# College Admission

Project 6

DESCRIPTION

**Background and Objective:**
Every year thousands of applications are being submitted by international students for admission in colleges of the USA. It becomes an iterative task for the Education Department to know the total number of applications received and then compare that data with the total number of applications successfully accepted and visas processed. Hence to make the entire process easy, the education department in the US analyze the factors that influence the admission of a student into colleges. The objective of this exercise is to analyse the same.

**Domain:** Education

**Dataset Description:**

| Attribute | Description |
| --- | --- |
| GRE | Graduate Record Exam Scores |
| GPA | Grade Point Average |
| Rank | It refers to the prestige of the undergraduate institution. The variable rank takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. |
| Admit | It is a response variable; admit/don't admit is a binary variable where 1 indicates that student is admitted and 0 indicates that student is not admitted. |
| SES | SES refers to socioeconomic status: 1 - low, 2 - medium, 3 - high. |
| Gender_male | Gender_male (0, 1) = 0 -> Female, 1 -> Male |
| Race | Race – 1, 2, and 3 represent Hispanic, Asian, and African-American |

**Analysis Tasks:** Analyze the historical data and determine the key drivers for admission.

**Predictive:**

- Find the missing values. (if any, perform missing value treatment)
- Find outliers (if any, then perform outlier treatment)
- Find the structure of the data set and if required, transform the numeric data type to factor and vice-versa.
- Find whether the data is normally distributed or not. Use the plot to determine the same.
- Normalize the data if not normally distributed.
- Use variable reduction techniques to identify significant variables.
- Run logistic model to determine the factors that influence the admission process of a student (Drop insignificant variables)
- Calculate the accuracy of the model and run validation techniques.

- Try other modelling techniques like decision tree and SVM and select a champion model
- Determine the accuracy rates for each kind of model
- Select the most accurate model
- Identify other Machine learning or statistical techniques

**Descriptive:**
Categorize the average of grade point into High, Medium, and Low (with admission probability percentages) and plot it on a point chart.
Cross grid for admission variables with GRE Categorization is shown below:

| GRE | Categorized |
|---------|-------------|
| 0-440 | Low |
| 440-580 | Medium |
| 580+ | High |