

Forage des Données - 8INF436 UQAC – Hiver 2025

TP individuel 2

Date de remise du travail : 21/03/2025 à 23H55

Modalité de dépôt : Espace de dépôt sur Moodle

Documents à rendre : Code source + rapport des résultats

Evaluation et Pondération : Barème : sur 100, Pondération : 0.2

Note importante : attention au plagiat !

Enoncé du TP : Forage des règles d'association sur le « Bank Marketing Data Set »

Dans le cadre de l'apprentissage non supervisé, ce TP s'intéresse particulièrement au forage des règles d'association à partir du « Bank Marketing Data Set » :

<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>

Cibler le bon public pour une campagne marketing peut faire économiser à une entreprise des milliers de dollars, si elle est menée dans la bonne direction (en tirant parti de l'analyse pour prendre une décision), ce qui conduira à un taux de conversion élevé. Les données du « Bank Marketing Data Set » sont liées à des campagnes de marketing direct (appels téléphoniques) d'une institution bancaire portugaise. L'objectif de l'analyse est de prédire si le client souscrira à un dépôt à terme (variable cible y).

L'objectif de ce TP est de trouver les caractéristiques fréquentes des clients qui souscrivent à un dépôt à terme. Les algorithmes **Apriori** et **PF-growth** seront utilisés pour extraire des règles qui peuvent aider au marketing ciblé, sur la base des données bancaires et des informations démographiques des clients.

Pour cela, vous êtes demandé de :

1- Préparer des données pour le forage des règles d'association :

a- **(Sur 15 points)** Discrétiser les variables continues (en intervalles ou en catégories)

b- **(Sur 15 points)** Transformer les données en format binaires (0/1 ou True/False) qui est le format d'entrée requis pour les algorithmes *Apriori* et *fpgrowth*. Pour cela, des techniques telles que « One Hot Encode » ou « get_dummies » (ou autres) peuvent être utilisées.

2- (Sur 30 points) Implémenter les algorithmes **Apriori** et **FP-Growth** et extraire les règles d'association tout en :

- Mentionnant le support minimum requis.
- Triant le résultat par confiance (tri descendant).
- Comparant le temps d'exécution des deux algorithmes.

3- (Sur 20 points) Reprendre la **question 2** et générer uniquement les règles d'association qui ont comme conséquence la variable cible **y** : souscription à un dépôt à terme ou non.

4- (Sur 20 points) Interpréter les antécédents des règles générées en **question 3** pour analyser les caractéristiques principales fréquentes chez un client qui s'inscrit à un dépôt à terme ou pas chez cette institution bancaire portugaise.

Bon travail