

3. Algoritmo de FOIL

Ejercicio 1: Identificar empleados en formación

Este ejercicio permite practicar cómo inducir reglas lógicas simples a partir de diferencias entre ejemplos positivos y negativos, tal como lo hace el algoritmo FOIL en su forma más básica.

Se tiene un conjunto de datos con personas que trabajan en una empresa. Algunas están en formación (aprendices) y otras no. Cada persona tiene atributos como:

- edad
- departamento
- nivel_educativo
- en_educacion (booleano: True o False)

→ no se agrega en atributos

Dataset de ejemplo:

datos = [

```
{ "edad": 22, "departamento": "IT", "nivel_educativo": "terciario", "en_educacion": True },  
{ "edad": 24, "departamento": "IT", "nivel_educativo": "universitario", "en_educacion": True },  
{ "edad": 21, "departamento": "RRHH", "nivel_educativo": "terciario", "en_educacion": True },  
{ "edad": 35, "departamento": "IT", "nivel_educativo": "universitario", "en_educacion": False },  
{ "edad": 40, "departamento": "Finanzas", "nivel_educativo": "maestría", "en_educacion": False },  
{ "edad": 29, "departamento": "RRHH", "nivel_educativo": "universitario", "en_educacion": False },  
{ "edad": 23, "departamento": "IT", "nivel_educativo": "terciario", "en_educacion": True },  
{ "edad": 38, "departamento": "Finanzas", "nivel_educativo": "universitario", "en_educacion": False }]
```

Objetivo del ejercicio:

1. Separar los ejemplos positivos ($\text{en_educacion} == \text{True}$) y negativos ($\text{en_educacion} == \text{False}$).
2. Para cada atributo, identificar los valores que aparecen en los positivos pero no en los negativos.
3. Desarrollar un programa en Python que muestre la Regla inducida para identificar a personas en formación:

Regla inducida para identificar a personas en formación:

- edad debe ser uno de: [24, 21, 22, 23]
- nivel_educativo debe ser uno de: ["terciario"]

Respuestas a las preguntas guía:

- ¿Qué valores de departamento aparecen solo en los positivos?
IT (RRHH está en ambos)
- ¿Qué niveles educativos son comunes en los positivos pero no en los negativos?
Terciario
- ¿Qué edades aparecen solo en los positivos?
21, 22, 23, 24

Ejercicio 2:

a) Desarrollar un programa en Python que obtenga el cálculo del FOIL Gain para la condición `nivel_educativo == 'terciario'`, con la siguiente salida:

Condición: `nivel_educativo == 'terciario'`

P (positivos antes) = 4

N (negativos antes) = 4

p (positivos después) = 3

n (negativos después) = 0

$p / (p + n) = 1.000$

$P / (P + N) = 0.500$

$\log_2(p / (p + n)) = 0.000$

$\log_2(P / (P + N)) = -1.000$

FOIL Gain = 3.000

b) Realizar el calculo de FOIL Gain manual para comprobar los cálculos.

Condición: `nivel_educativo == 'terciario'`

Valores utilizados:

Término	Valor
P (positivos antes)	4
N (negativos antes)	4
p (positivos después)	3
n (negativos después)	0

1. Fracción después de aplicar la condición:

$$\frac{p}{p+n} = \frac{3}{3+0} = \frac{3}{3} = 1$$

2. Fracción antes de aplicar la condición:

$$\frac{P}{P+N} = \frac{4}{4+4} = \frac{4}{8} = 0,5$$

3. Logaritmos:

$$\bullet \log_2 \left(\frac{p}{p+n} \right) = \underline{0,000}$$

$$\bullet \log_2 \left(\frac{P}{P+N} \right) = \underline{-1,000}$$

4. FOIL Gain:

$$3 \times [0 - (-1,000)] = 3 \times 1,000 = \underline{3,000}$$

5. Interpretación:

- La condición "nivel_educativo == Terciario" mejora notoriamente
- la identificación de empleados en formación.
- Separa los positivos (sin falsos negativos)

c) Desarrollar un programa en Python que obtenga el cálculo del FOIL Gain para la condición edad \leq 23, con la siguiente salida:

P (positivos antes) = 4
N (negativos antes) = 4
p (positivos después) = 3
n (negativos después) = 0
 $p / (p + n) = 1.000$
 $P / (P + N) = 0.500$
 $\log_2(p / (p + n)) = 0.000$
 $\log_2(P / (P + N)) = -1.000$
FOIL Gain = 3.000

d) Realizar el cálculo de FOIL Gain manual para comprobar los cálculos.

Condición: edad \leq 23

Valores utilizados:

Término	Valor
P (positivos antes)	4
N (negativos antes)	4
p (positivos después)	3
n (negativos después)	0

1. Fracción después de aplicar la condición:

$$\frac{p}{p - n} = \frac{3}{3 - 0} = \frac{3}{3} = 1$$

2. Fracción antes de aplicar la condición:

$$\frac{P}{P+N} = \frac{4}{4+4} = \frac{4}{8} = 0,5$$

3. Logaritmos:

$$\log_2 = (1) = 0$$

$$\log_2 = (0,5) = \underline{1,000}$$

4. FOIL Gain:

$$3 \times [0 - (-1)] = 3 \times 1 = \underline{3}$$

5. Interpretación

- La condición "edad ≤ 23 " separa los casos + y -
- Esto da el = FOIL Gain (3,000)
- Indica que ambas reglas son iguales de efectivas.