

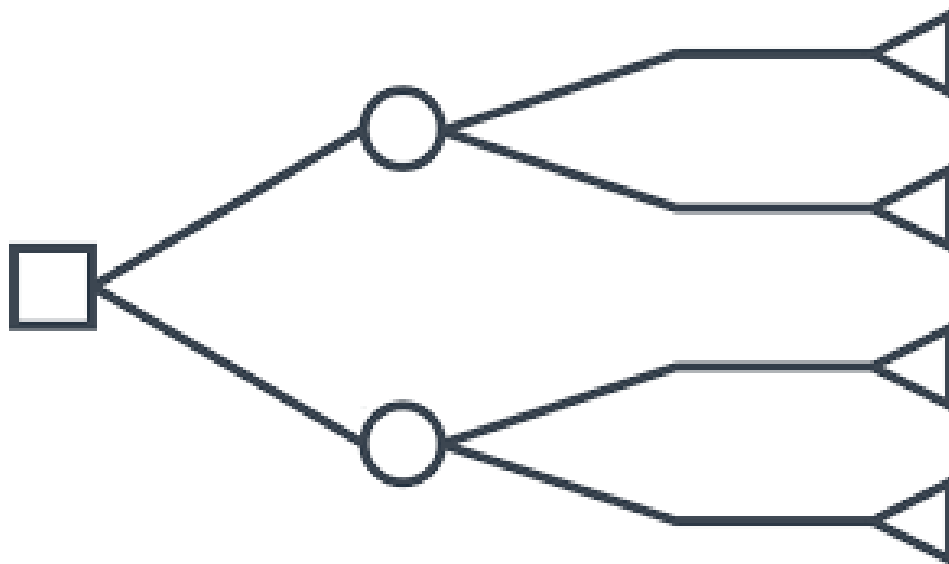


INSTITUTO TECNOLÓGICO BELTRÁN
Centro de Tecnología e Innovación

Empresa De Telecomunicaciones

Trabajo Práctico N.º 3

Árbol de Decisión



Nombre: Coral Tolazzi

Tema: Árbol de Decisión. Empresa de Telecomunicaciones.

Profesora: Yanina Ximena Scudero

Cuatrimestre y Año: 2 Cuatrimestre del 2025

Instituto tecnológico Beltrán

Procesamiento de Aprendizaje Automático

Construcción de un árbol de decisión

Una empresa de telecomunicaciones quiere predecir si un cliente aceptará una oferta de plan de datos móviles. Para ello, se dispone de un conjunto de datos con información de 10 clientes, incluyendo los siguientes atributos:

- Edad (en años)
- Nivel de uso mensual de datos (en GB)
- Tiene línea fija (Sí / No)
- Aceptó la oferta (Sí / No)

Datos del conjunto

ID	Edad	Uso de datos	Tiene línea fija	Aceptó oferta
1	24	2.5	No	No
2	38	6.0	Sí	Sí
3	29	3.0	No	No
4	45	8.0	Sí	Sí
5	52	7.5	Sí	Sí
6	33	4.0	No	No
7	41	5.5	Sí	Sí
8	27	2.0	No	No
9	36	6.5	Sí	Sí
10	31	3.5	No	No

Objetivo del ejercicio

1. Calcular la entropía del conjunto original.
2. Evaluar la ganancia de información para los atributos:
 - Edad (agrupada en rangos: Joven ≤ 30 , Adulto 31–50, Mayor > 50)
 - Tiene línea fija
 - Uso de datos (agrupado: Bajo ≤ 3 GB, Medio 3.1–6GB, Alto > 6 GB)
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.

1. Entropía del conjunto original

Fórmula de la entropía

Para un conjunto de datos con c clases posibles:

$$H(S) = - \sum_{i=1}^c p_i \log_2(p_i)$$

Donde:

- $H(S)$ es la entropía del conjunto S
- p_i es la proporción de elementos de la clase i

calcular proporciones:

$$p_{Si} = 5/10 = 0.5$$

$$p_{No} = 5/10 = 0.5$$

aplicar fórmula:

$$H(S) = -(0.5 * \log_2(0.5) + 0.5 * \log_2(0.5))$$

$$= - (0.5 * -1 + 0.5 * -1)$$

$$= -1$$

Entropía inicial = 1.0 (máxima incertidumbre).

2. Ganancia de información de cada atributo

a) Atributo Tiene línea fija

- Sí (5 personas): 4 aceptaron, 1 no

$$H = -4/5 \log_2 4/5 - 1/5 \log_2 1/5 \approx 0.722$$

- No (5 personas): 1 aceptó, 4 no

$$H = -1/5 \log_2 1/5 - 4/5 \log_2 4/5 \approx 0.722$$

Entropía ponderada:

$$H_{division} = 5/10(0.722) + 5/10(0.722) = 0.722$$

Ganancia:

$$Gain = 1.0 - 0.722 = 0.278$$

b) Atributo Edad (agrupada en Joven <30, Adulto 31–50, Mayor >50)

- Joven (<30): IDs 1,3,8,10 → 4 personas → 0 aceptaron, 4 no → $H = 0.0$
- Adulto (31–50): IDs 2,4,6,7 → 4 personas → 3 sí, 1 no →

$$H = -3/4 \log_2 3/4 - 1/4 \log_2 1/4 = 0.811$$

- Mayor (>50): IDs 5,9 → 2 personas → 2 sí, 0 no → $H = 0.0$

Entropía ponderada:

$$H_{\text{division}} = 4/10 (0) + 4/10(0.811) + 2/10 (0) = 0.324$$

Ganancia:

$$\text{Gain} = 1.0 - 0.324 = 0.676$$

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

- Bajo (≤3GB): IDs 1,3,10 → 3 personas → 0 sí, 3 no → $H = 0$
- Medio (3.1–6GB): IDs 6,7,8 → 3 personas → 2 no, 1 sí →

$$H = -2/3 \log_2 2/3 - 1/3 \log_2 1/3 = 0.918$$

- Alto (>6GB): IDs 2,4,5,9 → 4 personas → 4 sí, 0 no → $H = 0$

Entropía ponderada:

$$H_{\text{division}} = 3/10(0) + 3/10(0.918) + 4/10(0) = 0.275$$

Ganancia:

$$\text{Gain} = 1.0 - 0.275 = 0.725$$

Comparación de ganancias:

- Tiene linea fija = 0.278
- Edad = 0.676
- Uso de datos = 0.725

3. Construcción del árbol de decisión paso a paso

1. Nodo raíz: Uso de datos

- **Bajo ($\leq 3\text{GB}$)** → Todos dijeron **No** → nodo hoja **No**.
- **Alto ($> 6\text{GB}$)** → Todos dijeron **Sí** → nodo hoja **Sí**.
- **Medio ($3.1\text{--}6\text{GB}$)** → mezcla (2 No, 1 Sí) → seguir dividiendo.

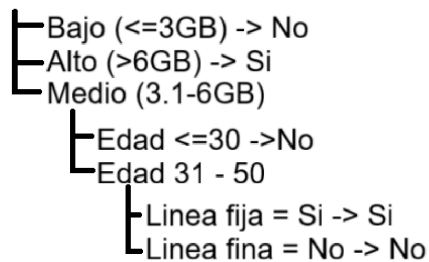
2. Subdividir el grupo Medio por Edad

- Edad Adulto (31–50): IDs 6 y 7 → (1 Sí, 1 No) → aún mezcla.
- Edad Joven (≤ 30): ID 8 → (No) → hoja **No**.

3. Subdividir Adultos con uso Medio por Línea fija

- ID 6: Línea fija = No → No
- ID 7: Línea fija = Sí → Sí

Uso de Datos



4. Conclusiones

- **El uso de datos es el factor más determinante.**
 - Si es bajo → no aceptan.
 - Si es alto → aceptan.
- En el rango intermedio, la **edad** y la **línea fija** permiten afinar la predicción.
- Este árbol permite segmentar clientes:
 - Jóvenes con poco uso → rechazan.
 - Mayores con mucho uso → aceptan.
 - Adultos con uso medio → depende de si tienen línea fija.