

ML-A2).

\* Title:- To find decision based on a given scenario from dataset using Decision tree classifier.

\* Problem Definition:- A dataset collected in a cosmetics shop showing details of response of a new lipstick. Use this dataset to build decision tree, with Buy's target variable to help buying lipstick in future.

\* Date of Execution:- 12/05/21

\* Objective:- 1. To understand how decision tree works on given dataset.

\* S/H & H/W required:- colab, Python, i5 64 bit Processor, OS-Linux, Keyboard, monitor.

\* Theory:-

1. In given dataset independent attributes are age, income, gender, marital status and buys is dependent.
2. Given data accepts Domain  $\{ \text{Age, Income, Gender, Marital status} \}$  and maps it to range  $\{ \text{Buys} \}$ . So it is supervised learning.
3. Categorical data takes fixed values from given set.  
eg. Age  $\in \{ <21, 21-35, >35 \}$ .  
 $0 < \text{Age} < 70$ .

## Why Decision Tree ?

A decision tree classifier is a binary tree where predictions are made by traversing the tree from root to leaf — at each node, we go left if a feature is less than a threshold, right otherwise. Finally, each leaf is associated with a class, which is the output of the predictor. Age can be classified into three classes = {<21 , 21-35 , >35 }

Income can be classified into={High , Medium , Low }

Gender can be classified into={Male , Female}

Marital status can be classified into={Single , Married }

As given data is categorical data and while finding output decisions are to be made we can use Decision Tree.

three types of nodes:

- A root node that has no incoming edges and zero or more outgoing edges.
- Internal nodes, each of which has exactly one incoming edge and two or more outgoing edges.
- Leaf or terminal nodes, each of which has exactly one incoming edge and no outgoing edges.

## Programmer's Perspective

Let S be our system

$S = \{ s, e, X, Y, fme, ffi \mid \phi \}$

s (start state)=system take input of the data  
s is constructor of class i.e.system()

e (end state) = system predict output

predict\_output() function gives mapping of domain to range

$X(\text{Input Set}) = \{X_1, X_2, X_3, X_4\}$

$X_1 \text{ (Age) } \{ X_1 \in [ <21, 21-35, 35 > ], 0 < X_1 < 70 \}$

$X_2 \text{ (Income) } \{ X_2 \in [ \text{Low}, \text{Medium}, \text{High} ],$   
     $0 < X_2 < 1000000$   
     $0 < \text{Low} < 100000$   
     $100000 \leq \text{Medium} < 500000$   
     $500000 \leq \text{High} < 1000000$   
}

$X_3 \text{ (Gender) } \{ X_3 \in [ \text{Male}, \text{Female} ] \}$

$X_4 \text{ (Marital Status) } \{ X_4 \in [ \text{Single}, \text{Married} ] \}$

Three types of nodes in Decision tree:

- A root node that has no incoming edges and zero or more outgoing edges.
- Internal nodes, each of which has exactly one incoming edge and two or more outgoing edges.
- Leaf or terminal nodes, each of which has exactly one incoming edge and no outgoing edges

Entropy - used to measure disorder or randomness

Mathematical formula is

$$e(s) = -p_i \log_2(p_i)$$

Example Output :

Enter age27

Enter income300000

Are you married(y,n): y

Select gender category:

1. male

2. female

1

Provided input: {'age\_21-35': 1, 'income\_medium': 1, 'marital\_status': 1, 'gender': 0}

Prediction: yes