



---

## THÉMATIQUE 7 - LES RÈGLES D'ASSOCIATION

---

Data Sciences Revision  
MENTION MATHÉMATIQUES ET INFORMATIQUE  
PARCOURS HPDA

10/10/2024

*Rédigé par :*

PAULY ALEXANDRE

alexandre.pauly@cy-tech.fr

L'objectif des règles d'association est de regrouper des objets qui vont naturellement ensemble pour définir des règles d'association du type : **SI condition ALORS résultat**.

Les règles d'association sont traditionnellement liées au secteur de la distribution, avec comme principale application «l'analyse du panier de la ménagère», c-a-d la recherche d'associations entre produits sur les tickets de caisse. La méthode recherche quels produits tendent à être achetés ensemble.

## 1 Confiance et support

Chaque transaction est constitué d'items. Ces items forment un ensemble de n éléments  $I = \text{Farine, Sucre, Lait, Chocolat, Œufs, Thé}$ . On construit un tableau récapitulatif des items par transaction :

Table 1: Table de vérité

	Farine	Sucre	Lait	Chocolat	Oeufs	Thé
Ticket 1	X	X	X		X	X
Ticket 2	X	X		X	X	
Ticket 3		X		X	X	
Ticket 4				X	X	X

Un tel tableau permet de déterminer la fréquence d'un produit,  $Frquence = \frac{nb\ de\ fois\ achet}{nb\ d'\ achats\ total}$ . On définit alors deux métriques :

Le **support** d'une règle  $X \rightarrow Y$  est la fréquence à laquelle les items X et Y apparaissent simultanément dans les transactions (\*ex :  $\text{Sup}(\text{Farine} \rightarrow \text{Sucre}) = 50\%$  signifie que les produits farine et sucre apparaissent dans 50% des transactions\*).

La **confiance** d'une règle  $X \rightarrow Y$  est le rapport entre son support et la fréquence de X (\*ex :  $\text{Conf}(\text{farine} \rightarrow \text{sucré}) = 100\%$  signifie que 100% des transactions contenant le produit farine contiennent aussi le produit sucre\*).

## 2 Le Lift ou Amélioration

Dans l'exemple précédent, on calcule le support et la confiance de la règle :  $\text{Sucre, Œufs} \rightarrow \text{Chocolat}$ . Donc si Sucre et Œufs apparaissent dans un ticket alors il y a 67% de chance de voir aussi Chocolat.

Le lift ou amélioration permet de comparer la prédiction d'un résultat avec une règle ou sans règle,  $Lift = \text{confiance} / \text{support}(\text{rsultat})$

- $Lift > 1$ , la règle améliore la prédiction

- $Lift < 1$ , la règle ne sert à rien

**Remarques :**

- $Lift(X -> Y) = \frac{Conf(X -> Y)}{freq(Y)} = \frac{freq(X,Y)}{freq(X)freq(Y)}$
- $Lift(X -> Y) = Lift(Y -> X)$