

# Decision Tree

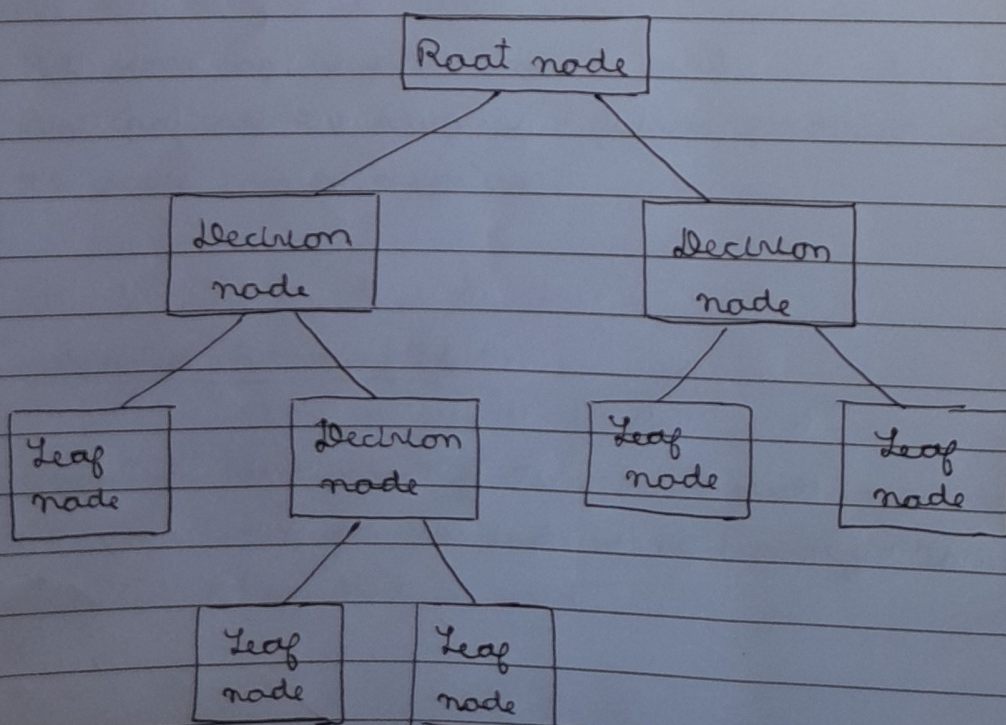
classmate

Date \_\_\_\_\_

Page \_\_\_\_\_

Decision Tree is a supervised ML algorithm. It is generally used for Classification, but it can also be used for Regression. It is a tree-like structure made from a series of feature-based splits. It starts with a root node and ends with leaf nodes.

- Root node - is present at the beginning of a decision tree. From this node, the population starts dividing according to features.
- Decision node - are the nodes we get after splitting the root node.
- Leaf node - are the nodes where further splitting is not possible.
- Sub-Tree - is a sub-section of a decision tree.





## ■ Entropy :-

Entropy is a measure of disorder / uncertainty / impurity in a dataset.

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

where,  $E \rightarrow$  entropy

$n \rightarrow$  total no. of classes

$p_i \rightarrow$  probability of class  $i$

For 2 class classification, ( $E_{\min} = 0$ ,  $E_{\max} = 1$ )

$$E = -p_1 \log_2(p_1) - p_2 \log_2(p_2)$$

For  $n$  classes, ( $E_{\min} = 0$ ,  $E_{\max} \geq 1$ )

$$E = -p_1 \log_2(p_1) - p_2 \log_2(p_2) - \dots - p_n \log_2(p_n)$$

## ■ Information gain :-

Information gain is the reduction of entropy after a dataset is split on the basis of a feature. It measures the quality of a split. The feature which gives the highest information gain is selected as a root / decision node.



$$I = E(P) - \{ \text{weighted avg.} * E(c) \}$$

where,  $I \rightarrow$  information gain

$E(P) \rightarrow$  entropy of parent dataset

$E(c) \rightarrow$  " " child " "

#### Decision Tree Algorithm :-

1. select the root node by finding the feature which has the highest information gain.
2. split the dataset into subsets based on this feature.
3. select the decision node by finding the feature which has the highest information gain.
4. split the dataset based on this feature.
5. continue step 3 and step 4 until the leaf node is reached (where entropy = 0).

#### Advantages of Decision Tree :-

- intuitive and easy to understand
- minimal data preparation is required
- cost of using the tree (predicting) is logarithmic.



## Disadvantages of Decision Tree :-

- prone to overfitting
- " " error for imbalanced datasets.