**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| **Suraj Kumar** ([surajkumar0892@gmail.com](mailto:surajkumar0892@gmail.com))   1. Data Wrangling: Introducing New Variables, Settling Date-Time, Cleaning data 2. Data Visualization: BarPlot, Scatter-Plot, Distribution of close price, Joint Plot, Line-Plot 3. Feature Engineering: Introduced new features, Introducing Dummy Variables, Correlation Analysis, Data Processing 4. Regression Models: Linear Regression, Lasso Regression, Ridge Regression, Decision Tree Regression.   **Shreya Ranjan** ([shreyasrivastav15@gmail.com](mailto:shreyasrivastav15@gmail.com))   1. Data Wrangling: Null values, Cleaning data, Introducing New Variables 2. Data Visualization: BoxPlot, DistPlot, Trend to close price, BarPlot, Line-Plot 3. Feature Engineering: Introduced new features, Introducing Dummy Variables, Correlation Analysis, Data Processing 4. Regression Models: Linear Regression, Lasso Regression, Ridge Regression, Decision Tree Regression. |
| **Please paste the GitHub Repo link.** |
| **Github Link**:- [**https://github.com/Shreyaranjan16/Yes-Bank-Stock-Analysis.git**](https://github.com/Shreyaranjan16/Yes-Bank-Stock-Analysis.git)  [**surajkumar089/YES-BANK-STOCK-CLOSING-PRICE-PREDICTION(github.com)**](https://github.com/surajkumar089/YES-BANK-STOCK-CLOSING-PRICE-PREDICTION) |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| The Dataset contains information about the YES BANK Stock Prediction.We have 185 rows and 5 columns in our Dataset. We are having columns- Date, Open, Close, High, Low.  In a first step, We have imported the dataset through pandas ’read\_csv’ then performed the Data wrangling over the raw data after that, we have divide the whole data set into various groups like feature engineering, Univariate analysis, Bivariate analysis and divide the model into various regressions.  As there is no null values in the dataset we didn't got a chance to replace or remove the null values. Also we have find out the outlier with the help of boxplot.  Then we have started with the EDA(exploratory data analysis), In which firstly we have gone through the opening and closing stock prices with the help of line graph for the last three years then we have drawn a scatter plot in which we have done yearly analysis of the opening and closing stock for all years.  After that we have done a graphical representation of the dependent and independent variables. And find out the relationship between dependent and independent variable.  Now, we have checked the correlation among each using the Heat map, there was a very high correlation among independent features which means high multicollinearity in our model. Due to the fact that each column is equally crucial for prediction, we are not deleting any columns.  After this we have divided our dataset into train and test data splitted into 70-30. Then, we fit our dataset into various models like Linear regression, Lasso regression, Ridge regression and Decision tree regression. And after analyzing the result from these regression model. The perfect fit model for this dataset is decision tree model. |