Name: Dev Bhangale
Roll No: 281059
Batch: A3

Assignment 7

Problem Statement:

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 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 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.

Dataset: Graduate Admissions Dataset

This dataset includes the following features:

- GRE Score (out of 340)
- TOEFL Score (out of 120)
- University Rating (out of 5)
- SOP (Statement of Purpose) Strength (out of 5)
- LOR (Letter of Recommendation) Strength (out of 5)

- CGPA (out of 10)
- Research Experience (0 = No, 1 = Yes)
- **Admitted** (Target Variable: 0 = Not Admitted, 1 = Admitted)

Objectives:

1. Perform necessary data preprocessing such as Label Encoding or transformation.
2. Prepare the dataset by splitting it into training and testing sets.
3. Train a Decision Tree Classifier to predict admission outcomes.
4. Evaluate the model using suitable classification metrics.

Resources Used:

- **Software:** Visual Studio Code, Jupyter Notebook
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

Theory:

Classification:

Classification is a supervised learning method used to assign class labels to data instances. In this practical, the aim is to classify whether a student will be admitted based on their academic profile.

Decision Tree Classifier:

A Decision Tree is a tree-like model used to make decisions:

- Internal nodes represent tests on features.
- Branches denote the result of those tests.
- Leaf nodes indicate final classification outcomes.

The tree is constructed by selecting the most significant features at each split to reduce classification error.

Methodology:

1. **Data Preprocessing:**
    - Load the dataset using Pandas.
    - Handle any missing values through removal or imputation.
    - Normalize or standardize features if necessary.
    - Apply Label Encoding to categorical variables (if any).
2. **Train-Test Split:**
    - Use Scikit-learn's `train_test_split` to divide the data into training (80%) and testing (20%) sets.
3. **Model Training:**
    - Train a `DecisionTreeClassifier` using the training data.
    - Tune hyperparameters like `max_depth` and `criterion` for better results.
4. **Model Evaluation:**
    - Generate a confusion matrix to visualize classification results.
    - Calculate Accuracy, Precision, Recall, and F1-score to assess the model's performance.

Conclusion:

- The Decision Tree model effectively classified students based on academic metrics and research experience.
- Evaluation metrics indicated the model's reliability and performance.
- Future improvements can include feature engineering and exploring other classifiers like Random Forest or Logistic Regression for better accuracy.