# Decision Tree

A Decision Tree is a supervised machine learning algorithm that is basically a tree shape diagram that shows a graphical representation of a problem with all possible solutions. It can be used in both classification and regression problems but is mostly used in classification problems.

## Decision Tree for Classification.

- A classification tree learns a sequence of if-else questions about individual features in order to infer the labels. In contrast to linear models, trees are able to capture non-linear relationships between features and labels. Trees don't require the features to be on the same scale through standardization.
- When a classification tree is trained on a labeled dataset, the tree learns patterns from the features in such a way as to produce the purest leafs.

# Building Blocks of Decision Tree

A decision tree is a data structure consisting of a hierarchy of individual units called nodes.

- Nodes: A node is a point that involves either a question or a prediction.
- Root: The root is the node at which the decision tree starts growing. It has no parent node and involves a question that gives rise to children nodes through branches.
- Internal node: A internal node is a node that has a parent.
- Leaf: One parent node, no children nodes ( A node that has no children is called a leaf ).
- Entropy: It is a measure of the randomness or unpredictability of the dataset.
- Information Gain (IG): It is the difference between the error n the root node and the leaf node. It is based on the degree of entropy after a dataset is split into an attribute. $I(parent\ node) - (Node(left) + Node(right))/N$

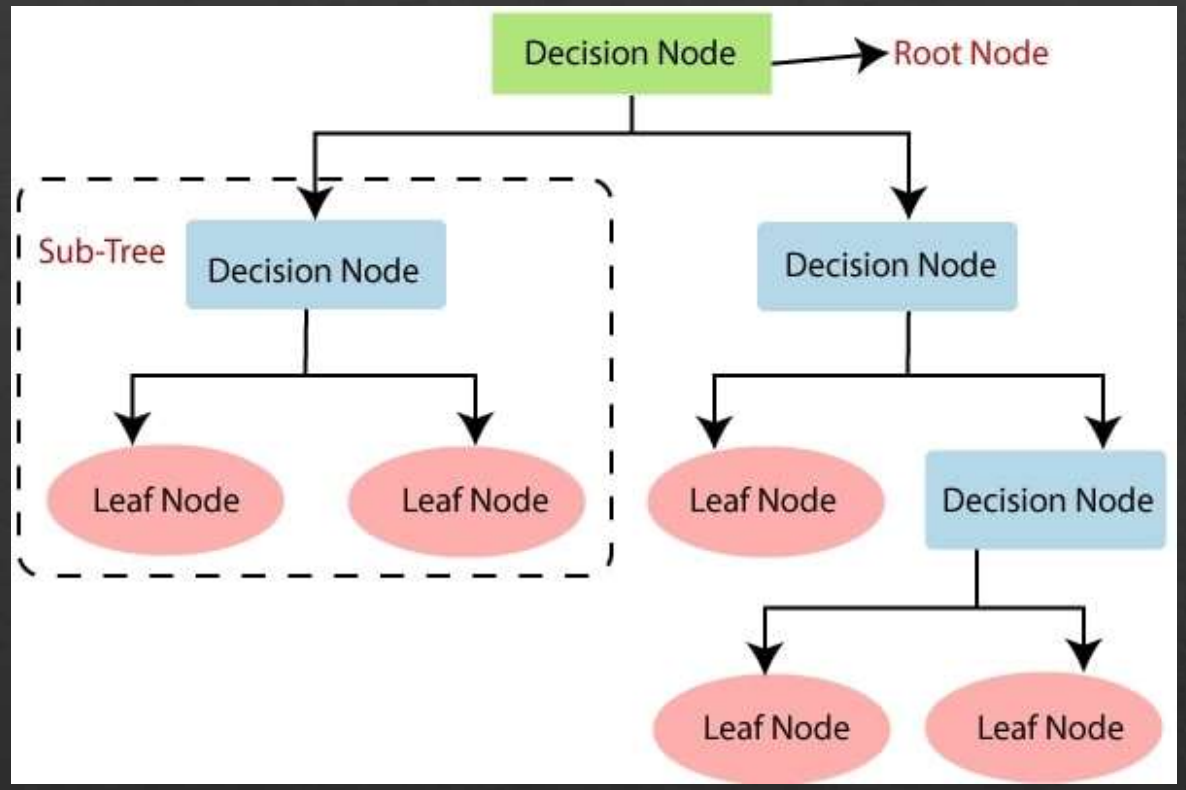
  High Entropy low IG and low entropy high IG.

## Criteria to measure the impurity of a node:

- Gini index.
- Entropy.

## Classification tree learning:

- Nodes are grown recursively.
- At each node, split the data based on feature and split_point to maximize IG (node).




Source: JavaPoint.com

## Why Use Decision Tree:

- Mimic Human brain.
- Logic can be understood easily because of the tree-like structure.

GitHub Link: https://github.com/AmeenUrRehman/Machine-Learning-Projects/tree/up-pages/Decision%20Tree

## Advantages:

- Decision tree is very useful for solving decision-related problems.
- There is less requirement for data cleaning compared to other algorithms

## Disadvantages:

- It contains lots of layers making it more complex.
- It may have an overfitting issue.