**Name:** Chandrakant Dattatrey Thakare

**Roll no:** 282014  **PRN:** 22310303

**Class :** SY CSE AI**Batch:** B1

**Assignment 7: Graduate Admissions Prediction using Classification**

**Problem Statement**

This assignment involves working on a dataset related to **Graduate Admissions**, where the goal is to help a counselor predict whether a student will be admitted to a university based on various academic parameters.

**Dataset Link**:  
<https://www.kaggle.com/mohansacharya/graduate-admissions>

The dataset contains the following features:

* **GRE Score** (out of 340)
* **TOEFL Score** (out of 120)
* **University Rating** (out of 5)
* **SOP Strength** (out of 5)
* **LOR Strength** (out of 5)
* **UG CGPA** (out of 10)
* **Research Experience** (0 = No, 1 = Yes)
* **Admitted** (0 = No, 1 = Yes) — **Target Variable**

The task is to use **Decision Tree Classification** to predict admission chances.

**Objective**

* Understand the classification technique using **Decision Trees**
* Preprocess and split data for training and testing
* Train a model to predict student admission
* Evaluate model performance using classification metrics

**Tools and Resources**

* **Software Used**: Google Colab / Jupyter Notebook
* **Libraries Used**: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

**Key Functions and Concepts Used**

* pd.read\_csv() – Load dataset
* LabelEncoder() – Encode categorical values
* train\_test\_split() – Split dataset into training and testing sets
* DecisionTreeClassifier() – Build classification model
* classification\_report(), confusion\_matrix(), accuracy\_score() – Evaluate model

**Methodology**

**1. Data Loading and Exploration**

* Load the dataset using **Pandas**
* Preview dataset structure and check for null values
* Understand the target variable: Admitted (binary classification)

**2. Data Preprocessing**

* Apply **Label Encoding** to categorical features (if any)
* Normalize or scale features if needed
* Check for any skewed or missing data and handle accordingly

**3. Data Preparation**

* Select relevant features:  
  Example: GRE Score, TOEFL Score, University Rating, CGPA, Research
* Define input (X) and output (y) variables
* Split data into **train** and **test sets** using train\_test\_split() function

**4. Model Building**

* Use **DecisionTreeClassifier()** from Scikit-learn
* Train the classifier using the training set
* Predict outcomes for the test set

**5. Model Evaluation**

* Evaluate model performance using:
  + **Confusion Matrix**
  + **Classification Report** (Precision, Recall, F1-score)
  + **Accuracy Score**

**Advantages of Decision Tree Classification**

* Easy to interpret and visualize
* Handles both numerical and categorical data
* Non-linear relationships can be modeled well
* Useful for explaining decision logic to non-technical users (like counselors)

**Challenges**

* Prone to **overfitting**, especially on small or noisy datasets
* Performance can vary based on the depth of the tree
* Needs proper **pruning** or **hyperparameter tuning** for best results

**Conclusion**

This assignment provided a practical understanding of **classification techniques** using **Decision Trees**. By completing this assignment, I:

* Learned how to preprocess and split academic data
* Built a machine learning model to predict student admissions
* Evaluated the model using appropriate metrics
* Understood how to support human decisions with machine predictions

This forms a strong foundation for more complex classifiers such as **Random Forests**, **SVM**, or **Logistic Regression**.