

### **Problem statement:**

The task at hand involves developing a predictive model to assist a counselor in determining whether a student will be admitted to foreign universities based on their GRE scores and academic performance. The dataset provided includes various features such as GRE scores, TOEFL scores, university rating, statement of purpose strength, letter of recommendation strength, undergraduate GPA, research experience, and admission status (admitted or not). To address this, the counselor seeks a machine learning model classifier, utilizing a Decision Tree algorithm, to predict the likelihood of admission for a given student.

### **Objective:**

The objective of this project is to develop a Decision Tree classifier to assist counselors in predicting student admissions to foreign universities based on their GRE scores and academic performance. By leveraging machine learning techniques, we aim to streamline the admissions process, providing counselors with a reliable tool to make informed decisions efficiently.

### **Software used:**

1. Python 3.x
2. Google Colab

**Libraries and packages used:** Pandas, sklearn, NumPy, Matplotlib

### **Methodology:**

**Classification:** Classification is the process of organizing a dataset into classes or categories. This can be applied to both structured and unstructured data, with the goal of predicting the class of given data points based on their features.

**Decision Tree:** A Decision Tree is a predictive model that utilizes a tree-like structure to represent decisions and their potential consequences. It starts with a root node and branches out into decision nodes, which are then further split into leaf nodes. Each node represents a decision based on a feature, and the leaves represent the final outcome or prediction.

**Entropy:** Entropy is a measure of the randomness or impurity in a dataset. In the context of Decision Trees, entropy is used to quantify the homogeneity of a sample. A sample with low entropy is more homogeneous, while a sample with high entropy is more diverse.

### **Constructing a Decision Tree:**

1. Calculate the entropy of the target variable.
2. Split the dataset based on different attributes and calculate the entropy for each branch.
3. Calculate the information gain for each attribute and select the attribute with the highest information gain as the decision node.
4. Continue this process recursively until all data is classified, with branches either becoming leaf nodes or further split.

**Pruning:** Pruning is a technique used to prevent overfitting in Decision Trees by removing nodes or branches that are not significant. It helps improve the performance of the tree by eliminating unnecessary complexity. Pruning can be done during tree construction (pre-pruning) or after the tree is built (post-pruning).

we will utilize these principles of Decision Trees, including entropy calculation, information gain, tree construction, pruning, and transformation to decision rules, to develop a predictive model for determining student admissions to foreign universities based on GRE scores and academic performance.

### **Advantages:**

- 1.Interpretability:** Decision trees offer a clear and intuitive representation of decision-making processes, making them easy to understand and interpret, even for non-technical users.
- 2.Simple Implementation:** Implementing decision trees requires minimal data preprocessing compared to other machine learning algorithms. They do not require feature scaling or transformation, simplifying the overall implementation process.

**3. Handling Non-linear Relationships:** Decision trees can effectively model complex, non-linear relationships between features and the target variable without the need for explicit feature engineering. This flexibility allows them to capture intricate patterns in the data.

**4. Feature Importance:** Decision trees provide insights into the importance of features for prediction. By analyzing the structure of the tree, users can identify which features have the most significant impact on the model's predictions, aiding in feature selection and interpretation.

**5. Versatility:** Decision trees can be used for both classification and regression tasks, offering a versatile solution to a wide range of predictive modeling problems. This adaptability makes them valuable tools in various domains and applications.

### **Disadvantages:**

**1. Overfitting:** Decision trees are prone to overfitting, especially with complex datasets or inadequate pruning techniques.

**2. Instability:** Small variations in the data can result in significantly different decision trees, making them less robust.

**3. Lack of Continuity:** Decision trees create disjoint regions in the feature space, leading to abrupt decision boundaries.

**4. Difficulty in Capturing Linear Relationships:** Decision trees struggle to capture linear relationships between features and the target variable, requiring complex structures or ensemble methods.

**5. Limited Handling of Missing Data:** Traditional decision tree algorithms may struggle with missing data, requiring preprocessing techniques that could introduce bias.

### **Applications with example:**

1. Credit Risk Assessment:

Example: A bank uses decision trees to assess the creditworthiness of loan applicants. Features such as income, credit history, and debt-to-income ratio are used to predict whether an applicant is likely to default on a loan. The decision tree helps the bank make informed decisions on approving or denying loan applications.

## 2. Medical Diagnosis:

Example: In healthcare, decision trees are used for medical diagnosis. For instance, a decision tree can be constructed to predict whether a patient has a certain medical condition based on symptoms, medical history, and diagnostic test results. Healthcare professionals can use this decision tree to assist in diagnosing diseases and recommending appropriate treatments.

## 3. Customer Churn Prediction:

Example: Telecom companies utilize decision trees to predict customer churn. By analyzing features such as usage patterns, customer demographics, and customer service interactions, a decision tree can determine which customers are at risk of switching to a competitor. This allows the company to take proactive measures, such as targeted marketing campaigns or personalized retention offers, to prevent churn and retain valuable customers.

## **Working / Algorithm:**

### Step 1: Initialization

- Select a decision tree algorithm.
- Instantiate the decision tree classifier with specified parameters.

### Step 2: Model Training

- Train the decision tree classifier using the training dataset ( $x_{\text{train}}$ ,  $y_{\text{train}}$ ).
- The algorithm recursively partitions the feature space based on the target variable to create a tree structure.

### Step 3: Prediction

- For each instance in the testing dataset ( $x_{\text{test}}$ ):

- Traverse the decision tree by following the learned rules.
- Determine the predicted class based on the final leaf node reached.

#### Step 4: Evaluation

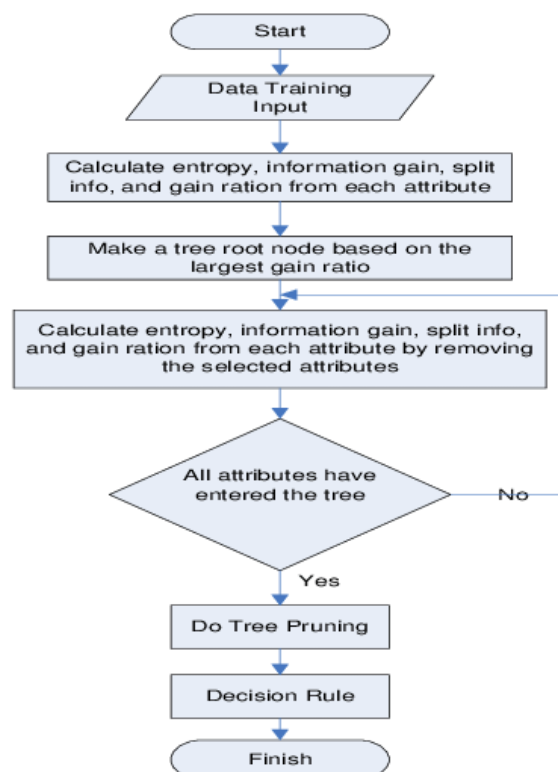
- Compare the predicted labels with the true labels from the testing dataset to assess model performance.
- Calculate evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

#### Step 5: Interpretation

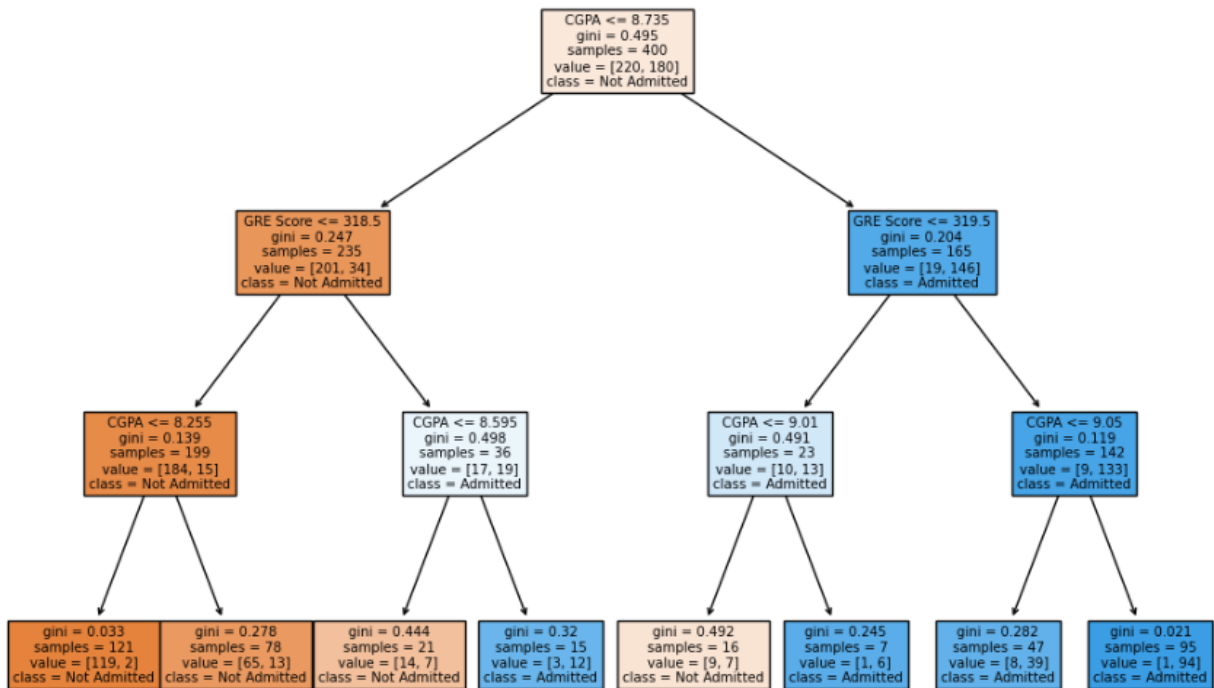
- Visualize the decision tree graphically to understand the rules learned by the model.
- Analyze feature importance to identify the most influential features in decision-making.

Step 8: The model is ready.

#### **Diagram:**



Decision Tree Max Depth = 3



## Conclusion:

In conclusion, the Decision Tree classifier for predicting student admissions to foreign universities based on GRE scores and academic performance offers a valuable solution for efficient decision-making. Its interpretability, simplicity, and versatility make it an effective tool for counselors. Through model evaluation, we gained insights into admission factors, guiding future improvements. The model contributes to transparent and equitable admission practices. Moving forward, ongoing monitoring, maintenance, and updates will ensure its continued relevance and effectiveness. The successful development and deployment of the Decision Tree classifier represents a significant advancement in improving admission procedures, leading to better outcomes for both students and educational institutions.