📝DECISION TREE CLASSIFICATION ALGORITHM

* It is a supervised learning technique that can be used for both classification and regression problems, but mostly it is preferred for solving classification problems.
* It is a tree-structured classifier, where internal nodes represent the features of the dataset, branches represent the decision rules, and each leaf node represents the outcome.
* In the decision tree, there are two nodes, which are the decision node and leaf node.
* Decision nodes are used to make any decision and have multiple branches, whereas leaf node is the output of those decisions and do not contain any further branches.
* The decision or the test are performed based on the features of the given dataset.
* It is a graphical representation for getting all the possible solutions to a problem/decision based on the given conditions.
* It is called decision tree because, like a tree, it starts with root node, which expands on further branches and construct a tree-like structure.
* To build a tree, we use CART algorithm which stands for Classification and Regression Tree Algorithm.
* A decision tree simply asks a question and based on the answer (Yes/No), it further split the tree into subtrees.

📒WHY USE DECISION TREE

* Decision trees usually mimic human thinking ability while making a decision, so it is easy to understand.
* The logic behind the decision tree can be easily understood because it shows a tree-like structure.

📝DECISION TREE TERMINOLOGIES

* Root Node
* Leaf Node
* Splitting
* Branch/Subtree
* Pruning
* Parent/Child Node

✍️HOW DOES THE DECISION TREE ALGORITHM WORKS?

1. Begin the tree with the root node, says S, which contain the complete dataset.
2. Find the best attribute in the dataset using Attribute Selection Measure (ASM).
3. Divide the S into subsets that contains possible values for the best attribute.
4. Generate the decision tree node which contains the best attribute.
5. Recursively make new decision tree using the subsets of the dataset created in step-3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

📝ATTRIBUTE SELECTION MEASURES

HOW TO SELECT THE BEST ATTRIBUTE FOR ROOT NODE AND FOR SUB NODE?

This kind of problem is solved by using technique called ASM. There are two popular techniques for ASM, which are:

* Information Gain
* Gini index

Entropy: is the metric to measure the impurity in a given attribute. It specifies randomness in the data.

Entropy(S) = -P(yes)log2(yes) – P(no)log2(no)