## **Random Forest**

Random Forest is a popular machine learning algorithm used for both classification and regression tasks. It's an ensemble learning method, which means it combines multiple simpler models to create a more powerful and robust predictive model.

#### **Ensemble of Decision Trees**

- A Random Forest is made up of many individual decision trees.
- Each tree is trained on a random subset of the data and features.

# **Bagging (Bootstrap Aggregating)**

- Random Forest uses a technique called bagging to create diverse trees.
- For each tree, a random sample of the training data is selected with replacement (bootstrap sampling).

#### **Feature Randomness**

- At each node split in a tree, only a random subset of features is considered.
- This introduces additional randomness and helps to decorrelate the trees.

# **Voting/Averaging**

- For classification tasks, each tree "votes" for a class, and the majority vote wins.
- For regression tasks, the predictions of all trees are averaged.

Random Forest 1

### **Hyperparameters**

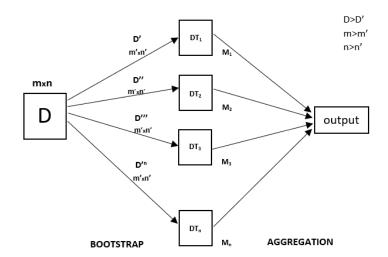
- · Number of trees in the forest.
- Maximum depth of trees.
- Minimum number of samples required to split an internal node.
- · Number of features to consider for the best split.

## Out-of-Bag (OOB) Error

 Random Forest can use the samples not included in the bootstrap sample for each tree to estimate the model's performance without a separate validation set.

### **Advantages**

- Reduces overfitting compared to individual decision trees.
- Handles high-dimensional data well.
- Can capture complex interactions between features.
- · Provides feature importance rankings.
- · Robust to outliers and noise.



Random Forest 2