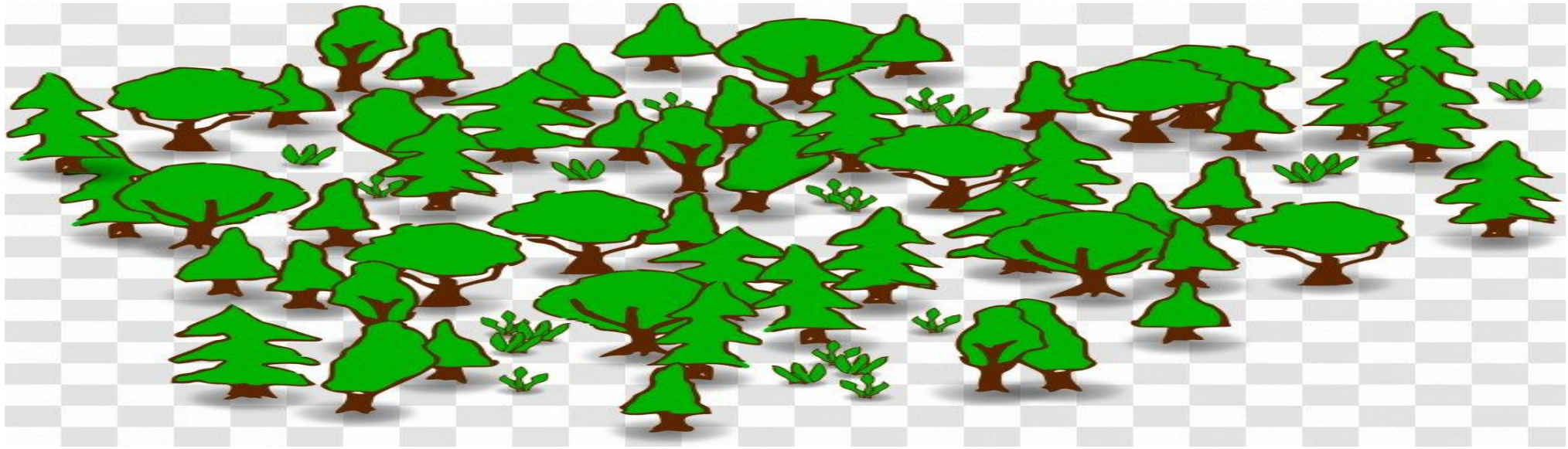


Random Forest Algorithm

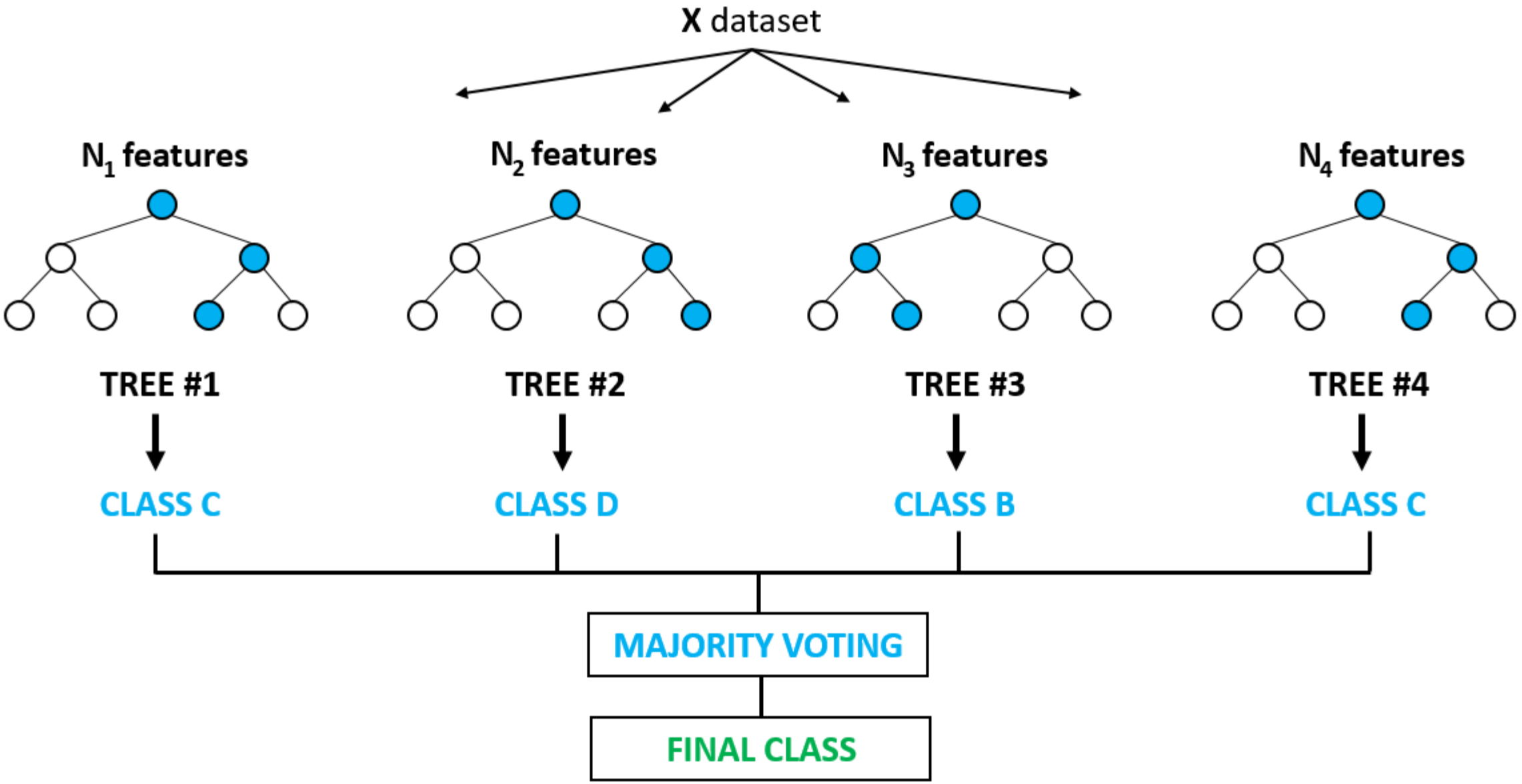
LEC.ASSIST : AHMED YOUSRY

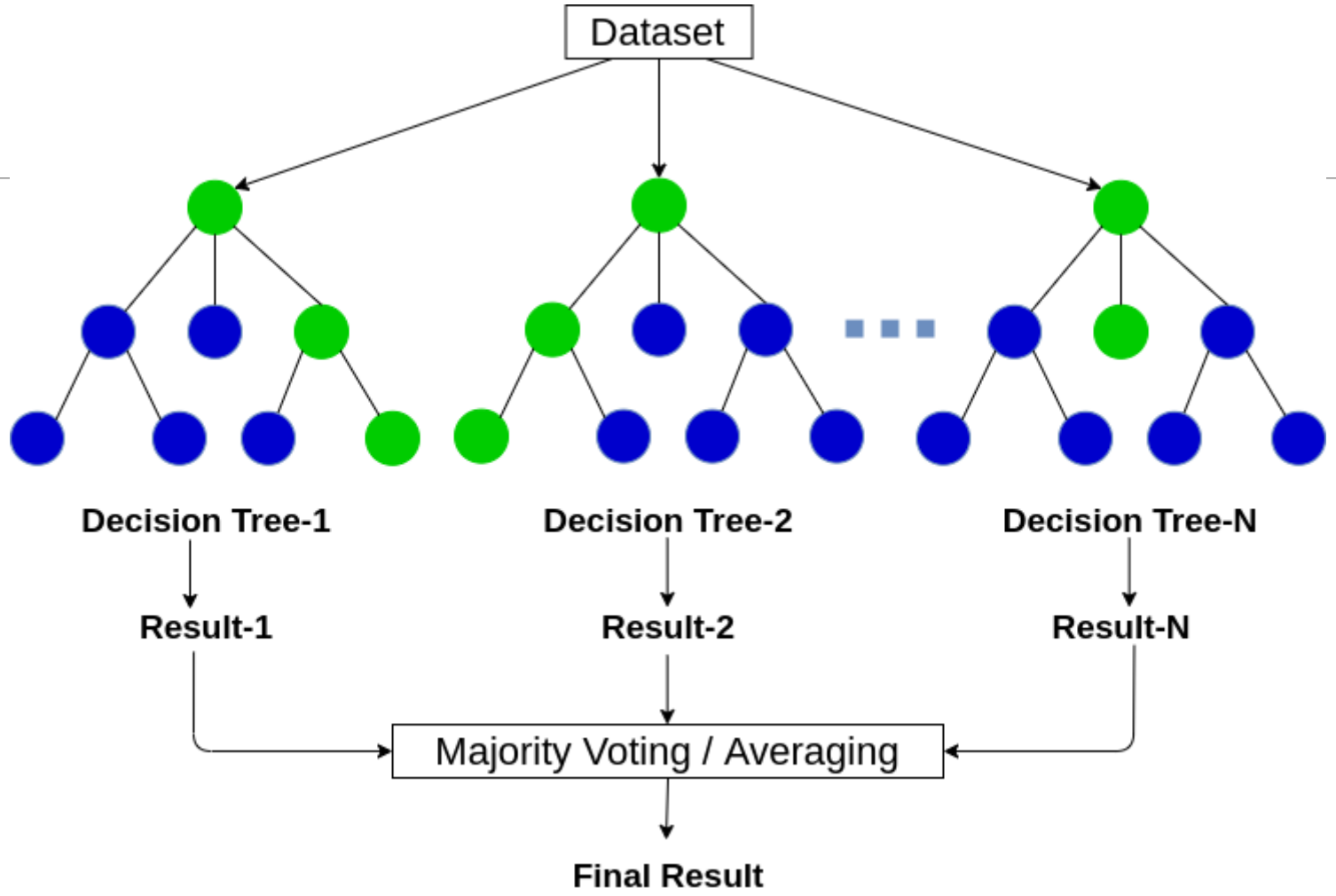
What is Random forest

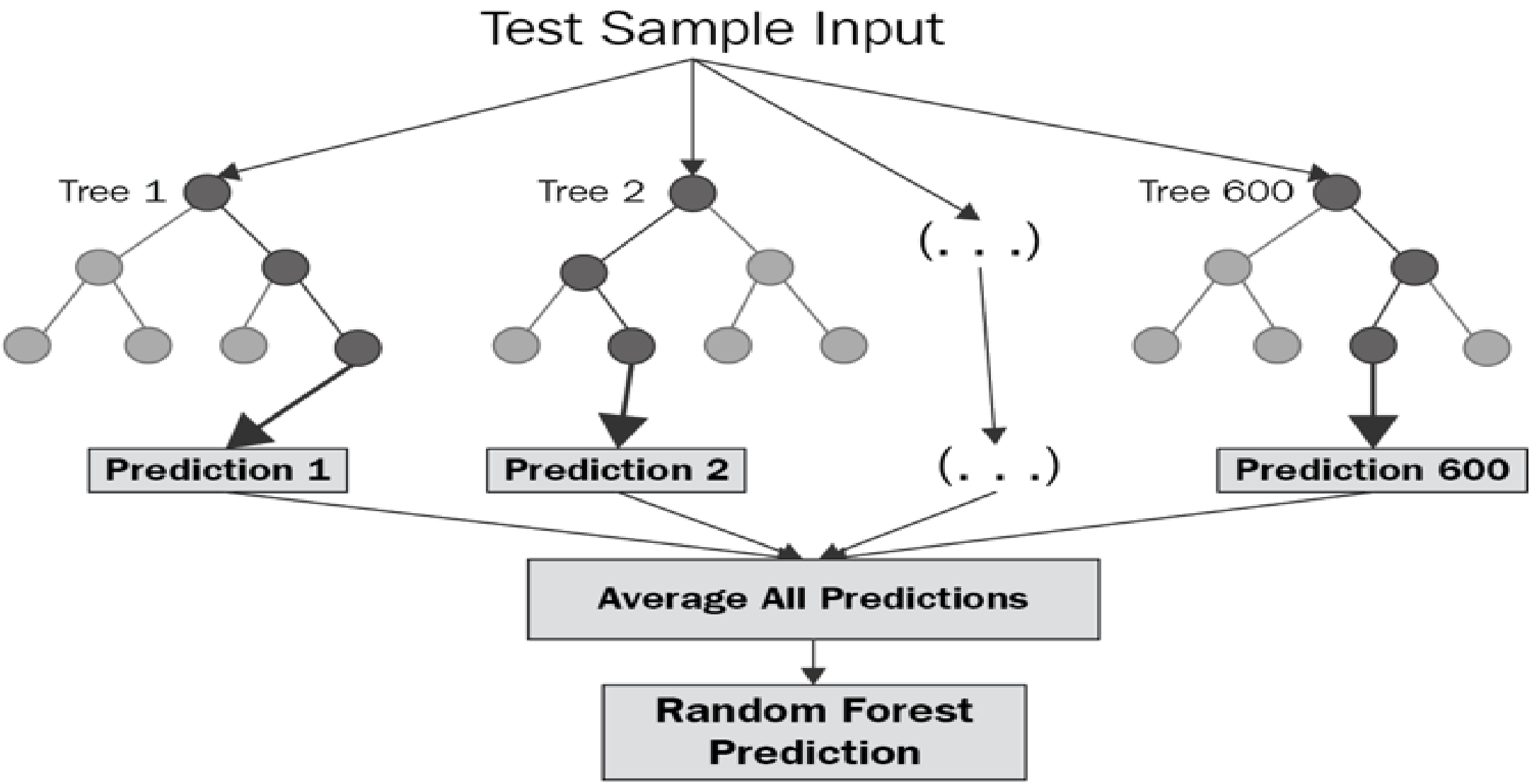
- Random forest is a supervised learning algorithm.
- The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method.
- The general idea of the bagging method is that a combination of learning models increases the overall result.

How it works

- Put simply: Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.
- One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems.


X dataset
N1 features
N2 features
N3 features
N4 features
TREE #1
TREE #2
TREE #3
TREE #4
CLASS C
CLASS D
CLASS B
CLASS C
MAJORITY VOTING
FINAL CLASS


Dataset
Decision Tree-1
Decision Tree-2
Decision Tree-N
Result-1
Result-2
Result-N
Majority Voting / Averaging
Final Result


Test Sample Input
Tree 1
Tree 2
(. . .)
Tree 600
Prediction 1
Prediction 2
(. . .)
Prediction 600
Average All Predictions
Random Forest Prediction

Difference between DT and RF

- The random forest algorithm randomly selects observations and features to build several decision trees and then averages the results.
- Another difference is "deep" decision trees might suffer from overfitting.
- Most of the time, random forest prevents this by creating random subsets of the features and building smaller trees using those subsets.

Thanks