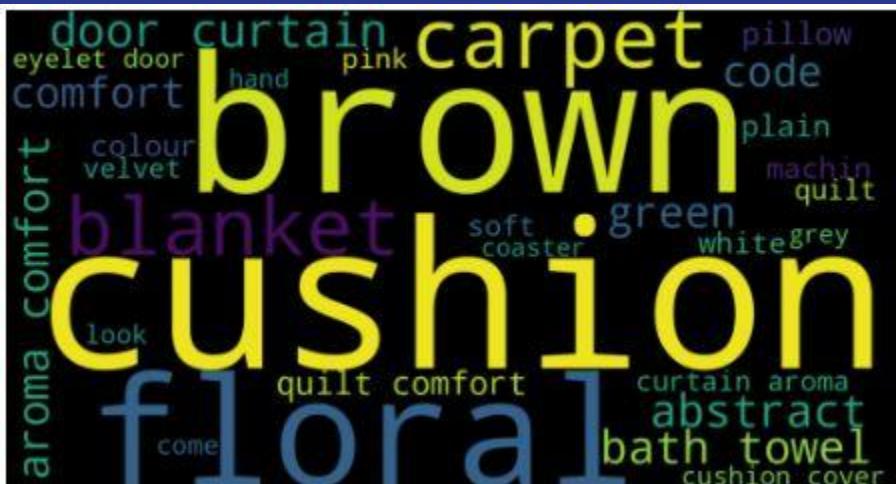


Classifiez automatiquement des biens de consommation



Données Textuelles

Images



Beauty and Personal Care



Home Decor & Festive Needs



Home Furnishing



Sommaire



- ❖ Problématique
- ❖ Les Données
- ❖ Traitement des données textuelles
- ❖ Traitement des données images
- ❖ Test API
- ❖ Le respect de la RGPD pour la collecte de données
- ❖ Conclusion et Suite du Projet

Problématique : Contexte

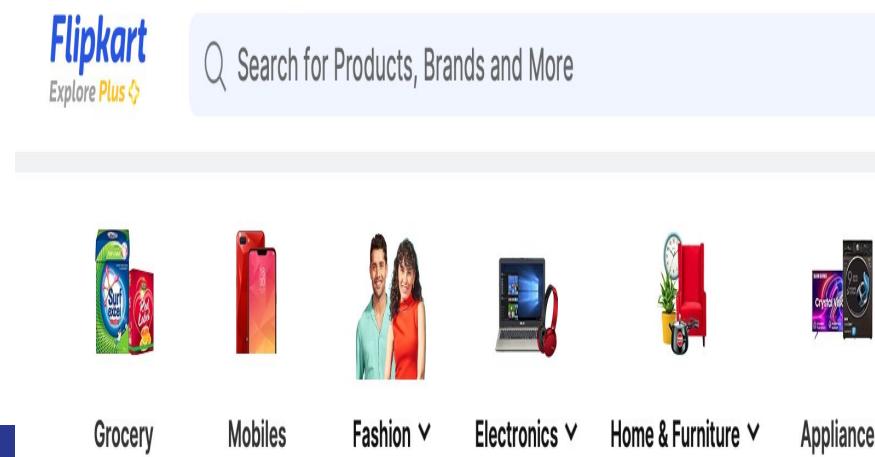


Mission :

Réaliser une **étude de faisabilité** d'un **moteur de classification** d'articles basé sur une image et une description pour l'**automatisation** de l'attribution de la **catégorie** de l'article pour l'entreprise *Place de Marché*.

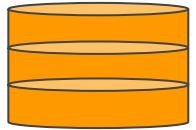
Objectif :

- ❖ **Améliorer** l'expérience des utilisateurs.
- ❖ **Fiabiliser** la catégorie des articles avec pertinence et précision.



Démarche

Pré-Traitement



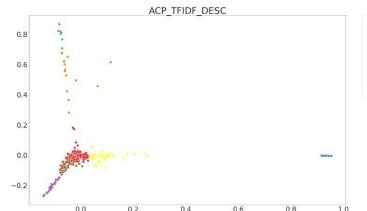
Données Textes

Features Extractions

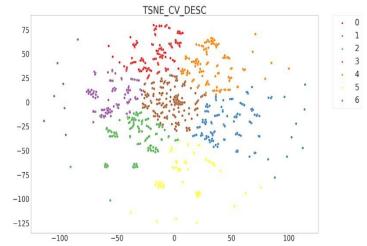
- Bags Of Words
- Words Embeddings
- Topics Embeddings

Réduction de dimensions

ACP

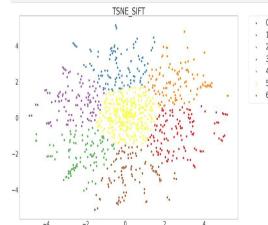


TSNE



Classification

Classification Non-Supervisée K Means



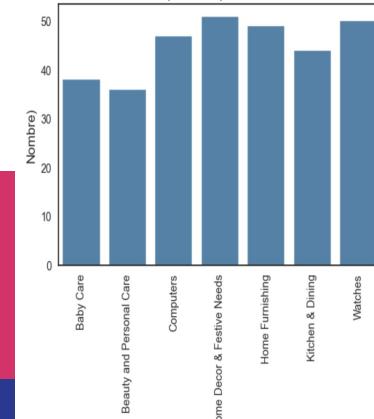
Evaluation

TSNE_XCEPTION

	0	1	2	3	4	5	8	4	0
Baby Care	26	2	0	5	8	4	0	0	0
Beauty and Personal Care	1	33	3	2	5	1	0	0	0
Computers	0	0	42	1	1	1	1	0	0
Home Decor & Festive Needs	1	0	2	37	3	2	0	0	0
Home Furnishing	6	0	1	2	36	0	0	0	0
Kitchen & Dining	1	2	1	0	1	40	0	0	0
Watches	0	0	0	2	0	0	0	43	0
	0	1	2	3	4	5	6	7	8

'Précision: 81.59%'

Répartition par clusters



Classification Supervisée : Pycaret

Les Données



TEXTES

product_name
Sathiya's Cotton Bath
Towel

description
Specifications of Sathiya's Cotton
Bath
Towel (3 Bath Towel, Red, Yellow,
Blue)
Bath Towel Features Machine
Washable Yes Material .../..

product_specifications
{"product_specification":>[{"key":>"Brand", "value":>"Elegance"}, {"key":>"Designed For", "value":>"Door"}, ...]

brand
Sathiya's

IMAGES



CIBLE

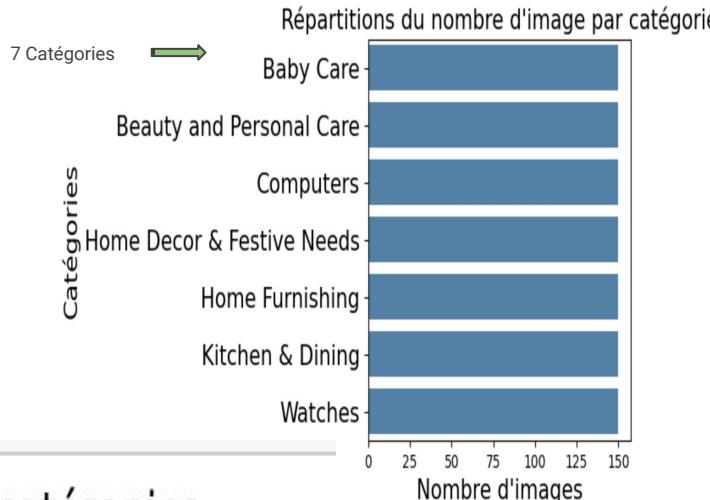
product_category_tree
"Baby Care >> Baby Bath & Skin >> Baby
Bath Towels >> Sathiya's Baby Bath
Towels >> Sathiya's Cotton Bath Towel (3
Bath Towel, Red, Y...")

Les Données : Niveaux des catégories



Niveau 1

150 images ou articles par catégorie



Categorie_1 : 7 catégories

Categorie_2 : 62 catégories

Categorie_3 : 241 catégories

Categorie_4 : 349 catégories

Categorie_5 : 297 catégories

Categorie_6 : 117 catégories

Categorie_7 : 57 catégories

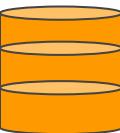
"Nombre de catégories contenant x% d'images pour la plage pour le niveau 1"

Plage	nb_données	%_données
(0, 10]	0	0.000000
(10, 50]	0	0.000000
(50, 100]	0	0.000000
(100, 150]	7	100.000000

Traitements des données Textes : NLP

Pré-Traitements

Données Textes



Description

Tokenisation



Normalisation



Stemmatisation



Lemmatisation

Features Extractions

Bags Of Words

- CountVectorizer
- TfidfVectorizer

Words Embeddings

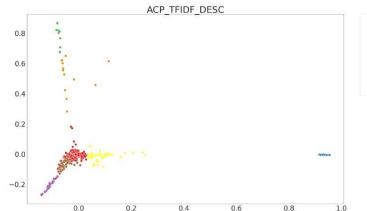
- Word2Vec

Sentence-Transformers

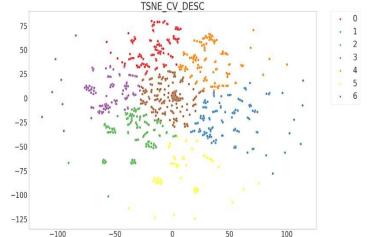
- BERT: roberta-large
- USE

Réduction de dimensions

ACP

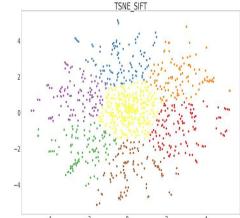


TSNE



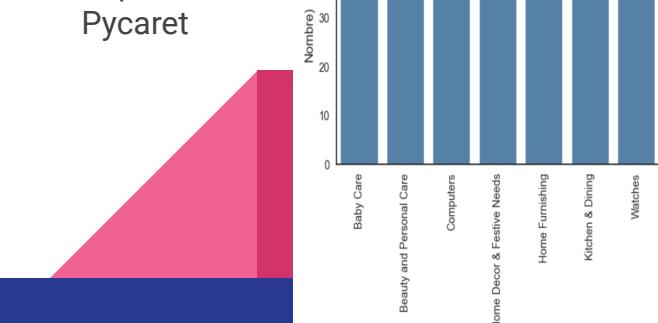
Classification

Classification Non-Supervisée K Means



Classification Supervisée : Pycaret

Répartition par clusters



Evaluation

	TSNE_XCEPTION						
Baby Care	26	2	0	5	8	4	0
Beauty and Personal Care	1	33	3	2	5	1	0
Computers	0	0	42	1	1	1	0
Home Decor & Festive Needs	1	0	2	37	3	2	0
Home Furnishing	6	0	1	2	36	0	0
Kitchen & Dining	1	2	1	0	1	40	0
Watches	0	0	0	2	0	0	43

'Précision: 81.59%'

Traitement des données Textes : NLP Pré-traitement

Action

Contenu

Wallmantra Large Vinyl Sticker Sticker (Pack of 1) Price: Rs. 1,896 Bring home this exclusive Piece of Wall Art to give your home a refreshing look it deserves ! Wall Decals are the latest trend, sweeping the world of interior design, as a quick and easy way to personalise and transform your home. We at Wallmantra use only the highest quality premium self-adhesive vinyl for our products to ensure you receive the best quality product. The sizes are in inches (Width x Height), rounded to the nearest inch. The size of the product is the final size that will be up on the walls . These are very easy to apply and remove. You can apply these yourself. We send ready to apply Decal/Sticker with 2 practice items with instruction manual. You can also see the how to apply video. These are not re-usable. Contact Wallmantra for more options Bring home this exclusive Piece of Wall Art to give your home a refreshing look it deserves ! Wall Decals are the latest trend, sweeping the world of interior design, as a quick and easy way to personalise and transform your home. We at Wallmantra use only the highest quality premium self-adhesive vinyl for our products to ensure you receive the best quality product. The sizes are in inches (Width x Height), rounded to the nearest inch. The size of the product is the final size that will be up on the walls . These are very easy to apply and remove. You can apply these yourself. We send ready to apply Decal/Sticker with 2 practice items with instruction manual. You can also see the how to apply video. These are not re-usable, Contact Wallmantra for more options

Texte original

wallmantra large vinyl sticker sticker (pack of 1) price: rs. 1,896 bring home this exclusive piece of wall art to give your home a refreshing look it deserves ! wall decals are the latest trend, sweeping the world of interior design, as a quick and easy way to personalise and transform your home. we at wallmantra use only the highest quality premium self-adhesive vinyl for our products to ensure you receive the best quality product. the sizes are in inches (width x height), rounded to the nearest inch. the size of the product is the final size that will be up on the walls . these are very easy to apply and remove. you can apply these yourself. we send ready to apply decal/sticker with 2 practice items with instruction manual, you can also see the how to apply video, these are not re-usable. contact wallmantra for more options bring home this exclusive piece of wall art to give your home a refreshing look it deserves ! wall decals are the latest trend, sweeping the world of interior design, as a quick and easy way to personalise and transform your home. we at wallmantra use only the highest quality premium self-adhesive vinyl for our products to ensure you receive the best quality product. the sizes are in inches (width x height), rounded to the nearest inch. the size of the product is the final size that will be up on the walls . these are very easy to apply and remove, you can apply these yourself, we send ready to apply decal/sticker with 2 practice items with instruction manual. you can also see the how to apply video, these are not re-usable, contact wallmantra for more options

Mettre en minuscule

['wallmantra', 'large', 'vinyl', 'sticker', 'sticker', 'pack', 'of', 'price', 'rs', 'bring', 'home', 'this', 'exclusive', 'piece', 'of', 'wall', 'art', 'to', 'give', 'your', 'home', 'a', 'refreshing', 'look', 'it', 'deserves', 'wall', 'decals', 'are', 'the', 'latest', 'trend', 'sweeping', 'the', 'world', 'of', 'interior', 'design', 'as', 'a', 'quick', 'and', 'easy', 'way', 'to', 'personalise', 'and', 'transform', 'your', 'home', 'we', 'at', 'wallmantra', 'use', 'only', 'the', 'highest', 'quality', 'premium', 'self', 'adhesive', 'vinyl', 'for', 'our', 'products', 'to', 'ensure', 'you', 'receive', 'the', 'best', 'quality', 'product', 'the', 'sizes', 'are', 'in', 'inches', 'width', 'x', 'height', 'rounded', 'to', 'the', 'nearest', 'inch', 'the', 'size', 'of', 'the', 'product', 'is', 'the', 'final', 'size', 'that', 'will', 'be', 'up', 'on', 'the', 'walls', 'these', 'are', 'very', 'easy', 'to', 'apply', 'and', 'remove', 'you', 'can', 'apply', 'these', 'yourself', 'we', 'send', 'ready', 'to', 'apply', 'decal', 'sticker', 'with', 'practice', 'items', 'with', 'instruction', 'manual', 'you', 'can', 'also', 'see', 'the', 'how', 'to', 'apply', 'video', 'these', 'are', 'not', 're', 'usable', 'contact', 'wallmantra', 'for', 'more', 'options']

Tokenisation

['wallmantra', 'large', 'vinyl', 'sticker', 'sticker', 'pack', 'price', 'bring', 'home', 'this', 'exclusive', 'piece', 'wall', 'art', 'give', 'your', 'home', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'are', 'the', 'latest', 'trend', 'sweeping', 'the', 'world', 'interior', 'design', 'quick', 'and', 'easy', 'way', 'personalise', 'and', 'transform', 'your', 'home', 'wallmantra', 'use', 'only', 'the', 'highest', 'quality', 'premium', 'self', 'adhesive', 'vinyl', 'for', 'our', 'products', 'ensure', 'you', 'receive', 'the', 'best', 'quality', 'product', 'the', 'sizes', 'are', 'inches', 'width', 'height', 'rounded', 'the', 'nearest', 'inch', 'the', 'size', 'the', 'product', 'the', 'final', 'size', 'that', 'will', 'the', 'walls', 'these', 'are', 'very', 'easy', 'to', 'apply', 'and', 'remove', 'you', 'can', 'apply', 'these', 'yourself', 'we', 'send', 'ready', 'to', 'apply', 'decal', 'sticker', 'with', 'practice', 'items', 'with', 'instruction', 'manual', 'you', 'can', 'also', 'see', 'the', 'how', 'to', 'apply', 'video', 'these', 'are', 'not', 're', 'usable', 'contact', 'wallmantra', 'for', 'more', 'options']

Supprimer tokens courts

Traitement des données Textes : NLP Pré-traitement

	<pre>[wallmantra, 'large', 'vinyl', 'sticker', 'sticker', 'pack', 'price', 'bring', 'home', 'exclusive', 'piece', 'wall', 'art', 'give', 'home', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'design', 'quick', 'easy', 'way', 'personalise', 'transform', 'home', 'width', 'height', 'rounded', 'nearest', 'inch', 'size', 'product', 'final', 'size', 'walls', 'easy', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'also', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options', 'bring', 'home', 'exclusive', 'piece', 'wall', 'art', 'give', 'home', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'design', 'quick', 'easy', 'way', 'personalise', 'transform', 'home', 'wallmantra', 'use', 'highest', 'quality', 'premium', 'self', 'adhesive', 'vinyl', 'products', 'ensure', 'receive', 'best', 'quality', 'product', 'sizes', 'inches', 'width', 'height', 'rounded', 'nearest', 'inch', 'size', 'product', 'final', 'size', 'walls', 'easy', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'also', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options']</pre>
supp_stopwords	<pre>[wallmantra, 'large', 'vinyl', 'sticker', 'sticker', 'bring', 'exclusive', 'piece', 'wall', 'art', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'quick', 'way', 'personalise', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'receive', 'sizes', 'inches', 'rounded', 'nearest', 'final', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options', 'bring', 'exclusive', 'piece', 'wall', 'art', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'quick', 'way', 'personalise', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'receive', 'sizes', 'inches', 'rounded', 'nearest', 'final', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options']</pre>
Suppressions 1er corpus communs	<pre>[wallmantra, 'large', 'vinyl', 'sticker', 'sticker', 'bring', 'exclusive', 'piece', 'wall', 'art', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'quick', 'way', 'personalise', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'receive', 'sizes', 'inches', 'rounded', 'nearest', 'final', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options', 'bring', 'exclusive', 'piece', 'wall', 'art', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'quick', 'way', 'personalise', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'receive', 'sizes', 'inches', 'rounded', 'nearest', 'final', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options']</pre>
Tokens en 1 phrase	wallmantra large vinyl sticker sticker bring exclusive piece wall art refreshing look deserves wall decals latest trend sweeping world interior quick way personalise transform wallmantra highest premium self adhesive vinyl ensure receive sizes inches rounded nearest final walls apply remove apply send ready apply decal sticker practice items instruction manual see apply video usable contact wallmantra options bring exclusive piece wall art refreshing look deserves wall decals latest trend sweeping world interior quick way personalise transform wallmantra highest premium self adhesive vinyl ensure receive sizes inches rounded nearest final walls apply remove apply send ready apply decal sticker practice items instruction manual see apply video usable contact wallmantra options
Tokenisation	<pre>[wallmantra, 'large', 'vinyl', 'sticker', 'sticker', 'bring', 'exclusive', 'piece', 'wall', 'art', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'quick', 'way', 'personalise', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'receive', 'sizes', 'inches', 'rounded', 'nearest', 'final', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options', 'bring', 'exclusive', 'piece', 'wall', 'art', 'refreshing', 'look', 'deserves', 'wall', 'decals', 'latest', 'trend', 'sweeping', 'world', 'interior', 'quick', 'way', 'personalise', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'receive', 'sizes', 'inches', 'rounded', 'nearest', 'final', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'practice', 'items', 'instruction', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options']</pre>
Supprimer 2 em corpus les mots frequents-rares	<pre>[wallmantra, 'large', 'vinyl', 'sticker', 'sticker', 'bring', 'exclusive', 'piece', 'art', 'refreshing', 'look', 'decals', 'latest', 'trend', 'world', 'interior', 'quick', 'way', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'sizes', 'inches', 'rounded', 'nearest', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'items', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options', 'bring', 'exclusive', 'piece', 'art', 'refreshing', 'look', 'decals', 'latest', 'trend', 'world', 'interior', 'quick', 'way', 'transform', 'wallmantra', 'highest', 'premium', 'self', 'adhesive', 'vinyl', 'ensure', 'sizes', 'inches', 'rounded', 'nearest', 'walls', 'apply', 'remove', 'apply', 'send', 'ready', 'apply', 'decal', 'sticker', 'items', 'manual', 'see', 'apply', 'video', 'usable', 'contact', 'wallmantra', 'options']</pre>

Traitement des données Textes : NLP Pré-traitement et fréquence des mots

Stemmatisation nearest wall send decal sticker item manual see video contact option bring art refresh look decal latest trend world interior quick way transform highest premium self vinyl size inch round nearest wall send decal sticker item manual see video contact option

```
['vinyl', 'sticker', 'sticker', 'bring', 'art', 'refresh', 'look', 'decal', 'latest', 'trend', 'world', 'interior', 'quick', 'way', 'transform', 'highest', 'premium', 'self', 'vinyl', 'size', 'inch', 'round', 'nearest', 'wall', 'send', 'decal', 'sticker', 'item', 'manual', 'see', 'video', 'contact', 'option', 'bring', 'art', 'refresh', 'look', 'decal', 'latest', 'trend', 'world', 'interior', 'quick', 'way', 'transform', 'highest', 'premium', 'self', 'vinyl', 'size', 'inch', 'round', 'nearest', 'wall', 'send', 'decal', 'sticker', 'item', 'manual', 'see', 'video', 'contact', 'option']
```

Lemmatisation vinyl sticker sticker bring art refresh look decal latest trend world interior quick way transform highest premium self vinyl size inch round nearest wall send decal sticker item manual see video contact option bring art refresh look decal latest trend world interior quick way transform highest premium self vinyl size inch round nearest wall send decal sticker item manual see video contact option

Home Furnishing



Baby Care



Traitements des données Textes : NLP

Pré-Traitements

Données Textes



Description
Stémmatisée
et Lemmatisée

Features Extractions

Bags Of Words

TfIdfVectorizer

CountVectorizer

Réduction de dimensions

ACP +
T-SNE

TSNE

Classification

Classification
Non-Supervisée
K Means

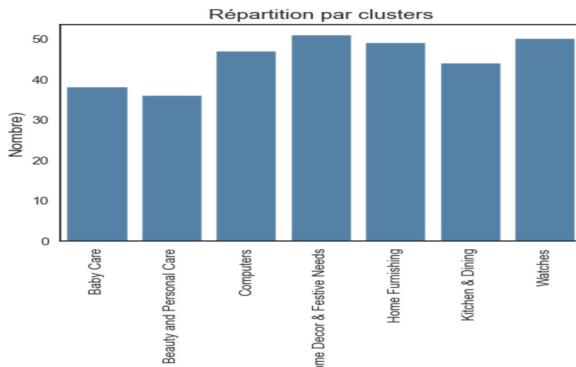
Classification
Supervisée :
Pycaret avec
BERT et **USE**

Evaluation

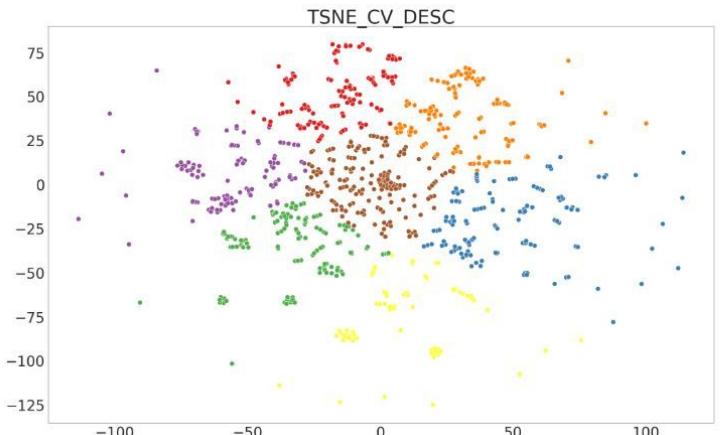
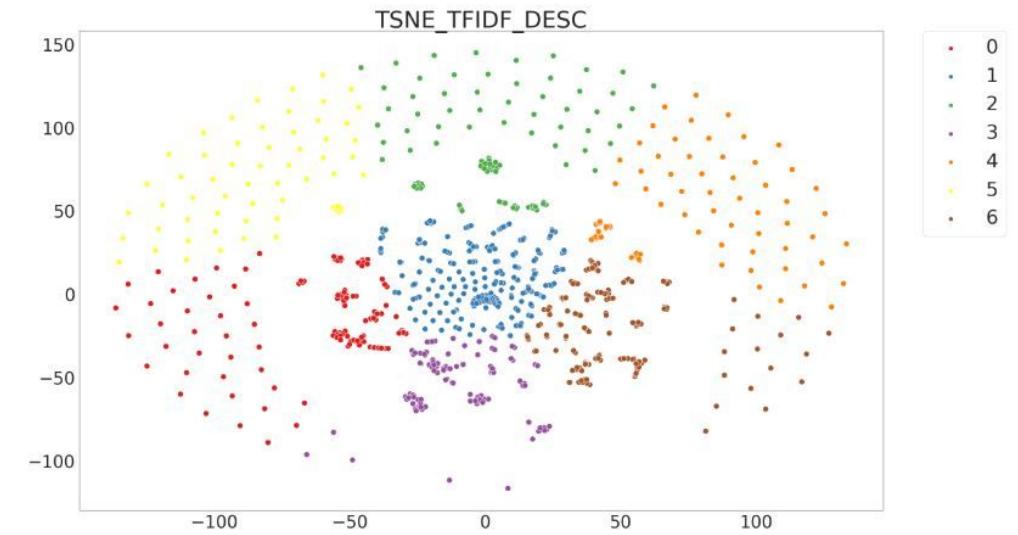
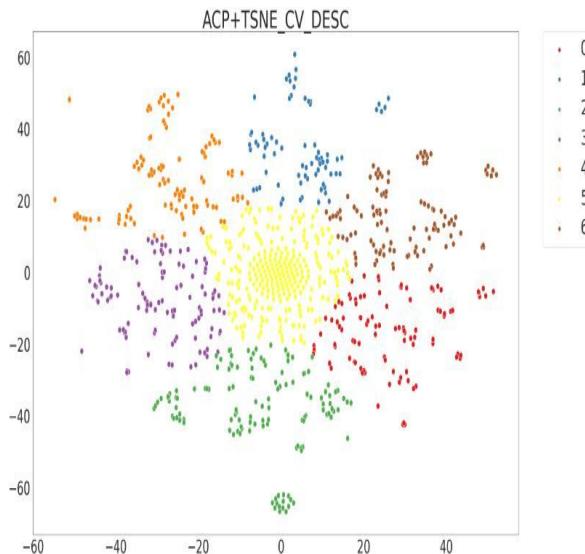
Évaluation
ARI
Accuracy

	TSNE_XCEPTION							
	Baby Care	26	2	0	5	8	4	0
Baby Care	1	33	3	2	5	1	0	0
Beauty and Personal Care	0	0	42	1	1	1	0	0
Computers	1	0	2	37	3	2	0	0
Home Decor & Festive Needs	6	0	1	2	36	0	0	0
Home Furnishing	1	2	1	0	1	40	0	0
Kitchen & Dining	0	0	0	2	0	0	43	0
Watches	0	1	2	3	4	5	0	0

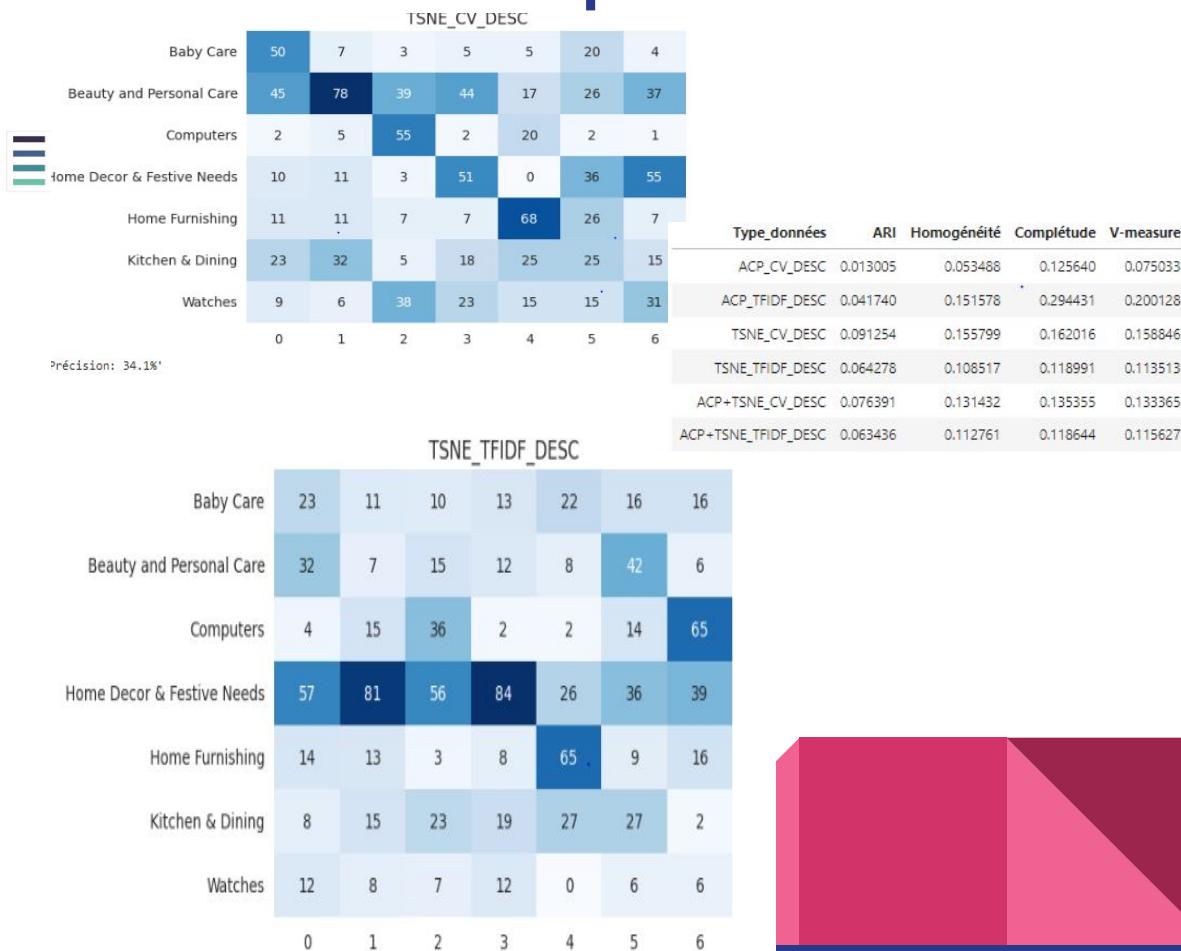
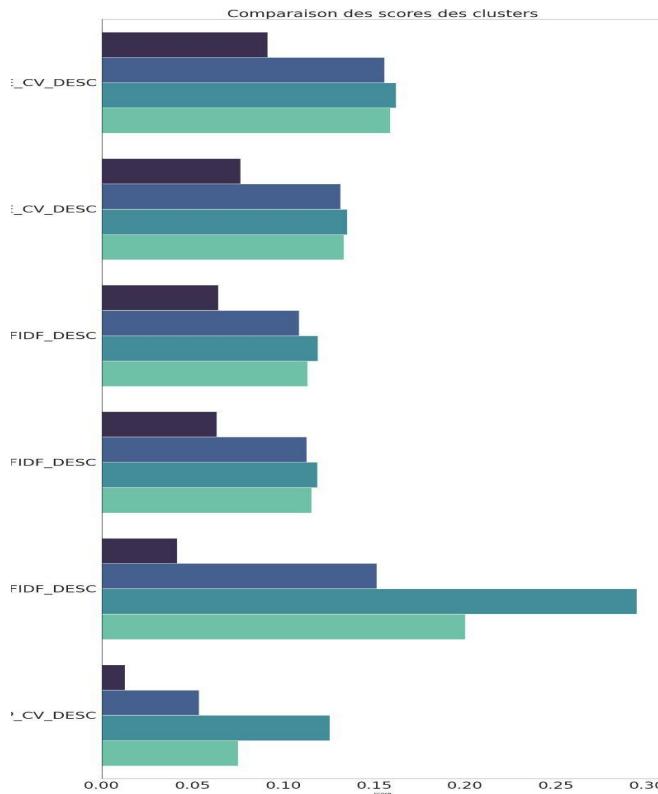
*Précision: 81.59%



Traitement des données Textes: Réduction dimensions

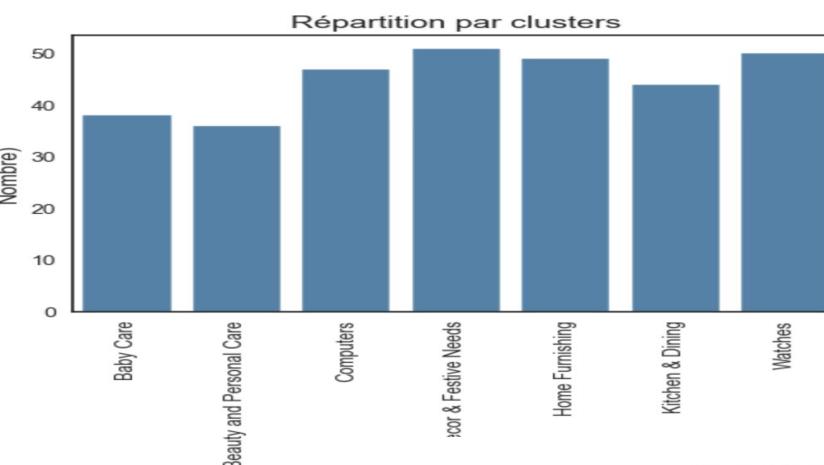


Traitements de textes : Classification Non-Supervisée K MEANS



Traitements de textes: Classification Supervisée Pycaret

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (S)
et	Extra Trees Classifier	0.7267	0.9147	0.7267	0.7374	0.7252	0.6811	0.6833	0.40
rf	Random Forest Classifier	0.6994	0.9127	0.6994	0.7076	0.6982	0.6493	0.6509	0.40
knn	K Neighbors Classifier	0.6912	0.8951	0.6912	0.7077	0.6907	0.6397	0.6424	0.21
catboost	CatBoost Classifier	0.6791	0.9122	0.6791	0.6908	0.6757	0.6257	0.6284	2.25
lightgbm	Light Gradient Boosting Machine	0.6573	0.9089	0.6573	0.6696	0.6570	0.6001	0.6019	1.20
dt	Decision Tree Classifier	0.6436	0.7919	0.6436	0.6592	0.6440	0.5842	0.5862	0.16
xgboost	Extreme Gradient Boosting	0.6395	0.9034	0.6395	0.6438	0.6354	0.5794	0.5812	0.36
gbc	Gradient Boosting Classifier	0.5906	0.8729	0.5906	0.5992	0.5859	0.5223	0.5249	1.25
qda	Quadratic Discriminant Analysis	0.3673	0.7354	0.3673	0.3358	0.3264	0.2619	0.2707	0.1600
nb	Naive Bayes	0.3305	0.7292	0.3305	0.3002	0.2892	0.2193	0.2263	0.1880
ada	Ada Boost Classifier	0.3212	0.7048	0.3212	0.3140	0.3001	0.2079	0.2138	0.2710
lr	Logistic Regression	0.2966	0.6929	0.2966	0.3199	0.2732	0.1797	0.1847	1.2060
lda	Linear Discriminant Analysis	0.2939	0.6924	0.2939	0.3301	0.2741	0.1764	0.1815	0.1800
ridge	Ridge Classifier	0.2477	0.0000	0.2477	0.1427	0.1753	0.1224	0.1339	0.1460
svm	SVM - Linear Kernel	0.2438	0.0000	0.2438	0.1955	0.1876	0.1176	0.1289	0.1630
dummy	Dummy Classifier	0.1361	0.5000	0.1361	0.0185	0.0326	0.0000	0.0000	0.1990



TSNE_ROBERTA_DESC									
	Baby Care	3	0	6	6	2	2		
Baby Care	26	3	0	6	6	2	2		
Beauty and Personal Care	3	25	4	5	2	1	5		
Computers	1	0	38	3	0	1	2		
Home Decor & Festive Needs	2	2	1	31	1	4	4		
Home Furnishing	4	0	1	0	39	0	1		
Kitchen & Dining	2	2	3	2	1	33	2		
Watches	0	4	0	4	0	3	34		
	0	1	2	3	4	5	6		

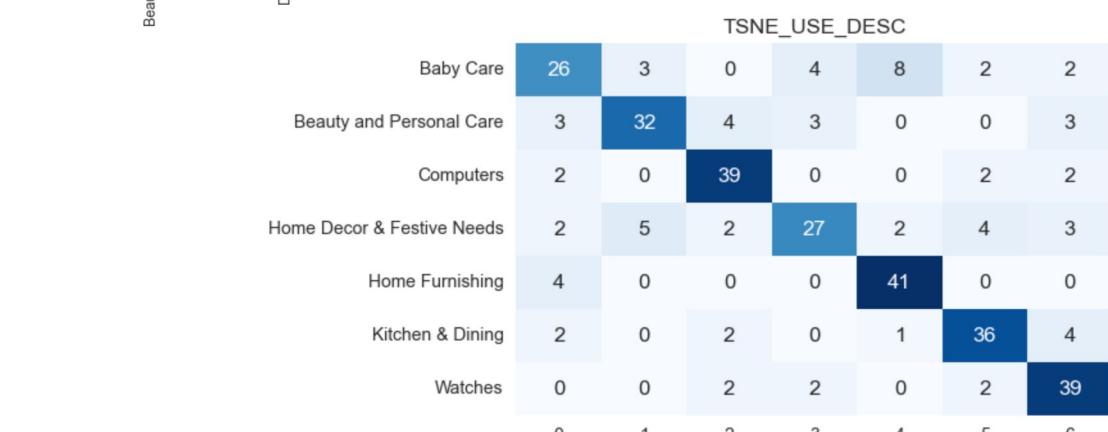
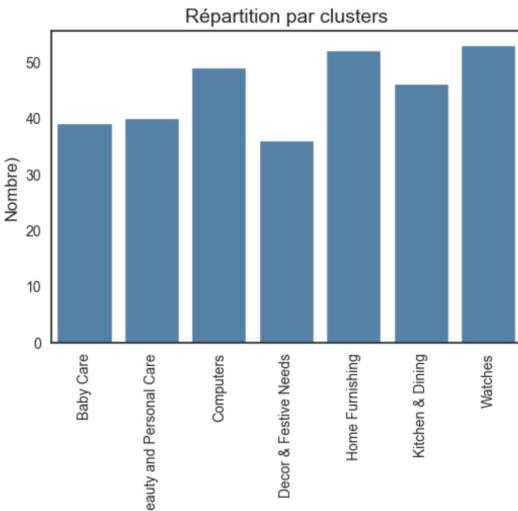


'Précision: 71.75%'

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Extra Trees Classifier	0.7175	0.9098	0	0	0.6704	0.6712	

Traitements de textes: Classification Supervisée Pycaret

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.7702	0.9393	0.7702	0.7802	0.7689	0.7318	0.7340	0.4990
rf	Random Forest Classifier	0.7484	0.9368	0.7484	0.7591	0.7476	0.7063	0.7083	0.5790
xgboost	Extreme Gradient Boosting	0.7430	0.9366	0.7430	0.7559	0.7429	0.7001	0.7021	0.3870
catboost	CatBoost Classifier	0.7389	0.9391	0.7389	0.7500	0.7377	0.6953	0.6974	2.4980
lightgbm	Light Gradient Boosting Machine	0.7267	0.9343	0.7267	0.7398	0.7261	0.6810	0.6831	1.7420
knn	K Neighbors Classifier	0.7185	0.9142	0.7185	0.7354	0.7177	0.6715	0.6746	0.3100
dt	Decision Tree Classifier	0.6967	0.8249	0.6967	0.7209	0.6965	0.6460	0.6496	0.2580
gbc	Gradient Boosting Classifier	0.6899	0.9177	0.6899	0.7023	0.6880	0.6382	0.6406	1.7320
qda	Quadratic Discriminant Analysis	0.4326	0.7728	0.4326	0.3924	0.3910	0.3379	0.3473	0.2610
nb	Naive Bayes	0.4041	0.7657	0.4041	0.3638	0.3669	0.3045	0.3117	0.2260
ada	Ada Boost Classifier	0.3876	0.7442	0.3876	0.3901	0.3642	0.2862	0.2933	0.3750
lr	Logistic Regression	0.2654	0.6807	0.2654	0.2394	0.2356	0.1432	0.1474	0.2750
lda	Linear Discriminant Analysis	0.2491	0.6786	0.2491	0.2296	0.2242	0.1242	0.1273	0.2320
ridge	Ridge Classifier	0.2286	0.0000	0.2286	0.1882	0.1954	0.1004	0.1046	0.2240
svm	SVM - Linear Kernel	0.2123	0.0000	0.2123	0.1658	0.1576	0.0821	0.0898	0.2420
dummy	Dummy Classifier	0.1361	0.5000	0.1361	0.0185	0.0326	0.0000	0.0000	0.2610



	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Extra Trees Classifier	0.7619	0.9194	0	0	0.7222	0.7234	

Précision: 76.19%

Traitements des images : Démarche

Pré-Traitements

Données Images



Originale



Correction Expo



Correction Contraste



Convertir niveaux de gris



Redimensionnement

Features Extractions

Bags Of Words

→ SIFT

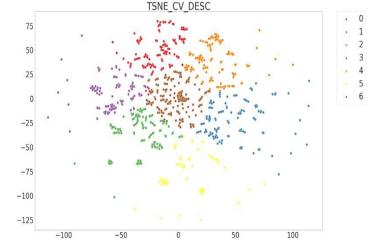


Keras CNN Transfert

- VGG16
- Xception

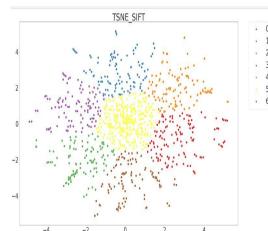
Réduction de dimensions

TSNE



Classification

Classification Non-Supervisée K Means

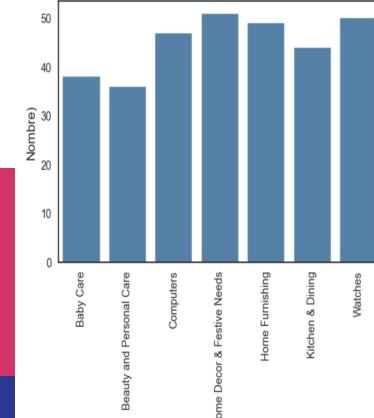


Evaluation

	TSNE_XCEPTION						
Baby Care	26	2	0	5	8	4	0
Beauty and Personal Care	1	33	3	2	5	1	0
Computers	0	0	42	1	1	1	0
Home Decor & Festive Needs	1	0	2	37	3	2	0
Home Furnishing	6	0	1	2	36	0	0
Kitchen & Dining	1	2	1	0	1	40	0
Watches	0	0	0	2	0	0	43
0	0	1	2	3	4	5	6

Évaluation ARI Accuracy

Répartition par clusters



Classification Supervisée : Pycaret

Traitements des images : Analyse Exploratoire

Quelques images

Beauty and Personal Care



Home Decor & Festive Needs

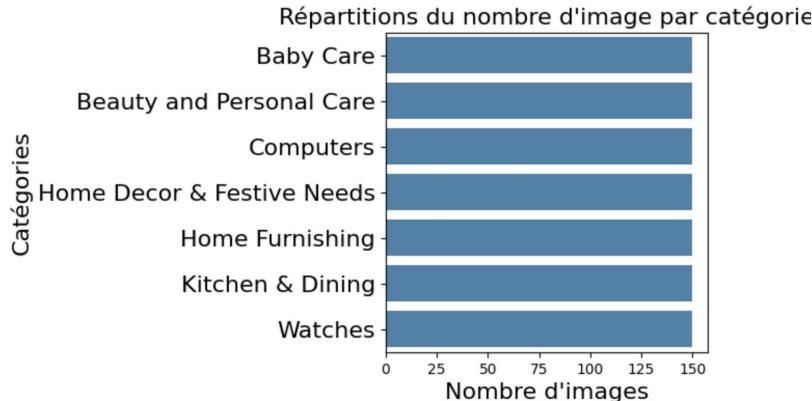


Home Furnishing

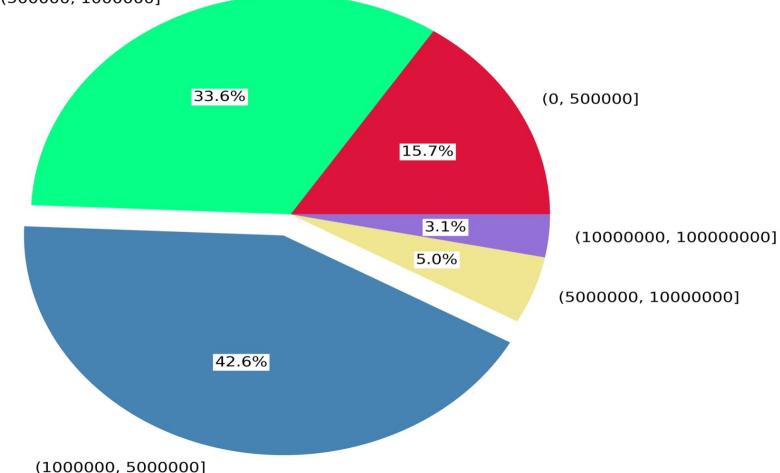


Tailles des images

Plage	nb_données	%_données
(0, 500000]	165	15.714286
(500000, 1000000]	353	33.619048
(1000000, 5000000]	447	42.571429
(5000000, 10000000]	52	4.952381
(10000000, 100000000]	33	3.142857

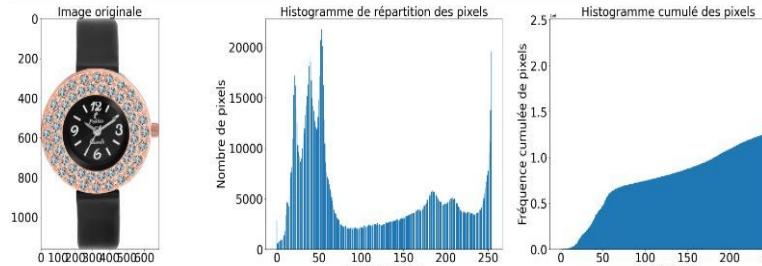


Répartition des tailles des images en pixel par plage

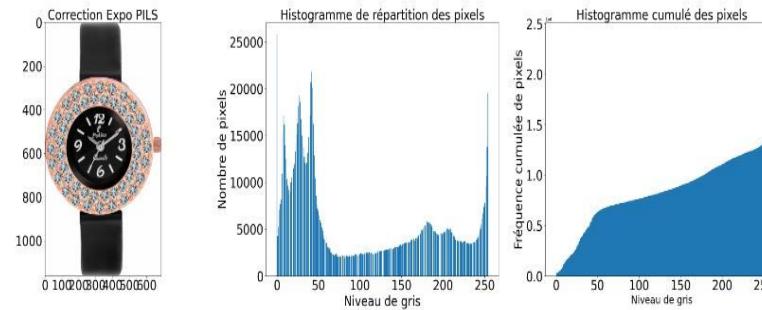


Traitements des images: Pré-traitement

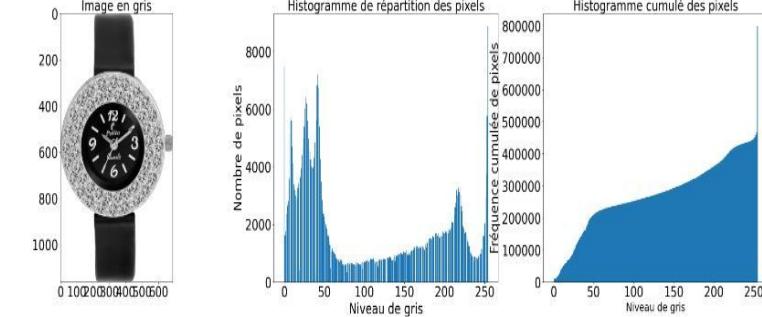
1



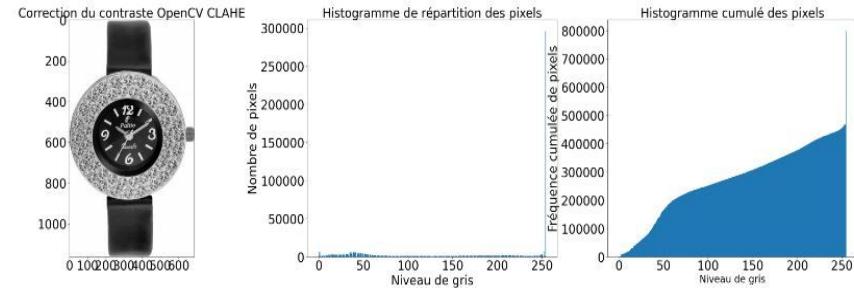
2



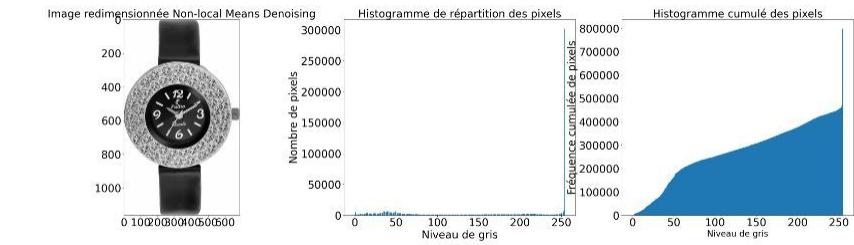
3



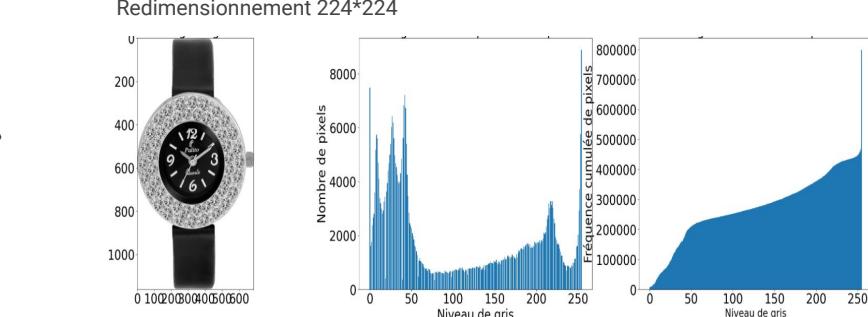
5



6

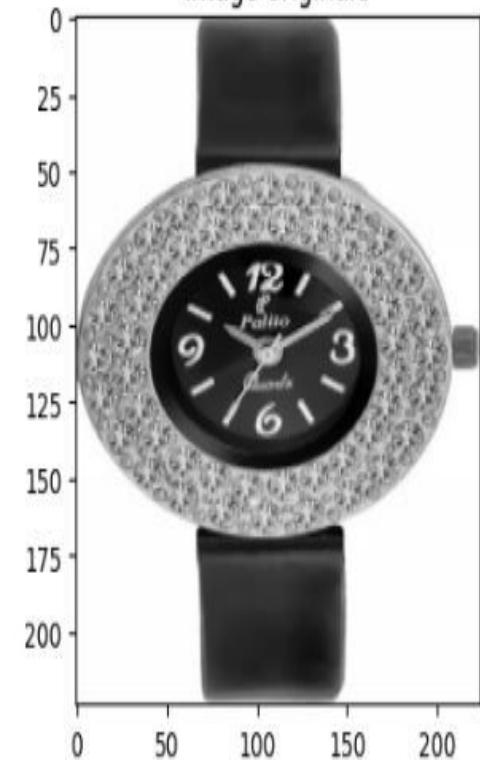


7



Traitements des images: SIFT Visual Words

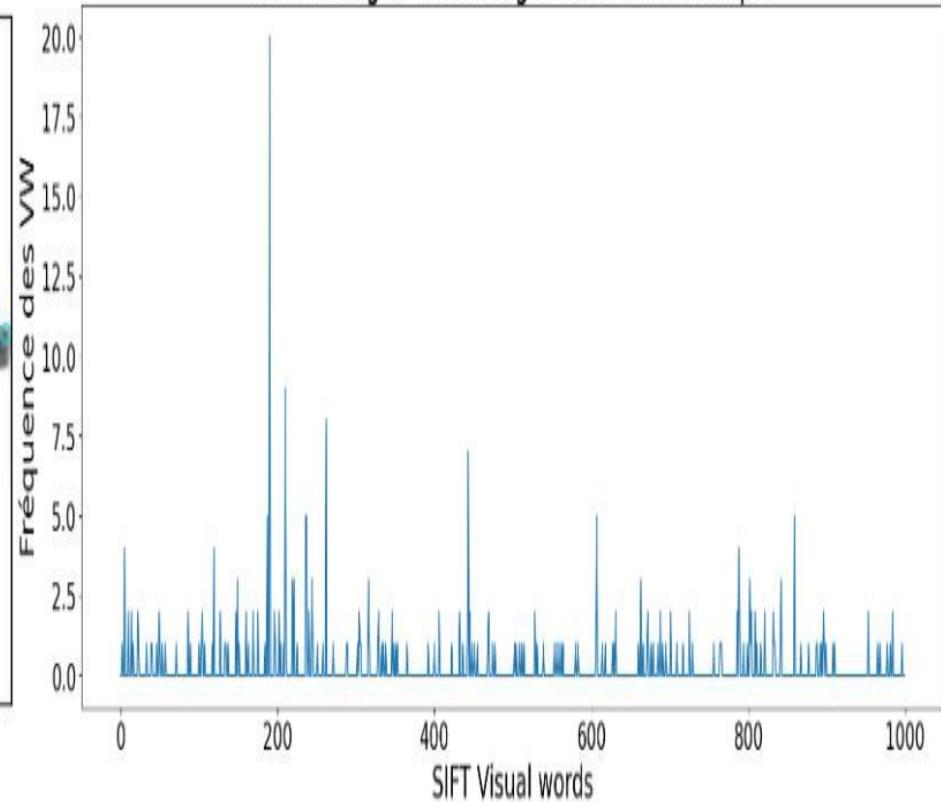
Image originale



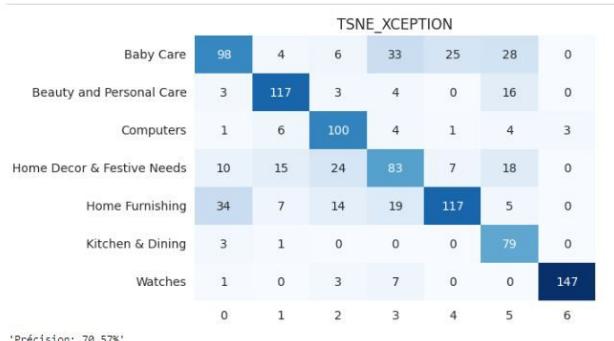
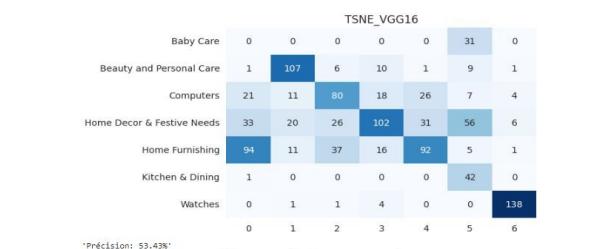
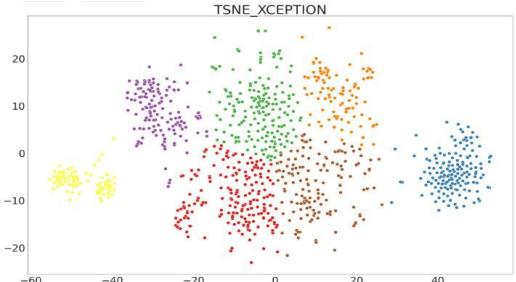
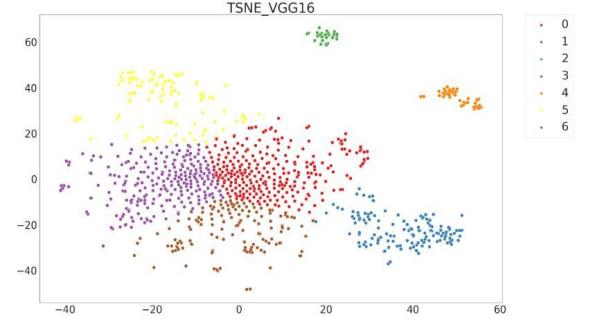
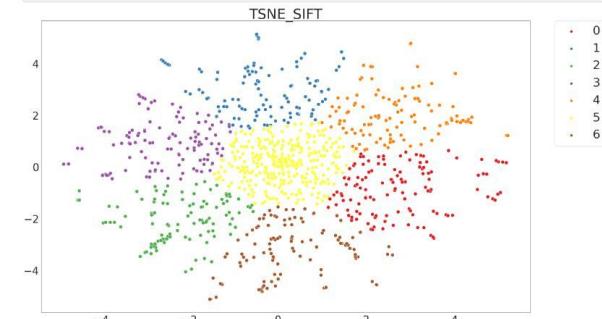
SIFT Points d'interets



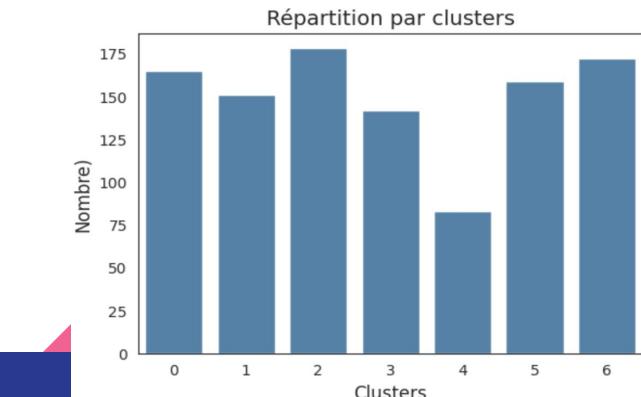
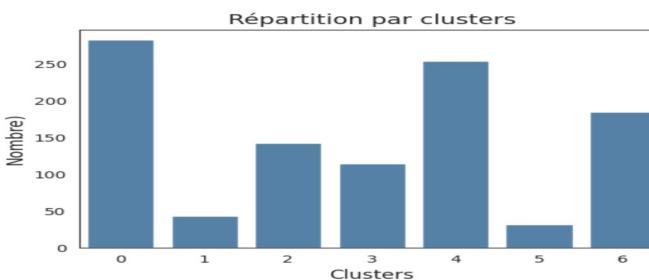
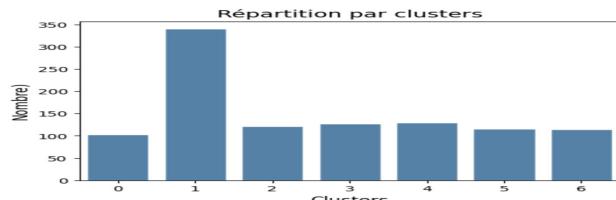
SIFT Histogramme Image Montre en exemple



Traitements des images: KMEANS

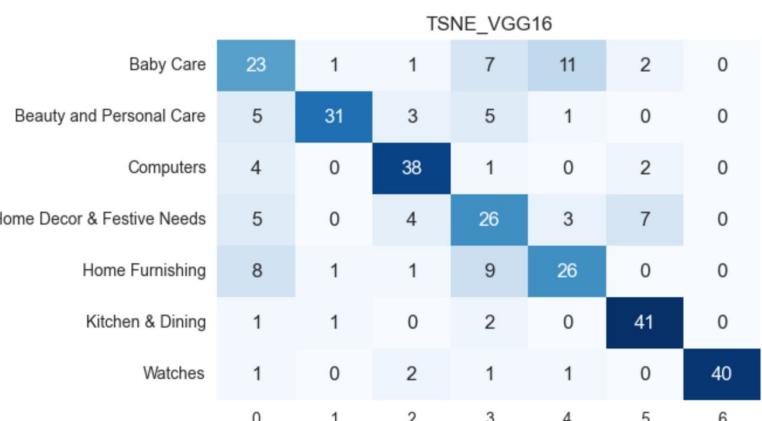
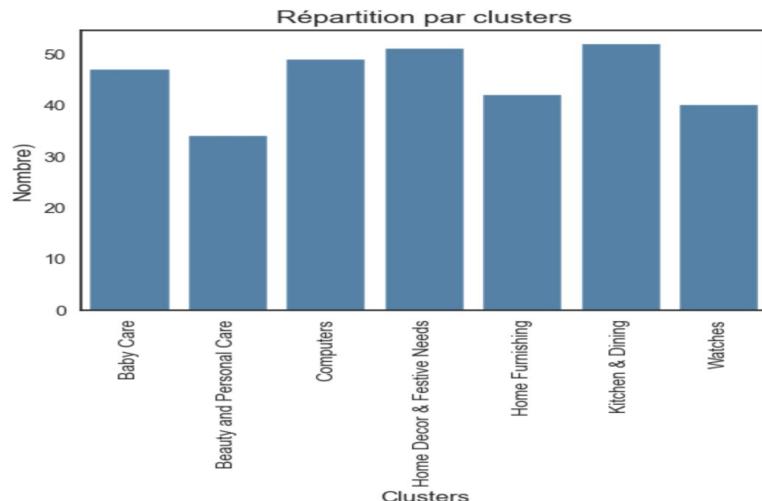


	Unnamed: 0	Type_données	ARI	Homogénéité	Complétude	V-measure
0	0	TSNE_SIFT	0.03551	0.06635	0.07051	0.06837
1	1	TSNE_VGG16	0.30362	0.39839	0.44094	0.41859
2	2	TSNE_XCEPTION	0.45847	0.52119	0.53006	0.52559



Traitements des Images: Classification Supervisée Pycaret

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
rf	Random Forest Classifier	0.7239	0.9191	0.7239	0.7480	0.7222	0.6779	0.6819	0.5910
catboost	CatBoost Classifier	0.7185	0.9266	0.7185	0.7393	0.7160	0.6716	0.6757	2.5580
et	Extra Trees Classifier	0.7144	0.9200	0.7144	0.7295	0.7108	0.6667	0.6702	0.4820
xgboost	Extreme Gradient Boosting	0.6941	0.9216	0.6941	0.7007	0.6903	0.6430	0.6453	0.5060
lightgbm	Light Gradient Boosting Machine	0.6940	0.9251	0.6940	0.7100	0.6887	0.6430	0.6470	1.5210
knn	K Neighbors Classifier	0.6927	0.8946	0.6927	0.7132	0.6926	0.6415	0.6447	0.2800
gbc	Gradient Boosting Classifier	0.6763	0.9123	0.6763	0.7009	0.6756	0.6222	0.6262	1.5580
dt	Decision Tree Classifier	0.6750	0.8105	0.6750	0.6899	0.6697	0.6208	0.6250	0.2840
nb	Naive Bayes	0.6191	0.8667	0.6191	0.6549	0.6193	0.5556	0.5607	0.2950
qda	Quadratic Discriminant Analysis	0.6191	0.8756	0.6191	0.6428	0.6160	0.5556	0.5602	0.2700
lr	Logistic Regression	0.5606	0.8293	0.5606	0.5655	0.5539	0.4874	0.4902	0.2660
lda	Linear Discriminant Analysis	0.5442	0.8291	0.5442	0.5350	0.5305	0.4682	0.4713	0.2310
ridge	Ridge Classifier	0.4067	0.0000	0.4067	0.3735	0.3296	0.3082	0.3334	0.2340
ada	Ada Boost Classifier	0.3414	0.7143	0.3414	0.3444	0.3172	0.2318	0.2470	0.3930
svm	SVM - Linear Kernel	0.3375	0.0000	0.3375	0.3399	0.3041	0.2279	0.2428	0.2670
dummy	Dummy Classifier	0.1361	0.5000	0.1361	0.0185	0.0326	0.0000	0.0000	0.2330
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	
0	Random Forest Classifier	0.7143	0.9173	0	0	0	0.6667	0.6677	



'Précision: 71.43%'

Traitements des images: Classification Supervisée Pycaret

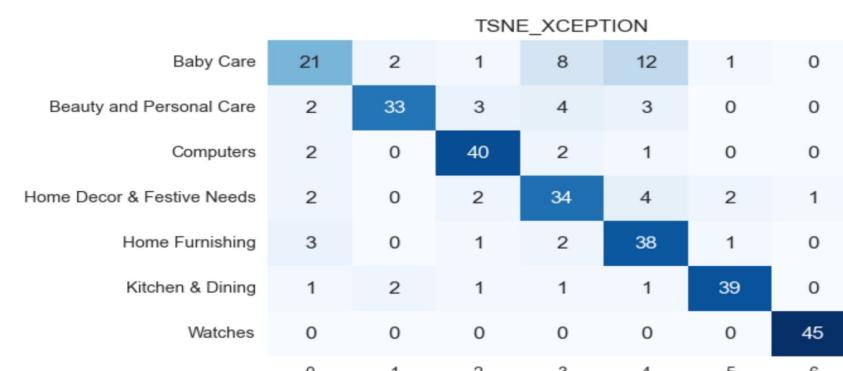
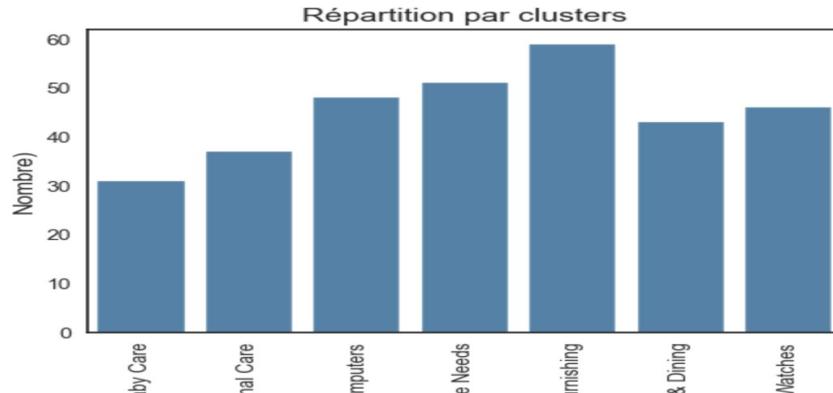
Amélioration des hyperparamètres de vgg16 avec : Epochs et batch_size

Accuracy sur les données tests de 77,19%

```
Epoch 1/10
6/6 [=====] - 2s 134ms/step - loss: 71.5828 - accuracy: 0.2859 - val_loss: 9.5168 - val_accuracy: 0.5856
Epoch 2/10
6/6 [=====] - 0s 73ms/step - loss: 3.4406 - accuracy: 0.7776 - val_loss: 9.6377 - val_accuracy: 0.6654
Epoch 3/10
6/6 [=====] - 0s 78ms/step - loss: 1.0122 - accuracy: 0.9072 - val_loss: 5.0429 - val_accuracy: 0.7376
Epoch 4/10
6/6 [=====] - 1s 98ms/step - loss: 0.2613 - accuracy: 0.9644 - val_loss: 4.8091 - val_accuracy: 0.7643
Epoch 5/10
6/6 [=====] - 0s 75ms/step - loss: 0.0286 - accuracy: 0.9949 - val_loss: 4.9635 - val_accuracy: 0.7681
Epoch 6/10
6/6 [=====] - 1s 99ms/step - loss: 0.0292 - accuracy: 0.9987 - val_loss: 5.2317 - val_accuracy: 0.7643
Epoch 7/10
6/6 [=====] - 0s 88ms/step - loss: 0.0170 - accuracy: 0.9987 - val_loss: 5.0888 - val_accuracy: 0.7643
Epoch 8/10
6/6 [=====] - 0s 78ms/step - loss: 4.7405e-04 - accuracy: 1.0000 - val_loss: 5.0657 - val_accuracy: 0.7719
Epoch 9/10
6/6 [=====] - 0s 66ms/step - loss: 8.2794e-05 - accuracy: 1.0000 - val_loss: 5.0692 - val_accuracy: 0.7719
Epoch 10/10
6/6 [=====] - 0s 76ms/step - loss: 6.9002e-05 - accuracy: 1.0000 - val_loss: 5.0722 - val_accuracy: 0.7719
```

Traitements des images : Classification Supervisée Pycaret

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.8476	0.9601	0.8476	0.8538	0.8446	0.8223	0.8243	0.5000
rf	Random Forest Classifier	0.8245	0.9577	0.8245	0.8336	0.8207	0.7953	0.7981	0.5800
catboost	CatBoost Classifier	0.8191	0.9664	0.8191	0.8273	0.8163	0.7889	0.7911	3.2110
knn	K Neighbors Classifier	0.8150	0.9470	0.8150	0.8236	0.8121	0.7842	0.7866	0.3280
xgboost	Extreme Gradient Boosting	0.8137	0.9641	0.8137	0.8229	0.8114	0.7826	0.7849	0.4100
lightgbm	Light Gradient Boosting Machine	0.8041	0.9609	0.8041	0.8136	0.8012	0.7715	0.7741	3.1520
gbc	Gradient Boosting Classifier	0.8001	0.9568	0.8001	0.8092	0.7983	0.7667	0.7688	1.8050
dt	Decision Tree Classifier	0.7891	0.8771	0.7891	0.8091	0.7874	0.7539	0.7583	0.4330
qda	Quadratic Discriminant Analysis	0.7470	0.9255	0.7470	0.7668	0.7410	0.7048	0.7097	0.2700
nb	Naive Bayes	0.7416	0.9228	0.7416	0.7570	0.7356	0.6985	0.7027	0.4150
lr	Logistic Regression	0.7076	0.8964	0.7076	0.7237	0.7021	0.6589	0.6633	0.4370
lda	Linear Discriminant Analysis	0.7061	0.8928	0.7061	0.7316	0.6992	0.6573	0.6631	0.2570
svm	SVM - Linear Kernel	0.4910	0.0000	0.4910	0.5274	0.4675	0.4063	0.4261	0.2940
ridge	Ridge Classifier	0.4910	0.0000	0.4910	0.3825	0.4093	0.4064	0.4296	0.3430
ada	Ada Boost Classifier	0.3592	0.7491	0.3592	0.3509	0.3118	0.2522	0.2732	0.3800
dummy	Dummy Classifier	0.1361	0.5000	0.1361	0.0185	0.0326	0.0000	0.0000	0.3770
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	
0	Extra Trees Classifier	0.7937	0.9423	0	0	0	0.7593	0.7615	



'Précision: 79.37%'

Test API : Protocole

- **API**

<https://rapidapi.com/edamam/api/edamam-food-and-grocery-database>

- **Code:** Snippet python fourni par cet API
- **Clé:** Fourni par l'API après avoir effectué une inscription (gratuit)
- **Collecte des produits:** À base de champagne
- **Label des produits:** foodId, label, category, foodContentsLabel, image

Unnamed: 0	foodId	label	category	foodContentsLabel	image
0	food_a656mk2a5dmqb2adiamu6beihduu	Champagne	Generic foods	NaN	https://www.edamam.com/food-img/a71/a718cf3c52...
1	food_b753ithamdb8psbt0w2k9aquo0c	Champagne Vinaigrette, Champagne	Packaged foods	OLIVE OIL; BALSAMIC VINEGAR; CHAMPAGNE VINEGAR...	NaN
2	food_b3dyababj054xobm68jzbghigqe	Champagne Vinaigrette, Champagne	Packaged foods	INGREDIENTS: WATER; CANOLA OIL; CHAMPAGNE VINE...	https://www.edamam.com/food-img/d88/d88b64d973...
3	food_a8e0ghsamvoc45bwa2ybsa3gken9	Champagne Vinaigrette, Champagne	Packaged foods	CANOLA AND SOYBEAN OIL; WHITE WINE (CONTAINS S...	NaN
4	food_an4jjeaucpus2a3u1ni8auhe7q9	Champagne Vinaigrette, Champagne	Packaged foods	WATER; CANOLA AND SOYBEAN OIL; WHITE WINE (CON...	NaN
5	food_bmu5dmkazwuvpaa5prh1daa8jxs0	Champagne Dressing, Champagne	Packaged foods	SOYBEAN OIL; WHITE WINE (PRESERVED WITH SULFIT...	https://www.edamam.com/food-img/ab2/ab2459f2a...
6	food_a1p44taoyv11ra0lic1qa8xculi	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_am5egz6aq3fjplaf8xpkdbc2asis	Champagne Truffles	Generic meals	butter; cocoa; sweetened condensed milk; vanil...	NaN
9	food_bc28rhiajk1fuvadvkfmekbouc0	Champagne Vinaigrette	Generic meals	champagne vinegar; olive oil; Dijon mustard; s...	NaN

Le respect de la RGPD pour la collecte de données via l'API

- ❖ **Licéité, loyauté et transparence:** Le droit de récupérer ces données a été accordé par l'API, et la transparence de celles-ci fut établie quant à son utilisation (lors de l'inscription).
- ❖ **Limitation des finalités:** Ce projet se concentre sur la collecte de données de certains aliments (champagne) et non toutes.
- ❖ **Minimisation des données:** Respectée car filtrage des données pour ne récupérer que les champs nécessaires (foodId, label, category, foodContentsLabel, image).
- ❖ **Exactitude:** Étant donné que les données ont été récupérées directement depuis l'API, on peut supposer qu'elles sont exactes.
- ❖ **Limitation de la conservation:** Ces données ne seront conservées que pour la durée du projet.

Conclusion et suite du projet



Meilleur Résultat pour le traitement des textes : TSNE_USE avec Description Stémmatisée et Lémmatisée

Accuracy_test	Accuracy_train
76,19%	77,02%

Meilleur Résultat pour le traitement des Images pré-traitées : TSNE_Xception

Accuracy_test	Accuracy_train
79,37%	84,76%

Jeu de données

- ❖ Ajouter des images (150 seulement par catégorie).
- ❖ Combiner les deux meilleurs algorithmes pour les textes et les images

Étiquettes des catégories

- ❖ Nombreuses erreurs entre les catégories Home Furnishing et Home Decor & Festive Needs.
- ❖ Vérifier les étiquettes avant d'envisager d'utiliser l'apprentissage supervisé