

# **ASSIGNMENT NO: 7**

## **Problem Statement-**

To implement Classification techniques for following scenario:

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.

The counselor of the firm is supposed to 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 a 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.

## **S/W Packages and Libraries used-**

Software Package: Python

Libraries Used:

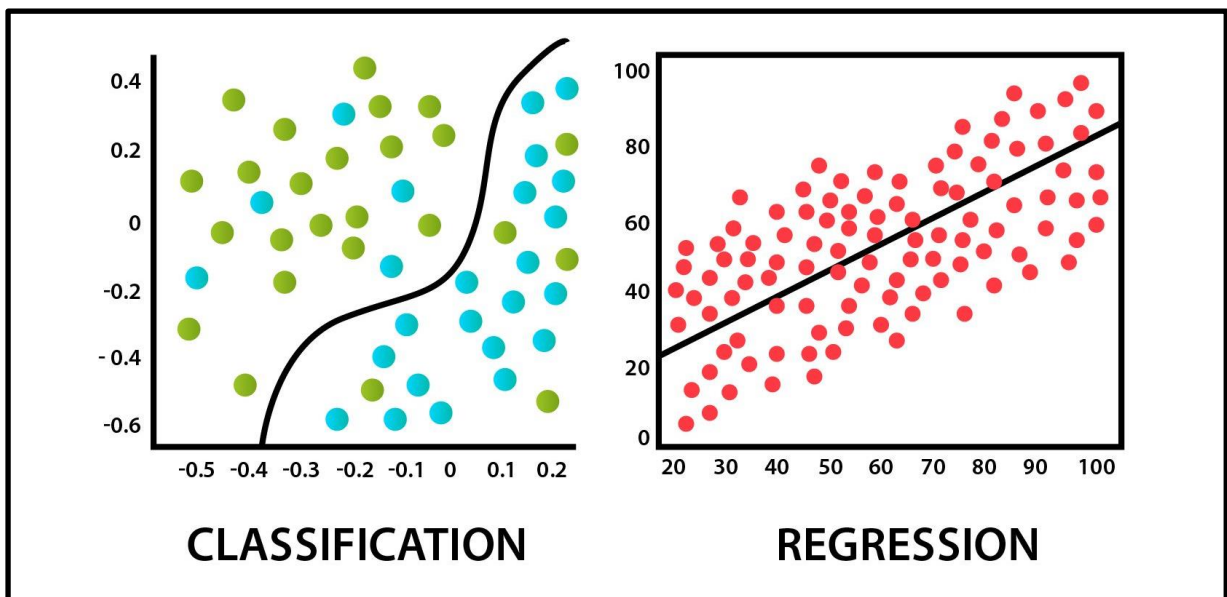
pandas: For data manipulation and analysis.

scikit-learn: For implementing machine learning algorithms, including Decision Tree Classifier.

matplotlib: For data visualization.

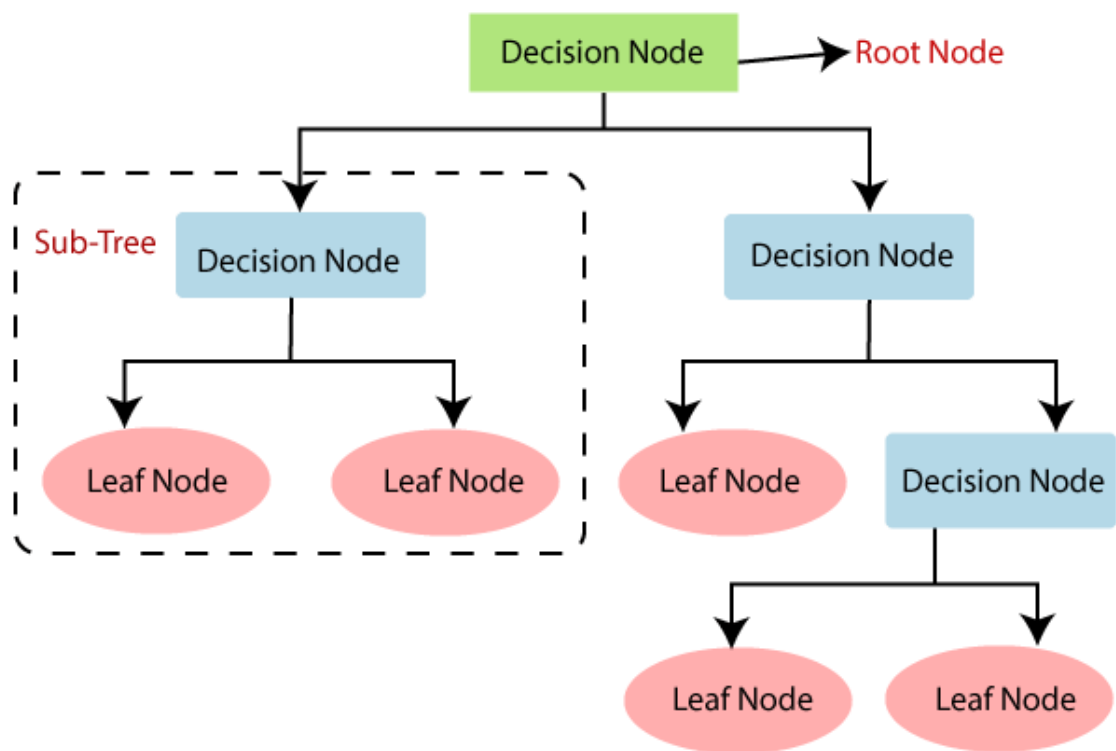
## Theory-

- Classification:
  - Classification is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data.
  - The process starts with predicting the class of given data points. The classes are often referred to as targets, labels, or categories.



- What is a Decision Tree?
  - It uses a flowchart like a tree structure to show the predictions that result from a series of feature-based splits.
  - It starts with a root node and ends with a decision made by leaves.

- Root Nodes – It is the node present at the beginning of a decision tree. From this node, the population starts dividing according to various features.
- Decision Nodes – the nodes we get after splitting the root nodes are called Decision Node
- Leaf Nodes – the nodes where further splitting is not possible are called leaf nodes or terminal nodes.
- Sub-tree – just like a small portion of a graph is called a sub-graph similarly a subsection of this decision tree is called a sub-tree.



---

## Methodology-

- Data Preprocessing:
  - Load the data.
  - Perform any necessary data transformations (e.g., label encoding for categorical variables).
  - Check for missing values and handle them if necessary.
- Data Preparation (Train-Test Split):
  - Split the data into training and testing sets to train the model on one set and evaluate its performance on another.
- Model Training:
  - Initialize a Decision Tree Classifier model.
  - Fit the model on the training data.
- Model Evaluation:
  - Predict admission outcomes for the test data.
  - Evaluate the model's performance using metrics like accuracy, precision, recall, and F1-score.
  - Visualize the results if necessary.

## **Applications-**

- The application of this classification technique is to predict whether a student will get admission to a foreign university based on their GRE score and academic performance.
- It can be used by education consultants, university admission offices, and students themselves to assess their chances of admission and make informed decisions.

## **Limitations-**

- Decision Tree models are prone to overfitting, especially on noisy data or data with a large number of features.
- They may not capture complex relationships between variables as effectively as other algorithms.
- Decision Trees are sensitive to small variations in the data, which can lead to different tree structures and results.
- The interpretability of the model might be limited if the tree becomes too large and complex.

## **Conclusion-**

- The Decision Tree Classifier can be used effectively to predict admission outcomes based on GRE scores and academic performance.
- It provides a simple and interpretable model that can help counselors and students make informed decisions about university admissions.
- However, it's important to be aware of its limitations and consider other machine learning techniques for more accurate predictions, especially in complex scenarios with a large number of features.