**Yes Bank Stock Closing Price Prediction**

**Suraj Kumar**

**Shreya Ranjan**

**Data Science Trainee**

**AlmaBetter, Bangalore**

**Abstract:**

The quantity of inventory that a company has on hand at the end of an accounting year is known as closing stock. Physical inventory counting is required to determine the amount of closing stock. The permanent inventory system can also use this information to calculate the final record of the amount of closing stock or inventory. Through retail banking and investment management services, it provides a wide range of unique solutions for corporate and retail consumers.

We used a set of stock price data from Yes Bank. This dataset includes 5 distinctive features that can be utilised for machine learning-based close price prediction. We have created a model that enables us to forecast stock values in the future.

We used all available models to predict the price of stocks, evaluated the results, and compared them for best accuracy and performance. We employed some of the greatest machine learning regression models.

**Keywords:** Stock Price prediction, Linear regression, Lasso regression, Ridge regression, Decision Tree regression.

1. **Problem Statement:**

In the Indian financial industry, Yes Bank is a well-known bank. Due to the Rana kapoor fraud case, it has been in the headlines since 2018. This made it interesting to investigate how it affected the company's stock prices and whether Time series models or other prediction models may be useful in such cases. Since the bank's inception, monthly stock prices have been collected in this dataset. Each month's closing, starting, highest, and lowest stock prices are included. The main goal is to forecast the stock's monthly closing price.

1. **Data Description:**

Understanding the data at a high level is important before carrying out any action on the dataset. It might be difficult to comprehend and interpret large data sets, depending on the type and scale of the data. By looking at a few of the initial and last rows after data loading, we examined the dataset. In our dataset, there are 185 rows and 5 columns of features, according to the shape analysis we performed. We noticed that the dataset contains several sorts of data, including float and object data. In order to forecast the future price of a stock, technical study analyzes previous stock prices, such as closing and opening prices, volume traded, adjacent close values, etc.

* **Date:** It shows the date of the investment (in our case we have month and year).
* **Open:** The price at which a stock began trading at the opening bell range is referred to as the "Open."
* **High:** High prices are the highest ones during a specific period of time.
* **Low:** Low prices relate to the lowest costs for a specific period of time.
* **Close:** Close describes a stock's price at the time the stock market shut down for the day.

1. **Introduction:**

In our project, building a prediction model for near price prediction, which focuses on short-term price prediction, is the goal of our project.

A stock market is an open marketplace where you can purchase and sell shares of firms that are publicly traded. The stocks sometimes referred to as equities, signify ownership in the business. The intermediary that makes it possible to buy and sell shares is the stock exchange. Stock markets help companies to raise capital.

* It promotes individual prosperity.
* An economic indicator is the state of the stock markets.
* People frequently utilise it as a source when investing in businesses with strong growth prospects.

We can determine the future worth of business stock and other financial assets traded on an exchange by utilising stock price prediction powered by machine learning.

The goal of stock price forecasting is to make substantial gains. It's challenging to make predictions about the future of the stock market. Other elements, such as biological and psychological components, rational and irrational behaviour, and so forth, play a role in the forecast. Share prices are dynamic and volatile due to a combination of all these variables. As a result, making highly accurate stock price predictions is quite challenging.

Modern stock price forecasting methods rely on sophisticated intelligent approaches based on either technical or fundamental analysis. In particular, the data size is large and non-linear for stock market analysis.

## **Exploratory Data Analysis:**

## **Data Cleaning: -**

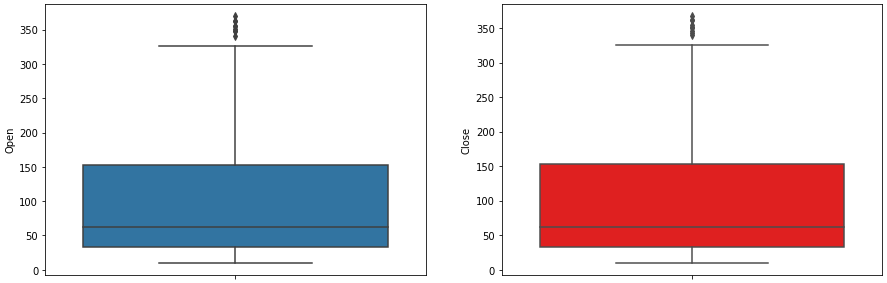
The given date in the data, which is in the MMM-YY format, is changed to the correct date of YYYY- MM-DD, and the given date column has the dtype object, which changes it to date and time format.

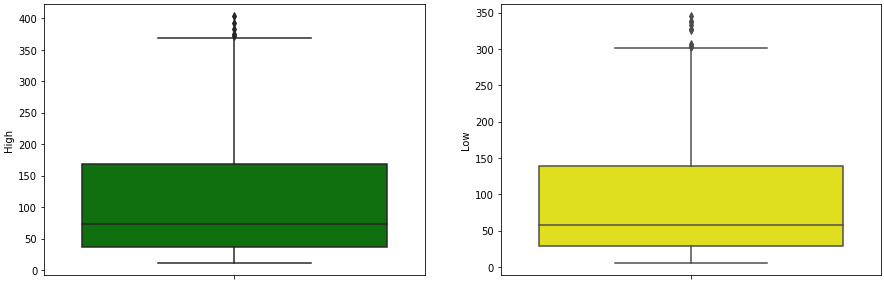
## **Null values Treatment:-**

Our dataset does not contain null values which might tend to interact with our accuracy; as a result, we dropped them at the start of our project to achieve a superior outcome.

## **Data Visualization:-**

The EDA components make data in a visual and graphical form more understandable. For data analysis, we mostly use the python modules matplotlib and seaborn. With graphs, the libraries are really helpful.

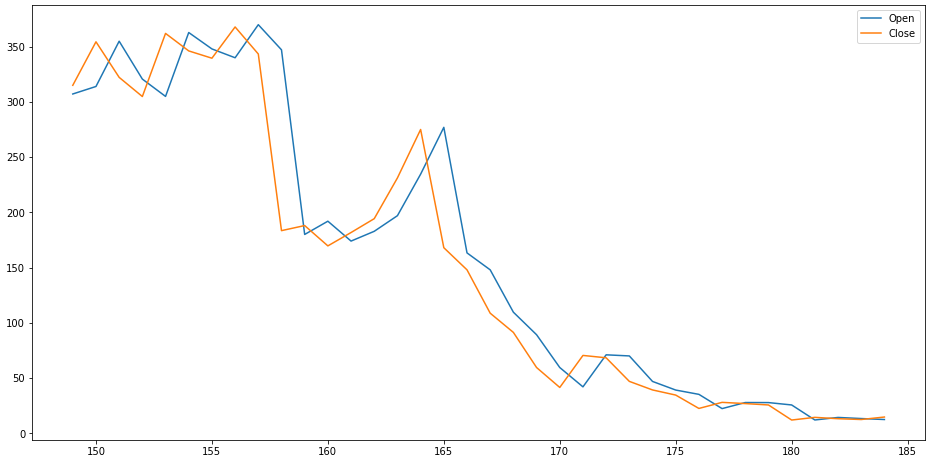




Outliners in Dataset

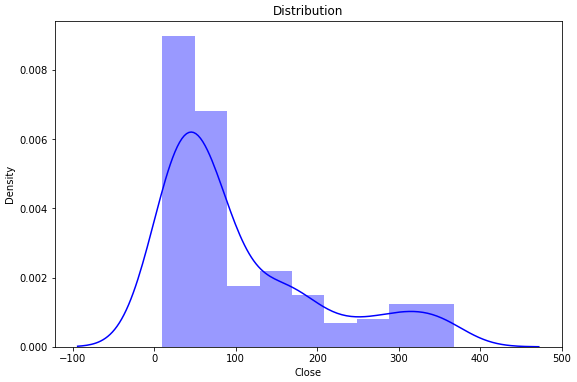
1. **Last three year record of opening and closing stock price:-**

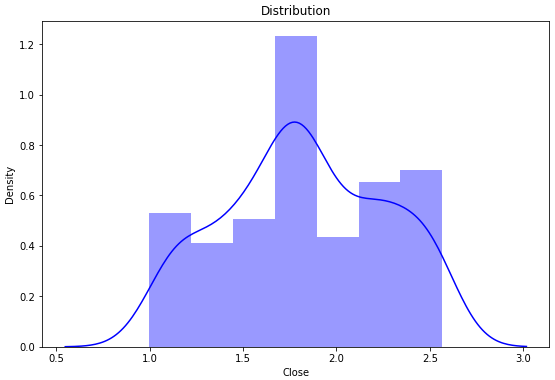
The line plot below shows that the stock price will continue to rise through 2018. But as a result of the Rana Kapoor fraud case, the stock price has continued to decline since 2018.



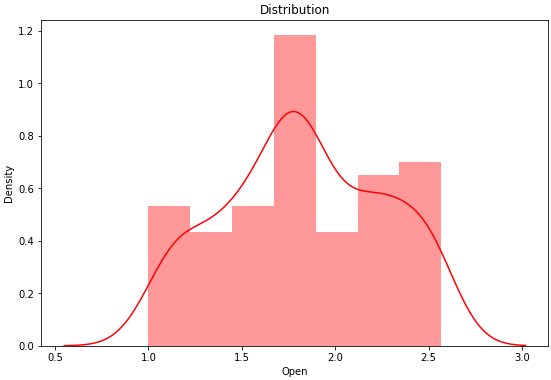
Plot Open vs Close price using line graph

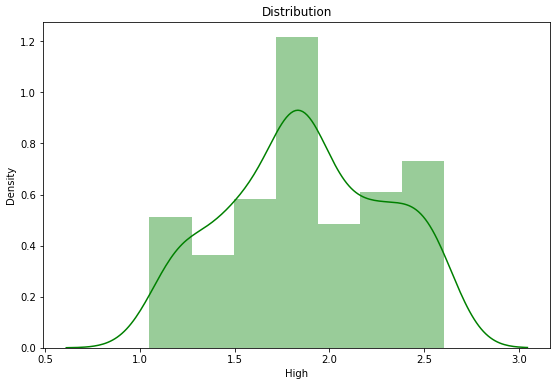
1. **Dependent variable of Close Price Stock:-**

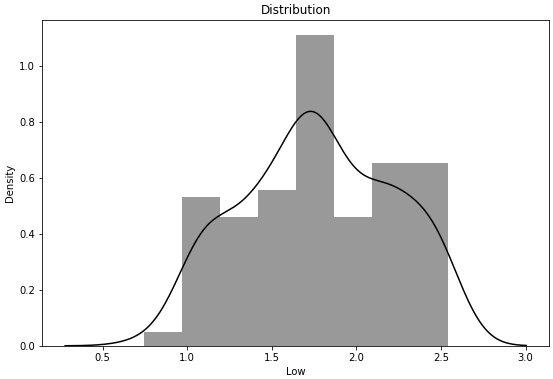
****

****

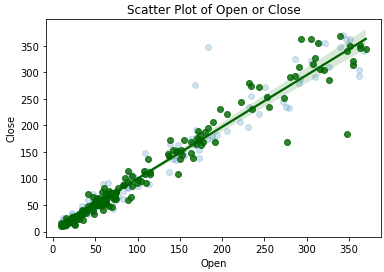
1. **Independent Variable Open , High and Low price of stock:**

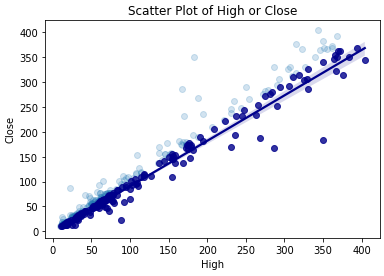
****

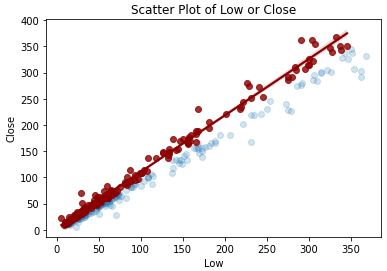
****

****

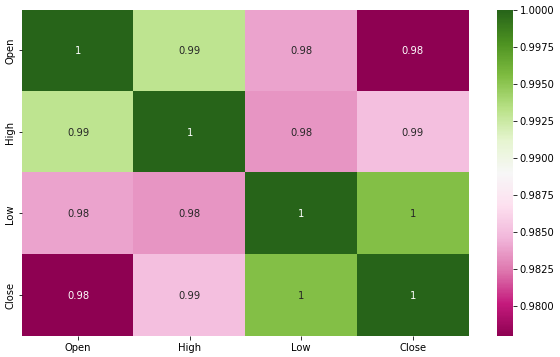
1. **relation between the Dependent Variable and Independent Variable:**

****

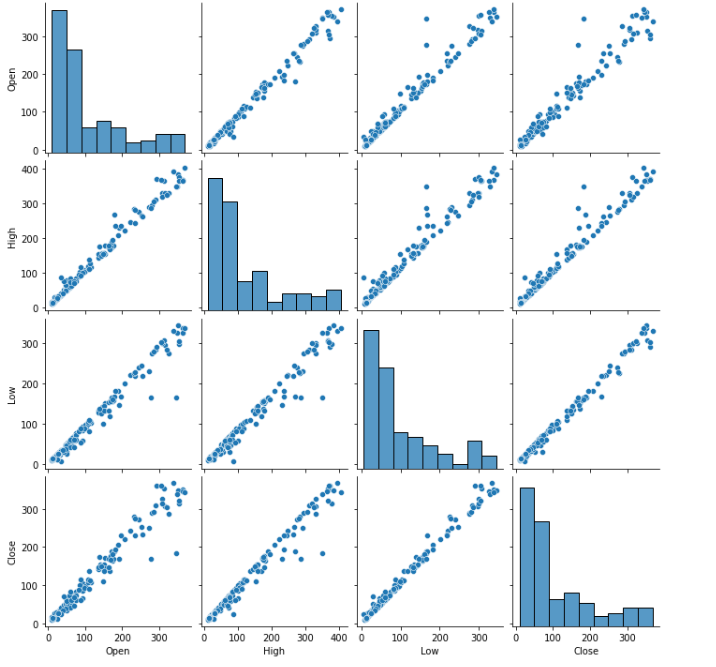
****

****

1. **Correlation Analysis with Heatmap:**

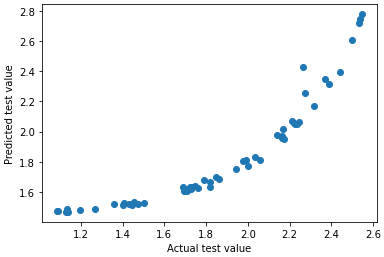
****

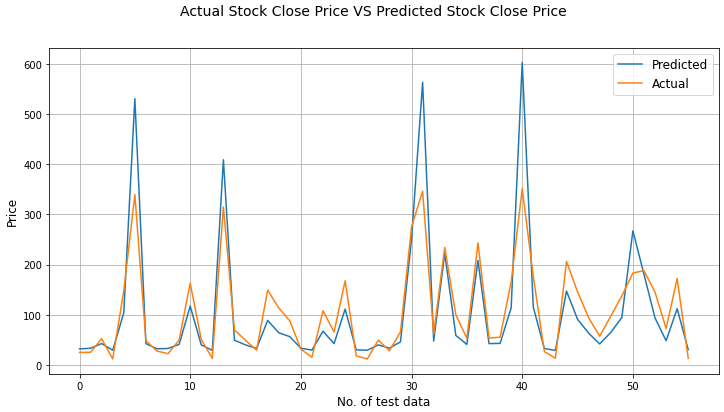
1. **Using pair grid to plot the entire dataframe while analyzing the relationship between the variables.**

****

1. **Linear Regression:**

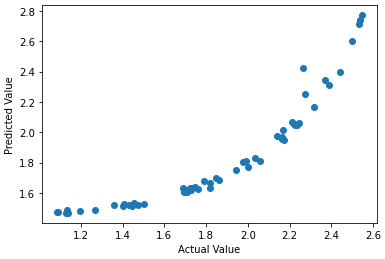
Linear regression is the most basic machine learning approach that can be applied to this data.The result of the linear regression model is an equation showing how the independent variables and dependent variable related to each other.

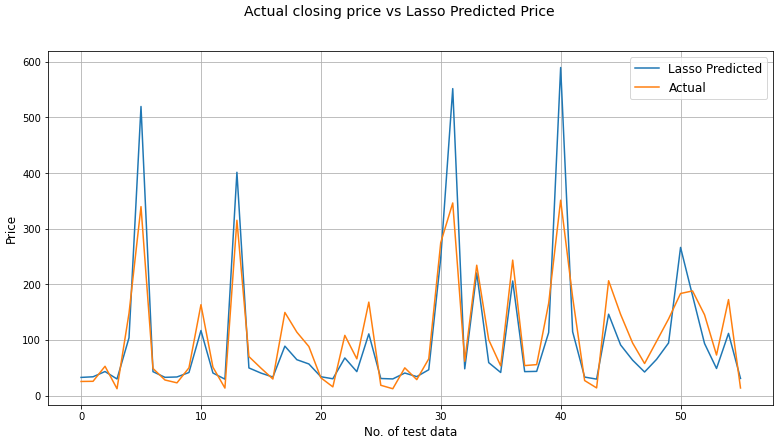




## **Lasso Regression:**

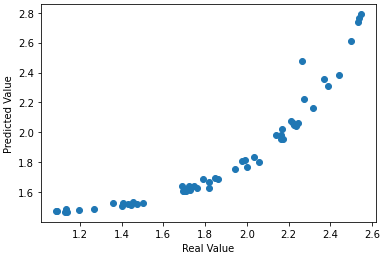
Lasso(least absolute shrinkage and selection operator) regression is another technique of Parameter estimation regression method. This method is usually used in machine learning for the selection of the subset of variables. It provides greater prediction accuracy as compared to other regression models. Lasso Regularization enhances the accessibility of models.

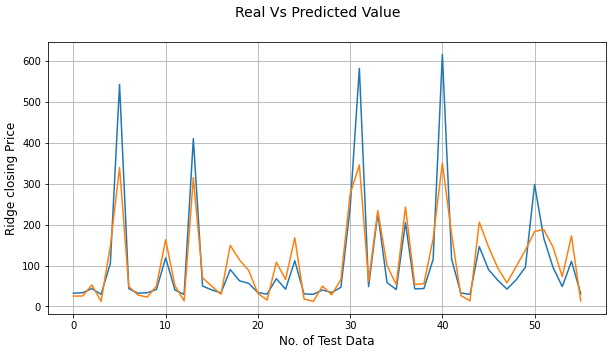




1. **Ridge Regression:**

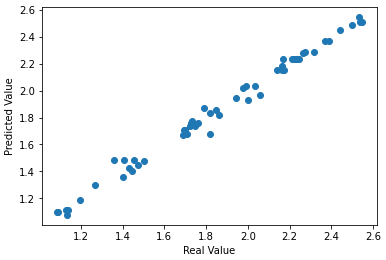
Ridge regression is a model-tuning technique that is used to analyse any multicollinear data. L2 regularisation is done using this technique. The projected values vary significantly from the actual values when the problem of multicollinearity is present, least-squares are unbiased, and variances are large.

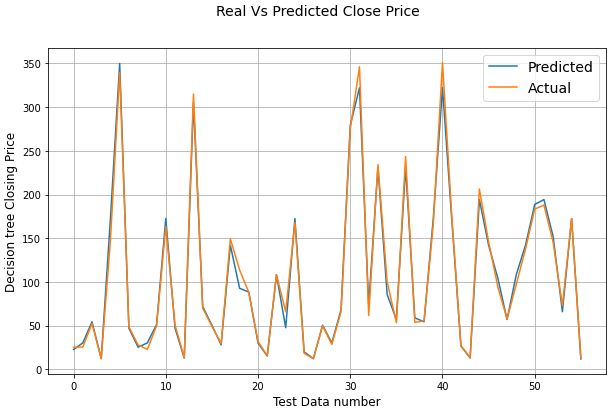




1. **Decision Tree Regression:**

Decision tree regression trains a model in the form of a tree to predict data in the future and generate useful continuous output by observing the properties of an item.





## **Conclusion:**

* At the conclusion of our exercise, that is where we are at. We started by importing our datasets, engaged in some data manipulation, data wrangling, and then conducted modelling (Linear regression, Lasso regression, Ridge regression & Decision Tree regression) where we made use of various graph types to gain a better understanding and come at a better conclusion.
* In our daily lives, the stock market has a surprising impact. It has a big impact on how quickly a nation's GDP grows.
* In this assignment, we learned the fundamentals of the stock market as well as machine learning stock price prediction techniques.
* After that we have done a graphical representation of the dependent and independent variables. And find out the relationship between dependent and independent variable.
* After conducting all of the predictive analysis, it was determined that the bank's stock price fluctuations were impacted by the fraud case, and based on model implementation, "linear regression" turned out to be the best performing model in this case, while "Lasso Regression" performed the worst of all.

## **References**:

* Geeks for Geeks
* Analytics Vidya
* My Great Learning

### 