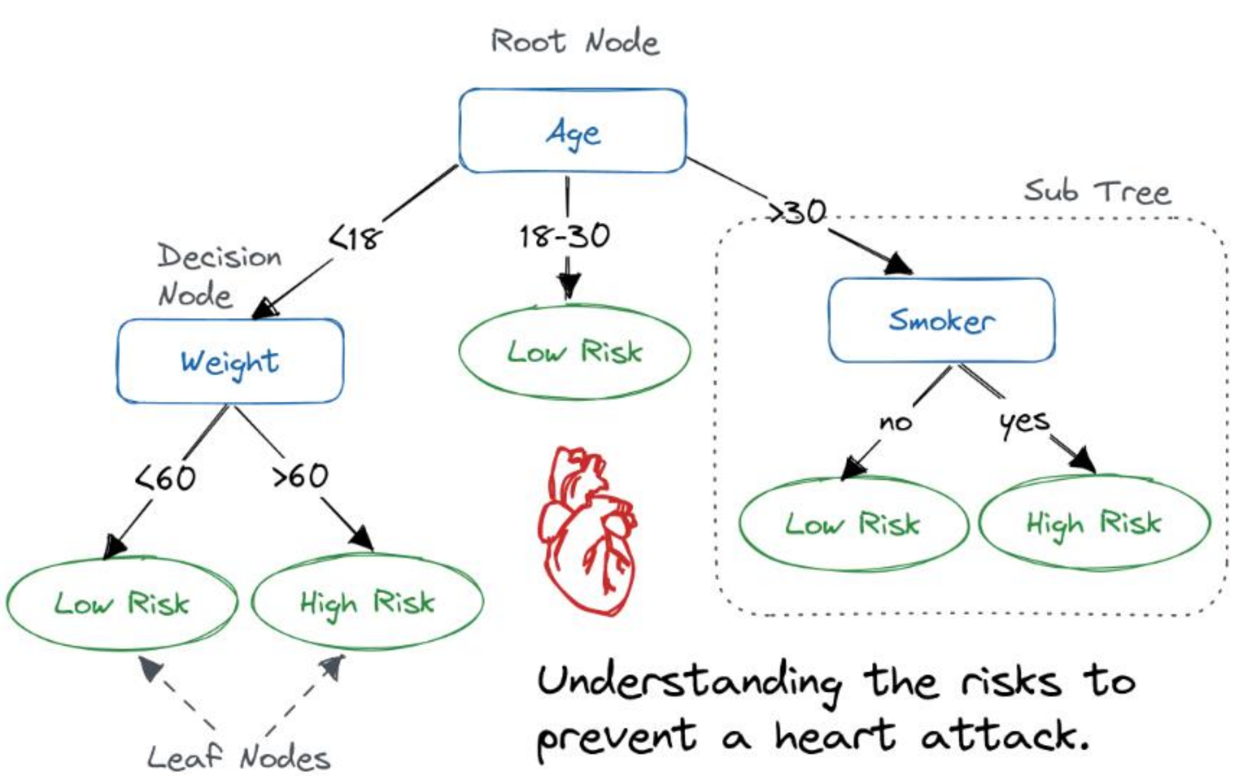
Day48 Decision Trees By: Loga Aswin

- It is a non-parametric supervised learning algorithm.
- It has a hierarchical tree structure, including a **root node, branches, internal nodes,** **and** **leaf nodes**.
- It serves as the foundation for classical machine learning algorithms like Random Forests, Bagging, and Boosted Decision Trees.
- The concept involves representing data as a tree, with internal nodes signifying tests on attributes (conditions).
- Each branch represents an outcome of the respective test.
- The leaf nodes, or terminal nodes, hold class labels.

Example:



Categories of Decision Trees:

- CART (Classification and Regression Trees):

  Utilizes the Gini Index as a metric for classification.

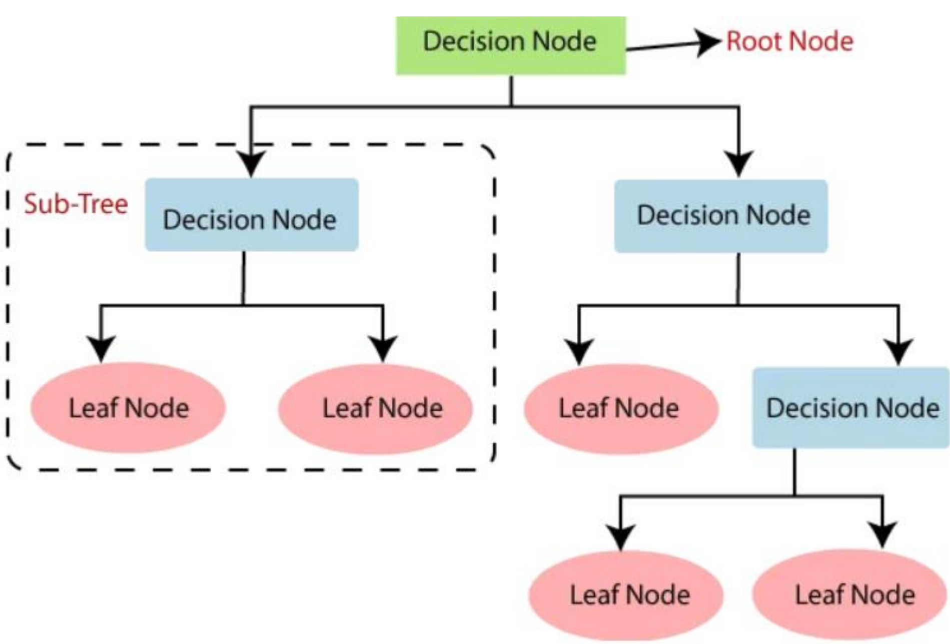
- ID3 (Iterative Dichotomiser 3):

  Employs the Entropy function and Information gain as metrics.

Terminologies:

- **Root Nodes:** The node at the beginning of a decision tree, where the population initiates division based on various features.
- **Decision Nodes:** Nodes resulting from the splitting of root nodes.
- **Leaf Nodes**: Nodes where further splitting is not possible; also known as leaf nodes or terminal nodes.
- **Branch/Sub-tree**: A sub-section of the decision tree.
- **Pruning**: Involves cutting down nodes to prevent overfitting.

Example:



How Does the Decision Tree Algorithm Work?

- Attribute Selection:

  Choose the best attribute using Attribute Selection Measures (ASM) to split the records.

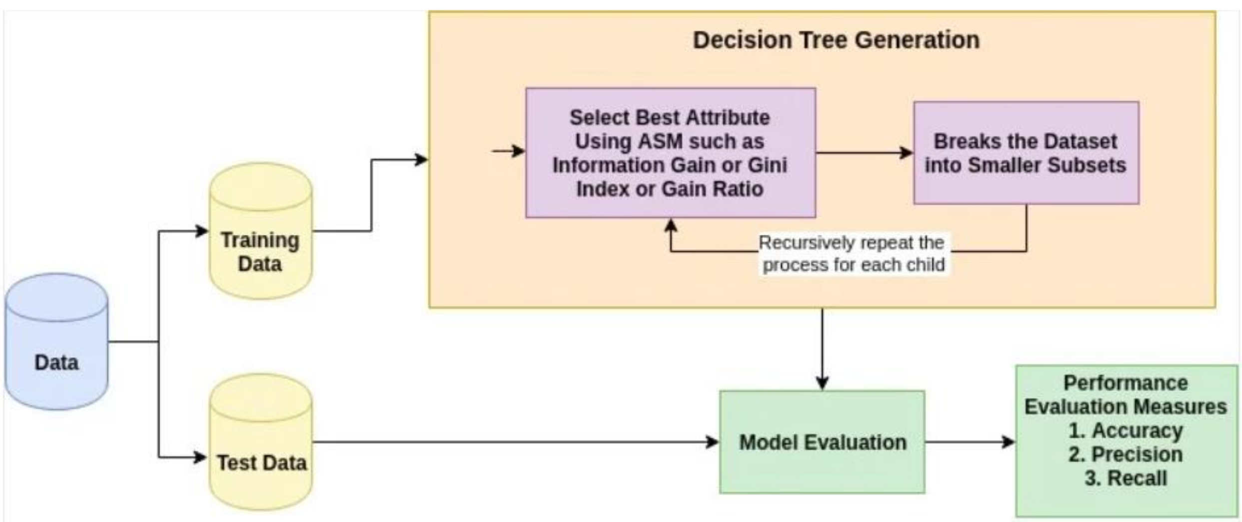
- Decision Node Creation:

  Make the selected attribute a decision node, breaking the dataset into smaller subsets.

- Recursive Tree Building:

  Begin building the tree by repeating the process recursively for each child until one of the conditions is met:

  - All tuples belong to the same attribute value.
  - There are no more remaining attributes.
  - There are no more instances.



Attribute Selection Measures:

In the process of implementing a Decision tree, a critical challenge emerges:

Determining the optimal attribute for both the root node and sub-nodes. To address this, an approach known as Attribute Selection Measure (ASM) is employed. This technique enables the selection of the most suitable attribute for the various nodes within the tree. Two widely used methods for ASM include:

- **Information Gain**
- **Gini Index**

1. Information Gain:

- Information gain quantifies the alteration in entropy following the division of a dataset by a specific attribute.
- In the decision tree construction process, nodes are split based on the value of information gain.
- The objective of a decision tree algorithm is to maximize information gain, prioritizing the split of nodes or attributes with the highest information gain.

formula:

Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)

Entropy:

Entropy serves as a metric for gauging impurity within a given attribute, indicating the level of randomness in the data.

Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)

2. Gini Index:

- The Gini index serves as a metric of impurity or purity in the context of creating a decision tree, particularly in the CART (Classification and Regression Tree) algorithm.
- Attributes with lower Gini indices are favored over those with higher Gini indices.

formula:

Gini Index= $1- \sum_j P_j^2$

Decision Tree Advantages:

- Simple to understand, mirroring human decision-making processes.
- Useful for solving decision-related problems.
- Encourages consideration of all possible outcomes.
- Requires less data cleaning compared to other algorithms.

Disadvantages:

- Complexity due to numerous layers.
- Potential overfitting issues, addressable with Random Forest.
- Computational complexity may increase with more class labels.