

Partimos de la siguiente tabla:

	trafico	cielo	viajar
1	Denso	soleado	no
2	normal	soleado	si
3	fluido	lluvia	si
4	fluido	soleado	si
5	normal	lluvia	no
6	fluido	lluvia	si
7	normal	soleado	si

El objetivo es sacar el árbol de decisiones. Tenemos que elegir el atributo que mayor ganancia aporte, por lo que tenemos que ver entre tráfico y cielo.

Antes de calcular cada ganancia, sacaremos la entropía del sistema:

$$E(S) = \sum_{i \in C} -p_i \lg_2 p_i$$

$$E(\text{viajar}) = -\frac{5}{7} \cdot \lg_2 \frac{5}{7} - \frac{2}{7} \cdot \lg_2 \frac{2}{7} = 0,8631$$

Ahora sacamos para cada atributo su ganancia.

Tráfico

$$\text{Prob}(\text{denso}) = 1/7 \quad \text{Prob}(\text{normal}) = \frac{3}{7} \quad \text{Prob}(\text{fluido}) = \frac{3}{7}$$

Sacamos las entropías:

$$E(\text{denso}) = -1 \lg_2 1 - 0 \cdot \lg_2 0 = 0$$

$$E(\text{normal}) = -\frac{2}{3} \lg_2 \frac{2}{3} - \frac{1}{3} \lg_2 \frac{1}{3} = 0,9183$$

$$E(\text{fluido}) = -1 \lg_2 1 - 0 \cdot \lg_2 0 = 0$$

Ahora sumamos:

$$H(\text{viajar} | \text{Tráfico}) = 0 \cdot \frac{1}{7} + 0,9183 \cdot \frac{3}{7} + 0 \cdot \frac{3}{7} = 0,3935$$

Ahora calculamos la ganancia:

$$G = 0,8631 - 0,3935 = 0,4696$$

Cielo

$$\text{Prob}(\text{lluvia}) = \frac{3}{7} \quad \text{Prob}(\text{sol}) = \frac{4}{7}$$

Entropías:

$$E(\text{lluvia}) = -\frac{2}{3} \lg_2 \frac{2}{3} - \frac{1}{3} \lg_2 \frac{1}{3} = 0,9183$$

$$E(\text{sol}) = -\frac{3}{4} \lg_2 \frac{3}{4} - \frac{1}{4} \lg_2 \frac{1}{4} = 0,8113$$

Se suma:

$$H(\text{viajar} | \text{cielo}) = \frac{3}{7} \cdot 0,9183 + \frac{4}{7} \cdot 0,8113 = 0,8571$$

Ahora la ganancia:

$$G = 0,8631 - 0,8571 = 0,006$$

→ Vemos que el atributo que más ganancia tiene es el tráfico

Ahora hacemos el caso recursivo



