

# *REGLAS DE ASOCIACIÓN*



# CONTEXTO...

- Tenemos dos tipos de problemas:
  - \* Predictivos (árboles, regresiones, redes neuronales)
  - \* Descriptivos (reglas de asociación, agrupamiento)
- Buscamos obtener relaciones entre los valores de los atributos de una base de datos (qué productos se compran juntos, qué películas se ven juntas, qué síntomas aparecen juntos)
- Queremos expresar patrones en los datos en forma de reglas:  
@ ANTECEDENTE -> CONSECUENTE  
(es decir, son del tipo IF -> THEN)"



# ENTONCES, QUÉ SON LAS REGLAS DE ASOCIACIÓN?

- Son proposiciones probabilísticas sobre la ocurrencia de ciertos estados en una base de datos
- Permiten expresar patrones de datos en forma 'entendible' para las personas
- Son APLICABLES A LA TOMA DE DECISIONES
- DEFINICIÓN:
  - \* Sea  $I$  el cjto de ítems de una base de datos  $D$ .
  - \* Una REGLA DE ASOCIACIÓN es una implicación de la forma

$$X \Rightarrow Y$$

donde  $X \subset I$ ,  $Y \subset I$ , y  $X \cap Y = \emptyset$ .



# EJEMPLO

	Vino	Gaseosa	Leche	Miel	Bizcochos	Galletas	Jugo
T1	1	1	0	0	0	1	0
T2	0	1	1	0	0	0	0
T3	0	0	0	1	1	1	0
T4	1	1	0	1	1	1	1
T5	0	0	0	0	0	1	0
T6	1	0	0	0	0	1	1
T7	0	1	1	1	1	0	0
T8	0	0	0	1	1	1	1
T9	1	1	0	0	1	0	1
T10	0	1	0	0	1	0	0

Si Bizcochos Y Miel entonces Galletas

Antecedente

Consecuente



# MÉTRICAS (expresan la calidad de una regla)

- El *Soporte* de una regla de asociación  $X \Rightarrow Y$  en una base de datos  $D$  está dado por la expresión.

$$Sup(X \Rightarrow Y) = \frac{|X \cup Y|}{|D|}$$

- La *Confianza* (confidence) de una regla de asociación  $X \Rightarrow Y$  está dada por la siguiente expresión.

$$Conf(X \Rightarrow Y) = \frac{Sup(X \Rightarrow Y)}{Sup(X)} = \frac{|X \cup Y|}{|X|}$$



# Ejemplo

N°	Ambiente	Temperatura	Humedad	Viento	Juega?
1	soleado	alta	alta	no	No
2	soleado	alta	alta	si	No
3	nublado	alta	alta	no	Si
4	lluvioso	media	alta	no	Si
5	lluvioso	baja	normal	no	Si
6	lluvioso	baja	normal	si	No
7	nublado	baja	normal	si	Si

**Si temperatura = alta entonces NO**

$$\text{soporte}(X \rightarrow Y) = \frac{|X|}{|D|} = \frac{3}{7}$$

$$\text{confianza}(X \rightarrow Y) = \frac{|X \rightarrow Y|}{\text{soporte}(X \rightarrow Y)} = \frac{2}{3}$$

# MÉTRICAS PARA REGLAS

- El SOPORTE es la frecuencia con que una regla dada aparece en la BBDD de la cual se extrae (proporción de transacciones que contienen al antecedente Y al consecuente)
- La CONFIANZA se refiere a la cantidad de veces que una regla dada resulta ser verdadera en la práctica (fiabilidad de la regla)
- **Confidence of  $A \rightarrow B = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$**
- Si la confianza es alta, entonces la regla es aplicable a la BBDD y podemos investigarla más a fondo (alta: criterio a determinar x el usuario)
- Una regla puede mostrar una alta ocurrencia en una BBDD pero aparecer menos cuando se la aplica: alto soporte pero baja confianza
- Una regla puede no destacar en el cjto de datos pero en el análisis continuo aparecer con mucha frecuencia: alta confianza pero bajo soporte
- El soporte de la regla  $A \rightarrow B$  es igual que para la regla  $B \rightarrow A$  pero la confianza es distinta para ambas

# MÉTRICAS (continuación)

\* LIFT (sin una traducción exacta al castellano, 'MEJORA' o 'INTERÉS') que es la proporción del soporte observado de un conjunto de productos respecto del soporte teórico de ese conjunto dado el supuesto de independencia (mide la correlación-dependencia entre item sets). Un valor de lift > 1 (positivamente correlacionado) indica que ese conjunto aparece una cantidad de veces superior a lo esperado bajo condiciones de independencia (por lo que se puede intuir que existe una relación que hace que los productos se encuentren en el conjunto más veces de lo normal). Un valor de lift < 1 (negativamente correlacionado) indica que ese conjunto aparece una cantidad de veces inferior a lo esperado bajo condiciones de independencia (por lo que se puede intuir que existe una relación que hace que los productos no estén formando parte del mismo conjunto más veces de lo normal). Lift=1 indica independencia entre item sets.

$$\text{Lift of } A \rightarrow B = \frac{\text{Support}(A \cup B)}{\text{Support}(A) \times \text{Support}(B)}$$

$$\text{Leverage of } A \rightarrow B = \text{Support}(A \cup B) - \text{Support}(A) \times \text{Support}(B)$$

\* LEVERAGE (apalancamiento ??): también mide la correlación entre item sets pero con una expresión distinta. Favorece item sets con soporte alto.

-Si leverage=0, A y B son independientes

-Si leverage<0, A y B están correlacionados negativamente (x ej, no se compran ambos juntos)

-Si leverage>0, A y B están correlacionados positivamente

<https://medium.com/@24littledino/association-mining-support-association-rules-and-confidence-60132a37e355>

<https://medium.com/@24littledino/association-mining-leverage-in-python-86b0e476edeb>

<https://medium.com/@24littledino/association-mining-lift-in-python-28c62a36678>





# EJEMPLO

	Apple	Bagel	Banana	Beer	Broccoli	Chocolate	Eggs	Grape	Guava	Ice cream	...	Orange	Pineapple	Pork	Potatoes	Spinach	Steak	Strawberry	Tomatoes	Watermelon	Yogurt
0	True	False	True	False	False	False	True	False	True	False	...	False	False	False	False	False	True	False	True	False	True
1	True	False	True	False	False	True	True	False	False	False	...	True	False	True	True	False	False	False	False	False	False
2	False	True	False	False	False	False	True	False	False	False	...	False	False	False	False	True	False	False	False	True	True
3	False	False	True	False	True	False	True	False	False	False	...	False	True	False	False	False	True	False	False	False	False
4	True	False	True	True	False	False	False	True	False	True	...	False	False	False	False	False	False	True	True	False	False

5 rows x 26 columns

```
frequent_items = apriori(df, min_support=0.3, use_colnames=True)
```

```
rules = association_rules(frequent_items, metric='confidence', min_threshold=0.4)
```

```
rules = rules.sort_values(by='leverage',
```

```
ascending = False)
```

```
rules
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
249	(Banana, Eggs, Milk)	(Onion, Apple)	0.4	0.4	0.4	1.000000	2.500000	0.24	inf
255	(Eggs, Apple)	(Onion, Banana, Milk)	0.4	0.4	0.4	1.000000	2.500000	0.24	inf
241	(Onion, Eggs, Apple)	(Banana, Milk)	0.4	0.4	0.4	1.000000	2.500000	0.24	inf
246	(Onion, Banana, Milk)	(Eggs, Apple)	0.4	0.4	0.4	1.000000	2.500000	0.24	inf
145	(Banana, Milk)	(Eggs, Apple)	0.4	0.4	0.4	1.000000	2.500000	0.24	inf
...	...	...	...	...	...	...	...	...	...
51	(Onion)	(Banana, Apple)	0.8	0.6	0.4	0.500000	0.833333	-0.08	0.8
50	(Banana, Apple)	(Onion)	0.6	0.8	0.4	0.666667	0.833333	-0.08	0.6
12	(Banana)	(Milk)	0.8	0.6	0.4	0.500000	0.833333	-0.08	0.8
13	(Milk)	(Banana)	0.6	0.8	0.4	0.666667	0.833333	-0.08	0.6
196	(Onion, Eggs, Milk)	(Banana)	0.6	0.8	0.4	0.666667	0.833333	-0.08	0.6

266 rows x 9 columns

Positively correlated

Negatively correlated

# ALGORITMOS Y LIBRERÍAS

- **APRIORI:** [https://es.wikipedia.org/wiki/Algoritmo\\_apriori](https://es.wikipedia.org/wiki/Algoritmo_apriori)
- **ECLAT:** <https://hands-on.cloud/implementation-of-eclat-algorithm-using-python/>  
<https://1library.co/article/reglas-de-asociaci%C3%B3n-contexto-de-investigaci%C3%B3n.zlg4o20l>
- **Apyori:** [pypi.org/project/apyori](https://pypi.org/project/apyori)
- **Mlxtend:** [pypi.org/project/mlxtend](https://pypi.org/project/mlxtend)  
<https://rasbt.github.io/mlxtend/>



***DUDAS, CONSULTAS???***

***GRACIAS POR SU  
ATENCIÓN!!!***



# ALGORITMOS y LIBRERÍAS

- APRIORI: [https://es.wikipedia.org/wiki/Algoritmo\\_apriori](https://es.wikipedia.org/wiki/Algoritmo_apriori)
- ECLAT: <https://1library.co/article/algoritmo-eclat-algoritmos-de-reglas-de-asociaci%C3%B3n.y8g21m5z>
- apyori: <https://pypi.org/project/apyori/>
- mlxtend: <https://pypi.org/project/mlxtend/>  
[https://rasbt.github.io/mlxtend/user\\_guide/frequent\\_patterns/apriori/](https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/)

