**Name:** Mansi Mahabdi

**Roll no:** 282076

**Batch:** B3

**ASSIGNMENT 7**

**Problem Statement:**

Assignment on Classification Techniques

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.

**Objective:**

The objective of this assignment is to build a classification model using a Decision Tree algorithm to help a counselor predict whether a student will be admitted to a foreign university based on key academic indicators such as GRE Score and Academic Performance.

This involves:

* Understanding and pre-processing the dataset (e.g., encoding categorical values, transforming data).
* Splitting the data into training and testing sets to ensure unbiased evaluation.
* Training a Decision Tree classifier to make binary predictions (Admitted: Yes/No).
* Evaluating the model’s performance using appropriate metrics (accuracy, confusion matrix, etc.).

Here I used dataset from Kaggle:

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

**S/W Packages and H/W apparatus used:**

Software used:

1. Python 3.x

2. Google Colab

Libraries and packages used: NumPy, Pandas, Seaborn

**Theory:**

**1. Introduction to Classification**

Classification is a supervised machine learning technique used to categorize data into predefined classes or labels. The goal is to learn a mapping function from input features (such as GRE score, GPA, etc.) to a discrete output label (like Admitted = 0 or 1).

In the context of this assignment, classification helps in predicting whether a student will be admitted or not, based on historical admission data**.**

**2. Types of Classification Problems**

* Binary Classification: Only two output classes (e.g., Admit = Yes or No).
* Multiclass Classification: More than two classes (e.g., grade levels A, B, C...).
* Multilabel Classification: Multiple labels per instance (e.g., tagging emails with multiple categories).

This assignment uses a binary classification approach.

**3. Common Classification Algorithms**

Several algorithms are used for classification, including:

* Logistic Regression
* Decision Trees
* Random Forests
* K-Nearest Neighbors (KNN)
* Naive Bayes
* Support Vector Machines (SVM)
* Neural Networks

Among these, Decision Tree is used in this assignment due to its simplicity and interpretability.

**4. Decision Tree Classifier**

A Decision Tree is a flowchart-like structure that recursively splits the dataset into branches to reach a decision.

Key Components:

Root Node: The first attribute split.

Internal Nodes: Contain decision rules based on features.

Leaf Nodes: Output class labels (e.g., Admitted = 1).

Splitting Criteria:

Gini Impurity: Measures impurity of a node.

Entropy & InformationGain: Measures the amount of information gained by a split.

**Advantages:**

* Easy to interpret and visualize
* Handles both numerical and categorical data
* Requires minimal data preprocessing

**Disadvantages:**

* Can overfit the data
* Less stable (small changes in data can lead to different trees)

**5. Model Evaluation for Classification**

To assess the performance of classification models, we use metrics such as:

Accuracy: Correct predictions out of total predictions

Precision: True Positives / (True Positives + False Positives)

Recall (Sensitivity): True Positives / (True Positives + False Negatives)

F1-Score: Harmonic mean of precision and recall

Confusion Matrix: Table showing TP, FP, TN, FN for better analysis

These metrics help understand how well the model distinguishes between admitted and non-admitted students.

**6. Real-World Relevance**

Classification techniques are widely used in real-world applications like:

Medical diagnosis (disease = yes/no)

Email spam detection

Credit approval

Student admission prediction (as in this assignment)

By using classification, institutions can make data-driven admission decisions more efficiently and accurately**.**

**Conclusion:**

In this assignment, we explored the application of **classification techniques**, particularly the **Decision Tree algorithm**, to predict student admissions based on academic scores and other relevant attributes. Classification plays a crucial role in supervised learning by enabling data-driven decisions where outcomes fall into discrete categories, such as "Admitted" or "Not Admitted."

The Decision Tree model proved to be an effective and interpretable choice for this binary classification task. It allowed us to identify key decision-making criteria, such as GRE score and GPA, and visualize how different academic profiles lead to different admission outcomes. Through proper data preprocessing, model training, and evaluation, we demonstrated how classification models can support admission counselors in making consistent, reliable, and unbiased predictions.

Overall, this assignment reinforced the importance of machine learning in educational analytics and highlighted how classification techniques can provide practical solutions to real-world decision-making problems.