

DATA SCIENCE WITH PYTHON : RANDOM FOREST ALGORITHM #1038

NAME: DEEPTHI M

SERIAL NUMBER: 172

BATCH NUMBER: 05

Random Forest Algorithm

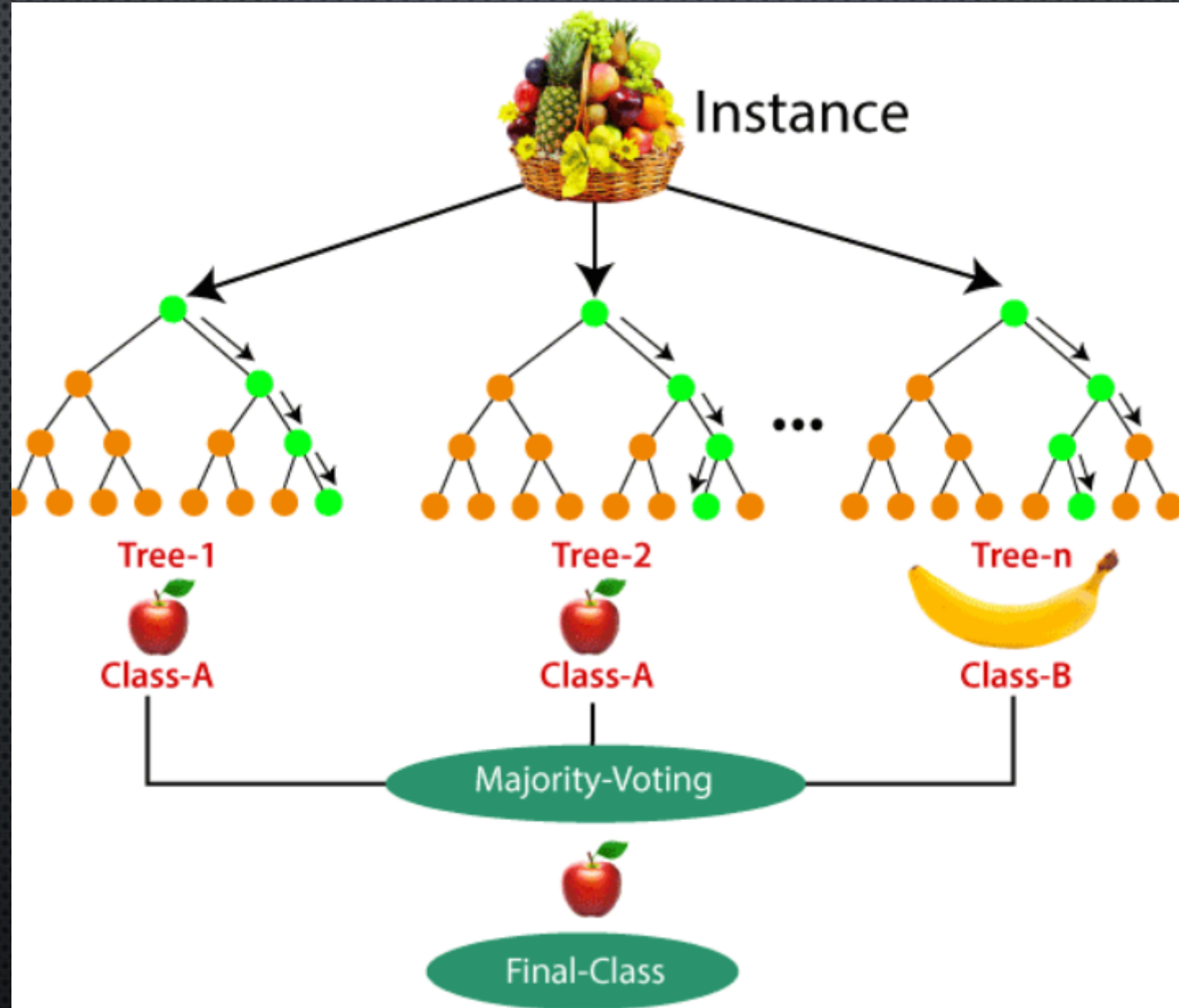
- It is a supervised learning model.
- It can be used for both Classification and Regression problems.
- It is based on the concept of ensemble learning.
- Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Two assumptions for a better Random forest classifier:

- 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.

Why use Random Forest?

- Less training time.
- Predicts output with high accuracy.
- It runs efficiently.
- It works well when data is missing.
- In order to prevent the overfitting, we use random forest model.



How does it works:

- It builds the decision tree and when new data point is added it assigns the categories which has the maximum votes.
- A decision tree is build on decisions or conditions.
- It builds various decisions tress.
- when new data point is added it assigns the categories which has the maximum votes.

Advantages:

- It is capable of performing both Classification and Regression tasks.
- It performs well even if data is missing.
- It enhances the accuracy of the model.
- Prevents the overfitting issue.

Disadvantages:

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