

DEPARTMENT OF COMPUTERE ENGINEERING

*SUBJECT: Machine Learning*

Name	Raj Kalpesh Mathuria
UID no.	2023300139

Experiment no.	
AIM :	Decision Trees
Theory:	<p>A <b>Decision Tree</b> is a supervised learning algorithm used for <b>classification and regression</b>.</p> <p>It works by recursively splitting the dataset based on feature values to create <b>homogeneous subsets</b>.</p> <p>For <b>numerical features</b>, splits are made using <b>thresholds</b>, and the best split is chosen by minimizing an impurity measure such as <b>Gini impurity</b> (or maximizing Information Gain).</p> <p>The process continues until nodes become pure or stopping criteria are met.</p>
Code:	<a href="https://github.com/CodeCraftsmanRaj/ML_Sem6">https://github.com/CodeCraftsmanRaj/ML_Sem6</a>
OUTPUT:	<p style="text-align: center;">Decision Tree Visualization</p> <pre>graph TD;     Root["Study_Hours &lt;= 6.5 gini = 0.444 samples = 6 value = [2, 4] class = Pass"]     True["gini = 0.0 samples = 2 value = [2, 0] class = Fail"]     False["gini = 0.0 samples = 4 value = [0, 4] class = Pass"]     Root -- true --&gt; True     Root -- false --&gt; False</pre>

## DEPARTMENT OF COMPUTERE ENGINEERING

## SUBJECT: Machine Learning

Name: RAS MATHURIA  
UID: 2023300139  
Div: B Batch: D

Page No. \_\_\_\_\_  
Date \_\_\_\_\_

Study Hours	Prev Exam	Attend (%)	Pass
10	81	61	1
7	82	71	1
6	59	74	0
9	67	98	1
2	31	76	0
10	89	77	1

Gini root  $P_{succ} = \frac{4}{6} = \frac{2}{3}$   $P_{fail} = \frac{2}{6} = \frac{1}{3}$   $Gini_{root} = 1 - (\frac{1}{3})^2 - (\frac{2}{3})^2 = 0.4444$

Feature A  
 $\rightarrow 2, 6, 7, 9, 10, 10$   $t = [\frac{2+6}{2}, \frac{6+7}{2}, \dots]$   
 $y \rightarrow 0, 0, 1, 1, 1, 1$   $t = [4, 6.5, 8, 9.5, 10]$   
 $t = 6.5$   $feat_A \leq 6.5$   
 YES  $y=0$  NO  $y=1$   $t = 6.5$   $t_{6.5} = 0.0$

Gini split = 0

Feature B  
 $\rightarrow 31, 51, 59, 67, 82, 89$   $t = [41, 55, 63, 74.5, 85.5]$   
 $y \rightarrow 0, 1, 0, 1, 1, 1$   $t \leq 41$   
 $t = 41$   $t \leq 41 \Rightarrow P_0 = \frac{4}{5}$   $P_1 = \frac{1}{5}$   
 $Gini_{weighted} = \frac{1}{6} \times 0 + \frac{5}{6} \times 0.32 = 0.267$   
 $t = (\frac{4}{5})^2 - (\frac{1}{5})^2 = 0.32$   
 $t_{63} = 0.222$   $t_{41} = 0.267$   $t_{55} = 0.417$   $\rightarrow$  Threshold *Similarly for feature C.*

Gini weighted min = Feature A = 0  $\Rightarrow$  Root Node

```

graph TD
    Root["Study - Hours <= 6.5<br/>gini = 0.4444"]
    Root -- TRUE --> Node0["y = 0<br/>FAIL<br/>GINI = 0"]
    Root -- FALSE --> Node1["y = 1<br/>PASS<br/>GINI = 0"]
  
```

## CONCLUSION:

In this experiment, a decision tree was manually constructed by calculating **Gini impurity** for different numerical feature split points.

For the feature **Prev\_Exam\_Score**, candidate thresholds were evaluated, and the split at **63.0** produced the **minimum weighted Gini impurity (0.222)**.

This shows that decision trees effectively identify the most informative feature thresholds to separate classes, and manual calculations help in understanding how trees make optimal decisions.