

Performance measurement

Pierre MASSÉ

May 8, 2020

1 Mesure de la performance du modèle

L'objet de ce notebook est d'illustrer la méthodologie de mesure de la performance du modèle.

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)

[46]: # imports and customization of display
import os
from functools import partial
import numpy as np
import pandas as pd
pd.options.display.min_rows = 6
pd.options.display.width=108
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score, cross_validate
from sklearn.pipeline import Pipeline
from matplotlib import pyplot as plt

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
from src.pimest import custom_accuracy
```

1.2 Acquisition des données

On récupère les données manuellement étiquetées et on les intègre dans un dataframe

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

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)
```

```
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= 37.4s
```

```
[3]:
```

	designation \
uid	
a0492df6-9c76-4303-8813-65ec5ccbfa70	Concentré liquide Asian en bouteille 980 ml CHEF
d183e914-db2f-4e2f-863a-a3b2d054c0b8	Pain burger curry 80 g CREATIV BURGER
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	Macaroni en sachet 500 g PANZANI
...	...
e67341d8-350f-46f4-9154-4dbbb8035621	PRÉPARATION POUR CRÈME BRÛLÉE BIO 6L
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3	Céréales instantanées en poudre saveur caramel...
Ofaad739-ea8c-4f03-b62e-51ee592a0546	FARINE DE BLÉ TYPE 45, 10KG
	ingredients \
uid	
a0492df6-9c76-4303-8813-65ec5ccbfa70	Eau, maltodextrine, sel, arômes, sucre, arôme ...
d183e914-db2f-4e2f-863a-a3b2d054c0b8	Farine de blé T65, eau, levure, vinaigre de ci...
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	- 100% Semoule de BLE dur de qualité supérieur...
...	...
e67341d8-350f-46f4-9154-4dbbb8035621	Sucre roux de canne*° (64%), amidon de maïs*, ...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3	Farine 87,1 % (Blé (GLUTEN), Blé hydrolysé (GL...
Ofaad739-ea8c-4f03-b62e-51ee592a0546	Farine de blé T45
	path \
uid	
a0492df6-9c76-4303-8813-65ec5ccbfa70	../../ground_truth/a0492df6-9c76-4303-8813-65e...
d183e914-db2f-4e2f-863a-a3b2d054c0b8	../../ground_truth/d183e914-db2f-4e2f-863a-a3b...
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	../../ground_truth/ab48a1ed-7a3d-4686-bb6d-ab4...
...	...
e67341d8-350f-46f4-9154-4dbbb8035621	../../ground_truth/e67341d8-350f-46f4-9154-4db...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3	../../ground_truth/a8f6f672-20ac-4ff8-a8f2-3bc...
Ofaad739-ea8c-4f03-b62e-51ee592a0546	../../ground_truth/Ofaad739-ea8c-4f03-b62e-51e...
	content \
uid	
a0492df6-9c76-4303-8813-65ec5ccbfa70	b'%PDF-1.5\r\n%\xb5\xb5\xb5\b5\r\n1 0 obj\r\n...
d183e914-db2f-4e2f-863a-a3b2d054c0b8	b'%PDF-1.5\r%\xe2\xe3\xcf\xd3\r\n4 0 obj\r<</L...
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	b'%PDF-1.4\n%\xc7\xec\x8f\xa2\n5 0 obj\n<</Len...
...	...
e67341d8-350f-46f4-9154-4dbbb8035621	b'%PDF-1.7\r\n%\xb5\xb5\xb5\b5\r\n1 0 obj\r\n...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3	b'%PDF-1.5\r\n%\xb5\xb5\xb5\b5\r\n1 0 obj\r\n...
Ofaad739-ea8c-4f03-b62e-51ee592a0546	b'%PDF-1.5\r\n%\xb5\xb5\b5\b5\r\n1 0 obj\r\n...
	text
uid	
a0492df6-9c76-4303-8813-65ec5ccbfa70	Concentré Liquide Asian CHEF® \n\nBouteille de...
d183e914-db2f-4e2f-863a-a3b2d054c0b8	
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8	Direction Qualité \n\n \n\n \n\nPATES ALIMENTA...
...	...
e67341d8-350f-46f4-9154-4dbbb8035621	FICHE TECHNIQUE \n\nCREME BRÛLÉE 6L \n\nREF : ...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3	81 rue de Sans Souci - CS13754 - 69576 Limones...
Ofaad739-ea8c-4f03-b62e-51ee592a0546	\n1050/10502066400 \n\n10502055300/1050202520...

[500 rows x 5 columns]

On splitte les textes en blocs de manière basique.

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

```
split_transfo = BlockSplitter(splitter_func=splitter)
splitted_df = split_transfo.fit_transform(texts_df)
splitted_df
```

Launching 8 processes.

```
[4]:
uid
a0492df6-9c76-4303-8813-65ec5ccbfa70 Concentré liquide Asian en bouteille 980 ml CHEF
d183e914-db2f-4e2f-863a-a3b2d054c0b8 Pain burger curry 80 g CREATIV BURGER
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8 Macaroni en sachet 500 g PANZANI
...
e67341d8-350f-46f4-9154-4dbbb8035621 PRÉPARATION POUR CRÈME BRÛLÉE BIO 6L
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3 Céréales instantanées en poudre saveur caramel...
0faad739-ea8c-4f03-b62e-51ee592a0546 FARINE DE BLÉ TYPE 45, 10KG

ingredients \
uid
a0492df6-9c76-4303-8813-65ec5ccbfa70 Eau, maltodextrine, sel, arômes, sucre, arôme ...
d183e914-db2f-4e2f-863a-a3b2d054c0b8 Farine de blé T65, eau, levure, vinaigre de ci...
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8 - 100% Semoule de BLE dur de qualité supérieur...
...
e67341d8-350f-46f4-9154-4dbbb8035621 Sucre roux de canne* (64%), amidon de maïs*, ...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3 Farine 87,1 % (Blé (GLUTEN), Blé hydrolysé (GL...
0faad739-ea8c-4f03-b62e-51ee592a0546 Farine de blé T45

path \
uid
a0492df6-9c76-4303-8813-65ec5ccbfa70 ../../ground_truth/a0492df6-9c76-4303-8813-65e...
d183e914-db2f-4e2f-863a-a3b2d054c0b8 ../../ground_truth/d183e914-db2f-4e2f-863a-a3b...
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8 ../../ground_truth/ab48a1ed-7a3d-4686-bb6d-ab4...
...
e67341d8-350f-46f4-9154-4dbbb8035621 ../../ground_truth/e67341d8-350f-46f4-9154-4db...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3 ../../ground_truth/a8f6f672-20ac-4ff8-a8f2-3bc...
0faad739-ea8c-4f03-b62e-51ee592a0546 ../../ground_truth/0faad739-ea8c-4f03-b62e-51e...

content \
uid
a0492df6-9c76-4303-8813-65ec5ccbfa70 b'%PDF-1.5\r\n%\xb5\xb5\xb5\r\n1 0 obj\r\n...
d183e914-db2f-4e2f-863a-a3b2d054c0b8 b'%PDF-1.5\r\n%\xe2\xe3\xcf\xd3\r\n4 0 obj\r\n<</L...
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8 b'%PDF-1.4\n%\xc7\xec\x8f\xa2\n5 0 obj\n<</Len...
...
e67341d8-350f-46f4-9154-4dbbb8035621 b'%PDF-1.7\r\n%\xb5\xb5\xb5\r\n1 0 obj\r\n...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3 b'%PDF-1.5\r\n%\xb5\xb5\xb5\r\n1 0 obj\r\n...
0faad739-ea8c-4f03-b62e-51ee592a0546 b'%PDF-1.5\r\n%\xb5\xb5\xb5\r\n1 0 obj\r\n...

text \
uid
a0492df6-9c76-4303-8813-65ec5ccbfa70 Concentré Liquide Asian CHEF® \n\nBouteille de...
d183e914-db2f-4e2f-863a-a3b2d054c0b8
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8 Direction Qualité \n\n \n\n \n\nPATES ALIMENTA...
...
e67341d8-350f-46f4-9154-4dbbb8035621 FICHE TECHNIQUE \n\nCREME BRÛLÉE 6L \n\nREF : ...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3 81 rue de Sans Souci - CS13754 - 69576 Limones...
0faad739-ea8c-4f03-b62e-51ee592a0546 \n1050/10502066400 \n\n10502055300/10502025200...

blocks
uid
a0492df6-9c76-4303-8813-65ec5ccbfa70 [Concentré Liquide Asian CHEF®, Bouteille de ...
d183e914-db2f-4e2f-863a-a3b2d054c0b8 [
]
ab48a1ed-7a3d-4686-bb6d-ab4f367cada8 [Direction Qualité , , , PATES ALIMENTAIRES ...
...
e67341d8-350f-46f4-9154-4dbbb8035621 [FICHE TECHNIQUE , CREME BRÛLÉE 6L , REF : NAP...
a8f6f672-20ac-4ff8-a8f2-3bc4306c8df3 [81 rue de Sans Souci - CS13754 - 69576 Limone...
0faad739-ea8c-4f03-b62e-51ee592a0546 [ \n1050/10502066400 , 10502055300/10502025200...
```

[500 rows x 6 columns]

1.3 Train/Test split, entraînement et tranformation

On effectue classiquement les étapes de train/test split, on entraîne le modèle sur le set d'entraînement et on le lance sur le set de test.

```
[5]: train, test = train_test_split(splitted_df, train_size=400, random_state=42)
model = SimilaritySelector(similarity='projection')
model.fit(train['blocks'], train['ingredients'])
predicted = pd.Series(model.predict(test['blocks']),
                      index=test.index,
                      name='predicted')
predicted = pd.concat([test['ingredients'], predicted], axis=1)
predicted
```

```
[5]:
```

	ingredients \
uid	
2892dd68-e3a6-474c-b543-3ebfd3490658	Café instantané, café torréfié moulu (3%).
a57c1561-b88e-4694-8bd8-55623f2afa17	Lentilles blondes
3634fb1e-ee79-41d1-8aaa-084c1fae5bd5	Poire 99,9%, antioxydant: acide ascorbique.
...	...
ebfc9e73-5d91-4b45-8331-8c8f9bed3bb3	Jus d'orange à base de concentré
c33aa83e-a502-4339-a8e0-c56db2e59e69	Farine de BLÉ, sucre, huile de colza,, cacao m...
54f40033-f9cf-411c-81a5-11974f6715aa	Piment rouge fort équeuté* (85%), cumin, ail m...
	predicted
uid	
2892dd68-e3a6-474c-b543-3ebfd3490658	- NESTLÉ a un système de management de la qual...
a57c1561-b88e-4694-8bd8-55623f2afa17	Cette fiche technique n'a pas de valeur contra...
3634fb1e-ee79-41d1-8aaa-084c1fae5bd5	Ce produit est une purée de fruits obtenue à p...
...	...
ebfc9e73-5d91-4b45-8331-8c8f9bed3bb3	\n \nVALEURS NUTRITIONNELLES pour 100mL / NUT...
c33aa83e-a502-4339-a8e0-c56db2e59e69	Ingrédients : Farine de BLÉ, sucre, huile de c...
54f40033-f9cf-411c-81a5-11974f6715aa	A) Ingrédients : \n \nPiment rouge fort équ...

[100 rows x 2 columns]

1.4 Mesure de la performance : Précision

1.4.1 Approche naïve

Dans cette première version, on calculera une précision brute, où seuls les strings parfaitement identiques sont considérés comme ok.

```
[6]: predicted['result'] = (predicted['ingredients'].fillna('') == predicted['predicted'].fillna(''))
predicted['result'].value_counts()
```

```
[6]: False    99
      True     1
      Name: result, dtype: int64
```

On a une précision très faible, 1%. L'unique liste d'ingrédients du set de test correctement prédite est la suivante :

```
[7]: print(predicted[predicted['result']].iloc[0, 0])
```

Sirop de glucose, sucre, eau, stabilisants (E440i, E440ii, E415), acidifiants (E330, E450i), conversateur (E202).

1.4.2 Cross-validation de l'approche naïve

Pour avoir une vision plus précise de la performance du modèle, on peut effectuer une cross-validation sur le set d'entraînement.

On commence par définir une fonction de scoring, qui pourra être appelée par la fonction standard de cross-validation de scikit-learn. Comme précédemment, il s'agit d'une fonction d'accuracy basique :

```
[8]: def accuracy_scorer(estim, X, y):
      y_pred = estim.predict(X)
      return((y_pred == y).mean())
```

On retrouve évidemment le même score que précédemment lorsqu'on utilise cette fonction sur le set de test :

```
[9]: accuracy_scorer(model, test.reset_index()['blocks'], test.reset_index()['ingredients'])
```

```
[9]: 0.01
```

Si on lance la cross-validation avec les paramètres par défaut (cv=5), on obtient le résultat suivant :

```
[11]: X = splitted_df.reset_index()['blocks'].copy()
      y = splitted_df.reset_index()['ingredients'].copy()

      cross_val = cross_validate(model,
                                X=X,
                                y=y,
                                scoring=accuracy_scorer,
                                )

      print(f'Strict accuracy yields a result of {np.mean(cross_val["test_score"]):.2%} +/-{np.
            ↳std(cross_val["test_score"]):.2%}')
      print(cross_val['test_score'])
```

```
Strict accuracy yields a result of 2.20% +/-0.75%
[0.03 0.03 0.02 0.01 0.02]
```

On voit que sur chacun des 5 folds (validation sur 80 produits), l'accuracy varie entre 1 et 3%.

Si on trace l'accuracy et la standard deviation pour plusieurs valeurs de cv, on obtient les résultats suivants :

```
[12]: x = [3, 5, 8, 10, 15, 20, 30, 50, 70, 100]
      mean = np.array([])
      std = np.array([])
      for n_cv in x:
          cross_val = cross_validate(model,
                                    X=splitted_df['blocks'],
                                    y=splitted_df['ingredients'],
                                    scoring=accuracy_scorer,
                                    cv=n_cv,
                                    )

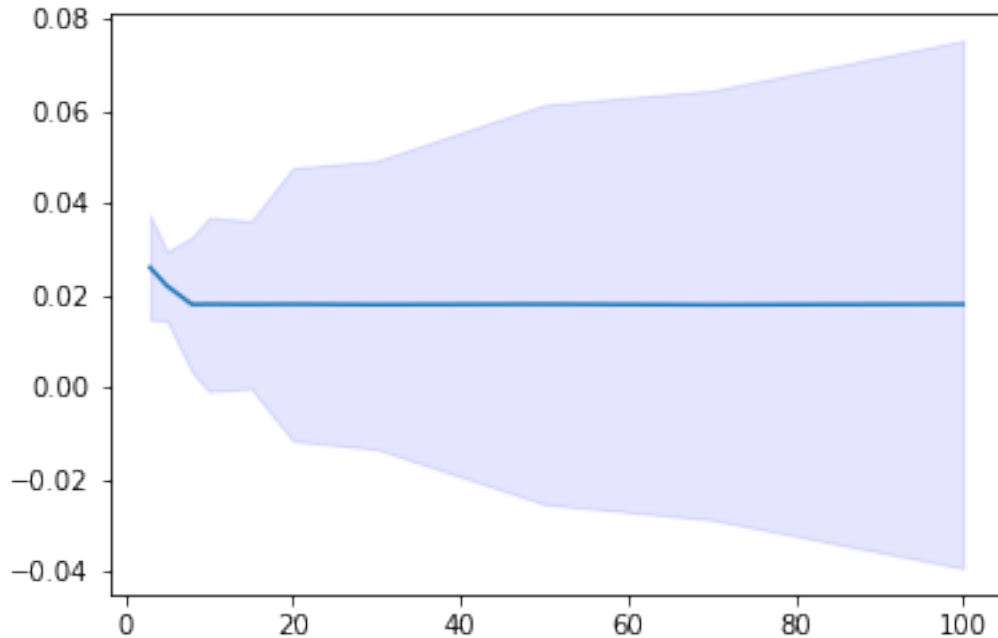
          mean = np.append(mean, [np.mean(cross_val['test_score'])], axis=0)
          std = np.append(std, [np.std(cross_val['test_score'])], axis=0)

      print('mean:', mean, '\nstandard dev:', std)

mean: [0.02598418 0.022      0.01795315 0.018      0.01794415 0.018
       0.01789216 0.018      0.01785714 0.018      ]
standard dev: [0.01126571 0.00748331 0.01470225 0.01886796 0.01816919 0.0295973
              0.03127858 0.04331282 0.04656573 0.05723635]
```

```
[13]: fig, ax = plt.subplots()
      ax.plot(x, mean)
      ax.fill_between(x, (mean - std), (mean + std), color='b', alpha=.1)
```

```
[13]: <matplotlib.collections.PolyCollection at 0x7fa08009e3a0>
```



Il apparaît que l'accuracy se situe aux alentours de 2%, avec un écart type important si on le compare à cette accuracy.

```
[14]: cross_val = cross_validate(model,
                                X=splitted_df['blocks'],
                                y=splitted_df['ingredients'],
                                scoring=accuracy_scorer,
                                cv=10,
                                )

print(f'Strict accuracy yields a result of {np.mean(cross_val["test_score"]):.2%} +/-{np.
      ↳std(cross_val["test_score"]):.2%}')
print(cross_val['test_score'])
```

```
Strict accuracy yields a result of 1.80% +/-1.89%
[0.04 0.  0.  0.06 0.02 0.02 0.  0.02 0.  0.02]
```

1.4.3 Ajout d'une étape de text-postprocessing

On utilise la fonction `custom_accuracy` définie dans le module `pimest` pour calculer l'accuracy avec du text processing. Elle prend en paramètre les mêmes attributs que le `CountVectorizer` de scikit-learn, en plus d'un attribut "tokenize" qui va tokenizer le résultat (pour prise en compte des whitespace et de la ponctuation).

```
[31]: custom_accuracy(model,
                      test['blocks'].fillna(''),
                      test['ingredients'].fillna(''),
                      tokenize=True,
                      strip_accents='unicode',
                      lowercase=True,
                      )
```

```
[31]: 0.14
```

L'accuracy est maintenant estimée à 14% (vs. 1%) sur le set de test, après entraînement sur le set d'entraînement.

On peut manuellement inspecter les blocks identique, en reproduisant le comportement de la fonction d'accuracy :

```
[49]: def text_processor(text, **kwargs):
        unused_model = CountVectorizer(**kwargs)
```

```

prepro = unused_model.build_preprocessor()
token = unused_model.build_tokenizer()
return(' '.join(token(prepro(text))))

partial_processor = partial(text_processor, strip_accents='unicode', lowercase=True)

```

```

[61]: prediction = model.predict(test['blocks'].fillna('')).rename('predicted')
processed_prediction = prediction.apply(partial_processor)
processed_prediction.head(3)

```

```

[61]: uid
2892dd68-e3a6-474c-b543-3ebfd3490658    nestle un systeme de management de la qualite ...
a57c1561-b88e-4694-8bd8-55623f2afa17    cette fiche technique pas de valeur contractue...
3634fb1e-ee79-41d1-8aaa-084c1fae5bd5    ce produit est une puree de fruits obtenue par...
Name: predicted, dtype: object

```

```

[62]: processed_ground_truth = test['ingredients'].fillna('').apply(partial_processor)
processed_ground_truth.head(3)

```

```

[62]: uid
2892dd68-e3a6-474c-b543-3ebfd3490658    cafe instantane cafe torrefie moulu
a57c1561-b88e-4694-8bd8-55623f2afa17    lentilles blondes
3634fb1e-ee79-41d1-8aaa-084c1fae5bd5    poire 99 antioxydant acide ascorbique
Name: ingredients, dtype: object

```

```

[63]: corrects = test.join(prediction).loc[processed_prediction == processed_ground_truth , ['ingredients',
↪ 'predicted']]
corrects

```

```

[63]:                                     ingredients \

uid
345591f4-d887-4ddc-bb40-21337fa9269d    Gésier de dinde émincé 50%, graisse de canard ...
13980d31-9002-457d-8d49-b451f08f473c    Edulcorants sorbitol, isomalt, sirop de maltit...
c3b6b4df-e586-4f10-8e58-15fbf0816acb    mini poivrons jaunes, eau, sucre, sel, affermi...
0481d91b-9653-42e7-b525-9dc9b87b06f2    Farine de BLE, huile de colza non hydrogénée, ...
484ac00a-a670-46a9-a9c4-5114174d9e3b    Pommes de terre 59,5 % - Céleris 40 % - Amidon...
49b11281-34ea-44b0-a11c-4ae21d4c58e3    NaN
d59d96cb-0230-4090-8220-78ce8496fd91    Amidon de maïs* - Lait écrémé* - Sel - Fécul...
b8cbe6f9-71d4-4e51-a169-1c163d49a561    Farine de FROMENT, poudre de LACTOSERUM, sucre...
a0492df6-9c76-4303-8813-65ec5ccbfa70    Eau, maltodextrine, sel, arômes, sucre, arôme ...
09e45b38-4da1-4eb5-888a-3ebd437a2291    OEUFs, farine de BLE, sucre, amidon de BLE, st...
4f83306f-66de-4545-9b12-7790b57b61ae    Sirop de glucose, sucre, eau, stabilisants (E4...
5cee689e-6fb1-493c-b232-1d8fb1f88a57    Flageolets verts. Jus : eau, sel, affermissant...
63968dc3-6e7c-4056-bd53-820c6cc925be    Carottes, eau, sucre, sel, vinaigre d'alcool, ...
dc536305-82fd-4afe-a472-5056ca0e21ea    Légumes 43,2 % (pomme de terre, oignon, carott...

                                     predicted

uid
345591f4-d887-4ddc-bb40-21337fa9269d    Gésier de dinde émincé 50%, graisse de canard...
13980d31-9002-457d-8d49-b451f08f473c    Edulcorants sorbitol, isomalt, sirop de maltit...
c3b6b4df-e586-4f10-8e58-15fbf0816acb    mini poivrons jaunes, eau, sucre, sel, affermi...
0481d91b-9653-42e7-b525-9dc9b87b06f2    Farine de BLE, huile de colza non hydrogénée, ...
484ac00a-a670-46a9-a9c4-5114174d9e3b    Pommes de terre 59,5 % - Céleris 40 % - Amidon...
49b11281-34ea-44b0-a11c-4ae21d4c58e3

d59d96cb-0230-4090-8220-78ce8496fd91    Amidon de maïs* - Lait écrémé* - Sel - Fécul...
b8cbe6f9-71d4-4e51-a169-1c163d49a561    Farine de FROMENT, poudre de LACTOSERUM, sucre...
a0492df6-9c76-4303-8813-65ec5ccbfa70    Eau, maltodextrine, sel, arômes, sucre, arôme ...
09e45b38-4da1-4eb5-888a-3ebd437a2291    OEUFs, farine de BLE, sucre, amidon de BLE, st...
4f83306f-66de-4545-9b12-7790b57b61ae    Sirop de glucose, sucre, eau, stabilisants (E4...
5cee689e-6fb1-493c-b232-1d8fb1f88a57    Flageolets verts. Jus : eau, sel, affermissant...
63968dc3-6e7c-4056-bd53-820c6cc925be    Carottes, eau, sucre, sel, vinaigre d'alcool, ...
dc536305-82fd-4afe-a472-5056ca0e21ea    Légumes 43,2 % (pomme de terre, oignon, carott...

```

```
[64]: with pd.option_context("max_colwidth", 100000):
    tex_str = (
        corrects.replace(r'\s*$', np.nan, regex=True)
        .to_latex(index=False,
                    index_names=False,
                    column_format='p{7cm}p{7cm}',
                    na_rep='<rien>',
                    longtable=False,
                    header=["Liste d'ingrédients cible", "Liste d'ingrédients prédite"],
                    # label='tbl:GT_postprocessed_corrects',
                    # caption="Prédictions identifiées comme correctes après postprocessing",
        )
        .replace(r'\textbackslash n', r' \newline ')
        .replace(r'\\', r'\\ \hline')
    )

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