

Machine Learning^{4DS}

Associativité – Sequence Mining

Mohamed Heny SELMI

medheny.selmi@esprit.tn

Enseignant et Responsable option Data Science à ESPRIT

objectifs

- ✓ Rechercher les associations consiste à rechercher les règles du type :
 - « Si pour un individu, la variable $A = X_a$, la variable $B = X_b$, etc, alors, dans 80% des cas, la variable $Z = X_z$, cette configuration se rencontrant pour 30 % des individus »
- ✓ Repérer des règles liant les données avec un bon niveau de probabilité
 - découverte de relations fines entre attributs (ou variables)
 - généralisation des dépendances fonctionnelles
- ✓ Mettre en évidence les produits / des articles achetés ensemble
- ✓ Transcrire la connaissance sous forme de règles d'association
- ✓ Règles du style : $\langle \text{si } [P(tid, X) := \text{prémisse}] \text{ alors } [P(tid, Y) := \text{conséquence}] \rangle$
- ✓ Différents types de règles
 - origine « panier de la ménagère »
 - étendues aux tables multiples, aux attributs continus, aux contenus textuels, etc.

Analyse des tickets de caisse

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			



Une observation = un caddie

Ne tenir compte que de la présence des produits : peu importe leur quantité

Dans un caddie : le nombre de produits est variable

La liste des produits est immense et variable

Tableau des transactions

- ✓ Mettre en évidence les produits / des articles achetés ensemble
- ✓ Transcrire la connaissance sous forme de règles d'association

si $[P(tid, X) := \text{prémisse}]$ **alors** $[P(tid, Y) := \text{conséquence}]$

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates

si Poulet **et** Moutarde **alors** Œufs **et** Pates

N° Transaction (Caddie)	Contenu du caddie	
6	Œufs	Pain

si Œufs **alors** Pain

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

	P1	P2	P3	P4	P5	P6	P7	P8
1								
2								
3								
4								
5								
6								
7								

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture



N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1							
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1						
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1					
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1				
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1	0			
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1	0	0		
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1	0	0	0	
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1	0	0	0	0
2								
3								
4								
5								
6								
7								

N° Transaction (Caddie)	Contenu du caddie			
1	Poulet	Moutarde	Œufs	Pates
2	Moutarde	Œufs		
3	Pain	Beurre	Poulet	
4	Pates			
5	Pain	Lait	Beurre	
6	Œufs	Pain		
7	Confiture			



	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1	0	0	0	0
2	0	1	1	0	0	0	0	0
3	1	0	0	0	1	1	0	0
4	0	0	0	1	0	0	0	0
5	0	0	0	0	1	1	1	0
6	0	0	1	0	1	0	0	0
7	0	0	0	0	0	0	0	1

désignation
P1 = Poulet
P2 = Moutarde
P3 = Œufs
P4 = Pates
P5 = Pain
P6 = Beurre
P7 = Lait
P8 = Confiture

Tableau binaire

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1	0	0	0	0
2	0	1	1	0	0	0	0	0
3	1	0	0	0	1	1	0	0
4	0	0	0	1	0	0	0	0
5	0	0	0	0	1	1	1	0
6	0	0	1	0	1	0	0	0
7	0	0	0	0	0	0	0	1



Dès que l'on peut se ramener à des données 0/1 :

Il est possible de construire des règles d'association

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince



Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1				
2				
3				

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince

Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1				
2				
3				

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince

Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1				
2				
3				

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince

Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1	1			
2				
3				

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince

Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1	1	0		
2				
3				

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince

Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1	1	0	1	
2				
3				

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince

Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1	1	0	1	0
2				
3				

Codage disjonctif complet – discrétisation

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince



Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1	1	0	1	0
2	0	1	0	1
3	0	1	1	0



Dès que l'on peut se ramener à des données 0/1 :

Il est possible de construire des règles d'association

Codage complet

- ✓ Il s'agit de détecter les cooccurrences des modalités (attribut = valeur)
- ✓ *Certaines associations sont impossibles par construction*

Observation	Taille = Petit	Taille = Grand	Corpulence = Mince	Corpulence = Enveloppé
1	1	0	1	0
2	0	1	0	1
3	0	1	1	0

	P1	P2	P3	P4	P5	P6	P7	P8
1	1	1	1	1	0	0	0	0
2	0	1	1	0	0	0	0	0
3	1	0	0	0	1	1	0	0
4	0	0	0	1	0	0	0	0
5	0	0	0	0	1	1	1	0
6	0	0	1	0	1	0	0	0
7	0	0	0	0	0	0	0	1

Passage en forme disjonctive complète

- **Catégoriel, qualitatif, discret** : type marché, entreprises, taux, appartenance, ...

	Marché
c ₁	Part.
c ₂	Autre
c ₃	Part.
c ₄	Part.
c ₅	Entr.



	P	E	A
c ₁	1		
c ₂			1
c ₃	1		
c ₄	1		
c ₅		1	

- **Continu, quantitatif** : virement, âge, température, consommation, pourcentage, ...

	Dom
c ₁	1100
c ₂	0
c ₃	2200
c ₄	800
c ₅	3800



	d0	d1	d2	d3	d4
c ₁			1		
c ₂	1				
c ₃				1	
c ₄		1			
c ₅					1

Critères probabilistes pour l'évaluation des règles d'association

Support

- ✓ indicateur de « **fiabilité** »
- ✓ probabilité absolue $P(X \cup Y)$
- ✓ $\frac{||X \cup Y||}{||BD||} = \%$
de transactions vérifiant la règle

Règle d'association :
p1 → p2

sup(R1) = 2 : en termes absolus
ou **sup**(R1) = 2 / 6 = **33%** : en termes relatifs

Conf(R1) = **sup**(R1) / **sup**(antécédant R1)
= **sup**(p1 → p2) / **sup**(p1) = 2 / 4 = **50 %**

Confiance

- ✓ Indicateur de « **précision** »
- ✓ probabilité conditionnelle $P(Y/X)$
- ✓ $\frac{||X \cup Y||}{||X||} = \%$
de transactions vérifiant l'implication

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1

Support

- ✓ indicateur de « fiabilité »
- ✓ probabilité absolue :
 - $P(X \cup Y)$
- ✓ $||X \cup Y|| / ||BD|| = \%$
de transactions vérifiant la règle

Règle d'association :
p1 → p2

$sup(R1) = 2$: en termes absolus
ou $sup(R1) = 2 / 6 = 33\%$: en termes relatifs

$Conf(R1) = sup(R1) / sup(\text{antécédant } R1)$
 $= sup(p1 \rightarrow p2) / sup(p1) = 2 / 4 = 50 \%$

Confiance

- ✓ Indicateur de « précision »
- ✓ probabilité conditionnelle :
 $P(Y/X)$
- ✓ $||X \cup Y|| / ||X|| = \%$
de transactions vérifiant l'implication

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1



« Bonne » règle = règle avec un support et une confiance élevés

Simulation



On va prendre le pourcentage de la confiance minimale = **65%**

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1

{p1,p3}

$p1 \rightarrow p3$: confiance = $4/4 = 100\%$

$p3 \rightarrow p1$: confiance = $4/5 = 80\%$



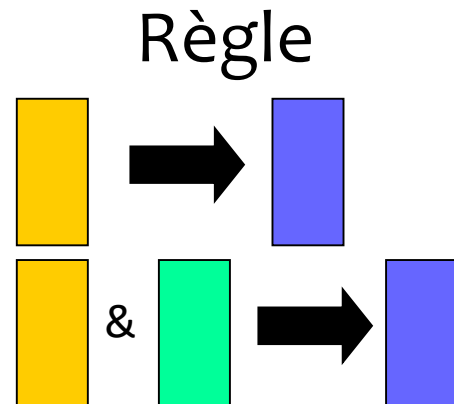
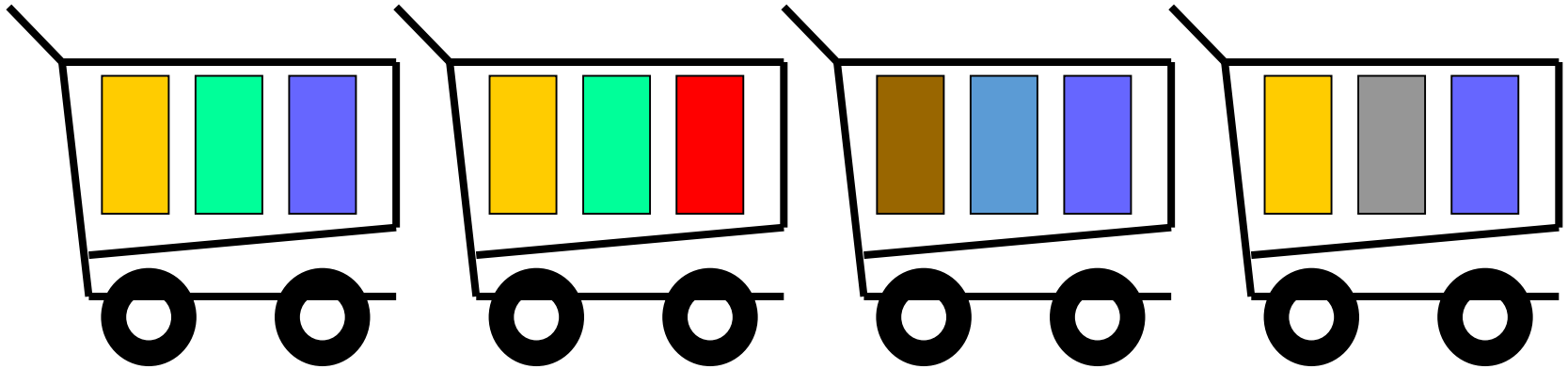
{p2,p3}

$p2 \rightarrow p3$: confiance = $3/3 = 100\%$

$p3 \rightarrow p2$: confiance = $3/5 = 60\%$



Analyse du panier de la ménagère



Support

$\frac{2}{4}$

$\frac{1}{4}$

Confiance

$\frac{2}{3}$

$\frac{1}{2}$

Utilité des règles d'association dans le *caddy analysis*



La distribution réussit à mieux cibler ses mailings



La course à la fidélisation des clients



Réductions personnalisées à la caisse



Profil-client



Le test des nouveaux produits



Le panier moyen



Le parcours magasin



Cartes de fidélité

Analyse des tickets

- { "crème" } → { "pain" }



ID	PRODUITS
1	pain, crème, eau
2	crème
3	pain, crème, vin
4	eau
5	crème, eau

Support = Prob. (crème et pain) :

$$\text{Sup} = \frac{\text{nom}(\text{tran. contenant crème et pain})}{\text{nom_total}(\text{tran.})} = \frac{2}{5} = 0.4$$

Confiance = Prob(crème et pain / crème) :

$$\text{Conf} = \frac{\text{nom}(\text{tran. contenant crème et pain})}{\text{nom}(\text{tran. contenant crème})} = \frac{2}{4} = 0.5 = \frac{\text{sup}(\text{crème et pain})}{\text{sup}(\text{crème})}$$

Support / Confiance

- « Si pour un individu, la variable $A = X_a$, la variable $B = X_b$, etc, alors, dans 80% des cas, la variable $Z = X_z$, cette configuration se rencontrant pour 30 % des individus »
- La valeur de 80% est appelée indice de confiance
- La valeur de 30% est appelée indice de support

- l'indice de confiance de $B \rightarrow E$:

$$3 / 4$$

- l'indice de support de $B \rightarrow E$:

$$3 / 5$$

T26	A	B	C	D	E
T1245	B	C	E	F	
T156	B	E			
T2356	A	B	D		
T145	C	D			

- Dans l'exemple précédent, on a :

- indice de confiance de l'association $C \rightarrow B$ est $2/3$

- indice de support = $2/5$.

- Or, probabilité $(B) = 0.8$, B est présent dans presque tous les tickets de caisse.

- Cette probabilité est supérieure à l'indice de confiance de

$C \rightarrow B$, ce qui fait que l'on ne gagne rien à utiliser la règle

$C \rightarrow B$ pour prédire B.

- Si l'on suppose aléatoirement qu'un ticket de caisse contient B, on n'a qu'1 chance / 5 de se tromper, contre 1 chance / 3 en appliquant la règle $C \rightarrow B$.

Démarche d'Extraction des règles d'association

- Paramètres : Fixer un degré d'exigence sur les règles à extraire
 - ✓ Support min. (exp. 2 transactions)
 - ✓ Confiance min. (exp. 75%)



L'idée est surtout de contrôler (limiter) le nombre de règles produites

Démarche : Construction en deux temps

- ✓ recherche des itemsets fréquents (support \geq support min.)
- ✓ à partir des itemsets fréquents, produire les règles (conf. \geq conf. min.)

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1

Quelques définitions :

- item = produit
- itemset = ensemble de produits (ex. $\{p1, p3\}$)
- $\text{sup}(\text{itemset})$ = nombre de transactions d'apparition simultanée des produits (ex. $\text{sup}\{p1, p3\} = 4$)
- $\text{card}(\text{itemset})$ = nombre de produits dans l'ensemble (ex. $\text{card}\{p1, p3\} = 2$)

Recherche des Itemsets Fréquents

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1



Combinaisons
possibles

$C_4^1 = 4$
 $C_4^2 = 6$
 $C_4^3 = 4$
 $C_4^4 = 1$

 $\Sigma = 15 = 2^4 - 1$

Itemsets de card = 1

Itemsets de card = 2

Itemsets de card = 3

...

Cas général : $2^J - 1$!!



Réduire l'exploration en éliminant certaines pistes

Simulation



On va prendre la valeur du Support minimal = **3**

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1



1-Itemsets	Support
{p1}	4
{p2}	3
{p3}	5
{p4}	1

$$L1 = \{ \{p1\}, \{p2\}, \{p3\} \}$$

$$C2 = \{ \{p1,p2\}, \{p1,p3\}, \{p2,p3\} \}$$

Simulation



On va prendre le Support minimal = 3

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1



2-Itemsets	Support
{p1,p2}	2
{p1,p3}	4
{p2,p3}	3

$$L_2 = \{ \{p1,p3\}, \{p2,p3\} \}$$

$$C_3 = \{ \{p1,p2,p3\} \}$$

Simulation



On va prendre le Support minimal = 3

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1



3-Itemsets	Support
{p1,p2, p3}	2

$$L_3 = \emptyset$$

$$C_4 = \emptyset$$

$$F = \{ \{p1\}, \{p2\}, \{p3\}, \{p1,p3\}, \{p2,p3\} \}$$

Les algorithmes du sequence mining

- Apriori
- Eclat
- GSP (Generalized Sequential Pattern)
- FreeSpan (Frequent pattern-projected Sequential pattern mining)
- PrefixSpan (Prefix-projected Sequential pattern mining)
- SPADE (Sequential PAttern Discovery using Equivalence classes)

Apriori

Algorithme Apriori [Agrawal94]

- Première passe :
 - recherche des 1-ensembles fréquents
 - un compteur par produits
- L'algorithme génère un candidat de taille k à partir de deux candidats de taille $k-1$ différents par le dernier élément
 - procédure apriori-gen
- Passe k :
 - comptage des k -ensemble fréquents candidats
 - sélection des bons candidats

Apriori – Fréquents itemsets

```
L1 = { frequent 1-ensemble };  
for (k = 2 ; Lk-1 ≠ ∅ ; k++) do  
{  
  Ck = apriori-gen(Lk-1); // Generate new candidates  
  foreach transactions t ∈ DB do  
    { // Counting  
      Ct = { subset(Ck, t) }; // get subsets of t candidates  
      foreach c ∈ Ct do c.count++;  
    }  
  Lk = { c ∈ Ck | c.count ≥ minsup }; // Filter candidates  
}  
F = {Lk};
```

Indicateur de pertinence des règles

mesure d'intérêt : *Lift d'une Règle*

- L'amélioration apportée par une règle, par rapport à une réponse au hasard est appelée « **lift** » et vaut :

$$\text{lift (règle)} = \text{confiance (règle)} / p(\text{résultat})$$
$$= p(\text{condition et résultat}) / [p(\text{condition}) \times p(\text{résultat})]$$

- Le lift est une bonne mesure de performance de la règle d'association.
- Le lift est la confiance de la règle divisée par la valeur espérée de la confiance.

Interprétation du lift :

- Un lift supérieur à 1 : Indique une corrélation positive
- Un lift de 1 indique une corrélation nulle
- Un lift inférieur à 1 : Indique une corrélation négative

T26	A	B	C	D	E
T1245	B	C	E	F	
T156	B	E			
T2356	A	B	D		
T145	C	D			

$$\text{lift}(C \rightarrow B) = 5/6 < 1$$



$$\text{lift}(B \rightarrow C) = 6/5 > 1$$



Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

TICKET 1
Farine
Sucre
Lait



Farine → Sucre	Sucre → Farine
Sucre → Lait	Lait → Sucre
Lait → Farine	Farine → Lait

Les **n items** sont :
 $I = \{\text{Farine, Sucre, Lait, Œuf, Chocolat, Thé}\}$

Critère d'extraction des règles
A partir d'un ensemble T de transactions, trouver toutes les règles :
avec un support **sup** > S_{min} et une confiance **conf** > C_{min} où S_{min} et C_{min} sont des seuils fixés a priori par l'utilisateur

Algorithme proposé par Agrawal et Srikant.



✓ Le principe de l'algorithme est de rechercher l'ensemble L_1 de tous les items apparaissant dans au moins $S_{\min} \times m$ transactions.

✓ Puis, parmi C_2 qui est le produit cartésien de L_1 avec lui-même, on construit l'ensemble L_2 de tous les couples d'items apparaissant dans au moins $S_{\min} \times m$ transactions.

✓ L'algorithme s'arrête quand L_k est vide.

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2
Sucre	3

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2
Sucre	3
Lait	1

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2
Sucre	3
Lait	1
Œuf	3

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2
Sucre	3
Lait	1
Œuf	3
Chocolat	3

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2
Sucre	3
Lait	1
Œuf	3
Chocolat	3
Thé	1

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2
Sucre	3
Lait	1
Œuf	3
Chocolat	3
Thé	1

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C1

item1	nombre
Farine	2 ✓
Sucre	3 ✓
Lait	1 ✗
Œuf	3 ✓
Chocolat	3 ✓
Thé	1 ✗

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat

On décide de retenir un taux de support de 30%



Construction de C2

item1	item2	nombre
Farine	Sucre	2
Farine	Œuf	
Farine	Chocolat	
Sucre	Œuf	
sucre	Chocolat	
Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat

On décide de retenir un taux de support de 30%



Construction de C2

item1	item2	nombre
Farine	Sucre	2
Farine	Œuf	
Farine	Chocolat	
Sucre	Œuf	
sucre	Chocolat	
Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat

On décide de retenir un taux de support de 30%



Construction de C2

item1	item2	nombre
Farine	Sucre	2
Farine	Œuf	1
Farine	Chocolat	
Sucre	Œuf	
sucre	Chocolat	
Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat

On décide de retenir un taux de support de 30%



Construction de C2

item1	item2	nombre
Farine	Sucre	2
Farine	Œuf	1
Farine	Chocolat	1
Sucre	Œuf	
sucre	Chocolat	
Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1	TICKET 2	TICKET 3	TICKET 4
Farine	Œuf	Farine	Œuf
Sucre	Sucre	Œuf	Chocolat
	Chocolat	Sucre	
		Chocolat	

On décide de retenir un taux de support de 30%



Construction de C2

item1	item2	nombre
Farine	Sucre	2
Farine	Œuf	1
Farine	Chocolat	1
Sucre	Œuf	2
sucre	Chocolat	
Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat

On décide de retenir un taux de support de 30%



Construction de C2

item1	item2	nombre
Farine	Sucre	2
Farine	Œuf	1
Farine	Chocolat	1
Sucre	Œuf	2
sucre	Chocolat	2
Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1	TICKET 2	TICKET 3	TICKET 4
Farine	Œuf	Farine	Œuf
Sucre	Sucre	Œuf	Chocolat
	Chocolat	Sucre	
		Chocolat	

On décide de retenir un taux de support de 30% 

Construction de C2

item1	item2	nombre
Farine	Sucre	2
Farine	Œuf	1
Farine	Chocolat	1
Sucre	Œuf	2
Sucre	Chocolat	2
Œuf	Chocolat	3

Etude de cas de recherche d'associations intéressantes

On décide de retenir un taux de support de 30%



Construction de C2

item1	item2	nombre	
Farine	Sucre	2	✓
Farine	Œuf	1	✗
Farine	Chocolat	1	✗
Sucre	Œuf	2	✓
Sucre	Chocolat	2	✓
Œuf	Chocolat	3	✓

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C₃

item1	item2	item3	nombre
Farine	Sucre	Œuf	
Farine	Sucre	Chocolat	
Sucre	Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C₃

item1	item2	item3	nombre
Farine	Sucre	Œuf	1
Farine	Sucre	Chocolat	
Sucre	Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C₃

item1	item2	item3	nombre
Farine	Sucre	Œuf	1
Farine	Sucre	Chocolat	1
Sucre	Œuf	Chocolat	

Etude de cas de recherche d'associations intéressantes

TICKET 1
Farine
Sucre
Lait

TICKET 2
Œuf
Sucre
Chocolat

TICKET 3
Farine
Œuf
Sucre
Chocolat

TICKET 4
Œuf
Chocolat
Thé

On décide de retenir un taux de support de 30%



Construction de C₃




item1	item2	item3	nombre
Farine	Sucre	Œuf	1
Farine	Sucre	Chocolat	1
Sucre	Œuf	Chocolat	2

Etude de cas de recherche d'associations intéressantes

TICKET 1	TICKET 2	TICKET 3	TICKET 4
Farine	Œuf	Farine	Œuf
Sucre	Sucre	Œuf	Chocolat
Lait	Chocolat	Sucre	Thé
		Chocolat	

On décide de retenir un taux de support de 30% 

Construction de C3

item1	item2	item3	nombre
Farine	Sucre	Œuf	1 
Farine	Sucre	Chocolat	1 
Sucre	Œuf	Chocolat	2 

L'ensemble L des sous-ensembles fréquents est l'union des ensembles L_1, \dots, L_K :

$L_1 = \{Farine, Sucre, Oeuf, Chocolat\}$
 $L_2 = \{(Farine, Sucre), (Sucre, Oeuf), (Sucre, Chocolat), (Oeuf, Chocolat)\}$
 $L_3 = \{(Sucre, Oeuf, Chocolat)\}$

- ✓ Pour chaque ensemble fréquent on construit des règles vérifiant la contrainte de seuil de confiance.
- ✓ Un algorithme simple pour produire des règles à partir d'un sous-ensemble fréquent f est de considérer tous les sous ensembles possibles g de f et de produire la règle $g \rightarrow (f-g)$ si la condition sur la confiance est vérifiée.
- ✓ Alors on commence par les règles ayant un seul conséquent, puis sur les règles retenues on génère les règles ayant deux conséquents.



L'ensemble L des sous-ensembles fréquents est l'union des ensembles L_1, \dots, L_K :

$L_1 = \{\text{Farine, Sucre, Oeuf, Chocolat}\}$

$L_2 = \{(\text{Farine, Sucre}), (\text{Sucre, Oeuf}), (\text{Sucre, Chocolat}), (\text{Oeuf, Chocolat})\}$

$L_3 = \{(\text{Sucre, Oeuf, Chocolat})\}$

Tableau des règles

Sous-ensemble	Règle	Support	Confiance
Farine, Sucre	Farine \rightarrow Sucre	2/4	2/2
	Sucre \rightarrow Farine	2/4	2/3
Sucre, Œuf	Sucre \rightarrow Œuf	2/4	2/3
	Œuf \rightarrow Sucre	2/4	2/3
Sucre, Chocolat	Sucre \rightarrow Chocolat	2/4	2/3
	Chocolat \rightarrow Sucre	2/4	2/3
Œuf, Chocolat	Œuf \rightarrow Chocolat	3/4	3/3
	Chocolat \rightarrow Œuf	3/4	3/3
Sucre, Œuf, Chocolat	Sucre \rightarrow Œuf + Chocolat	2/4	2/3
	Œuf + Chocolat \rightarrow Sucre	2/4	2/3
	Œuf \rightarrow Sucre + Chocolat	2/4	2/3
	Sucre + Chocolat \rightarrow Œuf	2/4	2/2
	Chocolat \rightarrow Sucre + Œuf	2/4	2/3
	Sucre + Œuf \rightarrow Chocolat	2/4	2/3

Tableau des règles

Sous-ensemble	Règle	Support	Confiance
Farine, Sucre	Farine \rightarrow Sucre	2/4	2/2
	Sucre \rightarrow Farine	2/4	2/3
Sucre, Œuf	Sucre \rightarrow Œuf	2/4	2/3
	Œuf \rightarrow Sucre	2/4	2/3
Sucre, Chocolat	Sucre \rightarrow Chocolat	2/4	2/3
	Chocolat \rightarrow Sucre	2/4	2/3
Œuf, Chocolat	Œuf \rightarrow Chocolat	3/4	3/3
	Chocolat \rightarrow Œuf	3/4	3/3
Sucre, Œuf, Chocolat	Sucre \rightarrow Œuf + Chocolat	2/4	2/3
	Œuf + Chocolat \rightarrow Sucre	2/4	2/3
	Œuf \rightarrow Sucre + Chocolat	2/4	2/3
	Sucre + Chocolat \rightarrow Œuf	2/4	2/2
	Chocolat \rightarrow Sucre + Œuf	2/4	2/3
	Sucre + Œuf \rightarrow Chocolat	2/4	2/3

Tableau des règles intéressantes

Sous-ensemble	Règle	Confiance	Lift
Farine, Sucre	Farine \rightarrow Sucre	2/2	4/3
Œuf, Chocolat	Œuf \rightarrow Chocolat	3/3	4/3
Œuf, Chocolat	Chocolat \rightarrow Œuf	3/3	4/3
Sucre, Œuf, Chocolat	Sucre + Chocolat \rightarrow Œuf	2/4	2/2



On a retenu les règles ayant une confiance égale à 1 et un lift supérieur à 1