

INTRODUCTION

Dans ce projet de fin semestre notre objectif est d'appliquer les algorithmes d'apprentissage vus tout au long du semestre sur les données agrégées afin de mettre en évidence des résultats intéressants.

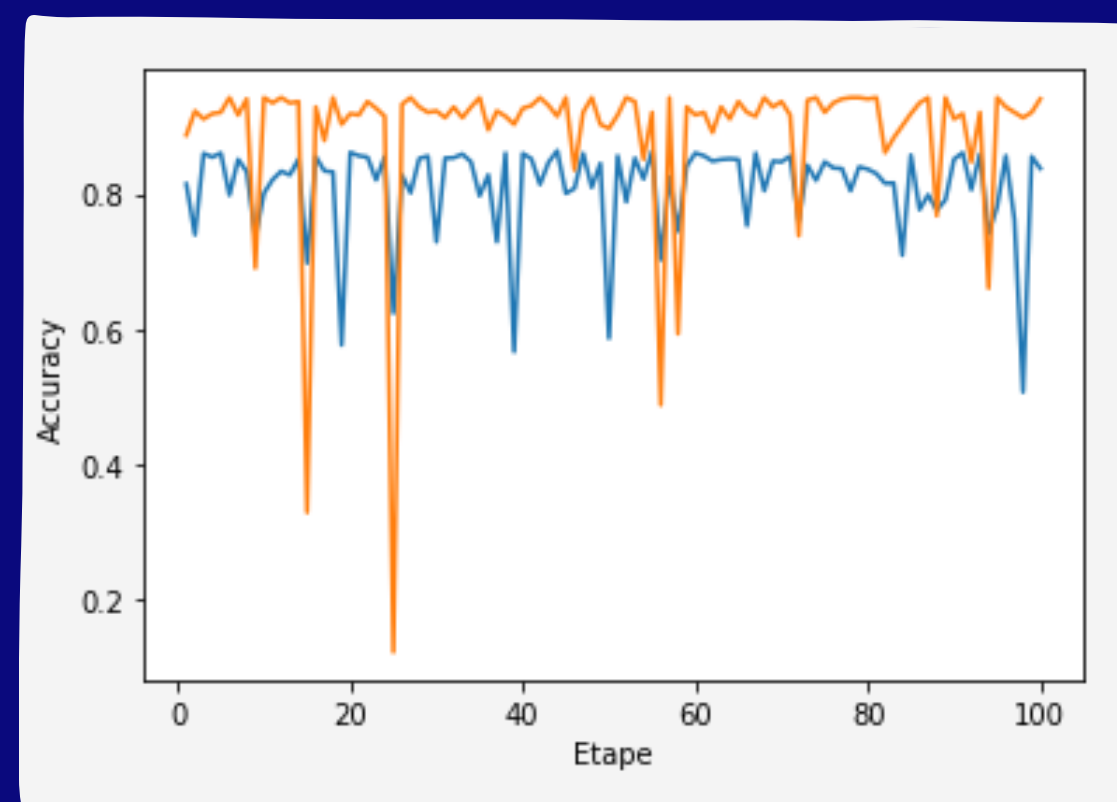
01. APPRENTISSAGE SUPERVISÉ

Classification des produits en fonction des données de leur cycle de vie pour prédire les aliments qui ont un plus grand impact sur le changement climatique, autrement dit éco-responsables.

01.1 METHODOLOGIE

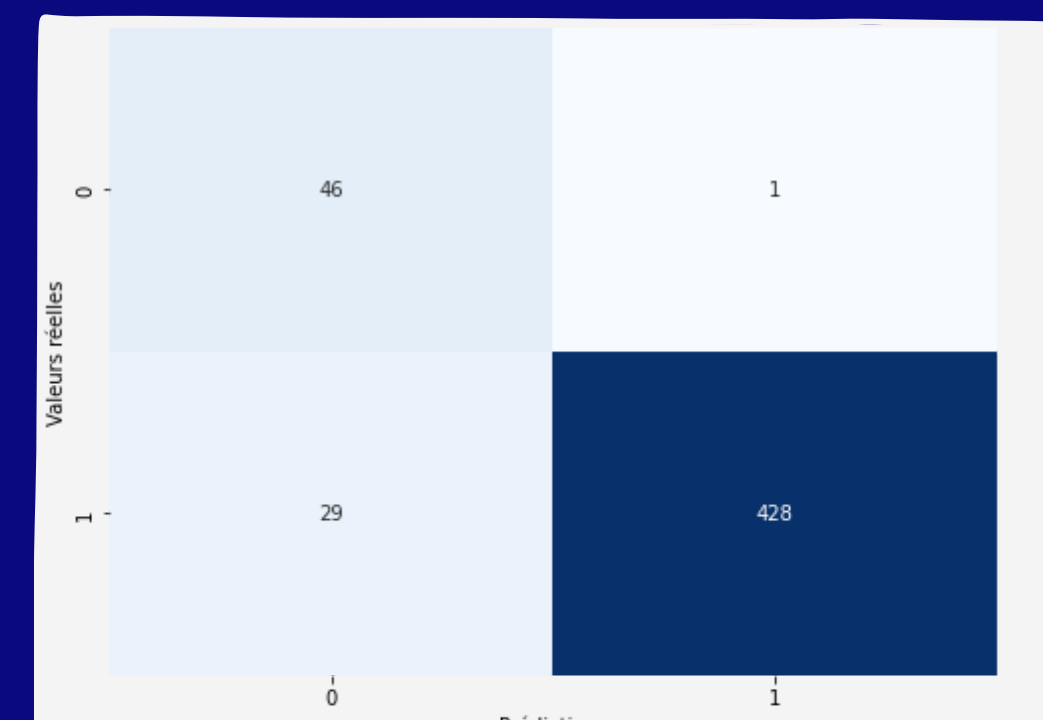
- Perceptron
- KNN
- Arbre de décision numérique

I- PERCEPTRON



On constate une performance d'environ 77%

Matrice de confusion

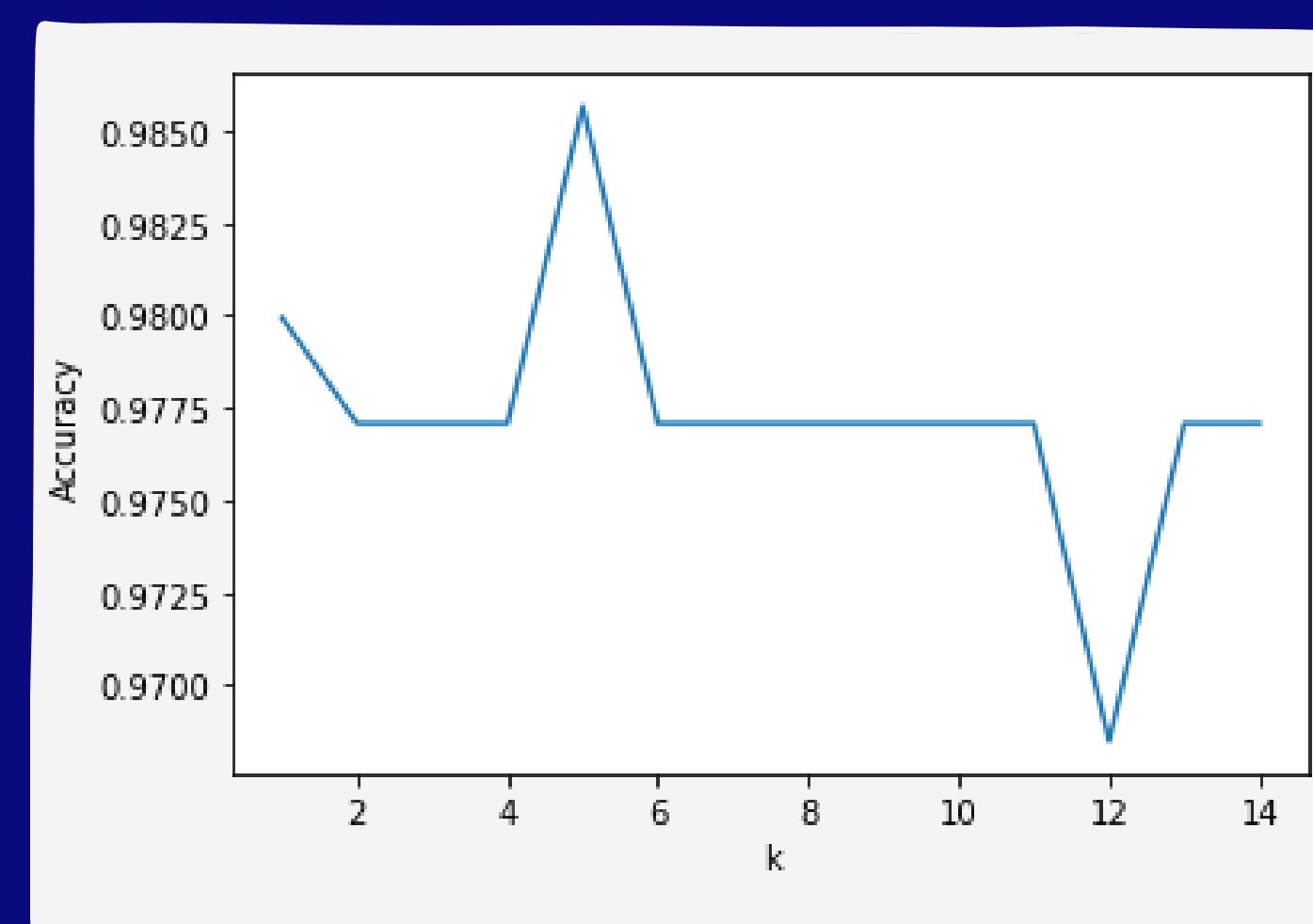


- Nous avons 14 faux-positifs et 309 vrai-positifs.
- Et 4 faux-négatif et 22 vrai-négatifs.

Résultats Perceptron

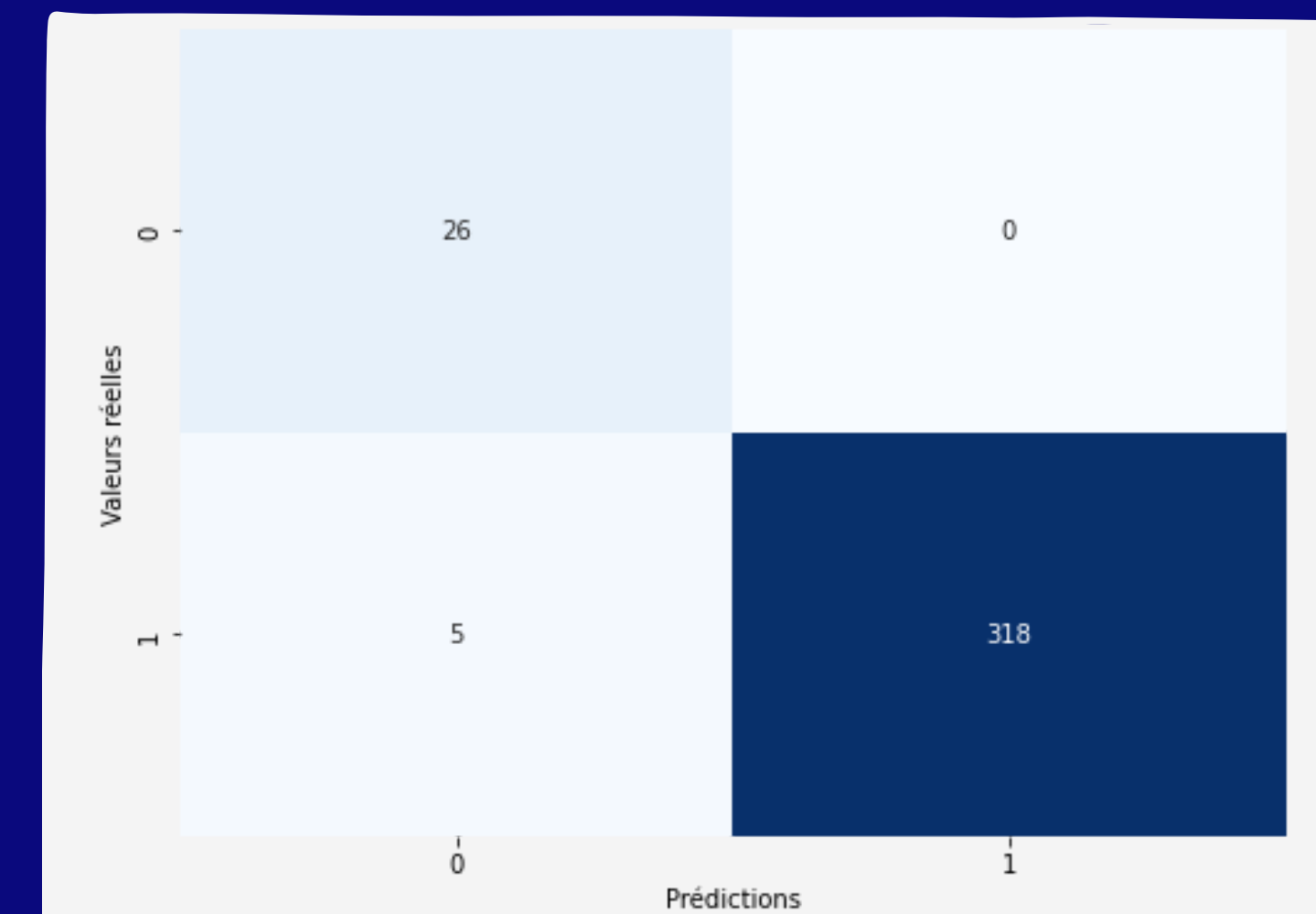
- validation croisée de 0.778
- AUC de 0.96
- Temps d'exécution sur 100 iterations de 32s

II-KNN



k_optimal = 5

Matrice de confusion



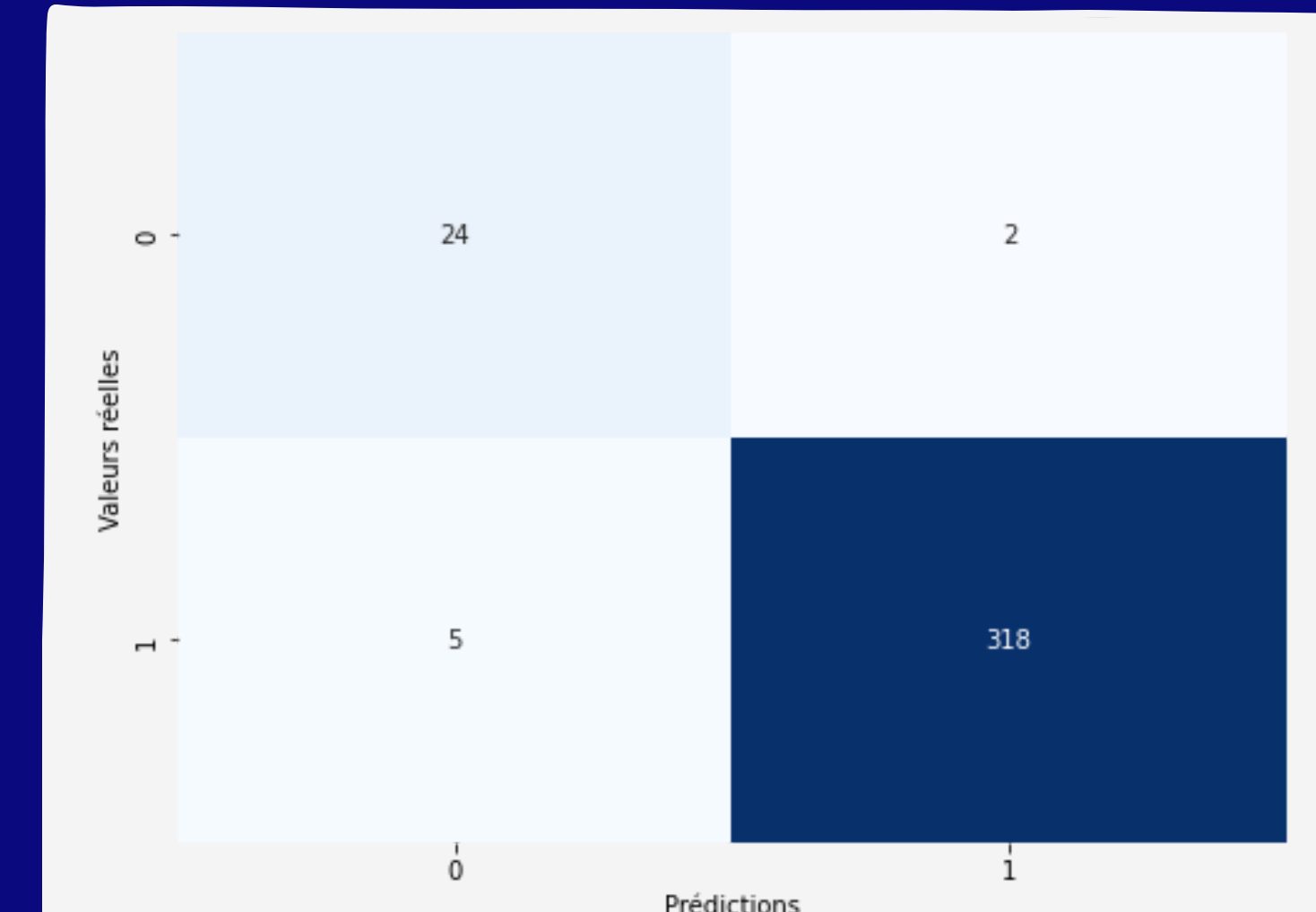
De meilleures prédictions que le modèle Perceptron

Résultats KNN

- validation croisée de 0.98
- AUC de 0.99
- Temps d'exécution sur 100*100 iterations de 47s

III-ARBRE DE DÉCISION NUMÉRIQUE

Matrice de confusion

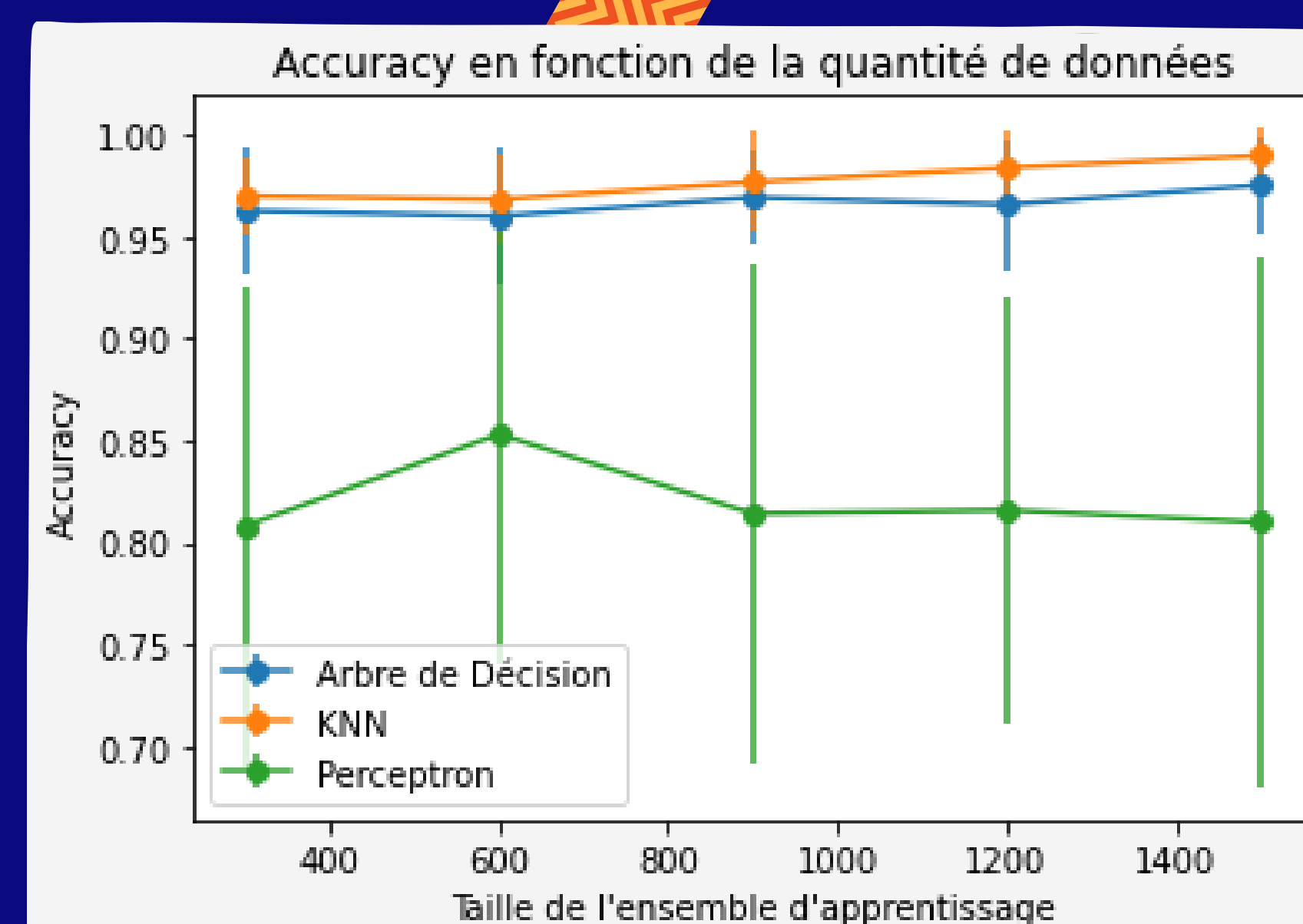


L'Arbre de Décision Numérique présente également une forte performance.

Résultats Arbre de Décision

- validation croisée de 0.98
- AUC de 0.99
- Temps d'exécution sur 100 iterations de 20s

01.2 COMPARAISON DES MODÈLES



02 APPRENTISSAGE NON SUPERVISÉ

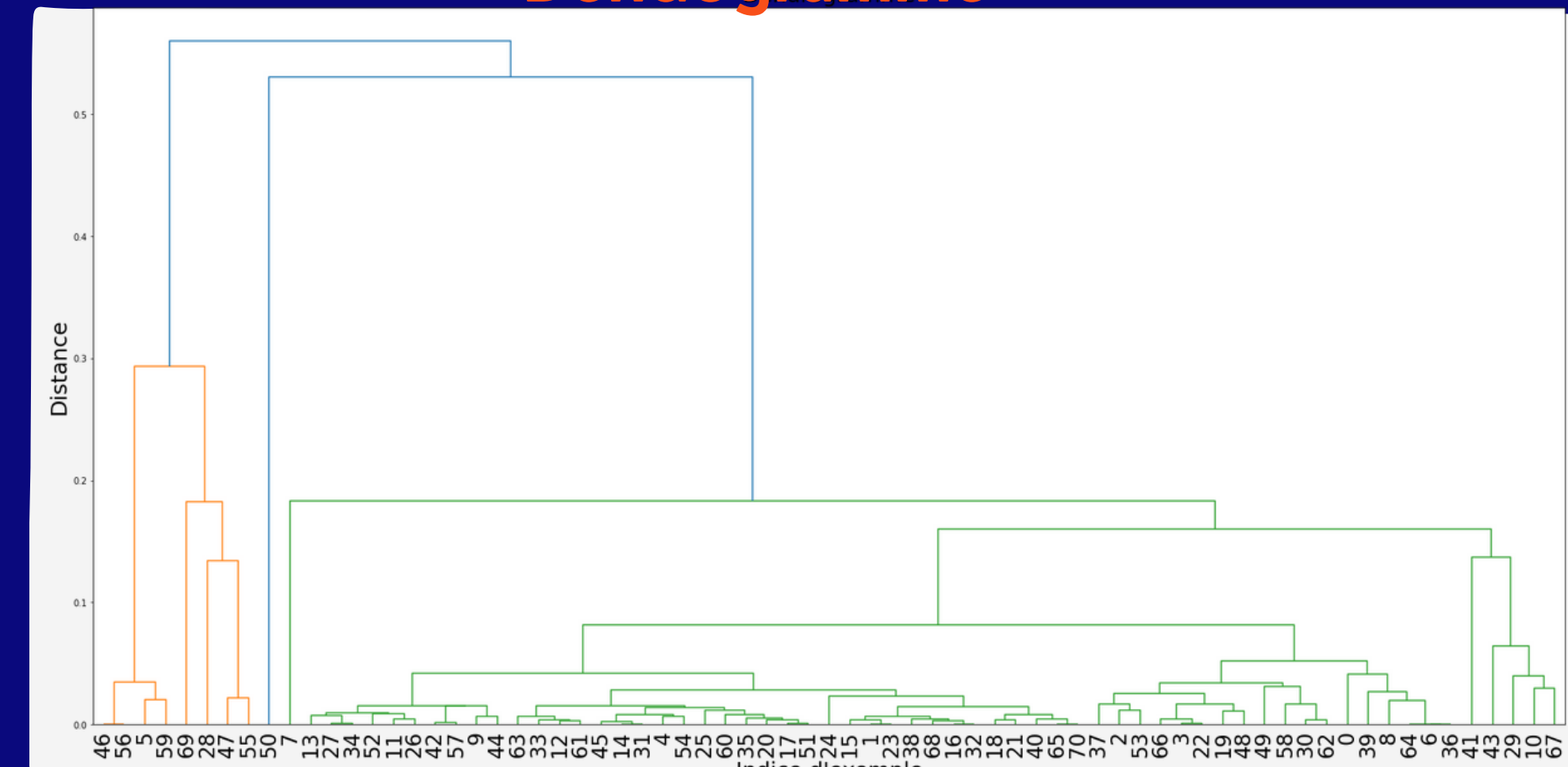
Classification des produits alimentaires en fonction de leurs impact sur la santé Humaines pour une application de régimes alimentaires.

02.1 METHODOLOGIE

- Clustering Hierarchique
- Agglomératif
- Kmeans

I-CLUSTERING HIERARCHIQUE AGGLOMÉRATIF

Dendrogramme



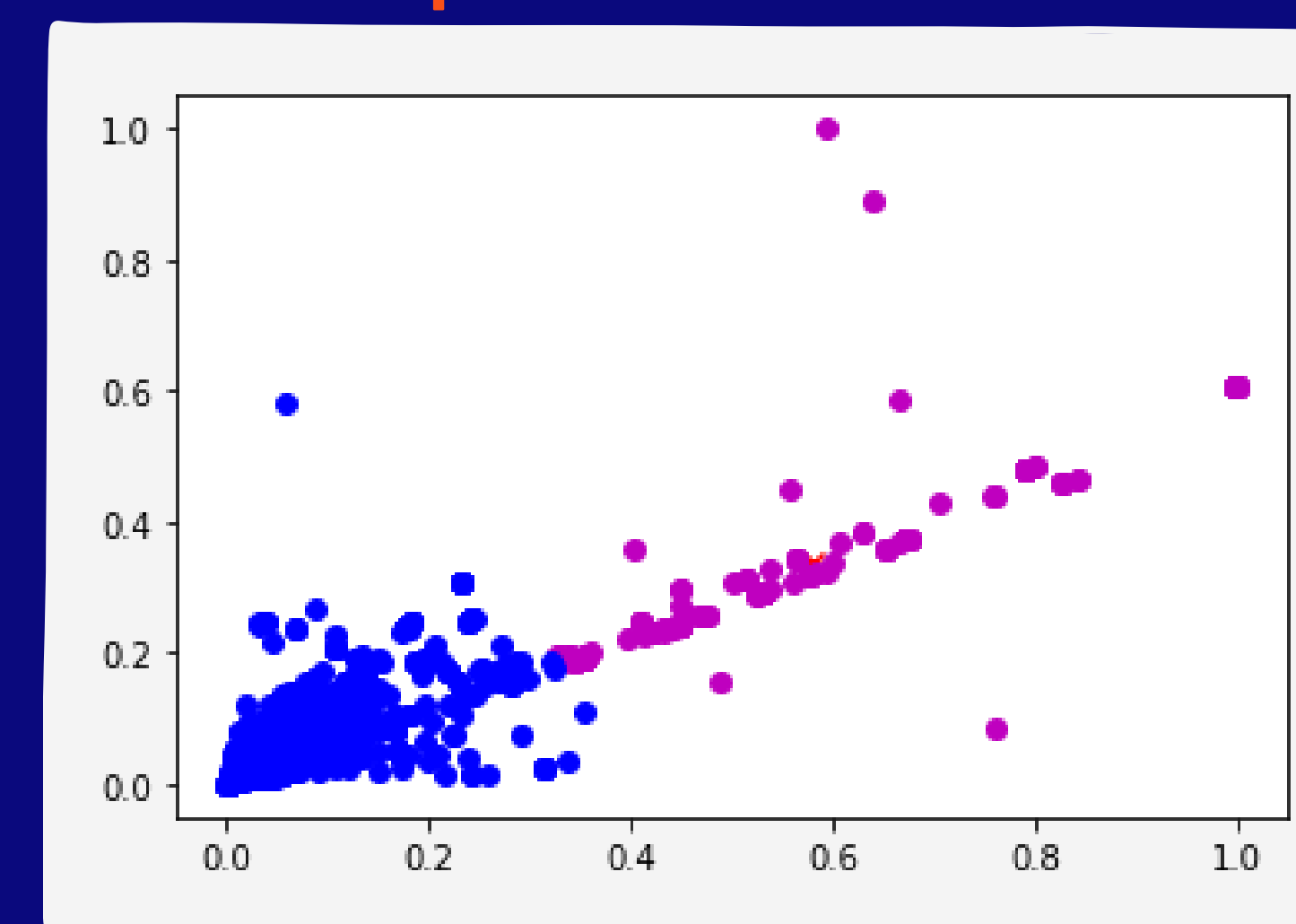
Le cluster le plus nocif : Sauce hollandaise préemballée, Sauce tomate à la viande ou Sauce bolognaise, ...

Résultats Clustering Hierarchique

L'algorithme de clustering hiérarchique a été utilisé pour regrouper les données en formant une hiérarchie de clusters. L'analyse des résultats révèle une structure hiérarchique des données, mettant en évidence les relations de similarité et de dissimilarité entre les différents groupes.

II-K-MEANS

Graphes résultant



Convergence de l'algorithme avec une inertie faible

Résultats Kmeans

- L'algorithme k-means a permis de regrouper les données en différents clusters en fonction de leurs caractéristiques similaires. L'analyse des résultats indique une distinction claire entre les clusters, soulignant ainsi la diversité et la séparation des données selon les critères spécifiés.
- Le cluster en Couleur Rose est plus nocif que le cluster Bleu

03. PROBLEME SUPPLÉMENTAIRE

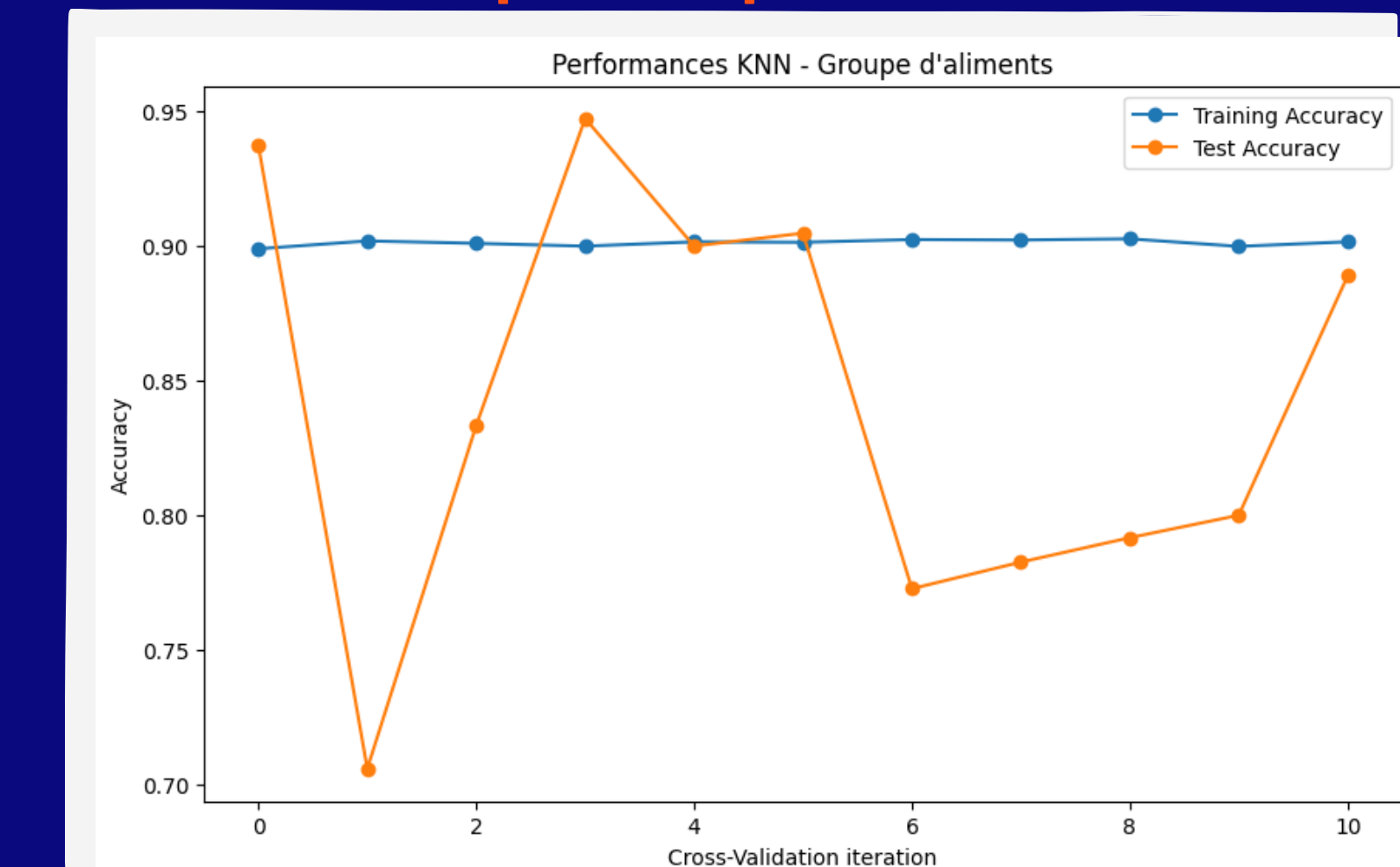
Peut-on prédire le groupe d'aliments des ingrédients à l'aide des données Data Ingrédient et Data étape?

03.1 METHODOLOGIE

- KNN MC
- K-Means

I-KNN MC

Graphes de performance



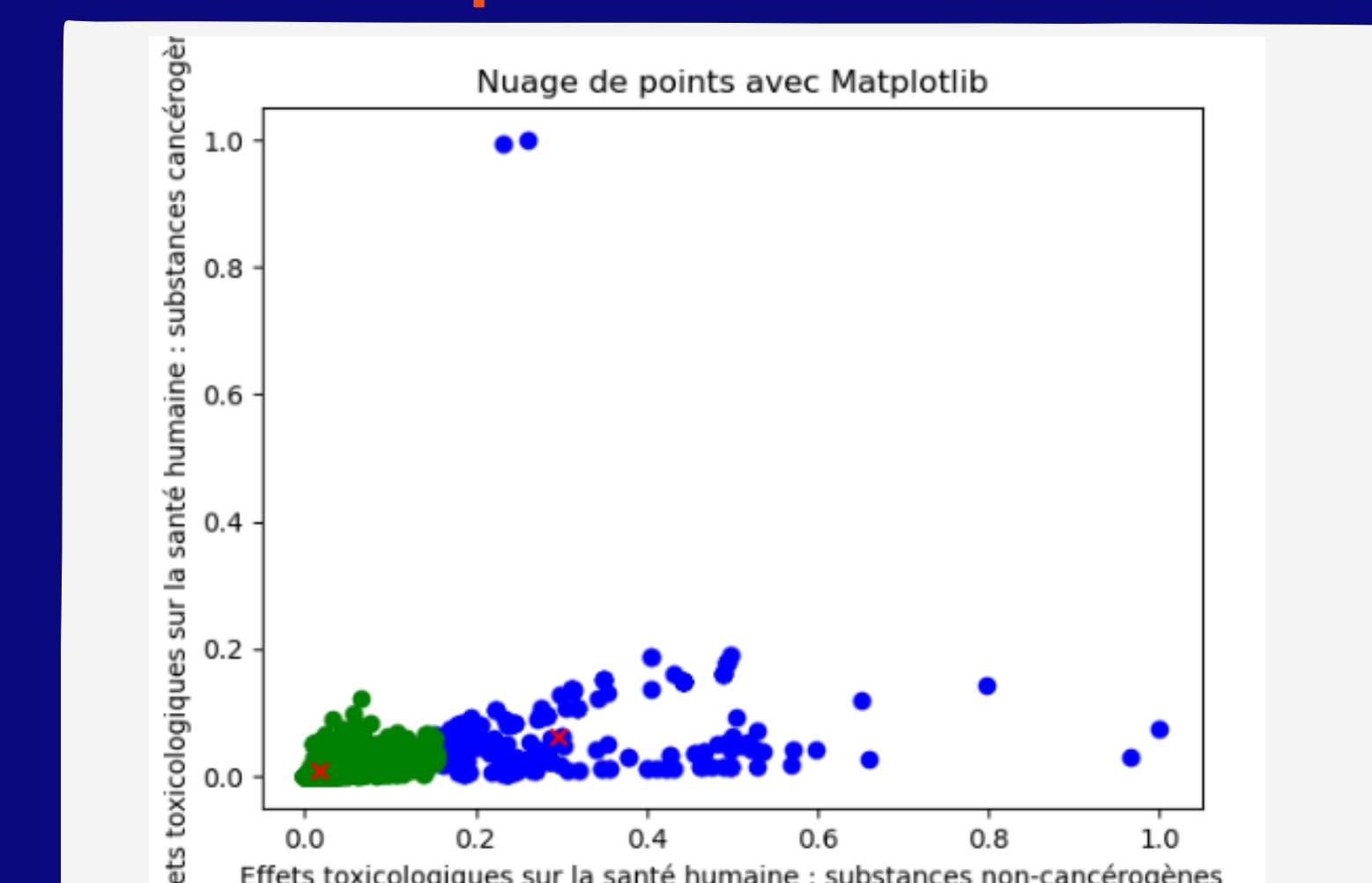
- Nous constatons un moyenne du train accuracy d'environ 0.90 et 0.80 pour le test
- on constate aussi une grosse instabilité dans les accuracy des tests.

Résultats KNN MC

- Moyenne d'accuracy test K optimal : 0.93
- Inefficace avec un grand nombre de dimensions
- Grande variance dans la précision du test

II- K-MEANS

Graphes résultant



- La visualisation devient difficile lorsqu'on a un nombre de clusters importants

Résultats Kmeans

- k means est inefficace pour gérer des classes de différentes tailles.
- une formation de 11 groupes. Les tailles des groupes varient de 74 à 549 éléments, avec un total de 2517 éléments