

Capstone Project - 2

YES Bank Stock Closing Price Prediction

Nitesh Gajakosh

Content

- Introduction
- Problem definition
- Data Summary
- EDA on given data
- Statistical Data Analysis
- Model implementation
- Model validation and selection
- Conclusion

Introduction

- Yes Bank Limited is an Indian private sector bank headquartered in Mumbai, India and was founded by Rana Kapoor and Ashok Kapur in 2004.
- It offers wide range of differentiated products for corporate and retail customers through retail banking and asset management services.
- On 5 March 2020, in an attempt to avoid the collapse of the bank, which had an excessive amount of bad loans, the Reserve Bank of India (RBI) took control of it.

- We have been Provided with a dataset of the monthly Stock Price of Yes Bank .
- The data has been provided from July 2005 till November 2020.
- The bank has been making headline due to recent default.
- We analysed the dataset and worked on predicting the stock closing price for the bank using other given parameter.

Problem definition

- 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.

Data Summary

Dataset contains Yes Bank stock monthly prices since it listed on stock exchange.

Following are its features:-

Date: In our data its monthly observation of stock since it listed.

Open: The price a stock when the stock exchange open for the day.

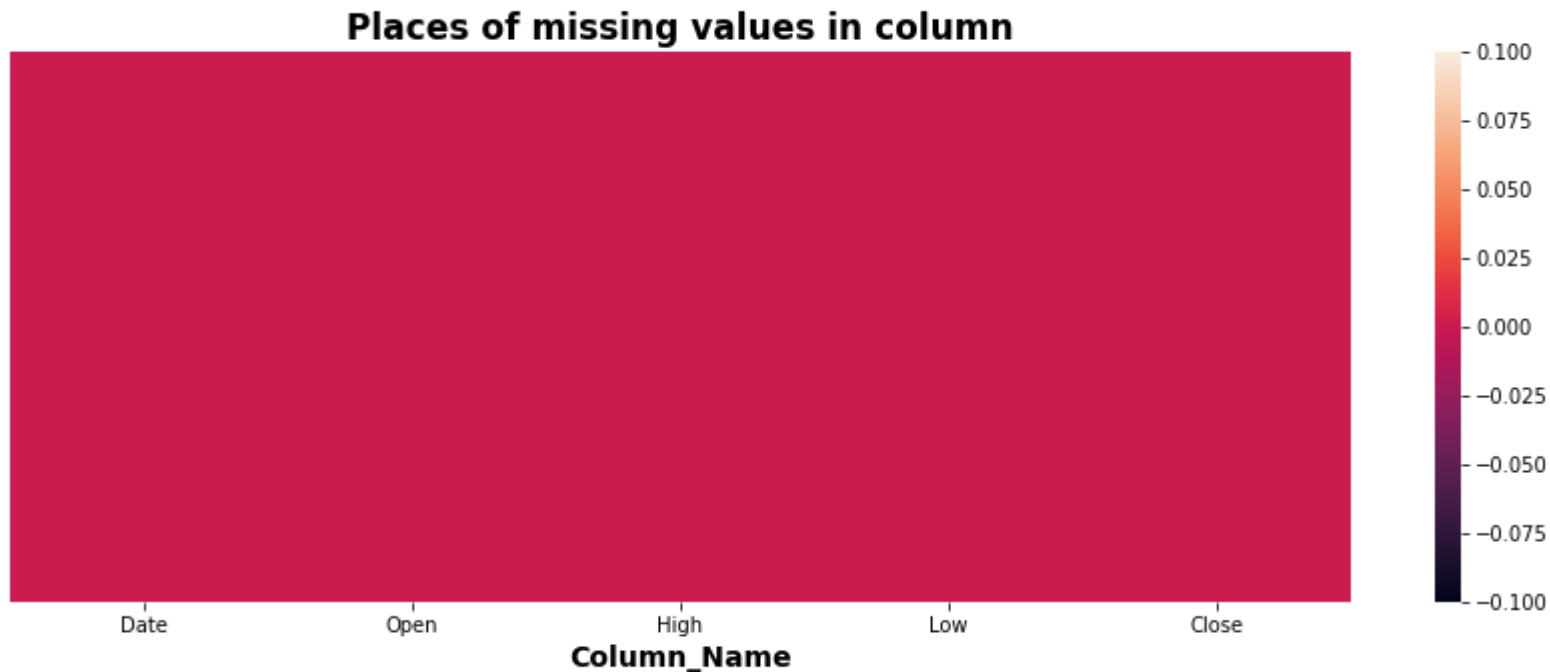
High: The maximum price of a stock attain during given period of time.

Low: The minimum price of a stock attain during given period of time.

Close: The price of a stock when the stock exchange closed

EDA

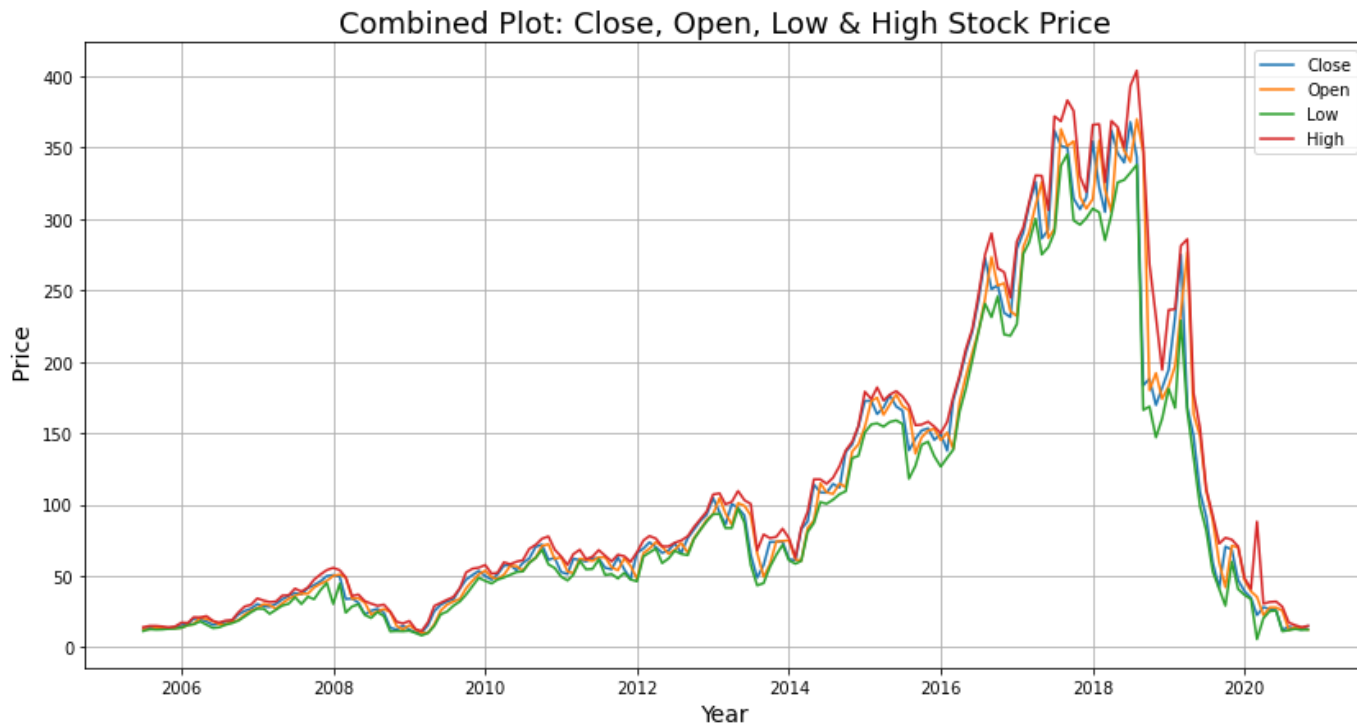
- We can see there is no null values in the given data set.



After 2018 yes bank stock price fall beacuse of fraud case regarding to Rana Kapoor.

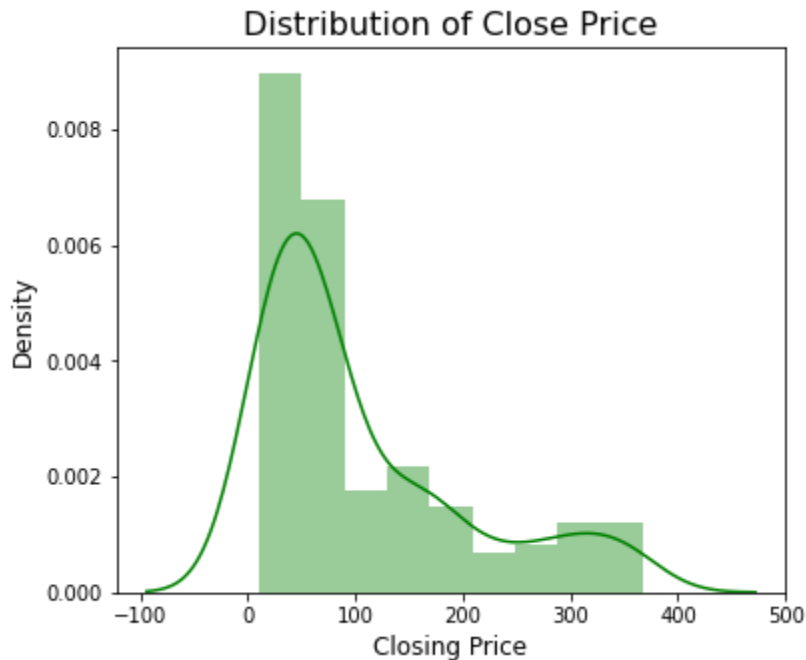


we can see in 2017 to 2019 there can be high action seen because of difference in high and low lines.



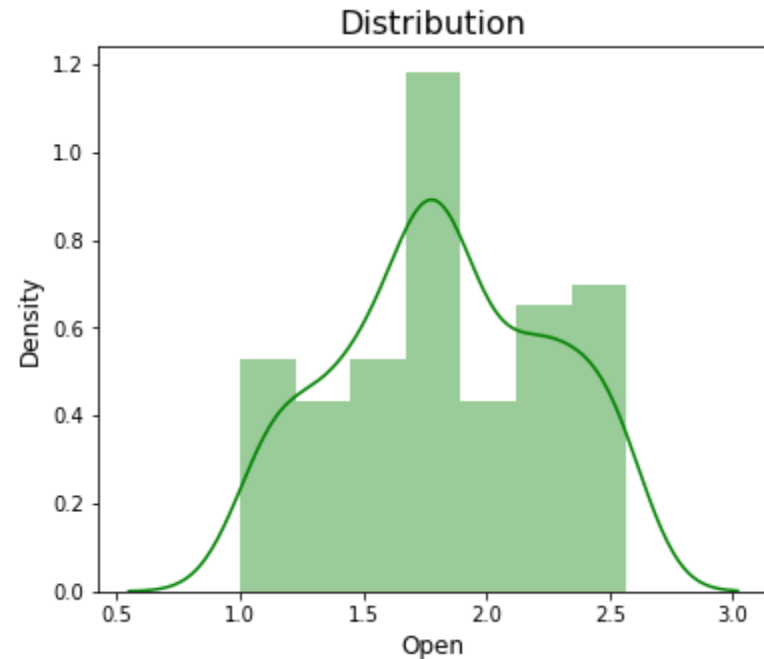
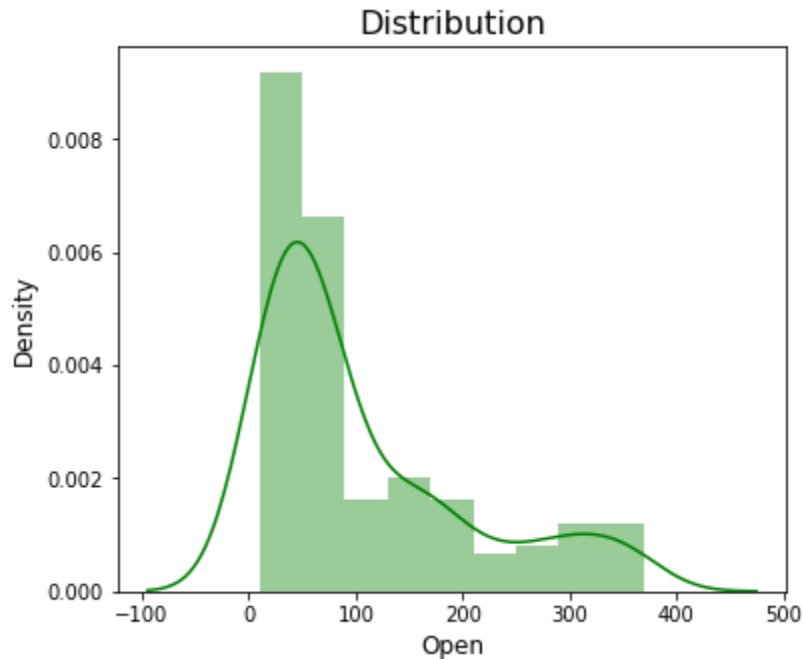
EDA

Distribution of Close Price is Positively Skewed need to go with log transformation to make it normal distribution.

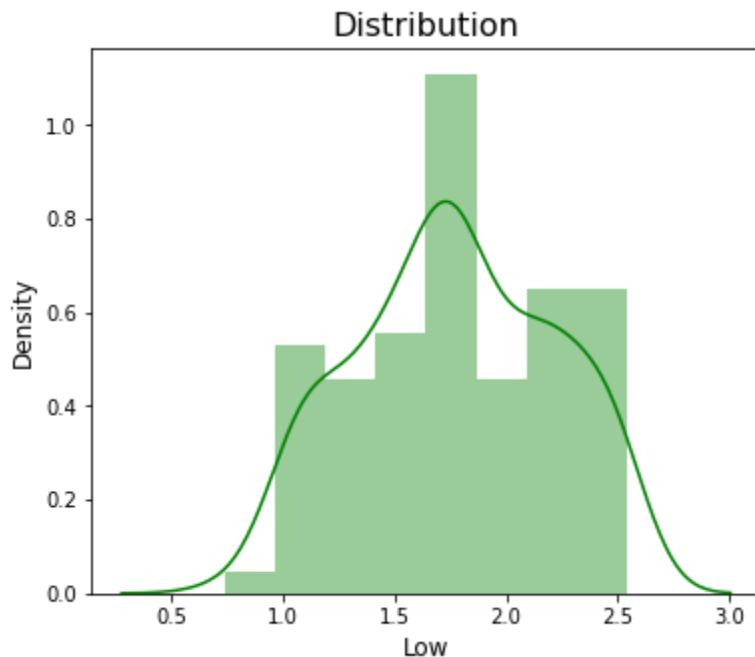
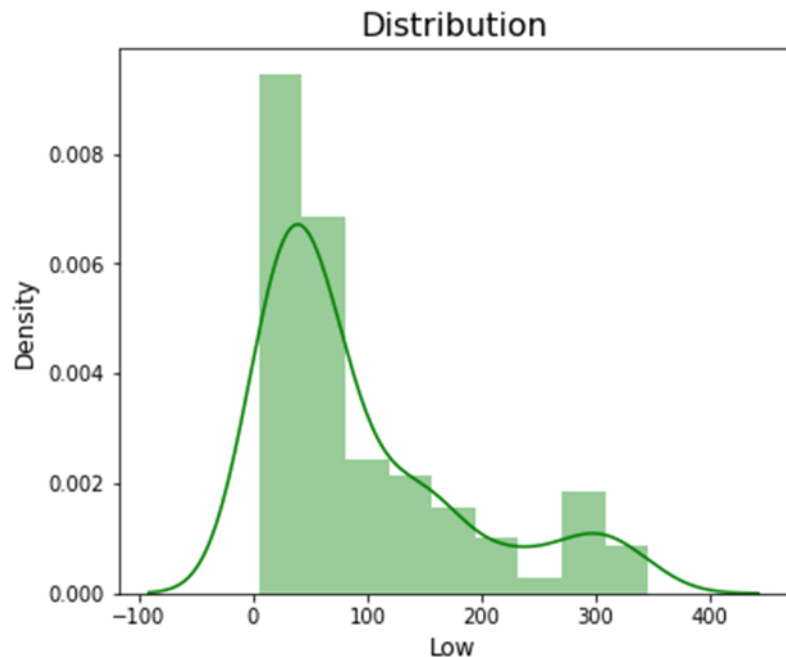


EDA

Distribution of Open Price is Positively Skewed need to go with log transformation to make it normal distribution.

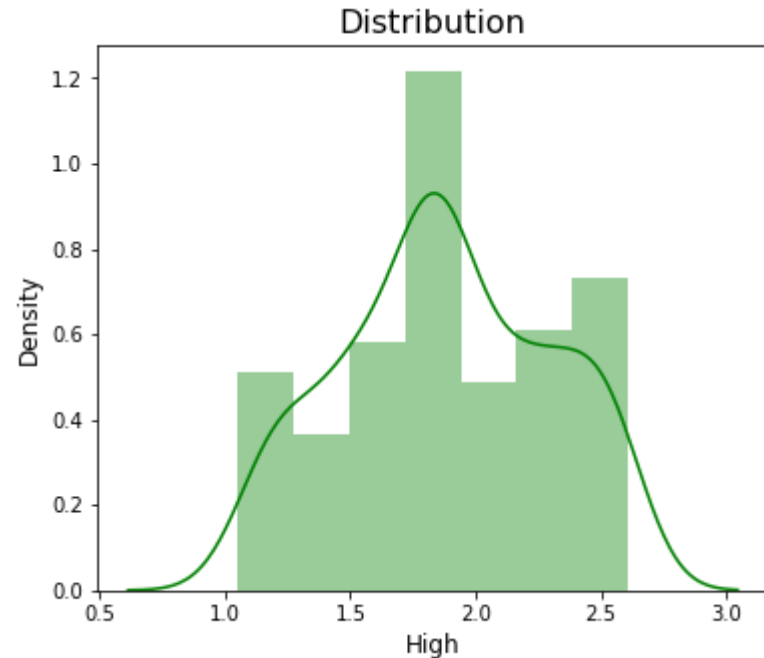
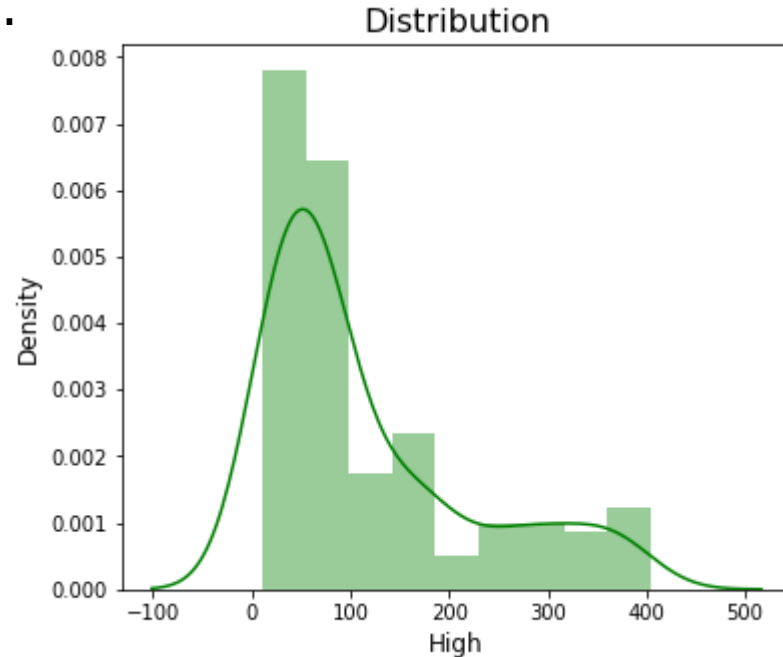


Distribution of Low Price is Positively Skewed need to go with log transformation to make it normal distribution.

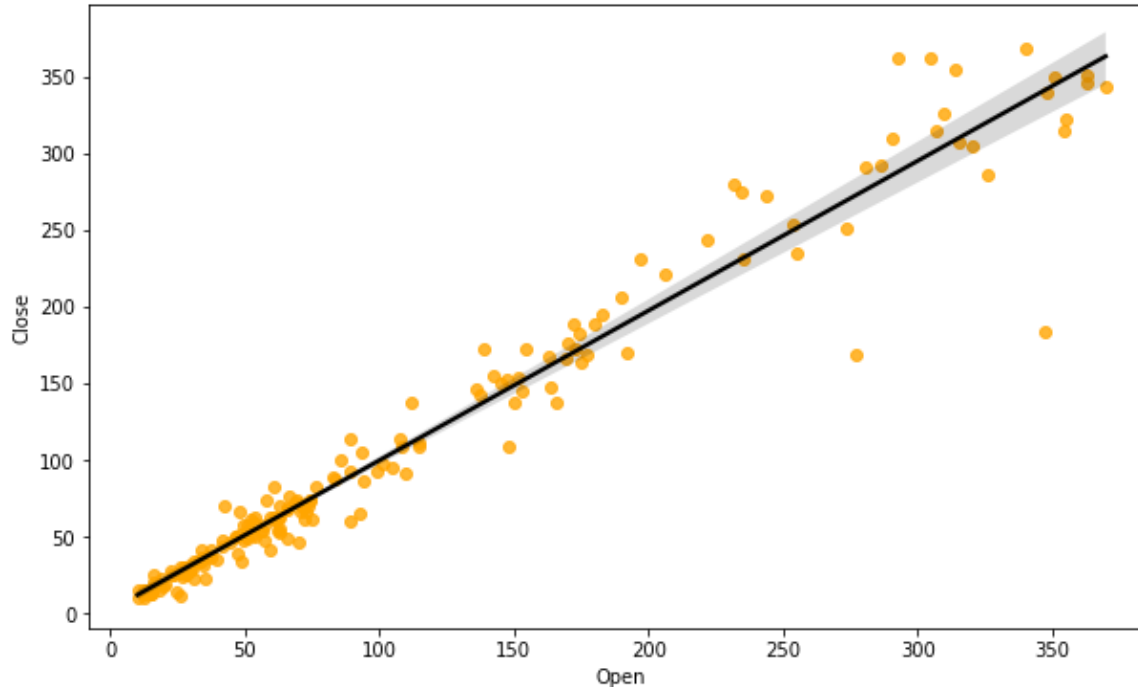


EDA

Distribution of High Price is Positively Skewed need to go with log transformation to make it normal distribution.

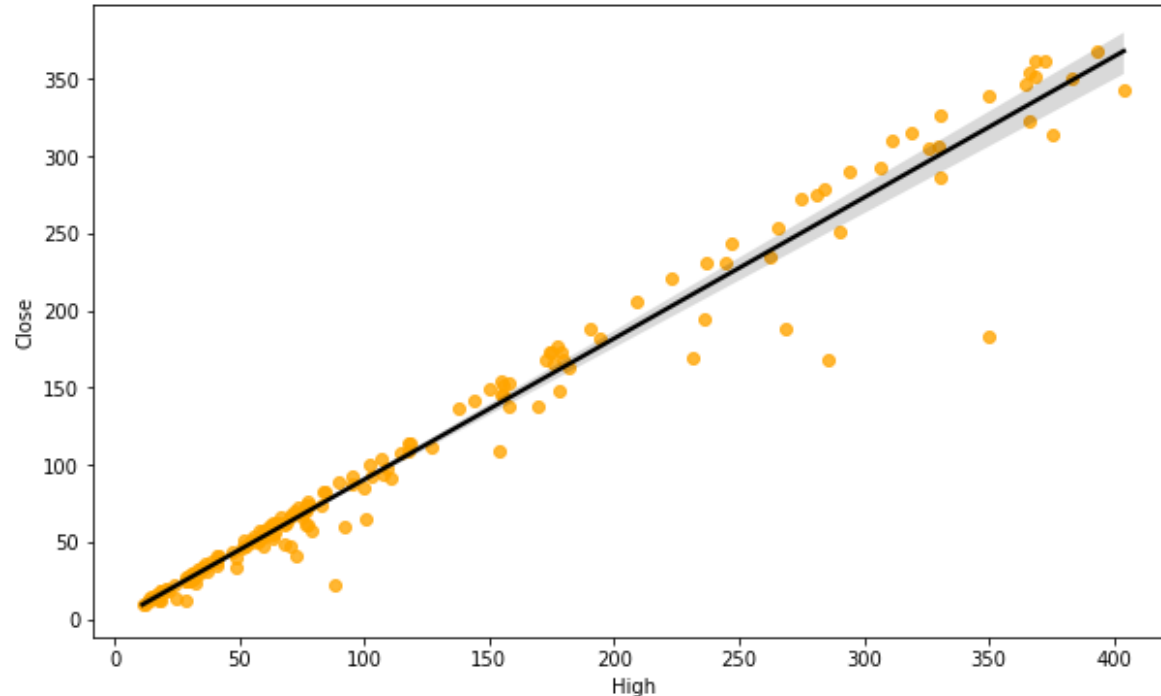


From Regression Plot Open Price are Positively Relation to target Variable Close Price.



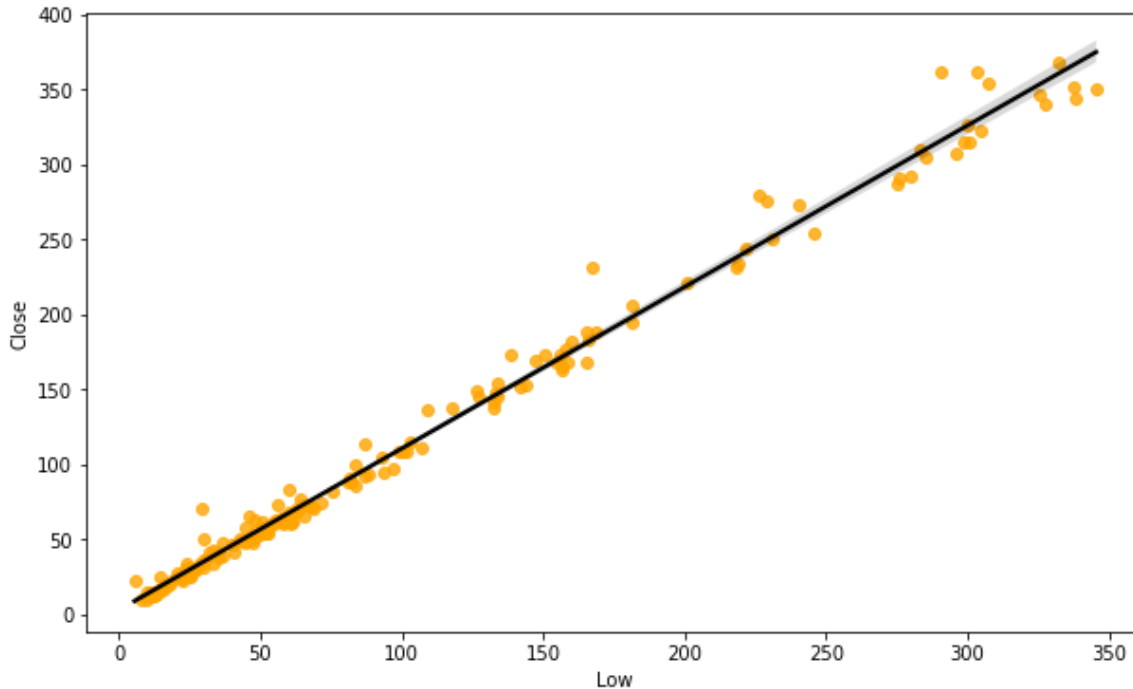
EDA

From Regression Plot High Price are Positively Relation to target Variable Close Price.



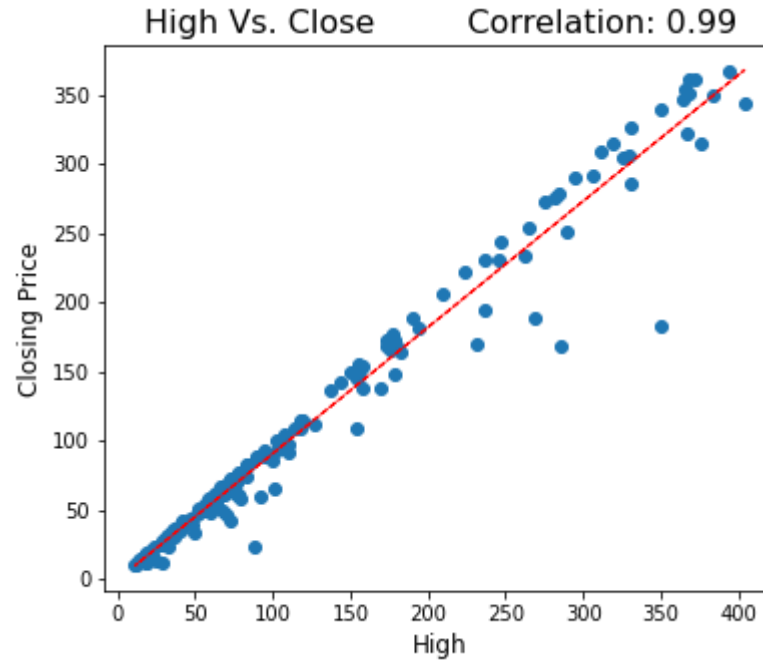
EDA

From Regression Plot Low Price are Positively Relation to target Variable Close Price.



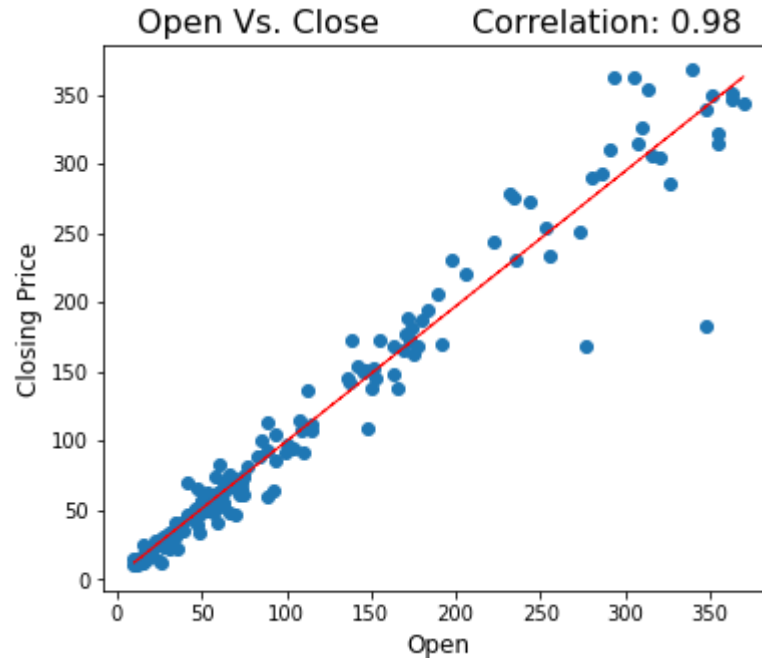
Statistical Data Analysis

High Price and Closing Price are positively co related .



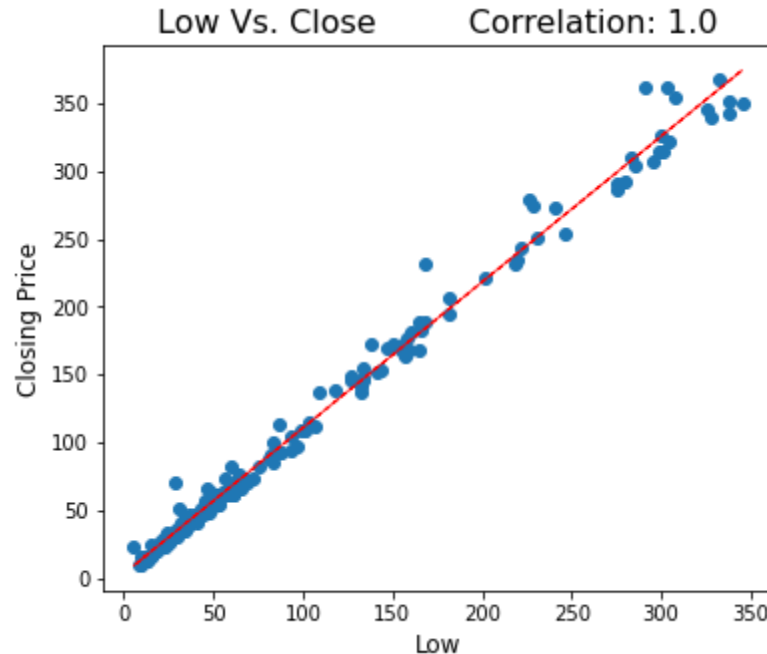
Statistical Data Analysis

Open Price and Closing Price are positively co related .



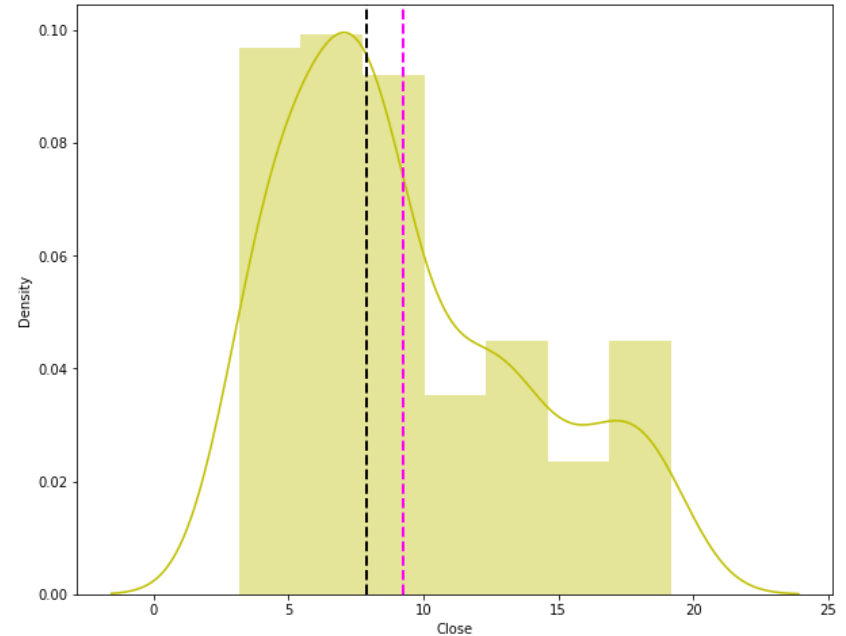
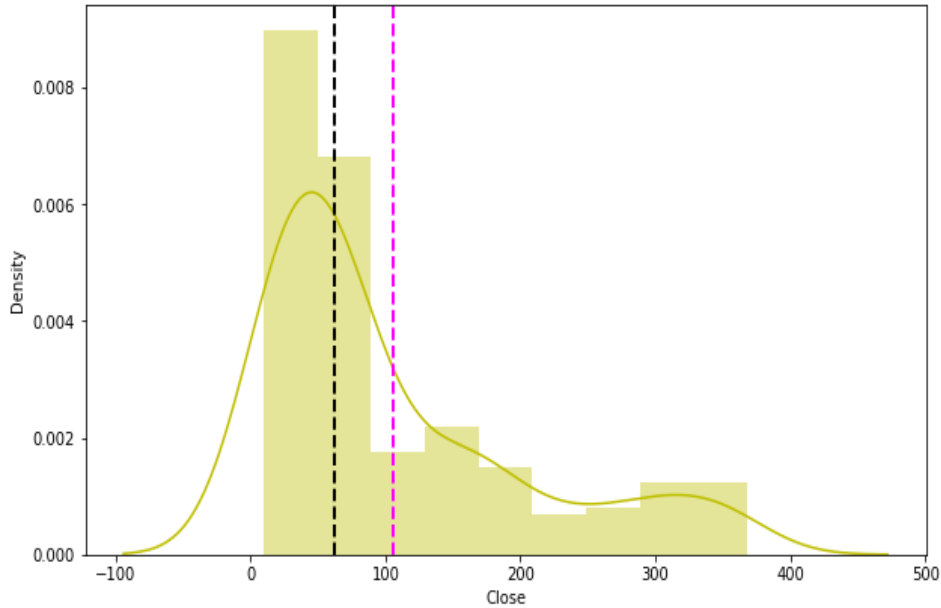
Statistical Data Analysis

Low Price and Closing Price are positively co related .

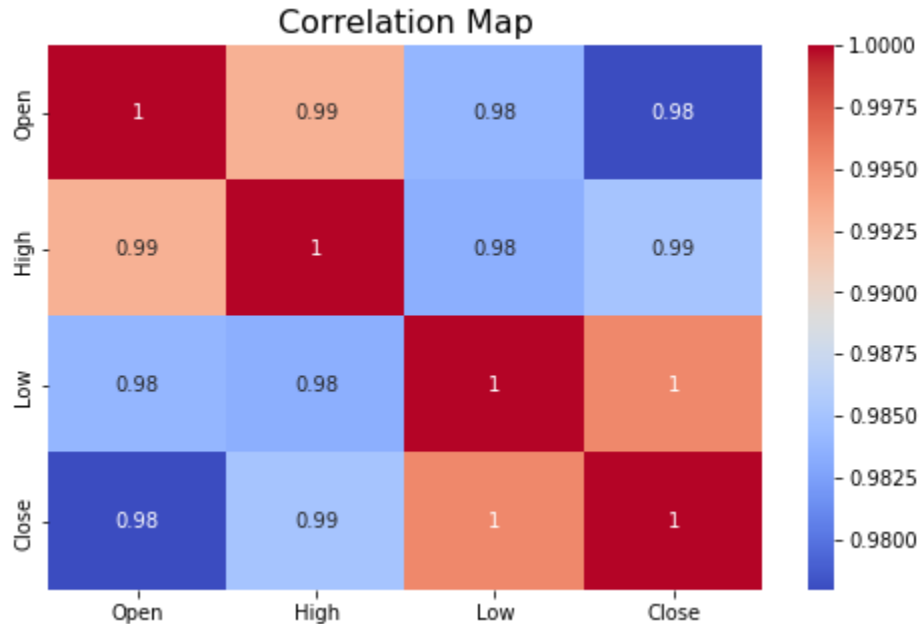


Features Tranformation

Closing Price of our features are Positively Skewed so we can apply sqrt,log10 or reciprocal tranformation on them to make it normal Distribution.



Correlation Map with Heatmap



Very high correlation between 'High Price' and 'Close Price'.

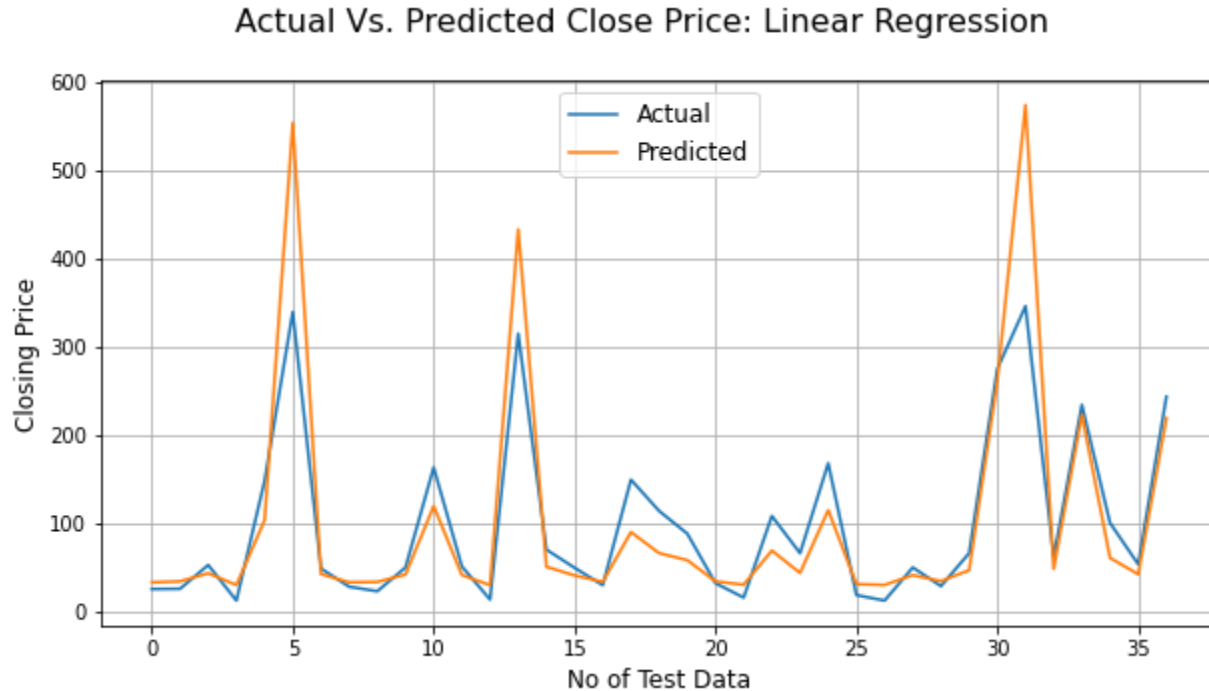
Applying ML Alorithms

Since we have to predict Closing Price . Hence we have to use regression algorithms.

Algorithms that we will use are:

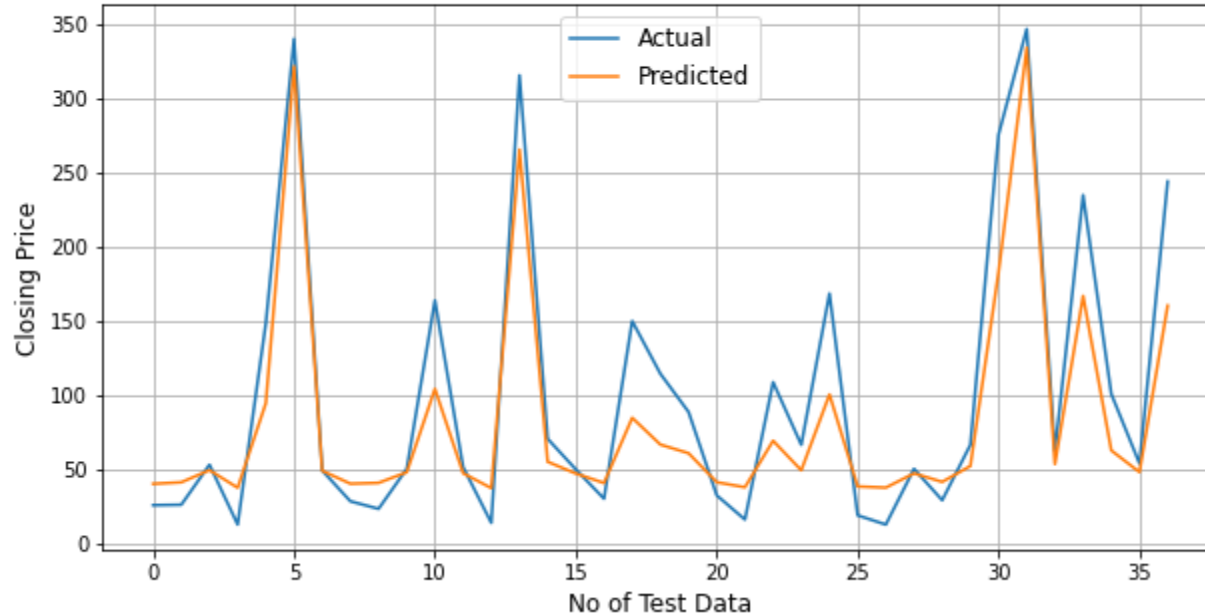
- Linear Regression
- Lasso Regression
- Ridge Regression
- Elastic Net Regression

Model implementation



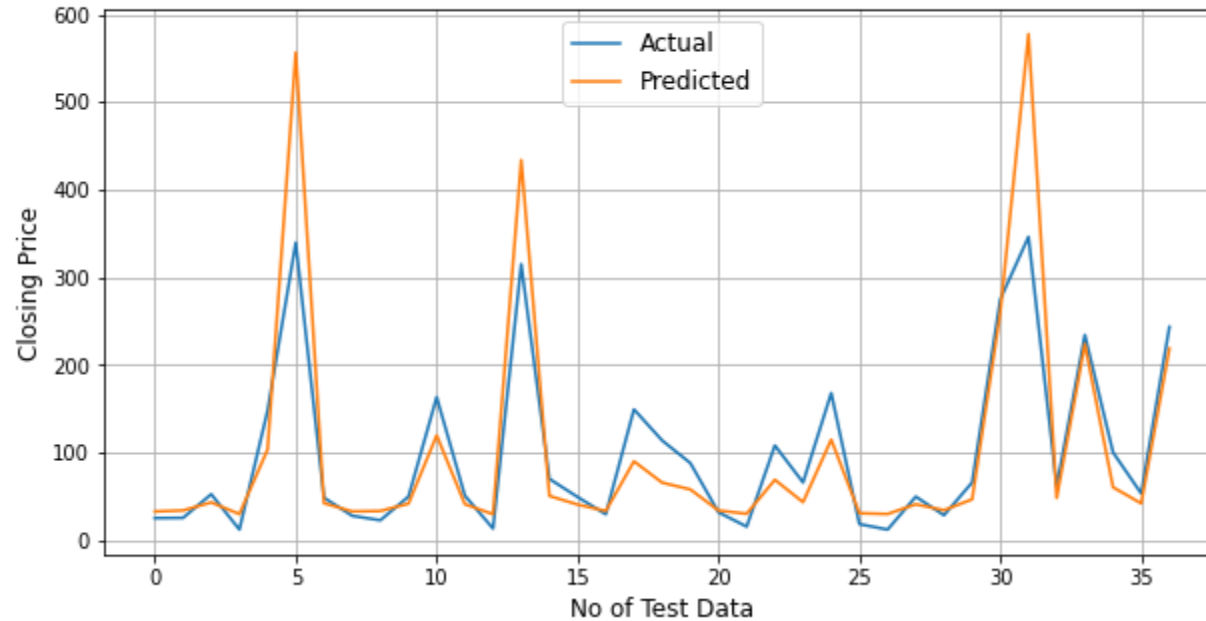
There is much value difference seen spikes of Predicted price and Actual Price.

Actual Vs. Predicted Close Price: Lasso Regression



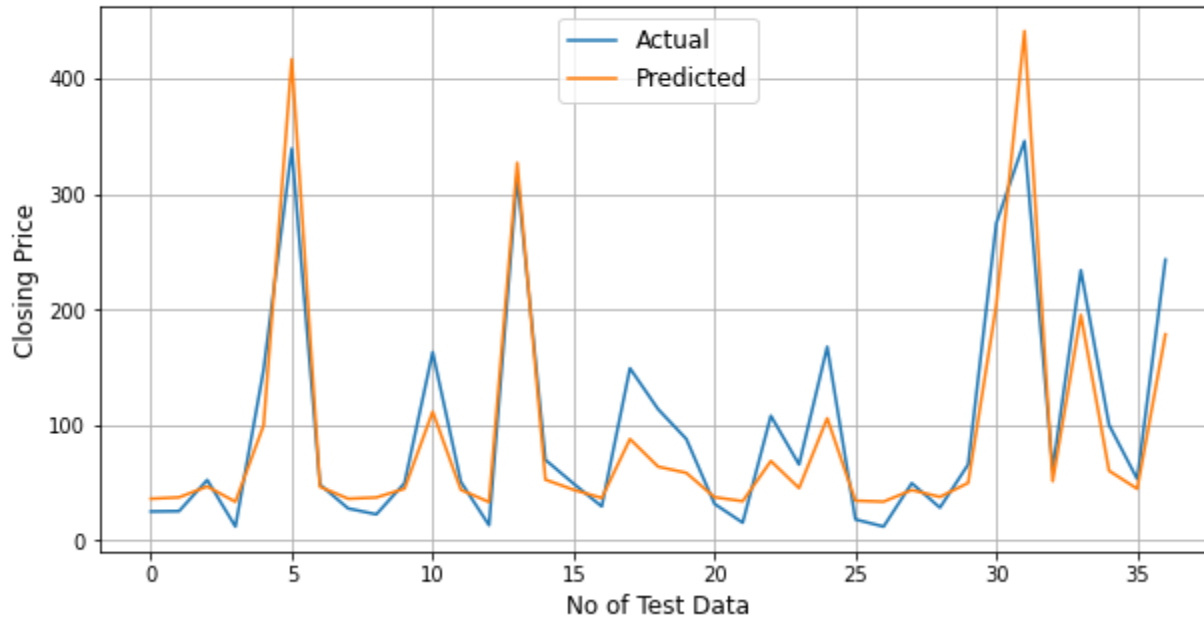
As we see spike get accurately predicted in lasso but flat prices predication varies much.

Actual Vs. Predicted Close Price: Ridge Regression



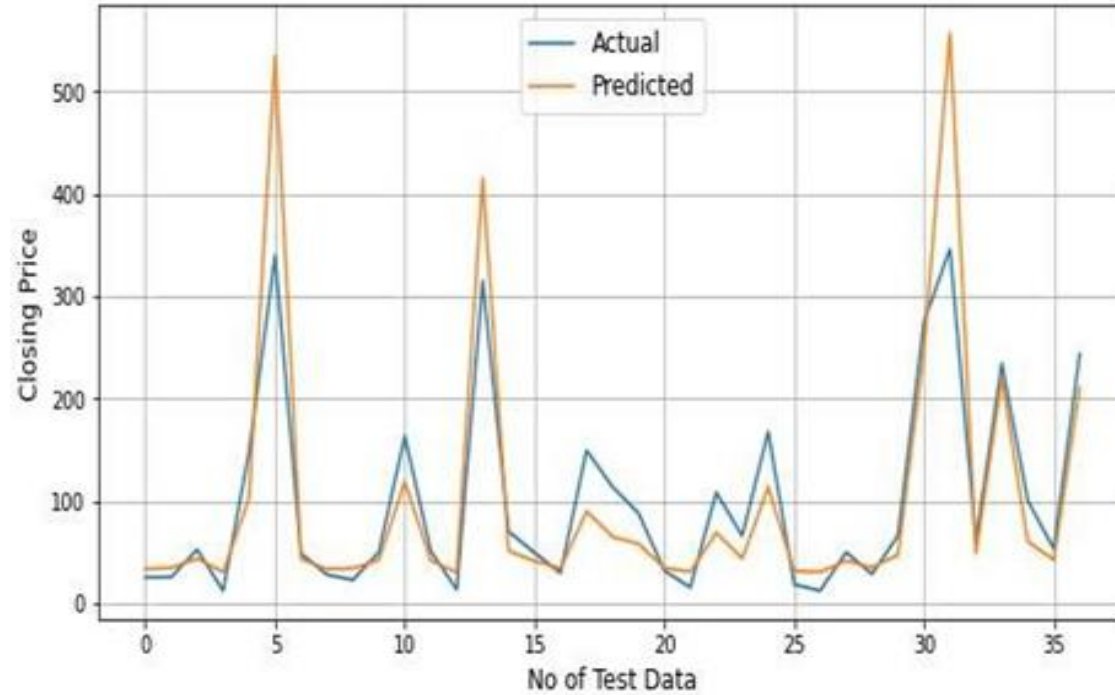
There is much value difference seen spikes of Predicted price and Actual Price.

Actual Vs. Predicted Close Price: Elastic Net



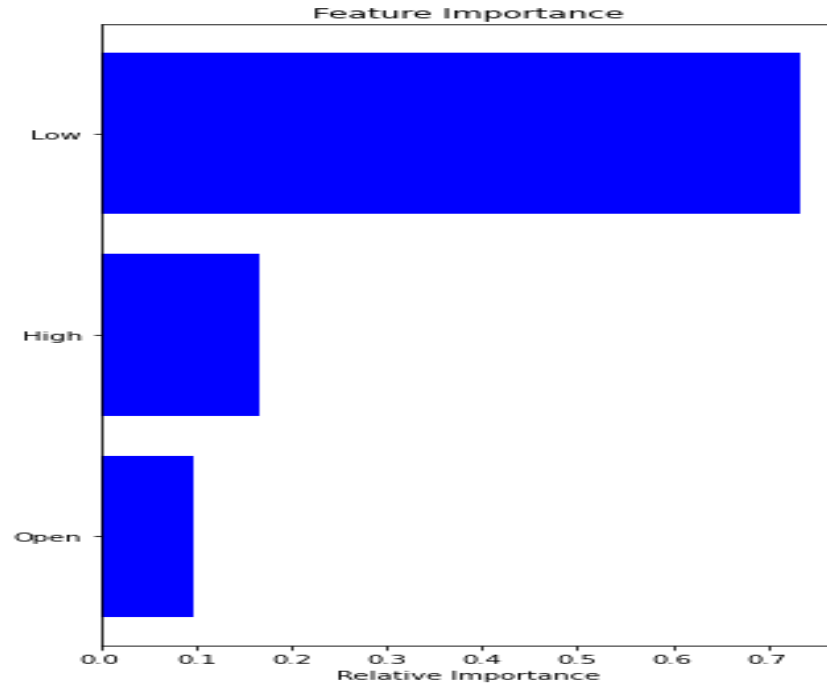
As we see spike get accurately predicted in Elastic Net but flat prices predication varies much.

Actual Vs. Predicted Close Price: Elastic Net After CV



As we see spike get accurately predicted after Hyperparameter Tuning flat prices predication varies.

Random Forest Regression



From the chart Low price are highly Correlated with Close Price
Compare to other parameter.

Model implementation

- Various types of linear model are implemented on data such as Linear regression, Ridge regression, Lasso regression, Elastic net regression.

Model	MSE (Train)	RMSE(Train)	R2(Train)	MSE (Test)	RMSE(Test)	R2(Test)
Linear Regression	0.034	0.183	0.815	0.032	0.178	0.823
Ridge	0.034	0.183	0.815	0.032	0.178	0.823
Lasso	0.044	0.209	0.760	0.044	0.209	0.755
Elastic Net	0.037	0.193	0.796	0.036	0.191	0.796
Elastic Net gridSearchcv	0.034	0.184	0.814	0.032	0.180	0.819

Table 1Result table of linear models.

Model validation and selection

- Best results on model is obtained from Random forest regressor and Elastic Net regressor.
- After implementation of Grid search CV on both models final accuracy (r2) for Elastic Net was more.

Model	MSE (Train)	RMSE(Train)	R2(Train)	MSE (Test)	RMSE(Test)	R2(Test)
Random forest regressor	0.000	0.019	0.998	0.001	0.038	0.992
Elastic Net gridsearchCV	0.034	0.184	0.814	0.032	0.180	0.819

Table 2 Result table after implementation of Grid search CV on Regression models.

Conclusion

- No overfitting is seen.
- Linear Regression and Ridge Regression are almost same R2 Score 82%.
- Random forest Regressor and ElasticNet gridsearchcv gives the highest R2 score of 99% and 81% respectively for Train Set and 82% for Test set.
- Loss Regression Show Less R2 Score.
- Target variable(dependent variable) strongly dependent on independent variables

Thank You...