**CS556 Mathematical Foundations of Machine Learning**

**Final Project Report**

CWID’s: **20011577** – Sai Bandla, **20013368** - Sai Ritwik Reddy Kosi Reddy, **20012150** -Sabitha Rachel Nazareth

**Project Title: College Admissions Predictor Using Linear Regression**

Mean Square Error for Linear Regression Model: **0.002249**

Mean Square Error for Linear Regression Model after PCA: **0.002249**

**Salient Project Features:**

From the dataset given to us. We have concluded that the features we must consider while building the model are: **1. GRE Score 2. TOEFL Score 3. University Rating 4. SOP 5. LOR 6. CGPA 7. Research** and the target variable we are trying to predict accurately is “**Chance of Admit**”

* We Analyzed the data, checked for null values and duplicate values and made sure that the data set is clean.
* We analyzed the features individually and checked for outliers.
* We didn’t eliminate any features from the given dataset because after checking the correlation of each attribute to the target variable we can confirm that disregarding any feature will lead to a decrease in the model’s efficiency.

Table

Description automatically generated

* We first split the data into training set(80%) and test set(20%) and fed the data into the linear regression model(from sklearn.linear\_model import LinearRegression) as shown in our code.
* Before scaling the data, we get a model accuracy of 88.4% and after scaling(using StandardScaler.Fit\_Transform()) we get an improved accuracy of 88.75%
* When we try to drop “Research”(which has least correlation) feature from the data frame accuracy decreased to 88.33% thus we concluded it was not helping the model.
* Next, we performed dimensionality reduction using principle component analysis(from sklearn.decomposition import PCA ,PCA(n\_components=2) ) and we have gotten same accuracy of 88.75%.
* We performed regression analysis using reg plot.