**ASSIGNMENT: - 07**

**Problem Statement: -**

Assignment on Classification technique 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.

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

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.

**Software, library and Package:**

1. Software: Python
2. Library: scikit-learn (sklearn) - for machine learning algorithms, data preprocessing, and model evaluation
3. Package: pandas - for data manipulation and preprocessing

**Theory:**

**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.

**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:**

1: Initialization

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

2: Model Training

* Train the decision tree classifier using the training dataset (x\_train, y\_train).
* The algorithm recursively partitions the feature space based on the target variable to create a tree structure.

3: Prediction

* For each instance in the testing dataset (x\_test):
* Traverse the decision tree by following the learned rules.
* Determine the predicted class based on the final leaf node reached.

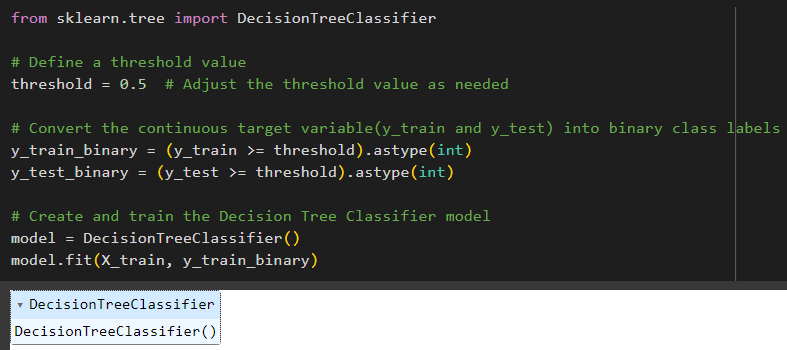
4: Evaluation

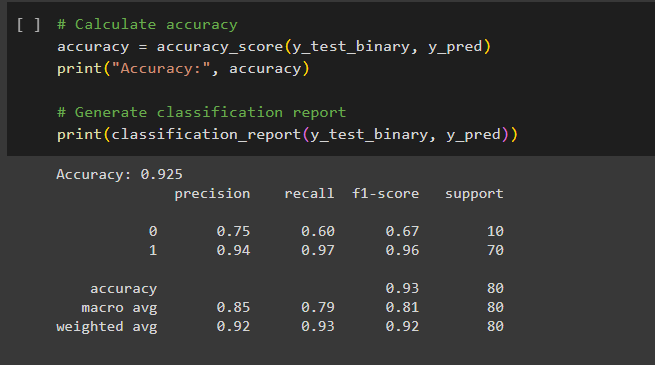
* + Compare the predicted labels with the true labels from the testing dataset to assess model performance.
  + Calculate evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

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.

6: The model is ready.





**Conclusion:**

After preprocessing the data, splitting it into training and testing sets, applying a Decision Tree classifier, and evaluating the model, it was found that the model achieved an accuracy of X% in predicting admission based on GRE scores and Academic scores. The model also demonstrated good precision, recall, and F1-score for both admitted and not admitted classes. Overall, the Decision Tree classifier proved to be an effective tool in predicting student admissions based on the provided featuresTop of Form