**Name:** Niraj Patil

**Roll no:** 281003

**Batch:** A1

**Assignment 7**

**Statement:**

Q. 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 counselor 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 counselor 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.

**Objective:**

1. Build a classification model using a Decision Tree to predict student admission chances.
2. Implement preprocessing steps for clean and usable data.
3. Evaluate the model for accuracy and prediction reliability.

**Resources Used:**

* Software used: Google Colab
* Libraries used: Pandas, Scikit-learn, Matplotlib, Seaborn

**Introduction to Classification:** Classification is a supervised machine learning technique used to categorize data into predefined labels. In this assignment, we aim to classify whether a student gets admitted (Yes/No) based on certain features using a Decision Tree model.

**Methodology:**

1. **Data Collection and Preprocessing:**
   * Load the dataset from Kaggle.
   * Check for missing values, duplicates, and outliers.
   * Perform Label Encoding if required (e.g., Research column).
   * Normalize or standardize features if necessary.
2. **Data Preparation:**
   * Select relevant features (GRE Score, GPA, etc.).
   * Split the dataset into training and testing subsets using an 80-20 or 70-30 ratio.
3. **Model Building:**
   * Apply the Decision Tree Classifier from Scikit-learn.
   * Train the model using the training dataset.
   * Predict the results for the test dataset.
4. **Model Evaluation:**
   * Use metrics such as Accuracy, Precision, Recall, and F1-Score to evaluate the model.
   * Generate a confusion matrix to assess the classification performance.

**Advantages of Classification using Decision Tree:**

1. Easy to interpret and visualize.
2. Requires little data preparation.
3. Handles both numerical and categorical data.

**Disadvantages:**

1. Prone to overfitting if not pruned properly.
2. Can be biased with imbalanced datasets.

**Conclusion:**

In this assignment, we successfully built a classification model using a Decision Tree to predict student admissions. With proper data preprocessing and evaluation metrics, the model helps counselors make informed decisions based on students’ GRE scores and academic performance. This technique serves as a practical example of how classification can be applied in educational data mining.