Yes Bank Stock Closing Price Prediction

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

Abstract- This project on ' Yes bank Stock Closing price prediction' is using Machine learning Regression. In this project, we had given the dataset for Yes bank stock closing price for the year 2005-2020. We had 185 rows and 5 features. Date, open, close, high, and low were the 5 features. The feature close is our target feature.

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 determine the YES bank's stock's future value on the national stock exchange. The advantage of a successful prediction of a stock's future price could result in insignificant profit. The efficient market hypothesis recommends that stock costs mirror all right now accessible data and any value changes that are not founded on recently uncovered data subsequently are unpredictable. We have to build

models which help us to predict the future stock prices.

    1. Data Descriptions
        A. **Date-** It describes the date when investors buy or sell the yes bank stock.
        B. **Open**- This feature tells us the date on which the investment is done. We have not given the date but month and year of the investment is given.
        C. **High**- This feature tells us the maximum price of that time.
        D. **Low**- This feature tells us the maximum price of that time.
        E. **Close**- This feature tells the close price for the same day of the stock.

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 dataset in which there were different columns. So we extracted out the null and 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 in to proper format of date.We have given date in mmm-yy format then it converted in proper format of yyyy-mm-dd and 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 as 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 into ridge regression implementing the best parameters best R^2 score we have 99.78% for Ridge regression model.

Conclusions :

- At first we do the data wrangling and then data cleaning and after that we do the EDA part.
- In EDA part we conclude from our dataset that
    - Stock close price decreased after year 2018 it is mainly because of Rana Kapoor case and hitted 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 high correlation of close price with other features.
    - All histogram plots show that all are right skewed.
    - From heatmap we can conclude that all the features show high correlation between 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.