

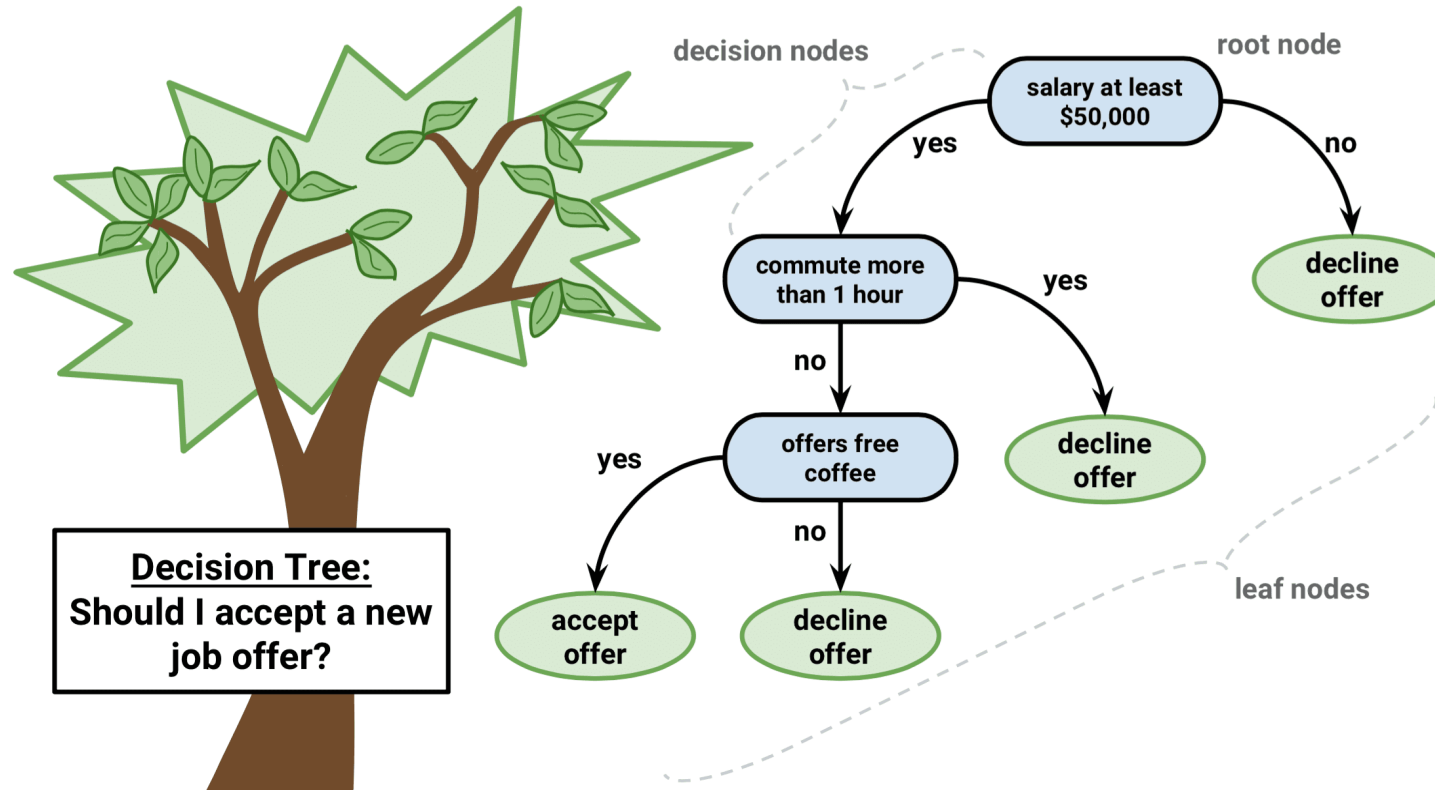
Random forest y árboles de clasificación y regresión (CART)

Hugo Andrés Dorado

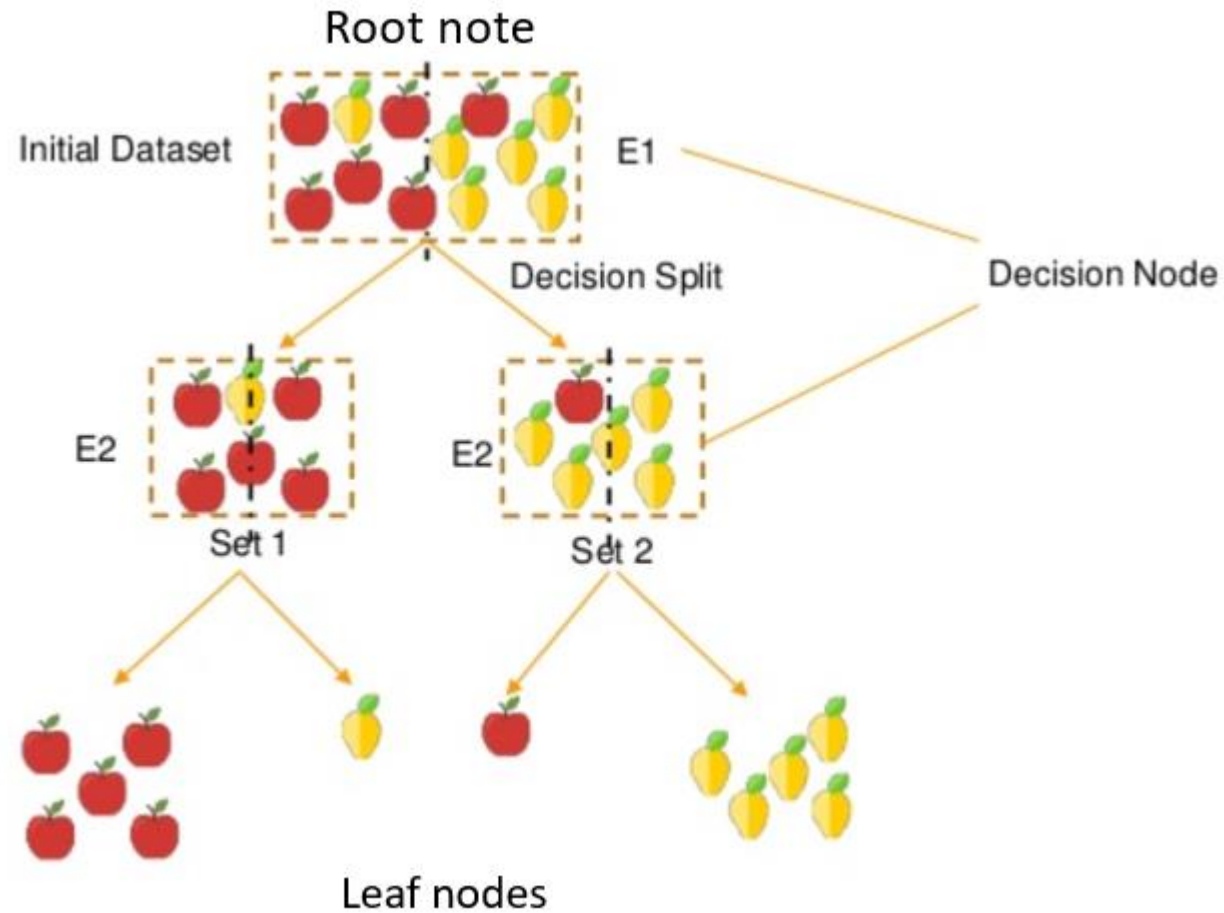
CART - Detalles importantes

- Conocido entre los modelos supervisados.
- Flexibles ante datos atípicos.
- Fácil de interpretar.
- Permite visualizar interacciones.
- Árboles de clasificación para variables de salida cualitativas, árboles de regresión para variables de salida cuantitativas.

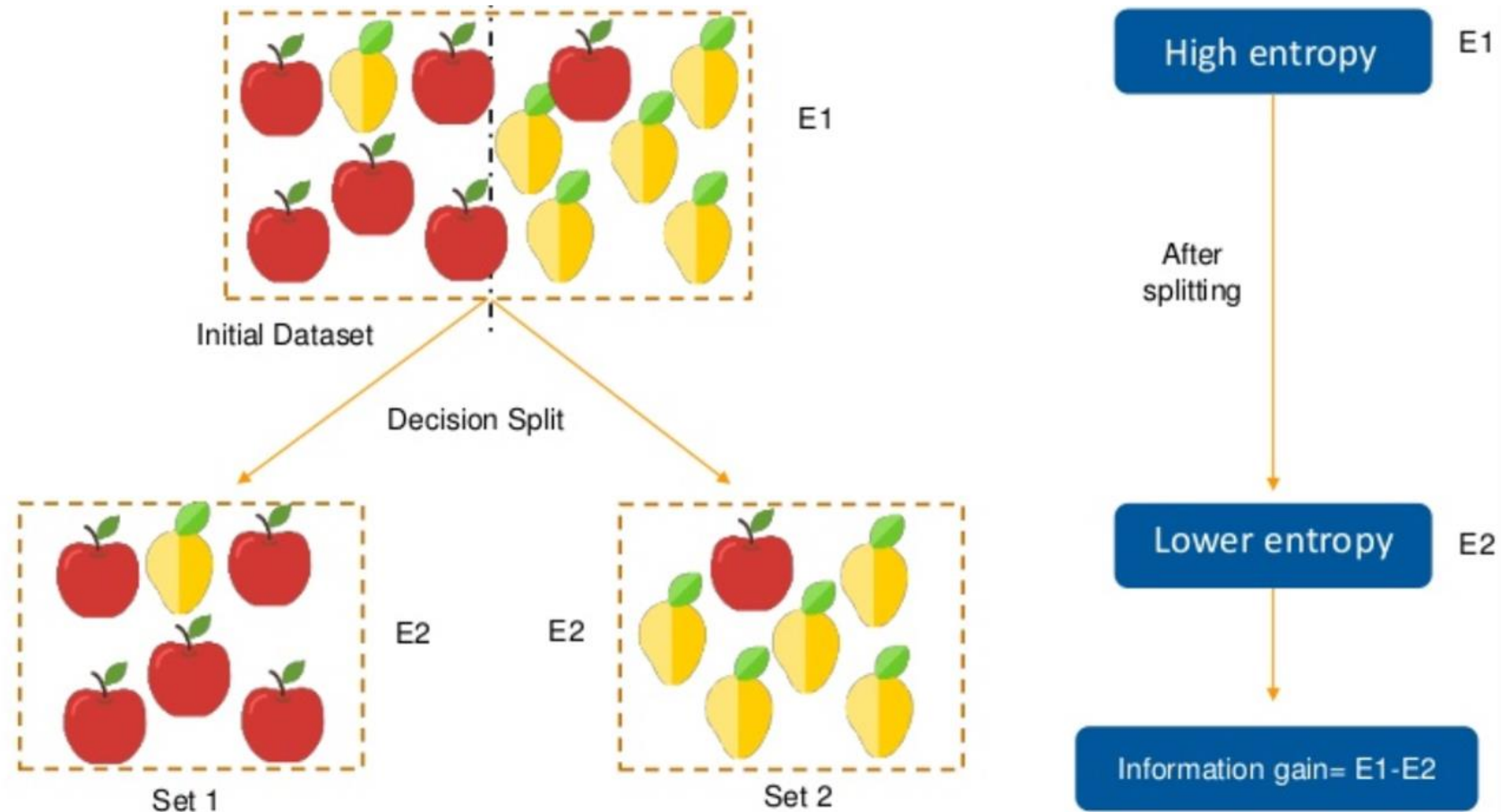
Ejemplo



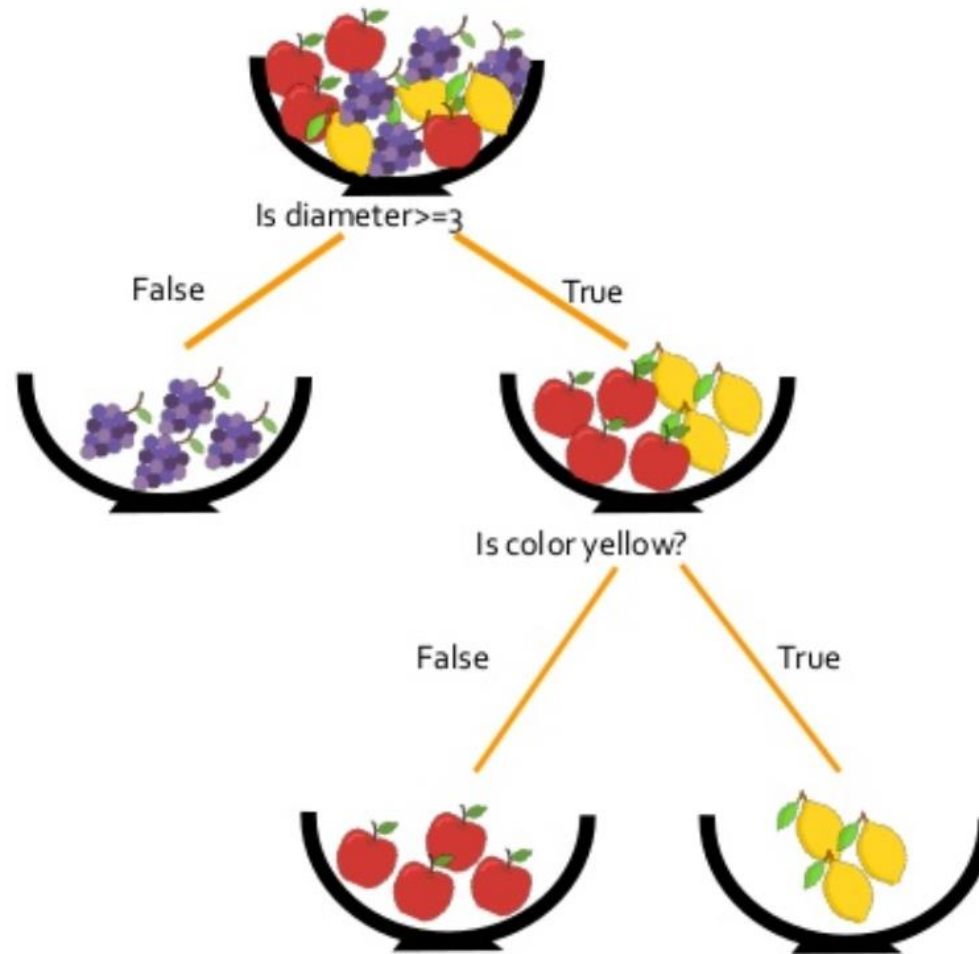
Componentes de un CART



Entropía y ganancia de información



Construcción del árbol

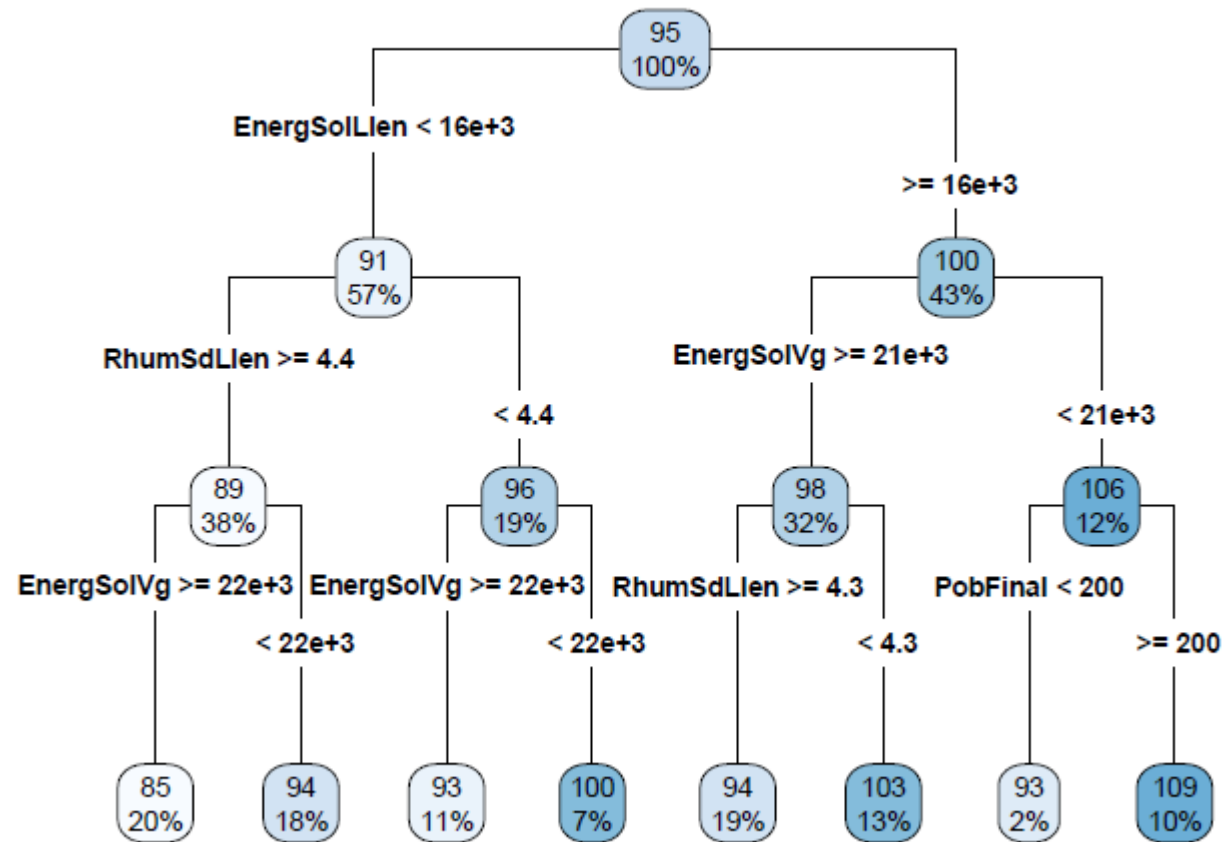


So the entropy in this case is now zero

Parámetros del CART en R – Función rpart

- **minsplit:** El número mínimo de observaciones en un nodo para para que este sea dividido (20).
- **minbucket:** Indica el número mínimo de observaciones en cualquier nodo terminal ($\text{minsplit}/3$).
- **Cp:** Parámetro de complejidad. Indica que si el criterio de impureza no es reducido reducido en mas de $\text{cp} * 100\%$ entonces detiene (0.01).
- **maxdepth:** que condiciona la profundidad máxima del árbol (30).

CART – Rendimiento de arroz de riego (QQ/MZ) Nicaragua



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



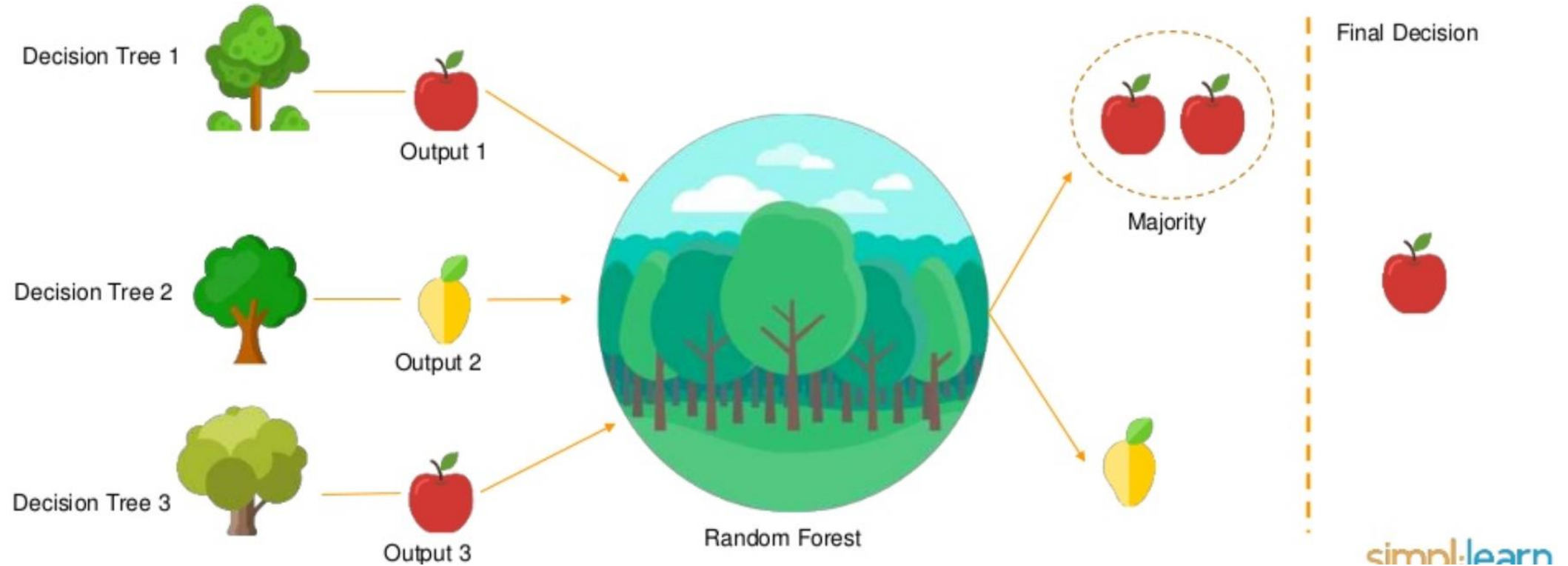
Estimates missing data

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

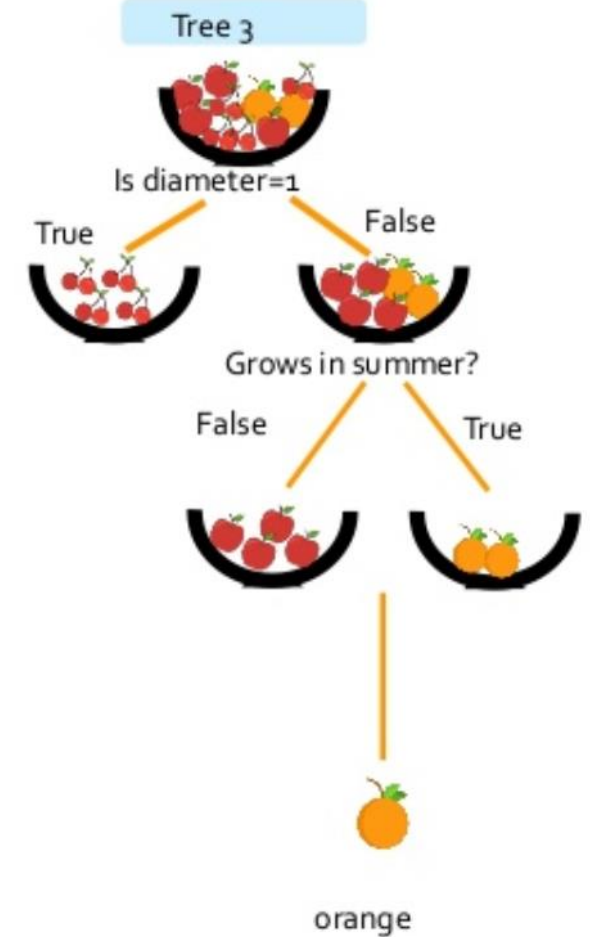
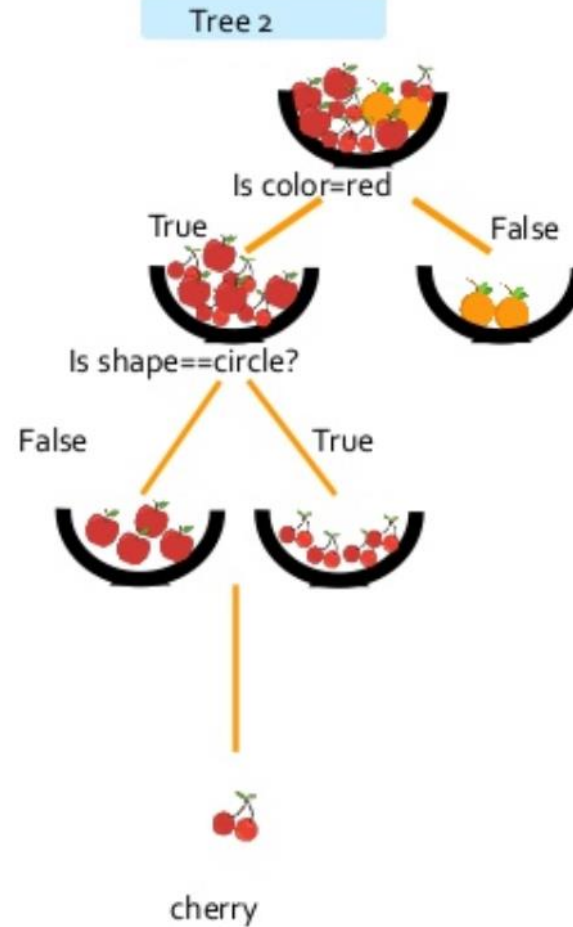
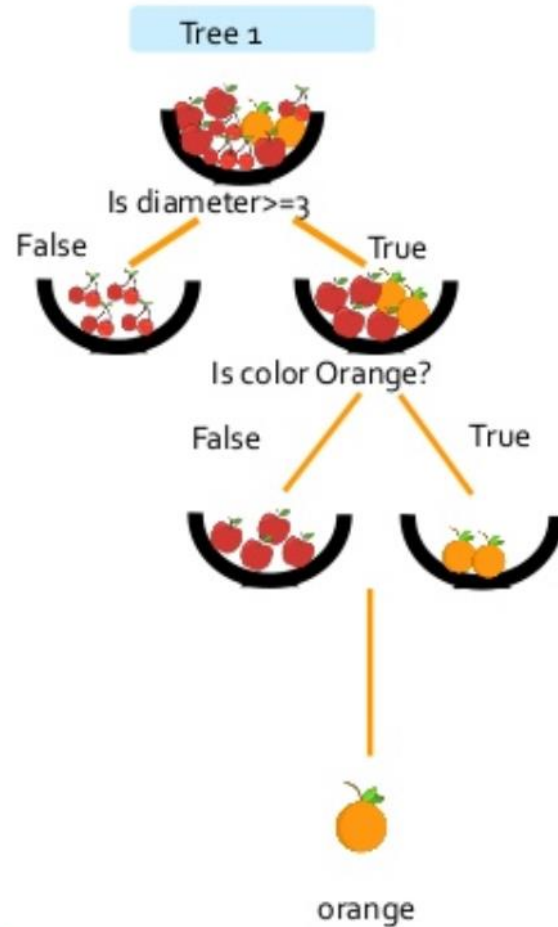
Random Forest

- Integración de varios árboles, contruidos por conjuntos de datos re muestrados y variables seleccionadas al azar.
- Hereda la flexibilidad de los CART en cuanto a manejo de datos faltantes.
- Pierde flexibilidad en interpretación.

Random Forest



Random Forest



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

Parámetros del random forest en R - Función randomForest

- **mtry**: Número de predictores considerado en cada partición.
- **ntrees**: Número de árboles generados dentro del ensamble.