

Capstone Project-II

Yes Bank Stock Closing Price Prediction

Pankaj Kumar Yadav,
Harshada Gore



About Project:

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.

POINTS FOR DISCUSSION :

Problem Statement

Introduction

Data Cleaning

Extra Data Analysis(EDA)

Transforming Data

Splitting Data

Data Modeling

Cross Validation & Hyperparameter Tunning

Conclusion



Problem Statement

- ❖ Perform regression analysis using multiple models to predict the closing price of the stock and compare the evaluation metrics for all of them to find the best model.
- ❖ Prediction of Yes Bank stock closing price.
- ❖ Getting accuracy score of several machine learning model.



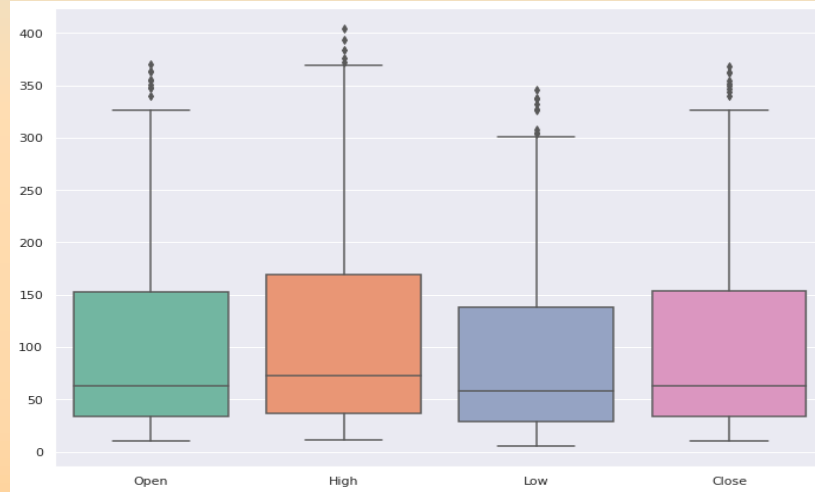
Introduction

- Data set - data Yes Bank Stock Prices - contains observations regarding open, close, high and low prices of the yes bank stock from July 2005 - November 2020.
1. **Date:** Monthly observation of stock prices since its inception.
 2. **Open:** The price of a stock when stock exchange market open for the day.
 3. **Close:** The price of a stock when stock exchange market closed for the day.
 4. **High:** The maximum price of a stock attained during given period of time.
 5. **Low:** The minimum price of a stock attained during given period of time.



Data Cleaning

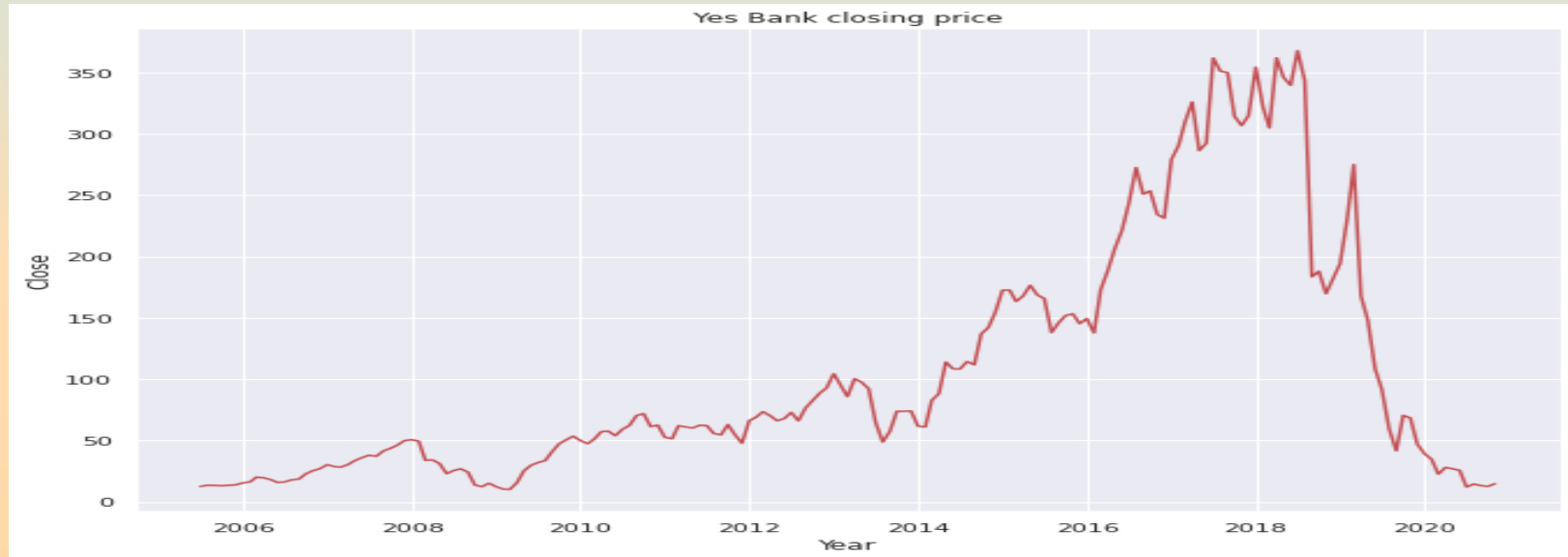
- Null Values Treatment
- Duplicated Values Treatment
- Missing Values Treatment
- Date Format Change (i.e. from Jul-05 to 2005-07-01)
- Checking outliers



- After successfully cleaning the dataset we have 185 columns and 5 rows.

Exploratory Data Analysis(EDA)

Let's see Year wise the Close price of stock of dataset through visualizing the data



- Now, We can easily see in above plot **the trend is increasing from 2009 to 2018** but after that the trend decreases. This is because of the fraud case of involving **Rana Kapoor**.

Exploratory Data Analysis(EDA)

Distribution of Closing Price :



- Distribution of closing price is right skewed.
- We need this distribution to be normal distribution for training algorithm.

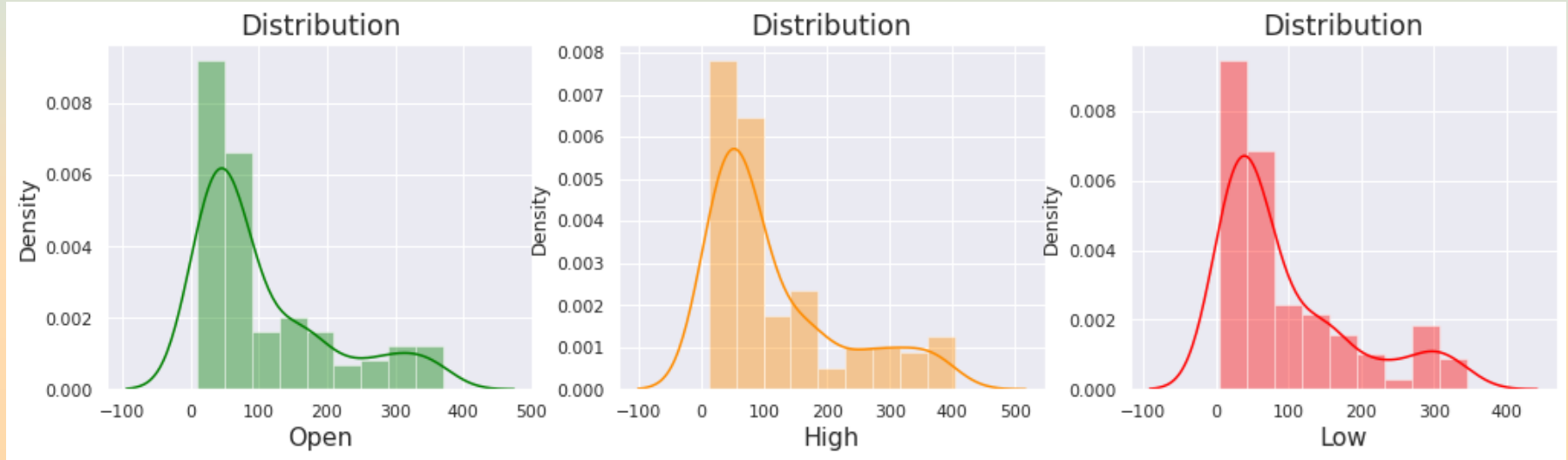
After Log Transformation :



- Now, distribution of closing price is more normal, after applying log transformation

Exploratory Data Analysis(EDA)

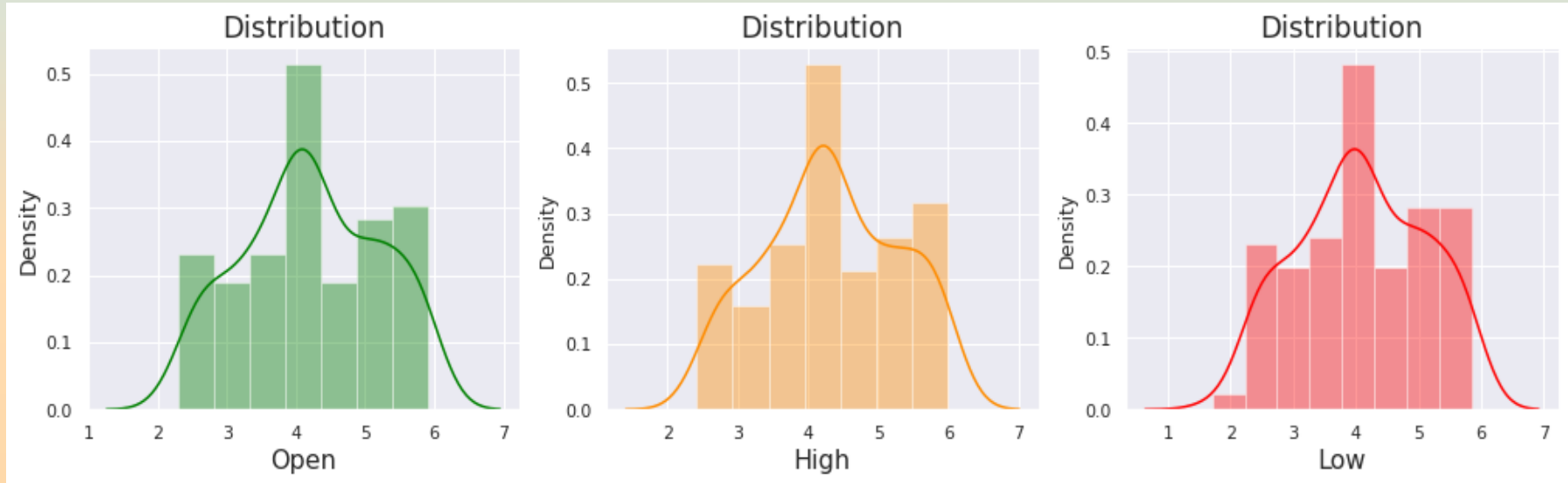
Distribution of Open, High & Low Price of a stock :



- Distribution of opening price, high price and low price are also right skewed.
- Log transformation applied to make this distribution normal.

Exploratory Data Analysis(EDA)

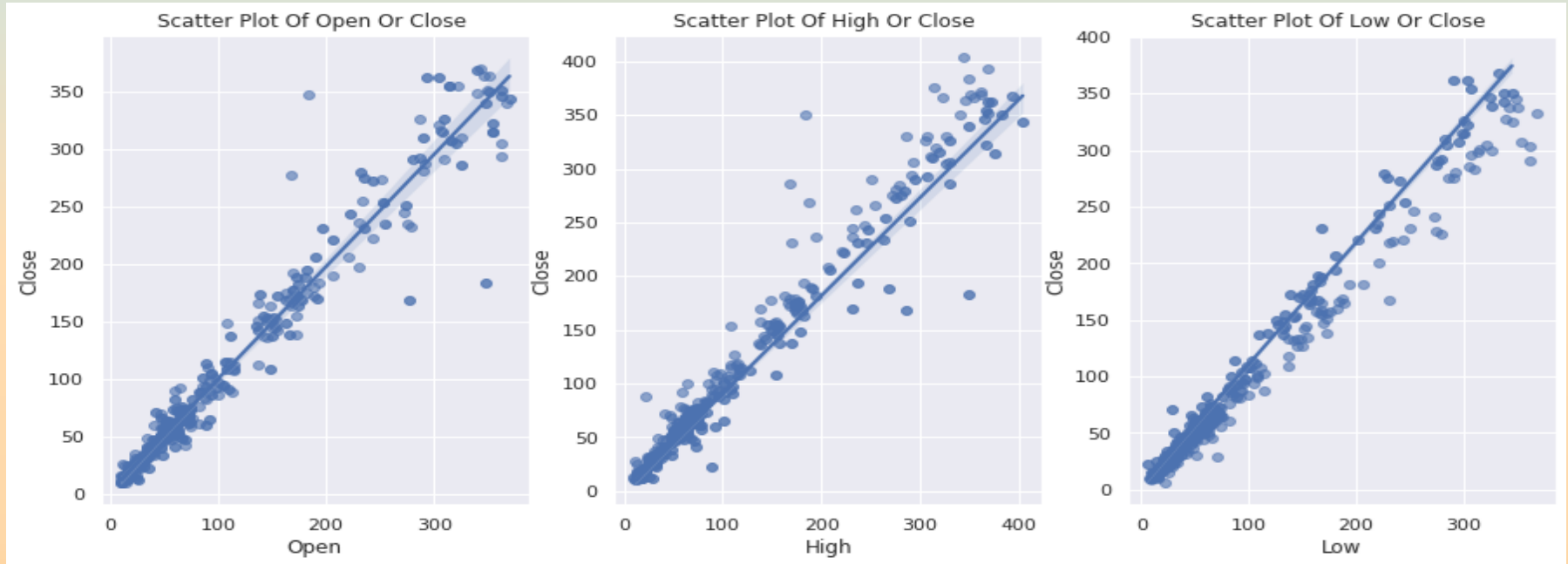
Distribution of Open, High & Low Price of a stock after Log Transformation :



- Distribution of opening price, high price and low price are now normal distribution.

Exploratory Data Analysis(EDA)

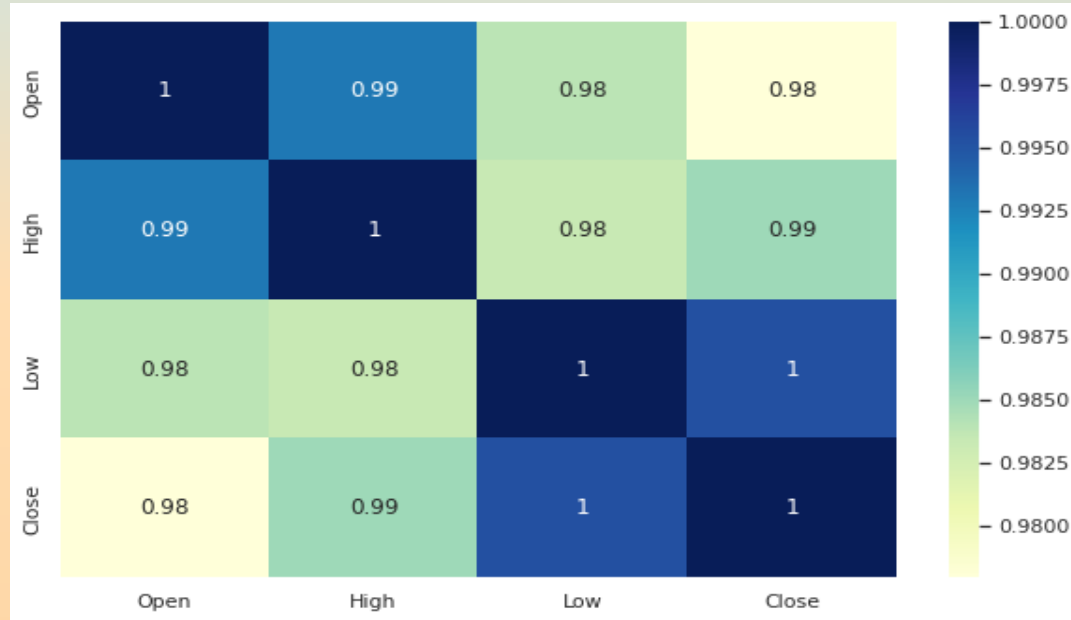
Bivariate Analysis with Scatter Plots



- Now find out the relation between the Dependent Variable and independent Variable
- scatter plot with Best Fit line

Exploratory Data Analysis(EDA)

Correlation:



- All the features are strongly correlated with each other.

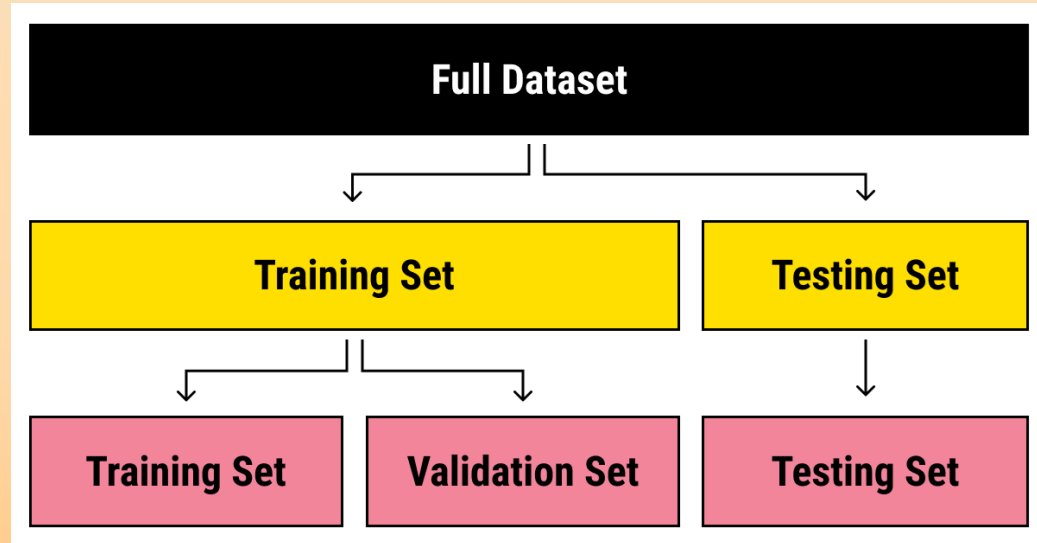
Transformation of Data

- To scale data into a uniform format that would allow us to utilize the data in a better way.
- For performing fitting and applying different algorithms to it.
- The basic goal was to enforce a level of consistency or uniformity to dataset.



Splitting Data

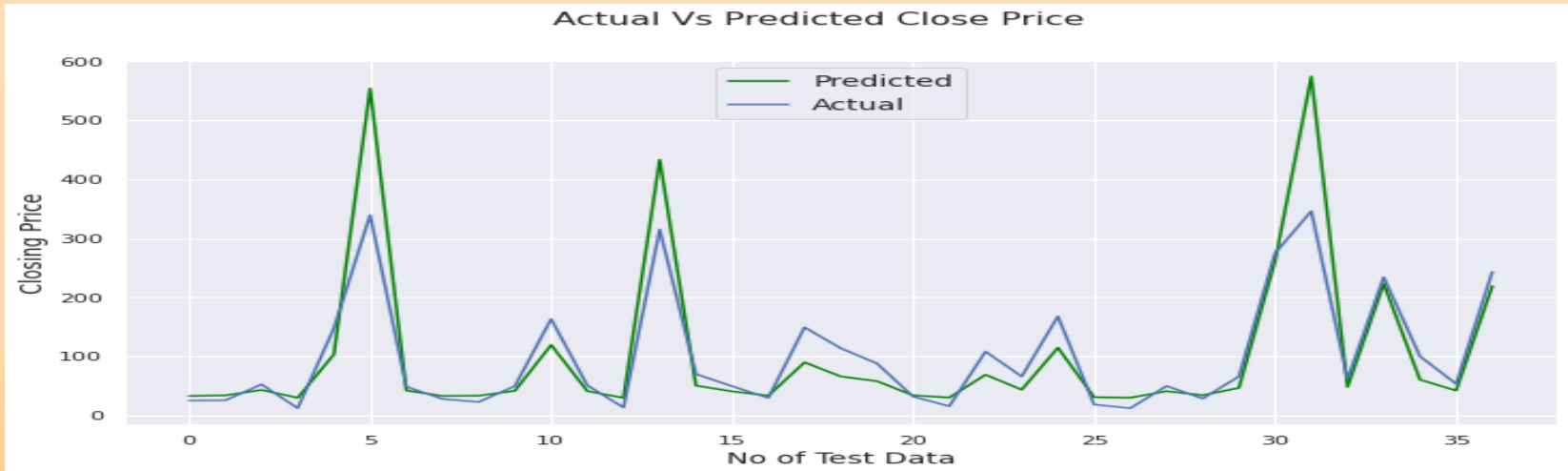
- Data splits into training dataset and testing dataset.
- Training dataset is for making algorithm learn and train model.
- Test dataset is for testing the performance of train model.
- Here 80% of data taken as training dataset & remaining 20% of dataset used for testing purpose.



Data Modeling

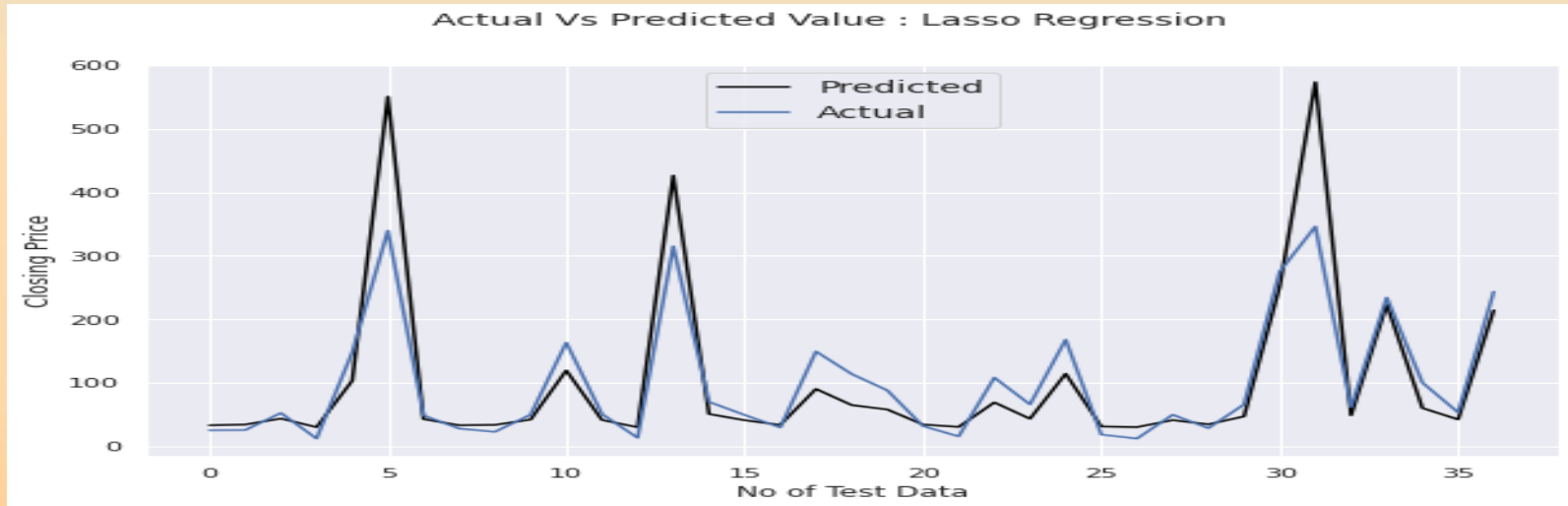
1. Linear Regression:

- Linear regression is one of the easiest and most popular Machine Learning algorithms.
- It is a statistical method that is used for predictive analysis.
- Linear regression algorithm shows a linear relationship between a dependent and independent variable; hence it is called as linear regression



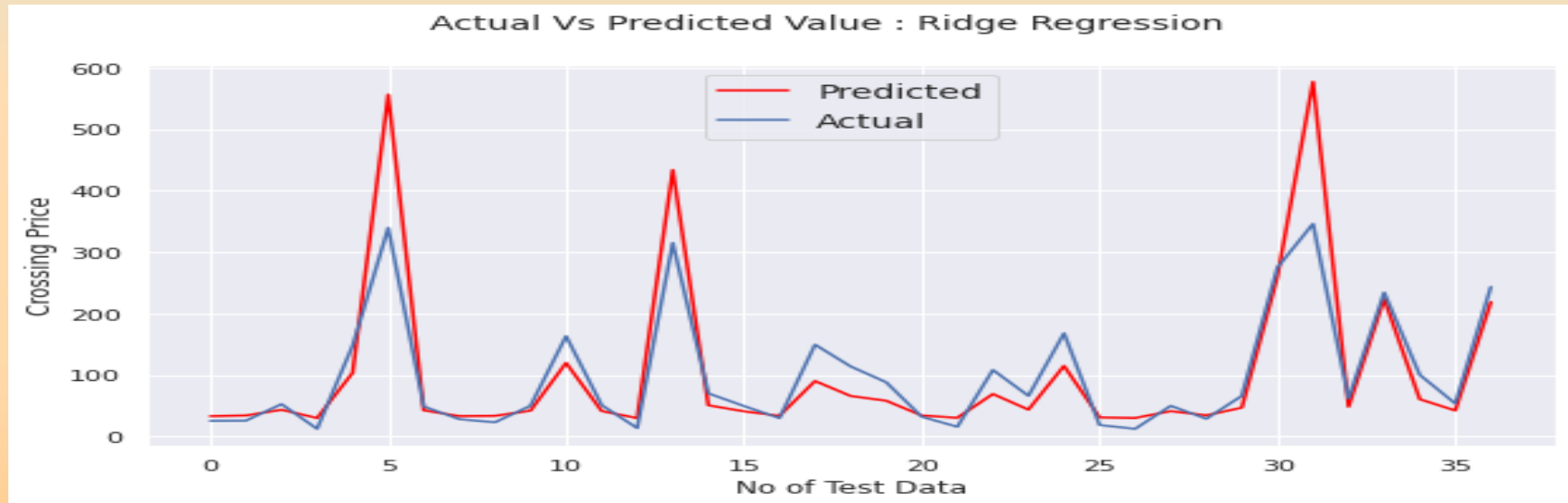
2. Lasso Regression:

- Lasso: Least Absolute Shrinkage and Selection operator.
- It is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.
- This method performs L1 regularization.



3. Ridge Regression:

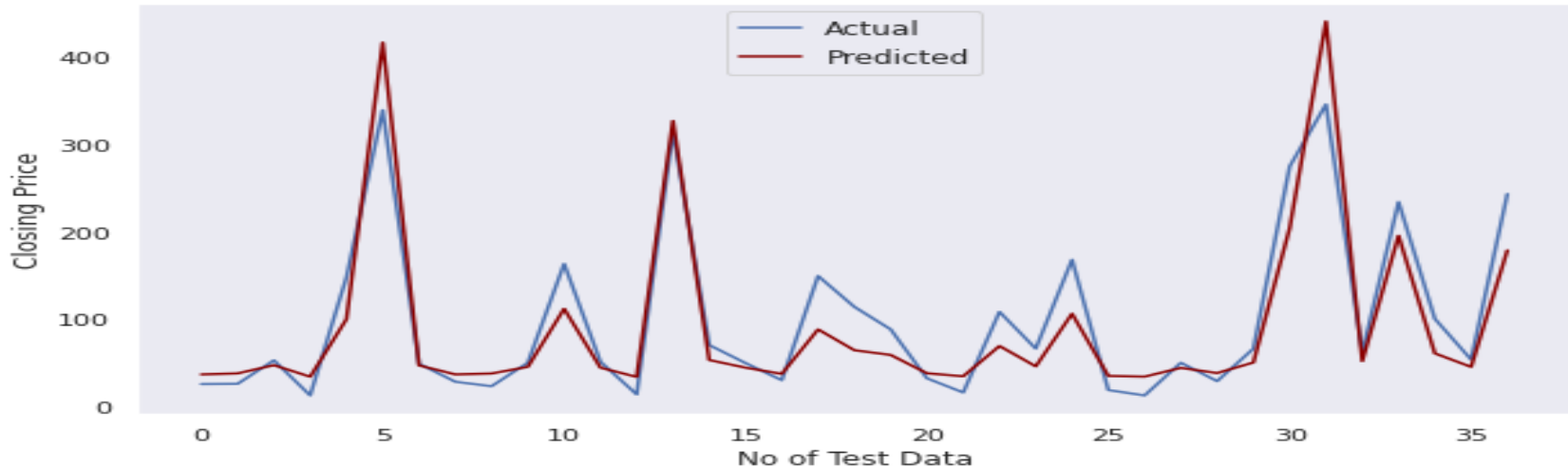
- Ridge regression is a model tuning method that is used to analyses any data that suffers from Multicollinearity.
- When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values.
- This method performs L2 regularization.



4. Elastic Net Regression:

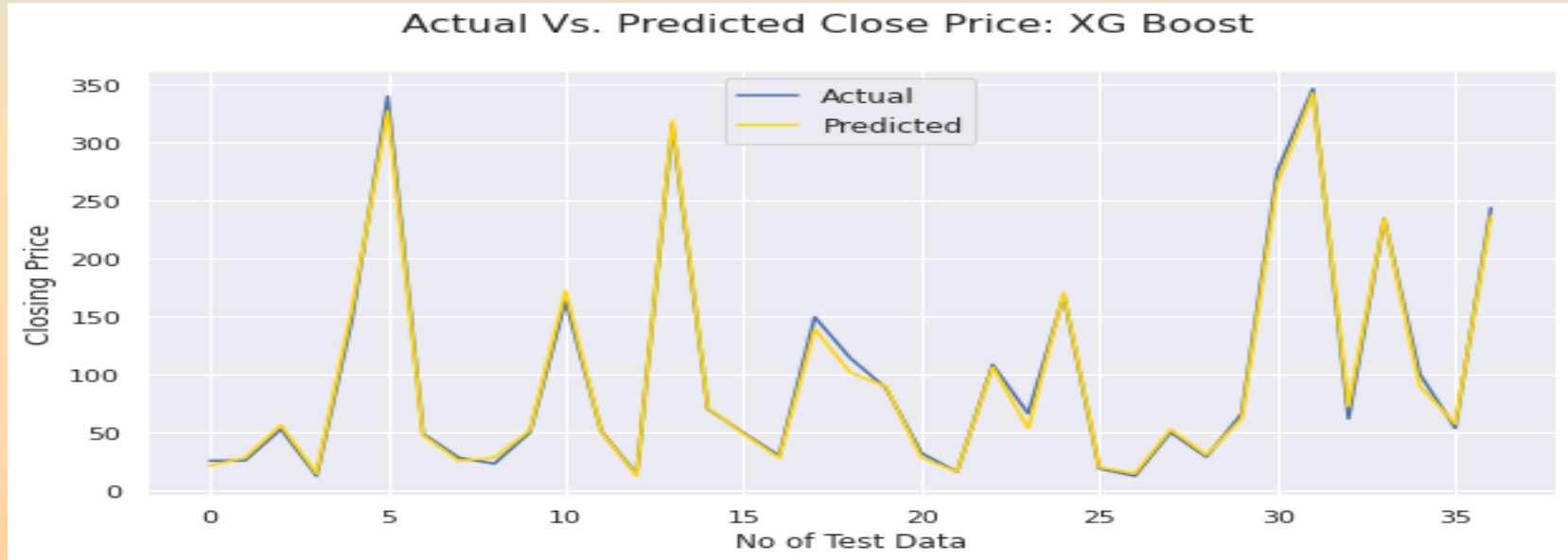
- Elastic net is a popular type of regularized linear regression that combines two popular penalties, specifically the L1 and L2 penalty functions.
- Elastic Net is an extension of linear regression that adds regularization penalties to the loss function during training.

Actual Vs. Predicted Close Price: Elastic Net



5. XGBoost Regressor:

- XGBoost stands for “Extreme Gradient Boosting”. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements Machine Learning algorithms under the Gradient Boosting framework. It provides a parallel tree boosting to solve many data science problems in a fast and accurate way.



Final Score Point Table of all Modals

Model_Name	MAE	MSE	RMSE	MAPE	Rsquare
XGBRegressor	0.030	0.002	0.039	1.956	0.991
LinearRegression	0.151	0.032	0.178	9.543	0.823
Ridge	0.152	0.032	0.179	9.580	0.820
Lasso	0.152	0.032	0.179	9.623	0.820
ElasticNet	0.157	0.036	0.191	10.240	0.796

Conclusion

- The popularity of stock closing is growing extremely rapidly day by day which encourage the researcher to find new methods if any fraud happens.
- This technique is used for prediction and is not only helpful to researchers to predict future stock closing prices or any fraud happen or not but also helps investors or any person who dealing with the stock market in order to prediction of model with good accuracy.
- In this work we use the linear regression technique, lasso regression, ridge regression, elastic net regression, and XGBoost Regression technique. these five models give us the following results
- High, low, and open are directly correlate with the closing price of stocks
- Target variable (dependent variable) is strongly dependent on independent variables
- Xgboost regression is the best model for yes bank stock closing price data this model used for further prediction



thank you!