

L'impacte de sa consommation alimentaire sur l'environnement

Une analyse des données Agribalyse

Données étapes : On utilise la somme de l'impacte environnementale de pour chaque étape de production. On fini avec 6 dimensions.

Nom des colonnes d'entrées (sommés):

Agriculture
Transformation
Emballage
Transport
Supermarché et distribution
Consommation

Étapes

Apprentissage Supervisé

Pre-traitement des données :

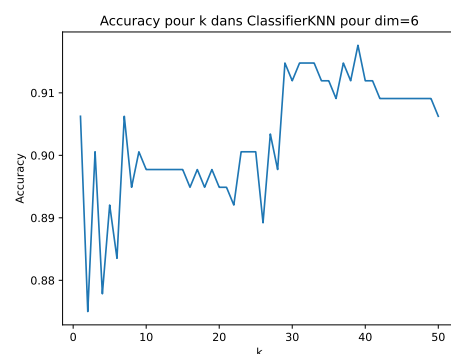
- Élimination des données de trop faible qualité selon le critère DQR (DQR < 3.0)
- Normalisation des données

Attribution de label :

Nous calculons un nouveau score pour chaque étape de la distribution des aliments a partir de la somme de ces étapes. Nous regardons ensuite le médian et attribuons dans une nouvelle colonne « Environment Classe »

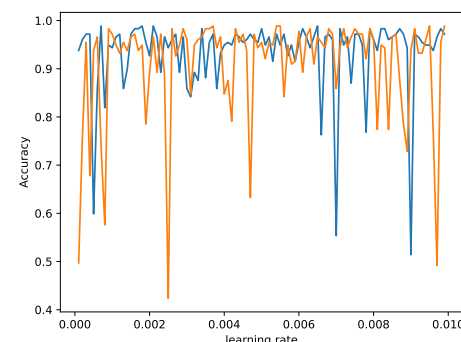
Environment Classe = 1 : Si > médian
-1 : Si ≤ médian

KNN



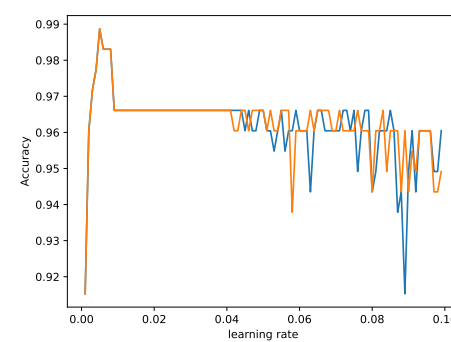
Meilleur k : 39
Accuracy moyenne : 0.933

Perceptron



Meilleur learning rate : 0.0037
Accuracy moyenne : 0.810

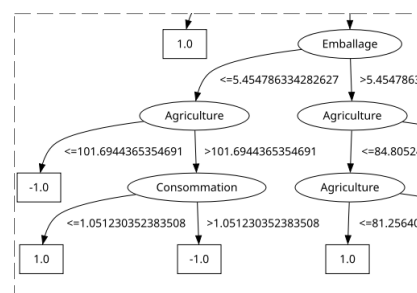
Perceptron Biases



Meilleur learning rate : 0.005
Accuracy moyenne : 0.832

Arbres

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



Accuracy moyenne : 0.930

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	0.933
Perceptron	0.810
Perceptron Biases	0.832
Arbres	0.930

Apprentissage non-supervisé

Clustering hiérarchique

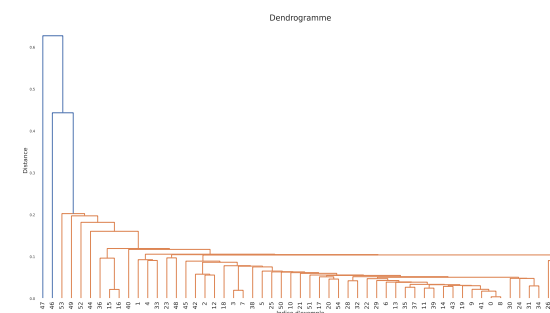
Pre-traitement des données :

- Normalisation des données
- GroupeBy sous-groupe alimentaire

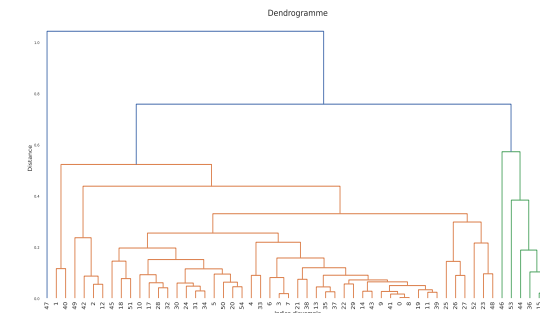
Sous-groupes uniques :

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

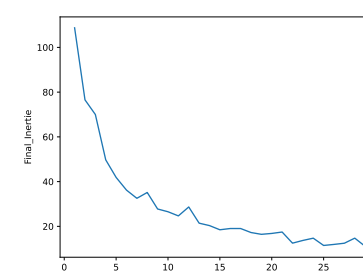
Linkage simple



Linkage complet

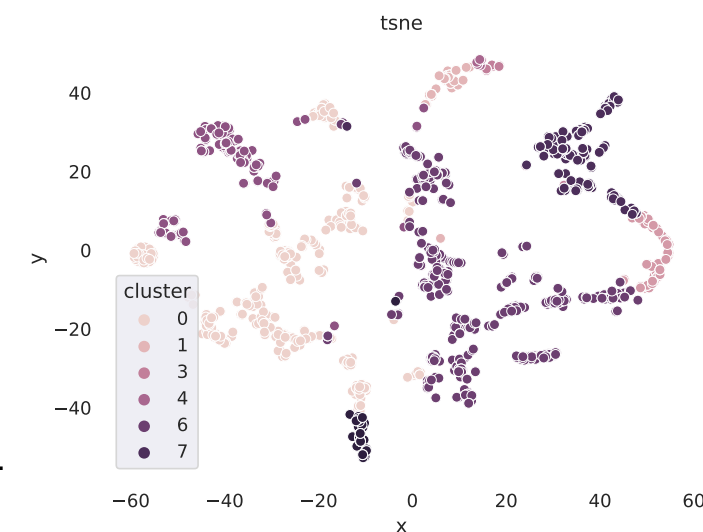


K-means



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=9**



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