As the name suggests, ***"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."*** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

**The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.**

The below diagram explains the working of the Random Forest algorithm:



Applications of Random Forest

There are mainly four sectors where Random forest mostly used:

1. **Banking:** Banking sector mostly uses this algorithm for the identification of loan risk.
2. **Medicine:** With the help of this algorithm, disease trends and risks of the disease can be identified.
3. **Land Use:** We can identify the areas of similar land use by this algorithm.
4. **Marketing:** Marketing trends can be identified using this algorithm.

# Hyperparameters of a Random Forest

Below is the list of the most important parameters and below that is a more refined section on how to improve prediction power and your model training phase easier.

**max\_depth:** The maximum depth of the tree - meaning the longest path between the root node and the leaf node.

**min\_sample\_split:** The minimum number of samples required to split an internal node:where the default = 2

**max\_leaf\_nodes:** This is the maximum number of leaf nodes a decision tree can have..

**min\_samples\_leaf:** This is the minimum number of samples required to be at a leaf node where the default = 1

**n\_estimators:** This is the number of trees in the forest. Suggested 64-128

**max\_sample:** This determines the fraction of the original dataset that is given to any individual tree.

**max\_features:** This is the number of features to consider when looking for the best split. Log2(N + 1), square root of n

**bootstrap:** If this is set as False, the whole dataset is used to build each tree, but it is set as Default.

**criterion:** The function to measure the quality of a split

**Out Of Bag : Test** Data

BootStraping – Random Selection of rows during decision tree creation.