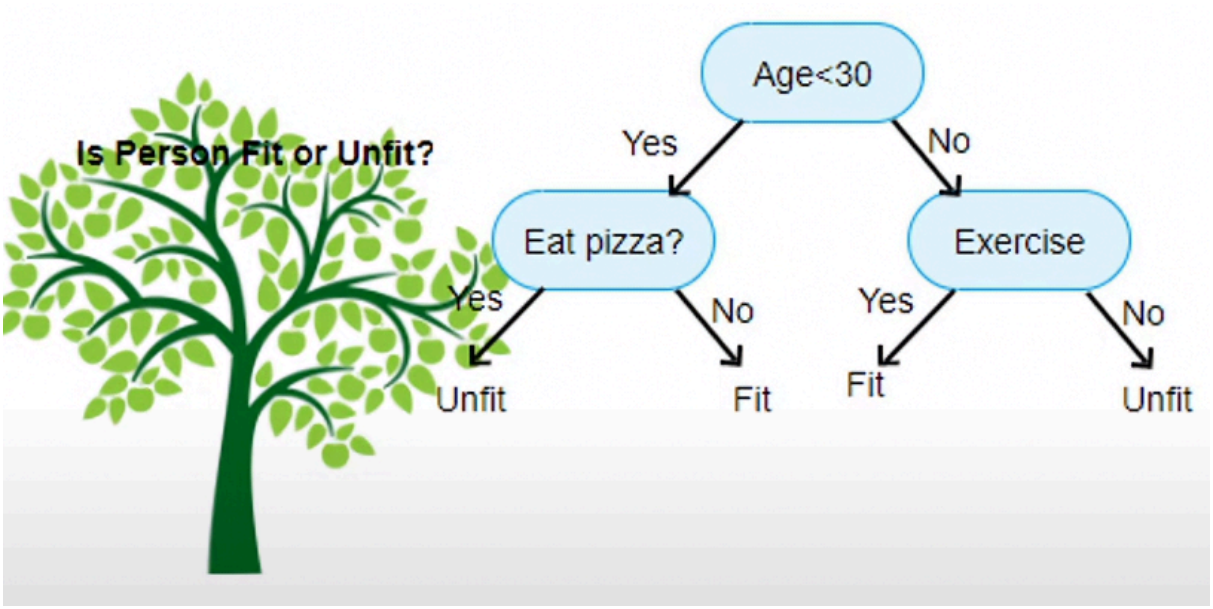
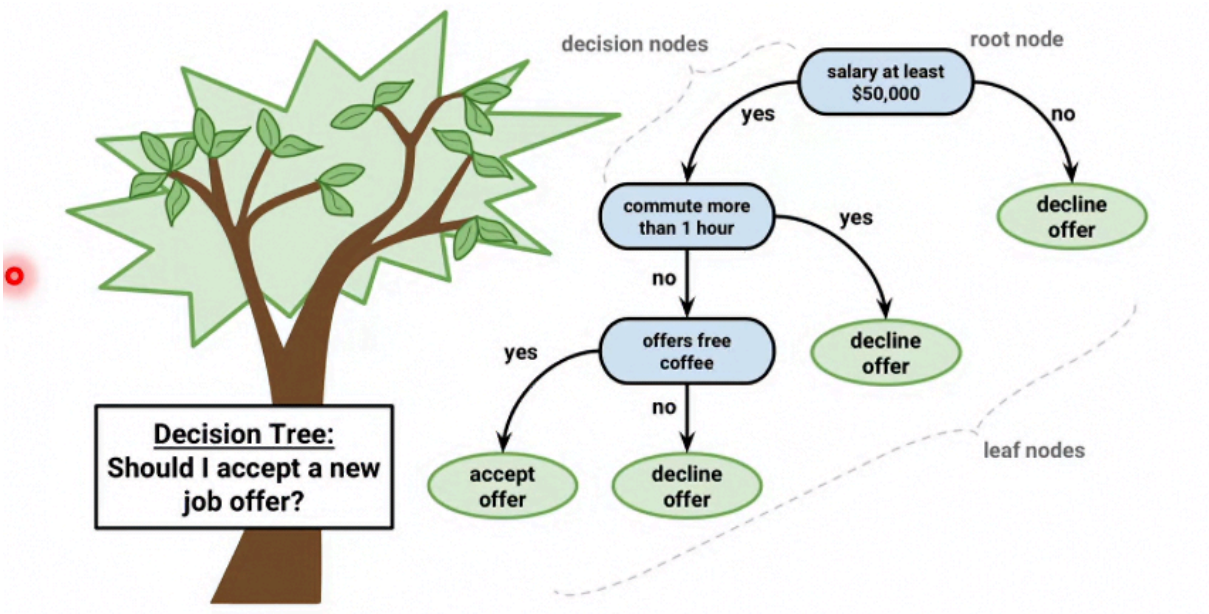
Decision Tree

Created	@August 8, 2025
Edited	@August 9, 2025 4:24 AM
Archive	☐

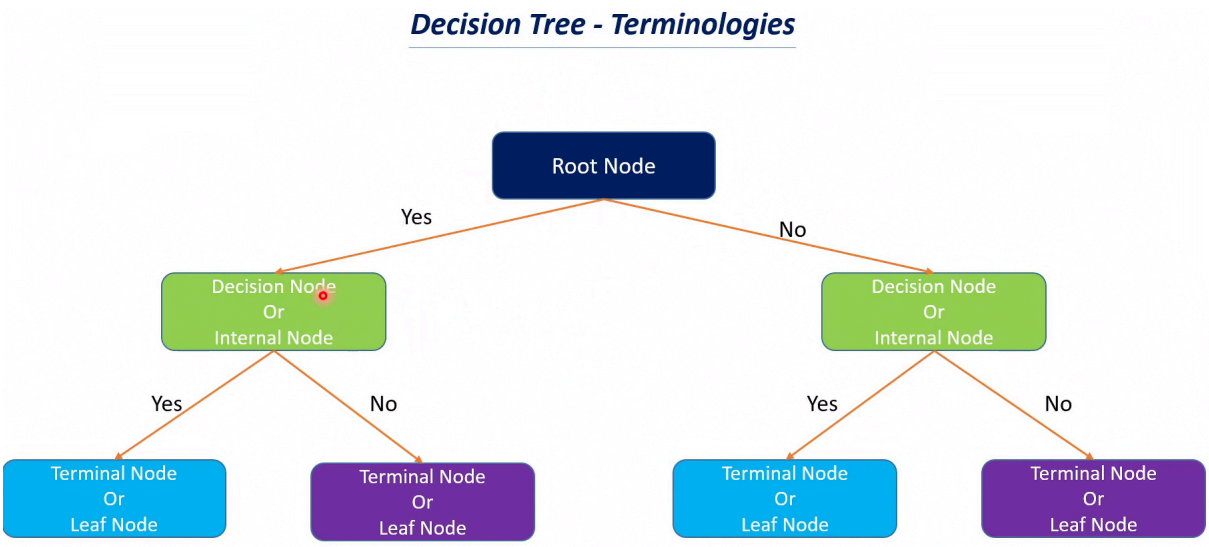
1. Supervised Learning Model
2. Used for both classification and regression
3. Builds Decision Nodes at each step
4. Basis of tree-based models

Decision Tree - examples:





Terminologies:



Advantages:

1. used both for classification and regression
2. Easy to interpret
3. It does not require normalization or scaling
4. Not sensitive to outliers

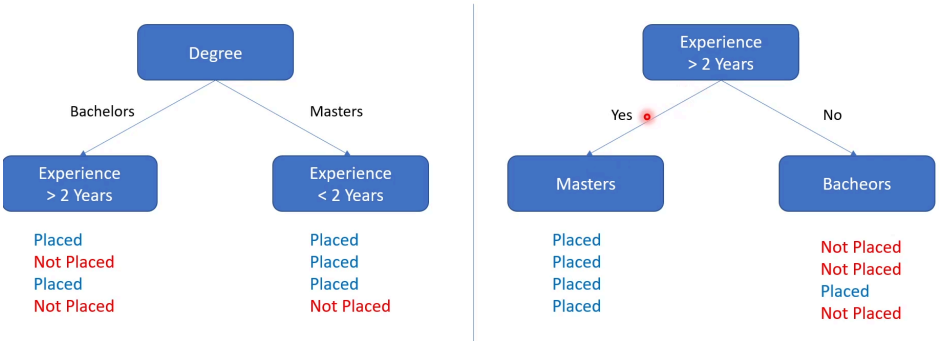
Disadvantages:

1. Overfitting issue
2. Small changes in data alter the tree structure causing instability
3. Training time is relatively higher

Entropy:

Problem Statement: Build a Decision Tree to determine whether a person will **get a Job or not** based on their **Degree** & **Years of Experience**.

Degree	Experience in Years	Placed / Not Placed
Masters	2	Placed
Bachelors	0	Not Placed
Masters	3	Placed
Masters	1	Not Placed
Bachelors	2	Placed
Masters	3	Placed
Bachelors	0	Not Placed
Bachelors	1	Not Placed



to get the best root node, entropy comes into play.

Entropy is the quantitative measure of the randomness of the information being processed.

A high value of Entropy means that the randomness in the system is high and thus making accurate predictions is tough.

A low value of Entropy means that the randomness in the system is low and thus making accurate predictions easier.

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

c --> number of classes

p_i --> Probability of i^{th} class

Information Gain:

Information gain is the the measure of how much information a feature provides about a class. Low entropy leads to increased information gain and high entropy leads to low information gain.

It computes the difference between the entropy before split and average entropy after split of that dataset based on a given feature.

$$\text{Information gain (T, F)} = \text{Entropy}(T) - \sum_{v \in F} \frac{|T_v|}{T} . \textit{Entropy}(T)$$

- T = the complete training dataset
- F = the feature/attribute being evaluated for the split
- v = a specific value that feature F can take
- T_v = {x ∈ T | x[F] = v} (the subset of examples where feature F equals value v

Gini Impurity:

The split made in a decision trees is said to be pure if all the data points are accurately separated into different classes.

Gini impurity measures the likelihood that a randomly selected data point would be incorrectly classified by a specific node.

$$G = \sum_{i=1}^{C} p(i) * (1 - p(i))$$