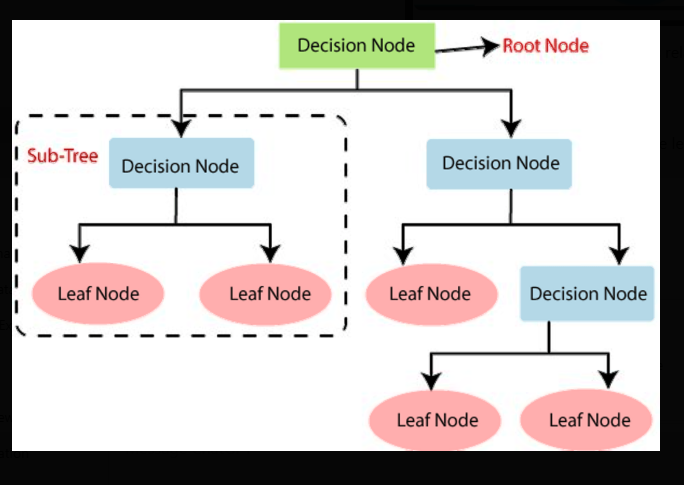
Decision Matrix

* A **Decision Tree** is a popular supervised machine learning algorithm used for both **classification** and **regression** tasks.
* The tree breaks down a dataset into smaller subsets, incrementally building the tree structure.
* The final output is a tree consisting of **decision nodes** and **leaf nodes**.
* It emphasizes that decision trees can handle both **categorical** and **numerical** data
* 
* **Root Node :** The top-most node, representing the first feature split. It initiates the decision-making process.
* **Decision Nodes :** These nodes represent points where the data is further split based on feature values.
* **Leaf Nodes:** These are the terminal nodes that provide the final outcome or prediction (classification/regression result).
* **Sub-Tree:** A subset of the tree that acts as its own smaller decision tree within the larger structure.
* **1. Regression Tree**
* **Purpose:** Used when the target variable is **continuous** and **quantitative**.
* **Examples:**
  + Predicting **rainfall** (in mm)
  + Estimating **revenue** (in $)
  + Predicting **exam scores**
* **Splitting Metric:** Often uses **Mean Squared Error (MSE)** or **Mean Absolute Error (MAE)** for splits.

### **2. Classification Tree**

### **Purpose:** Used when the target variable is **categorical** and **discrete**.

### **Examples:**

* + Predicting **High** or **Low** risk
  + Classifying as **Win** or **Loss**
  + Determining if a patient is **Healthy** or **Unhealthy**
* **Splitting Metric:** Uses **Gini Impurity** or **Entropy (Information Gain)**.

### **Advantages:**

### Simple to understand and visualize.

### Requires little data preprocessing.

* Can handle both numerical and categorical data.
* **Disadvantages:**
* Prone to overfitting (solved using techniques like **pruning** or using **Random Forests**).
* Can be biased with imbalanced data.

### **Applications of Decision Trees:**

1. Finance

2. Healthcare

3. Retail & E-commerce