

Arboles de decisión y Bosques Aleatorios

Hugo Andrés Dorado B.

Juan Camilo Rivera

Ventajas del random forest



No overfitting

Use of multiple trees
reduce the risk of
overfitting

Training time is less



High accuracy

Runs efficiently on large
database

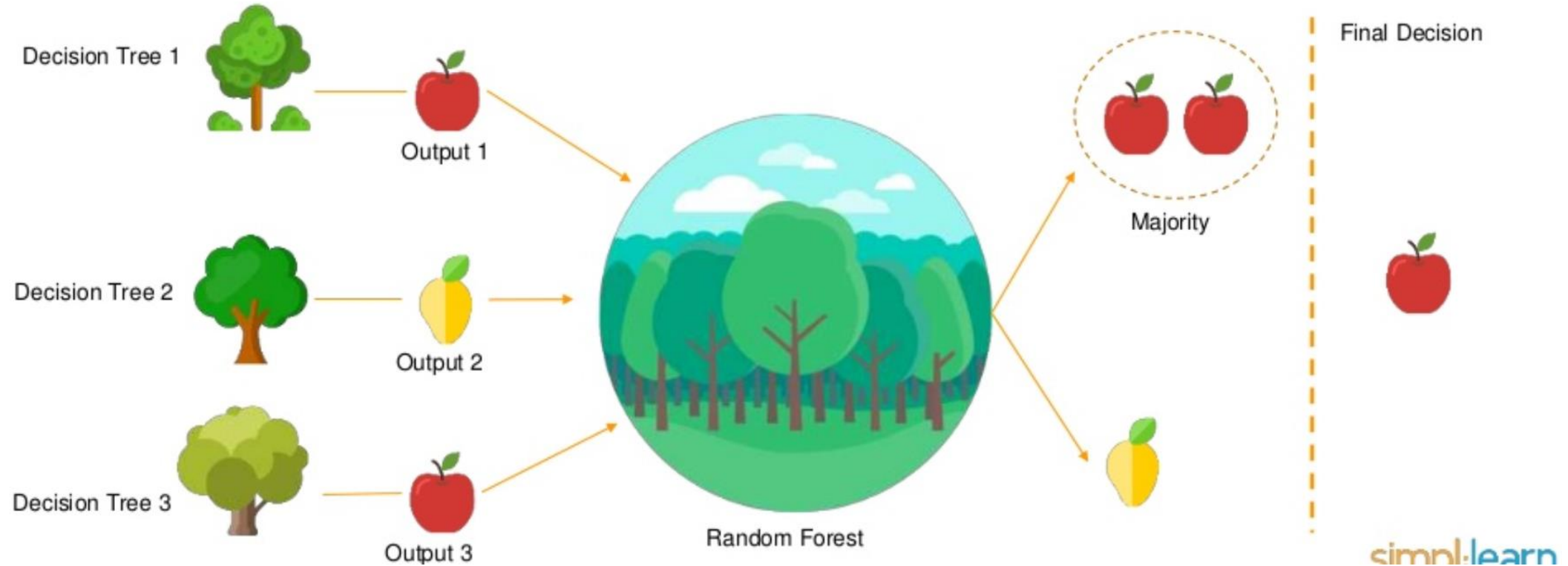
For large data, it
produces highly
accurate predictions



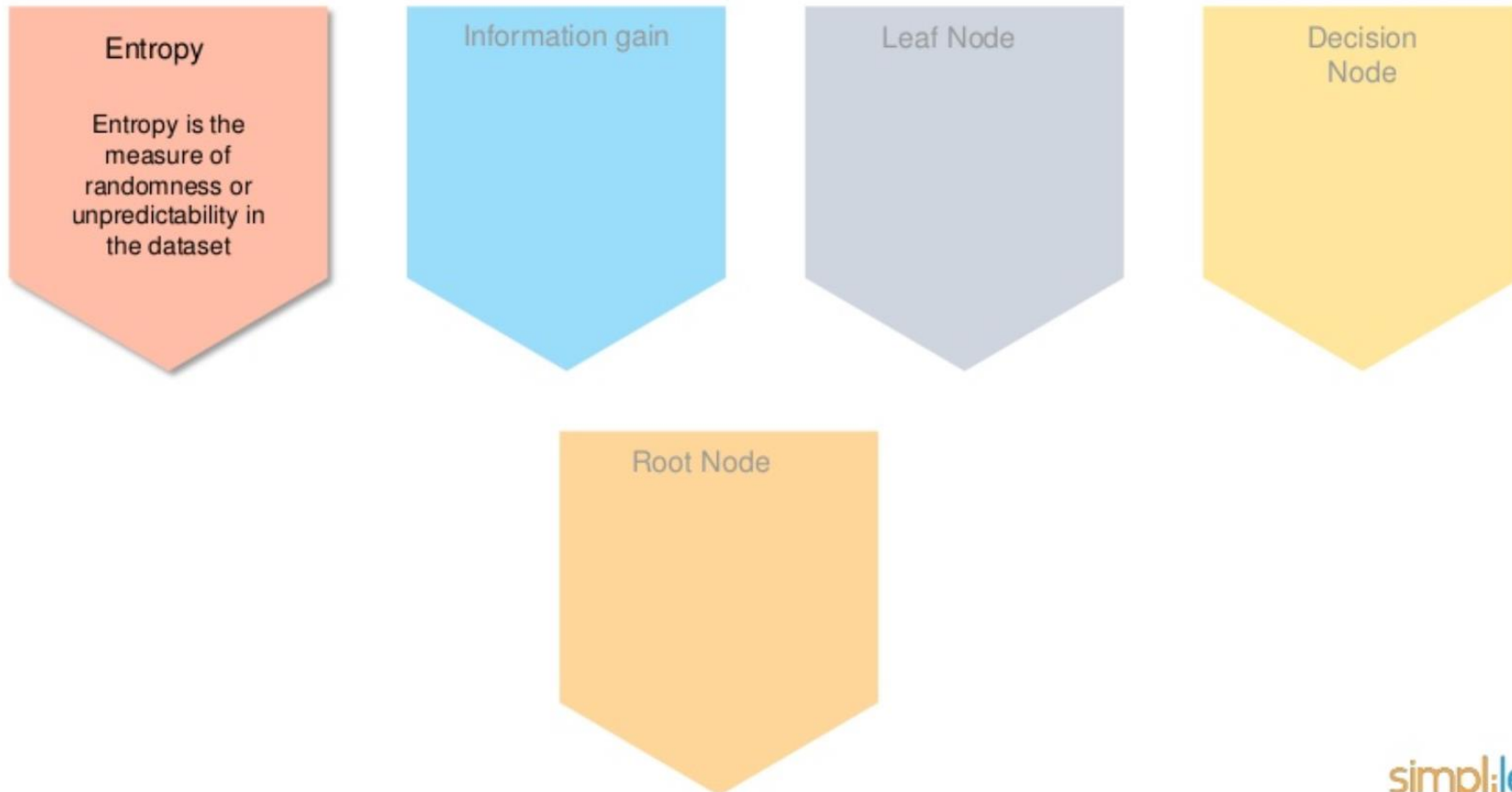
Estimates missing data

Random Forest
can maintain
accuracy when a
large proportion of
data is missing

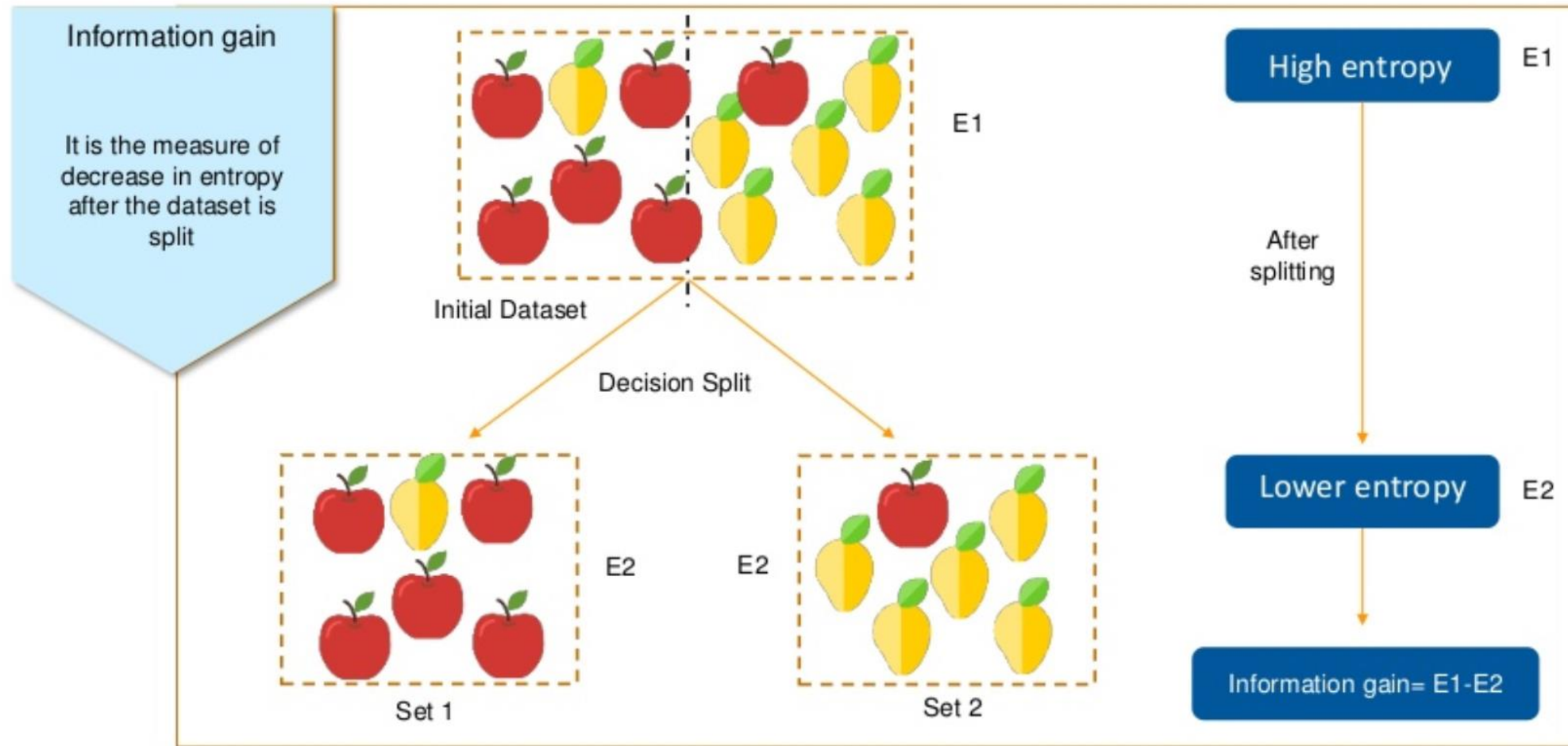
Como funciona el random forest?



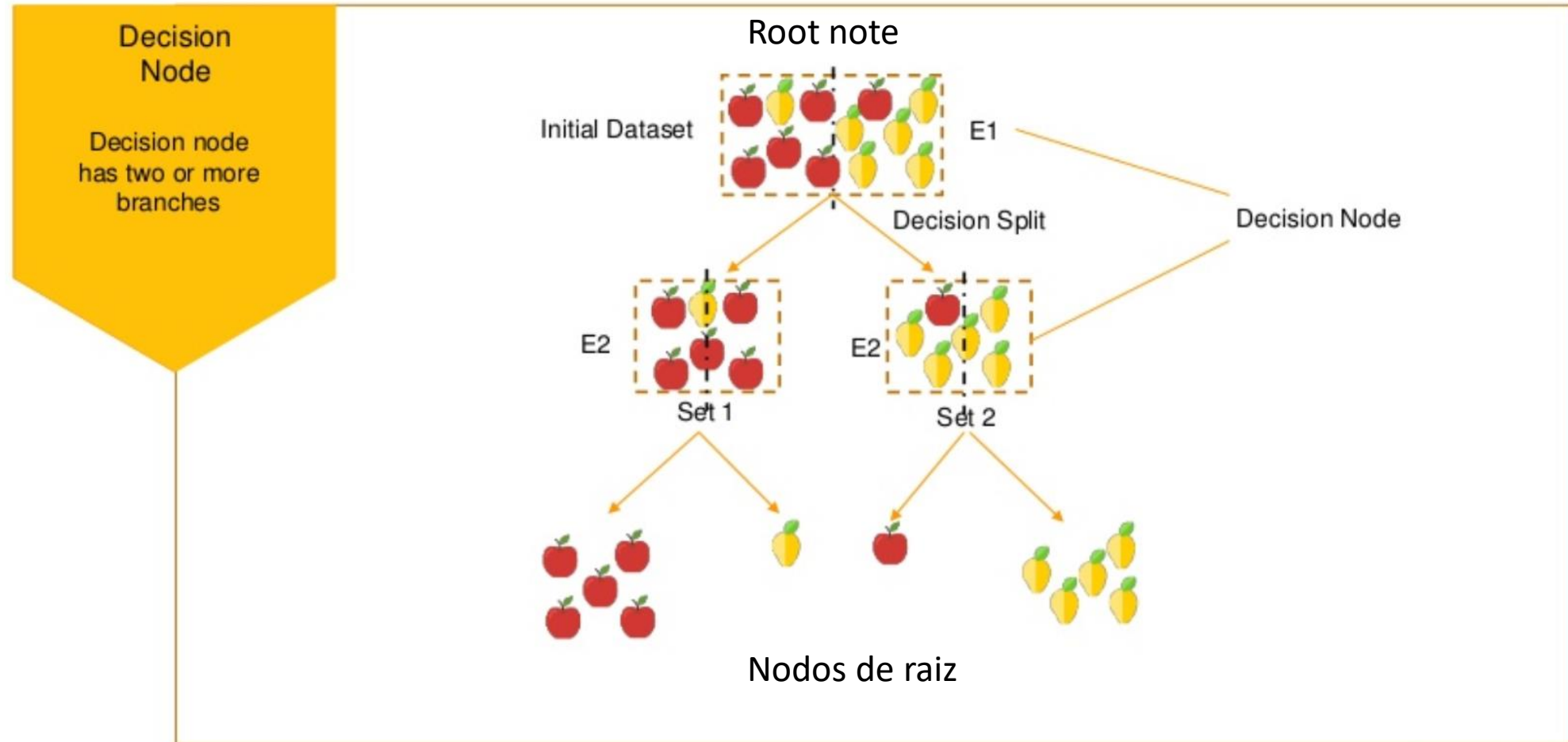
Conceptos importantes de los árboles



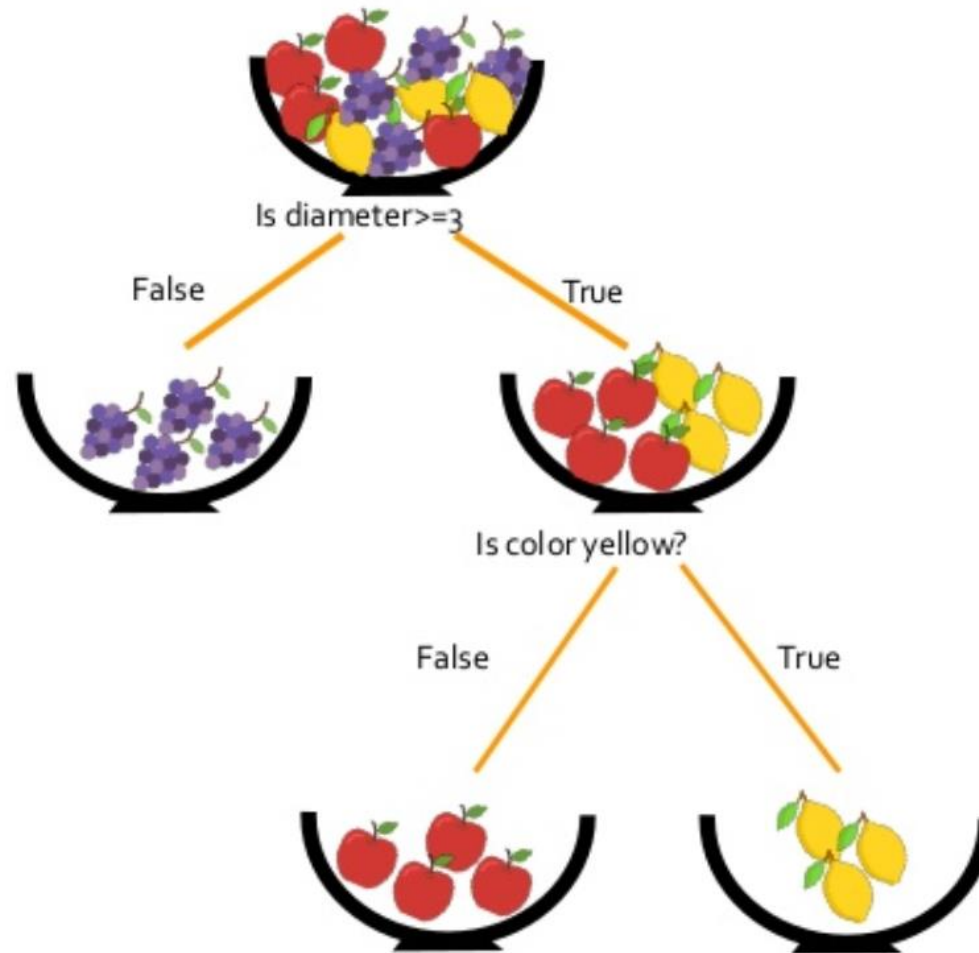
Entropía y Ganancia de Información



Nodos (Raíz, decisión y hojas)



Ejemplo de construcción de un árbol



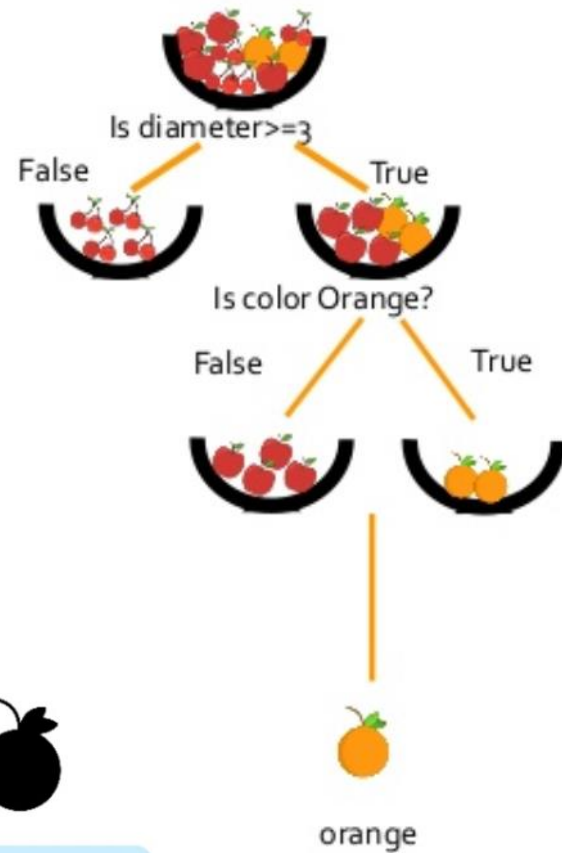
So the entropy in this case is now zero

Random forest

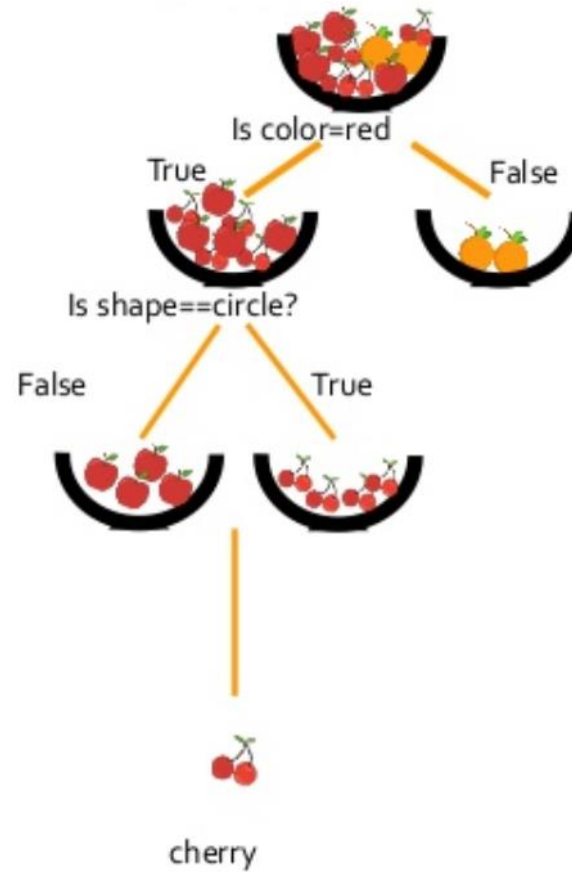
Paramétros:

- mtry: Número de variables
- ntrees: Número de árboels

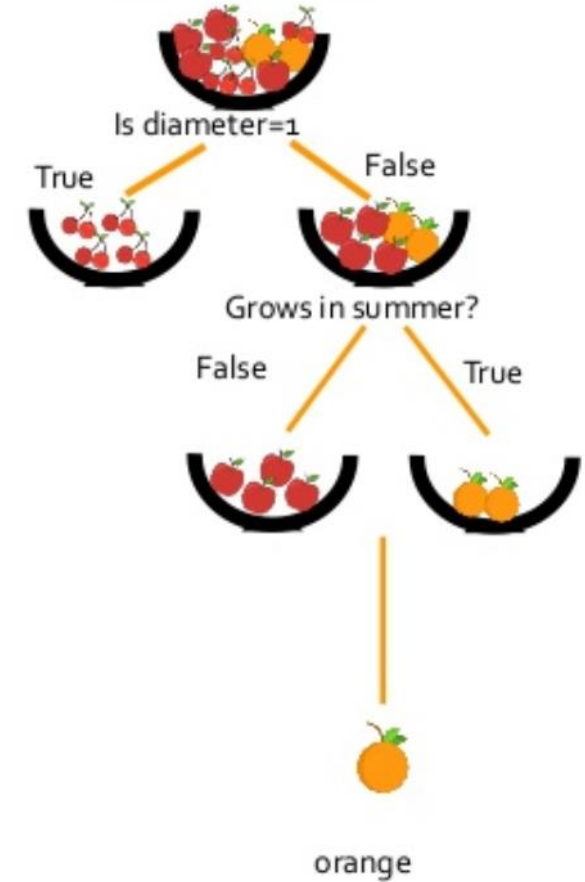
Tree 1



Tree 2



Tree 3



Diameter = 3
Colour = orange
Grows in summer = yes
SHAPE = CIRCLE

Bibliografía.