PROJECT REPORT

YES BANK STOCK CLOSING PRICE PREDICTION

21CSC305P

Submitted by-:

YASH SARIN [Reg No.:RA2211026030015]
AKSHAY MALVIYA [Reg No.:RA2211026030034]
RISHABH JHA [Reg No.:RA2211026030062]
SAARANSH YADAV [Reg No.:RA2211026030063]

Under the guidance of

Ms. Deepinder Kaur
(Assistant Professor, Department of Computer Science & Engineering)

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

SRM
INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

SRM Institute of Science and Technology, NCR Campus

Odd Semester- 2024-25

BONAFIDE CERTIFICATE

CERTIFIED TO BE THE BONAFIDE RECORD OF PROJECT REPORT ON "**YES BANK STOCK CLOSING PRICE PREDICTION**" DONE BY **YASH SARIN [Reg No.: RA2211026030015], AKSHAY MALVIYA [Reg No.: RA2211026030034], RISHABH JHA [Reg No.: RA2211026030062], SAARANSH YADAV [Reg No.: RA2211026030063]** B. TECH DEGREE COURSE IN SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, DELHI-NCR CAMPUS FOR THE DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, IN MACHINE LEARNING DURING THE ACADEMIC YEAR 2024-2025.

Ms. Deepinder Kaur
(Assistant Professor)

Head of the Department
(CSE)

Submitted for End Semester examination held on / / at SRM INSTITUTE OF SCIENCE & TECHNOLOGY, DELHI-NCR Campus.

Internal Examiner-I

Internal Examiner-II

INDEX

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 4 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 Overview

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.

Yes Bank Objective:

Yes Bank is a well-known bank in the Indian financial domain. It has been in the headlines since 2018 as a result of the Rana Kapoor fraud case. Due to this, it was interesting to observe how it affected the company's stock prices and whether Time series models or other prediction models could properly reflect for such circumstances. Since the bank's founding, this dataset has included closing, starting, highest, and lowest stock prices for each month.

We have 185 rows and 5 columns in our dataset. Here our dependent variable is Close and Independent variable is Open, High and Low.

Date :- It denotes the month and year for a specific pricing.

Open :- The price at which a stock started trading that month is referred to as the "Open."

High :- The highest price for that particular month.

Low :- It describes the monthly minimum price.

Close:- It refers to the final trading price for that month, which we have to predict using regression.

Independent or input variables -: ‘Open ’, ‘ High ’,‘ Low’

Dependent or target variable -: ‘Close’

‘Date ’is only useful for EDA purpose & don't have any influence for closing price prediction.

Problem Statement: 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.

Proposed Solution:

1. Data Understanding and Preparation

- Dataset Review:
    - Examine the provided dataset containing monthly stock prices (closing, opening, highest, lowest) since the bank's inception.
    - Ensure data consistency, handle missing values, and potentially outliers.
- Feature Selection:
    - Focus primarily on the closing price as the target variable for prediction.
    - Other features like opening, highest, and lowest prices can be used for additional analysis or feature engineering.

2. Exploratory Data Analysis (EDA)

- Visualize the trends, seasonality, and any underlying patterns in the closing prices over time.
- Use statistical techniques like autocorrelation function and partial autocorrelation function plots to understand dependencies.

3. Model Training :

- Spilt data into training and validation sets.
- Train model on historical data and validate using out of sample data to assess performance.

4. Deployment:

- Implement the model in a production environment for ongoing predictions.
- Monitor model performance and retrain periodically if necessary.

5. Model Evaluation

- Use metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to evaluate model accuracy.
- Compare different models to choose the best performing one.

6. Result

Technology Used

Hardware & Software specification

- Processor (CPU): For personal use and small-scale applications, a modern dual-core or quad-core processor should suffice.
- Memory (RAM): For personal use, Windows 10/11 operating system ,6 GB of RAM is typically adequate. For small to medium-sized businesses or more extensive personal usage, 8 GB or more is recommended, Python 3.11/3.10, Jupyter Notebook.
- Library required to build the model:

- NumPy
- Panda
- Matplotlib
- Seaborn
- Datetime
- Sklearn

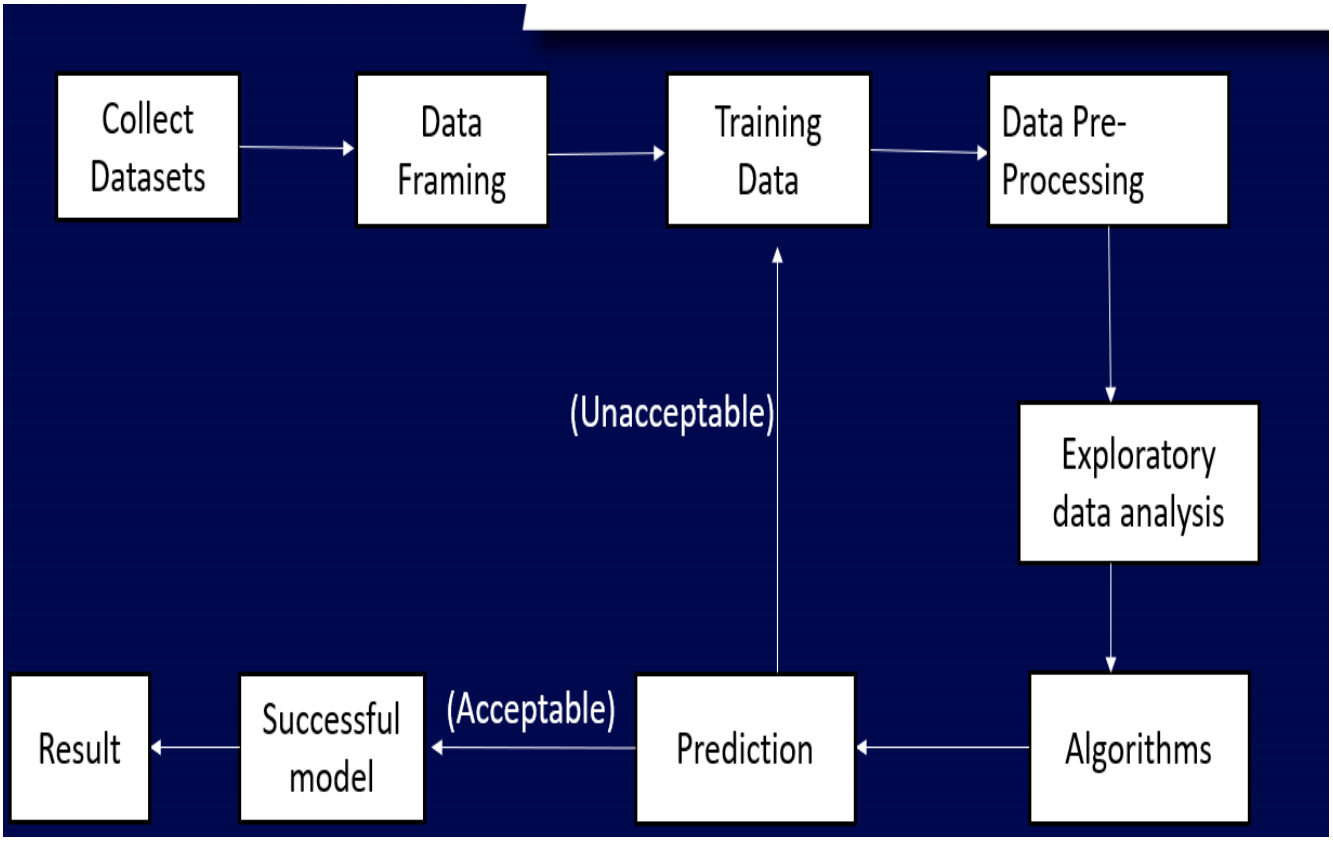
ADVANTAGES

- Data-Driven Insights: Machine learning models can analyze vast amounts of historical data, identifying patterns and trends that may not be apparent through traditional analysis.
- Predictive Accuracy: With the right algorithms and features, machine learning can improve the accuracy of stock price predictions, helping investors make informed decisions.
- Real-Time Analysis: Machine learning systems can process and analyze data in real time, allowing for timely trading decisions based on current market conditions.
- Automated Trading: Implementing machine learning models can lead to automated trading strategies, which can execute trades based on predictions without human intervention.
- Risk Management: By predicting stock price movements, machine learning can help in assessing risks, allowing investors to manage their portfolios more effectively.
- Sentiment Analysis: Integrating sentiment analysis from news articles, social media, and financial reports can provide additional context for stock predictions, capturing market sentiment.
- Adaptability: Machine learning models can adapt to changing market conditions over time, improving their predictions as new data becomes available.
- Feature Selection: Advanced techniques can help in identifying the most relevant factors influencing stock prices, leading to more focused and effective models.
- Backtesting Capabilities: Machine learning models can be backtested against historical data to evaluate their performance before applying them to real-world scenarios.
- Visualization Tools: Incorporating machine learning can lead to better data visualization, helping stakeholders understand trends and predictions intuitively

DRAWBACKS

- Data Quality and Availability: Accurate predictions depend on high-quality historical data. Incomplete, noisy, or biased data can lead to poor model performance.
- Overfitting: Machine learning models may become too complex, capturing noise instead of underlying trends, resulting in poor generalization to unseen data.
- Feature Selection Challenges: Identifying the right features (variables) to include can be difficult, and irrelevant or redundant features can degrade model performance.
- Model Interpretability: Some machine learning models (like deep learning) are complex and difficult to interpret, making it hard for investors to understand the rationale behind predictions.
- Dependence on Historical Trends: Machine learning models primarily rely on historical data, which may not always predict future movements, especially in rapidly changing markets.
- Computational Resources: Training complex machine learning models can require significant computational power and resources, which may not be accessible to all individuals or institutions.
- Regulatory and Ethical Concerns: Automated trading based on predictions may raise regulatory issues or ethical concerns, particularly in terms of market manipulation.
- Investment Risks: Relying solely on machine learning predictions can lead to substantial financial risk if predictions are incorrect, as stock trading inherently involves uncertainty

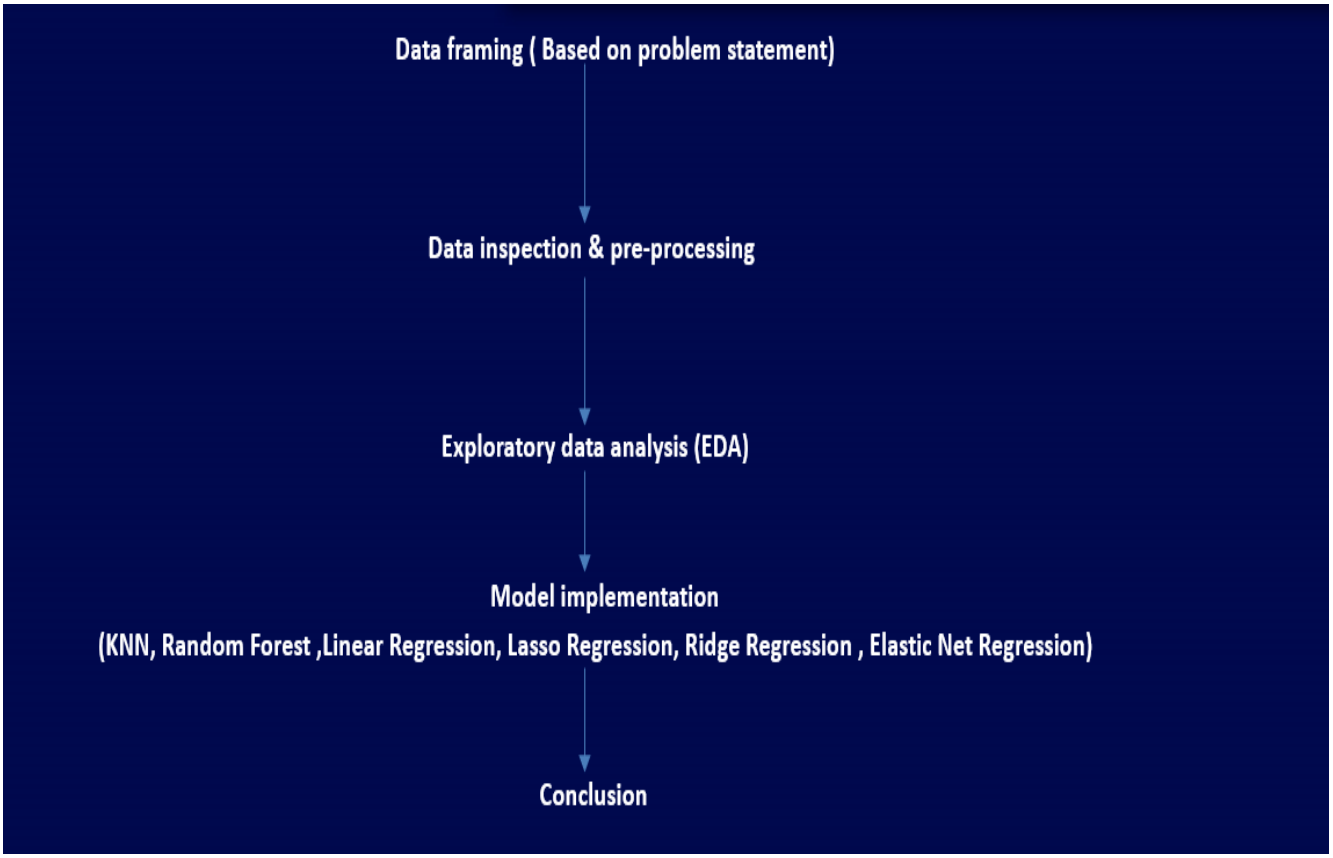
ARCHITECTURAL DIAGRAM



Algorithm & Deployment

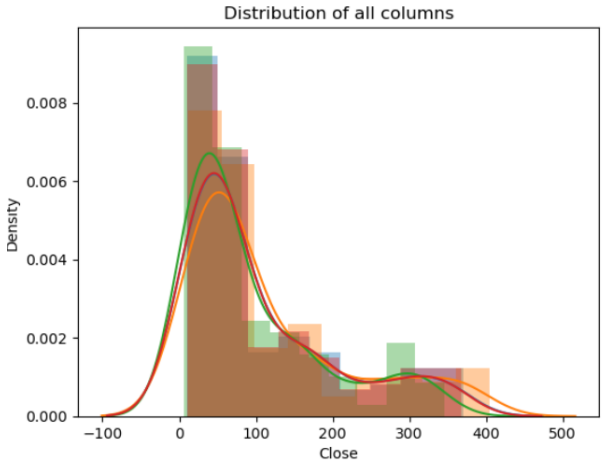
- Step 1- Dataset- Collection of dataset from Kaggle.
- Step 2- Data Framing – Define the success metrics used , define the ideal outcome and models goals. Or what is observed and what should be predicted based on Problem Statement.
- Step 3-Data Pre-Processing- Data processing is a crucial step in the machine learning (ML) , as it prepares the data for use in building and training the models.
- Step 4- Exploratory data analysis (EDA)- Summarize the main characteristics according to the dataset.
- Step 5- K-Nearest Neighbours (KNN)-The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until function evaluation. Since this algorithm relies on distance for classification, normalizing the training data can improve its accuracy dramatically.
- Step 6- Random Forest: Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and aggregation , commonly known as bagging.
- Step 7- Implementation of Regression- We create 4 regression models for our data such as Linear Regression, Lasso Regression , Ridge Regression , Elastic Net Regression.
- Step 8- Conclusion

METHODOLOGY



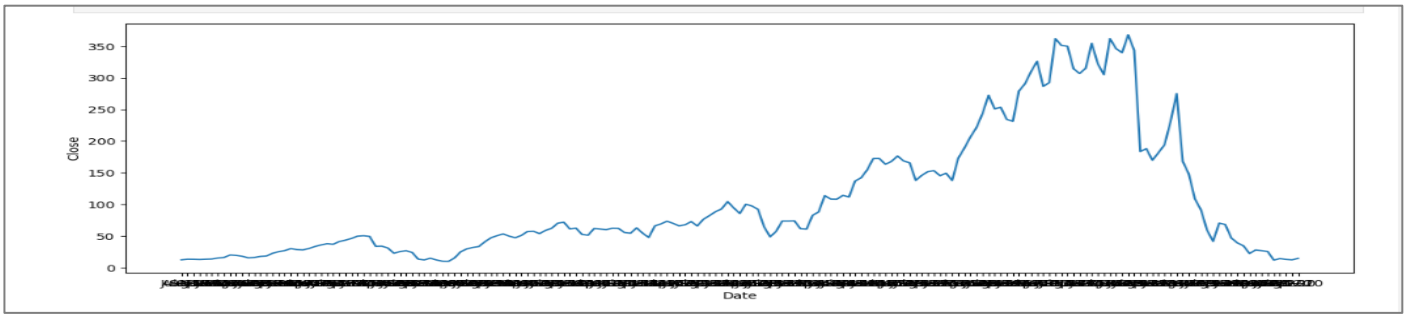
RESULTS

Visualization of Distribution of all Columns of Closing Price:



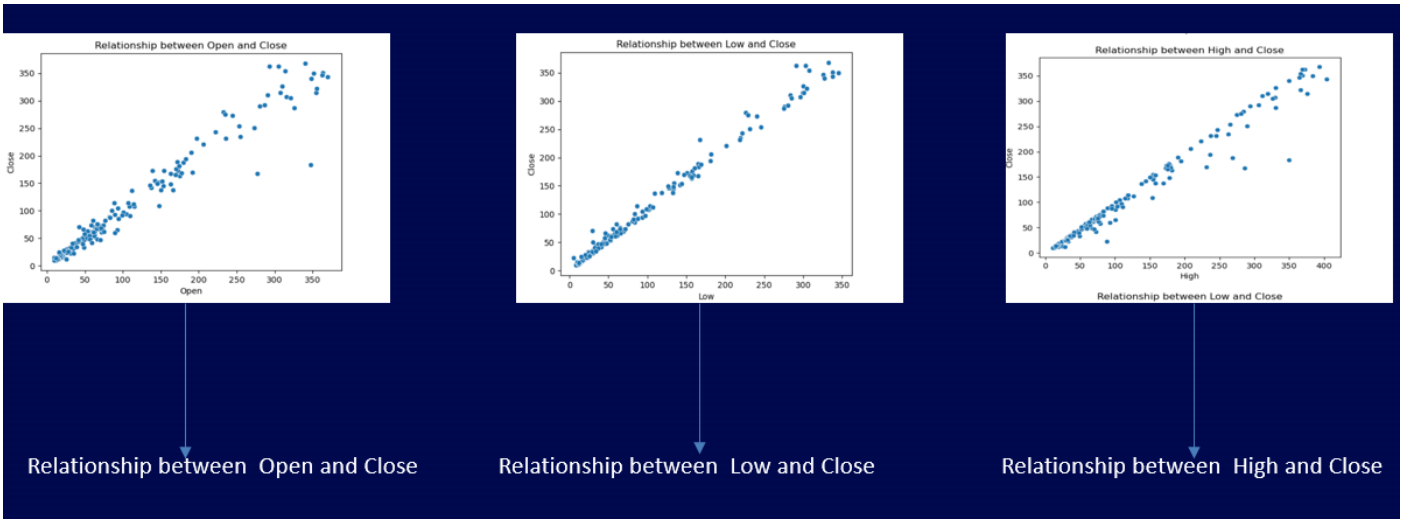
The above graph shows that they are not normally distributed. The mean and median should be equal for perfect normal distribution curve. So we log transform all the features to normal distribution.

Bivariate Analysis:



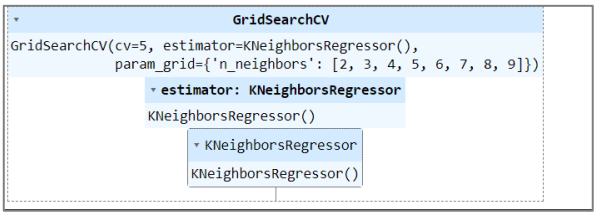
From the above line plot, We conclude that the stock price is keep on increasing till 2018.But after 2018, the stock price is kept on decreasing due the fraud case involving Rana Kapoor.

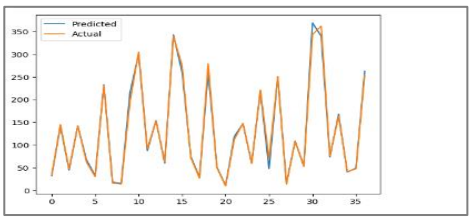
Visualization of Relationship of all Datasets of Closing Price



Correlation analysis:

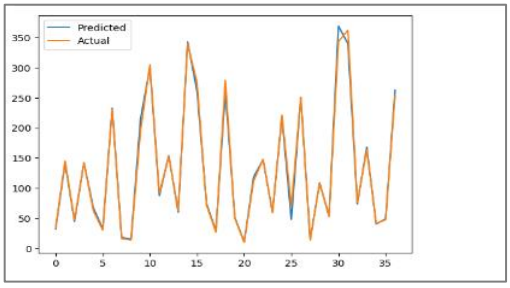
KNN:





Random forest :

```
GridSearchCV
GridSearchCV(cv=7, estimator=RandomForestRegressor(),
             param_grid={'criterion': ['squared_error', 'absolute_error',
                                       'friedman_mse', 'poisson'],
                         'max_features': ['sqrt', 'log2', None],
                         'n_estimators': [100, 200, 300]})
estimator: RandomForestRegressor
RandomForestRegressor()
RandomForestRegressor
RandomForestRegressor()
```



Linear Regression:

```
LinearRegression
LinearRegression()
```

Ridge Regression:

```
ridge =Ridge(alpha= 150)
ridge.fit(x_train,y_train)
```

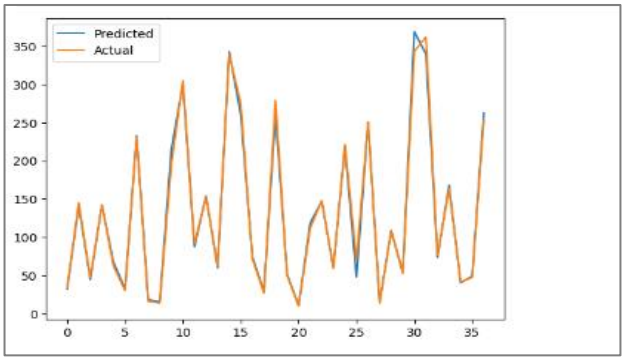
```
Ridge
Ridge(alpha=150)
```

Lasso Regression:

```
lasso =Lasso(alpha= 1.6)

lasso.fit(x_train,y_train)
```

```
Lasso
Lasso(alpha=1.6)
```



CONCLUSION

- The trend of the price of Yes Bank's stock increased until 2018 and then Close, Open, High, Low price decreased.
- Based on the open vs. close price graph, we concluded that Yes Bank's stock fell significantly after 2018.
- Both duplicate and null values are absent, as we have seen. But object data type values are available for the Date feature.
- The dependent and independent values were found to be linearly related.
- The data contained a significant amount of multicollinearity.
- Decision Tree regression Is best model for yes bank stock closing price data this model use for further prediction
- Visualization has allowed us to notice that the closing price of the stock has suddenly fallen starting in 2018. It seems reasonable that the Yes Bank stock price was significantly impacted by the Rana Kapoor case fraud.
- KNN performed the worst out of all.
- In this work, we create 4 regression models for our data:-

1. Linear Regression ● 2. Lasso Regression ● 3. Ridge Regression 4. Elastic Net Regression

- These four models gives us the following results: High, Low, Open are directly correlate with the closing price of the stocks.

REFERENCE

- https://www.kaggle.com/datasets/simranjain17/yes-bank-stock-prices
- https://www.youtube.com/watch?app=desktop&v=lx-dcCr0JCI
- https://www.indiapropertydekho.com/article/154/yes-bank-share-price-target