3I026 – Projet

Analyse de données Agribalyse

BESSAD Melina – KLICH NourElHouda

Introduction

Ce projet s'intéresse à l'analyse des données de la base Agribalyse représentant des données de référence sur les impacts environnementaux des produits agricoles et alimentaires. On s'intéresse à :

- La ***classification supervisée*** des aliments selon leur note de qualité (DQR)
- ***Catégorisation non-supervisée*** des aliments d'un seul groupe en utilisant des indicateurs environnementaux.

Classification supervisée

Classification binaire :

On a prédit la note de qualité (DQR) des aliments. Un aliment peut avoir une note entre 1 et 5. Si la note est inférieure ou égale à 3, l'aliment est de bonne qualité et sa classe est donc de 1, sinon il est de mauvaise qualité et sa classe est donc de -1. Nous avons utilisé les classifieur ***Perceptron*** et ***KNN*** avec validation croisée sur la base Synthèse.

Classifieur	Accuracy (test)
Perceptron	74 %
Knn (k=7)	81 %

Classification multi-classes :

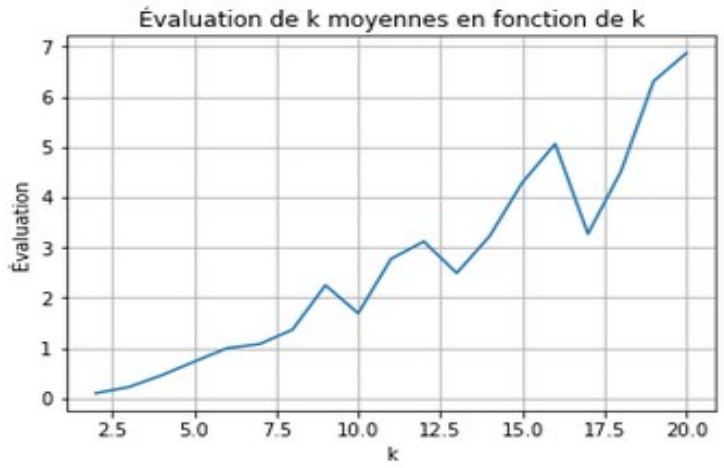
On a essayé de spécifier un peu plus la note de qualité de nos aliments. Pour cela nous avons prédit les 04 classes précisant le niveau de qualité des aliments allant de1 très bon à 5 mauvais en utilisant le classifieur ***arbre de décision numérique***.

Classifieur	Accuracy (test)
Arbre de décision numérique	92,69 %

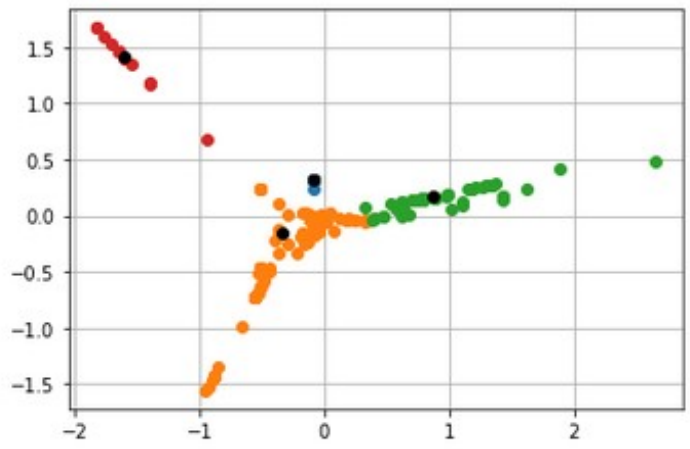
Nos accuracies étant très bonnes (>70%), on déduit donc que les indicateurs environnementaux ont bien un impact sur la qualité des aliments.

Catégorisation non- supervisée

On a décidé de catégoriser les aliments d'un seul groupe d'aliments qui est le groupe « *aides culinaires et ingrédients divers* » en utilisant la méthode des ***k-moyennes***. Pour ce faire, nous avons évalué plusieurs valeurs de k avec l'index de Dunn et avons avons choisi la meilleure qui est k = 4.



Avec une projection 2D , on obtient le graphe de classes ci-dessous pour k=4 (les centroïdes sont représentes en noir)



On a aussi catégorisé les aliments avec la méthode de ***clustering hiérarchique.***
Ci-dessous le dendrogramme résultant :