

Projet LU3IN026: Etude de la base de données Agribalyse 3.1

Problématique

- Supervisé: Prédire le groupe d'aliment auquel appartient un aliment selon ses caractéristiques
- Non supervisé: Regrouper les aliments du groupe 'viandes, œufs, poissons' par impact environnemental

Partie Supervisée

1. Création d'une base d'entraînement et d'une base de test
2. Evaluation des algorithmes avec la validation croisée en 10
3. Détermination des paramètres optimaux
4. Entraînement sur la base de test

KNN

aides culinaires et ingrédients divers	aliments infantiles	boissons	entrées et plats composés	fruits, légumes, légumineuses et oléagineux	glaces et sorbets	lait et produits laitiers	matières grasses	produits céréaliers	produits sucrés	viandes, œufs, poissons
0.961688	0.995455	0.966452	0.934416	0.935714	1.000000	0.961688	0.990260	0.931169	0.991558	0.959091
0.934609	0.992026	0.956938	0.872408	0.907496	0.998405	0.936204	0.980861	0.896332	0.971292	0.923445

KNN Multiclasse

Performance entraînement = 0.8674971687429218

Performance test = 0.8708946772366931

Perceptron

aides culinaires et ingrédients divers	aliments infantiles	boissons	entrées et plats composés	fruits, légumes, légumineuses et oléagineux	glaces et sorbets	lait et produits laitiers	matières grasses	produits céréaliers	produits sucrés	viandes, œufs, poissons
0.949351	0.992857	0.903226	0.822078	0.855844	1.000000	0.879221	0.983117	0.788312	0.981818	0.612338
0.901116	0.988836	0.909091	0.712919	0.830941	0.998405	0.306220	0.976077	0.835726	0.972887	0.765550

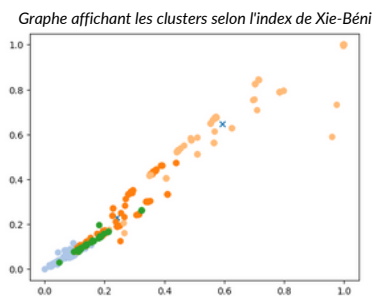
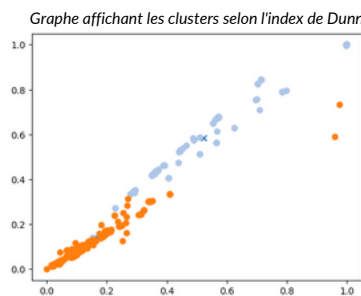
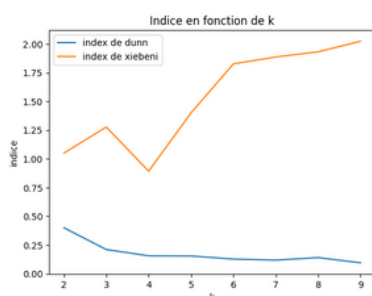
Arbre de décision

Performance moyenne = 0.618 - Ecart type: 0.176

Partie Non Supervisée

K-Moyennes

1. Pré-traitement des données
2. Détermination du nombre de clusters optimal avec l'indice de Dunn et Xie-Béni



Clustering hiérarchique

1. Création d'un jeu de données avec 50 exemples pris aléatoirement
2. Application du clustering hiérarchique simple