

Étude de faisabilité d'un moteur de classification d'articles basé sur une image et une description

- Baby Care
- Beauty and Personal Care
- Computers
- Home Decor & Festive Needs
- Home Furnishing
- Kitchen & Dining
- Watches

Marc Sellam

Projet 6 Data Scientist 01/2023

Sommaire

- Mission
- Les données fournies
- Analyses des données
- Plan d'analyses
- Analyse du texte
- Analyse des images
- Analyse images + texte
- Conclusion

Page 1.

Pages 2.

Pages 3 et 4.

Page 5.

Pages 6 à 9.

Pages 10 à 15.

Page 16.

Page 17.

Mission

L'entreprise Place de marché souhaite lancer une marketplace e-commerce.

Les vendeurs y proposent **des articles** en postant **une photo et une description**.

Ils attribuent eux-mêmes les **catégories des articles**, et cela de manière peu fiable.

L'entreprise souhaite étudier la faisabilité d'un **moteur de classification** des articles en différentes catégories, avec un niveau de précision suffisant.

Les données fournies

Place de marché fourni un fichier compressé :

- Un fichier .csv (1050, 15).
- Un dossier : 1050 photos.

Analyse des données

15 variables disponibles
4 variables sélectionnées

	% manquants	type	unique	échantillon des 1 premières valeurs
uniq_id	0.000000	object	1050	[55b85ea15a1536d46b7190ad6fff8ce7]
crawl_timestamp	0.000000	object	149	[2016-04-30 03:22:56 +0000]
product_url	0.000000	object	1050	[http://www.flipkart.com/elegance-polyester-multicolor-abstract-eyelet-door-curtain/p/itmeg7bkzdgthcts?pid=CRNEG7BKMFFYHQ8Z]
product_name	0.000000	object	1050	[Elegance Polyester Multicolor Abstract Eyelet Door Curtain]
product_category_tree	0.000000	object	642	[["Home Furnishing >> Curtains & Accessories >> Curtains >> Elegance Polyester Multicolor Abstract Eyelet Do..."]]
pid	0.000000	object	1050	[CRNEG7BKMFFYHQ8Z]
image	0.000000	object	1050	[55b85ea15a1536d46b7190ad6fff8ce7.jpg]
is_FK_Advantage_product	0.000000	bool	2	[False]
description	0.000000	object	1050	[Key Features of Elegance Polyester Multicolor Abstract Eyelet Door Curtain Floral Curtain,Elegance Polyester Multicolor Abstract Eyelet Door Curtain (213 cm in Height, Pack of 2) Price: Rs. 899 This curtain enhances the look of the interiors.This curtain is made from 100% high quality polyester fabric.It features an eyelet style stitch with Metal Ring.It makes the room environment romantic and loving.This curtain is ant- wrinkle and anti shrinkage and have elegant apparence.Give your home a bright and modernistic appeal with these designs. The surreal attention is sure to steal hearts. These contemporary eyelet and valance curtains slide smoothly so when you draw them apart first thing in the morning to welcome the bright sun rays you want to wish good morning to the whole world and when you draw them close in the evening, you create the most special moments of joyous beauty given by the soothing prints. Bring home the elegant curtain that softly filters light in your room so that you get the right amount of sunlight. Specifications of Elegance Polyester Multicolor Abstract Eyelet Door Curtain (213 cm in Height, Pack of 2) General Brand Elegance Designed For Door Type Eyelet Model Name Abstract Polyester Door Curtain Set Of 2 Model ID Duster25 Color Multicolor Dimensions Length 213 cm In the Box Number of Contents in Sales Package Pack of 2 Sales Package 2 Curtains Body & Design Material Polyester]
product_rating	0.000000	object	27	[No rating available]
overall_rating	0.000000	object	27	[No rating available]
retail_price	0.095238	float64	355	[1899.0]
discounted_price	0.095238	float64	425	[899.0]
product_specifications	0.095238	object	985	[{"product_specification"=>{"key"=>"Brand", "value"=>"Elegance"}, {"key"=>"Designed For", "value"=>"Door"}, {"key"=>"Type", "value"=>"Eyelet"}, {"key"=>"Model Name", "value"=>"Abstract Polyester Door Curtain Set Of 2"}, {"key"=>"Model ID", "value"=>"Duster25"}, {"key"=>"Color", "value"=>"Multicolor"}, {"key"=>"Length", "value"=>"213 cm"}, {"key"=>"Number of Contents in Sales Package", "value"=>"Pack of 2"}, {"key"=>"Sales Package", "value"=>"2 Curtains"}, {"key"=>"Material", "value"=>"Polyester"}]]]
brand	32.190476	object	491	[Elegance]

Analyse des données

4 variables retenues

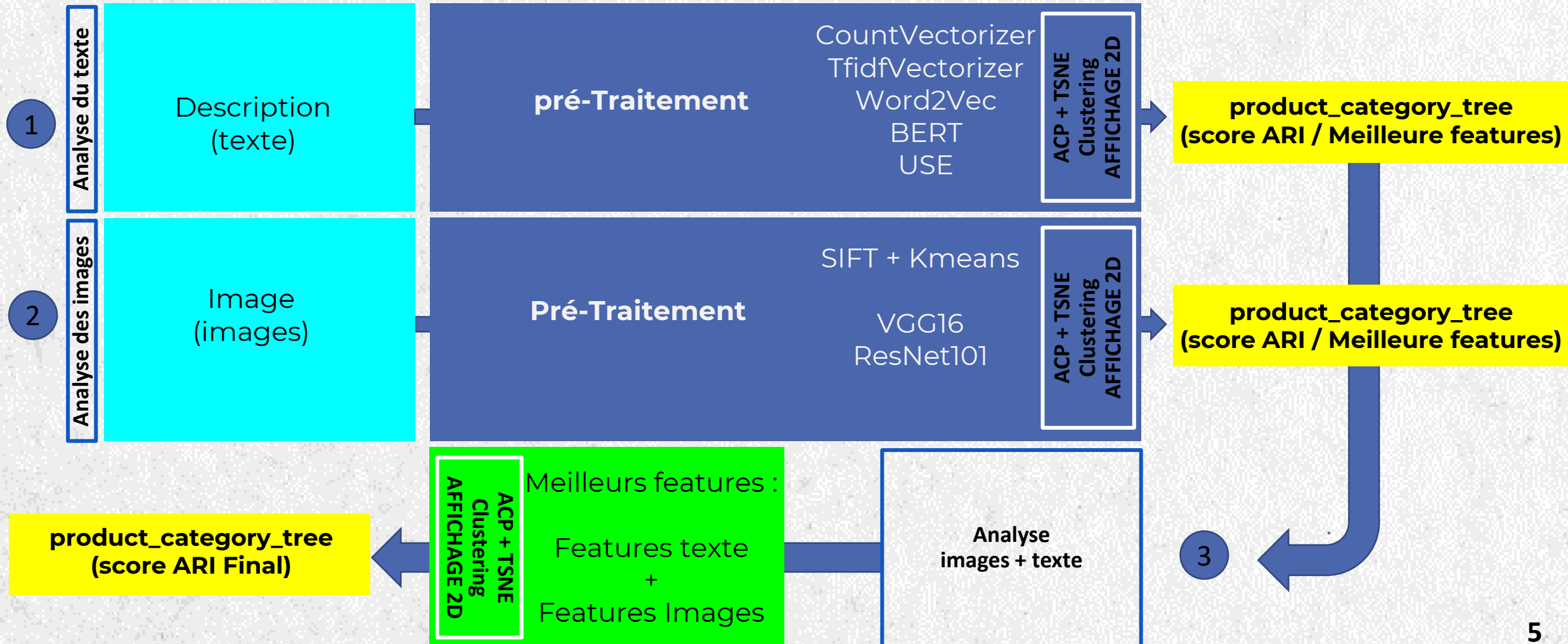
	% manquants	type	unique	échantillon des 1 premières valeurs
uniq_id	0.0	object	1050	['55b85ea15a1536d46b7190ad6fff8ce7']
product_category_tree	0.0	object	7	['Home Furnishing']
image	0.0	object	1050	['55b85ea15a1536d46b7190ad6fff8ce7.jpg']
description	0.0	object	1050	['Key Features of Elegance Polyester Multicolor Abstract Eyelet Door Curtain Floral Curtain,Elegance Polyester Multicolor Abstract Eyelet Door Curtain (213 cm in Height, Pack of 2) Price: Rs. 899 This curtain enhances the look of the interiors.This curtain is made from 100% high quality polyester fabric.It features an eyelet style stitch with Metal Ring.It makes the room environment romantic and loving.This curtain is ant- wrinkle and anti shrinkage and have elegant apparence.Give your home a bright and modernistic appeal with these designs. The surreal attention is sure to steal hearts. These contemporary eyelet and valance curtains slide smoothly so when you draw them apart first thing in the morning to welcome the bright sun rays you want to wish good morning to the whole world and when you draw them close in the evening, you create the most special moments of joyous beauty given by the soothing prints. Bring home the elegant curtain that softly filters light in your room so that you get the right amount of sunlight.,Specifications of Elegance Polyester Multicolor Abstract Eyelet Door Curtain (213 cm in Height, Pack of 2) General Brand Elegance Designed For Door Type Eyelet Model Name Abstract Polyester Door Curtain Set Of 2 Model ID Duster25 Color Multicolor Dimensions Length 213 cm In the Box Number of Contents in Sales Package Pack of 2 Sales Package 2 Curtains Body & Design Material Polyester']

```
data1['product_category_tree'].value_counts()
```

```
Computers          150
Baby Care          150
Kitchen & Dining    150
Beauty and Personal Care  150
Home Furnishing     150
Watches            150
Home Decor & Festive Needs  150
Name: product_category_tree, dtype: int64
```

La répartition des catégories est équilibrée

Plan d'analyses



1

Analyse du texte

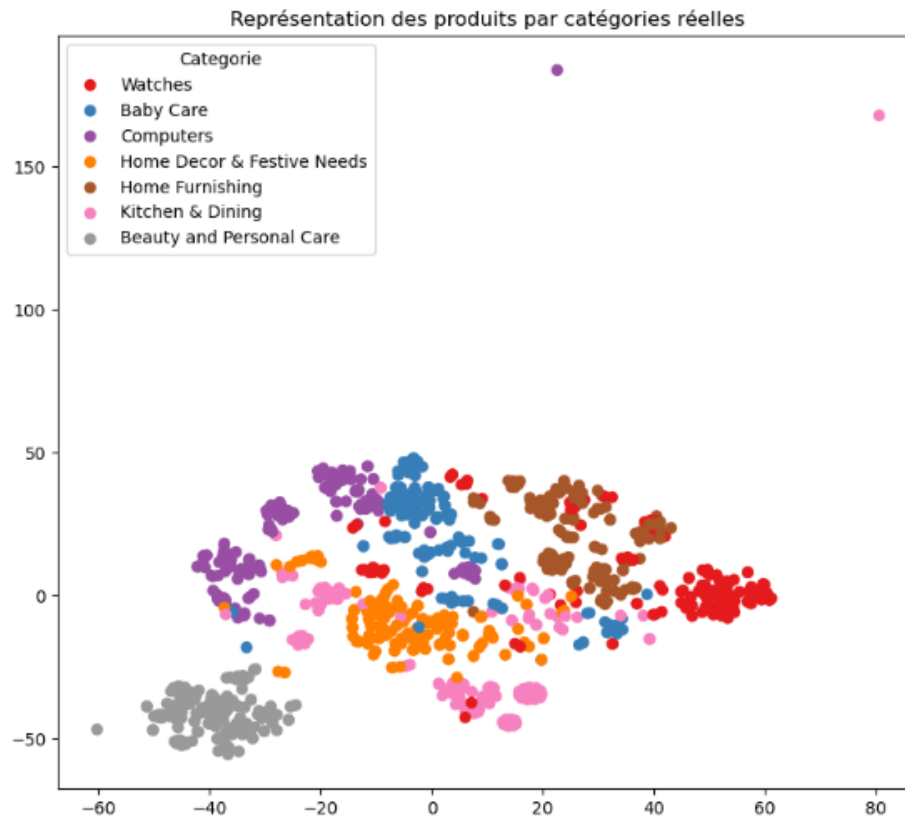
Résultat obtenu par l'approche CountVectorizer et Tf-idf

CountVectorizer :

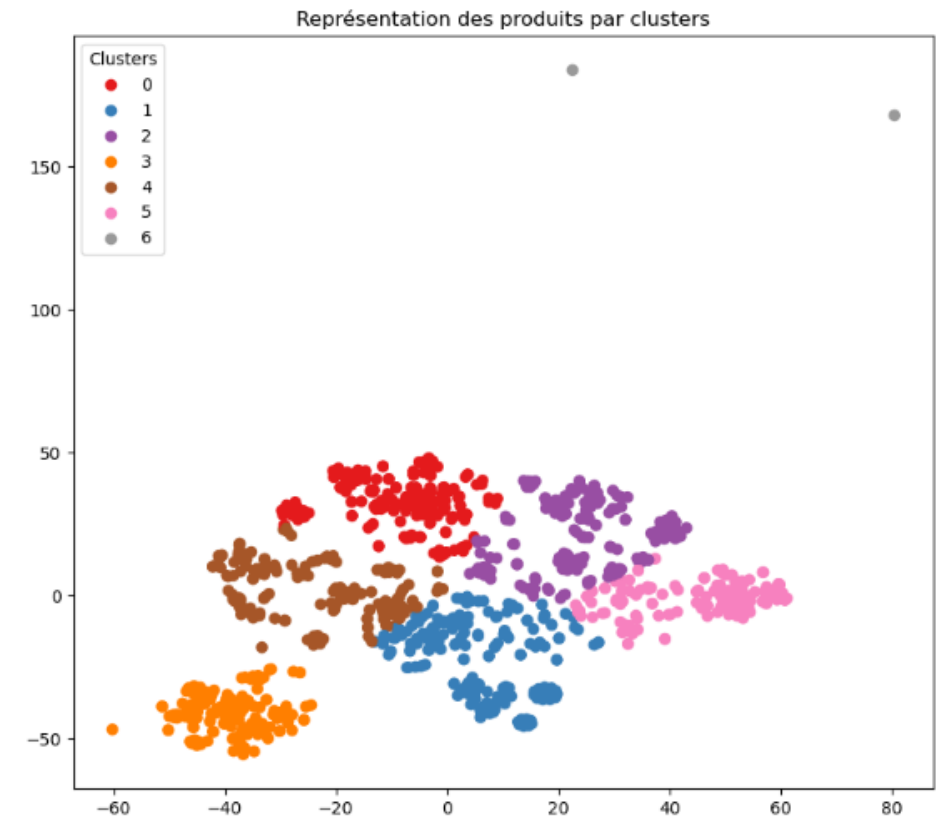
ARI : 0.3544 time : 5.0

Tf-idf :

ARI : 0.4284 time : 4.0



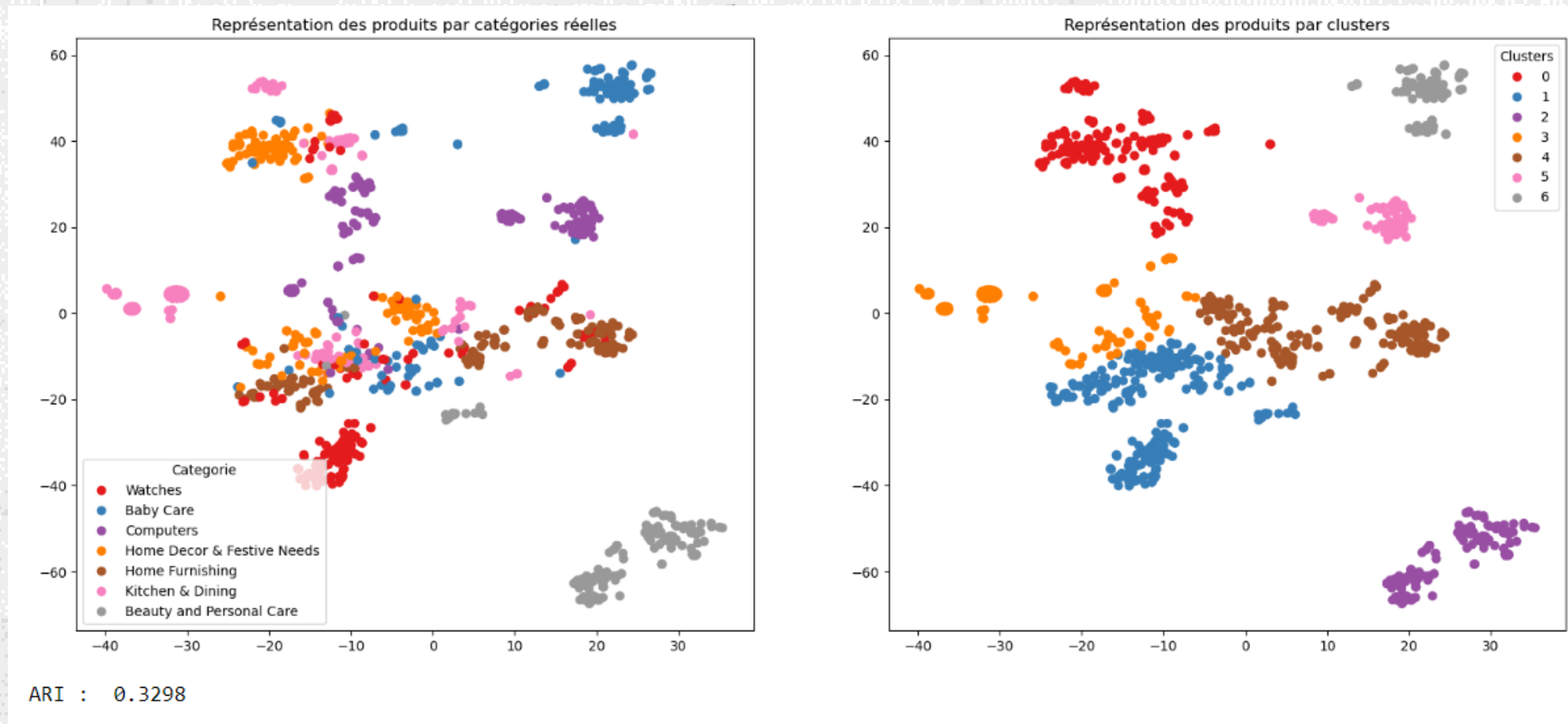
ARI : 0.4284



1

Analyse du texte

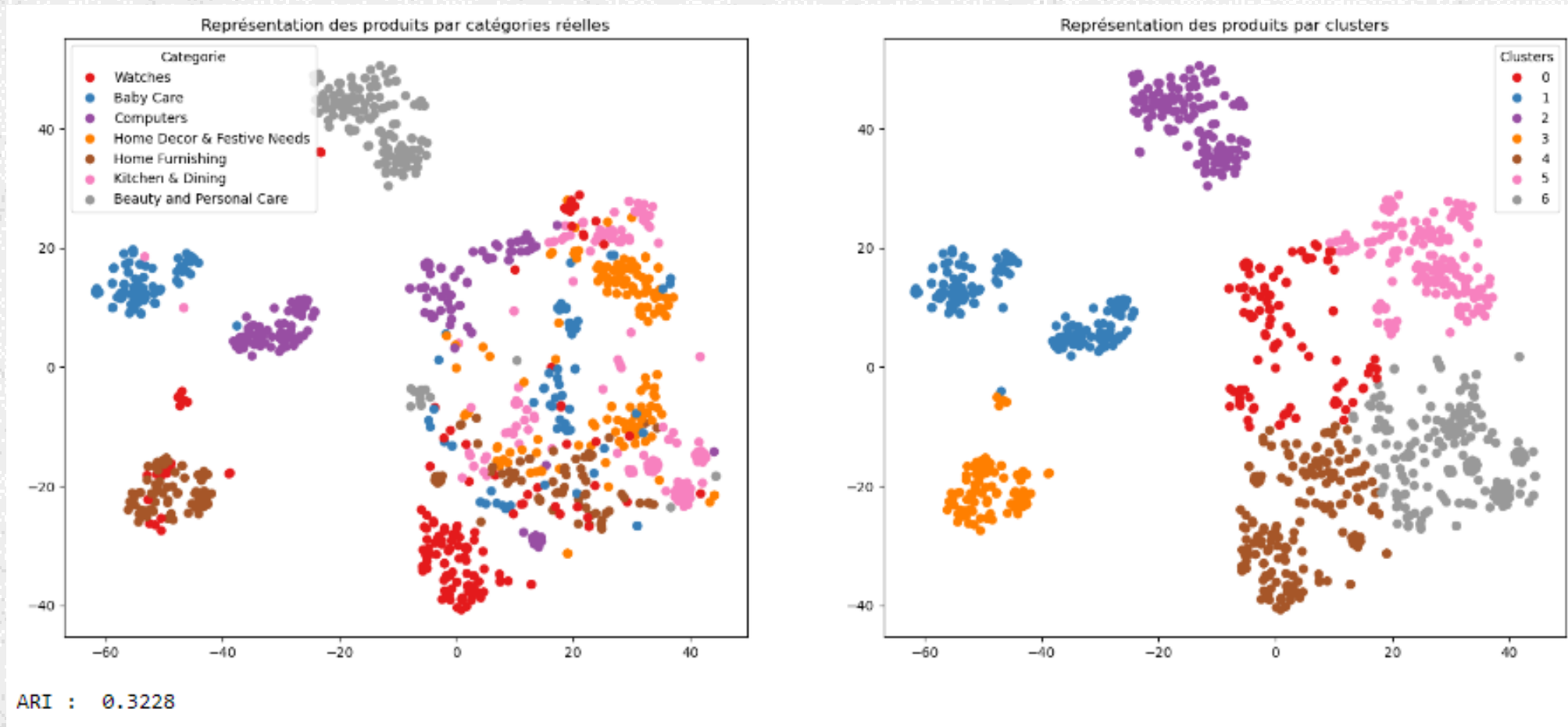
Résultat obtenu par l'approche Word2Vec



1

Analyse du texte

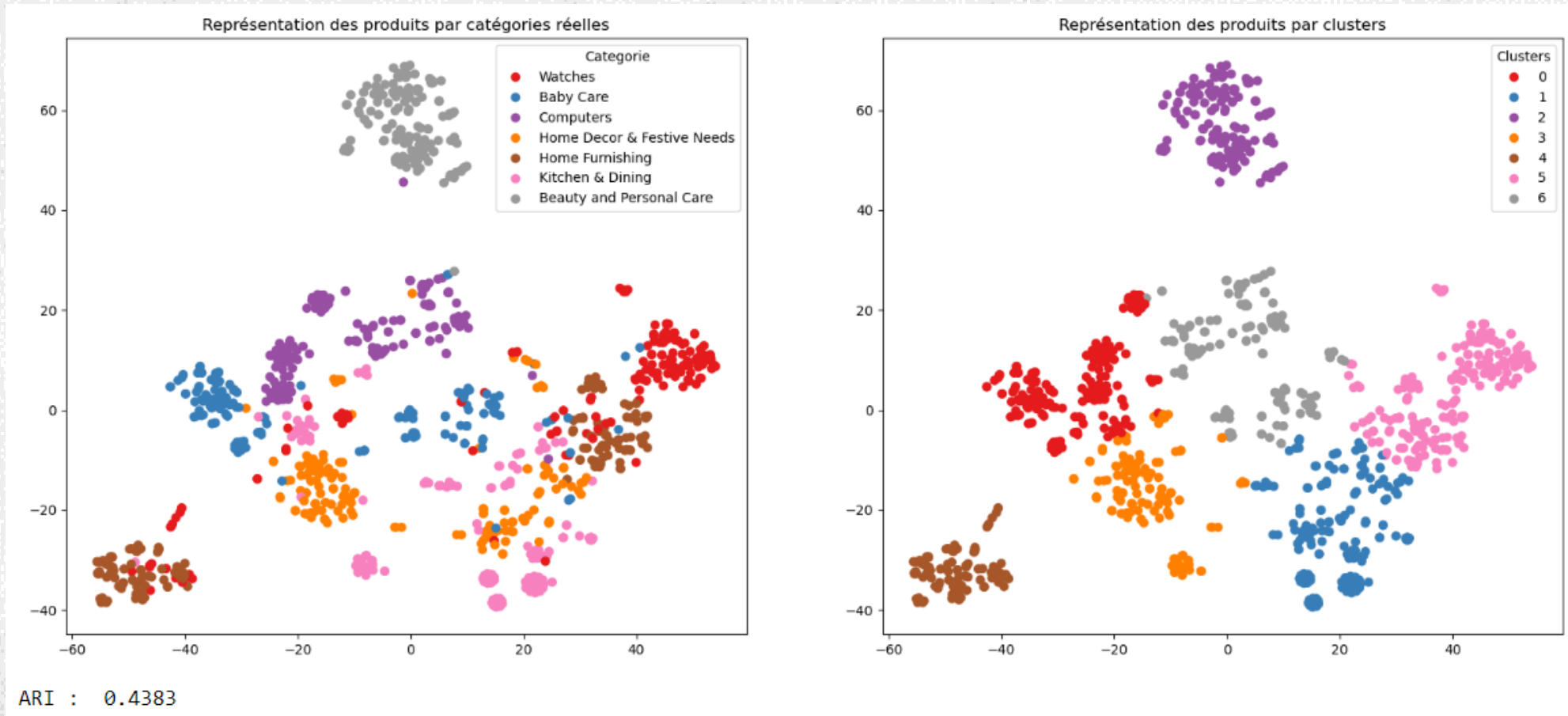
Résultat obtenu par l'approche BERT



1

Analyse du texte

Résultat obtenu par l'approche USE



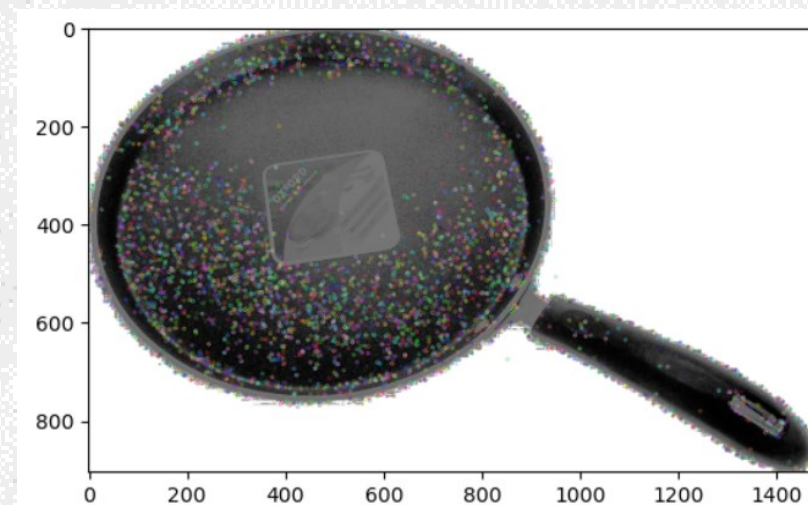
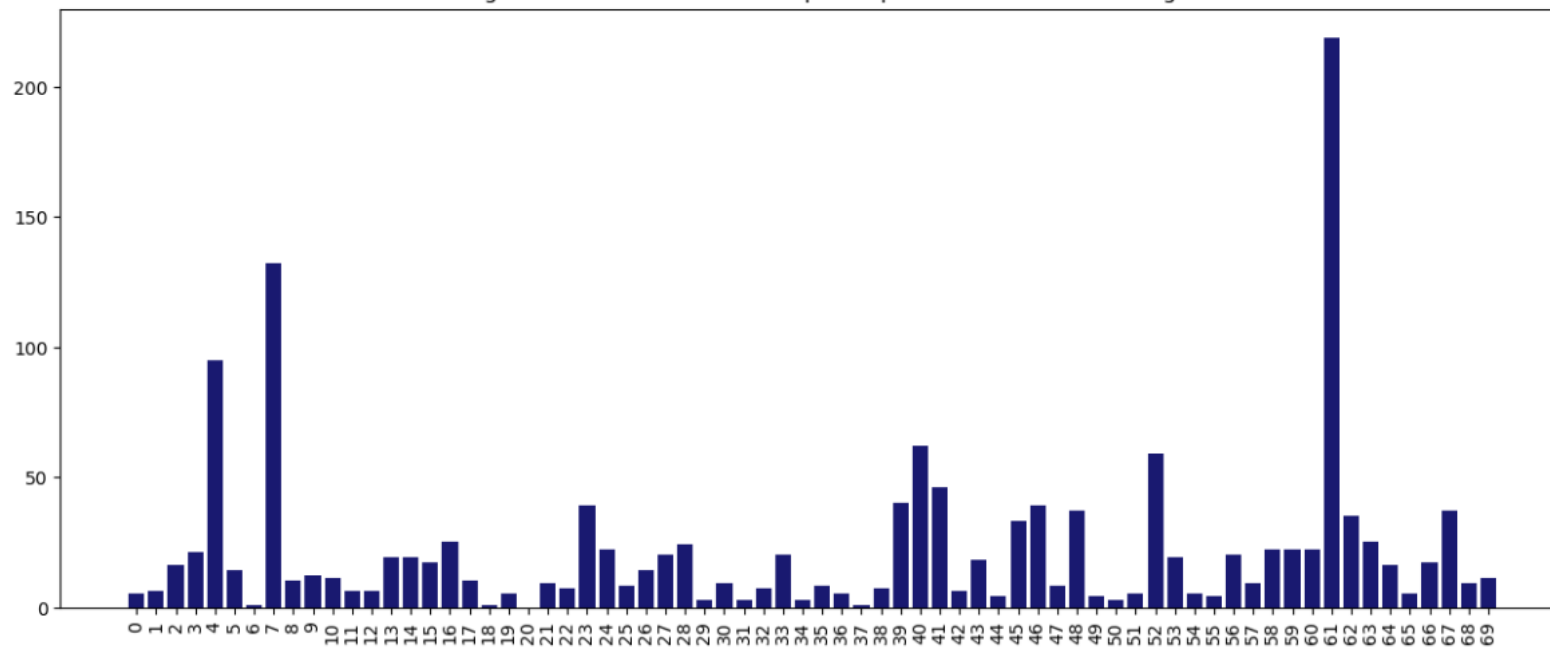
2 Analyse des images

Approche SIFT / KMeans

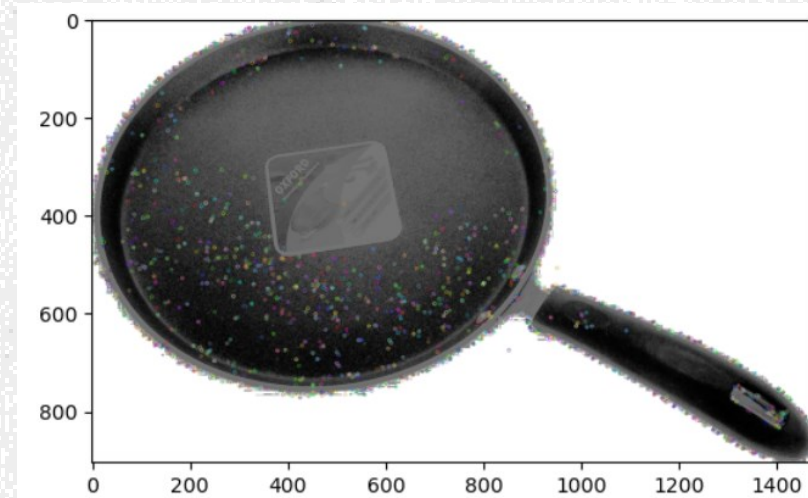
1500 descripteurs SIFT par image

1 417 165 descripteurs pour toutes les images

Histogramme du nombre de descripteurs par clusters de la 87e image



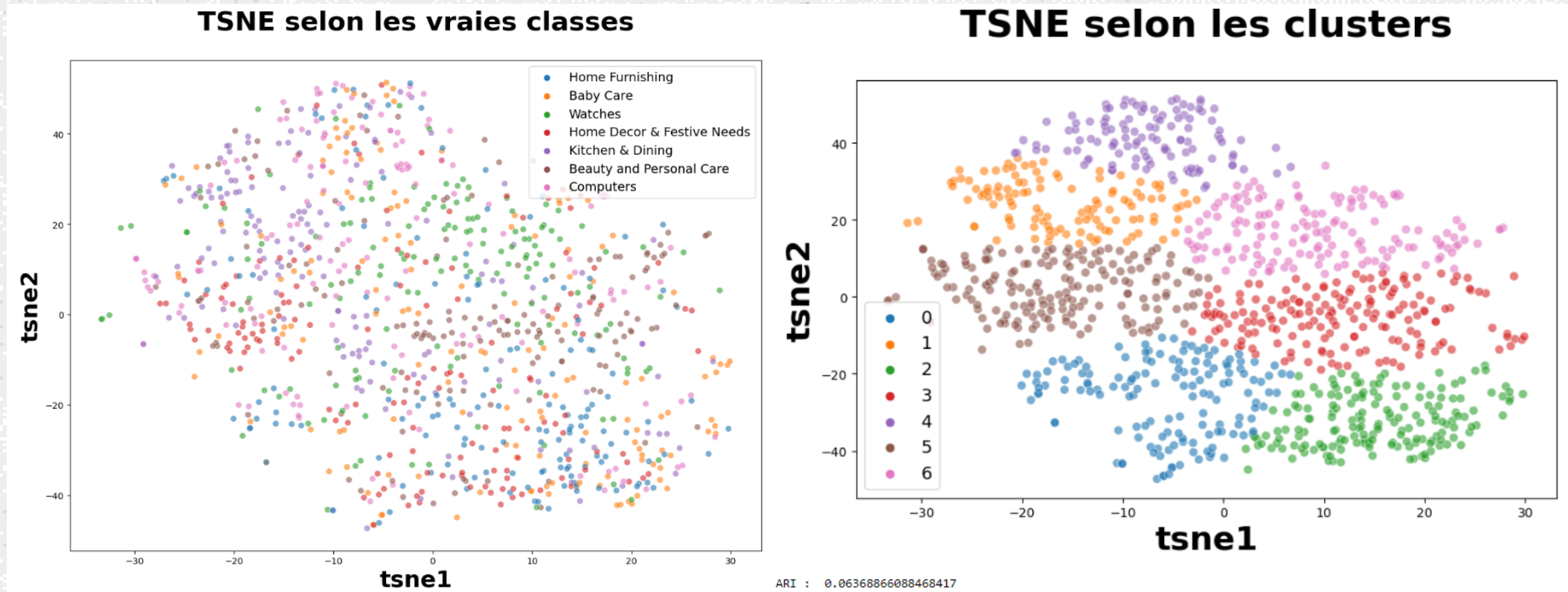
Descripteurs : (4656, 128)



Descripteurs : (1501, 128)

2 Analyse des images

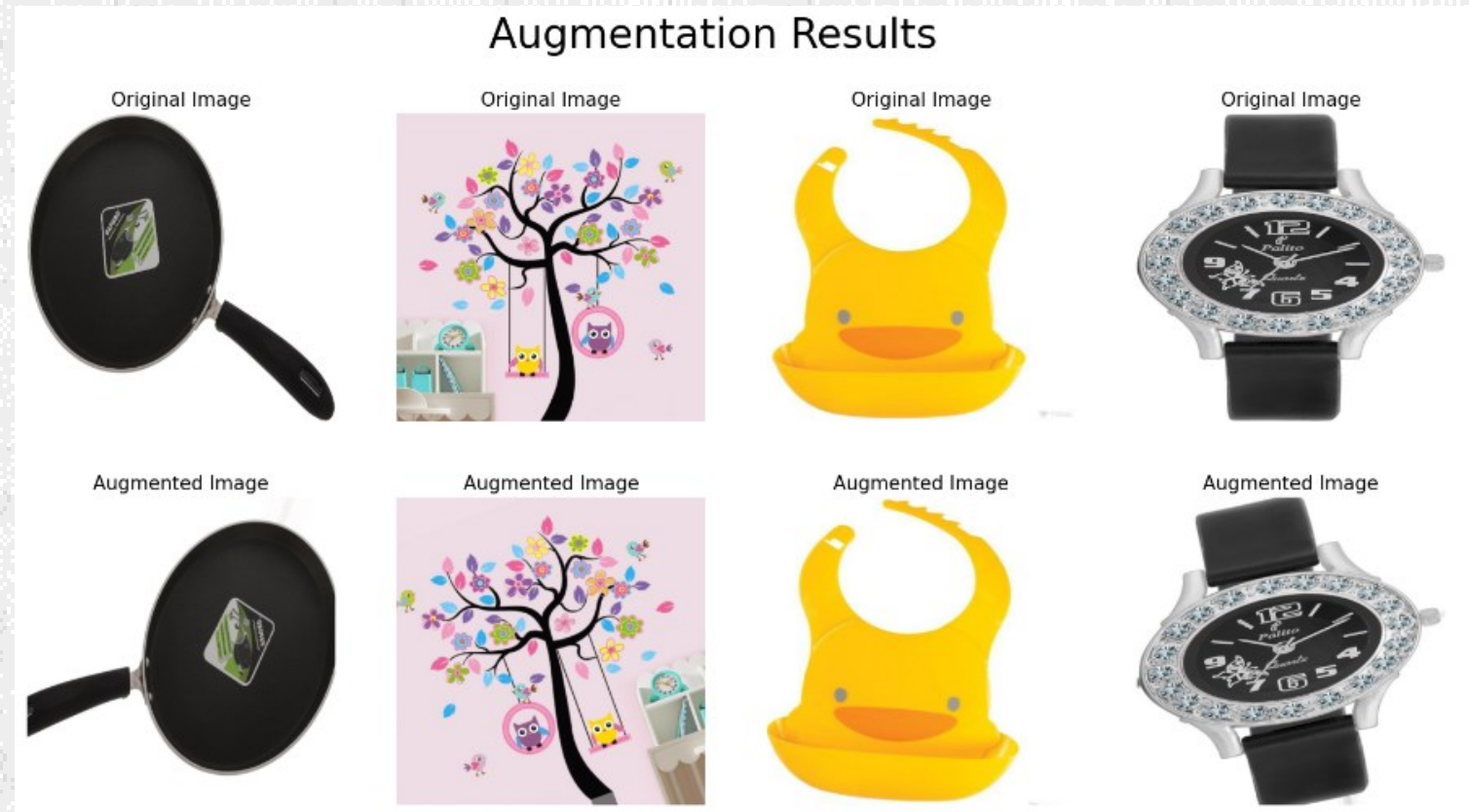
Résultat obtenu par l'approche SIFT



2 Analyse des images

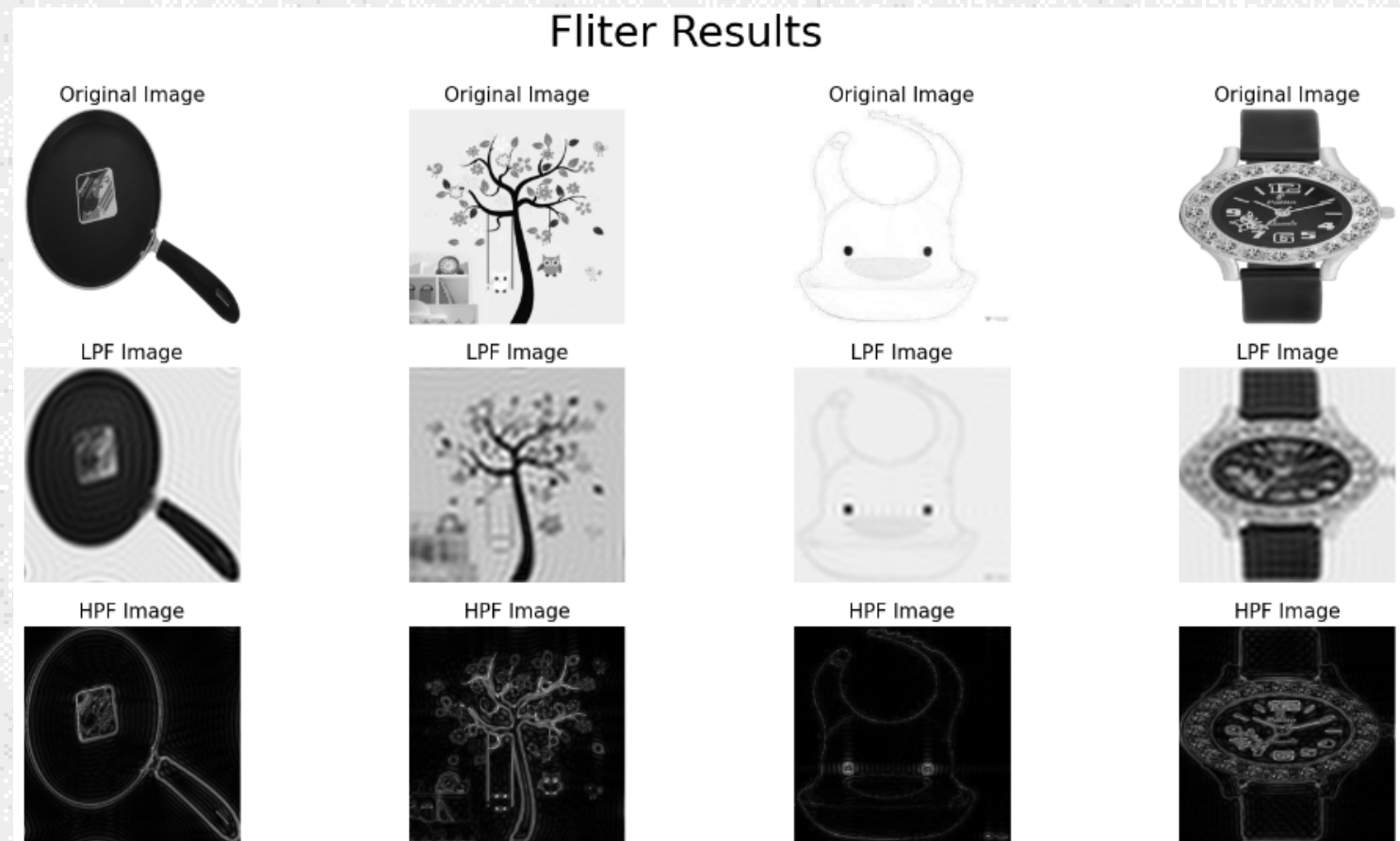
Approches VGG16 et ResNet101

Exemples d'augmentation



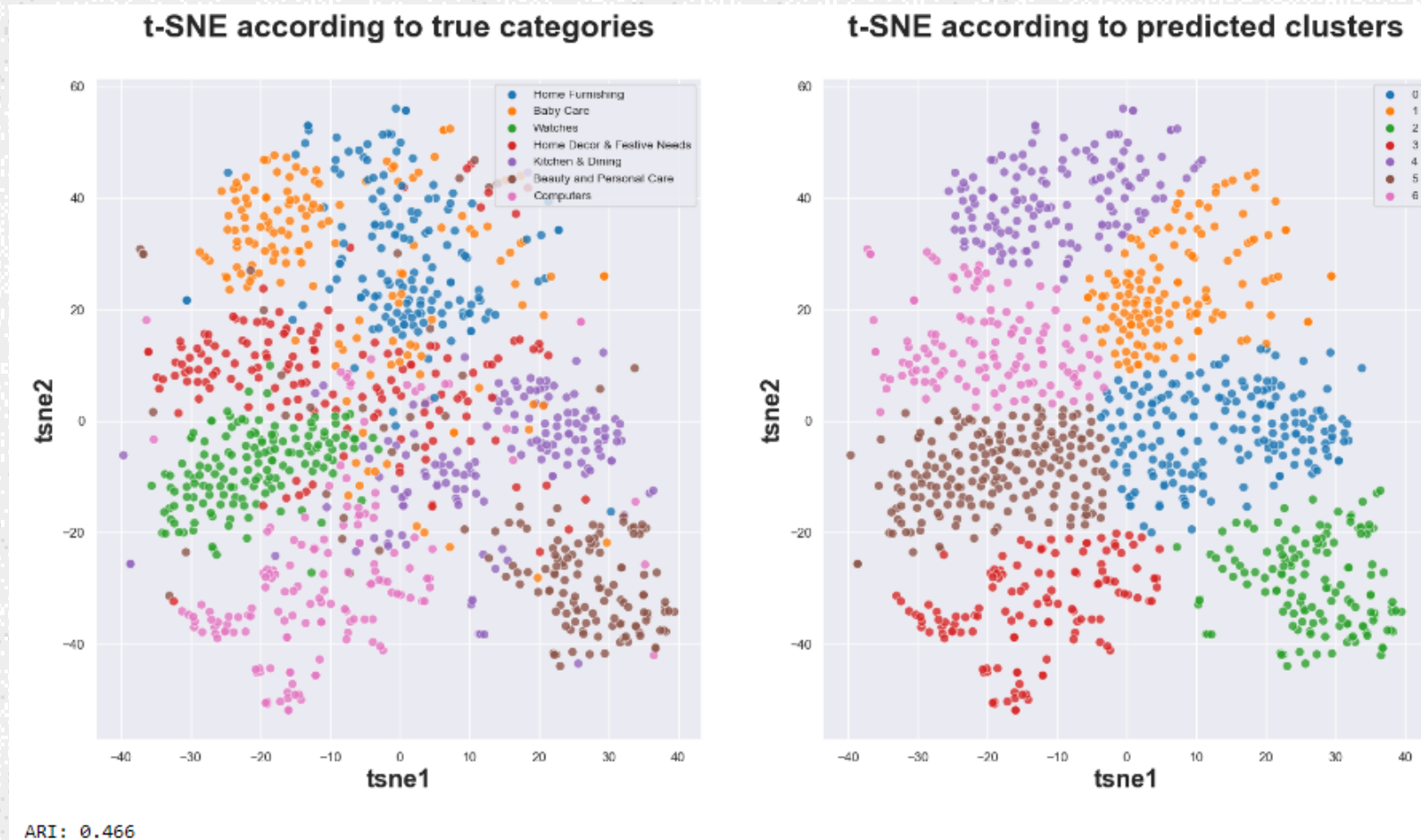
2 Analyse des images

Exemples d'application de
filtres LPF et HPF



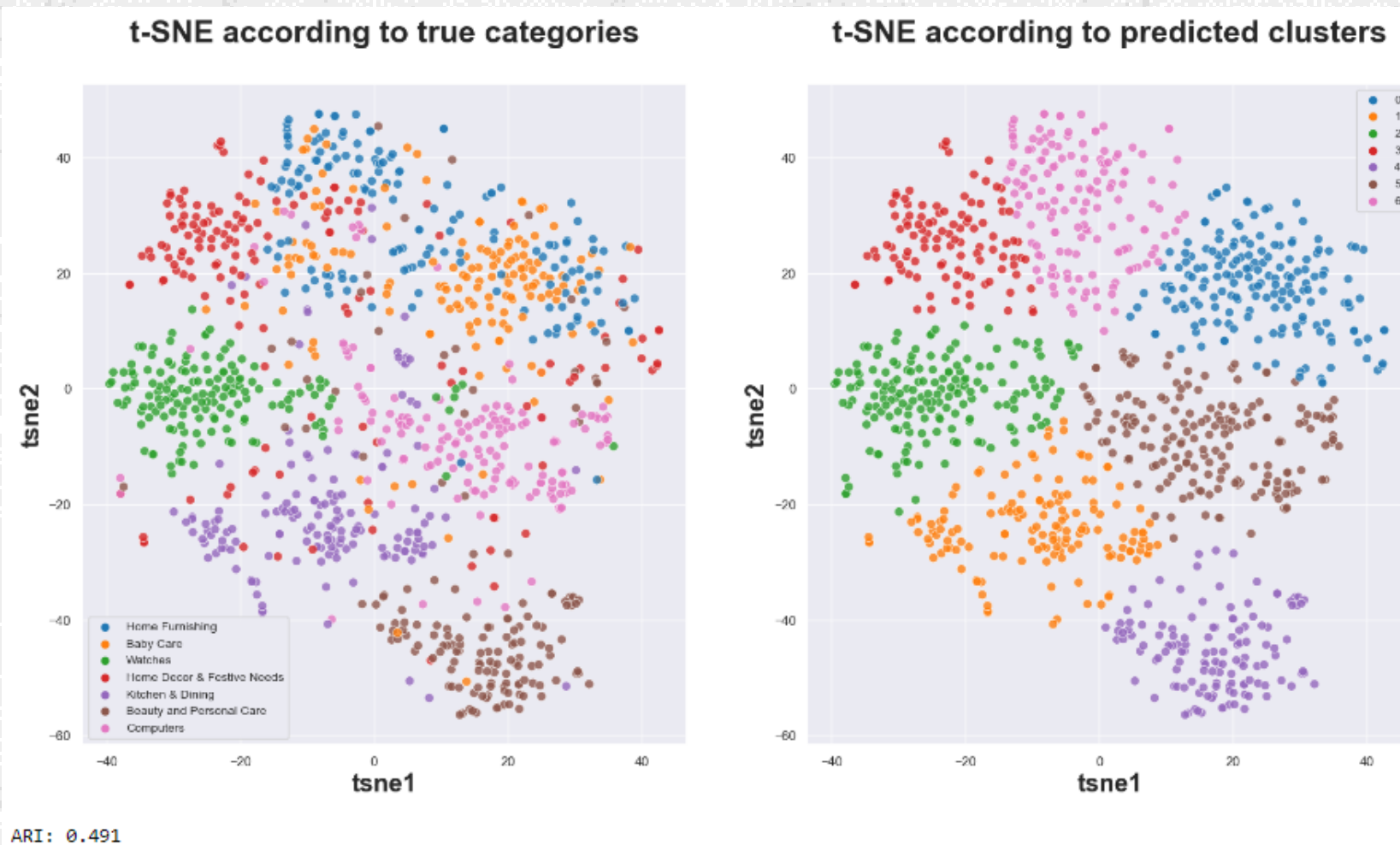
2 Analyse des images

Résultat obtenu par l'approche ResNet101



2 Analyse des images

Résultat obtenu par l'approche VGG16

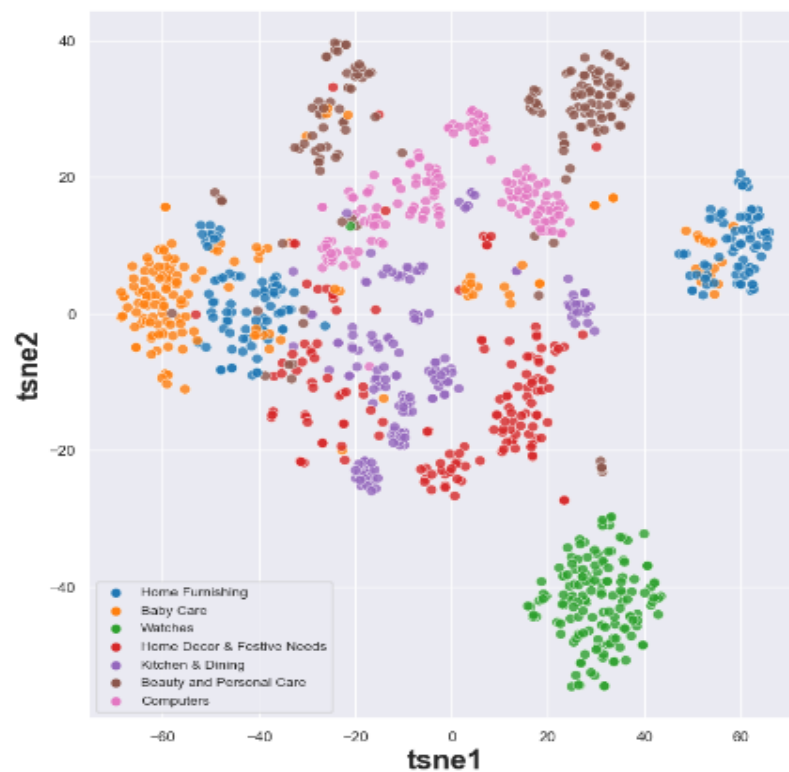


3 Analyse images + texte

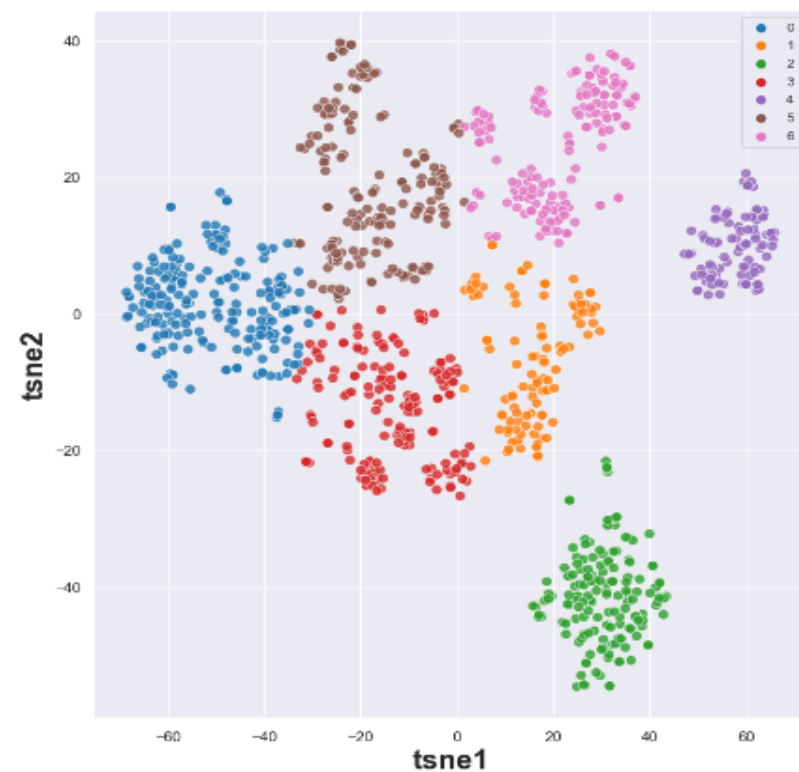
Résultat obtenu par l'approche VGG16 + USE

Dataset dimensions before PCA: (1050, 1024)
Dataset dimensions after PCA: (1050, 623)
Processing time: 6.0 seconds
ARI: 0.441

t-SNE according to true categories



t-SNE according to predicted clusters



Conclusion

Le score de similarité final obtenu à partir des images et des descriptions afin de permettre de regrouper les produits de mêmes catégories n'a pas été suffisamment satisfaisant pour prétendre à une réelle efficacité de cette approche de modélisation.