Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1) **Mohd Danish:** **Email:** mdanish63364@gmail.com
   1) Feature Engineering:
      - Introduced new features
      - Introducing Dummy Variables
   2) Data Visualization:
      - Trend of Close price
      - Distribution of Close price
      - Heatmap
   3) VIF
   4) Regression Analysis:
      - Linear Regression
      - Lasso
      - Ridge
      - ElasticNet
   5) Group Colab
2) **Abdul Rahman Talha:** **Email:** rahman88talha@gmail.com
   1) Correlation Analysis
      - Between Independent Variables
      - Between Dependent and Independent Variables
   2) Data Visualization:
      - JointPlot
      - Distplot
   3) Regression Analysis:
      - Linear Regression
      - Lasso
      - Ridge
   4) PPT
3) **Huzaifa Khan:** **Email:** huzaifakhan2974@gmail.com
   1) Data Munging:
      - Introducing New variables
      - Settling DateTime
   2) Data Visualization:
      - Distplot
      - Boxplot
      - Scatter Plot
   3) Regression Analysis:
      - Linear Regression
      - Lasso
      - Ridge
      - ElasticNet

4) **Arbaz Malik:** **Email:** malikarbaaz267@gmail.com

1) Data Munging:
   - Feature Engineering
2) Data Visualization
   - Barplot
   - Scatter-Plot
3) Regression Analysis
   - Linear Regression
   - Lasso
   - Ridge
   - ElasticNet

Please paste the GitHub Repo link.

Github Link:- Mdanish2020/Yes_Bank_Stock_closing_Price_Prediction (github.com)

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Who would have not heard about Yes Bank and Rana Kapoor's fraud when he invested 3700 crores in short-term debentures in DHFL, when news came out in the market and fraud revealed all the trusted shareholders sold their stakes and the stock price has fallen from 391 to 175.
Our objective is to build a regression model that can predict the close price of next month.

Our 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 trend of stock closing price, distribution of dependent variable 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's and ElasticNet's 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 Hyperparameter Tuning and Cross-validation.
- R2 and Adjusted R2 are around 95% and 91% in each model.