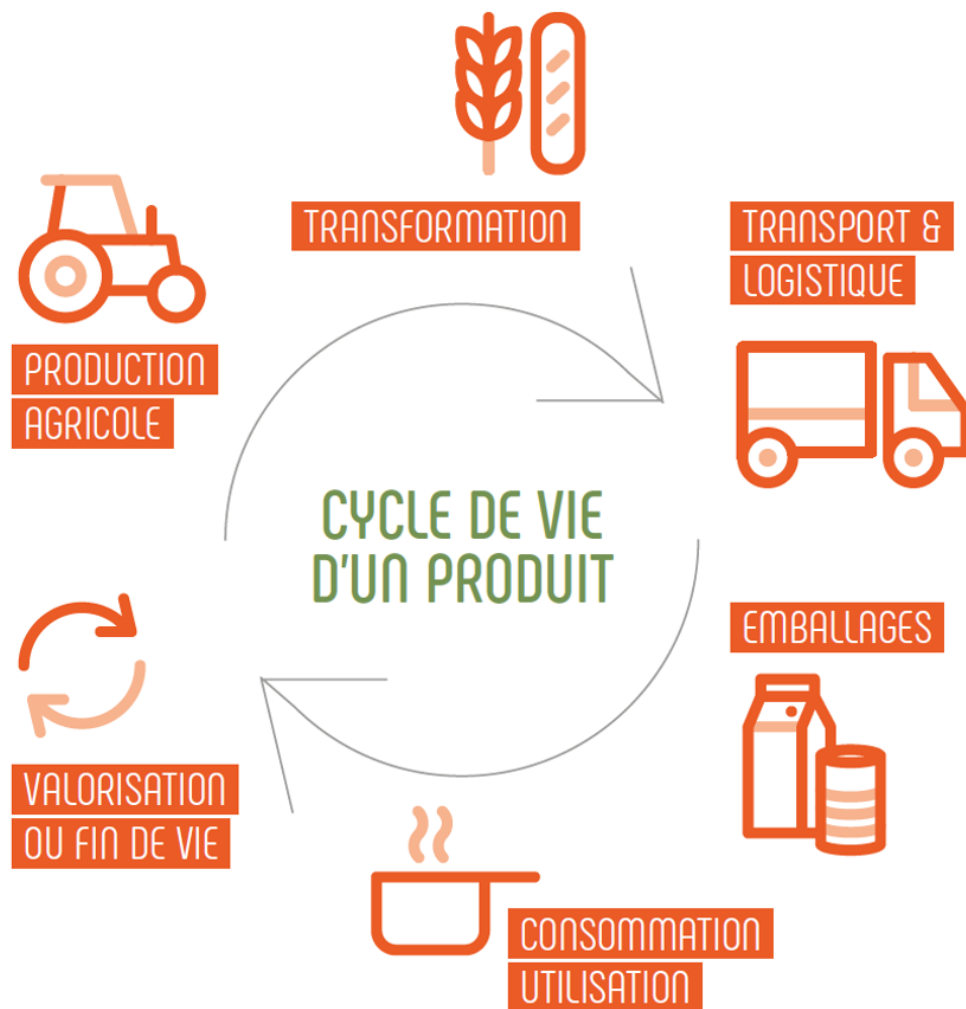


# Rapport DALAS

Saliou BARRY  
Zhile ZHANG

Date

Soutenance orale  
le 24 mai 2024



# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Collecte des données</b>	<b>3</b>
2.1	Méthodologie de Web Scraping . . . . .	3
2.2	Description des données . . . . .	4
<b>3</b>	<b>Pretraitement des données</b>	<b>4</b>
<b>4</b>	<b>Exploration des données</b>	<b>5</b>
4.1	Distribution du score EF . . . . .	5
4.2	Clustering par impact . . . . .	6
4.3	Correlation et Reduction de dimensionnalité . . . . .	9
4.4	No IDEA FOR NOW . . . . .	11
4.4.1	Fusion des données . . . . .	13
4.4.2	ACV . . . . .	13
4.4.3	viandes cuites . . . . .	13
4.5	Viandes Crues . . . . .	15
<b>5</b>	<b>Implémentation des Modèles pour la Prédiction de l'Impact Environnemental</b>	<b>16</b>
5.1	Préparation des Données . . . . .	16
5.2	Modèles Implémentés . . . . .	16
5.2.1	Régression Logistique . . . . .	16
5.2.2	Support Vector Machine (SVM) . . . . .	17
5.2.3	Random Forest (RF) . . . . .	18
5.2.4	Decision Tree (DT) . . . . .	19
5.2.5	K-Nearest Neighbors (KNN) . . . . .	20
<b>6</b>	<b>Test de Significativité entre les Modèles</b>	<b>22</b>
6.0.1	Résultats du Test de Significativité . . . . .	22
<b>7</b>	<b>Conclusion</b>	<b>23</b>

# 1 Introduction

L'impact environnemental de notre alimentation est devenu un enjeu majeur dans la lutte contre le changement climatique et la préservation de notre planète. Les choix que nous faisons à chaque repas ont des répercussions sur les ressources naturelles, les émissions de gaz à effet de serre et la biodiversité. Dans ce contexte, l'analyse des données agribalyse de l'ADEME (Agence de la Transition Écologique) offre une opportunité précieuse pour évaluer et comprendre l'empreinte écologique de nos habitudes alimentaires.

Ce rapport vise à déterminer l'impact spécifique de différents aliments sur l'environnement en utilisant des données scientifiques rigoureuses. Nous examinerons les différentes étapes du cycle de vie des produits alimentaires, de la production à la consommation, en passant par la transformation, l'emballage et le transport. Grâce à cette analyse, nous espérons identifier les domaines clés où des mesures d'atténuation peuvent être prises pour réduire notre empreinte écologique.

## 2 Collecte des données

Dans notre quête pour comprendre l'impact environnemental de notre alimentation, nous nous sommes tournés vers le jeu de données Agribalyse fourni par l'Agence de la Transition Écologique (ADEME) en France. Agribalyse propose une collection complète de données d'analyse du cycle de vie (ACV) pour divers produits alimentaires, nous permettant ainsi d'explorer la relation complexe entre nos choix alimentaires et leurs conséquences écologiques.

### 2.1 Méthodologie de Web Scraping

Pour rassembler ce précieux ensemble de données, nous avons utilisé des techniques de web scraping. Voici comment nous avons procédé :

#### 1. Exploration Initiale :

- Nous avons accédé à l'API Agribalyse via l'URL suivante : <https://data.ademe.fr/data-fair>
- Le nom du jeu de données a été dynamiquement défini en fonction de notre domaine de recherche (détails des ingrédients, détails des étapes ou synthèse).

#### 2. En-têtes et Paramètres :

- Nous avons configuré les en-têtes appropriés (y compris un user-agent) pour simuler une requête de navigateur web.
- Les paramètres tels que `size`, `page`, `q_mode` et `inializedAt` ont été ajustés pour récupérer les données de manière efficace.

#### 3. Récupération des Données :

- Nous avons itéré à travers les résultats paginés, récupérant les informations pour chaque ligne.

- Le schéma a fourni les noms de colonnes nécessaires et leurs descriptions.

#### 4. Stockage des Données :

- Nous avons enregistré les données récupérées dans un fichier CSV, en veillant à ce que les noms de colonnes soient correctement représentés.
- Chaque ligne correspondait à un produit alimentaire spécifique ou à une étape de son cycle de vie.

## 2.2 Description des données

Nous avons récupéré les données en utilisant le scraping sur trois tableaux distincts : les données de synthèse, les données par étape, et les données par ingrédient.

- **Données de Synthèse** : Fournissent une vue d'ensemble de l'impact environnemental global de chaque produit alimentaire, incluant des indicateurs tels que les émissions de gaz à effet de serre, la consommation d'eau et l'utilisation des sols. Les dimensions de ces données sont (2500, 32).
- **Données par Étape** : Détaillent le cycle de vie des produits alimentaires, mettant en lumière l'impact environnemental à chaque étape du processus, de la production à la consommation. Les dimensions de ces données sont (2500, 116).
- **Données par Ingrédient** : Fournissent des informations spécifiques sur l'impact écologique de chaque composant des produits alimentaires, permettant une analyse approfondie des choix d'ingrédients pour des recettes plus durables. Les dimensions de ces données sont (8840, 27).

## 3 Prétraitement des données

Avant d'entamer l'analyse des données environnementales d'Agribalyse, nous avons procédé à un prétraitement afin d'assurer la qualité et la fiabilité de nos résultats. Un aspect crucial de ce prétraitement est la sélection des données en fonction du **Data Quality Ratio (DQR)**. Le **DQR** est une mesure de qualité attribuée à chaque produit agricole et alimentaire dans Agribalyse, notée de 1 (très bon) à 5 (très mauvais). Voici quelques points importants à retenir sur le DQR :

- La **Commission Européenne** recommande la prudence lors de l'utilisation de données avec un DQR supérieur à 3.
- Dans la base de données Agribalyse, environ **67 %** des données ont un DQR jugé bon ou très bon (de 1 à 3).

Voici les étapes spécifiques que nous avons suivies pour prétraiter les données Agribalyse:

### 1. Suppression des Doublons :

- Nous avons éliminé les doublons dans les données de synthèse (**df\_synt**) et les données par étape (**df\_etape**), en nous basant sur la colonne "Code CIQUAL".
- Cette étape garantit que chaque produit est représenté de manière unique dans notre analyse.

### 2. Remplissage des Valeurs Manquantes :

- Nous avons identifié les colonnes numériques (de type float64) dans les deux ensembles de données.
- Pour chaque colonne, nous avons rempli les valeurs manquantes avec la moyenne des valeurs existantes, cela a permis d'éviter les biais liés aux données manquantes.

### 3. Filtrage par DQR :

- Nous avons conservé uniquement les lignes où le DQR est inférieur à 4.
- Cette décision est en accord avec la recommandation précédente et assure que nous utilisons des données de qualité pour nos analyses.

## 4 Exploration des données

### 4.1 Distribution du score EF

Le score EF est un indicateur environnemental, qui évalue l'impact écologique d'un produit agricole ou alimentaire. Il est obtenu selon la méthodologie PEF (Product Environmental Footprint) par l'agrégation des 16 indicateurs d'impact. Plus ce score est proche de zéro, plus l'empreinte environnementale du produit est faible.

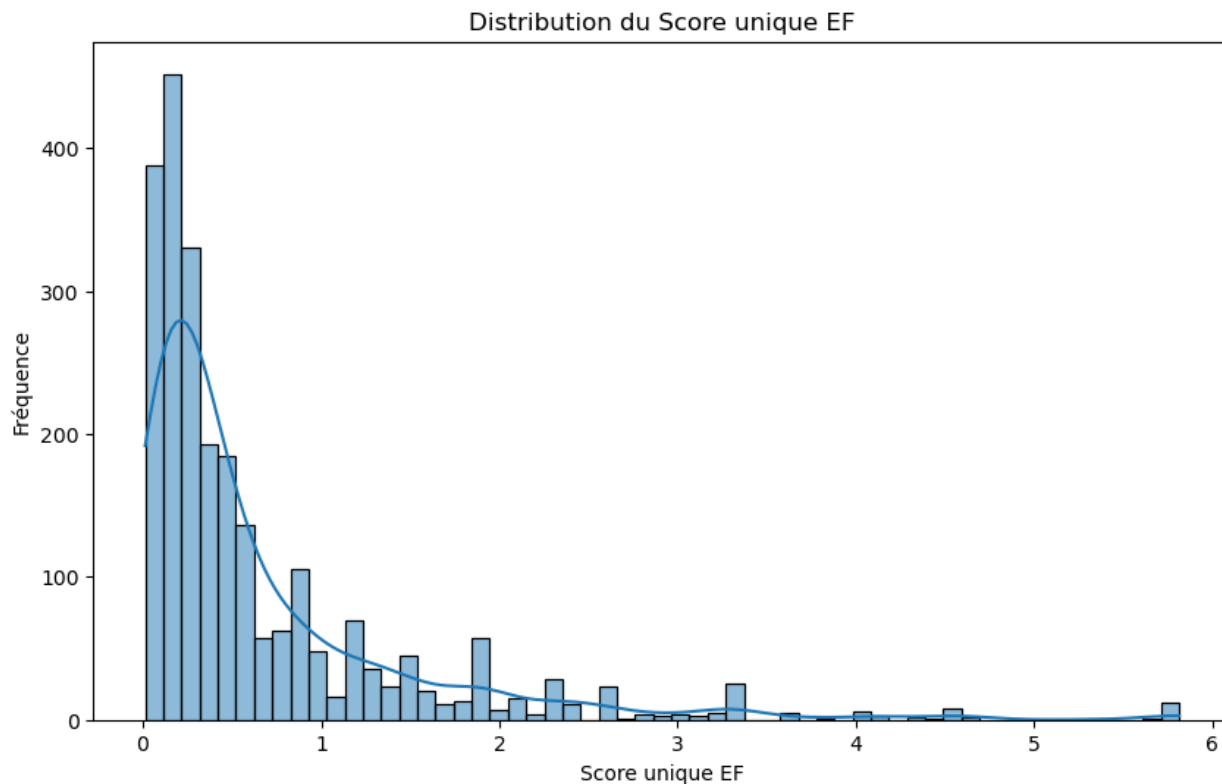


Figure 1: Distribution du Score unique EF

Cet histogramme représente la distribution du score unique EF parmi les produits alimentaires, en fonction de leur fréquence dans notre ensemble de données synthèse. Nous observons que les produits alimentaires ayant un faible impact sur l’environnement sont très nombreux par rapport à ceux qui ont un impact plus élevé.

## 4.2 Clustering par impact

Afin de mieux cibler les aliments ayant le plus grand impact environnemental et ainsi déterminer des pistes d’amélioration pour la préservation de notre environnement, nous avons entrepris de les catégoriser par clusters. Pour ce faire, nous avons opté pour l’algorithme K-Means, en utilisant les 16 indicateurs des données de synthèse.

Pour déterminer le nombre optimal de clusters, nous avons utilisé la méthode du coude (inertie). Nous avons tracé l’inertie en fonction du nombre de clusters et choisi le point où l’inertie commence à diminuer de manière moins significative.

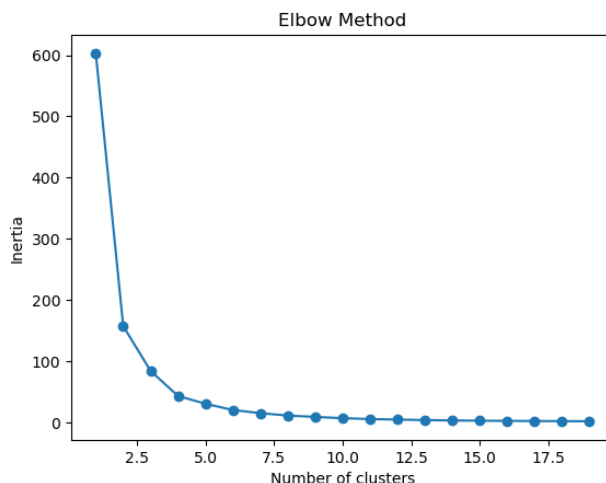


Figure 2: Elbow Methode

Après avoir analysé l’inertie, nous avons déterminé que 5 *clusters* étaient appropriés pour regrouper les aliments en fonction de leurs caractéristiques environnementales. Ainsi, nous avons appliqué l’algorithme **Kmeans** avec 5 clusters. Pour visualiser ces clusters dans un espace de dimension réduite, nous avons un **t-SNE**. Le graphe ci-dessous illustre la visualisation des clusters obtenus. . Ensuite, nous avons calculé la moyenne de l’impact

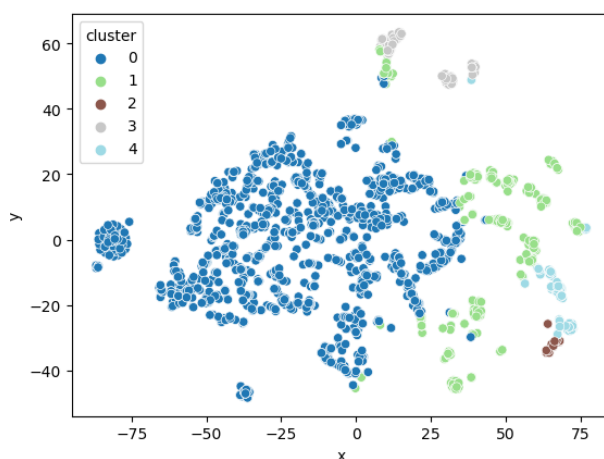


Figure 3: T-SNE

environnemental de chaque cluster en utilisant la colonne ”Score unique EF”. Cette analyse nous a permis de comprendre les variations d’impact environnemental entre les différents clusters, comme le montre le graphe ci-dessous.

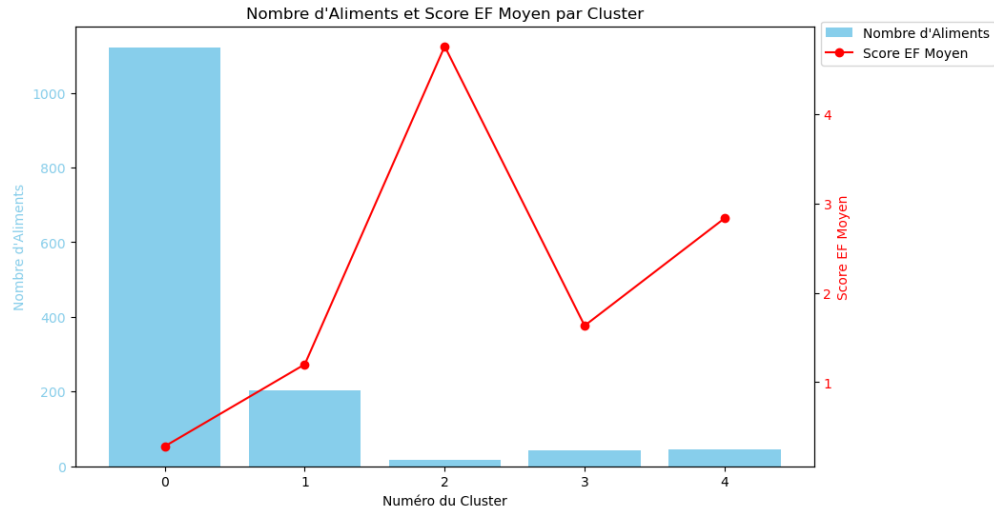


Figure 4: Nombre d'aliment et Score EF moyen par cluster

Pour la suite de notre analyse, nous nous sommes focalisé spécifiquement sur les aliments ayant le plus fort impact environnemental, le cluster 2 dans notre cas. Nous avons examiné les différents sous-groupes d'aliments de ce cluster afin d'évaluer leur impact individuel sur l'environnement. Pour ce faire, nous avons calculé la moyenne de l'impact environnemental de chaque sous-groupe en fonction du score EF. Le graphe ci-dessous illustre cette répartition et met en évidence les niveaux d'impact de chaque sous-groupe sur l'environnement.

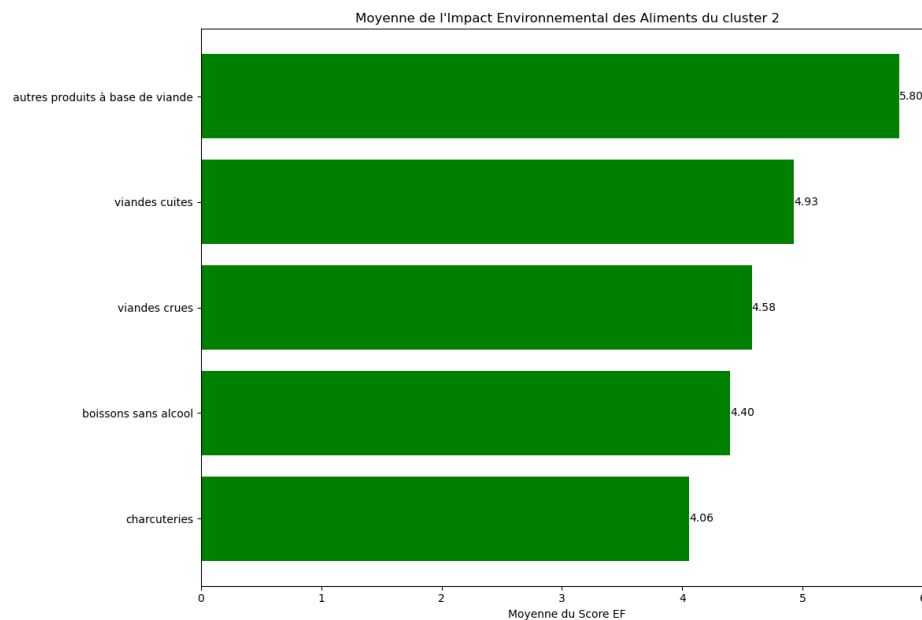


Figure 5: Nombre d'aliment et Score EF moyen par cluster



### 4.3 Correlation et Reduction de dimensionnalité

Dans notre étude sur l'impact environnemental des aliments, nous avons exploré la corrélation entre différents indicateurs. L'objectif était de réduire la dimensionnalité de nos données tout en conservant les informations essentielles. Après avoir examiné la corrélation entre chaque paire d'indicateurs, nous avons constaté des corrélations significatives entre plusieurs d'entre eux comme le montre le graphe ci-dessous.

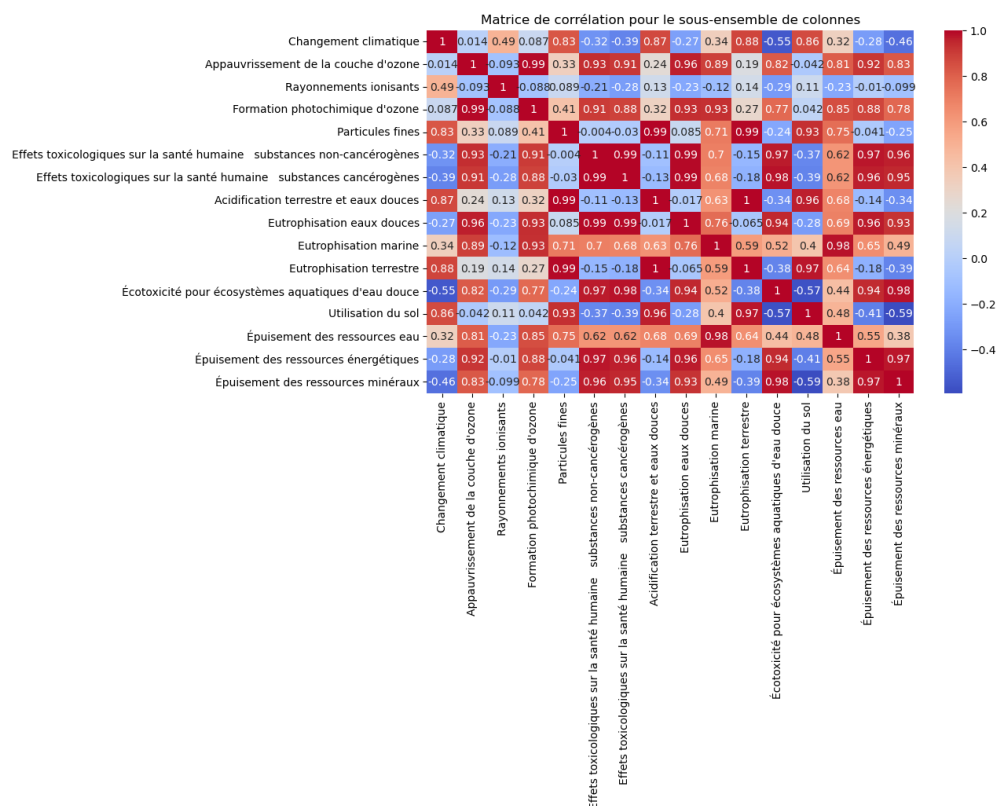


Figure 6: Matrice de corrélation par pair d'indicateur

Cette observation a renforcé notre décision d'appliquer une analyse en composantes principales (PCA) pour réduire la dimensionnalité de nos données tout en préservant au maximum leur variance.

Le tableau récapitulatif ci-dessous, issu de notre analyse PCA, illustre la proportion de variance expliquée par chaque composante principale. Au dessus de ce tableau, le graphe des proportions de variance expliquées en pourcentage offre une visualisation claire de la manière dont chaque composante principale contribue à la variabilité totale des données.

Table 1: summary\_table

Dimension	Explained Variance	Explained Variance Ratio	Cumul Explained Variance Ratio
1	6.778787e-01	5.129386e-01	0.512939
2	5.156703e-01	3.901984e-01	0.903137
3	1.199902e-01	9.079441e-02	0.993931
16	3.269859e-34	2.474243e-34	1.000000

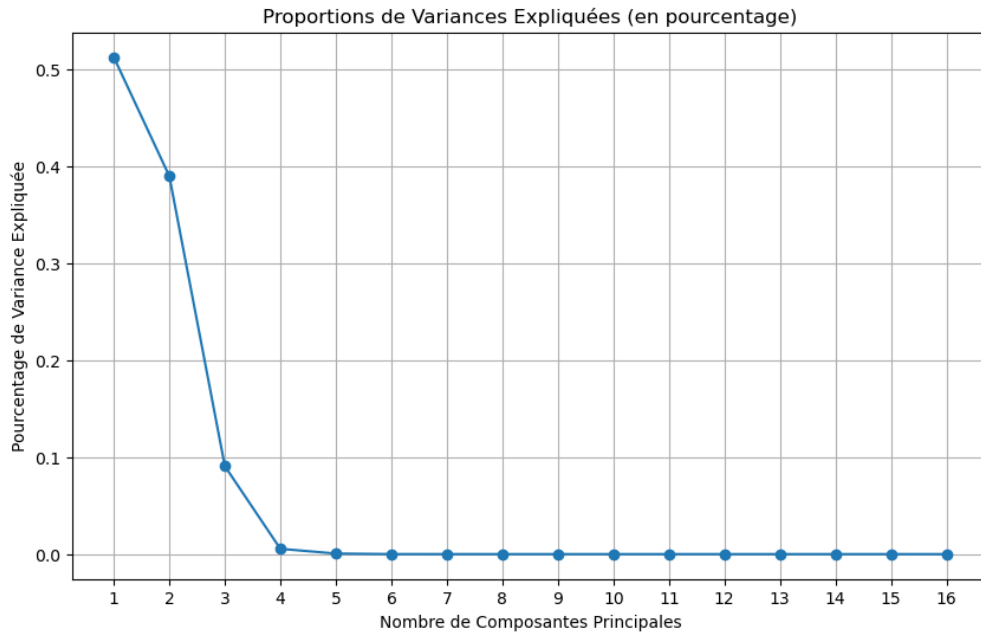


Figure 7: Proportions de variance expliquées

Au vu des résultats obtenus, il est remarquable que seulement **2** ou **3** composantes principales soient nécessaires pour obtenir une représentation presque parfaite de nos données. Cette observation suggère que la structure sous-jacente de nos données peut être réduite à seulement deux ou trois dimensions tout en conservant l'essentiel de l'information.

Afin de mieux visualiser la structure des données dans un espace réduit, nous avons projeté nos données sur les deux premières composantes principales. Cette projection nous a permis d'observer les relations entre les variables et les observations dans un plan bidimensionnel, facilitant ainsi l'interprétation des résultats de notre analyse en composantes principales.

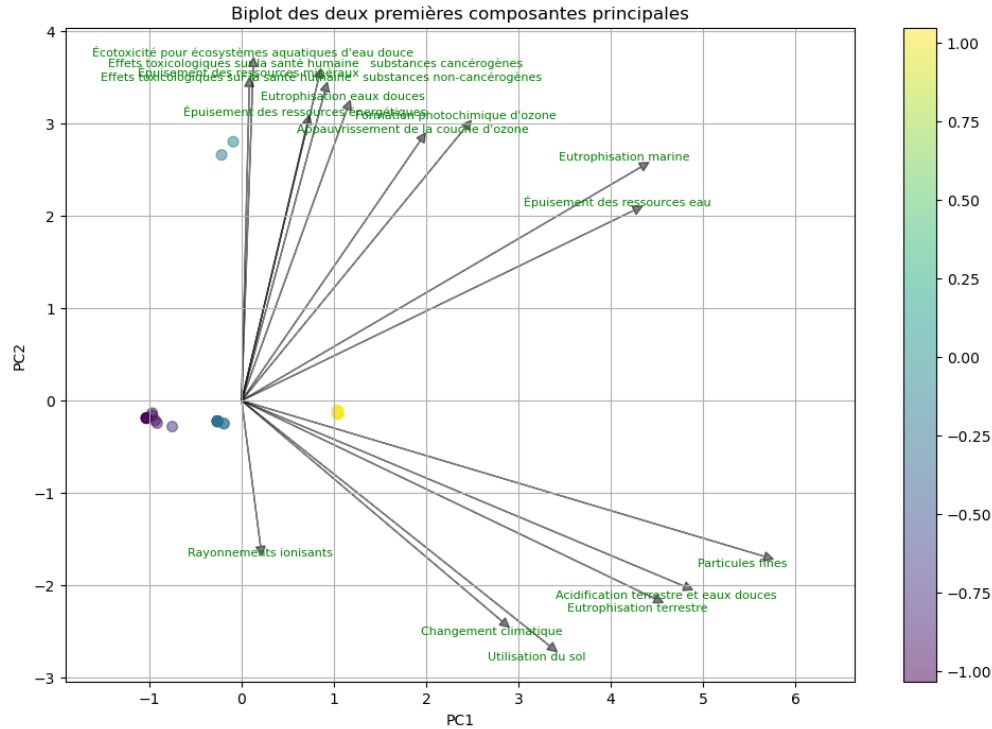


Figure 8: Projection des données

#### 4.4 No IDEA FOR NOW

Jusqu'à présent, notre analyse s'est concentrée sur la détermination de l'impact de chaque sous-groupe d'aliments en fonction du score EF. Cependant, cette approche, bien qu'instructive, ne fournit pas une compréhension approfondie des facteurs spécifiques qui contribuent à l'impact environnemental élevé de chaque sous-groupe d'aliments.

Pour obtenir des insights plus approfondis, nous avons dirigé notre attention vers une analyse plus granulaire. Nous avons cherché à identifier parmi tous les sous-groupes d'aliments présents dans le cluster le plus impactant, quels sont les indicateurs environnementaux qui sont les plus affectés par chaque sous-groupe d'aliments.

Pour explorer cette question, nous avons généré un graphe illustrant l'impact de chaque indicateur environnemental au sein des sous-groupes d'aliments dans le cluster le plus impactant. Il est important de noter que les données ont été normalisées.

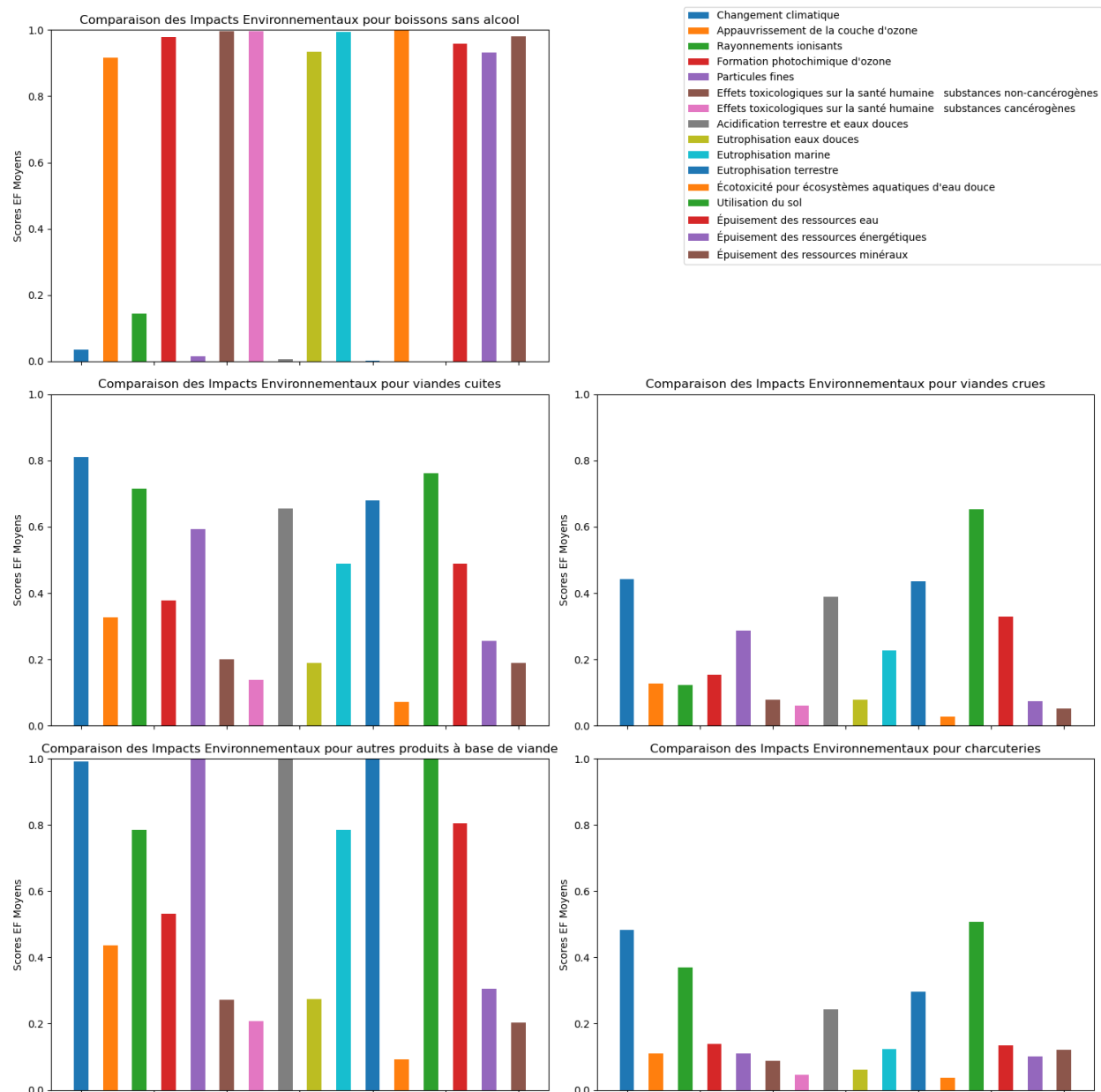


Figure 9: Impact de chaque sous-groupe sur les différents indicateurs

Grâce à cette approche, nous avons pu cibler spécifiquement les aspects à analyser, nous évitant ainsi de devoir examiner des indicateurs qui ne sont pas réellement impactés par un sous-groupe alimentaire en particulier. Dans cette optique, nous avons décidé d'établir un seuil pour sélectionner les indicateurs ayant un impact supérieur ou égal à 0.75. Ce choix de seuil nous permet de nous concentrer sur les indicateurs ayant une influence significative sur l'empreinte environnementale des sous-groupes alimentaires, facilitant ainsi l'identification des principaux facteurs à prendre en compte dans nos recommandations pour une alimentation plus durable.

#### 4.4.1 Fusion des données

Afin d’obtenir une vision plus complète et détaillée de chaque indicateur, depuis l’agriculture jusqu’à la consommation, nous avons jugé nécessaire de fusionner les données synthétiques avec celles des étapes du cycle de vie des produits alimentaires. Cette fusion, réalisée en utilisant la colonne CIQUAL comme référence, nous permet d’enrichir notre analyse en intégrant des informations spécifiques sur chaque étape du processus, de la production à la consommation. En combinant ces deux ensembles de données, nous sommes en mesure de mieux comprendre l’impact environnemental de chaque indicateur tout au long du cycle de vie des produits alimentaires, renforçant ainsi la pertinence et la profondeur de notre analyse

#### 4.4.2 ACV

Nous avons cherché à sélectionner un ou des sous-groupe d’aliments représentatifs pour approfondir notre analyse. Bien que ”les autres produits à base de viande” semblait être le sous-groupe le plus impactant dans le cluster initial, il s’est avéré peu représentatif car il ne comptait qu’un seul exemplaire. De même, les sous-groupes ”boissons sans alcool” et ”charcuteries” ne comptaient que deux exemplaires chacun. En revanche, la majorité des produits appartenaient aux sous-groupes ”viandes cuites” et ”viandes crues”. Nous nous sommes donc concentrés exclusivement sur ces deux sous-groupes, considérés comme plus représentatifs de la diversité des produits alimentaires et de leur impact environnemental dans le cluster.

#### 4.4.3 viandes cuites

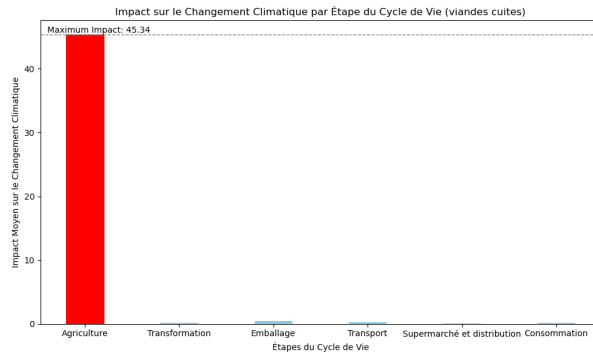
Nous avons réalisé un test d’ANOVA afin d’évaluer s’il existe des différences significatives entre les différentes étapes du cycle de vie d’un indicateur, notamment l’agriculture, la transformation, le transport, l’emballage, la distribution en supermarché et la consommation. Le tableau ci-dessous présente les résultats obtenus pour chaque indicateur.

Table 2: Test d’ANOVA

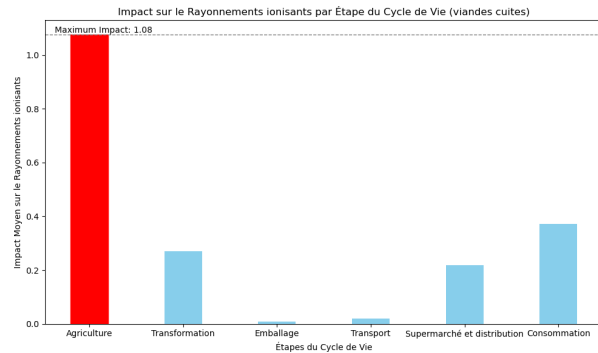
Indicateur	F-value	P-value	Différence Significative
Changement climatique	1485.95	$8.21 \times 10^{-98}$	Vrai
Rayonnements ionisants	108.52	$1.82 \times 10^{-40}$	Vrai
Acidification terrestre et eaux douces	494.26	$1.11 \times 10^{-72}$	Vrai
Eutrophisation terrestre	490.67	$1.62 \times 10^{-72}$	Vrai
Utilisation du sol	288.38	$9.37 \times 10^{-61}$	Vrai

Ces résultats indiquent une différence significative dans les moyennes des indicateurs entre les différentes étapes du cycle de vie des produits et soulignent l’importance d’analyser spécifiquement les impacts environnementaux à chaque étape de la chaîne alimentaire.

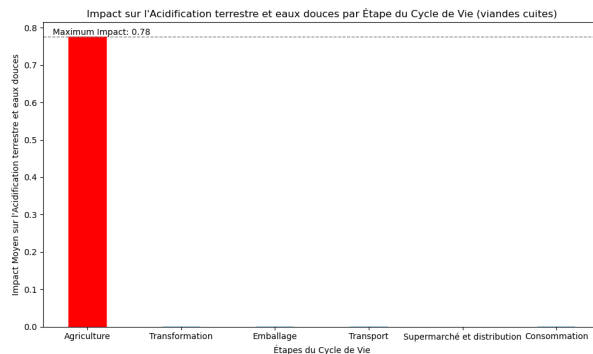
Les graphiques suivants fournissent des informations détaillées sur l'Analyse du Cycle de Vie (ACV) de chaque indicateur, permettant ainsi une compréhension approfondie de leur impact environnemental à différentes étapes du cycle de vie des produits.



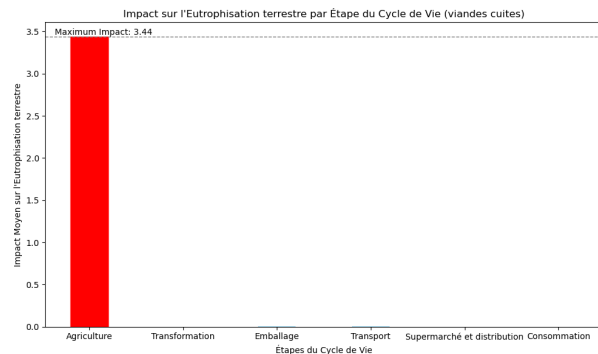
(a) Changement Climatique



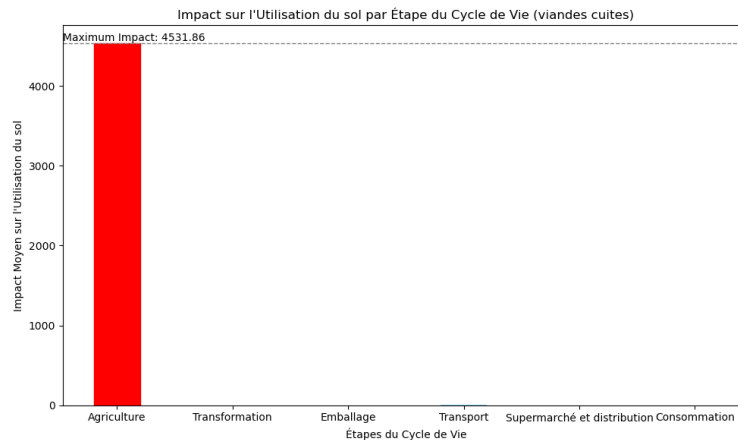
(b) Rayonnement ionisant



(c) Acidification terrestre et eaux douces



(d) Eutrophisation terrestre



(e) Utilisation du sol

En analysant les graphiques, nous constatons que, en moyenne, 97% des impacts environnementaux des viandes cuites se situent lors de la phase de production agricole. Les étapes de transformation, de transport et d'emballage ont un poids moindre en comparaison. Cependant, il est intéressant de noter que la transformation, la consommation et la distribution ont tout de même un impact sur le rayonnement ionisant, bien que ce dernier reste très faible par rapport à la phase de production agricole.

## 4.5 Viandes Crues

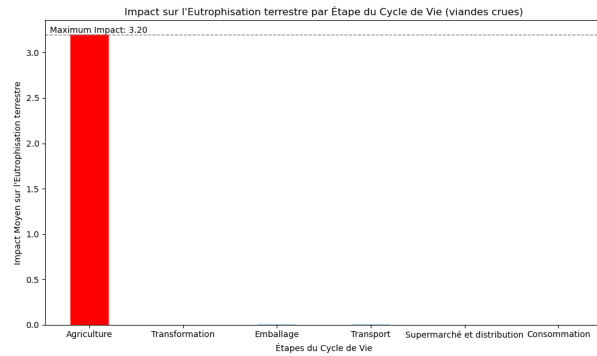
nous avons étendu notre analyse aux viandes crues, suivant la même méthodologie que pour les viandes cuites. Ci-dessous, vous trouverez le tableau présentant les résultats d'ANOVA pour cette analyse, ainsi que les graphiques illustrant l'Analyse du Cycle de Vie (ACV) pour les viandes crues.

Table 3: ANOVA Test

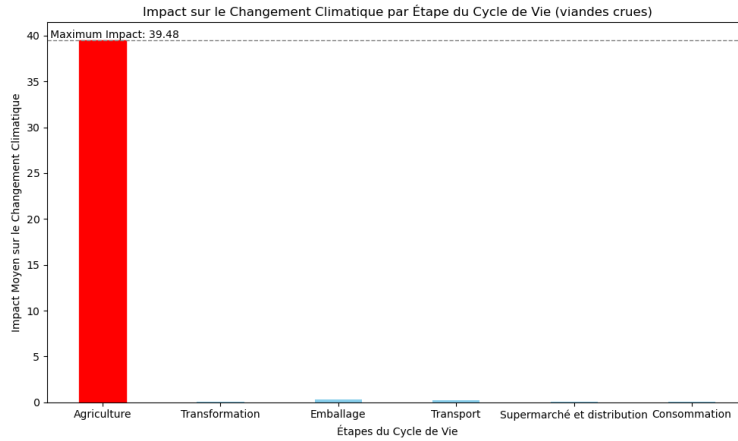
Impact	F-value	P-value	Différence Significative
Changement climatique	$2.02 \times 10^8$	$8.00 \times 10^{-154}$	Vrai
Eutrophisation terrestre	$1.30 \times 10^{12}$	$8.25 \times 10^{-23}$	Vrai
Utilisation du sol	$1.12 \times 10^{13}$	$1.74 \times 10^{-253}$	Vrai



(a) Changement Climatique



(b) Eutrophisation terrestre



(c) Utilisation du sol

Nos résultats révèlent que les viandes crues partagent des similitudes dans leurs impacts environnementaux avec les viandes cuites, notamment en ce qui concerne la prépondérance de l'impact environnemental lors de la phase de production agricole.

## 5 Implémentation des Modèles pour la Prédiction de l'Impact Environnemental

Dans cette section, nous abordons l'implémentation des modèles de prédiction pour évaluer l'impact environnemental en fonction des indicateurs. Notre objectif est de déterminer si nous pouvons prédire la classe d'impact (impact ou pas d'impact) en utilisant le **score unique EF** comme label. Pour ce faire, nous avons divisé le score EF en deux catégories en fonction de la moyenne de ses valeurs : 0 pour "pas d'impact" et 1 pour "impact", nous avons obtenu 1211 exemples pour la classe 1 et 1212 pour la classe 0.

### 5.1 Préparation des Données

Nous avons utilisé les données qui avaient été appliquées au PCA, en retenant uniquement les deux axes principaux. Les données ont ensuite été divisées en ensembles d'entraînement et de test.

### 5.2 Modèles Implémentés

Nous avons implémenté cinq modèles différents :

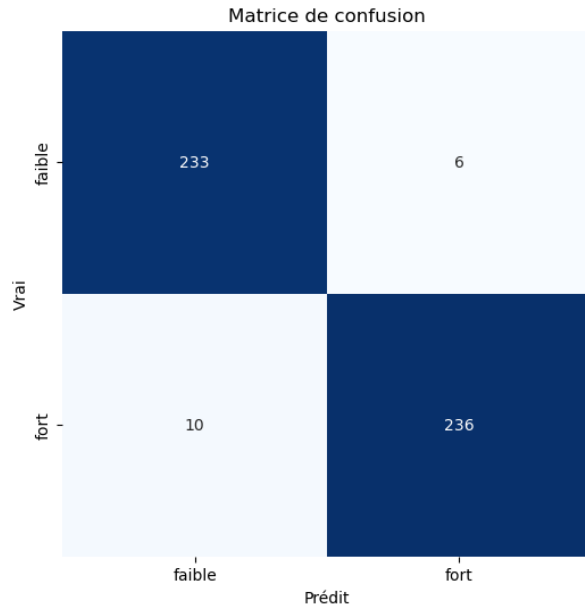
#### 5.2.1 Régression Logistique

- Nous avons utilisé la régression logistique pour prédire la classe d'impact en utilisant une cross validation.
- Accuracy : 97.7% sur le jeu de donnée de test.

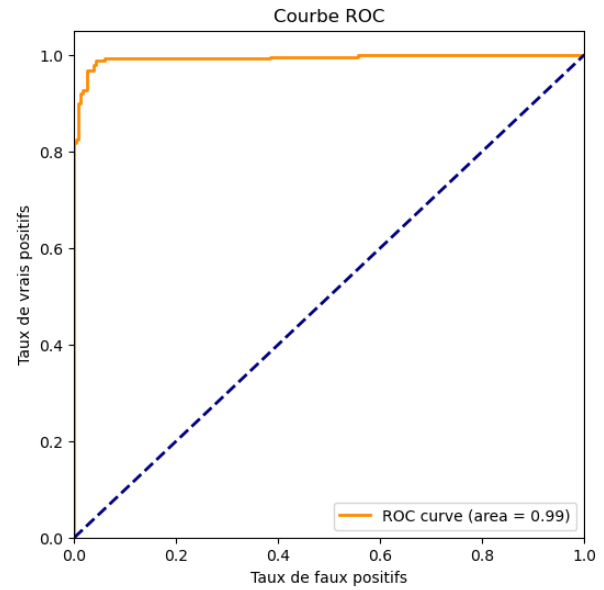
Table 4: Résultats des métriques de classification

Classe	Précision	Rappel	F1-score	Support
Faible	0.96	0.97	0.97	239
Fort	0.98	0.96	0.97	246
Précision globale			0.97	485
Moyenne pondérée	0.97	0.97	0.97	485





(a) Matrice de confusion



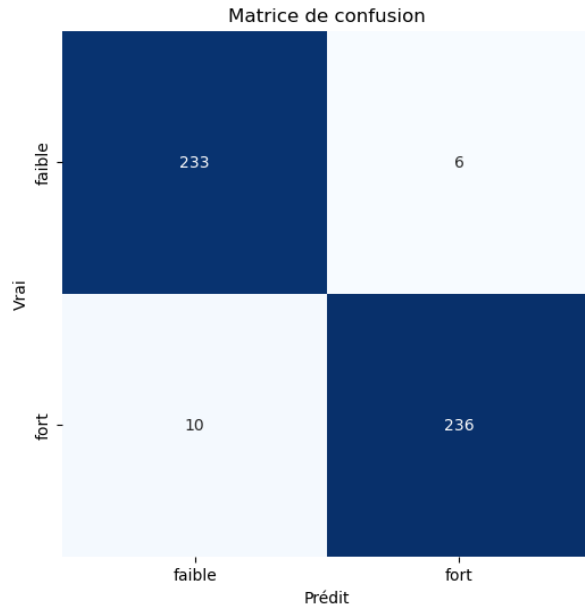
(b) Courbe Roc

### 5.2.2 Support Vector Machine (SVM)

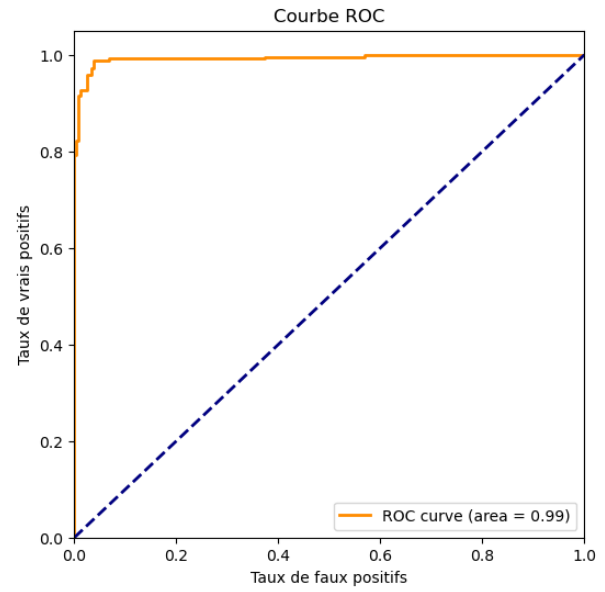
- Nous avons effectué une recherche des meilleurs paramètres pour le modèle SVM en utilisant un *GridSearch*.
- Meilleurs paramètres :  $C = 1$ , kernel = "linear".
- Accuracy : 96.7% sur le jeu de donnée de test.

Table 5: Résultats des métriques de classification

Classe	Précision	Rappel	F1-score	Support
Faible	0.96	0.97	0.97	239
Fort	0.98	0.96	0.97	246
Précision globale			0.97	485
Moyenne pondérée	0.97	0.97	0.97	485



(a) Matrice de confusion



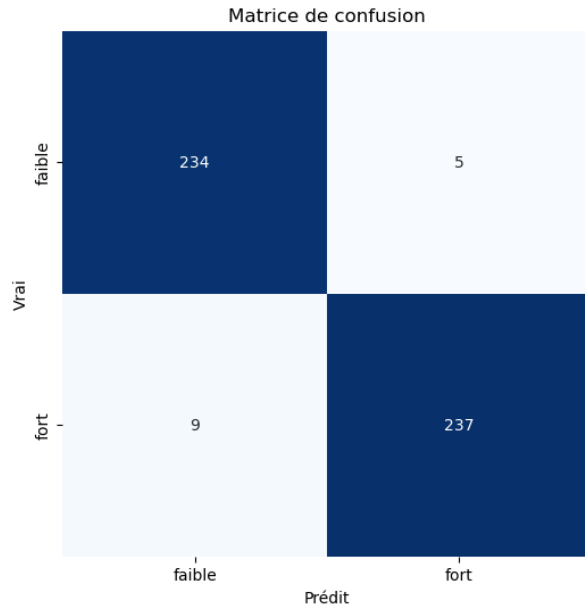
(b) Courbe Roc

### 5.2.3 Random Forest (RF)

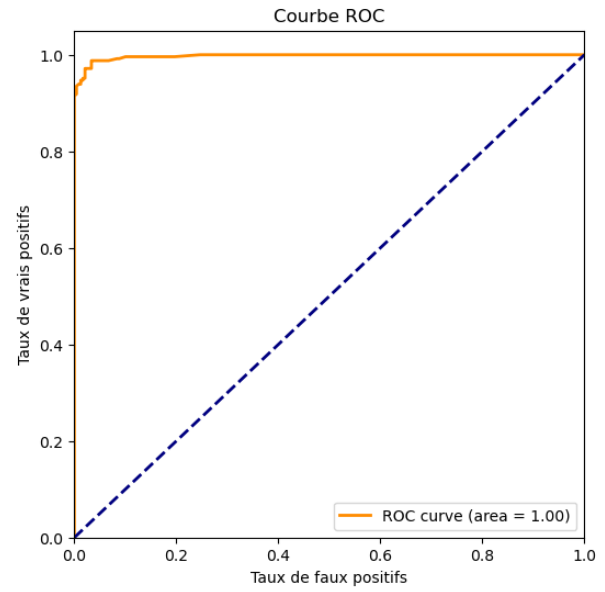
- Nous avons entraîné un modèle Random Forest pour la prédiction.
- Accuracy : 97% sur le jeu de donnée test.

Table 6: Résultats des métriques de classification

Classe	Précision	Rappel	F1-score	Support
Faible	0.96	0.98	0.97	239
Fort	0.98	0.96	0.97	246
Précision globale			0.97	485
Moyenne pondérée	0.97	0.97	0.97	485
weighted avg	0.97	0.97	0.97	485



(a) Matrice de confusion



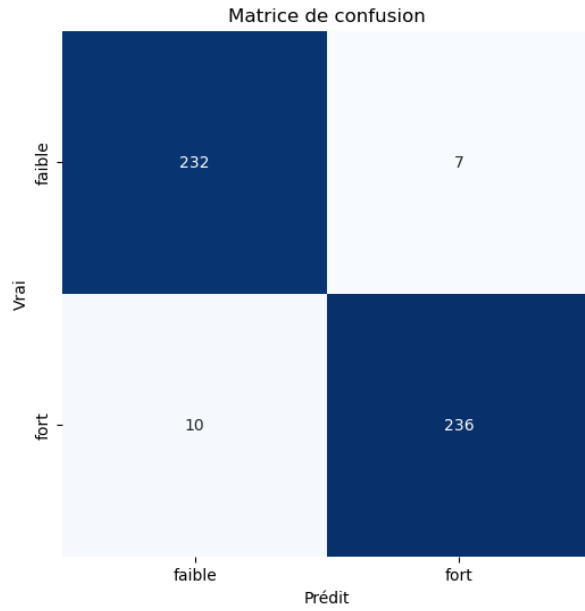
(b) Courbe Roc

#### 5.2.4 Decision Tree (DT)

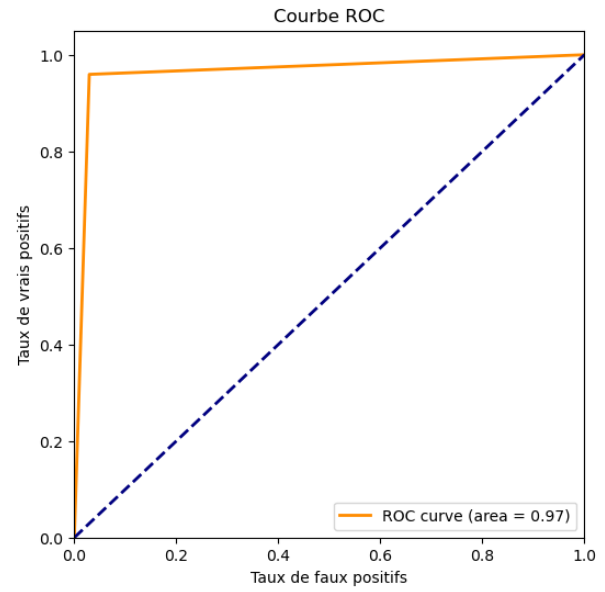
- Nous avons utilisé un arbre de décision pour la prédiction.
- Accuracy : 96% sur le jeu de donnée test.

Table 7: Résultats des métriques de classification

Classe	Précision	Rappel	F1-score	Support
Faible	0.96	0.97	0.96	239
Fort	0.97	0.96	0.97	246
Précision globale			0.96	485
Moyenne pondérée	0.96	0.97	0.96	485
weighted avg	0.97	0.96	0.96	485



(a) Matrice de confusion



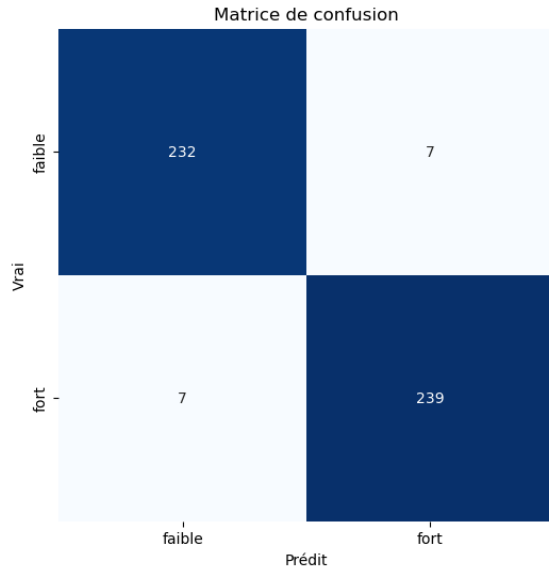
(b) Courbe Roc

### 5.2.5 K-Nearest Neighbors (KNN)

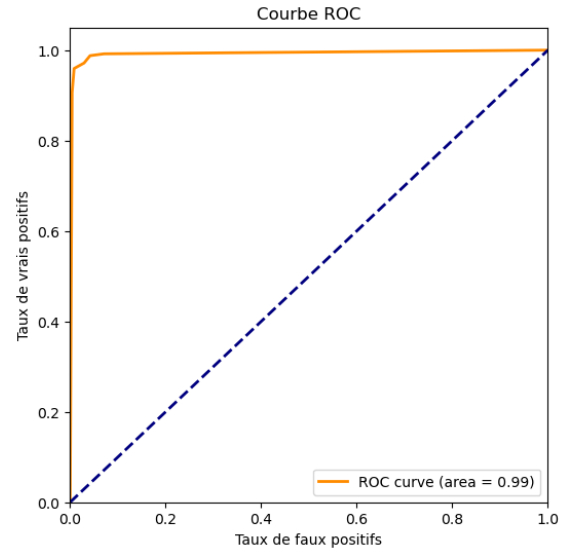
- Nous avons entraîné un modèle KNN pour la prédiction.
- Accuracy : 97% sur le jeu de donnée test.

Table 8: Résultats des métriques de classification

Classe	Précision	Rappel	F1-score	Support
Faible	0.97	0.97	0.97	239
Fort	0.97	0.97	0.97	246
Précision globale			0.97	485
Moyenne pondérée	0.97	0.97	0.97	485
weighted avg	0.97	0.97	0.97	485



(a) Matrice de confusion



(b) Courbe Roc

Pour comparer les courbes ROC des différents modèles, nous les avons toutes superposées sur un même graphique.

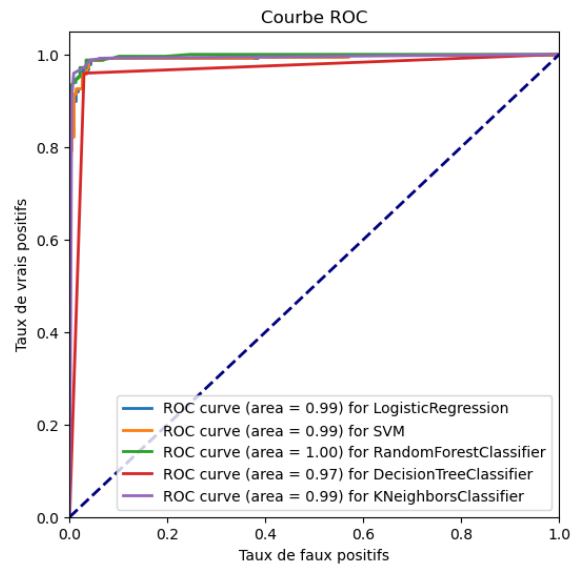


Figure 17: Courbes Roc

## 6 Test de Significativité entre les Modèles

Après avoir obtenu les résultats des différents modèles de classification, nous avons cherché à évaluer leur performance de manière plus approfondie. Pour cela, nous avons calculé la précision de chaque modèle au niveau de chaque instance de l'ensemble de test.

Ensuite, nous avons effectué un test de significativité pour comparer la performance des modèles deux à deux. Nous avons utilisé le test t de Student apparié pour déterminer s'il existait des différences significatives entre les performances des différents modèles. L'hypothèse nulle ( $H_0$ ) de ce test était que les performances des deux modèles étaient équivalentes, tandis que l'hypothèse alternative ( $H_1$ ) était qu'il y avait une différence significative entre les performances.

### 6.0.1 Résultats du Test de Significativité

Les résultats du test de significativité sont présentés dans le tableau ci-dessous. Les colonnes "Model-1" et "Model-2" indiquent les modèles comparés, "t-statistic" représente la statistique t calculée, "p-value" est la valeur p associée au test, et "Hypothesis Result" indique si l'hypothèse nulle a été rejetée ou non à un niveau de signification de 5%.

Table 9: Résultats du test de significativité entre les modèles

Model-1	Model-2	t-statistic	p-value	Hypothesis Result
Logistic Regression	SVM	NaN	NaN	Fail to reject $H_0$
Logistic Regression	Random Forest	0.000	1.000	Fail to reject $H_0$
Logistic Regression	Decision Tree	0.229	0.819	Fail to reject $H_0$
Logistic Regression	KNN	-0.500	0.618	Fail to reject $H_0$
SVM	Random Forest	0.000	1.000	Fail to reject $H_0$
SVM	Decision Tree	0.229	0.819	Fail to reject $H_0$
SVM	KNN	-0.500	0.618	Fail to reject $H_0$
Random Forest	Decision Tree	0.447	0.655	Fail to reject $H_0$
Random Forest	KNN	-0.632	0.528	Fail to reject $H_0$
Decision Tree	KNN	-0.904	0.366	Fail to reject $H_0$

Les résultats montrent qu'il n'y a pas de différences significatives entre les performances des différents modèles. Aucune des paires de modèles comparées n'a permis de rejeter l'hypothèse nulle, indiquant que les performances des modèles sont équivalentes dans l'ensemble.

## 7 Conclusion

Notre analyse approfondie de l’impact environnemental des aliments a permis de mettre en évidence plusieurs points clés. En premier lieu, l’application de l’algorithme K-means avec 5 clusters, visualisée par t-SNE, a révélé des variations significatives dans l’impact environnemental entre les différents groupes d’aliments.

En se concentrant sur les aliments ayant le plus fort impact environnemental, le cluster 2, nous avons identifié les sous-groupes spécifiques et leur contribution individuelle à l’impact environnemental en utilisant le score EF. Cette approche a permis de cibler les aspects critiques à analyser pour chaque sous-groupe d’aliments.

L’étude des corrélations entre les différents indicateurs environnementaux a justifié l’utilisation de l’analyse en composantes principales (PCA) pour réduire la dimensionnalité des données tout en préservant l’essentiel de l’information. Les résultats indiquent que seulement deux ou trois composantes principales suffisent pour représenter efficacement la variabilité des données.

De plus, l’analyse a révélé que les impacts environnementaux principaux des sous-groupes d’aliments les plus impactants étaient principalement associés à la phase agricole. Cette découverte souligne la nécessité de réviser et d’améliorer les pratiques agricoles actuelles pour réduire l’empreinte écologique des productions alimentaires.

En conclusion, cette étude a non seulement mis en lumière les variations d’impact environnemental entre les différents clusters d’aliments, mais a également démontré l’efficacité des techniques de réduction de dimensionnalité pour simplifier l’analyse des données complexes. Ces insights offrent une base solide pour des actions futures visant à réduire l’impact environnemental des produits alimentaires, notamment en repensant les pratiques agricoles utilisées pendant la phase agricole.