

Et si on essayait de manger écologique?

Introduction

Nous avons traité deux problématiques:

- 1)Quels sont les aliments les plus éco-responsables?
- 2)Peut-on prédire le groupe alimentaire d'un aliment?

data source

Nos données proviennent de la base Agribalyse. Nous avons exploité la sous base: Synthèse et nous avons gardé que les lignes qui ont un DQR<3(ie. les plus fiables)

Expérimentation

prob 2

Apprentissage Supervisé:

- ▶ KNN
- ▶ arbre de décision

prob 1

Apprentissage Non-Supervisé:

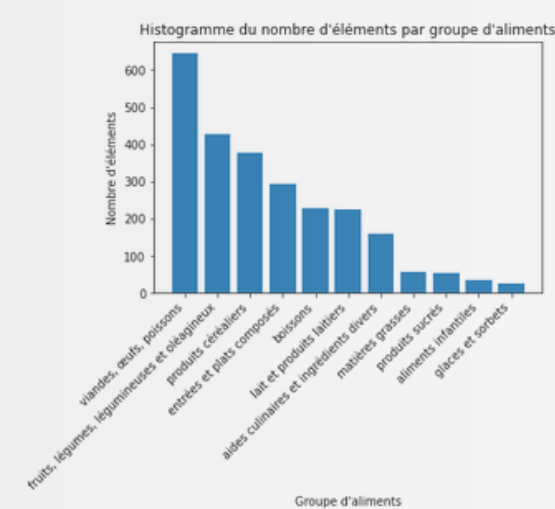
- ▶ K-mean
- ▶ CHA(centroid, average, complete, simple)

On ne considère que les colonnes (changement climatique, couche d'ozone, particules fines) pour leur fiabilité et robustesse.

- ▶ **KNN et arbre de décision** : description: [chang_clim, oz, part_f] label=sous-gr alimentaire (pour l'arbre on prend un échantillon)
- ▶ **K-mean et CHA**: description: [chang_clim, oz, part_f] (pour le CHA on prend un petit échantillon)

Résultats et Analyse

Partie 1 : Comprendre la base

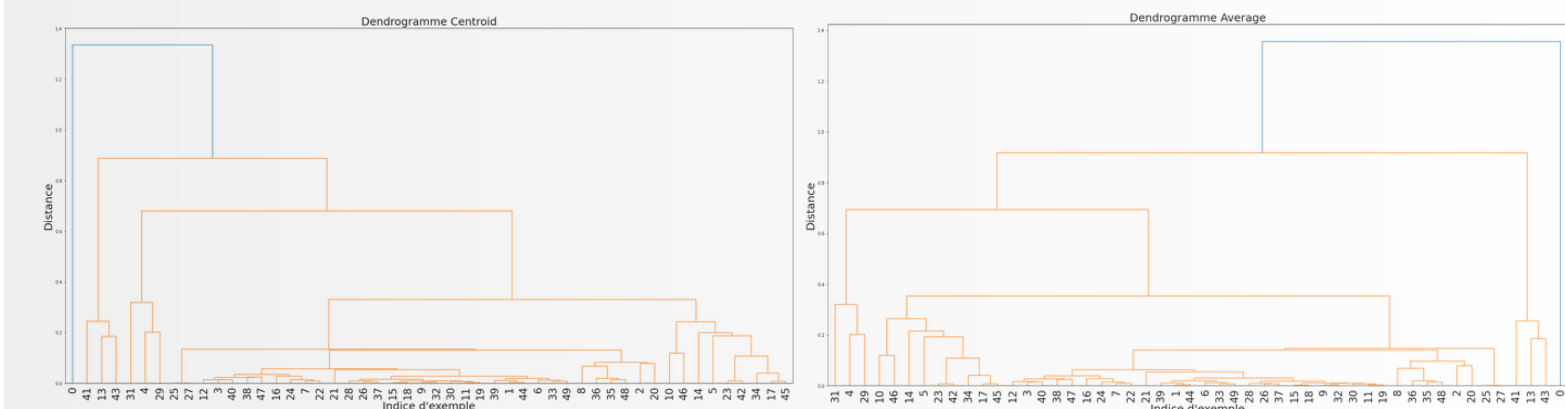


Les groupes alimentaires par ordre croissant de danger environnemental:

fruits, légumes, légumineuses et oléagineux - Toxicité moyenne : 0.06
entrées et plats composés - Toxicité moyenne : 0.14
boissons - Toxicité moyenne : 0.14
viandes, œufs, poissons - Toxicité moyenne : 0.22
aliments infantiles - Toxicité moyenne : 0.23
produits céréaliers - Toxicité moyenne : 0.25
aides culinaires et ingrédients divers - Toxicité moyenne : 0.25
lait et produits laitiers - Toxicité moyenne : 0.26
matières grasses - Toxicité moyenne : 0.28
produits sucrés - Toxicité moyenne : 0.50
glaces et sorbets - Toxicité moyenne : 0.50

Partie 2 : répondre aux problématiques

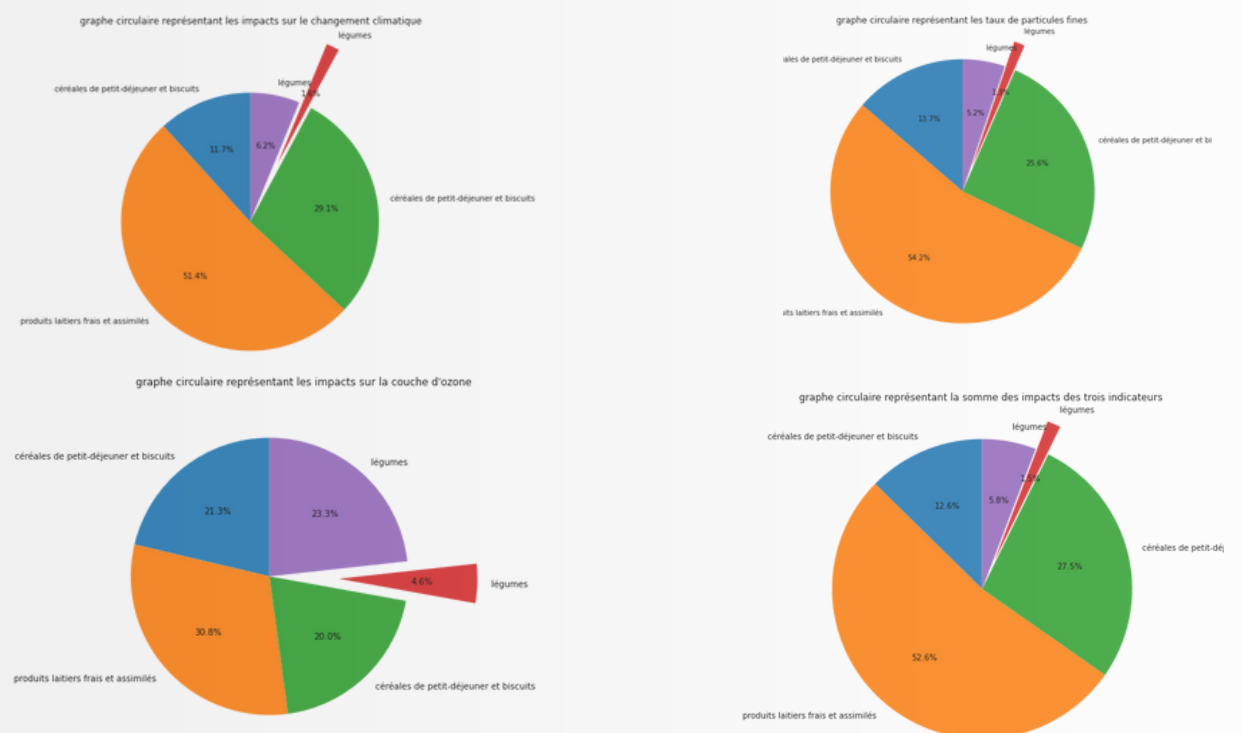
CHA sur un échantillon de 50 exemples



au seuil 0.5 il y a 3 groupes distincts:

le n° du groupe le plus écologique (moyenne d'impact minimale) est: 1
Il contient entre autres : 'Eau minérale Ardesy', 'Blé dur précuit' ...etc

Diagrammes circulaires



- ➔ Le cluster le plus écologique : Cluster 3 Taille : 913
Sous-groupe alimentaire majoritaire : – Nom : légumes – Pourcentage : 12.92%



CHA et K-mean permettent d'isoler le cluster d'aliments le plus éco-responsable



KNN

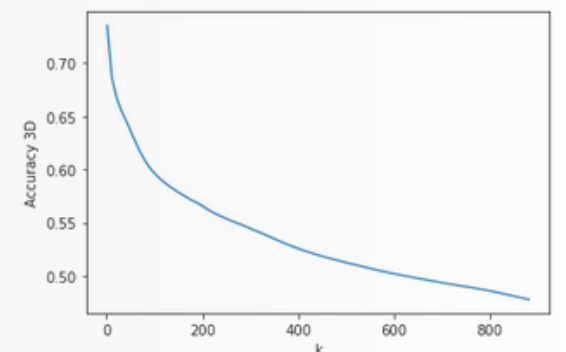
- ▶ Passage de labels catégorielles à numériques
- ➔ Aides culinaires et ingrédients divers ----> 0 Aliments infantiles ----> 1etc

On applique le KNN sur X et Y et on trouve le k-optimal avec la validation croisée

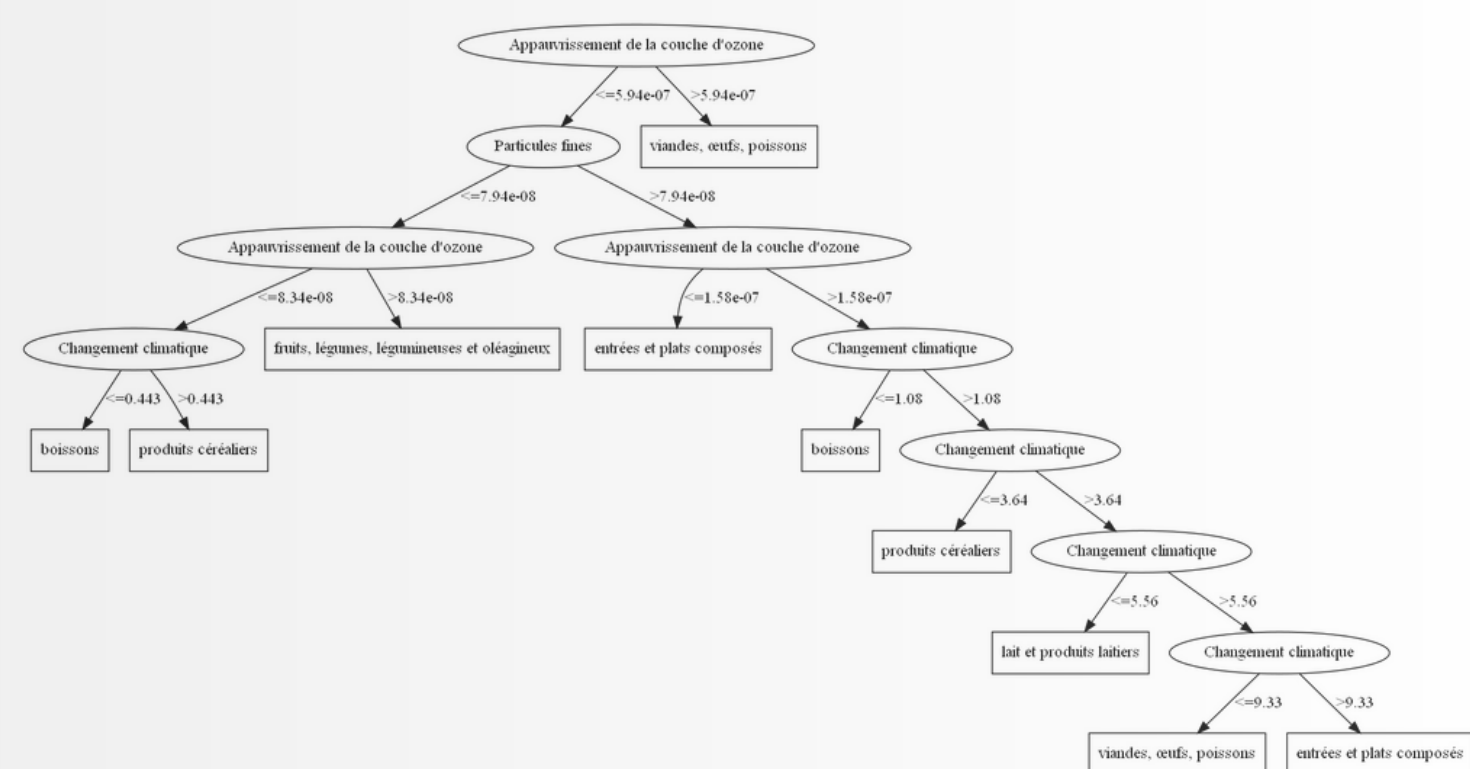
- ➔ L'accuracy maximale = 0.773
L'accuracy moyenne = 0.736

- ▶ Application sur 10 aliments

- ➔ Le taux de réussite est de 100%



L'arbre de décision sur un échantillon de 20 exemples



- ▶ Validation croisée

- ➔ L'accuracy maximale = 0.727
L'accuracy moyenne = 0.539



Les modèles d'apprentissage supervisés que nous avons utilisés permettent de prédire le groupe alimentaire d'un aliment connaissant les valeurs de ses indicateurs environnementaux.

Merci



K-mean:

- ▶ Recherche du meilleur k dans l'intervalle (2,20) en utilisant l'index de Dunn
- ➔ Meilleur k : 5