Roll No. 41163

Page No.
Date: / /

Assignment A2

Execution Date: 17/02/2021
Submission Date: 19/05/2021

Title: Decision Tree Classifier

Problem Statement:

A dataset collected in cosmetics shop showing details of customers and whether they responded to special offer to buy a new lipstick is shown in table below. Use this dataset to build a decision tree, with Buys as the target variable, to help in buying lipsticks in future. Find the root node of decision tree. According to the decision tree you have made from previous training dataset: what is the decision for test data:
[Age <21, Income = Low, Gender = Female, marital status = Married]

Objective:

- Learn how to apply decision tree classifier to find the root node of decision tree
- Make decisions based on decision tree.

Outcome:

On completing this assignment, students will be able to implement code for creating a decision tree for given dataset.

Page No.
Date: / /

Software Requirements:

- Jupyter Notebook
- Python 3.8.5
- 64 bit operating system

Hardware Requirements:

- Computer with 64 bit processor

Theory:

Decision Tree:

A decision tree is a flowchart like structure in which each internal node represents a test on an attribute, each branch represents a class label i.e. outcome of the test. The path from root to leaf represents classification rules.

A decision tree consists of 3 types of nodes

1) Decision Node
   - Represented by squares

2) Chance Node
   - Represented by circle

3) End Node
   - Represented by triangle.

Page No.
Date: / /

Algorithm used:

Here ID3 (Iterative Dichotomiser 3) is used to construct decision tree. ID3 iteratively dichotomizes (divides) features into two or more groups at each step. It uses a top-down approach to construct the tree.

Steps:

1) Calculate the entropy of every attribute using the dataset.

$$entropy = -\sum pi * \log_2(pi); \quad i = 1 \text{ to } n$$

2) Split the set into subsets using the attributes for which entropy is minimum.

3) Make a decision tree node containing that attribute.

4) Recurse on subset using remaining attributes.

5) If all rows belong to the same class, make current node as a leaf node with class as its label.

Page No.
Date: / /

Test Case:

	Description	Expected O/P	Actual O/P
1)	Use given data to construct decision tree.	Tree constructed.	Successful
2)	Predict class for given test data	Class correctly predicted	Successful

Conclusion:

Thus, we have built decision tree for the given data and predicted class for given test data.