

# Classifiez automatiquement des biens de consommation

Yonss JOSE

Parcours: Data Scientist



23/12/2021

# Résumé du projet

"Place de marché", qui souhaite lancer une marketplace e-commerce. Sur la place de marché, des vendeurs proposent des articles à des acheteurs en postant une photo et une description

Pour l'instant, l'attribution de la catégorie 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.

## Demande de l'entreprise

- Etudier la faisabilité d'un moteur de classification des articles en différentes catégories

## Objectifs

- Rendre l'expérience utilisateur des vendeurs et des acheteurs la plus fluide possible
- Dans l'optique d'un passage à l'échelle, il devient nécessaire d'automatiser cette tâche

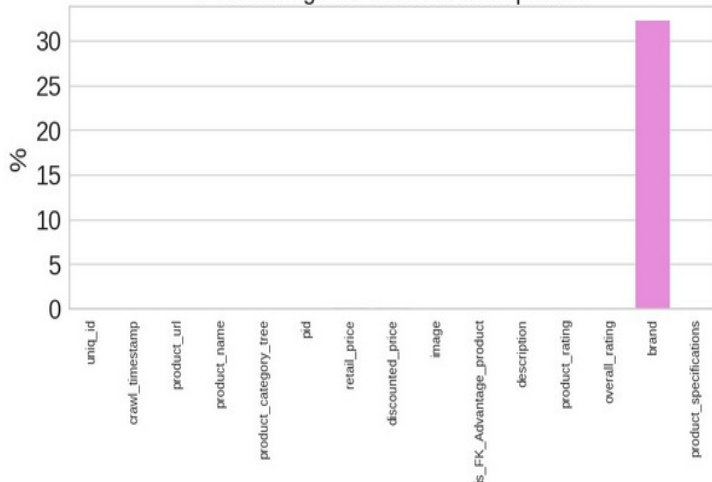
# Données

Données uniformément réparties

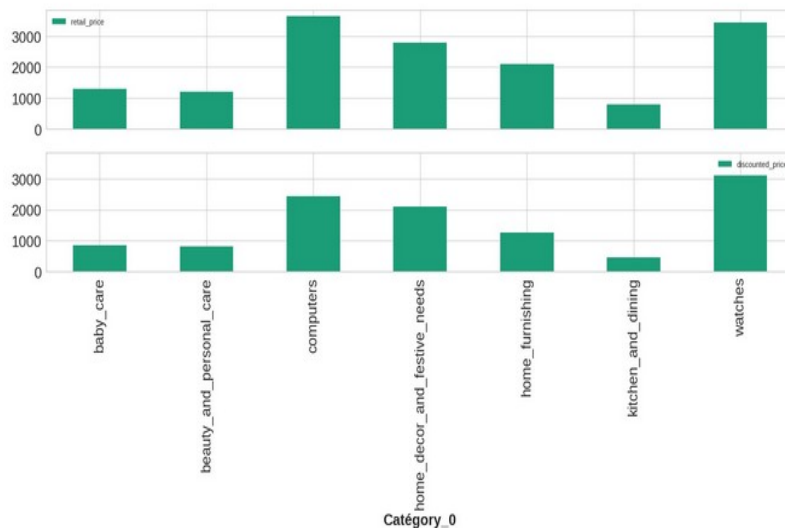
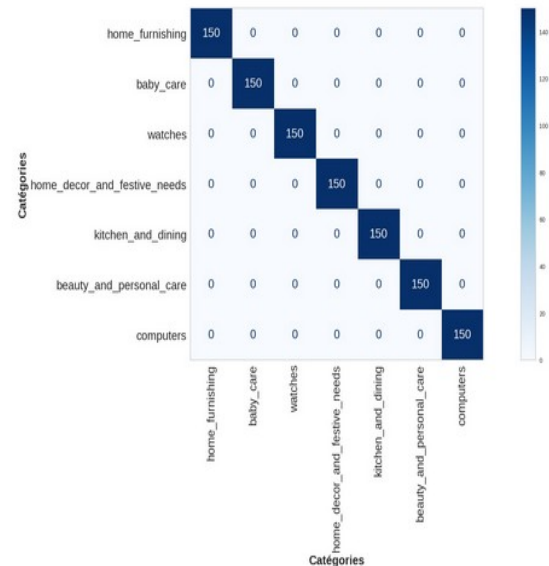
## Nettoyage des données

- Supprimer variable brand
- Remplacer null par Unknown

Pourcentage de données manquantes



Variables	Valeur Manquantes
uniq_id	0
crawl_timestamp	0
product_url	0
product_name	0
product_category_tree	0
pid	0
retail_price	1
discounted_price	1
image	0
is_FK_Advantage_product	0
description	0
product_rating	0
overall_rating	0
brand	338
product_specifications	1



# Classification d'articles basé sur la description

## Prétraitement

- Tokenization - Division en chaîne de caractères.
- Suppression des stopwords
- Suppression des mots vides

## Réduction de dimension

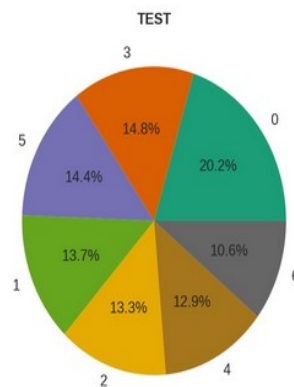
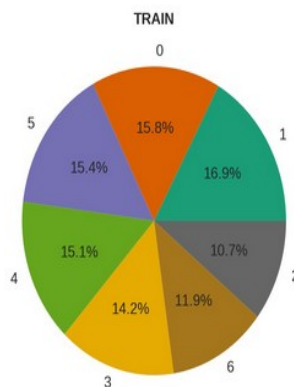
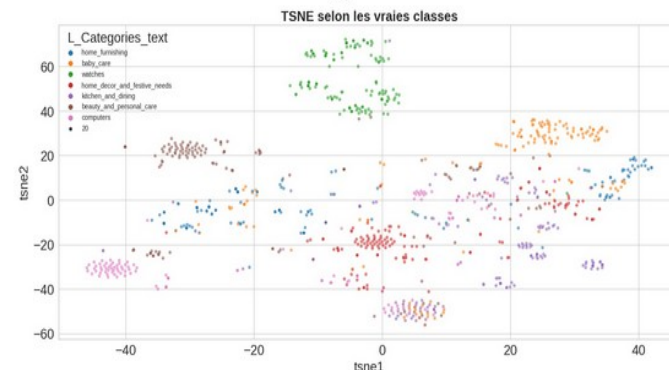
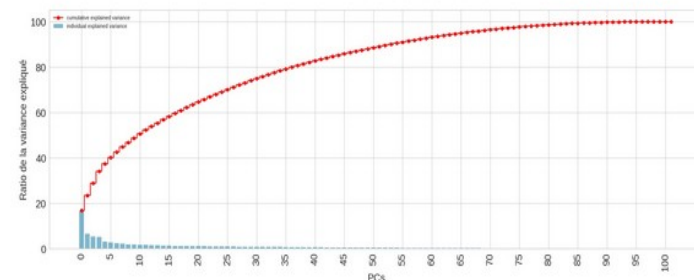
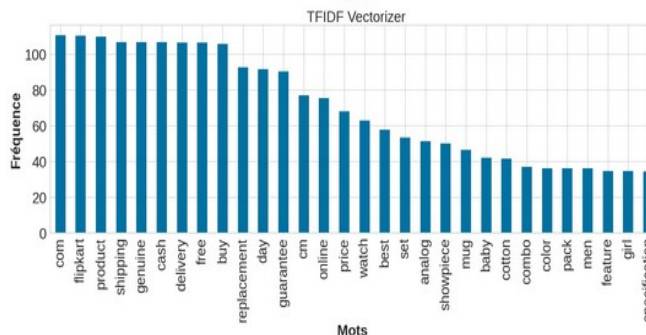
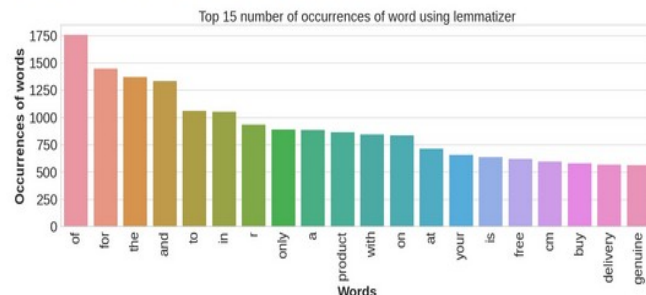
- PCA

## Classification

- KMeans

## Note à retenir

- Oui, est il possible de faire un moteur de classification des articles basés sur la description

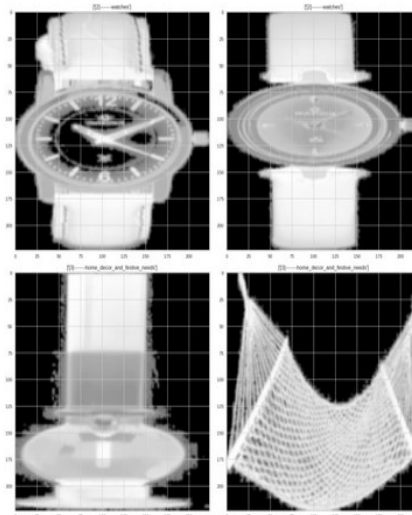




# Classification d'articles basé sur l'image

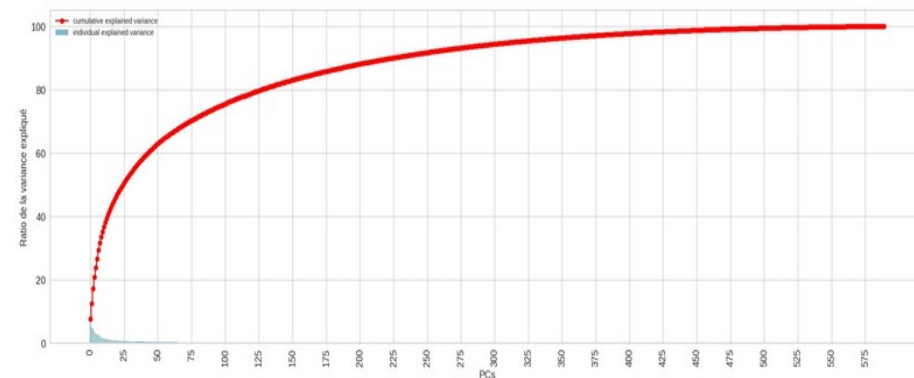
## Prétraitement

- Convertir en gris
- Convertir à la même taille
- Amélioration des contrastes
- Débruitage
- Filtre gaussien
- Trouver les points clés et les descripteurs

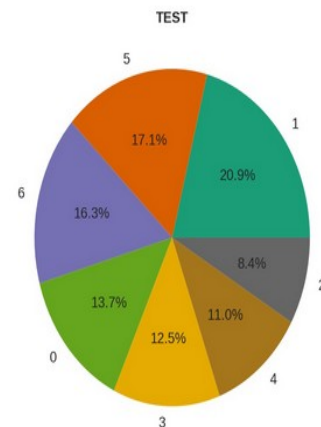
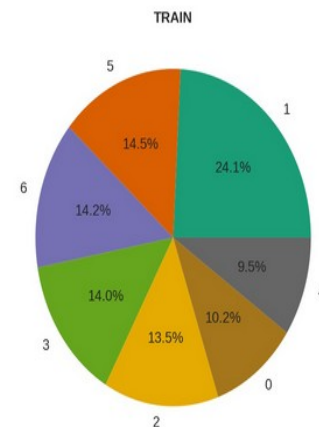
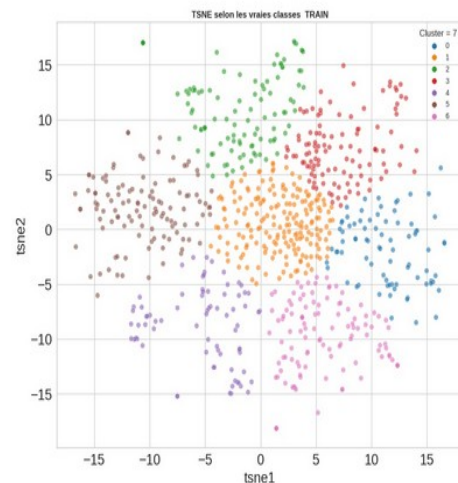


## Création des features des images

- Sift

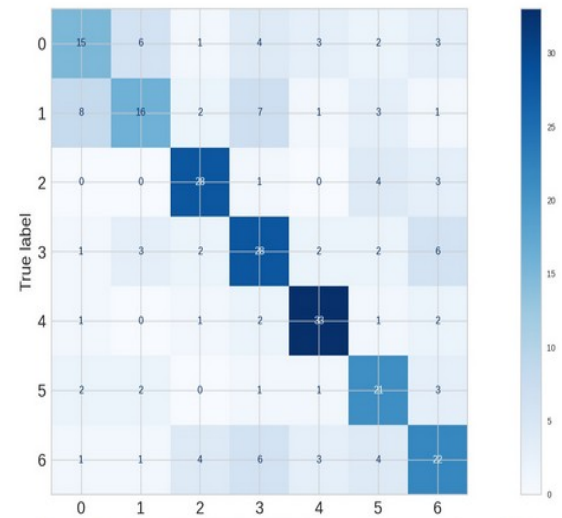
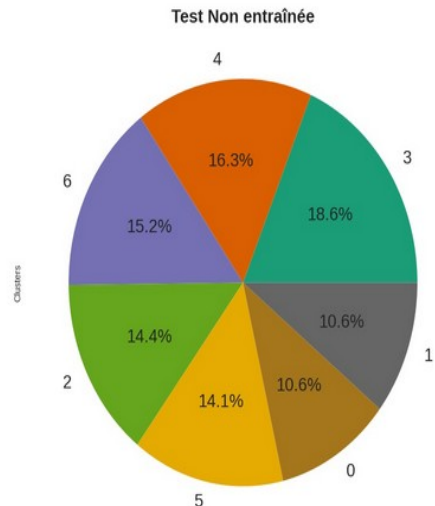
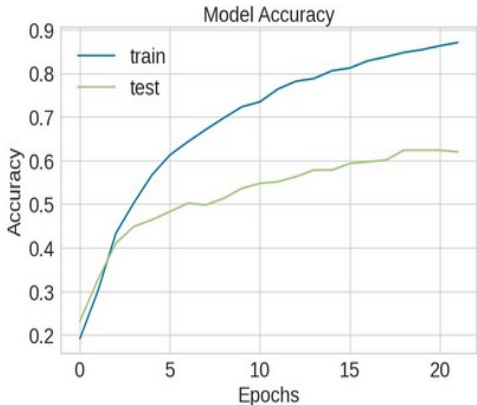


## Classification avec KMeans

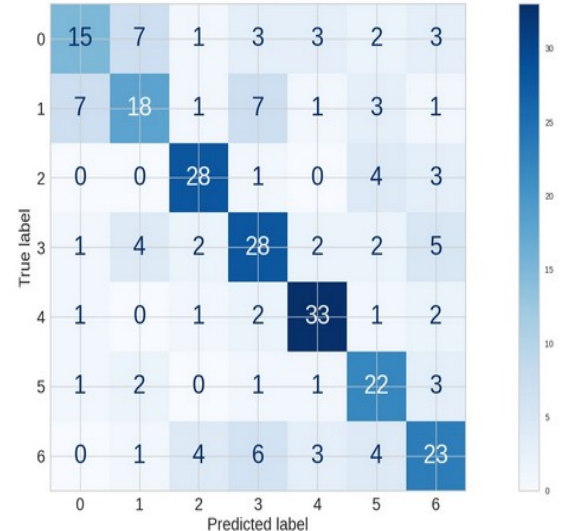
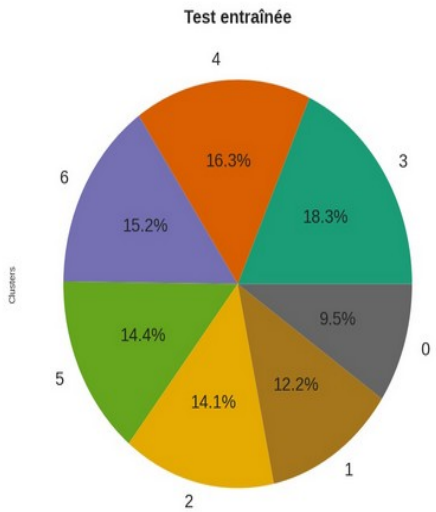
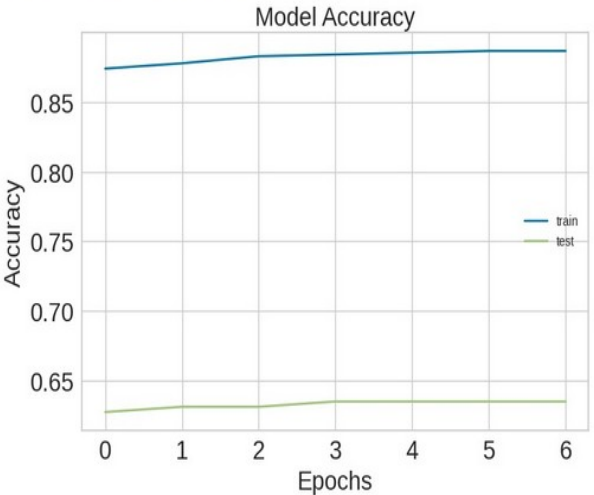


# Classification d'image avec CNN - VGG16

## Modèle non entraînée



## Modèle avec 3 couches entraînée



## Conclusion

- Oui, est il possible de faire un moteur de classification des articles basés sur la description