

## Projet : Étude de l'eco-score

L'eco-score est un indicateur permettant de calculer l'impact écologique d'un aliment. Il est développé par divers acteurs du secteur de l'alimentation durable avec le soutien d'acteurs tels que l'ADEME ou Zero Waste France. Cet indicateur est calculé à partir de la base de données AGRIBALYSE. Le programme AGRIBALYSE, mené par l'ADEME et l'INRAE, produit des méthodologies et des données sur l'évaluation environnementale des produits agricoles et alimentaires. Chaque aliment est évalué sur l'ensemble de son cycle de vie et selon une analyse multicritère. Les différentes valeurs de cette analyse sont ensuite agrégées en utilisant une moyenne pondérée ; le score ainsi produit est normalisé à l'aide d'une formule mathématique ; enfin, l'indicateur est obtenu en ajoutant divers bonus-malus. La méthodologie détaillée est disponible sur ce site.

Soit  $\mathcal{A}$  l'ensemble des aliments. Les éléments de  $\mathcal{A}$  sont évalués sur les 16 critères suivants :

1. L'impact sur le réchauffement climatique, mesuré en kilogrammes de dioxyde de carbone équivalent par kilogramme de produit alimentaire ( $\text{kgCO}_{2eq}/\text{kg}$ ).
2. L'impact sur l'appauvrissement de la couche d'ozone, mesuré en kilogrammes de trichlorofluorométhane équivalent par kilogramme de produit alimentaire ( $\text{kgCFC}_{11eq}/\text{kg}$ ).
3. Le niveau de radiations ionisantes, mesuré en kiloBecquerels U-235 équivalent par kilogramme de produit alimentaire ( $\text{kBq}_{eq}/\text{kg}$ ).
4. Le niveau de formation photochimique d'ozone, mesuré en kilogrammes de composés organiques volatils non-méthaniques par kilogramme de produit alimentaire ( $\text{kgNMVOC}_{eq}/\text{kg}$ ).
5. Le niveau d'émissions de particules fines, mesuré en augmentation de l'incidence de maladies par kilogramme de produit alimentaire (disease inc./kg).
6. Les effets toxicologiques sur la santé humaine (substances non-cancérogènes), mesuré en unités toxiques comparatives pour l'humain par kilogramme de produit alimentaire ( $\text{CTUh}/\text{kg}$ ).
7. Les effets toxicologiques sur la santé humaine (substances cancérogènes), mesuré en unités toxiques comparatives pour l'humain par kilogramme de produit alimentaire ( $\text{CTUh}/\text{kg}$ ).
8. L'acidification terrestre et des eaux douces, mesuré en moles d'ions hydrogène par kilogramme de produit alimentaire ( $\text{mol H}_{eq}^+/\text{kg}$ ).
9. L'eutrophisation des eaux douces, mesuré en kilogrammes de phosphore équivalent par kilogramme de produit alimentaire ( $\text{kgP}_{eq}/\text{kg}$ ).
10. L'eutrophisation des eaux marines, mesuré en kilogrammes d'azote équivalent par kilogramme de produit alimentaire ( $\text{kgN}_{eq}/\text{kg}$ ).
11. L'eutrophisation terrestre, mesuré en moles d'azote équivalent par kilogramme de produit alimentaire ( $\text{mol N}_{eq}/\text{kg}$ ).
12. Le niveau d'écotoxicité pour les écosystèmes aquatiques d'eau douce, mesuré en unités toxiques comparatives pour l'environnement par kilogramme de produit alimentaire ( $\text{CTUe}/\text{kg}$ ).
13. Le niveau d'utilisation du sol, mesuré en points eco-indicateur par kilogramme de produit alimentaire ( $\text{Pt}/\text{kg}$ ).
14. L'impact sur l'épuisement des ressources en eau, mesuré en mètres cubes par kilogramme de produit alimentaire ( $\text{m}^3/\text{kg}$ ).
15. L'impact sur l'épuisement des ressources énergétiques, mesuré en MégaJoules par kilogramme de produit alimentaire ( $\text{MJ}/\text{kg}$ ).

16. L'impact de l'empreinte matière, mesuré en kilogrammes antimoine équivalent par kilogramme de produit alimentaire ( $\text{kgSb}_{eq}/\text{kg}$ ).

Une alternative  $X \in \mathcal{A}$  est donc exprimée sous la forme d'un tuple de 16 éléments  $(x_i)_{i \in [16]}$ . Notez que chaque critère est à minimiser, i.e., une valeur plus faible est plus désirable qu'une valeur plus élevée. Dans ce projet, nous allons analyser plusieurs relations binaires définies sur  $\mathcal{A}$  afin de classer ses éléments et nous interroger sur les choix réalisés dans la définition de l'eco-score. Pour ce faire nous allons utiliser un programme informatique en Python, et les données du projet AGRIBALYSE.

- 1 Télécharger le fichier AGRIBALYSE3.1\_partie agriculture\_conv\_vf.xlsx sur ce site. Ces données sont disponibles sous licence ETALAB. Pour récupérer ces données dans votre programme python, vous pouvez utiliser le package Python *pandas*, et le bout de code suivant :

```
1 import pandas as pd
2 exc = pd.ExcelFile("AGRIBALYSE3.1_partie agriculture_conv_vf.xlsx")
3 df = pd.read_excel(exc, 'AGB_agri_conv', usecols = [0]+[i for i in range(3,20)])
4 data = df[2:].values
```

- 2 Normaliser les différents critères en divisant chaque valeur par les facteurs suivants :

Critère	1	2	3	4	5	6
Facteur de norm.	$7.55 \times 10^3$	$5.23 \times 10^{-2}$	$4.22 \times 10^3$	$4.09 \times 10$	$5.95 \times 10^{-4}$	$1.29 \times 10^{-4}$
Critère	7	8	9	10	11	12
Facteur de norm.	$1.73 \times 10^{-5}$	$5.56 \times 10$	1.61	$1.95 \times 10$	$1.77 \times 10^2$	$5.67 \times 10^4$
Critère	13	14	15	16		
Facteur de norm.	$8.19 \times 10^5$	$1.15 \times 10^4$	$6.5 \times 10^4$	$6.36 \times 10^{-2}$		

- 3 En utilisant les trois critères : impact sur le dérèglement climatique, impact sur l'épuisement des ressources en eau, et impact sur l'épuisement des ressources matières (critères 1, 14, et 16). Réaliser des graphiques pour illustrer les liens entre ces différents critères (e.g., avec le critère en abscisse et le critère en ordonnée).

**Pareto dominance.** Nous allons comparer les alternatives vis-à-vis de la Pareto dominance. Une alternative  $X = (x_i)_{i \in [16]} \in \mathcal{A}$  Pareto-domine une alternative  $Y = (y_i)_{i \in [16]} \in \mathcal{A}$  ssi  $\bigwedge_{i \in [16]} x_i \leq y_i$  et  $\bigvee_{i \in [16]} x_i < y_i$ . Dit autrement  $X$  doit être “aussi bien” que  $Y$  sur tous les critères et strictement meilleure sur au moins un critère.

- 4 Quelles sont les propriétés vérifiées par la Pareto Dominance ? À quelle structure particulière, vue en cours, correspond-elle ? Vous devez apporter des justifications pour les propriétés vérifiées et celles non vérifiées.
- 5 Ecrire une méthode afin de déterminer si une alternative  $X$  Pareto-domine une alternative  $Y$ .
- 6 Réaliser des méthodes pour vérifier les propriétés identifiées à la question 4 devant être vérifiées par la Pareto dominance (une méthode par propriété identifiée).
- 7 Calculer le pourcentage d'ensembles  $\{X, Y\}$  de deux alternatives avec  $X \in \mathcal{A} \neq Y \in \mathcal{A}$  tels que  $X$  Pareto-domine  $Y$ . Comment ce pourcentage évolue-t-il si les alternatives n'étaient évaluées que sur les critères 1, 14, et 16 ?

**Somme pondérée.** L'eco-score agrège ensuite les 16 scores de la partie précédente à l'aide d'une somme pondérée, ce qui donne un score en points eco-indicateur. Le jeu de poids utilisé, noté  $\hat{w}$ , est donné ci-dessous :

Critère	1	2	3	4	5	6	7	8
Poids	$21.06 \times 10^{-2}$	$6.31 \times 10^{-2}$	$5.01 \times 10^{-2}$	$4.78 \times 10^{-2}$	$8.96 \times 10^{-2}$	$1.84 \times 10^{-2}$	$2.13 \times 10^{-2}$	$6.2 \times 10^{-2}$
Critère	9	10	11	12	13	14	15	16
Poids	$2.8 \times 10^{-2}$	$2.96 \times 10^{-2}$	$3.71 \times 10^{-2}$	$1.92 \times 10^{-2}$	$7.94 \times 10^{-2}$	$8.51 \times 10^{-2}$	$8.32 \times 10^{-2}$	$7.55 \times 10^{-2}$

- 8 Calculer les taux de substitutions entre les critères 1, 14, et 16 et donner les interprétations de ces taux de substitution.

Soit  $\text{SP}_w(X) = \sum_{i \in [16]} w_i x_i$ , le score de l'alternative  $X \in \mathcal{A}$  lorsque l'on considère la somme pondérée avec le jeu de poids  $w$ . Soit  $X, Y$  une paire d'alternatives telles que  $\text{SP}_{\hat{w}}(X) < \text{SP}_{\hat{w}}(Y)$ . L'alternative  $X$  est donc jugée “plus écologique” que  $Y$ . On cherche à estimer la robustesse de cette comparaison vis-à-vis du jeu de poids  $w$ .

- 9 Proposer un programme linéaire permettant de trouver le jeu de poids  $w$  (poids positifs, sommant à 1), tel que  $\text{SP}_w(X) \geq \text{SP}_w(Y)$ , et que  $\|w - \hat{w}\|_1$  soit minimale.
- 10 Implémenter ce programme linéaire dans une méthode, nommée `L1_inv`, qui étant donné  $X, Y$  une paire d'alternatives telles que  $\text{SP}_{\hat{w}}(X) < \text{SP}_{\hat{w}}(Y)$ , retourne la valeur optimale du programme linéaire précédant. Vous utiliserez le package Python `mip`.
- 11 Réaliser un graphique associant à chaque valeur  $x \in [0, 2]$ , le nombre de paires  $(X, Y)$  telles que  $\text{L1\_inv}(X, Y) \leq x$ . Quelle critique peut-on faire sur l'interprétation de ce graphique ?

Soit  $X = (x_i)_{i \in [16]}$  une alternative et  $\sigma$  la permutation de  $[16]$  telle que

$$x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(16)}.$$

On note  $\vec{X} = (x_{\sigma(i)})$  le tuple obtenu à partir de  $X$  en réarrangeant les éléments de  $X$  selon la permutation  $\sigma$ . La *moyenne pondérée ordonnée* de jeu de poids  $w$  de  $X$  est définie par

$$\text{MPO}_w(X) = \sum_{i \in [16]} w_i x_{\sigma(i)}.$$

La *moyenne géométrique pondérée* de jeu de poids  $w$  de  $X$  est définie par

$$\text{MG}_w(X) = \sqrt[\sum w_i]{\prod_{i \in [16]} x_i^{w_i}}.$$

- 12 Implémenter une méthode permettant d'évaluer les différentes alternatives par une moyenne pondérée ordonnée. Implémenter une méthode permettant d'évaluer les différentes alternatives par une moyenne géométrique pondérée. Proposer quelques jeux de poids qui vous semblent appropriés.
- 13 Quelle propriété peut être satisfaite par ces deux agrégateurs qui n'est pas satisfaite par la somme pondérée ?

Les différents agrégateurs précédents définissent implicitement des ordres forts. La distance de Kendall-tau est définie sur les ordres stricts totaux. Pour notre problème, étant donnés deux ordres forts  $R$  et  $R'$ , nous proposons de définir la distance de Kendall-Tau  $d_{KT}(R, R')$  de la manière suivante :

$$d_{KT}(R, R') = \sum_{(i,j) \in \mathcal{A}^2} 0.5(iRj \wedge jR'i) + 0.5((iRj \wedge \neg(iR'j)) \vee (\neg(jRi) \wedge jR'i)).$$

Dit d'une autre manière, chaque paire  $i, j$  induit une pénalité de 1 si la préférence est inversée entre  $R$  et  $R'$  et de 0.5 si l'un des ordres forts exprime une préférence mais pas l'autre.

- 14 Montrer que  $d_{KT}$  définit bien une distance sur les ordres forts définis sur  $\mathcal{A}$ .
- 15 Implémenter des méthodes afin de calculer les distances  $d_{KT}$  entre les ordres forts résultant des différents agrégateurs explorés précédemment. Rapporter vos résultats pour différents jeux de poids.

**Normalisation logarithmique de l’eco-score.** Après l’obtention de la somme pondérée  $\mathbf{SP}_{\hat{w}}$ , l’eco-score se ramène à un score entre 0 et 100 en utilisant la formule suivante :

$$\mathbf{score}(x) = 100 - \frac{\ln(10x + 1)}{\ln(2 + \frac{1}{100x^4})} \times 20.$$

- 16 Que pensez vous de la formule logarithmique utilisée ? Quels avantages et désavantages y voyez vous ? Essayez de proposer une alternative (sans la tester).

En fait, l’ecoscore est plus utilisé pour une tâche de catégorisation que pour une tâche de rangement. À partir du score précédent, une catégorie A, B, C, D, ou E est assignée à l’alternative en fonction de son score. Une catégorie A pour un score dans  $[100, 80)$ , B pour un score dans  $[80, 60)$ , C pour un score dans  $[60, 40)$ , D pour un score dans  $[40, 20)$ , E pour un score dans  $[20, 0)$ . Les scores de 80, 60, 40 et 20 sont obtenus (approximativement) pour les valeurs de 0.28, 0.43, 0.75, et 1.5 respectivement.

- 17 Réaliser des tests similaires à ceux réalisés aux questions 9-11, pour voir combien d’alternatives peuvent changer de catégorie si le jeu de poids  $\hat{w}$  est modifié.

**Rendu.** Le rendu de votre projet doit comprendre un rapport et une archive de votre code. On attend bien sûr un code bien organisé et commenté, muni d’un fichier ReadMe. Votre rapport contiendra les réponses aux questions 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 16, et 17.