

DECISSION TREES

RANDOM FOREST

Mr.Gangadhar Immadi

immadi.gangadhar@gmail.com

9986789040

Random Forest Classifier

- Construct **multiple decision trees** rather than only one tree which outperform over single tree
- It is an **ensemble** method of decision trees generated on a randomly split dataset.
- Also known as the **forest**.
- The individual decision trees are generated using information **gain, gain ratio, and Gini index** for each attribute.
- Each tree **depends on an independent random sample**.
- Each tree votes and the most **popular class** is chosen as the final result.
- It is **simpler and more powerful** compared to the other non-linear classification algorithms.

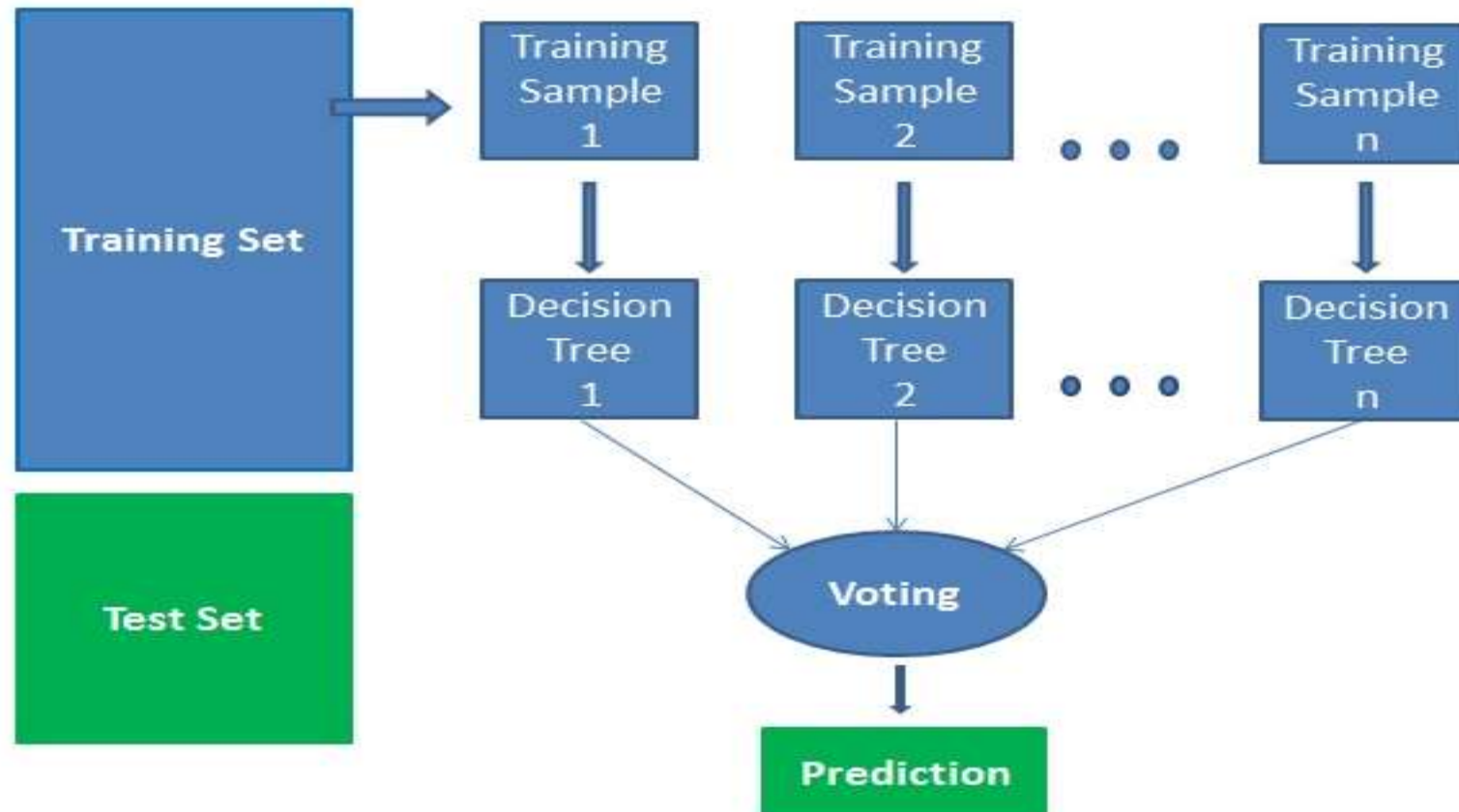
- **Definition** :A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.
- The reason for this wonderful effect is that the trees **protect each other** from their **individual errors**

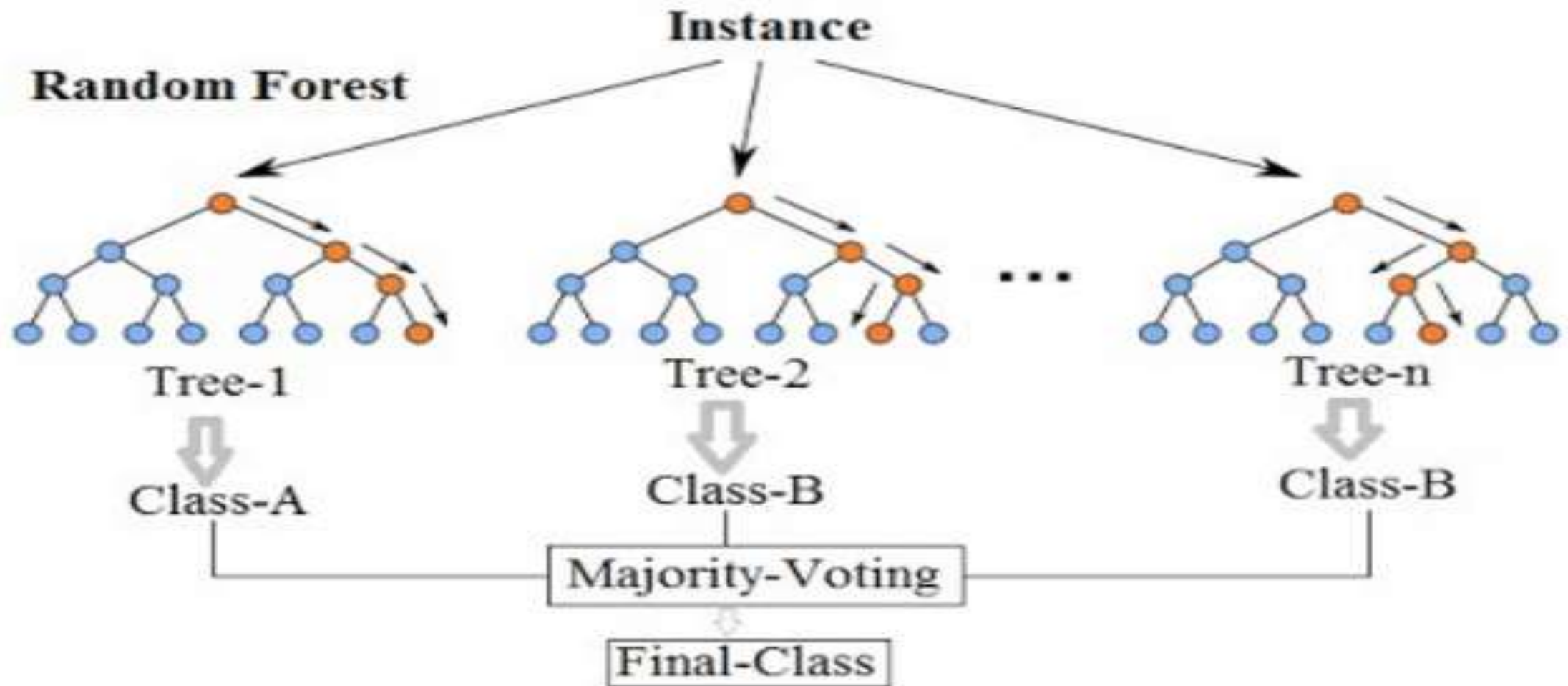
Prerequisites :

- There needs to be some actual signal in our features – **Feature Selection**
- The predictions (and therefore the errors) made by the individual trees need to have **low correlations** with each other.

Algorithm

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





Advantages

- Highly accurate and robust method
- It does not suffer from the over fitting problem. The main reason is that it takes the average of all the predictions, which cancels out the biases.
- Used in both classification and regression problems.
- Random forests can also handle missing values. There are two ways to handle these: Median & weighted average of missing values.
- Facility to select suitable features

Disadvantages

- Slow in generating predictions
- Difficult to interpret compared to a decision tree

Random Forests vs Decision Trees

Decision Trees	Random Forest
Computationally Faster	Computationally Slow
Suffers from Over fitting	Less or No Over fitting
Single Tree	Multiple Tree
Easy to interpret the predictions	Difficult to interpret
Decision rules can be generated	Decision rules cannot be generated