

**IFT 599/799 – Science de données**  
**TP2 : Règles d'associations**  
**Date de remise : le jeudi 4 novembre 2021**

Ce TP porte sur l'extraction des règles d'associations. Il consiste à développer l'étude d'une base de données bancaire à l'aide des techniques d'analyse d'association. Le but de l'étude est d'identifier des profils d'épargnants parmi les clients et identifier des non épargnants qui seraient intéressés d'acheter des produits d'épargne.

**Analyse d'une base de données bancaire**

Les données du TP :

- "bank-data" fournies en deux formats, .csv et .ARFF (format de Weka).
- Explication des attributs : Tableau 1.

Le travail peut être réalisé par Weka. Vous pouvez aussi le faire avec un langage de votre choix. Par exemple, si vous voulez le faire avec Python, la page suivante vous donne des informations sur l'extraction des règles en utilisant l'algorithme Apriori avec Python.

<https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/>

Un épargnant est une personne ayant soit un « mortgage » soit un plan « pep ». Les tâches à effectuer pour ce TP sont :

- 1- Extraire des règles significatives (intéressantes) pour décrire les profils des épargnants. Les seuils comme min-conf et min-lift n'étant pas fournis, vous devez chercher des règles ayant les plus grandes valeurs de *confiance* et de « *lift* ». Le but n'est pas d'extraire TOUTES les règles possibles, une dizaine, s'il y en a, serait suffisant;
- 2- Identifier des clients non-épargnants correspondant aux profils extraits afin de considérer ces clients comme susceptibles de devenir des épargnants ;

Vous utilisez l'algorithme *Apriori* pour l'extraction des "itemsets" fréquents. Votre rapport doit décrire les différentes étapes de traitement incluant les prétraitements effectués, les résultats obtenus, vos commentaires et votre conclusion.

Vous avez une certaine liberté pour développer votre propre solution. Vous devez faire preuve d'un peu d'imagination. Il n'y a pas UNE meilleure solution. Toutefois, le bon sens est de mise.

TABLE 1 – Explication des attributs

id	a unique identification number
age	age of customer in years (numeric)
sex	MALE / FEMALE
region	inner_city/rural/suburban/town
income	income of customer (numeric)
married	is the customer married (YES/NO)
children	number of children (numeric)
car	does the customer own a car (YES/NO)
save_acct	does the customer have a saving account (YES/NO)
current_acct	does the customer have a current account (YES/NO)
mortgage	does the customer have a mortgage (YES/NO)
pep	did the customer buy a PEP (Personal Equity Plan) after the last mailing (YES/NO)

### Points Importants :

1. Le TP doit être fait seul ou en équipe de deux personnes;
2. La date limite de remise officielle est le jeudi 4 novembre 23h59;
3. Soignez votre rapport, une pénalité de 10% pourrait être appliquée pour un rapport mal écrit; Votre rapport ne doit pas dépasser 5 pages + la page de couverture (pas indispensable);
4. Les fichiers à soumettre sont votre programme(s) et le rapport (en Word ou pdf).  
**Ne pas soumettre les données!**
5. N'oubliez pas de vous identifier. Indiquez votre nom et matricule (ou cip) dans chacun des fichiers que vous soumettez. La remise doit être faite par <http://turnin.dinf.usherbrooke.ca>