**CAPSTONE PROJECT TECHNICAL DOCUMENT**

**Yes Bank Stock Price Prediction**

By- **Mohd Suhel**

[mohdsuhail2460@gmail.com](mailto:mohdsuhail2460@gmail.com)

Table Of Contents:

1. Abstract
2. Introduction
3. Problem Statement
4. Data Description
5. Exploratory Data Analysis
6. Data Visualization
7. ML Regression model(linear regressor,lasso,elastic net)
8. Conclusion

Abstract :

Yes Bank is a banking company that was founded in 2004 that offers a wide range of differentiated products for its corporate and retail customers through retail banking and asset management services. It is also a publically traded company. That provides an opportunity for anyone to invest in Yes bank and become a shareholder. But at the same time, it means that the valuation of the company is now in the hands of investors and speculators as share prices are often heavily impacted by public opinion.

We have used yes bank stock price data set. This dataset contains 5 different features that can be used for predicting close price prediction using machine learning. We have built machine learning regression model for price prediction. We have used some of best models.

Introduction :

YES bank stands for Youth Enterprise Scheme Bank. Stock market is one of the major fields that attracts people, thus stock market price prediction is always a hot topic for researchers from both financial and technical domains. In our project our objective is to build a prediction model for close price prediction. A stock market is a public market where you can buy and sell shares for publicly listed companies.

Stock Price Prediction using machine learning helps you get an estimate of value of company stock going forward and other financial assets traded on an exchange.

The entire idea of predicting stock prices is to gain significant profits. Predicting how the stock market will perform is a hard task to do. There are numerous other factors involved in the prediction, such as the psychological factor - namely crowd behavior etc. All these factors combine to make share prices very difficult to predict with high accuracy.

Problem Statements :

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 any 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 Description :

Before performing any operation on the dataset, it is important to understand the data. After loading the data, we observed the dataset by checking a few of the first and last rows. We checked the shape of the dataset and there are 185 rows and 5 features columns in our dataset.

Lets understand the features present in our dataset.

• 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 at the end of the considered time period.

Exploratory Data Analysis :

A) Data format : -

The Given Date in data is of Month-year format (mmm-yy) is converted to proper date of YYYY-MM-DD and given date column has data type as object converting it into date time format.

Data Visualization :

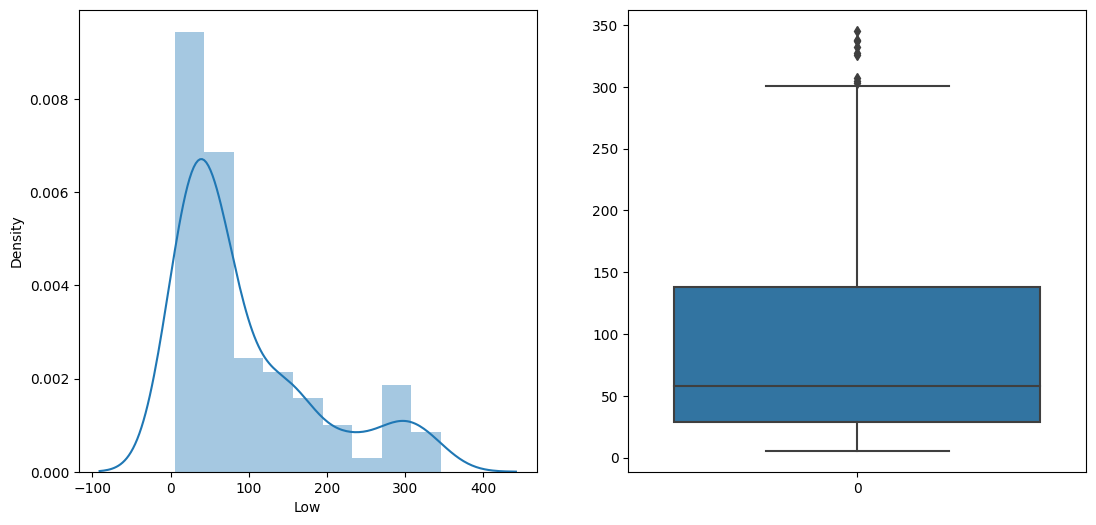
#### 1 . Plot of Closing Price (Univariate) :



From the above chart we can see that there was a good price rise from July 2005 till 2018 and it was a good time to invest as there was a up trend. The price between 5 Rs - 10 Rs in around July 2005 and reached a peak price of above 350 Rs in the year 2018.

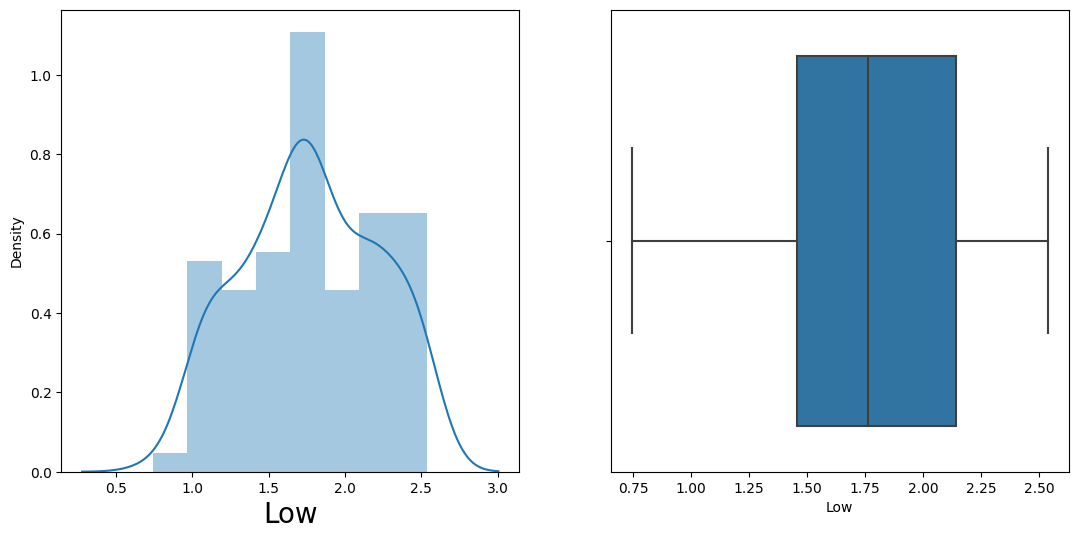
After Rana Kapoor fraud which was done in 2018 the stock price started to fall and there was a severe down trend, in around 2020 the price reached its July 2005 rate of around 5 Rs - 10 Rs.

2 . Box and Dist plot of independent variable for detecting outliers and skewness :



In our yes bank stock market dataset all the features have positively skewed distributions.

3 . Treatment of outliers and skewness :

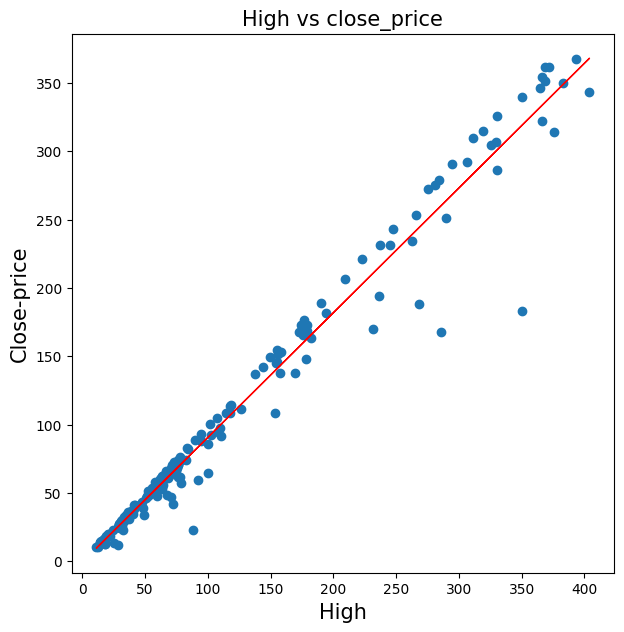


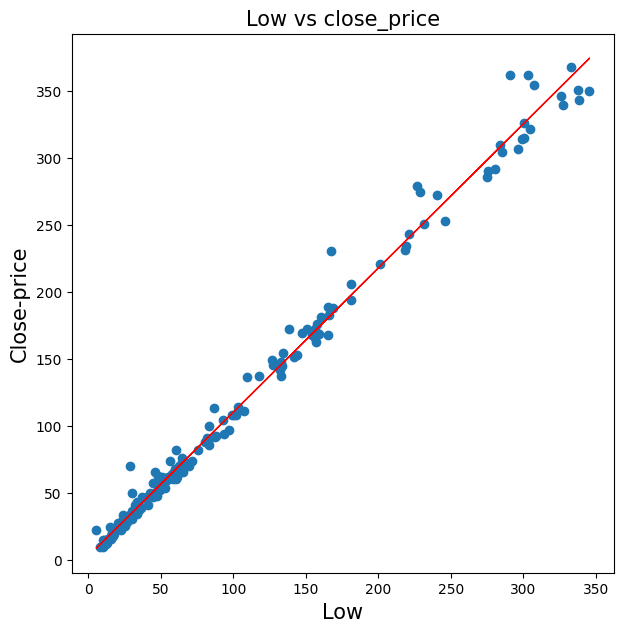
This charts show that our data variable's distplot and box plot when we performed log base 10 transformation for remove skewness.

4 . Correlation

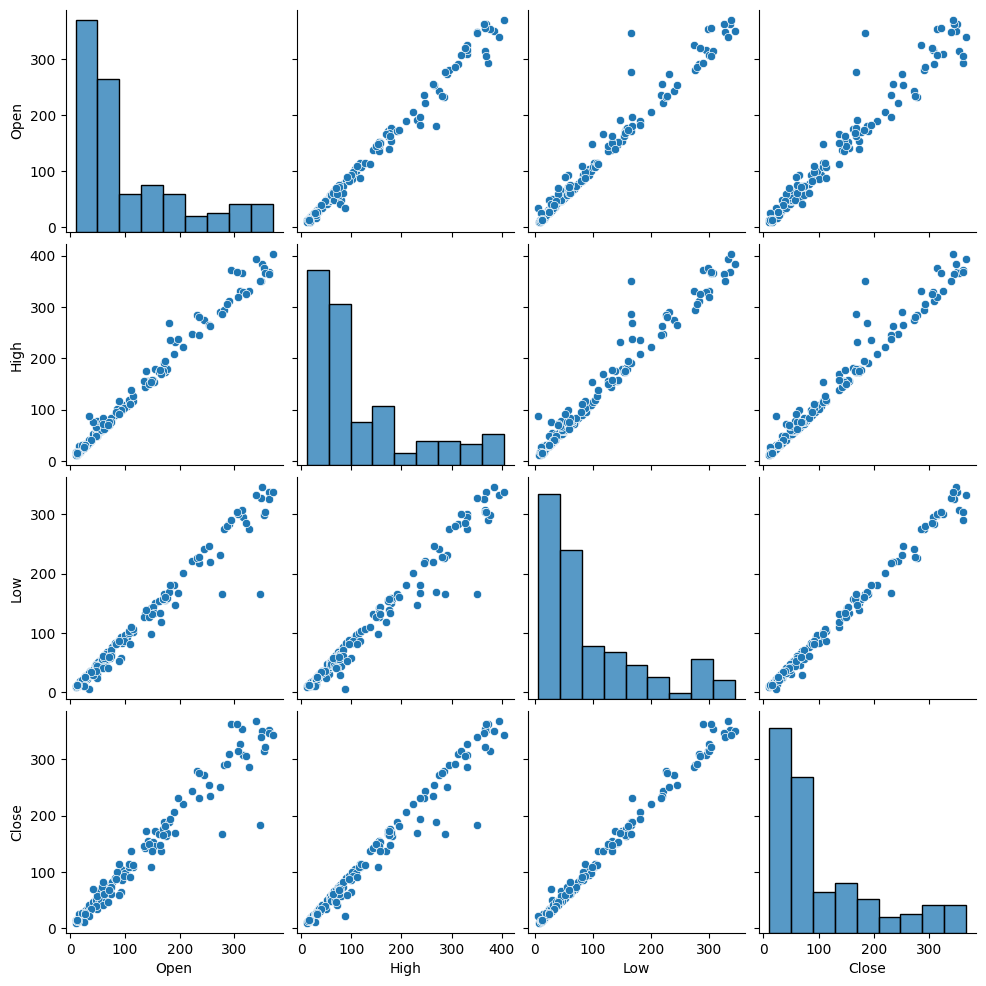
In the context of supervised learning, it can help determine the essential predictors when the bivariate analysis is done by plotting one variable against another. The graphs below depict that there is high correlation between dependent (Close) and independent variable.







5 . Pairplot :



Modelling :

1 . Linear Regression :

Linear regression is one of the easiest and most popular Machine Learning algorithms.

It works best when there is a linear relationship between dependent and independent variables.

Conclusion of Linear Regression:

Test MSE : 0.0008146348135191463

Test RMSE : 0.028541808168354478

Test R2 : 0.9956268809883747

Test Adjusted R2 : 0.9952293247145906

2 . Lasso Regression :

The goal of lasso regression is to obtain the subset of predictors that minimizes prediction error for a quantitative response variable. It does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Lasso performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.

Conclusion of Lasso Regression :

Lasso Test MSE : 0.0027290093067341534

Lasso Test RMSE : 0.052239920623352344

Lasso Test R2 : 0.9853501442804457

Lasso Test Adjusted R2 : 0.9840183392150316

3 . Elastic Net Regression :

Elastic net regression works in a manner that takes the best of lasso and ridge regressions. It adds up the penalty terms for regularization in lasso and ridge(L1 and L2) and uses that for regularization. It is used for regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.

Conclusion of Elastic Net Regression :

ElasticNet Test MSE : 0.002664776671232623

ElasticNet Test RMSE : 0.05162147490369316

ElasticNet Test R2 : 0.9856949576309396

ElasticNet Test Adjusted R2 : 0.9843944992337523

Conclusion :

1. Using data visualization on our target variable, we can clearly see the impact of 2018 fraud case involving Rana Kapoor as the stock prices decline dramatically during that period.
2. After loading the dataset, we found that there are no null values in our dataset nor any duplicate data.
3. There are some outliers in our features however this being a very small dataset, dropping those instances will lead to loss of information.
4. We found that the distribution of all our variables is positively skewed. so we performed log transformation on them.
5. There is a high correlation between the dependent and independent variables. This is a signal that our dependent variable is highly dependent on our features and can be predicted accurately from them.
6. We found that there is a rather high correlation between our independent variables. This multicollinearity is however unavoidable here as the dataset is very small.
7. We implemented several models on our dataset in order to be able to predict the closing price and found that all our models are performing remarkably well and Linear Regressor is the best performing model which R2 score value of 99.5% .
8. All of the implemented models on Cross-Validation and Hyperparameter Technique performed quite well on our data giving us the R2 Score of over 99%.
9. With our model making predictions with such high accuracy, we can confidently deploy this model for further predictive tasks using future data.