Yes Bank Stock Closing Price Prediction

Team- Data Minds
Team members- Uday Kant and Sonu Kumar

Abstract-

In this given project based on machine learning on the topic 'yes bank stock closing price prediction', we use linear regression to make a model which predicts the stock closing price. For this project, we had given the dataset of yes bank stock price for the year 2005-2020.

In this given dataset 185 rows and 5 features are present as Date, Open, High, Low, and Close price. Our target feature and also independent variable is closing price and the rest features such as open, low, and high are the dependent variables.

Keywords- Machine learning algorithm, Regression, Python.

Problem Statements:
1. Introduction
2. Data Summary
3. Analysis of data
4. Data cleaning
5. Data Preprocessing
6. Model Training
7. Challenges
8. Conclusion

Introduction :

To calculate the future value of the YES bank's stock on the national stock exchange by making a machine learning model of linear regression. The benefit of correctly predicting the price of a stock could lead to a small profit. According to the efficient market theory, stock prices should reflect all currently available data, and any value changes that are not based on recently discovered data will therefore be unpredictable. We have to create models that enable us to forecast future stock prices.

A. Data Descriptions

1. We have given the YES BANK stock price dataset which has the monthly stock price of the bank. In our dataset, we have given multiple variables like date, open, high, low, and close.
2. The column open and close depicts us the starting and final price at which the stock is traded on a particular month.
3. High and low represents the maximum and minimum price of the shape for the month.
4. Date- date describes the date of investment done i.e in which month and year investment occurs

Steps involved in this Project

Step 1:

In the first step, we wrote the python programming to find some of the results. We had different types of datasets in which there were different columns. So we extracted unique values. Then we wrote the programming to draw the graph.

Step 2:

In Data cleaning, we are importing DateTime so that we can convert the date into the proper format of a date. We have given the date in mmm-yy format then it converted in proper format of yyyy-mm-dd and the given date column has dtype as object converting it into date time format.

Step 3:

Data Visualization-In this step, we saw data visualization for the given dataset. We have analyzed the bivariate, variable, and statistical analysis.

Step 4:
In the last step, we trained our model. We distributed our dataset into two categories. 80 percent of the data were distributed for training and the rest 20 percent were distributed for the test dataset.
We used a Linear regression model for this project. A linear relationship between a dependent (y)(in our case is Close Price) and one or more independent (in our case Open, Low, high) variables, hence called linear regression.

Then we applied Lasso and Ridge regression for this dataset. In Lasso regression the R squared value for the test data was 99.7%. This is almost the same as in the score from the training dataset which proves that in a dataset we achieve the best fit model.
After applying this model to ridge regression implementing the best parameters and best R^2 score we have 99.78% for the Ridge regression model.

Conclusions :
- At first, we do the data wrangling and then data cleaning and after that, we do the EDA part.
- In the EDA part, we conclude from our dataset that
    - Stock close price decreased after the year 2018 it is mainly because of Rana Kapoor's case which hit the stock price badly.
    - The graph for Yes bank opening price and Yes bank closing price has the same result.
    - The point is that the stock price of the YES BANK falls down after the year 2018 and it is not beneficial for investors to invest their money.
    - From the scatter plot we can conclude that bivariate analysis shows a high correlation of close price with other features.
    - All histogram plots show that all are right-skewed.
    - From the heatmap, we can conclude that all the features show high correlation with each other.
- We implemented linear regression and the accuracy of our linear regression model is 99.78%.
- After that we visualize the performance of our linear regression model and the graph shows that we achieve the almost best fit model for our dataset.