

Random Forest Algorithm:

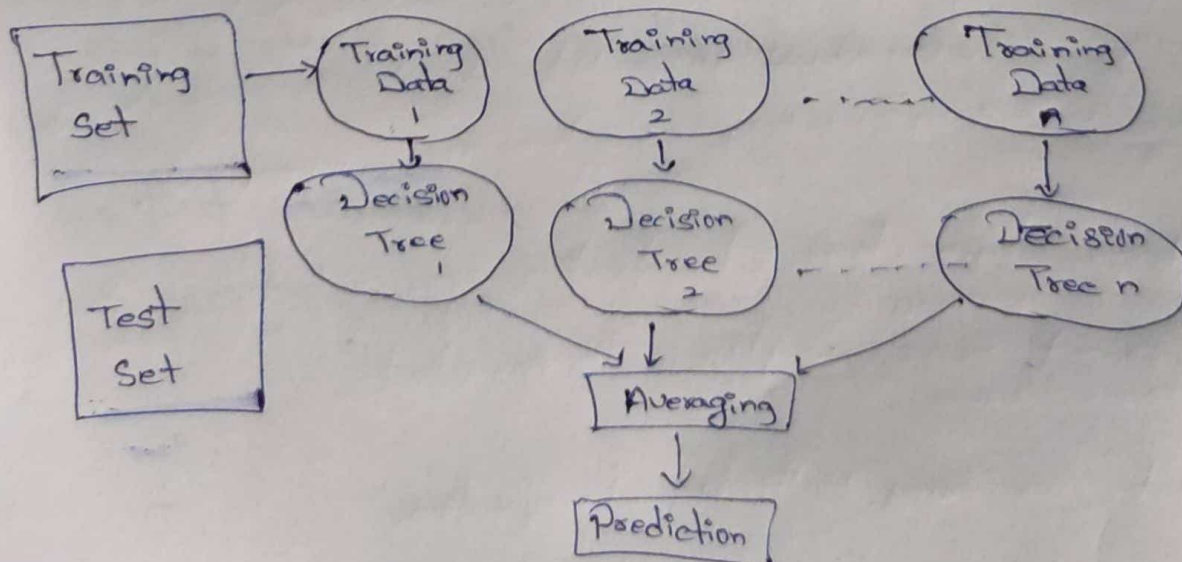
- Supervised Machine Learning Algorithm
- Used for both classification & Regression
- Based on the Concept of Ensemble Learning, which is a process of Combining Multiple classifiers to solve a Complex problem & To Improve the performance of the Model.

Random Forest ⇒ Classifier that Contains a number of Decision trees on various subsets of the given dataset and takes the average to Improve the predictive Accuracy of that Dataset.

- Takes the Prediction from each tree & based on the Majority votes of predictions, and it predicts the final Output.

→ Greater Number of trees in the forest Leads to higher Accuracy & prevents the problem of Overfitting.

Working:



Why use Random Forest?

- Takes Less Training time
- Predicts Output with high Accuracy.
- Maintain Accuracy when a large proportion of Data is Missing.

process:

- Step 1 → Select Random k Data points from the Training Set.
- Step 2 → Build the Decision Tree associated with the Selected Data points.
- Step 3 → Choose the number N for Decision Trees that want to build.
- Step 4 → Repeat step 1 & 2.
- Step 5 → For new Data points, find the predictions of each Decision Tree, & Assign the new Data points to the Category that wins the Majority votes.

Advantages:

- Capable of performing both Classification & Regression tasks.
- Capable of Handling Large Datasets with high Dimensionality.
- Enhances the Accuracy of the Model & prevents the Overfitting issue.

Implementation steps:-

- 1) Data pre-processing step
- 2) Fitting the Random forest Algorithm to the Training Set.
- 3) Predicting the test result.
- 4) Test Accuracy of the result.
- 5) Visualizing the test set result.