**Name:** Aayush Avinash Bokde  
**Class:** CSE (AI) B Division  
**Roll No.:** 282016  
**PRN No.:** 22310316  
**Subject:** Machine Learning

**Assignment 7 – Admission Prediction using Decision Tree Classifier**

**Objective**

The goal of this assignment was to build a classification model using a Decision Tree Classifier that can help educational counselors predict whether a student will be admitted (1) or not admitted (0) to a foreign university. The prediction is based on academic factors like GRE scores, TOEFL scores, CGPA, and others.

**Dataset Description**

The dataset consists of various student-related academic features:

* GRE Score (out of 340)
* TOEFL Score (out of 120)
* University Rating (1 to 5)
* Statement of Purpose (SOP) strength (1 to 5)
* Letter of Recommendation (LOR) strength (1 to 5)
* CGPA (out of 10)
* Research Experience (0 = No, 1 = Yes)
* Chance of Admit (continuous values between 0 and 1)

To make the model suitable for classification, the "Chance of Admit" column was converted into a binary label where values less than 0.75 were considered "Not Admitted" (0) and values equal to or greater than 0.75 were considered "Admitted" (1).

**1. Data Preprocessing**

The dataset was first loaded and inspected for any missing values. It was found that there were no null values in the dataset. The "Serial No." column was identified as irrelevant and removed to avoid unnecessary noise in the model.

A new target column named **Admitted** was created by converting the continuous "Chance of Admit" scores into binary values. This helped in transforming the problem into a clear classification task.

Although not deeply explored in this instance, visualization tools such as histograms and scatter plots were available for exploratory data analysis to understand score distributions and potential correlations.

**2. Data Preparation**

After preprocessing, only the relevant features were kept. These included GRE Score, TOEFL Score, CGPA, University Rating, SOP, LOR, and Research. The target variable was the newly created binary column, Admitted.

The dataset was then split into training and testing sets, with 80% of the data used for training the model and 20% for testing. This ensures that the model is trained on the majority of the data but still has unseen data to validate its performance.

**3. Model Building**

The classification model used was a **Decision Tree Classifier**. The Gini index was selected as the splitting criterion, and the maximum depth of the tree was limited to 4 in order to control complexity and prevent overfitting. The model was trained using the training data, and predictions were generated using the test data.

**4. Model Evaluation**

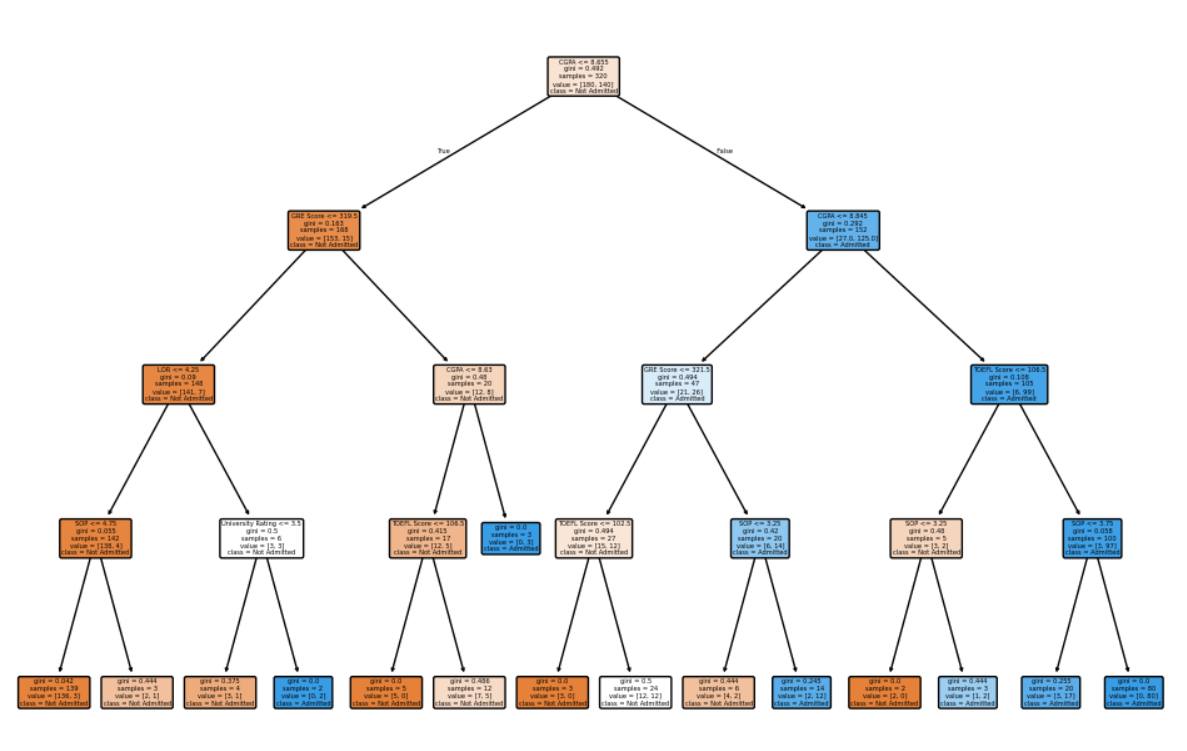
The model performance was evaluated using standard classification metrics:

| **Metric** | **Value** |
| --- | --- |
| Accuracy Score | 0.85 |
| Precision Score | 0.744 |
| Recall Score | 0.970 |
| F1 Score | 0.842 |

A **confusion matrix** was used to better understand the number of correct and incorrect predictions for both classes (Admitted and Not Admitted). Additionally, a **classification report** provided more detailed class-wise metrics, confirming that the model was especially strong in identifying students who are likely to be admitted, as shown by its high recall.

**5. Decision Tree Visualization**

The trained decision tree was visualized to show how the model makes decisions. The visualization provided a clear view of the feature splits, decision paths, and the importance of different academic indicators in classifying students.



**6. Additional Observations & Enhancements**

* The class distribution was verified to ensure that there was no major imbalance, which could bias the model.
* Though the current model used a basic setup, **hyperparameter tuning** such as adjusting the maximum depth and minimum samples required for a split could be explored to further improve model performance.
* Decision trees inherently provide **feature importance**, giving insights into which academic factors contribute most to the admission decision.
* To improve performance and generalizability, alternative ensemble models such as **Random Forest**, **Gradient Boosting**, or **XGBoost** could be tested.
* Implementing **cross-validation** would help obtain a more reliable estimate of the model’s performance across different data subsets.

**Conclusion**

The Decision Tree Classifier achieved an accuracy of 85%, demonstrating its effectiveness in predicting admission outcomes. With a high recall score, the model is particularly good at identifying students who are likely to be admitted. This model can serve as a valuable decision-support tool for educational counselors. Future enhancements using advanced techniques and model tuning could further boost predictive accuracy and robustness.