



Projet 6: Classifiez automatiquement des biens de consommation

Anissa TALEB

Février 2024



objectif :

- ⇒ une étude de faisabilité d'un moteur de classification d'articles, basé sur une image et une description, pour l'automatisation de l'attribution de la catégorie de l'article.
- ⇒ classification supervisée à partir des images
- ⇒ test d'API



PLAN :

I Etude de faisabilité de classification :

- 1 Données textuelles
- 2 Données images

II Classification supervisée des images

III Test d'API



I Etude de faisabilité de classification :

1 Données textuelles



Analyse des descriptions textuelles :

1 Prétraitement des données : NLP

- sur les deux colonnes “product_category_tree ” et “description”
- Création de bag of word (tokenization, stopWords, lowerCase, lemmatization)
⇒ TF_iDF , Word2vec et CountVectorizer

Analyse des descriptions textuelles :

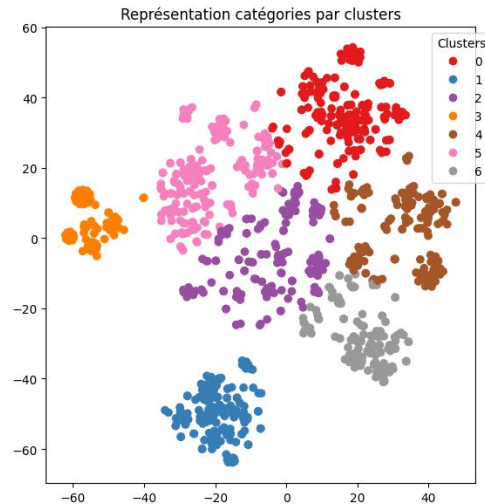
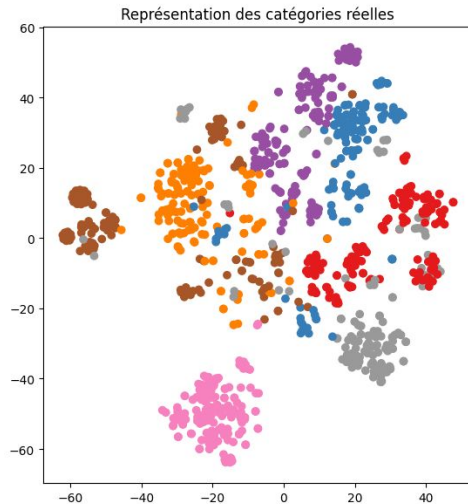
2 Création des vecteurs : CountVectorizer et TF-idf

- Bag of words `max length bow : 379`
- Réduction de dimension : APC et TSNE
- Calcul de l'indice de similarité : ARI

```
➡ CountVectorizer :  
-----  
ARI : 0.4174 time : 18.0  
  
Tf-idf :  
-----  
ARI : 0.451 time : 16.0
```

Analyse des descriptions textuelles :

- Analyse graphique : TF-idf



⇒ Classification par
vectorisation/bag of words
est possible !



Analyse des descriptions textuelles :

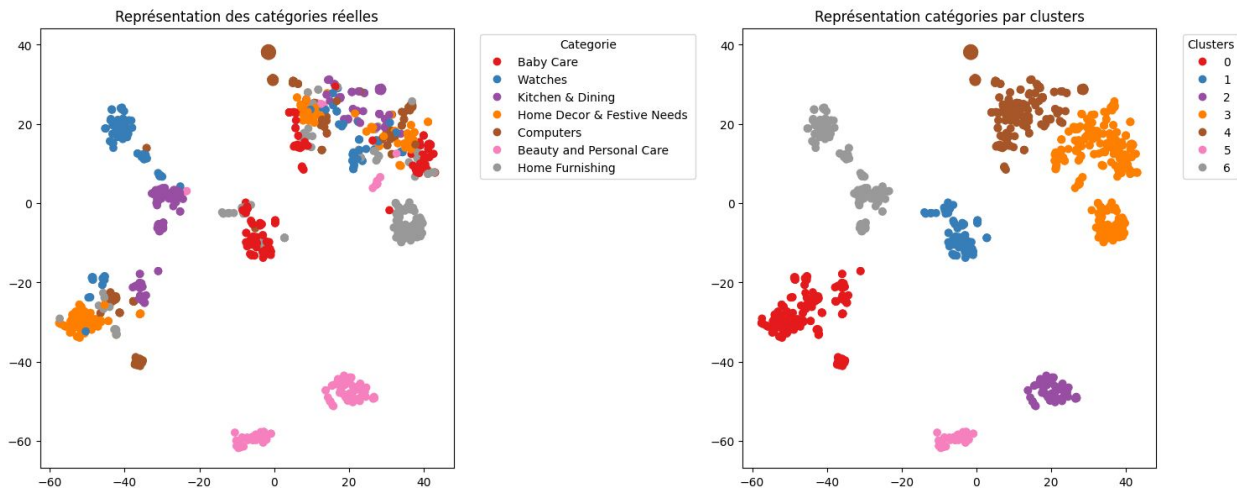
3 Word2Vec : import tensorflow.keras

- Création de modèle
- Création de la matrice d'embedding
- Création du modèle d'embedding
- Exécution du modèle
- Réduction de dimension APC et TSN-E et Calcul de l'ARI

```
ARI : 0.2159 time : 13.0
```


Analyse des descriptions textuelles :

- Analyse graphique



⇒ faisabilité médiocre
en vue de ARI < 0.4



Analyse des descriptions textuelles :

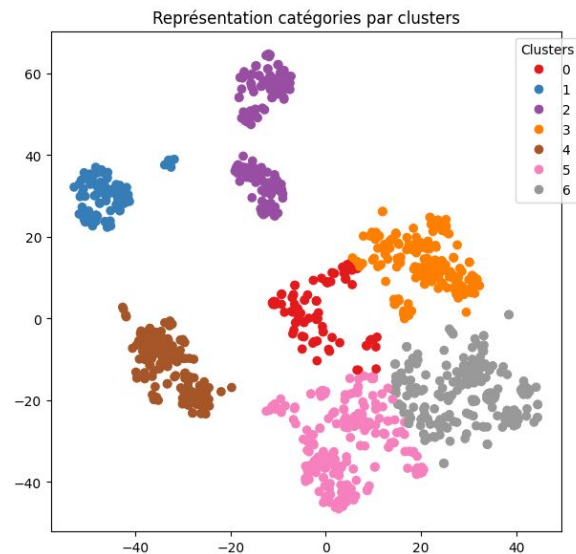
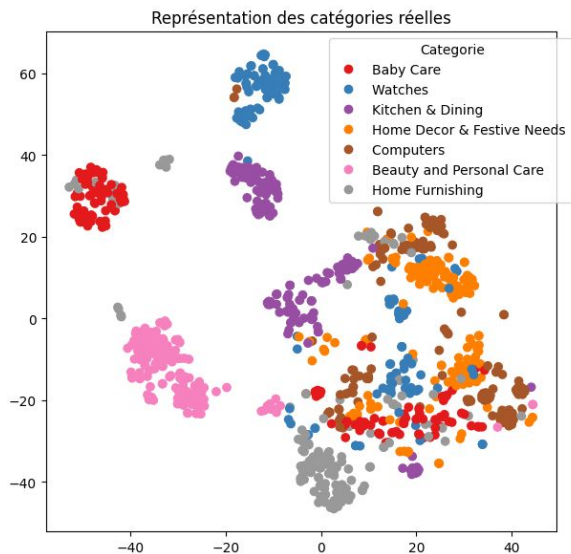
4 BERT HuggingFace :

- Création des features
- Réduction de dimension APC et TSN-E et Calcul de l'ARI

➡ ARI : 0.3203 time : 14.0

Analyse des descriptions textuelles :

- Analyse graphique:





Analyse des descriptions textuelles :

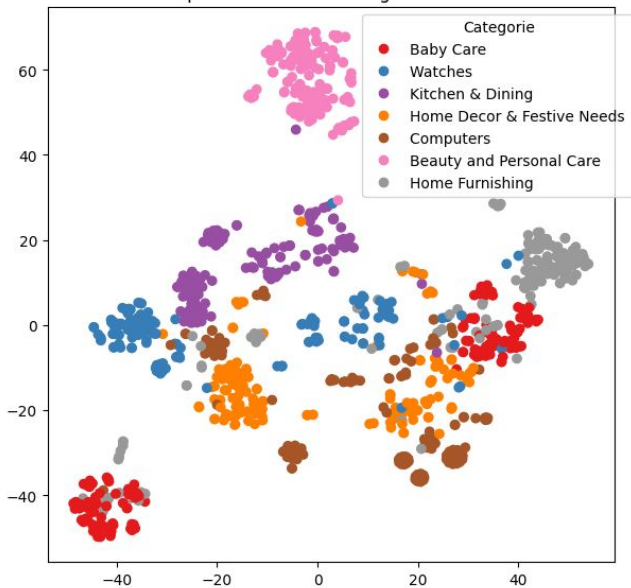
5 USE - Universal Sentence Encoder

- Création des features
- Réduction de dimension APC et TSN-E et Calcul de l'ARI

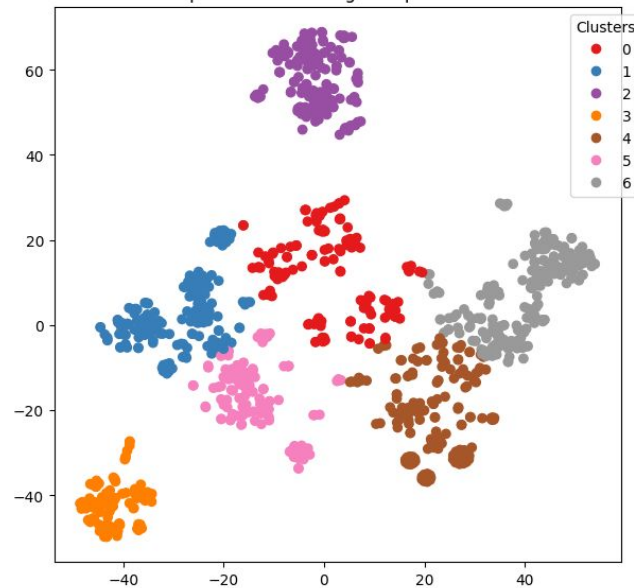
➞ ARI : 0.4338 time : 14.0

Analyse des descriptions textuelles :

Représentation des catégories réelles



Représentation catégories par clusters



⇒ Classification possible !



Analyse des descriptions textuelles : ARI

- CountVectorizer = 0.42
- Tf-idf = 0.45
- Word2Vec = 0.21
- BERT = 0.32
- USE = 0.43



Analyse des descriptions textuelles : ARI

- CountVectorizer = 0.42
- Tf-idf = 0.45 ⇐
- Word2Vec = 0.21
- BERT = 0.32
- USE = 0.43



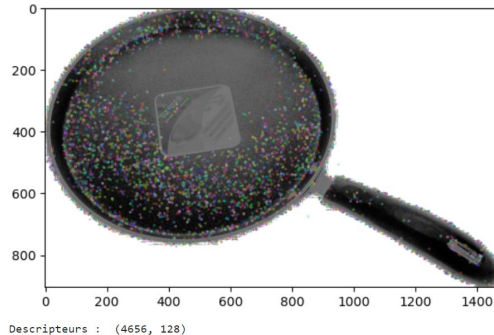
I Etude de faisabilité de classification :

2 Données images

Analyse des images :

1 Approche SIFT :

- filtrer les images en niveau de gris + equalisation
- création d'une liste de descripteurs par image



Nombre de descripteurs : (517351, 128)

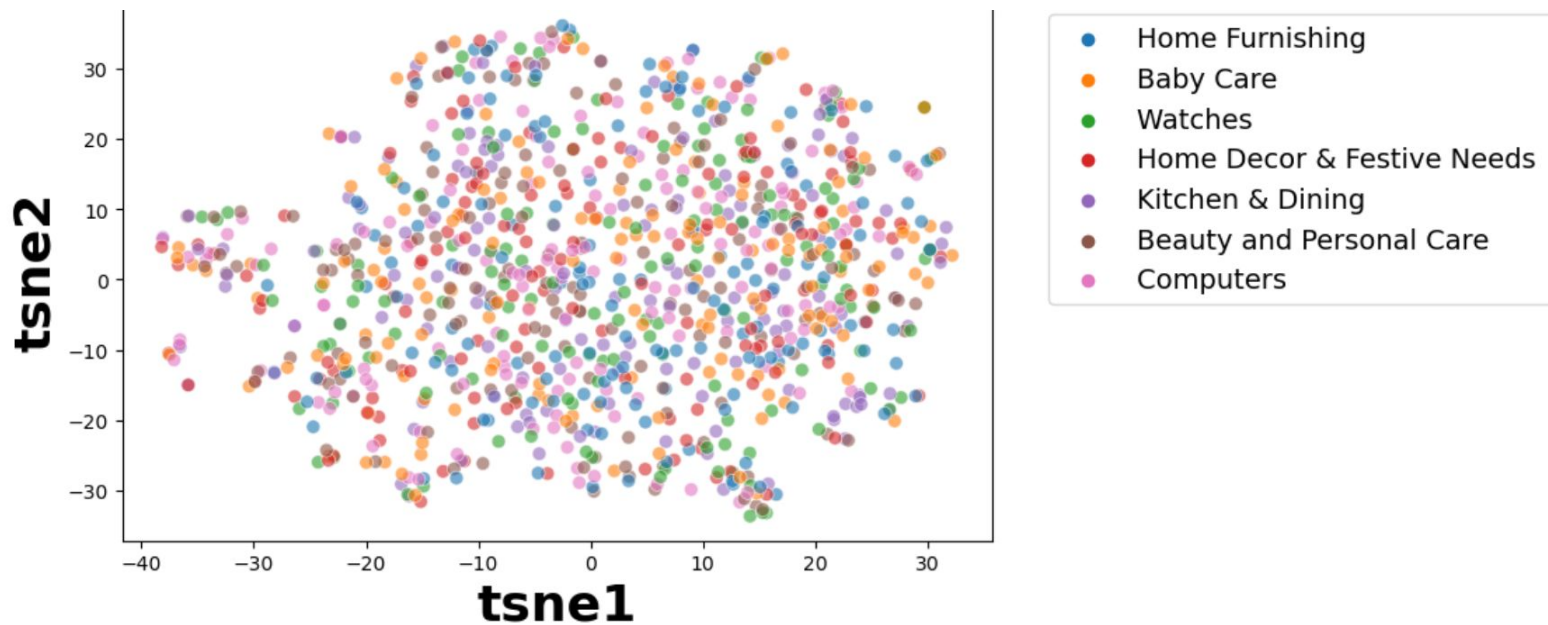


Analyse des images :

1 Approche SIFT :

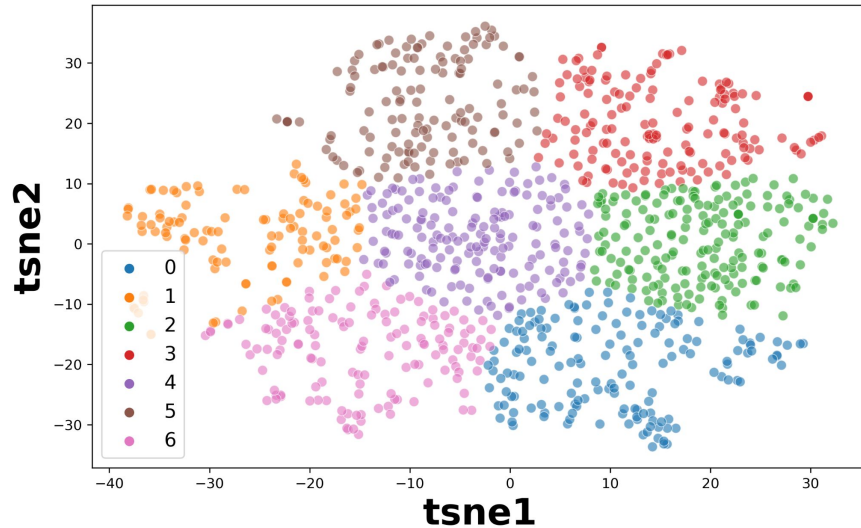
- création d'une liste de descripteurs pour l'ensemble des images
- créer les clusters de descripteurs (MiniBatchKMeans)

Analyse des images : Selon les vraies classes



Analyse des images : Selon les clusters

TSNE selon les clusters



ARI : 0.0012437155160732376

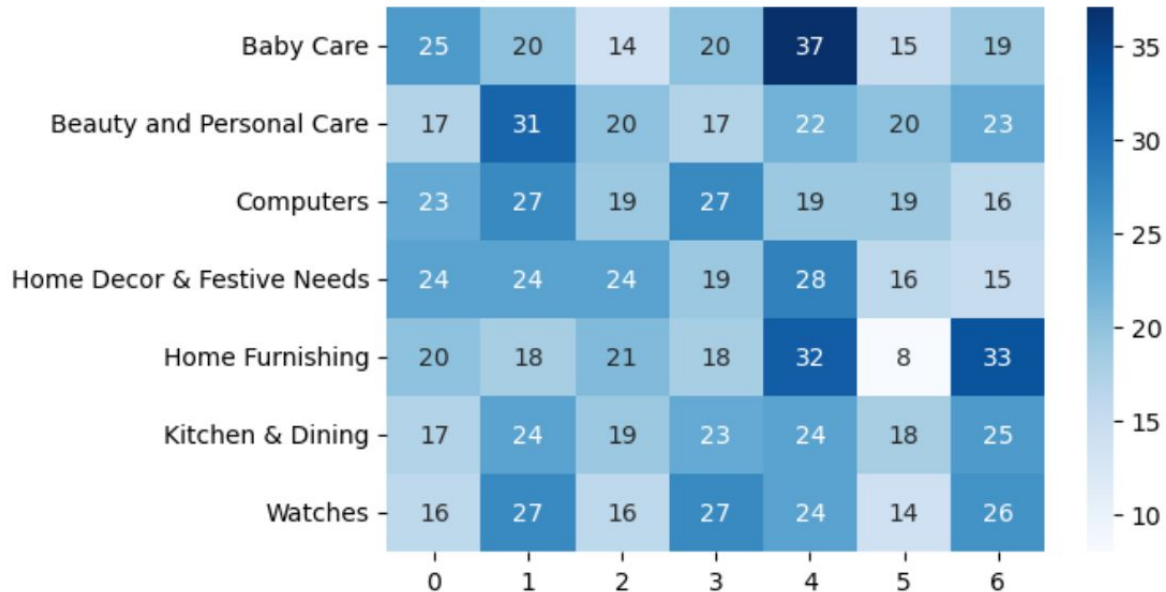
⇒ approche très peu fiable !



Analyse des images : Matrice de confusion

	precision	recall	f1-score	support
0	0.18	0.17	0.17	150
1	0.18	0.21	0.19	150
2	0.14	0.13	0.13	150
3	0.13	0.13	0.13	150
4	0.17	0.21	0.19	150
5	0.16	0.12	0.14	150
6	0.17	0.17	0.17	150
accuracy			0.16	1050
macro avg	0.16	0.16	0.16	1050
weighted avg	0.16	0.16	0.16	1050

Analyse des images : Matrice de confusion





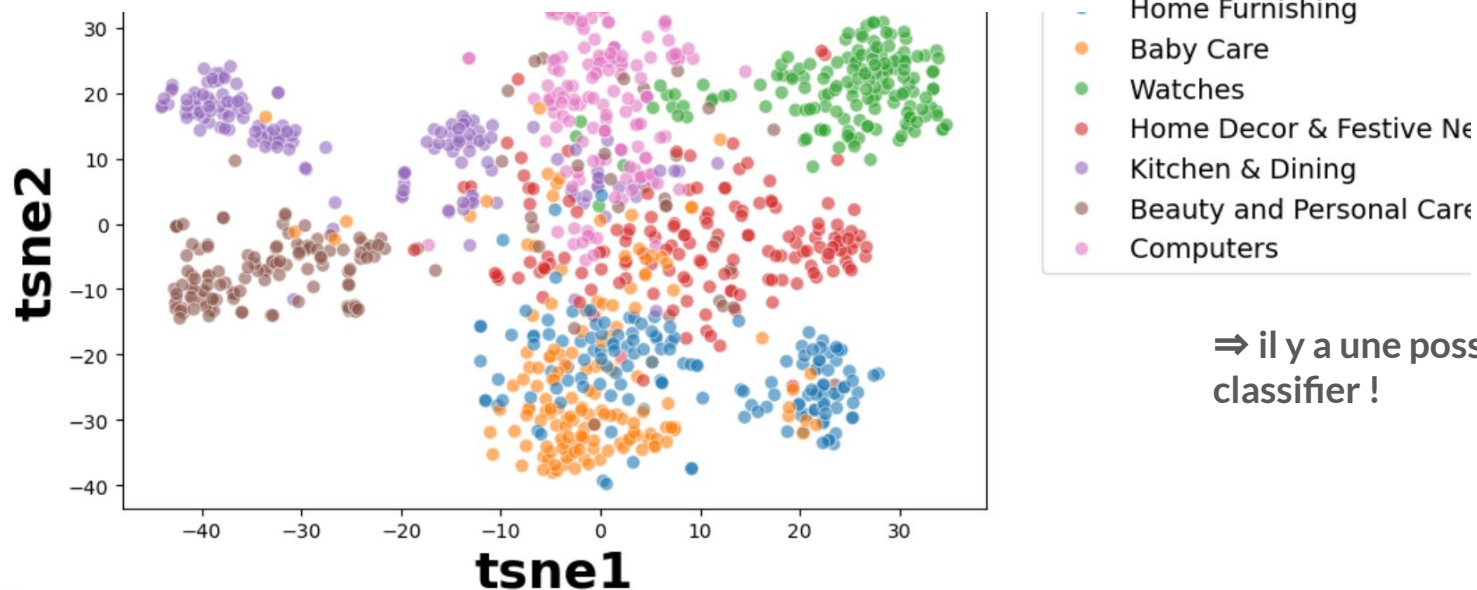
Analyse des images :

1 Approche Transfer learning CNN : VGG16

- **Création du modèle pré-entraîné**
- **Création des features des images**
- **Réduction de dimension PCA et TSNE**
- **Création de clusters à partir du T-SNE**

Analyse des images : Approche Transfer learning

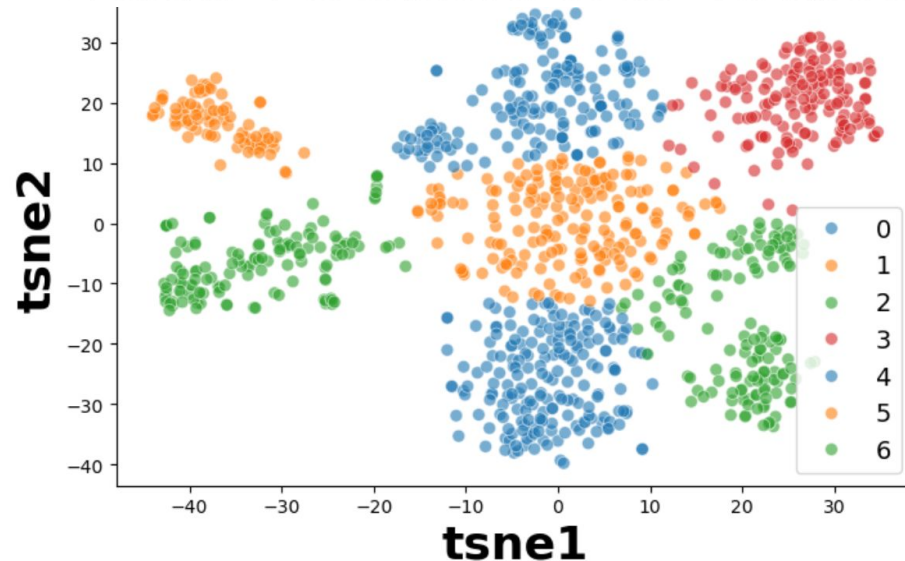
TSNE selon les vraies classes



⇒ il y a une possibilité de classifier !

Analyse des images : Approche Transfer learning

TSNE selon les clusters



⇒
cla



Analyse des images :

2 Approche Transfer learning CNN : VGG16

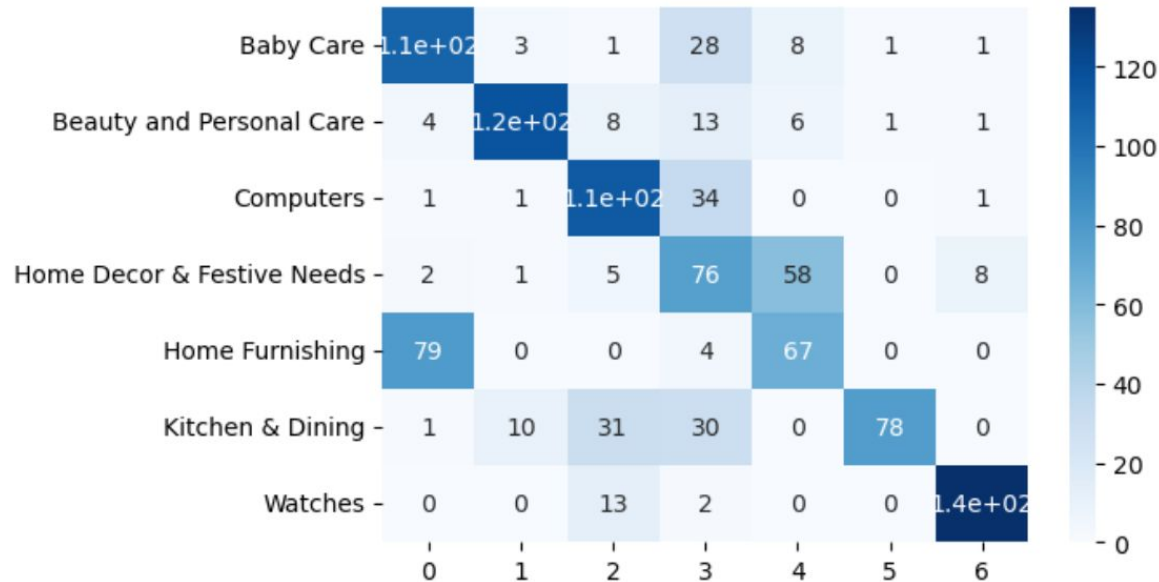
- Calcul de l'ARI ARI : 0.45283450222669647



Analyse des images : Matrice de confusion

	precision	recall	f1-score	support
0	0.55	0.72	0.63	150
1	0.89	0.78	0.83	150
2	0.66	0.75	0.70	150
3	0.41	0.51	0.45	150
4	0.48	0.45	0.46	150
5	0.97	0.52	0.68	150
6	0.92	0.90	0.91	150
accuracy			0.66	1050
macro avg	0.70	0.66	0.67	1050
weighted avg	0.70	0.66	0.67	1050

Analyse des images : Matrice de confusion





Analyse des images : Comparaison

- **SIFT : ARI = 0,0012**
- **Transfert learning : ARI = 0,452**



Analyse des images : Comparaison

- **SIFT : ARI = 0,0012**
- **Transfert learning : ARI = 0,452 ⇐**

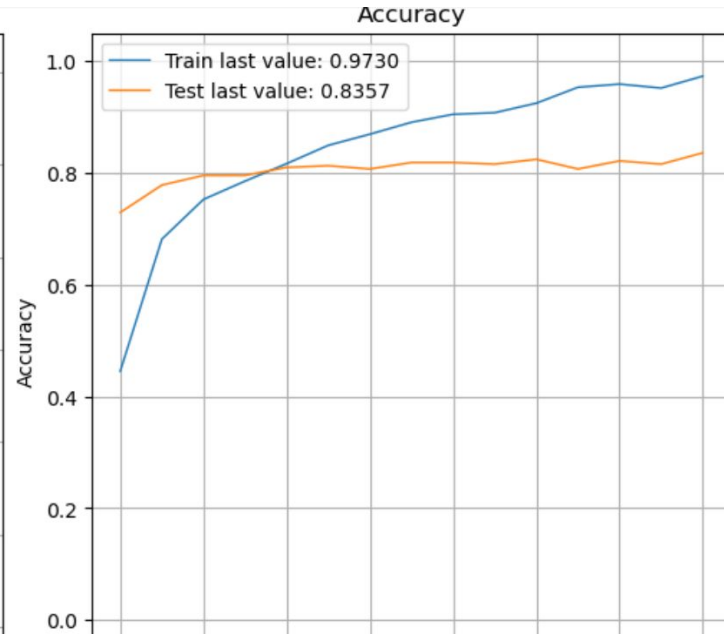
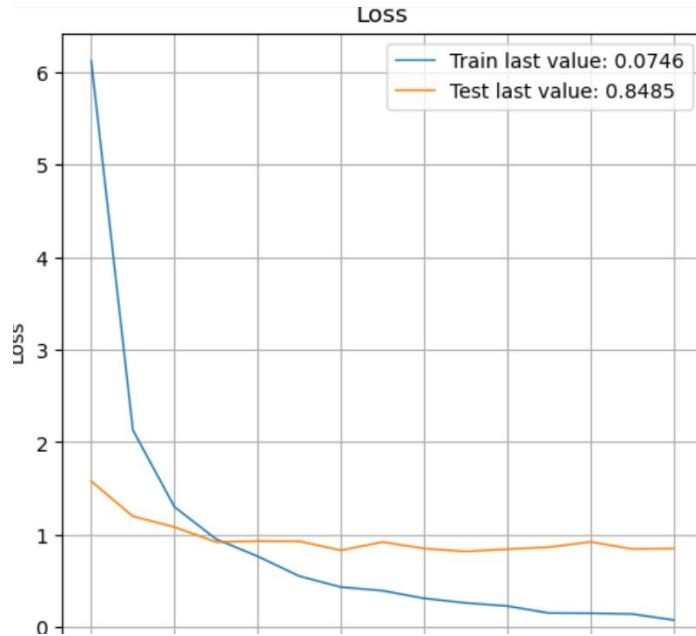


Analyse des images :

3 Classification supervisée :

- Création de modèle de classification VGG16
⇒ Récupération modèle pré-entraîné
- Entraînement/Validation/test

Analyse des images :

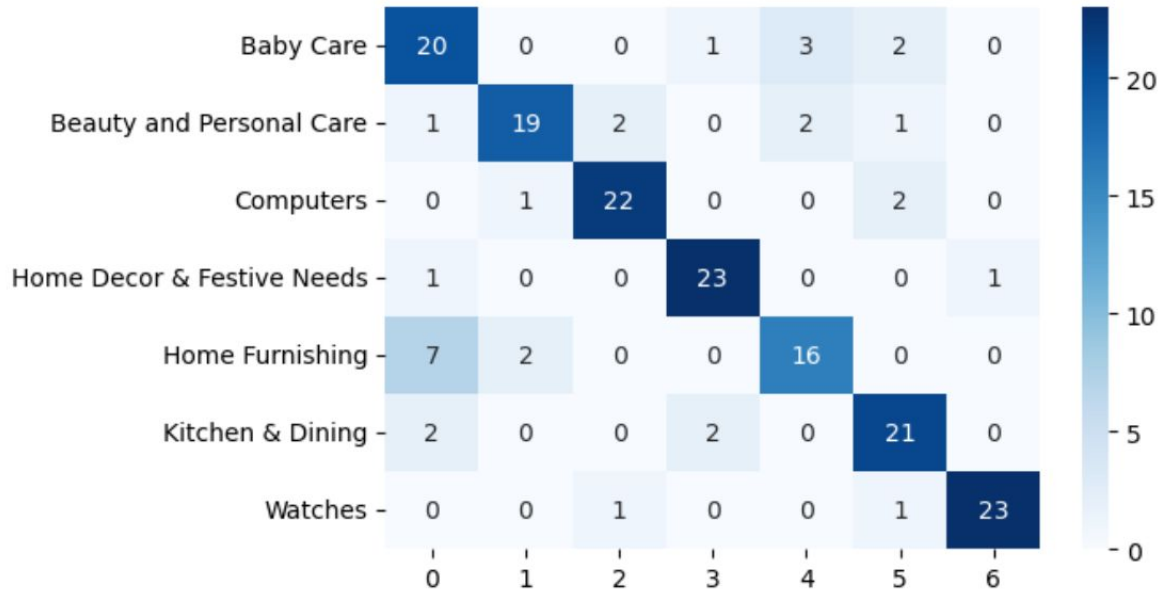




Analyse des images : Classification supervisée

	precision	recall	f1-score	support
0	0.65	0.77	0.70	26
1	0.86	0.76	0.81	25
2	0.88	0.88	0.88	25
3	0.88	0.92	0.90	25
4	0.76	0.64	0.70	25
5	0.78	0.84	0.81	25
6	0.96	0.92	0.94	25
accuracy			0.82	176
macro avg	0.82	0.82	0.82	176
weighted avg	0.82	0.82	0.82	176

Analyse des images : Classification supervisée



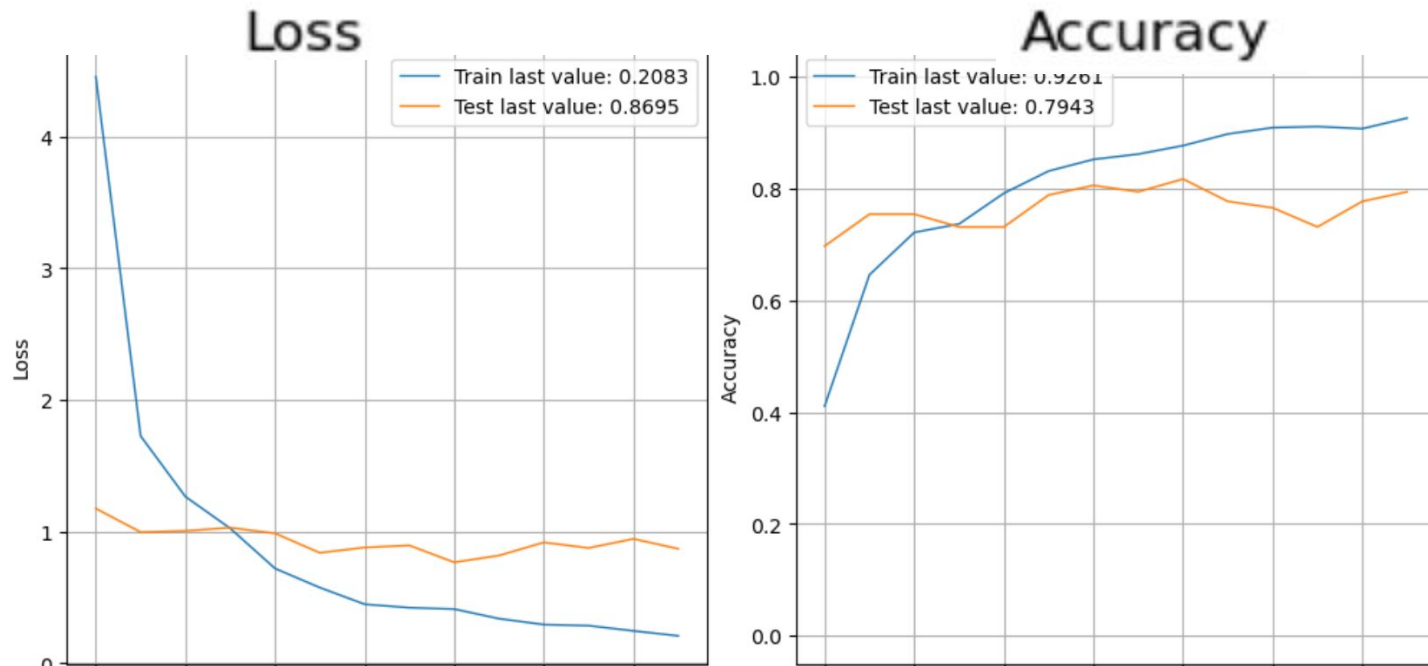


Analyse des images :

3 Classification supervisée avec Data augmentation :

- **ImageDataGenerato**

Analyse des images :



III Test API : épicerie fine :

V1 (Current) ▾

🔍 Search endpoints

Food Request Step 1 - Parser

GET /api/food-database/v2/parser

Food Request Step 2 - Nutrients

Food Search Autocomplete

< GET /api/food-database...

Test Endpoint

\d+(?:\d*)?)|(?:(?:[+]?[d+](?:\d*)?)[+])

numbers.

See calcium for an example for how to fill in the range.

ingr

champagne

STRING

OPTIONAL

A keyword search parameter to be found in the food name. **REQUIRED** if 'upc' and 'brand' are not specified. **NOT REQUIRED** if 'brand' is specified. **DO NOT POPULATE** if 'upc' is specified.

nutrition-type

cooking

STRING

OPTIONAL

Select between cooking and food

Code Snippets

Example Responses

Results

200 Success

Body Headers

▼ {} 4 keys

text: "champagne"

▼ parsed: [] 1 item

▶ 0: {} 1 key

▼ hints: [] 20 items

▶ 0: {} 2 keys

▶ 1: {} 2 keys

▶ 2: {} 2 keys

▶ 3: {} 2 keys

▶ 4: {} 2 keys



API : épicerie fine :

```
85]: liste_10 =[response['hints'][indice]['food'] for indice in range(10)]
```

```
86]: result=[[element.get('foodId', 'NONE'),  
             element.get('label', 'NONE'),  
             element.get('category', 'NONE'),  
             element.get('foodContentsLabel', 'NONE'),  
             element.get('image', 'NONE')] for element in liste_10 ]
```

```
87]: import pandas as pd  
df= pd.DataFrame(result , columns=['foodId','label','category','foosContentsLable','image'])
```

[13]:

Unnamed: 0		foodId	label	category	foosContentsLable	image
0	0	food_a656mk2a5dmqb2adiamu6beihduu	Champagne	Generic foods	NONE	https://www.edamam.com/food-img/a71/a718cf3c52...
1	1	food_b753ithamdb8psbt0w2k9aquo06c	Champagne Vinaigrette, Champagne	Packaged foods	OLIVE OIL; BALSAMIC VINEGAR; CHAMPAGNE VINEGAR...	NONE
2	2	food_b3dyababjo54xobm6r8jzbghjgqe	Champagne Vinaigrette, Champagne	Packaged foods	INGREDIENTS: WATER; CANOLA OIL; CHAMPAGNE VINE...	https://www.edamam.com/food-img/d88/d88b64d973...
3	3	food_a9e0ghsamvoc45bwa2ybsa3gken9	Champagne Vinaigrette, Champagne	Packaged foods	CANOLA AND SOYBEAN OIL; WHITE WINE (CONTAINS S...	NONE
4	4	food_an4jjueaucpus2a3u1ni8auhe7q9	Champagne Vinaigrette, Champagne	Packaged foods	WATER; CANOLA AND SOYBEAN OIL; WHITE WINE (CON...	NONE
5	5	food_bmu5dmkazwuvpaa5prh1daa8jxs0	Champagne Dressing, Champagne	Packaged foods	SOYBEAN OIL; WHITE WINE (PRESERVED WITH SULFIT...	https://www.edamam.com/food-img/ab2/ab2459fc2a...
6	6	food_alpl44taoyv11ra0lic1qa8xculi	Champagne Buttercream	Generic meals	sugar; butter; shortening; vanilla; champagne;...	NONE
7	7	food_byap67hab6evc3a0f9w1oag3s0qf	Champagne Sorbet	Generic meals	Sugar; Lemon juice; brandy; Champagne; Peach	NONE
8	8	food_am5egz6aq3fpjlaf8xpkdbc2asis	Champagne Truffles	Generic meals	butter; cocoa; sweetened condensed milk; vanil...	NONE