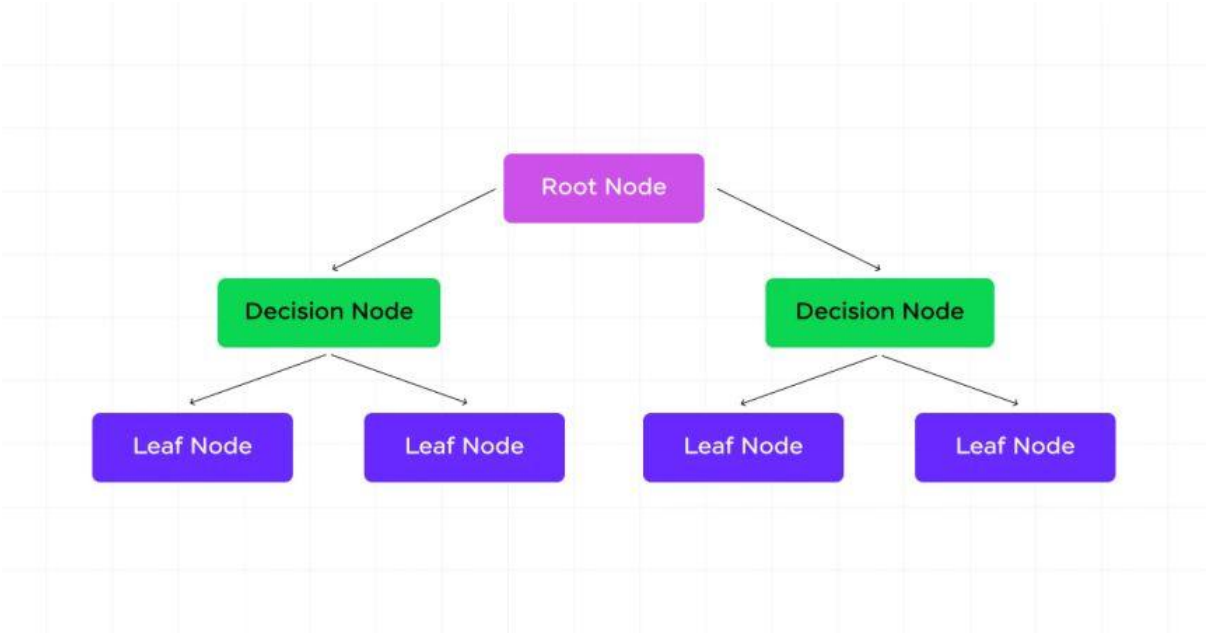
#DECISION TREE

- ➔ Supervised learning algo primarily used for classification and regression tasks.
- ➔ The model decisions as a tree-like structure, where the internal nodes represent features (attributes), the branches represent decision rules, and the leaf nodes represent the outcome (class label or values).
- ➔ A decision tree represents choices and their results in the form of a tree.
- ➔ The nodes represent an event or choice and the edges of the graph represent the decision rules or conditions.

#PRINCIPLE OF DECISION TREE

- ➔ Decision tree can be used to divide a set of items into n predetermined classes based on a specified criterion.
- ➔ They belong to a class of recursive partitioning algo that are simple to describe and implement.

#Core concepts of decision tree



1.Structure:

-> Root node: Represents the entire dataset, which is then split into two or more homogenous sets.

-> Internal node: Represents a test on a specific feature (example: “is age >30?”)

-> Leaf Node (Terminal node): Represents the final classification or regression value (the decision).

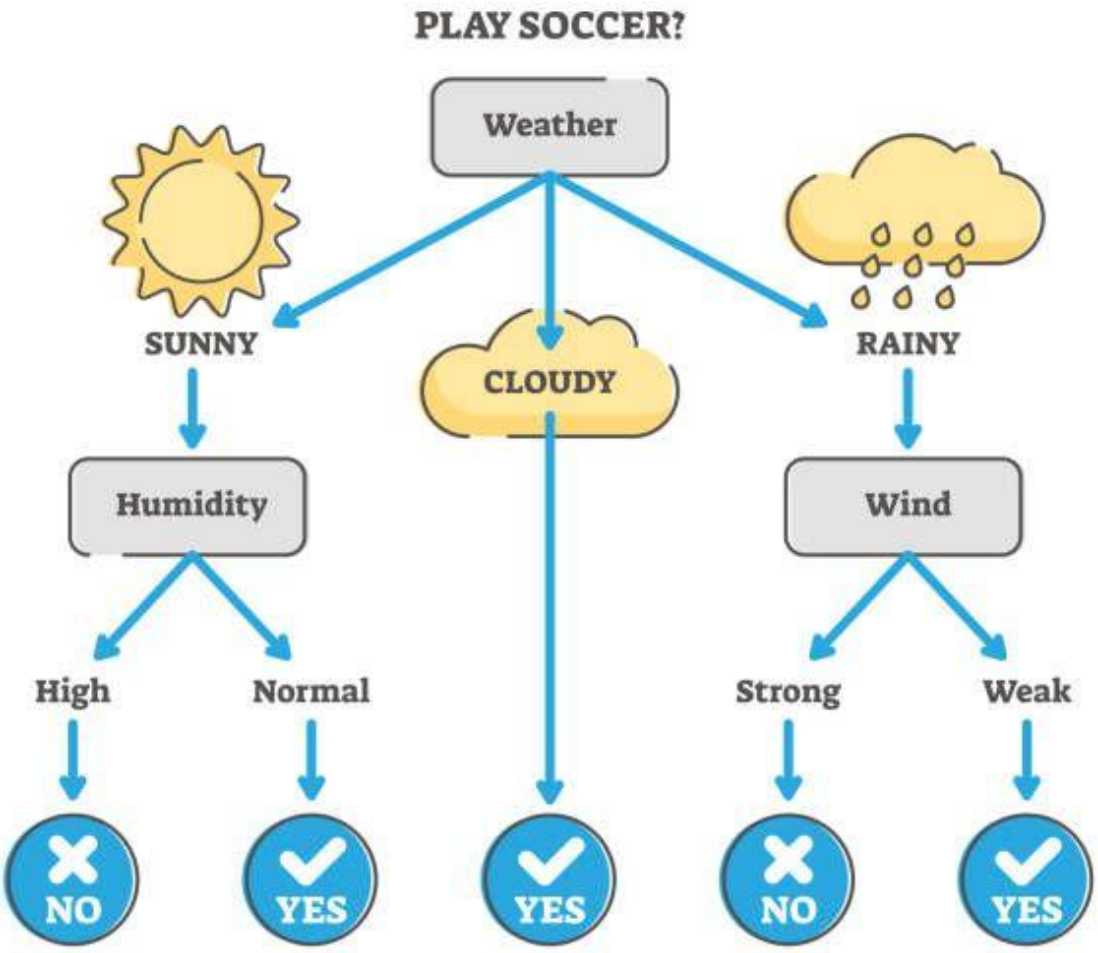
2. Splitting Criteria (how to build a tree):

-> The core challenge in building a decision tree is deciding which feature to split on and what threshold to use at each node.

->This is determined by a metric that measures the impurity or disorder of the nodes.

-> The goal is to maximize the info gain.

DECISION TREE



STEP1: SELECT A VARIABLE WHICH BEST SEPERATES THE CLASSES. SET THIS VARIABLE AS THE ROOT NODE.

STEP2: DIVIDE EACH INDIVIDUAL VARIABLE INTO GIVEN CLASSES THEREBY GENERATING NEW NODES.

DECISION TREE ARE BASED ON FORWARDING SELECTION MECHANISM, SO AFTER SPLITTING IS CREATED, IT CANNOT BE REVISITED.

STEP3: AGAIN, SELECT A VARIABLE WHICH BEST SEPERATES THE CLASSES.

STEP4: REPEAT 2 AND 3 FOR EACH NODE GENERATED UNTIL FURTHER SEPEARTION OF INDIVIDUALS IS NOT POSSIBLE.

DECISION TREE IS SUPERVISED LEARNING PREDICTIVE MODEL THAT USES RULES TO CALCULATE A TARGET VALUE.

TARGET CAN BE → **CONTINUOUS VARIABLE (->** FOR REGRESSION TREES.

-> TREES ARE KNOWN AS CONTINOUS VARIABLE D.T.

-> EXAMPLE: PREDICT PRICE OF A PRODUCT.

CATEGORICAL VARIABLE

- ➔ FOR CLASSIFICATION TREE.
- ➔ TREES ARE ALSO KNOWN AS CATEGORICAL VARIABLE D.T.

EXAMPLE: CHECK FOR YES OR NO.

#PRUNING A TREE

WHEN THE DECISION TREE BECOMES VERY DEEP -> PRUNED TO REMOVE IRRELEVANT NODES IN THE LEAVES.

- ➔ PRUNING AVOIDS CREATING VERY SMALL NODES WITH NO REAL STATSTICAL SIGNIFICANCE.
- ➔ AN ALGO BASED ON DECISION TREE IS GOOD IF IT CREATES A LARGEST-SIZED TREE AND AUTOMATICALLY PRUNES IT AFTER DETECTING THE OPTIMAL PRUNING THRESHOLD.

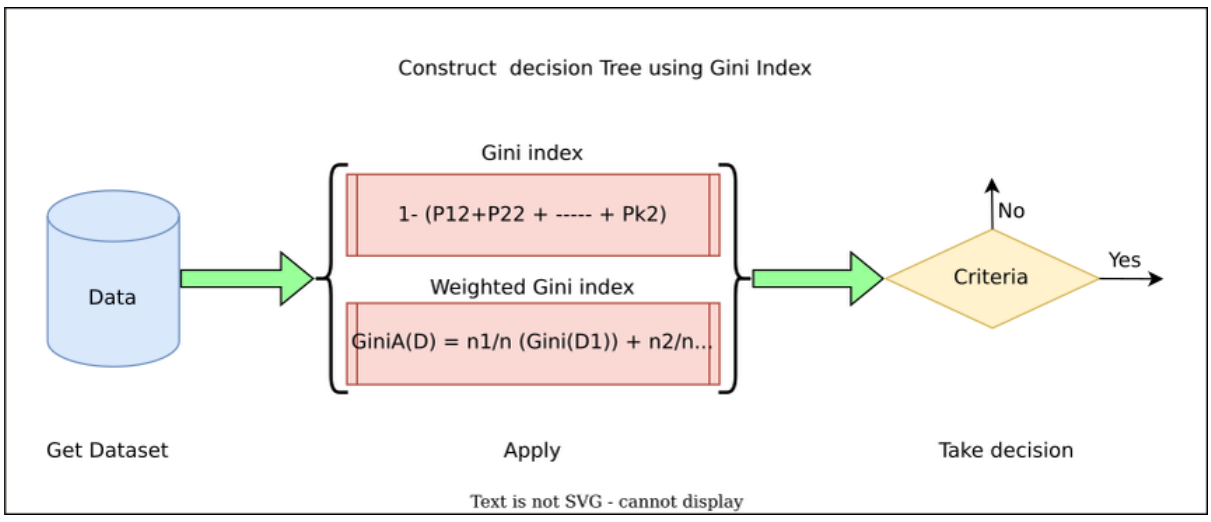
#DECISION TREE CLASSIFIER

IS CAPABLE OF BOTH BINARY (WHERE LABELS ARE [-1,1] CLASISFICATION AND MULTICLASS (WHERE LABELS ARE [0….. K-1] CLASSIFICATION.

 # MEASURES USED FOR SPLIT

1.GINI INDEX

$$Gini = 1 - \sum_{i=1}^{j} P(i)^2$$



2. ENTROPY

$$E = -\sum_{i=1}^{n} p_i \, log_2(p_i)$$

$$Information\ Gain = Entropy_{\ parent} - Entropy_{\ children}$$

$$E = -(P_{pass} \, log_2(P_{pass}) + P_{fail} \, log_2(P_{fail})$$

Advantages and Disadvantages of Trees Decision trees

1. Trees give a visual schema of the relationship of variables used for classification and hence are more explainable. The hierarchy of the tree provides insight into variable importance.
2. At times they can actually mirror decision making processes.
3. White box model which is explainable and we can track back to each result of the model. This is in contrast to black box models such as neural networks.
4. In general there is less need to prepare and clean data such as normalization and one hot encoding of categorical variables and missing values.

Note the Sklearn implementation currently does not support categorical variables, so we do need to create dummy variables. Similarly it does not support missing values. But both can be handled in theory.

5. Model can be validated statistically

Disadvantages

1. Prone to overfitting and hence lower predictive accuracy

2. Decision trees can be unstable because small variations in the data might result in a completely different tree being generated. This problem for example can be mitigated by using decision trees within an ensemble
3. Can be non-robust, i.e., a small change in the data can cause a large change in the final estimated tree
4. Predictions are approximate, based on relevant terminal nodes. Hence it it may not be the best method to extrapolate the results of the model to unseen cases.
5. Decision tree learners create biased trees if some classes dominate. It is required to balance the dataset prior to fitting with the decision tree.