S. B. JAIN INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NAGPUR.

Practical No. 8

Aim: Implementation of decision tree for sample data.

Name of Student: Shrutika Pradeep Bagdi

Roll No.: CS22130

Semester/Year: V/III

Academic Session: 2024-2025

Date of Performance:

Date of Submission:

AIM: Implementation of decision tree for sample data.

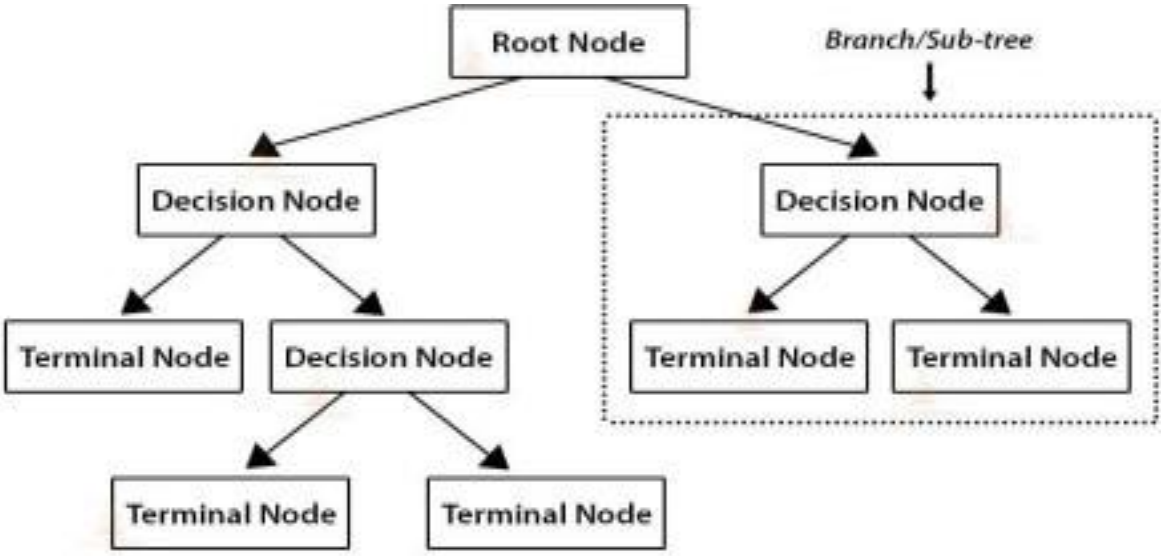
OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- To be able to understand the concept of a decision tree.
- To be able to implement a decision tree for given sample data.

THEORY:

Decision Tree is the most powerful and popular tool for classification and prediction. A decision tree is flowchart-like representation of data that graphically resembles a tree that has been drawn upside down. In this analogy, the root of the tree is a decision that has to be made, the tree's branches are actions that can be taken and the tree's leaves are potential decision outcomes. The purpose of a decision tree is to partition a large dataset into subsets that contain instances with similar values in order to understand the likely outcomes of specific options.



Important Terminology related to Decision Trees.

1. **Root Node:** It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
2. **Splitting:** It is a process of dividing a node into two or more sub-nodes.

3. **Decision Node:** When a sub-node splits into further sub-nodes, then it is called the decision node.
4. **Leaf / Terminal Node:** Nodes that do not split are called Leaf or Terminal nodes.
5. **Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
6. **Branch / Sub-Tree:** A subsection of the entire tree is called a branch or sub-tree.
7. **Parent and Child Node:** A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.

Advantages:

- Compared to other algorithms, decision trees requires less effort for data preparation during pre-processing.
- A decision tree does not require normalization of data.
- A decision tree does not require scaling of data as well.
- Missing values in the data also do NOT affect the process of building a decision tree to any considerable extent.
- A Decision tree model is very intuitive and easy to explain to technical teams as well as stakeholders.

Disadvantage:

- A small change in the data can cause a large change in the structure of the decision tree causing instability.
- For a Decision tree sometimes, calculation can go far more complex compared to other algorithms.
- Decision trees often involves higher time to train the model.
- Decision tree training is relatively expensive as the complexity and time taken are more. •

The Decision Tree algorithm is inadequate for applying regression and predicting continuous values.

PROGRAM CODE:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("Position_Salaries.csv")
import numpy as np
from sklearn.tree import  DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state=0)
X = data.iloc[ :,1:2].values       #feature selection
Y = data.iloc[ :,2].values
regressor.fit(X,Y)
pred = regressor.predict([[5],[6]])
print(pred)
```

INPUT & OUTPUT:

Sr. No.	INPUT	OUTPUT
1.	pred = regressor.predict([[5],[6]])	[110000. 150000.]
2.	pred = regressor.predict([[6],[7]])	[150000. 200000.]

CONCLUSION:

I successfully implemented a decision tree for sample data.

DISCUSSION QUESTIONS:

1. Why is the decision tree used?
2. What is the problem with decision tree?
3. Is decision tree affected by noise?
4. What type of dataset is good for decision tree?
5. What algorithm do decision trees use?

REFERENCES:

- *https://byjus.com/maths/magic-square/*
- *https://www.geeksforgeeks.org/magic-square/*
- *https://mathworld.wolfram.com/MagicSquare.html*