

Name: Prachee Prasad
Class: CSE-AI SY A
PRN: 22311748
Roll no.: 281060

Assignment 7

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 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.

Objective

The objective of this assignment is to build a machine learning classifier to help admission counselors predict whether a student will get admitted to a foreign university based on their GRE score and academic records. The model uses a Decision Tree classifier and is evaluated on its predictive performance.

Software and Libraries Used

Software:

Operating System: Windows/Linux

Programming Language: Python

Environment: Jupyter Notebook / Google Colab

Python Libraries:

pandas – for data manipulation

numpy – for numerical operations

sklearn.tree – for Decision Tree Classifier

sklearn.model_selection – for train-test split

sklearn.metrics – for model evaluation (accuracy, confusion matrix, classification report)

matplotlib and seaborn – for data visualization

Theory

What is a Decision Tree?

A Decision Tree is a supervised machine learning algorithm used for classification and regression problems. It mimics human decision-making by breaking down a problem into smaller, manageable parts through a tree-like structure. Each internal node represents a decision based on a feature, each branch represents an outcome, and each leaf node represents a class label.

Advantages of Decision Trees:

Easy to understand and interpret

Requires little data preprocessing

Can handle both numerical and categorical data

Capable of modeling non-linear relationships

Limitations of Decision Trees:

Prone to overfitting, especially on small datasets

Can be unstable to small variations in data

Less accurate compared to ensemble methods like Random Forest

Methodology

Data Collection

The dataset was downloaded from:

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

It includes student academic information such as GRE, TOEFL, University Rating, SOP, LOR, CGPA, Research experience, and Admission status.

Data Preprocessing

Checked for missing values and null entries

No categorical variables needed label encoding in this dataset

Scaled/normalized features if needed for improved performance

Feature Selection

Selected GRE Score, TOEFL Score, University Rating, CGPA, and Research as features

Admitted (0/1) used as the target variable

Train-Test Split

Used `train_test_split` to divide data into training and testing sets (e.g., 80:20 ratio)

Model Building

Applied `DecisionTreeClassifier` from `sklearn.tree`

Trained the model using the training data

Model Evaluation

Evaluated model performance using:

Accuracy Score

Confusion Matrix

Classification Report (Precision, Recall, F1 Score)

Conclusion

In this assignment, a Decision Tree classifier was successfully built and trained to predict whether a student will be admitted to a foreign university based on academic credentials. The model can assist counselors in making informed decisions during the admissions counseling process. While the Decision Tree provided interpretable results, further tuning or using ensemble techniques (like Random Forest) could enhance accuracy and reduce overfitting.