Capstone Project-2

Yes Bank Stock Closing Price Prediction

Team

Name: WebCrawlers

Team Members

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

Contents

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

Introduction

- Yes Bank Limited is an Indian Private Sector Bank headquartered in Mumbai, India, and was founded by Rana Kapoor and Ashok Kapoor in 2004. It offers a wide range of banking and financial products for corporate and retail customers through retail banking and asset management services.

- 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

This dataset contains Monthly Stock Prices of the Yes Bank and includes Open, High, Low and Close stock prices of every month. The main objective is to Predict the Future Stock Closing Price of the Month.

Data Pre-processing

- First check for duplicate values and missing values and treat them if any values present .
- Next, we check the datatype of the features present in our dataset and transform them if necessary .

Data Splitting

We have Yes Bank monthly stock price dataset. It has following features (Columns):

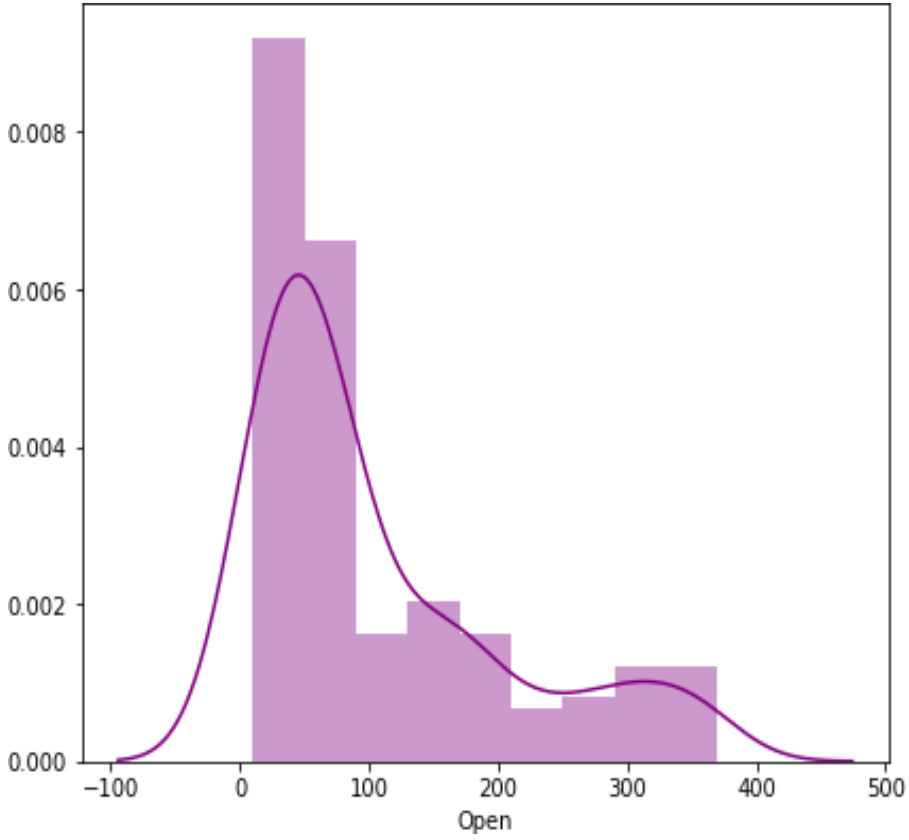
1) Open: Opening price of the stock of particular day
2) High: It's the highest price at which a stock traded during a period
3) Low: It’s the lowest price at which stock traded during a period
4) Close: Closing price of a stock at the end of a Trading Day
5) Date: We will use it as a index

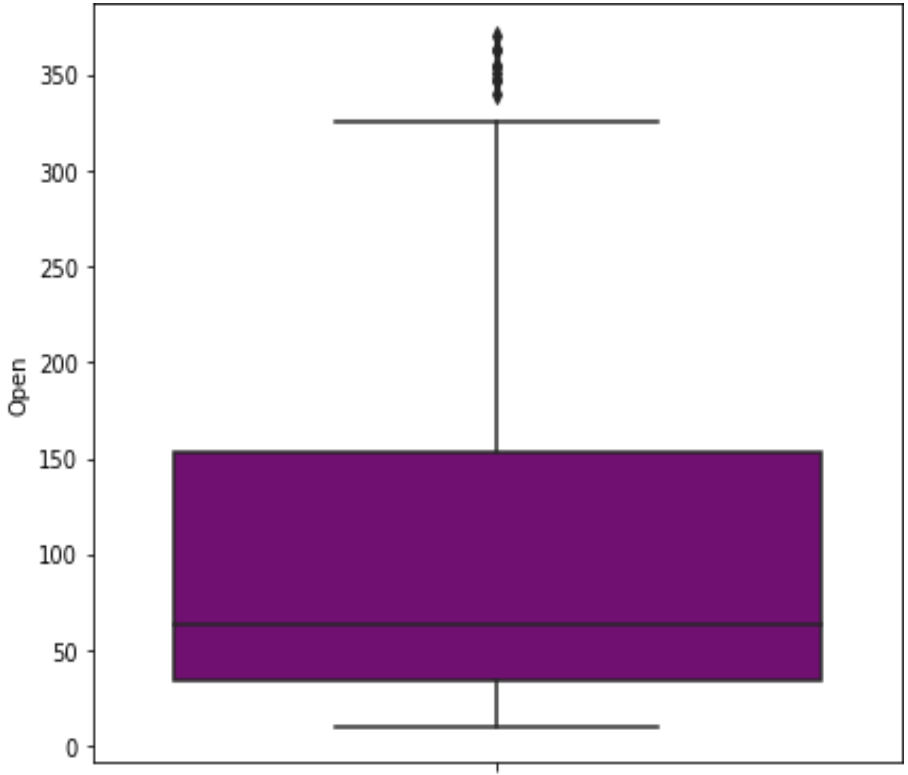
Note: ‘Close’ will be our Dependent variable & Others will be independent.

Exploratory Data Analysis(EDA)

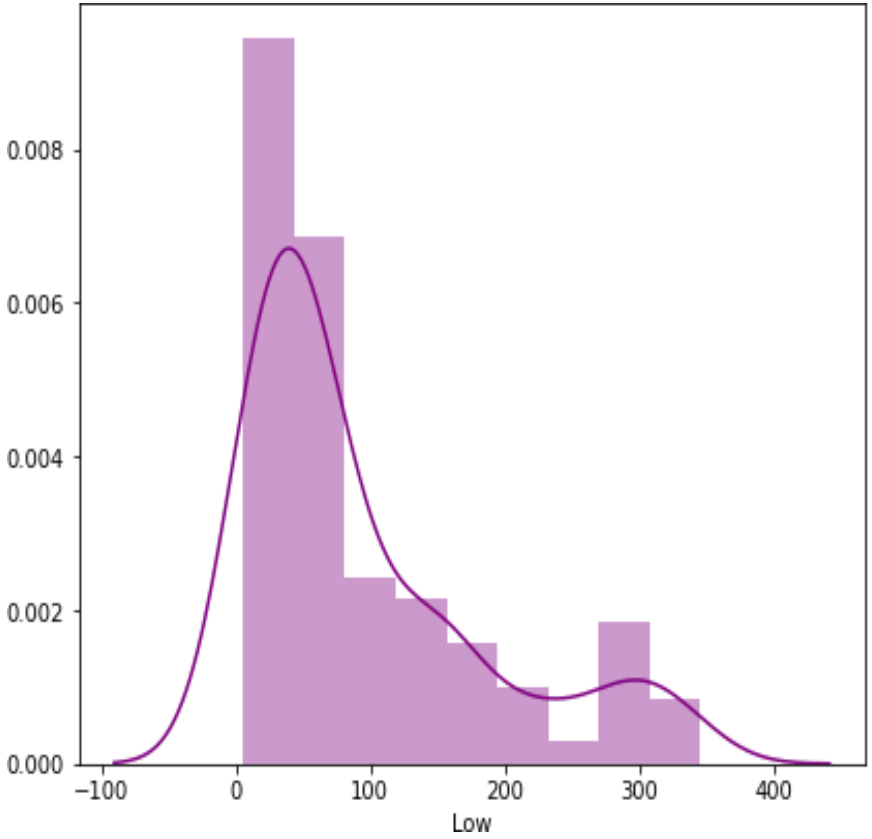
Closing Price vs Date

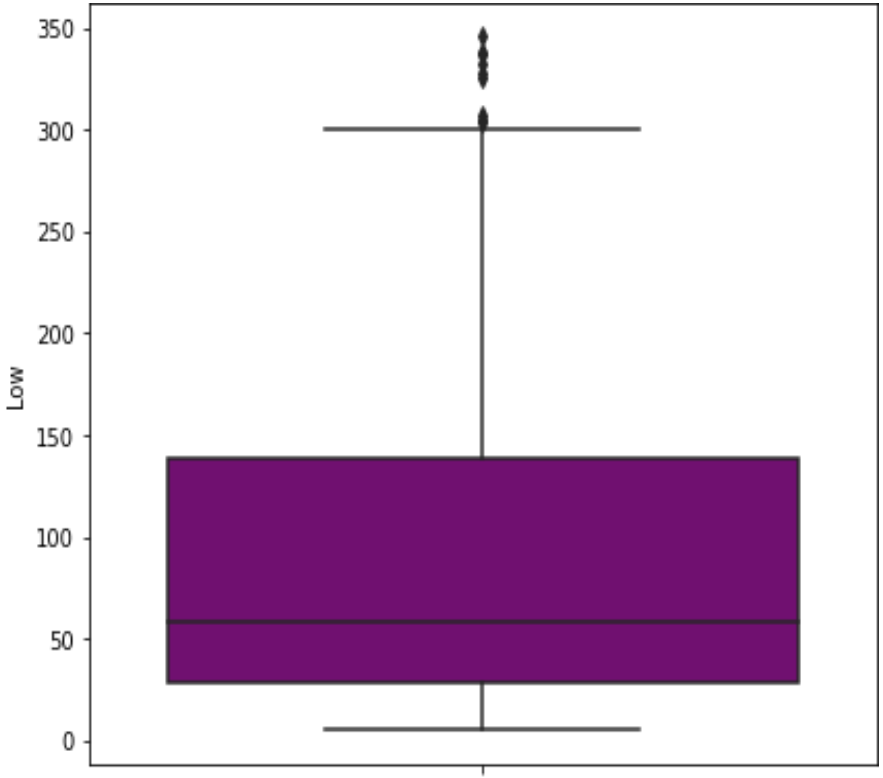
Distribution of ‘Open’



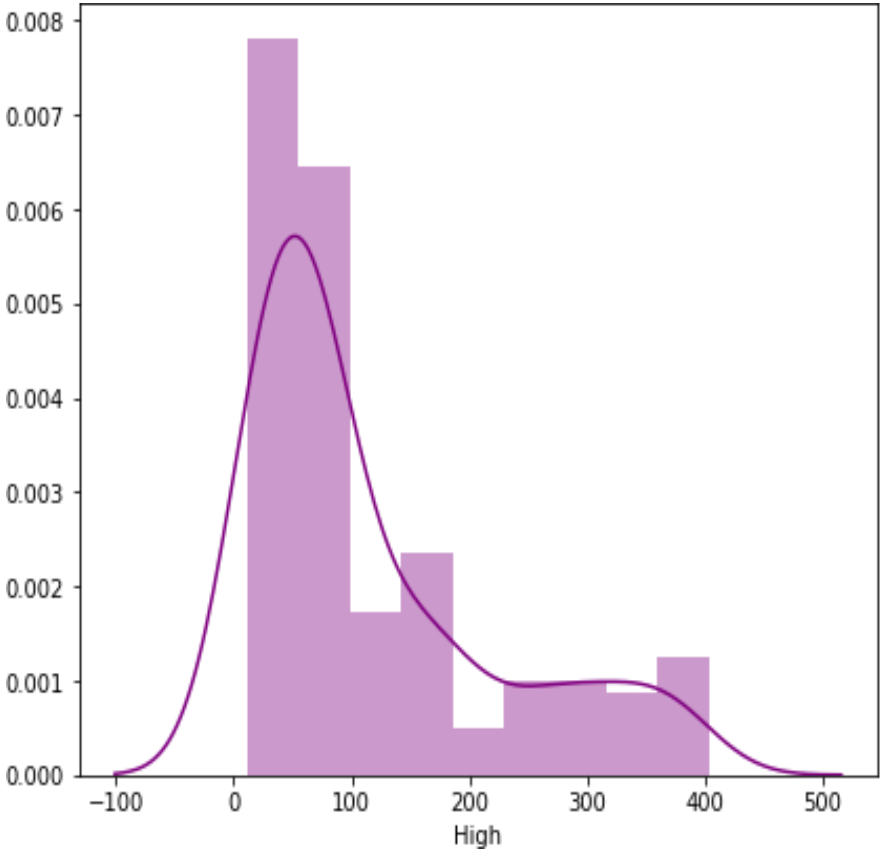


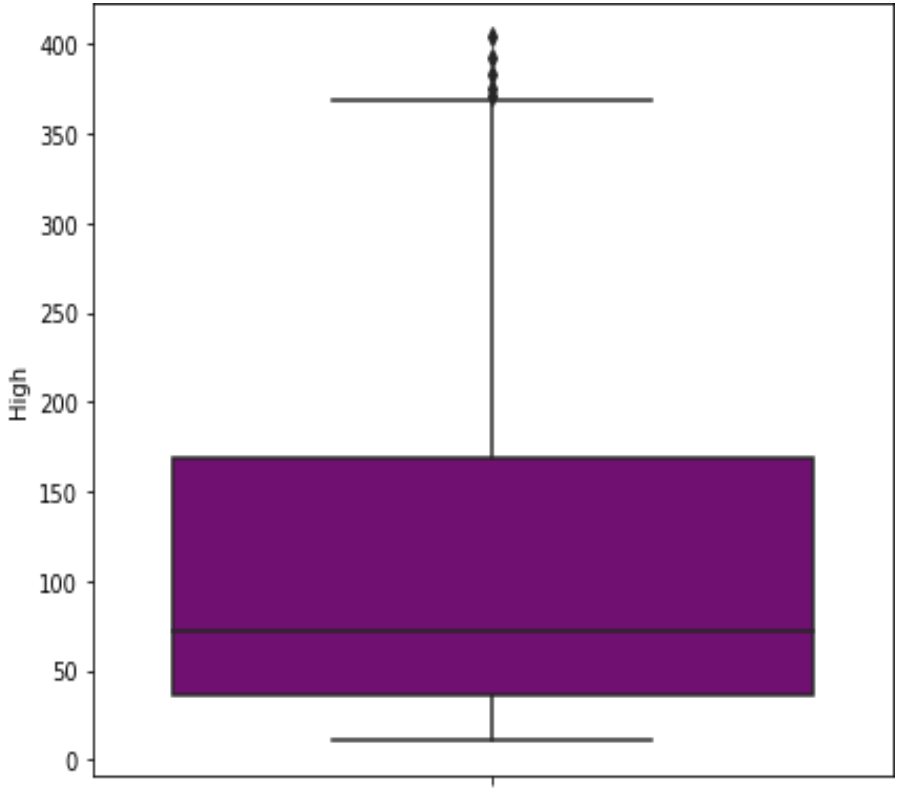
Distribution of ‘Low’





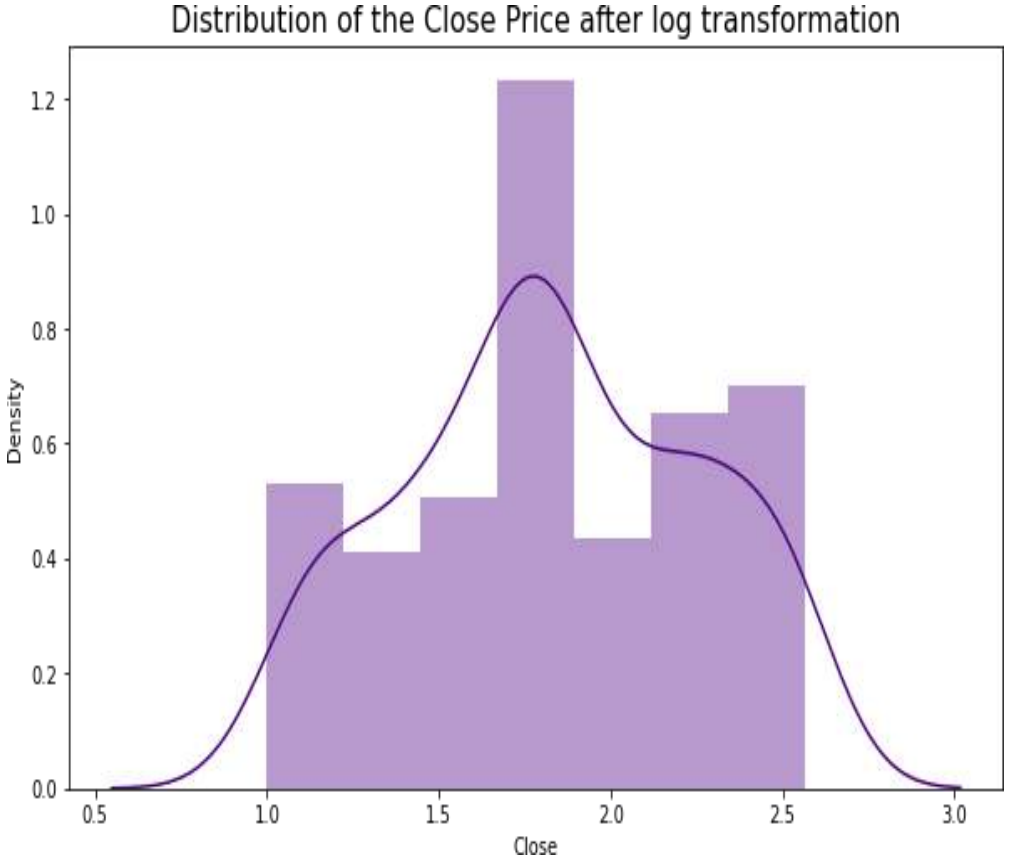
Distribution of ‘High’





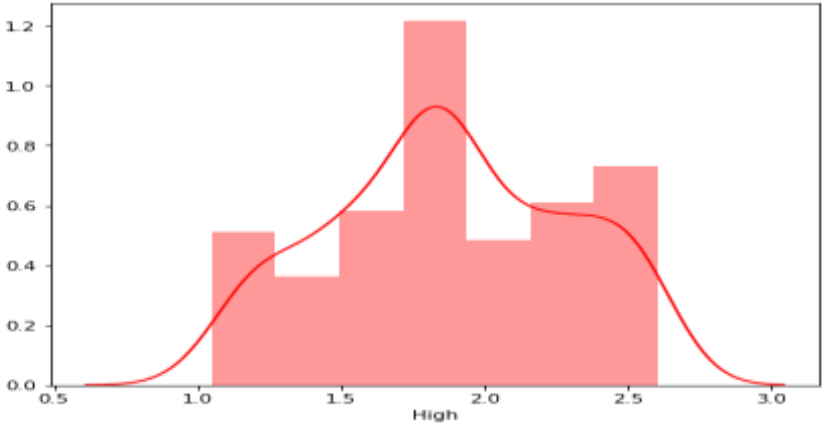
Distribution of ‘Close’

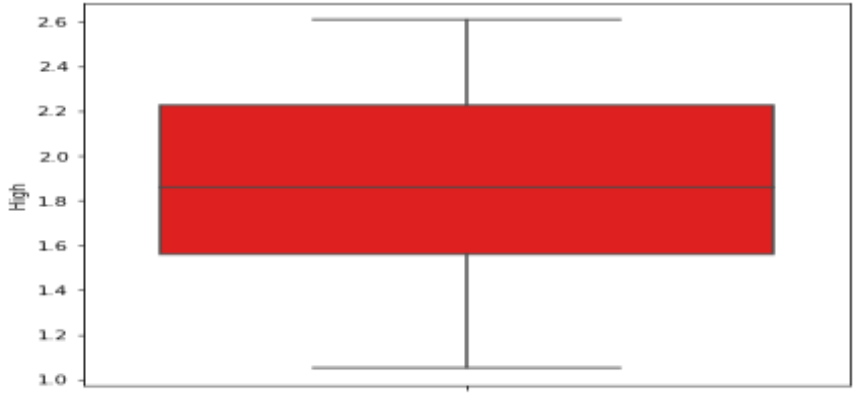
Transformation of Data

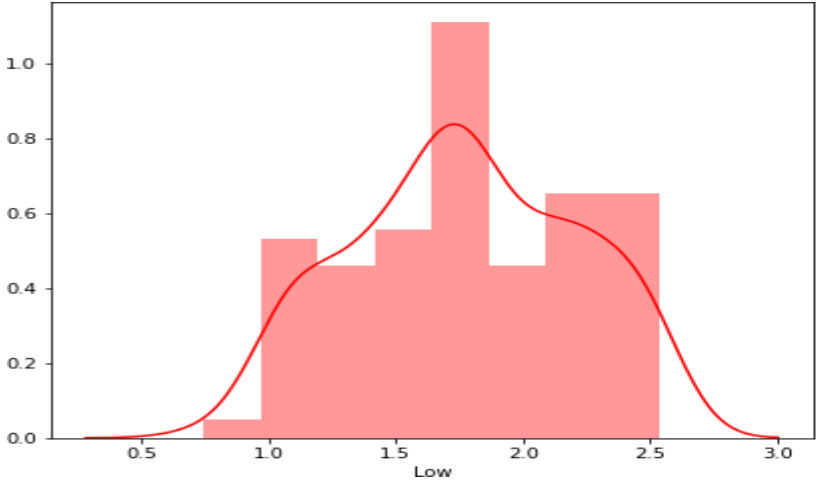


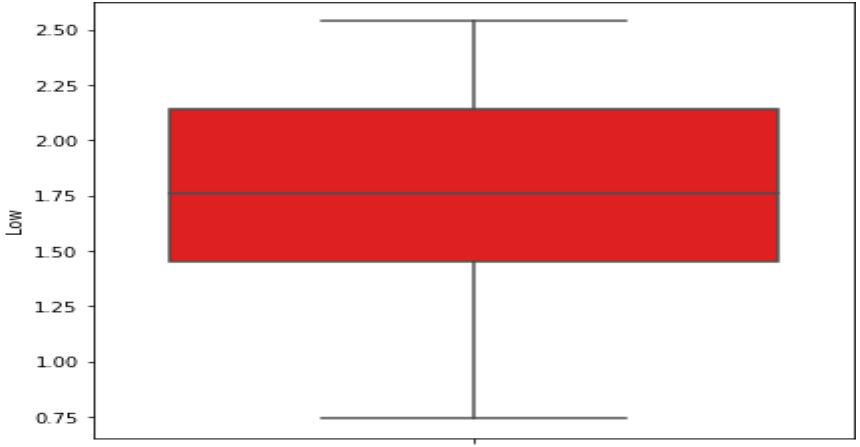
- From EDA We observed that data was found to be skewed. So we transformed the data to make it uniform before passing it into our machine learning models.
- We applied log transformation. The figure shows the distribution of close price after a log transformation.

Continue…

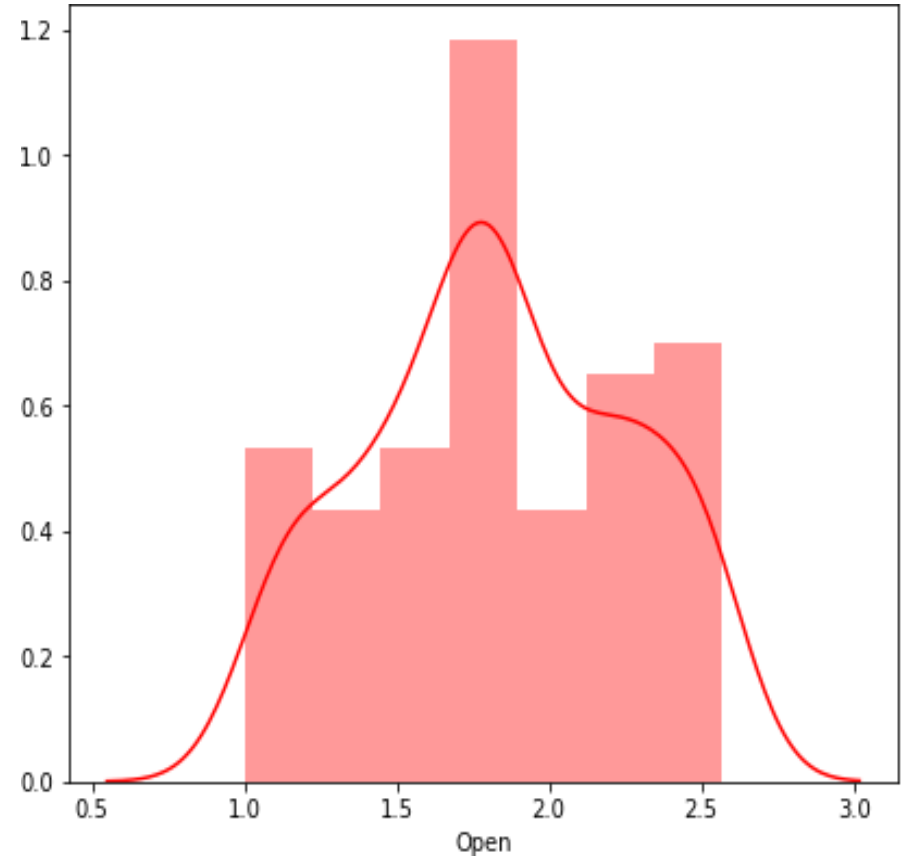

1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
High

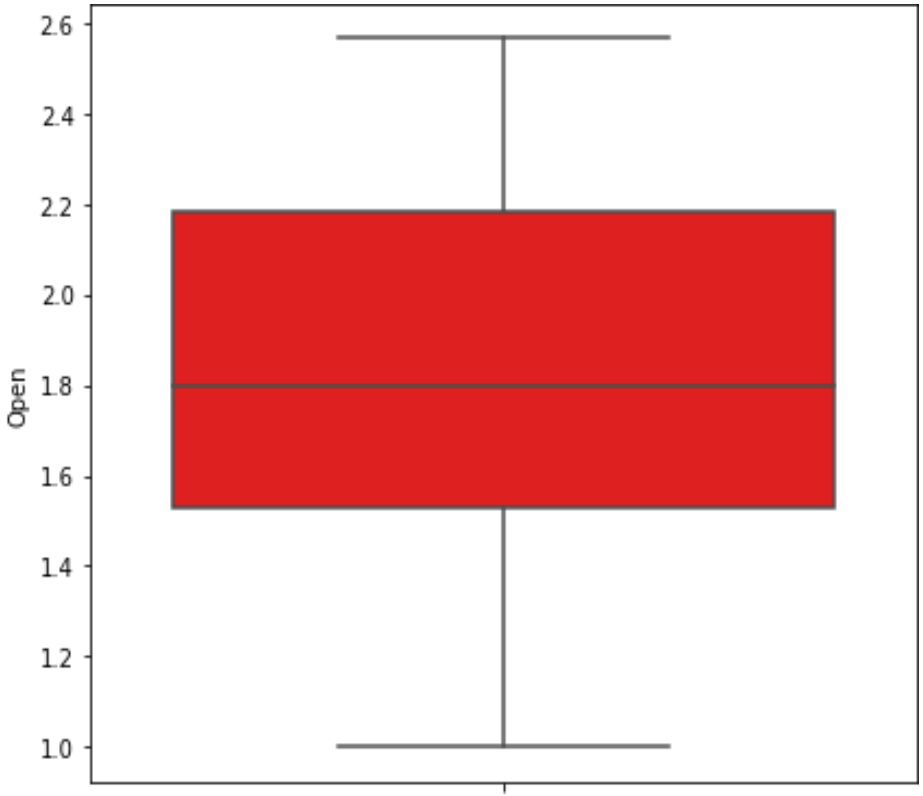

2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
High


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

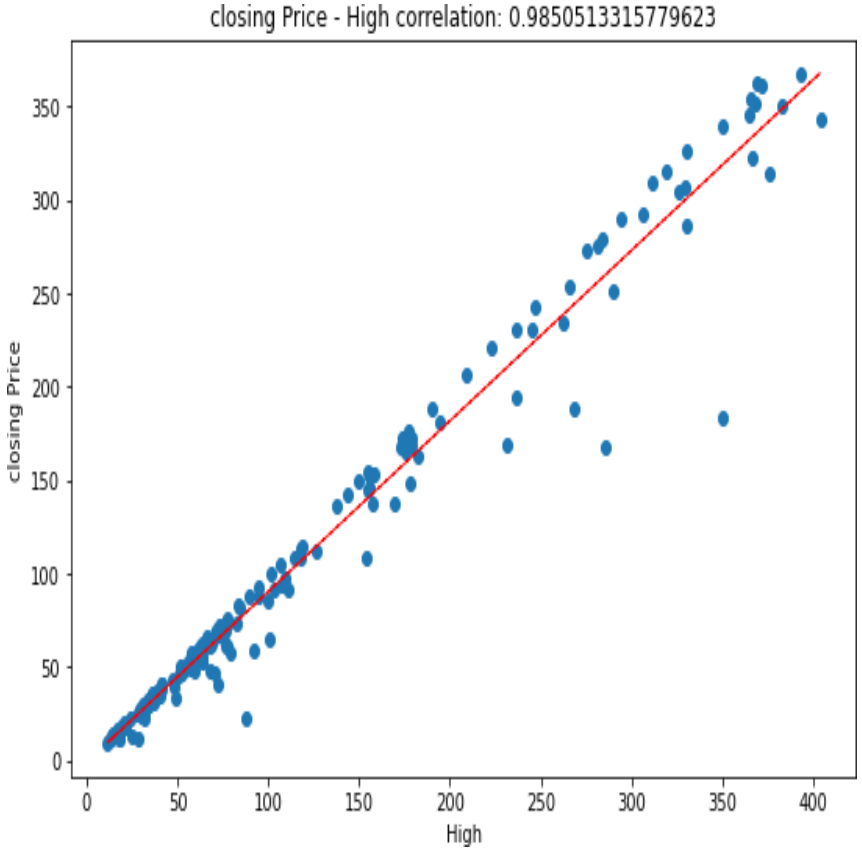

2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
Low

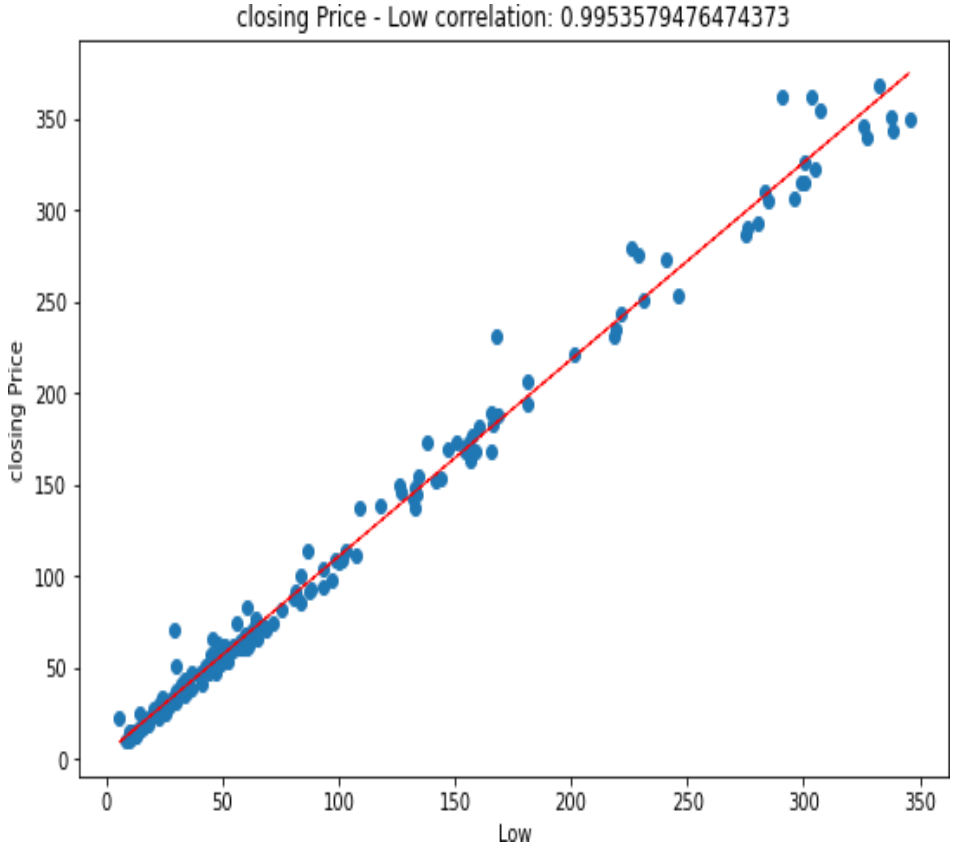
Continue...


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

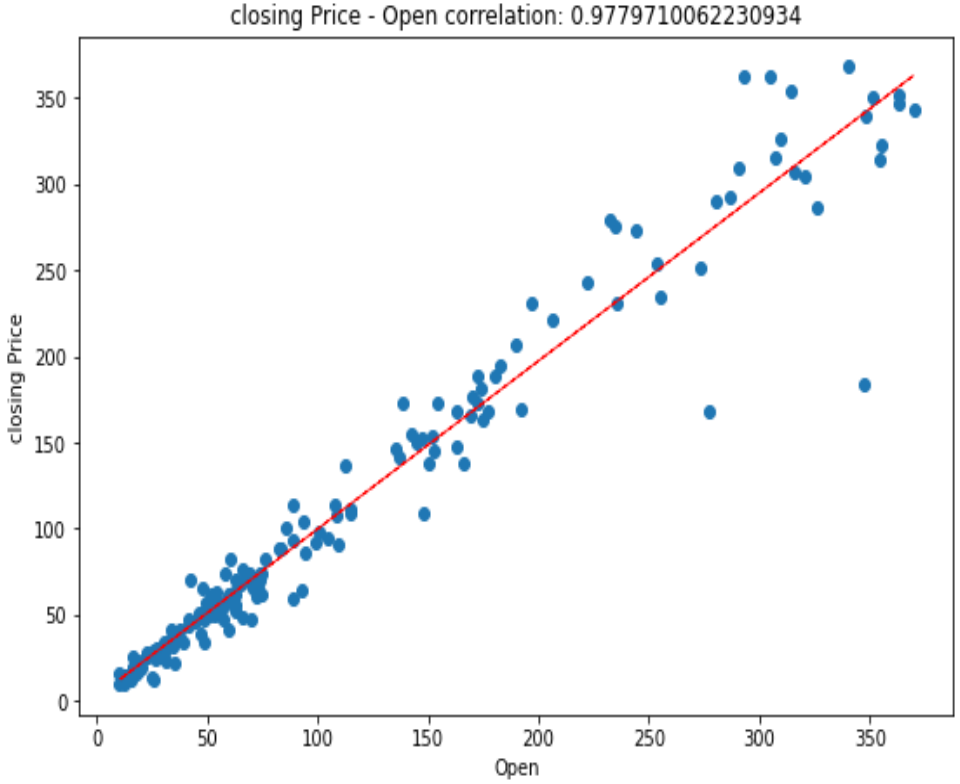

2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
Open

Correlation of ‘Closing Price’ with Independent Features





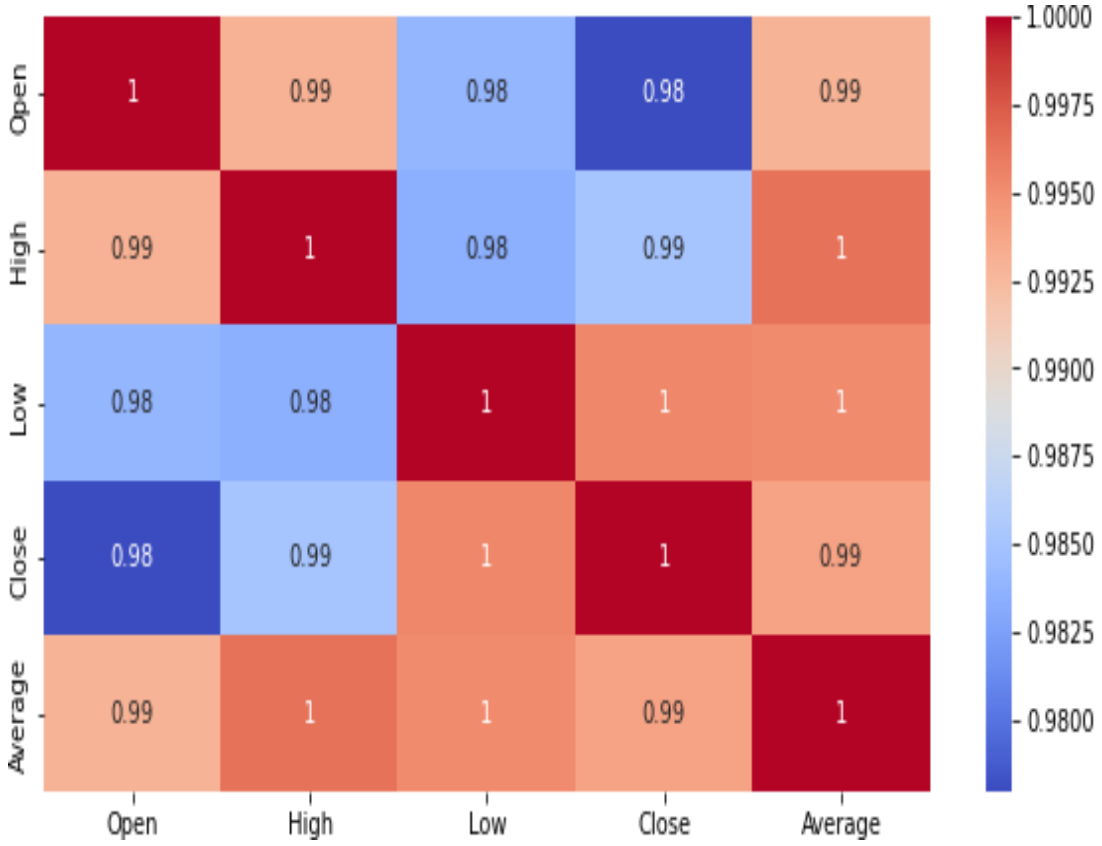
Continue…



- 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

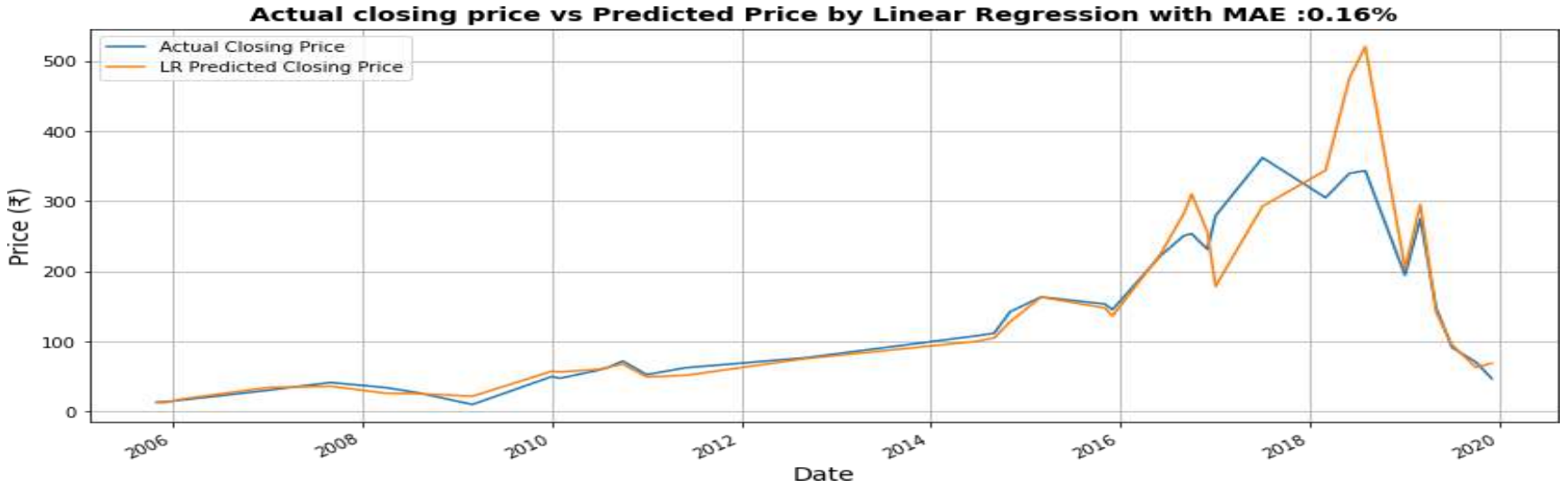
- The correlation matrix helps us visualize the correlation of each parameter with respect to every other parameter.
- The shades changes from the highest to lowest (or vice versa) correlations.
- We observed from matrix that dependent variable (close price) is highly correlated with all the other independent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



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
  - Root mean squared error.

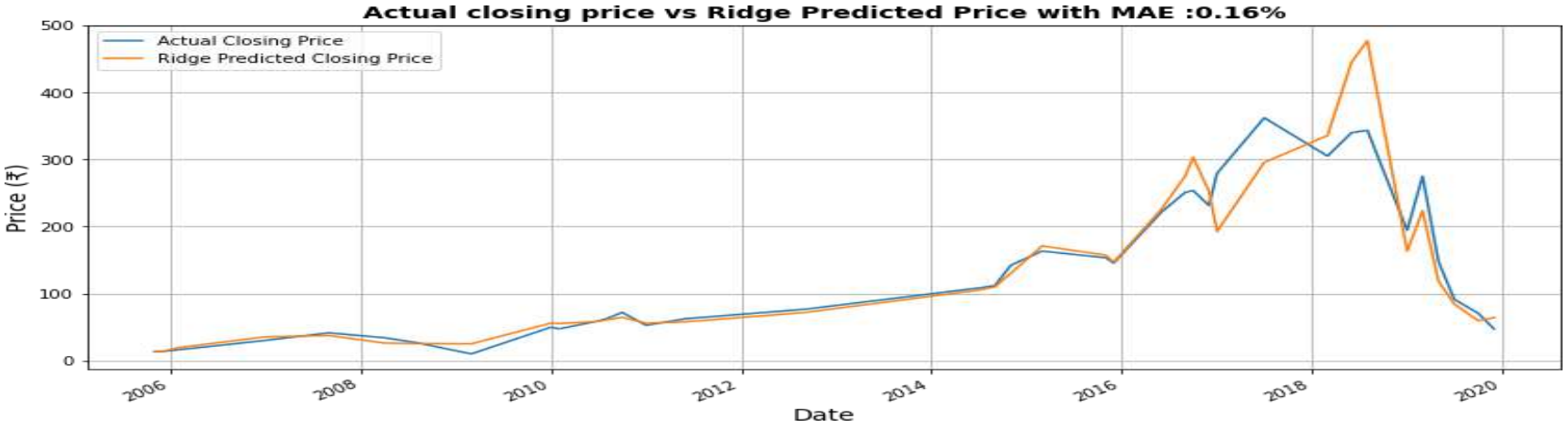
Linear Regression



Explanations:

- Linear Model predicted the close price with 0.16% Mean Absolute error. It has training accuracy 80.19%.
- R2 tells us that our independent variable is able to describe 83% of our dependent variable.
- Adjusted R2 is about 82.17%.

Ridge Regression



Explanations:

• Ridge predicted the close price with 0.16% Mean Absolute error. Having training accuracy 80.81%.

• R2 is about 82.99% which means model's independent features is able to describe 82.99% of our dependent variable.

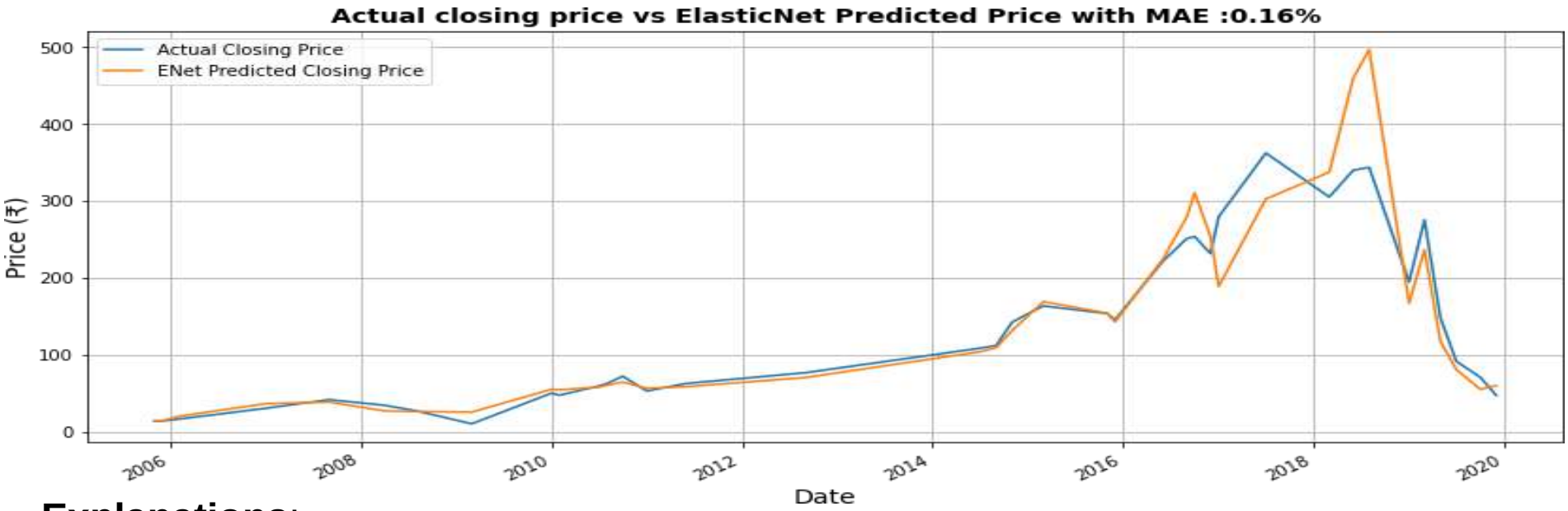
•Adjusted R2 is about 81.44%. We'll consider adjusted R2 because we have too many independent features

Lasso Regression

Explanations:

- ❑ Lasso predicted the close price with 0.17% Mean Absolute error. Having training accuracy 80.91%.
- ❑ Here, R2 is about 82.96% which means models' independent features is able to describe 82.96% of our dependent variable.
- ❑ Adjusted R2 is about 81.41%. 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 78.68%.
- ❑ R2 tells us that our independent is able to describe 83.04% of our dependent variable.
- ❑ Adjusted R2 is about 81.50%.

Comparison 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 ElasticNet.

All four models have good R2 and Adjusted R2.

Final Matrix:

	Linear Regression	Ridge	Lasso	Elastic Net
MSE	0.031981	0.031687	0.031737	0.031596
RMSE	0.178831	0.178010	0.178148	0.177753
R2	0.828322	0.829896	0.829630	0.830386
Adjusted_R2	0.812715	0.814432	0.814142	0.814967
Training Accuracy	0.809115	0.808140	0.809094	0.786767

Conclusion

- Target Variable is strongly dependent on Independent Variables.
- Linear Regression and Lasso are performing better than other models with training accuracy 80.91% and 80.90% respectively.
- Apart from Linear Regression and Lasso, Ridge and Elastic Net is also performing better but they have less training accuracy.
- Ridge and ElasticNet is performing far much better after Applying Hyperparameter Tuning and Cross-validation, it is because we have small set of datasets.
- R2 and AdjustedR2 is around 82% and 81% in each model.