NAME: Kshitij Vinod Sali

CLASS & DIV: TE - A

ROLL NO.: 35059

ML

UA-03

Title: Assignment on classification Technique.

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 last (0=no, 1=yes). Admitted is the target variable. Data set available on kaggle (The last column of the dataset needs to be changed to 0 or 1) Data set: https://www.kaggle.com/mohansacharya/graduate. admission.

The counselor of firm is supposed check whether student will get an admission or not based on his/her GRE Score & 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. Apply data preprocessing (Label Encoding Data Transformation...) techniques if necessary. Perform data-preparation (Train-Test split). Apply machine learning algorithm and evaluate mode.

Course Outcome:

C314448.1: Implement different supervised & unsupervised learning algorithm.

C314448.2: Evaluate performance of machine learning algorithms for real-world applications.

Pre-requisite: Python, Discrete Structure.

Practical Assignment Objective: Understand the basis fundamental elements of machine learning to work on machine learning classification algorithm.

Requirements: Python programming, Jupiter Notebook, Google Colab.

Theory:

1) Classification.

→ It is a supervised learning technique used to categorize data into predefined classes or labels. The model is trained on labeled data and predicts the class of new, unseen data. It is widely used in applications like spam detection, medical diagnosis & image recognition. Popular classification algorithms include decision trees, k-nearest neighbors & support vector machines.

2) Types of classification

→ i) Binary classification: Where the data is categorized into two classes (e.g. spam or not spam).

ii) Multi-class classification: Where the data is categ-

orized into more than two classes (e.g. classify fruits into apples, bananas or oranges).

These can be further extended to multi-label classification, where each instance can belong to multiple classes.

3) What is decision tree classifier?

→ A decision tree classifier is a tree like model that splits data into branches based on features values to predict a class label. Each internal node represents a decision on a feature & each leaf node corresponds to a class label. It is easy to interpret & widely used for both classification & regression tasks due to its visual simplicity.

4) What is entropy & gini index, information gain in short?

→ i) Entropy: Measures the impurity or disorder in data, with lower entropy indicating purer subsets

ii) Gini Index: Measures the probability of incorrectly classifying a randomly chosen element with values ranging from 0 (pure) to 1.

iii) Information Gain: Measures the reduction in entropy or Gini Index after splitting the data based on a feature, helping.

5) Algorithmic steps of decision tree classifier.

→ i) Select the best feature to split the data using criteria like information gain or Gini Index.

ii) Split the dataset into subsets based on the

chosen feature.

iii) Recursively apply the process to the subsets until all data points are correctly classified or a stopping criterion is met.

iv) Assign a class label to the leaf nodes based on majority voting or pure class.

Algorithm:

i) Initialize the classifier to be used.

ii) Train the classifier: All classifiers in scikit-learn uses a fit (x, y) method to fit model (training) for given train data X and train label y.

iii) Predict the target: Given an unlabelled observation x, the predict (x) returns predicted label y.

iv) Evaluate the classifier model.

Input: We have given students dataset to build a machine learning model classifier using decision tree to predict whether a student will get admission or not.

Output: Students are able to design classification model & predicting whether a student will get admission or not.

Inference: Thus we have studied about basics of classification & to design classification model using different techniques.