A Decision Tree is a simple representation for classifying examples. It is a Supervised Machine Learning where the data is continuously split according to a certain parameter.

Decision Tree consists of :

1. **Nodes** : Test for the value of a certain attribute.
2. **Edges/ Branch** : Correspond to the outcome of a test and connect to the next node or leaf.
3. **Leaf nodes** : Terminal nodes that predict the outcome (represent class labels or class distribution).

Advantages of Classification with Decision Trees:

1. Inexpensive to construct.
2. Extremely fast at classifying unknown records.

Disadvantages of Classification with Decision Trees:

1. Easy to overfit.
2. Decision tree models are often biased toward splits on features having a large number of levels.

Applications of Decision trees in real life :

1. Financial analysis (Customer Satisfaction with a product or service).
2. Astronomy (classify galaxies).

1. 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.
- According to the value of information gain, we split the node and build the decision tree.

- A decision tree algorithm always tries to maximize the value of information gain, and a node/attribute having the highest information gain is split first. It can be calculated using the below formula:

1. Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)

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

`Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)`

Where,

- S= Total number of samples
- P(yes)= probability of yes
- P(no)= probability of no

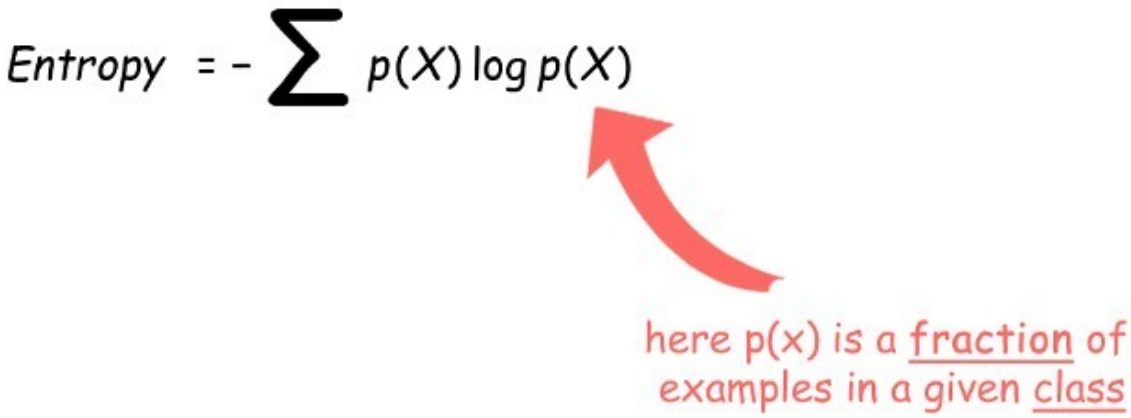
2. Gini Index:

- Gini index is a measure of impurity or purity used while creating a decision tree in the CART(Classification and Regression Tree) algorithm.
- An attribute with the low Gini index should be preferred as compared to the high Gini index.
- It only creates binary splits, and the CART algorithm uses the Gini index to create binary splits.
- Gini index can be calculated using the below formula:

$$\text{Gini Index} = 1 - \sum_j P_j^2$$

Entropy is the measures of impurity, disorder or uncertainty in a bunch of examples.

Entropy controls how a Decision Tree decides to ***split*** *the data. It actually effects how a* ***Decision Tree*** *draws its boundaries.*



Pruning:

Pruning is a process of deleting the unnecessary nodes from a tree in order to get the optimal decision tree.

Random Forest Classifier:

The Random Forest is also known as Decision Tree Forest. It is one of the popular decision tree-based ensemble models. The accuracy of these models is higher than other decision trees. This algorithm is used for classification.

In a random forest, we create a large number of decision trees, and in each decision tree, every observation is fed. The final output is the most common outcome for each observation. We take a majority vote for each classification model by feeding a new observation into all the trees.

An error estimate is made for cases that were not used when constructing the tree. This is called an out-of-bag(OOB) error estimate mentioned as a percentage.

Support Vector Classifier:

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane.

These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The hyperparameters
n estimators, max features, min sample leaf are used for increasing the predictive
power. The hyperparameters n jobs, random state, oob score are used for increas-
ing the model’s speed.