



# POORNIMA FOUNDATION

## LECTURE NOTES

Campus: PCE Course: BTECH in CSE Class/Section: III Yr. Section- A Date: 28-02-21  
Name of Faculty: Praveen Kumar Yadav Name of Subject: Machine Learning Code: 6CS4-02  
Date (Prep.): 28-02-21 Date (Del.): 18-03-21 Unit No.: 1 Lect. No.: 69

**OBJECTIVE:** To be written before taking the lecture (Pl. write in bullet points the main topics/concepts etc., which will be taught in this lecture)

Random Forest Algorithm in ML

**IMPORTANT & RELEVANT QUESTIONS:**

what are disadvantage of decision tree?

**FEED BACK QUESTIONS (AFTER 20 MINUTES):**

why Random Forest algo<sup>n</sup> is required?

**OUTCOME OF THE DELIVERED LECTURE:** To be written after taking the lecture (Pl. write in bullet points about students' feedback on this lecture, level of understanding of this lecture by students etc.)

good.

**REFERENCES:** Text/Ref. Book with Page No. and relevant Internet Websites:

scikit learn with ML



# POORNIMA

## COLLEGE OF ENGINEERING

### DETAILED LECTURE NOTES

PAGE NO. ....

#### Random Forest Algorithm in Machine Learning:-

A random forest is actually, just a bunch of decision trees bundled together (that's why it's called a forest).

For eg- Consider a following dataset

COLOR	DIAMETER	LABEL
RED	3	APPLE
YELLOW	3	LEMON
PURPLE	1	Grapes
RED	3	APPLE
YELLOW	3	LEMON
PURPLE	1	GRAPES

CONDITIONS

COLOR == PURPLE?

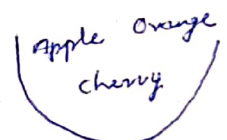
DIAMETER = 3

COLOR == YELLOW?

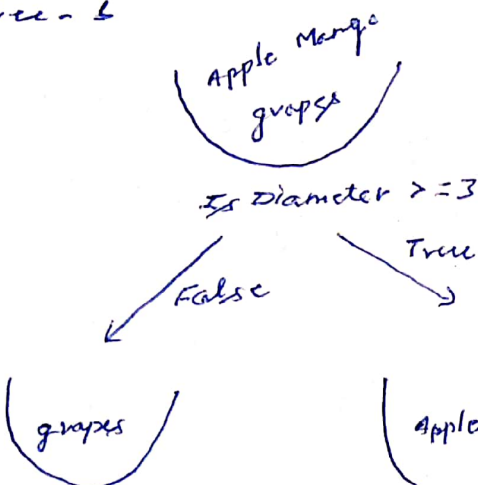
COLOR == RED?

DIAMETER = 1

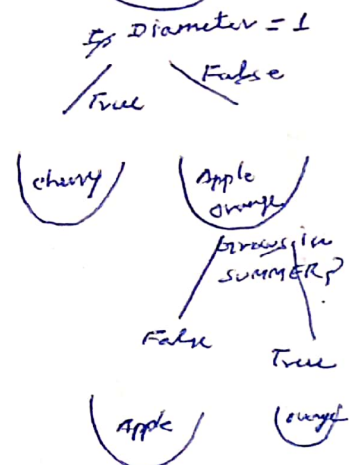
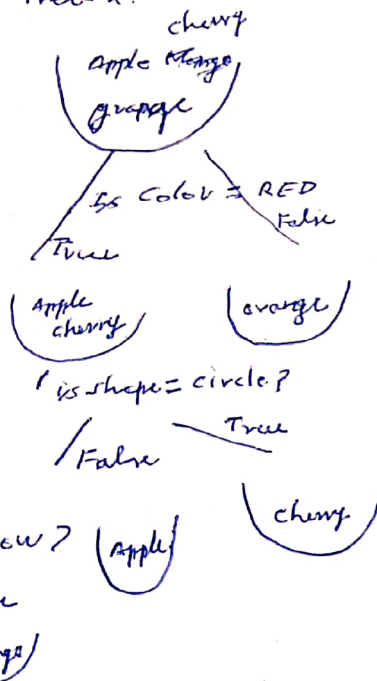
Tree-3



Tree-1



Tree-2



Algo<sup>m</sup>. Sample with replacement,  $n$  training examples from the dataset.

Step 2. Train a decision tree on the  $n$ -samples.

Step 3. Repeat  $t$ -times for same  $T$ .

To make a predictions using this model with  $t$ -trees, we aggregate the predictions from the individual decision trees

- Take the majority vote if our trees produce class labels (like colors).

- Take the average if our trees produce numerical values (e.g. when predicting temperature, price etc).

This technique is called as **bagging** or **bootstrap aggregating**.

**Bagging**- Bagged decision trees have only one param- $t$ , the no. of trees.

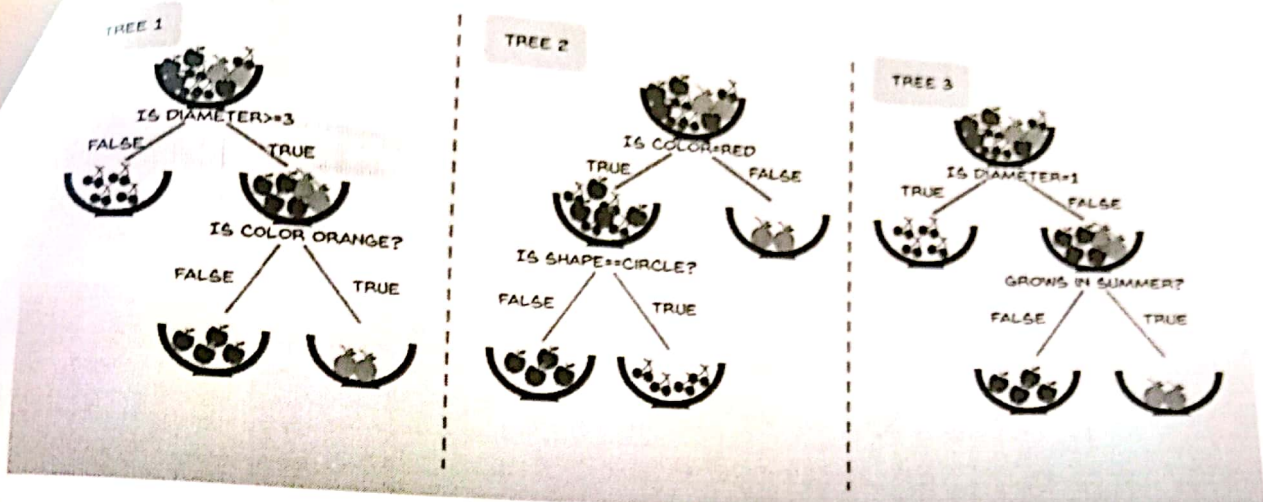
Random Forest have a second parameter that controls how many features to try when finding the best split.

Here Instead of trying All features every time, we can only try a subset of features usually  $\sqrt{p}$  or  $p/3$ .

This technique is sometime referred as **feature bagging**.



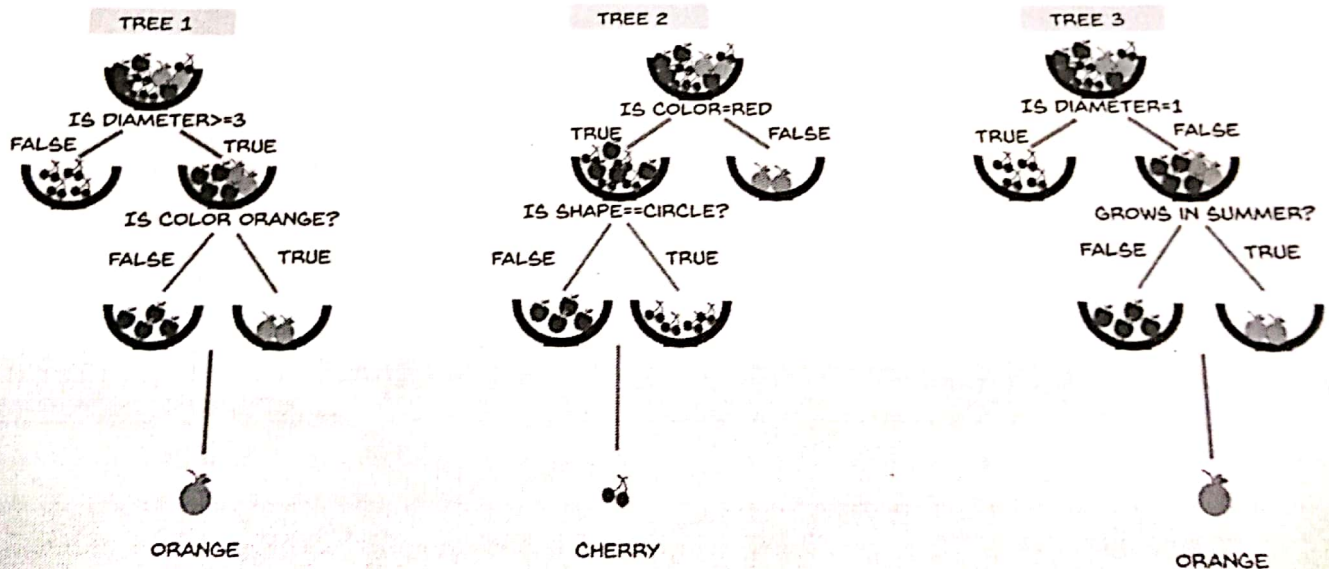
...decision tree would be-



Let's take a new fruit and find the name of the fruit.



DIAMETER = 3  
COLOUR = ORANGE  
GROWS IN SUMMER = YES  
SHAPE = CIRCLE





# POORNIMA

## COLLEGE OF ENGINEERING

### DETAILED LECTURE NOTES

PAGE NO. ....

So a Random forest or Random decision forest is a method that operates by constructing multiple decision tree during training phase.

The decision of the majority of the trees is chosen by random forest as the final decision.

#### Advantages of Random Forest-

1. it reduces overfitting in decision trees and helps to improve the accuracy.
2. Flexible for classification and regression problems.
3. work well for categorical and continuous values.
4. automates missing values present in data.

#### Disadvantage of Random Forest algorithm:-

1. Requires much computational power as well as resources as it build numerous trees to combine their outputs.

2. It also requires much time for training as it combined a lot of decision trees to determine the class.