

L'ALGORITHME DE CALCUL DU NUTRI-SCORE  
DES PRODUITS ALIMENTAIRES EST-IL TRANSPARENT ?

PARTIE 2

Année académique 2019/2020

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 du Nutri-score, un système 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 vu comme un problème d'Aide Multicritères à la Décision

Le calcul du Nutri-Score comme un problème d'Aide à la Décision Multicritères où :

- L'ensemble  $X$  des alternatives correspond à l'ensemble des produits à analyser.
- Les six critères (constituant l'ensemble  $N$ ) à prendre en compte pour une quantité de 100 g seront :
  1. **La valeur énergétique** (Kcal/KJ) (critère à minimiser)
  2. **La quantité d'acides gras saturés** (g) (critère à minimiser)
  3. **La quantité de sucres** (g) (critère à minimiser)
  4. **La quantité de Sodium** (g/mg) (critère à maximiser)
  5. **La quantité de protéines** (g) (critère à maximiser)
  6. **La quantité de Fibres** (g) (critère à maximiser)

## 2 Le Nutri-Score expliqué par une démarche de classification multicritère : la méthode ELECTRE TRI

Dans cette partie, au lieu d'attribuer une note à chaque aliment, nous allons plutôt l'affecter à une classe à l'aide d'une approche ne nécessitant pas une normalisation des critères : 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.

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-stricte 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  $\lambda$ ). 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  $\lambda$ , 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  $C_{i-1}$ .

## 3 Travail minimum à faire

- Le Nutri-score peut-il être expliqué par une ou plusieurs fonctions d'utilités additive ?
- Le Nutri-score peut-il être expliqué par un ou plusieurs modèles de classification de type ELECTRE TRI (avec par exemple des profils fixés, des profils déterminés par programmation linéaire, ... ) ?
- D'autres modèles de décision peuvent-ils expliquer le Nutri-Score d'un aliment (exemple : arbre de décision, ... ) ?
- Faire une analyse minutieuse de chaque modèle construit, ainsi qu'une étude comparative de ces modèles. Ces analyses et/ou études pourront, par exemple, s'appuyer sur des éléments suivants :
  - ★ La validation des modèles obtenus à l'aide de nouveaux aliments (taux d'erreurs, ... )
  - ★ La présentation et l'explication claire des modèles construits à un consommateur quelconque est-il possible (graphiques expliquant les fonctions d'utilités, ... ) ?
  - ★ Existe-t-il un minimum de connaissances (académique, non académique, ... ) requis pour une compréhension du Nutri-Score ?
- Toute initiative supplémentaire pour la réalisation du projet est encouragée et sera appréciée.
- **Date limite d'envoi du rapport et des fichiers : Dimanche 8 Décembre 2019 à 23h59**. Tout retard sera sanctionné.
- La soutenance des projets (devant toute la promotion), d'une quinzaine de minutes par groupe, est prévue le **Vendredi 29 Novembre 2019 à 8h30**.