# **Algorithm Assignment – Random Forest**

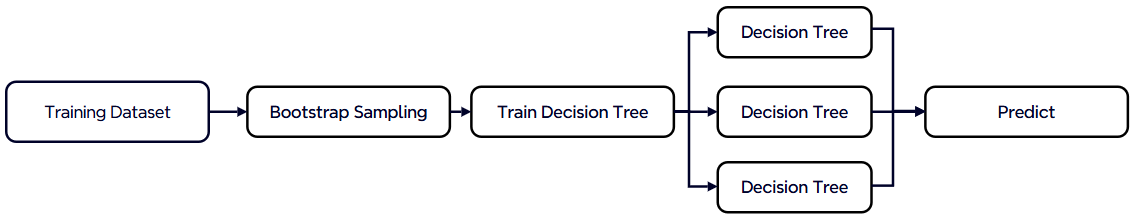
## **Introduction**

Random Forest is a powerful ensemble learning method that combines multiple decision trees to improve prediction accuracy and robustness. It follows the Bagging (Bootstrap Aggregating) principle by creating diverse subsets of training data through random sampling with replacement. Each tree is trained independently and uses a randomly selected subset of features at each split.

For classification, the final output is determined by majority voting; for regression, it is the average of all tree outputs. Random Forest effectively reduces overfitting and performs well on both structured and unstructured data.

## **Workflow of the Random Forest Algorithm**

The training dataset is used to generate multiple bootstrap samples. Each sample trains a decision tree using a random subset of features at each split. The final prediction is made by aggregating the outputs of all trees using majority voting (for classification) or averaging (for regression).



## **Core Mechanism of Random Forest**

The following table summarizes the key mechanisms that make Random Forest effective as an ensemble learning algorithm.

|  |  |
| --- | --- |
| **Component** | **Description** |
| Base learner | Each tree in the forest is a decision tree, typically grown without pruning. |
| Data diversity | Bootstrap sampling generates different datasets for training each tree. |
| Feature diversity | At each split, only a random subset of features is considered to find the best split. |
| Aggregation | For classification, predictions are made by majority vote; for regression, by averaging outputs. |

## **Comparison: Decision Tree vs. Random Forest**

The table below highlights the main differences between a single decision tree and the Random Forest algorithm.

|  |  |  |
| --- | --- | --- |
| **Aspect** | **Decision Tree** | **Random Forest** |
| Overfitting | Easy to overfit | Less likely to overfit |
| Accuracy | Moderate | High |
| Interpretability | High | Medium to low |

## **Equations**

**Classification**:

**Regression**: