
Analyse du panier du marché

Nous considérons dans ce TP la base "Groceries" disponible sur "http://www.salemmarafi.com/wp-content/uploads/2014/03/groceries.csv". Cette base, sous format csv, représente une collection de d'items achetés dans une épicerie regroupés par panier (transaction). Chaque ligne représente un panier (une transaction) et chaque colonne représente un item acheté. Nous souhaitons analyser cet ensemble en vue d'extraire les règles solides pertinentes.

Travail 1 Analyse du format de la base

1. Ouvrir la base dans un éditeur de textes,
2. Vérifier la structure des lignes et des colonnes,
3. Peut-on utiliser la base telle qu'elle sur weka ? Pourquoi ?
4. Peut-on l'utiliser avec l'algorithme Apriori ? Pourquoi ?
5. Proposer les modification à apporter sur la base,
6. Proposer les étapes et les outils nécessaires à sa transformation en un fichier "arff".

Travail 2 Préparation

On vous demande de transformer la base en une base Weka (arff) en tenant compte de :

1. Le nombre de colonnes dans chaque ligne,
2. L'ordre des items dans chaque ligne,
3. Les entêtes des colonnes.

Travail 3 Analyse du panier

Sur weka :

1. Ouvrir la base précédente.
2. Lancer l'algorithme apriori.
3. Donner le nombre de motifs fréquents.
4. Donner la longueur maximale des motifs fréquents.
5. Trouver les vingt meilleures règles.
6. Commenter.