

Unidad 7

Reglas de asociación

Modelizado de Minería de Datos - Q22025



Introducción

Imagina que estás en un supermercado.

Al analizar los tickets de compra (transacciones), podrías notar patrones interesantes.

Por ejemplo, quizás observes que muchas veces, las personas que compran pan también compran mantequilla.

O que quienes adquieren pañales para bebé, a menudo compran toallitas húmedas.





Concepto

- Concepto Central: Reglas de Asociación basadas en el Análisis de Cestas de Mercado.
 - Forma: {Antecedente} \Rightarrow {Consecuente}.
- Algoritmo Principal: APRIORI. Utiliza el umbral de Soporte Mínimo (Min_Support) para encontrar ítems que se compran juntos frecuentemente (Frequent Itemsets).



Concepto

- Métricas para Reglas Interesantes:
 - Soporte (Support): Mide la popularidad de la regla completa ($A \cup B$).
 - Confianza (Confidence): Mide la fiabilidad ($P(B | A)$).
 - Elevación (Lift): Mide la fuerza de la asociación (si A y B son comprados juntos más de lo esperado por azar. Lift>1 es valioso).



Concepto

Aplicación Estratégica

- Sistemas de recomendación
- diseño de promociones
- optimización de la disposición de productos en tiendas físicas y virtuales.



Algoritmo APRIORI

- 
- Objetivo: Descubrir Conjuntos de Ítems Frecuentes (Frequent Itemsets) y generar Reglas de Asociación a partir de datos transaccionales (ej., tickets de compra).
 - Mecanismo Central: El Principio de Apriori (Propiedad de Cierre Hacia Abajo)
 - Concepto: Si un conjunto de ítems es frecuente, TODOS sus subconjuntos también deben ser frecuentes.
 - Poda de Eficiencia: Si el conjunto $\{A,B\}$ no es frecuente, el algoritmo inmediatamente ignora todos sus superconjuntos (ej., $\{A,B,C\}$). Esto poda el espacio de búsqueda y hace el algoritmo eficiente.



Algoritmo APRIORI - Fases de Modelado y Parámetros

El modelado con APRIORI se divide en dos fases distintas, cada una controlada por sus propios hiperparámetros:

- Fase 1: Encontrar Conjuntos Frecuentes
 - min_support
 - max_len
- Fase 2: Generar Reglas de Asociación
 - min_threshold
 - metric

Consideraciones del Analista y Limitaciones

Título	APRIORI: Consideraciones Clave para el Analista
Tipo de Datos (Input)	Categórico y Binario.
Formato Requerido	Una matriz de transacciones donde cada celda es 1 (el ítem está en la canasta) o 0 (no está).
Escalado de Datos	NO APLICA. El algoritmo no utiliza distancias ni valores continuos, por lo que el escalado es irrelevante.
Sensibilidad a Outliers	No es sensible.
Mecanismo	Los ítems o transacciones raros son automáticamente descartados si no alcanzan el umbral de min_support , evitando que afecten la generación de reglas.
Ventajas (Cuándo Usar) <input checked="" type="checkbox"/>	Fácil Interpretación: Reglas sencillas y directas para Marketing. Garantía: Encuentra todos los conjuntos frecuentes que cumplen el soporte.
Desafíos y Desventajas ✗	Intensivo en Cómputo: Se vuelve muy lento si la base de datos es muy grande o si min_support se establece muy bajo (riesgo de <i>combinatorial explosion</i>). Selección de Umbrales: Elegir el valor óptimo de min_support es difícil; un mal ajuste puede omitir reglas sutiles o generar miles de reglas inútiles.



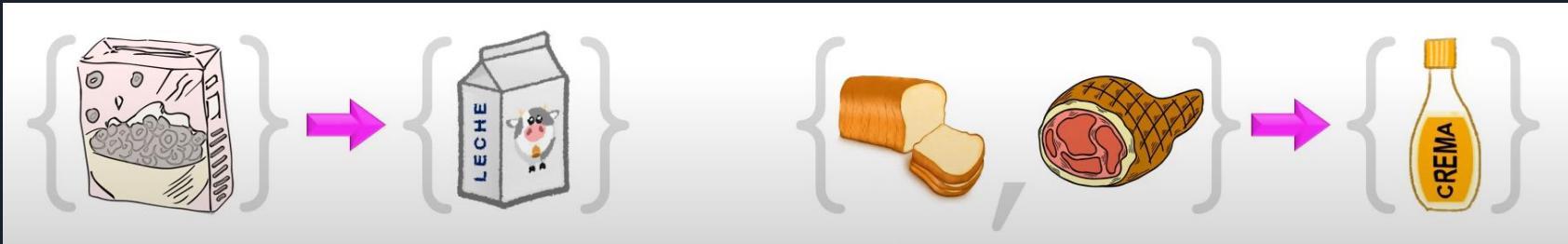
Ej: Qué productos se compran en conjunto?

Determinaremos patrones de ventas dentro de las transacciones de un supermercado.
Podremos definir promociones o estrategia de ventas para maximizar ganancias.

Ej: Qué productos se compran en conjunto?

El resultado de APRIORI son las reglas de asociación, indican patrones de co-ocurrencia de elementos dentro de conjunto de colecciones. Apriori nos ayuda a obtener este tipo de reglas en base a la probabilidad de que un elemento esté presente dado que otros elementos lo están.

- “*si compro cereal, entonces compraré leche*”.
- “*si compro pan y jamón, entonces compraré crema*”

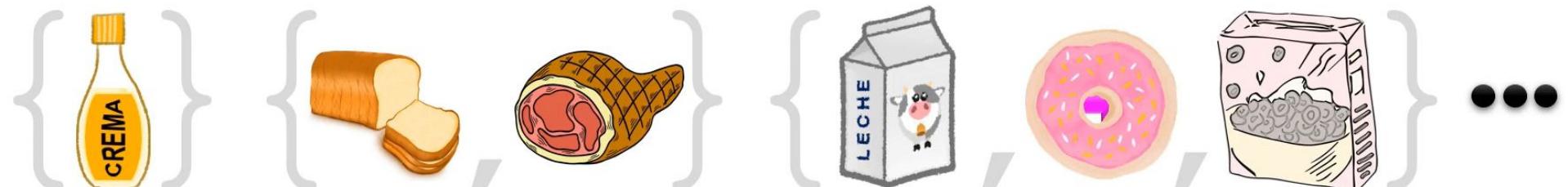


Aplicaría para todo caso de venta de productos

Cómo funciona el algoritmo?

Paso 1. **Conjuntos de elementos frecuentes**: combinaciones de elementos con una **frecuencia mínima**

Por ejemplo, un conjunto frecuente de solo crema , de pan y jamon... y así tener múltiples conjuntos.



Cómo funciona el algoritmo?

Paso 2. **Reglas**: expresan la probabilidad de co-ocurrencia de los elementos

Por ejemplo, un conjunto frecuente de solo crema , de pan y jamón... y así tener múltiples conjuntos.

Nuestra regla “si, entonces” tendrá dos partes: “antecedente” y “consecuente”



Cómo funciona el algoritmo?

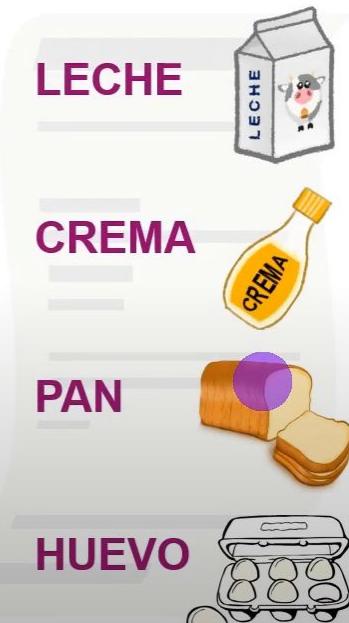
Compra #1



Compra #2



Compra #3



Cómo funciona el algoritmo?

Compra #1, analizaremos una cardinalidad > 1 para evitar analizar muchísimos escenarios.

Definamos los potenciales escenarios



Cómo funciona el algoritmo?

Compra #2, tenemos mayor cantidad de elementos

Definamos los potenciales escenarios (acotamos la cardinalidad a 3 elementos)

Compra #2



- {CEREAL, LECHE}
- {CEREAL, CREMA}
- {CEREAL, HUEVO}
- {LECHE, CREMA}
- {LECHE, HUEVO}
- {CREMA, HUEVO}
- {CEREAL, LECHE, CREMA}
- {CEREAL, LECHE, HUEVO}
- {CEREAL, CREMA, HUEVO}
- {LECHE, CREMA, HUEVO}

Soporte de Conjuntos de elementos

Cómo podemos determinar la relevancia de estos escenarios? Para eso tenemos el “soporte” (valor de 0 a 1)

$$Soporte(X) = \frac{\text{Num. de transacciones con } X}{n}$$

Número de Transacciones

Soporte de Conjuntos de elementos

$$Soporte(X) = \frac{\text{Num. de transacciones con } X}{n}$$

Conjunto de Elementos Número de Transacciones

Compra #1

{LECHE, HUEVO}

{LECHE, PAN}

{PAN, HUEVO}

{LECHE, PAN, HUEVO}

Compra #2

{CEREAL, LECHE}

{CEREAL, CREMA}

{CEREAL, HUEVO}

{LECHE, CREMA}

{LECHE, HUEVO}

{CREMA, HUEVO}

{CEREAL, LECHE, CREMA}

{CEREAL, LECHE, HUEVO}

{CEREAL, CREMA, HUEVO}

{LECHE, CREMA, HUEVO}

Compra #3

{LECHE, CREMA}

{LECHE, PAN}

{LECHE, HUEVO}

{CREMA, PAN}

{CREMA, HUEVO}

{PAN, HUEVO}

{LECHE, CREMA, PAN}

{LECHE, CREMA, HUEVO}

{LECHE, PAN, HUEVO}

{CREMA, PAN, HUEVO}

Soporte de los Conjuntos

{LECHE, HUEVO}

= 3/3 = 1.00

Soporte de Conjuntos de elementos

$$\text{Soporte}(X) = \frac{\text{Num. de transacciones con } X}{n}$$

Conjunto de Elementos

Número de Transacciones

Compra #1

{LECHE, HUEVO}
{LECHE, PAN}
{PAN, HUEVO}
{LECHE, PAN, HUEVO}

Compra #2

{CEREAL, LECHE}
{CEREAL, CREMA}
{CEREAL, HUEVO}
{LECHE, CREMA}
{LECHE, HUEVO}
{CREMA, HUEVO}
{CEREAL, LECHE, CREMA}
{CEREAL, LECHE, HUEVO}
{CEREAL, CREMA, HUEVO}
{LECHE, CREMA, HUEVO}

Compra #3

{LECHE, CREMA}
{LECHE, PAN}
{LECHE, HUEVO}
{CREMA, PAN}
{CREMA, HUEVO}
{PAN, HUEVO}
{LECHE, CREMA, PAN}
{LECHE, CREMA, HUEVO}
{LECHE, PAN, HUEVO}
{CREMA, PAN, HUEVO}

Soporte de los Conjuntos

{LECHE, HUEVO}	= 3/3 = 1.00
{LECHE, PAN}	= 2/3 = 0.66
{PAN, HUEVO}	= 2/3 = 0.66
{LECHE, PAN, HUEVO}	= 2/3 = 0.66
{LECHE, CREMA}	= 2/3 = 0.66
{CREMA, HUEVO}	= 2/3 = 0.66
{LECHE, CREMA, HUEVO}	= 2/3 = 0.66
Todos los demás	= 1/3 = 0.33



Soporte de Conjuntos de elementos

A modo de ejemplo, vamos a trabajar sobre dos conjuntos frecuentes. Notar diferencia entre antecedente y consecuente. Tienen un “buen” soporte pero, qué tan importante serán estas reglas?

{LECHE, PAN}

Si {LECHE} → {PAN}
Si {PAN} → {LECHE}

{LECHE, CREMA, HUEVO}

Si {LECHE, CREMA} → {HUEVO}
Si {LECHE, HUEVO} → {CREMA}
Si {CREMA, HUEVO} → {LECHE}
Si {LECHE} → {CREMA, HUEVO}
Si {CREMA} → {LECHE, HUEVO}
Si {HUEVO} → {LECHE, CREMA}



Soporte de Conjuntos de elementos



Soporte($A \rightarrow C$):
Qué tan frecuente es la regla

A : Antecedente
C : Consecuente



Confianza($A \rightarrow C$):
Qué tanto aparece el consecuente en las transacciones que contienen al antecedente



Lift($A \rightarrow C$):
Fuerza de la regla sobre su ocurrencia aleatoria

Levantamiento

[confianza \geq 0.6 y soporte \geq 0.6]

Sobre nuestro dominio podríamos identificar un total de 16 potenciales reglas (reforzadas ahora con la confianza)

Conjuntos Frecuentes	REGLAS	Conjuntos Frecuentes	REGLAS
{LECHE, PAN}	{LECHE} \rightarrow {PAN} {PAN} \rightarrow {LECHE}	{CREMA, HUEVO}	{CREMA} \rightarrow {HUEVO} {HUEVO} \rightarrow {CREMA}
{LECHE, HUEVO}	{LECHE} \rightarrow {HUEVO} {HUEVO} \rightarrow {LECHE}	{LECHE, PAN, HUEVO}	{LECHE, PAN} \rightarrow {HUEVO} {LECHE, HUEVO} \rightarrow {PAN} {PAN, HUEVO} \rightarrow {LECHE}
{PAN, HUEVO}	{PAN} \rightarrow {HUEVO} {HUEVO} \rightarrow {PAN}	{LECHE, CREMA, HUEVO}	{LECHE, CREMA} \rightarrow {HUEVO} {LECHE, HUEVO} \rightarrow {CREMA} {CREMA, HUEVO} \rightarrow {LECHE}
{LECHE, CREMA}	{LECHE} \rightarrow {CREMA} {CREMA} \rightarrow {LECHE}		

Soporte: qué tan frecuente es esta regla?

$$\text{Soporte}(A \rightarrow C) = \frac{\text{Conteo}(A \cup C)}{n}$$

Número de Transacciones

SOPORTE DE LAS REGLAS

$$\{LECHE\} \rightarrow \{PAN\} \quad \frac{\text{Conteo}(\{LECHE\} \cup \{PAN\})}{n} = \frac{2}{3} = 0.6\bar{6}$$

$$\{PAN\} \rightarrow \{LECHE\} \quad \frac{\text{Conteo}(\{PAN\} \cup \{LECHE\})}{n} = \frac{2}{3} = 0.6\bar{6}$$

$$\{LECHE, CREMA\} \rightarrow \{HUEVO\} \quad \frac{\text{Conteo}(\{LECHE, CREMA\} \cup \{HUEVO\})}{n} = \frac{2}{3} = 0.6\bar{6}$$

$$\{LECHE, HUEVO\} \rightarrow \{CREMA\} \quad \frac{\text{Conteo}(\{LECHE, HUEVO\} \cup \{CREMA\})}{n} = \frac{2}{3} = 0.6\bar{6}$$



Compra #1



Compra #2



Compra #3

Confianza: qué tanto aparece “C” en las transacciones que contienen “A”?

$$\text{Confianza}(A \rightarrow C) = \frac{\text{Conteo}(\text{Antecedente} \cup \text{Consecuente})}{\text{Conteo}(\text{Antecedente})}$$

CONFIANZA DE LAS REGLAS

$$\{LECHE\} \rightarrow \{PAN\} \quad \frac{\text{Conteo}(\{LECHE\} \cup \{PAN\})}{\text{Conteo}(\{LECHE\})} = \frac{2}{3} = \mathbf{0.6\bar{6}}$$

$$\{PAN\} \rightarrow \{LECHE\} \quad \frac{\text{Conteo}(\{PAN\} \cup \{LECHE\})}{\text{Conteo}(\{PAN\})} = \frac{2}{2} = \mathbf{1.00}$$

$$\{LECHE, CREMA\} \rightarrow \{HUEVO\} \quad \frac{\text{Conteo}(\{LECHE, CREMA\} \cup \{HUEVO\})}{\text{Conteo}(\{LECHE, CREMA\})} = \frac{2}{2} = \mathbf{1.00}$$

$$\{LECHE, HUEVO\} \rightarrow \{CREMA\} \quad \frac{\text{Conteo}(\{LECHE, HUEVO\} \cup \{CREMA\})}{\text{Conteo}(\{LECHE, HUEVO\})} = \frac{2}{3} = \mathbf{0.6\bar{6}}$$



Lift: Fuerza de la regla sobre su ocurrencia aleatoria

> 1 mejor!

$$\text{Lift}(A \rightarrow C) = \frac{\text{Soporte(Antecedente} \rightarrow \text{Consecuente})}{\text{Soporte(Antecedente)} * \text{Soporte(Consecuente)}}$$

LEVANTAMIENTO DE LAS REGLAS

$$\{LECHE\} \rightarrow \{PAN\} \quad \frac{\text{Soporte}(\{LECHE\} \rightarrow \{PAN\})}{\text{Soporte}(\{LECHE\}) * \text{Soporte}(\{PAN\})} = \frac{0.66}{(3/3) * (2/3)} = 1$$

Soporte de la regla



Compra #1

$$\{PAN\} \rightarrow \{LECHE\} \quad \frac{\text{Soporte}(\{PAN\} \rightarrow \{LECHE\})}{\text{Soporte}(\{PAN\}) * \text{Soporte}(\{LECHE\})} = \frac{0.66}{(2/3) * (3/3)} = 1$$

Soporte del conjunto frecuente



Compra #2

$$\{LECHE, CREMA\} \rightarrow \{HUEVO\} \quad \frac{\text{Soporte}(\{LECHE, CREMA\} \rightarrow \{HUEVO\})}{\text{Soporte}(\{LECHE, CREMA\}) * \text{Soporte}(\{HUEVO\})} = \frac{0.66}{(2/3) * (3/3)} = 1$$



Compra #3

$$\{LECHE, HUEVO\} \rightarrow \{CREMA\} \quad \frac{\text{Soporte}(\{LECHE, HUEVO\} \rightarrow \{CREMA\})}{\text{Soporte}(\{LECHE, HUEVO\}) * \text{Soporte}(\{CREMA\})} = \frac{0.66}{(3/3) * (2/3)} = 1$$

Otro ejemplo

Soporte mínimo (min_support): 40%, el conjunto de ítems debe aparecer en al menos el 40% de las transacciones, que en este caso son $5 * 0.4 = 2$ transacciones

Nº transacción	items_comprados
1	Pan, Leche, manteca
2	Pan, manteca
3	Leche, Pañales, Cerveza
4	Pan, Leche, Pañales
5	Pan, manteca, Cerveza





Paso 1

Se “escanean” todas las transacciones y se cuenta la frecuencia de cada ítem individual (ocurrencia ítem / total de transacciones).

ítem	cantidad
Cerveza	2/5
Leche	3/5
Manteca	3/5
Pan	4/5
Pañales	2/5



Paso 2

Generación de conjuntos de ítems frecuentes (C) de tamaño 1 (L1)

Cada ítem que cumpla con el soporte mínimo se convierte en un **conjunto L1**:

Se calcula el soporte →

item	cantidad
Cerveza	$2/5 = 40\%$
Leche	$3/5 = 60\%$
Manteca	$3/5 = 60\%$
Pan	$4/5 = 80\%$
Pañales	$2/5 = 40\%$

min_support: 40% (2 transacciones)

$$\text{Support}(A \implies B) = \frac{\text{Número de transacciones que contienen } A \cup B}{\text{Número total de transacciones}}$$



Resultado del Paso 2

C1, (conjuntos de ítems frecuentes de tamaño L1):

C1 = {{Cerveza}, {Leche}, {Manteca}, {Pan}, {Pañales}}



Paso 3

Generación de conjuntos de ítems frecuentes (C) de tamaño 2 (L2)

Combinamos los conjuntos de ítems frecuentes de L1 para generar pares de ítems:

{Cerveza}, {Leche}, {Manteca}, {Pan}, {Pañales}

Paso 3

Nº transacción	items_comprados
1	Pan, Leche, manteca
2	Pan, manteca
3	Leche, Pañales, Cerveza
4	Pan, Leche, Pañales
5	Pan, manteca, Cerveza

item	cantidad
Cerveza, Leche	$1/5 = 20\%$
Cerveza, Manteca	$1/5 = 20\%$
Cerveza, Pan	$1/5 = 20\%$
Cerveza, Pañales	$1/5 = 20\%$
Leche, manteca	$1/5 = 20\%$
Leche, Pan	$2/5 = 40\%$
Leche, Pañales	$2/5 = 40\%$
Manteca, Pan	$3/5 = 60\%$
Manteca, Pañales	$0/5 = 0\%$
Pan, Pañales	$1/5 = 20\%$

$$C1 = \{\text{Cerveza}, \text{Leche}, \text{Manteca}, \text{Pan}, \text{Pañales}\}$$

Paso 3

Se calcula el soporte

Nº transacción	items_comprados
1	Pan, Leche, manteca
2	Pan, manteca
3	Leche, Pañales, Cerveza
4	Pan, Leche, Pañales
5	Pan, manteca, Cerveza

item	cantidad
Cerveza, Leche	$1/5 = 20\%$
Cerveza, Manteca	$1/5 = 20\%$
Cerveza, Pan	$1/5 = 20\%$
Cerveza, Pañales	$1/5 = 20\%$
Leche, manteca	$1/5 = 20\%$
Leche, Pan	$2/5 = 40\%$
Leche, Pañales	$2/5 = 40\%$
Manteca, Pan	$3/5 = 60\%$
Manteca, Pañales	$0/5 = 0\%$
Pan, Pañales	$1/5 = 20\%$

$$C1 = \{\text{Cerveza}, \text{Leche}, \text{Manteca}, \text{Pan}, \text{Pañales}\}$$



Resultado del Paso 3

C2, (conjuntos de ítems frecuentes de tamaño L2):

$C2 = \{\{Leche, Pan\}, \{Leche, Pañales\}, \{Manteca, Pan\}\}$



Paso 4

C3, (conjuntos de ítems frecuentes de tamaño L3)

Combinamos los conjuntos de ítems frecuentes de L2 cuyo primer ítem (en orden alfabético) es el mismo:

$$C2 = \{\{Leche, Pan\}, \{Leche, Pañales\}, \{Manteca, Pan\}\}$$

De {Leche, Pan}, {Leche, Pañales} => {Leche, Pan, Pañales}

De {Manteca, Pan}, => ninguna combinación

$$C3 = \{\{Leche, Pan, Pañales\}\}$$



Paso 4

Se calcula el soporte



Nº transacción	items_comprados
1	Pan, Leche, manteca
2	Pan, manteca
3	Leche, Pañales, Cerveza
4	Pan, Leche, Pañales
5	Pan, manteca, Cerveza

ítem	cantidad
Leche, Pan, Pañales	1/5 = 20%

Terminación: El algoritmo se detiene porque no se pueden generar más conjuntos de ítems frecuentes.

$$C3 = \{\text{Leche, Pan, Pañales}\}$$



Paso 5



La generación de reglas se extiende a todos los conjuntos de ítems frecuentes de tamaño 2 o superior.

Los conjuntos de tamaño 1 (L1) no pueden generar reglas de asociación porque no tienen un antecedente y un consecuente distintos.

Ahora, a partir de los conjuntos de ítems frecuentes encontrados (L1 y L2), podemos generar reglas.



Paso 5

Ahora, a partir de los conjuntos de ítems frecuentes encontrados (L1 y L2), podemos generar reglas.

item	cantidad
Leche, Pan	$2/5 = 40\%$
Leche, Pañales	$2/5 = 40\%$
Manteca, Pan	$3/5 = 60\%$



Paso 5

Evaluamos la confianza de cada regla

Conjunto	Regla	Conf. (soporteAUB/SoporteA)
Leche, Pan	Leche, Pan	$2/3 = 0.667$
	Pan, Leche	$2/4 = 0.5$
Leche, Pañales	Leche, Pañales	$2/3 = 0.667$
	Pañales, Leche	$2/2 = 1$
Manteca, Pan	Manteca, Pan	$3/3 = 1$
	Pan, Manteca	$3/4 = 0.75$

Paso 5

Evaluamos la confianza de cada regla

Conjunto	Regla	Conf. (soporteAUB/SoporteA)
Leche, Pan	Leche, Pan	$2/3 = 0.667$
	Pan, Leche	$2/4 = 0.5$
Leche, Pañales	Leche, Pañales	$2/3 = 0.667$
	Pañales, Leche	$2/2 = 1$
Manteca, Pan	Manteca, Pan	$3/3 = 1$
	Pan, Manteca	$3/4 = 0.75$

Paso 5

Evaluamos el lift

Conjunto	Regla	Conf. (soporteAUB/Sop orteA)	Lift (Conf. A=>B/soporteB)
Leche, Pan	Leche, Pan	$2/3 = 0.667$	$0.667/0.8 = 0.833$
Leche, Pañales	Leche, Pañales	$2/3 = 0.667$	$0.667/0.4 = 1.667$
	Pañales, Leche	$2/2 = 1$	$1/0.6 = 1.667$
Manteca, Pan	Manteca, Pan	$3/3 = 1$	$1/0.8 = 1.25$
	Pan, Manteca	$3/4 = 0.75$	$0.75/0.6 = 1.25$

Paso 6

Por último, la convicción

Conjunto	Regla	Conf. (soporte AUB / soporte A)	Lift (Conf. A=>B / soporte B)	Conv. ((1 - soporte B) / (1- Conf. A=>B))
Leche, Pan	Leche, Pan	$2/3 = 0.667$	$0.667/0.8 = 0.833$	$1-0.8 / 1-0.667 = 0.6$
Leche, Pañales	Leche, Pañales	$2/3 = 0.667$	$0.667/0.4 = 1.667$	$1-0.4 / 1-0.667 = 1.8$
	Pañales, Leche	$2/2 = 1$	$1/0.6 = 1.667$	$1-0.6 / 1-1 (\text{inf})$
Manteca, Pan	Manteca, Pan	$3/3 = 1$	$1/0.8 = 1.25$	$1-0.8 / 1-1$
	Pan, Manteca	$3/4 = 0.75$	$0.75/0.6 = 1.25$	$1-0.6 / 1-0.75 = 1.6$