



## **PROJET :**

### Calcul des informations nutritionnelles de produits alimentaires

Dans le cadre du cours : Transparence des algorithmes

Encadrée par : Brice MAYAG

Elèves :

Riccardo **FIGLIOZZI**

Giovanni Ivan **INDIVERI**

Chloé **PATRAS**

## Introduction

Pour ce projet nous avons utilisé trois bases de données différentes (BDD1, BDD2 et BDD3). Dans la première partie on décrit les trois bases des données qu'on a utilisées pour l'analyse des différentes méthodes. Ensuite nous présenterons pour chaque indicateur les différents résultats obtenus et après nous le comparerons entre eux. Finalement, nous évaluerons la transparence des indicateurs considérés et l'approche avec une méthodologie différente.

# SOMMAIRE

<b>Création des bases de données et paramètres</b>	<b>4</b>
1.1 Création BDD1	4
1.2 Création BDD2	5
1.3 Création BDD3	7
<b>Implémentation de l'algorithme ELECTRE TRI</b>	<b>8</b>
2.1 Résultats pour la BDD1	8
2.2 Résultats pour la BDD2	9
2.3 Résultats pour la BDD3	10
<b>Analyse des autres méthodes</b>	<b>12</b>
3.1 Nutri-Score	12
3.2 Classification Nova	13
3.3 Application YUKA	14
3.4 Feux Tricolore	15
<b>Comparaison des méthodes</b>	<b>16</b>
4.1 Premier rapport : BDD1	16
4.2 Second rapport : BDD2	17
4.3 Troisième rapport : BDD3	19
<b>Etude de la transparence</b>	<b>20</b>
5.1 Loyauté	20
5.2 Équité	20
5.3 Responsabilité	21
5.4 Explicabilité	22
5.5 Intelligibilité	23
5.6 Vue d'ensemble	23
<b>Approche par le Machine Learning</b>	<b>24</b>
6.1 Arbre de décision	24
6.2 Réseau de neurones	24
6.3 Clustering	24
<b>Conclusion</b>	<b>24</b>
<b>Annexes</b>	<b>25</b>

# 1. Création des bases de données et paramètres

## 1.1 Création BDD1

Ce premier jeu de données "OpenFood\_Petales" est donné par le professeur encadrant. Il est composé de 320 produits céréales du petit-déjeuner. On y retrouve le score Nutri-Score.

Pour ce fichier, les poids et les profils ont été donnés également dans le sujet. Nous avons donc réalisé nos analyses sur ces informations.

Critères	Energy	Saturated Fat	Sugars	Fiber	Proteins	Sodium
Poids	1	1	1	2	2	1
Profil 1	100	0	0	100	100	0
Profil 2	1550	11	0.8	11	10	0.3
Profil 3	1650	14	1	8	7	0.4
Profil 4	1750	17	1.7	5	4	0.5
Profil 5	1850	20	4	2.5	3	0.6
Profil 6	10000	100	100	0	0	100

## 1.2 Création BDD2

Pour la création de la base des données BDD2 nous avons pris les données du jeu de données (Open Facts food data simplified) fournis par le professeur. Ici nous avons choisi les différentes catégories des produits solides avec 1500 produits de telle sorte que chaque catégorie du Nutri-score soit représentée par au moins 150 produits et aussi pour la classe NOVA de chaque aliment.

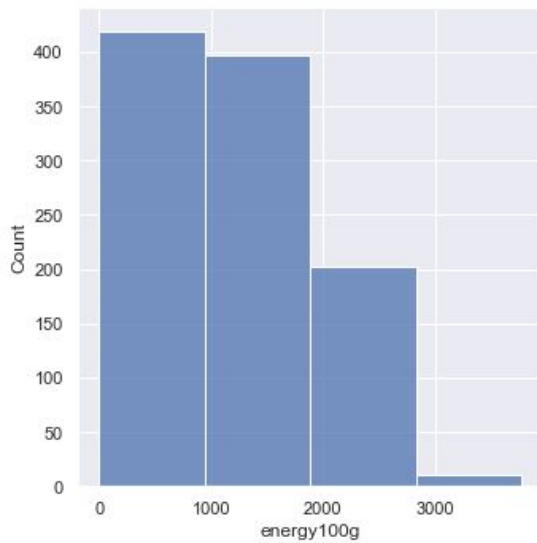
Après nous avons nettoyé nos données parce qu'il y avait des champs non renseignés puis nous avons choisi aléatoirement les aliments entre les catégories des produits solides. Ensuite on a ajusté les poids en regardant une publication du WHO (World Health Organization)<sup>1</sup>, nous avons choisi les valeurs suivantes pour les attributs :

Critères	Energy	Saturated Fat	Sugars	Fiber	Proteins	Sodium
Poids	1.5	1	1	2	2	1

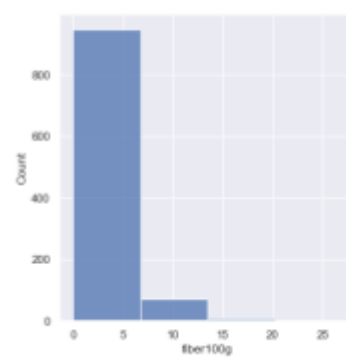
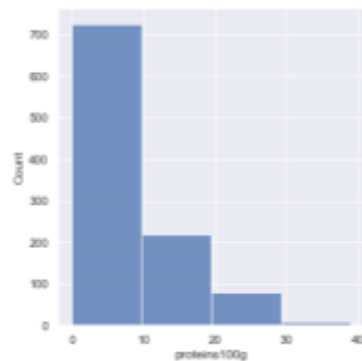
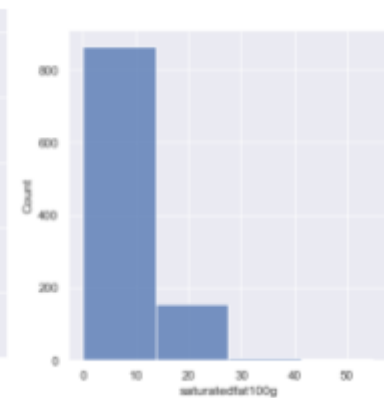
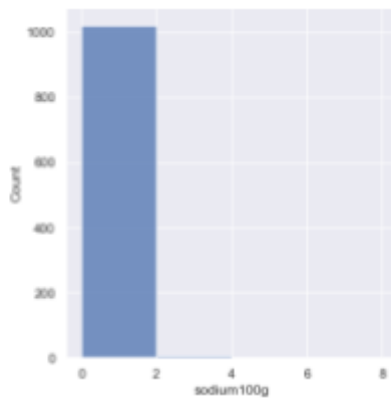
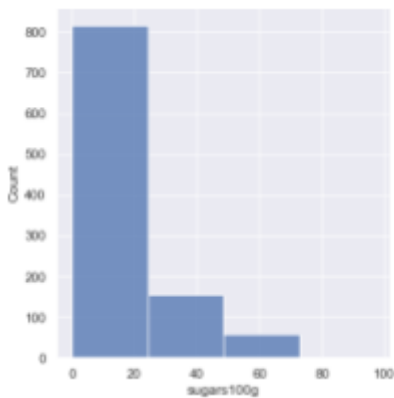
---

<sup>1</sup> <https://www.who.int/news-room/fact-sheets/detail/healthy-diet>

Puis on a choisi les différents profils pour chaque groupe. Pour établir ces profils nous avons décidé de regarder la distribution de chacune des caractéristiques et d'en sortir cinq intervalles.



Par exemple pour établir les profils de l'attribut energy\_100g, nous avons choisi ici de



Ci-dessus, la représentation de la distribution pour chacune des autres caractéristiques.

Nous obtenons les profils suivants :

Critères	Energy	Saturated Fat	Sugars	Fiber	Proteins	Sodium
b5	0	0	0	100	100	0
b4	900	13	22	2	29	20
b3	1800	27	48	4	19	14
b2	2700	41	70	6	10	6
b1	10000	100	100	50	0	0

### 1.3 Création BDD3

La troisième base de données alimentaire doit contenir 60 produits de telle sorte que le nutri-score soit représenté par au moins 10 produits de chaque catégorie. Il sera renseigné pour chacun des produits, la classe YUKA et la classe NOVA. De plus, il est indiqué le nombre d'additifs que possède un produit et un label bio ou non.

Nous avons donc décidé pour cette base de donnée de prendre en compte les additifs. En effet, d'après nos recherches, l'algorithme YUKA prend en compte cette information pour son choix de catégorie. Nous avons donc pensé qu'il serait préférable pour notre algorithme ELECTRE TRI d'en faire de même pour cette comparaison. De plus, nous avons spécifié si le produit possède un label bio. Cette caractéristique nous servira notamment dans l'analyse des résultats sur le modèle Power BI.

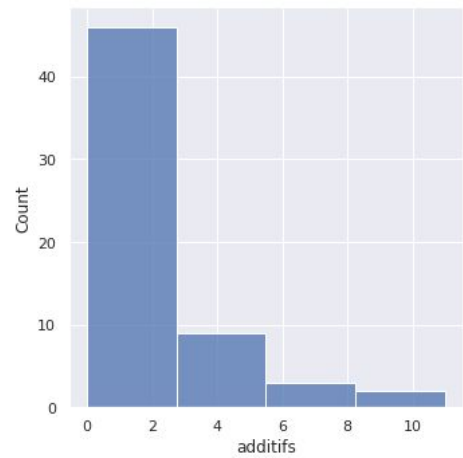
Pour déterminer les poids de ce nouveau critère, nous nous sommes appuyé également sur les recommandations de l'OMS<sup>2</sup> sur :

Critères	Energy	Saturate d Fat	Sugars	Fiber	Proteins	Sodium	Additif
Poids	1.5	1	1	2	2	1	1.5

<sup>2</sup> <https://www.who.int/news-room/fact-sheets/detail/food-additives>

Nous avons décidé de garder les profils établis dans la base de données 2 car la base de données 3 est composée seulement de 60 produits donc il n'y a pas assez de données pour établir un résultat fiable statistiquement.

Afin de connaître les profils des additifs nous avons procédé comme précédemment, c'est-à-dire en observant la distribution de l'additif.



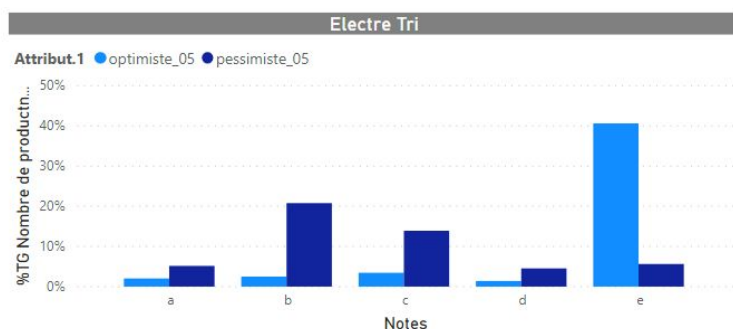
Critères	Energy	Saturated Fat	Sugars	Fiber	Proteins	Sodium	Additif
b5	0	0	0	100	100	0	50
b4	900	13	22	2	29	20	8
b3	1800	27	48	4	19	14	5
b2	2700	41	70	6	10	6	2
b1	10000	100	100	50	0	0	0

## 2.Implémentation de l'algorithme ELECTRE TRI

Nous avons implémenté la méthode optimiste et pessimiste d'ELECTRE TRI sur python. On a utilisé les différents lambda suggérés dans le sujet afin d'obtenir nos résultats. Dans cette partie, nous présentons un aperçu de la distribution des méthodes optimiste et pessimiste pour chacun des lambda par base de données.

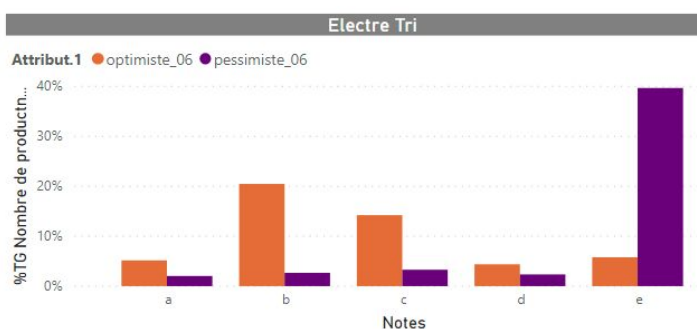
### 2.1 Résultats pour la BDD1

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.5** :



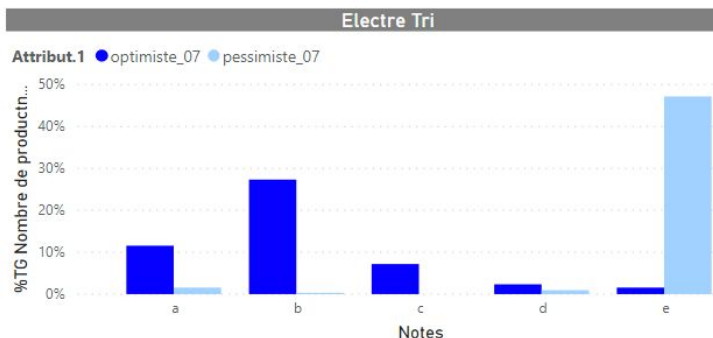
Commentaire : Nous pouvons constater que la classification optimiste attribue des notes majoritairement négatives dans le cas où il y a un lambda égale à 0.5. En revanche, la pessimiste attribue davantage de note B mais reste assez répartie dans l'ensemble des notes.

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.6** :



Commentaire : Dans le cas où lambda est égale à 0.6, la méthode pessimiste classifie de manière plus sévère les produits. On peut voir qu'il attribue une grande partie des notes à la catégorie e. On peut remarquer que la distribution est inversée comparée à celle où le lambda est égale à 0.5.

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.7** :

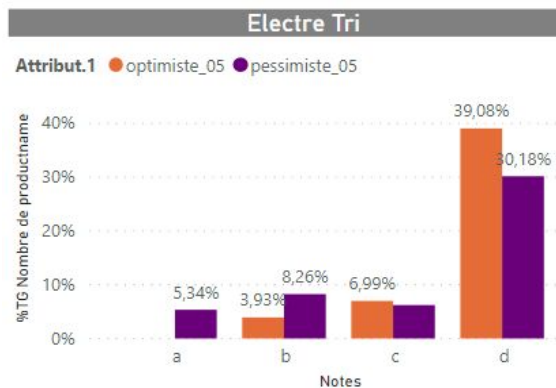


Commentaire : On observe que la méthode optimiste avec un lambda 0.7 ressemble à la précédente mais elle attribue majoritaire des notes a et b. En revanche, la classification pessimiste de ELECTRE TRI donne la note la plus négative la plupart du temps.



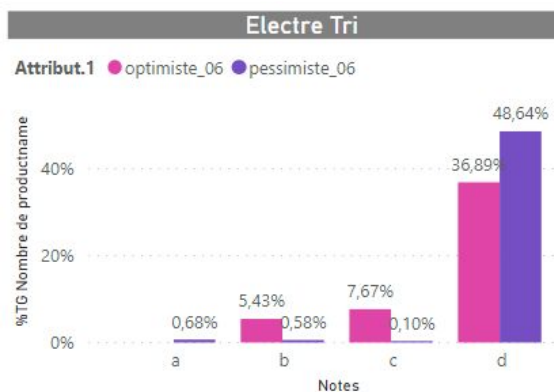
## 2.2 Résultats pour la BDD2

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.5** :



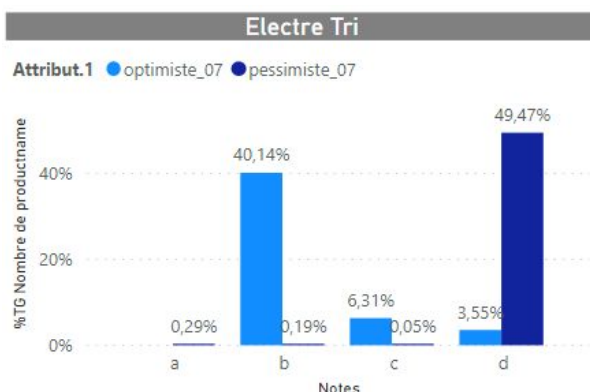
Commentaire : On constate que la classification optimiste attribue des notes majoritairement négatives dans le cas où il y a un lambda égale à 0.5. De même, la pessimiste attribue davantage de note D mais avec une proportion plus modérée que la classification optimiste.

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.6** :



Commentaire : Dans le cas où lambda est égale à 0.6, la méthode pessimiste classifie de manière plus sévère les produits. On peut voir qu'il attribue une grande partie des notes à la catégorie D. On constate que la distribution est inversée comparée à celle où le lambda est égale à 0.5 pour la note de E.

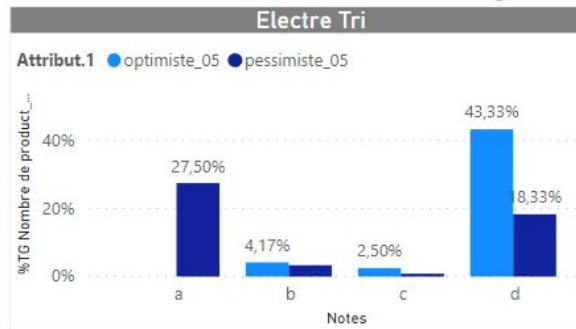
Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.7** :



Commentaire : La méthode optimiste avec un lambda 0.7 attribue majoritaire des notes de catégorie B. En revanche, la classification pessimiste de ELECTRE TRI donne la note la plus négative la plupart du temps, c'est à dire dans 49,47% des cas sur 50% (en effet au vue des 2 méthodes, chacun possède 50% des parts des votes).

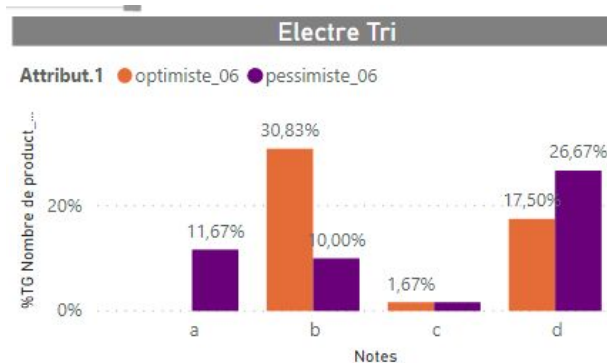
## 2.3 Résultats pour la BDD3

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.5** :



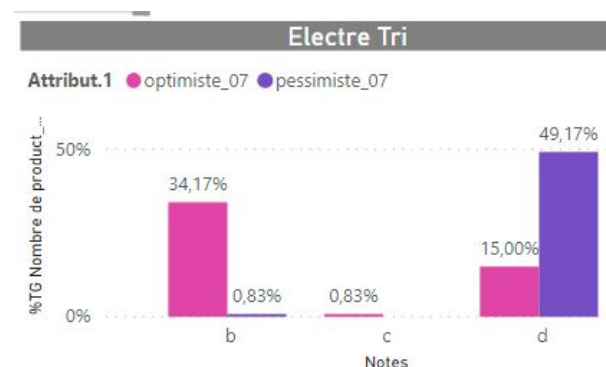
Commentaire : La classification optimiste attribue des notes majoritairement négatives dans le cas où il y a un lambda égale à 0.5. En revanche, la méthode pessimiste attribue davantage de note A mais avec une bonne partie de note D.

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.6** :



Commentaire : Quand le lambda est égale à 0.6, la méthode pessimiste classifie de manière plus sévère les produits. Cependant la différence n'est pas si éloigné de la méthode optimiste qui attribue majoritairement des notes de B ou D. On constate que la pessimiste accorde des bonnes notes également.

Ci dessous, la distribution de la classification de la méthode Electre tri avec un lambda égale à **0.7** :



Commentaire : On observe que la méthode optimiste avec un lambda 0.7 à accentuer la tendance du lambda 0.6, c'est-à-dire que les notes attribuées sont majoritairement B ou D. En revanche, la classification pessimiste de ELECTRE TRI donne la note la plus négative en grande partie.

## 2.4 Conclusion ELECTRE TRI

En comparant les résultats précédents, nous observons que l'attribution des notes est très corrélée au choix du  $\lambda$ . En effet, nous constatons que la distribution des bases de données sont assez semblables quand le  $\lambda$  est identique.

On constate que si  $\lambda$  est égale à 0.5, la méthode optimiste attribue majoritairement plus de mauvaise note que la méthode pessimiste et inversement pour un  $\lambda$  égale à 0.7. Les deux méthodes sont plus modérées pour un  $\lambda$  égale à 0.6.

### 3. Analyse des autres méthodes

Nous allons présenter d'autres méthodes de catégorisation de produit, puis montrer les résultats sur nos différentes bases de données.

#### 3.1 Nutri-Score

Le Nutri-Score est un logo de 5 couleurs présent sur les emballages recommandé par le gouvernement français. Avec une lettre et une couleur, il informe les consommateurs sur la qualité nutritionnelle d'un produit.

L'échelle possède 5 niveaux allant :

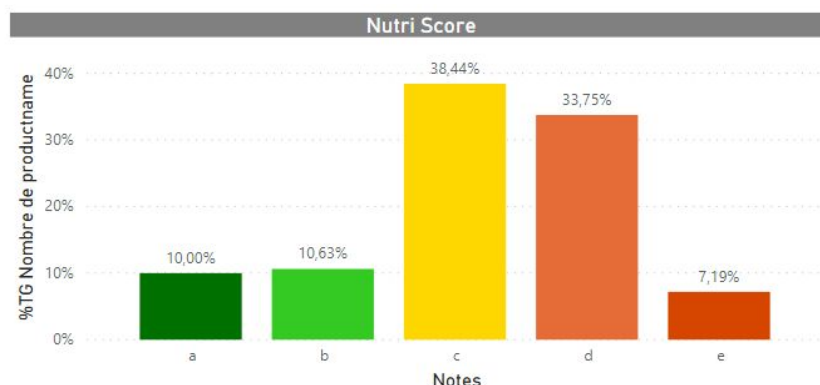
- du produit le plus favorable sur le plan nutritionnel (classé A)
- au produit le moins favorable sur le plan nutritionnel (classé E)

Les caractéristiques sur 100g d'un produit utilisé pour la notation Nutri score sont les suivants : l'énergie, le sucre simple, l'acide gras saturés, le sodium, les fibres et les protéines.

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

#### BDD1 :

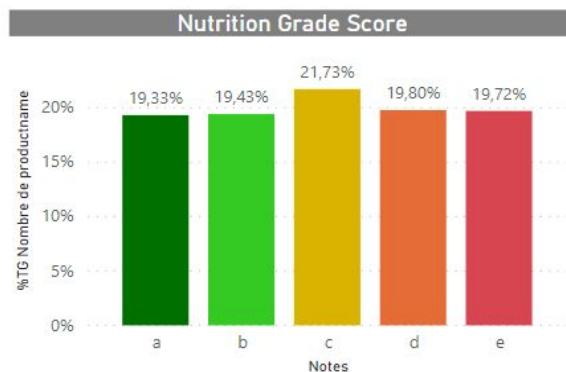
Ci dessous la distribution du Nutri-Score sur la BDD1 :



Commentaire : Nous remarquons que la distribution des produits sur les catégories a, b et e sont assez équivalentes (entre 7 et 11%). En revanche, il y a plus de produits dans les classes c et d, qui avoisine un pourcentage de 34% à 39%.

## BDD2 :

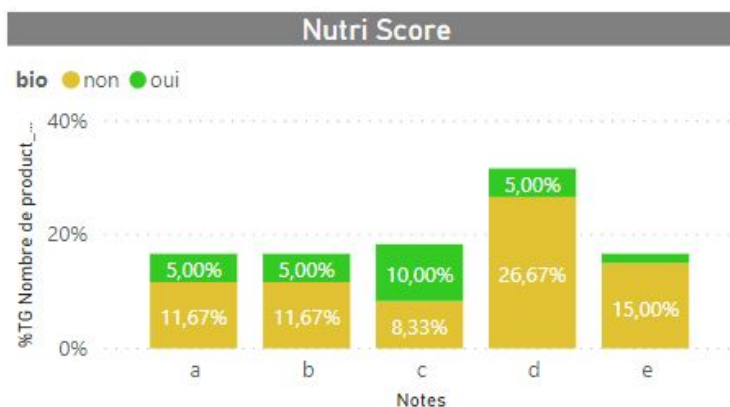
Ci dessous la distribution du Nutri-Score sur la BDD2 :



Commentaire : Pour constituer la base de données 2, nous avons dû prendre au moins 150 produits pour chaque catégorie, c'est pour cela que la répartition est plutôt équivalente entre les notes.

## BDD3 :

Ci dessous la distribution du Nutri-Score sur la BDD3 :



Commentaire : Nous avons considéré la présence de produits bio pour Nutri-Score, afin de faire une comparaison avec Yuka. Il y a une distribution assez équitables des notes avec une attribution un peu plus forte pour la note de D. On constate que le NutriScore n'attribue pas forcément les produits bio à des notes particulières, même s' ils sont légèrement moins présents dans la catégorie D.

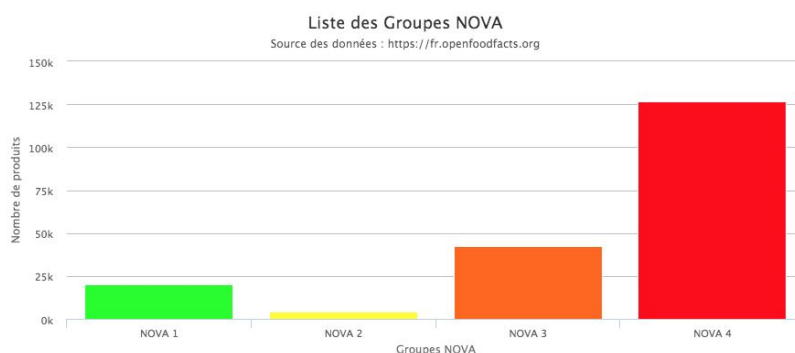
## 3.2 Classification Nova

La classification NOVA classe les aliments selon leur degré de transformation. De ce fait, elle permet de mieux maîtriser les problèmes de santé chroniques pouvant être liés à la prise de poids (obésité, diabète de type 2, maladies cardio-vasculaires...) en contrôlant les aliments de nos produits.



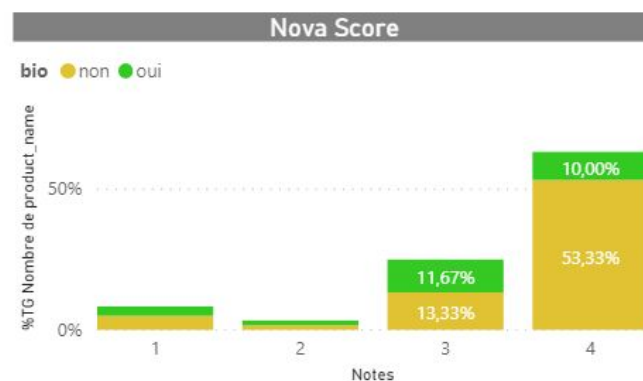
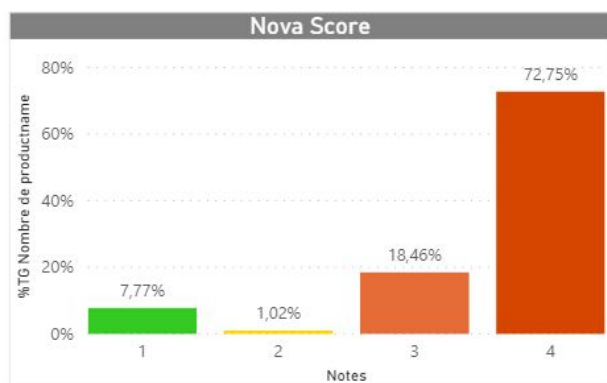
**Remarque :** Très peu d'aliments sont classés "2" avec la classification NOVA. Nous pouvons constater ceci grâce à une distribution des produits répertoriés sur le site OpenFoodFact (graphique du site ci-dessous). De plus, la plupart des produits de cette classe sont des produits d'épicerie (sel, huiles, ...)

Dans nos données de la base de données 3, il est donc normal qu'il n'y ait que très peu de classe "2" notamment car YUKA calcul peu de score pour les produits "épicerie".



### BDD2 et BDD3:

Ci dessous la distribution du Nova sur la BDD2 et BDD3 :



**Commentaire :** Comme on peut le voir ci-dessus la distribution pour la classification NOVA dans les BDD2 et BDD3 est la même que la distribution affichée sur le site de OpenFoodFact. Une grande partie des produits a pour note attribuée "4". De plus, peu de produits obtiennent la note de "2" pour les raisons expliquées précédemment.

### 3.3 Application YUKA

Pour utiliser l'application, il faut lire le code-barres d'un produit avec son téléphone.

L'application accède ainsi aux détails de la composition du produit et retourne une note sous forme de couleur allant du vert au rouge.

- **Excellent**
- **Bon**
- **Médiocre**
- **Mauvais**

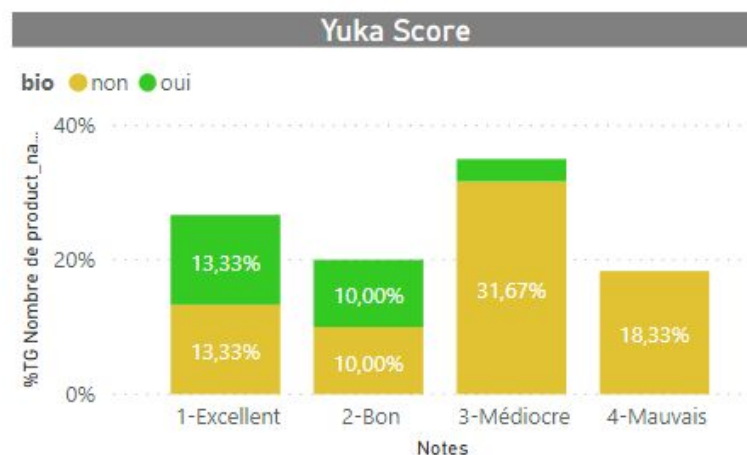
Pour la notation des produits alimentaires, Yuka se base sur trois critères :

- La qualité nutritionnelle représente 60% de la note.
- La présence d'additifs représente 30% de la note.
- La dimension biologique représente 10% de la note.



#### BDD3 :

Afin d'obtenir le classement Yuka, nous avons scanné 60 produits de type alimentaire solide.



Commentaire : Si l'on prend toute notre base de donnée en considération (bio et non bio) la distribution est assez répartie, avec une faible majorité pour la note "médiocre". Cependant, comme nous l'avons souligné précédemment, Yuka considère la présence de produits bio, ici on peut constater que les produits bio sont classifiés avec une meilleure note. Ce graphique nous permet de réaliser l'impact que possède les labels bio sur un produit pour la classification YUKA.

### 3.4 Feux Tricolore

Le système Feux tricolore permet d'étiqueter chaque nutriment dans l'une des trois couleurs en fonction de sa teneur dans l'aliment pour 100 g.

Des seuils sont définis pour chacun des quatre nutriments afin de les classer en trois couleurs : rouge (high), orange (medium) et vert (low).

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

#### BDD3 :

Ci dessous les résultats obtenus sur la BDD3 :

product\_name

- ☐ Abbracci mulino bianco
- ☐ Amande a tartiner Bjorg Bio
- ☐ Baiocchi - Mulino Bianco
- ☒ Beurre Moulé demi-sel - Paysan Br...
- ☐ Biscuit langue de chat Franprix
- ☐ boudoir bossard
- ☐ Boulettes de boeuf bolognaise Big...
- ☐ Boulgour courgettes cranberries ...

#### Feux Tricolore

feu\_tri

Attribut

	lipides_100g
	saturated-fat_100g
	sel_100g
	sugars_100g

Commentaire : Dans ce cas on affiche la classe de chaque attribut pour le produit sélectionné. Nous avons respecté



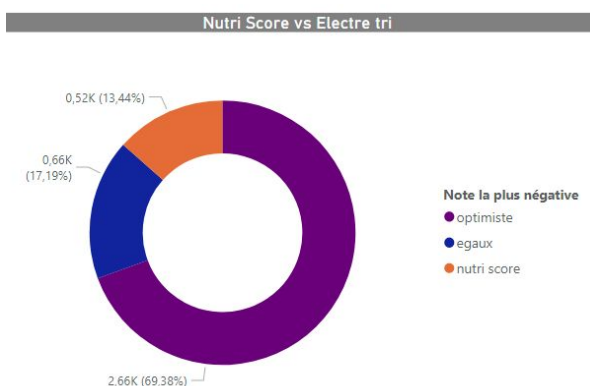
## 4. Comparaison des méthodes

Afin de rendre la comparaison plus dynamique, nous avons produit un rapport PowerBI qui permet de rendre interactif nos choix de comparaison.

### 4.1 Premier rapport : BDD1

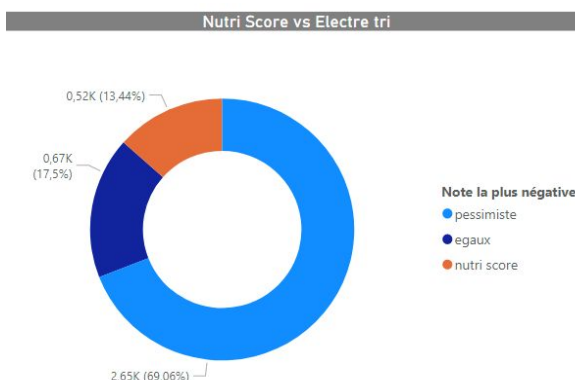
Pour la base de données "Pétales", nous avons les résultats de classification du Nutri-Score et de l'Electre tri (pessimiste et optimiste). Il sera présenté dans le graphique circulaire la proportion de chacun la proportion de la classification qui note le plus sévèrement pour chaque méthode. La catégorie "égaux" représente le cas où au moins deux méthodes attribuent la même note étant la plus sévère.

Ci dessous, le résultat pour un lambda égale à **0.5** :



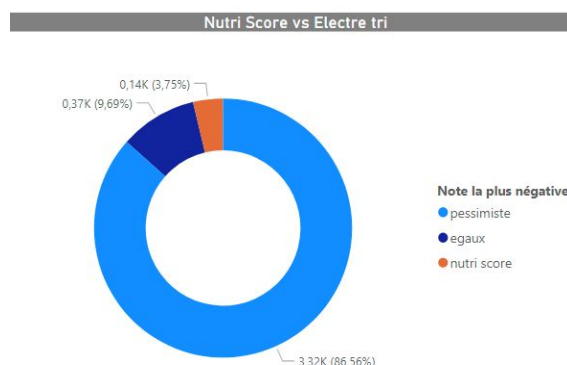
Commentaire : Nous constatons que la classification "optimiste" est dans quasiment 70% des cas la plus sévère. Le nutri-score attribue la plus mauvaise note dans 13% des cas.

Ci dessous, le résultat pour un lambda égale à **0.6** :



Commentaire : Si nous regardons le graphique précédent, nous remarquons que le ratio pessimiste et optimiste est inversé.

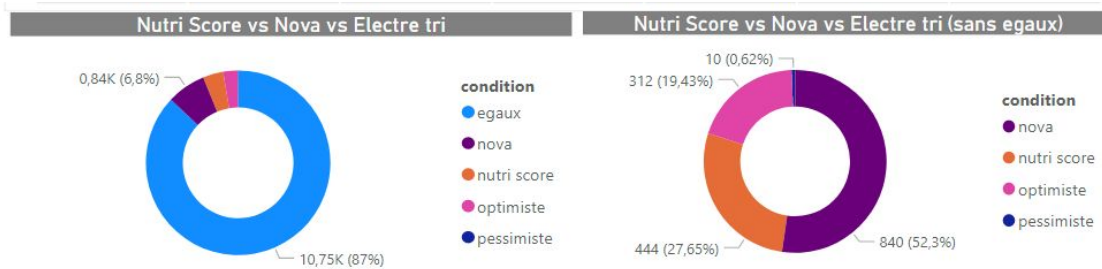
Ci-dessous, le résultat pour un lambda égale à **0.7** :



Commentaire : Pour cette dernière comparaison, on observe une forte proportion de note plus sévère pour la méthode du Electre tri pessimiste avec une faible part pour le nutri-score.

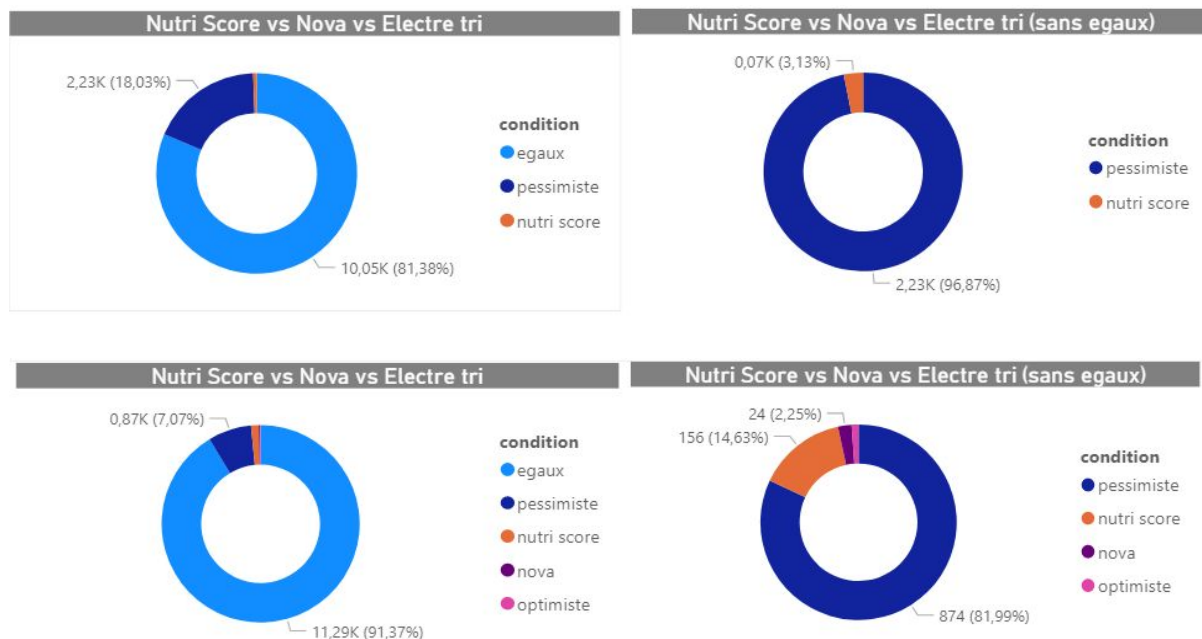
## 4.2 Second rapport : BDD2

Ci dessous, le résultat pour un lambda égale à **0.5** :



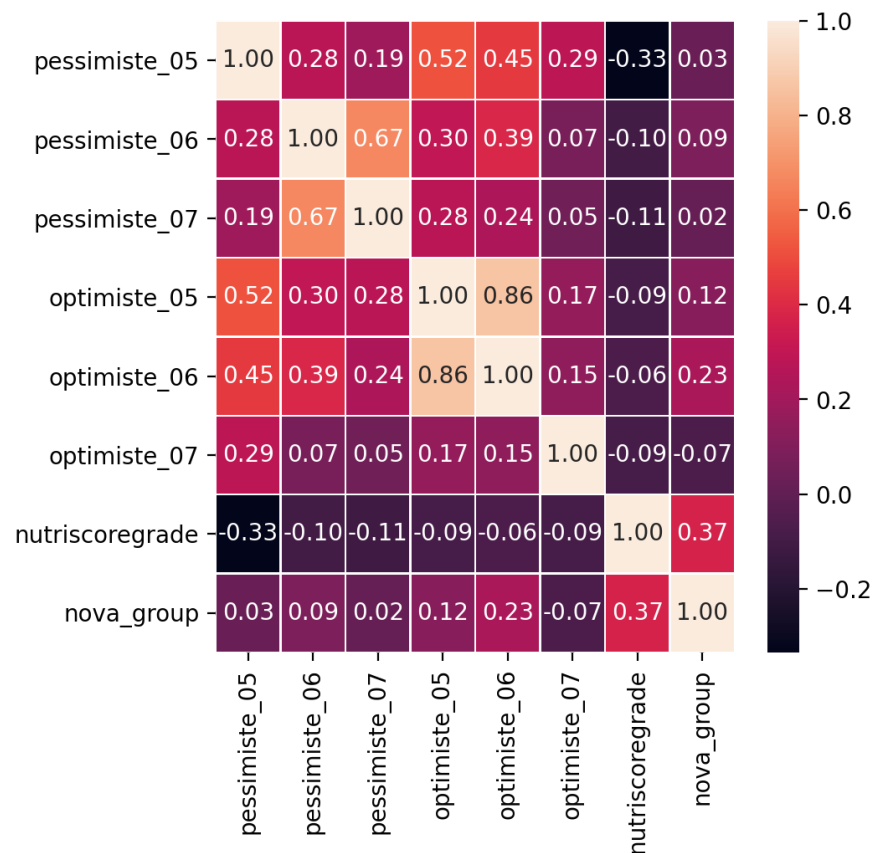
Commentaire : Dans la première figure nous constatons que la partie égale constitue 87% de la classification. À droite, nous effectuons un “zoom” sur les résultats non “égaux”. On peut noter que la méthode qui attribue le plus souvent le score le plus sévère est Nova suivi par Nutri Score et Electre tri optimiste.

Ci dessous, le résultat pour un lambda égale à **0.6** et **0.7** :



Commentaire : Dans ces cas, la méthode “pessimiste” est le score le plus sévère, on peut noter que le pourcentage va augmenter avec un lambda plus élevé. De plus, on constate que la méthode pessimiste a un pourcentage plus élevé également sur le graphique de gauche également.

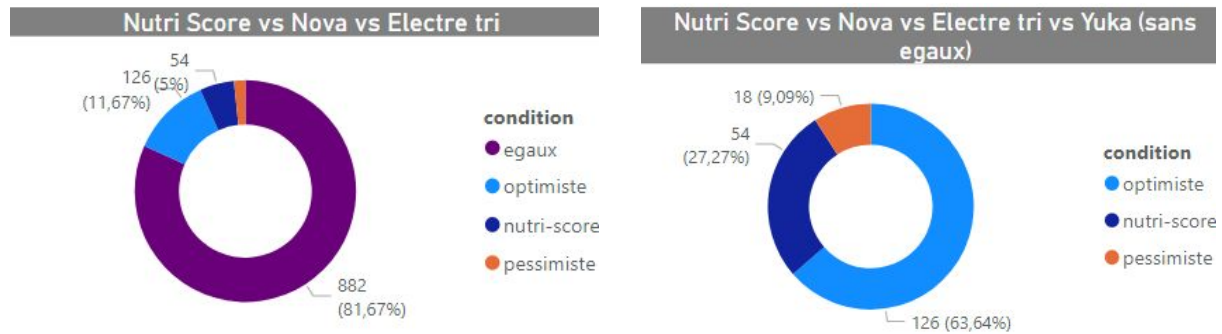
Ci-dessous, la matrice de corrélation :



Comme on peut le voir dans la matrice, il n'y a pas de grande corrélation entre ELECTRE TRI, nutrition score et le groupe nova. La principale corrélation est entre l'ELECTRE TRI pessimiste avec un lambda égal à 0,6 et 0,7 ce qui est plutôt "normal" dans le fait que c'est la même méthode.

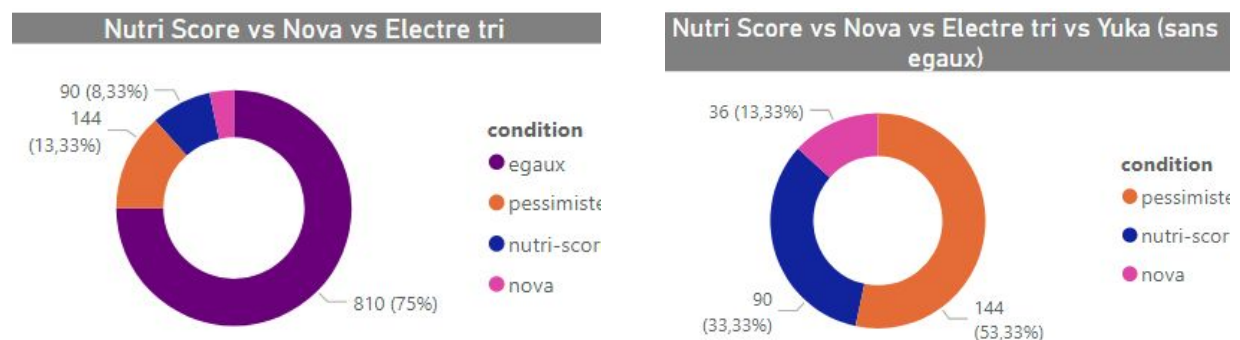
## 4.3 Troisième rapport : BDD3

Ci-dessous, le résultat pour un lambda égale à **0.5** :



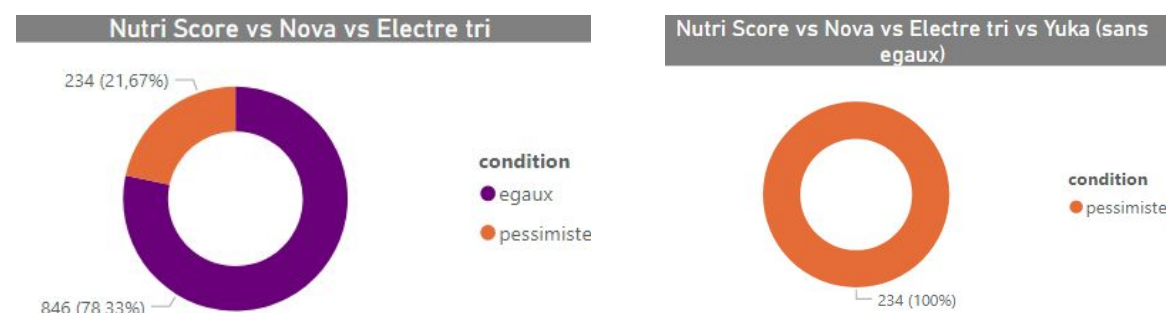
Commentaire : Nous constatons que la classification “optimiste” est dans quasiment 64% des cas la plus sévère et le deuxième est Nutri-Score (27%).

Ci-dessous, le résultat pour un lambda égale à **0.6** :



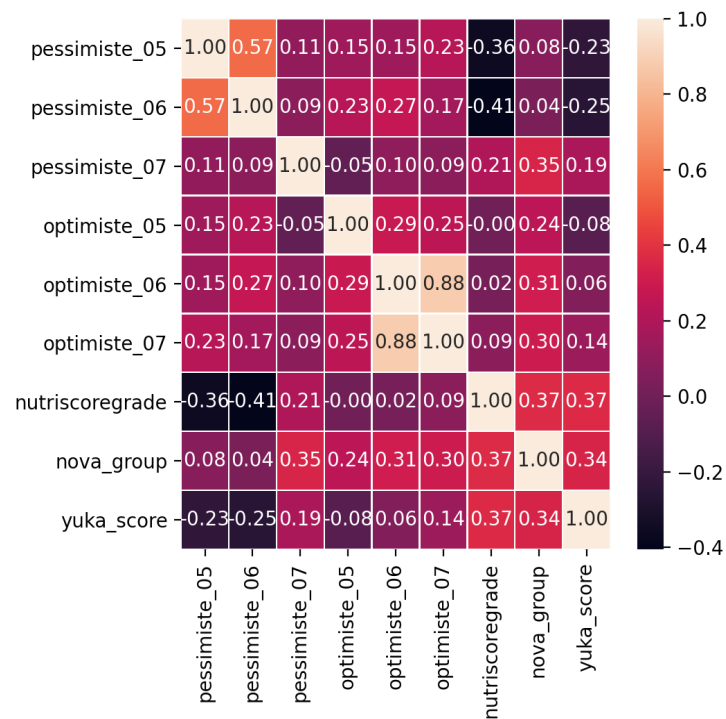
Commentaire : Dans ces cas la méthode “pessimiste” est la plus sévère par rapport aux autres indicateurs en augmentant le lambda à 0.6.

Ci-dessous, le résultat pour un lambda égale à **0.7** :



Commentaire : Ici on voit que la méthode “pessimiste” est dans 100% de cas la plus sévère en terme d'attribution de note..

Ci-dessous, la matrice de corrélation :



Dans ce cas, il n'y a pas de corrélations particulières. En effet, la base ne comporte que 60 éléments, les résultats ne sont pas particulièrement fiables.

## 5. Etude de la transparence

Le texte<sup>3</sup> de loi pour une République numérique de 2016, prévoit trois obligations pour les administrations concernant la transparence de ses algorithmes :

1. Obligation d'indiquer aux intéressés qu'un algorithme est utilisé et quels sont leurs droits
2. Administrations doivent publier les principes de fonctionnement des principaux traitements quand ils fondent des décisions administratives individuelles
3. Fournir à l'individu concerné un ensemble d'informations concernant l'algorithme, son fonctionnement en détail et sous forme intelligible ainsi que les données traitées pour son cas spécifique

Dans notre cours, nous avons tenté de définir la transparence à travers cinq propriétés que nous allons vérifier sur les méthodes étudiées. Pour savoir si chacune des méthodes répondent à ces propriétés, nous nous sommes renseigné à travers des sites web dont nous stipulons les sources en annexe à la fin de notre rapport.

### 5.1 Loyauté

On peut qualifier un algorithme de loyal si la fonctionnalité connue du fournisseur est la même que celle qui est affichée pour l'utilisateur. C'est-à-dire qu'il n'y a pas de fonctionnalité dissimulée par le fournisseur.

L'ensemble des algorithmes sont loyaux car leurs fonctionnalités sont explicitées, c'est-à-dire de permettre de classer les aliments selon leurs catégories prédéfinies avec des critères explicités.

L'application YUKA a des fonctionnalités en plus qui sont détaillées sur leur site internet (lien en annexe).

### 5.2 Équité

Un algorithme est dit équitable si sa conception n'induit pas, de manière consciente ou inconsciente, un effet discriminant ou un biais à l'égard d'une catégorie particulière d'une population.

#### **NOVA**

La méthode ne nous paraît pas équitable car le principe de l'algorithme de NOVA, comme spécifié précédemment, attribue une note négative aux aliments transformés. Cependant cela ne nous permet pas d'être fiable sur tous les aliments, car certains d'entre eux contiennent les nutriments nécessaires pour notre organisme. Il y a donc une défavorisation des aliments transformés.

---

<sup>3</sup> <https://dataanalyticspost.com/lindispensable-transparence-des-algorithmes-publics/>

## **YUKA**

L'algorithme utilisé favorise les produits ne contenant pas d'additif et qui possède un label biologique. La méthode n'est donc pas équitable.

## **NUTRI-SCORE**

La méthode de classification du nutri-score est équitable.

## **FEUX TRICOLORE**

Le feu tricolore est un algorithme qui n'est pas équitable car les codes couleurs ne prennent pas en compte les nutriments, vitamines et minéraux dont la consommation est recommandée.

## **ELECTRE TRI**

La méthode de classification du electre tri est équitable.

# 5.3 Responsabilité

Un algorithme est dit responsable si les concepteurs et les utilisateurs peuvent justifier les résultats du système, qui dépendent exclusivement de la personne qui les met en œuvre. La méthode est évaluée selon différents critères qui sont: les sources de données, les formules utilisées, la manière dont les données ont été nettoyées et traitées.

## **NOVA**

Les critères et le traitement utilisés pour la classification des produits sont connus et chaque groupe est expliqué en détail. Mais il n'est pas possible de connaître avec précision les formules de cette méthode, cependant il est possible de savoir les données utilisées et les télécharger.

## **YUKA**

La base de données est connue, de plus il est possible de vérifier comment elle a été constituée, ainsi que contrôler les informations renseignées. De plus, les informations d'un produit sont mises à jour. Cependant, la méthode de classification n'est pas exactement donnée au public, nous n'avons qu'une échelle de proportion (cf partie 3.3).

## **NUTRI-SCORE**

Pour ce type de méthode il y a beaucoup d'informations sur les formules et sur le calcul fait afin de classer et il est aussi possible de comprendre les différents profils adoptés. La base de données et les traitements utilisés pendant la collecte des données sont bien disponibles.

## **FEUX TRICOLORES**

La responsabilité de la méthode de feux tricolores est similaire à la méthode Nutri-score, puisqu'il est possible de connaître les formules et les calculs. De plus, on sait comment ont été collectées et traitées les données.

## **ELECTRE TRI**

L'algorithme et son fonctionnement sont responsables dans la façon de comprendre comment ont été faites les différentes fonctions et calculs. Les bases de données utilisées sont disponibles comme les traitements fait pour le nettoyage des données.

## **5.4 Explicabilité**

Une décision algorithmique est dite explicable s'il est possible d'en rendre compte explicitement à partir de données et caractéristiques connues de la situation. Autrement dit, s'il est possible de mettre en relation les valeurs prises par certaines variables (les caractéristiques) et leurs conséquences sur la prévision.

### **YUKA**

L'algorithme utilisé favorise les produits ne contenant pas d'additif et qui sont bio selon un pourcentage d'importance. Le problème étant que le choix de cette pondération ne repose sur aucune preuve scientifique.

### **NUTRI-SCORE**

La méthode peut être qualifiée de explicable car nous disposons des détails mathématiques qui permettent de calculer le score.

### **FEUX TRICOLEURE**

Le fonctionnement est explicable car les quantités qui catégorisent les couleurs sont expliquées.

### **NOVA**

Explicable car l'algorithme associe le produit à un groupe selon si il est transformé ou pas.

## **ELECTRE TRI**

Les formules permettant la classification sont données. Cependant, la justification des choix des poids et profils, qui sont renseignés dans la partie 1 de notre rapport, ne nous paraît pas "explicable". En effet, le fait que l'on ait choisi les profils selon la distribution rend la catégorisation assez subjective.

## **5.5 Intelligibilité**

Un algorithme est intelligible si le concepteur est capable de comprendre son fonctionnement et de vérifier que les propriétés désirées sont bien satisfaites. En effet, cela porte sur la capacité des concepteurs à expliquer l'algorithme qu'ils conçoivent.

L'ensemble des algorithmes sont intelligibles car la personne qui a conçu comprend correcteur le fonctionnement et peut aussi vérifier que les résultats obtenus sont cohérents avec la description des différentes méthodes.



## 5.6 Vue d'ensemble

Ci dessous un tableau récapitulatif sur la validation des différentes caractéristiques de transparence selon la méthode de classification.

Méthodes de classification	Loyauté	Equité	Responsabilité	Explicabilité	Intelligibilité
Nutri-Score	✓	✓	✓	✓	✓
Nova	✓	X	X	✓	✓
Yuka	✓	X	X	X	✓
Feux tricolores	✓	X	✓	✓	✓
Electre tri	X	✓	✓	X	X

## 6.Approche par le Machine Learning

### 6.1 Arbre de décision

Dans un modèle d'apprentissage supervisé comme un arbre de décision, l'algorithme apprend sur un ensemble de données étiquetées, fournissant un modèle pour classifier des nouvelles données.

Bien que l'arbre de décision soit l'une des méthodes les plus compréhensibles, si l'on applique cette méthode à notre base de données, la classification est obtenue par le choix de Nutri-Score comme classe (label) de référence, donc il y a un biais.

### 6.2 Réseaux de neurones

Les réseaux de neurones sont comme des boîtes noires, ils prennent des données d'entrée et renvoient une sortie. Par conséquent, nous ne sommes pas en mesure d'expliquer ce qui se passe à l'intérieur de ces réseaux ; également, ces dernières années, plusieurs études ont vu le jour pour tenter d'expliquer le fonctionnement des réseaux de neurones (explainability AI). Nous avons donc conclu que cette approche n'est pas très transparente.

### 6.3 Clustering

Le Clustering est un algorithme non supervisé qui tente, en extrayant lui-même des caractéristiques et des modèles, d'obtenir des groupes homogènes. Dans notre cas, nous souhaitons détecter des clusters pour l'implémentation de la méthode ELECTRE-TRI, cependant les clusters obtenus ne sont pas bien séparés.

# Conclusion

Ce projet nous a permis de constater que les différentes typologies des indicateurs utilisés pour les produits alimentaires ont parfois des biais. Par exemple, la méthode Yuka privilégie les produits bio ou Nova les produits non transformés. Également nous avons vu que pour des scores analysés, certaines méthodes manquent de transparence. Il est donc possible d'envisager ces indicateurs mais il ne faut pas se fier entièrement à leur utilisation car ils présentent plusieurs points critiques.

# Références

## **NUTRI-SCORE :**

<https://www.santepubliquefrance.fr/determinants-de-sante/nutrition-et-activite-physique/articles/nutri-score>

<https://www.mangerbouger.fr/Manger-mieux/Comment-manger-mieux/Comment-comprendre-les-informations-nutritionnelles/Qu-est-ce-que-le-Nutri-Score>

## **NOVA :**

<https://fr.openfoodfacts.org/nova>

<https://siga.care/blog/classification-nova-et-classification-siga/>

<https://www.jequilibre.fr/nutrition-societe/outils-de-classification-nutritionnelle/nutriscore-yuka-nova-sont-ils-efficaces/>

## **YUKA :**

<https://yuka.io/questions-cat/notation-alimentaire/>

<https://lehub.laposte.fr/tendances/effet-yuka-transparence-impose-aux-marques>

## **FEU TRICOLERE :**

[http://alimentation-sante.org/wp-content/uploads/2015/01/Fiche\\_FeuxTricolores\\_1501.pdf](http://alimentation-sante.org/wp-content/uploads/2015/01/Fiche_FeuxTricolores_1501.pdf)

<https://sante.lefigaro.fr/actualite/2014/01/29/21911-conseils-sur-aliments-sous-forme-feux-tricolores>

## **TRANSPARENCE :**

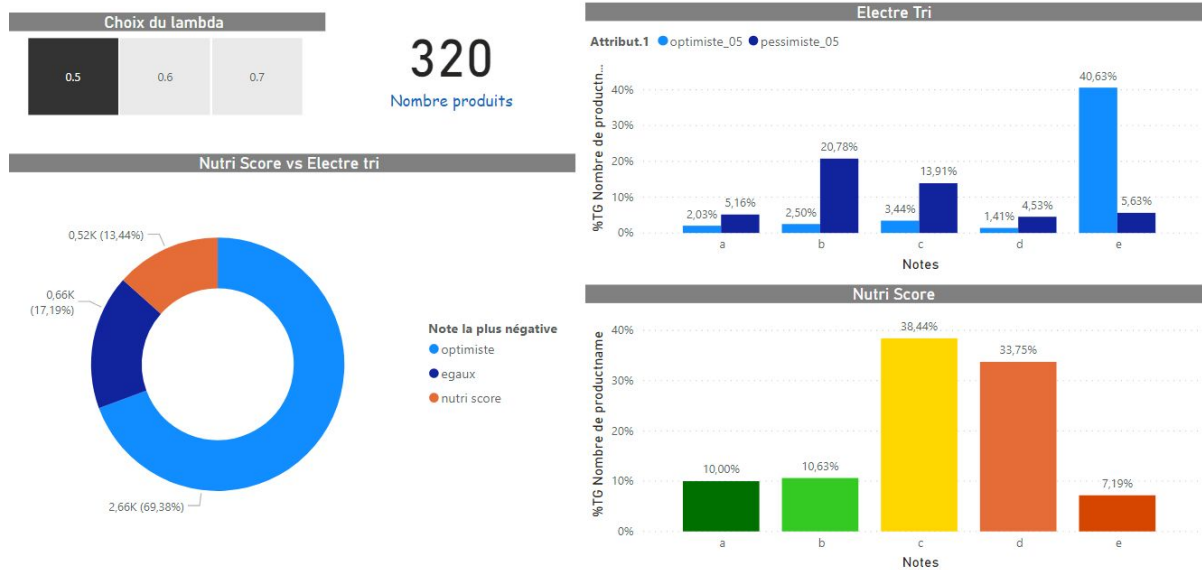
<https://www.cairn.info/journal-etudes-2001-6-page-751.htm>

# Annexes

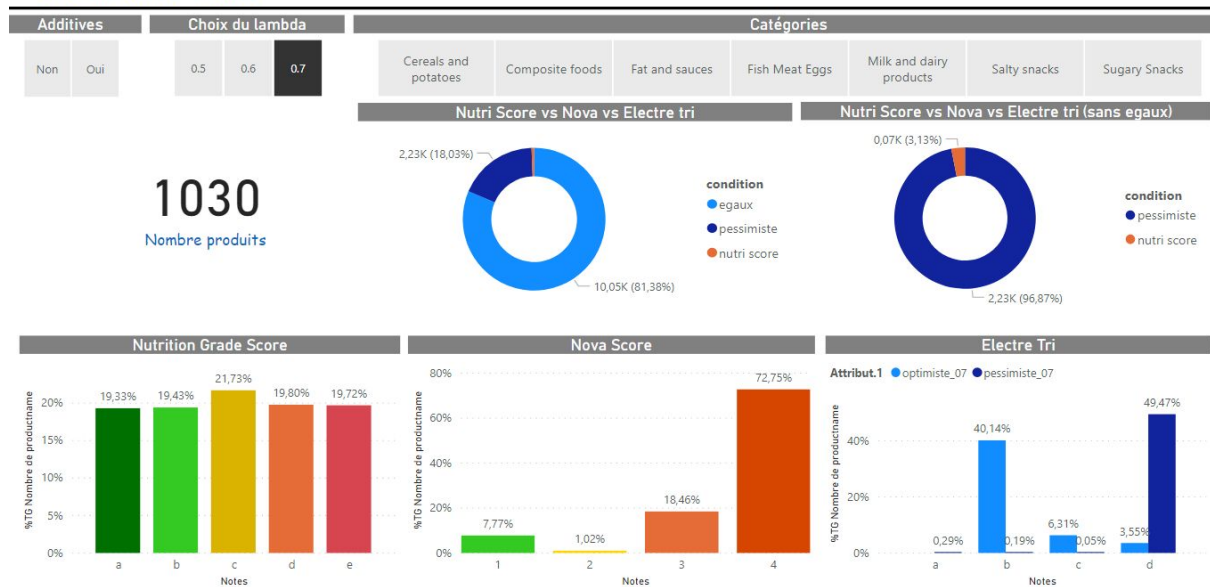
Capture d'écran du modèle power BI



## Base de données de pétales



## Base de données de 1000 produits



## Base de données de 60 produits



## Base de données de 60 produits

