**Capstone Project -Regression**

|  |
| --- |
| **Team Member’s Name, Email and Contribution: Individual** |
| **Name : Rishi Chaturvedi Email:rishichaturvedi012@gmail.com**     1. Data Cleaning :-    * Dealing with null values, duplicate data and outliers present in our data. 2. Exploratory Data Analysis :-    * Plotting the dependent variable and distributions of dependent and independent variables.    * Checking and visualizing the correlation between our dependent and independent variables.    * Visualizing the relationship between each pair of our variables. 3. Data Preprocessing & Feature Engineering :-    * Checking for and Dealing with multicollinearity present in our dataset.    * Applying the log transform to deal with positively skewed data.    * Scaling the data and splitting it into train and test sets. 4. Model Implementation :-    * Fitting various models on our data and optimizing them via cross-validation.    * Using these models to make predictions on test and train data. The Models implemented are :-      1. Linear Regression      2. Lasso Regression      3. Ridge Regression      4. Elastic Net Regression 5. Data Visualization :-    * Using several kinds of charts like Line chart, scatter plot, heat map, pair plot, distplot, box plot etc.  to better visualize data and understand correlation and trends. 6. Model performance comparison :-    * Comparison of all implemented models using various Regression evaluation metrics like Mean absolute error, Mean squared error, RMSE, R-squared and Adjusted R-squared. 7. Conclusion :-    * Drawing some insights from the data and the predictions made by our various predictive models on unseen (test) data. |
| **GitHub Repo links.** |
| Github Link:-  Drive folder link :- |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  **Problem statement :-**  Yes Bank is a well-known Indian bank headquartered in Mumbai, India and was founded by ***Rana Kapoor***and **Ashok Kapoor** in **2004**. It offers a wide range of differentiated products for corporate and retail customers through retail banking and asset management services.  Yes Bank is a publicly traded company listed on the stock market and is therefore subject to the ups and downs of the stock market cycle.  The stock market is driven by speculation. The investors decide on buying or selling shares of a company based on its performance and its reputation. Public opinion has a huge impact on stock market prices.  Which is why when the news of the fraud case involving Rana Kapoor broke in 2018, the stock price of Yes bank went down significantly.  Here we are presented with the stock market price data of Yes bank and our job is to try and predict stock’s closing price of the month. This data contains the date, lowest, highest and closing price details.  Our approach is to fit a machine learning model on this past data and try to predict the closing price for new unseen data using the parameters learned during training.  This way, we can get our model to learn the trends present in the data during training and use that information during prediction.  We will apply various Regression Models for this task such as: Linear Regression, Lasso Regression, Ridge Regression, and Elastic Net Regression.  **Conclusions :-**     * The popularity of stock closing is growing extremely rapidly day by day which encourages researchers to find new methods if any fraud happens. * This technique used for prediction is not only helpful to researchers to predict future stock closing prices or any fraud happens or not but also helps investors or any person who is dealing with the stock market in order to predict the model with good accuracy. * In this work we use linear regression technique, lasso regression, ridge regression, elastic net regression and XGBoost Regression technique. these five models gives us the following results * High, low, open are directly correlate with the closing price of stocks * Target variable(dependent variable) strongly dependent on independent variables   Xgboost regression is best model for yes bank stock closing price data this model use for further prediction |
|  |