Day 51 Random Forest Concept By: Loga Aswin

Overview:

- Random forests stand out as a well-known supervised machine learning algorithm.
- They find application in supervised machine learning scenarios characterized by a labeled target variable.
- Suitable for both regression (numeric target variable) and classification (categorical target variable) problems.
- As an ensemble method, random forests consolidate predictions from multiple models.
- The constituent models within the random forest ensemble are decision trees.

Why we use Random Forest?

- It provides higher accuracy through cross validation.

- Random forest algorithm can be used for both classifications and regression task.

- It has the power to handle a large data set with higher dimensionality.

- If there are more trees, it won't allow over-fitting trees in the model.

- Random forest classifier will handle the missing values and maintain the accuracy of a large proportion of data.

- It has the power to handle a large data set with higher dimensionality.

Working of Random Forest Algorithm:

The Random Forest operates through a two-phase process. Initially, it creates the random forest by combining N decision trees, and subsequently, it makes predictions for each tree formed in the first phase.

The working process is outlined in the following steps and diagram:

Step-1: Random Data Selection

Randomly select K data points from the training set.

Step-2: Decision Tree Construction

Build decision trees associated with the selected data points, forming subsets.

Step-3: Decision Tree Number Specification

Choose the number N for decision trees that need to be constructed.

Step-4: Iterative Process

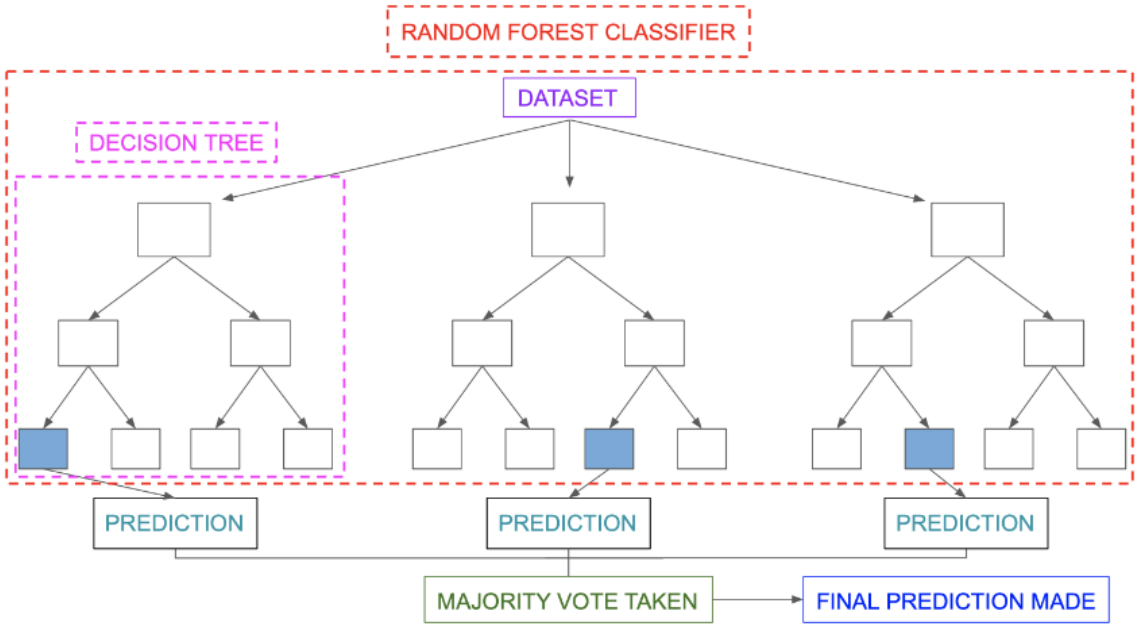
Repeat Steps 1 and 2 for the specified number of decision trees (N).

Step-5: Prediction Aggregation

For new data points, obtain predictions from each decision tree.

Assign the new data points to the category that receives the majority votes among the decision trees.

This process ensures the creation of a robust and diverse ensemble of decision trees, contributing to the Random Forest's predictive power.



Let’s take an another Example ….

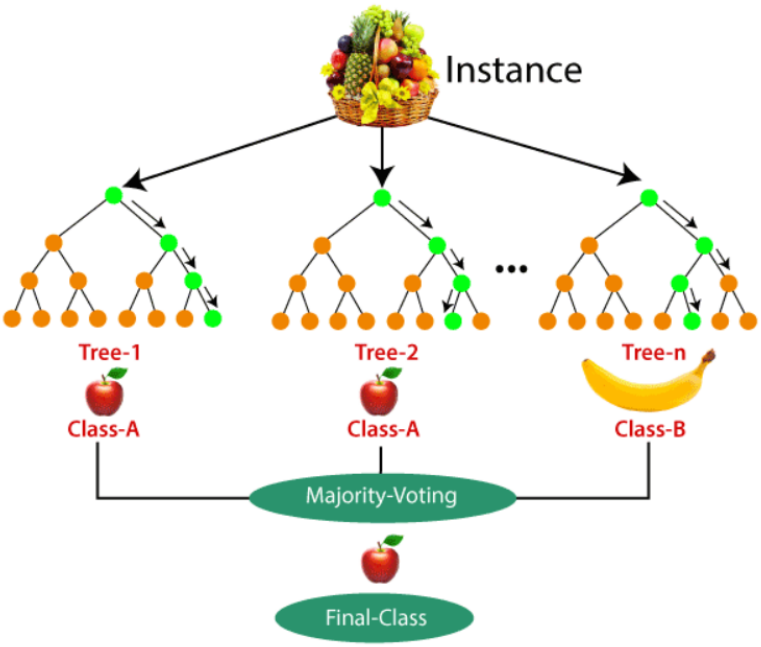


Image Source: Javatpoint

Applications of Random Forest:

- Banking sector: Identification of loan risks.
- Medicine: Disease trend identification and risk assessment.
- Land Use: Identification of areas with similar land use.
- Marketing: Identification of marketing trends.

Advantages of Random Forest:

- Performs Classification and Regression tasks.
- Handles large datasets with high dimensionality efficiently.
- Enhances model accuracy and prevents overfitting.
- Provides feature importance for better insights.

Disadvantages of Random Forest:

- Less suitable for regression tasks.
- Reduced interpretability due to ensemble nature.

[Note: For the Implementation Part: Visit My Github Page]