Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name : NISARGA C **Email: nisargac5577@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, heatmap, pair plot, distplot, boxplot 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.

Please paste the GitHub Repo link.

Github Link:- https://github.com/Nisargaavi/Yes-Bank-Stock-Closing-Price-Prediction

Drive folder link :-
https://drive.google.com/drive/folders/1ByC6U1J71Ij_QxH8GWKAsZx53dZ5EOUS?usp=share_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 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 fraud case involving Rana Kapoor broke in 2018, 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, Elastic Net Regression.

Conclusions :-

- Using data visualization on our target variable, we can clearly see the impact of 2018 fraud case involving Rana Kapoor as the stock prices decline dramatically during that period.
- There is a high correlation between the dependent and independent variables. This is a signal that our dependent variable is highly dependent on our features and can be predicted accurately from them.
- We implemented several models on our dataset in order to be able to predict the closing price and f ound that Elastic Net regressor is the best performing model with Adjusted R2 score value of 0.99 32 and it scores well on all evaluation metrics**.**
- All of the implemented models performed quite well on our data giving us the accuracy of over 99 %.
- We checked for presence of Heterodasceticity in our dataset by plotting the residuals against the Elastic Net model predicted value and found that there is no Heterodasceticity present. Our model is performing well on all data-points.
- With our model making predictions with such high accuracy even on unseen test data , we can confidently deploy this model for further predictive tasks using future real data.
- There are some outliers in our features however this being a very small dataset, dropping those inst ances will lead to loss of information.
- We found that there is a rather high correlation between our independent variables. This multicolli nearity however is unavoidable here as the dataset is very small.
- We found that the distribution of all our variables is positively skewed. so we performed log transformation on them.

……..