

EXPERIMENT NO : 4

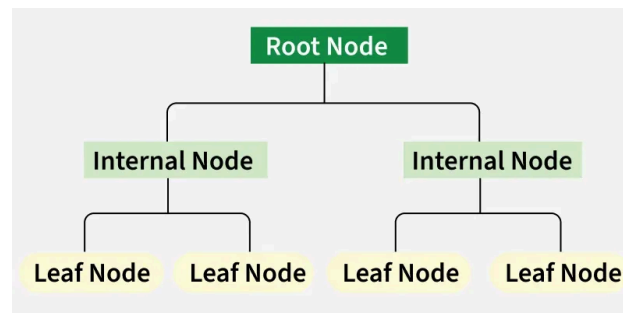
Aim: Implement Decision Tree classifier models to perform supervised classification and evaluate model performance

Software used : Google Colab

Theory:

A Decision Tree helps us make decisions by showing different options and how they are related. It has a tree-like structure that starts with one main question called the root node which represents the entire dataset. From there, the tree branches out into different possibilities based on features in the data.

- **Root Node:** Starting point representing the whole dataset.
- **Branches:** Lines connecting nodes showing the flow from one decision to another.
- **Internal Nodes:** Points where decisions are made based on data features.
- **Leaf Nodes:** End points of the tree where the final decision or prediction is made.



A Decision Tree also helps with decision-making by showing possible outcomes clearly. By looking at the "branches" we can quickly compare options and figure out the best choice.

There are mainly two types of Decision Trees based on the target variable:

1. **Classification Trees:** Used for predicting categorical outcomes like spam or not spam. These trees split the data based on features to classify data into predefined categories.
2. **Regression Trees:** Used for predicting continuous outcomes like predicting house prices. Instead of assigning categories, it provides numerical predictions based on the input features.

Splitting Criteria in Decision Trees

In a Decision Tree, the process of splitting data at each node is important. The splitting criteria finds the best feature to split the data on. Common splitting criteria include [Gini Impurity and Entropy](#).

- **Gini Impurity:** This criterion measures how "impure" a node is. The lower the Gini Impurity the better the feature splits the data into distinct categories.
- **Entropy:** This measures the amount of uncertainty or disorder in the data. The tree tries to reduce the entropy by splitting the data on features that provide the most information about the target variable.

Advantages of Decision Trees

- **Easy to Understand:** Decision Trees are visual which makes it easy to follow the decision-making process.
- **Versatility:** Can be used for both classification and regression problems.
- **No Need for Feature Scaling:** Unlike many machine learning models, it doesn't require us to scale or normalize our data.
- **Handles Non-linear Relationships:** It captures complex, non-linear relationships between features and outcomes effectively.
- **Interpretability:** The tree structure is easy to interpret and helps in allowing users to understand the reasoning behind each decision.
- **Handles Missing Data:** It can handle missing values by using strategies like assigning the most common value or ignoring missing data during splits.

Conclusion : Hence, we learned about implement Decision Tree classifier models to perform supervised classification and evaluate model performance