

gt_based_model

Pierre MASSÉ

June 3, 2020

1 Modèle basé sur les données manuellement étiquetées

L'objet de ce notebook est de mettre en place le modèle basé sur les données manuellement étiquetées.

1.1 Récupération des données

1.1.1 Préambule technique

```
[1]: # setting up sys.path for relative imports
from pathlib import Path
import sys
project_root = str(Path(sys.path[0]).parents[1].absolute())
if project_root not in sys.path:
    sys.path.append(project_root)
```

```
[2]: # imports and customization of display
import os
import pandas as pd
pd.options.display.min_rows = 6
pd.options.display.width = 108
pd.options.display.latex.repr = True
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline

from src.pimapi import Requester
from src.pimest import ContentGetter
from src.pimest import PathGetter
from src.pimest import PDFContentParser
from src.pimest import BlockSplitter
from src.pimest import SimilaritySelector
```

1.1.2 Chargement du fichier des données manuellement étiquetées

On commence par charger le fichier csv contenant les données manuellement étiquetées.

```
[3]: ground_truth_df = pd.read_csv(Path('.') / '..' / 'ground_truth' / 'manually_labelled_ground_truth.csv',
                                   sep=';',
                                   encoding='latin-1',
                                   index_col='uid')
ground_truth_df.head()
```

[3]:

uid	designation	ingredients
a0492df6-9c76-4303-8813-65ec5ccbfa70	Concentré liquide Asian en bouteille 980 ml CHEF	Eau, maltodextrine, sel, arômes, suc
d183e914-db2f-4e2f-863a-a3b2d054c0b8	Pain burger curry 80 g CREATIV BURGER	Farine de blé T65, eau, levure, vina
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	Macaroni en sachet 500 g PANZANI	- 100% Semoule de BLE dur de qualité
528d4be3-425c-4f8b-8a87-12f1bc645ddd	Fève de Tonka en sachet 100 g COMPTOIR COLONIAL	fève de tonka (graines ridées de 25
51b38427-b2ea-4c56-93e8-4242361ef31b	Caviar d'aubergine en pot 500 g PUGET RESTAURA...	Aubergine 60,5% (aubergine, huile de

```
[4]: ground_truth_uids = list(ground_truth_df.index)
      ground_truth_uids[:5]
```

```
[4]: ['a0492df6-9c76-4303-8813-65ec5ccbfa70',
      'd183e914-db2f-4e2f-863a-a3b2d054c0b8',
      'ab48a1ed-7a3d-4686-bb6d-ab4f367cada8',
      '528d4be3-425c-4f8b-8a87-12f1bc645ddd',
      '51b38427-b2ea-4c56-93e8-4242361ef31b']
```

1.1.3 Pipeline d'acquisition du contenu des données

On commence par construire un premier pipeline d'acquisition des données. Il fonctionne en 3 étapes : -
détermination du chemin vers lequel aller chercher les fiches techniques - récupération du contenu binaire du
fichier - conversion de ce contenu binaire en texte

```
[5]: acqui_pipe = Pipeline([('PathGetter', PathGetter(ground_truth_uids=ground_truth_uids,
                                                       train_set_path=Path('.') / '..' / 'ground_truth',
                                                       ground_truth_path=Path('.') / '..' / 'ground_truth',
                                                       )),
                           ('ContentGetter', ContentGetter(missing_file='to_nan')),
                           ('ContentParser', PDFContentParser(none_content='to_empty'))],
                           verbose=True)
```

```
[6]: texts_df = acqui_pipe.fit_transform(ground_truth_df)
      texts_df
```

```
[Pipeline] ... (step 1 of 3) Processing PathGetter, total= 0.1s
[Pipeline] ... (step 2 of 3) Processing ContentGetter, total= 0.1s
Launching 8 processes.
[Pipeline] ... (step 3 of 3) Processing ContentParser, total= 37.2s
```

```
[6]:
```

uid	designation	ingredients
a0492df6-9c76-4303-8813-65ec5ccbfa70	Concentré liquide Asian en bouteille 980 ml CHEF	Eau, maltodextrine, sel, arômes, sucres
d183e914-db2f-4e2f-863a-a3b2d054c0b8	Pain burger curry 80 g CREATIV BURGER	Farine de blé T65, eau, levure, vinaigre
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	Macaroni en sachet 500 g PANZANI	- 100% Semoule de BLE dur de qualité
528d4be3-425c-4f8b-8a87-12f1bc645ddd	Fève de Tonka en sachet 100 g COMPTOIR COLONIAL	fève de tonka (graines ridées de 25%)
51b38427-b2ea-4c56-93e8-4242361ef31b	Caviar d'aubergine en pot 500 g PUGET RESTAURANT	Aubergine 60,5% (aubergine, huile de colza)
e4c8c61f-a401-4384-8128-181447e5bdd2	Confitures assorties en bocaux 30 g BONNE MAMAN	Confiture de myrtilles et de cassis
50766d10-6135-4958-b743-1cddfcfb7c230	Spécialité pomme saveur biscuitée en boîte 5/1...	Pommes 95%, sirop de glucose-fructose
bb9f3995-57b3-429d-b075-2d81a90e406f	Coco Pops en sachet 500 g KELLOGG'S	Riz, Sucre, Cacao maigre en poudre, jus d'orange issu de l'agriculture biologique
52bf5f5a-79f2-4652-8eac-2b80c697269b	PUR JUS ORANGE DE MEDITERRANÉE 3L	Persil
04235024-80f3-46c2-bad0-aae0d5fab024	Persil feuilles en boîte 80 g LA CASE AUX EPICES	Pommes de terre, huile de tournesol
93a5d344-4ab0-48de-bbb0-fec8117d07a2	Chips saveur poulet braisé en sachet 25 g SIBELL	Sucre, huiles végétales (Colza), noix
440ca0b4-1c79-433f-ac8d-6e2ae77d3288	Pâte à tartiner noisette en bouteille squeezezer...	Sucre, eau, vinaigre d'alcool, purée d'aubergine
7b9cf3b3-bf4b-408c-b8c8-807a09745979	Sauce Sweet chili en flacon souple 875 ml HEINZ	Huile d'olive vierge extra 99,5%, Arôme
b638ab96-427c-4d5f-9a19-74684eec3b8e	PRÉPARATION CULINAIRE À BASE D'HUILE D'OLIVE E...	Huile de colza, extrait naturel de piment
8bb101c0-8c6e-4665-8310-1a2d13257b2e	Huile pimentée spicy'piz en dose 3 ml SAVEURS ...	Piment rouge fort équeuté* (85%), curcuma
54f40033-f9cf-411c-81a5-11974f6715aa	Harissa en seau de 3 kg PROVENCE OLIVES	SUCRE DE CANNE, EAU, AROME, COLORANTS
2405356e-ecc9-404f-8279-16fddf168b73	1883 VANILLE 100 CL	Baie des Batak
c57be8e4-30e0-417b-90fa-b60d9f9ece48	Baie des Batak en sachet 250 g TERRE EXOTIQUE	Filets de maquereaux (61 %), eau, colorants
6622442d-d88f-4b90-9414-b91c9e8e5296	Filet de maquereau à la tomate en poche 4/4 FU...	Mélange de Blés de pays recommandés
507b428e-e99d-464b-b9d3-50629efe4355	Farine de blé type 00 pour pizza en sachet 1 k...	Lait entier (88%) ; sucre ; poudre de cacao
545e8bc5-15fd-458a-8c9f-b18faecc2016	GOURDES AU CHOCOLAT P'TITS TOQUES YABON 4X80G	Sirop de glucose-fructose, farine de blé
b45b3942-accd-4115-8f1f-736bdb42a35a	CHAMONIX*2 25G	Thon Albacore (92%), eau (7%), sel (2%)
6c07fd47-116d-4531-a164-d4f4040f212e	Thon Albacore en morceaux au naturel en poche ...	Farine de BLÉ 67 %, sucre, huile de colza
ca711137-4644-45b7-9928-456a09d9746a	THE BISCUITS POCKET 22 GR	Crème frangipane 39% (sucre, BEURRE, œufs)
e70d347d-5e91-4e0f-b396-b9bb4b09da29	Galette des rois à la crème frangipane avec fê...	TAGADA\nsucre; sirop de glucose; gélatine
70500268-802d-4211-93ba-9edbf6e0e7a3	COLIS KERMESSE 2019 A.P " 22+8"	eau\nsucre\netrait de thé noir (1,4%)
60ab6264-3aae-4f82-8828-9abd606683f5	Ice Tea citron-citron vert en bouteille 50 cl ...	semoule de blé dur blanche biologique
380ee776-b92f-4378-a14b-c33622327b09	Couscous blanc BIO en sac 5 kg MARKAL	Riz à grain long précuit (65%), feuilles de menthe
7a8f6966-3710-4e3e-b656-af4eec271d1	Feuilles de vigne farcies au riz en boîte 2/1 ...	jus de fruits à base de concentrés de fruits
5f3aba4a-b591-43b6-af5d-0d3304b263e3	Spécialité au citron en bouteille verre 70 cl ...	Olives vertes 98%denoyautées , huile d'olive
adb90cc7-4b91-4771-b505-61a207cc1f96	Olives vertes piquantes dénoyautées en seau 3,...	Poivrons rouge et vert, tomate en poches
3caac27c-704b-4f83-9203-482219d73d02	Mélange d'épices italien en boîte 200 g DUCROS	Sucre, texturants : alginate de sodium
9b4cd824-bbb4-4cc9-be85-790988a803fd	Entremet parfum caramel en sachet 1 kg NUTRACY	cèpes 70% (Boletus edulis et respectueux)
046cddb1f-1915-4916-8874-902cc5ec73be	Crème de cèpes à la truffe en boîte 4/4 DEMETRA	Sirop de glucose, purée de fraises (10%)
8f7c1e03-64c4-41cd-b557-53e559a39290	Sauce dessert fraise en bouteille 1 kg NESTLÉ ...	Sirop de glucose-fructose, sucre, eau
b3ba5d5bb-78cf-40ce-97a3-b7847977e6e7	Mini marshmallows en boîte 150 g VAHINE	Eau, graines de moutarde 16%, vinaigre
a6c2eeb0-0046-4c1f-998c-d90bbd560f16	Moutarde en stick 10,5 g HEINZ	Sucre, gélifiant : alginate de sodium
7b4b2030-63f3-416c-b3a7-b20204a748ef	Entremet parfum vanille en sachet 750 g NUTRACY	Thon, eau, sel
7e5ff57f-5a13-46bf-bb1c-b59b40727515	Thon listao en morceaux au naturel en boîte 4/...	Sirop de glucose, sucre, eau, stabilisateurs
4f83306f-66de-4545-9b12-7790b57b61ae	Nappage miroir neutre en seau 7 kg ANCEL	Maltodextrine de blé - Amidon de maïs
f254910f-8562-431e-a499-cae5e67158b7	Sauce diable en boîte 800 g NEFF MADA	Cassis 89,3%*1, sucre.\n*1 : Variations
acba6c83-385c-405a-ac35-96e3c2828683	Purée de cassis sucrée en poche 1 kg LEONCE BLANC	eau de source 80%, jus de fruits à base de concentrés
bcdf718d-968f-4ad0-bad5-0a7384b8ab52	Oasis orange en bouteille 2 L OASIS	Sucre, amidon de FROMENT, farine de blé
e6eb7635-748d-424f-a1e2-b47a010a31e0	Muffin Gourmet en sac 10 kg COMPLET	Miso (95%) (eau, fèves de soja, riz, sel)
30778190-c91f-4942-bf6c-2b07640d4d12	Miso en poudre en boîte 250 g ARIAKE	farine de BLE enrichie (farine de BLE + vitamines)
007f266f-fc64-4bd0-98d1-0d1e5650ec5b	Tarte sablée sucrée au beurre en forme de cur...	Farine de blé, beurre en poudre (7 %) et blé dur biologique (triticum durum).
8afeb736-1b04-4aa7-9b25-fa534a177019	Beignet au beurre en sachet 10 kg COMPLET	Amidon de maïs - Arômes - Sirop de glucose
100c3211-9dc2-48e9-a2d6-0a7384b8ab52	Semoule de blé fine BIO en sac 5 kg MARKAL	Eau, graines de moutarde, vinaigre de cidre
42c360ae-7bcc-4671-81b9-f6170eadd007	Jus de rôti de boeuf hyposodé en boîte 800 g N...	Céleri râpé, eau, vinaigre d'alcool, sel
ecc3c180-51dd-4070-bdfe-abbd3344626d	Moutarde forte de Dijon en tube 105 g REINE DE...	cheddar fondu (44%) (eau, lactosérum, sel)
7f8c0f20-50d9-4b18-9c25-4074f10b6f3e	Céleri râpé en boîte 5/1 EPISAVEURS	Huile de colza, morue (32%), eau, poivre
19f93f6d-2168-43c5-9713-4b6e564f51185	CHEDDAR CHEESE SAUCE 1.420 KG	Olives, eau, sel, stabilisateur de couleur
321363df-147c-4e4d-bf1a-c982c0ce9951	Brandade de morue en boîte 5/1 RAYMOND GEOFFROY	Huile de colza (48 %), sucre, eau, vinaigre
b7a5df3b-42b4-4cde-ae39-18f0e7a5f776	OLIVES NOIRES CONFITES DENOYAUTEES 4/4	Purée de tomates mi réduite (64%), soja jaune biologique (glycine max)
81ee31fb-37cc-4093-89cd-5aeaae7c1add	Sauce algérienne en flacon 850 g COLONA	Pomme 77% (origine 100% France), abricot
41da4d6f-7e9f-4f95-bf2a-2acd7f138cd9	Ketchup en flacon souple 970 ml CALIFORNIA	Farine de BLÉ - Sucre - OEUFS - Huile d'olive
00a76c54-b2f6-4c4d-a97e-bd9a9cd2f112	Protéine de soja BIO en sac 10 kg MARKAL	Poires, sirop de glucose-fructose, arômes
9f7f946c-dd35-41ae-815e-ad3181e96d5d	Spécialité pomme-abricot allégée en coupelle 9...	Pois chiches, eau, sel.
ceeb8b03-8e6d-43f3-9594-410085ce71b3	Barre pâtissière 800 g COQUELIN	sucré, pâte de cacao, beurre de cacao
820a65ca-5112-404c-a79a-953db0c4e989	Compote poire allégée en boîte 5/1 LA PULPE	Amidon modifié de pomme de terre - Lactose
7c1d47f4-a2ca-4579-9687-a4e31d97647f	Pois chiches en boîte 1/2 MAINGOURD	Farine de FROMENT, poudre de LACTOSE
ce71ed2f-fb3f-48bb-b2c5-b9b911f78adb	Bâton chocolat noir 40% cacao en boîte 1,6 kg ...	Mirabelles (55%), eau, acidifiant : citrique
6c7987c9-22f1-40d5-99cf-927c31c7db5c	Potage aux 9 légumes en sachet 20 g NEFF MADA	Eau, poissons 35%, langoustines 5% (langoustine)
b8cbe6f9-71d4-4e51-a169-1c163d49a561	Pancake en sachet 1 kg COMPLET	FRAMBOISE\Nmaltodextrine, carraghénane
ade044cf-a734-4633-b1fa-8f724e943aeb	Mirabelle dénoyautée au naturel en boîte 2/1 O...	Légumes (pommes de terre (14%), carottes (14%),
oe736da3-faa6-4e57-a777-e354e29ac219	Soupe de poissons et langoustines au sel de Gu...	NaN
4c8fcdcb-1c40-479f-84f7-589bd764b59e	PRÉPARATION DÉSHYDRATÉE POUR FLAN 5 PARFUMS ÉD...	NaN
2b055200-0c27-4274-b976-7ec71de9ad15	Saladière parisienne en coupelle 220 g SAUPIQUET	Pommes 89%, purée de fraises à base de concentrés
21233a00-bc20-40fc-acb9-ee2e2321cac2	San Pellegrino Limonata en canette 33 cl SAN P...	Eau, graines de moutarde, vinaigre, curcuma
9fa94fea-6315-4439-b976-0730d37522c9	Chapelure Batter Mix en sac de 4,2 kg GRIFFITH...	Croustilles de céréales enrobées de sucre
656618d-3bdd-43d8-ab0c-de51894621f9	SPÉCIALITÉ POMME-FRAISE ALLÉGÉE EN SUCRES EN B...	
b988521e-1191-472f-ad90-5afc8c2a57b8	MOUTARDE DE DIJON SEAU 5KG - EDMOND FALLLOT	
846db626-d8a9-49b2-be8d-7d724adeaff3	Assortiment Mix Turbine 3 saveurs en colis MON...	

On peut afficher quelques textes récupérés par le pipeline :

```
[19]: with pd.option_context("max_colwidth", 1000):  
      pass  
      # print(texts_df.sample(3, random_state=42)['text'])  
      # (texts_df.sample(3, random_state=42)['text']  
      #      .to_latex(Path('.') / 'tbls' / 'processed_FT.tex',  
      #               index=False,  
      #               index_names=False,  
      #               column_format='p{\linewidth}',  
      #               na_rep='-',  
      #               escape=True,  
      #               )  
      # )
```

1.2 Découpage en blocs

On découpe les longs textes en blocs. Chaque texte devient une liste de strings plus court.

```
[22]: def splitter(text):  
      return(text.split('\n\n'))
```

```
[23]: split_transfo = BlockSplitter(splitter_func=splitter)  
      splitted_df = split_transfo.fit_transform(texts_df)  
      splitted_df
```

Launching 8 processes.

```
[23]:
```

uid	designation	ingredients
a0492df6-9c76-4303-8813-65ec5ccbfa70	Concentré liquide Asian en bouteille 980 ml CHEF	Eau, maltodextrine, sel, arômes, sucres
d183e914-db2f-4e2f-863a-a3b2d054c0b8	Pain burger curry 80 g CREATIV BURGER	Farine de blé T65, eau, levure, vinaigre
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	Macaroni en sachet 500 g PANZANI	- 100% Semoule de BLE dur de qualité
528d4be3-425c-4f8b-8a87-12f1bc645ddd	Fève de Tonka en sachet 100 g COMPTOIR COLONIAL	fève de tonka (graines ridées de 25%)
51b38427-b2ea-4c56-93e8-4242361ef31b	Caviar d'aubergine en pot 500 g PUGET RESTAURANT	Aubergine 60,5% (aubergine, huile de colza)
e4c8c61f-a401-4384-8128-181447e5bdd2	Confitures assorties en bocaux 30 g BONNE MAMAN	Confiture de myrtilles et de cassis
50766d10-6135-4958-b743-1cddfcfb7c230	Spécialité pomme saveur biscuitée en boîte 5/1...	Pommes 95%, sirop de glucose-fructose
bb9f3995-57b3-429d-b075-2d81a90e406f	Coco Pops en sachet 500 g KELLOGG'S	Riz, Sucre, Cacao maigre en poudre, jus d'orange issu de l'agriculture biologique
52bf5f5a-79f2-4652-8eac-2b80c697269b	PUR JUS ORANGE DE MEDITERRANÉE 3L	Persil
04235024-80f3-46c2-bad0-aae0d5fab024	Persil feuilles en boîte 80 g LA CASE AUX EPICES	Pommes de terre, huile de tournesol
93a5d344-4ab0-48de-bbb0-fec8117d07a2	Chips saveur poulet braisé en sachet 25 g SIBELL	Sucre, huiles végétales (Colza), noisette
440ca0b4-1c79-433f-ac8d-6e2ae77d3288	Pâte à tartiner noisette en bouteille squeezezer...	Sucre, eau, vinaigre d'alcool, purée d'aubergine
7b9cf3b3-bf4b-408c-b8c8-807a09745979	Sauce Sweet chili en flacon souple 875 ml HEINZ	Huile d'olive vierge extra 99,5%, Arôme naturel de piment
b638ab96-427c-4d5f-9a19-74684eec3b8e	PRÉPARATION CULINAIRE À BASE D'HUILE D'OLIVE E...	Huile de colza, extrait naturel de piment
8bb101c0-8c6e-4665-8310-1a2d13257b2e	Huile pimentée spicy'piz en dose 3 ml SAVEURS ...	Piment rouge fort équeuté* (85%), curcuma
54f40033-f9cf-411c-81a5-11974f6715aa	Harissa en seau de 3 kg PROVENCE OLIVES	SUCRE DE CANNE, EAU, AROME, COLORANTS
2405356e-ecc9-404f-8279-16fddf168b73	1883 VANILLE 100 CL	Baie des Batak
c57be8e4-30e0-417b-90fa-b60d9f9ece48	Baie des Batak en sachet 250 g TERRE EXOTIQUE	Filets de maquereaux (61 %), eau, colorants
6622442d-d88f-4b90-9414-b91c9e8e5296	Filet de maquereau à la tomate en poche 4/4 FU...	Mélange de Blés de pays recommandés
507b428e-e99d-464b-b9d3-50629efe4355	Farine de blé type 00 pour pizza en sachet 1 k...	Lait entier (88%) ; sucre ; poudre de cacao
545e8bc5-15fd-458a-8c9f-b18faecc2016	GOURDES AU CHOCOLAT P'TITS TOQUES YABON 4X80G	Sirop de glucose-fructose, farine de blé
b45b3942-accd-4115-8f1f-736dbd42a35a	CHAMONIX*2 25G	Thon Albacore (92%), eau (7%), sel (2%)
6c07fd47-116d-4531-a164-d4f4040f212e	Thon Albacore en morceaux au naturel en poche ...	Farine de BLÉ 67 %, sucre, huile de colza
ca711137-4644-45b7-9928-456a09d9746a	THE BISCUITS POCKET 22 GR	Crème frangipane 39% (sucre, BEURRE, œufs)
e70d347d-5e91-4e0f-b396-b9bb4b09da29	Galette des rois à la crème frangipane avec fê...	TAGADA\nsucre; sirop de glucose; gélatine
70500268-802d-4211-93ba-9edbf6e0e7a3	COLIS KERMESSE 2019 A.P " 22+8"	eau\nsucre\netrait de thé noir (1,4%)
60ab6264-3aae-4f82-8828-9abd606683f5	Ice Tea citron-citron vert en bouteille 50 cl ...	semoule de blé dur blanche biologique
380ee776-b92f-4378-a14b-c33622327b09	Couscous blanc BIO en sac 5 kg MARKAL	Riz à grain long précuit (65%), feuilles de menthe
7a8f6966-3710-4e3e-b656-af4eec271d1	Feuilles de vigne farcies au riz en boîte 2/1 ...	jus de fruits à base de concentrés de fruits
5f3aba4a-b591-43b6-af5d-0d3304b263e3	Spécialité au citron en bouteille verre 70 cl ...	Olives vertes 98%denoyautées , huile d'olive
adb90cc7-4b91-4771-b505-61a207cc1f96	Olives vertes piquantes dénoyautées en seau 3,...	Poivrons rouge et vert, tomate en poches
3caac27c-704b-4f83-9203-482219d73d02	Mélange d'épices italien en boîte 200 g DUCROS	Sucre, texturants : alginate de sodium
9b4cd824-bbb4-4cc9-be85-790988a803fd	Entremet parfum caramel en sachet 1 kg NUTRACY	cèpes 70% (Boletus edulis et respectueux)
046cddb1f-1915-4916-8874-902cc5ec73be	Crème de cèpes à la truffe en boîte 4/4 DEMETRA	Sirop de glucose, purée de fraises
8f7c1e03-64c4-41cd-b557-53e559a39290	Sauce dessert fraise en bouteille 1 kg NESTLÉ ...	Sirop de glucose-fructose, sucre, eau
b3ba5dbb-78cf-40ce-97a3-b7847977e6e7	Mini marshmallows en boîte 150 g VAHINE	Eau, graines de moutarde 16%, vinaigre
a6c2eeb0-0046-4c1f-998c-d90bbd560f16	Moutarde en stick 10,5 g HEINZ	Sucre, gélifiant : alginate de sodium
7b4b2030-63f3-416c-b3a7-b20204a748ef	Entremet parfum vanille en sachet 750 g NUTRACY	Thon, eau, sel
7e5ff57f-5a13-46bf-bb1c-b59b40727515	Thon listao en morceaux au naturel en boîte 4/...	Sirop de glucose, sucre, eau, stabilisateurs
4f83306f-66de-4545-9b12-7790b57b61ae	Nappage miroir neutre en seau 7 kg ANCEL	Maltodextrine de blé - Amidon de maïs
f254910f-8562-431e-a499-cae5e67158b7	Sauce diable en boîte 800 g NEFF MADA	Cassis 89,3%*1, sucre.\n*1 : Variations
acba6c83-385c-405a-ac35-96e3c2828683	Purée de cassis sucrée en poche 1 kg LEONCE BLANC	eau de source 80%, jus de fruits à base de concentrés
bcdf718d-968f-4ad0-bad5-0a7384b8ab52	Oasis orange en bouteille 2 L OASIS	Sucre, amidon de FROMENT, farine de blé
e6eb7635-748d-424f-a1e2-b47a010a31e0	Muffin Gourmet en sac 10 kg COMPLET	Miso (95%) (eau, fèves de soja, riz)
30778190-c91f-4942-bf6c-2b07640d4d12	Miso en poudre en boîte 250 g ARIAKE	farine de BLE enrichie (farine de BLE)
007f266f-fc64-4bd0-98d1-0d1e5650ec5b	Tarte sablée sucrée au beurre en forme de cur...	Farine de blé, beurre en poudre (7 %)
8afeb736-1b04-4aa7-9b25-fa534a177019	Beignet au beurre en sachet 10 kg COMPLET	blé dur biologique (triticum durum).
100c3211-9dc2-48e9-a2d6-0a7384b8ab52	Semoule de blé fine BIO en sac 5 kg MARKAL	Amidon de maïs - Arômes - Sirop de glucose
42c360ae-7bcc-4671-81b9-f6170eadd007	Jus de rôti de boeuf hyposodé en boîte 800 g N...	Eau, graines de moutarde, vinaigre de vin
ecc3c180-51dd-4070-bdfe-abbd3344626d	Moutarde forte de Dijon en tube 105 g REINE DE...	Céleri rave, eau, vinaigre d'alcool, sel
7f8c0f20-50d9-4b18-9c25-4074f10b6f3e	Céleri râpé en boîte 5/1 EPISAVEURS	cheddar fondu (44%) (eau, lactosérum)
19f93f6d-2168-43c5-9713-4b6e564f51185	CHEDDAR CHEESE SAUCE 1.420 KG	Huile de colza, morue (32%), eau, poivre
321363df-147c-4e4d-bf1a-c982c0ce9951	Brandade de morue en boîte 5/1 RAYMOND GEOFFROY	Olives, eau, sel, stabilisateur de couleur
b7a5df3b-42b4-4cde-ae39-18f0e7a5f776	OLIVES NOIRES CONFITES DENOYAUTEES 4/4	Huile de colza (48 %), sucre, eau, vinaigre
81ee31fb-37cc-4093-89cd-5aeaae7c1add	Sauce algérienne en flacon 850 g COLONA	Purée de tomates mi réduite (64%), soja jaune biologique (glycine max)
41da4d6f-7e9f-4f95-bf2a-2acd7138cd9	Ketchup en flacon souple 970 ml CALIFORNIA	Pomme 77% (origine 100% France), abricot
00a76c54-b2f6-4c4d-a97e-bd9a9cd2f112	Protéine de soja BIO en sac 10 kg MARKAL	Farine de BLÉ - Sucre - OEUFS - Huile
9f7f946c-dd35-41ae-815e-ad3181e96d5d	Spécialité pomme-abricot allégée en coupelle 9...	Poires, sirop de glucose-fructose, arômes
ceeb8b03-8e6d-43f3-9594-410085ce71b3	Barre pâtissière 800 g COQUELIN	Pois chiches, eau, sel.
820a65ca-5112-404c-a79a-953db0c4e989	Compote poire allégée en boîte 5/1 LA PULPE	sucré, pâte de cacao, beurre de cacao
7c1d47f4-a2ca-4579-9687-a4e31d97647f	Pois chiches en boîte 1/2 MAINGOURD	Amidon modifié de pomme de terre - Lactose
ce71ed2f-fb3f-48bb-b2c5-b9b911f78adb	Bâton chocolat noir 40% cacao en boîte 1,6 kg ...	Farine de FROMENT, poudre de LACTOSE
6c7987c9-22f1-40d5-99cf-927c31c7db5c	Potage aux 9 légumes en sachet 20 g NEFF MADA	Mirabelles (55%), eau, acidifiant : citrique
b8cbe6f9-71d4-4e51-a169-1c163d49a561	Pancake en sachet 1 kg COMPLET	Eau, poissons 35%, langoustines 5% (cristallin)
ade044cf-a734-4633-b1fa-8f724e943aeb	Mirabelle dénoyautée au naturel en boîte 2/1 O...	FRAMBOISE\Nmaltodextrine, carraghénane
oe736da3-faa6-4e57-a777-e354e29ac219	Soupe de poissons et langoustines au sel de Gu...	Légumes (pommes de terre (14%), carottes
4c8fcdcb-1c40-479f-84f7-589bd764b59e	PRÉPARATION DÉSHYDRATÉE POUR FLAN 5 PARFUMS ÉD...	NaN
2b055200-0c27-4274-b976-7ec71de9ad15	Saladière parisienne en coupelle 220 g SAUPIQUET	NaN
21233a00-bc20-40fc-acb9-ee2e2321cac2	San Pellegrino Limonata en canette 33 cl SAN P...	Pommes 89%, purée de fraises à base de concentrés
9fa94fea-6315-4439-b976-0730d37522c9	Chapelure Batter Mix en sac de 4,2 kg GRIFFITH...	Eau, graines de moutarde, vinaigre, sel
656618d-3bdd-43d8-ab0c-de51894621f9	SPÉCIALITÉ POMME-FRAISE ALLÉGÉE EN SUCRES EN B...	Croustilles de céréales enrobées de sucre
b988521e-1191-472f-ad90-5afc8c2a57b8	MOUTARDE DE DIJON SEAU 5KG - EDMOND FALLLOT	
846db626-d8a9-49b2-be8d-7d724adeaff3	Assortiment Mix Turbine 3 saveurs en colis MON...	

On peut afficher un exemple de texte découpé en blocs :

```
[24]: sep = '\n-----\n'
sample = splitted_df.sample(1, random_state=39)['blocks'].iloc[0]
print(sep.join(sample))

tex_str = (
pd.DataFrame(sample, columns=['Blocs'])
.to_latex(column_format='p{10cm}',
index=False,
index_names=False,
escape=True,
)
.replace(r'\textbackslash n', '\\\newline ')
)

#with open(Path('.') / 'tbls' / 'block_example.tex', mode='w') as file:
#    file.write(sep.join(sample).replace('\n', r' \newline '))
```

30/12/19

Date d'impression :

Remarque :

Les informations contenues dans cette fiche technique sont données de bonne foi, en l'état actuel de nos connaissances, et selon les indications communiquées par le producteur ou le fournisseur. Il appartient au client de vérifier la conformité de la marchandise par rapport à l'usage qu'il en fait.

Création :

12/06/12

12 rue René Cassin
37390 NOTRE DAME

Tél :

02 47 85 55 00

Fax :02 47 41 33 32

FICHE TECHNIQUE

Mélange du trappeur, 70 g

Trapper blend, 70g

Code article KEREX

Nom latin (si disponible)

/ EAN Code

Code barre

/ KEREX Code

/ (Latin name)

TEEPTRAPPEUR

X

3760063322262

Poids net

Poids brut

Origine

/ net weight

/ gross weight

/ Origin

0,07 Kilogramme

0,125 Kilogramme
CANADA

/ General information

Informations générales
DLUO conseillée / "Best before date" recommended
Nomenclature douanière / Customs code
Conditions idéales de stockage
/ Conditions of storage
Ingrédients :

Conserver dans un endroit frais et sec
Store in a cool dry place

5 ans / 5 years
0910999900

Sucre, poivre noir, coriandre, légumes déshydratés (ail, oignon, poivron rouge), sel de mer, sucre d'érable, arôme d'érable naturel, huile végétale (canola)
Sugar, black pepper, coriander, dehydrated vegetables (garlic, onion, red bell pepper), sea salt, maple sugar, natural maple aroma, vegetable oil (canola)

/ Ingredients

Contaminants / Contaminating
Ionisation / Irradiation

OGM / GMO

Pesticides/ Pesticides

Métaux Lourds

/ Heavy Metals

Allergènes et leurs dérivés (si présents)
/ Allergens (if existing)

Conformité à la directive 1999/2/CE (22/02/99)
Produit non ionisé et ne contenant pas d'ingrédients ionisés.
Not irradiated
accordingly with the Reg 1999/2/CE (22/02/99).
Free from GMO
Ne contient pas d'OGM, est non soumis à l'étiquetage sur les OGM
Conforme à la directive 396/2005 /CE
In accordance with Reg 396/2005 /CE.
Conforme au règlement 1881/2006 /CE
In accordance with Reg 1881/2006 /CE..

Gluten
Crustacés
Oeufs
Poisson
Soja
Lait
Fruits à coque - Arachides
Céleri
Moutarde
Sésame
Sulfites
Lupin
Mollusques

/ Gluten
/ Crustaceans

```

/ Eggs
/ Fish
/ Soy
/ Milk
/ Peanuts and Treenuts
/ Celery and celeriac
/ Mustarde
/ Sésame
/ Sulphites
/ Lupin
/ Shellfish
-----
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
Absence
-----
Caractères microbiologiques
-----
/ Microbiological characteristics
-----
Microorganismes aérobies 30 °C
Escherichia coli
Salmonelles
Levures
Moisissures
Aflatoxine Total
Aflatoxine B1
-----
/ Total plat count (APC)
E. Coli
/
/ Salmonella
/ Yeasts
/ Moulds
/ Total aflatoxin
B1 aflatoxin
/
-----
NF V05-051 < 6 000 000 / g
NF V08-053 < 10 / g
NF V08-052 Absence dans 25g
NF V08-059 < 10 000 / g
NF V08-059 < 10 000 / g
Kit Enzymatique < 10 ppb
Kit Enzymatique < 5 ppb
-----

```

1.3 Train / Test split

On procède au découpage en un jeu d'entrainement et un jeu de test en gardant 400 produits pour l'entrainement et 100 produits pour le test :

```
[25]: train, test = train_test_split(splitted_df, train_size=400, random_state=42)
```


1.4 Entraînement sur le jeu d'entraînement

On entraîne un modèle `SimilaritySelector`, sur le set d'entraînement :

```
[26]: model = SimilaritySelector()
```

```
[27]: model.fit(train['blocks'], train['ingredients'])
```

```
[27]: <src.pimest.SimilaritySelector at 0x7f0ebad7bb80>
```

```
[28]: len(model.count_vect.vocabulary_)
```

```
[28]: 1204
```

```
[29]: predicted = pd.Series(model.predict(test['blocks']),
                           index=test.index,
                           name='predicted'
                           )
predicted = pd.concat([test['ingredients'], predicted], axis=1)
predicted
```

```
[29]:
```

uid	ingredients	predicted
2892dd68-e3a6-474c-b543-3ebfd3490658 a57c1561-b88e-4694-8bd8-55623f2afa17 3634fb1e-ee79-41d1-8aaa-084c1fae55bd5 345591f4-d887-4ddc-bb40-21337fa9269d 13980d31-9002-457d-8d49-b451f08f473c 74297717-3fa8-4aed-95cc-e8737c1a6157 9458a288-04f0-49fc-9d92-d88c696ab085 6e976147-adeb-4d2d-925a-cb7c58c111a2 21233a00-bc20-40fc-acb9-ee2e2321cac2 c2ef743e-f3f2-4e8a-aab0-1e6cbeb71666 04235024-80f3-46c2-bad0-aae0d5fab024 c3b6b4df-e586-4f10-8e58-15fbf0816acb 8d833da0-610f-4045-9877-adf8ad9ea56c 209431f8-9436-49cb-a6fe-16225deb5c3f aeda7bf1-4d4d-44bb-8518-2e91fc6fd76c 0481d91b-9653-42e7-b525-9dc9b87b06f2 7c709a5e-b913-4a02-9396-a6469b09482a adb90cc7-4b91-4771-b505-61a207cc1f9e 484ac00a-a670-46a9-a9c4-5114174d9e3b d8b0687e-5e25-4ff9-9384-087f993218bc 7f622727-e4ad-45cc-9af4-4509acf91154 46211a29-279c-4c8f-9f06-635d61f83d66 49b11281-34ea-44b0-a11c-4ae21d4c58e3 194419d0-d9f2-4799-81ac-d9e3aa77fd27 c6fd7472-f363-4aaf-8702-bd56e2e139de 8065d46f-0731-4018-9bcb-f654466badb0 e67341d8-350f-46f4-9154-4dbbb8035621 09cf97db-4cf9-4885-a2b2-da65a4cfe5c3 1868226c-812c-438d-bba4-6b046537d0cd 2cd5567a-bee6-46f2-a4bd-19d2b42ec001 d59d96cb-0230-4090-8220-78ce8496fd91 de22b1fa-039d-479e-b36e-608946b75bb8 8dec0469-c9f5-4139-be25-fc2a58959444 2ca5dc9e-8058-499a-affe-3ec9c06d55b7 ab48aled-7a3d-4686-bb6d-ab4f367cada8 3b7deab3-349b-4bd0-b4d8-c4ba53e61b58 22663c1d-2580-40b3-b4a0-ed8ba1054980 6566e18d-3bdd-43bd-ab0c-de51894621f9 eadb972c-6623-472d-a11d-489a7faf6f11 b8cbe6f9-71d4-4e51-a169-1c163d49a561 e51b7fd6-d878-47f8-a36b-f10f8d4087bd 78665d97-164a-4c23-956b-3e41f5f7e423 6dc9277e-3cd7-415b-a4c7-733ec3a0ba705 f4c16bbd-130d-4e12-b2a0-541853dd560d 046cdb1f-1915-4916-8874-902cc5ec73be c91c8d53-4260-4d9c-851e-7756672a8941 a0492df6-9c76-4303-8813-65ec5ccbfa70 440ca0b4-1c79-433f-ac8d-6e2ae77d3288 f1982954-f091-4aa1-addb-6659062e0f66 6c07fd47-116d-4531-a164-d4f4040f212e 846db626-d8a9-49b2-be8d-7d724adeaff3 f45db604-11ad-4756-aeab-3a5a1a34f914 c8b299c5-22d3-40ae-8df3-ad638507803f 0eb6fa2e-c9e5-4d99-984e-625b1e6bb136 09e45b38-4da1-4eb5-888a-3ebd437a2291 41da4d6f-7e9f-4f95-bf2a-2acd7138cd9 14c9068c-633a-4eea-bb4d-6ac5de0e9a32 6622442d-d88f-4b90-9414-b91c9e8e5296 a2418174-e16a-41e0-ac14-c87208fb3529 5adc7512-6168-4966-ae3f-f6ec133bf56e b4b57208-278a-489e-9814-f8a5ae5f44cf 7508bd85-f350-4d81-a7f4-954411ba089e 064d3b22-c1cc-41bd-bd03-56e8829bd4cd c5344fe1-ac16-4630-824e-6036ae491399 2286f782-9d2e-410f-84c4-4ab88003a002 4283306f-66de-4545-9b12-7790b57b61ae 42094d22-5907-4089-bdaf-0a19ff09b3fe 8afeb736-1b04-4aa7-9b25-fa534a177019 eb5bb1e3-0a86-46cb-9d76-3a965cf20c4b 5cee689e-6fb1-493c-b232-1d8fb1f88a57 3624db37-f88e-47c0-9e9c-bb55163bfbaca 1268d9ef-d848-41ad-973c-7f9e694a4625 6bdb201d-5879-423b-96b8-3ae88b857818	Café instantané, café torréfié moulu (3%). Lentilles blondes Poire 99,9%, antioxydant: acide ascorbique. Gésier de dinde émincé 50%, graisse de canard ... Edulcorants sorbitol, isomalt, sirop de maltit... sucre, amidon modifié, LACTOSÉRUM en poudre, d... paprika, poivre noir, coriandre, cumin, ail se... Maltodextrine, amidon de maïs, sel, farine de ... NaN Farine de BLÉ 20% - Huile de colza - OEUFs ent... Persil mini poivrons jaunes, eau, sucre, sel, affermi... 100% haricots blancs bio Poudre de lait écrémé, anti-agglomérant : E551... HAPPY CHERRY\nsirop de glucose; sucre; gélatin... Farine de BLE, huile de colza non hydrogénée, ... SIROP DE GLUCOSE / SUCRE / AROMES / ACIDIFIANT... Olives vertes 98%denoyautées, huile tournesol... Pommes de terre 59,5 % - Céleris 40 % - Amidon... Eau ; jus de fruit de la passion ; sucre ; épa... Eau, huile de tournesol, beurre 9,5 %, jaune d... pâte de cacao 50,5% ; sucre 47,5% ; beurre de ... NaN Pommes de terre, eau, sel. Farine de pois chiche Lentilles, eau, sel, arôme naturel (céleri) Sucre roux de canne* (64%), amidon de maïs*, ... Chocolat au lait (sucre, beurre de cacao, LAIT... pâte à tartiner aux noisettes et au cacao 81,5... Sucre, farine de FROMENT, poudre d'OEUFs entie... Amidon de maïs* - Lait écrémé* - Sel - Féculé ... NOISETTES entières décortiquées. Sucre; sirop de glucose; graisse de palme; hum... 100% Arabica - 100% Semoule de BLE dur de qualité supérieur... Lait écrémé 59.5%, lait en poudre reconstitué ... Lait écrémé 72%, caramel 10% crème, sucre, ami... Pommes 89%, purée de fraises à base de concentr... - Soja fermenté naturellement (soja, sel, eau)... Farine de FROMENT, poudre de LACTOSERUM, sucre... Débris de truffes d'hiver, jus de truffes, sel Maltodextrine, amidon modifié de pomme de terr... Graine de coriandre (19,5%), ail déshydraté, c... - 100% Semoule de BLE dur de qualité supérieur... cèpes 70% (Boletus edulis et respective famill... Haricots blancs issus de l'Agricultrue Biologique Eau, maltodextrine, sel, arômes, sucre, arôme ... Sucre, huiles végétales (Colza), noisettes (1... NaN Thon Albacore (92%), eau (7%), sel (1%) Croustilles de céréales enrobées de chocolat a... sucre; sirop de glucose; dextrose; gélatine; a... Huile de colza, eau, vinaigre d'alcool, jaunes... sel, exhausteurs de goût : glutamate, inosinat... OEUFs, farine de BLE, sucre, amidon de BLE, st... Purée de tomates mi réduite (64%), sucre, vina... Suillus luteus (40%), Volvariella volvacea (25... Filets de maquereaux (61 %), eau, concentré de... Salicornes de culture, eau, sel, acide citrique semoule de blé dur supérieure et de l'eau Semoule de blé dur (71%), lentilles vertes pré... NaN Cerises noires 65%, sucre, gélifiant : pectine... Eau, purée de tomate à base de concentré*, suc... Pêches et poires en cubes (avec leur jus d'ori... Sirop de glucose, sucre, eau, stabilisants (E4... - 100% Semoule de blé dur de qualité supérieure Farine de blé, beurre en poudre (7 %), sucre, ... SUCRE DE CANNE, EAU, JUS DE KIWI A BASE DE CON... Flageolets verts. Jus : eau, sel, affermissant... Huile de tournesol, huile de tournesol haute... Graine de courge, canneberge, graine de tourne... vinaigre d'alcool, piment jalapeno, eau, sel,...	- NESTLÉ a un système de management Cette fiche technique n'a pas de va Ce produit est une purée de fruits Gésier de dinde émincé 50%, graisse Edulcorants sorbitol, isomalt, siro Z16005 / 002\nsucre, amidon modifié Gluten\nCrustacés\nOeufs\nPoisson\nMaltodextrine, amidon de maïs, sel, Boisson gazeuse aromatisée au jus d LISTE D'INGREDIENTS\nLes ingrédient Céréales contenant du gluten (à sav mini poivrons jaunes, eau, sucre, s Les légumes secs sont mis en avant DESCRIPTION DU PRODUIT : Poudre de Ingrédients: sucre; sirop de glucos Farine de BLE, huile de colza non h MENTIONS D'ÉTIQUETAGE- LABELS MENT \n Ingrédients \nOlives vertes 98%d Pommes de terre 59,5 % - Céleris 40 Céréales contenant du gluten (2) \n Une sauce onctueuse et savoureuse a Composition typique (Données inappr Anhydride sulfureux et sulfites en Une poudre superfine sans grumeaux Pour 100 g*AR**Pour 100 g*AR**Energ * Produits issus de l'Agriculture B Ingrédients : Chocolat au lait (suc Ingrédients : pâte à tartiner aux r Sucre, farine de FROMENT, poudre d' Amidon de maïs* - Lait écrémé* - Se Les noisettes décortiquées entières gammemarkueargumentation commercial Conforme à : \n- Décret « Pâtes Ali Dénomination légale : Crème dessert Dénomination légale : Crème dessert Liste ingrédients : Pommes 89%, pur Céréales contenant du gluten \net d Farine de FROMENT, poudre de LACTOS \nOrigine Truffes et Sel : ... INFORMATIONS PRODUIT \n \nDénominat GRAINS\nMais (grain) et produits à Conforme à : \n- Décret « Pâtes Ali CODE DU PRODUIT: \nNOME DU PRODUIT: Cette fiche technique n'a pas de va Eau, maltodextrine, sel, arômes, su \nCOD. PRODUIT : CPGLLAB 49 \nDESCR 10\n70 THON ALBACORE MORCEAUX NATUREL EN \nCroustilles de céréales enrobées Ingrédients: sucre; sirop de glucos Huile de colza, eau, vinaigre d'alco Dénomination légale du produit : bo OEUFs, farine de BLE, sucre, amidon Liste des ingrédients : Purée de to Suillus luteus (40%), Volvariella v \nFormat : \n \nLieu de fabrication Se consomment en légumes d'accompag Ingrédients: semoule de blé dur sup Semoule de blé dur (71%), lentilles Boisson gazeuse aromatisée au jus d Ingrédients : Cerises noires 65%, s de maïs, sel, conservateurs : E202- INGREDIENTS: Pêches et poires en cu Sirop de glucose, sucre, eau, stabi \n- 100% Semoule de blé dur de qual Farine de blé, beurre en poudre (7 LISTE DES INGREDIENTS\nSUCRE DE CAN Flageolets verts. Jus : eau, sel, :

```
[30]: predicted['pred_len'] = predicted['predicted'].apply(len)
      sub_sample = predicted.loc[predicted['pred_len'] <= 500, ['ingredients', 'predicted']]
      sub_sample
```

[30]:

uid	ingredients	predicted
2892dd68-e3a6-474c-b543-3ebfd3490658 3634fb1e-ee79-41d1-8aaa-084c1fae5bd5 345591f4-d887-4ddc-bb40-21337fa9269d 13980d31-9002-457d-8d49-b451f08f473c 74297717-3fa8-4aed-95cc-e8737c1a6157 9458a288-04f0-49fc-9d92-d88c696ab085 6e976147-adeb-4d2d-925a-cb7c58c111a2 21233a00-bc20-40fc-acb9-ee2e2321cac2 04235024-80f3-46c2-bad0-aae0d5fab024 c3b6b4df-e586-4f10-8e58-15fbf0816acb 8d833da0-610f-4045-9877-adf8ad9ea56c aeda7bf1-4d4d-44bb-8518-2e91fc6fd76c 0481d91b-9653-42e7-b525-9dc9b87b06f2 7c709a5e-b913-4a02-9396-a6469b09482a adb90cc7-4b91-4771-b505-61a207cc1f96 484ac00a-a670-46a9-a9c4-5114174d9e3b d8b0687e-5e25-4ff9-9384-087f993218bc 46211a29-279c-4c8f-9f06-635d61f83d66 49b11281-34ea-44b0-a11c-4ae21d4c58e3 194419d0-d9f2-4799-81ac-d9e3aa77fd27 c6fd7472-f363-4aaf-8702-bd56e2e139de 8065d46f-0731-4018-9bcb-f654466badb0 e67341d8-350f-46f4-9154-4dbbb8035621 1868226c-812c-438d-bba4-6b046537d0cd 2cd5567a-bee6-46f2-a4bd-19d2b42ec001 d59d96cb-0230-4090-8220-78ce8496fd91 8dec0469-c9f5-4139-be25-efa258959444 22663c1d-2580-40b3-b4a0-ed8ba1054980 6566e18d-3bdd-43d8-ab0c-de51894621f9 b8cbe6f9-71d4-4e51-a169-1c163d49a561 e51b7fd6-d878-47f8-a36b-f10f8d4087bd 046c9db1f-1915-4916-8874-902cc5ec73be a049d2df-9c76-4303-8813-65ec5ccbf70 f1982954-f091-4aa1-addb-6659062e0f66 6c07fd47-116d-4531-a164-d4f4040f212e f45db604-11ad-4756-aeab-3a5a1a34f914 c8b299c5-22d3-40ae-8df3-ad638507803f 09e45b38-4da1-4eb5-888a-3ebd437a2291 41da4d6f-7e9f-4f95-bf2a-2acd7138cd9 14c9068c-633a-4eea-bb4a-6ac5de0e9a32 6622442d-d88f-4b90-9414-b91c9e8e5296 a2418174-e16a-41e0-ac14-c87208fb3529 5adc7512-6168-4966-ae3f-f6ec133bf56e 75088d85-f350-4d81-a7f4-954411ba089e 064d3b22-c1cc-41bd-bd03-56e8829bd4cd c5344fe1-ac16-4630-824e-6036ae491399 2286f782-9d2e-410f-84c4-4ab88003a002 4f83306f-66de-4545-9b12-7790b57b61ae 42094f22-5907-4089-bdaf-0a19ff09b3fe 8afeb736-1b04-4aa7-9b25-fa534a177019 e5ebb1e3-0a86-46cb-9d76-3a965cf20c4b 5cee689e-fbf1-b232-1d8f-b1f88a57 3624dd37-f88e-47c0-9e9c-bb5163bfbaca 1268d9ef-d848-41ad-973c-7f9e694a4625 6bdb201d-5879-423b-96b8-3ae88b857818 1de02b1c-f17e-4d46-b90f-3a6c37ecf6aa 536361db-1bbb-4e64-aa53-d970eeac7db2 e521bd01-f2bb-4e00-9ae0-0151a1c7a047 8278ffc7-c82b-44b9-9e4d-9150943faaf1 63968dc3-6e7c-4056-bd53-820c6cc925be d1be6f74-1e0e-4631-bb4a-6b16b9fc908f 64836d33-5b4f-42ff-862a-31ee19cefda4 bb77c7b0-9c63-4869-9ab1-823ba1158f53 b7d7621a-fcdd-4487-9b38-e07fae698c4a dc536305-82fd-4afe-a472-5056ca0e21ea df1caa23-9714-4659-803b-33501d64eead 272246a4-d79d-47fc-a475-7b873b6bb47f 6dfae8fd-6111-4a57-862e-c20a39a195e0 8f9f7f4e-da12-4cc2-9d02-b9eb06d659f4 9ef0d351-4982-4a2d-88a9-85573dc396dc ebfc9e73-5d91-4b45-8331-8c8f9bed3bb3 c33aa83e-a502-4339-a8e0-c56db2e59e69 54f40033-f9cf-411c-81a5-11974f6715aa	Café instantané, café torréfié moulu (3%). Poire 99,9%, antioxydant: acide ascorbique. Gésier de dinde émincé 50%, graisse de canard ... Edulcorants sorbitol, isomalt, sirop de maltit... sucre, amidon modifié, LACTOSÉRUM en poudre, d... paprika, poivre noir, coriandre, cumin, ail se... Maltodextrine, amidon de maïs, sel, farine de ... NaN Persil mini poivrons jaunes, eau, sucre, sel, affermi... 100% haricots blancs bio HAPPY CHERRY\nsirop de glucose; sucre; gélatin... Farine de BLE, huile de colza non hydrogénée, ... SIROP DE GLUCOSE / SUCRE / AROMES / ACIDIFIANT... Olives vertes 98%denoyautées , huile tournesol... Pommes de terre 59,5 % - Céleris 40 % - Amidon... Eau ; jus de fruit de la passion ; sucre ; épa... pâte de cacao 50,5% ; sucre 47,5% ; beurre de ... NaN Pommes de terre, eau, sel. Farine de pois chiche Lentilles, eau, sel, arôme naturel (céleri) Sucre roux de canne* (64%), amidon de maïs*, ... pâte à tartiner aux noisettes et au cacao 81,5... Sucre, farine de FROMENT, poudre d'OEUFs entie... Amidon de maïs* - Lait écrémé* - Sel - Féculé ... Sucre; sirop de glucose; graisse de palme; hum... Lait écrémé 72%, caramel 10% crème, sucre, ami... Pommes 89%, purée de fraises à base de concentr... Farine de FROMENT, poudre de LACTOSERUM, sucre... Débris de truffes d'hiver, jus de truffes, sel cèpes 70% (Boletus edulis et respective famill... Eau, maltodextrine, sel, arômes, sucre, arôme ... NaN Thon Albacore (92%), eau (7%), sel (1%) sucre; sirop de glucose; dextrose; gélatine; a... Huile de colza, eau, vinaigre d'alcool, jaunes... OEUFs, farine de BLE, sucre, amidon de BLE, st... Purée de tomates mi réduite (64%), sucre, vina... Suillus luteus (40%), Volvariella volvacea (25... Filets de maquereaux (61 %), eau, concentré de... Salicornes de culture, eau, sel, acide citrique semoule de blé dur supérieure et de l'eau NaN Cerises noires 65%, sucre, gélifiant : pectine... Eau, purée de tomate à base de concentré*, suc... Pêches et poires en cubes (avec leur jus d'ori... Sirop de glucose, sucre, eau, stabilisants (E4... - 100% Semoule de blé dur de qualité supérieure Farine de blé, beurre en poudre (7 %), sucre, ... SUCRE DE CANNE, EAU, JUS DE KIWI A BASE DE CON... Flageolets verts. Jus : eau, sel, affermissant... Huile de tournesol, huile de tournesol haute... Graine de courge, canneberge, graine de tourne... vinaigre d'alcool, piment jalapeno, eau, sel, ... AMANDES décortiquées Sucre, amidon de maïs, arôme vanille Sucre, cacao maigre en poudre (beurre de cacao... Tofu* 76.6% (eau, soja* dépelliculé 20.7%, gél... Carottes, eau, sucre, sel, vinaigre d'alcool, ... sucre*, LAIT en poudre*, beurre de cacao*, pât... NaN Tilleul (100%). Eau, haricots verts, sel. Légumes 43,2 % (pomme de terre, oignon, carott... sucre, pâte de cacao, beurre de cacao, cacao m... Maltodextrine, édulcorant : sucralose (1,23%). Farine de BLÉ 63%, eau, sucre, huile de colza,... Graines de sésame 37%, chocolat de couverture ... Sirop de glucose-fructose, framboises 35%, suc... Jus d'orange à base de concentré Farine de BLÉ, sucre, huile de colza,, cacao m... Piment rouge fort équeuté* (85%), cumin, ail m...	- NESTLÉ a un système de management Ce produit est une purée de fruits Gésier de dinde émincé 50%, graisse Edulcorants sorbitol, isomalt, siro Z16005 / 002\nsucre, amidon modifié Gluten\nCrustacés\nŒufs\nPoisson\nMaltodextrine, amidon de maïs, sel, Boisson gazeuse aromatisée au jus d Céréales contenant du gluten (à sav mini poivrons jaunes, eau, sucre, s Les légumes secs sont mis en avant Ingrédients: sucre; sirop de glucos Farine de BLE, huile de colza non h MENTIONS D'ÉTIQUETAGE- LABELS MENT \n Ingrédients \nOlives vertes 98%d Pommes de terre 59,5 % - Céleris 40 Céréales contenant du gluten (2) \n Composition typique (Données inappr Anhydride sulfureux et sulfites en Une poudre superfine sans grumeaux Pour 100 g*AR**Pour 100 g*AR**Energ * Produits issus de l'Agriculture B Ingrédients : pâte à tartiner aux r Sucre, farine de FROMENT, poudre d' Amidon de maïs* - Lait écrémé* - Se Dénomination légale : Crème dessert Liste ingrédients : Pommes 89%, pur Farine de FROMENT, poudre de LACTOS \nOrigine Truffes et Sel : ... CODE DU PRODUIT: \nNOME DU PRODUIT: Eau, maltodextrine, sel, arômes, su 10\n70 THON ALBACORE MORCEAUX NATUREL EN Ingrédients: sucre; sirop de glucos Huile de colza, eau, vinaigre d'alco OEUFs, farine de BLE, sucre, amidon Liste des ingrédients : Purée de to Suillus luteus (40%), Volvariella v \nFormat : \n \nLieu de fabrication Se consomment en légumes d'accompag Ingrédients: semoule de blé dur sup Boisson gazeuse aromatisée au jus d Ingrédients : Cerises noires 65%, s de maïs, sel, conservateurs : E202- INGREDIENTS: Pêches et poires en cu Sirop de glucose, sucre, eau, stabi \n- 100% Semoule de blé dur de qual Farine de blé, beurre en poudre (7 LISTE DES INGREDIENTS\nSUCRE DE CAN Flageolets verts. Jus : eau, sel, s Gluten\nCrustacés\nŒufs\nPoisson\nListe d'ingrédients : vinaigre d'al Valeur énergétique : \n Matières gr ajouter le produit à la préparation Sucre, cacao maigre en poudre (beur Si le produit est vendu en Espagne Carottes, eau, sucre, sel, vinaigre Liste des Ingrédients:\nsucre*, LAIT Les points forts de ce riz \n* Un p Dans le cadre des recommandations d Égoutter, ne pas\nrincer. Faire sau Légumes 43,2 % (pomme de terre, oig Liste des Ingrédients:\nsucre, pâte liste des ingrédients :\nMaltodextr Farine de BLÉ 63%, eau, sucre, huil Croquant sésame chocolat\nBarre cro Liste ingrédients : Sirop de glucos \n\nVALEURS NUTRITIONNELLES pour 1 Ingrédients : Farine de BLÉ, sucre, A) Ingrédients : \n \n Piment rouge

On constitue une table pour intégration dans le rapport :

```
[31]: with pd.option_context("max_colwidth", 100000):
      tex_str = (
          sub_sample.sample(20, random_state=41)
              .replace(r'\s*$', np.nan, regex=True)
              .to_latex(index=False,
                        index_names=False,
                        column_format='p{7cm}p{7cm}',
                        na_rep='<rien>',
                        longtable=True,
                        header=["Liste d'ingrédients cible", "Liste d'ingrédients prédite"],
                        label='tbl:GT_prediction_sample',
                        caption="Extrait des résultats de la prédiction",

                        )
              .replace(r'\textbackslash n', r' \newline ')
              .replace(r'\\', r'\\ \hline')
      )

# with open(Path('.') / 'tbls' / 'GT_prediction_sample.tex', 'w') as file:
#     file.write(tex_str)
```