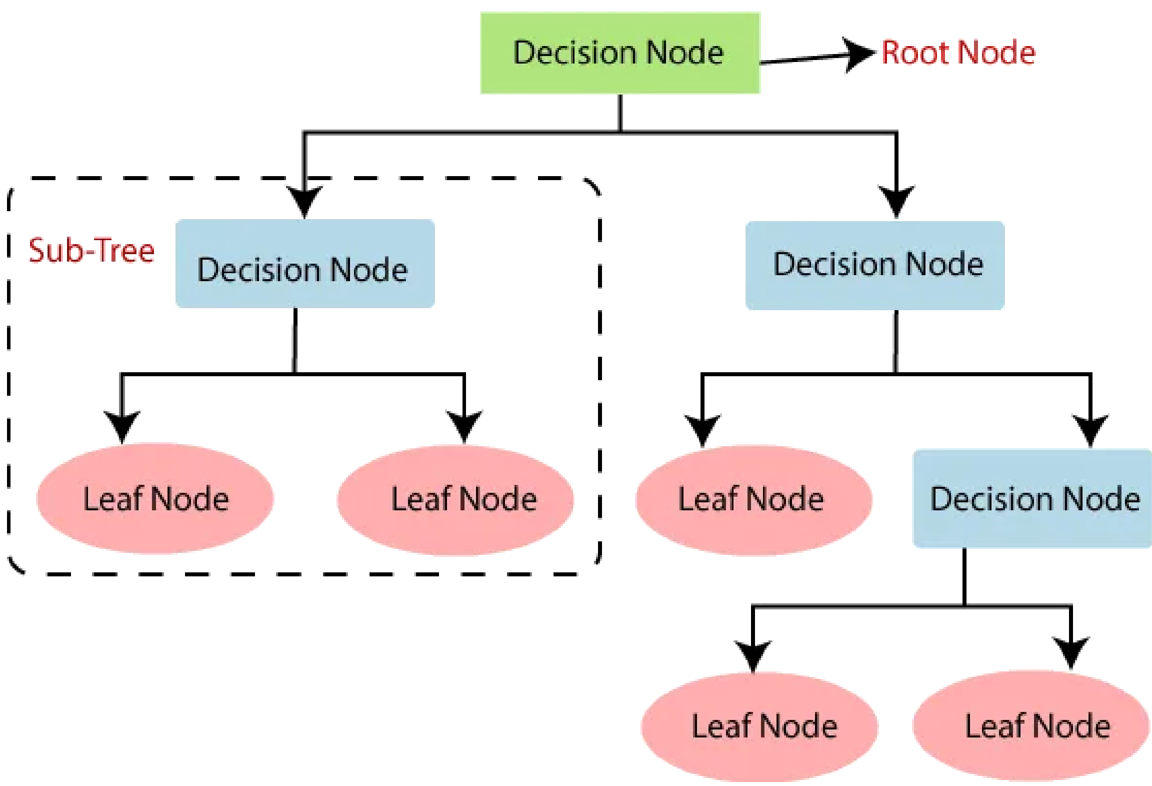
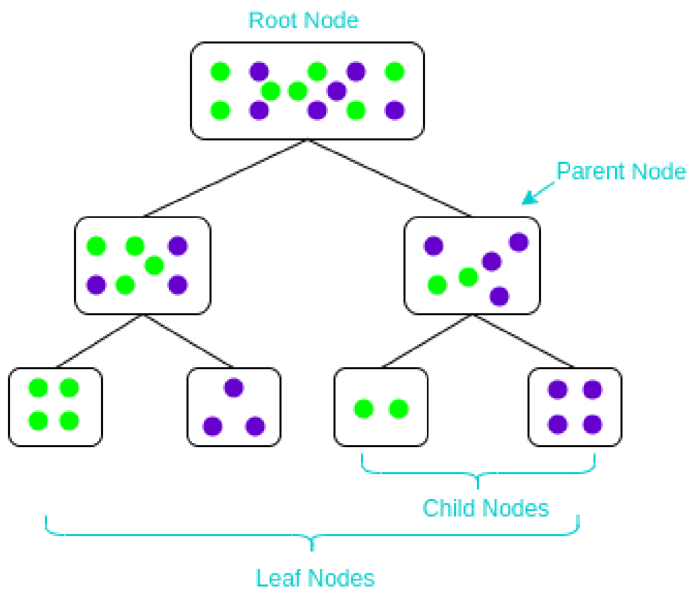
3- Supervised Classification Models - Decision Tree

By: eng.Esraa Madhi

What is decision tree ?



- **Root Node:** This is the initial node of a decision tree where the data begins to be segmented based on various attributes.
- **Decision Nodes:** These nodes are formed following the split of the root node, where further decisions on data division are made.
- **Leaf Nodes:** Also known as terminal nodes, these are the endpoints of a decision tree where no further splitting occurs.
- **Branch/Sub-tree:** Similar to a sub-graph in graph theory, a sub-tree refers to a subsection of the overall decision tree.



Overview about How Decision Trees work:

https://youtu.be/zs6yHVtxyv8

https://mlu-explain.github.io/decision-tree/

Decision Trees •

How does it works?

1. Starts at the root node
2. Splits data into groups (based on some criteria, we will see this later)
3. Set a decision at node
4. Move the data along the respective branches
5. Repeat the process until a stopping criterion is met (max levels/depth reached, min samples left to split, nothing left to split, etc)

Decision Trees in Math:

- **The entropy** can be used to quantify the impurity of a collection of labeled data points: a node containing multiple classes is impure whereas a node including only one class is pure.

$$\text{Entropy}(S) = -\sum_{i=1}^{c} p_i \log_2(p_i)$$

  Where:

  - *S* represents the dataset or a subset of the dataset.
  - *c* is the number of different classes or labels within *S*.
  - P_i is the proportion (or probability) of the elements belonging to class *i* in *S*.

  (Best Case): The entropy is minimum when all examples in a set belong to a single class. In this case, the entropy is 0

  (Worst Case): The entropy is maximum when the examples are evenly distributed across all the classes, the entropy is $\log_2(C)$

- Information gain: Used to determine which feature/attribute gives us the maximum information about a class.

$$\text{Information Gain} = \text{Parent Entropy} - \text{weighted average} * \text{Children Entropy}$$

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up

Target : Return
which features order should I use ?

Calculation of information gain for `Past Trend`.

- **Parent Entropy** = - (6/10)log(6/10) - (4/10)log(4/10) = 0.97
- **Children Entropy:**
  - (Past Trend=Positive):

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Positive	Low	High	Up
Positive	High	High	Up
Positive	Low	Low	Down
Positive	Low	Low	Down
Positive	High	High	Up

- If (Past Trend = Positive & Return = Up), probability = 4/6

        - If (Past Trend = Positive & Return = Down), probability = 2/6
            - Entropy = - (4/6)log(4/6) - (2/6)log(2/6) = 0.91
    - (Past Trend=Negative):

Past Trend	Open Interest	Trading Volume	Return
Negative	High	Low	Down
Negative	Low	High	Down
Negative	High	High	Down
Negative	Low	High	Down

        - If (Past Trend = Negative & Return = Up), probability = 0
        - If (Past Trend = Negative & Return = Down), probability = 4/4
            - Entropy = - ((0)log(0) - (4/4)log(4/4)) = 0
- Entropy of `Past Trend`:
    - P(Past Trend=Positive): 6/10
    - P(Past Trend=Negative): 4/10
    - information gain for Past Trend = 0.97 - (6/10)*0.91 - (4/10)*0 = 0.42

Why Use Decision Trees?

There are various algorithms in Machine learning. Below are the two reasons for using the Decision tree:

- *Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.*
- 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.

Pruning: Getting an Optimal Decision tree

Pruning is a process of deleting the unnecessary nodes from a tree in order to get the optimal decision tree.
A too-large tree increases the risk of **overfitting**, and a small tree may not capture all the important features of the dataset. Therefore, a technique that decreases the size of the learning tree without reducing accuracy is known as Pruning.

Resources:

- https://medium.com/@MrBam44/decision-trees-91f61a42c724
- https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/
- https://rakendd.medium.com/building-decision-trees-and-its-math-711862eea1c0