Concevez une application au service de la santé publique

3 mars 2023 Yoann Poupart

Food-Tracker

Story:

→ Suivre ses habitudes

alimentaires

→ Mesurer la qualité des apports

Potential features:

- → Ajout d'aliment
- → Ajout de recette
- → Planification des repas
- → Liste de course

Features:

- → Suivi des AJR (statistiques sur la durée)
- → Recherche & recommandation d'aliments



Éléments clés des données

Composition:

- → Indication précise sur les habitudes alimentaires
- **→** Suivi détaillé des AJR

Nom et catégorie:

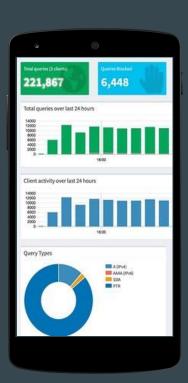
- **→** Recherche d'aliments
- → Recommandation et regroupement d'aliments

<u>Énergie:</u>

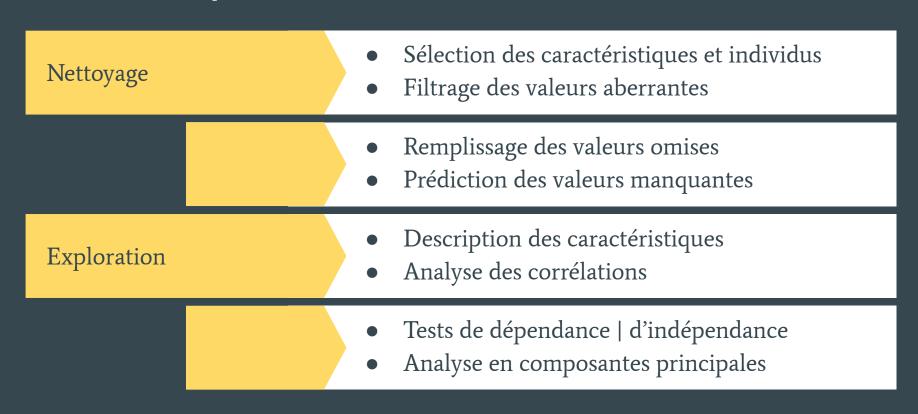
- → Résumé synthétique d'un aliment
- → Suivi globale des <u>AJR</u>

Mesure de qualité:

- → Qualité des aliments (nutri-score|nutri-grade)
- → Qualité de la production (eco-score)



Démarche expérimentale



Nettoyage des données

Caractéristiques

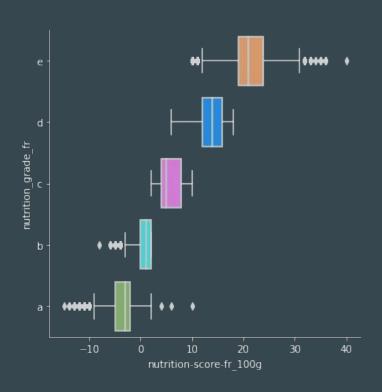
Sélection

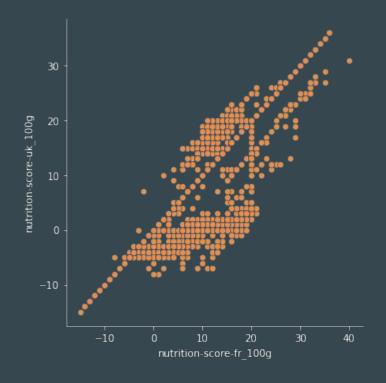
- Composition nutritionnelle
 - ➡ Suivi détaillé des AJR
- Mesures globale (énergie, nutri-grade, nutri-score)
 - → Suivi de la qualité et des AJR
- Nom et différentes catégories
 - → Regroupement d'aliments

Filtrage

- Caractéristiques relativement remplies (seuil à 1%)
 - → Réduire la sparsité (Perte de variables intéressantes eco-score)
- Fusion de caractéristiques
 - → Remplir des valeurs manquantes

Corrélations des caractéristiques





Corrélations des caractéristiques

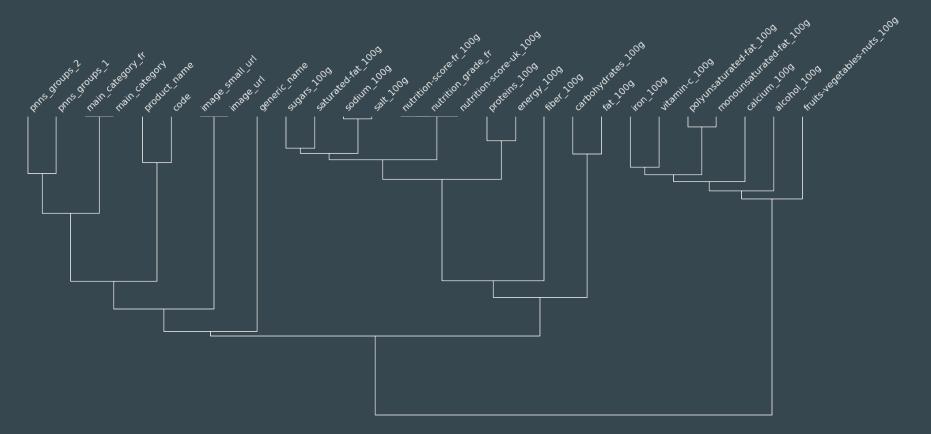
Interprétation:

- → Caractéristiques corrélées mais différentes
- → Donnent des informations légèrements différentes

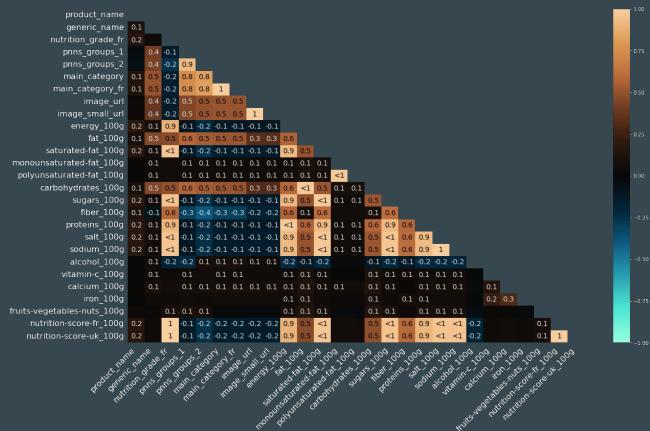
Procédure adoptée:

- **→** Garder les 3 scores
- → Créer un nutri-score moyen

Corrélations des valeurs manquantes



Corrélations des valeurs manquantes



Corrélations des valeurs manquantes

Interprétation:

- → Valeurs manquantes corrélées par nutriments similaires
- → Relève de l'omission plus que de l'erreur

Procédure adoptée:

→ Nutriment omis remplacé par 0

Individus

Sélection

- Aliments vendus en France
 - → Application française
- Aliments nommés
 - → Affichage pour l'application
- Aliments uniques

Filtrage

- Valeurs aberrantes
 - → Interval des caractéristiques (positif, energie<3765.6, nutriment<100)
- Cohérence des caractéristiques
 - → Lien caractéristiques & sous-caractéristiques
 - → Total inférieur à 100

Outliers

Interprétation:

- → Nutriments: distributions à queues lourdes
- → Certains aliments sont des outliers par nature (sel, huile, ...)

Procédure adoptée:

→ Pas de filtrage des outliers

Prédiction des valeurs manquantes

Variable catégorielle (nutri-grade)

- Classifieur randomForest
 - → Robuste aux outliers
 - **→** Interprétable
- Encodage numérique
 - → Par indiçage
 - **⇒** Encodage one-hot

Variables numériques

- Imputeur KNN
 - → Prédictions simultanées
 - \times Lent et sensible aux outliers
- Régresseur randomForest
 - **→** Robuste aux outliers
 - **→** Interprétable

Prédiction des valeurs manquantes

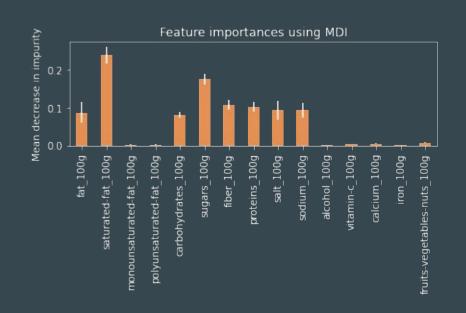
Résultats:

- → RandomForest (RF) meilleure que KNN
- → Overfitting (avec 20 arbres et profondeur max à 20)

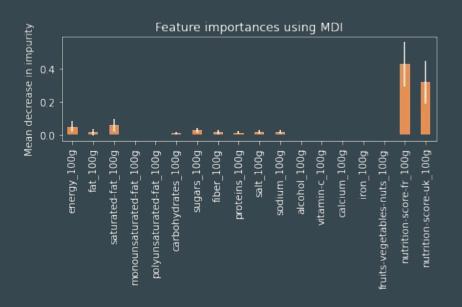
	Train	Test			
KNN		D2: 0.040			
KININ		R ² : 0.919			
RF	R²: 0.975	R²: 0.921			
RF (put)	A: 0.980 - F1: 0.978	A: 0.877 - F1: 0.873			
(nut)	A: 1.000 - F1: 1.000	A: 0.999 - F1: 0.999			
│RF │					

Prédiction des valeurs manquantes

Classifieur (RF) à partir des nutriments

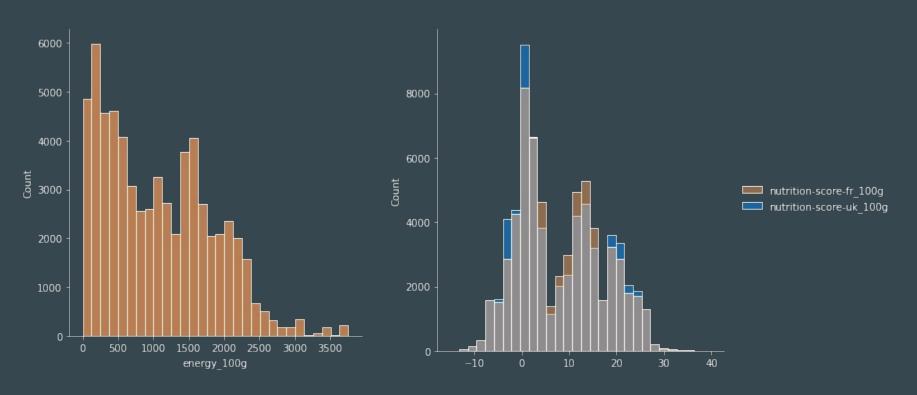


Classifieur (RF) à partir des variables numériques

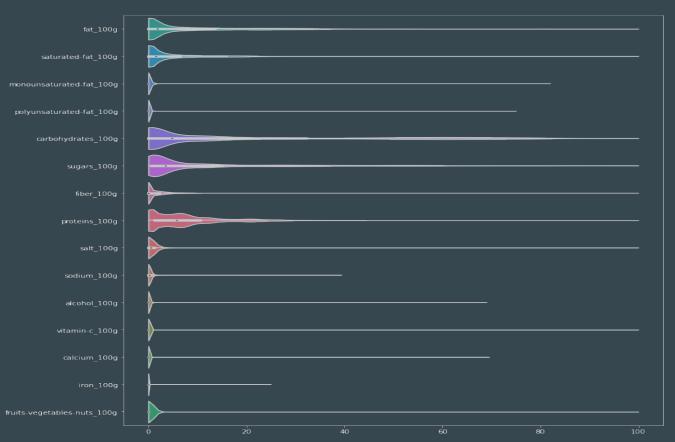


Exploration des données

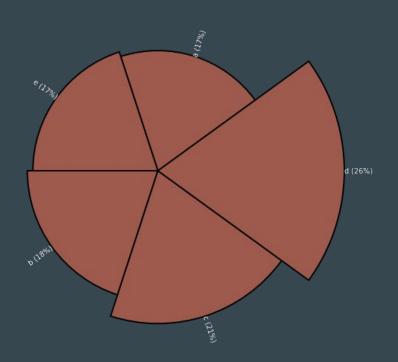
Numerical features

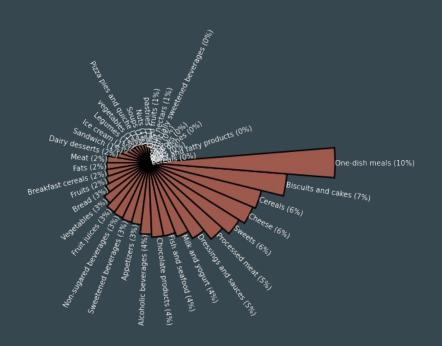


Numerical features

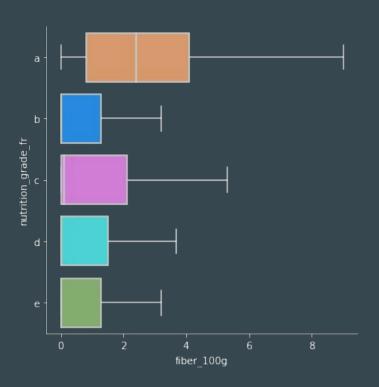


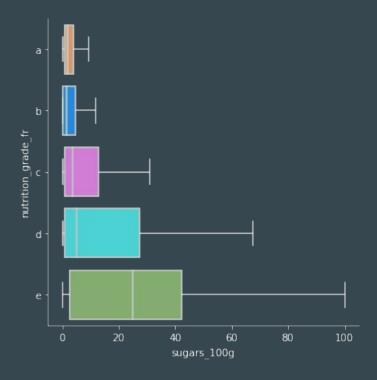
Categorical features



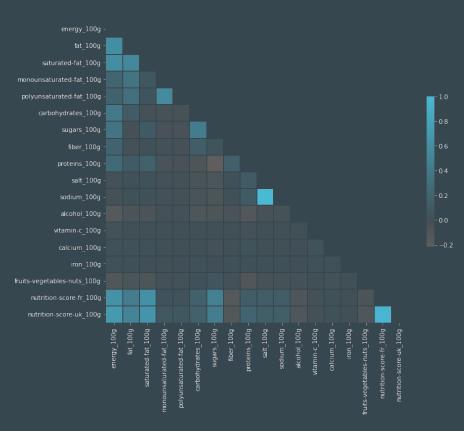


Corrélations qualitatives





Corrélations quantitatives



Corrélations quantitatives

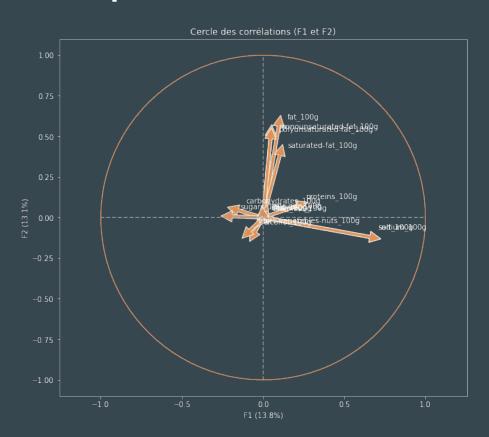
LR	Énergie - Nutri-score (moyen)	R ² : 0.445	p-value: 0.000		
	Nutri-score (fr) - Nutri-score (uk)	R ² : 0.928	p-value: 0.000		
χ²	Nutri-grade - Catégorie	-	p-value: 0.000		
ANOVA	Nutri-grade - Nutri-score (fr)	η²: 0.914	p-value: 0.000		
	Nutri-grade - Énergie	η²: 0.338	p-value: 0.000		
	Catégorie - Nutri-score (fr)	η²: 0.391	p-value: 0.000		
	Catégorie - Énergie	η²: 0.496	p-value: 0.000		

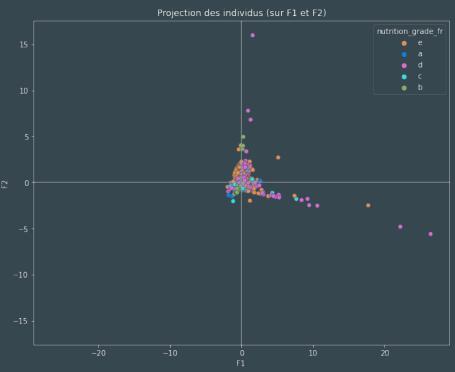
Test de normalité: Énergie (p-value: 0.000) - Nutri-score (p-value: 0.000)

ACP partielle

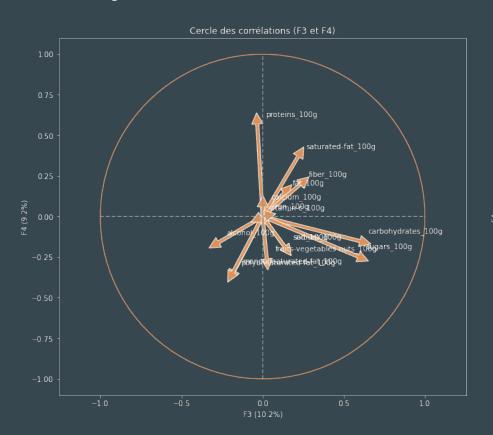
fat_100g -	0.10	0.56	0.13	0.15	-0.22	-0.00	0.01	0.03	0.01	0.14
saturated-fat_100g -	0.10	0.38	0.22	0.37	-0.44	-0.01	-0.01	0.07	0.03	0.21
monounsaturated-fat_100g -	0.05	0.50	-0.17	-0.33	0.23	0.00	-0.00	-0.04	-0.01	-0.05
polyunsaturated-fat_100g -	0.05	0.48	-0.18	-0.34	0.25	0.00	-0.00	-0.04	-0.02	-0.06
carbohydrates_100g -	-0.15	0.04	0.60	-0.15	0.12	0.01	0.06	-0.16	-0.03	-0.01
sugars_100g -	-0.19	0.01	0.59	-0.25	-0.12	0.01	0.02	-0.08	-0.02	-0.04
fiber_100g -	0.01	-0.00	0.23	0.20	0.67	-0.05	0.02	-0.20	0.00	0.44
proteins_100g -	0.21	0.07	-0.03	0.57	0.28	-0.02	-0.04	0.06	0.01	0.04
salt_100g -	0.66	-0.12	0.14	-0.18	-0.03	0.01	0.01	-0.02	-0.01	0.02
sodium_100g -	0.66	-0.12	0.14	-0.18	-0.03	0.01	0.01	-0.02	-0.01	0.02
alcohol_100g -	-0.05	-0.09	-0.27	-0.16	-0.27	-0.01	0.24	-0.50	-0.08	0.66
vitamin-c_100g -	-0.00	-0.00	-0.01	-0.00	0.00	0.69	-0.18	-0.16	0.68	0.03
calcium_100g -	0.00	0.01	0.00	0.06	0.02	0.69	-0.03	0.03	-0.71	-0.00
iron_100g -	-0.00	0.00	0.01	0.00	0.07	0.18	0.91	0.35	0.14	-0.01
fruits-vegetables-nuts_100g -	-0.08	-0.08	0.03	-0.26	0.03	0.03	-0.28	0.72	0.04	0.55
	F1	F2	F3	F4	F5	F 6	F7	F8	F9	F10

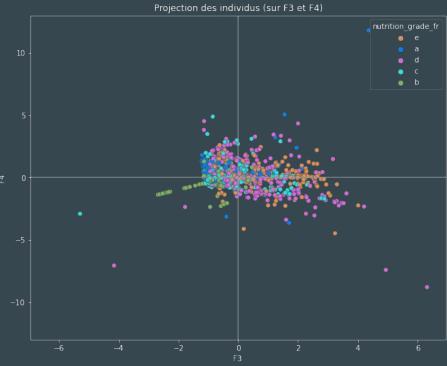
ACP partielle





ACP partielle



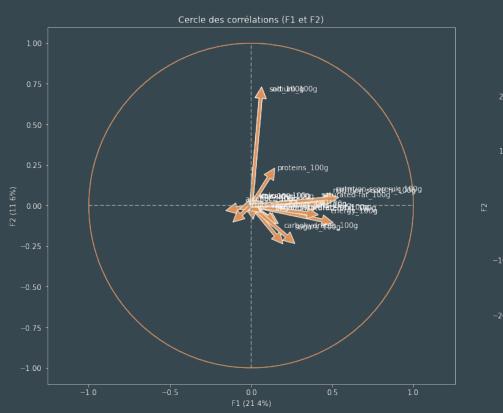


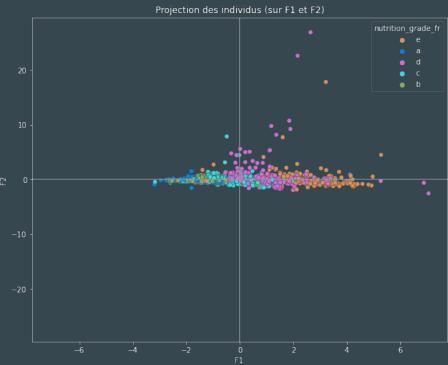
ACP totale

										5.002
energy_100g -	0.44	-0.09	0.01	-0.02	0.23	-0.02	0.03	-0.05	0.00	0.09
fat_100g -	0.35	-0.05	0.32	-0.06	0.02	-0.00	-0.01	0.03	0.01	0.10
saturated-fat_100g -	0.38	0.01	0.05	-0.28	-0.14	-0.01	-0.03	0.05	0.02	0.16
monounsaturated-fat_100g -	0.11	-0.07	0.55	0.32	0.06	0.01	0.00	-0.01	-0.01	-0.06
polyunsaturated-fat_100g -	0.10	-0.07	0.55	0.32	0.06	0.01	0.00	-0.01	-0.01	-0.07
carbohydrates_100g -	0.15	-0.18	-0.29	0.36	0.36	0.01	0.08	-0.13	-0.03	-0.05
sugars_100g -	0.22	-0.19	-0.40	0.40	-0.01	0.01	0.02	-0.06	-0.03	-0.07
fiber_100g -	0.00	-0.02	-0.04	-0.05	0.70	-0.04	0.05	-0.13	0.01	0.37
proteins_100g -	0.11	0.17	0.07	-0.52	0.33	-0.00	-0.06	0.08	0.01	0.03
salt_100g -	0.06	0.66	-0.03	0.22	0.04	0.00	0.01	-0.02	-0.01	0.02
sodium_100g -	0.06	0.66	-0.03	0.22	0.04	0.00	0.01	-0.02	-0.01	0.02
alcohol_100g -	-0.09	-0.02	0.05	0.02	-0.31	-0.06	0.37	-0.51	-0.09	0.65
vitamin-c_100g -	-0.00	-0.00	0.00	0.00	-0.01	0.69	-0.12	-0.23	0.68	0.03
calcium_100g -	0.01	0.00	0.01	-0.05	0.02	0.70	-0.03	0.01	-0.71	0.01
iron_100g -	0.00	-0.00	-0.00	0.01	0.04	0.18	0.82	0.52	0.14	-0.01
fruits-vegetables-nuts_100g -	-0.06	-0.06	-0.06	0.24	-0.07	0.05	-0.39	0.60	0.05	0.62
nutrition-score-fr_100g -	0.45	0.03	-0.16	-0.02	-0.23	0.00	-0.00	0.02	0.00	-0.01
nutrition-score-uk_100g -	0.47	0.04	-0.11	-0.04	-0.19	0.00	-0.00	0.01	0.00	-0.01
	F1	F2	F3	F4	F5	F <mark>6</mark>	F7	F8	F9	F10

- 0.75 - -0.75

ACP totale





Merci de votre attention.

•••

Des questions?