

Projet Kantar •

Maxime Buisson, Léa Margery, Dorian Penso



<https://github.com/scia-assistant/kantar-LMD>

Sommaire.

- 1 ● Méthodes de clusterisation
- 2 ● Présentation des clusters
- 3 ● Méthodes de réaffectation
- 4 ● Classement obtenu

Sommaire.

1

Méthodes de clusterisation

Méthodes de clusterisation •

Voici, les différentes méthodes de clusterisation que nous avons testés :

- KNN
- Hierarchical
- Birch
- DBSCAN

Après avoir effectué une **PCA**, nous avons constaté une amélioration significative des résultats, ce qui nous a conduit à *l'utiliser systématiquement sur tous les modèles*.

Pour évaluer les différentes méthodes, nous nous basons sur les **ratios inter/intra-cluster**, pour une séparation optimale des clusters, et l'**indice de silhouette**, reflétant une bonne cohésion intra-cluster.

Voici les résultats que nous avons obtenus pour chaque méthode:

Méthode	(k)	Ratio inter/intra	Silhouette
K-means	3	1.5156	0.324
Hiérarchique	3	1.3204	0.295
BIRCH	3	1.3264	0.305
DBSCAN	-	0.0002	Non applicable

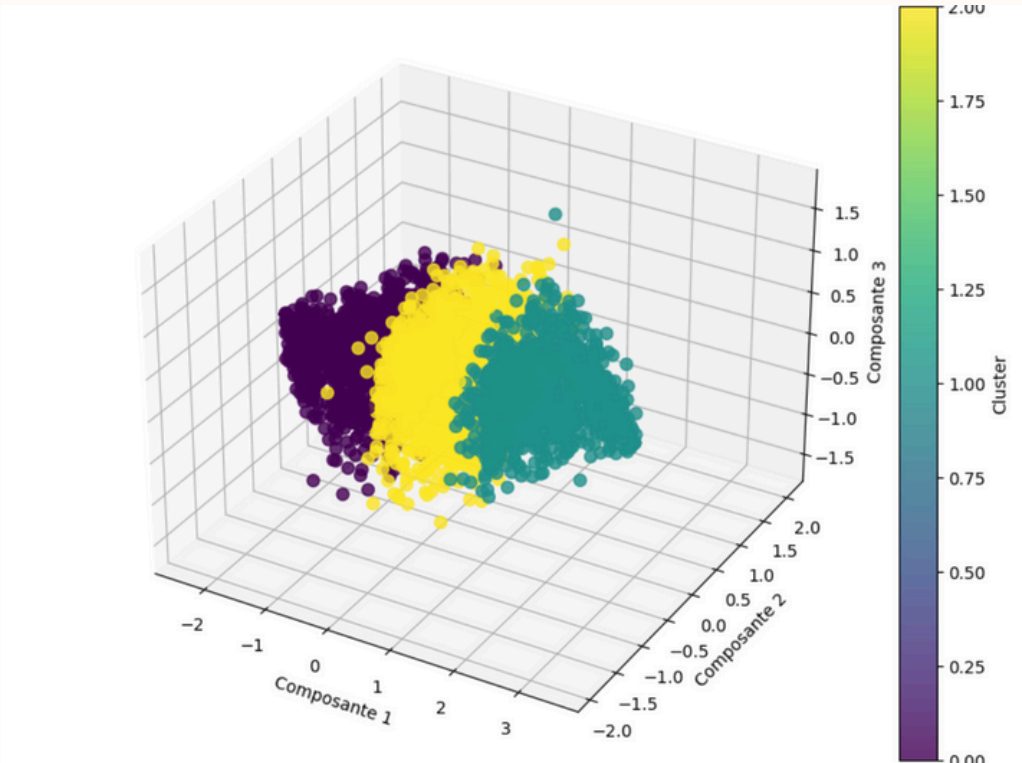
Résultats avec PCA sur l'attitude des consommateurs

Méthode	(k)	Ratio inter/intra	Silhouette
K-means	5	3.0109	0.4911
Hiérarchique	4	1.5323	0.4705
BIRCH	6	2.0966	0.4508
DBSCAN	-	0.3271	0.3047

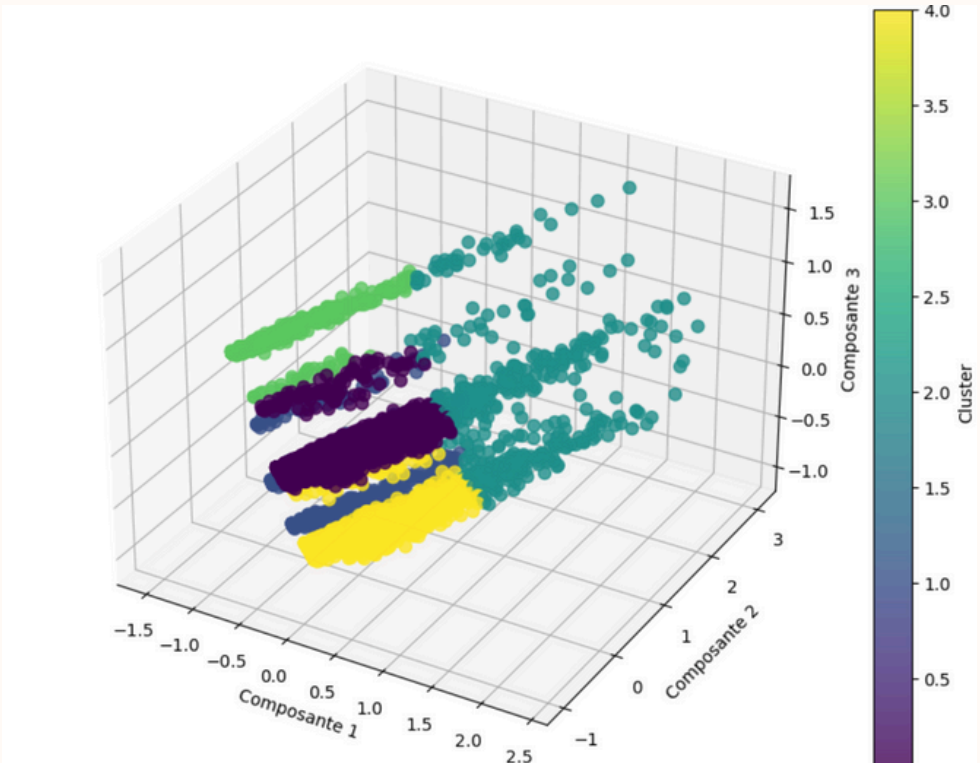
Résultats avec PCA sur les usages de consommateurs

Les stratégies que nous avons retenues avec leur visualisation sont:

- Sur l'attitude des consommateurs: **KNN avec PCA et k=3**
- Sur les usage des consommateurs: **KNN avec PCA et k=5**



Clusters avec KNN k=3 et PCA (attitude)



Clusters avec KNN k=5 et PCA (usages)

Sommaire.



2

Présentation des clusters

Clusters obtenus •

Attitudes



Les passionnés

Nous avons un premier cluster composé de 1385 personnes qui aiment fortement tout ce qui concerne le jardinage, les espaces verts et tout ce qui se rapproche de près ou de loin à ce domaine.



Les apaisés

Le deuxième cluster est composé de 1006 personnes qui n'aiment pas le jardinage mais qui trouvent quand même une certaine beauté et du plaisir dans les espaces verts.



Les paresseux

Le troisième cluster composé de 2609 personnes qui aiment fortement tout ce qui concerne le jardinage, les espaces verts et tout ce qui se rapproche de près ou de loin à ce domaine. Cependant, ils ne voient pas l'intérêt de s'intéresser et de faire des recherches sur le sujet.

Clusters obtenus •

Usages 1/2



Le cluster 1 regroupe des personnes peu engagées dans les activités numériques liées au jardinage, mais qui entretiennent un petit espace extérieur.



Le cluster 2 regroupe des personnes qui disposent principalement d'une terrasse et ne s'investissent pas particulièrement dans l'entretien ou le jardinage. Ce sont des individus peu engagés dans le jardinage ou les espaces extérieurs.



Le cluster 3 regroupe des personnes intéressées par le jardinage, disposant d'espaces extérieurs modérés et qui utilisent des ressources numériques pour s'informer ou s'inspirer.

Clusters obtenus •

Usages 2/2



Le cluster 4 regroupe des personnes vivant probablement en milieu urbain, disposant uniquement de balcons, et peu engagées dans les ressources numériques liées au jardinage.



Le cluster 5 regroupe des personnes disposant d'espaces extérieurs relativement grands mais qui ne s'engagent pas dans les ressources numériques. Elles se concentrent sur des activités pratiques plutôt que sur l'inspiration en ligne.

Sommaire.



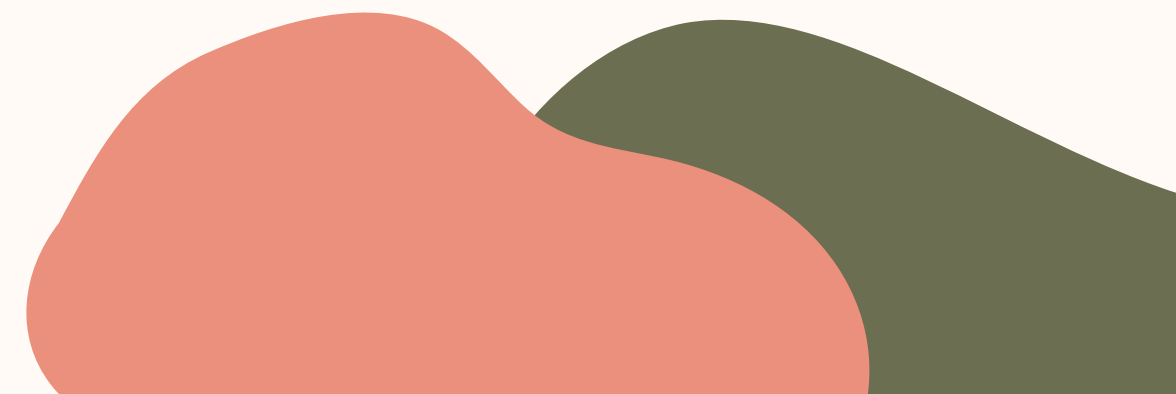
3

Méthodes de réaffectation

Méthodes de réaffectation •

Analyse des caractéristiques

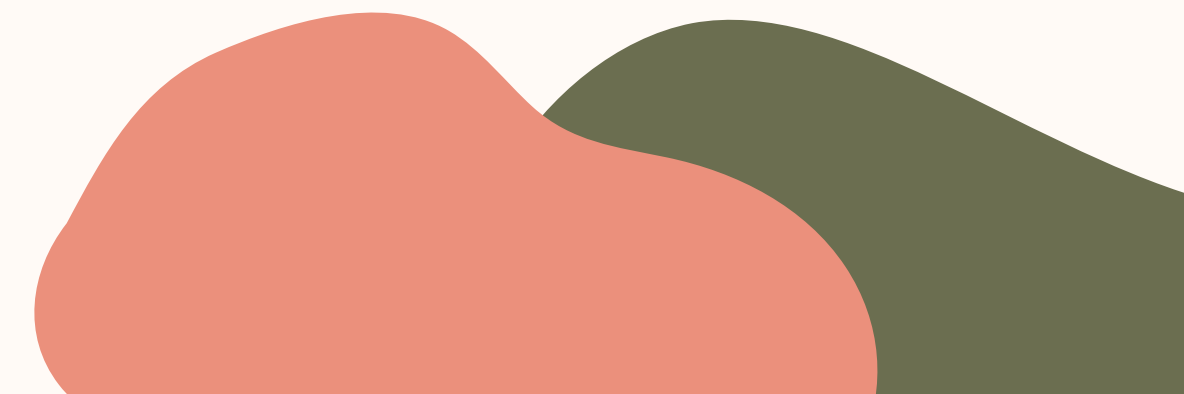
- Entraîner un modèle *Random Forest* sur les données disponibles.
- Identifier les caractéristiques les plus importantes à partir du modèle entraîné.
- Analyser et observer l'impact des caractéristiques sélectionnées sur les résultats.
- Réentraîner le modèle avec la liste des caractéristiques les plus importantes pour comparer les performances.



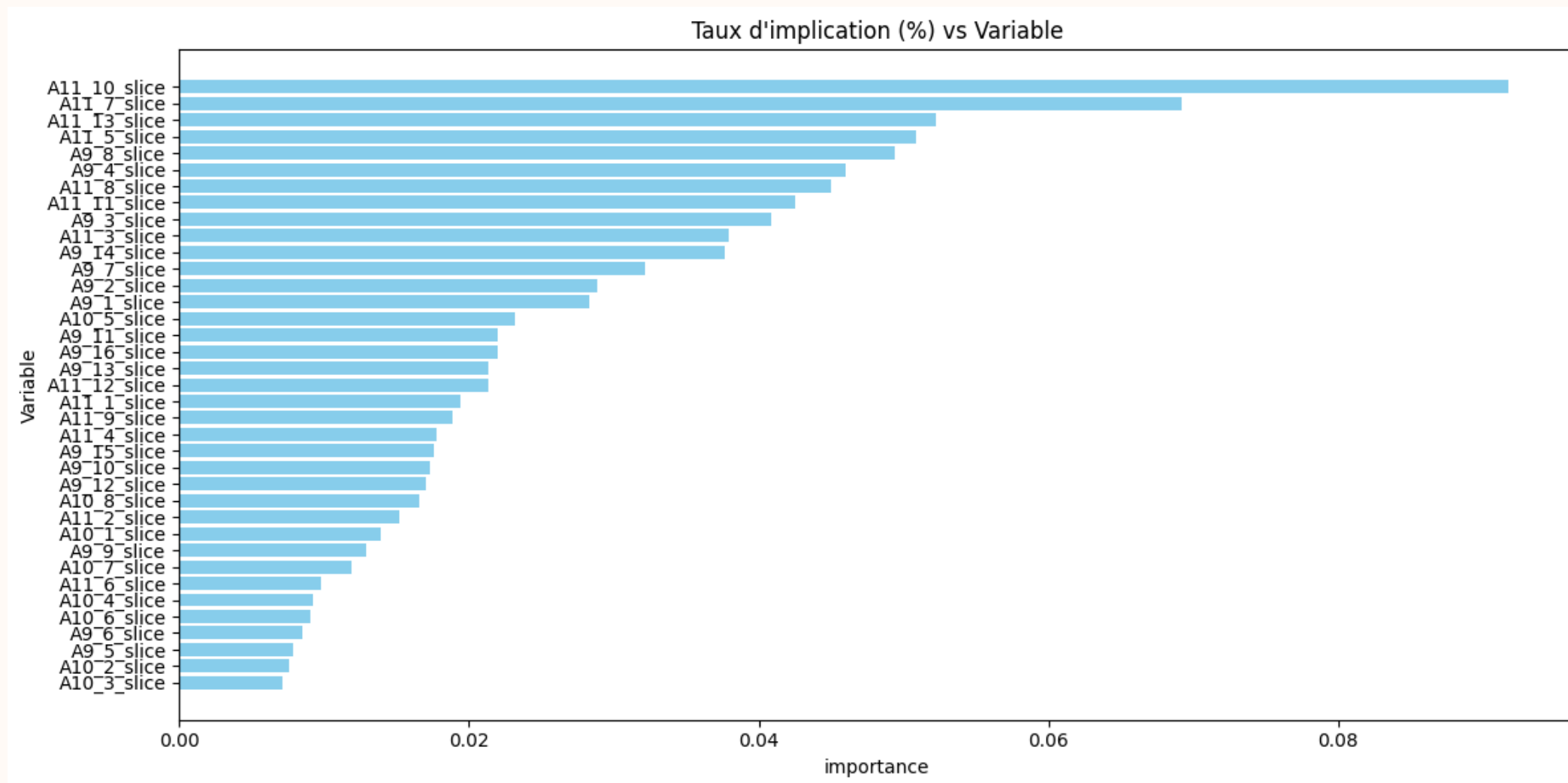
Méthodes de réaffectation •

Analyse des caractéristiques: **ATTITUDES**

Analyse des caractéristiques: **USAGES**



Méthodes de réaffectation •



Les questions les plus importantes:

Pour vous, les espaces extérieurs de votre logement (jardin, terrasses, balcons) sont ... ?

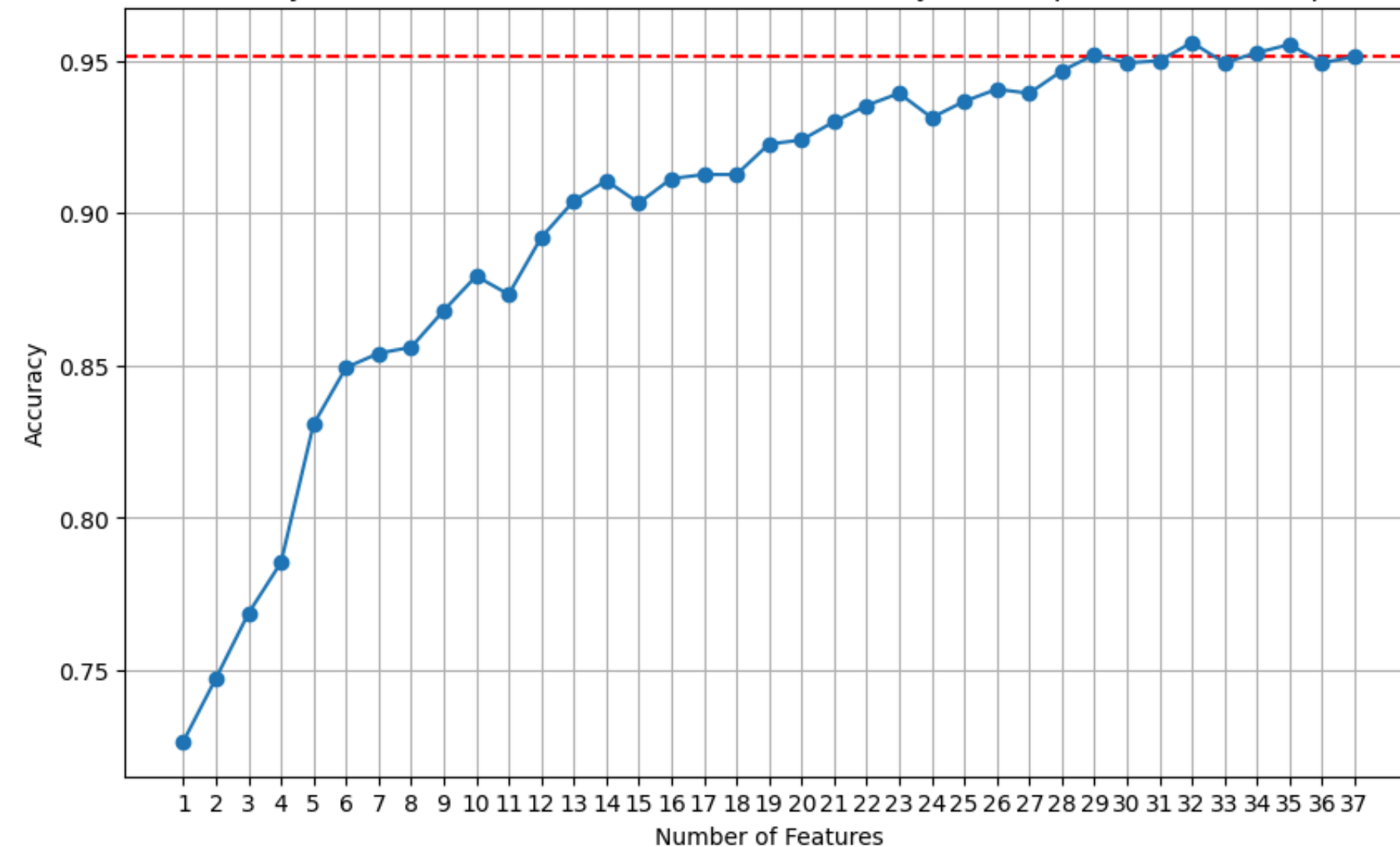
- Votre contact privilégié avec le vert, la nature

Pour vous, les espaces extérieurs de votre logement (jardin, terrasses, balcons) sont ... ?

- Un moyen de réaliser jusqu'au bout quelque chose de vos propres mains
- Un moyen vous permettant de rester en bonne santé (en récoltant vos propres fruits et légumes par exemple)

Méthodes de réaffectation •

Evolution of Accuracy with Incremental Feature Addition (sorted by most important to least important feature)

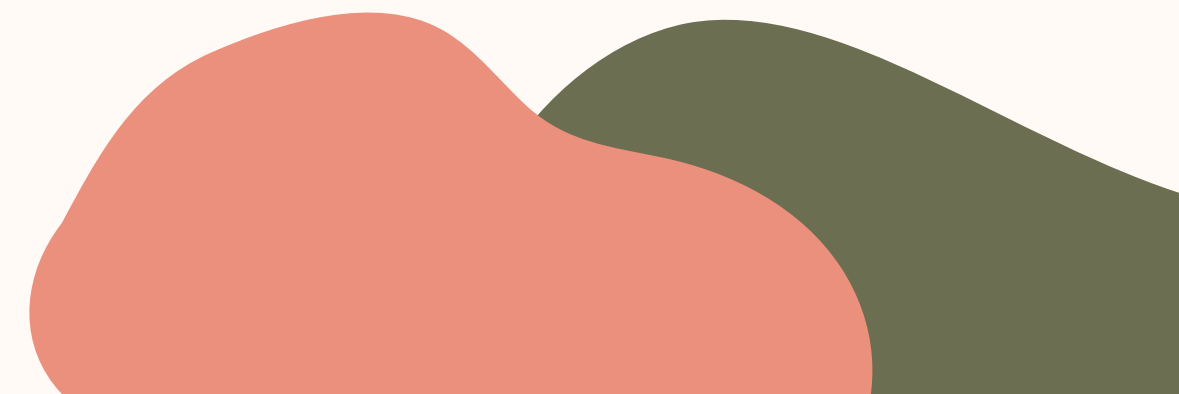


- 95 % accuracy sur toutes les caractéristiques,
- + 90 % avec seulement avec au moins 13 golden questions

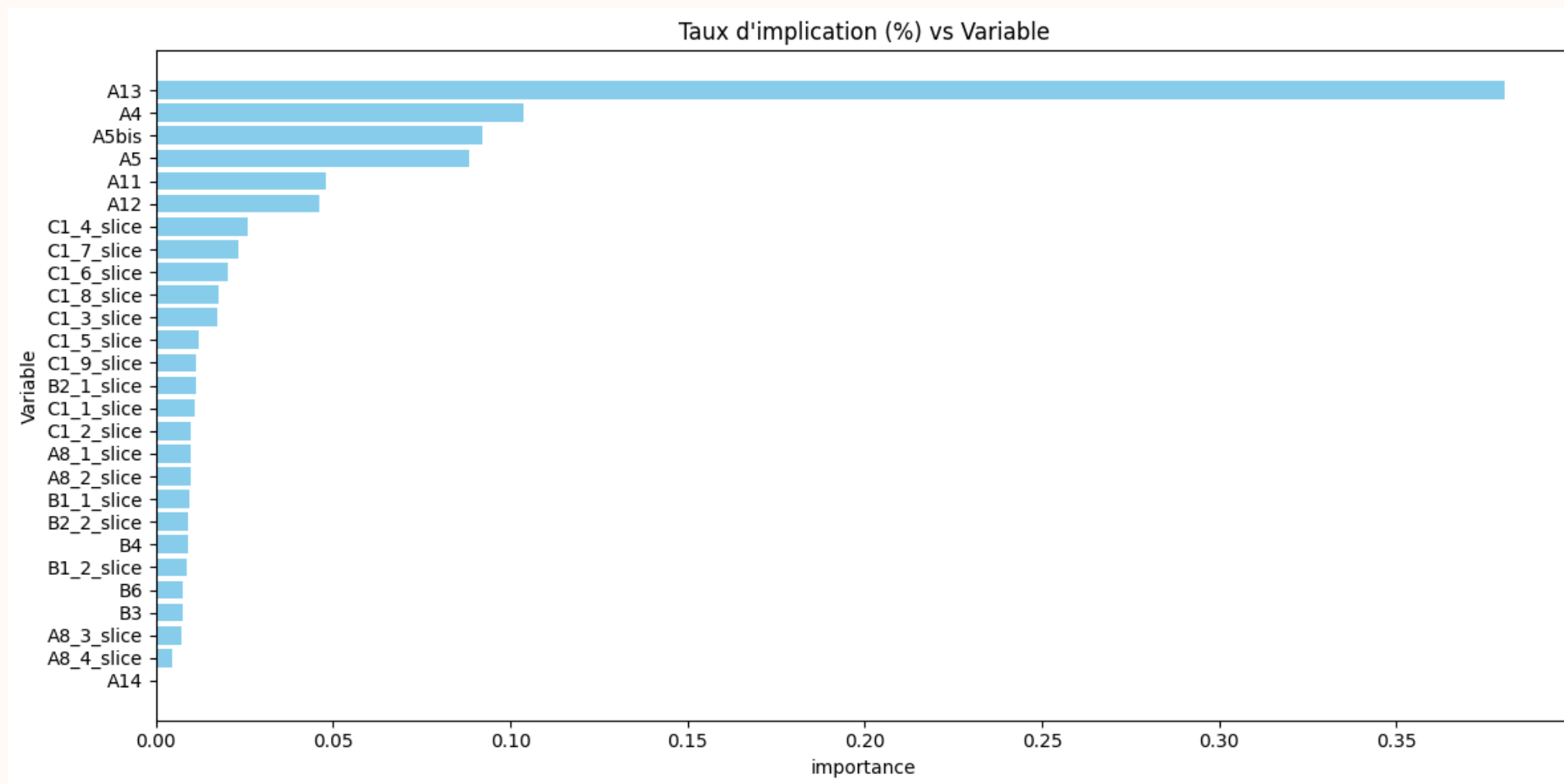
Méthodes de réaffectation •

Analyse des caractéristiques: ATTITUDES

Analyse des caractéristiques: USAGES



Méthodes de réaffectation •

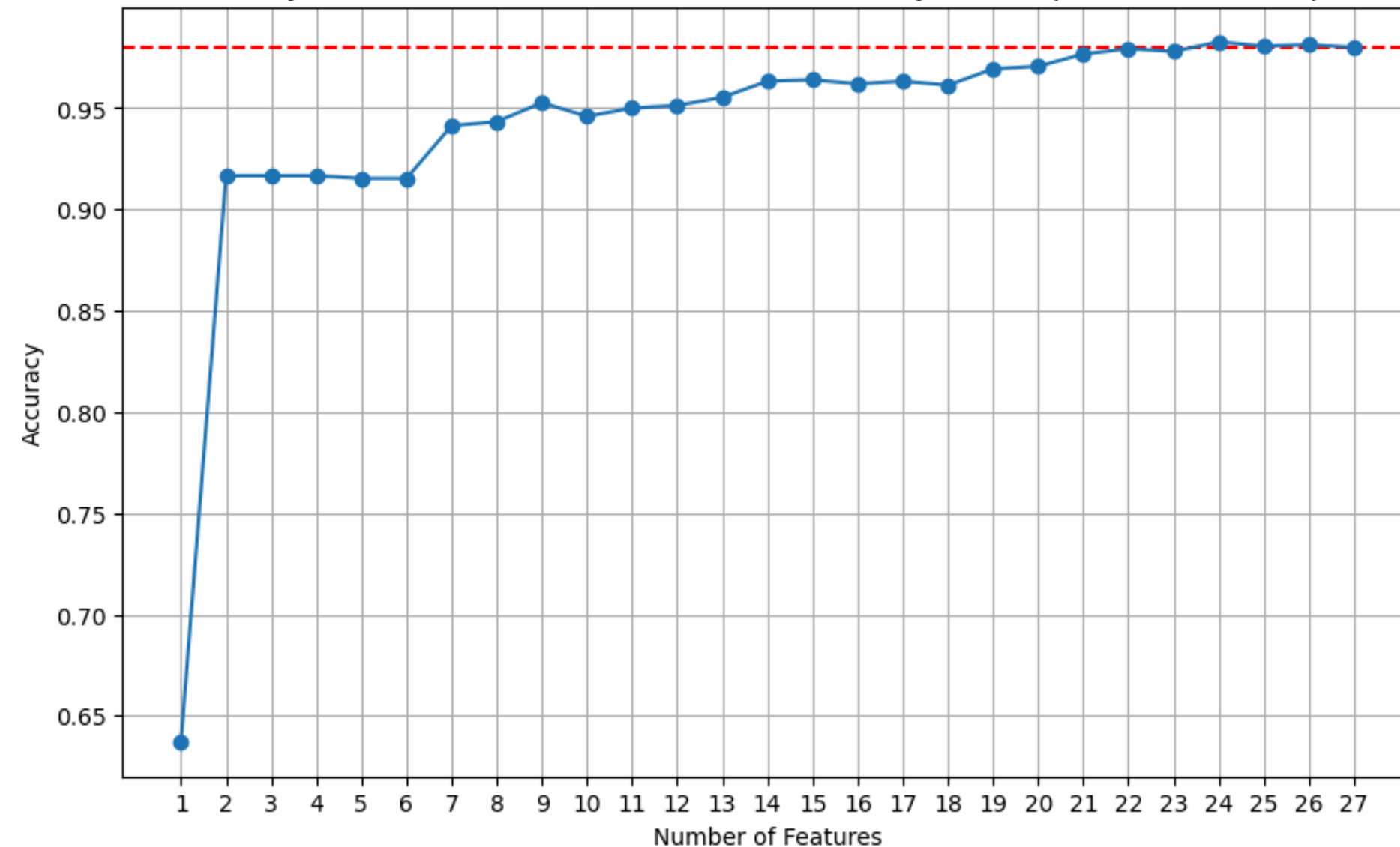


Les questions les plus importantes portent sur:

- Entretien de jardin/terrasse/balcon - Terrasse
- TYPE D'ESPACE
- Taille de la terrasse/balcon

Méthodes de réaffectation •

Evolution of Accuracy with Incremental Feature Addition (sorted by most important to least important feature)



- 97 % accuracy sur toutes les caractéristiques,
- + 90 % avec seulement 2 questions

Sommaire.



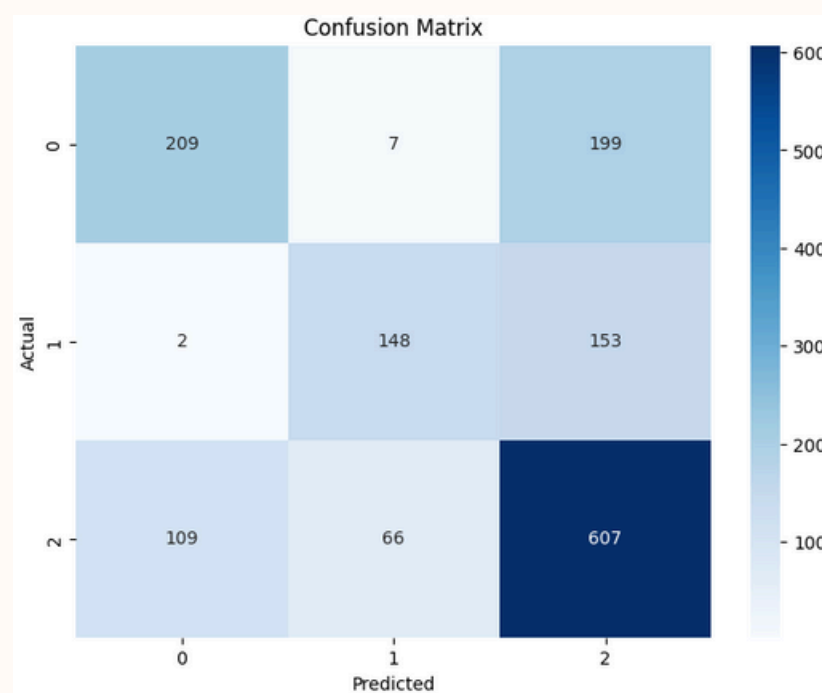
4

Classement obtenu

Classement obtenu •

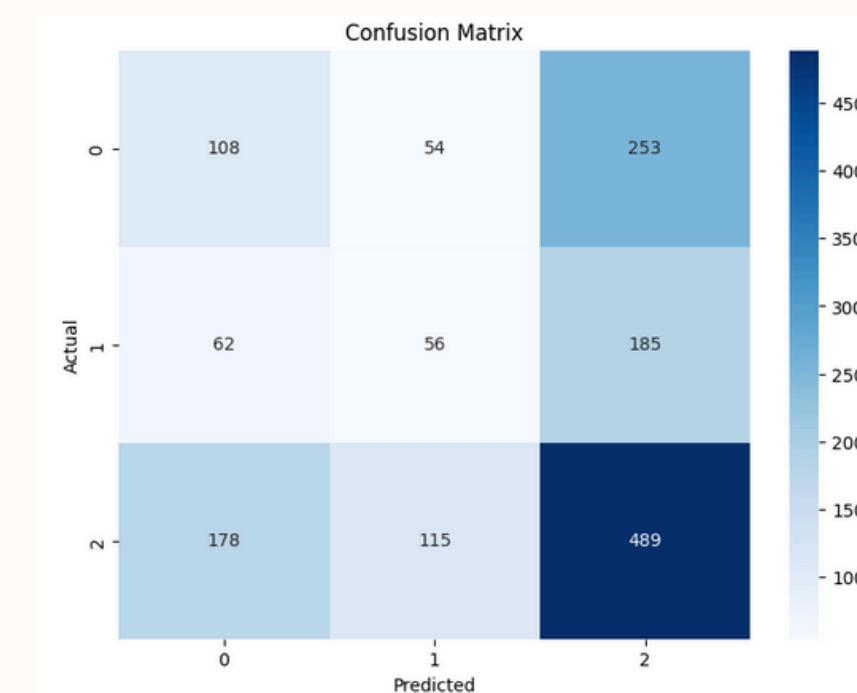
Usages entraînée sur le clustering des attitudes

- Précision : 64.3 %
- Interprétation : Ce résultat est correct, mais montre des difficultés à différencier certaines classes. Par exemple :
 - La classe 0 a un rappel de seulement 50 %, indiquant que beaucoup de ses éléments sont mal classés, notamment dans la classe 2.
 - La classe 2 est la mieux prédite avec un rappel de 78 % et un F1-score de 70 %, ce qui en fait la classe la plus robuste pour ce modèle.
 - La classe 1 est mal différenciée, avec une précision correcte (67 %) mais un rappel faible (49 %).



D'autres données entraînées sur le clustering des attitudes

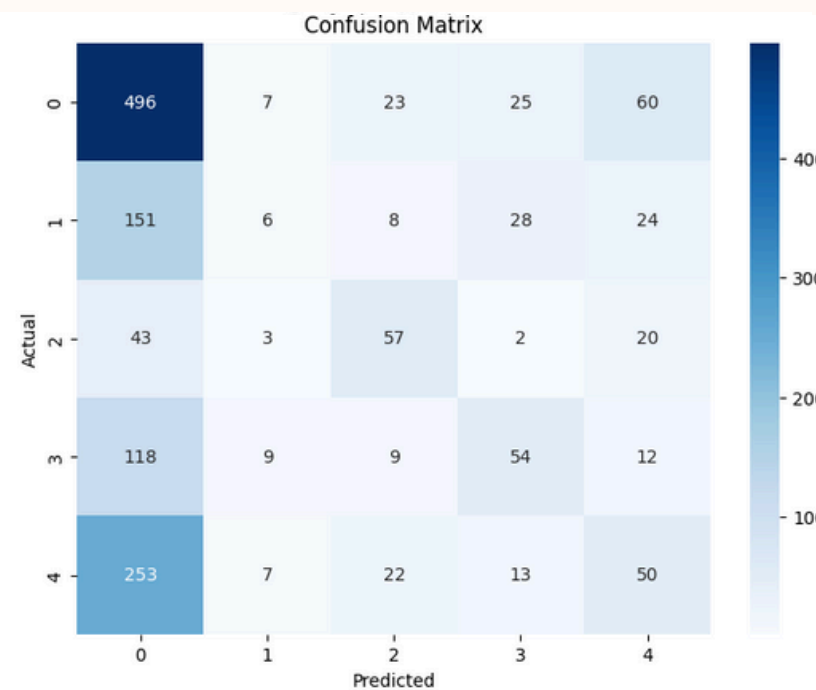
- Précision : 43.5 %
- Interprétation : Ce résultat montre des performances faibles, notamment pour les classes 0 et 1. Seule la classe 2 atteint un niveau de performance légèrement acceptable :
 - Classe 2 : La mieux prédite, avec un rappel de 63 % et un F1-score de 57 %, mais des erreurs significatives persistent.
 - Classe 0 et 1 : Très mal classées, avec des rappels de seulement 26 % et 18 %, ce qui montre des difficultés majeures à différencier ces classes.



Classement obtenu

Données attitude entraînées sur le clustering des usages

- Précision : 44.2 %
- Interprétation : Ce résultat indique des difficultés significatives dans la classification. Voici les points clés :
 - Classe 0 : La mieux prédite avec un rappel de 81 %, mais une précision de 47 %, ce qui montre de nombreuses erreurs malgré une bonne identification des instances positives.
 - Classe 1 et 4 : Les plus problématiques avec des rappels très faibles (3 % pour la classe 1, 14 % pour la classe 4), indiquant une mauvaise différenciation.
 - Classe 2 et 3 : Performances moyennes, mais encore insuffisantes, avec un rappel respectif de 46 % et 27 %.



Autres données entraînées sur le clustering des usages

- Précision : 43.3 %
- Interprétation : Le modèle montre des difficultés significatives pour différencier plusieurs classes. Les résultats sont faibles pour toutes les classes sauf la classe 3, qui est la mieux prédite :
 - Classe 3 : La meilleure performance, avec un rappel de 71 % et un F1-score de 63 %, ce qui montre que le modèle la distingue mieux des autres.
 - Classe 0 : Performances moyennes avec un rappel de 59 %, mais une précision faible (51 %), indiquant une confusion importante avec d'autres classes.
 - Classes 1, 2, et 4 : Très mal prédites, avec des rappels inférieurs à 30 %.

