Random Forest Explained

Random Forest is a popular ensemble machine learning algorithm, primarily used for classification and regression tasks. It builds multiple decision trees (hence the term "forest") and merges their results to improve accuracy and control overfitting.

The key concepts behind Random Forest are:

Ensemble of Decision Trees: Instead of relying on a single decision tree, Random Forest creates multiple decision trees during training. Each tree is trained on a random subset of the training data (both features and samples).

Bagging: This is a technique used to create each tree from a different random subset of data. It reduces variance by averaging multiple predictions.

Voting/ Averaging: For classification, the final result is based on a majority vote of the trees' individual predictions. For regression, it takes the average of predictions from all trees.

Advantages:

Reduces overfitting compared to a single decision tree.

Works well with large datasets and can handle missing data.

Handles both classification and regression problems.