Name: Rugved Sawant
Roll No: 281065
Batch: A3

Assignment 7

Statement

Q. This assignment involves building a **classification model** to predict whether a student will get **admission** based on GRE score and academic profile. The dataset includes features such as GRE Scores, TOEFL Scores, University Rating, SOP & LOR Strength, GPA, Research Experience, and Admission Status (Target: 0 or 1).

Dataset Link:
Graduate Admissions Dataset – Kaggle

A **counselor** needs assistance in predicting admission chances using GRE score and academic performance. To support this decision-making, a **Decision Tree classifier** will be applied.

Tasks:

a) Apply **data preprocessing** (Label Encoding, Transformation, etc.)
b) Perform **train-test splitting**
c) Implement **Decision Tree** classifier
d) **Evaluate** the model performance

Objective

1. Build a **Decision Tree** classification model to predict admission chances.

2. Apply necessary **data preprocessing** for accurate modeling.

3. Evaluate model using standard **classification metrics**.

Resources Used

- **Software**: Google Colab

- **Libraries**: Pandas, Scikit-learn, Matplotlib, Seaborn

Introduction to Classification

Classification is a **supervised machine learning** approach where input data is assigned to a category or class label. In this task, we classify whether a student gets **admitted (Yes/No)** based on their academic indicators.

Methodology

1. **Data Collection & Preprocessing**

    o   Load dataset from Kaggle

    o   Check and handle missing values, duplicates, and outliers

    o   Perform **Label Encoding** for categorical variables like Research

    o   Normalize/scale features if needed

2. **Data Preparation**

   o Select important features: GRE Score, TOEFL Score, University Rating, GPA, Research

   o Define Admitted as the **target**

   o Use **train_test_split** (80:20 or 70:30)

3. **Model Building**

   o Apply DecisionTreeClassifier from sklearn.tree

   o Fit the model on training data

   o Predict on test data

4. **Model Evaluation**

   o Evaluate using **Accuracy**, **Precision**, **Recall**, **F1-Score**

   o Generate and interpret a **Confusion Matrix**

Advantages of Decision Tree Classification

- Easy to understand and **visualize**

- Handles both **categorical and numerical** data

- Requires **minimal data preprocessing**

Disadvantages

- Can **overfit** if not pruned

- May perform poorly with **imbalanced** data

Conclusion

In this assignment, a **Decision Tree classifier** was successfully developed to predict student admissions. With proper preprocessing and evaluation, the model can assist counselors in making reliable decisions based on a student's GRE score and academic background. This showcases the power of **classification techniques** in the field of **educational data mining**.

Output