

## Projet Final : Etude de la base de données Agribalyse 3.1

### Problématiques

1. Quelle est la meilleure approche pour prédire la classe d'un aliment en relation avec le changement climatique ?
2. Comment peut-on utiliser le clustering non-supervisé pour identifier les profils environnementaux des produits alimentaires dans la base de données Agribalyse ?

### Prétraitement des données

#### **Agribalyse 3 - Synthèse (supervisé) :**

Les éléments dont le DQR est supérieur à 3 sont supprimés. On a retiré toutes les colonnes inutiles à la prédiction du changement climatique.

#### **Agribalyse 3 - Ingrédients (non-supervisé) :**

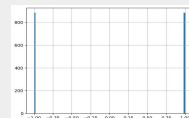
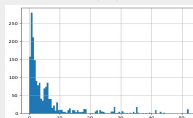
On a retiré toutes les colonnes de variables qualitatives. On a normalisé les valeurs pour faciliter l'utilisation de l'algorithme K-Means.

### Supervisé

Découpage des données en 3 bases, 60% pour la base d'apprentissage, 20% pour la base de test et de validation. Validation croisée sur la base de validation. Entraînement sur la base d'apprentissage. Détermination de l'accuracy d'un classifieur sur la base de tests.

Classes :

- supérieur à la médiane : 1 (haut facteur de changement climatique)
- inférieur ou égale à la médiane : -1 (facteur négligeable)



Algorithme	Accuracy
KNN	0.94
Perceptron	0.52
Perceptron Biases	0.5
Arbre de décision numérique	0.98

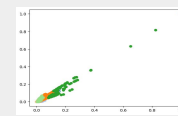
### Non-supervisé

Détermination du K optimal pour l'exécution de l'algorithme K-Means grâce à l'index de Dunn.

On trouve K = 5.

Application de l'algorithme K-Means pour former 5 clusters.

On affiche la répartition des catégories de chaque cluster.



### **Conclusion:**

Supervisé

Non-supervisé: