# Yes Bank Stock Closing Price Prediction

# ASHUTOSH MAURYA

**Data science**

**Alma Better-Cohort AZADI**

**Introduction**

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 to the past methods.

**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.

activities.

**Data Summary**

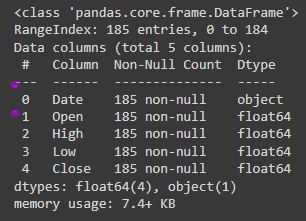
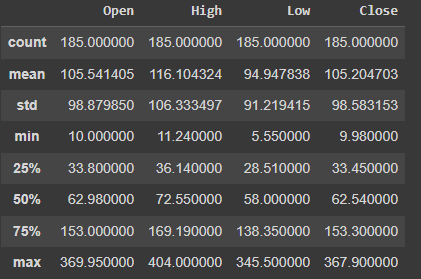
The dataset of YES BANK has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month of around 180 observations. It contains the following features.

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 180 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 an individual stock when the stock exchange closed for the day.

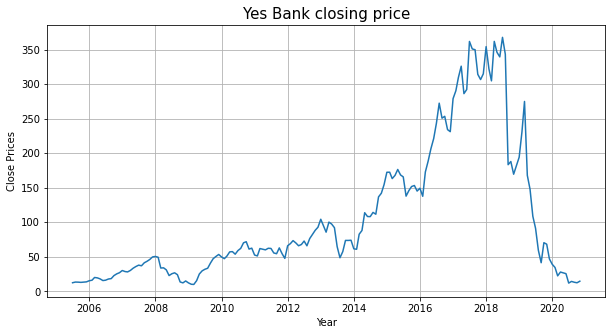
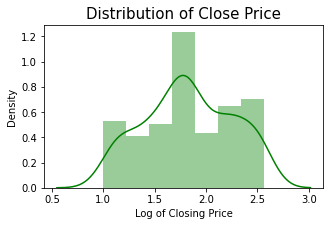
**Data Understanding**

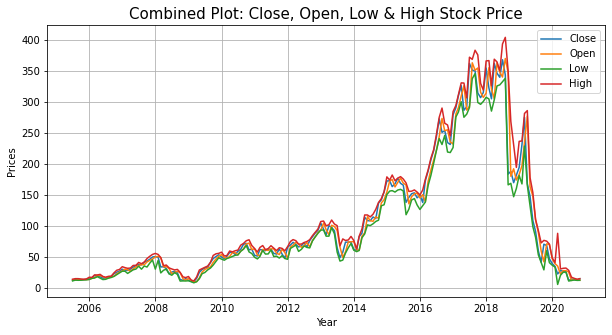
After mounting data, start preprocessing of dataset.

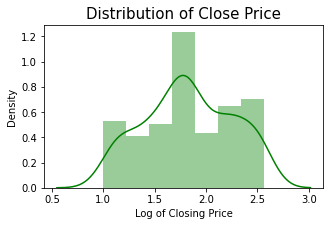
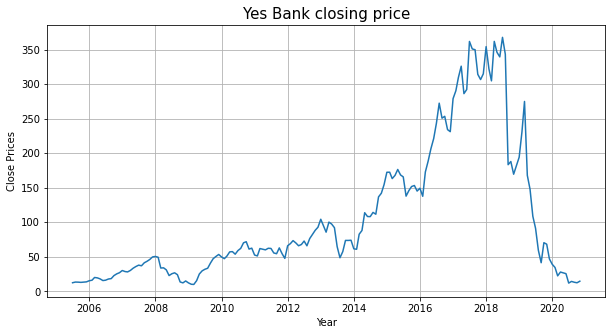
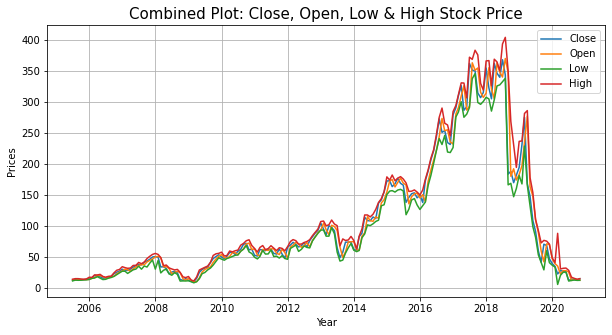
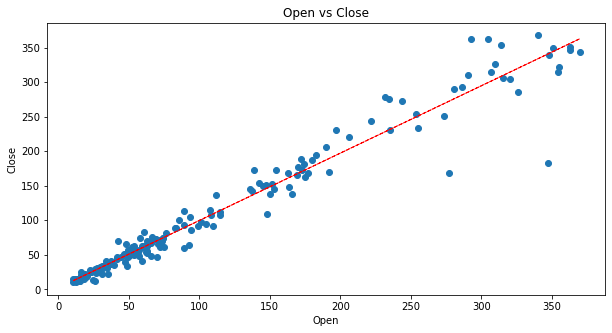
* Dataset has Open, High,Low,Close price of the stock in every month.
* Dataset available form Jul'2005 to Nov'2020.
* We have total 185 entries
* Non-null values
* convert string object to date time object
* There are 0 missing values in this data set.
* For Analysis, select only Closing price.

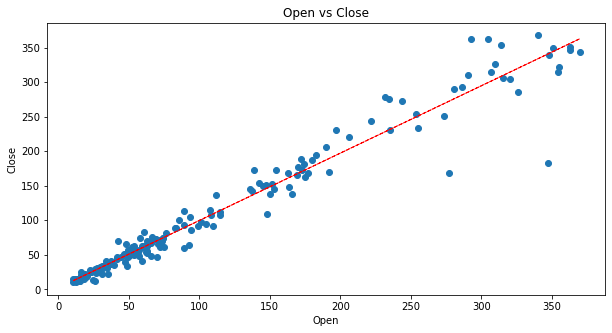
**Data Visualization/EDA-**

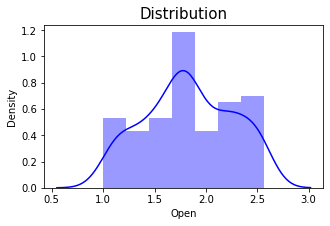
* Plotting close price vs date
* Plotting Distribution plot for closing price
* Make normal distribution applying log transformation
* Plotting distribution for each numeric features.
* Scatter plot to see relationship between dependent variable and independent variable
* Correlation heat map

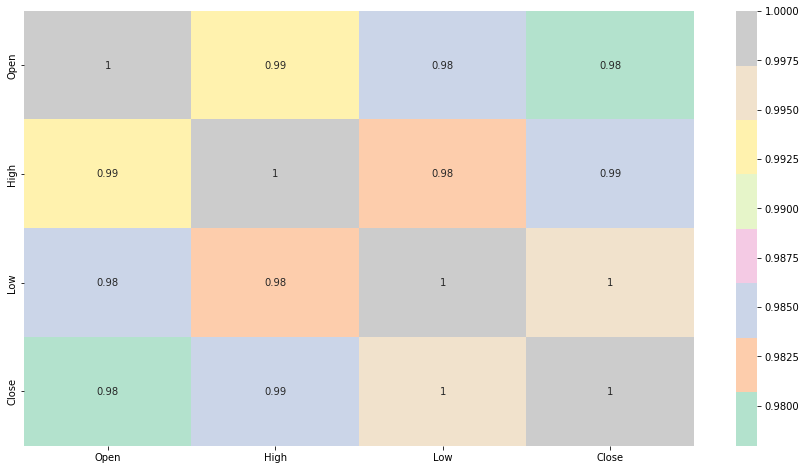
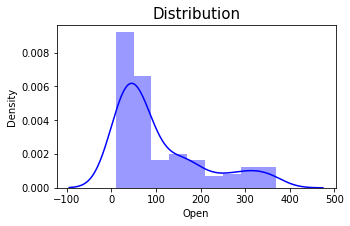








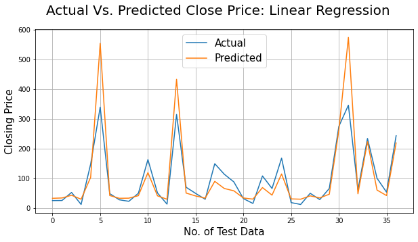




**Transforming & Splitting the data**

* Creating a new feature based on average of other features in the dataset.
* Established X as Independent Variable values & Y as Dependent Variable values
* Splitting our data into Dependent and Independent Variables
* Normalization and training the data
* We get Train (80%) & Test (20%) Split

**Model Building**

Regression -Regression model returns an equation that determines the relationship between the independent variables and the dependent variable.

Here I used four regression model-

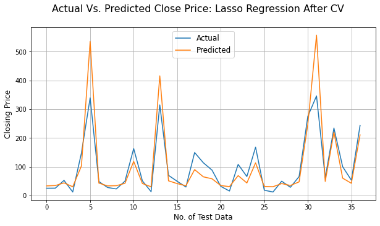
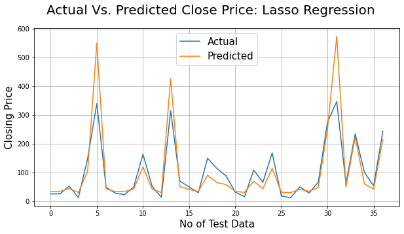
**Linear Regression**

\* Fit Linear Regression

\* Train & test Accuracy

\* Evaluation Metrics of Linear Regression

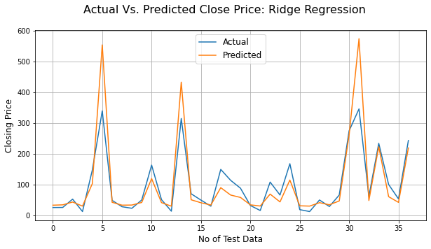
**Lasso Regression**

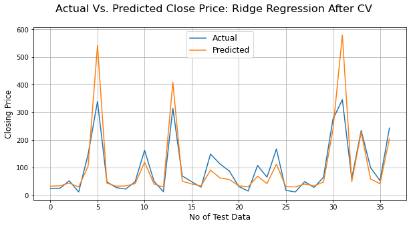
\* Fit Linear Regression

\* Train & test Accuracy

\* Evaluation Metrics of Linear Regression

\* Cross Validation on Lasso Regression



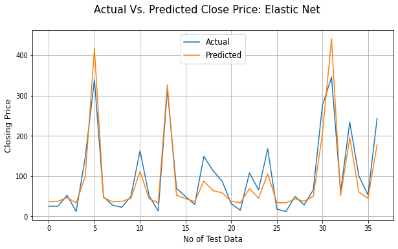
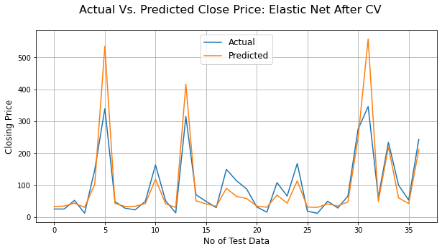
**Ridge Regression**

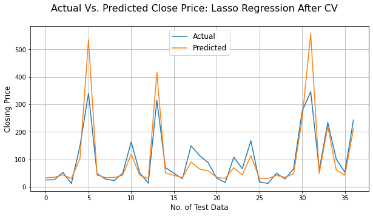
\* Fit Linear Regression

\* Train & test Accuracy

\* Evaluation Metrics of Linear Regression

\* Cross Validation on Lasso Regression

**Elastic Net**

\* Fit Linear Regression

\* Train & test Accuracy

\* Evaluation Metrics of Linear Regression

\* Cross Validation on Lasso Regression

**Evaluation Metrics Comparison**

After Completing regression, we get below result. And maximum Accuracy of the dataset is 82%.

**Conclusion**

* We got a maximum accuracy of 82%.
* Linear, lasso and ridge regression show almost same R squared values.
* Whereas elastic net model shows lowest R squared value and high MSE, RMSE, MAE & MAPE.
* Close, Open and high price of stock are strongly correlated with each other.