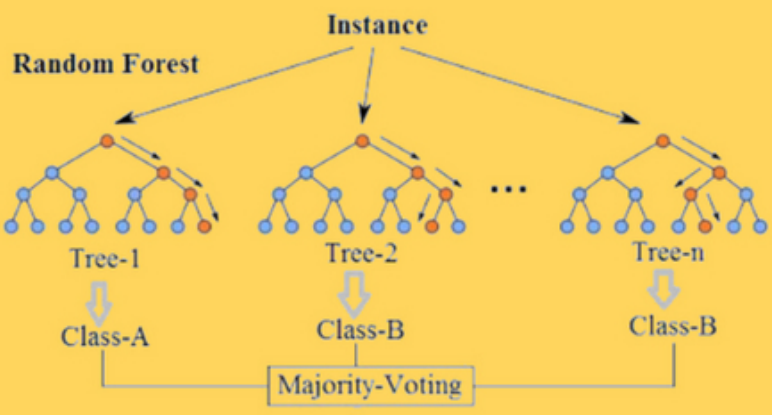
RANDOM FOREST

AN INTUITION TO RANDOM FOREST

RANDOM FORESTS ARE SUPERVISED ENSEMBLE-LEARNING MODELS USED FOR CLASSIFICATION AND REGRESSION.

Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

WHAT IS THE RANDOM FOREST ALGORITHM?



Ensemble learning models aggregate multiple machine learning models, allowing for overall better performance.

The logic behind this is that each of the models used is weak when employed on its own, but strong when put together in an ensemble. In the case of Random Forests, a large number of Decision Trees, acting as the "weak" factors, are used and their outputs are aggregated, with the result representing the "strong" ensemble.

HOW DOES IT WORK?

There are two steps in the Random Forest algorithm, one is random forest creation, the other is to make a prediction from the random forest classifier created in the first step.

THE DIFFERENCE BETWEEN THE RANDOM FOREST ALGORITHM AND THE DECISION TREE ALGORITHM IS THAT IN RANDOM FOREST, THE PROCESSES OF FINDING THE ROOT NODE AND SPLITTING THE FEATURE NODES WILL RUN RANDOMLY.

CREATION

Each tree is grown as follows:

1. If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. This sample will be the training set for growing the tree.
2. If there are M input variables, a number is specified such that at each node, m variables are selected at random out of the M and the best split on this m is used to split the node.

PREDECTION

The random forest prediction is broken down in the below steps :

1. Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target)
2. Calculate the votes for each predicted target
3. Consider the high voted predicted target as the final prediction from the random forest algorithm