



Instituto Politécnico Nacional

Escuela Superior de Cómputo



Unidad de Aprendizaje: Data Mining

Ejercicio: Reglas de asociación

Equipo 4

Grupo: 3CV11

Profesora: Ocampo Botello Fabiola

Fecha de entrega: 24 de mayo de 2021

Integrantes:

Arévalo Andrade Miguel Ángel

Castelán Hidalgo Jorge Bernardo

Cortés Castillo Daniela

Solano Guzmán César Alberto

Índice

Conjunto de datos.....	3
Representación tabular.....	4
Primera iteración (algoritmo apriori).....	5
Segunda iteración.....	6
Tercera iteración.....	7
Poda.....	8
Generación de las reglas de asociación.....	10
Evaluación de las reglas de asociación.....	14

Desarrollo

Considere el siguiente conjunto de elementos:

$I = \{\text{agua, azúcar, chocolate, huevo, leche, mantequilla, pan}\}$

Nota: Los nombres de los productos deben estar ordenados ascendentemente para la identificación de patrones frecuentes (join).

Listado de transacciones

No.	Productos
1	{agua, azúcar, mantequilla, pan}
2	{huevo, leche, mantequilla, pan}
3	{agua, azúcar, chocolate, huevo}
4	{agua, azúcar, huevo, leche, pan}
5	{azúcar, huevo, mantequilla, pan}
6	{chocolate, leche}
7	{mantequilla, pan}
8	{chocolate, leche, mantequilla, pan}
9	{azúcar, chocolate, huevo, leche, mantequilla, pan}
10	{chocolate, huevo, leche}

Representación tabular

$I = \{\text{agua, azúcar, chocolate, huevo, leche, mantequilla, pan}\}$

No	agua	azúcar	chocolate	huevo	leche	mantequilla	pan
1	1	1	0	0	0	1	1
2	0	0	0	1	1	1	1
3	1	1	1	1	0	0	0
4	1	1	0	1	1	0	1
5	0	1	0	1	0	1	1
6	0	0	1	0	1	0	0
7	0	0	0	0	0	1	1
8	0	0	1	0	1	1	1
9	0	1	1	1	1	1	1
10	0	0	1	1	1	0	0
Total	3	4	5	6	6	6	7

Generar las reglas de asociación y la evaluación de las mismas.

Se establece **$fi(\phi)=3$, $min_sup=3$, $min_conf=50\%$**

PRIMERA ETAPA. IDENTIFICACIÓN DE PATRONES FRECUENTES **(ALGORITMO APRIORI)**

PRIMERA ITERACIÓN

C1 =

Cumple que $\text{min_sup}=3$, $\text{min_sup} \geq \text{fi}(\phi) = 3$

Itemset	Soporte
Agua	3
Azúcar	4
Chocolate	5
Huevo	6
Leche	6
Mantequilla	6
Pan	7

L1 =

Itemset	Soporte
Agua	3
Azúcar	4
Chocolate	5
Huevo	6
Leche	6
Mantequilla	6
Pan	7

SEGUNDA ITERACIÓN

C2 =

Itemset	Soporte
{Agua , Azúcar}	3
{Agua,Chocolate}	1
{Agua, Huevo}	2
{Agua, Leche}	1
{Agua, Mantequilla}	1
{Agua, Pan}	2
{Azúcar , Chocolate}	2
{Azúcar , Huevo}	4
{Azúcar , Leche}	2
{Azúcar , Mantequilla}	3
{Azúcar , Pan}	4
{Chocolate , Huevo}	3
{Chocolate , Leche}	4
{Chocolate, Mantequilla}	2
{Chocolate , Pan}	2
{Huevo, Leche}	4
{Huevo, Mantequilla}	3
{Huevo, Pan}	4
{Leche, Mantequilla}	3
{Leche, Pan}	4
{Mantequilla, Pan}	6

Se analiza si las ocurrencias cumplen con $\text{min_sup}=3$. $\text{min_sup}=3 \geq f_i(\Phi)=3$.

L2=

Itemset	Soporte
{Agua , Azúcar}	3
{Azúcar , Huevo}	4
{Azúcar , Mantequilla}	3
{Azúcar , Pan}	4
{Chocolate , Huevo}	3
{Chocolate , Leche}	4
{Huevo, Leche}	4
{Huevo, Mantequilla}	3
{Huevo, Pan}	4
{Leche, Mantequilla}	3
{Leche, Pan}	4
{Mantequilla, Pan}	6

TERCERA ITERACIÓN

Se generan los 3-itemset candidatos para el conjunto C3, lo cual se realiza a partir de L2.

<u>Itemset</u>
{Azúcar, Huevo, Mantequilla}
{Azúcar, Mantequilla, Pan}
{Azúcar, Huevo, Pan}
{Chocolate, Huevo, Leche}
{Huevo, Leche, Mantequilla}
{Huevo, Leche, Pan}
{Huevo, Mantequilla, Pan}
{Leche, Mantequilla, Pan}

Paso 2. Poda, hay que revisar que se cumpla la propiedad a priori que establece: Si un conjunto de elementos Z no es frecuente, entonces para cualquier elemento A, Z, U A no será frecuente (Larose & Larose, 2015)

Subconjunto	Frecuencia	Decisión
{Azúcar, Huevo, Mantequilla}		<u>No se poda</u>
{Azúcar , Huevo}	4	
{Azúcar , Mantequilla}	3	
{Huevo, Mantequilla}	3	
{Azúcar, Mantequilla, Pan}		<u>No se poda</u>
{Azúcar, Mantequilla}	3	
{Azúcar, Pan}	4	
{Mantequilla, Pan}	6	
{Azúcar, Huevo, Pan}		<u>No se poda</u>
{Azúcar, Huevo}	4	
{Azúcar, Pan}	4	
{Huevo, Pan}	4	
{Chocolate, Huevo, Leche}		<u>No se poda</u>
{Chocolate, Huevo}	3	
{Chocolate, Leche}	4	
{Huevo, Leche}	4	
{Huevo, Leche, Mantequilla}		<u>No se poda</u>
{Huevo, Leche}	4	
{Huevo, Mantequilla}	3	
{Leche, Mantequilla}	3	
{Huevo, Leche, Pan}		<u>No se poda</u>
{Huevo, Leche}	4	
{Huevo, Pan}	4	
{Pan, Leche}	4	

{Huevo, Mantequilla, Pan}		<u>No se poda</u>
{Huevo, Mantequilla}	3	
{Huevo, Pan}	4	
{Mantequilla, Pan}	6	
{Leche, Mantequilla, Pan}		<u>No se poda</u>
{Leche, Mantequilla}	3	
{Leche, Pan}	4	
{Mantequilla, Pan}	6	

L3=

Itemset
{Azúcar, Huevo, Mantequilla}
{Azúcar, Mantequilla, Pan}
{Azúcar, Huevo, Pan}
{Chocolate, Huevo, Leche}
{Huevo, Leche, Mantequilla}
{Huevo, Leche, Pan}
{Huevo, Mantequilla, Pan}
{Leche, Mantequilla, Pan}

SEGUNDA ETAPA. GENERACIÓN DE LAS REGLAS DE ASOCIACIÓN

Regla	Soporte	Confianza
{Azúcar, Huevo} -> {Mantequilla}	3/10=0.3	2/4 = 50%
{Azúcar, Mantequilla} -> {Huevo}	3/10=0.3	2/3 = 66.66%
{Huevo, Mantequilla} -> Azúcar	3/10=0.3	2/3 = 66.66%
{Azúcar} ->Huevo, Mantequilla	3/10=0.3	2/4 = 50 %
{Huevo}-> Azúcar, Mantequilla	3/10=0.3	2/6 = 33.33%
{Mantequilla} -> Huevo, Azúcar	3/10=0.3	2/6 = 33.33%
{Azúcar, Mantequilla} -> Pan	3/10 = 0.3	3/3 = 100 %
{Azúcar, Pan} -> Mantequilla	3/10 = 0.3	3/4 = 75%
{Mantequilla, Pan} -> Azúcar	3/10 = 0.3	3/6 = 66.66%
{Azúcar} -> Mantequilla, Pan	3/10 = 0.3	3/4 = 75 %
{Mantequilla} -> Azúcar, Pan	3/10 = 0.3	3/6 = 66.66%
{Pan} -> Mantequilla, Azúcar	3/10 = 0.3	3/7 = 42.85%
{Azúcar, Huevo}-> Pan	3/10 = 0.3	3/4 = 75%
{Azúcar, Pan}-> Huevo	3/10 = 0.3	3/4 = 75%
{Pan, Huevo}-> Azúcar	3/10 = 0.3	3/4 = 75%
{Azúcar}-> Pan, Huevo	3/10 = 0.3	3/4 = 75%
{Huevo}-> Azúcar, Pan	3/10 = 0.3	3/6 = 50%
{Pan}-> Azucar, Huevo	3/10 = 0.3	3/7 = 42.85%
{Chocolate, Huevo } -> {Leche}	3/10=0.3	2/3 = 66.66%
{Chocolate, Leche} -> {Huevo}	3/10=0.3	2/4 = 50%
{Huevo, Leche} -> {Chocolate}	3/10=0.3	2/4 = 50%

{Chocolate} -> {Huevo, Leche}	3/10=0.3	2/5= 40%
{Huevo} -> {Chocolate, Leche}	3/10=0.3	2/6 = 33.33%
{Leche} -> {Chocolate, Huevo }	3/10=0.3	2/6 = 33.33%
{Huevo, Leche} -> {Mantequilla}	3/10=0.3	2/4= 50%
{Huevo, Mantequilla} -> {Leche}	3/10=0.3	2/3=66.66%
{Leche, Mantequilla} -> {Huevo}	3/10=0.3	2/3=66.66%
{Huevo} -> {Leche, Mantequilla}	3/10=0.3	2/6=33.33%
{Leche} -> {Huevo, Mantequilla}	3/10=0.3	2/6=33.33%
{Mantequilla} -> {Huevo, Leche}	3/10=0.3	2/6=33.33%
{Huevo,Leche} - > Pan	3/10=0.3	3/4=75%
{Huevo, Pan} -> Leche	3/10=0.3	3/4=75%
{Pan, Leche} -> Huevo	3/10=0.3	3/4=75%
{Huevo} -> Leche,Pan	3/10=0.3	3/6=50%
{Leche} -> Pan, Huevo	3/10=0.3	3/6=50%
{Pan} -> Leche, Huevo	3/10=0.3	3/7=42.8%
{Huevo, Mantequilla} -> Pan	3/10 = 0.3	3/3 =100 %
{Huevo, Pan} -> Mantequilla	3/10 = 0.3	3/4 = 75%
{Mantequilla, Pan} -> Huevo	3/10 = 0.3	3/6 = 50%
{Pan} -> Huevo, Mantequilla	3/10 = 0.3	3/7 = 42.85%
{Mantequilla} -> Huevo, Pan	3/10 = 0.3	3/6 = 50%
{Huevo} -> Pan, Mantequilla	3/10 = 0.3	3/6 = 50%
{Leche, Mantequilla} -> Pan	3/10 = 0.3	3/3 = 100%
{Leche,Pan}-> Mantequilla	3/10 = 0.3	3/4= 75%
{Pan, Mantequilla} -> Leche	3/10 = 0.3	3/6 = 50%

{Leche}-> Mantequilla, Pan	$3/10 = 0.3$	$3/6 = 50\%$
{Mantequilla} -> Leche, Pan	$3/10 = 0.3$	$3/6 = 50\%$
{Pan} -> Leche, Mantequilla	$3/10 = 0.3$	$3/7 = 42.85\%$

Reglas de asociación finales considerando una Confianza del 50%

Regla	Soporte	Confianza
{Azúcar, Huevo} -> {Mantequilla}	$3/10=0.3$	$2/4 = 50\%$
{Azúcar, Mantequilla} -> {Huevo}	$3/10=0.3$	$2/3 = 66.66\%$
{Huevo, Mantequilla} -> Azúcar	$3/10=0.3$	$2/3 = 66.66\%$
{Azúcar} ->Huevo, Mantequilla	$3/10=0.3$	$2/4 = 50 \%$
{Azúcar, Mantequilla} -> Pan	$3/10 = 0.3$	$3/3 = 100 \%$
{Azúcar, Pan} -> Mantequilla	$3/10 = 0.3$	$\frac{3}{4} = 75\%$
{Mantequilla, Pan} -> Azúcar	$3/10 = 0.3$	$3/6 = 66.66\%$
{Azúcar} -> Mantequilla, Pan	$3/10 = 0.3$	$\frac{3}{4} = 75 \%$
{Mantequilla} -> Azúcar, Pan	$3/10 = 0.3$	$3/6 = 66.66\%$
{Azúcar, Huevo}-> Pan	$3/10 = 0.3$	$3/4 = 75\%$
{Azúcar, Pan}-> Huevo	$3/10 = 0.3$	$3/4 = 75\%$
{Pan, Huevo}-> Azúcar	$3/10 = 0.3$	$3/4 = 75\%$
{Azúcar}-> Pan, Huevo	$3/10 = 0.3$	$3/4 = 75\%$
{Huevo}-> Azúcar, Pan	$3/10 = 0.3$	$3/6 = 50\%$
{Chocolate, Huevo } -> {Leche}	$3/10=0.3$	$2/3 = 66.66\%$
{Chocolate, Leche} -> {Huevo}	$3/10=0.3$	$2/4 = 50\%$
{Huevo, Leche} -> {Chocolate}	$3/10=0.3$	$2/4 = 50\%$

{Huevo, Mantequilla} -> {Leche}	$3/10=0.3$	$\frac{2}{3}=66.66\%$
{Leche, Mantequilla} -> {Huevo}	$3/10=0.3$	$\frac{2}{3}=66.66\%$
{Huevo, Pan} -> Leche	$3/10=0.3$	$\frac{3}{4}=75\%$
{Pan, Leche} -> Huevo	$3/10=0.3$	$\frac{3}{4}=75\%$
{Huevo} -> Leche,Pan	$3/10=0.3$	$3/6=50\%$
{Leche} -> Pan, Huevo	$3/10=0.3$	$3/6=50\%$
{Huevo, Mantequilla} -> Pan	$3/10 = 0.3$	$3/3 = 100 \%$
{Huevo, Pan} -> Mantequilla	$3/10 = 0.3$	$3/4 = 75\%$
{Mantequilla, Pan} -> Huevo	$3/10 = 0.3$	$3/6 = 50\%$
{Mantequilla} -> Huevo, Pan	$3/10 = 0.3$	$3/6 = 50\%$
{Huevo} -> Pan, Mantequilla	$3/10 = 0.3$	$3/6 = 50\%$
{Leche, Mantequilla} -> Pan	$3/10 = 0.3$	$3/3 = 100\%$
{Leche,Pan}-> Mantequilla	$3/10 = 0.3$	$3/4= 75\%$
{Pan, Mantequilla} -> Leche	$3/10 = 0.3$	$3/6 = 50\%$
{Leche}-> Mantequilla, Pan	$3/10 = 0.3$	$3/6 = 50\%$
{Mantequilla} -> Leche, Pan	$3/10 = 0.3$	$3/6 = 50\%$

TERCERA ETAPA. EVALUACIÓN DE LAS REGLAS DE ASOCIACIÓN

Para evaluar las reglas de asociación generadas, se aplica el Lift.

Calculando el ratio entre la confianza de la regla y el soporte del itemset en el consecuente de la regla. Según Larose & Larose

$\text{Lift} = \text{Confianza de la regla} / \text{Proporción a priori del consecuente}.$

Lift de las reglas generadas:

Número total de transacciones = 10

Proporción a priori = $3/10 = 0.3$

- 1) Regla 1: {Azúcar, Huevo} -> {Mantequilla}
 $\text{Lift} = 0.5/0.3 = 1.16$
- 2) Regla 2: {Azúcar, Mantequilla} -> {Huevo}
 $\text{Lift} = 0.66/0.3 = 2.22$
- 3) Regla 3: {Huevo, Mantequilla} -> Azúcar
 $\text{Lift} = 0.66/0.3 = 2.22$
- 4) Regla 4: {Azúcar} -> Huevo, Mantequilla
 $\text{Lift} = 0.5/0.3 = 1.16$
- 5) Regla 5: {Azúcar, Mantequilla} -> Pan
 $\text{Lift} = 1.0/0.3 = 3.33$

La regla 5 indica que cada vez que las personas compran mantequilla y azúcar, tienen 3.33 más de probabilidad de comprar pan que el resto de clientes.

- 6) Regla 6: {Azúcar, Pan} -> Mantequilla
 $\text{Lift} = 0.75/0.3 = 2.5$

La regla 6 indica que cada vez que las personas compran pan y azúcar, tienen 2.5 más de probabilidad de comprar mantequilla que el resto de clientes. El pan con mantequilla azucarada es todo un éxito.

- 7) Regla 7: {Mantequilla, Pan} -> Azúcar
 $\text{Lift} = 0.66/0.3 = 2.22$
- 8) Regla 8: {Azúcar} -> Mantequilla, Pan
 $\text{Lift} = 0.75/0.3 = 2.5$
- 9) Regla 9: {Mantequilla} -> Azúcar, Pan
 $\text{Lift} = 0.66/0.3 = 2.22$

- 10) Regla 10: {Azúcar, Huevo} -> Pan
Lift = $0.75/0.3 = 2.5$
- 11) Regla 11: {Azúcar, Pan} -> Huevo
Lift = $0.75/0.3 = 2.5$
- 12) Regla 12: {Pan, Huevo} -> Azúcar
Lift = $0.75/0.3 = 2.5$
- 13) Regla 13: {Azúcar} -> Pan, Huevo
Lift = $0.75/0.3 = 2.5$
- 14) Regla 14: {Huevo} -> Azúcar, Pan
Lift = $0.5/0.3 = 1.666$
- 15) Regla 15: {Chocolate, Huevo} -> {Leche}
Lift = $0.66/0.3 = 2.22$
- 16) Regla 16: {Chocolate, Leche} -> {Huevo}
Lift = $0.5/0.3 = 1.16$
- 17) Regla 17: {Huevo, Leche} -> {Chocolate}
Lift = $0.5/0.3 = 1.16$
- 18) Regla 18: {Huevo, Leche} -> {Mantequilla}
Lift = $0.5/0.3 = 1.16$
- 19) Regla 19: {Huevo, Mantequilla} -> {Leche}
Lift = $0.66/0.3 = 2.22$
- 20) Regla 20: {Leche, Mantequilla} -> {Huevo}
Lift = $0.66/0.3 = 2.22$
- 21) Regla 21: {Huevo, Leche} -> Pan
Lift = $0.75/0.3 = 2.5$
- 22) Regla 22: {Huevo, Pan} -> Leche
Lift = $0.75/0.3 = 2.5$
- 23) Regla 23: {Pan, Leche} -> Huevo
Lift = $0.75/0.3 = 2.5$

Podemos notar una fuerte relación entre la compra de pan, leche y huevo. Pues, al comprar un par de cualquiera de estos productos, el cliente tiene 2.5 más de probabilidad de adquirir el tercero restante.

- 24) Regla 24: {Huevo} -> Leche, Pan
Lift = $0.5/0.3 = 1.16$
- 25) Regla 25: {Leche} -> Pan, Huevo
Lift = $0.5/0.3 = 1.16$
- 26) Regla 26: {Huevo, Mantequilla} -> Pan

$$\text{Lift} = 1.0/0.3 = 3.33$$

Cuando los clientes compran mantequilla y huevo, hay 3.33 veces más de probabilidad de que compren pan para desayunar.

27) Regla 27: {Huevo, Pan} -> Mantequilla
 $\text{Lift} = 0.75/0.3 = 2.5$

28) Regla 28: {Mantequilla, Pan} -> Huevo
 $\text{Lift} = 0.5/0.3 = 1.16$

29) Regla 29: {Mantequilla} -> Huevo, Pan
 $\text{Lift} = 0.5/0.3 = 1.16$

30) Regla 30: {Huevo} -> Pan, Mantequilla
 $\text{Lift} = 0.5/0.3 = 1.16$

31) Regla 31: {Leche, Mantequilla} -> Pan
 $\text{Lift} = 1.0/0.3 = 3.33$

Los clientes que compran leche y mantequilla tienen 3.33 veces más probabilidad de comprar pan.

32) Regla 32: {Leche, Pan} -> Mantequilla
 $\text{Lift} = 0.75/0.3 = 2.5$

33) Regla 33: {Pan, Mantequilla} -> Leche
 $\text{Lift} = 0.5/0.3 = 1.16$

34) Regla 34: {Leche} -> Mantequilla, Pan
 $\text{Lift} = 0.5/0.3 = 1.16$

35) Regla 35: {Mantequilla} -> Leche, Pan
 $\text{Lift} = 0.5/0.3 = 1.16$