

<b>Matière Apprentissage artificiel</b>
<b>Classe 3 IDL</b>
<b>Chargé cours Amel BORGI</b>
<b>Chargé TP Arafet ZOUARI</b>
<b>Année universitaire 2022 - 2023</b>

## TP 1 : Les règles d'associations

### Données transactionnelles:

Le fichier "market\_basket.txt" est un fichier texte qui se présente sous la forme d'une base transactionnelle (voir – "Règles d'association – Données transactionnelles", décembre 2010 –pour les spécificités de cette organisation des données).

Nous avons un tableau de données sur deux colonnes : la première correspond aux identifiants de transaction, la seconde aux noms des produits. Voici les premières lignes du fichier :

ID	Product
1	Peaches
2	Vegetable_Oil
2	Frozen_Corn
3	Plums
4	Pancake_Mix
5	Cheese
6	Cauliflower
7	2pct_Milk
8	98pct_Fat_Free_Hamburger
8	Potato_Chips
8	Sesame_Oil
8	Ice_Cream_Sandwich

### Travail demandé:

- 1- Créer Un DataFrame en utilisant les données de fichier "market\_basket.txt" qui contient les identifiants des caddies et les produits associées.
- 2-Afficher les 10 premières lignes du DataFrame.
- 3-Afficher les dimensions du dataframe.

- 4-Écrire un script python qui permet de Construire un table binaire indiquant la présence de chaque produit au niveau des caddies (True:1 si le produit est présent dans le caddie et 0 dans le cas réciproque).
- 5-Tester la bibliothèque pandas.crosstab pour construire la table binaire et vérifier que vous avez les mêmes résultats de votre script.
- 6-Afficher les 30 premières transactions et les 3 premiers produits.
- 7-Écrire un script python de la fonction a\_priori() qui permet l'extraction des itemsets les plus fréquents. ( on définit un **min\_supp=0.025** et un longueur maximum de **4** produits)
- 8-Afficher les 15 premiers itemsets.
- 9-Ecrire une fonction is\_inclus() qui permet de vérifier si un sous-ensemble items est inclus dans l'ensemble x.
- 10-Afficher les itemsets comprenant le produit 'Aspirin'.
- 11-Afficher les itemsets contenant Aspirin et Eggs.
- 12-Nous produisons les règles à partir des itemsets fréquents. Elles peuvent être très nombreuses,nous en limitons la prolifération en définissant un seuil minimal (**min\_threshold = 0.75**) sur une mesure d'intérêt, en l'occurrence la confiance dans notre exemple (metric = "confidence").  
Utiliser la bibliothèque mlxtend.frequent\_patterns pour générer les règles d'associations.
- 13-Afficher les 5 premières règles.
- 14-Filtrer les règles en affichant celles qui présentent un LIFT supérieur ou égal à 7.
- 15-Filtrer les règles en affichant celles menant au conséquent {'2pct\_milk'}.