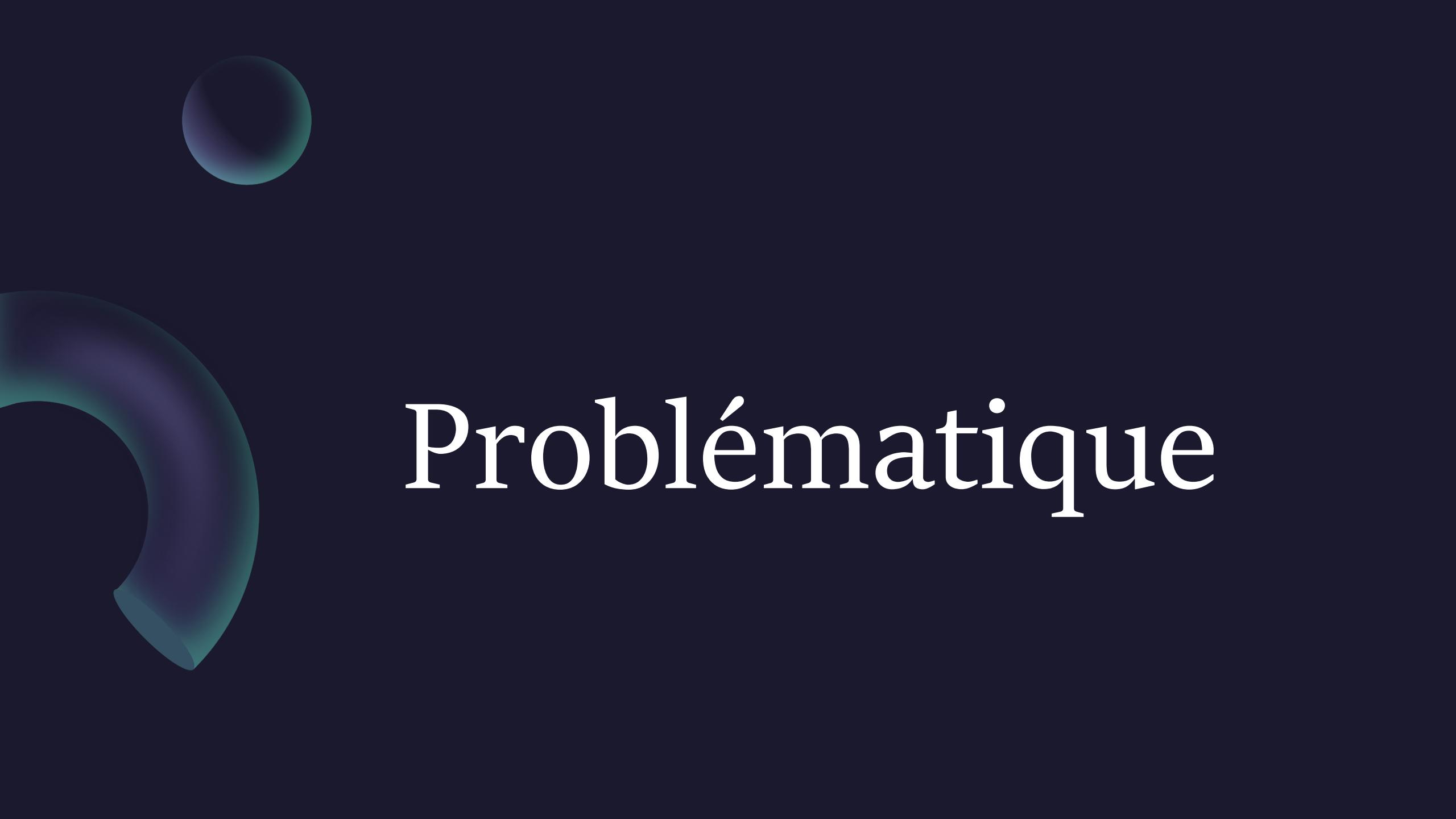


Marketplace

Moteur de recommandation



The background features two large, semi-transparent circles. One circle is positioned in the lower-left foreground, showing a gradient from dark purple at the top to teal at the bottom. The other circle is located in the upper-left background, showing a similar gradient from dark purple to teal.

Problématique

LE BESOIN : PRÉDIRE
AUTOMATIQUEMENT
LA CATÉGORIE D'UN
ARTICLE

Un gain de temps

Un gain d'argent

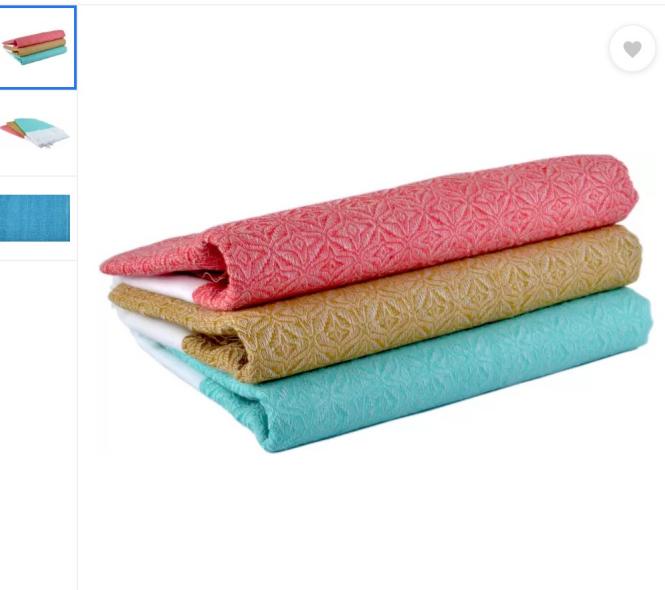
Une meilleure
expérience utilisateur

LA DATA : 1050 biens de consommation

Flipkart
Explore Plus

Search for products, brands and more

Login Become a Seller More Cart



Highlights

- Towel Type: Bath
- Size: 60 inch x 30 inch
- 3 Towels
- GSM: 500

Services

- 7 Days Return Policy
- Cash on Delivery available

Seller

Sathiyas 3.6 ★

Super Coin

For every ₹100 Spent,
you earn ⚡ 2 SuperCoins

Max 50 coins per order

Description

Comfort towel for drying the hair and body after a bath or shower. Cotton is sustainable,durable and absorbent fabric. So it nourish the skin.

Specifications

Sales Package	3 Bath Towel
Model Number	asvtwl322
Model Name	Sathiyas cotton bath towel

ADD TO CART **BUY NOW**

NOS CATEGORIES D'ARTICLE (150 x 7)



Accessoires
pour bébés



Meubles



Montres



Décoration
d'intérieur



Objets de
cuisine

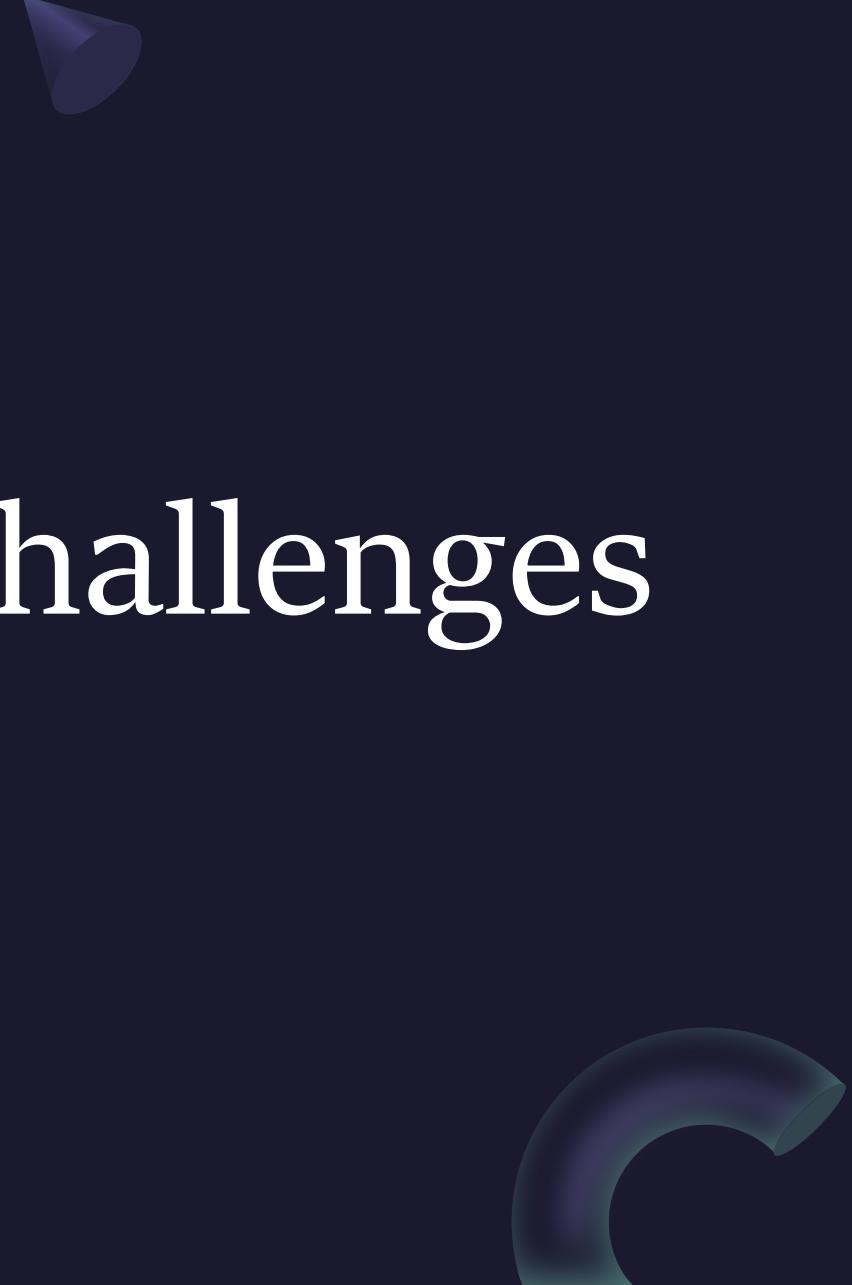


Produits de
beauté



Matériel
informatique

Challenges



Peu de data
par catégorie

Choix d'une
classification
non supervisée



Pistes de modélisation

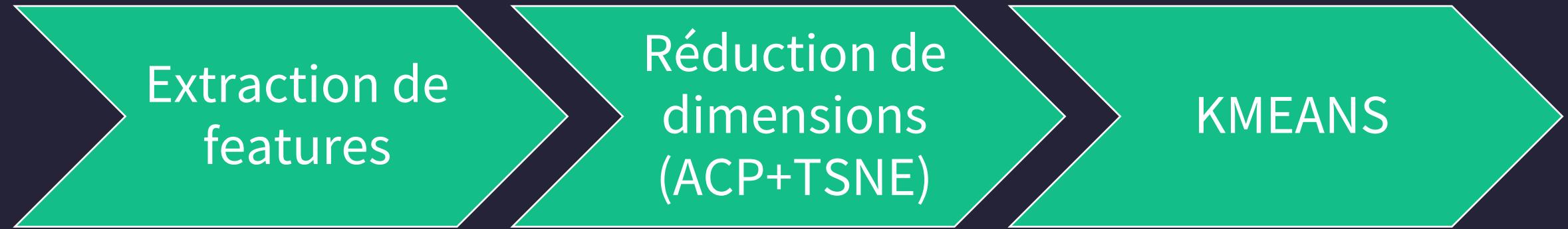
APPROCHES DE MODELISATION



APPROCHE VISUELLE

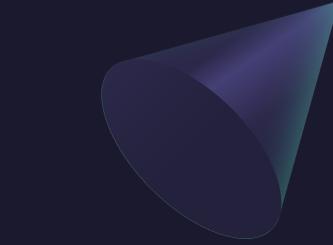


APPROCHE TEXTUELLE

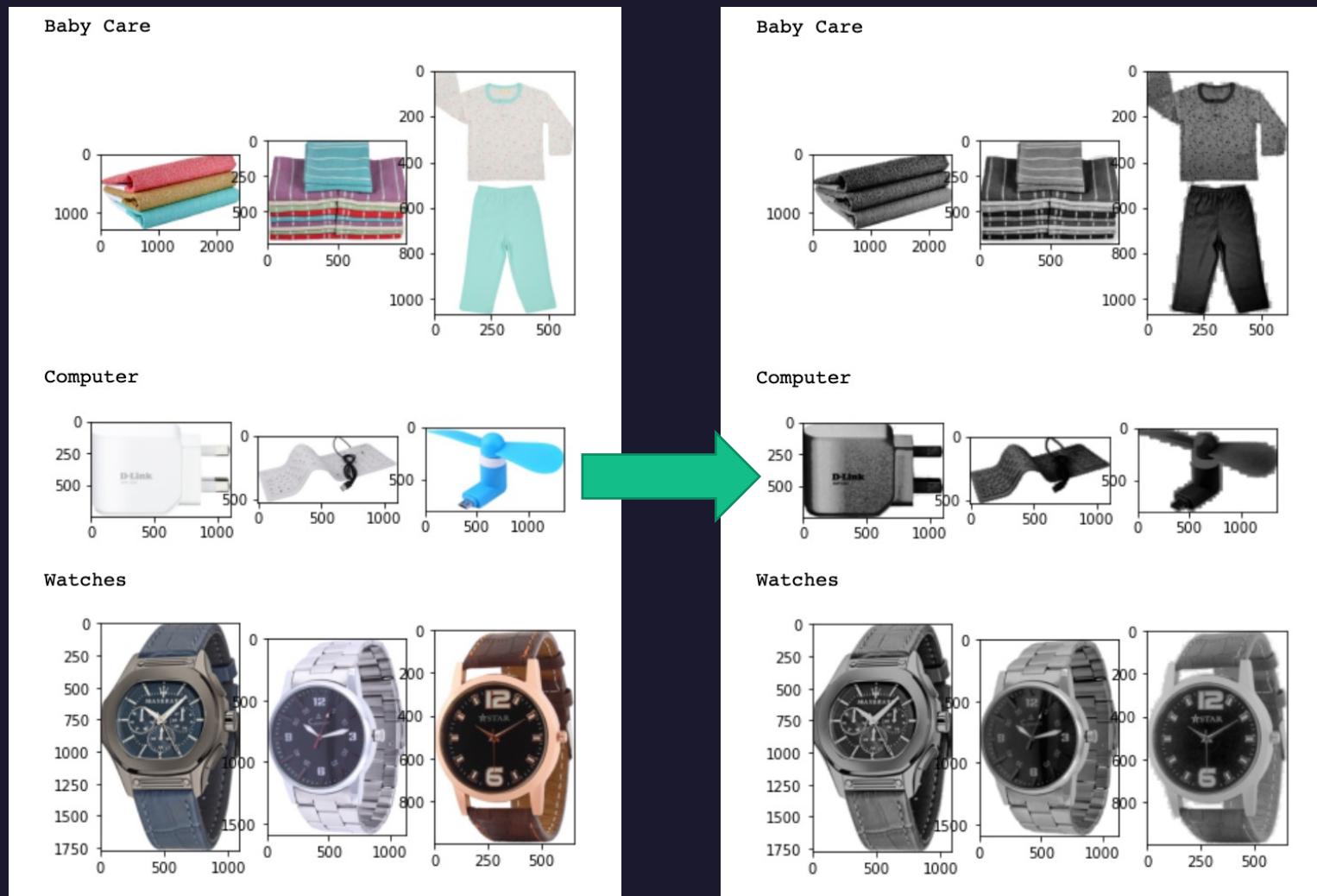


Idée générale

Approche visuelle



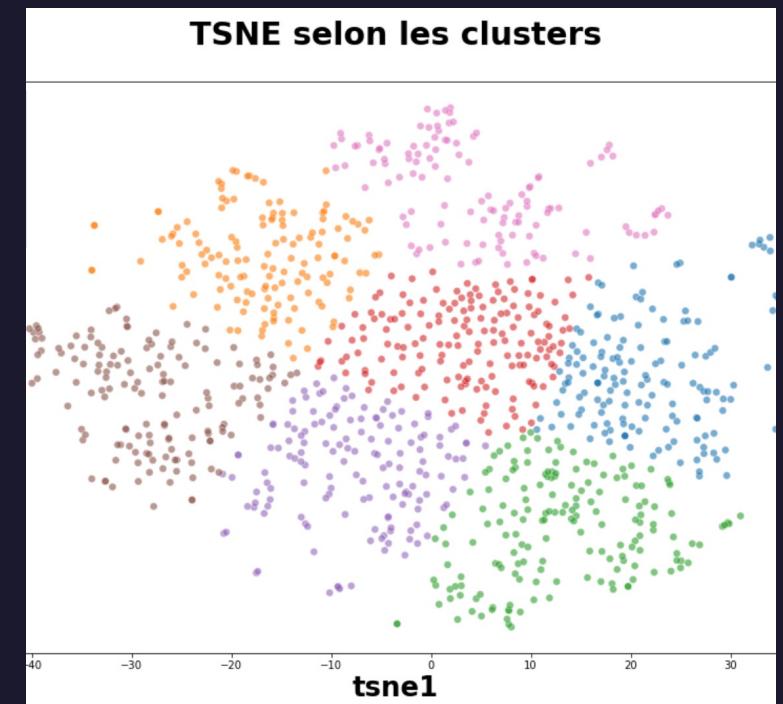
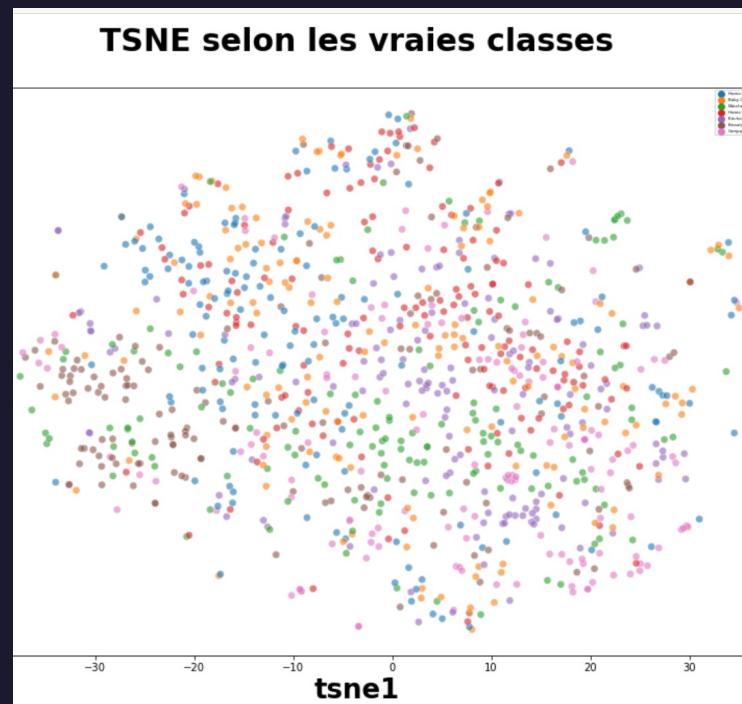
Nettoyage des images



- Passage au gris
- ajustement du contraste

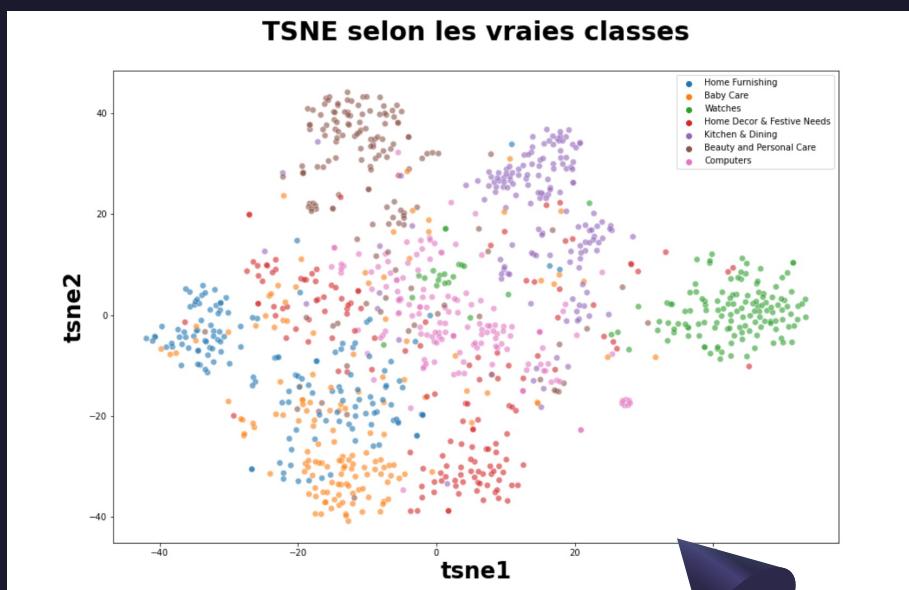
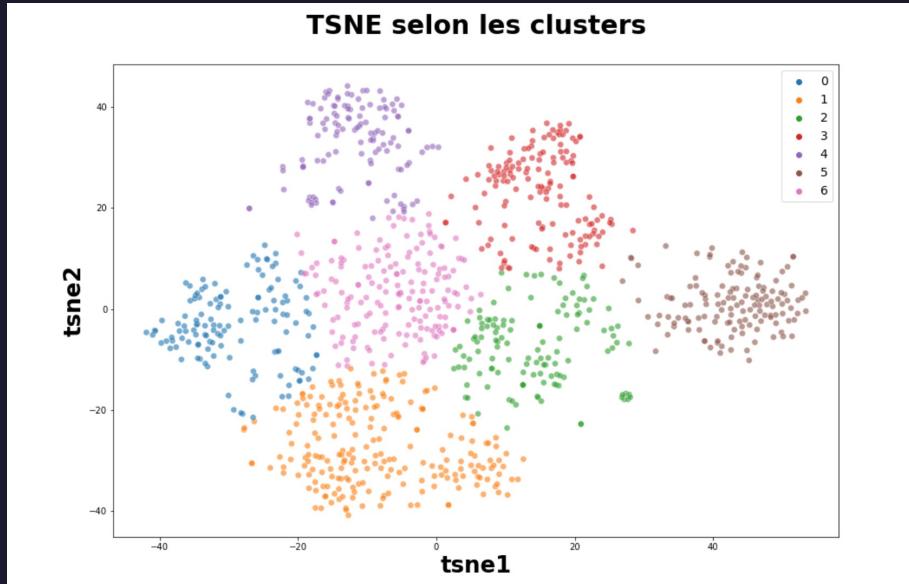
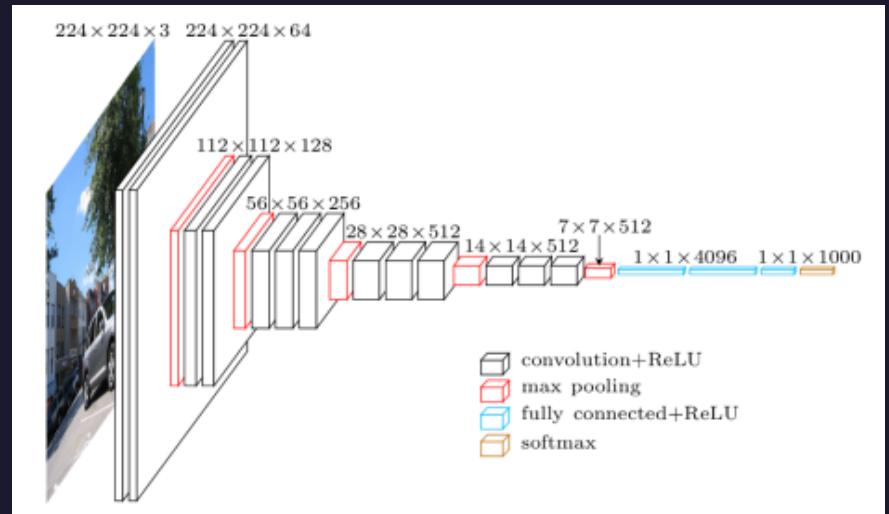
SIFT

Résultat mauvais : ARI=0.06



CNN

- ARI=0.4



Approche textuelle



Nettoyage des descriptions

- Passage en minuscules
- Élimination stop words
- lemmatization

Pré-processing

```
Entrée [48]: test=df[ "description "][1]
test

Out[48]: 'Specifications of Sathiyas Cotton Bath Towel (3 Bath Towel, Red, Yellow, Blue) Bath Towel Features Machine Washable Yes Material Cotton Design Self Design General Brand Sathiyas Type Bath Towel GSM 500 Model Name Sathiyas cotton bath towel Ideal For Men, Women, Boys, Girls Model ID asvtwl322 Color Red, Yellow, Blue Size Medium Dimensions Length 30 i nch Width 60 inch In the Box Number of Contents in Sales Package 3 Sales Package 3 Bath Towel'

Entrée [49]: def tokenizer_fct(sentence) :
    # print(sentence)
    sentence_clean = sentence.replace('-', ' ').replace('+', ' ').replace('/', ' ').replace('#', ' ')
    word_tokens = word_tokenize(sentence_clean)
    return word_tokens

stop_w = list(set(stopwords.words('english')) + ['[', ']', ',', '.', ':', '?', '(', ')'])

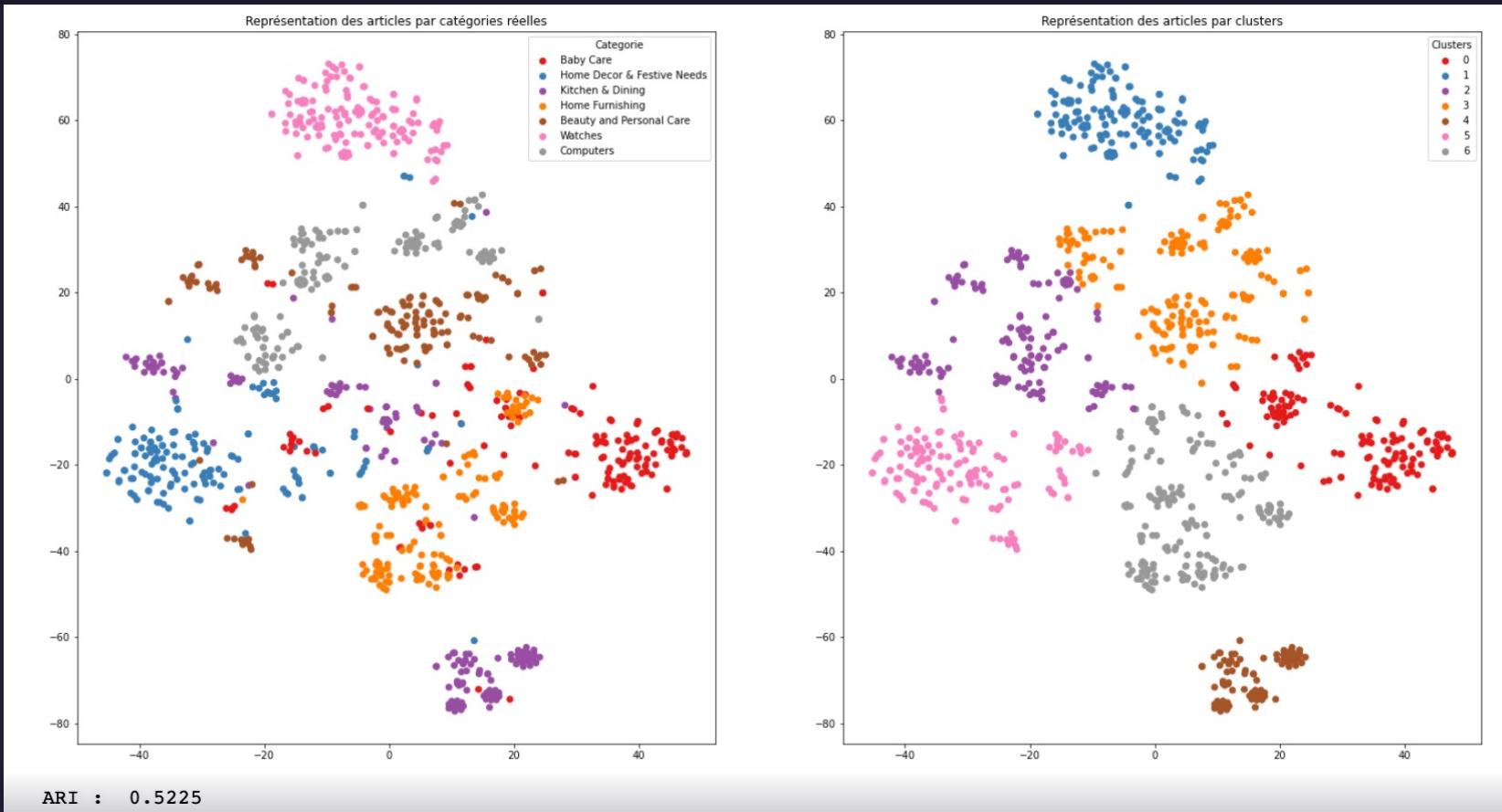
def stop_word_filter_fct(list_words) :
    filtered_w = [w for w in list_words if not w in stop_w]
    filtered_w2 = [w for w in filtered_w if len(w) > 2]
    return filtered_w2

# En minuscule
def lower_start_fct(list_words) :
    lw = [w.lower() for w in list_words]
    return lw

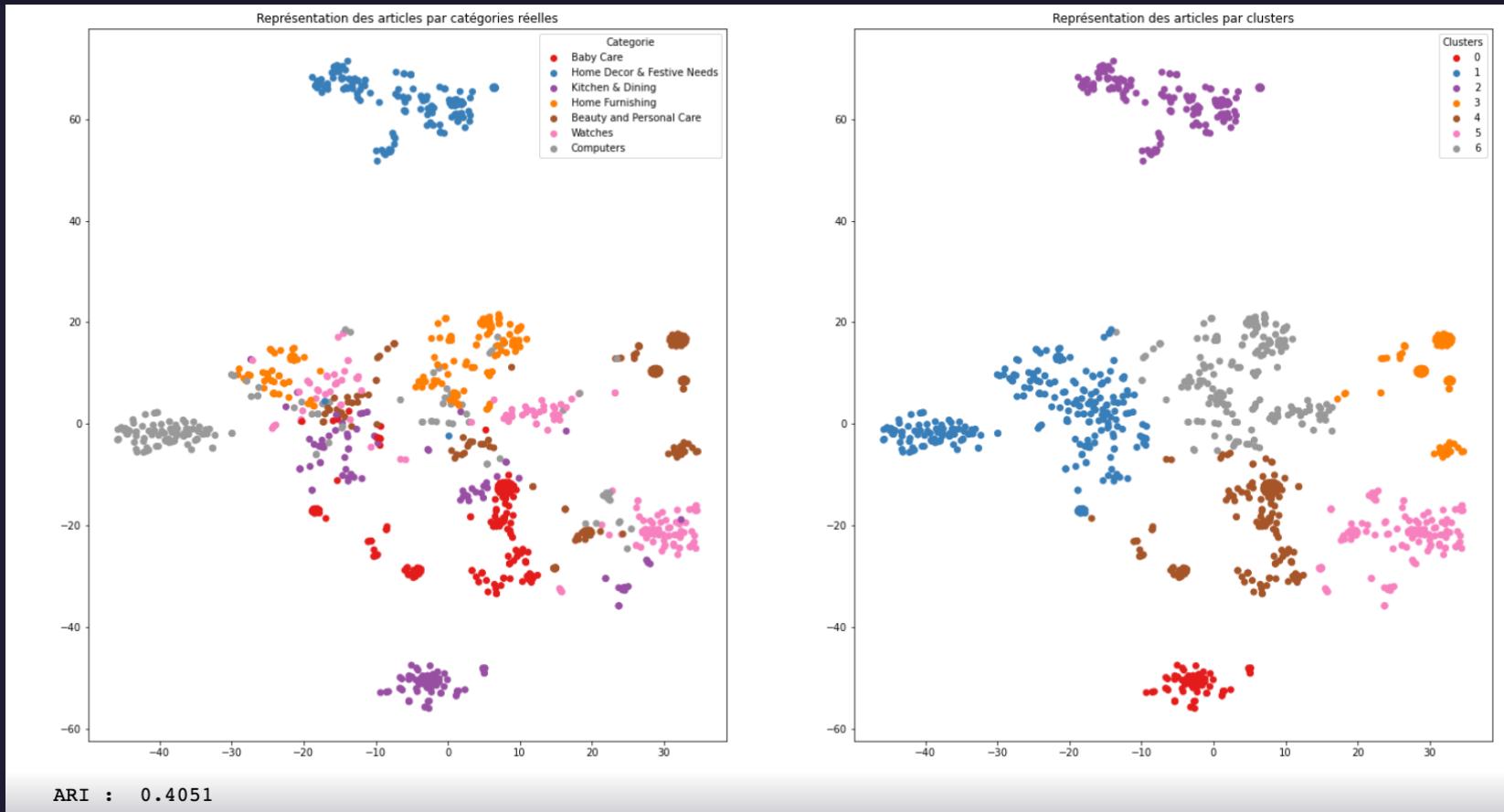
def lemma_fct(list_words) :
    lemmatizer = WordNetLemmatizer()
    lem_w = [lemmatizer.lemmatize(w) for w in list_words]
    return lem_w

def sans_chiffre(toto):
    number=["0","1","2","3","4","5","6","7","8","9"]
    for i in range(10):
        toto=[ x for x in toto if number[i] not in x ]
    return(toto)
```

Bag of word, ARI = 0.52



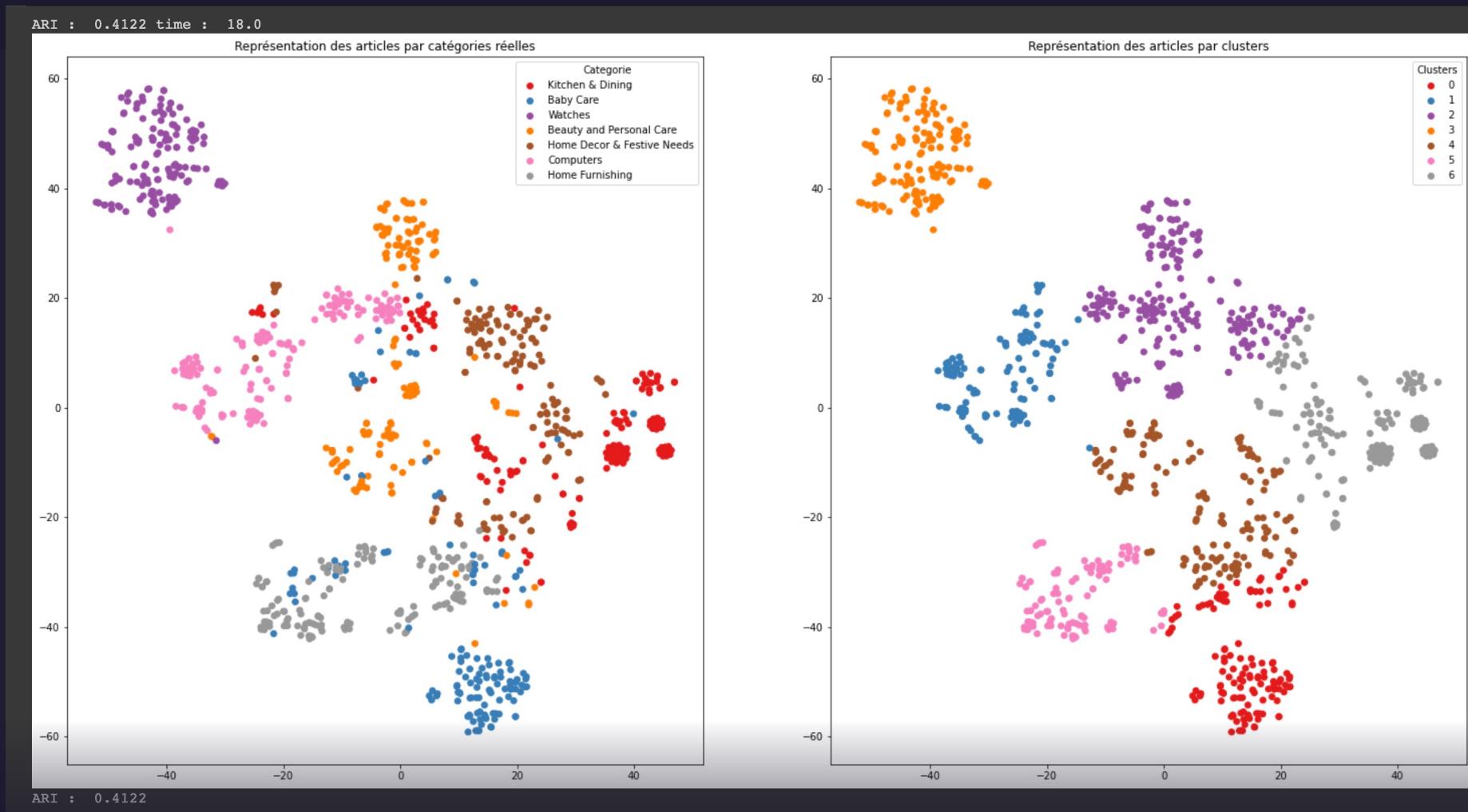
Word2vec, ARI=0.40



Bert, ARI=0.36



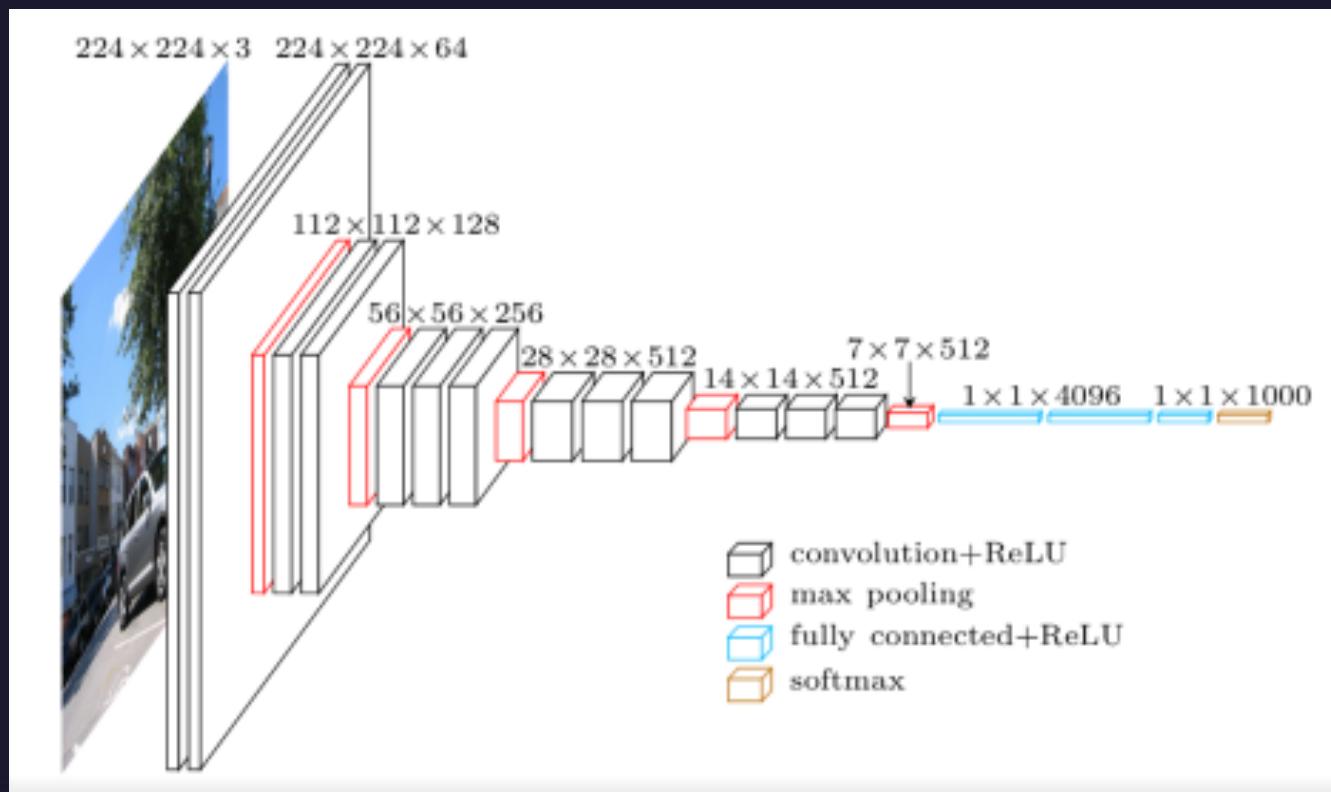
USE, ARI = 0.41



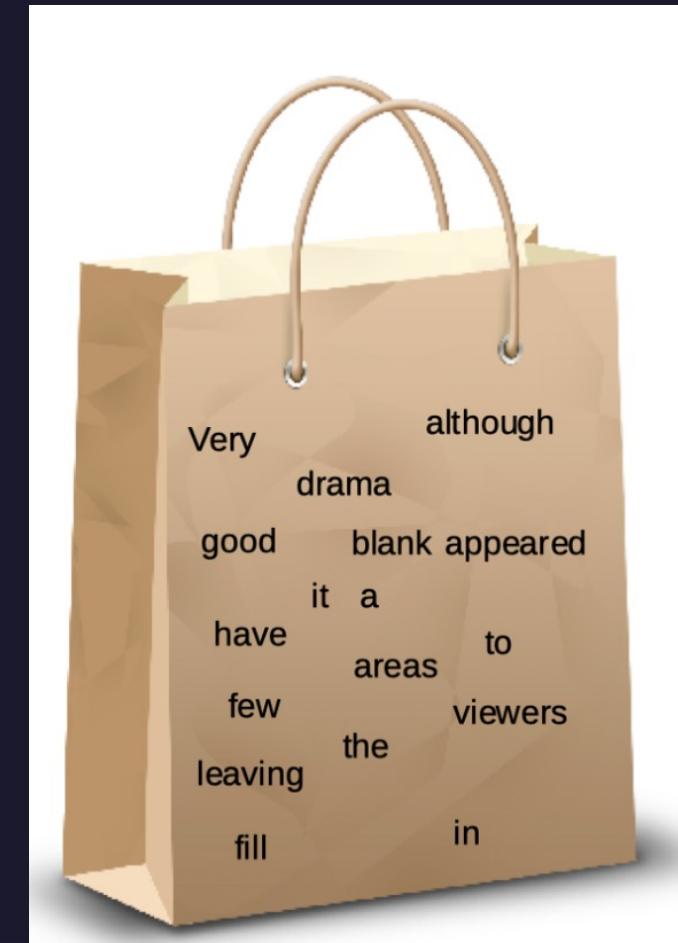
Conclusion

Meilleurs modèles :

CNN, ARI=0.40



Bag of word : ARI = 0.52



Faisabilité de la classification

Catégories	Approche visuelle	Approche textuelle	faisabilité
Accessoires pour bébés	65 %	72 %	oui
Meubles	53 %	87 %	oui
Montres	85 %	100 %	oui
Décoration d'intérieur	39 %	79 %	oui
Objets de cuisine	80 %	49 %	oui
Produits de beauté	73 %	63 %	oui
Matériel informatique	55 %	63 %	compliqué

Axes d'amélioration:

Redéfinition
des
catégories

Utilisation
des feedbacks
des clients

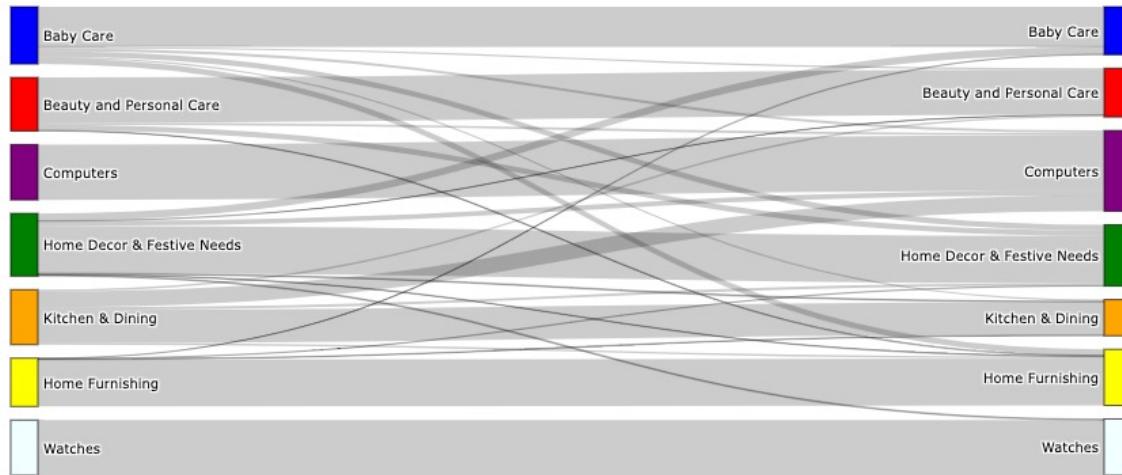
Approche
supervisée



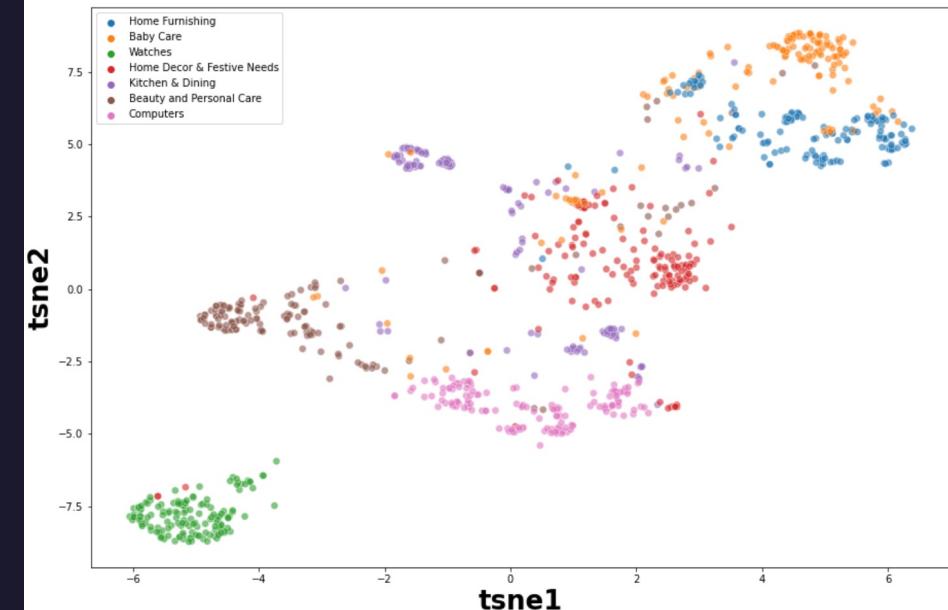
FIN

Approche ensembliste, ARI=0.65 et Accuracy=0.83

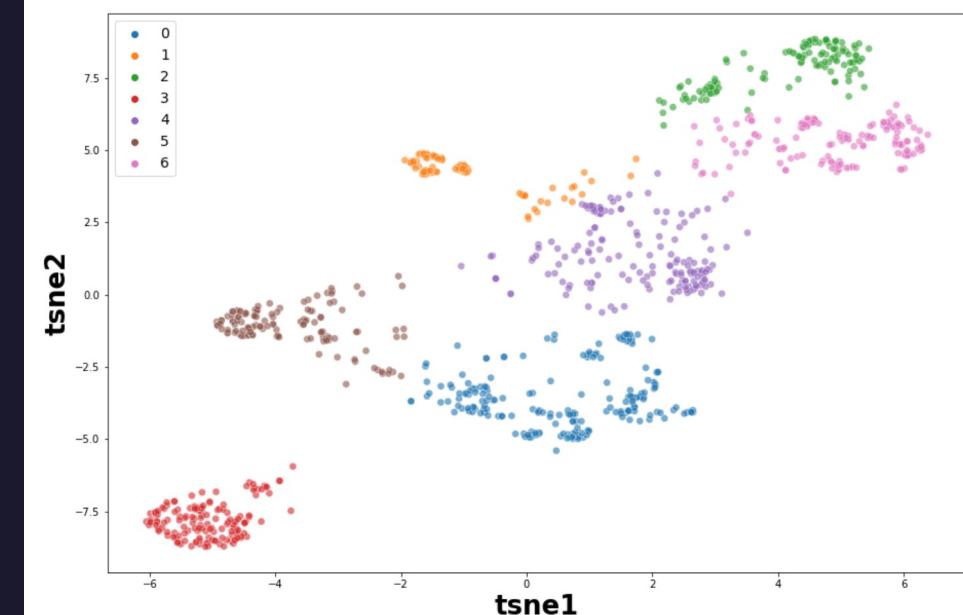
Diagramme de Sankey



TSNE selon les vraies classes



TSNE selon les clusters



ARI : 0.656137544390499