Decision Tree

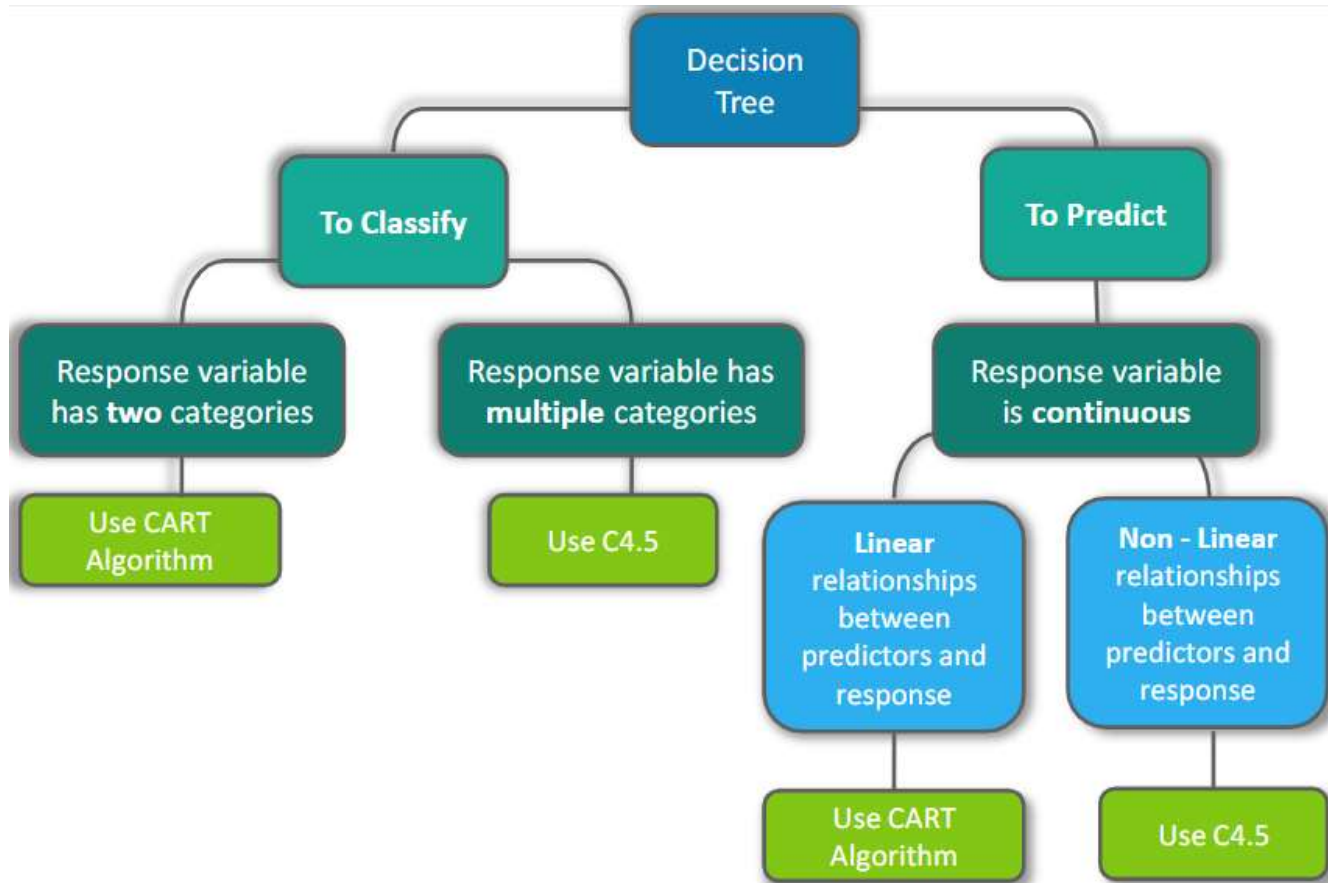| Colour	Diameter	Label
Green	3	Mango
Yellow	3	Mango
Red	1	Grape
Yellow	3	Lemon
Red	1	Grape

Categorical Attribute

Class

Numerical Attribute

The dataset is not perfectly separable as 2nd and

4th example have same feature but different label

When and Where

Decision Tree Terminology

Pruning

Opposite of Splitting, basically pruning is removing of unwanted branches from the tree

Branch/Sub Tree

Formed by splitting the tree/node

Splitting

Dividing the root node/sub node into different parts on the basis of some condition

Parent/Child Node

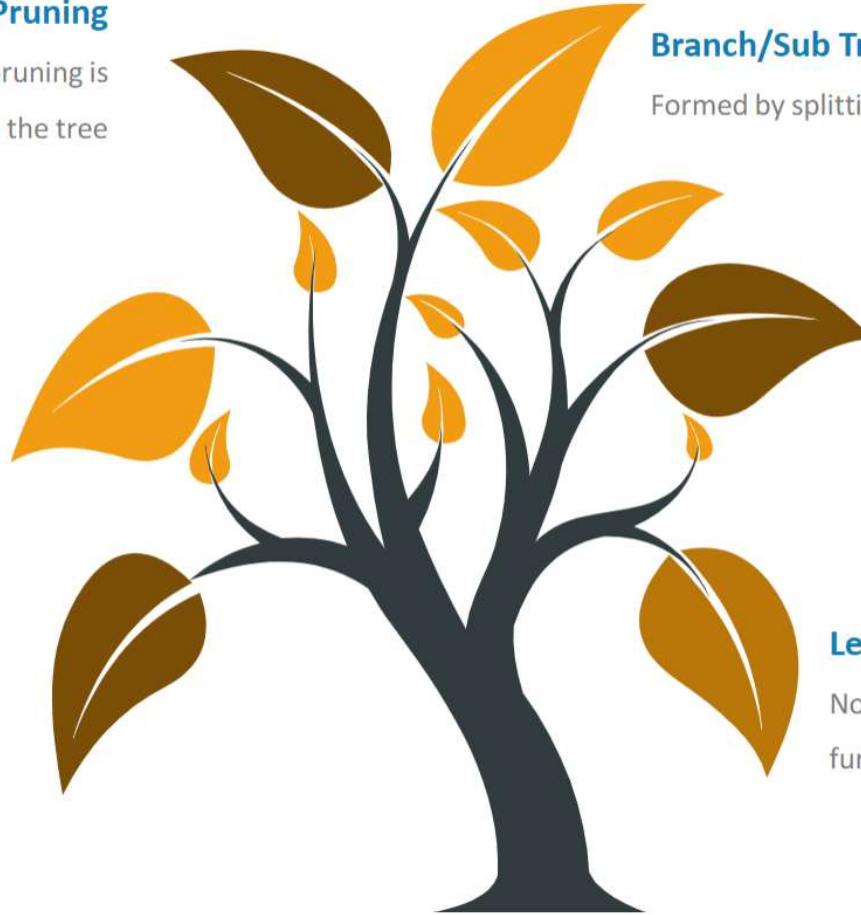Root node is the parent node and all the other nodes branched from it is known as child node

Leaf Node

Node cannot be further segregated into further nodes

Root Node

It represents the entire population or sample and this further gets divided into two or more homogenous sets

Skill AP
A P S S D C

THE RECOMMENDATION ENGINE FOR THE APP STORE OR FOR GOOGLE PLAY

- Our task is to recommend to people the app they're most likely to download, and we should do this based on previous data.

- Our previous data is this table with six people each in a row, and the columns are their gender, male or female, their occupation, work or study, and the app they downloaded.

- The options for the app are Pokémon Go, WhatsApp, and Snapchat.

- So, the model we'll create will take the first two columns and guess the third one.

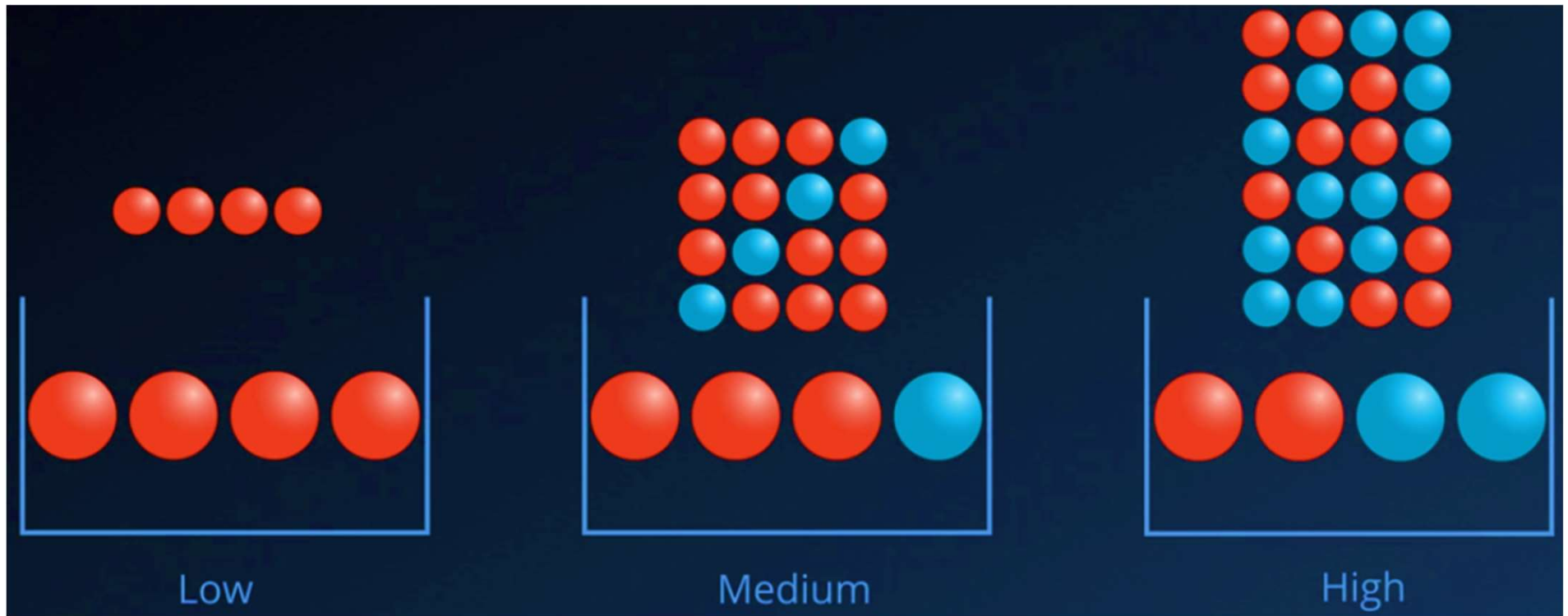| Gender	Occupation	App
F	Study	Pokemon Go
F	Work	WhatsApp
M	Work	Snapchat
F	Work	WhatsApp
M	Study	Pokemon Go
M	Study	Pokemon Go

Entropy

Entropy comes from physics, **Entropy,** as it relates to machine learning, it is a measure of the randomness in the information being processed. The higher the entropy, the harder it is to draw any conclusions from that information. Flipping a coin is an example of an action that provides information that is random
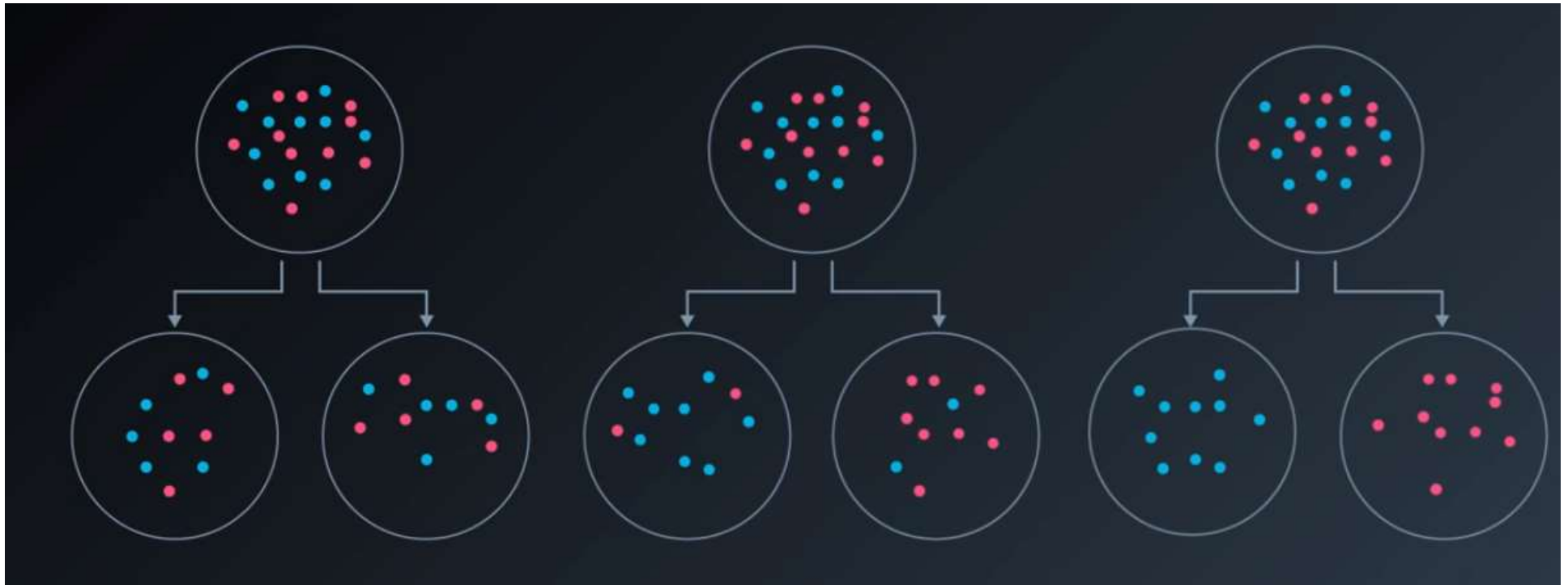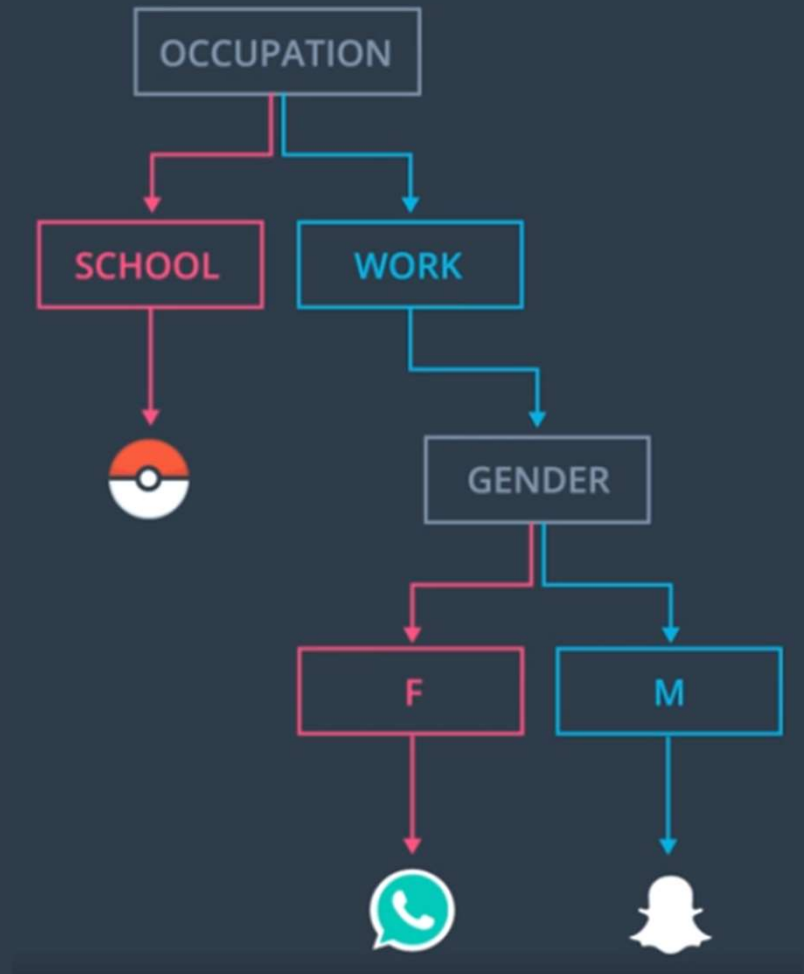
Entropy

Low   Medium   High
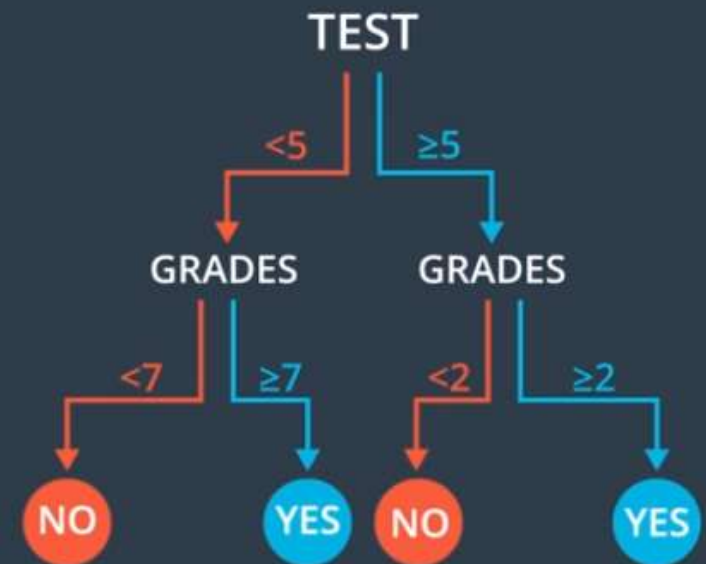
Entropy Formulae
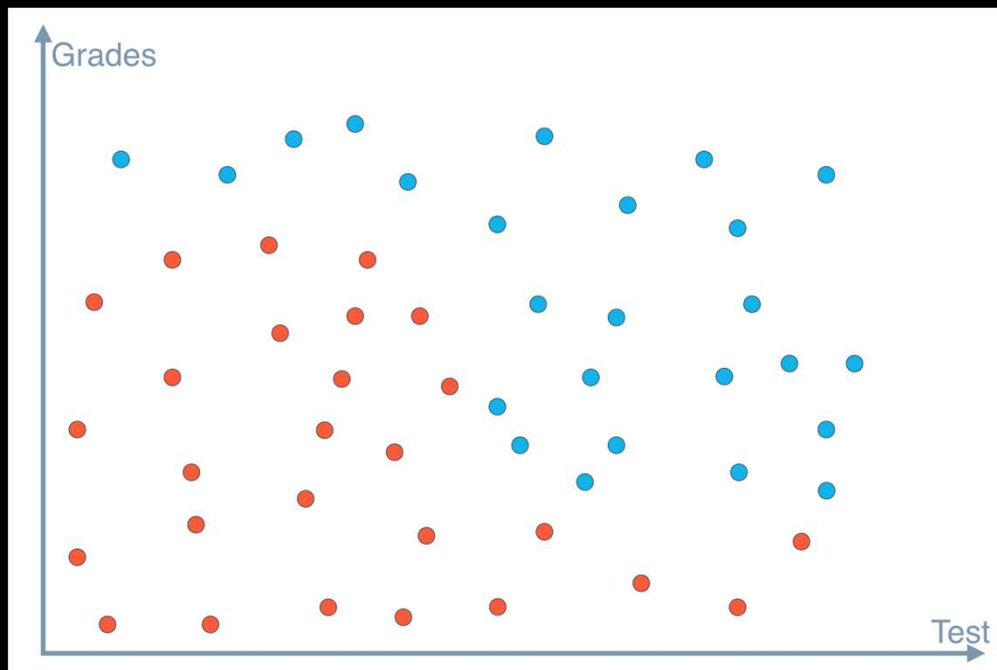
$$-\sum_{i=0}^{n} P_i \log P_i$$

Information Gain

Working Model

Decision tree as regressor

Student data test vs Grades

Decision Tree

Gender	Occupation	App
F	Study	Pokemon Go
F	Work	WhatsApp
M	Work	Snapchat
F	Work	WhatsApp
M	Study	Pokemon Go
M	Study	Pokemon Go

Overfitting Problems in Decision Trees

Random Forrest

Random Forrest

Random Forrest