



Pecotche Andres
Lucas Carballo

Minería de Datos usando Sistemas Inteligentes

PRÁCTICA 5 PARTE A – REGLAS CLASIFICACIÓN.

PARTE A - REGLAS DE CLASIFICACIÓN

A)

Archivo → globos.xlsx

Atributo → Inflado

ZeroR

útil para determinar una base de performance, se clasifica inflado=no por ser clase mayoritaria.

	<i>si</i>	<i>no</i>	<i>Predicción</i>	
<i>si</i>	6	10	Valor positivo	0.625
<i>no</i>	0	0	Valor negativo	0.0

OneR

Errores

+ Atributo TAMAÑO

SI (TAMAÑO=MEDIANO) ENTONCES INFLADO=NO	0/4
SI (TAMAÑO=CHICO) ENTONCES INFLADO=SI	3/6
SI (TAMAÑO=GRANDE) ENTONCES INFLADO=SI	3/6

Error total = 6/16

+ Atributo COLOR

SI (COLOR=AMARILLO) ENTONCES INFLADO=SI	4/8
SI (COLOR=ROJO) ENTONCES INFLADO=NO	2/8

Error total = 6/16

+ Atributo ESTIRA

SI (ESTIRA=SI) ENTONCES INFLADO=SI	2/8
SI (ESTIRA=NO) ENTONCES INFLADO=NO	0/8

Error total=2/16

+ Atributo EDAD

SI (EDAD=NIÑO) ENTONCES INFLADO=NO	2/8
SI (EDAD=ADULTO) ENTONCES INFLADO=SI	4/8

Error total=6/16

Procedemos a seleccionar el atributo (ESTIRA) por tener el menor error posible 2/16.

B) PRISM

Datos:

Orden	Color	Tamaño	Se estira	Edad	Inflado
1	Amarillo	Chico	Si	Adulto	Inflado
2	Amarillo	Chico	Si	Niño	Inflado
3	Amarillo	Chico	No	Adulto	No Inflado
4	Amarillo	Chico	No	Niño	No Inflado
5	Amarillo	Grande	Si	Adulto	Inflado
6	Amarillo	Grande	Si	Niño	Inflado
7	Amarillo	Mediano	No	Adulto	No Inflado
8	Amarillo	Mediano	No	Niño	No Inflado
9	Rojo	Chico	Si	Adulto	Inflado
10	Rojo	Chico	Si	Niño	No Inflado
11	Rojo	Mediano	No	Adulto	No Inflado
12	Rojo	Mediano	No	Niño	No Inflado
13	Rojo	Grande	Si	Adulto	Inflado
14	Rojo	Grande	Si	Niño	No Inflado
15	Rojo	Grande	No	Adulto	No Inflado
16	Rojo	Grande	No	Niño	No Inflado

Atributo a utilizar → Inflado

-----INFLADO=SI-----

Fijamos la regla de antemano (Inflado=SI).

Color	Amarillo	4/8
	Rojo	2/8
Tamaño	Chico	3/6
	Mediano	0/4
	Grande	3/6
Estira	Si	6/8

	No	0/8
Edad	Niño	2/8
	Adulto	4/8

Elegimos la regla Estira=Si por ser la de mayor precisión, pero no es perfecta, por lo que debemos seguir.

Orden ▾	Color ▾	Tamaño ▾	Se estira ▾	Edad ▾	Inflado? ▾
1	Amarillo	Chico	Si	Adulto	Inflado
2	Amarillo	Chico	Si	Niño	Inflado
5	Amarillo	Grande	Si	Adulto	Inflado
6	Amarillo	Grande	Si	Niño	Inflado
9	Rojo	Chico	Si	Adulto	Inflado
10	Rojo	Chico	Si	Niño	No Inflado
13	Rojo	Grande	Si	Adulto	Inflado
14	Rojo	Grande	Si	Niño	No Inflado

Color	Amarillo	4/4
	Rojo	2/4
Tamaño	Chico	3/4
	Mediano	0/0
	Grande	3/4
Edad	Niño	2/4
	Adulto	4/4

La regla ahora es perfecta.

SI (ESTIRA="SI") Y (COLOR="AMARILLO") ENTONCES INFLADO=SI.

Ahora procedemos a eliminar los datos de la tabla y seguimos con la siguiente iteración:

Orden ▾	Color ▾	Tamaño ▾	Se estira ▾	Edad ▾	Inflado? ▾
9	Rojo	Chico	Si	Adulto	Inflado
10	Rojo	Chico	Si	Niño	No Inflado
13	Rojo	Grande	Si	Adulto	Inflado
14	Rojo	Grande	Si	Niño	No Inflado

Como se observa en la tabla, quedan dos casos por cubrir.

Tamaño	Chico	1/2
	Grande	1/2
Edad	Niño	0/2
	Adulto	2/2

SI (ESTIRA="SI") Y (EDAD="ADULTO") ENTONCES INFLADO=SI.

Ahora sí, no quedan más casos por cubrir.

Observamos que ya se cubrieron el 100% de los casos en que inflado=si.

-----**INFLADO=NO**-----

Una vez terminada el proceso de creación de reglas para inflado=si, procedemos a usar todos los datos nuevamente y crear reglas para **inflado="no"**

Color	Amarillo	4/8
	Rojo	6/8
Tamaño	Chico	3/6
	Mediano	4/4
	Grande	3/6
Estira	Si	2/8
	No	8/8
Edad	Niño	6/8
	Adulto	4/8

Procedemos a eliminar de la tabla los datos que cumplen con la regla **SI(Estira=NO) ENTONCES INFLADO=NO**. :Se elige esta por abarcar más ejemplos que el atributo mediano.

Orden ▼	Color ▼	Tamaño ▼	Se estira ▼	Edad ▼	Inflado ▼
5	Amarillo	Grande	Si	Adulto	Inflado
6	Amarillo	Grande	Si	Niño	Inflado
13	Rojo	Grande	Si	Adulto	Inflado
14	Rojo	Grande	Si	Niño	No Inflado
1	Amarillo	Chico	Si	Adulto	Inflado
2	Amarillo	Chico	Si	Niño	Inflado
9	Rojo	Chico	Si	Adulto	Inflado
10	Rojo	Chico	Si	Niño	No Inflado

Color	Amarillo	0/4
	Rojo	2/4
Tamaño	Chico	1/4
	Grande	1/4
Estira	Si	2/8
Edad	Niño	2/4
	Adulto	0/4

Elegimos color=rojo, pero no es perfecto por lo tanto seguimos:

Color	Amarillo	0/4
Tamaño	Chico	1/4
	Grande	1/4
Estira	Si	2/8
Edad	Niño	2/4
	Adulto	0/4

Procedemos a elegir Edad=niño. Una vez que elegimos **Color rojo y edad niño**, se vuelve perfecta, por lo tanto la regla es: SI (COLOR="ROJO") Y (EDAD=NIÑO) ENTONCES INFLADO=NO.

Eliminamos los respectivos casos en la tabla y como se ve, se cubrieron el 100% de los casos para inflado="NO".

REGLAS

- SI (ESTIRA="SI") Y (COLOR="AMARILLO") ENTONCES INFLADO=SI.
 - SI (ESTIRA="SI") Y (EDAD="ADULTO") ENTONCES INFLADO=SI.
 - SI(ESTIRA=NO) ENTONCES INFLADO=NO.
 - SI (COLOR="ROJO") Y (EDAD=NIÑO) ENTONCES INFLADO=NO.
-

C)

Si comenzara con Inflado=NO el modelo quedaría de la siguiente manera:

- SI(ESTIRA=NO) ENTONCES INFLADO=NO.
- SI (COLOR="ROJO") Y (EDAD=NIÑO) ENTONCES INFLADO=NO.
- SI (ESTIRA="SI") Y (COLOR="AMARILLO") ENTONCES INFLADO=SI.
- SI (ESTIRA="SI") Y (EDAD="ADULTO") ENTONCES INFLADO=SI.

Las reglas para Inflado = NO e INFLADO = SI, ya fueron evaluadas en el inciso B, solo cambia el orden.

E)

Si invertimos el orden de las reglas obtenidas:

- SI (COLOR="ROJO") Y (EDAD=NIÑO) ENTONCES INFLADO=NO.
- SI(ESTIRA=NO) ENTONCES INFLADO=NO.
- SI (ESTIRA="SI") Y (EDAD="ADULTO") ENTONCES INFLADO=SI.
- SI (ESTIRA="SI") Y (COLOR="AMARILLO") ENTONCES INFLADO=SI.

Si analizamos los ejemplos usados en la generación del modelo con estas reglas:

Orden ▾	Color ▾	Tamaño ▾	Se estira ▾	Edad ▾	Inflado: ▾
1	Amarillo	Chico	Si	Adulto	Inflado
2	Amarillo	Chico	Si	Niño	Inflado
3	Amarillo	Chico	No	Adulto	No Inflado
4	Amarillo	Chico	No	Niño	No Inflado
5	Amarillo	Grande	Si	Adulto	Inflado
6	Amarillo	Grande	Si	Niño	Inflado
7	Amarillo	Mediano	No	Adulto	No Inflado
8	Amarillo	Mediano	No	Niño	No Inflado
9	Rojo	Chico	Si	Adulto	Inflado
10	Rojo	Chico	Si	Niño	No Inflado
11	Rojo	Mediano	No	Adulto	No Inflado
12	Rojo	Mediano	No	Niño	No Inflado
13	Rojo	Grande	Si	Adulto	Inflado
14	Rojo	Grande	Si	Niño	No Inflado
15	Rojo	Grande	No	Adulto	No Inflado
16	Rojo	Grande	No	Niño	No Inflado

1_

- Es color rojo y niño? NO
- se estira = no? NO
- se estira = si y edad = adulto? SI → **inflado** (correcto)

2_

- Es color rojo y niño? NO
- se estira = no? NO
- se estira = si y edad = adulto? NO
- se estira = si y color = amarillo? SI → **inflado** (correcto)

3_

- Es color rojo y niño? NO
- se estira = no? SI → **No inflado** (correcto)

4_

- Es color rojo y niño? NO
- se estira = no? SI → **No inflado** (correcto)

5_

- Es color rojo y niño? NO
- se estira = no? NO
- se estira = si y edad = adulto? SI → **inflado** (correcto)

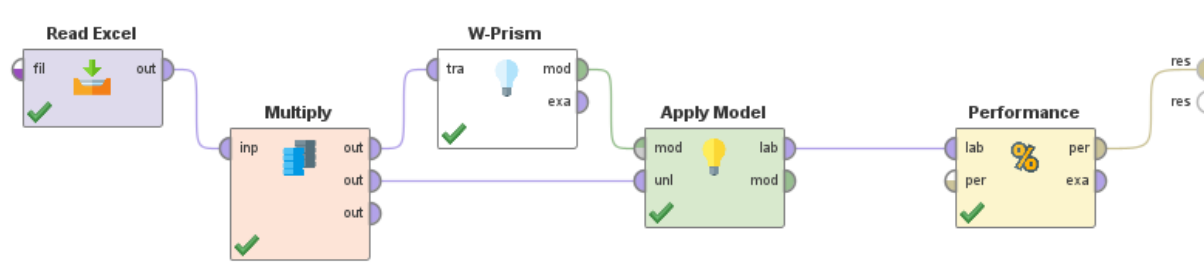
6_

- Es color rojo y niño? NO
- se estira = no? NO
- se estira = si y edad = adulto? NO
- se estira = si y color = amarillo? SI → **inflado** (correcto)

- 7_
- Es color rojo y niño? NO
- se estira = no? SI → **No Inflado** (correcto)
- 8_
- Es color rojo y niño? NO
- se estira = no? SI → **No Inflado** (correcto)
- 9_
- Es color rojo y niño? NO
- se estira = no? NO
- se estira = si y edad = adulto? SI → **Inflado** (correcto)
- 10_
- Es color rojo y niño? SI → **Inflado** (correcto)
- 11_
- Es color rojo y niño? NO
- se estira = no? SI → **No Inflado** (correcto)
- 12_
- Es color rojo y niño? SI → **No Inflado** (correcto)
- 13_
- Es color rojo y niño? NO
- se estira = no? NO
- se estira = si y edad = adulto? SI → **inflado** (correcto)
- 14_
- Es color rojo y niño? NO
- se estira = no? SI → **No Inflado** (correcto)
- 15_
- Es color rojo y niño? NO
- se estira = no? SI → **No inflado** (correcto)
- 16_
- Es color rojo y niño? SI → **No Inflado** (correcto)

Precisión 100% evaluando con las reglas en orden invertido

Precisión con el orden original (calculado en rapidminer):



En la tabla original se descartó el atributo “orden” y se fijó como atributo etiqueta “Inflado”.

Los resultados arrojados corresponden con las reglas halladas en los incisos anteriores:

W-Prism

Prism rules

```
If Se estira? = Si
    and Color = Amarillo then Inflado
If Se estira? = Si
    and Edad = Adulto then Inflado
If Se estira? = No then No Inflado
If Color = Rojo
    and Edad = Niño then No Inflado
```

Y la precisión de las mismas (con los mismos datos de entrenamiento):

accuracy: 100.00%

	true Inflado	true No Inflado	class precision
pred. Inflado	6	0	100.00%
pred. No Inflado	0	10	100.00%
class recall	100.00%	100.00%	

Como se puede observar la precisión del modelo es del 100%.

En conclusión, independientemente del orden, las reglas nos brindan el 100% de precisión evaluando los ejemplos usados en la generación del modelo. Esto sucede ya que las reglas, independientemente del orden, cubren todos los ejemplos de forma certera, es decir:

*-Si se estira = si y edad = adulto → **inflado** aplica para TODOS los ejemplos de la muestra, independientemente si antes se evaluó cualquier otra regla.*

*-Si se estira = si y edad = amarillo → **inflado** tambien aplica para TODOS los ejemplos de la muestra.*

*Si se estira = no → **no inflado**. Nuevamente también es una regla que aplica para todos los modelos de la muestra, sin necesidad de filtrar ningún ejemplo previamente con otra regla.*

Sí color=rojo y edad=niño → no inflado. También es una afirmación correcta para todos los ejemplos de la muestra que sean color rojo y niños.

En general esto no es así siempre, en algunos casos invertir el orden de las reglas podría llevar a predicciones incorrectas. Ya que las reglas se evalúan en orden, y la regla superior, condiciona el valor de ciertos atributos para las inferiores.