CAPSTONE PROJECT

Yes Bank Stock Closing Price Prediction

Done By:

PARUL SRIVASTAVA

CONTENT

INTRODUCTION

- Yes Bank is a commercial bank which provides variety of services like corporate banking, internet banking and personal banking for its customers.

- In 2020 to avoid the bank from collapsing because of its huge amount of loan it was taken over by the Reserve Bank of India.

- Because of this, it became interesting to see how that affected the stock prices of the bank and whether any other predictive models can do justice to such situations.

PROBLEM DESCRIPTION

- The dataset contains multiple variables - date, open, high, low and close. The column date contains the month and the year of the price of the share.

- The columns Open and Close represent the starting and final price at which the stock is traded in a particular month.

- High and Low represent the maximum and minimum price of the share for the month.

- The profit or loss calculation is usually determined by the closing price of a stock for the month; hence we will consider the closing price as the target variable.

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.

- The main objective is to perform regression analysis, which could help us in predicting stock's closing price of the month.

- The objective is also to apply different models and to check whether the price can be predicted using it and with how much accuracy.

- The dataset has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month.

OBJECTIVE

1. Data Analysing: -Performing Data Wrangling, Data Gathering, processing and transforming.
2. Data Cleaning: -Finding missing values and removing null/replicated values.
3. Drawing Inferences from the data.
4. Data Transformation
5. Performing Regression Analysis using: -
- Linear Regression
- Support Vector Machine
- Lasso Regression
- Ridge Regression
- K- Nearest Neighbour Regression
- Elastic Net Regression

EXPLORATORY DATA ANALYSIS

Plotting graph for closing price with respect to the year: -
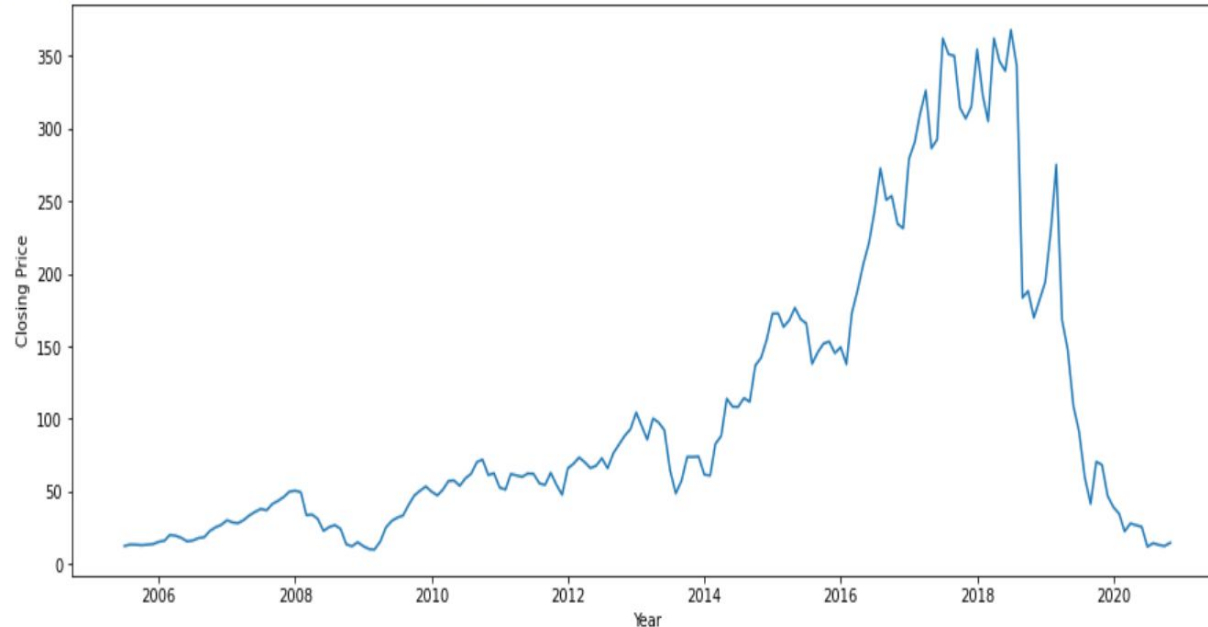
❑ The observations made from this is that the closing price increased from the year 2014 to 2018.

❑ And then there is huge downfall in the closing price after the year 2018.

❑ Here we can see that the stocks were high in the year 2017 and 2018 and then the downfall occurred because of the fraud case.

EXPLORATORY DATA ANALYSIS

Pair plot for showing correlation between variables: -
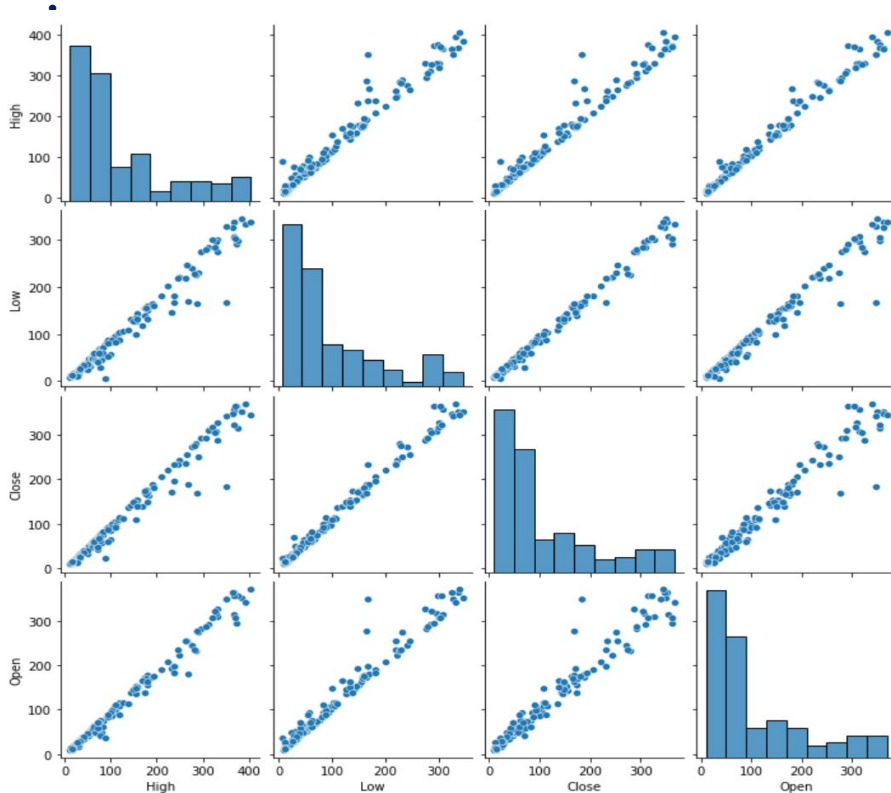
- ❑ From this pair plot the inference made is that the relationship between the variables namely- Open, Close, Low, High.

- ❑ It tells about how increase in one may lead to increase or decrease in the other.

- ❑ The correlation between all the variables as mentioned above is shown in the pairplot.

EXPLORATORY DATA ANALYSIS

Plotting heatmap to show correlation: -
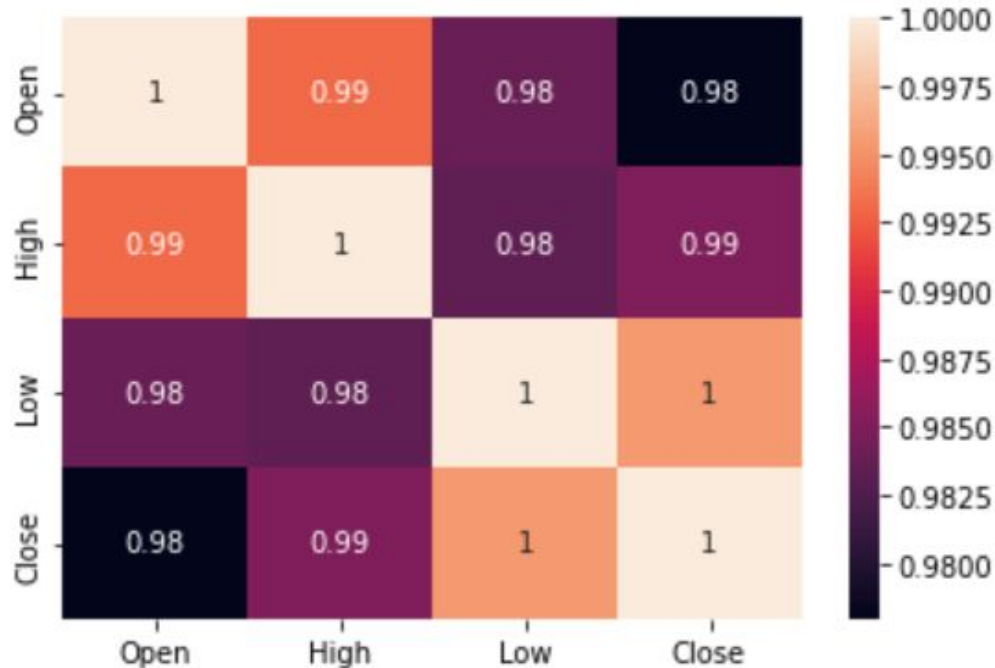
- ❏ From this heatmap plot the inference made is that there are very high correlation between independent variables.

- ❏ It results in the multicollinearity.

- ❏ High multicollinearity sometimes leads to unpredictability in the models and it is best practice to reduce it as much as possible.

EXPLORATORY DATA ANALYSIS

Distribution of dependent variable close price of stock:
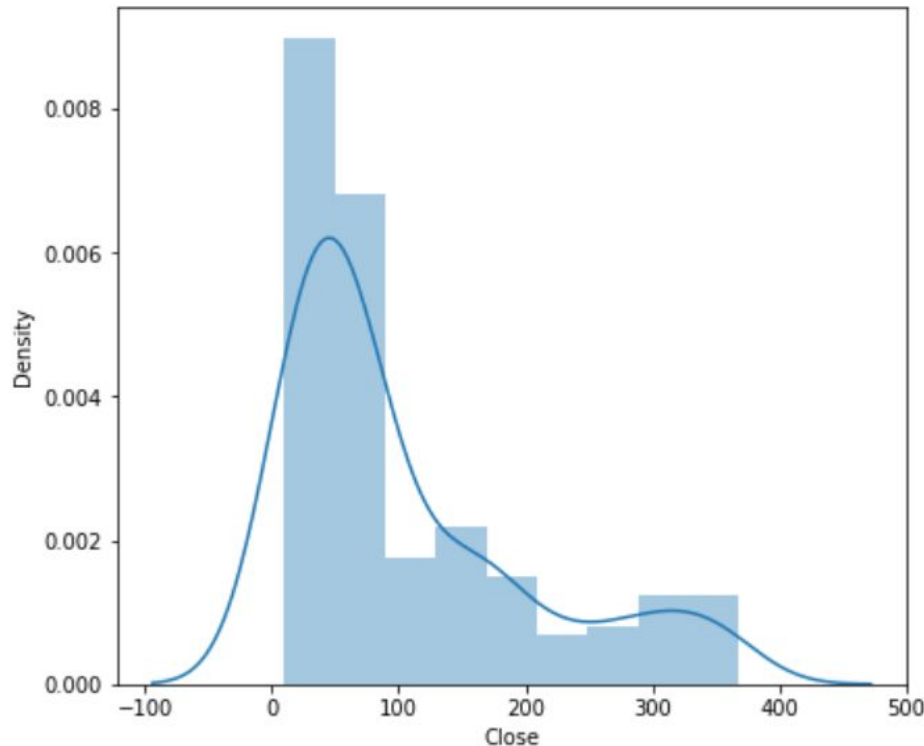
- ❏ From this displot the inference drawn that the distribution is rightly skewed.

- ❏ The rightly skewed distribution can be corrected by applying log10 transformation.

- ❏ And as a result, we got the following plot as shown and it needs to be transformed which can be done using the process of normalization, we will perform log normalization.

EXPLORATORY DATA ANALYSIS

After Transformation distribution of dependent variable close price of stock: -
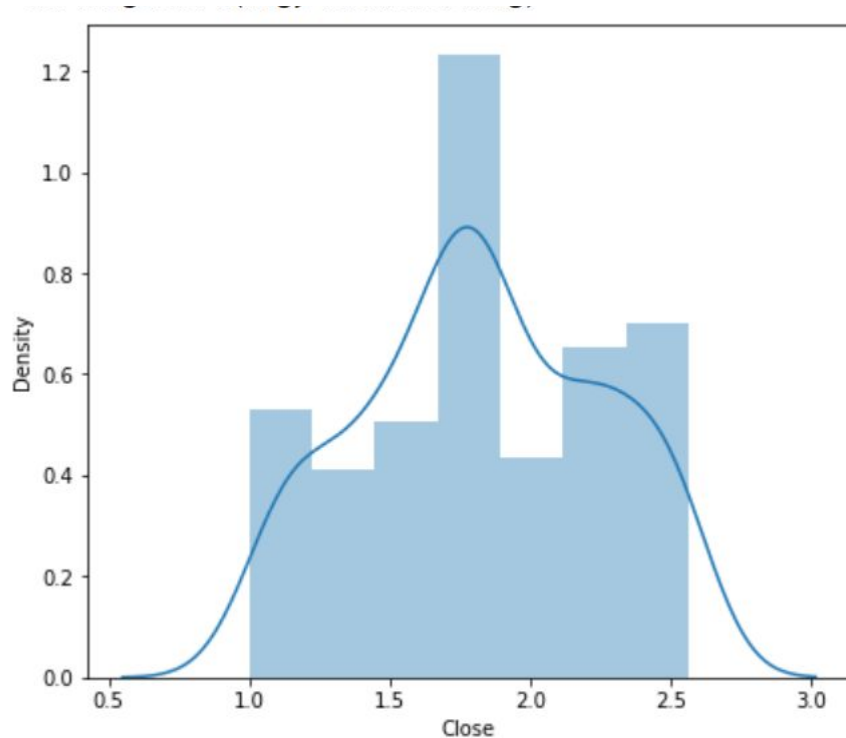
- ❑ From this displot the problem of rightly skewed is resolved.

- ❑ Using log10 transformation the displot for the same is shown and it gives more accuracy and better results.

DATA TRANSFORMATION

AI

DATA SPLITTING:

- ❏ X=Independent Variable (high, low, open)

- ❏ Y= Dependent Variable (close),

- ❏ Normalization must be performed to make the data have same scale which further makes data easy to process.

- ❏ Normalization is one of the best feature scaling techniques which is the most popular and widely used by many.

- ❏ Normalization is used when the data is skewed and when data needs to be transformed for its better understanding.
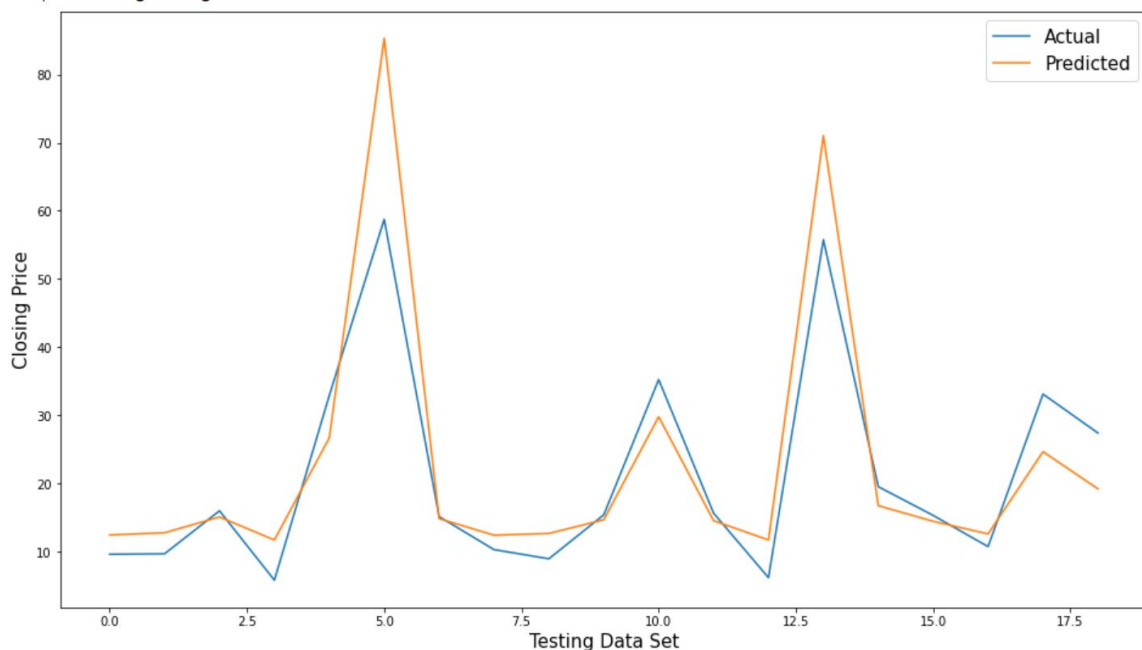
LINEAR REGRESSION ANALYSIS

AI

ACTUAL V/S PREDICTED CLOSED PRICE LINEAR REGRESSION-
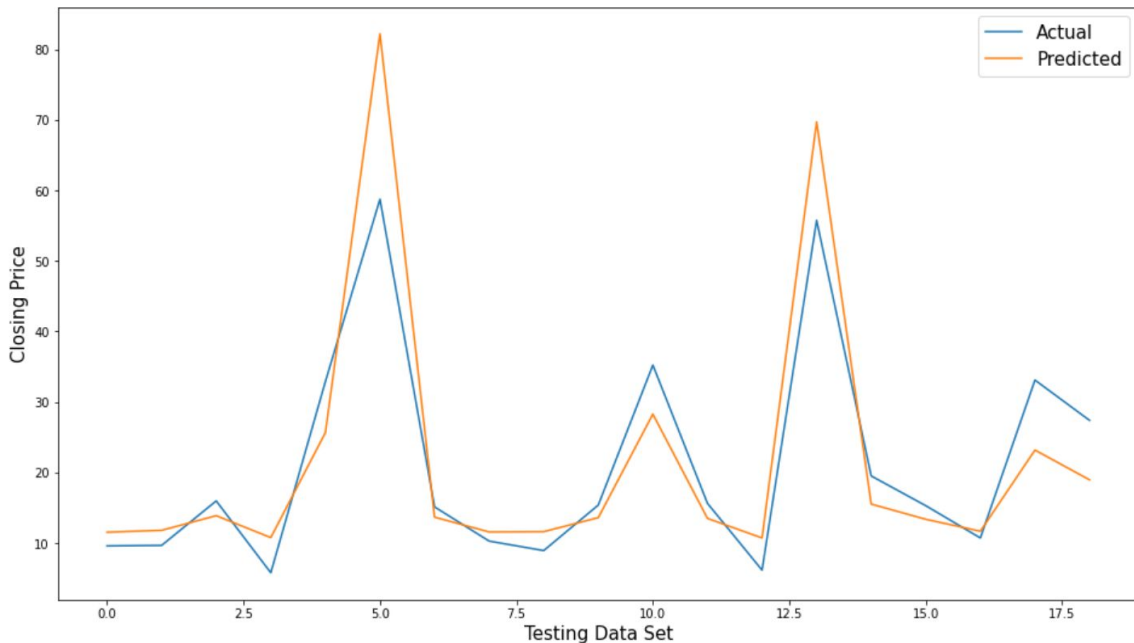
❑ Linear Regression is machine learning algorithm that is based on supervised learning, it performs regression task which targets prediction values based on independent variables.

❑ Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x).

❑ We got an R2 score with 80% Accuracy in linear regression.

SUPPORT VECTOR MACHINE

AI

ACTUAL V/S PREDICTED CLOSED PRICE SUPPORT VECTOR MACHINE ANALYSIS

❑ Support Vector Machine is another machine learning algorithm that is based on supervised learning.

❑ It performs both regression and classification task.

❑ We got an R2 score of 74% in support vector machine.

RIDGE REGRESSION

AI

ACTUAL V/S PREDICTED CLOSED PRICE RIDGE REGRESSION-
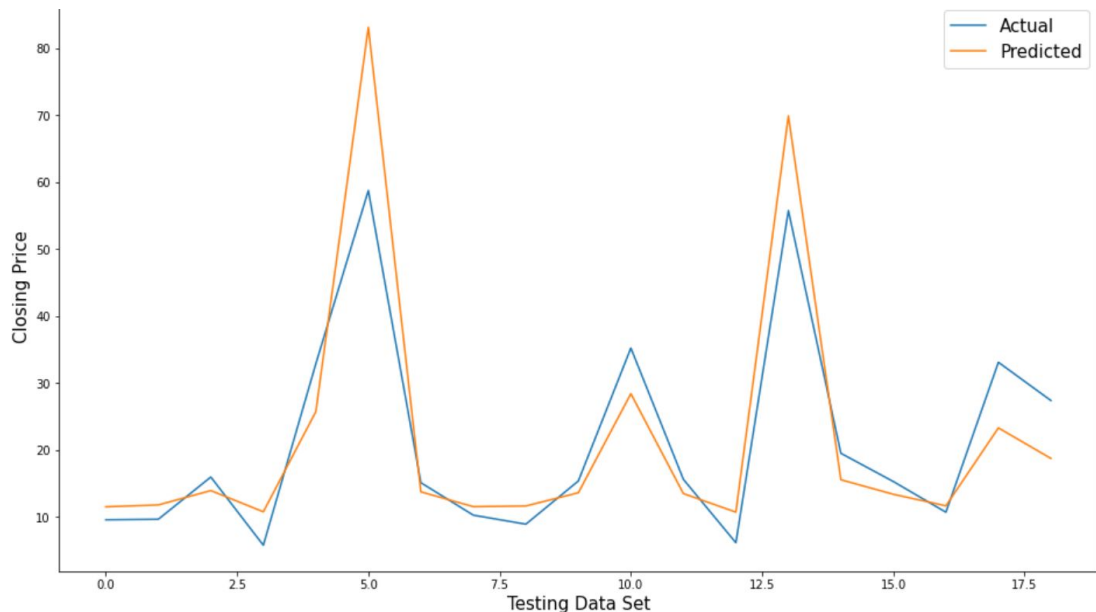
- ❏ It is also called as regularized linear regression; it comprises of the factor alpha which controls the regularization of model.

- ❏ It works on shrinking the coefficients and it is achieved by imposing the squared penalty on their size.

- ❏ We got an R2 score with 77% Accuracy in ridge regression.

LASSO REGRESSION

AI

ACTUAL V/S PREDICTED CLOSED PRICE LASSO REGRESSION-
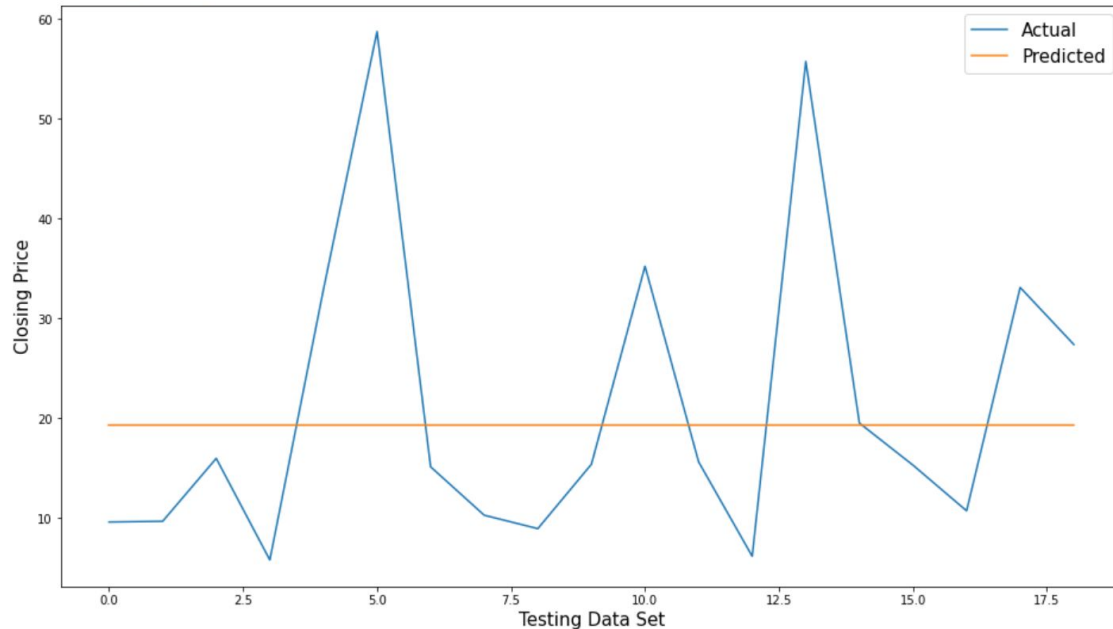
❏ Lasso stands for "Least Absolute Shrinkage and Selection Operator". It is another linear model derived from linear regression.

❏ It also uses the property called as the shrinkage where the coefficients are shrunk towards zero.

❏ Following is the graph showing lasso regression.

K-NEAREST NEIGHBOR

**ACTUAL V/S PREDICTED CLOSED PRICE
K-NEAREST NEIGHBOR ANALYSIS**
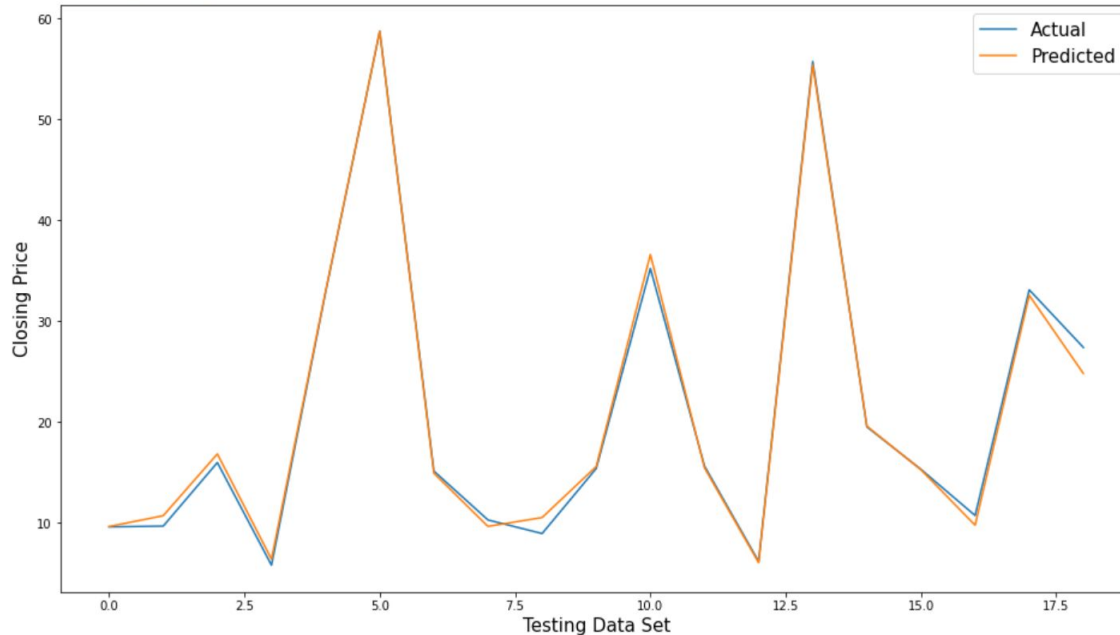
- ❑ Yet another supervised machine learning algorithm. It can be used to solve both classification and regression problems. It works on the basis that similar thing exists in the neighbour or close proximity.

- ❑ It basically takes k nearest neighbour and takes average of them. We got an R2 score of 98% using KNN.

- ❑ The following is the graph for actual v/s predicted in the case of k-nearest neighbour.
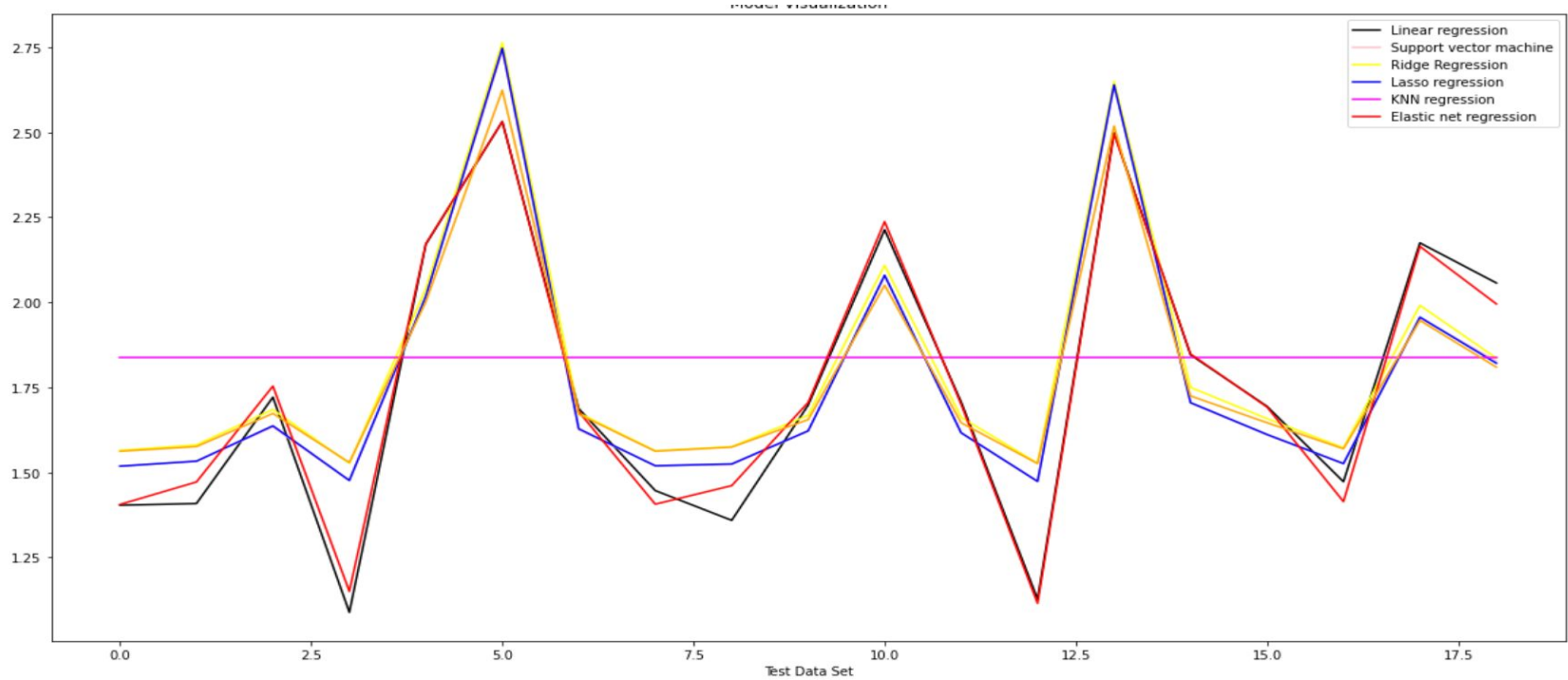
ELASTIC NET REGRESSION

ACTUAL V/S PREDICTED CLOSED PRICE ELASTIC NET REGRESSION-
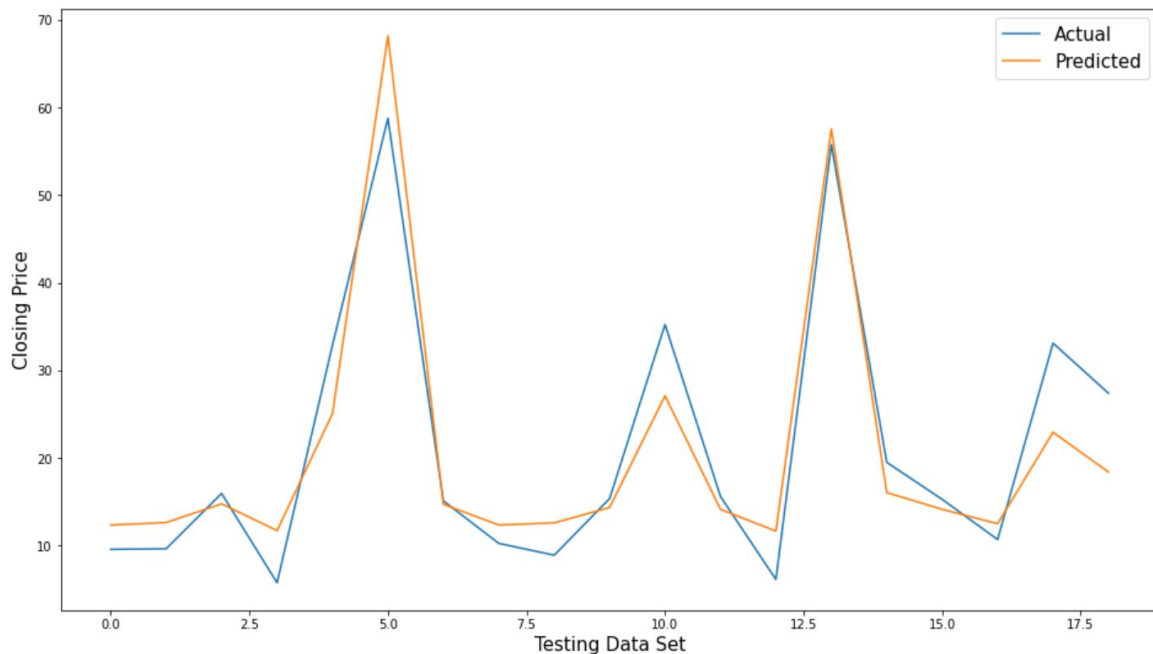
❑ It uses the penalties from both lasso and ridge regression to regularize the models.

❑ It further allows us to tune alpha=0 and alpha =1 in case of ridge and lasso regression.

❑ So, it allows us to choose between 0 and 1 to optimize elastic net regression. We were able to achieve an R2 score of 75%.

ALL MODELS VISUALIZATION

The chart above is showing all regression models such as linear regression, support vector machine, lasso regression, ridge regression, k-nearest neighbour and elastic net regression.

CONCLUSIONS

- The mean square error values for linear regression, support vector machine, ridge regression and elastic test prediction were somewhat same.

- The root mean square values of linear regression and ridge regression were somewhat same.

- Lasso regression gave highest mean square and highest root mean square value.

- While on the other hand lasso regression gave lowest mean absolute error an $R2$ score.

- The highest $R2$ value was of K-nearest neighbour as it gave 98% accuracy.

- Also, for mean absolute error K-nearest neighbour gave the highest accuracy of 99%.

- Accuracy was mostly same in the case of elastic net regression and support vector machine.

- The maximum accuracy attained was 99%.

REFERENCES

For this capstone project we have taken help from sources as mentioned:-

❏ Almabetter Video Lectures and live classes.

❏ GeeksforGeeks Website.

❏ Towards Data Science Website.

THANK YOU