College Application

By Group 57:
Shivam Patel : 20012448
Darsh Patel : 20012159
Anurag Athawale : 20012263

A.) Mean Square Error Training and Testing:

- ❖ The Mean Squared Error measures how close a regression line is to a set of data points. It is a risk function corresponding to the expected value of the squared error loss. Mean square error is calculated by taking the average, specifically the mean, of errors squared from data as it relates to a function.

- Training: Linear Regression MSE = 0.0041742938784538304
  Linear Regression PCA MSE = 0.004244813045478053

- TestIng: Linear Regression MSE = 0.003967271299117479
  Linear Regression PCA MSE = 0.003608167795482937

- Scores: Linear Regression Model Score = 0.7640600485218347
  Linear Regression PCA Model Score = 0.8420154227358346

B.) Salient Features :

- We chose linear regression model because our task was to find/ predict the value given some features and SVM is used for classification problem.

- We implemented models using different combinations of columns from the given data set and choose the features with highest impact on the model.

C.) Analysis :
- Using Standard Scalar of Sklearn we have scaled our data.

- We have used 80% of the data for training and 20% for testing for both the models.

- After training linear regression model we got 0.00417 MSE of training and 0.0039 mse of testing with model score of 0.764 which is great. However, performing PCA and using 2 principle components we increased our model score to 0.842 while our testing error reduced to 0.0036 which is excellent. (Training error remain almost same)
- After seeing both model scores we can conclude that dimensionality reduction using PCA greatly improves our model performance. Following are the scatter plot for the LR and PCA using LR model.