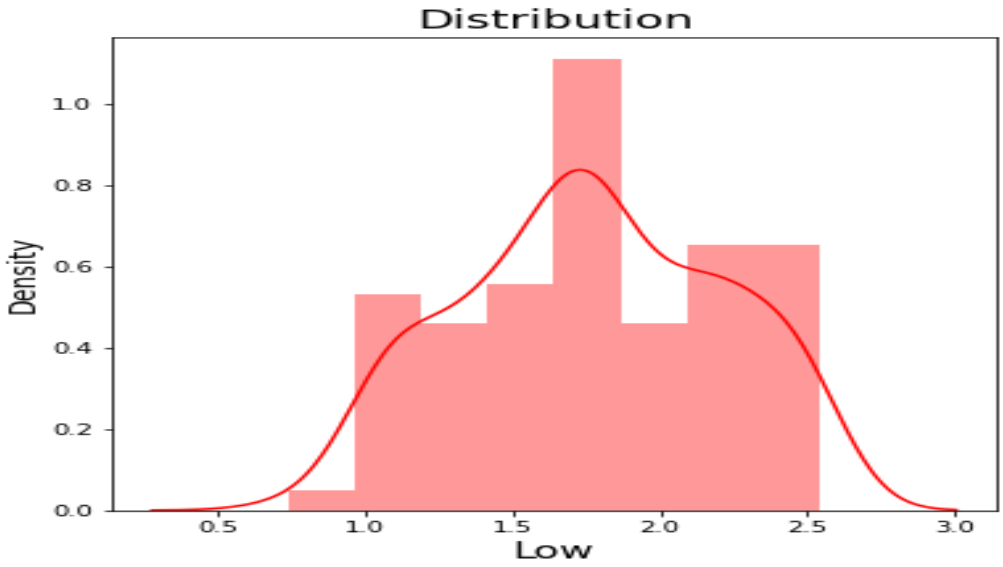



Distribution
Density
Open
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0

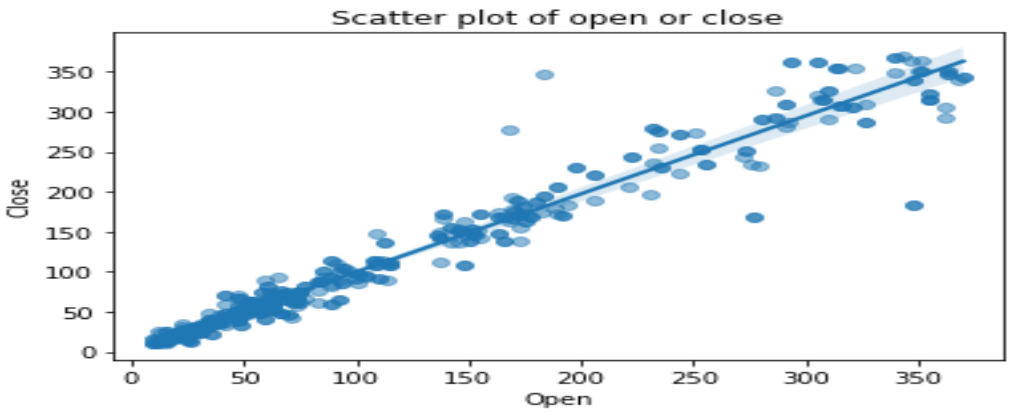

Distribution
Density
High
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
−100
0
100
200
300
400
500

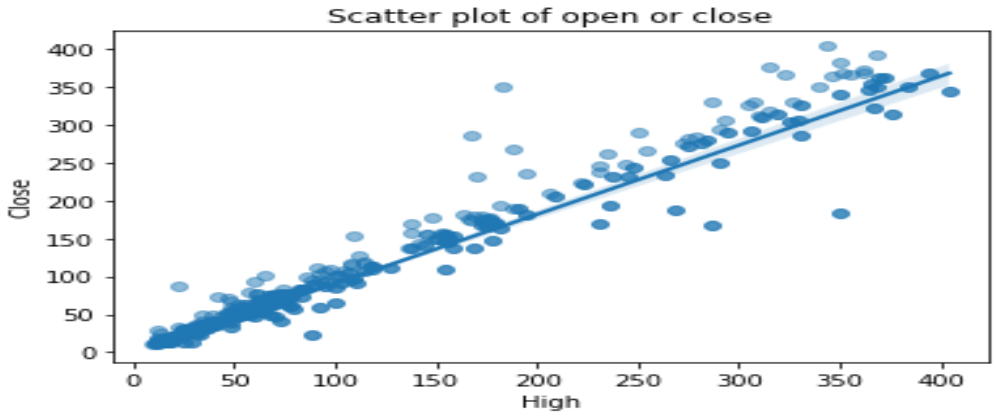

Distribution
Density
High
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0

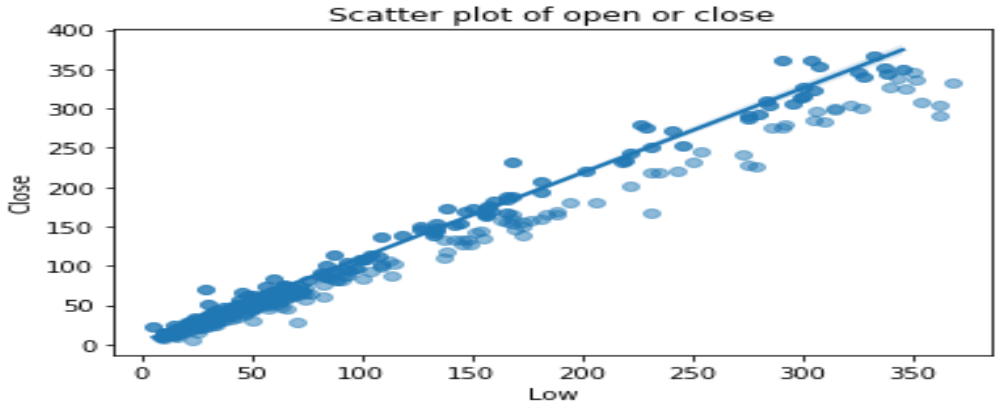




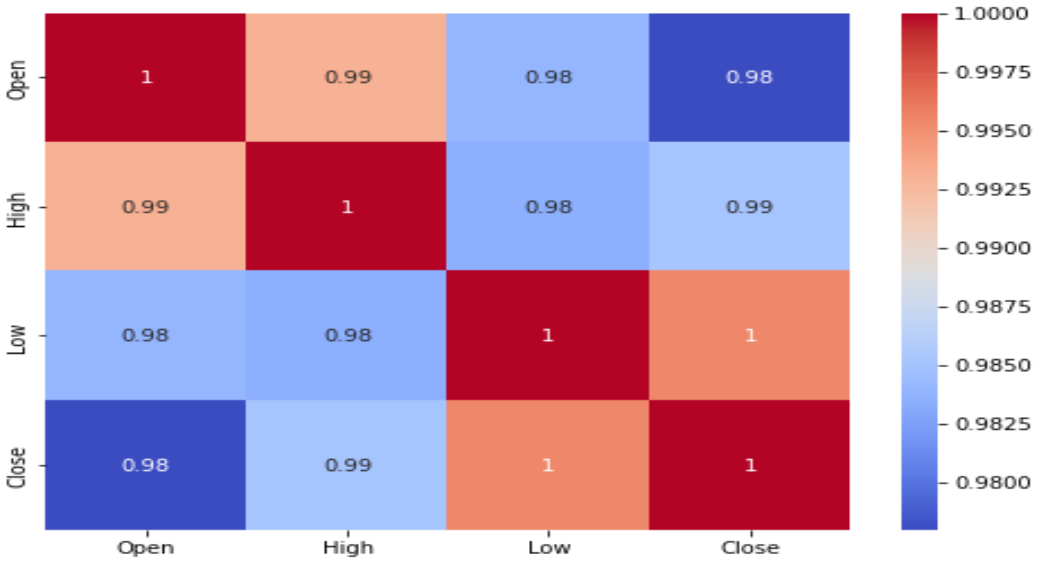
Relation between Dependent and Independent Variable:

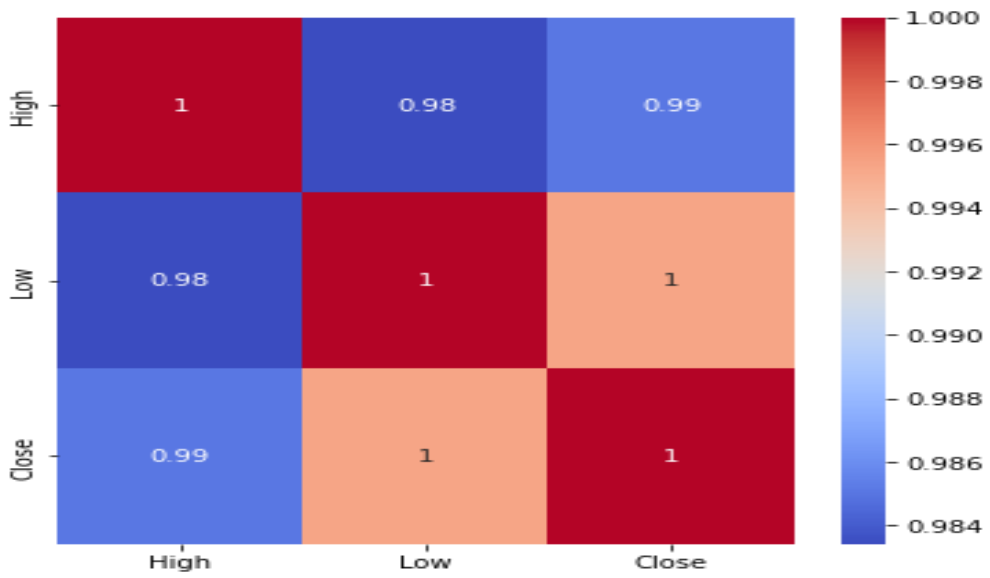






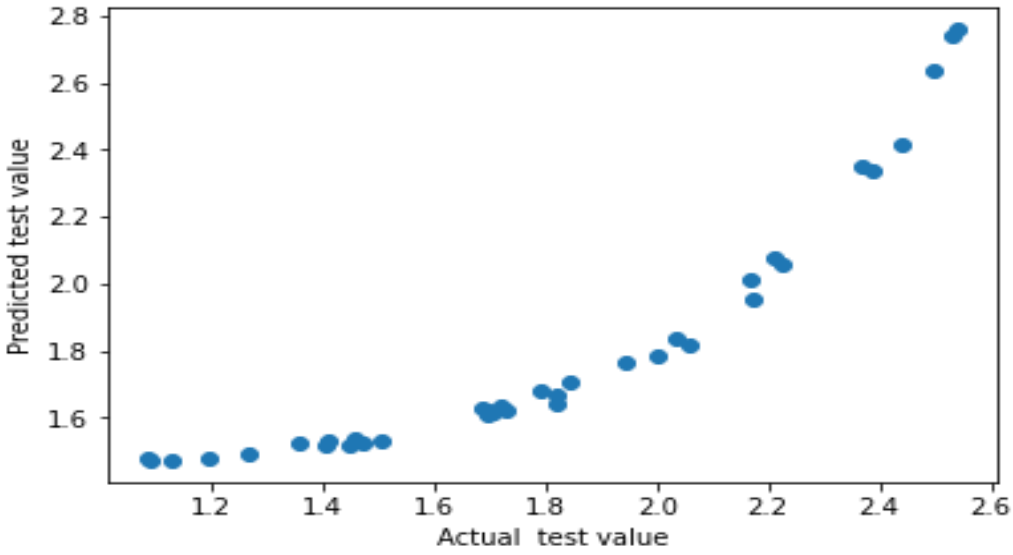
Corelation With Heatmap:





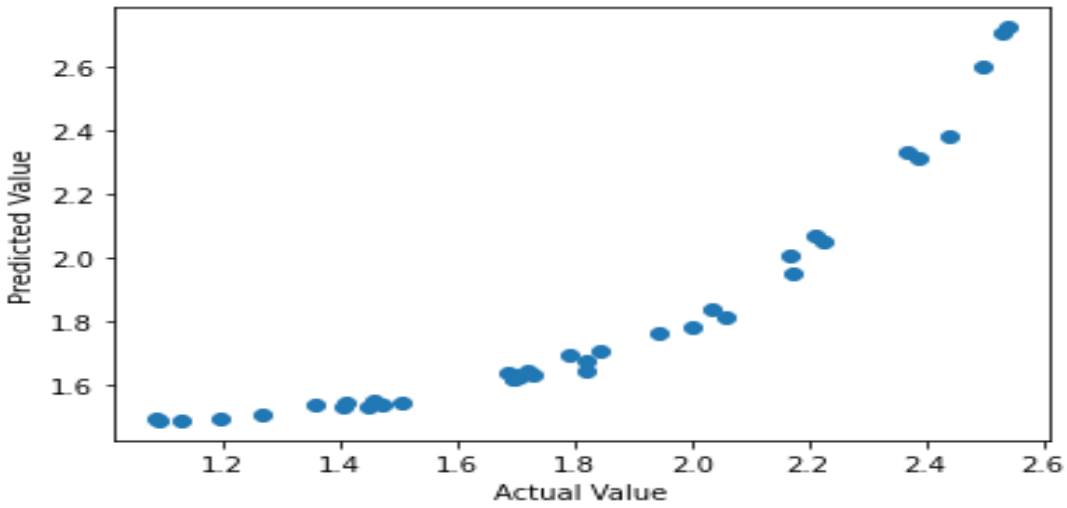
Linear Regression:

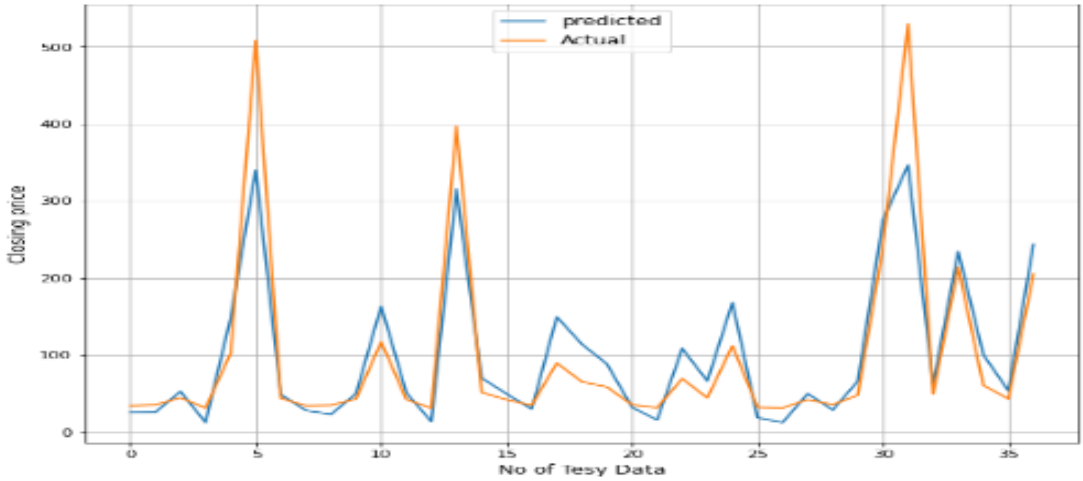
The most basic machine learning algorithm that can be implemented on this data is linear regression. The linear regression model returns an equation that determines the relationship between the independent variables and the dependent variable.



Lasso Regression:

Lasso Regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models(i.e.models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicolinearity or when you want to automate certain parts of model selection, like variable selection elimination.





Cross-validation, sometimes called rotation estimation or out-of-sample testing, is any of various similar model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set.

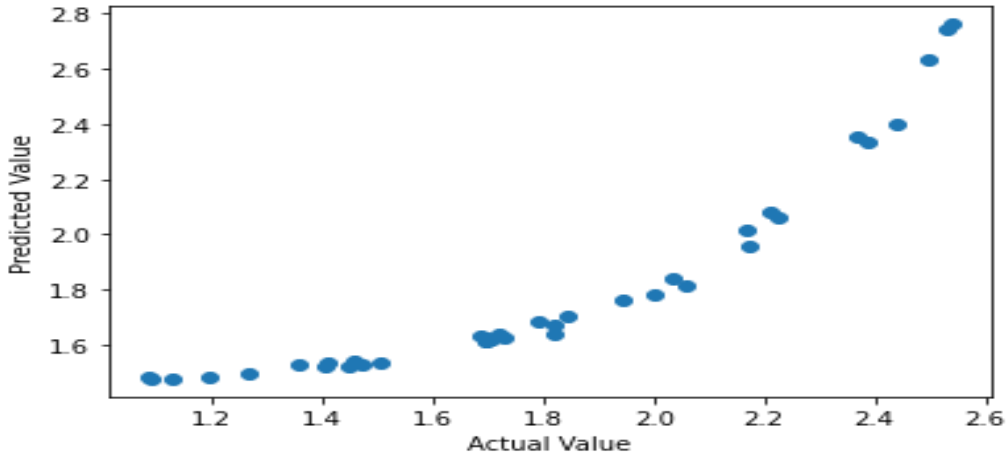
Cross-validation is a resampling method that uses different portions of the data to test and train a model on different iterations.

It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice.

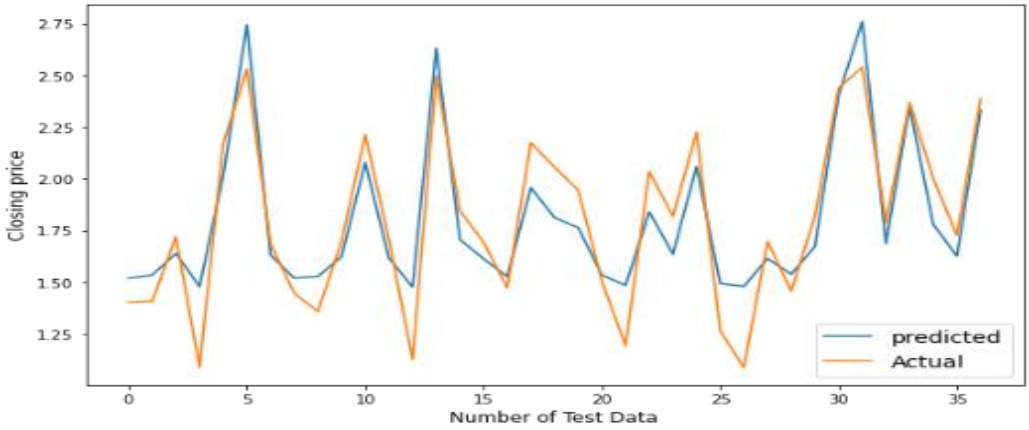
In a prediction problem, a model is usually given a dataset of *known data* on which training is run (*training dataset*), and a dataset of *unknown data* (or *first seen* data) against which the model is tested (called the validation dataset or *testing set*).

The goal of cross-validation is to test the model's ability to predict new data that was not used in estimating it, in order to flag problems like overfitting or selection bias and to give an insight on how the model will generalize to an independent dataset (i.e., an unknown dataset, for instance from a real problem).

To reduce variability, in most methods multiple rounds of cross-validation are performed using different partitions, and the validation results are combined (e.g. averaged) over the rounds to give an estimate of the model's predictive performance.

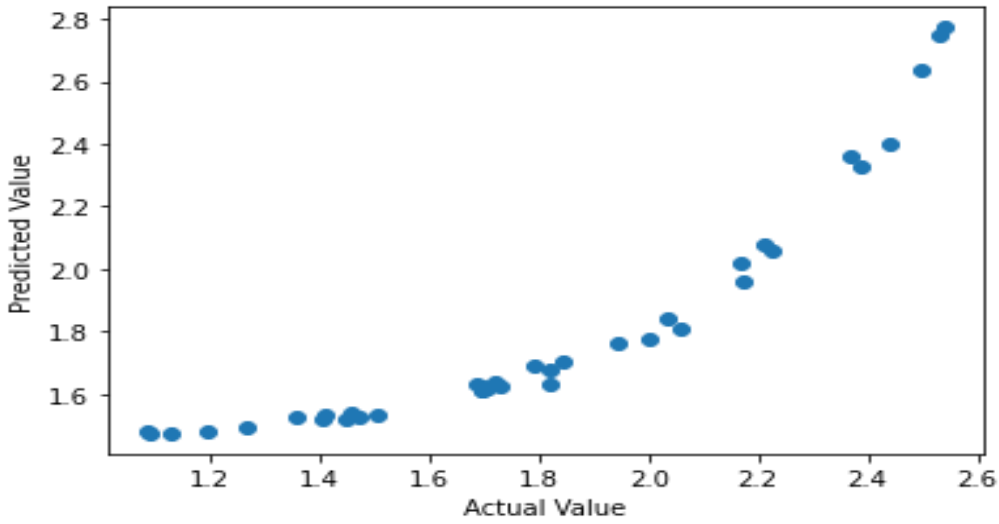


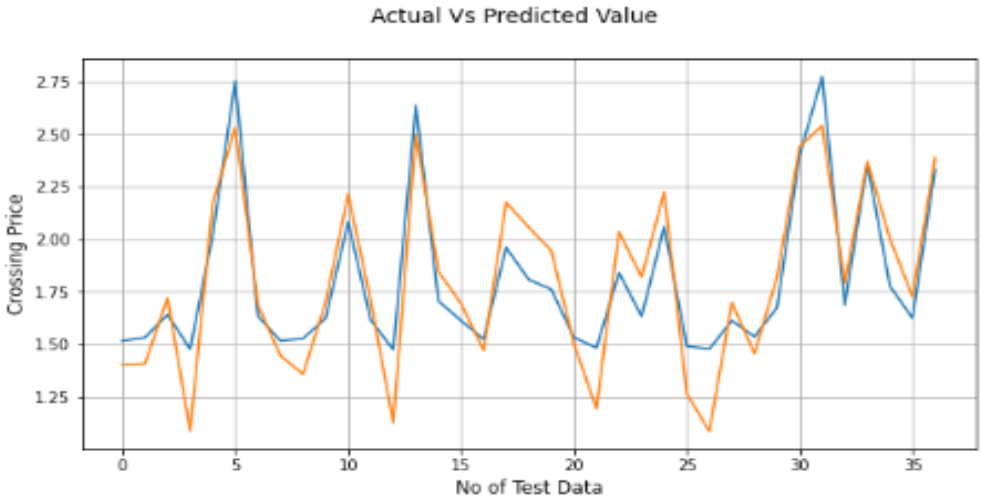
Actual vs Predicted price



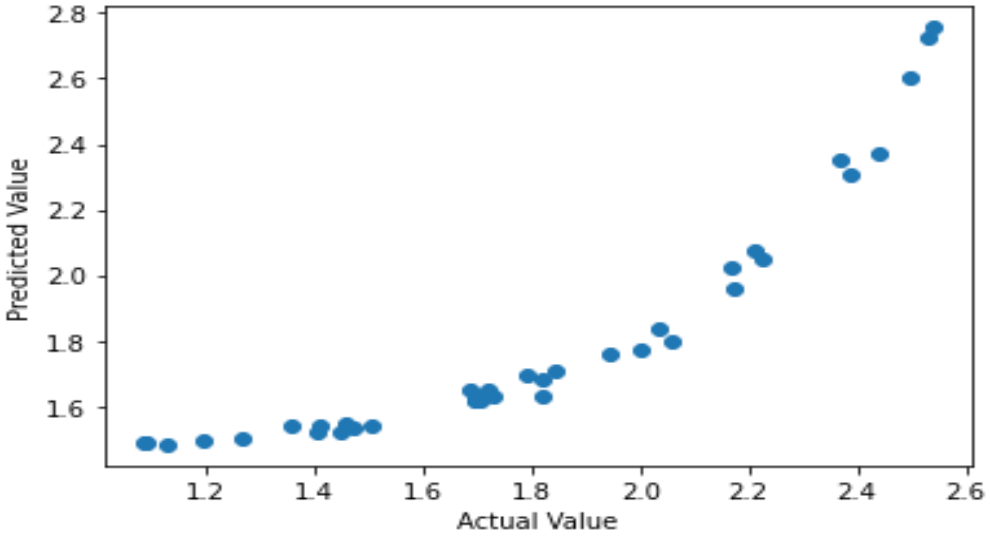
Ridge Regression:

Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where linearly independent variables are highly correlated.





Cross Validation for Ridge Regression:



CONCLUSION:

- The popularity of stock closing is growing extremely rapidly day by day which encourage researcher to find new methods if any fraud happens

- This technique is used for prediction is not only helpful to researchers to predict future stock closing prices or any fraud happen or not but also helps investors or any person who dealing with the stock market in order to prediction of model with good accuracy.

- In this work we use linear regression technique, lasso regression, ridge regression and elastic net regression technique. these four models gives us the following results

- High, low, open are directly correlate with the closing price of stocks.

- Target variable(dependent variable) strongly dependent on independent variables.

- We get maximum accuracy of 82%.