

# TP EPITA – KANTAR

## 14 et 21 novembre 2024

A partir des données fournies par Kantar :

- fic\_epita\_kantar\_codes.cs  
ou
- fic\_epita\_kantar\_labels.csv

Il vous est demandé de réaliser :

### 1/ Clustering

- Une clusterisation des 5000 individus en utilisant les variables A9 A10 et A11 (en orange dans le fichier variables\_fichier.xlsx)
- Une clusterisation des 5000 individus en utilisant les variables A11 A12 A13 A14 A4 A5 A5bis A8\_1\_slice A8\_2\_slice A8\_3\_slice A8\_4\_slice B1\_1\_slice B1\_2\_slice B2\_1\_slice B2\_2\_slice B3 B4 B6 C1\_1\_slice C1\_2\_slice C1\_3\_slice C1\_4\_slice C1\_5\_slice C1\_6\_slice C1\_7\_slice C1\_8\_slice C1\_9\_slice (en vert dans le fichier variables\_fichier.xlsx)

Vous devrez justifier du choix de la méthode de clusterisation (ou des méthodes) utilisée(s) et expliquer les raisons du choix de la version de clustering choisie (nb de groupes). Calcul des variances intra groupes et inter groupes. (ratio)  
Pour chacun des groupes finaux, un descriptif succinct devra être fourni.

### 2/ Réaffectation des individus dans les groupes avec variables actives

A partir des variables utilisées pour construire les clusters, vous devrez établir un algorithme permettant de réaffecter les individus dans les groupes (échantillon d'apprentissage / échantillon test). La réaffectation devra minimiser le nombre de variables utilisées en maximisant le % de bon classement. (Golden questions)

L'objectif est de déterminer un algorithme qui pourra être utilisé sur une autre enquête. (des nouveaux individus répondront aux « golden question » et pourront être affectés dans les différents groupes.

### **3/ Réaffectation des individus dans les groupes avec variables illustratives**

Pour chacune des 2 clusterisations, vous devez affecter les individus dans les groupes en utilisant des variables illustratives.

- Pour la segmentation variables en Orange
  - o Affectation avec les variables en Vert
  - o Affectation avec les variables suivantes rs3 rs5 rs6 RS1 RS191 RS192 RS193 RS102RECAP rs11recap2 RS11recap RS193bis RS2Recap RS56Recap RS2 RS11 RS102
- Pour la segmentation variables en Vert
  - o Affectation avec les variables en Orange
  - o Affectation avec les variables suivantes rs3 rs5 rs6 RS1 RS191 RS192 RS193 RS102RECAP rs11recap2 RS11recap RS193bis RS2Recap RS56Recap RS2 RS11 RS102

L'objectif est de déterminer des algorithmes qui pourront être utilisés sur un fichier annexe ou fichier client dans lequel nous ne disposons que des variables sélectionnées. Déterminer la capacité d'affectation les individus avec ces variables illustratives.

#### **Document à fournir :**

Fichier doc ou pdf présentant :

- La et les méthodes de clusterisation choisies avec justification
- La présentation des clusters obtenus pour les 2 clusterisations
- La ou les méthodes de réaffectation utilisées avec justification.
- Le % de bon classement obtenu dans les 4 cas.

Joindre les codes des programmes

Document à envoyer à david.mermoud@kantarc.com