



SNS COLLEGE OF TECHNOLOGY

Coimbatore-37.

An Autonomous Institution



COURSE NAME : 19CSE301-INTRODUCTION TO DATA SCIENCE

III YEAR/ V SEMESTER

UNIT – V REPLICABILITY

Topic: Random Forest

Mrs.B.Sumathi

Assistant Professor

Department of Computer Science and Engineering



Introduction

- Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique.
- It can be used for both Classification and Regression problems in ML.
- It is based on the concept of **ensemble learning**, which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

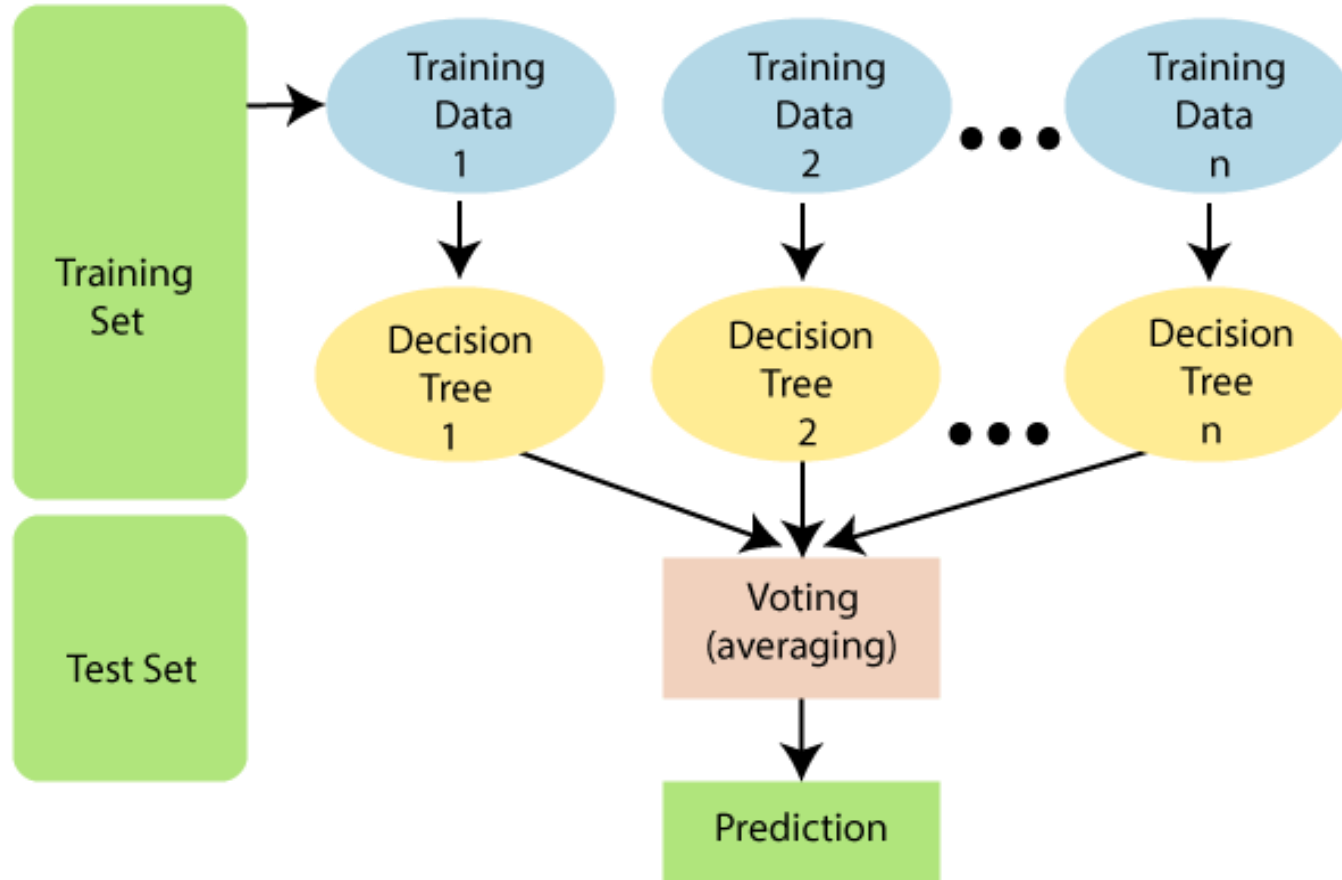


Introduction

- *"Random Forest is a 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."*
- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.



Example





Assumption & Steps

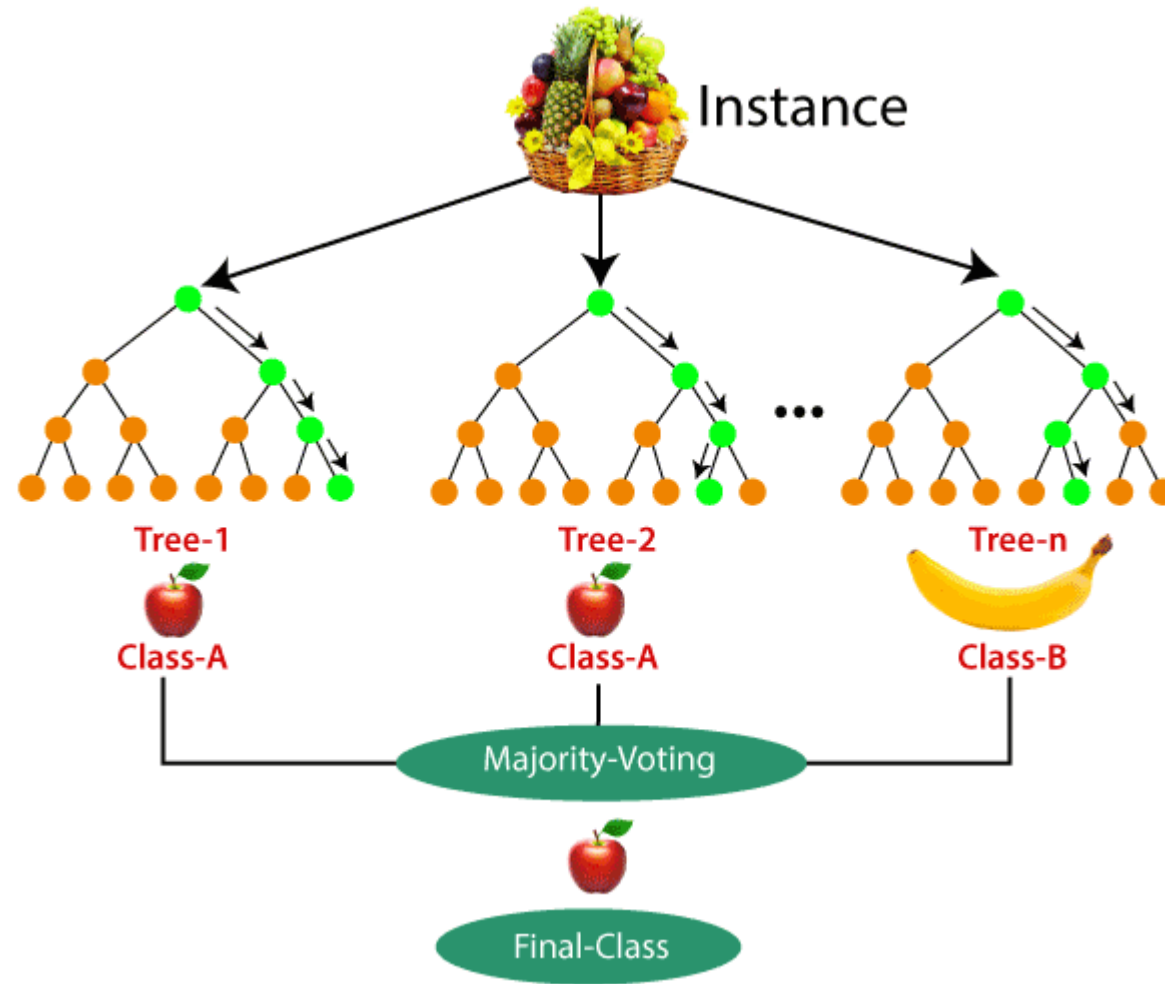
- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

Steps:

- **Step-1:** Select random K data points from the training set.
- **Step-2:** Build the decision trees associated with the selected data points (Subsets).
- **Step-3:** Choose the number N for decision trees that you want to build.
- **Step-4:** Repeat Step 1 & 2.
- **Step-5:** For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.



Example





Advantage and Disadvantage

Advantage:

- Random Forest is capable of performing both Classification and Regression tasks.
- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

Disadvantage:

- Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.



References

- 1 Tom M. Mitchell, “Machine Learning”, McGraw-Hill Education (India) Private Limited, 2013.
- 2 Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning: with Applications in R”, Springer; First Edition 2013.
- 3 P. Flach, —Machine Learning: The art and science of algorithms that make sense of data, Cambridge University Press, 2012.



Thank
You