

L'impacte de sa consommation alimentaire sur l'environnement

Une analyse des données Agribalyse

Données Ingrédients: Ici, on se base sur 4 facteurs environnementaux pour essayer de prédire l'**Écotoxicité pour écosystèmes aquatiques d'eau douce**.

Données d'entrées :
Acidification terrestre et eaux douces
Eutrophisation eaux douces
Eutrophisation marine
Eutrophisation terrestre
Écotoxicité pour écosystèmes aquatiques d'eau douce

Ingrédients

Apprentissage Supervisé

Apprentissage non-supervisé

Pré-traitement

Pre-traitement des données :

- Correction des données (décalage des colonnes)
- Normalisation des données

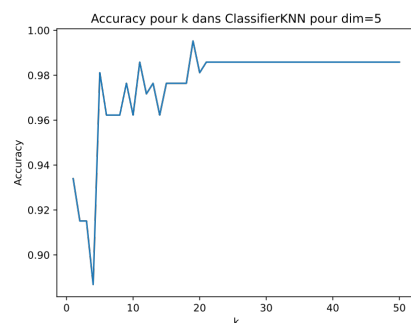
Attribution de label :

Nous regardons le médian de la colonne « Écotoxicité pour écosystèmes aquatiques d'eau douce. » douce pour y attribuer un label.

Écotoxicité pour écosystèmes aquatiques d'eau douce. = 1 : Si \leq médian
-1 : Si $>$ médian

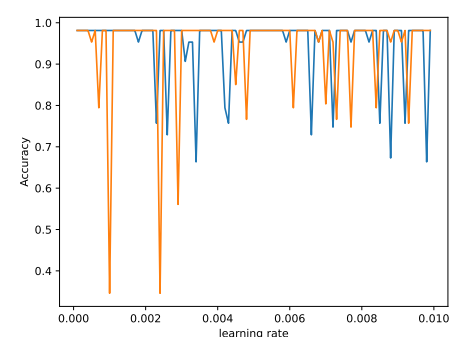
Méthodes

KNN



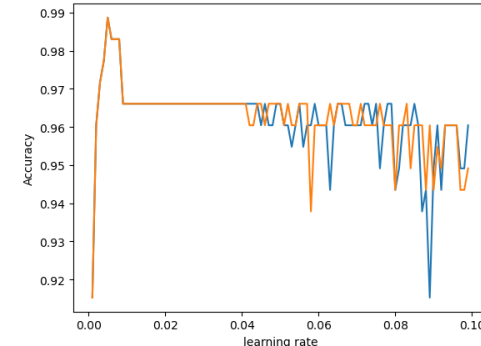
Meilleur k : 19
Accuracy moyenne : 0.989

Perceptron



Meilleur learning rate : 0.0001
Accuracy moyenne : 0.984

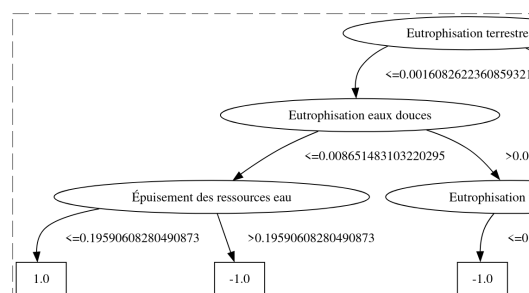
Perceptron Biases



Meilleur learning rate : 0.0051
Accuracy moyenne : 0.989

Arbres

Du au nombre d'éléments classes, il est impossible d'afficher l'arbre complet de décision. En voici un extrait :



Accuracy moyenne: 0.980

Évaluation

Des tests de validation croisée en 10 ont été réalisés sur chacune des 4 méthodes pour évaluer la performance de prédiction

Méthode	Accuracy
KNN	0.989
Perceptron	0.984
Perceptron Biases	0.989
Arbres	0.980

Clustering hiérarchique

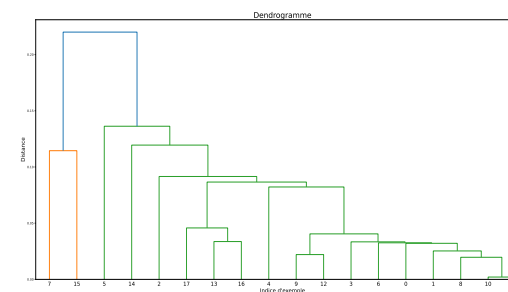
Pre-traitement des données :

- Normalisation des données
- GroupeBy sous-groupe alimentaire

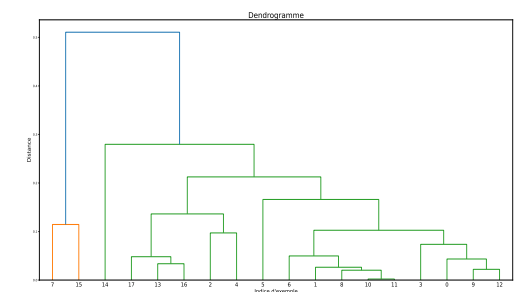
Sous-groupes uniques :

- aides culinaires et ingrédients divers, aides culinaires
- aides culinaires et ingrédients divers, algues
-
- viandes, œufs, poissons, œufs

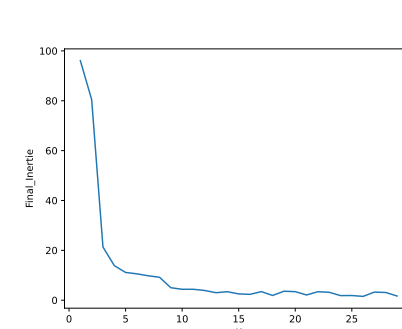
Linkage simple



Linkage complet

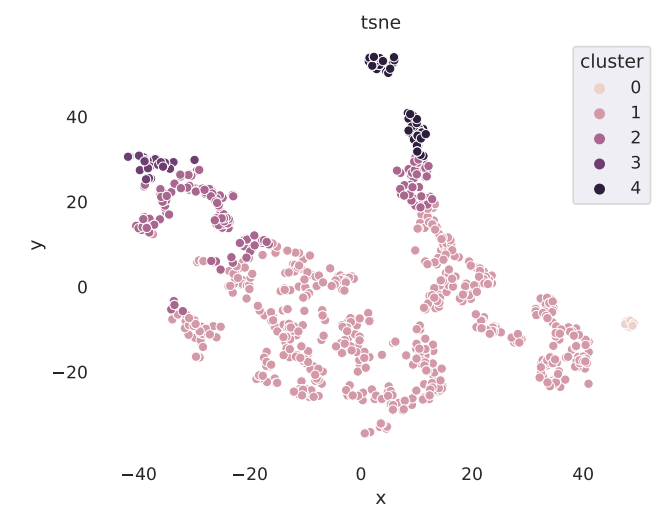


K-means



Pour trouver le meilleur k, on utilise la « Elbow method ». Ceci est la méthode standard utilisé dans l'analyse K-means.

Meilleur K trouve : **k=5**



On utilise la méthode TSNE pour faciliter l'affichage des clusters.