

Projet 6

**Classifiez automatiquement des
biens de consommation**

Rédigé et Présenté par

Etudiant : GASSUC CEDRIC

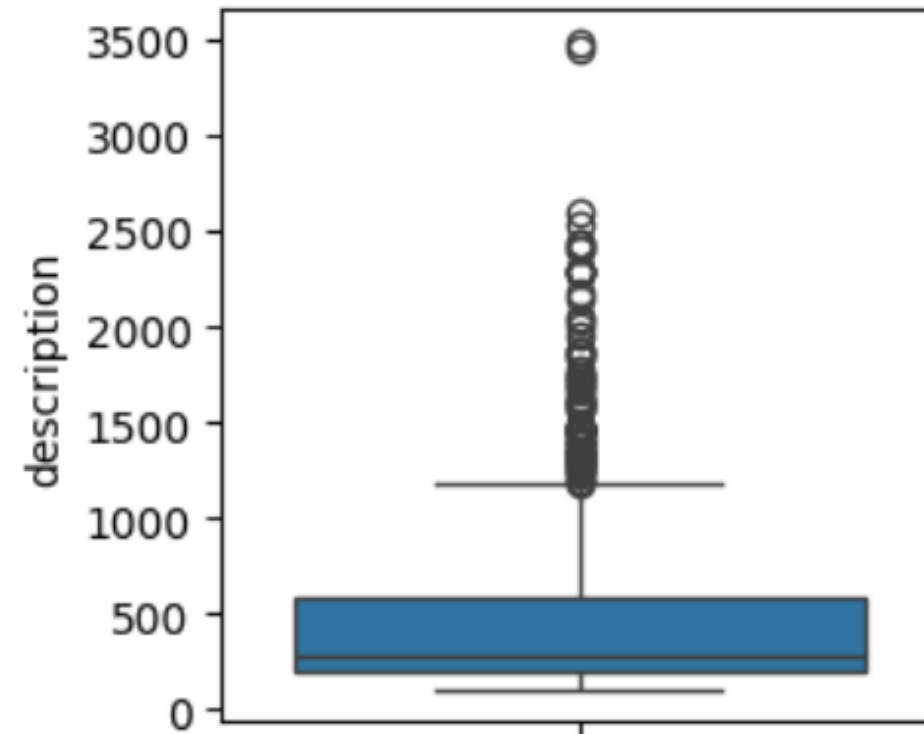
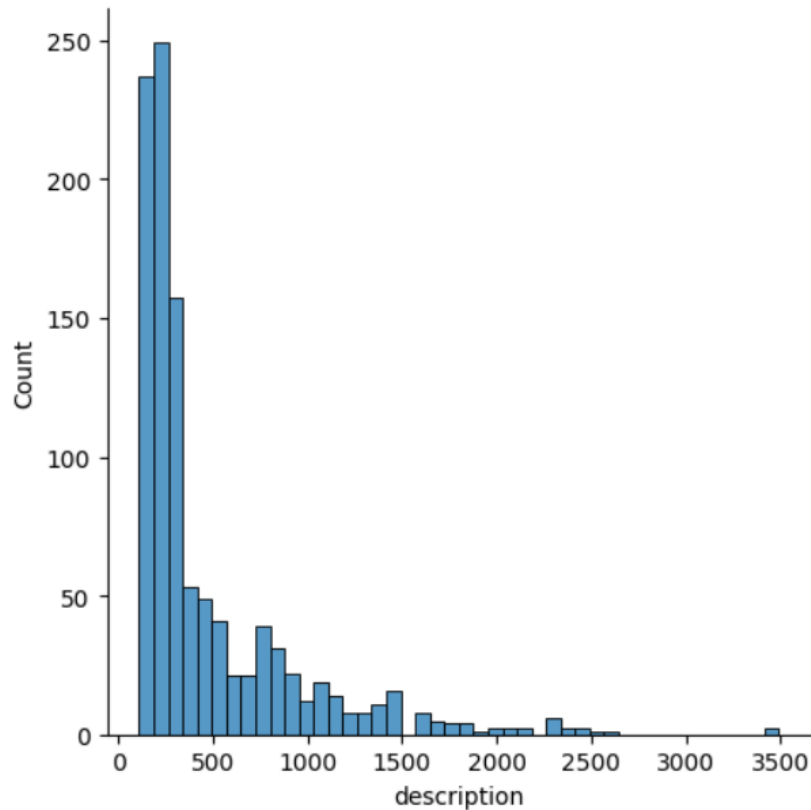
Mentor : Adrien Chambord

Plan

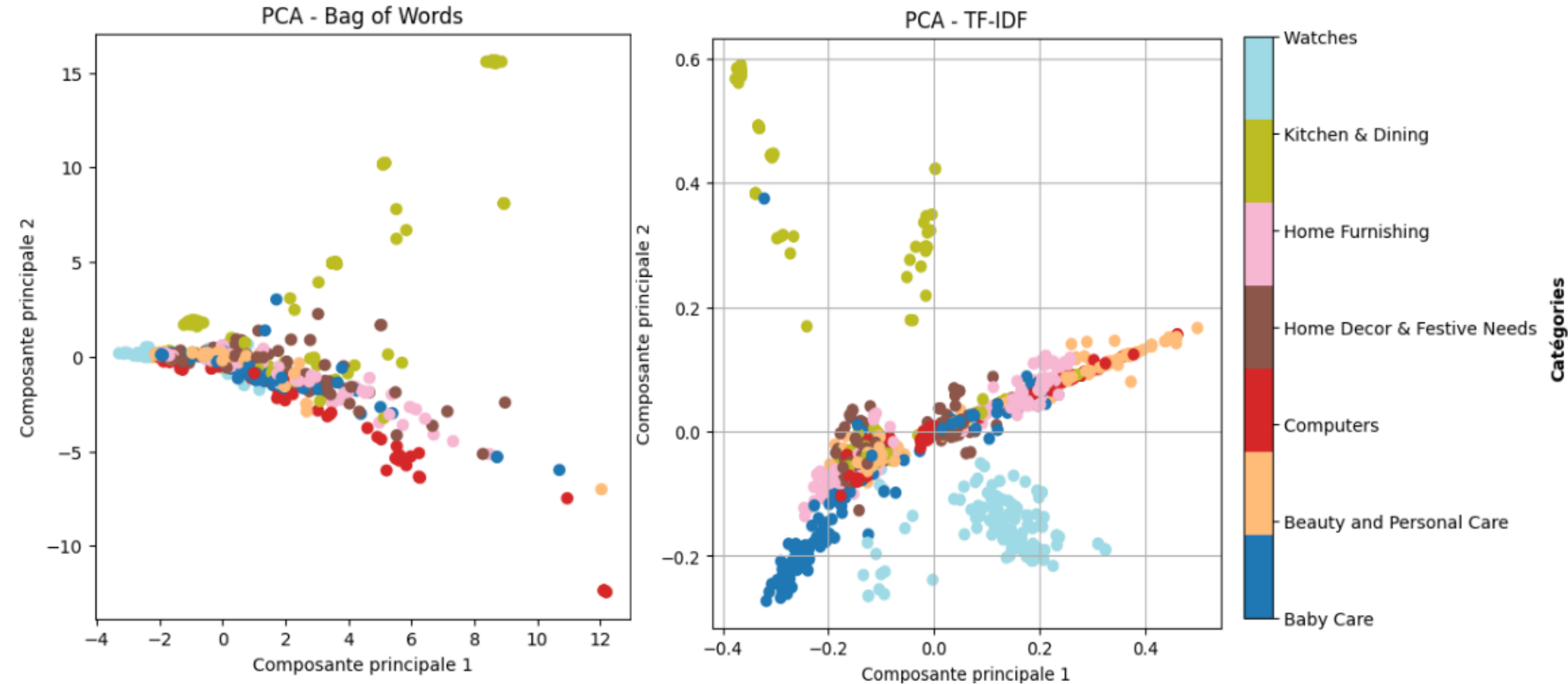
- ① Prétraitement des textes, méthode NLP basique (Bag of Words, TfIdf) : nettoyage, stemming, lemmatization
- ② Prétraitement des textes, méthodes Word Embeddings : Word2Vec, BERT, Universal Sentence Encoder
- ③ Analysez les images via différentes méthodes de transformation (niveau de gris, filtrage bruit, contraste)
- ④ Méthodes de réduction de dimension sur les images encodées et comparer les résultats avec les catégories réelles
- ⑤ Traitement d'image avec des méthodes avancées réseau de neurone pré-entraînés(avec data augmentation)
- ⑥ Comparez vos résultats avec ceux des méthodes de traitement d'image plus basiques.
- ⑦ CONCLUSION

I Prétraitement des textes, méthode NLP basique (Bag of Words, TfIdf) : nettoyage, stemming, lemmatization

1 - Nettoyage, stemming, lemmatization



Comparaison entre deux différentes méthodes d'encodage Bag of Words & TF-IDF

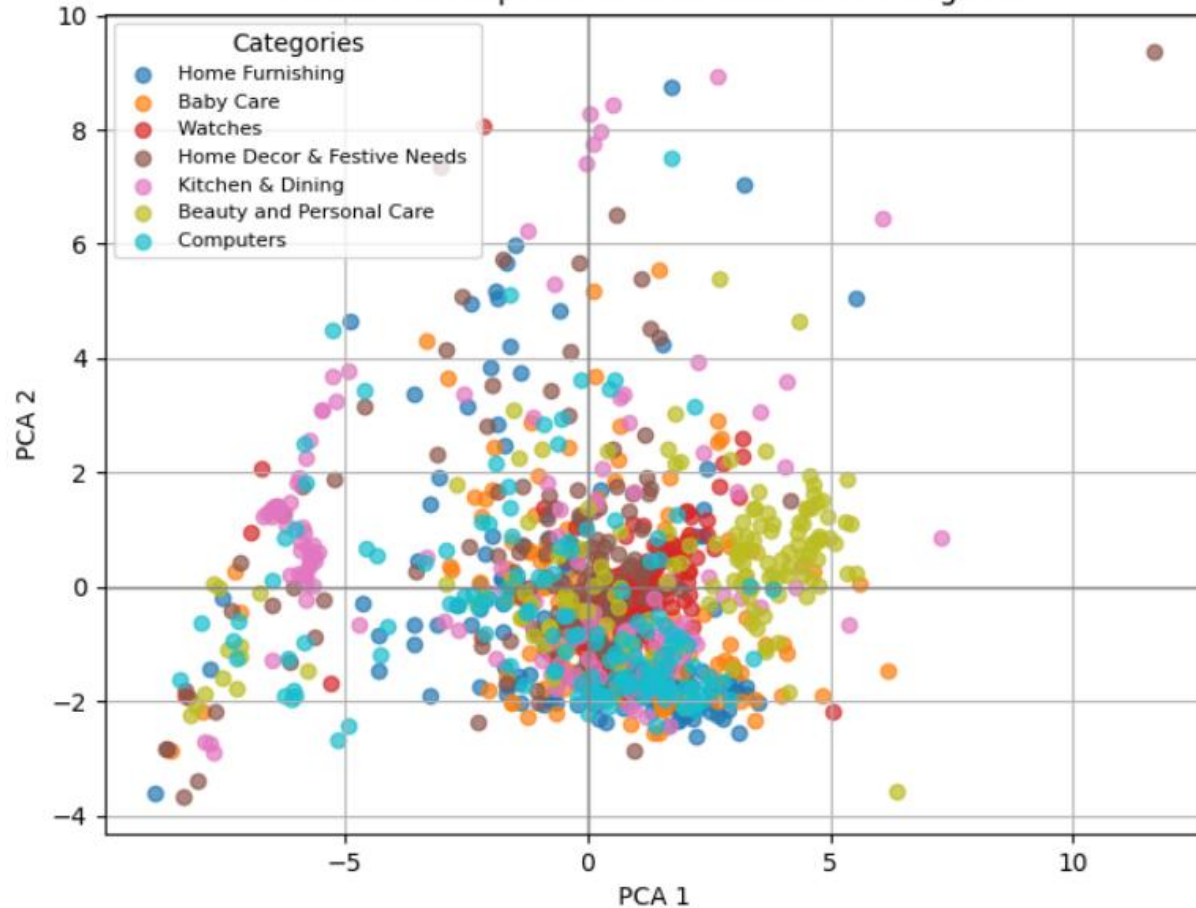


ARI méthode Bag of Words : 0,05

ARI méthode TF-IDF: 0,21

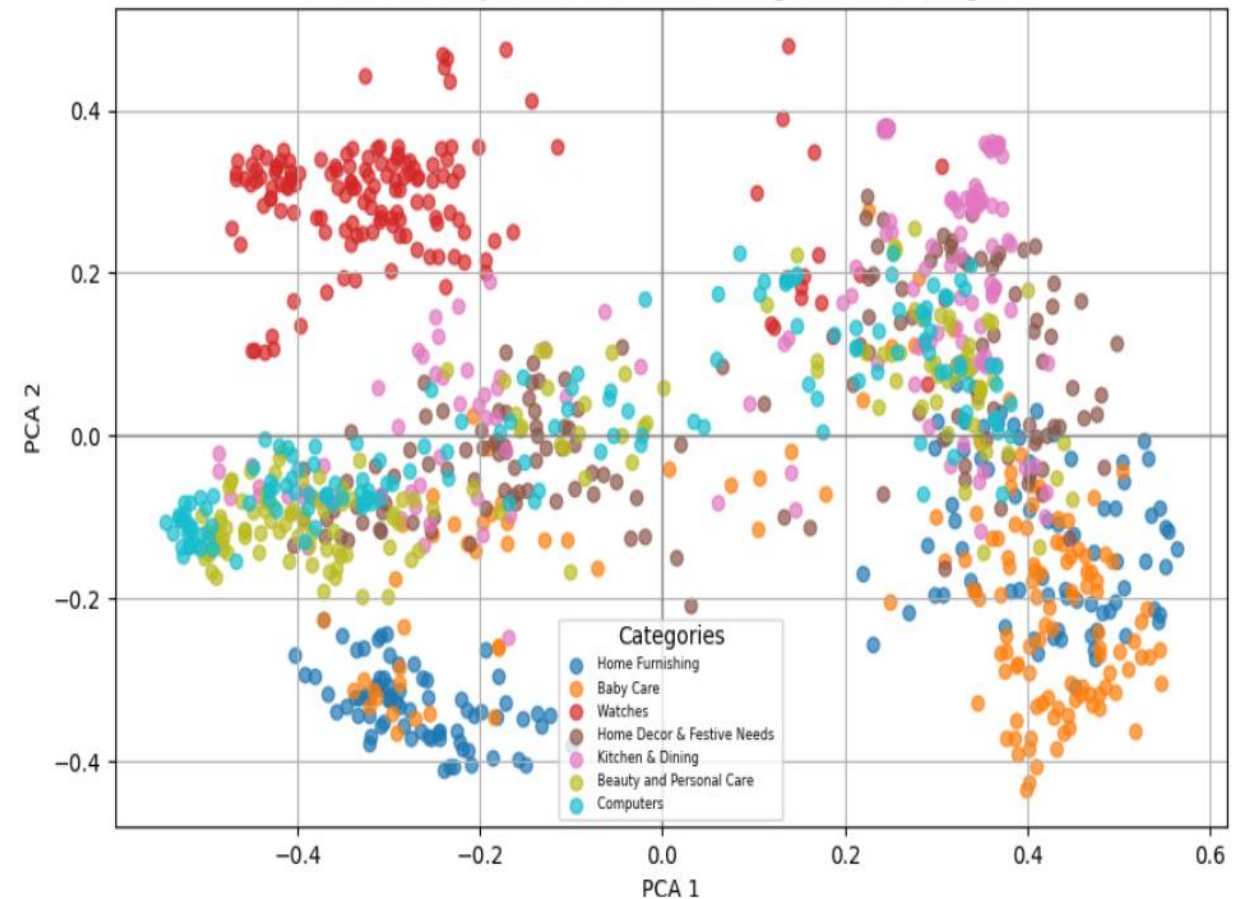
Comparaison entre trois différentes méthodes d'encodage embedding BERT, USE, Word2vec

PCA des descriptions codées BERT avec catégories



ARI méthode Bert: 0.12

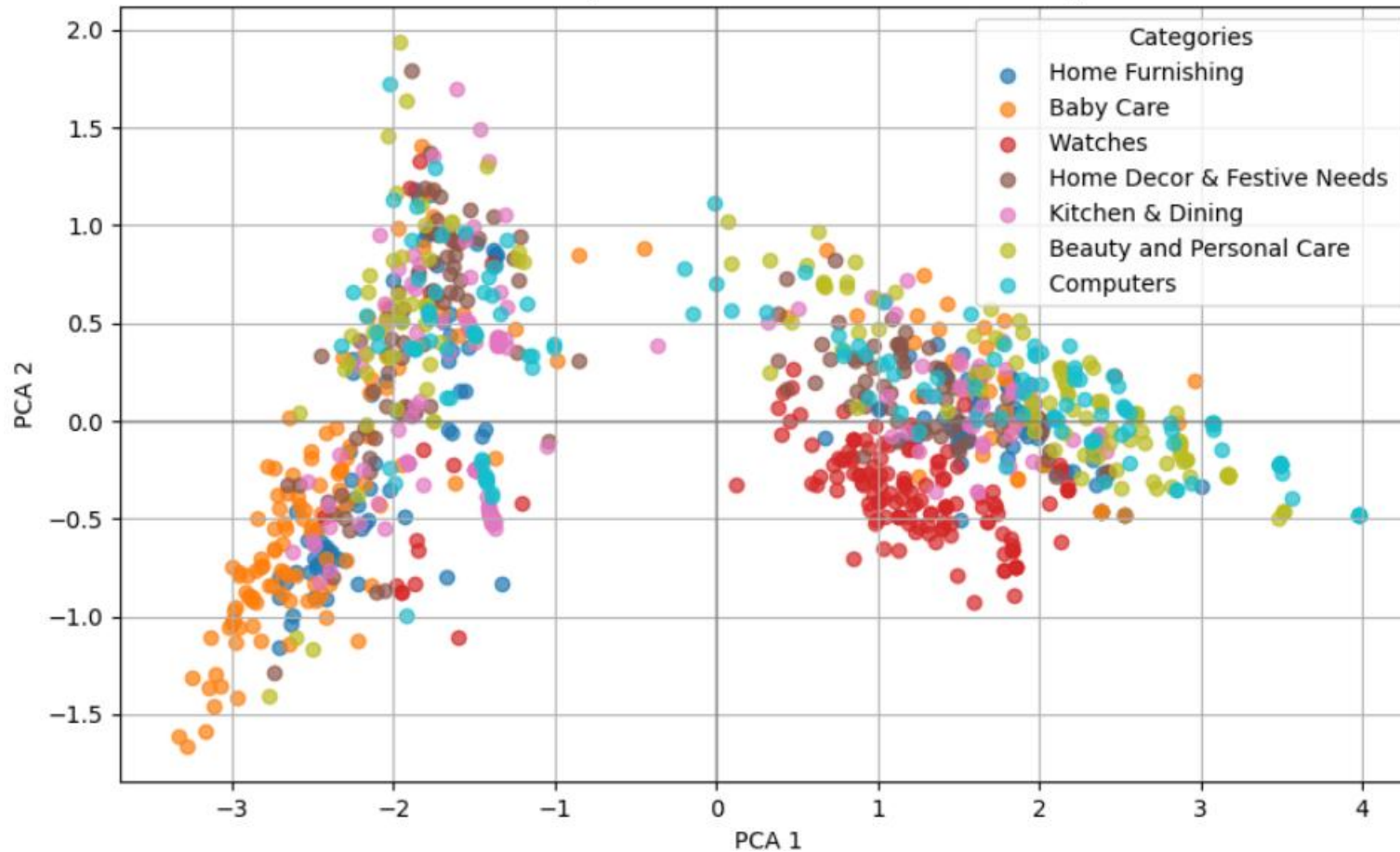
PCA des descriptions codées embedding USE avec catégories



ARI méthode USE: 0.27

Encodage Word2vec

PCA des descriptions codées Word2vec avec catégories



**Classement méthodes
d'encodage en fonction
de ARI**

- 1 – Embedding USE**
- 2 – Embedding BERT**
- 3 – Embedding Word2vec**

ARI méthode Word2vec: -0.14

Partie1: Analyses d'images (méthode simple)

Etape de transformation d'images



Image réelle



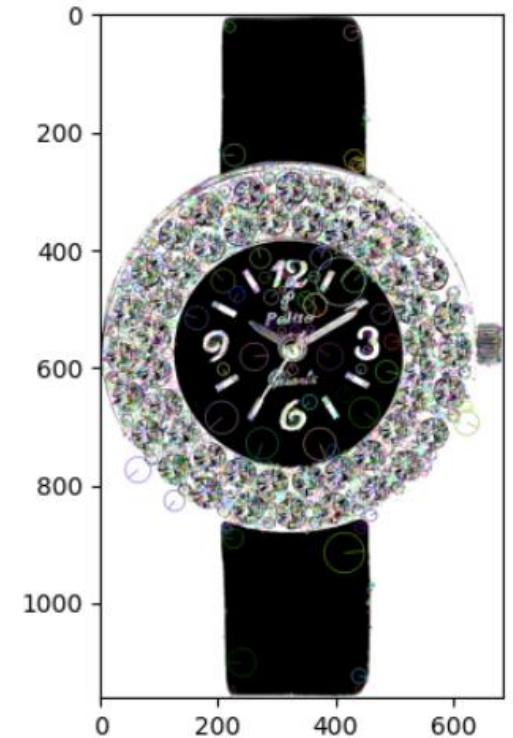
Niveau de gris



Filtrage du bruit



contraste

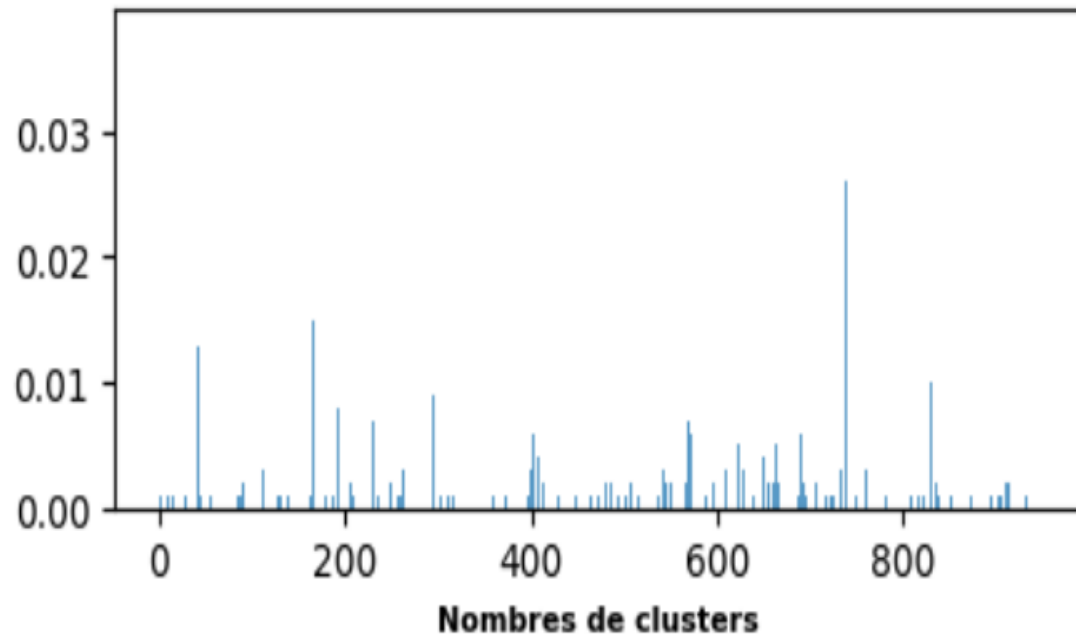


Descripteurs sift

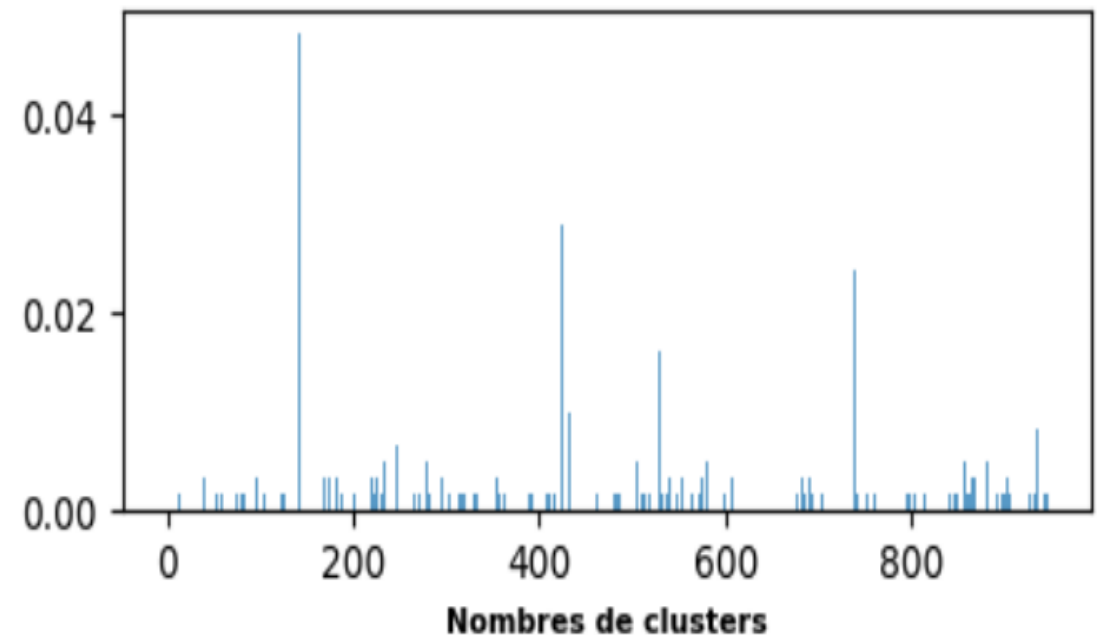
Histogrammes des images

Comptage pour chaque numéro de cluster du nombre de descripteurs de l'image

Histogramme de image 009099b1f6e1e8f893ec29a7023153c4.jpg



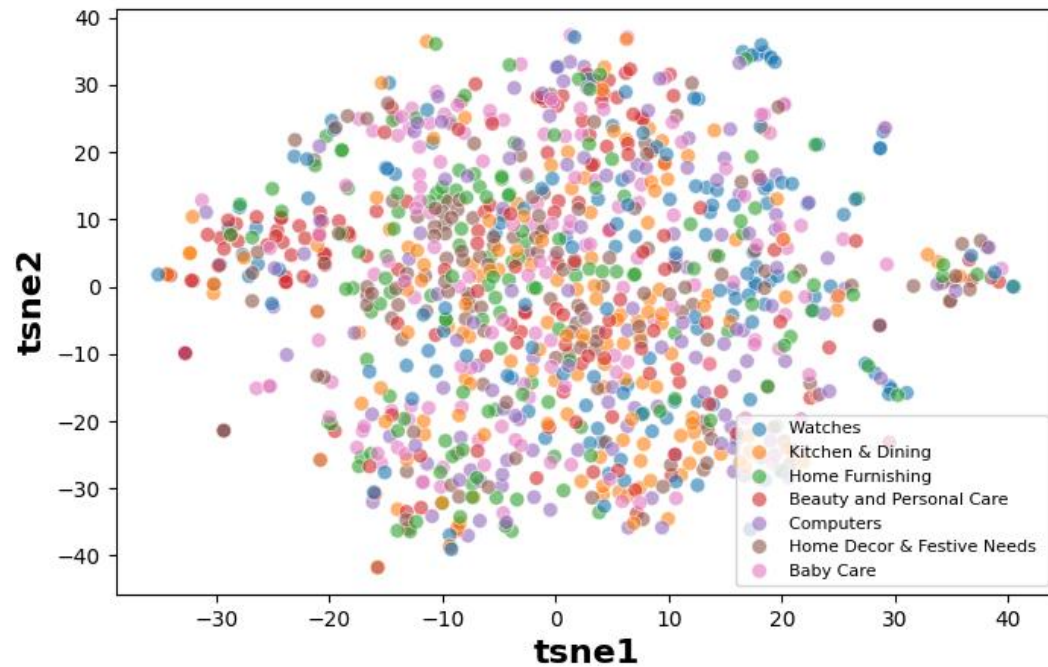
Histogramme de image 0096e89cc25a8b96fb9808716406fe94.jpg



Nombre Maximale de descripteurs sift: 947

Réduction de dimension t-SNE

TSNE selon les vraies classes



TSNE selon les clusters

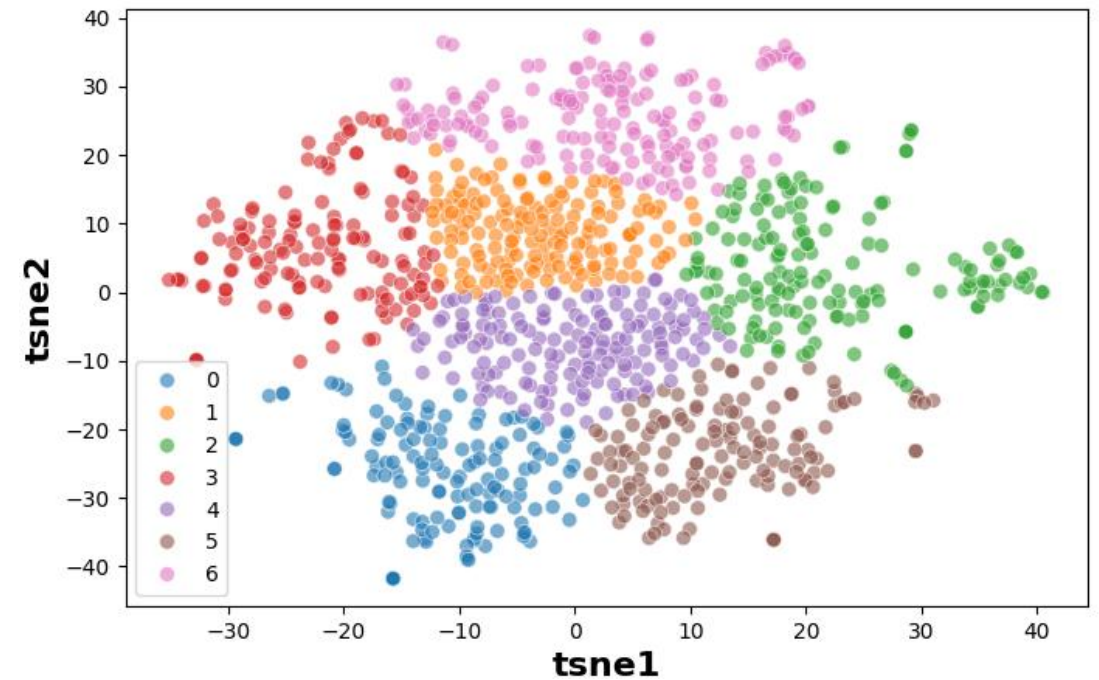
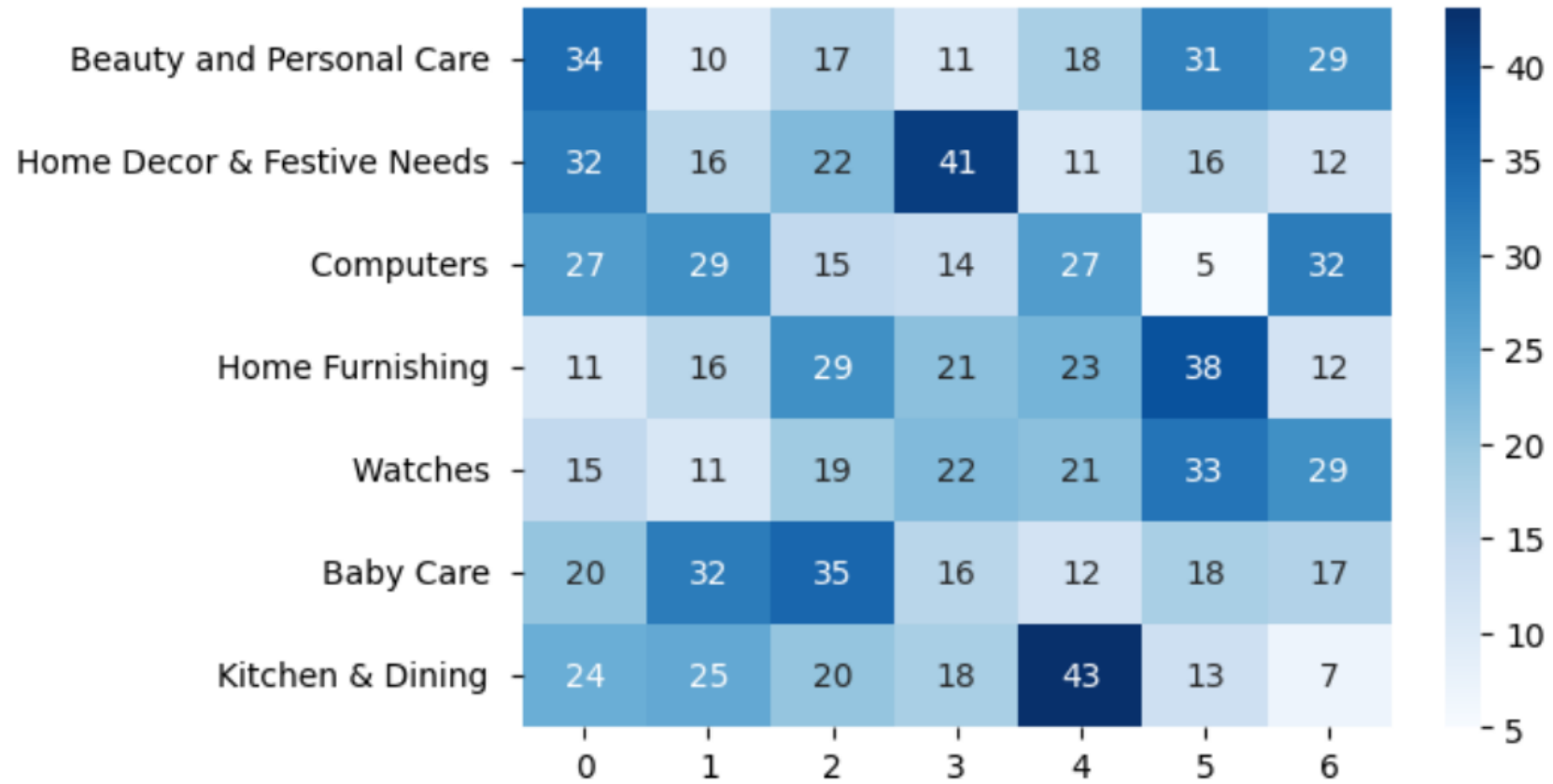


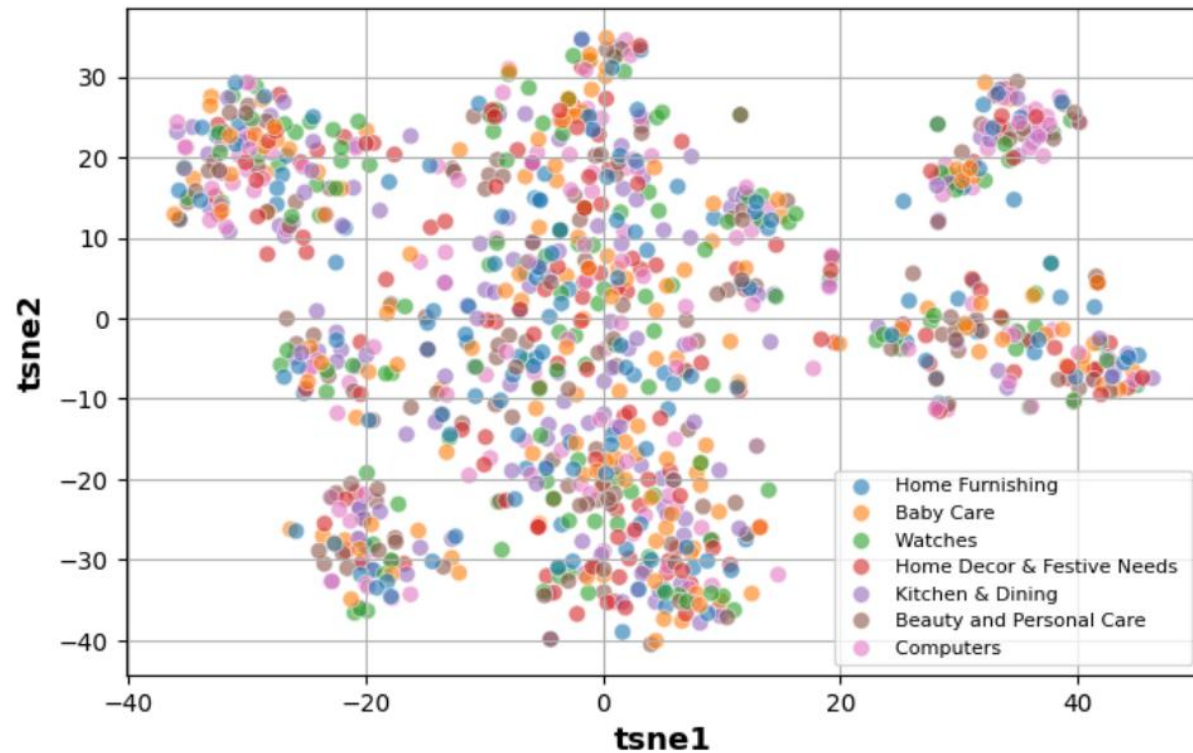
Tableau de contingence



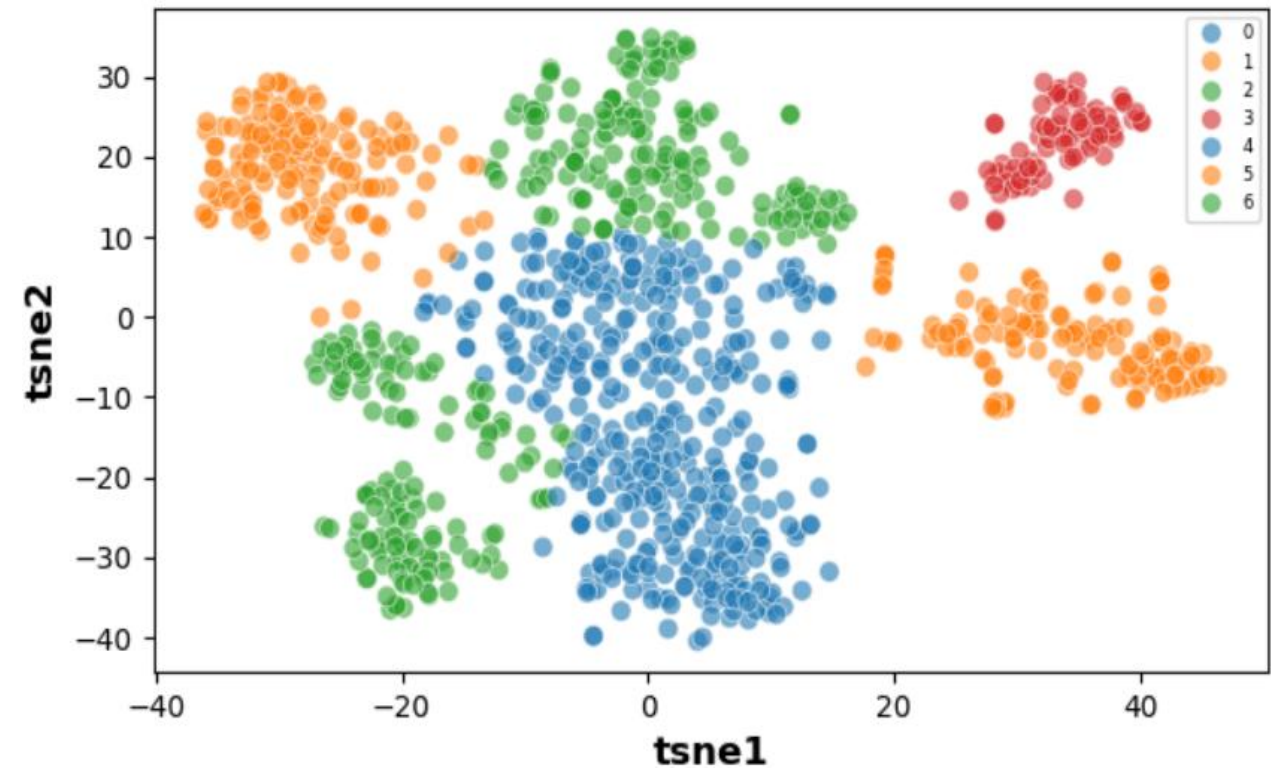
Partie 2: Modèle pré-entraîné VGG16

Réduction de Dimension

TSNE selon les vraies classes

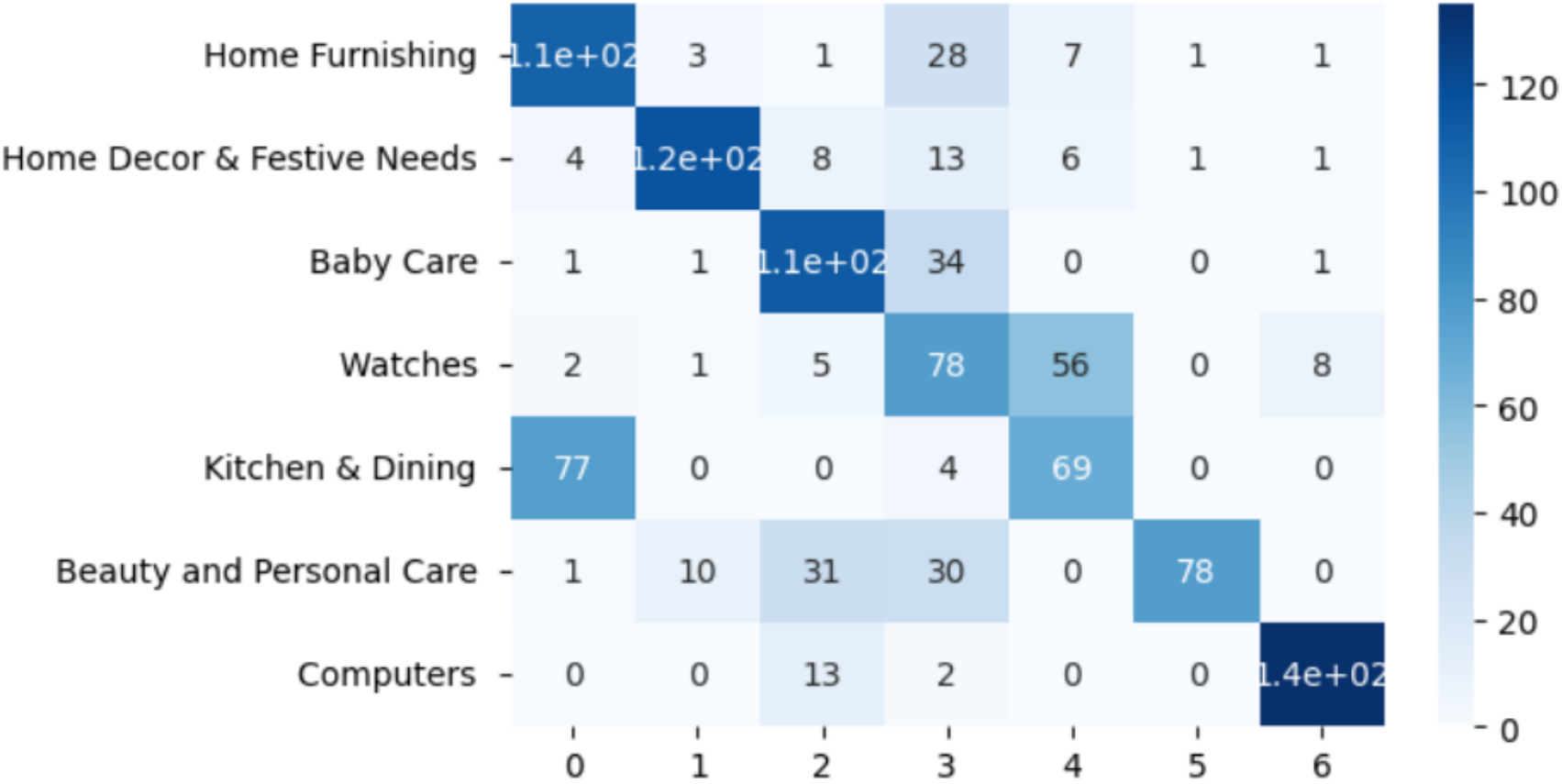


TSNE selon les clusters



L'indice de rand ajusté(ARI) : 0.45

Tableau de contingence



Classification Supervisée

1- Approche simple par préparation initiale de l'ensemble des images avant classification supervisée

Performance du modèle

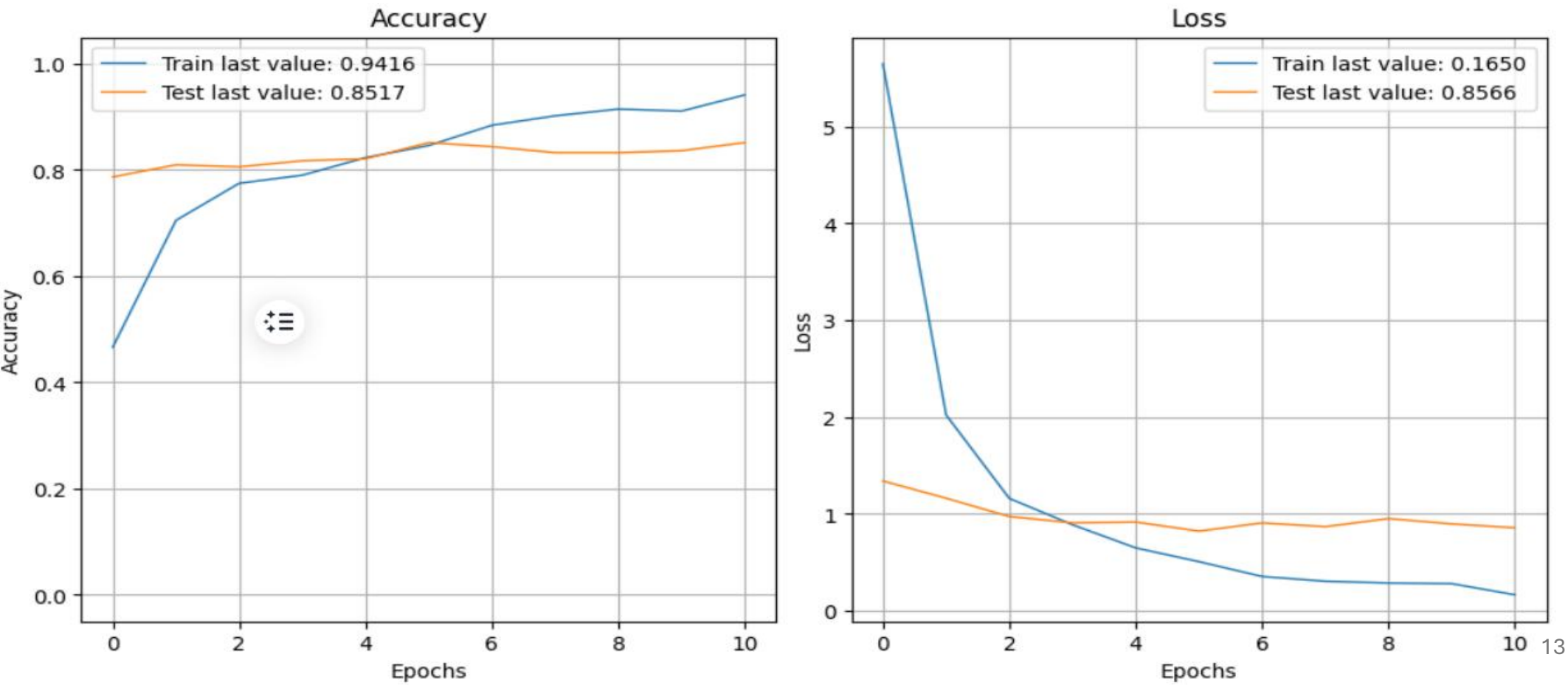
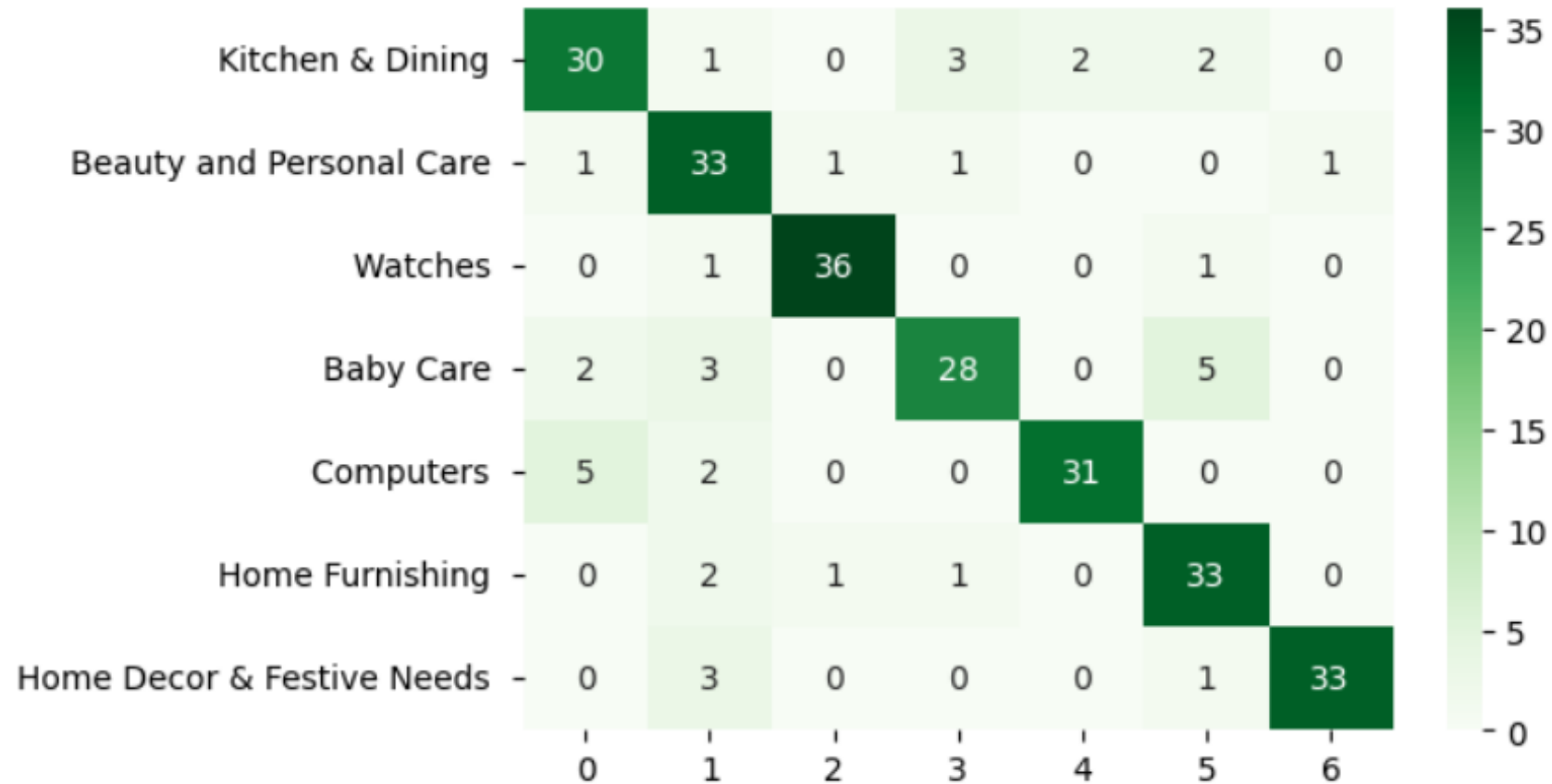


Tableau de contingente



Classification Supervisée

Modèle pré-entraîné Solution puissante de classification d'image, permettant de tirer parti de réseaux de neurones préalablement entraînés sur de larges ensembles de données.

Méthodes

- ☐ 1- Extraction des caractéristiques
- ☐ 2- Fine-tuning

Extraction des caractéristiques des images

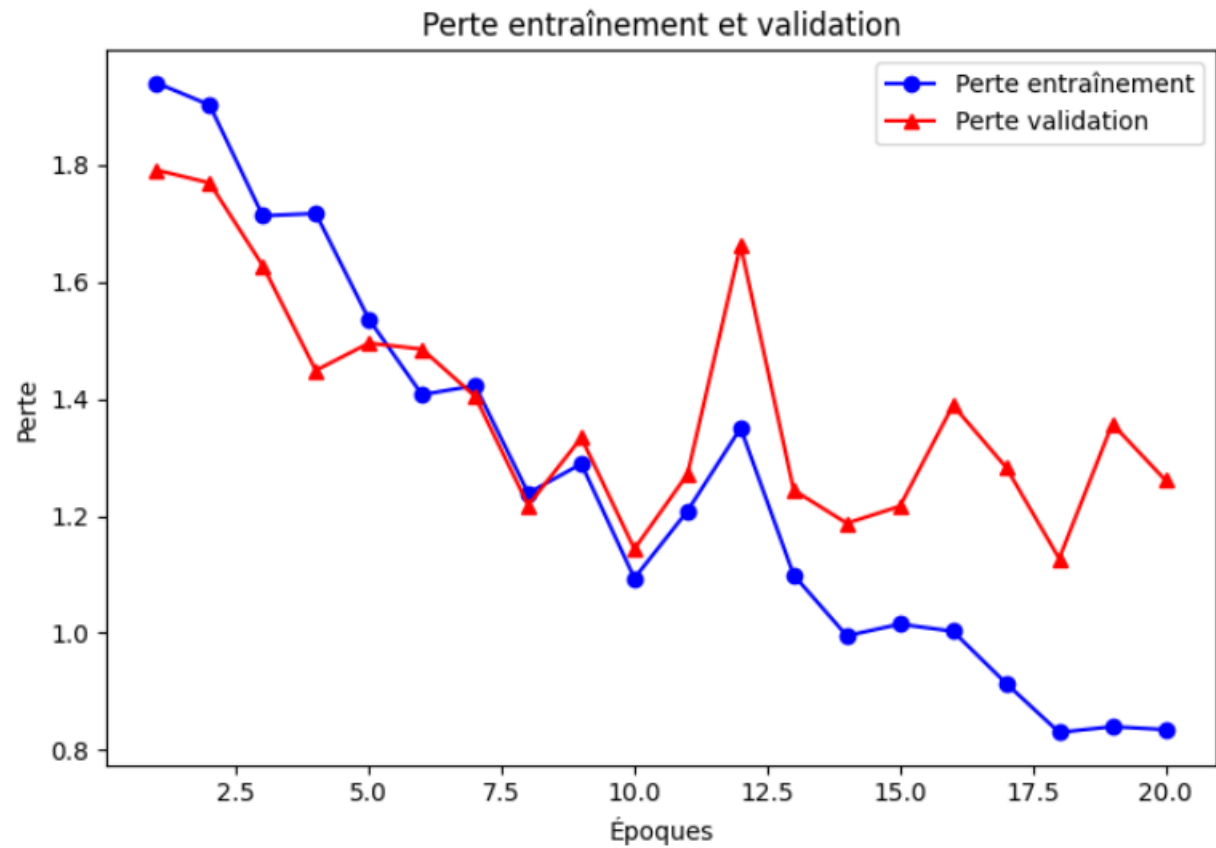
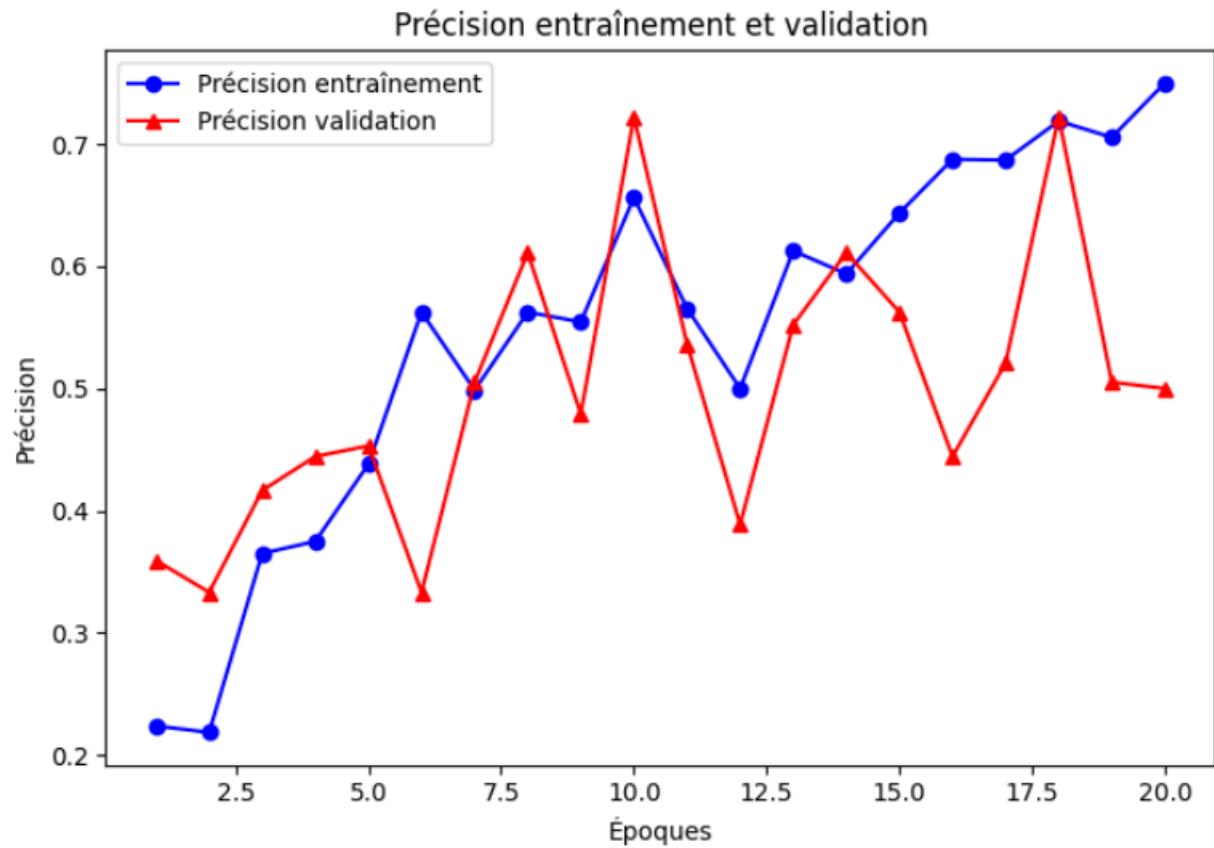
Les couches convolutives du modèle pré entraîné sont utilisées pour extraire des caractéristiques des images, qui sont ensuite utilisées par un classificateur pour effectuer la classification.

Etapes:

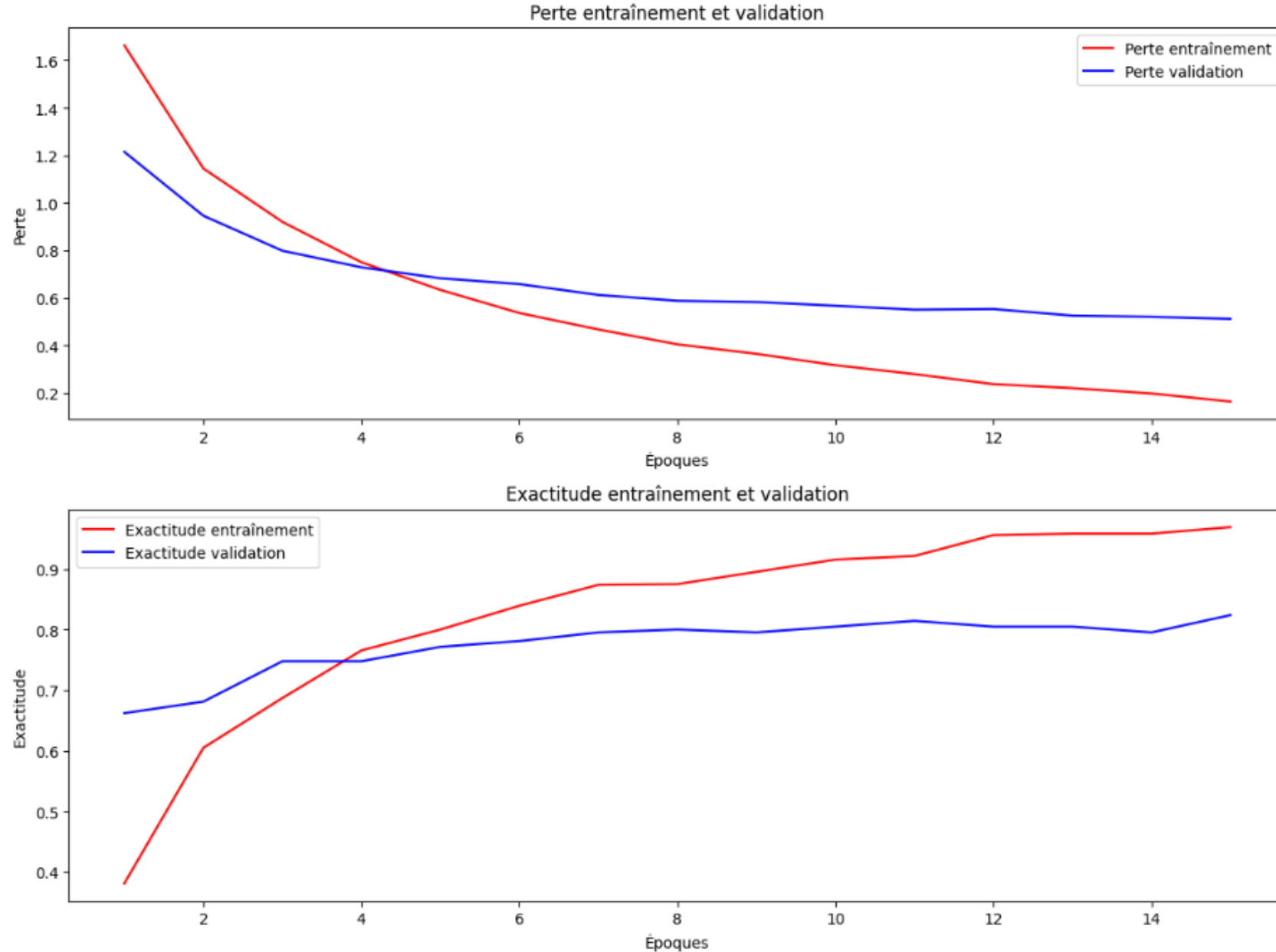
- 1. Charger un modèle pré entraîné sans la couche de classification finale.**
- 2. Geler les poids des couches convolutives pour éviter de les entraîner à nouveau.**
- 3. Ajouter une ou plusieurs couches de classification.**
- 4. Entraîner uniquement les nouvelles couches de classification avec les nouvelles données.**

2- Approche par Data generator

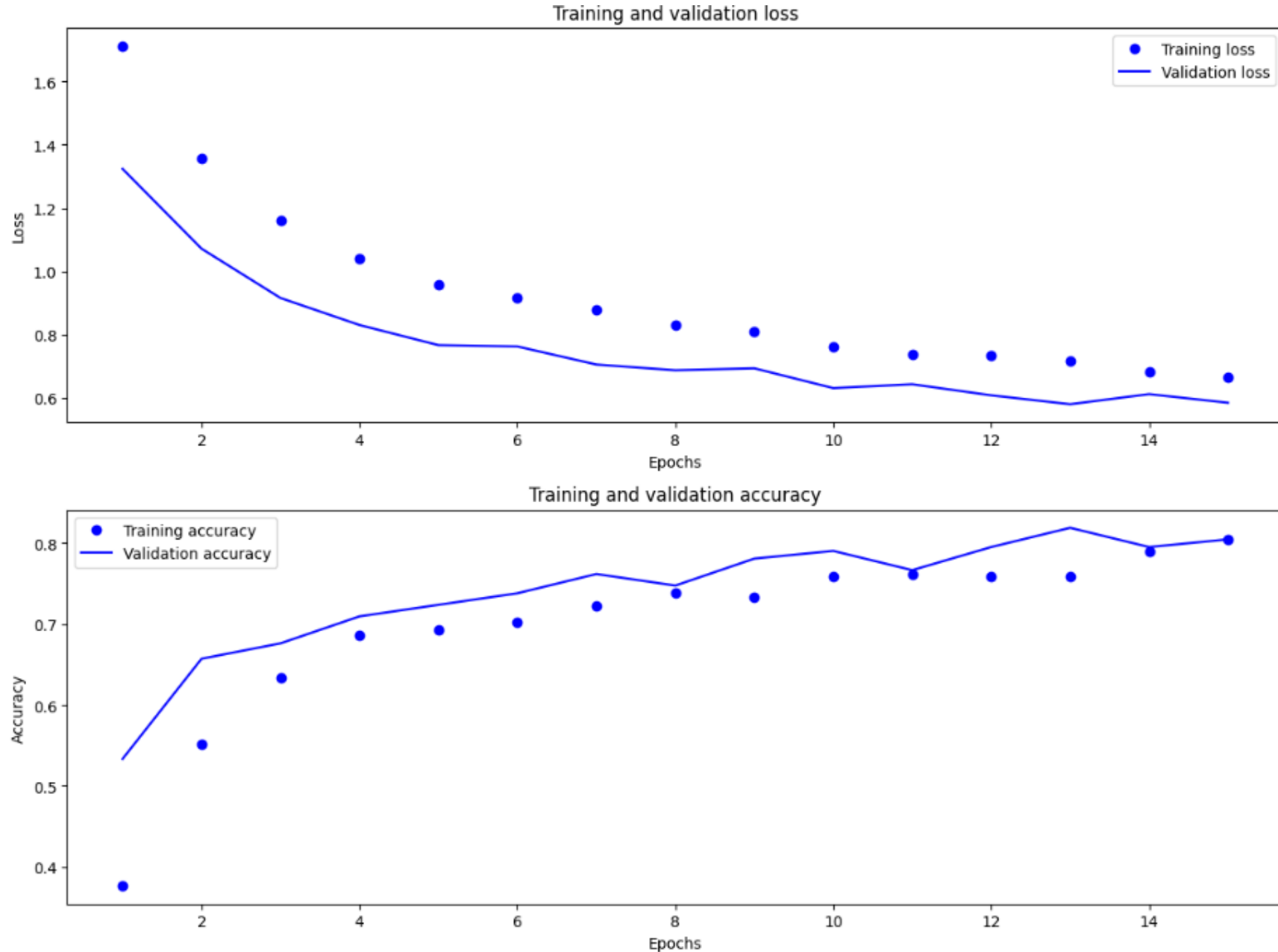
Performance du modèle



Performance du modèle sans data augmentation



Performance du modèle avec data augmentation



Test d'API de collecte de donnée

| | foodId | label | category | foodContentsLabel | image |
|---|---------------|---|---|---|---|
| 0 | 3039820510250 | Vivien Paille Lentilles vertes le paquet de 500 g | en:plant-based-foods-and-beverages, en:plant-b... | Lentilles vertes. | https://images.openfoodfacts.org/images/produ... |
| 1 | 3292070010264 | Betteraves de Champagne & chèvre crémeux, poin... | en:spreads, en:salted-spreads | Pois chiches 44% - purée de betterave rouge 37... | https://images.openfoodfacts.org/images/produ... |
| 2 | 4056489276845 | Miel de luzerne de la champagne | en:breakfasts, en:spreads, en:sweet-spreads, e... | NaN | https://images.openfoodfacts.org/images/produ... |
| 3 | 3560070962334 | Miel des terres blanches de Champagne | en:breakfasts, en:spreads, en:sweet-spreads, e... | Miel de la Champagne (France). | https://images.openfoodfacts.org/images/produ... |
| 4 | 4820097815556 | Splashes of champagne | NaN | NaN | https://images.openfoodfacts.org/images/produ... |
| 5 | 3049610004104 | Veuve Clicquot Champagne Ponsardin Brut | en:beverages, en:alcoholic-beverages, en:wines... | Champagne | https://images.openfoodfacts.org/images/produ... |
| 6 | 3282946015837 | Nicolas Feuillatte | en:beverages, en:alcoholic-beverages, en:wines... | Champagne, Contient des _sulfites_ | https://images.openfoodfacts.org/images/produ... |
| 7 | 3760091726964 | Lentilles roses bio | en:plant-based-foods-and-beverages, en:plant-b... | Lentilles | https://images.openfoodfacts.org/images/produ... |
| 8 | 3114080034057 | Champagne rosé | en:beverages, en:alcoholic-beverages, en:wines... | NaN | https://images.openfoodfacts.org/images/produ... |
| 9 | 3596710488803 | Miel de fleurs de la Champagne - 375g | en:breakfasts, en:spreads, en:sweet-spreads, e... | Miel de fleurs de la Champagne - Récolté en Fr... | https://images.openfoodfacts.org/images/produ... |

Conclusion

**Merci Pour
Votre
Attention**