

Dataset 1

Muestra	Cielo	Temperatura	Humedad	Viento	JugarTennis
\mathbf{x}_1	Sol	Alta	Alta	Débil	-
\mathbf{x}_2	Sol	Alta	Alta	Fuerte	-
\mathbf{x}_3	Nubes	Alta	Alta	Débil	+
\mathbf{x}_4	Lluvia	Media	Alta	Débil	+
\mathbf{x}_5	Lluvia	Baja	Normal	Débil	+
\mathbf{x}_6	Lluvia	Baja	Normal	Fuerte	-
\mathbf{x}_7	Nubes	Baja	Normal	Fuerte	+
\mathbf{x}_8	Sol	Media	Alta	Débil	-
\mathbf{x}_9	Sol	Baja	Normal	Débil	+
\mathbf{x}_{10}	Lluvia	Media	Normal	Débil	+
\mathbf{x}_{11}	Sol	Media	Normal	Fuerte	+
\mathbf{x}_{12}	Nubes	Media	Alta	Fuerte	+
\mathbf{x}_{13}	Nubes	Alta	Normal	Débil	+
\mathbf{x}_{14}	Lluvia	Media	Alta	Fuerte	-

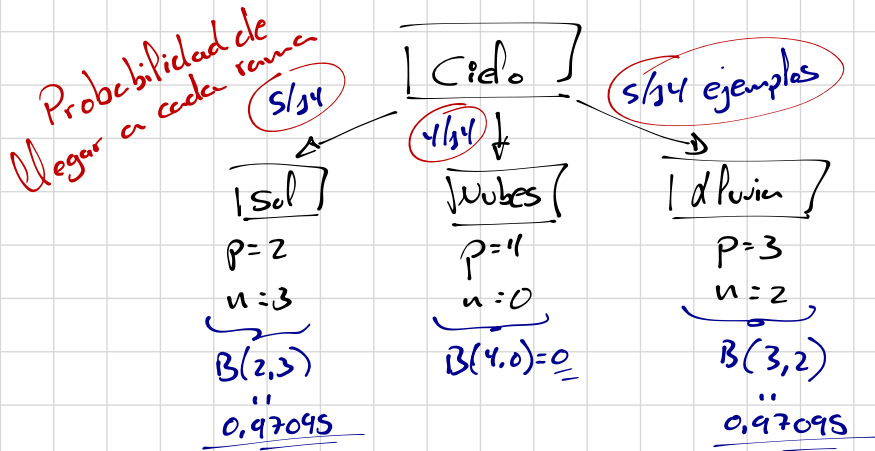
Dataset 1

Muestra	Cielo	Temperatura	Humedad	Viento	JugarTennis
x_1	Sol	Alta	Alta	Débil	-
x_2	Sol	Alta	Alta	Fuerte	-
x_3	Nubes	Alta	Alta	Débil	+
x_4	Lluvia	Media	Alta	Débil	+
x_5	Lluvia	Baja	Normal	Débil	+
x_6	Lluvia	Baja	Normal	Fuerte	-
x_7	Nubes	Baja	Normal	Fuerte	+
x_8	Sol	Media	Alta	Débil	-
x_9	Sol	Baja	Normal	Débil	+
x_{10}	Lluvia	Media	Normal	Débil	+
x_{11}	Sol	Media	Normal	Fuerte	+
x_{12}	Nubes	Media	Alta	Fuerte	+
x_{13}	Nubes	Alta	Normal	Débil	+
x_{14}	Lluvia	Media	Alta	Fuerte	-

Formula Entropía Binaria →

- Analizamos por cuál de los atributos nos merece tomar como raíz.

Por ejemplo, empezamos con el atributo "Cielo".



$$\left[\text{Entropía Inicial} \right]$$

$$\rightarrow \left. \begin{matrix} p=9 \\ n=5 \end{matrix} \right\} B(9,5) = \underline{\underline{0.940}}$$

$$\text{Entropía Promedio} = \frac{5}{14} \cdot B(2,3) + \frac{4}{14} \cdot B(4,0) + \frac{5}{14} \cdot B(3,2) = \underline{\underline{0.693}}$$

- Por tanto, la ganancia de información (IG).

$$IG(\text{Cielo}) = \underline{\underline{0.940}} - \underline{\underline{0.693}} = \underline{\underline{0.247 \text{ bits}}}$$

Inf. Inicial
(Entropía Inicial)

¡Nota!

→ De entre todos los atributos queremos quedarnos con los atributos que más información nos aporte.

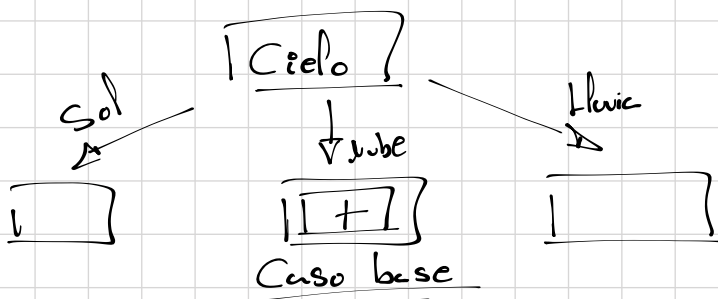
(Hacer lo mismo con el resto de atributos).

$$IG(\text{temp}) = 0.440 - 0.911 = \underline{\underline{0.003 \text{ bits}}}$$

$$IG(\text{Humedad}) = 0.940 - 0.788 = \underline{\underline{0.152 \text{ bits}}}$$

$$IG(\text{Viento}) = 0.940 - 0.892 = \underline{\underline{0.048 \text{ bits}}}$$

- Por tanto, el atributo que más información da el "Cielo". Será nuestro nodo raíz.

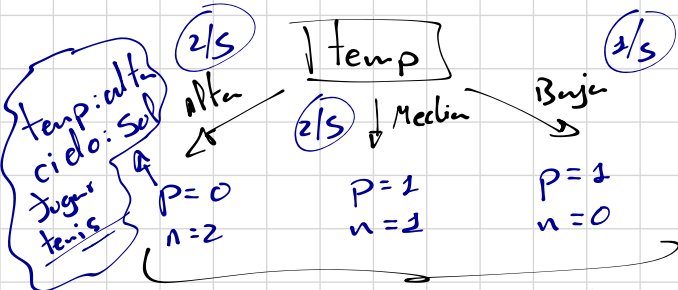


• Rama Cielo = Sol

Entropía Inicial \Rightarrow $\left. \begin{matrix} p=2 \\ n=3 \end{matrix} \right\} B(2,3) = 0,970$

dos casos de cielo que están sucediendo

• Dividiremos con el atributo "temp"



$$\text{Promedio} = \frac{2}{5} \cdot B(0,2) + \frac{2}{5} \cdot B(1,1) + \frac{1}{5} \cdot B(1,0) = \underline{0,4}$$

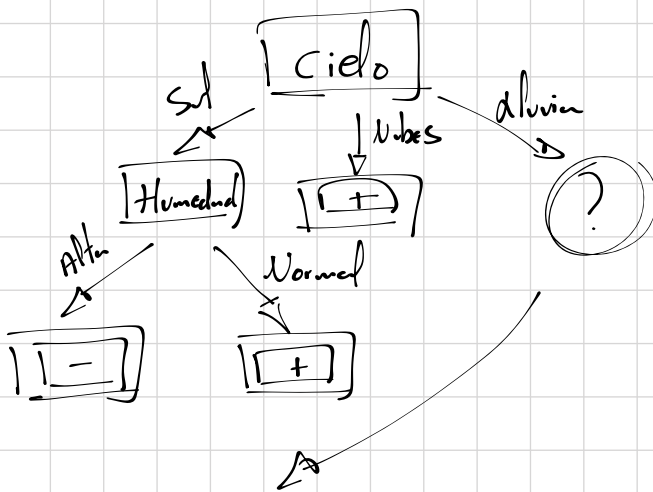
$$IG(\text{temp}) = 0,970 - 0,4 = \underline{0,570 \text{ bits}}$$

• (Lo mismo para el resto de atributos).

$$IG(\text{humedad}) = 0,970 - 0 = \underline{0,970}$$

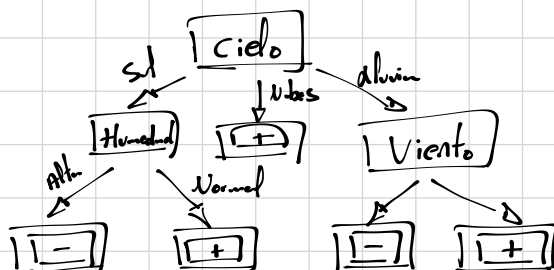
(Ha salido el máximo.

Por tanto, podemos deducir que no habrá otro atributo que nos aporte más información).



• Rama Cielo = lluvia.

• (Tras el análisis del árbol)



Casos Base

- Cuando uno de los atributos, su "p" o "n" es cero.
- Cuando la variable de un atributo son todos los casos o positivos o negativos.
- Ambos valores ("p" y "n") de un atributo son cero.