

TD N° 2
Approches non supervisées

Nous disposons des données suivantes :

IdClient	Produits achetés
1	A,D
2	B,C,E
3	A,E
4	C,D
5	C,D,E
6	A,B,D
7	A,C,E
8	A,B,C
9	B,D

1 Clustering

- 1) Face à ce type de données, quelle mesure de distance vous semble la plus appropriée ?
- 2) Appliquez l'algorithme des k-moyennes ($k=2$) sur les 6 premiers clients de la base.
- 3) Appliquez l'algorithme DBSCAN($\text{minPts}=2$, $\text{eps} = 2$) sur les 6 premiers clients de la base.
- 4) Appliquez une CAH (méthode du lien unique) sur les 6 premiers clients de la base.

2 Calcul de motifs fréquents

- 1) Calculez les règles d'association avec $\text{minSupp} = 20\%$ et $\text{minConf} = 50\%$
- 2) Quels sont les itemsets fréquents, maximaux et fermés ?