

# L'impacte de sa consommation alimentaire sur l'environnement

Une analyse des données Agribalyse

Données synthèse : Ce poster explique en détail notre analyse des données synthèses Ici, toutes les colonnes de facteurs environnementaux sont utilise.

## Synthèse

### Pré-traitement

### Méthodes

## Apprentissage Supervisé

### Pre-traitement des données :

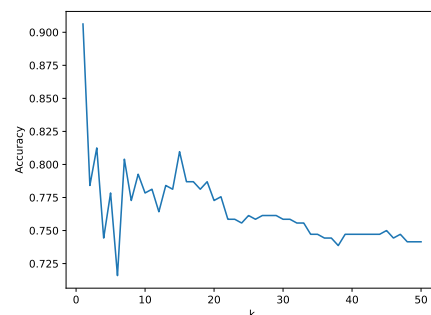
- Élimination des données de trop faible qualité selon le critère DQR (DQR < 3.0)
- Normalisation des données

### Attribution de label :

Nous calculons un nouveau score « Environnement Total » à partir de la somme des facteurs environnementaux. Nous regardons ensuite la médiane et attribuons dans une nouvelle colonne « Environment Classe »

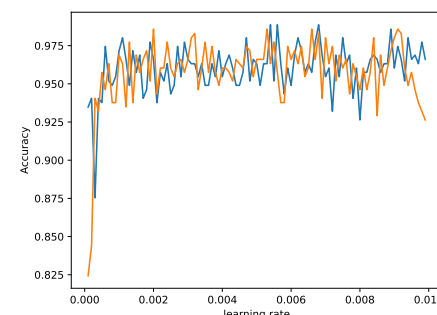
Environment Classe = 1 : Si > median  
-1 : Si ≤ median

### KNN



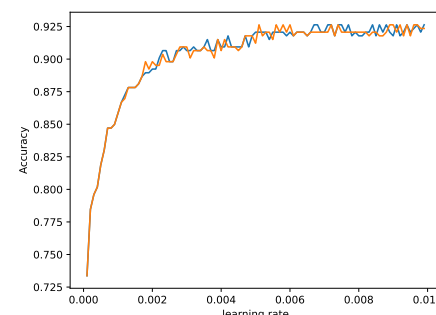
Meilleur k : 1  
Accuracy moyenne : 0.927

### Perceptron



Meilleur learning rate : 0.0007  
Accuracy moyenne : 0.979

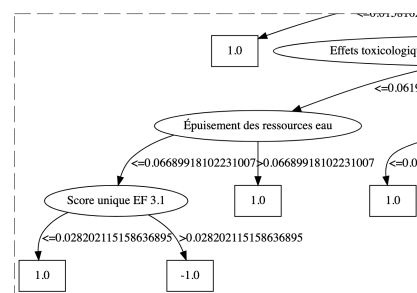
### Perceptron Biases



Meilleur learning rate : 0.0051  
Accuracy moyenne : 0.979

### Arbres

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



Accuracy moyenne : 0.981

## 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	0.927
Perceptron	0.979
Perceptron Biases	0.961
<b>Arbres</b>	<b>0.981</b>

## Apprentissage non-supervisé

### Clustering hiérarchique

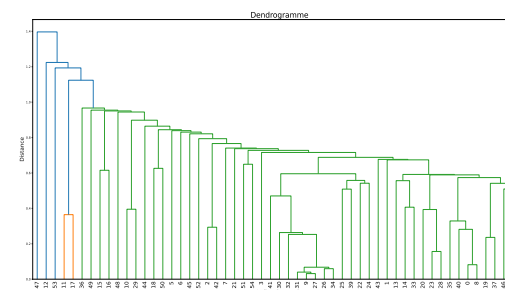
### Pre-traitement des données :

- Normalisation des données
- GroupeBy sous-groupe alimentaire

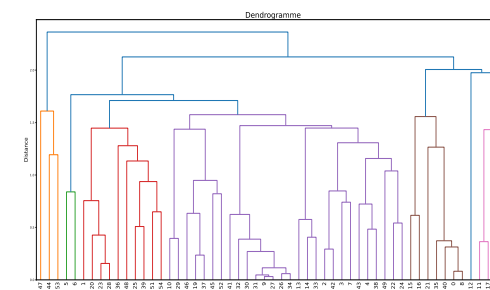
### Sous-groupes uniques :

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

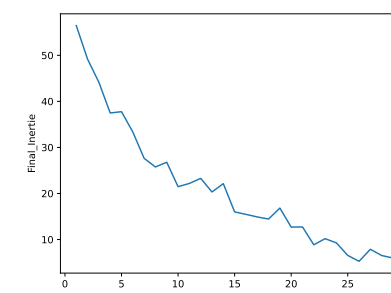
### Linkage simple



### Linkage complet

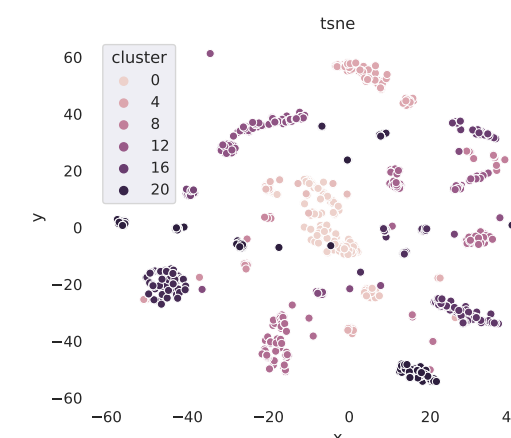


### K-means



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=22**



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