Capstone Project Submission
Yes Bank Stock Closing Price Prediction

Team Member's Name, Email and Contribution:
Rajat Chaudhary (rajat.25.chaudhary@gmail.com) 　1. Data Munging: Introducing New variables, Settling Date-Time 　2. Data Visualization: Distplot , Boxplot Scatter Plot, Trend of Close price, Distribution of Close price ,Heatmap 　3. Feature Engineering: Introduced new features ,Introducing Dummy Variables 　4. Regression Analysis: Linear Regression ,Lasso , Ridge, ElasticNet Anukriti Shakyawar (shakyawaranukriti@gmail.com) 　1. Data Munging: Feature Engineering 　2. Feature Engineering: Introduced new features ,Introducing Dummy Variables 　3. Data Visualization: Trend of Close price ,Distribution of Close price ,Heatmap 　4. Regression Analysis: Linear Regression ,Lasso, Ridge ,ElasticNet Raman Kumar (ramank445522@gmail.com) 　1. Correlation Analysis 　2. Data Visualization: JointPlot ,Distplot 　3. Regression Analysis: Linear Regression ,Lasso , Ridge Deepmala Srivastava (srvdeepmala@gmail.com) 　1. Data Visualization :Barplot ,Scatter-Plot 　2. Regression Analysis :Linear Regression , Lasso ,Ridge ,ElasticNet
Github Link:- https://github.com/anukriti2602/Yes-Bank-Stock-Closing-Price-Prediction
Our objective is to build a regression model that can predict the close price of next month. First Step was to import the dataset through Pandas 'read_csv' then data wrangling and feature engineering in our dataset. We did not get into the situation to remove NA values because there are 0 null values in the Yes Bank dataset. Next, EDA(exploratory Data Analysis) in which trends of stock closing price,

distribution of dependent variables have been examined. Plotted histogram of all variables with mean and median(Axline) to check measures of central tendency is close to each other or far. Then log transformation has been applied on each variable, which led to a conclusion: to normalize right-skewed data perform log transformation.

Now, the correlation has been checked among each other through heatmap, there was a very high correlation among independent features means high multicollinearity in our model, so to check how high multicollinearity is VIF(Variation Inflation Factor) has been checked based on VIF, three features had to drop from the dataset to prevent the wrong prediction.

Introduced Dummy variables with year column and with these dummy variables total independent variables became 17. By applying log transformation on close price and z-score on all 17 independent variables passing it to the next step which is to train models.

Prepared independent and dependent variables for the train test split method.

Applied Linear model, Ridge regression, Lasso regression and ElasticNet all the models are performing in a better way but Linear Model and Lasso are performing in a better way in comparison to Ridge and ElasticNet but Ridge regression and ElasticNet performance improved by applying cross-validation and Hyperparameter tuning.

Conclusions:
- **Target Variable is strongly dependent on Independent Variables.**
- **Linear Regression and Lasso are performing better than other models with training accuracy *94.0359% and 94.45777% respectively.**
- **Apart from Linear Regression and Lasso, Ridge and Elastic Net is also performing better but they have less training accuracy.**
- **Ridge and ElasticNet are performing far much better after Applying HyperparameterTuning and Cross-validation.**
- **R2 and Adjusted R2 are around 95% and 91% in each model**