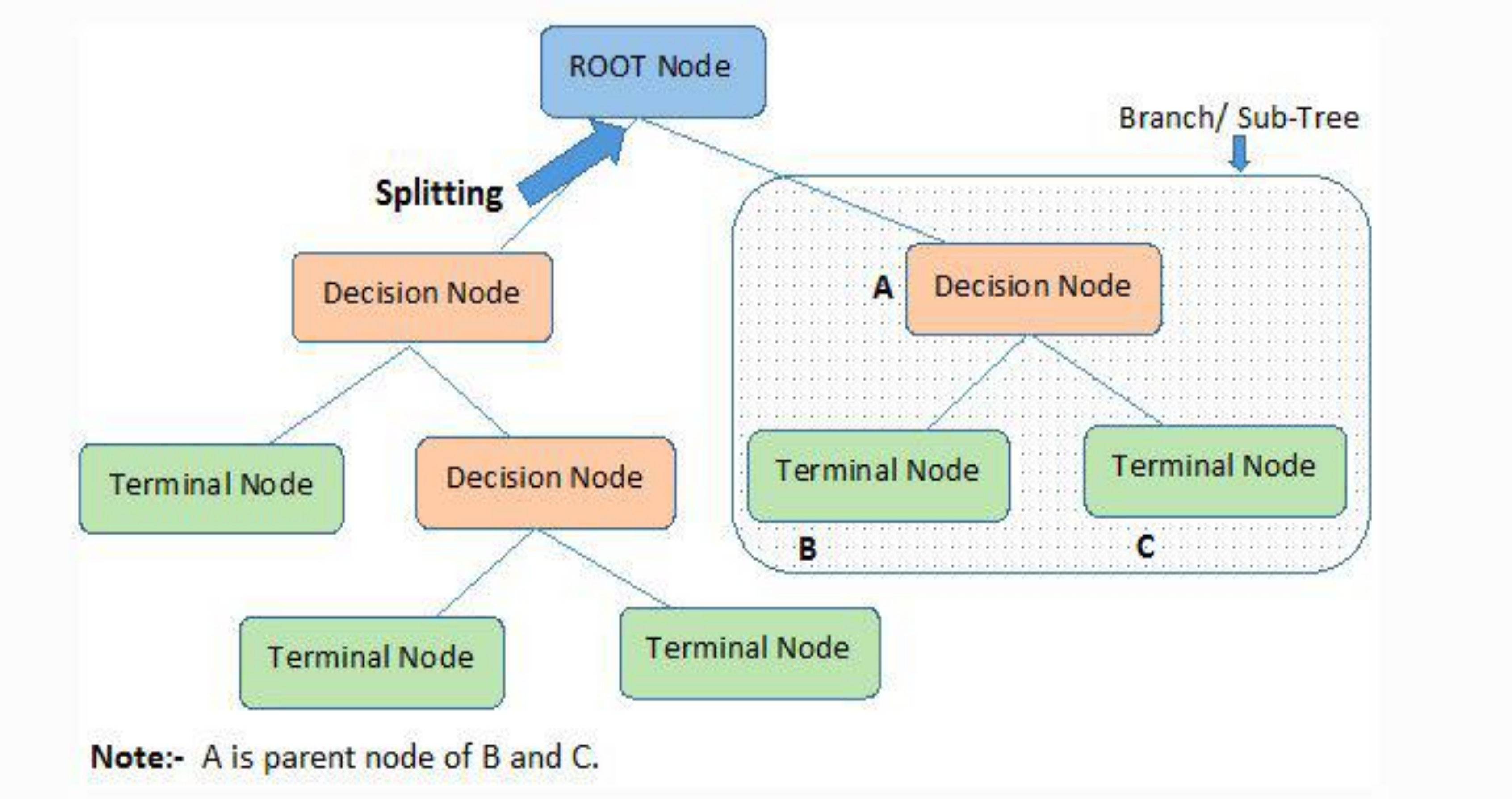
DECISION TREE ALGORITHM

DECISION TREE

- Decision tree builds models in the form of a tree structure and it helps to make certain decisions by monitoring the different attributes provided.
- It breaks down a dataset into smaller and smaller subsets

Structure of a decision tree



Terminologies related to decision tree

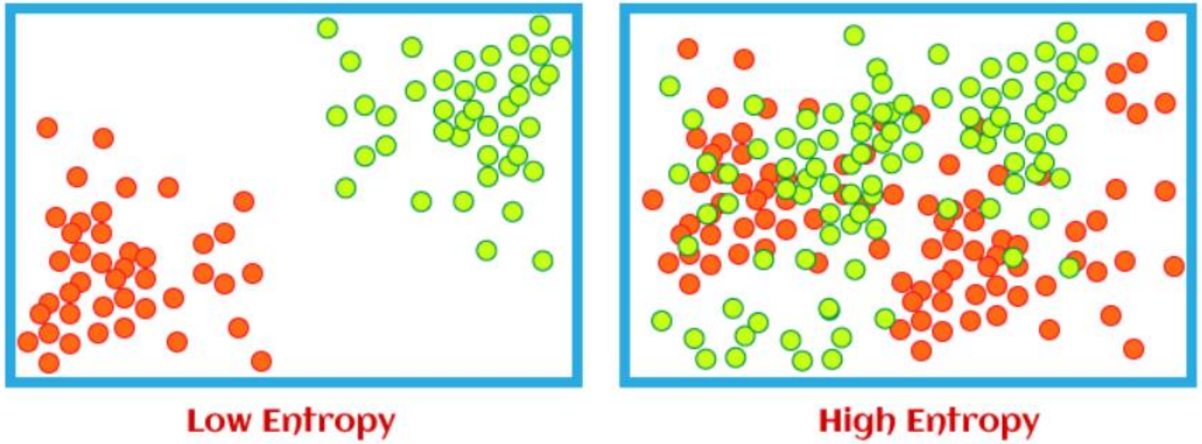
- **Root Node**: It represents the entire population and can be further divided into two or more sets.
- **Splitting**: Dividing a node into two or more sub-nodes.
- **Decision node**: When a root node splits into further sub-nodes, it is called the decision node
- **Leaf/Terminal node**: Nodes that do not split further.
- **Pruning**: Removing sub-nodes of a decision node is called pruning; opposite of splitting.

- **Branch/sub-tree**: A sub-section of an entire tree

- **Parent and Child node**: A node which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node

Entropy

- Entropy is the measurement of disorder or impurities in the information processed in machine learning. It determines how a decision tree chooses to split data.



Mathematical Formula for Entropy

$$\boxed{E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i}$$

Pi- probability of an element/class ‘i’ in our data.

Example:

let’s say we only have two classes , a positive class and a negative class. Therefore ‘i’ here could be either + or (-).
So if we had a total of 100 data points in our dataset with 30 belonging to the positive class and 70 belonging to the negative class then ‘P+’ would be 3/10 and ‘P-’ would be 7/10.
Entropy for this class is given by:

$$-\frac{3}{10} \times \log_2\left(\frac{3}{10}\right) - \frac{7}{10} \times \log_2\left(\frac{7}{10}\right) \approx 0.88$$

MSIGMA GOKULAM

Information Gain

- The concept of entropy plays an important role in measuring the information gain.
- Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.
- It calculates how much information a feature provides us about a class.

$$\boxed{\text{Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)}}$$

S= Total number of samples

Sample dataset

Outlook	Temperature	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rainy	Mild	High	Weak	Yes
Rainy	Cool	Normal	Weak	Yes
Rainy	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rainy	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rainy	Mild	High	Strong	No

Outlook	**Temperature**	**Humidity**	**Wind**	**Play Tennis**
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Weather	Temperature	Humidity	Wind	Play Tennis
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No