**ÁRBOLES DE DECISIÓN**

Es un modelo que predice un valor de salida (clase o número) a partir de una serie de reglas en forma de ramificaciones.

* Cada nodo interno representa una presunta sobre una variable (feature)
* Cara rama representa la respuesta a esa pregunta
* Cada nodo hoja representa una predicción (una clase o un valor numérico)

1. Tipos de árbol
   1. Clasificación -> salida categórica (si/no, A/B/C)
   2. Regresión -> salida numérica
2. Como se construye un árbol

El algoritmo cart (clasification and regresión tres) divide los datos de manera recursiva

* Se busca todas las posibles divisiones en cada variable
* Se elige la división que mejor separa las clases o reduce la variabilidad (dependiendo del tipo de problema)
* Se repite el proceso en cada subconjunto hasta cumplir una condición de parada (profundidad máxima, número mínimo d muestras, etc)

1. Criterios de división

Los árboles usan medidas para decidir la mejor partición

Clasificación:

índice de Gini -> mide la impureza

entropía -> mide la incertidumbre

regresión

MSE -> se busca la partición que minimiza la varianza de los valores en cada nodo

1. Ventajas
   1. Muy fáciles de interpretar
   2. No requiere normalizar datos ni escalar variables
   3. Puede manejar variables categóricas y numéricas
   4. Captura relaciones lineales
2. Desventajas
   1. Sobreajuste (overfitting) si el árbol crece demasiado (profundidad alta)
   2. Sensible a pequeñas variaciones en los datos (inestables)
   3. Tiende a ser menos precisos que métodos más avanzados
3. Técnicas de control
   1. Máxima profundidad (max\_deph): limita el tamaño del árbol
   2. Mínimo de muestras por nodo 8min\_samples\_split/min\_samples\_leaf): evita ramas con muy ócos datos
   3. Poda (pruning): corta ramas que no a portan mucho valor predictivo