

Problématique : Quels sont les produits alimentaires les plus et les moins impactants sur l'environnement selon les indicateurs ACV (Analyse du Cycle de Vie) ?

Sous-questions :

- Quels sont les principaux facteurs qui influencent les impacts environnementaux des produits alimentaires ?
- Comment se répartissent les impacts environnementaux entre les différentes étapes du cycle de vie des produits alimentaires ?
- Quelles sont les différences d'impacts environnementaux entre les produits issus de l'agriculture conventionnelle et ceux issus de l'agriculture biologique ?
- Comment varient les impacts environnementaux des produits alimentaires selon les modes de consommation (quantité, fréquence) ?

Méthodes d'analyses :

1ère étape : réduction de dimensionnalité

- Analyse en composantes principales (ACP) : Nous allons réduire la dimensionnalité des données et identifier les axes de variation les plus importants, en utilisant des critères de sélection du nombre de composantes (variance expliquée, critère de Kaiser).
- T-SNE : Nous allons également visualiser les données multidimensionnelles dans un espace à deux dimensions, en ajustant le paramètre de perplexité pour obtenir une représentation optimale.

2ème étape : clustering

- K-means : Nous allons grouper les produits alimentaires en fonction de leurs profils environnementaux, et pour déterminer le nombre de cluster nécessaire on utilisera peut être un arbre de décision pour voir un peu les niveau de branches qu'on pourrait garder pour former des clusters (pas sûr que ça soit une très bonne idée mais on essayera quand même)
- Ensuite on utilise des techniques de visualisation de données (graphiques, cartes, tableaux de bord) pour comparer les indicateurs ACV des groupes de produits.

3ème étape : prédiction

- Analyse prédictive : On essayera de prédire ou de classer un impact environnemental en fonction de variables explicatives (type de culture, pratiques agricoles, etc.), en utilisant des techniques de validation croisée et de sélection de variables pour choisir le meilleur modèle parmi ceux-là (régression linéaire, régression logistique, SVM).

Traitement de données

Notre objectif est d'enrichir l'analyse environnementale des produits agricoles et alimentaires en croisant les données de Agribalyse avec d'autres sources de données pertinentes. Pour cela, nous allons procéder comme suit :

- Nous allons d'abord nettoyer les données en les normalisant et en imputant les valeurs manquantes ou aberrantes.
- Nous allons ensuite intégrer les données de Agribalyse avec les autres sources de données, en utilisant des identifiants communs ou des noms d'aliments pour les relier entre elles, et en résolvant les éventuels problèmes de compatibilité et de cohérence des données.

➔ Voici le plan que nous envisageons de suivre, sous réserve que vous le validiez bien entendu 😊. Il est probable que des ajustements soient nécessaires en cours de route, avec des analyses supplémentaires ou moins selon les données disponibles, mais le concept général restera le même.