

Data mining
 Série N° 1

Objectif de la série : Data Mining, règles d'associations, arbres de décision.

Exercice 1 : Algorithme Apriori

Soit D la base de transactions suivante :

T1	1 3 4
T2	2 3 5
T3	1 2 3 5
T4	2 5

1. Appliquer l'algorithme Apriori pour extraire les sous-ensembles d'itemsets fréquents avec $\text{minSupp} = 2$.
2. Générer les règles d'association avec $\text{minConf} = 1$

Exercice 2 : Extraction de règles d'associations

Vous avez 6 tablettes {T1,T2,T3,T4,T5,T6}. Chaque tablette est décrite par une liste de caractéristiques:

Tablette1(T1): GPS , LCD, 3G , Bluetooth (B), Wi-Fi (WF), USB, HDMI

Tablette 2(T2) : Wi-Fi (WF), USB,

Tablette 3(T3) : LCD, Bluetooth (B)

Tablette 4(T4) : Bluetooth (B), Wi-Fi(WF)

Tablette 5(T5) : GPS, LCD, Bluetooth (B), Wi-Fi

Tablette 6 (T6): GPS, 3G , Bluetooth (B), Wi-Fi (WF), USB, HDMI

1. Appliquer l'algorithme APRIORI pour rechercher les itemsets fréquents ayant un support supérieur à 35%.
2. Donner les règles d'association les plus pertinentes, avec une confiance minimum de 50%.

Exercice 3 : Extraction de règles d'associations

Soit le problème suivant :

Produits = {A, B, C, D, E, F}

Transactions = {AB, ABCD, ABD, ABDF, ACDE, BCDF}

Transaction	A	B	C	D	E	F
T1	1	1	0	0	0	0
T2	1	1	1	1	0	0
T3	1	1	0	1	0	0
T4	1	1	0	1	0	1
T5	1	0	1	1	1	0
T6	0	1	1	1	0	1

Avec : MINSUP = $\frac{1}{2}$, MINCONF= 75%

1. Donner l'ensemble des ItemSets.
2. Donner l'ensemble des règles.
3. Peut-on préciser les règles les plus pertinentes ? commenter et justifier ta réponse.

Exercice 4 : Extraction de règles d'associations

Le tableau suivant donne la description de cinq transactions au niveau d'un point de vente dans une épicerie. Chaque transaction identifie le sous-ensemble de produits achetés simultanément.

T1 : jus d'orange, eau minérale

T2 : lait, jus d'orange, nettoyant de vitre

T3 : jus d'orange, détergent

T4 : jus d'orange, détergent, eau minérale

T5 : nettoyant de vitre, eau minérale

1. Procédez à l'extraction des règles d'associations les plus pertinentes en entachant chacune d'elle de son degré de confiance et de support