DECISION TREE

A decision tree is a supervised learning algorithm used for both classification and regression tasks. It has a hierarchical tree structure which consists of a root node, branches, internal nodes and leaf nodes. It works like a flowchart help to make decisions step by step where:

- Internal nodes represent attribute tests
- Branches represent attribute values
- Leaf nodes represent final decisions or predictions.

Decision trees are widely used due to their interpretability, flexibility and low preprocessing needs.
A decision tree splits the dataset based on feature values to create pure subsets ideally all items in a group belong to the same class. Each leaf node of the tree corresponds to a class label and the internal nodes are feature-based decision points. Let'S understand this with an example

Information Gain and Gini Index in Decision Tree

Till now we have discovered the basic intuition and approach of how decision tree works, so lets just move to the attribute selection measure of decision tree. We have two popular attribute selection measures used:

1 InformationGain

Information Gain tells us how useful a question (or feature) is for splitting data into groups. It measures how much the uncertainty decreases after the split. A good question will create clearer groups and the feature with the highest Information Gain is chosen to make the decision.

For example if we split a dataset of people into "Young" and "Old" based on age and all young people bought the product while all old people did not, the Information Gain would be high because the split perfectly separates the two groups with no uncertainty left

- Suppose S is a set of instances A is an attribute, Sv is the subset of S represents an individual value that the attribute A can take and Values (A) is the set of all possible values of A then

$$Gain(S,A) = Entropy(S) - \sum^{A} |S_v|/|S| . Entropy(S_v)$$

Entropy: is the measure of uncertainty of a random variable it characterizes the impurity of an arbitrary collection of examples. The higher the entropy more the information content.

For example if a dataset has an equal number of "Yes" and "No" outcomes (like 3 people who bought a product and 3 who didn't), the entropy is high because it's uncertain which outcome to predict