

# Classifiez automatiquement des biens de consommation

Parcours Data Scientist | projet 6

**Rim BAHROUN**

Avril 2023



# PROBLEMATIQUE & MISSION



Computers



Beauty care



Kitchen &  
dining



Home  
decor



Watches



Baby care



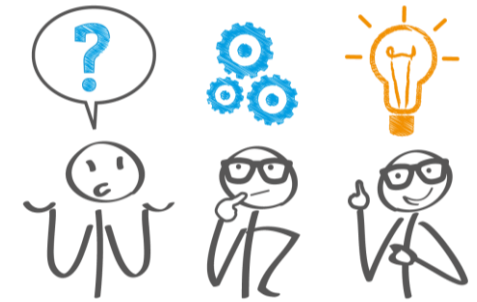
Home  
Furnishing

**Place de marché:** marketplace e-commerce

Des vendeurs proposent des articles à des acheteurs.

Un article: **une photo** et **une description**.

Attribution **manuelle** de la catégorie d'un article.



## Mission

1. **Etudier la faisabilité** d'un **moteur de classification automatique** des articles en catégories en se basant sur **une image et une description**.
2. Réaliser la **classification supervisée** des produits à partir des images.
3. Tester la collecte de produits à base de 'champagne' via **une API**.

## Objectifs

Faciliter l'expérience des utilisateurs

Fiabiliser l'attribution de la catégorie

Passage à l'échelle





## PLAN DE LA PRESENTATION

1

**Préparation des données**

2

**Faisabilité via données textuelles**

3

**Faisabilité via données visuelles**

4

**Classification via données visuelles**

5

**Collecte de produits via API**

6

**Conclusion**



# Préparation du jeu de données

- Données à disposition : 1 fichiers .csv

1050 produits et 15 variables.



```
<class 'pandas.core.frame.DataFrame'>
```

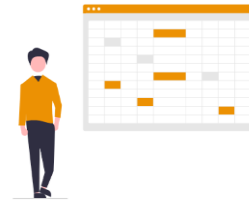
```
RangeIndex: 1050 entries, 0 to 1049
```

```
Data columns (total 15 columns):
```

#	Column	Non-Null Count	Dtype
0	uniq_id	1050 non-null	object
1	crawl_timestamp	1050 non-null	object
2	product_url	1050 non-null	object
3	product_name	1050 non-null	object
4	product_category_tree	1050 non-null	object
5	pid	1050 non-null	object
6	retail_price	1049 non-null	float64
7	discounted_price	1049 non-null	float64
8	image	1050 non-null	object
9	is_FK_Advantage_product	1050 non-null	bool
10	description	1050 non-null	object
11	product_rating	1050 non-null	object
12	overall_rating	1050 non-null	object
13	brand	712 non-null	object
14	product_specifications	1049 non-null	object

```
dtypes: bool(1), float64(2), object(12)
```

```
memory usage: 116.0+ KB
```



```
data_ = data.loc[:, ['uniq_id', 'product_name', 'product_category_tree', 'image', 'description']]
data_.head(2)
```

	uniq_id	product_name	product_category_tree	image	description
0	55b85ea15a1536d46b7190ad6fff8ce7	Elegance Polyester Multicolor Abstract Eyelet ...	["Home Furnishing >> Curtains & Accessories >>...	55b85ea15a1536d46b7190ad6fff8ce7.jpg	Key Features of Elegance Polyester Multicolor ...

```
uniq_id:
55b85ea15a1536d46b7190ad6fff8ce7
```

```
product_name:
Elegance Polyester Multicolor Abstract Eyelet Door Curtain
```

```
product_category_tree:
["Home Furnishing >> Curtains & Accessories >> Curtains >> Elegance Polyester Multicolor Abstract Eyelet Do..."]
```

```
image:
55b85ea15a1536d46b7190ad6fff8ce7.jpg
```

```
description:
```

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 interior. 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 anti-wrinkle and anti-shrinkage and has elegant appearance. 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



# Préparation du jeu de données

- Analyse des catégories

Arborescence complète d'un article.

```
product_category_tree:  
["Home Furnishing >> Curtains & Accessories >> Curtains >> Elegance Polyester Multicolor Abstract Eyelet Do..."]
```



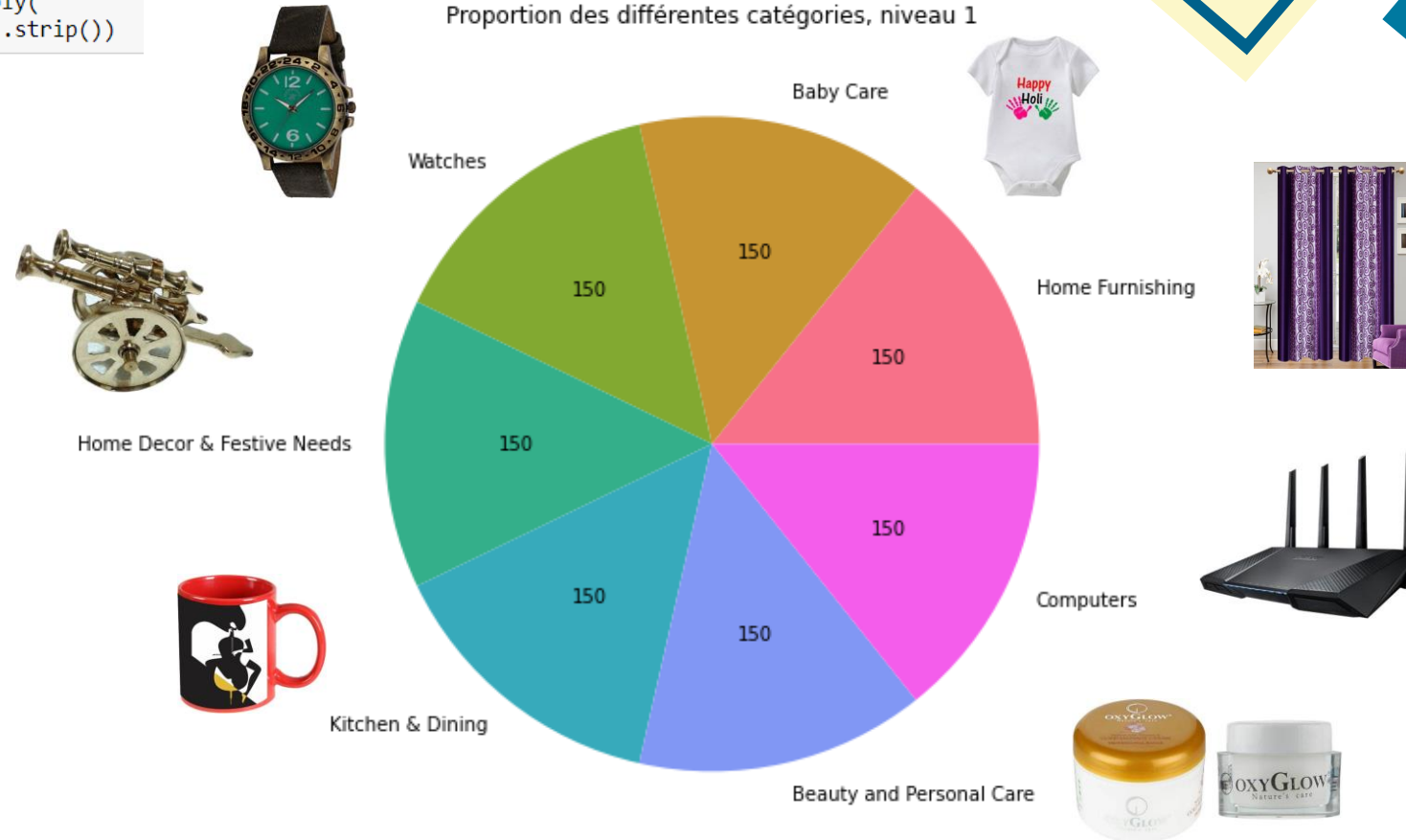
```
# 'categorie_l1' le premier niveau de catégorie du produit  
data['categorie_l1'] = data.product_category_tree.apply(  
    lambda x: x.strip('"').strip('"').split('>>')[0].strip())
```

```
data.categorie_l1.value_counts()
```

Home Furnishing	150
Baby Care	150
Watches	150
Home Decor & Festive Needs	150
Kitchen & Dining	150
Beauty and Personal Care	150
Computers	150

Name: categorie\_l1\_, dtype: int64

Proportion des différentes catégories, niveau 1







## PLAN DE LA PRESENTATION

1

**Préparation des données**

2

**Faisabilité via données textuelles**

3

**Faisabilité via données visuelles**

4

**Classification via données visuelles**

5

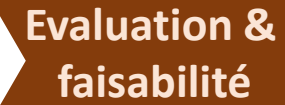
**Collecte de produits via API**

6

**Conclusion**



# Faisabilité via les données textuelles



## Lemmatiser

## Tf-idf

# USE

## ARI score



	uniq_id	product_name	description	name_description_	categorie_l1_
0	55b85ea15a1536d46b7190ad6fff8ce7	Elegance Polyester Multicolor Abstract Eyelet ...	Key Features of Elegance Polyester Multicolor ...	Elegance Polyester Multicolor Abstract Eyelet ...	Home Furnishing

## PRETRAITEMENT

**Nettoyage:** suppression de la ponctuation, mots de liaison, chiffres, mise en minuscules, ...

**Tokeniser:** découper un texte en entités plus petites appelées tokens

**Stemmer:** découper la fin des mots dans afin de ne conserver que la racine du mot.

**Lemmatiser:** consiste à sélectionner uniquement la forme canonique d'un mot

```
data.name_description_[0]
```

'Elegance Polyester Multicolor Abstract Eyelet Door Curtain 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 eye let style stitch with Metal Ring. It makes the room environment romantic and loving. This curtain is ant- wrinkle and anti shrink age and have elegant apparance. 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 m orning to welcome the bright sun rays you want to wish good morning to the whole world and when you draw them close in the even ing, you create the most special moments of joyous beauty given by the soothing prints. Bring home the elegant curtain that sof tly filters light in your room so that you get the right amount of sunlight., Specifications of Elegance Polyester Multicolor Ab stract Eyelet Door Curtain (213 cm in Height, Pack of 2) General Brand Elegance Designed For Door Type Eyelet Model Name Abstra ct 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'



```
preprocess_text_lem_stem(data.name_description_[0])
```

'eleg polyst multicolor abstract eyelet door curtain key featur eleg polyst multicolor abstract eyelet door curtain floral cu rtain polyst multicolor abstract eyelet door curtain height pack price curtain enhanc look interior curtain made high qualiti polyst fabric featur eyelet style stitch metal ring make room environ romant love curtain ant wrinkl anti shrinkag eleg appar home bright modernist appeal design surreal attent sure steal heart contemporari eyelet valanc curtain slide smoothli draw apar t first thing morn welcom bright sun ray want wish good morn whole world draw close even creat special moment joyou beauti give n sooth print bring home eleg curtain softli filter light room get right amount sunlight eleg polyst multicolor abstract eyele t door curtain height pack gener brand eleg design door type eyelet model name abstract polyst door curtain set model color mu lticolor dimens length box number content sale packag pack sale packag curtain bodi design materi polyst'



# EVALUATION & FAISABILITE

## Bag of word

Counter

Tf-idf

**Nettoyage:** suppression des mots qui apparaissent plus de 90% ou moins de 0,5% dans le corpus (**name\_description**)

CountVectorizer

Sparse matrix 1050 x 735



TSNE - 2D

1050 x 2

K-means

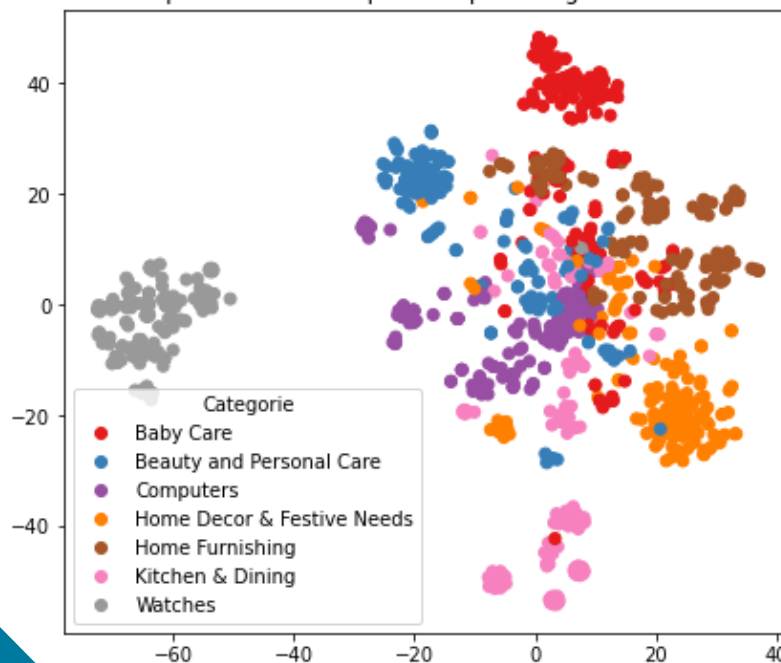
Visualisation  
2D

ARI score

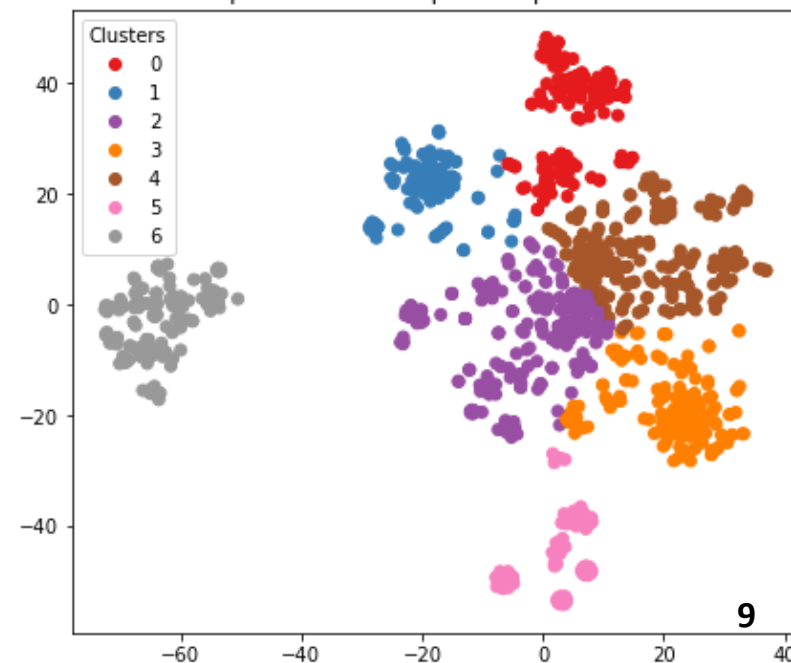
ARI-score 0,49

Temps 11s

Représentation des produits par catégories réelles



Représentation des produits par clusters



# EVALUATION & FAISABILITE

## Bag of word

Counter

Tf-idf

**Nettoyage:** suppression des mots qui apparaissent plus de 90% ou moins de 0,5% dans le corpus (**name\_description**)

TfidfVectorizer

Sparse matrix 1050 x 735



TSNE - 2D

K-means

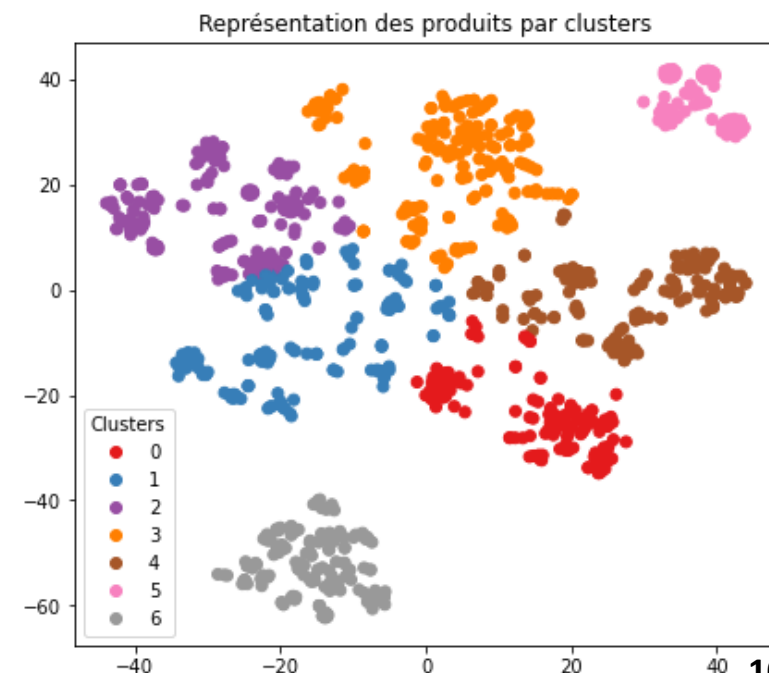
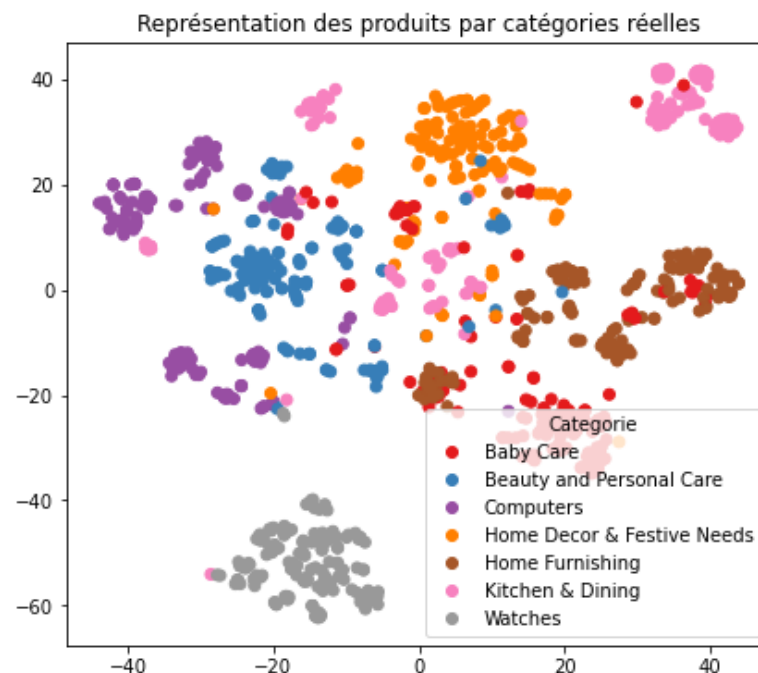
Visualisation  
2D

ARI score

1050 x 2

ARI-score 0,52

Temps 10s



Faisabilité via les données  
textuelles

# EVALUATION & FAISABILITE

Word/sentences  
embedding

Word2Vec

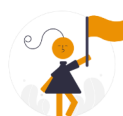
BERT

USE

name\_description

Word2Vec

Full matrix 1050 x 300



PCA(0,99)

1050 x 27

TSNE - 2D

1050 x 2

K-means

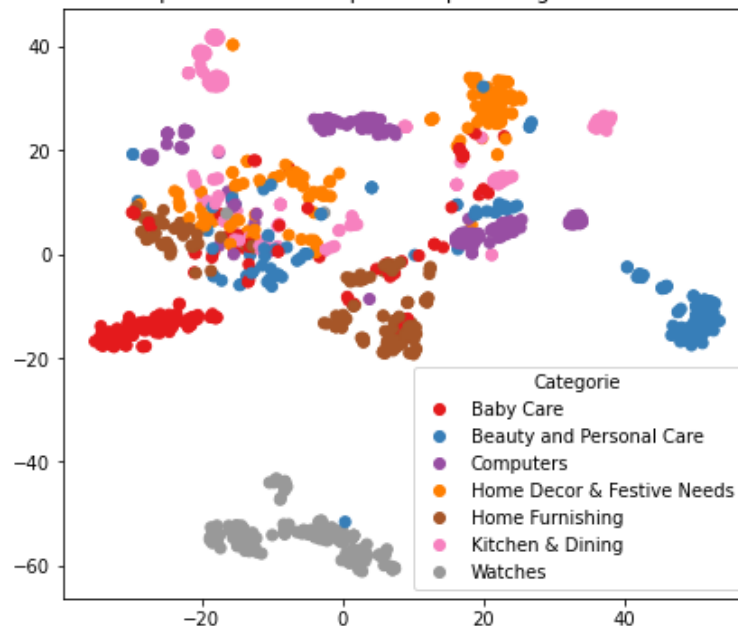
Visualisation  
2D

ARI score

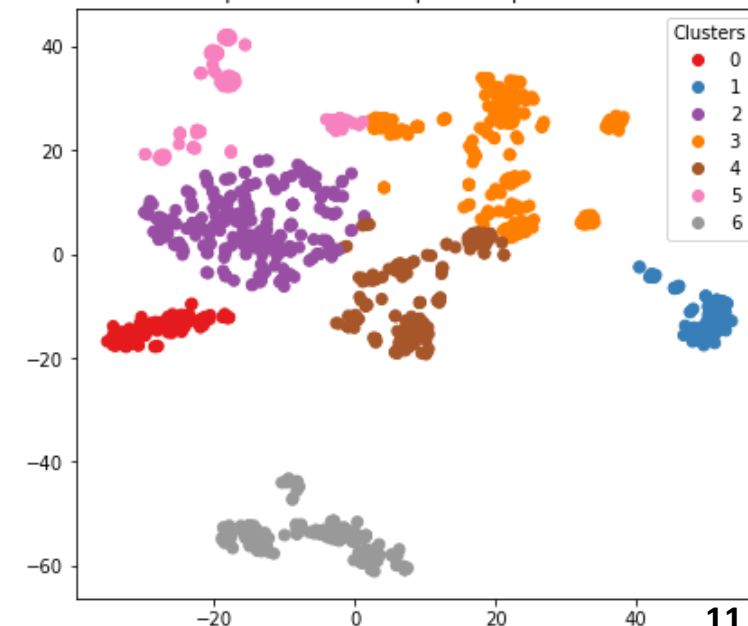
ARI-score 0,36

Temps 10s

Représentation des produits par catégories réelles



Représentation des produits par clusters



Faisabilité via les données  
textuelles

# EVALUATION & FAISABILITE

Word/sentences  
embedding

Word2Vec

BERT

USE

Product\_name

BERT

Full matrix 1050 x 768



PCA(0,99)

1050 x 387

TSNE - 2D

1050 x 2

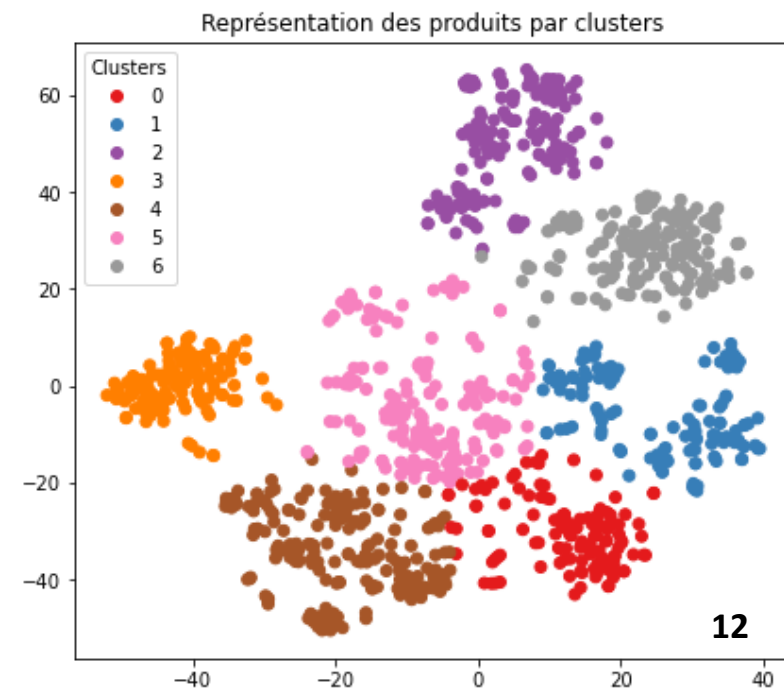
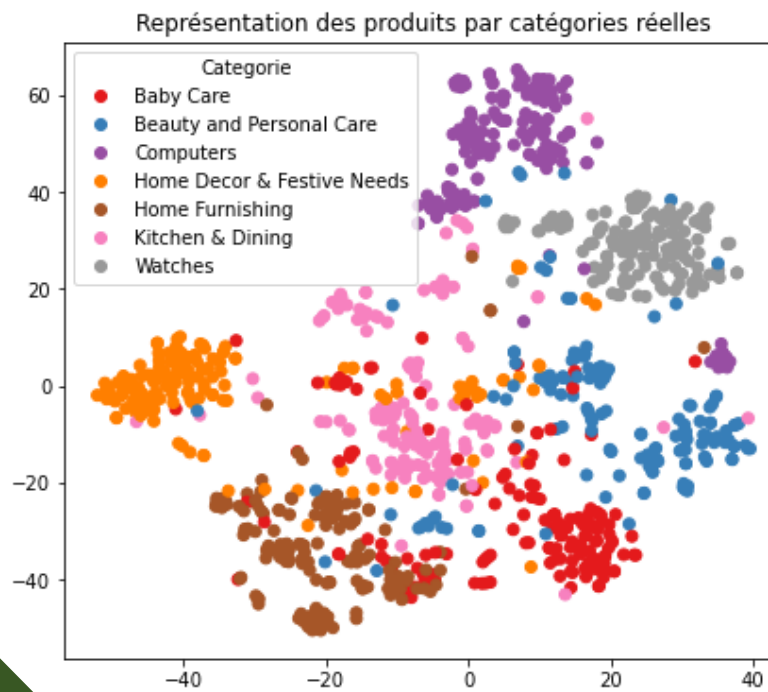
K-means

Visualisation  
2D

ARI score

ARI-score 0,63

Temps 10s



# EVALUATION & FAISABILITE

Word/sentences  
embedding

Word2Vec

BERT

USE

Product\_name

USE

Full matrix 1050 x 512



PCA(0,99)

1050 x 303

TSNE - 2D

1050 x 2

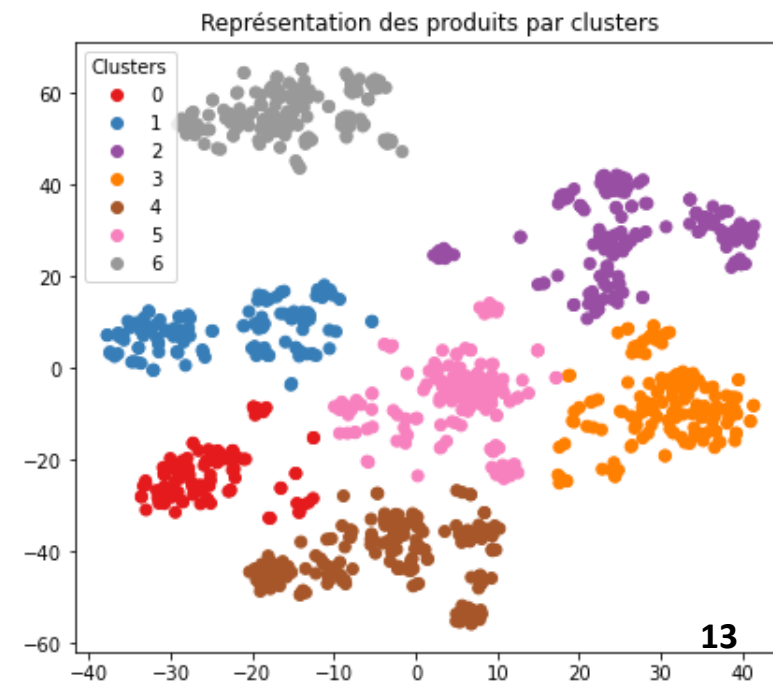
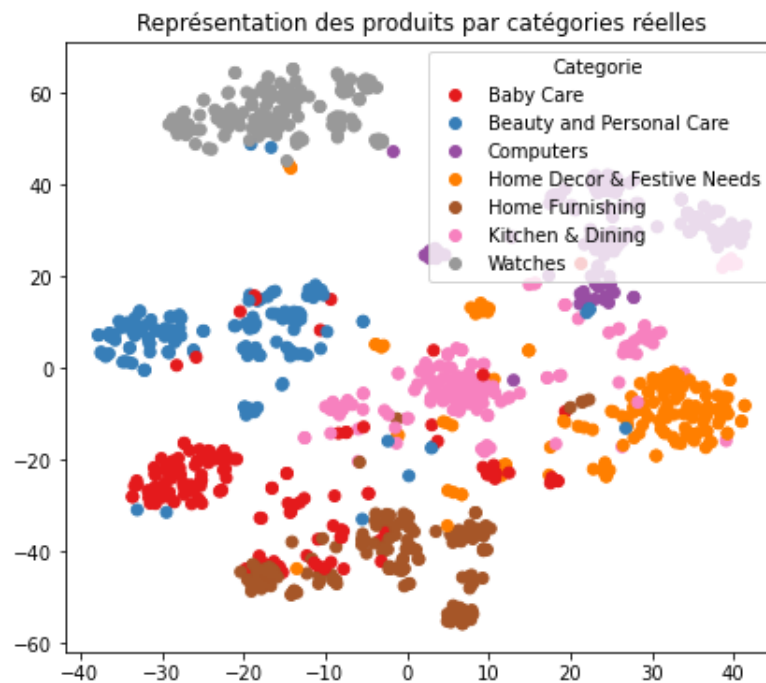
K-means

Visualisation  
2D

ARI score

ARI-score 0,69

Temps 10s



Faisabilité via les données  
textuelles



# EVALUATION & FAISABILITE



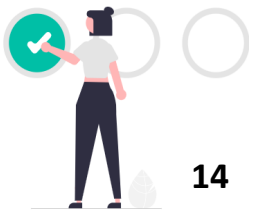
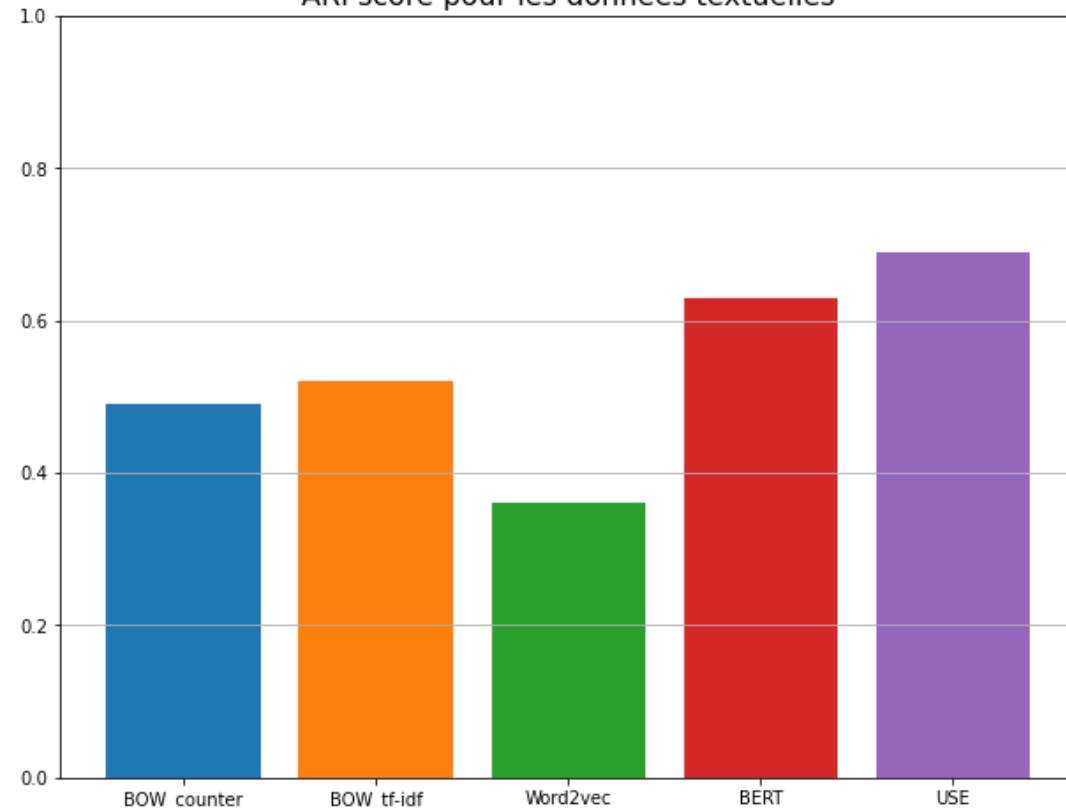
CONCLUSION

L'attribution des catégories des produits **peut être automatisée** à partir des données **textuelles**

Faisabilité via les données textuelles



ARI score pour les données textuelles



## PLAN DE LA PRESENTATION

1

**Préparation des données**

2

**Faisabilité via données textuelles**

3

**Faisabilité via données visuelles**

4

**Classification via données visuelles**

5

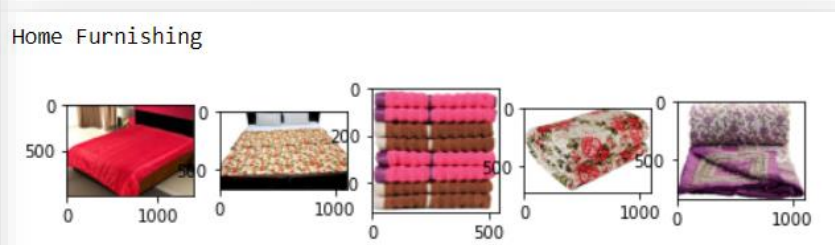
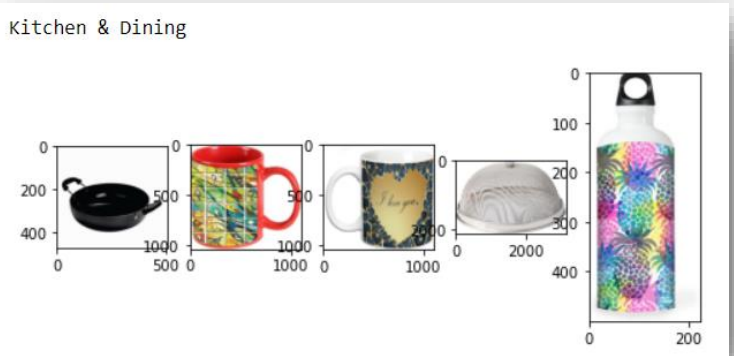
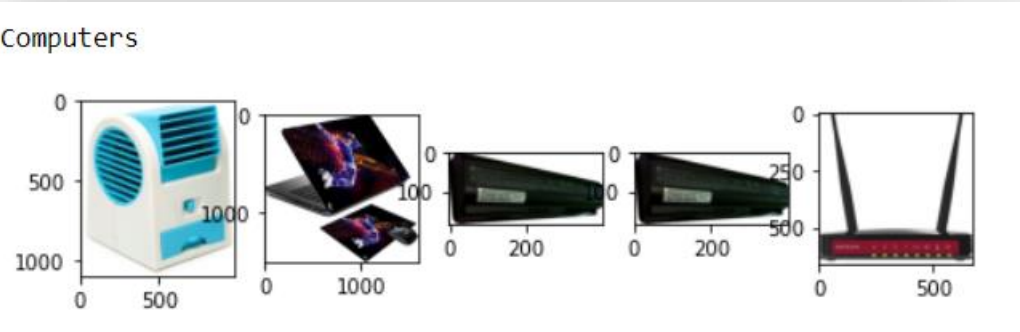
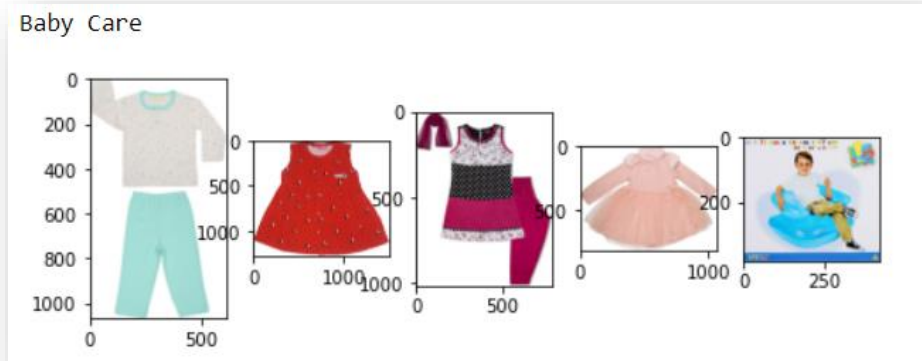
**Collecte de produits via API**

6

**Conclusion**



Exemples d'images



	uniq_id	image_path	label_name	label
0	55b85ea15a1536d46b7190ad6fff8ce7	./data/Images/55b85ea15a1536d46b7190ad6fff8ce7...	Home Furnishing	4

# Démarche

Faisabilité via les données visuelles



Prétraitement

Features  
extraction

Réduction de  
dimension

Classification &  
visualisation

Evaluation &  
faisabilité

Redimensionnement

Egalisation

Passage en gris

Filtrage du bruit

Bag of images

SIFT



Transfer Learning

CNN-VGG16

TensorFlow

K Keras

PCA

TSNE - 2D

K-means

visualisation - 2D

ARI score



	uniq_id	image_path	label_name	label
0	55b85ea15a1536d46b7190ad6fff8ce7	./data/Images/55b85ea15a1536d46b7190ad6fff8ce7...	Home Furnishing	4

# PRETRAITEMENT & DESCRIPTEURS SIFT



Redimensionnement

Egalisation

Passage en gris

Filtrage du bruit

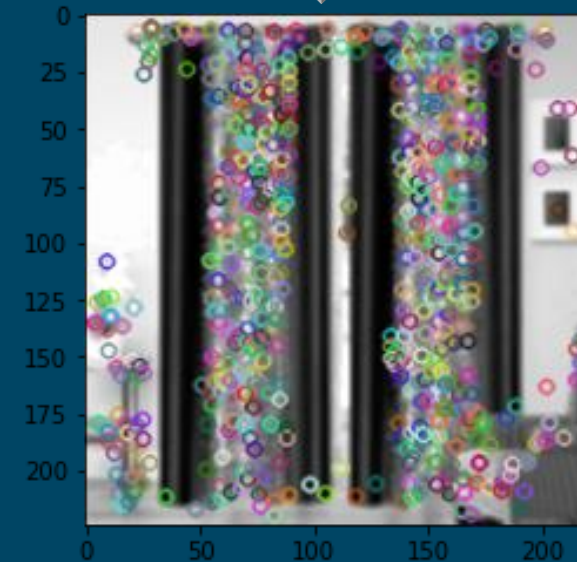
Nombre de descripteurs : (368797, 128)

temps de traitement SIFT descriptor : 44.20 secondes

Faisabilité via les données  
visuelles



Descripteurs : (619, 128)





# EVALUATION & FAISABILITE

Bag of images

SIFT

Redimensionnement,  
égalisation, passage en gris

SIFT

1050 x 607



PCA (0,97)

TSNE - 2D

K-means

Visualisation  
2D

ARI  
score

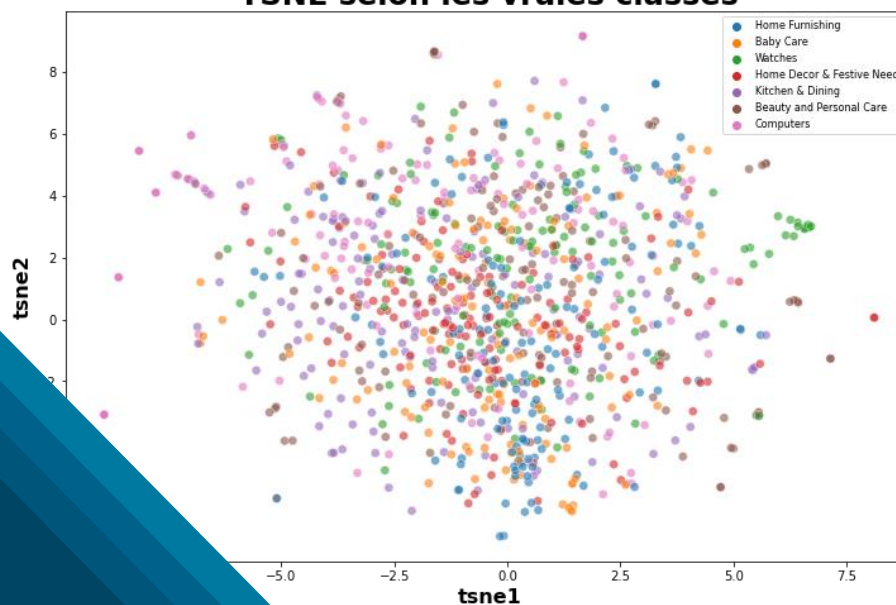
1050 x 454

1050 x 2

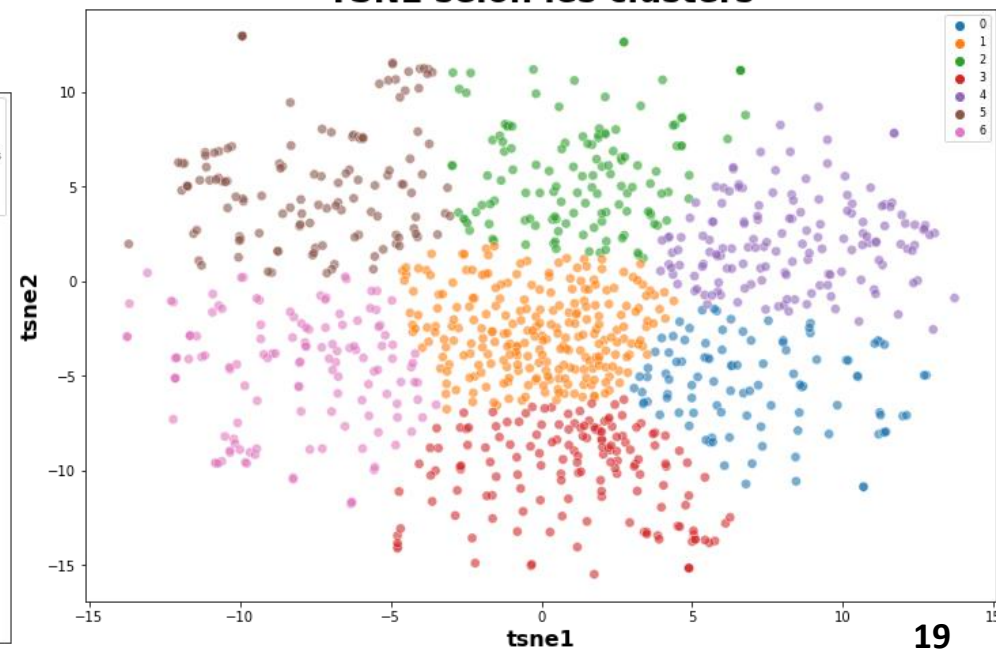


ARI-score 0,04

TSNE selon les vraies classes



TSNE selon les clusters



Faisabilité via les données  
visuelles

# EVALUATION & FAISABILITE

Transfer Learning

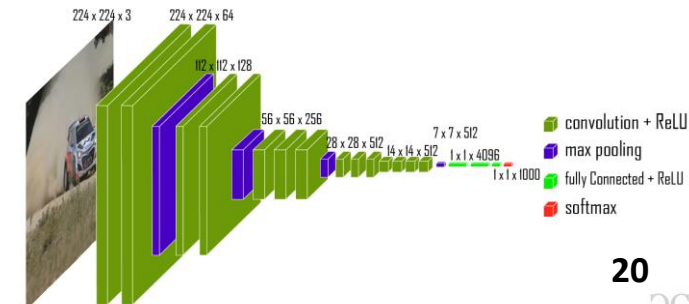
CNN – VGG16

Faisabilité via les données visuelles

 TensorFlow  
 Keras

Layer (type)	Output Shape	Param
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312

fc3 (dense) — (None, 1000)



# EVALUATION & FAISABILITE

Transfer Learning

CNN – VGG16

 TensorFlow

Preprocess\_input

CNN-VGG16



Faisabilité via les données visuelles

1050 x 4096  Keras



PCA (0,99)

1050 x 850

TSNE - 2D

1050 x 2

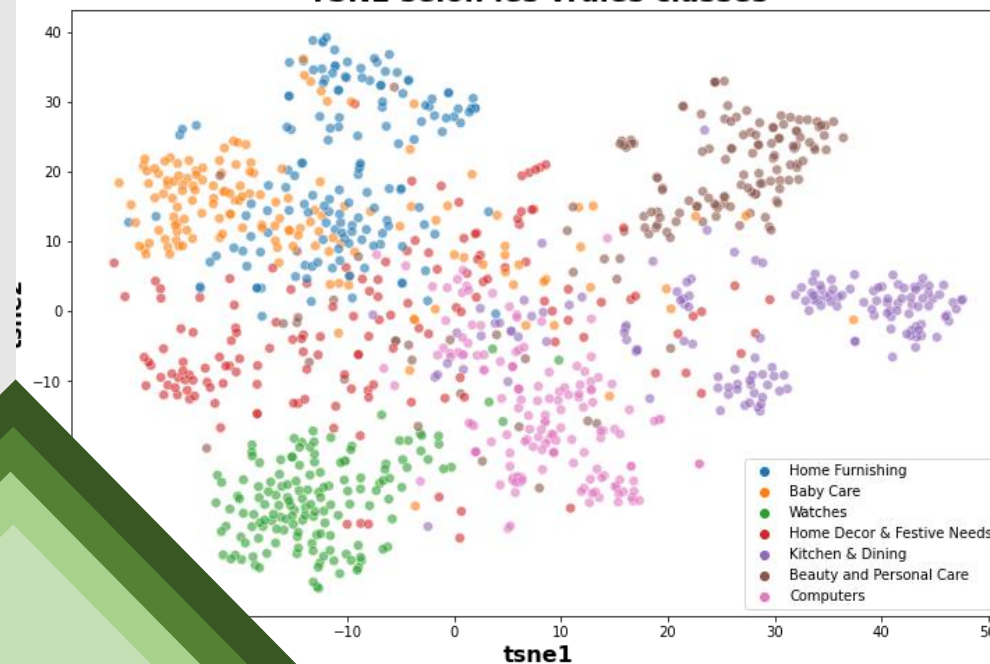
K-means

Visualisation 2D

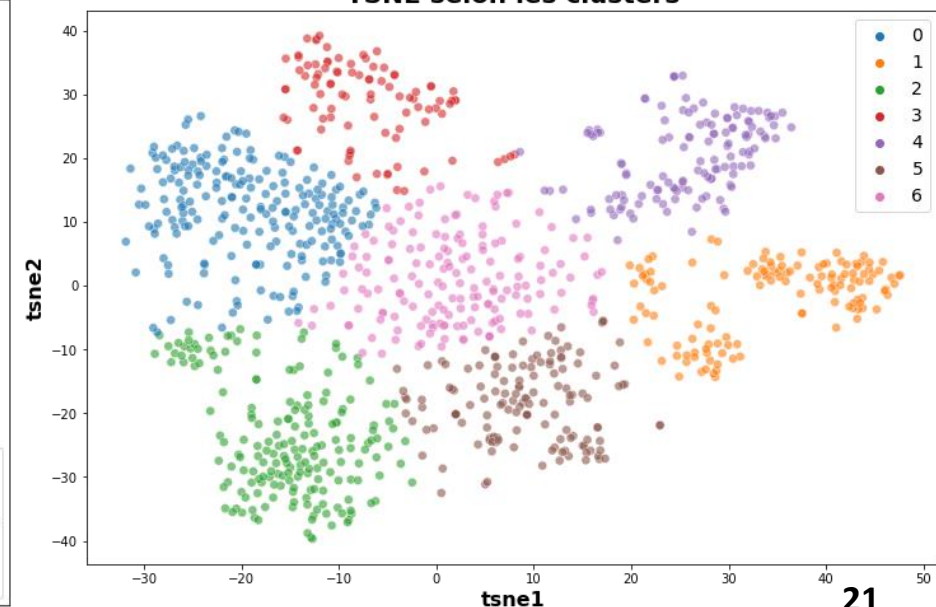
ARI score

ARI-score 0,46

TSNE selon les vraies classes



TSNE selon les clusters



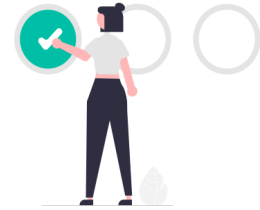
# EVALUATION & FAISABILITE

Faisabilité via les données  
visuelles



CONCLUSION

L'attribution des catégories des produits **peut être automatisée** à partir  
des données **visuelles**





# EVALUATION & FAISABILITE

USE

ARI-score 0,69

+

CNN – VGG16

ARI-score 0,46

CNN-VGG16

USE

1050 x 4608

PCA (0,99)

TSNE - 2D

K-means

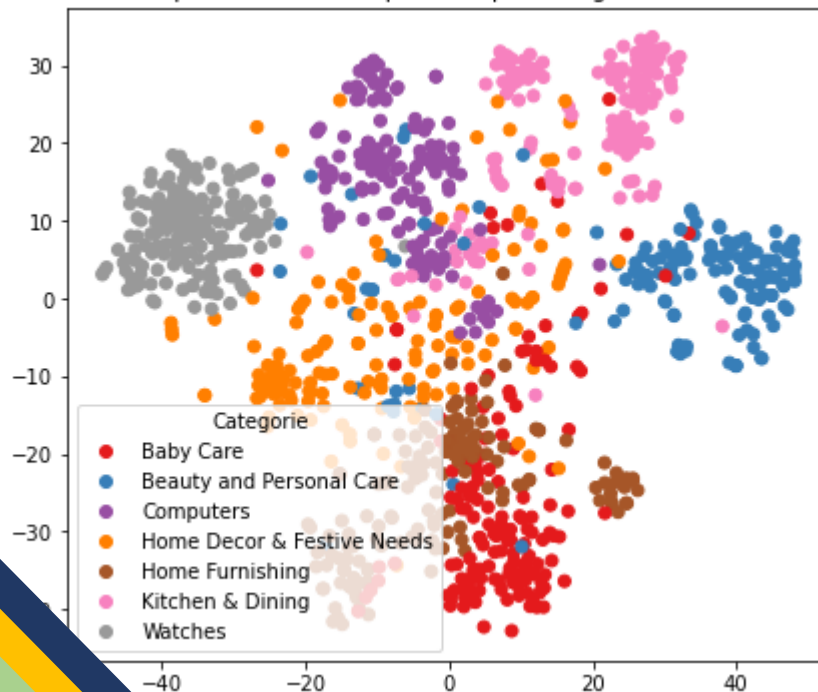
Visualisation  
2D

ARI  
score

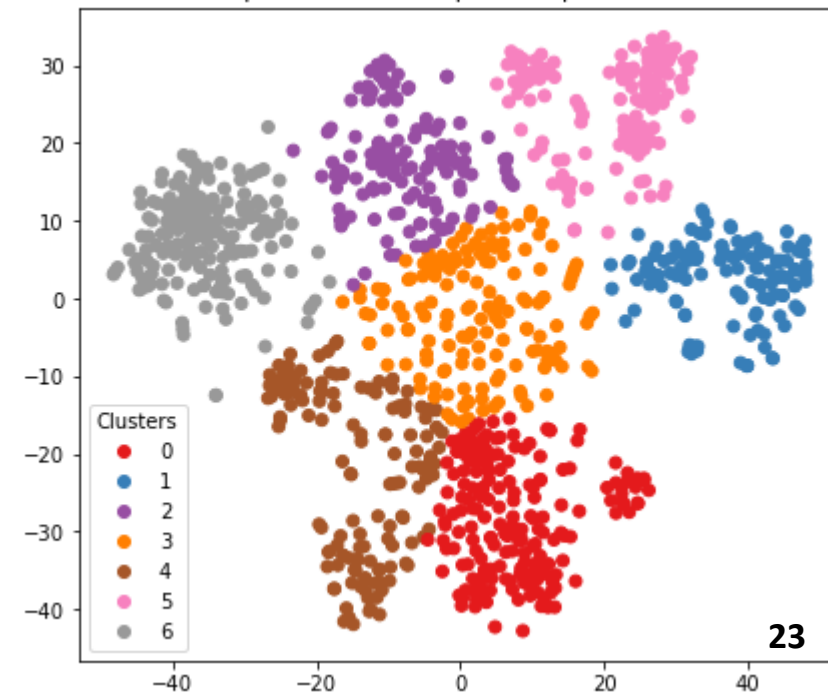
1050 x 869

ARI-score 0,54

Représentation des produits par catégories réelles



Représentation des produits par clusters



Faisabilité via les données  
visuelles et textuelles





## PLAN DE LA PRESENTATION

1

**Préparation des données**

2

**Faisabilité via données textuelles**

3

**Faisabilité via données visuelles**

4

**Classification via données visuelles**

5

**Collecte de produits via API**

6

**Conclusion**



# Modèle: VGG16 transfert Learning

Classification  
supervisée via les  
donnée visuelles

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312
Fc3 (dense)	(None, 1000)	



Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 512)	0
dense (Dense)	(None, 256)	131328
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 7)	1799



## Quatre approches pour la préparation d'images :

### Approche 1

Une approche simple par préparation initiale de l'ensemble des images avant classification supervisée.

### Approche 2

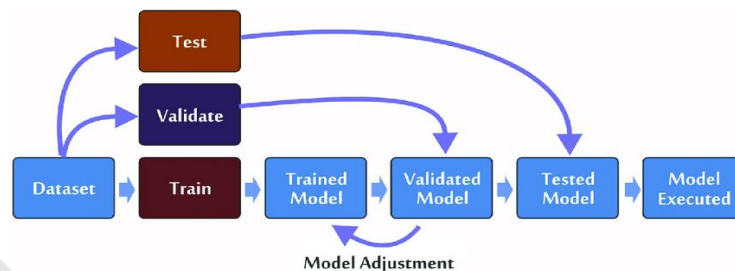
Une approche par **data generator**, permettant facilement la **data augmentation**. Les images sont directement récupérées à la volée dans le répertoire des images.

### Approche 3

**Une approche récente** proposée par **DataSet**, sans data augmentation.

### Approche 4

**Une approche récente** par **DataSet**, **avec data augmentation** intégrée au modèle : layer en début de modèle.



	uniq_id	image_path	label_name	label
0	55b85ea15a1536d46b7190ad6fff8ce7	./data/Images/55b85ea15a1536d46b7190ad6fff8ce7...	Home Furnishing	4

# CLASSIFICATION SUPERVISEE

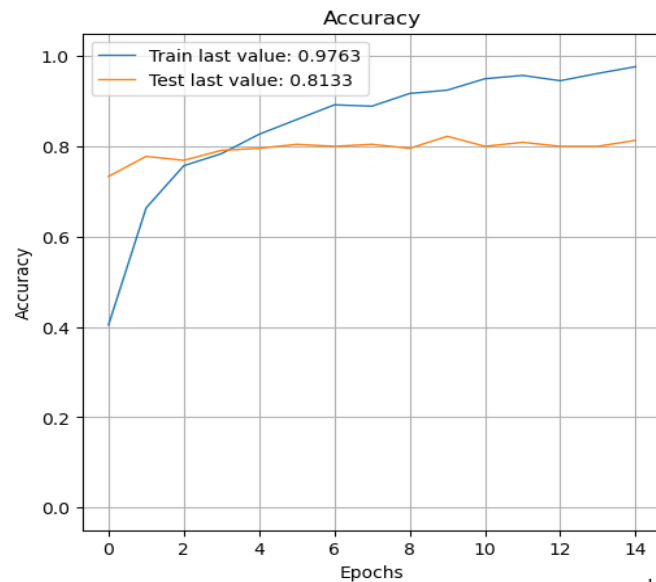
Approche 1

Approche 2

Approche 3

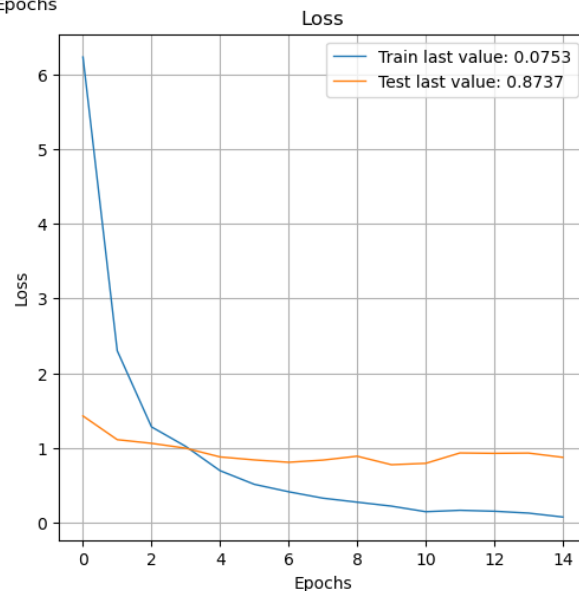
Approche 4

Une approche simple par préparation initiale de l'ensemble des images avant classification supervisée



Accuracy:

Training: 1  
Validation: 0,81  
Test: 0,79



Baby Care	26	1	0	0	6	0	0
Beauty and Personal Care	0	29	3	1	0	1	0
Computers	0	2	25	2	0	1	0
Home Decor & Festive Needs	2	0	0	23	2	3	2
Home Furnishing	5	1	0	0	25	0	0
Kitchen & Dining	0	2	0	2	1	27	0
Watches	0	0	3	0	0	0	30
	0	1	2	3	4	5	6

# CLASSIFICATION SUPERVISEE

Approche 1

Approche 2

Approche 3

Approche 4

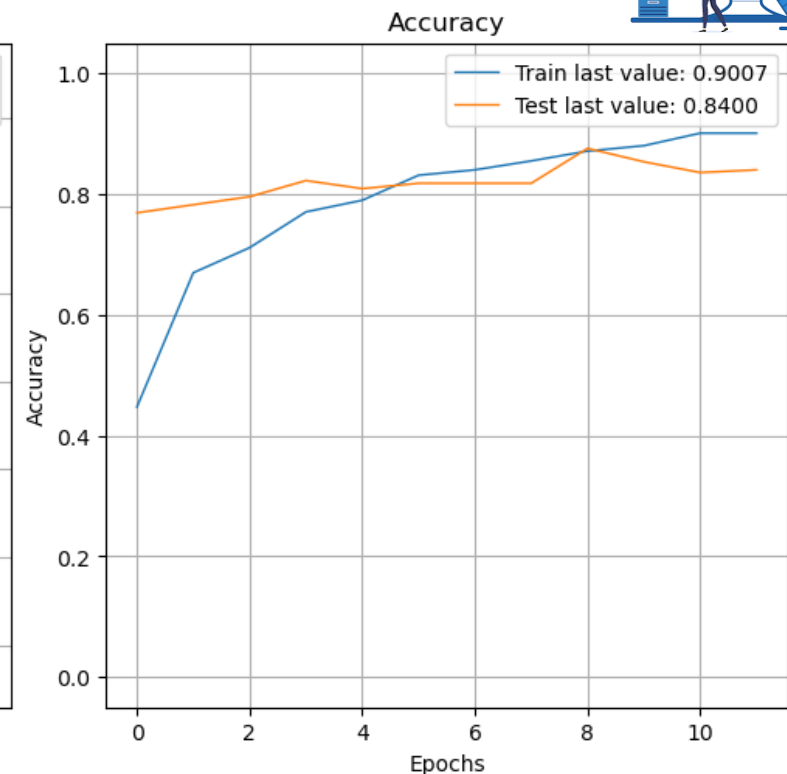
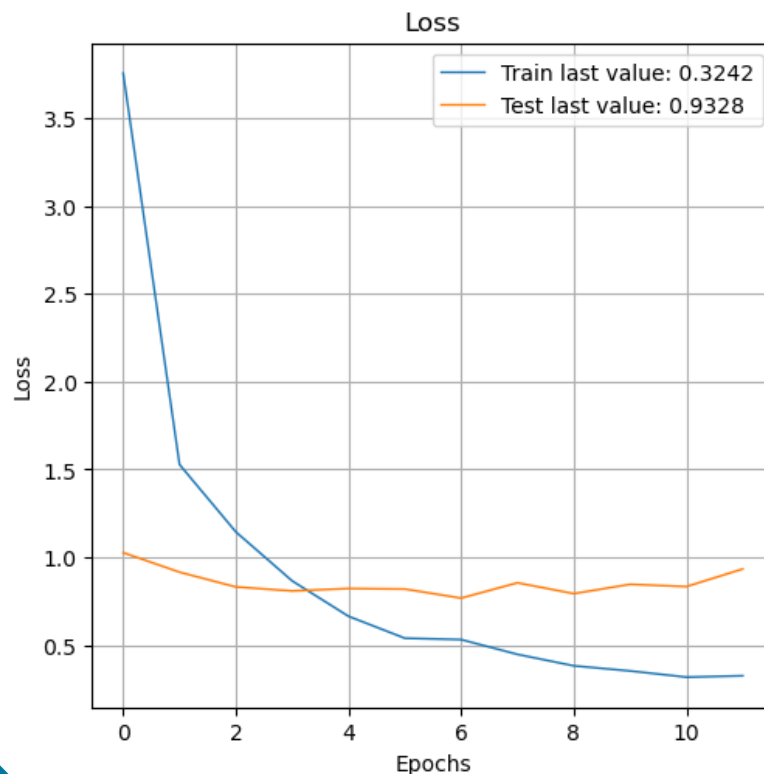
Une approche par **data generator**, permettant facilement la **data augmentation** (rotation, translation, retournement). Les images sont directement récupérées à la volée dans le répertoire des images

## Accuracy:



**Training: 0,97**  
**Validation: 0,83**  
**Test: 0,78**

Réduction de l'overfitting avec la data augmentation.





# CLASSIFICATION SUPERVISEE

Approche 1

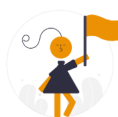
Approche 2

Approche 3

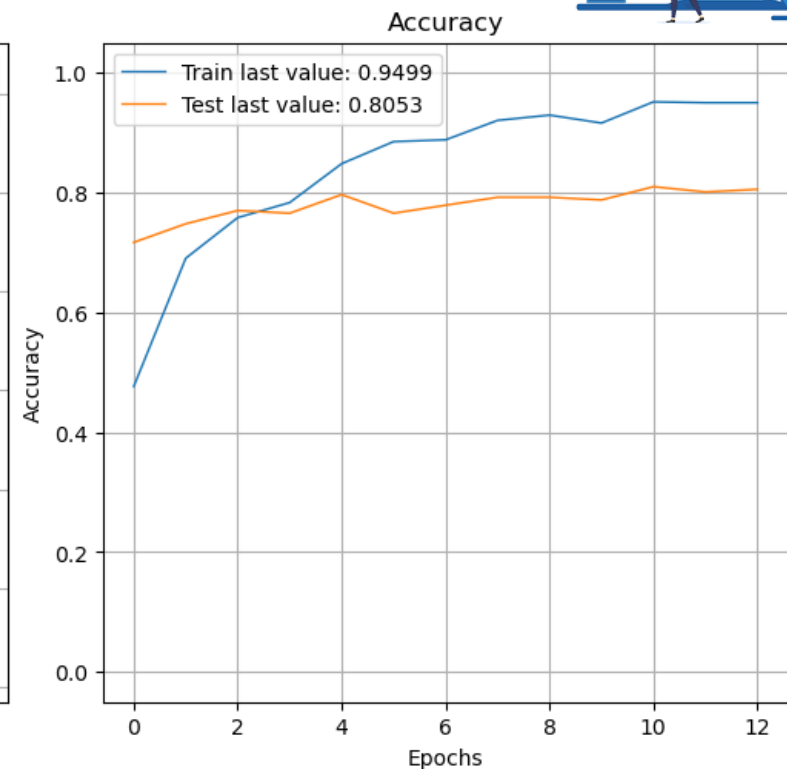
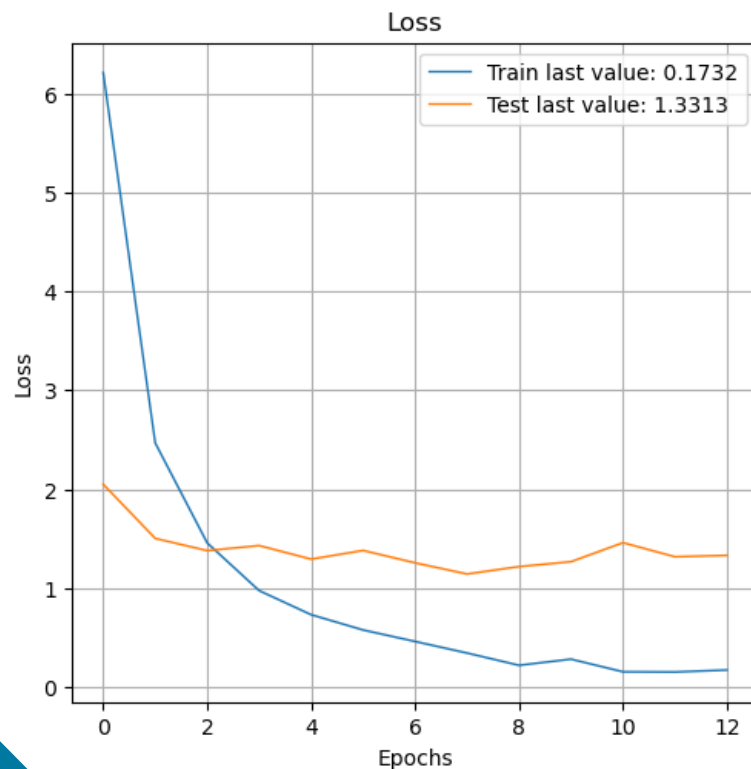
Approche 4

Une **approche récente** proposée par  
Tensorflow.org par **DataSet**, sans **data  
augmentation**

## Accuracy:



**Training: 0,99**  
**Validation: 0,80**  
**Test: 0,80**



# CLASSIFICATION SUPERVISEE

Approche 1

Approche 2

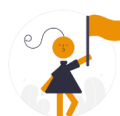
Approche 3

Approche 4

Une approche récente par **DataSet**, avec **data augmentation** intégrée au modèle (rotation, translation, zoom): layer en début de modèle

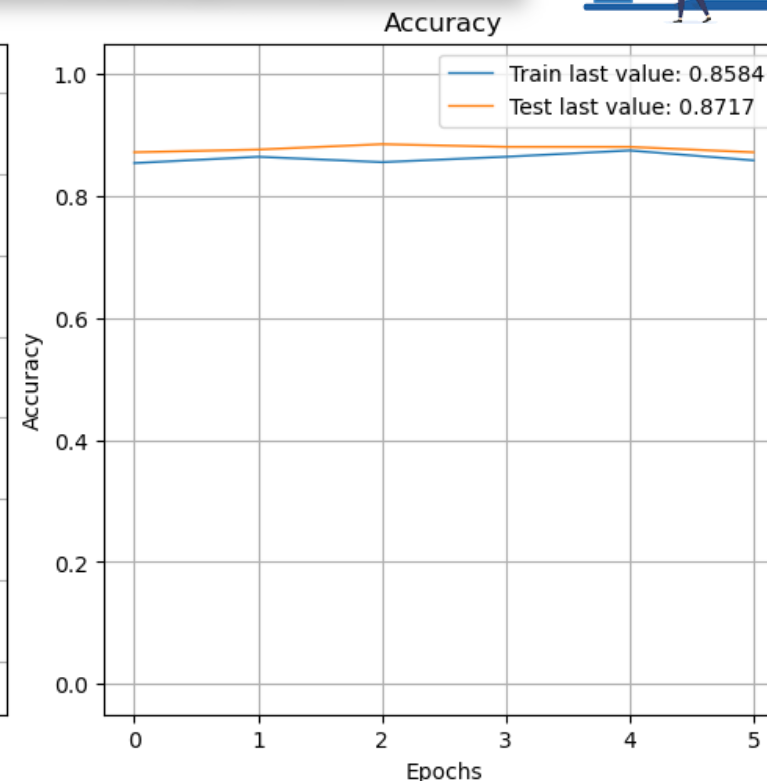
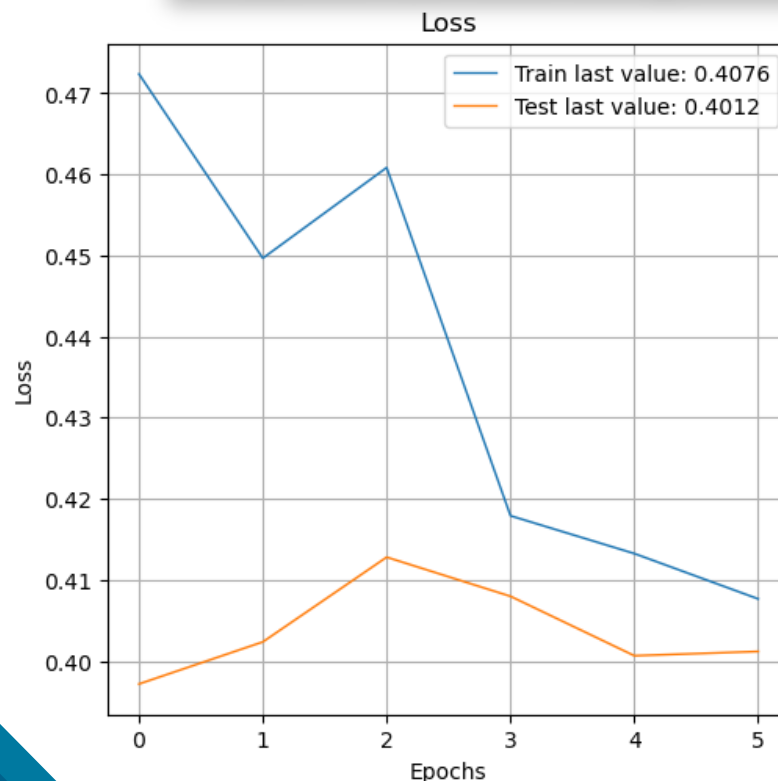
Classification supervisée via les données visuelles

## Accuracy:



**Training: 0,92**  
**Validation: 0,87**  
**Test: 0,86**

Réduction de l'overfitting avec la data augmentation.

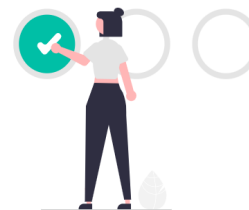


# CLASSIFICATION SUPERVISEE

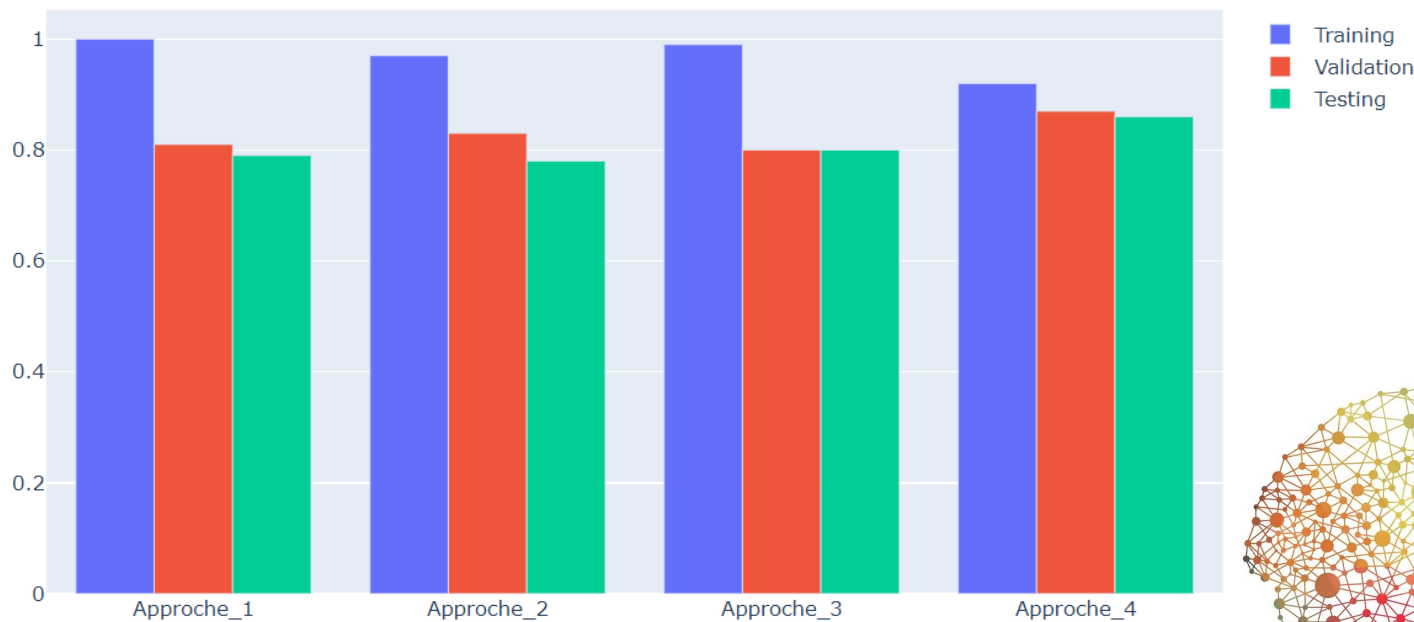


CONCLUSION

Classification supervisée via  
les donnée visuelles



La **classification automatique** des produits en catégories à partir des données **visuelles** est précise à **86%**





## PLAN DE LA PRESENTATION

1

**Préparation des données**

2

**Faisabilité via données textuelles**

3

**Faisabilité via données visuelles**

4

**Classification via données visuelles**

5

**Collecte de produits via API**

6

**Conclusion**



Pourrais-tu tester la collecte de produits à base de “champagne” via l’[API](#)? Je souhaiterais que tu puisses nous fournir une extraction des 10 premiers produits dans un fichier “.csv”, contenant pour chaque produit les données suivantes : **foodId**, **label**, **category**, **foodContentsLabel**, **image**.

# Collecte de produits via API



```
import requests
import pandas as pd

url = "https://edamam-food-and-grocery-database.p.rapidapi.com/parser"

querystring = {"ingr": "champagne"}

headers = {
    "X-RapidAPI-Key": "bd8c2f188cmshf5d0f1a1c8718c6p1c7f46jsn370366720478",
    "X-RapidAPI-Host": "edamam-food-and-grocery-database.p.rapidapi.com"
}

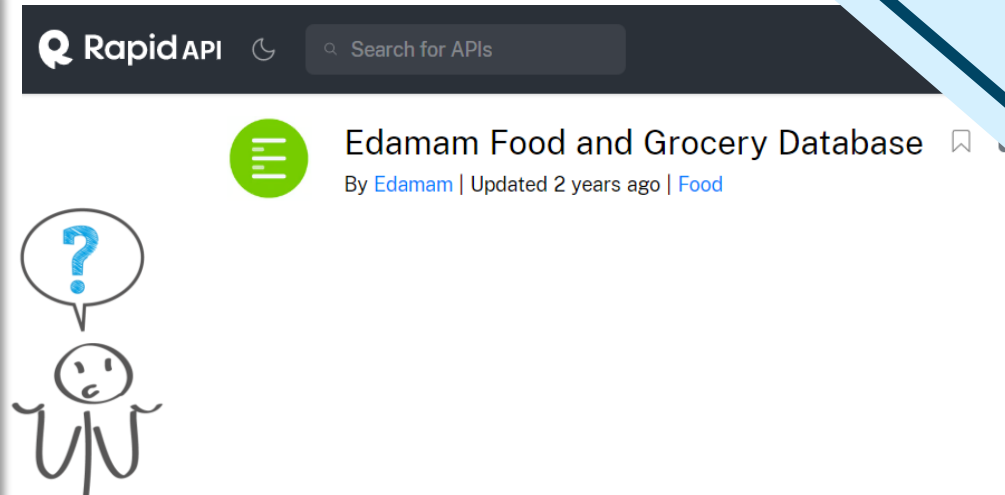
# Appel à l'API
response = requests.request("GET", url, headers=headers, params=querystring)

# Extraction des 10 premiers produits et création d'un dataframe
products = response.json()["hints"][:10]

# Extract required data from each product
data = []
for product in products:
    food = product['food']
    data.append({
        'foodId': food['foodId'],
        'label': food['label'],
        'category': food['category'],
        'foodContentsLabel': food.get('foodContentsLabel', ''),
        'image': food.get('image', '')
    })

# Create a pandas dataframe from the extracted data
df = pd.DataFrame(data, columns=['foodId', 'label', 'category', 'foodContentsLabel', 'image'])

# Save the dataframe to a CSV file
df.to_csv('products.csv', index=False)
```



	foodId	label	category	foodContentsLabel	image
0	food_a656mk2a5dmqb2adiamu6beihduu	Champagne	Generic foods	NaN	https://www.edamam.com/food-img/a71/a718cf3c52...
1	food_b753ithamdb8psbt0w2k9aquo06c	Champagne Vinaigrette, Champagne	Packaged foods	OLIVE OIL; BALSAMIC VINEGAR; CHAMPAGNE VINEGAR...	NaN
2	food_b3dyababjo54xobm6r8jzbghjqe	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...	NaN
4	food_an4jueaucpus2a3u1ni8auhe7q9	Champagne Vinaigrette, Champagne	Packaged foods	WATER; CANOLA AND SOYBEAN OIL; WHITE WINE (CON...	NaN
5	food_bmu5dmkazwvupaa5prh1daa8jxs0	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,...	NaN
7	food_byap67hab6evc3a0f9w1oag3s0qf	Champagne Sorbet	Generic meals	Sugar; Lemon juice; brandy; Champagne; Peach	NaN
8	food_am5egz6aq3fpjla8xpkdbc2asis	Champagne Truffles	Generic meals	butter; cocoa; sweetened condensed milk; vanil...	NaN
9	food_bcz8rhiajk1fuva0vkfmeakbouc0	Champagne Vinaigrette	Generic meals	champagne vinegar; olive oil; Dijon mustard; s...	NaN



## PLAN DE LA PRESENTATION

1

**Préparation des données**

2

**Faisabilité via données textuelles**

3

**Faisabilité via données visuelles**

4

**Classification via données visuelles**

5

**Collecte de produits via API**

6

**Conclusion**



# CONCLUSION



Computers



Beauty care



Kitchen & dining



Home decor



Watches

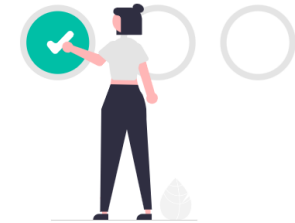


Baby care



Home Furnishing

- Prétraiter des données **image & texte**
- **Représenter graphiquement** des données à grandes dimensions
- Mettre en œuvre des techniques de **réduction de dimension**
- Collecter des données en recensant une **API**
- Elaboration un modèle **d'apprentissage profond**
- **Évaluer** la performance des modèles d'apprentissage profond
- Utiliser des techniques **d'augmentation des données**





**Merci de votre attention**

*Rim BAHROUN*