Capstone Project-2

Yes Bank Stock Closing Price Prediction

Team

Name: Web Crawlers

Members:

1. Deepmala Srivastava
2. Raman Kumar
3. Rajat Chaudhary
4. Anukriti Shakyawar

Contents

- Introduction
- Objective
- Data Preprocessing
- Data Summary
- Exploratory data analysis
- Regression Analysis
  - Linear Regression
  - Ridge Regression
  - Lasso Regression
  - ElasticNet Regression
- Conclusion

Introduction

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

- We used Regression Analysis to predict the future stock price of this company. Starting with linear regression, and then move on to Ridge Regression, Lasso Regression and ElasticNet Regression

Objective

The main objective is to predict the stock's closing price of the month.

Data Summary

We have Yes Bank monthly stock price dataset. Dataset has 185 entries & 5 columns. It has following features (Columns):

1) Open: Opening price of the stock of particular day

2) High: It's the highest price at which a stock traded during a period

3) Low: It’s the lowest price at which stock traded during a period

4) Close: Closing price of a stock at the end of a trading

5) Date: We will use it as a index

Note: ‘Close’ will be our Dependent variable & Others will be independent.

Data Pre-processing

After data inspection we found:

- ❑ No missing values in any instances.
- ❑ No duplicates values in any instances.
- ❑ Date have values in object data type. So, we need to convert it into proper date format.

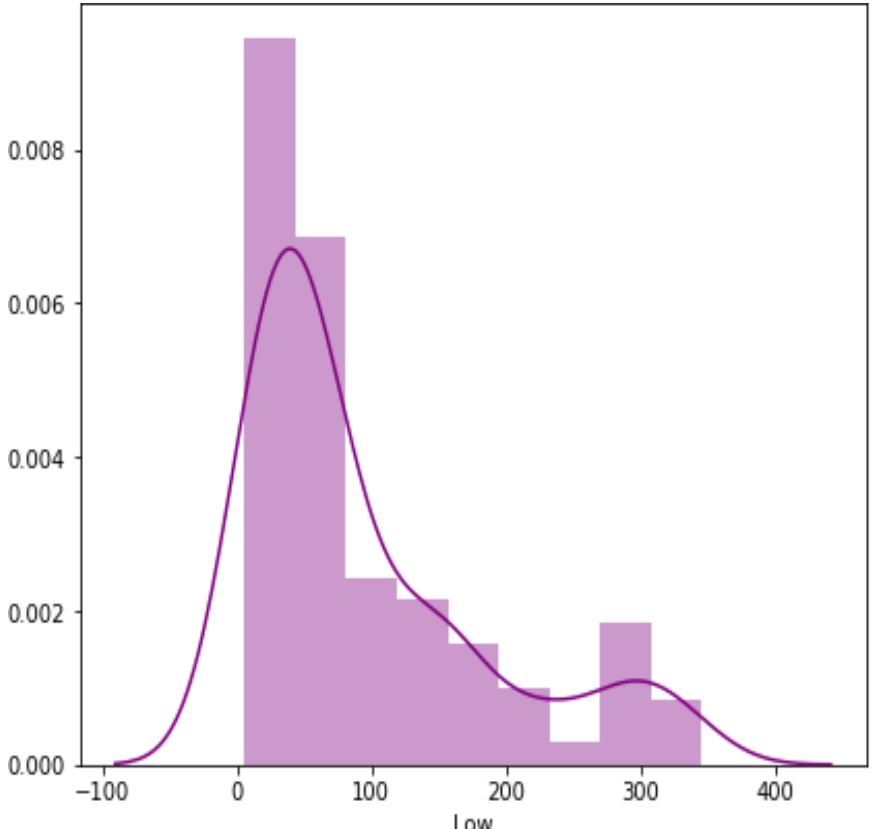
Exploratory Data Analysis(EDA)

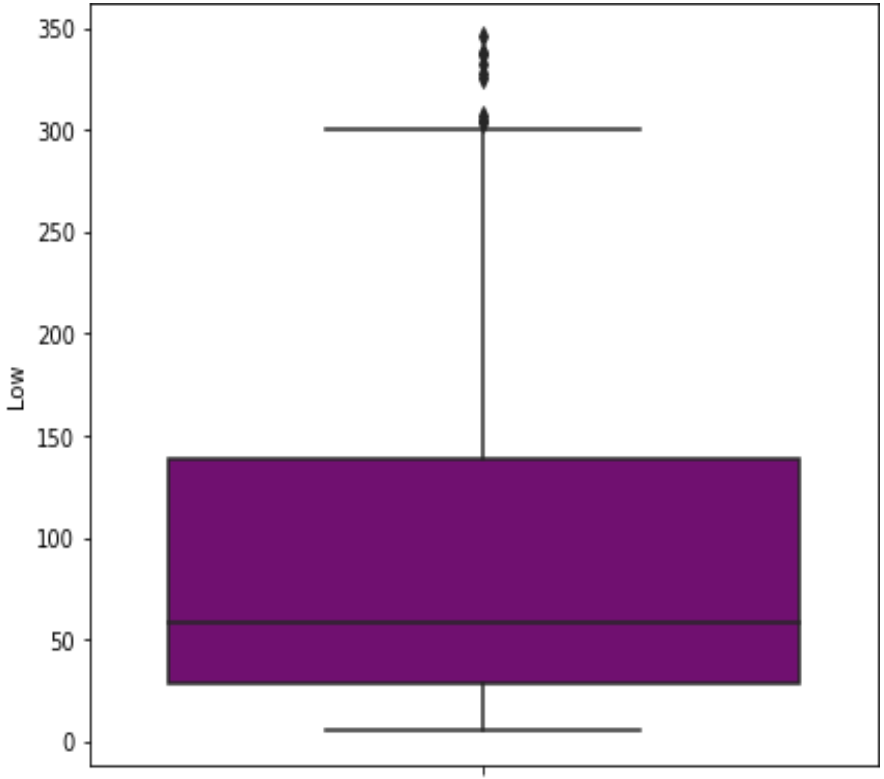
Visualize dependent variable

This plot of Closing prices of different dates shows that fluctuation in prices regarding different time-duration. After 2018 there is sudden fall in the stock closing price. It makes sense how severly Rana Kapoor case fraud affected the price of Yes bank stocks.

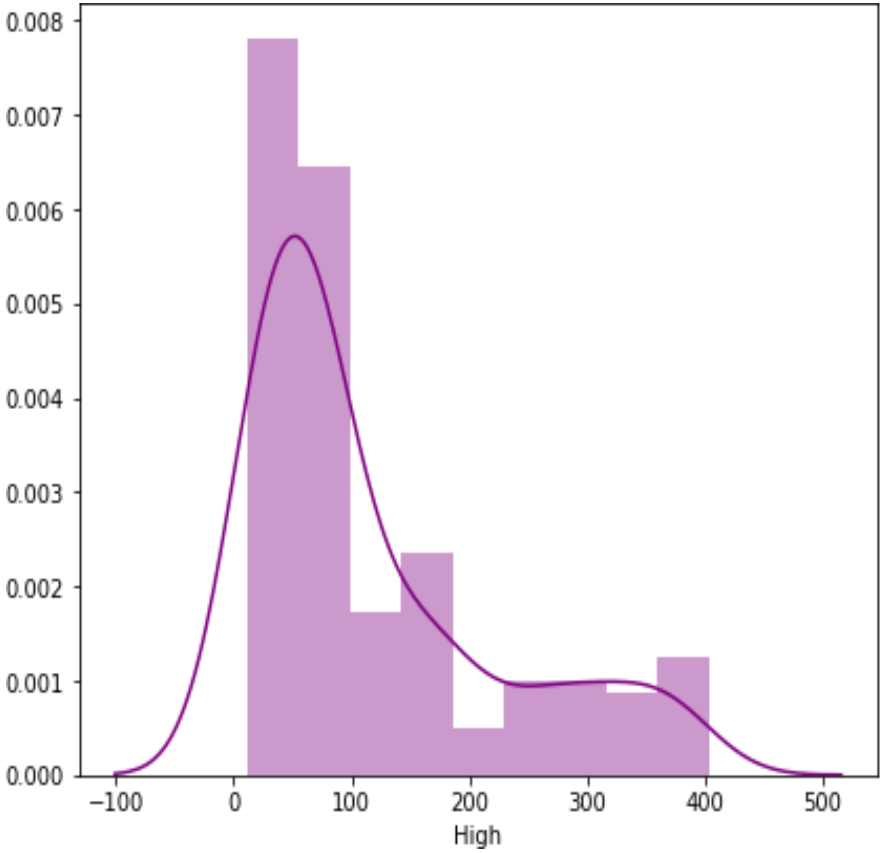
Distribution of Independent variables

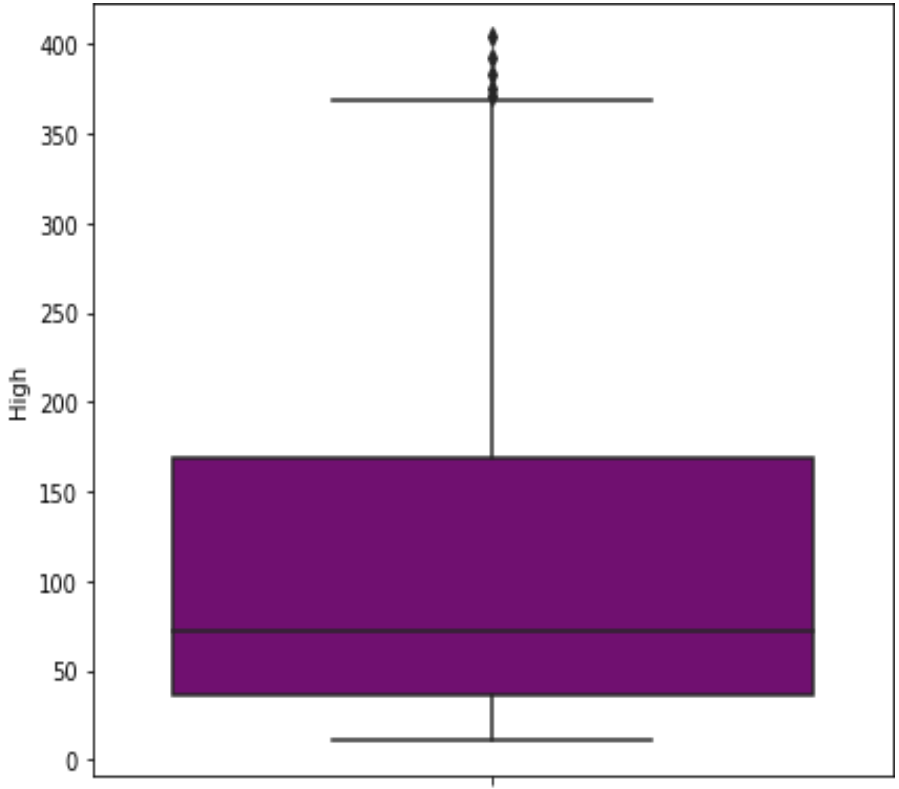
Distribution of ‘Low’



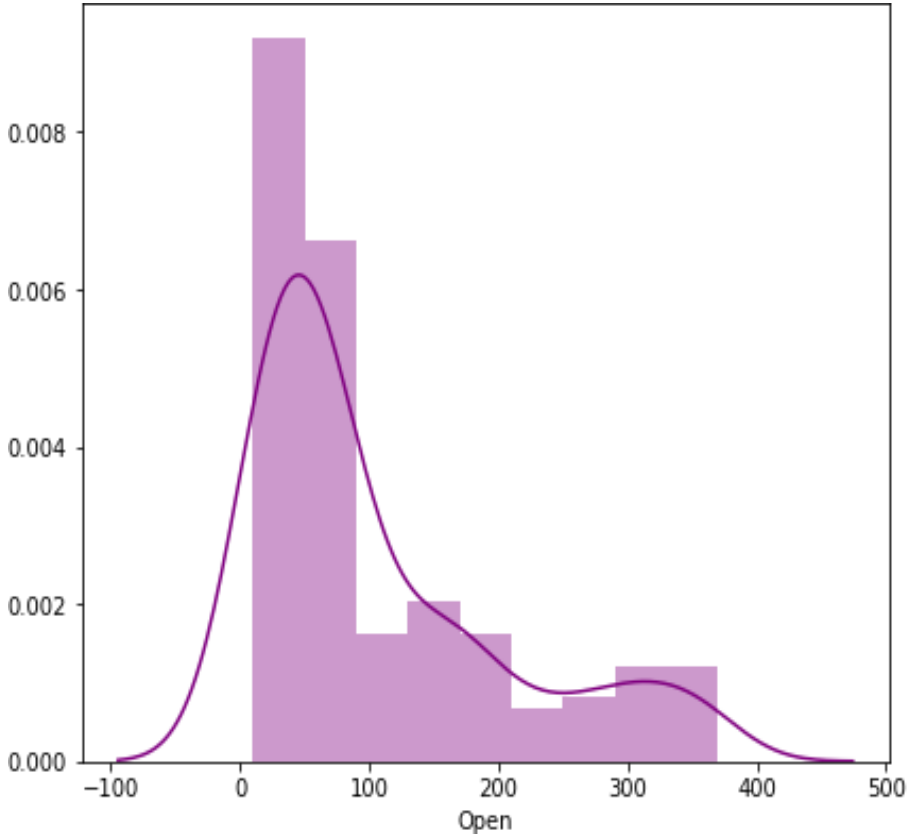


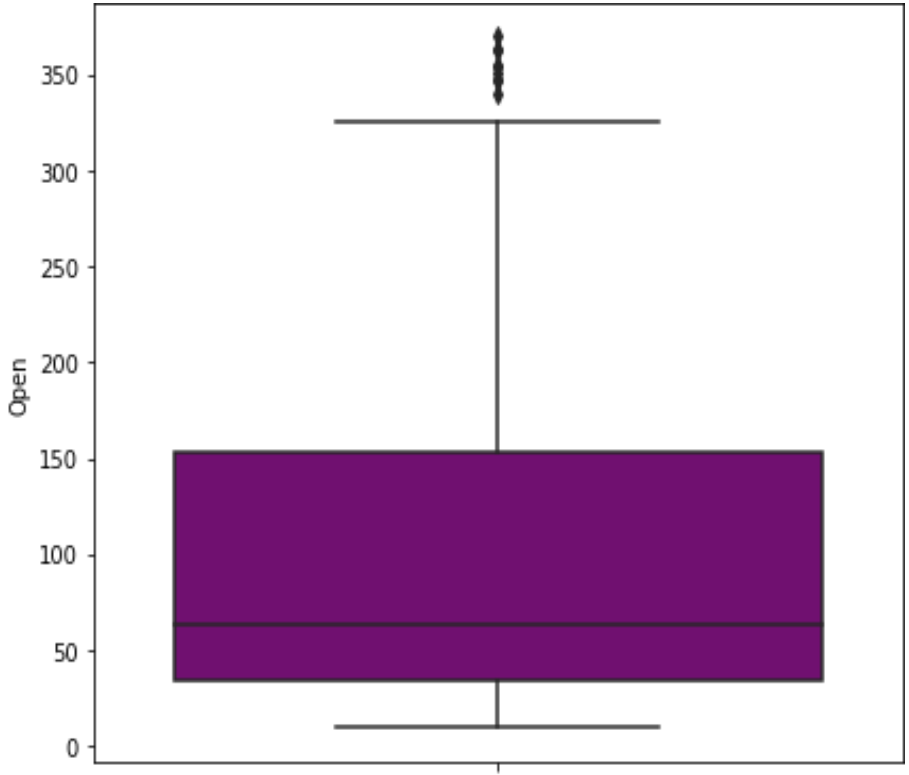
Distribution of ‘High’





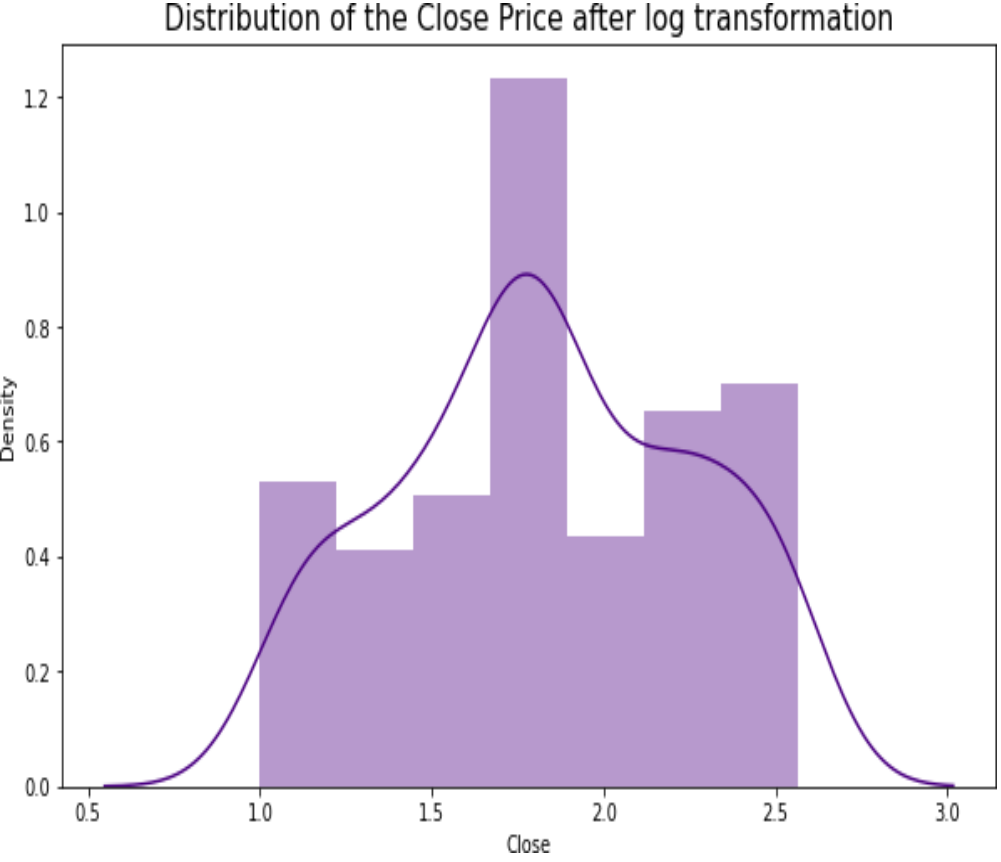
Distribution of ‘Open’





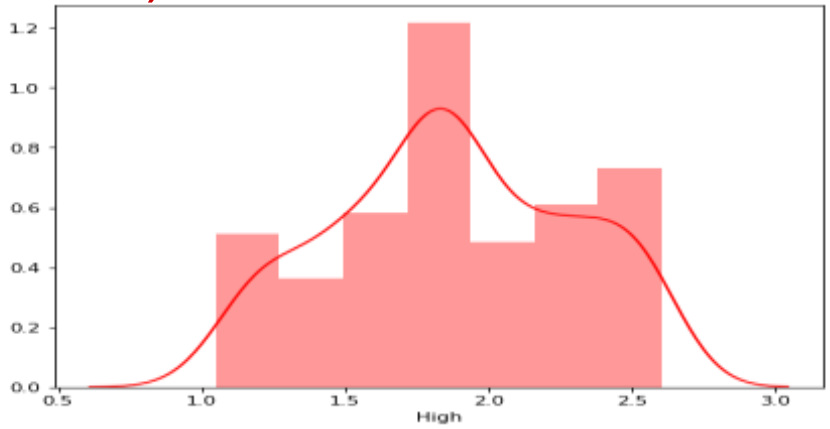
Distribution of Dependent Variable

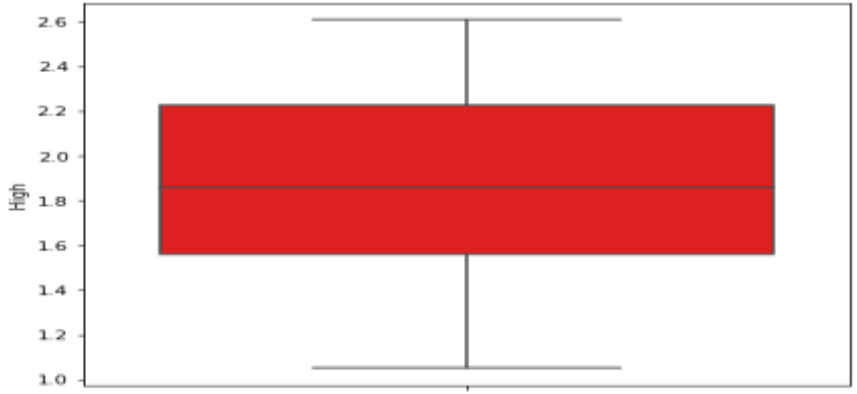
Transformation of Data

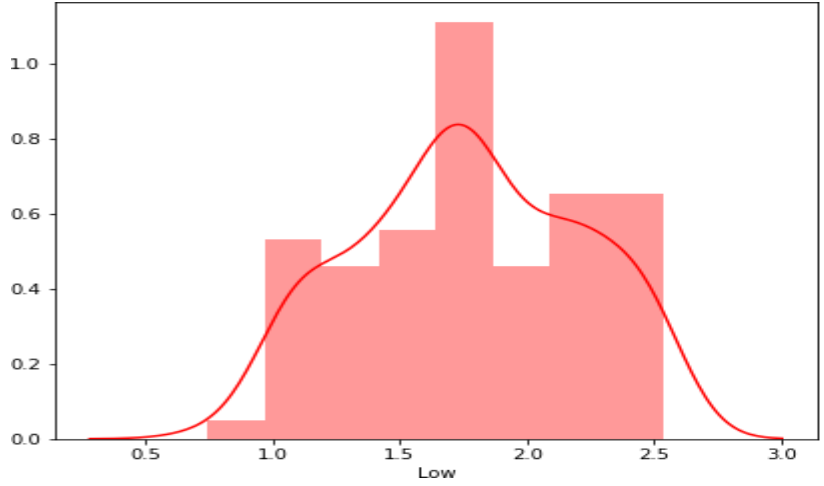


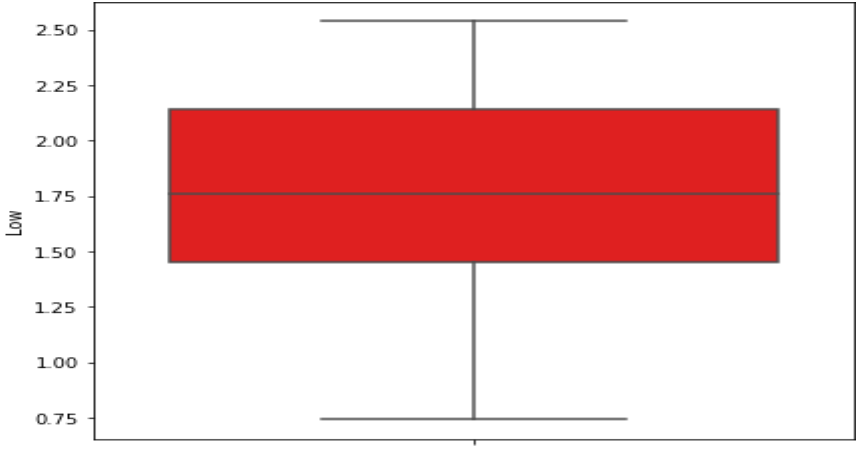
- From EDA We observed that data was found to be skewed. So we transformed the data to make it uniform before passing it into our machine learning models.
- We applied log transformation. The figure shows the distribution of close price after a log transformation.

(Cont..)

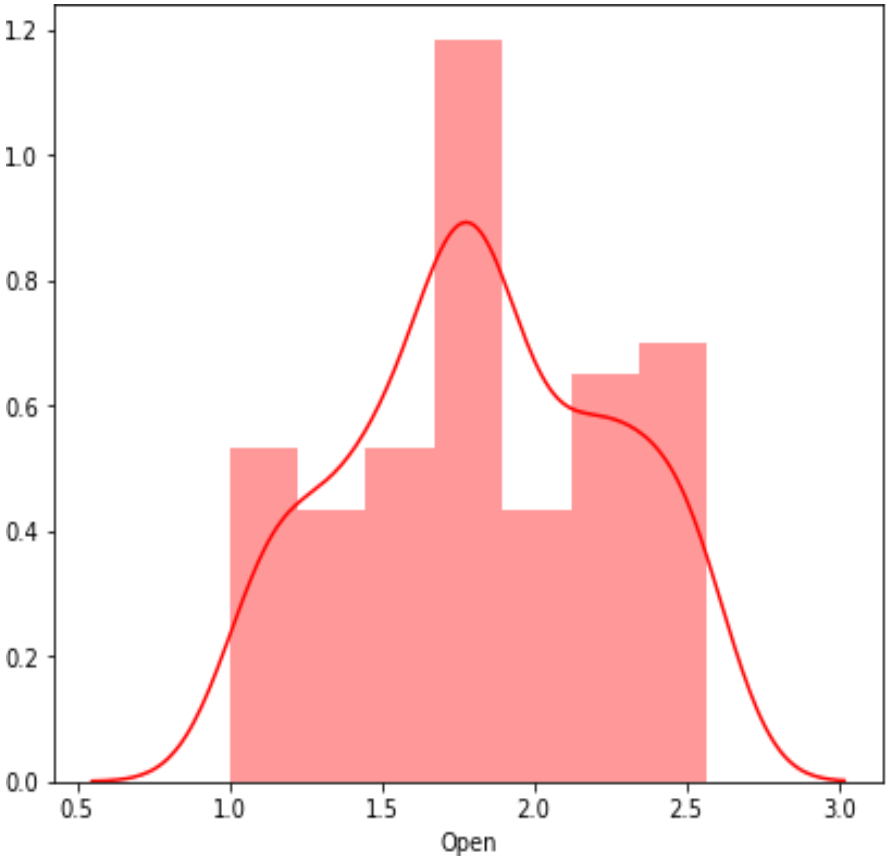

1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
High

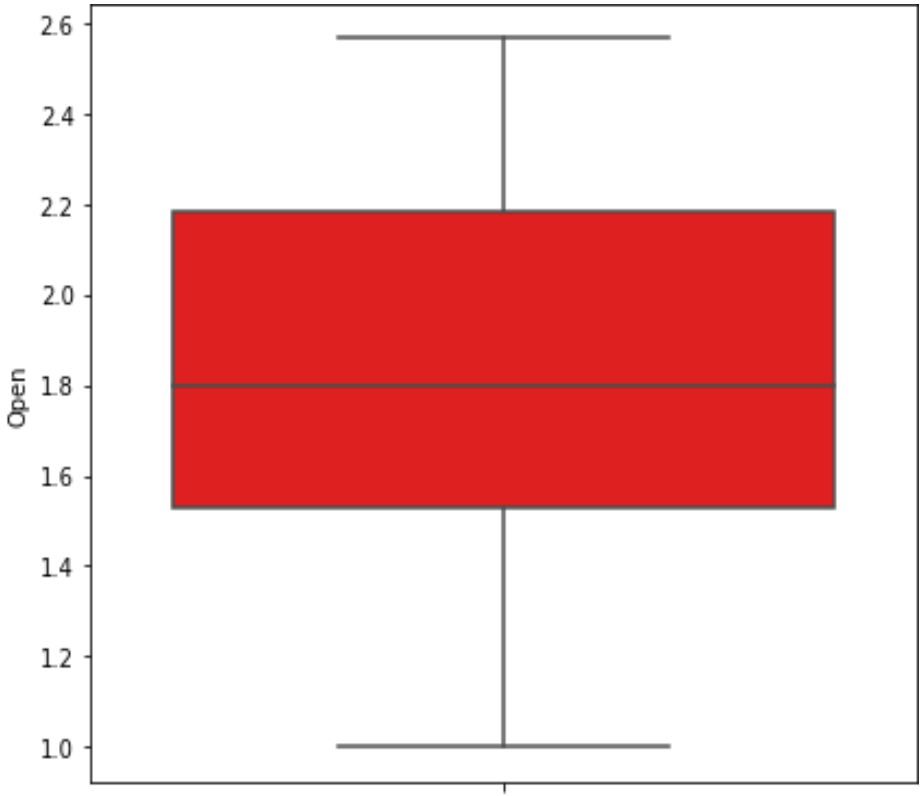

2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
High


1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Low

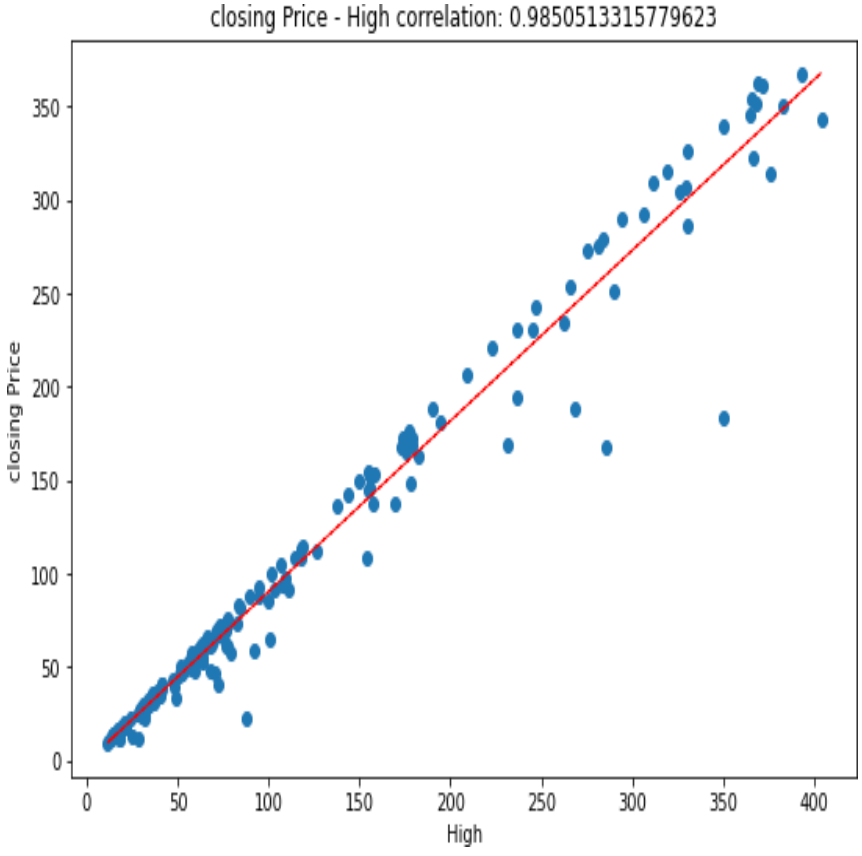

2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
Low

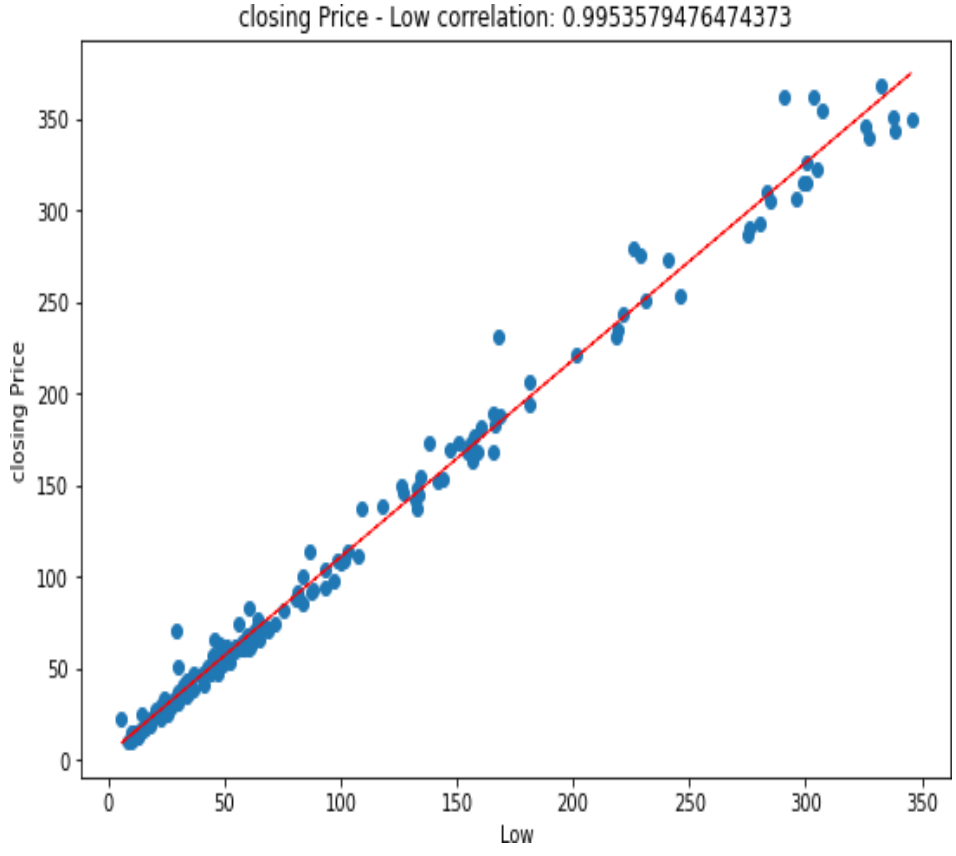
(Cont..)


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
Open

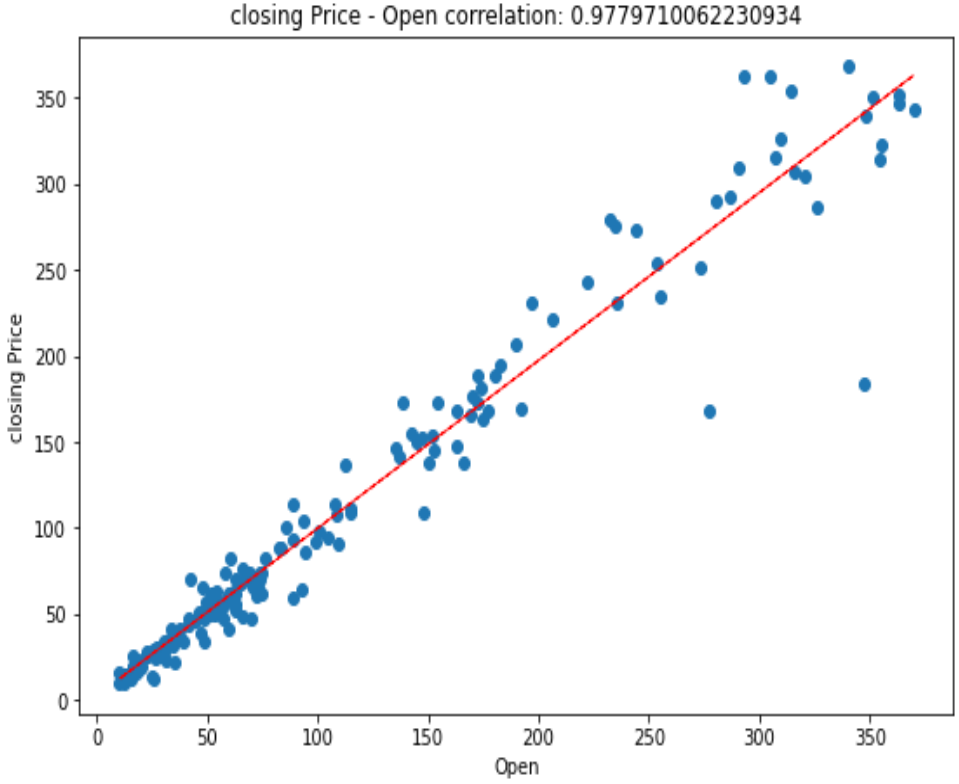

2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
Open

Correlation of ‘Closing Price’ with Independent Features:





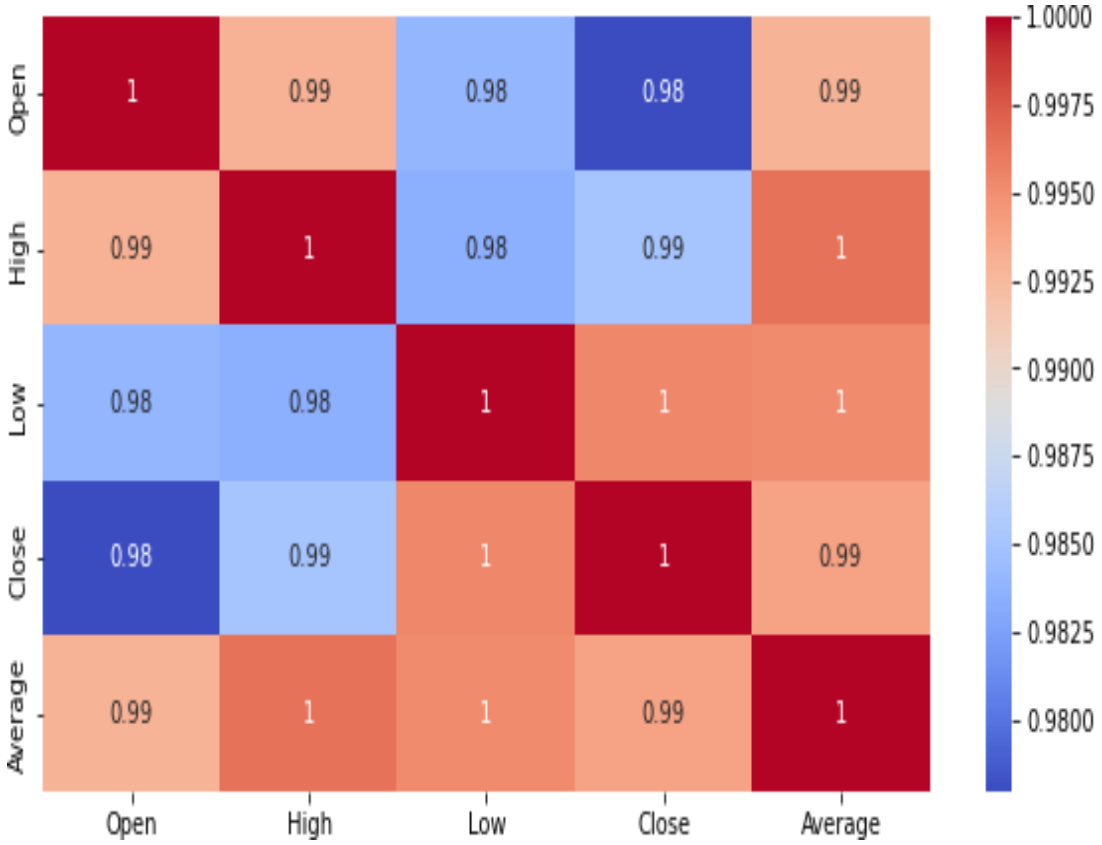
(Cont..)



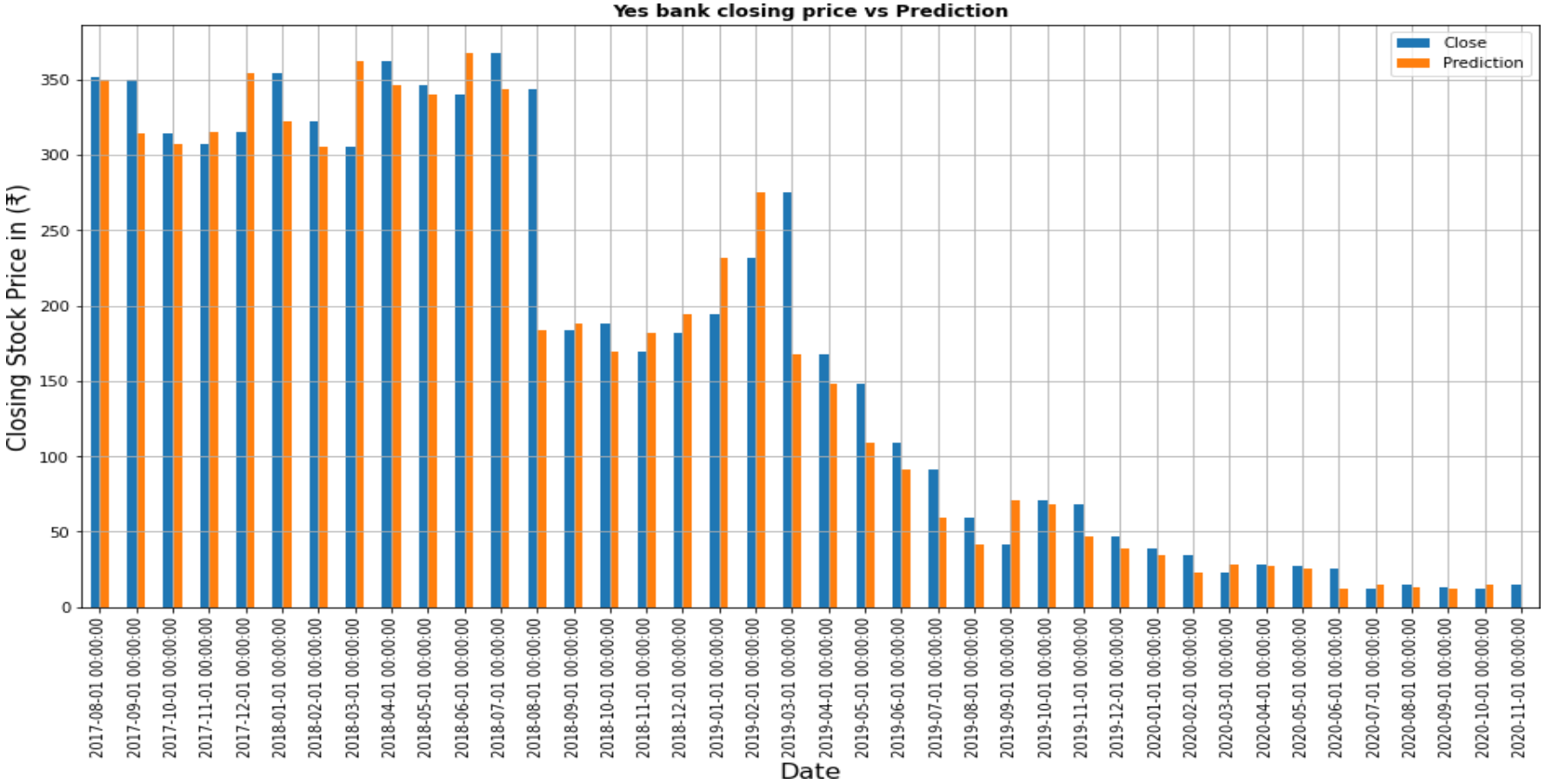
- There is linear relation and high correlation between each independent variables and dependent variable.
- The correlation is 0.985, 0.995, 0.978 .
- This suggests a high level of correlation, e.g. a value above 0.5 and close to 1.0.

Correlation Matrix

- The correlation matrix helps us visualize the correlation of each parameter with respect to every other parameter.
- The shades changes from the highest to lowest (or vice versa) correlations.
- We observed from matrix that dependent variable (close price) is highly correlated with all the other independent variables



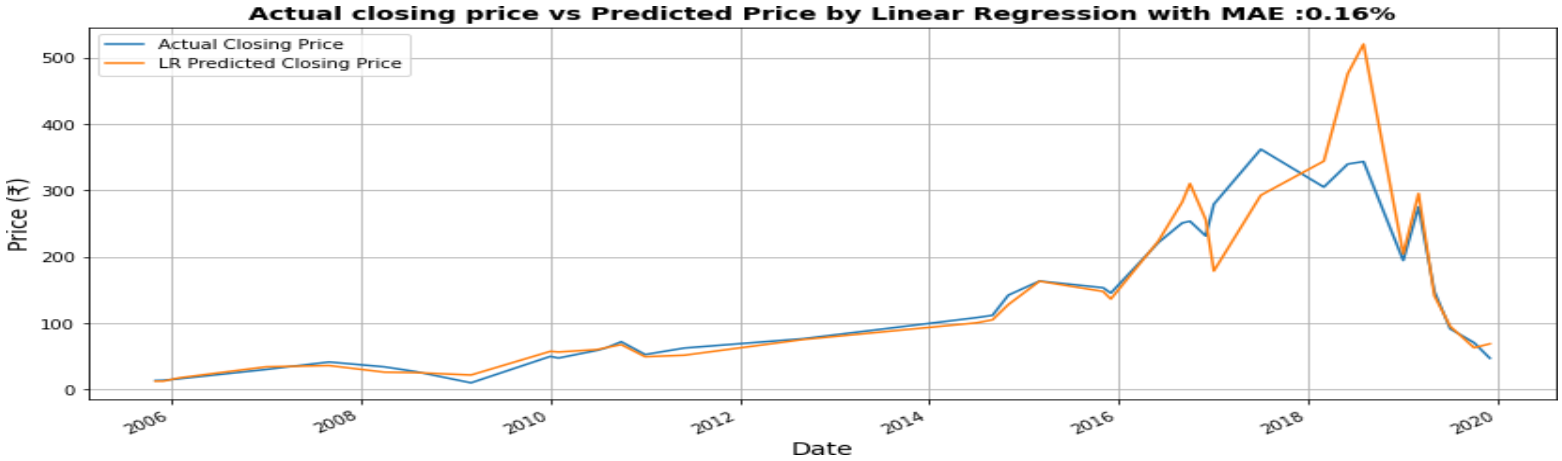
Bar Graph Comparison between Actual and predicted Price (price predicted by Shift() function



Model Selection

- We passed the data into following algorithm :
  - Linear Regression Model
  - Lasso Regression
  - Ridge Regression
  - ElasticNet
- We checked the performance of the model across various parameters.
- Then we decided our best models on the basis of following metrics
  - R2
  - Adjusted R2
  - Training Accuracy
  - Mean squared error and Root mean squared error.

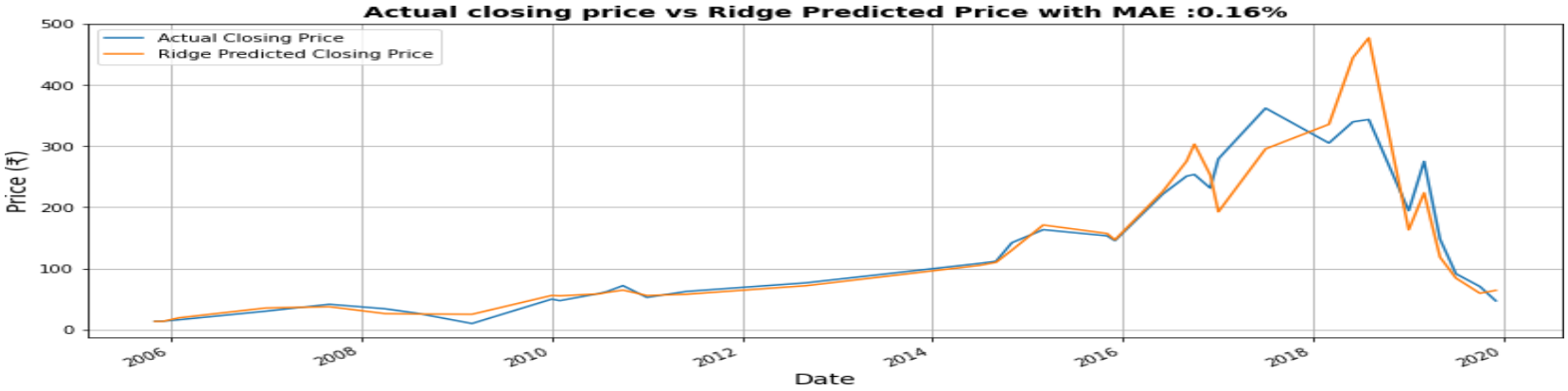
Linear Regression



Explanations:

- Linear Model predicted the close price with 0.16% Mean Absolute error. It has training accuracy 94.03%.
- R2 tells us that our independent variable is able to describe 95% of our dependent variable.
- Adjusted R2 is about 91.44%, just because we consider 17 independent features adjusted R2 would be the best matrix to consider.

Ridge Regression



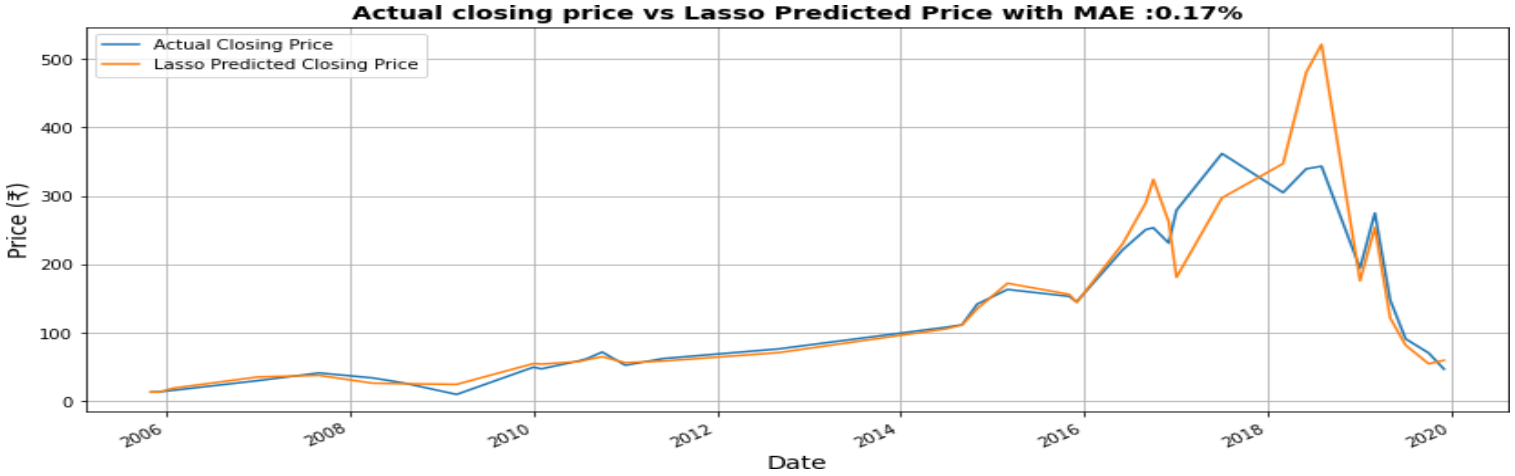
Explanations:

• Ridge predicted the close price with 0.16% Mean Absolute error. Having training accuracy 94.57%.

• R2 is about 95.25% which means model's independent features is able to describe 95.25% of our dependent variable.

•Adjusted R2 is about 91%. We'll consider adjusted R2 because we have too many independent features

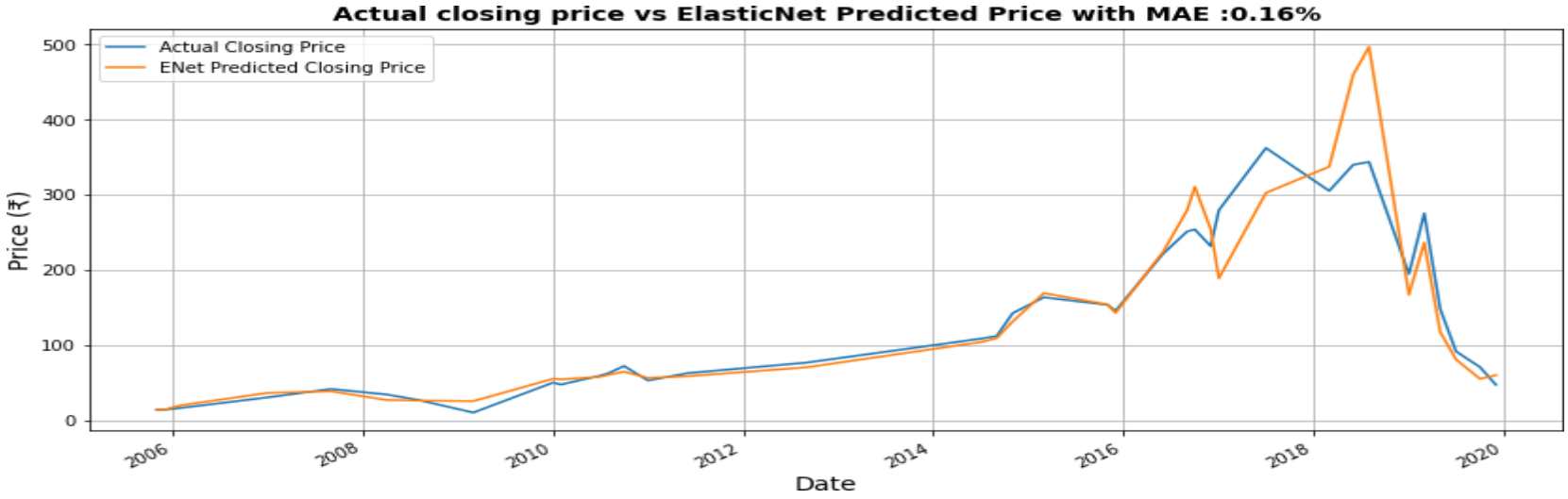
Lasso Regression



Explanations:

- ❑ Lasso predicted the close price with 0.17% Mean Absolute error. Having training accuracy 94.57%.
- ❑ Here, R2 is about 94.96% which means models' independent features is able to describe 94.96% of our dependent variable.
- ❑ Adjusted R2 is about 90.46%. We'll consider adjusted R2 because we have too many independent features.

Elastic Net Regression Before Cross Validation:



Explanations:

- ❑ Our Linear Model predicted the close price with 0.16% Mean Absolute error. Having training accuracy 82.64%.
- ❑ R2 tells us that our independent is able to describe 95.11% of our dependent variable.
- ❑ Adjusted R2 is about 90.74%, just because we consider 17 independent features adjusted R2 would be the best matrix to consider.

Comparispn among all Model predictions in one graph:

Final Explanation: In this combined comparison graph among actual closing price and Predicted closing price predicted by all four models, Linear Regression and Lasso are predicting closing price of next month better than Ridge and Elastic Net.

All four models have good R2 and Adjusted R2.

Final Matrix:

	Linear Regression	Ridge	Lasso	Elastic Net
MSE	0.008368	0.008848	0.009377	0.009096
RMSE	0.091477	0.094061	0.096833	0.095375
R2	0.955079	0.952505	0.949664	0.951169
Adjusted_R2	0.914887	0.910009	0.904627	0.907478
Training Accuíacy	0.940359	0.945655	0.945777	0.826402

Conclusion

- Target Variable is strongly dependent on independent variables.
- Linear Regression and Lasso are performing better than other models with training accuracy **94.0359%** and **94.7881%** respectively.
- Apart from Linear Regression and Lasso, Ridge and Elastic Net is also performing better but they have less training accuracy.
- Ridge and Elastic Net is performing far much better after applying Hyperparameter Tuning and Cross-validation, it is because we have small set of datasets.
- R2 and AdjustedR2 are around 95% and 91% in each model.