

A PROPOS DE LA TRANSPARENCE DES ALGORITHMES SUR LE CALCUL DES INFORMATIONS NUTRITIONNELLES DE PRODUITS ALIMENTAIRES

Année académique 2022/2023

L'objectif de ce projet est de sensibiliser les étudiants à une démarche d'analyse, d'interprétation et de transparence des modèles d'évaluation utilisés pour l'élaboration des classements ou classification de produits, d'organismes, de personnes, etc. Nous examinerons ici le cas des produits alimentaires où il existe plusieurs systèmes d'étiquetage nutritionnel pour faciliter le choix d'achat du consommateur, au regard de la composition nutritionnelle des produits. Ce travail sera réalisé à l'aide d'algorithmes implémentés en langage Python.

1 Le Nutri-score

1.1 Introduction

Le Nutri-Score, est un système d'étiquetage nutritionnel qui répartit les produits en cinq classes (5 couleurs), du vert (catégorie A) pour les produits de très bonne qualité nutritionnelle au rouge (catégorie E) pour ceux dont il vaut mieux limiter la consommation. Il a été mis en place en janvier 2016, par le gouvernement français, dans le cadre de la loi de modernisation du système de santé. Son but est de faciliter le choix d'achat du consommateur, au regard de la composition nutritionnelle des produits. Ainsi, il favorise le choix de produits plus sains par les consommateurs et par conséquent, participe à la lutte contre l'augmentation des maladies cardiovasculaires, de l'obésité et du diabète.

Le logo Nutri-Score a été conçu par Santé publique France, à la demande de la Direction générale de la santé, en s'appuyant sur les travaux de l'équipe du Professeur Serge Hercberg (Université Paris 13), les expertises de l'Anses (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail) et du Haut Conseil de Santé Publique.



FIGURE 1 – Logo Nutri-score

L'ensemble des articles scientifiques et documents publiés relatifs au Nutri-Score est accessible sur le site internet <https://solidarites-sante.gouv.fr/prevention-en-sante/preserver-sa-sante/nutrition/article/articles-scientifiques-et-documents-publies-relatifs-au-nutri-score>.

1.2 Méthode de calcul du Nutriscore

Cette section est rédigée à partir du Rapport d'appui scientifique et technique de l'Anses sur le Nutri-score¹.

Le Nutri-score permet de calculer le score nutritionnel des produits alimentaires en se basant sur :

- **Les nutriments ou apports à favoriser** : fibres, protéines, fruits, légumes, légumineuses, noix ;
- **Les nutriments ou apports à limiter** : énergie (kcal), acides gras saturés, sucres, sel.

Cette note est proposée après analyse d'une portion de 100 grammes ou 100 ml s'il s'agit de boisson. Plus formellement, le Nutri-score est basé sur le score nutritionnel des aliments, tel que défini par Rayner et al.², est un score intégrant :

- Une composante dite «négative», calculée à partir des teneurs en nutriments dont la consommation doit être limitée : énergie (kJ/100g), sucres simples (g/100g), acides gras saturés (g/100g) et sodium (mg/100g) ;
- Une composante dite «positive», calculée en intégrant les teneurs en nutriment dont la consommation est recommandée : fibres (g/100g), protéines (g/100g) ;
- Une deuxième composante «positive», calculée à partir des teneurs d'une catégorie spécifique d'aliments : les fruits/légumes/fruits à coque (g/100g)

Les composantes "positives" et "négatives" sont chacune associées à un score plus ou moins important, en fonction de la composition nutritionnelle de l'aliment considéré :

- De 0 à 10 pour les nutriments de la composante "négative" (Voir Tableau de la Figure 3) ;
- De 0 à 5 pour les éléments de la composante "positive" (Voir Tableau de la Figure 2).

Ainsi, le score associé à la composante "négative" peut théoriquement aller de 0 à 40. Celui de la composante "positive" peut quant à lui théoriquement aller de 0 à 15.

Points	Fruits, leg (%)	Fibres (g)	Protéine (g)
		AOAC	
0	≤ 40	≤ 0.9	≤ 1,6
1	> 40	> 0.9	> 1,6
2	> 60	> 1.9	> 3,2
3	-	> 2.8	> 4,8
4	-	> 3.7	> 6,4
5	> 80	> 4.7	> 8,0

FIGURE 2 – Table de Calcul des points attribués à chacun des éléments de la composante dite "positive" selon la méthodologie développée par Rayner et al.

Pour chaque aliment, le score nutritionnel est ensuite calculé selon la méthodologie présentée à la Figure 1 5.

Dans la plupart des cas, le score est calculé en retranchant le score de la composante "positive" à celui de la composante "négative" (voir Figure 4).

$$\text{Nutri-Score} = \text{nutriments à limiter (a)} - \text{nutriments à favoriser (b)} \quad (1)$$

Cependant, si le score de la composante "négative" est supérieur ou égal à 11 et que la teneur en fruits/légumes/fruits à coque ne dépasse pas 80%, alors les protéines ne sont plus prises en compte dans le calcul du score nutritionnel.

1. Évaluation de la faisabilité du calcul d'un score nutritionnel tel qu'élaboré par Rayner et al. (pages 12-15 et 27-30). Mars 2015
 2. Application of the Nutrient profiling model : Definition of "fruit, vegetables and nuts" and guidance on quantifying the fruit, vegetable and nut content of a processed product - Peter Scarborough, Mike Rayner, Anna Boxer and Lynn Stockley - British Heart Foundation - Health Promotion Research Group, Department of Public Health, University of Oxford - December 2005.

Points	Valeur énergétique (kJ/100g)	Acides gras saturés (g/100g)	Sucres (g/100g)	Sodium (mg/100g)
0	≤335	≤1	≤4,5	≤90
1	>335	>1	>4,5	>90
2	>670	>2	>9	>180
3	>1005	>3	>13,5	>270
4	>1340	>4	>18	>360
5	>1675	>5	>22,5	>450
6	>2010	>6	>27	>540
7	>2345	>7	>31	>630
8	>2680	>8	>36	>720
9	>3015	>9	>40	>810
10	>3350	>10	>45	>900

FIGURE 3 – Table de Calcul des points attribués à chacun des nutriments de la composante dite “négative” selon la méthodologie développée par Rayner et al.

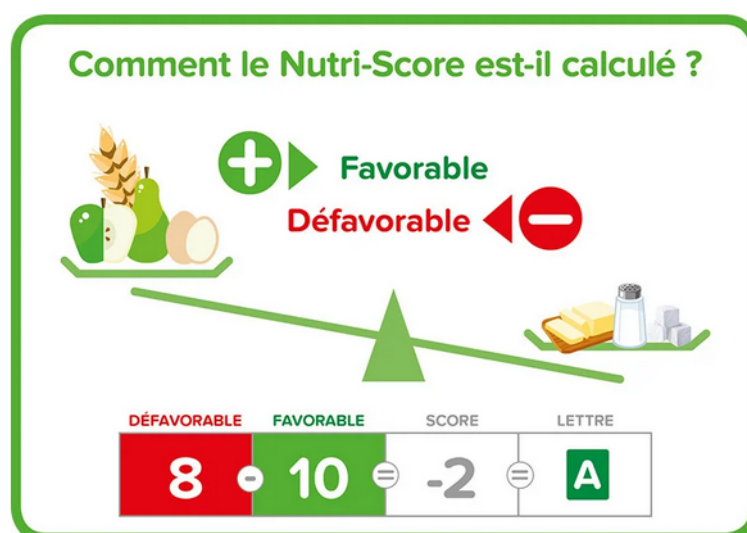


FIGURE 4 – Formule générale du Nutri-score

Le score nutritionnel final peut ainsi théoriquement varier de -15 à +40 en fonction des produits. **Plus il est faible, plus le produit est considéré comme ayant un profil nutritionnel favorable.**

L'affectation d'un produit à une classe, à partir de son score Nutri-score, se fait suivant le tableau de la Figure 6 où sont mentionnés les seuils des 5 catégories.

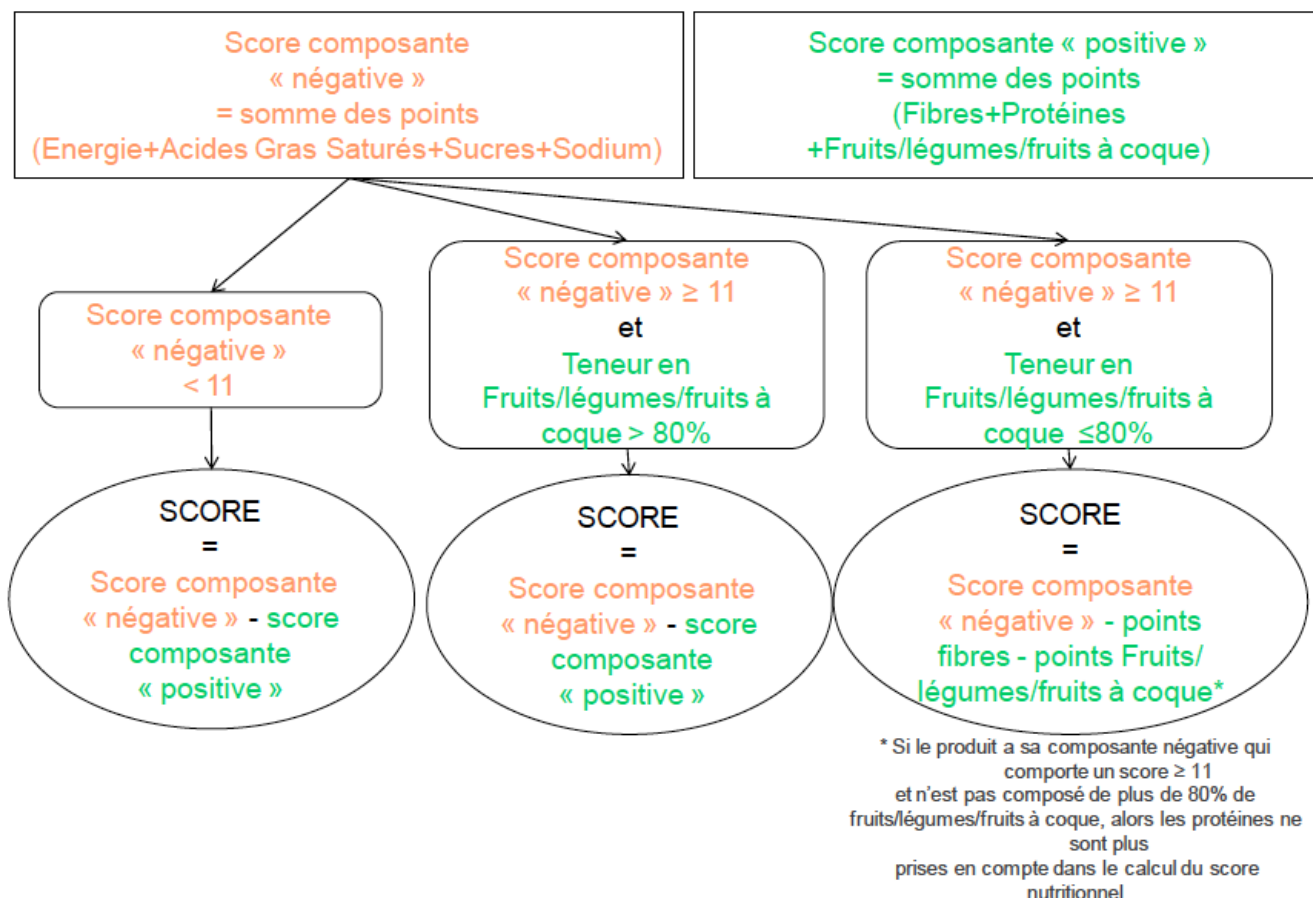


FIGURE 5 – Calcul du score nutritionnel selon la méthodologie développée par Rayner et al.

Le Nutri-Score est exprimé à l'aide d'une lettre allant de A à E

Aliments solides (points)	Boissons (points)	Nutri-score
-15 à -1	Eau	A B C D E
0 à 2	≤ 1	A B C D E
3 à 10	2 à 5	A B C D E
11 à 18	6 à 9	A B C D E
19 à 40	10 à 40	A B C D E

FIGURE 6 – Définition des 5 classes utilisées dans le cadre du Nutri-score

1.3 Exemples de calcul de Nutri-score

Ci-dessous, deux exemples de calcul du Nutri-score concernant le “Yaourt Skyr nature bio - Les 2 Vaches - 480 g” et le “Biscottes 6 Céréales - Heudebert - 300 g”. Ces données et calculs sont issus du site internet <https://fr.openfoodfacts.org>.

Informations nutritionnelles	Tel que vendu pour 100 g / 100 ml	
Énergie (kJ)	?	
Énergie (kcal)	57 kcal	
Énergie	238 kJ (57 kcal)	
Matières grasses / Lipides	0,2 g	
Acides gras saturés	0,1 g	
Glucides	3,9 g	
Sucres	3,9 g	
Protéines	10 g	
Sel	0,09 g	
Sodium	0,036 g	
Calcium	110 mg	
Fruits, légumes, noix et huiles de colza, noix et olive (estimation par analyse de la liste des ingrédients)	0 %	

Ce produit n'est pas considéré comme une boisson pour le calcul du Nutri-Score.

Points positifs : 5

- **Protéines : 5 / 5** (valeur : 10, valeur arrondie : 10)
- **Fibres alimentaires : 0 / 5** (valeur : 0, valeur arrondie : 0)
- **Fruits, légumes, noix et huiles de colza / noix / olive : 0 / 5** (valeur : 0, valeur arrondie : 0)

Points négatifs : 0

- **Énergie : 0 / 10** (valeur : 238, valeur arrondie : 238)
- **Sucres : 0 / 10** (valeur : 3,9, valeur arrondie : 3,9)
- **Graisses saturées : 0 / 10** (valeur : 0,1, valeur arrondie : 0,1)
- **Sodium : 0 / 10** (valeur : 36, valeur arrondie : 36)

Les points pour les protéines sont comptés car les points négatifs sont inférieurs à 11.

Score nutritionnel : -5 (0 - 5)

Nutri-Score : A

FIGURE 7 – Calcul du Nutri-score Yaourt Skyr nature bio - Les 2 Vaches - 480 g

Informations nutritionnelles	Tel que vendu pour 100 g / 100 ml	
Énergie (kJ)	1 722 kJ	
Énergie (kcal)	408 kcal	
Énergie	1 722 kJ (408 kcal)	
Matières grasses / Lipides	8 g	
Acides gras saturés	2,4 g	
Glucides	68 g	
Sucres	3,2 g	
Fibres alimentaires	5,2 g	
Protéines	14 g	
Sel	1,61 g	
Sodium	0,644 g	
Fruits, légumes, noix et huiles de colza, noix et olive (estimation par analyse de la liste des ingrédients)	0 %	

Ce produit n'est pas considéré comme une boisson pour le calcul du Nutri-Score.

Points positifs : 5

- **Protéines : 5 / 5** (valeur : 14, valeur arrondie : 14)
- **Fibres alimentaires : 5 / 5** (valeur : 5,2, valeur arrondie : 5,2)
- **Fruits, légumes, noix et huiles de colza / noix / olive : 0 / 5** (valeur : 0, valeur arrondie : 0)

Points négatifs : 14

- **Énergie : 5 / 10** (valeur : 1722, valeur arrondie : 1722)
- **Sucres : 0 / 10** (valeur : 3,2, valeur arrondie : 3,2)
- **Graisses saturées : 2 / 10** (valeur : 2,4, valeur arrondie : 2,4)
- **Sodium : 7 / 10** (valeur : 644, valeur arrondie : 644)

Les points pour les protéines ne sont pas comptés car les points négatifs sont supérieurs ou égaux à 11.

Score nutritionnel : 9 (14 - 5)

Nutri-Score : C

FIGURE 8 – Calcul du Nutri-score Biscottes 6 Céréales Heudebert - 300 g

1.4 Le Nutri-score vu comme un problème d’Aide Multicritères à la Décision

Dans la suite, nous considérerons le calcul du Nutri-Score comme un problème d’Aide à la Décision Multicritères où :

- L'ensemble X des alternatives correspond à un ensemble des produits à analyser.
- Les sept critères (constituant l'ensemble N) à prendre en compte pour une quantité de 100 g seront :

1. en : **La valeur énergétique** (Kcal/KJ) (critère à minimiser)
2. sa : **La quantité d'acides gras saturés** (g) (critère à minimiser)
3. su : **La quantité de sucres** (g) (critère à minimiser)
4. so : **La quantité de Sodium** (g/mg) (critère à minimiser)
5. pr : **La quantité de protéines** (g) (critère à maximiser)
6. fi : **La quantité de Fibres** (g) (critère à maximiser)
7. fr : **La teneur en Fruits/Légumes/Fruits à coques** (critère à maximiser)

2 Le système de feux tricolore

Un système de feux tricolores (rouge, orange, vert) permet de repérer d'un coup d'oeil la quantité de gras, de graisses saturés, de sucres et de sel contenue dans les produits alimentaires. Ce système de feux tricolores a été établi par les autorités sanitaires de Grande Bretagne (Food Standards Agency - FSA). Il est utilisé par certains fabricants de manière volontaire en Grande Bretagne, mais a été rejeté par la Commission Européenne en 2010.

Sur le site internet d' Open Food Facts, <https://fr.openfoodfacts.org/>, lorsque les valeurs nutritionnelles sont connues, les feux tricolores sont affichés sur les fiches produits.

La formule de calcul définissant les couleurs des feux est décrite par le Tableau suivant (voir Figure 9). Pour les boissons, toutes les valeurs sont divisées par 2.

pour 100g	 Faible Quantité	 Quantité modérée	 Quantité élevée
Lipides	jusqu'à 3g	de 3g à 20g	plus de 20g
Acides gras saturés	jusqu'à 1,5g	de 1,5g à 5g	plus de 5g
Sucres	jusqu'à 5g	de 5g à 12,5g	plus de 12,5g
Sel	jusqu'à 0,3g	de 0,3g à 1,5g	plus de 1,5g

FIGURE 9 – Processus de calcul des feux tricolores pour les aliments

3 Élaboration d'une base de données de produits alimentaires

Chaque étudiant élaborera une base de données, en format Excel, constituée d'au moins 100 aliments, de telle sorte que chaque catégorie du Nutri-score originel soit représentée par au moins 15 aliments. Le feu tricolore associé à chacun de ces produits devra être présent dans la base de données.

Le site internet de l'association OPEN FOOD FACT <https://fr.openfoodfacts.org/> peut vous aider à la constitution de la base de données.

4 Travail minimal à faire

4.1 Élaboration des modèles de Nutri-score basés sur l'approche Electre Tri

- Construire les fonctions `PessimisticmajoritySorting` et `OptimisticmajoritySorting` basées respectivement sur les procédures d'affectation Pessimiste et Optimiste de la méthode ELECTRE TRI appliquées au cas du Nutri-score. Ces fonctions renverront un fichier Excel contenant les produits alimentaires et les classes auxquelles ils ont été affectés.
- Le fichier Excel contenant les produits alimentaire sera une des entrées de vos fonctions.
- ★ Les poids associés aux sept critères du Nutri-Score, et utilisés dans un premier temps dans les tests, sont donnés par le Tableau 1.

	1- Énergie	2-Acide Gras sat.	3-Sucre	4- Sodium	5- Protéine	6- Fibre	7- Teneur en fruits. . .
Poids w_i	2	2	2	2	1	1	1

TABLE 1 – Des poids associés au critère du Nutri-Score

- ★ Dans un deuxième temps, chaque étudiant définira, en le justifiant, pour ses propres tests, un autre jeu de poids.
- Chaque étudiant proposera une méthode d'identification des 6 profils limites données par le Tableau 2. Les profils construits devront être justifiés. Ils seront dédiés à une application d'ELECTRE TRI à la base de données élaborée précédemment par vos soins. Ne pas oublier que l'idée est d'avoir, au final, une démarche transparente.
- Vous testerez vos fonctions, au moins avec les trois seuils majoritaires $\lambda = 0.5$, $\lambda = 0.6$ et $\lambda = 0.7$.
- En fait, les deux fonctions d'ELECTRE TRI implémentées vous permettent d'obtenir vos propres modèles de classification des produits alimentaires.
- Déterminer toutes les affectations obtenues avec votre base de données. Comparer vos résultats avec les affectations données par le Nutri-score et la classification des feux tricolores (en utilisant par exemple une matrice de confusion, proportion d'éléments appartenant à une classe, tableaux de bord, . . .).

4.2 Autres questions

- Quelle analyse faites-vous de la transparence de vos modèles élaborés ci-dessus, à travers les cinq propriétés vues en cours (loyauté, équité, responsabilité, explicabilité, intelligibilité).
- D'autres modèles de décision, comme les arbres de décision, peuvent-ils être plus transparents que les modèles étudiés ci-dessus ? Pourquoi ?
- Toute initiative supplémentaire pour la réalisation du projet est encouragée et sera appréciée.

π^6 : Borne supérieure des profils limites
 π^5 : Profil limite entre A et B
 π^4 : Profil limite entre B et C
 π^3 : Profil limite entre C et D
 π^2 : Profil limite entre D et E
 π^1 : Borne inférieure des profils

TABLE 2 – Profils à déterminer

4.3 Date limite d'envoi

Rapport écrit (à déposer dans le canal Rattrapage 2022-2023 de l'équipe Teams du cours, sur la plate-forme Teams), des fichiers Excel et fichiers .py soigneusement commentés : **Lundi 11 septembre 2023 à 23h59**. Tout retard sera sanctionné.

5 Description de la méthode ELECTRE TRI

La méthode ELECTRE TRI. Cette dernière est une approche d'Aide MultiCritère à la Décision qui vise à résoudre des problèmes d'affectation (classification) d'objets dans des catégories prédéfinies.

Dans le cadre du Nutri-score, nous retenons ici comme catégories ou classes, les 5 catégories du Nutri-Score : 'A' (catégorie C_5), 'B' (catégorie C_4), 'C' (C_3), 'D' (catégorie C_2) et 'E' (C_1). La méthode ELECTRE-TRI consiste alors à affecter chaque aliment à une de ces 5 catégories.

On va également supposer que chaque catégorie C_i est délimitée par une frontière supérieure notée b_{i+1} et une frontière inférieure b_i . Les frontières b_{i+1} et b_i sont appelées "profils" et représentent des aliments de référence qui peuvent être fictives. Il y a une dominance paréto-strictte entre b_{i+1} et b_i . Nous aurons donc :

- C_1 délimitée par b_2 et b_1 ;
- C_2 délimitée par b_3 et b_2 ;
- C_3 délimitée par b_4 et b_3 .
- C_4 délimitée par b_5 et b_4 .
- C_5 délimitée par b_6 et b_5 .

Ainsi, comme l'affectation se fait dans 5 catégories distinctes, le profil b_3 représente la frontière entre les classes "A" et "B", et b_2 la frontière entre les classes "B" et "C".

Le principe de la méthode ELECTRE TRI consiste, non pas à comparer les aliments entre eux, mais à les comparer aux six aliments de référence b_6, b_5, b_4, b_3, b_2 et b_1 dont les évaluations sur chaque critère seront à déterminer. **Attention : dans ce tableau, les profils b_6 et b_1 doivent toujours avoir des valeurs extrêmes, ne pouvant jamais être atteintes, sur chaque critère.**

Ainsi, l'affectation d'un aliment à une catégorie dépendra de sa comparaison aux profils b_6, b_5, b_4, b_3, b_2 et b_1 . Plus formellement, l'affectation d'un aliment dans les catégories se base sur le concept de sur-classement. On dira qu'un aliment H surclasse le profil b_i (respectivement le profil b_i surclasse l'aliment H) et on note $H \mathcal{S} b_i$ (respectivement $b_i \mathcal{S} H$) si H est au moins aussi bon que b_i (respectivement b_i est au moins aussi bon que H) sur une majorité de les critères (la majorité étant définie par un seuil de majorité λ). Les étapes de la méthode ELECTRE TRI se décrivent comme suit :

Étape 1 : Détermination des indices de concordance partiels

Pour chaque critère j , l'indice de concordance partiel entre l'aliment H et le profil b_i est donné par :

- Si la fonction g_j est à maximiser :

$$c_j(H, b_i) = \begin{cases} 1 & \text{si } g_j(H) \geq g_j(b_i) \\ 0 & \text{sinon} \end{cases} \quad c_j(b_i, H) = \begin{cases} 1 & \text{si } g_j(b_i) \geq g_j(H) \\ 0 & \text{sinon} \end{cases}$$

- Si la fonction g_j est à minimiser :

$$c_j(H, b_i) = \begin{cases} 1 & \text{si } g_j(b_i) \geq g_j(H) \\ 0 & \text{sinon} \end{cases} \quad c_j(b_i, H) = \begin{cases} 1 & \text{si } g_j(H) \geq g_j(b_i) \\ 0 & \text{sinon} \end{cases}$$

avec $g_j(H)$ et $g_j(b_i)$ représentant respectivement le score de H et b_i sur le critère j . Dans notre cas, ce score représente l'évaluation qualitative attribuée à H et b_i . Par conséquent, la comparaison $g_j(H) \geq g_j(b_i)$ signifie simplement que la valeur qualitative de H est au moins aussi bonne que celle de b_i .

Étape 2 : Détermination des indices de concordance globaux

L'indice de concordance global entre l'aliment H et le profil b_i est donné par la formule suivante :

$$C(H, b_i) = \frac{\sum_{j=1}^n k_j c_j(H, b_i)}{\sum_{j=1}^n k_j} \quad C(b_i, H) = \frac{\sum_{j=1}^n k_j c_j(b_i, H)}{\sum_{j=1}^n k_j}$$

où k_j est le poids du critère j et n le nombre de critères.

Étape 3 : Détermination de la relation de surclassement \mathcal{S}

La relation de surclassement se définit à l'aide de l'indice de coupe λ , appelé seuil de majorité (en général supérieur à 50%), qui représente le paramètre déterminant la situation de préférence entre l'aliment H et le profil b_i . Ainsi pour l'aliment H et un profil b_i :

- H surclasse b_i et on notera $H \mathcal{S} b_i$ si et seulement si $C(H, b_i) \geq \lambda$.
- b_i surclasse H et on notera $b_i \mathcal{S} H$ si et seulement si $C(b_i, H) \geq \lambda$.

Étape 4 : Procédures d'affectation

Supposons qu'on dispose de r catégories C_1, \dots, C_r où chaque catégorie C_i est délimitées par deux profils b_i et b_{i+1} .

Deux procédures d'affectation de l'aliment H sont possibles :

- **Procédure pessimiste** : pour chaque aliment H , faire décroître les indices des profils de r jusqu'au premier indice k tel que $H \mathcal{S} b_k$. L'aliment H est alors affecté à la catégorie C_k .

Dans notre exemple, cette procédure revient à comparer successivement H à b_6, b_5, b_4, b_3, b_2 et b_1 . Si H surclasse b_i , i.e., $H \mathcal{S} b_i$, alors H est affecté à la catégorie C_i .

- **Procédure optimiste** : pour chaque aliment H , faire croître les indices des profils de 1 jusqu'au premier indice k tel que $b_k \mathcal{S} H$ et $\text{non}(H \mathcal{S} b_k)$. L'aliment H est alors affecté à la catégorie C_{k-1} .

Dans notre exemple, cette procédure revient à comparer successivement b_1, b_2, b_3, b_4, b_5 et b_6 à H . Si $b_i \mathcal{S} H$ et $\text{non}(H \mathcal{S} b_i)$ alors H est affecté à la catégorie C_{i-1} .