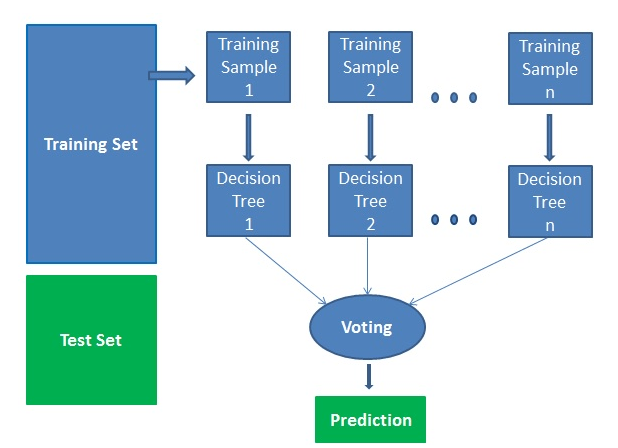
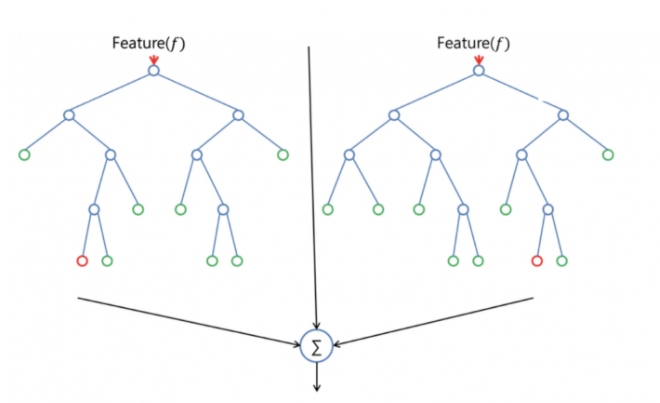
**Random forest Algorithm**

Random forest is a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forest create decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

How does it work?

* Select random samples from a given dataset.
* Construct a decision tree for each sample and get a prediction result from each decision tree.
* Perform a vote for each predicted result.
* Select the prediction result with the most votes as the final prediction.





**Advantages**

* Random forest is considered as a highly accurate and robust method because of the number of decision trees participating in the process.
* It does not suffer from the overfitting problem. The main reason is that it takes the average of all the predictions, which cancels out the biases.
* The algorithm can be used in both classification and regression problems.
* Random forests can also handle missing values. There are two ways to handle these: using median values to replace continuous variables and computing the proximity-weighted average of missing values.
* You can get the relative feature importance, which helps in selecting the most contributing features for the classifier.

**Disadvantages**

* Random forest is slow in generating predictions because it has multiple decision trees. Whenever it makes a prediction, all the trees in the forest must make a prediction for the same given input and then perform voting on it. This whole process is time-consuming.
* The model is difficult to interpret compared to a decision tree, where you can easily decide by following the path in the tree.

**Finding important features**

Random forest offers a good feature selection indicator. Scikit-learn provides an extra variable with the model, which shows the relative importance or contribution of each feature in the prediction. It automatically computes the relevance score of each feature in the training phase. Then it scales the relevance down so that the sum of all scores is 1.

**Hyperparamter tuning**

**Some of them are:**

n\_estimators : number of trees the algorithm builds before taking the maximum voting or taking the averages of predictions.

**max\_depth**: int or None, optional (default=None)

The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min\_samples\_split samples.

**min\_samples\_split**: int, float, optional (default=2)

The minimum number of samples required to split an internal node.

**min\_samples\_leaf**: int, float, optional (default=1)

The minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min\_samples\_leaf training samples in each of the left and right branches. This may have the effect of smoothing the model, especially in regression.

**n\_jobs**: how many processors it can use. If it has a value of one, it can only use one processor. A value of “-1” means that there is no limit.

**oob\_score**: It is a random forest cross-validation method. About one-third of the data is not used to train the model and can be used to evaluate its performance. These samples are called the out-of-bag samples. It's very similar to the leave-one-out-cross-validation method, but almost no additional computational burden goes along with it.

**Use Cases**

* The random forest algorithm is used in a lot of different fields, like banking, the stock market, medicine and e-commerce.
* In finance, for example, it is used to detect customers more likely to repay their debt on time or use a bank's services more frequently. In this domain it is also used to detect fraudsters out to scam the bank. In trading, the algorithm can be used to determine a stock's future behavior.
* In the healthcare domain it is used to identify the correct combination of components in medicine and to analyze a patient’s medical history to identify diseases.
* Random forest is used in e-commerce to determine whether a customer will like the product or not.