

Projet 2 : Concevez une application au service de la santé publique

Claire Gayral
Avril 2021

Introduction - Appel à projet

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

Analyse
exploratoire

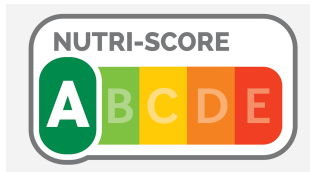
Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

- Santé Publique France
- Open Food Facts
- Nutri-score



from <https://agriculture.gouv.fr>

Introduction - Plan

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

- 1 Idée d'application
- 2 Nettoyage des données
- 3 Analyse exploratoire
 - Analyse univariée
 - Analyse multivariée
 - Réduction de dimension
 - Prédiction du nutri-score

1 Idée d'application

2 Nettoyage des données

3 Analyse exploratoire

Comment faire manger les enfants correctement ?

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

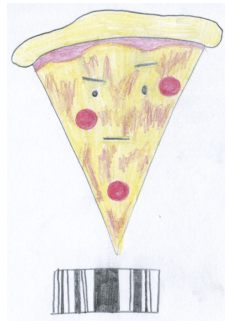
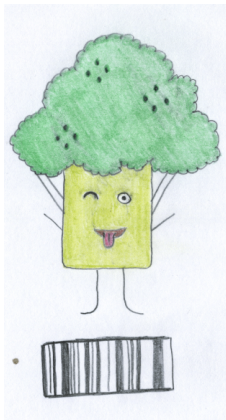
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



→ Apprendre à bien manger devient un jeu d'enfant

Une application simple d'utilisation

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

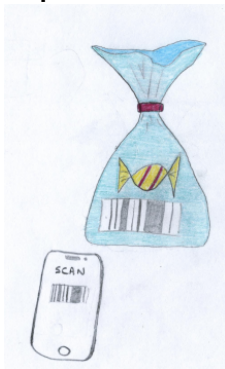
Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

1. Scanner le
produit :



2. Entrer la quantité
:

— — g
— — unités
— — cL

Valeur ajoutée

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

- Suivi des habitudes alimentaires
- Comparaison avec les conseils de l'OMS
- Ludique et éducatif

Éléments techniques

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

- Inférence du nutri-score à partir des informations de scan
- Création d'un score moyen par repas, avec une moyenne et écart-type
- Représentations graphiques facilement interprétables

- 1 Idée d'application
- 2 Nettoyage des données
- 3 Analyse exploratoire

Prétraitement et nettoyage

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

646576 produits et 184 variables

- Suppression variable sans intérêt
- Doublons
- Valeurs aberrantes
- Imputation valeurs manquantes

86150807 valeurs manquantes (72%)

Prétraitement et nettoyage

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

646576 produits et **43** variables

- **Suppression variable sans intérêt**
- Doublons
- Valeurs aberrantes
- Imputation valeurs manquantes

Retrait de variables :
—80 non numériques
—46 vides

56408865 valeurs manquantes (63%)

Prétraitement et nettoyage

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

460471 produits et 43 variables

- Suppression variable sans intérêt
- **Doublons**
- Valeurs aberrantes
- Imputation valeurs manquantes

Retrait de produits :
–186105 fusionnés

12932209 valeurs manquantes (65%)

Prétraitement et nettoyage

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

363519 produits et **32** variables

- Suppression variable sans intérêt
- Doublons
- **Valeurs aberrantes**
- Imputation valeurs manquantes

Retrait :
—96914 produits
—14 variables

5778219 valeurs manquantes (50%)

Prétraitement et nettoyage

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

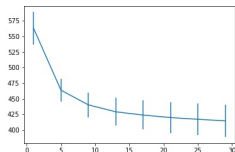
Réduction de
dimension

Prédiction du
nutri-score

363519 produits et 32 variables

- Suppression variable sans intérêt
- Doublons
- Valeurs aberrantes
- Imputation valeurs manquantes

Par KNN avec 12 voisins



1 Idée d'application

2 Nettoyage des données

3 Analyse exploratoire

- Analyse univariée
- Analyse multivariée
- Réduction de dimension
- Prédiction du nutri-score

Nutri-score : Histogramme

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

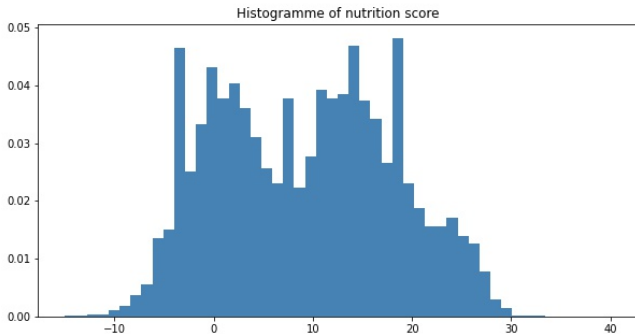
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



Nutri-score : Histogramme

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

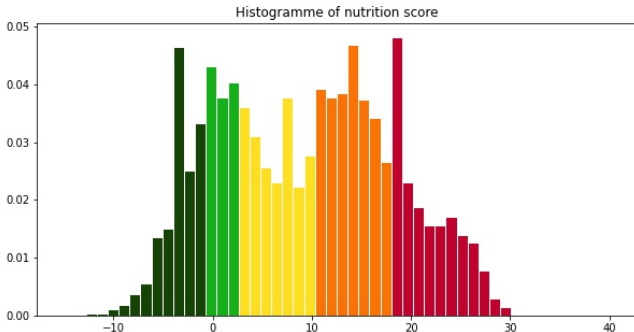
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



Classification en lettre sur le document officiel

Nutri-score : Diagramme circulaire

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

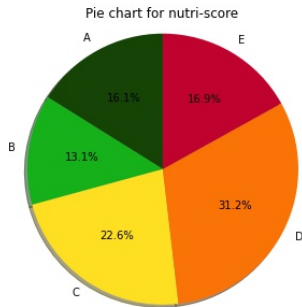
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



Et les autres variables ?

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

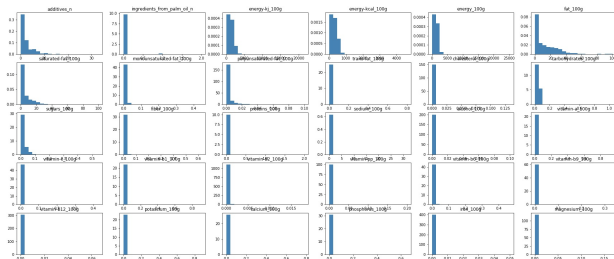
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



Matrice des corrélations

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

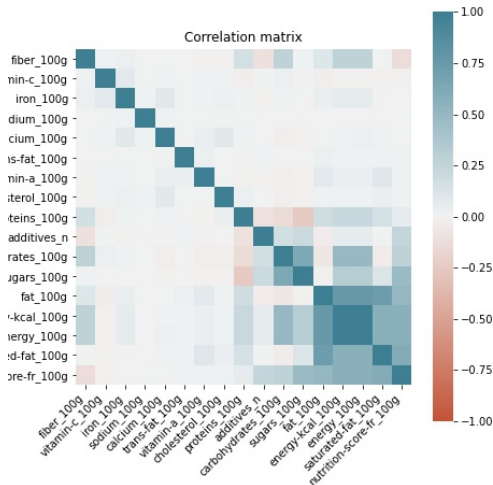
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



Boîte à moustache

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

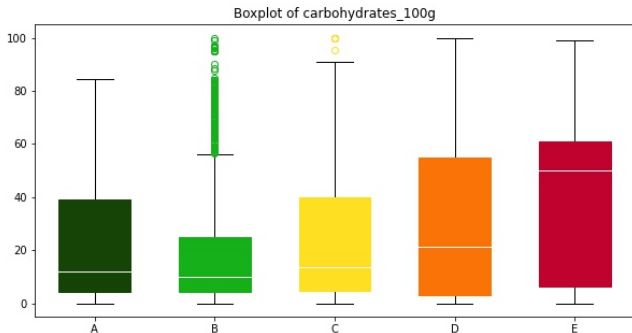
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



PCA - Cercle des corrélations

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

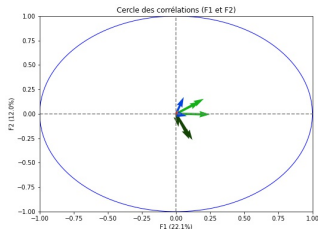
Analyse
exploratoire

Analyse univariée

Analyse multivariée

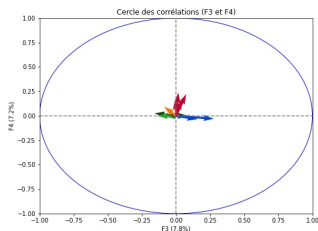
Réduction de
dimension

Prédiction du
nutri-score



Clustering legend

- proteins_100g, fiber_100g,
- iron_100g, calcium_100g, vitamin-c_100g, cholesterol_100g,
- vitamin-a_100g, sodium_100g, trans-fat_100g,
- saturated-fat_100g, fat_100g, energy_100g, energy-kcal_100g,
- sugars_100g, carbohydrates_100g, additives_n,



Clustering legend

- proteins_100g, fiber_100g,
- iron_100g, calcium_100g, vitamin-c_100g, cholesterol_100g,
- vitamin-a_100g, sodium_100g, trans-fat_100g,
- saturated-fat_100g, fat_100g, energy_100g, energy-kcal_100g,
- sugars_100g, carbohydrates_100g, additives_n,

PCA - Projection des produits

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

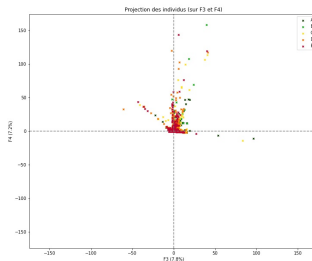
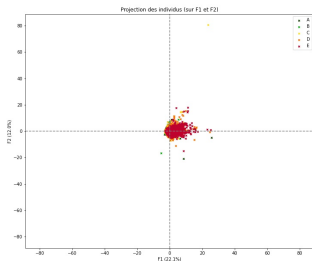
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



PCA - Projection des produits

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

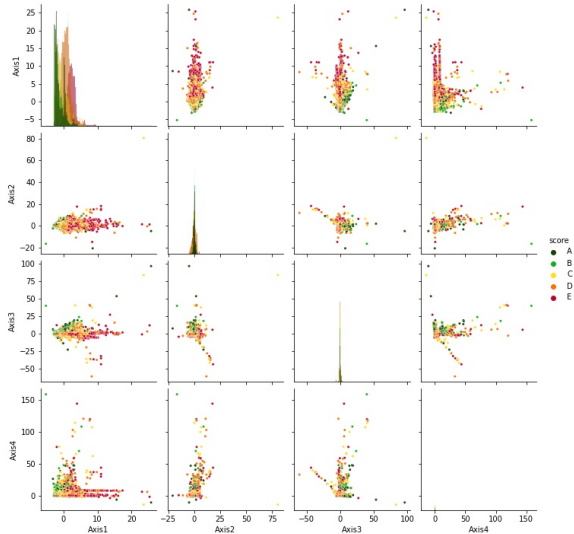
Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



ANOVA : score qualitatif

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

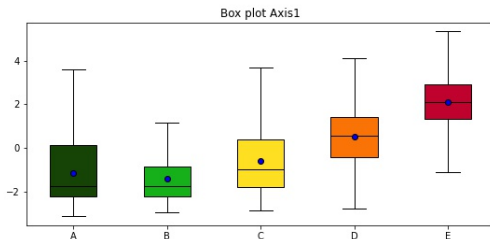
Analyse
exploratoire

Analyse univariée

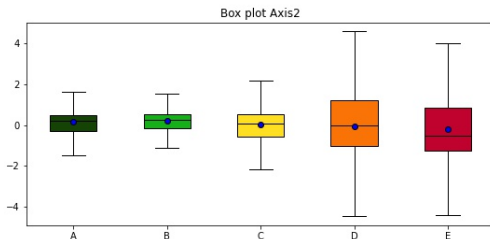
Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score



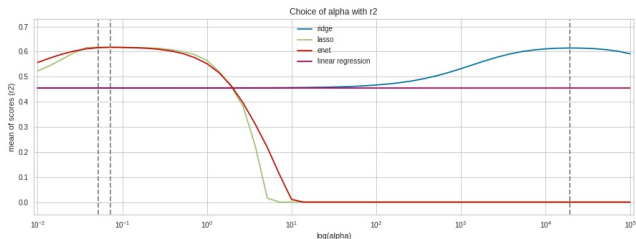
$$\eta^2 = 0.39$$



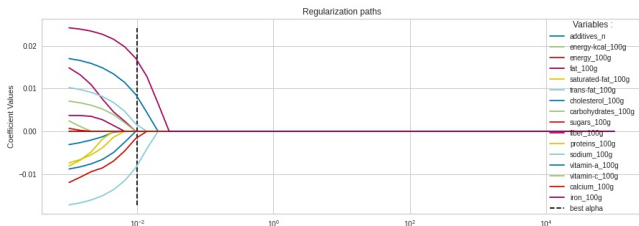
$$\eta^2 = 0.01$$

Régression Linéaire : score quantitatif

Choix du paramètre de régularisation



Chemin de régularisation :



Régression Linéaire : score quantitatif

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idée d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

Choix du modèle

Execution time for Linear Regression (seconds)

	simple LR	Ridge	Lasso	Elastic-net
All variables	0.2	3	21	10
12 most correlated	0.1	3	11	9
PCA 10 axis	0.1	3	5	5

R^2 Error for Linear Regression on best alpha

	simple LR	Ridge	Lasso	Elastic-net
All variables	0.6190	0.6080	0.6089	0.6084
12 most correlated	0.6159	0.6207	0.6207	0.6207
PCA 10 axis	0.5599	0.5562	0.5588	0.5574

→ Modèle linéaire simple

Conclusion

Projet 2 :
Concevez une
application au
service de la
santé publique

Claire Gayral

Idee d'application

Nettoyage des
données

Analyse
exploratoire

Analyse univariée

Analyse multivariée

Réduction de
dimension

Prédiction du
nutri-score

- Possibilité de prédire le score sur les produits non labellés
 - ANOVA
 - Régression Linéaire
- Jeu de données riche en information
- Intégrer une question d'écologie ?

Merci pour votre écoute !