Projet 6: Classification automatique des biens de consommation

Maodo FALL

OpenClassrooms

Soutenance du projet
14 décembre 2024

Sommaire

Sommaire

Place de Marché

Problématique

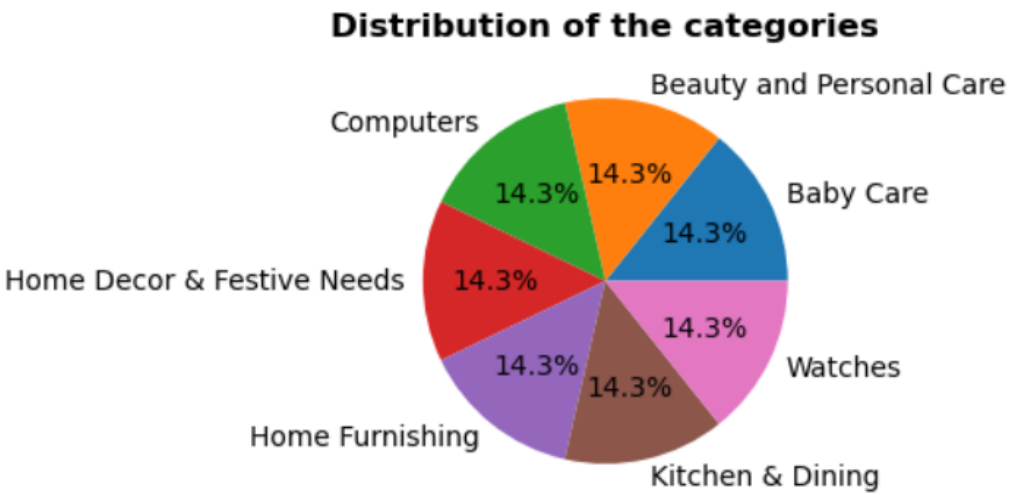
Problématique :
"Place de Marché" souhaite lancer une marketplace e-commerce. Sur leur site, des vendeurs proposent des articles à des acheteurs en postant une photo et une description. Pour l'instant, la catégorisation d'un article est effectuée manuellement par les vendeurs, et est donc peu fiable. De plus, le volume des articles est pour l'instant très petit. L'entreprise a pour objectif d'automatiser cette tâche d'attribution de la catégorie.

Missions :

- Etudier la faisabililté d'un moteur de classification.
- Classification supervisée d'images.
- Collecte de nouveaux produits via une API.

Jeu de données

- Le jeu de données contient 1050 produits et 15 colonnes.
- On se focalise plutôt sur la description et l'image des produits.
- La colonne des catégories est équilibrée.



Nettoyage des données textes

Les données brutes de texte ont été passées sous plusieurs étapes de traitement pour les nettoyer et les traiter.

- nettoyage de texte.
- Suppression des mots les plus fréquents du vocabulaire.
- Ramener les tokens à leur racine.

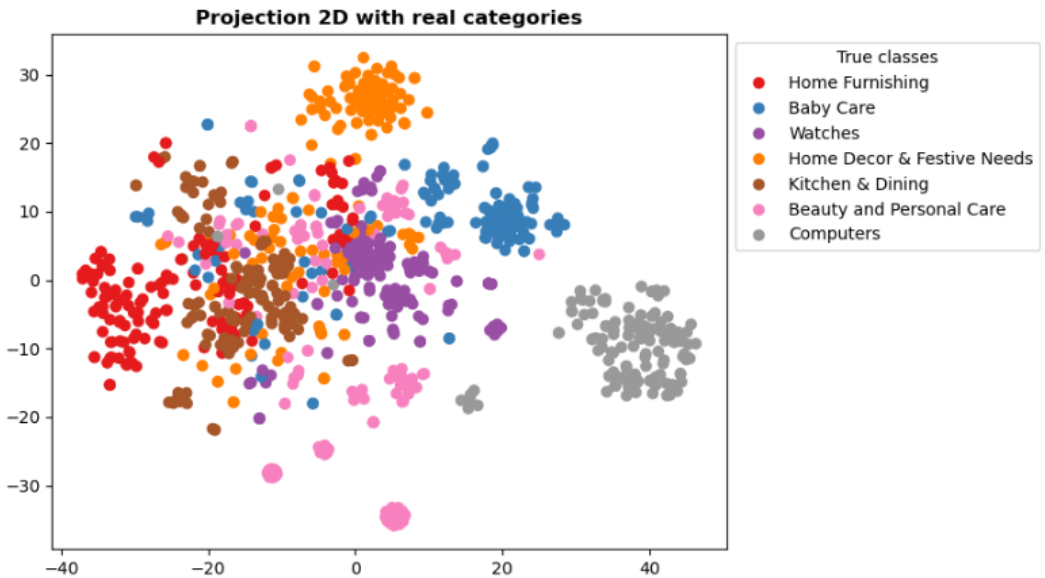
Exemple de données d'image

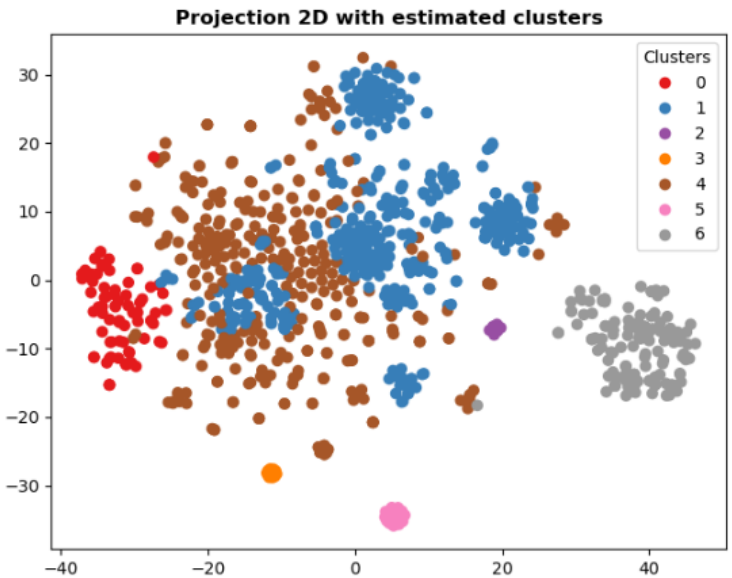
Exemple d'images de la catégorie Watches :

Sommaire

Approche Count-Vectorizer

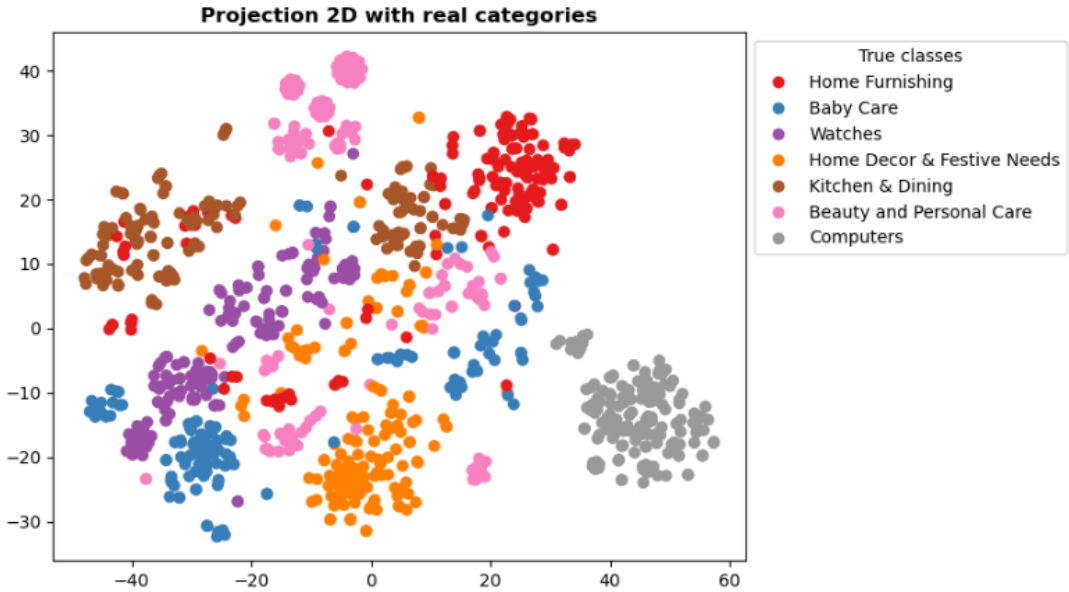
Le score ARI de mesure de similarité est de : 0.162

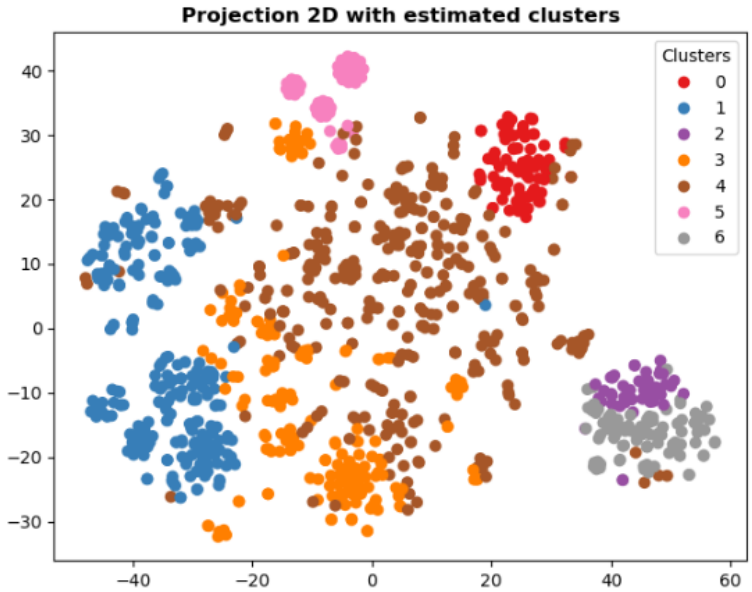




Approche TF-IDF

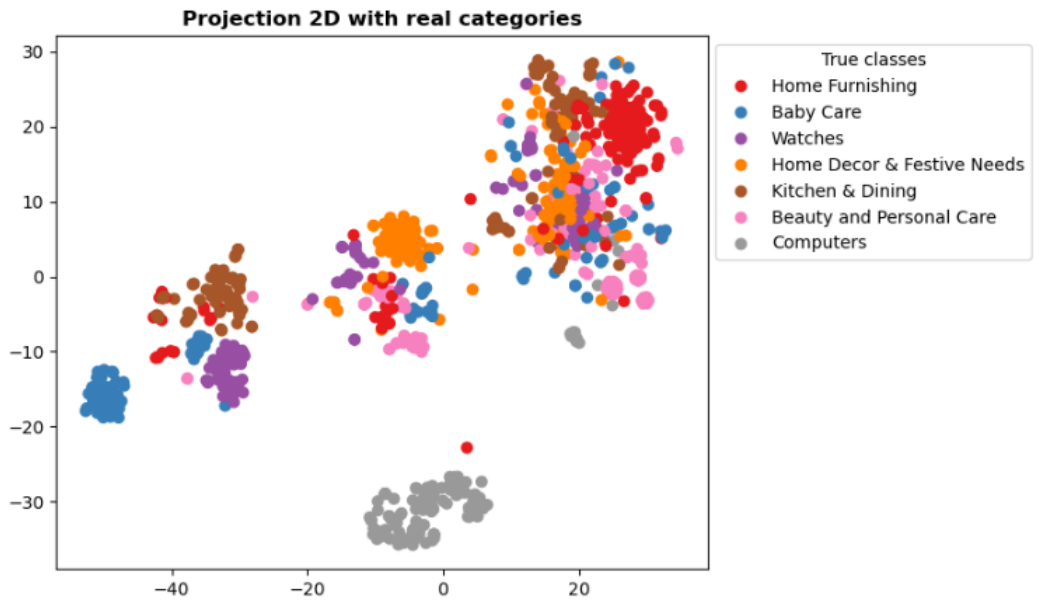
Le score ARI de mesure de similarité est de : 0.191

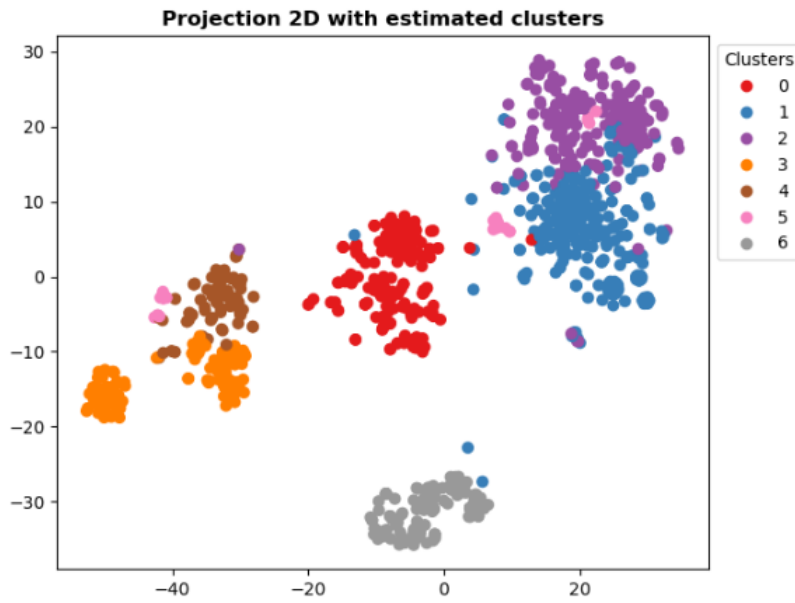




Approche Word2Vec

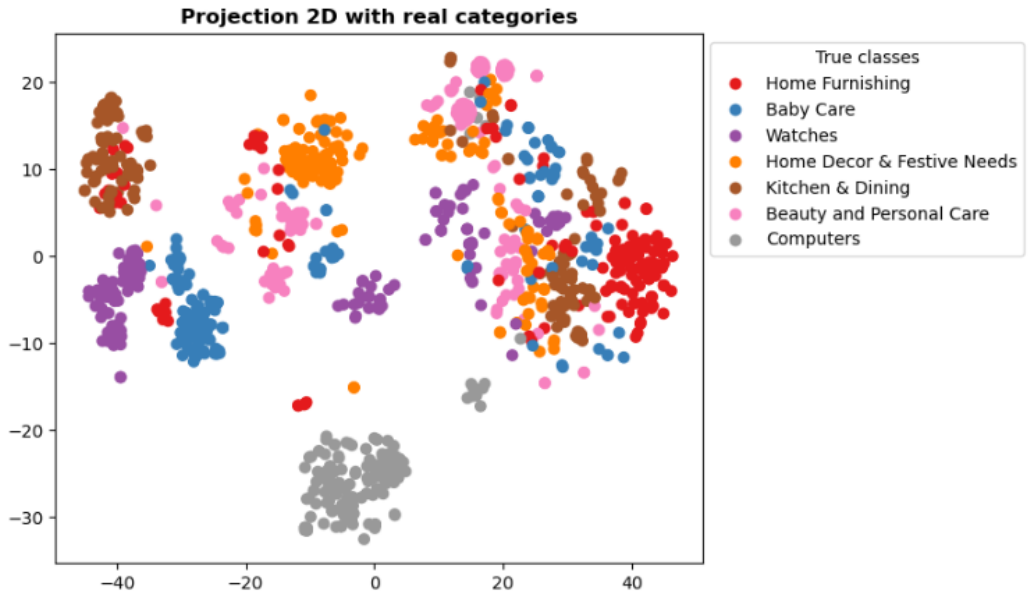
Le score ARI de mesure de similarité est de : 0.261

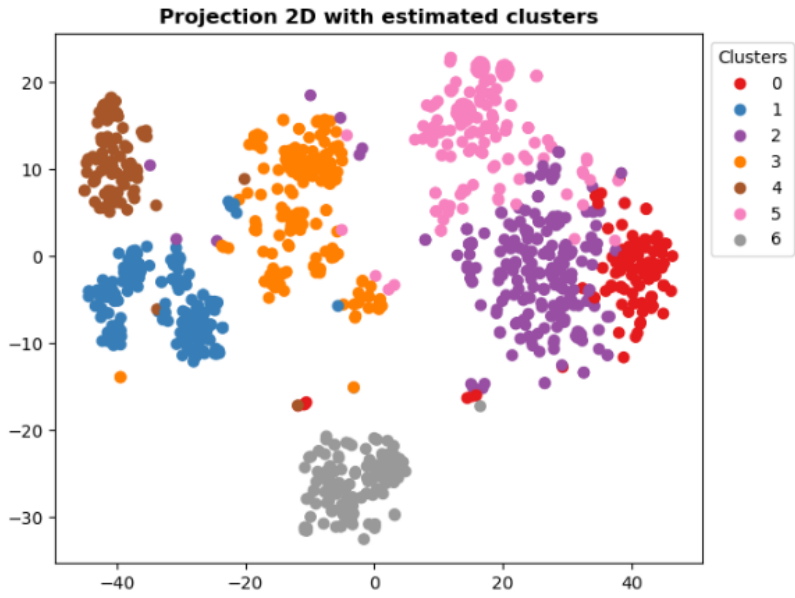




Approche BERT

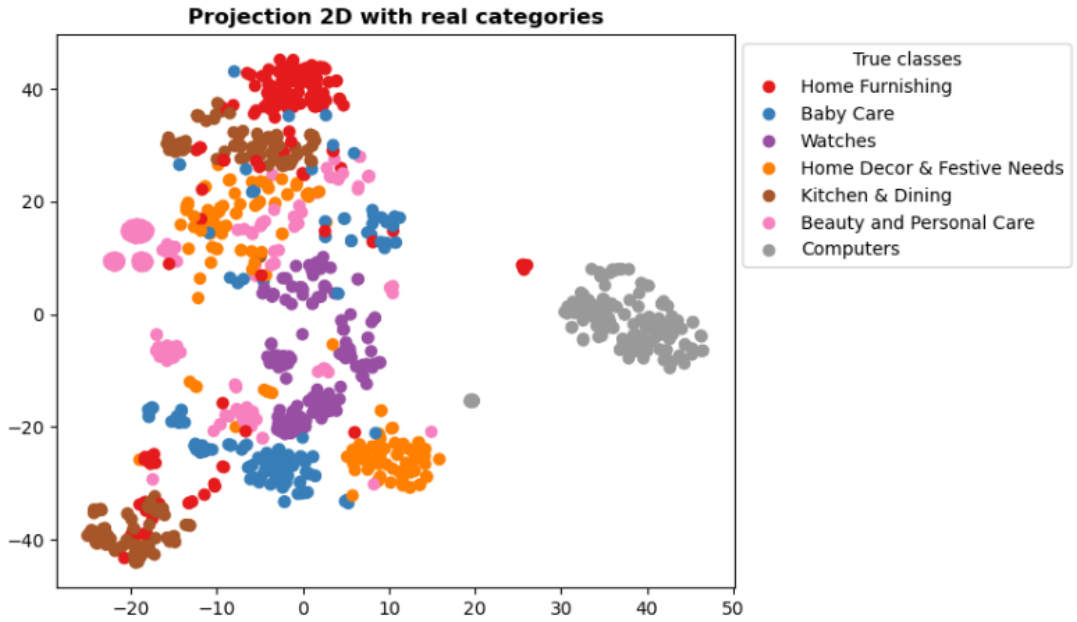
Le score ARI de mesure de similarité est de : 0.294

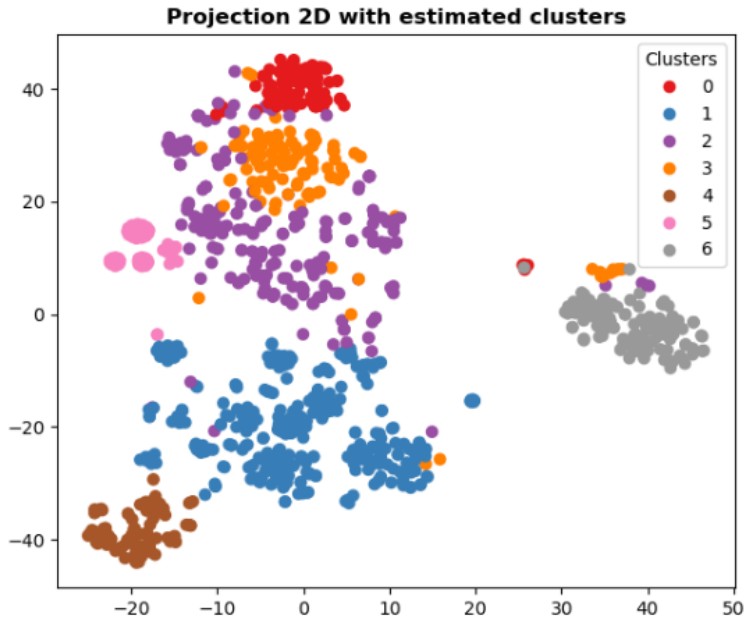




Approche USE

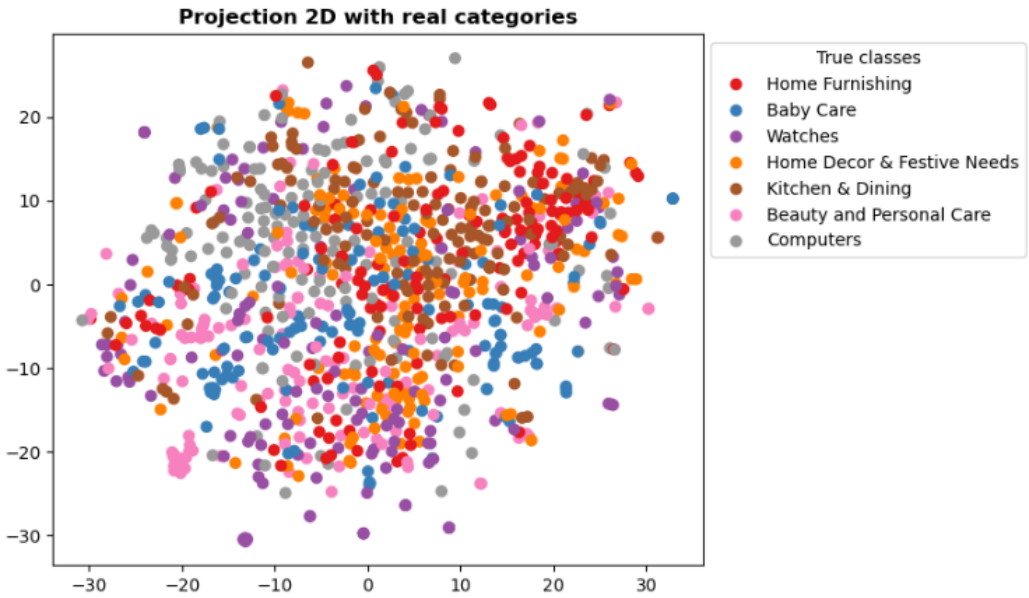
```
Le score ARI de mesure de similarité est de : 0.263
```

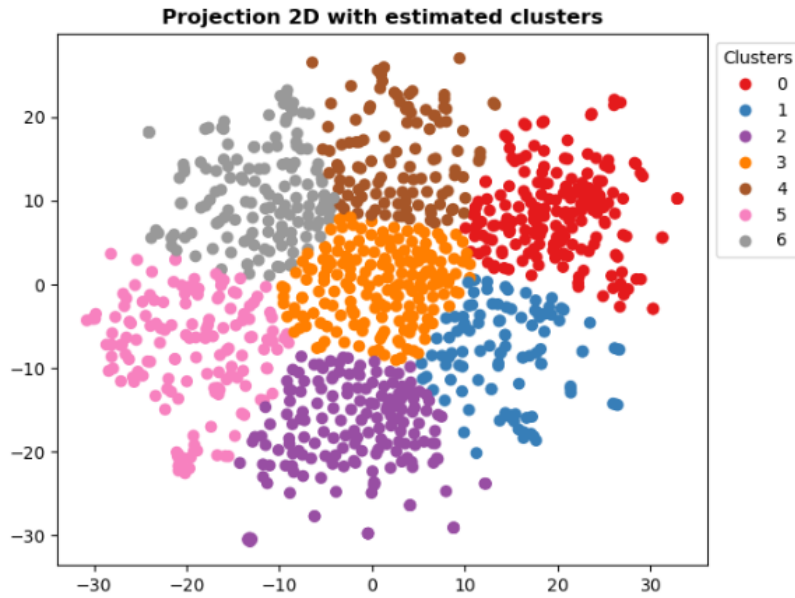




Approche SIFT

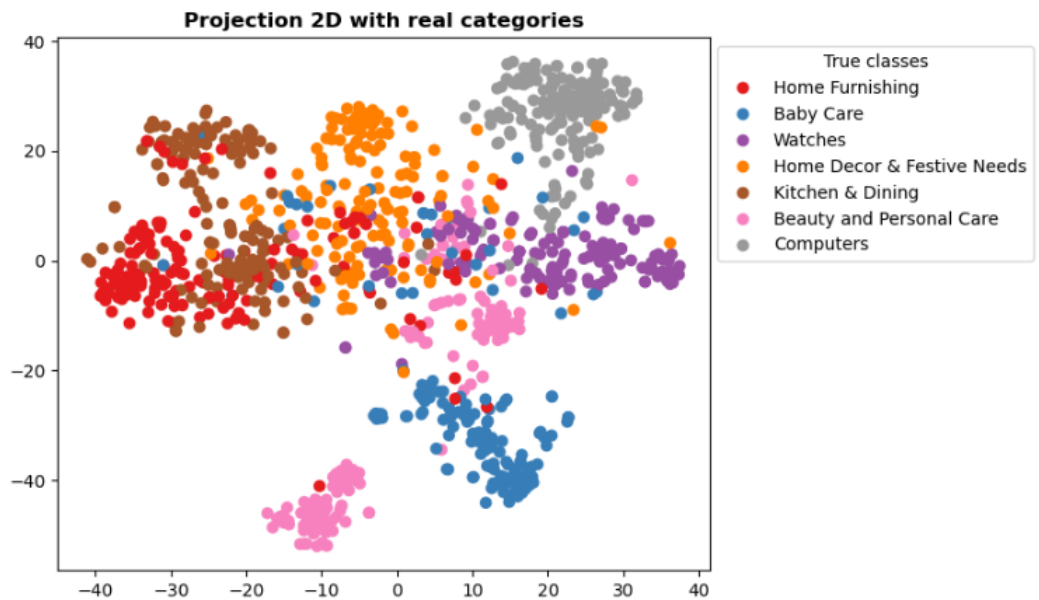
Le score ARI de mesure de similarité est de : 0.061

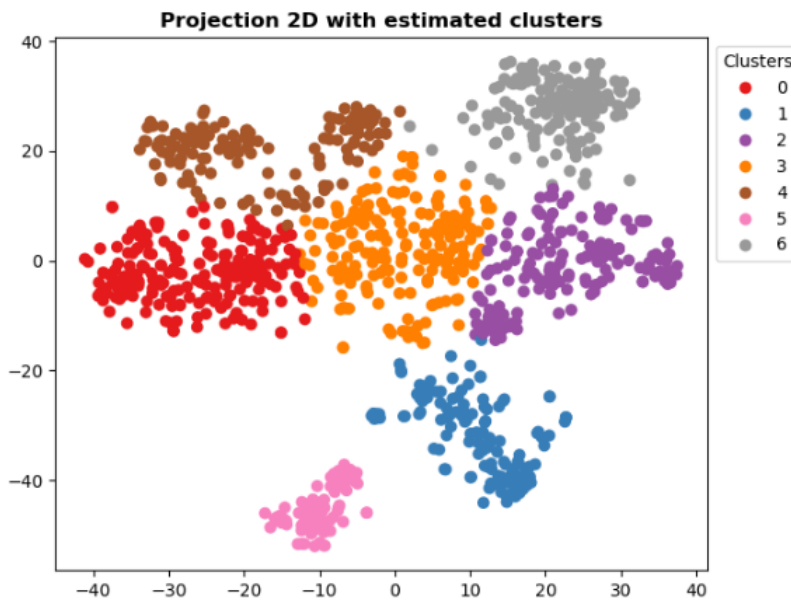




Approche ConvNet - Transfert Learning avec VGG16

Le score ARI de mesure de similarité est de : 0.451



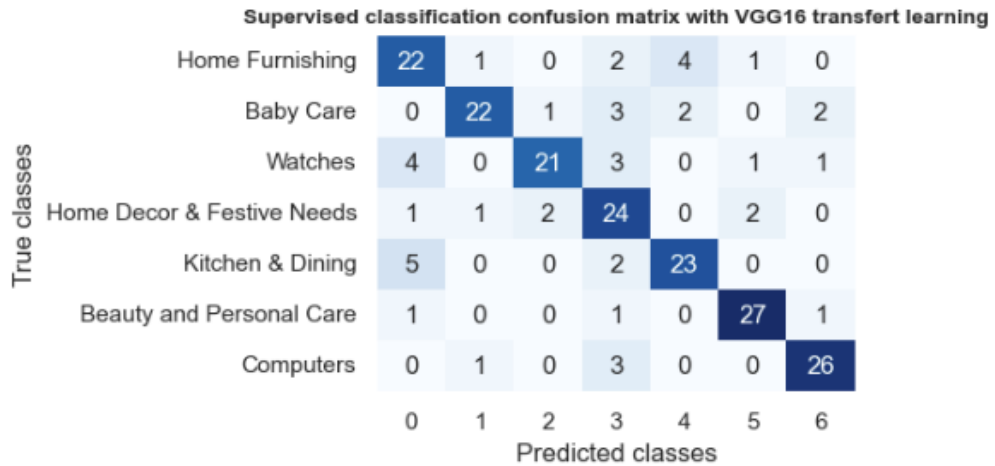


Sommaire

Approche simple par préparation initiale des images

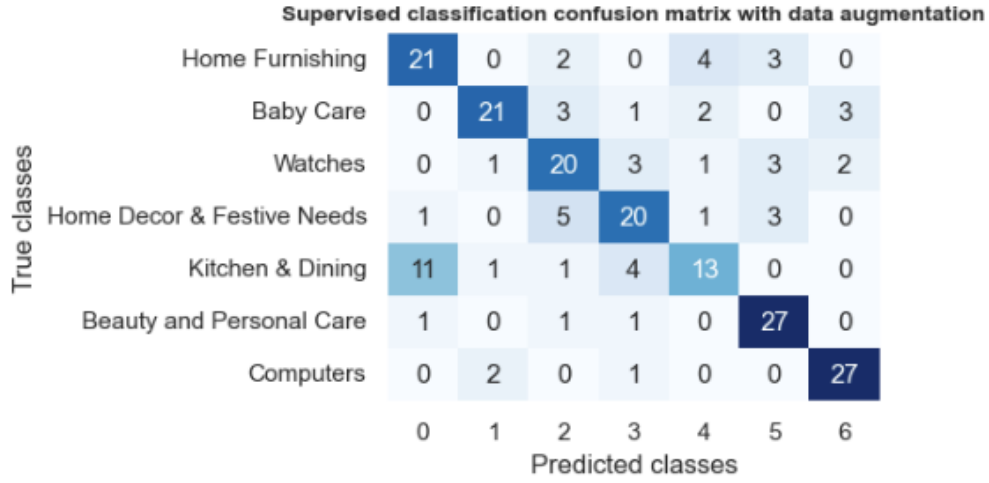
Classification avec Transfert Learning avec VGG16 :

	precision	recall	f1-score	support
0	0.67	0.73	0.70	30
1	0.88	0.73	0.80	30
2	0.88	0.70	0.78	30
3	0.63	0.80	0.71	30
4	0.79	0.77	0.78	30
5	0.87	0.90	0.89	30
6	0.87	0.87	0.87	30
accuracy			0.79	210
macro avg	0.80	0.79	0.79	210
weighted avg	0.80	0.79	0.79	210



Modèle VGG16 avec data augmentation

	precision	recall	f1-score	support
0	0.62	0.70	0.66	30
1	0.84	0.70	0.76	30
2	0.62	0.67	0.65	30
3	0.67	0.67	0.67	30
4	0.62	0.43	0.51	30
5	0.75	0.90	0.82	30
6	0.84	0.90	0.87	30
accuracy			0.71	210
macro avg	0.71	0.71	0.70	210
weighted avg	0.71	0.71	0.70	210



Sommaire

Résultats de l'appel de l'API food-database du site edamam

	foodId	label	category	foodContentsLabel	image
0	food_a656mk2a5dmqb2adiamu6beihduu	Champagne	Generic foods	None	https://www.edamam.com/food-img/a71/a718cf3c52...
1	food_b753ithamdb8psbt0w2k9aquo06c	Champagne Vinaigrette, Champagne	Packaged foods	OLIVE OIL; BALSAMIC VINEGAR; CHAMPAGNE VINEGAR...	None
2	food_b3dyababjo54xobm6r8jzbghjgqe	Champagne Vinaigrette, Champagne	Packaged foods	INGREDIENTS: WATER; CANOLA OIL; CHAMPAGNE VINE...	https://www.edamam.com/food-img/d88/d88b64d973...
3	food_a9e0ghsamvoc45bwa2ybsa3gken9	Champagne Vinaigrette, Champagne	Packaged foods	CANOLA AND SOYBEAN OIL; WHITE WINE (CONTAINS S...	None
4	food_an4jjueaucpus2a3u1ni8auhe7q9	Champagne Vinaigrette, Champagne	Packaged foods	WATER; CANOLA AND SOYBEAN OIL; WHITE WINE (CON...	None
5	food_bmu5dmkazwuvpaa5prh1daa8jxs0	Champagne Dressing, Champagne	Packaged foods	SOYBEAN OIL; WHITE WINE (PRESERVED WITH SULFIT...	https://www.edamam.com/food-img/ab2/ab2459fc2a...
6	food_alpl44taoyv11ra0lic1qa8xculi	Champagne Buttercream	Generic meals	sugar; butter; shortening; vanilla; champagne;...	None
7	food_am5egz6aq3fpjlaf8xpkdbc2asis	Champagne Truffles	Generic meals	butter; cocoa; sweetened condensed milk; vanil...	None
8	food_bcz8rhiajk1fuva0vkfmeakbouc0	Champagne Vinaigrette	Generic meals	champagne vinegar; olive oil; Dijon mustard; s...	None
9	food_a79xmnya6togreaeukbroa0thhh0	Champagne Chicken	Generic meals	Flour; Salt; Pepper; Boneless, Skinless Chicke...	None

Sommaire

Conclusion

- Le moteur de classification est bien réalisable au vu des résultats obtenus.
- L'approche CNN est l'algorithme qui obtient de meilleurs clusters et le meilleur score ARI, et elle fonctionne sur le principe de Transfert Learning. Le temps d'exécution de cette méthode reste néanmoins assez long.
- La classification supervisée avec data augmentation n'a pas permis d'améliorer les résultats de la classification sans augmentation.
- Les nouvelles données collectées pourront nous servir de données de test de notre modèle. Ce qui nous permettra d'utiliser l'ensemble du jeu de données initial pour entraîner le modèle.

MERCI BEAUCOUP !