Decision Tree

A Decision Tree is a flowchart-like structure for decision-making or prediction. It consists of nodes (decisions/tests on attributes), branches (outcomes of these decisions), and leaf nodes (final outcomes/predictions). Each internal node corresponds to a test on an attribute, each branch to the test's result, and each leaf node to a class label or a continuous value.

How Decision Trees Work:

1. Selecting the Best Attribute: The best attribute to split the data is selected using metrics like entropy or information gain.
2. Splitting the Dataset: The dataset is split into subsets based on the selected attribute.
3. Repeating the Process: This process is recursively repeated for each subset, creating new nodes until a stopping criterion is met (e.g., all instances in a node belong to the same class, or a predefined depth is reached).

Metrics for Splitting:

- Entropy: Measures uncertainty or impurity in the dataset.
  $Entropy = -\sum_{i=1} ^{n} p_i \log_2(p_i)$ $Entropy=−∑i=1npilog2(pi)$
  where pi is the probability of an instance being classified into a particular class.
- Information Gain: Measures the reduction in entropy after splitting the dataset.
  $InformationGain = Entropy_{parent} - \sum_{i=1}^{n} \left(\frac{|D_i|}{|D|} \times Entropy(D_i)\right)$ $InformationGain=Entropyparent−∑i=1n(|D||Di|×Entropy(Di))$
  where Di is the subset after splitting by an attribute.

Advantages of Decision Trees:

- Simplicity and Interpretability: Easy to understand and mirrors human decision-making.
- Versatility: Applicable for both classification and regression tasks.
- No Need for Feature Scaling: No requirement for data normalization or scaling.
- Handles Non-linear Relationships: Capable of capturing non-linear relationships.

Disadvantages of Decision Trees:

- Overfitting: Prone to overfitting, especially with deep trees.
- Instability: Small data variations can produce a completely different tree.
- Bias towards Features with More Levels: Features with more levels can dominate the tree structure.