

**Name: Adhija Bachhav**

**Roll No: 281053**

**PRN: 22311656**

**Batch: A-3**

## **ML Assignment 7**

### **Problem Statement:**

Assignment on Classification technique Every year many students give the GRE exam to get admission in foreign Universities. The data set contains GRE Scores (out of 340), TOEFL Scores (out of 120), University Rating (out of 5), Statement of Purpose strength (out of 5), Letter of Recommendation strength (out of 5), Undergraduate GPA (out of 10), Research Experience (0=no, 1=yes), Admitted (0=no, 1=yes). Admitted is the target variable.

Data Set: <https://www.kaggle.com/mohansacharya/graduate-admissions>

The counsellor of the firm is supposed check whether the student will get an admission or not based on his/her GRE score and Academic Score. So, to help the counsellor to take appropriate decisions build a machine learning model classifier using Decision tree to predict whether a student will get admission or not.

- a) Apply Data pre-processing (Label Encoding, Data Transformation....) techniques if necessary.
- b) Perform data-preparation (Train-Test Split)
- c) Apply Machine Learning Algorithm
- d) Evaluate Model.

### **Introduction:**

Every year, numerous students take the GRE (Graduate Record Examination) to seek admission into foreign universities. However, predicting a student's admission can be challenging due to various factors like GRE scores, TOEFL scores, university ratings, and academic performance.

To assist counsellors in making informed decisions, we aim to develop a **Machine Learning-based classification model** that predicts whether a student will be admitted based on their **GRE scores and Academic Scores**. The dataset includes attributes such as TOEFL scores, University Rating, Statement of Purpose (SOP) strength, Letter of Recommendation (LOR) strength, Undergraduate GPA, and Research Experience.

This assignment uses a **Decision Tree Classifier** to predict student admissions, ensuring interpretability and ease of decision-making.

### **Objective**

The primary objective of this project is to **build a Decision Tree classification model** that predicts whether a student will be admitted (1) or not (0) based on their GRE scores and academic performance. The model will help **counsellors** make quick and reliable admission decisions.

The key steps involved in this project are:

- **Data Preprocessing:** Handling missing values, encoding categorical variables, and feature selection.

- **Data Splitting:** Dividing the dataset into training and testing sets.
- **Model Training:** Applying the Decision Tree algorithm.
- **Model Evaluation:** Assessing accuracy using metrics like Confusion Matrix, Precision, Recall, and F1-score.

## Theory

### 1. Decision Tree Algorithm

A **Decision Tree** is a supervised learning algorithm used for classification and regression tasks. It works by **splitting** the data into subsets based on feature values using conditions in the form of **if-else rules**.

#### Working of Decision Tree:

1. **Choosing the Best Feature:** The dataset is divided based on a feature that maximizes information gain or minimizes impurity. Common measures used:
  - **Gini Impurity:** Measures how often a randomly chosen element would be incorrectly classified.
  - **Entropy (Information Gain):** Measures the uncertainty in the dataset.
2. **Splitting the Data:** The algorithm repeatedly partitions the data until all the samples belong to a single class or further splitting does not add value.
3. **Decision Making:** The model follows the learned tree structure to classify new instances based on feature values.

### 2. Data Preprocessing & Feature Engineering

- **Creating the Target Variable:** Since the dataset has Chance of Admit (ranging from 0 to 1) instead of Admitted (0/1), we create a new column:
  - **Admitted = 1** if Chance of Admit > 0.7, else **Admitted = 0**.
- **Feature Selection:** Since the counsellor focuses on **GRE Score and Academic Performance**, we select:
  - GRE Score
  - Undergraduate GPA
- **Train-Test Split:** The dataset is split into **80% training** and **20% testing**.

### 3. Model Training & Evaluation

- The Decision Tree model is trained using the **Gini Impurity criterion**.
- Evaluation is done using:
  - **Accuracy** =  $(TP + TN) / \text{Total Samples}$
  - **Confusion Matrix** (True Positives, False Positives, etc.)
  - **Precision, Recall, F1-score**

## **Conclusion**

In this assignment we successfully built a **Decision Tree classifier** to predict student admissions. The model was trained using GRE scores and academic performance data. The results show that a Decision Tree is effective for admission prediction, providing interpretable rules for decision-making.