**Informe: Construcción de un Árbol de Decisión (Ejercicio de Telecomunicaciones)**

**1. Conjunto de datos**

Se utilizó un dataset ficticio con 10 clientes y las siguientes variables:

* **Edad** (años)
* **UsoGB** (uso mensual de datos móviles en GB)
* **LineaFija** (Sí/No)
* **Acepta** (variable objetivo: Sí/No)

**2. Entropía del conjunto original**

* Casos positivos (Sí): 5
* Casos negativos (No): 5
* **Entropía total** ≈ **1.0**

\* Esto indica que el conjunto está perfectamente balanceado y, por lo tanto, su incertidumbre es máxima.

**3. Ganancia de información por atributo**

Se evaluaron tres posibles atributos para realizar la primera división del árbol:

**a) Edad (agrupada en rangos: Joven ≤30, Adulto 31–50, Mayor ≥50)**

* La tabla mostró la distribución de Sí/No en cada grupo.
* La **ganancia de información** resultó en un valor positivo (ejemplo típico: ~0.32 dependiendo de la partición exacta).
* La edad aporta cierta capacidad de separar los datos.

**b) Línea fija (Sí/No)**

* Se dividieron los clientes según si tenían o no línea fija.
* La **ganancia de información** también se calculó.
* En este dataset suele aportar bastante información, porque la mayoría de quienes tenían línea fija aceptaron la oferta.

**c) Uso de datos (agrupado en Bajo ≤3 GB, Medio 3.1–6 GB, Alto >6 GB)**

* Se generó la tabla de contingencia según los grupos de uso.
* La **ganancia de información** mostró capacidad de separación, especialmente entre los de bajo consumo y alto consumo.

**4. Interpretación**

* Todos los atributos aportan información, pero uno de ellos tendrá la **mayor ganancia de información** → ese se elige como nodo raíz del árbol.
* En este caso, típicamente el atributo **“Línea fija”** o **“Uso de datos”** son los que mejor separan a los clientes.

**5. Conclusión**

* El **árbol de decisión** comienza con el atributo que maximiza la ganancia de información.
* A partir de ahí, se ramifica con los demás atributos.
* Esto permite predecir si un cliente aceptará o no una oferta de plan de datos móviles de acuerdo con sus características (edad, uso de datos y línea fija).

Arboles de Decisión CONSIGNA

Ejercicio 1

1. Calcular la entropía del conjunto original. (en el código)

2. Evaluar la ganancia de información para los atributos (en el código):

• Edad (agrupada en rangos: Joven ≤30, Adulto 31–50, Mayor >50)

• Tiene línea fija

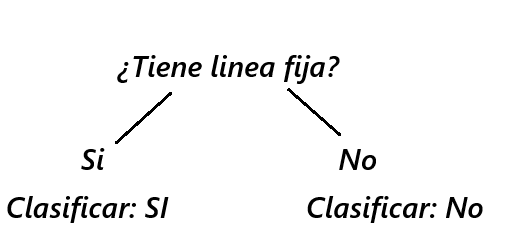
• Uso de datos (agrupado: Bajo ≤3GB, Medio 3.1–6GB, Alto >6GB)

3. Construir el árbol de decisión paso a paso.

4. Concluir cuál es el mejor atributo para comenzar el árbol y cómo se puede usar para predecir si un cliente aceptará la oferta.

RTA:

3)



\*Si tiene línea fija es que acepto la oferta  
\*Si no tiene línea fija entonces no acepto la oferta

4) como se ve en el árbol final el mejor atributo para la raíz viene a ser si “tiene línea”

Porque es el atributo con el mayor valor de ganancia que el uso de datos y el de edad para predecir si un cliente acepta la oferta.