

PROJET 6

Classifiez automatiquement des biens de consommation

Projected sales of main products in 2013

Distribution of market share among the major industry players

Distribution of market share among the major industry players: IT & C and BN & T was 74% and 26% percent respectively. A further change in the economic situation in the market will be characterized by a more equal distribution of market share major players

Share of market activity

Changes in the activity of the active and passive market is characterized by a more equal distribution of market share major players

Projected sales of main products in 2013

Passive market share

Problématique

Vous êtes Data Scientist au sein de l'entreprise "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.

Pour rendre l'expérience utilisateur des vendeurs (faciliter la mise en ligne de nouveaux articles) et des acheteurs (faciliter la recherche de produits) la plus fluide possible et dans l'optique d'un passage à l'échelle, il devient nécessaire d'automatiser cette tâche.

Linda, lead data scientist, vous demande donc d'étudier la **faisabilité** d'un moteur de classification des articles en différentes catégories, avec un niveau de précision suffisant.



Données et Mission

Linda vous a fourni un premier jeu de données d'articles avec le lien pour télécharger la photo et une description associée.

Votre mission est de réaliser une première 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.

Vous analyserez le jeu de données en réalisant un prétraitement des images et des descriptions des produits, une réduction de dimension, puis un clustering.

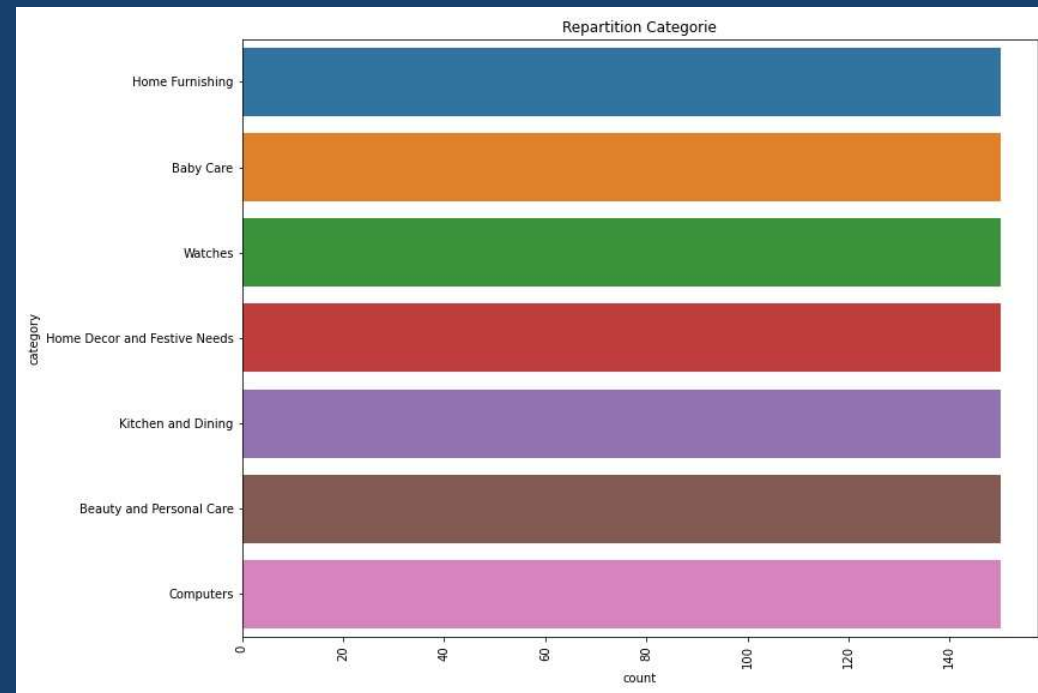
Les résultats du clustering seront présentés sous la forme d'une représentation en deux dimensions à déterminer, qui illustrera le fait que les caractéristiques extraites permettent de regrouper des produits de même catégorie.

Axes de recherche

- deux type de données : texte et images
- apprentissage non supervisé et supervisé
- affichage 2D et 3D des clusters trouvés
- conserve 20% des données pour obtenir un score et comparer les algorithmes

Analyse Exploratoire Données

- 1050 lignes
- 7 catégories uniformément réparties, il faut donc obtenir une prédiction supérieure à $1/7$ soit 0,143
- pas de données manquantes pour description nom produit et image



Traitement du TEXTE

- concaténation du Nom du produit et de la description
- passage en minuscule
- tokenisation du texte
- suppression de la ponctuation et des mots courants
- comparaison lemmatisation et stemmatisation
- comparaison Bag Of Words et TF-IDF

Traitement du TEXTE

1479 caracteres : 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 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

#####

MINUSCULE

1479 caracteres : 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 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

#####

TOKENIZE

259 mots : ['elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', 'key', 'features', 'of', 'elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', 'floral', 'c', 'urtain', ',', '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', 'an', 'd', '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']

Traitement du TEXTE

STOPWORDS

```
169 mots : ['elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', 'key', 'features', 'elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', 'floral', 'curtain', 'elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', '213', 'cm', 'height', 'pack', '2', 'price', 'rs', '899', 'curtain', 'enhances', 'look', 'interiors', 'curtain', 'made', '100', 'high', 'quality', 'polyester', 'fabric', 'features', 'eyelet', 'style', 'stitch', 'metal', 'ring', 'makes', 'room', 'environment', 'romantic', 'loving', 'curtain', 'ant-', 'wrinkle', 'anti', 'shrinkage', 'elegant', 'apparance', 'give', 'home', 'bright', 'modernistic', 'appeal', 'designs', 'surreal', 'attention', 'sure', 'steal', 'hearts', 'contemporary', 'eyelet', 'valance', 'curtains', 'slide', 'smoothly', 'draw', 'apart', 'first', 'thing', 'morning', 'welcome', 'bright', 'sun', 'rays', 'want', 'wish', 'good', 'morning', 'whole', 'world', 'draw', 'close', 'evening', 'create', 'special', 'moments', 'joyous', 'beauty', 'given', 'soothing', 'prints', 'bring', 'home', 'elegant', 'curtain', 'softly', 'filters', 'light', 'room', 'get', 'right', 'amount', 'sunlight.', 'specifications', 'elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', '213', 'cm', 'height', 'pack', '2', 'general', 'brand', 'elegance', 'designed', 'door', 'type', 'eyelet', 'model', 'name', 'abstract', 'polyester', 'door', 'curtain', 'set', '2', 'model', 'id', 'duster25', 'color', 'multicolor', 'dimensions', 'length', '213', 'cm', 'box', 'number', 'contents', 'sales', 'package', 'pack', '2', 'sales', 'package', '2', 'curtains', 'body', 'design', 'material', 'polyester']
```

LEMMATIZE

```
169 mots : ['elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', 'key', 'feature', 'elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', 'floral', 'curtain', 'elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', '213', 'cm', 'height', 'pack', '2', 'price', 'r', '899', 'curtain', 'enhance', 'look', 'interior', 'curtain', 'make', '100', 'high', 'quality', 'polyester', 'fabric', 'feature', 'eyelet', 'style', 'stitch', 'metal', 'ring', 'make', 'room', 'environment', 'romantic', 'love', 'curtain', 'ant-', 'wrinkle', 'anti', 'shrinkage', 'elegant', 'apparance', 'give', 'home', 'bright', 'modernistic', 'appeal', 'design', 'surreal', 'attention', 'sure', 'steal', 'heart', 'contemporary', 'eyelet', 'valance', 'curtain', 'slide', 'smoothly', 'draw', 'apart', 'first', 'thing', 'morning', 'welcome', 'bright', 'sun', 'ray', 'want', 'wish', 'good', 'morning', 'whole', 'world', 'draw', 'close', 'even', 'create', 'special', 'moment', 'joyous', 'beauty', 'give', 'soothe', 'print', 'bring', 'home', 'elegant', 'curtain', 'softly', 'filter', 'light', 'room', 'get', 'right', 'amount', 'sunlight.', 'specification', 'elegance', 'polyester', 'multicolor', 'abstract', 'eyelet', 'door', 'curtain', '213', 'cm', 'height', 'pack', '2', 'general', 'brand', 'elegance', 'design', 'door', 'type', 'eyelet', 'model', 'name', 'abstract', 'polyester', 'door', 'curtain', 'set', '2', 'model', 'id', 'duster25', 'color', 'multicolor', 'dimension', 'length', '213', 'cm', 'box', 'number', 'content', 'sale', 'package', 'pack', '2', 'sale', 'package', '2', 'curtain', 'body', 'design', 'material', 'polyester']
```


Traitement du TEXTE

Comparaison lemmatisation et stemmatisation

MOT	LEMMATIZE	STEM PORTER	STEM LANCASTE
elegance	elegance	eleg	eleg
polyester	polyester	polyest	polyest
multicolor	multicolor	multicolor	multicol
abstract	abstract	abstract	abstract
eyelet	eyelet	eyelet	eyelet
door	door	door	door
curtain	curtain	curtain	curtain
key	key	key	key
features	feature	featur	feat
elegance	elegance	eleg	eleg
polyester	polyester	polyest	polyest
multicolor	multicolor	multicolor	multicol
abstract	abstract	abstract	abstract
eyelet	eyelet	eyelet	eyelet
door	door	door	door
curtain	curtain	curtain	curtain
floral	floral	floral	flor
curtain	curtain	curtain	curtain
elegance	elegance	eleg	eleg
polyester	polyester	polyest	polyest
multicolor	multicolor	multicolor	multicol
abstract	abstract	abstract	abstract
eyelet	eyelet	eyelet	eyelet
door	door	door	door
curtain	curtain	curtain	curtain
213	213	213	213
cm	cm	cm	cm
height	height	height	height
pack	pack	pack	pack
2	2	2	2
price	price	price	pric
rs	r	rs	rs
899	899	899	899
curtain	curtain	curtain	curtain
enhances	enhance	enhanc	enh
look	look	look	look

Traitement du TEXTE

Bag of Words

Exemple première ligne jeu de données avec 2-grams (bigrams)

1479 mots : 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 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

	100	100 high	213	213 cm	899	899 curtain	abstract	abstract eyelet	abstract polyester	amount	amount sunlight	ant	ant wrinkle	anti	anti shrinkage	apart	apart first	apparance	apparance give	appeal	appeal design	attention	attention sure	beauty	beauty give	body	body design
0	1	1	3	3	1	1	5	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

Traitement du TEXTE

TF-IDF (Term-Frequency - Inverse Document Frequency)

Exemple première ligne jeu de données avec 2-grams (bigrams)

1479 mots : 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 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

	100	100 high	213	213 cm	899	899 curtain	abstract	abstract eyelet	abstract polyester	amount	amount sunlight	ant	ant wrinkle	anti	anti shrinkage	apart	apart first	apparence	apparence give	appeal	appeal design	attention
0	0.025906	0.045392	0.126035	0.130465	0.034741	0.048076	0.137352	0.181569	0.048076	0.043488	0.048076	0.043488	0.048076	0.037424	0.045392	0.045392	0.048076	0.048076	0.048076	0.036218	0.048076	0.042012

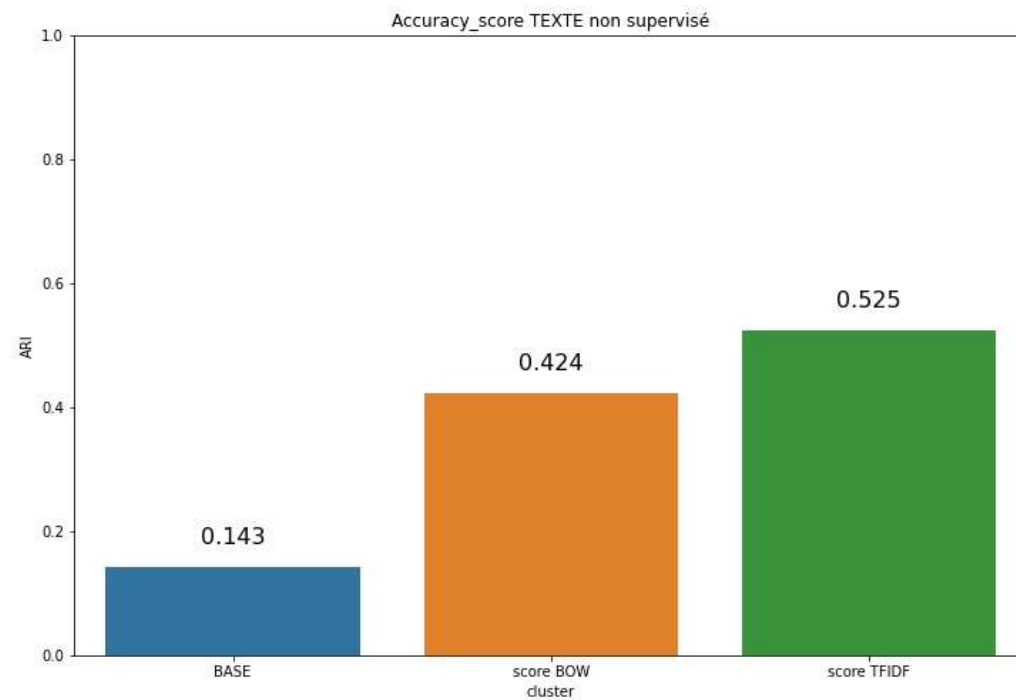
Traitement du TEXTE

Score accuracy non supervisé

BASE : 0,143

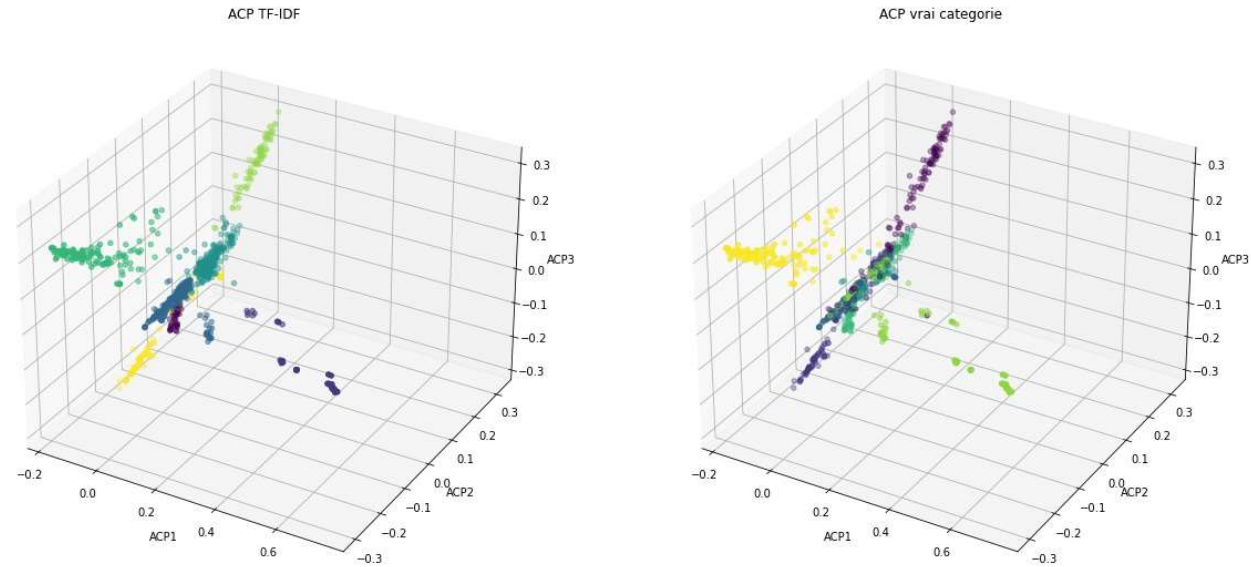
BOW : 0,424

Tf-itdf : 0,525

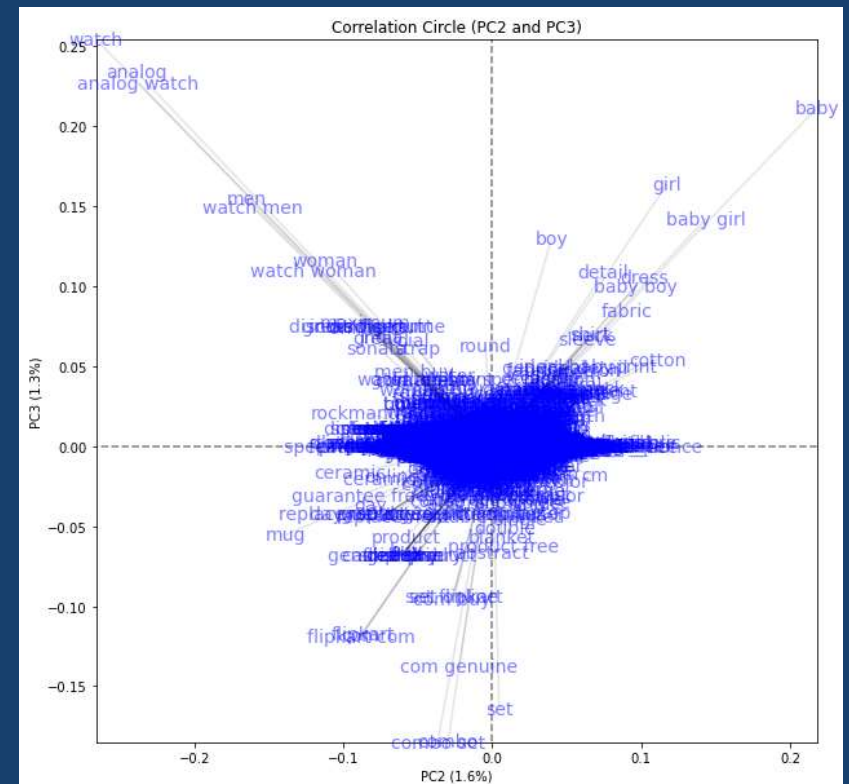


Traitement du TEXTE

Affichage 3D
après PCA

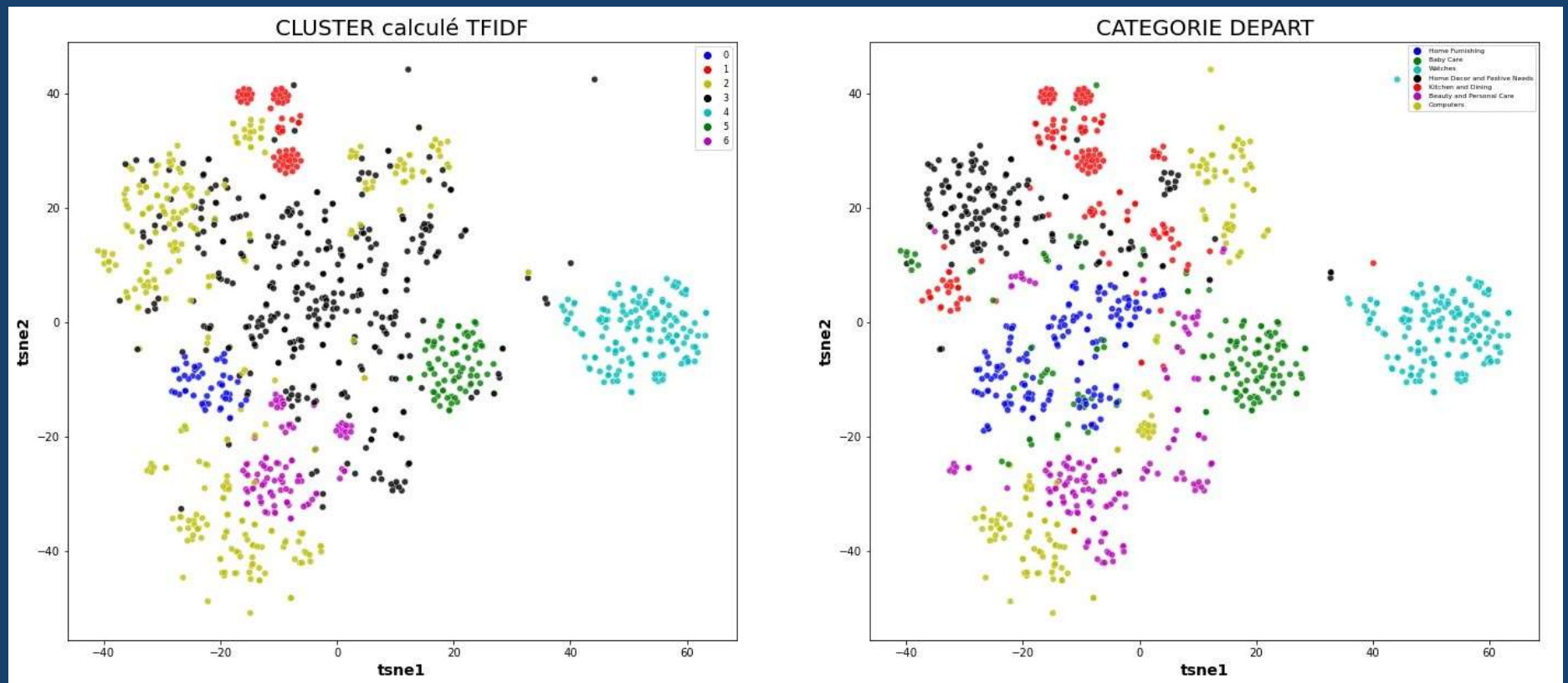


Affichage 3D après PCA



Traitement du TEXTE

Affichage 2D
après tSNE



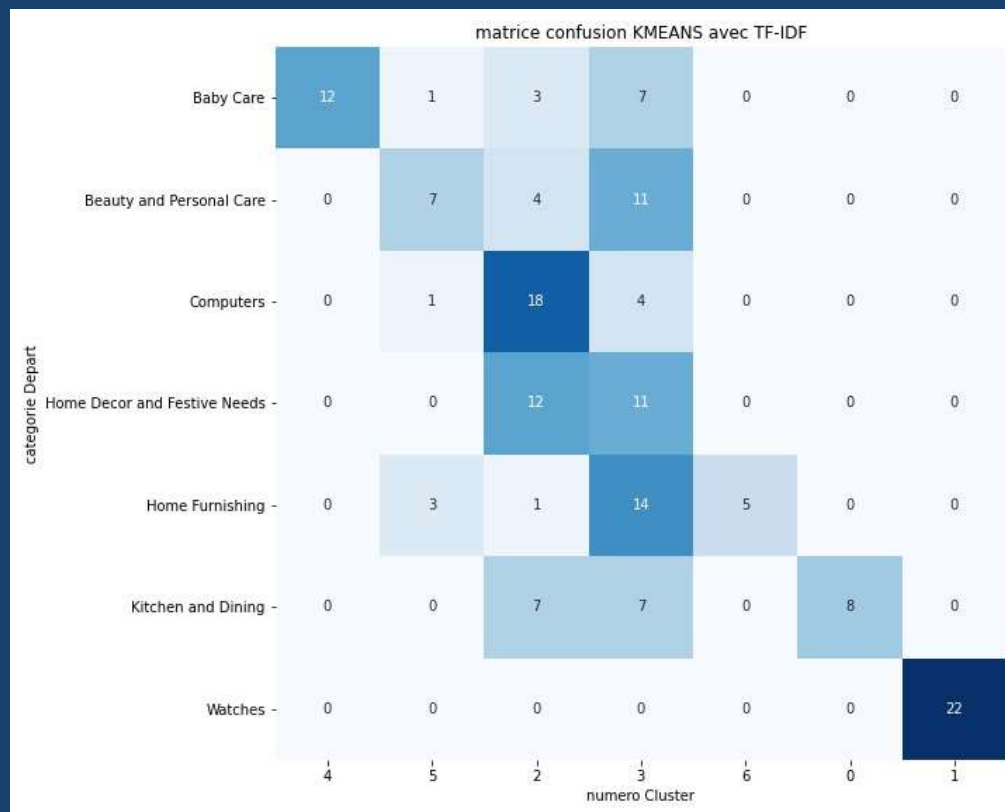
Traitement du TEXTE

Matrice confusion

Tf-idf

Accuracy_score 0,525

	precision	recall	f1-score
0	1.00	0.52	0.69
1	0.58	0.32	0.41
2	0.40	0.78	0.53
3	0.20	0.48	0.29
4	1.00	0.22	0.36
5	1.00	0.36	0.53
6	1.00	1.00	1.00
accuracy			0.53



Traitement du TEXTE

Prediction amazon.co.uk

```
'prediction_watches' : ['BENYAR Mens Watches Analogue Quartz Chronograph Date 3ATM Waterproof Watch Business Sport Leather Strap Wrist Watch for Men'],  
'prediction_furnishing' : ['Classic Sofa Chairs Lazy Lounger Bean Bag Cover without Fillings for Home Garden Lounge Living Room'],  
'prediction_kitchen' : ['Premier Housewares 1900261 Clear Bin Bags For Recycling Storage Biodegradable Kitchen Bin Set Recycling Box Small Bin Bags Pack of 3 bags 31 x 30 x 44'],  
'prediction_computers' : ['Complete set of 21.5in Monitor and HP 800G1 Quad Core i5-4570 8GB 240GB SSD WiFi Windows 10 64-Bit Desktop PC Computer (Renewed)'],  
'prediction_festive' : ['Inventox Water Beads 20000 Pcs Multipurpose Magic Balls, Grow upto 15mm in Water, Ideal for Sensory Play, Home Decor, Vase Filler, Multi Color Set with 05 Balloons'],  
'prediction_beauty' : ['Garnier Ultimate Blends Hair Food Aloe 3-in-1, Moisturising Hair Mask, Conditioning Treatment, Leave-in Conditioner for Normal Hair, Vegan Formula, 98% Natural Ingredients, No Silicones, 390 ml'],  
'prediction_baby' : ['NimNik Baby Sling Wrap Carrier - Natural Cotton Multi Position Soft Sling for Newborns Infants from Birth Grey']
```

```
prediction_watches ['BENYAR Mens Watches Analogue Quartz Chronograph Date 3ATM Waterproof Watch Business Sport Leather Strap Wrist Watch for Men']  
cluster_predit : [5]  
  
prediction_furnishing ['Classic Sofa Chairs Lazy Lounger Bean Bag Cover without Fillings for Home Garden Lounge Living Room']  
cluster_predit : [0]  
  
prediction_kitchen ['Premier Housewares 1900261 Clear Bin Bags For Recycling Storage Biodegradable Kitchen Bin Set Recycling Box Small Bin Bags Pack of 3 bags 31 x 30 x 44']  
cluster_predit : [6]  
  
prediction_computers ['Complete set of 21.5in Monitor and HP 800G1 Quad Core i5-4570 8GB 240GB SSD WiFi Windows 10 64-Bit Desktop PC Computer (Renewed)']  
cluster_predit : [3]  
  
prediction_festive ['Inventox Water Beads 20000 Pcs Multipurpose Magic Balls, Grow upto 15mm in Water, Ideal for Sensory Play, Home Decor, Vase Filler, Multi Color Set with 05 Balloons']  
cluster_predit : [6]  
  
prediction_beauty ['Garnier Ultimate Blends Hair Food Aloe 3-in-1, Moisturising Hair Mask, Conditioning Treatment, Leave-in Conditioner for Normal Hair, Vegan Formula, 98% Natural Ingredients, No Silicones, 390 ml']  
cluster_predit : [6]  
  
prediction_baby ['NimNik Baby Sling Wrap Carrier - Natural Cotton Multi Position Soft Sling for Newborns Infants from Birth Grey']  
cluster_predit : [6]
```

Traitement du TEXTE

- Apprentissage supervisé, on commence par un PCA en gardant 99% variance cela nous permet de passer de 26077 features à 959
- Essai OneVsAll DecisionTree, XGBOOST et RandomForest
- Essai SVM
- meilleur score pour SVM avec kernel rbf (0,92)

Traitement du TEXTE

Prediction amazon.co.uk

```
'prediction_watches' : ['BENYAR Mens Watches Analogue Quartz Chronograph Date 3ATM Waterproof Watch Business Sport Leather Strap Wrist Watch for Men'],  
'prediction_furnishing' : ['Classic Sofa Chairs Lazy Lounger Bean Bag Cover without Fillings for Home Garden Lounge Living Room'],  
'prediction_kitchen' : ['Premier Housewares 1900261 Clear Bin Bags For Recycling Storage Biodegradable Kitchen Bin Set Recycling Box Small Bin Bags Pack of 3 bags 31 x 30 x 44'],  
'prediction_computers' : ['Complete set of 21.5in Monitor and HP 800G1 Quad Core i5-4570 8GB 240GB SSD WiFi Windows 10 64-Bit Desktop PC Computer (Renewed)'],  
'prediction_festive' : ['Inventox Water Beads 20000 Pcs Multipurpose Magic Balls, Grow upto 15mm in Water, Ideal for Sensory Play, Home Decor, Vase Filler, Multi Color Set with 05 Balloons'],  
'prediction_beauty' : ['Garnier Ultimate Blends Hair Food Aloe 3-in-1, Moisturising Hair Mask, Conditioning Treatment, Leave-in Conditioner for Normal Hair, Vegan Formula, 98% Natural Ingredients, 390 ml'],  
'prediction_baby' : ['NimNik Baby Sling Wrap Carrier - Natural Cotton Multi Position Soft Sling for Newborns Infants from Birth Grey']
```

```
prediction_watches ['BENYAR Mens Watches Analogue Quartz Chronograph Date 3ATM Waterproof Watch Business Sport Leather Strap Wrist Watch for Men']  
categorie predite : ['Watches']
```

```
prediction_furnishing ['Classic Sofa Chairs Lazy Lounger Bean Bag Cover without Fillings for Home Garden Lounge Living Room']  
categorie predite : ['Home Furnishing']
```

```
prediction_kitchen ['Premier Housewares 1900261 Clear Bin Bags For Recycling Storage Biodegradable Kitchen Bin Set Recycling Box Small Bin Bags Pack of 3 bags 31 x 30 x 44']  
categorie predite : ['Kitchen and Dining']
```

```
prediction_computers ['Complete set of 21.5in Monitor and HP 800G1 Quad Core i5-4570 8GB 240GB SSD WiFi Windows 10 64-Bit Desktop PC Computer (Renewed)']  
categorie predite : ['Computers']
```

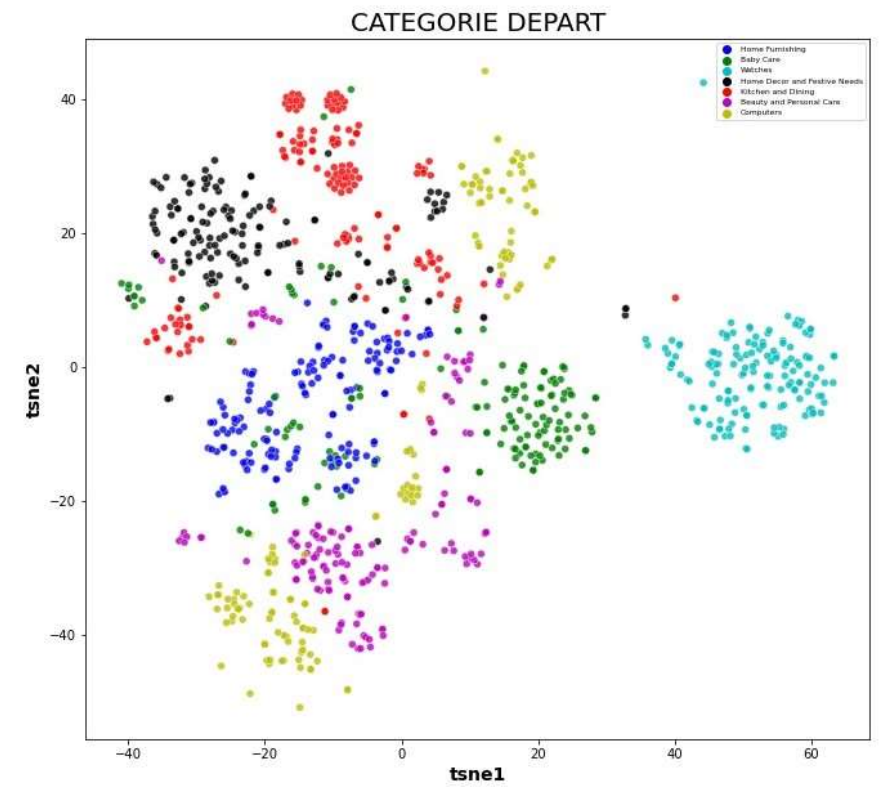
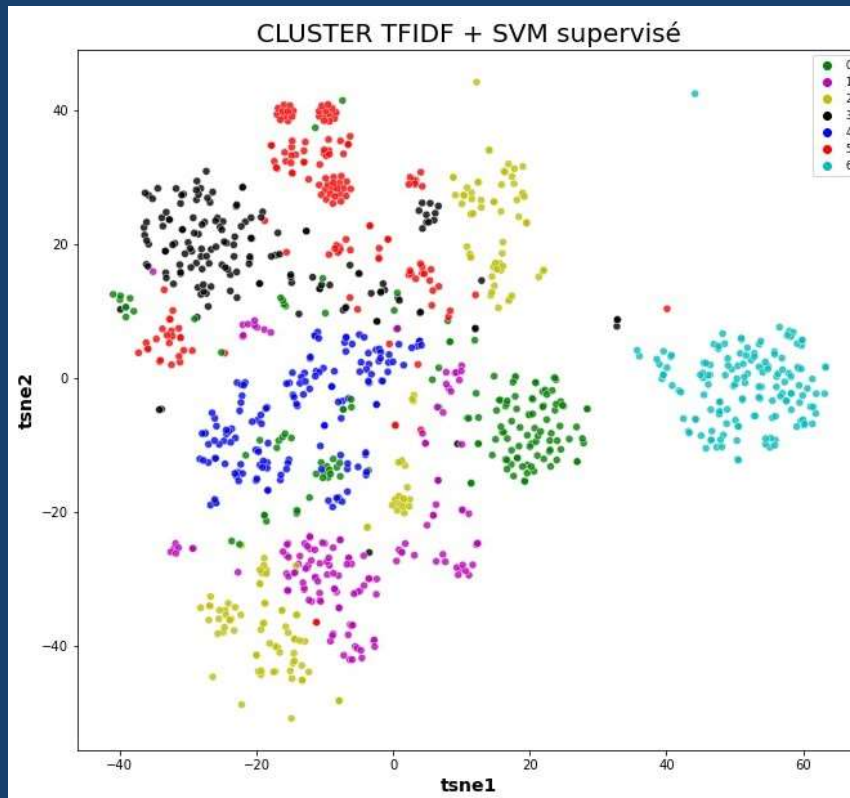
```
prediction_festive ['Inventox Water Beads 20000 Pcs Multipurpose Magic Balls, Grow upto 15mm in Water, Ideal for Sensory Play, Home Decor, Vase Filler, Multi Color Set with 05 Balloons']  
categorie predite : ['Baby Care']
```

```
prediction_beauty ['Garnier Ultimate Blends Hair Food Aloe 3-in-1, Moisturising Hair Mask, Conditioning Treatment, Leave-in Conditioner for Normal Hair, Vegan Formula, 98% Natural Ingredients, No Silicones, 390 ml']  
categorie predite : ['Beauty and Personal Care']
```

```
prediction_baby ['NimNik Baby Sling Wrap Carrier - Natural Cotton Multi Position Soft Sling for Newborns Infants from Birth Grey']  
categorie predite : ['Baby Care']
```

Traitement du TEXTE

Affichage 2D
après tSNE



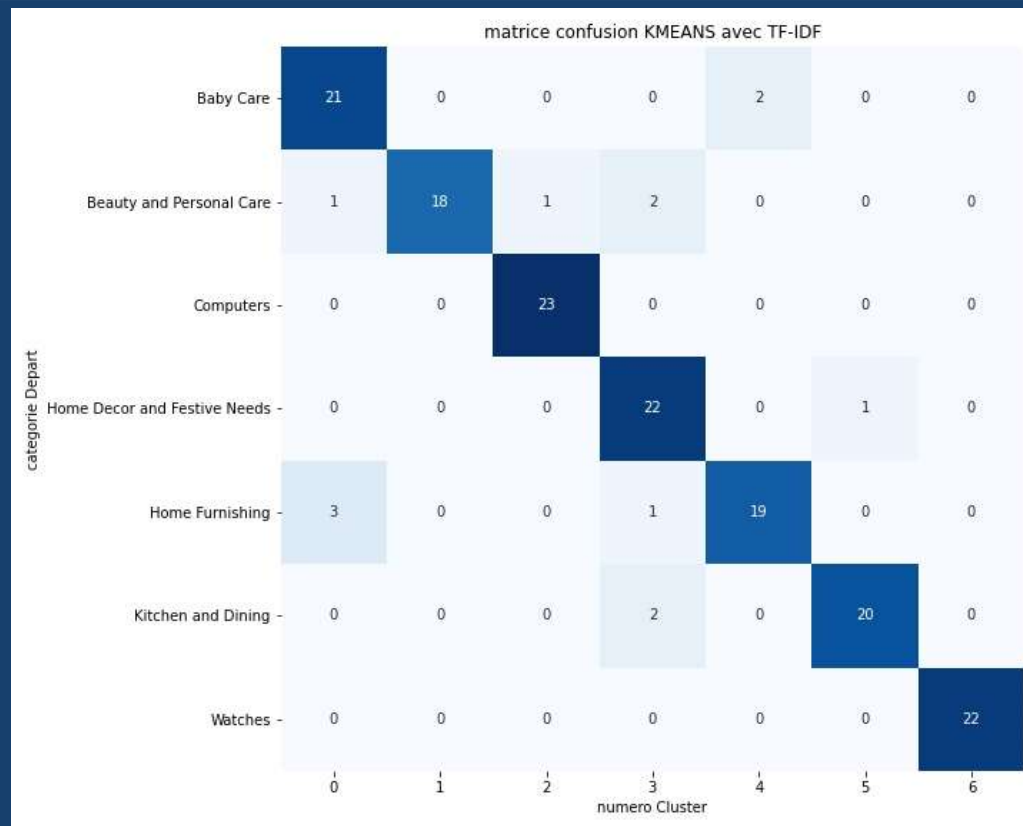
Traitement du TEXTE

Matrice confusion

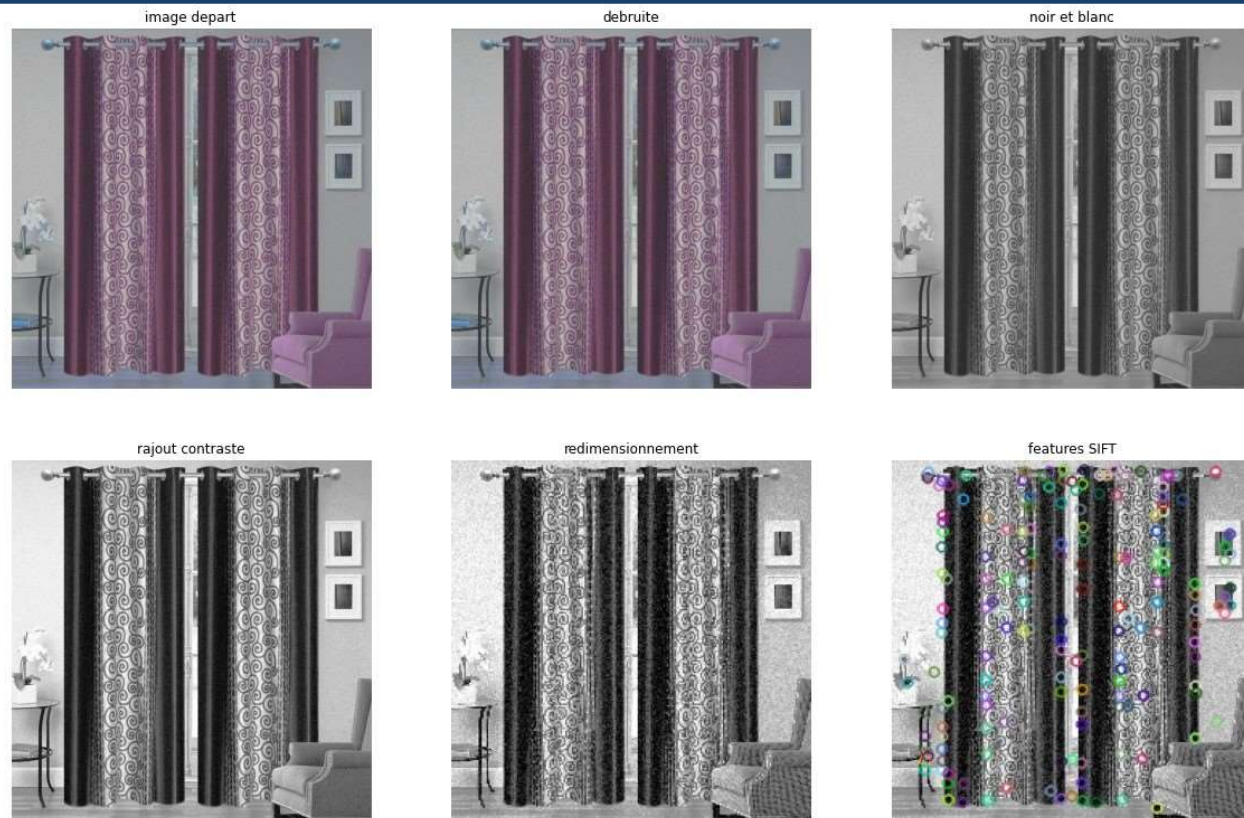
Tf-idf

Accuracy_score 0,918

	precision	recall	f1-score
0	0.84	0.91	0.87
1	1.00	0.82	0.90
2	0.96	1.00	0.98
3	0.81	0.96	0.88
4	0.90	0.83	0.86
5	0.95	0.91	0.93
6	1.00	1.00	1.00
accuracy			0.92



Traitement des IMAGES



Traitement des IMAGES

image depart



debruite



noir et blanc



rajout contraste



redimensionnement



features SIFT



Traitement des IMAGES



Traitement des IMAGES

Créations des descripteurs de chaque image

- Pour chaque image passage en gris et equalisation
- création d'une liste de descripteurs par image qui sera utilisée pour réaliser les histogrammes
- création d'une liste de descripteurs pour l'ensemble des images qui sera utilisé pour créer les clusters de descripteurs

Création des clusters de descripteur

Pour chaque image :

- prédiction des numéros de cluster de chaque descripteur
- création d'un histogramme = comptage pour chaque numéro de cluster du nombre de descripteurs de l'image

Features d'une image = Histogramme d'une image = Comptage pour une image du nombre de descripteurs par cluster

Utilisation d'une ACP gardant 99% de variance (543 à 431 features)

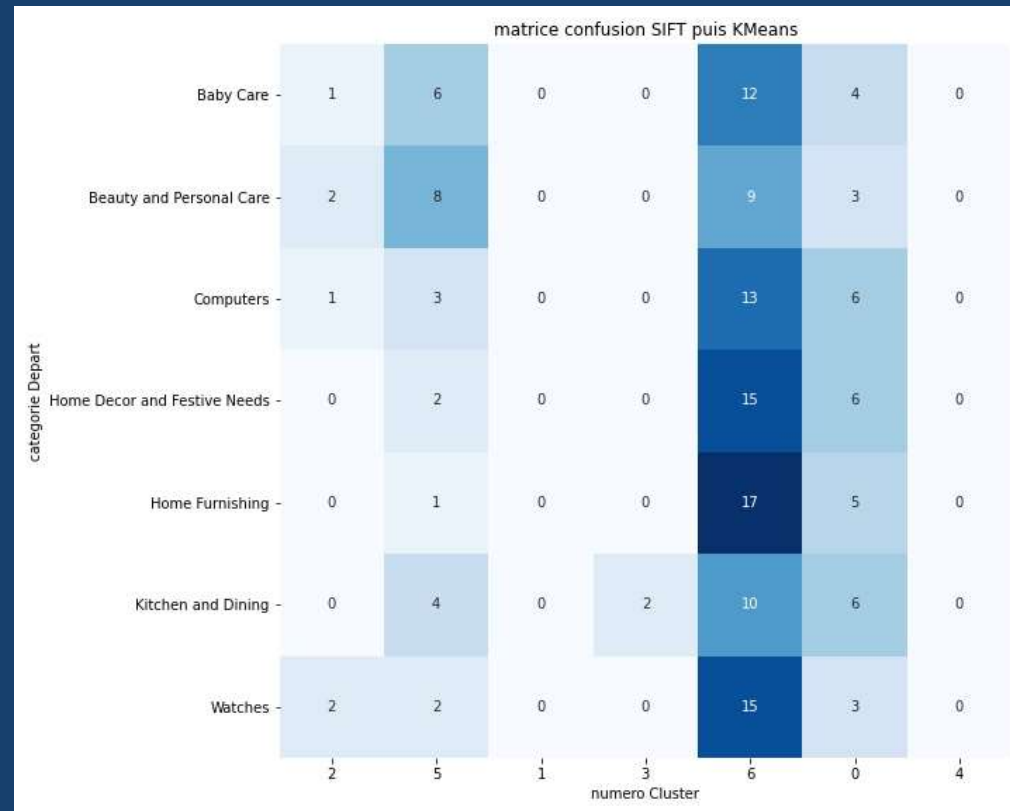
Traitement des IMAGES

Matrice confusion

SIFT

Accuracy_score 0,203

	precision	recall	f1-score
0	0.17	0.04	0.07
1	0.31	0.36	0.33
2	0.00	0.00	0.00
3	0.00	0.00	0.00
4	0.19	0.74	0.30
5	0.18	0.27	0.22
6	0.00	0.00	0.00
accuracy			0.20



Traitement des IMAGES

Essai en supervisé après réduction ACP, utilisation d'un SVM avec un kernel poly



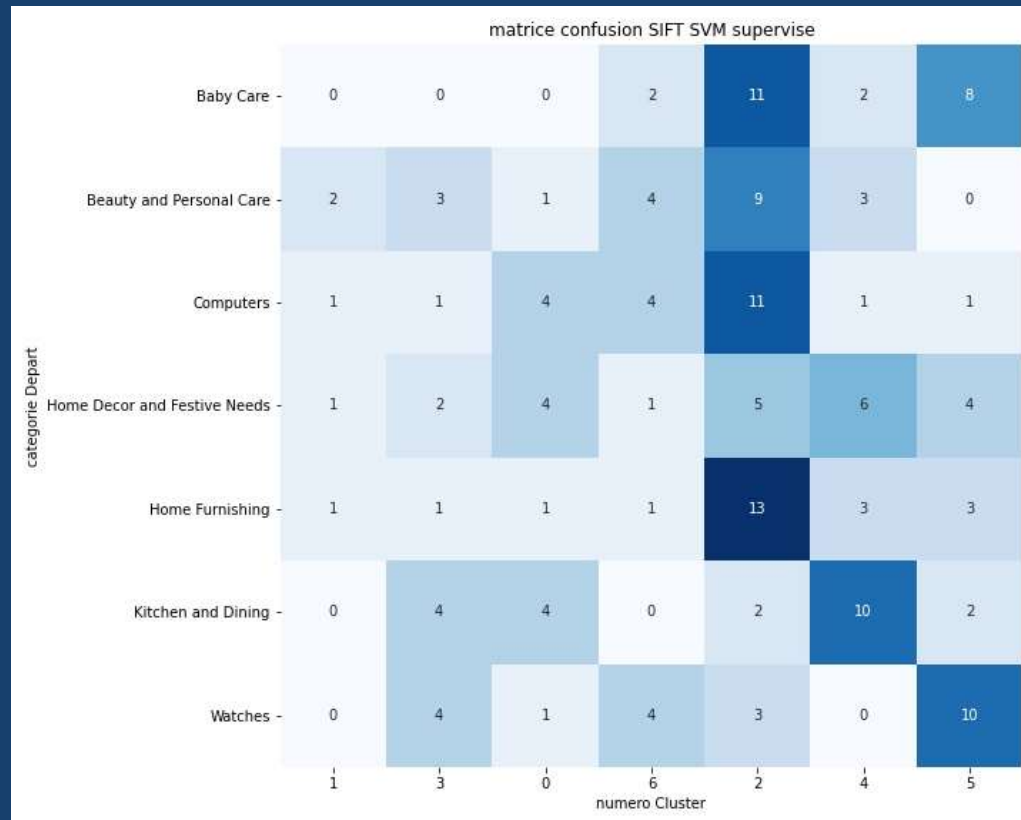
Traitement des IMAGES

Matrice confusion

SIFT supervisé

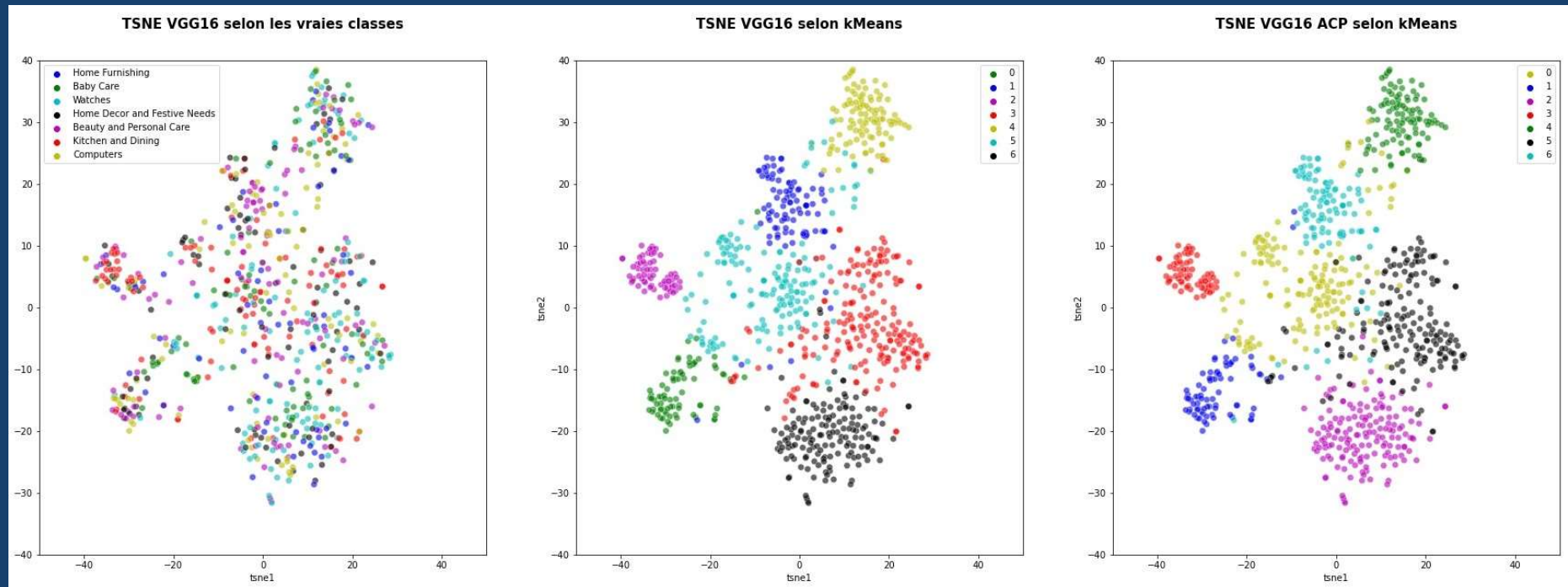
Accuracy_score 0,259

	precision	recall	f1-score
0	0.00	0.00	0.00
1	0.20	0.14	0.16
2	0.27	0.17	0.21
3	0.06	0.04	0.05
4	0.24	0.57	0.34
5	0.40	0.45	0.43
6	0.36	0.45	0.40
accuracy			0.26



Traitement des IMAGES

transfer learning avec un réseau VGG16 avec extraction des features, ACP et KMEANS.



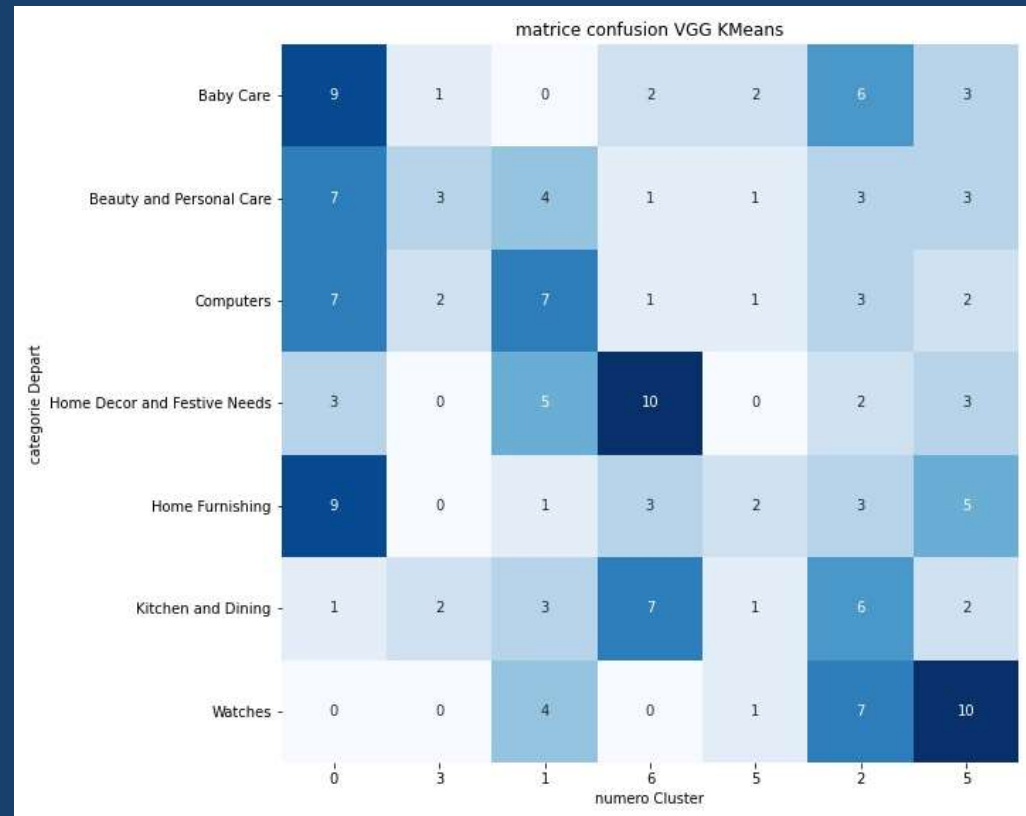
Traitement des IMAGES

Matrice confusion

VGG extraction features + Kmeans

Accuracy_score 0,297

	precision	recall	f1-score
0	0.25	0.39	0.31
1	0.38	0.14	0.20
2	0.29	0.30	0.30
3	0.42	0.43	0.43
4	0.25	0.09	0.13
5	0.20	0.27	0.23
6	0.36	0.45	0.40
accuracy			0.30

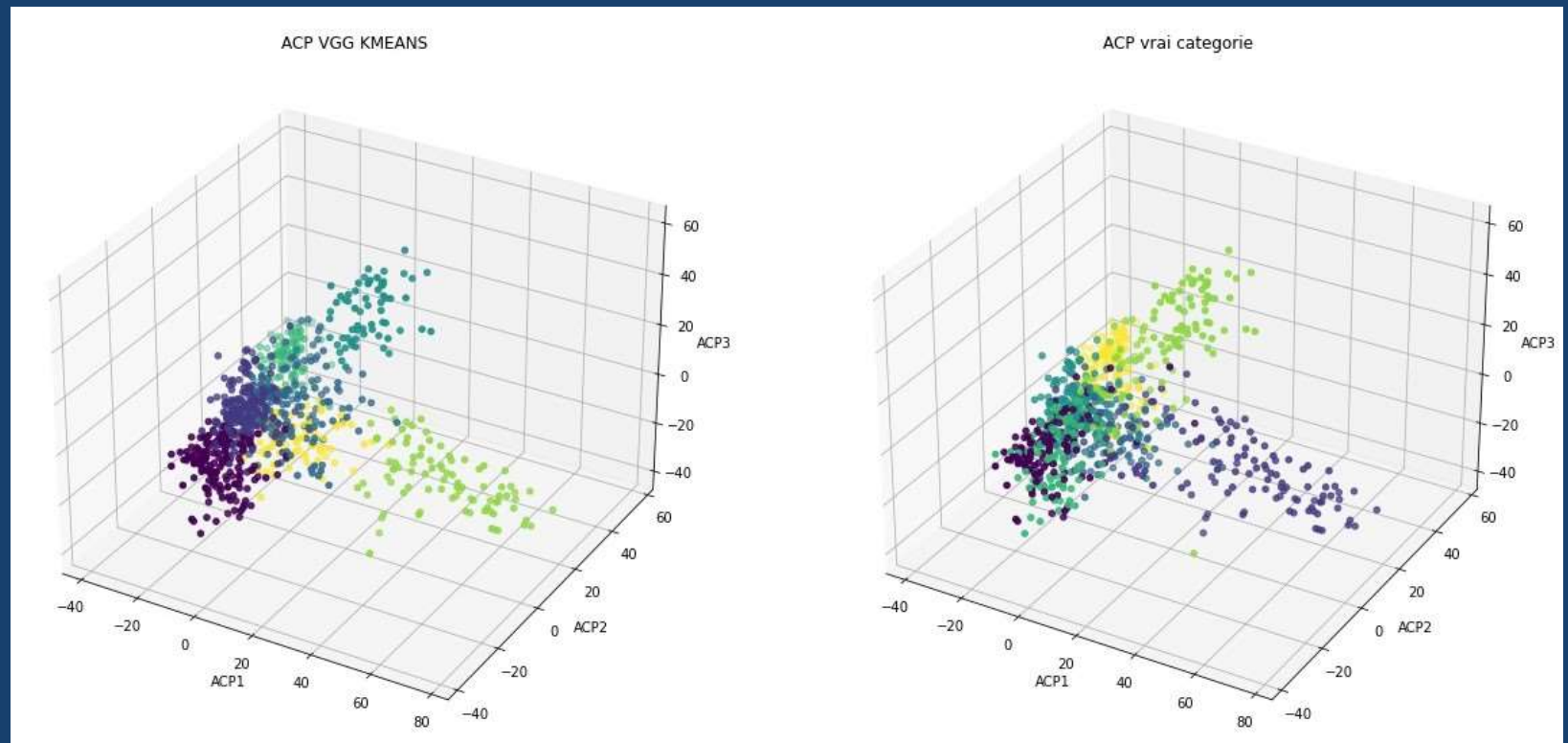


Traitement des IMAGES

Matrice confusion

VGG extraction features
+ Kmeans

Accuracy_score 0,297



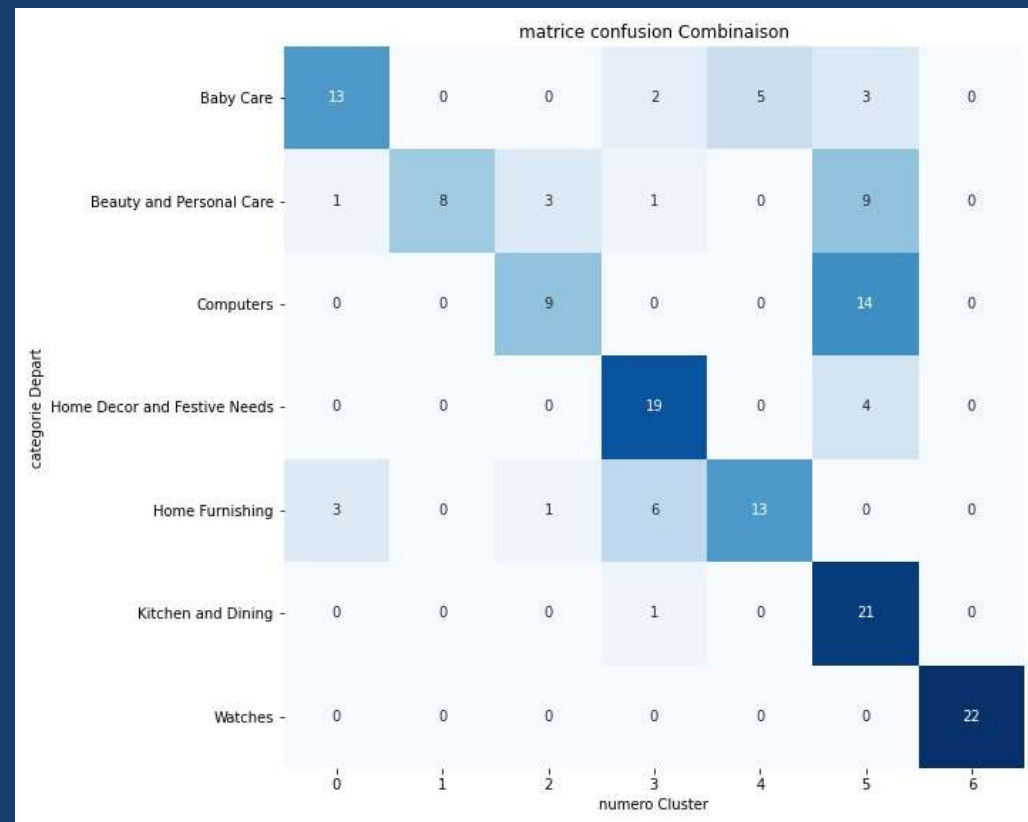
Combinaison Texte Images

Matrice confusion

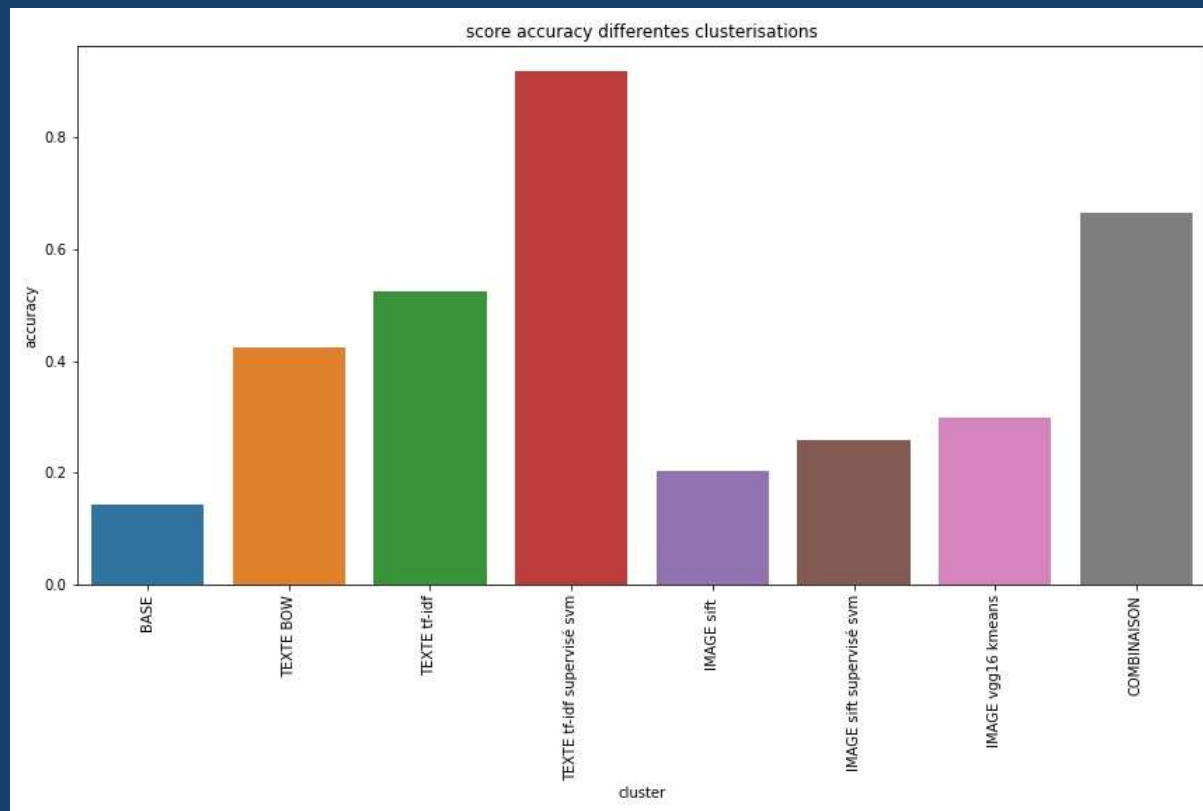
texte TF-IDF + image VGG Kmeans

Accuracy_score 0,665

	precision	recall	f1-score
0	0.76	0.57	0.65
1	1.00	0.36	0.53
2	0.69	0.39	0.50
3	0.66	0.83	0.73
4	0.72	0.57	0.63
5	0.41	0.95	0.58
6	1.00	1.00	1.00
accuracy			0.66



Combinaison Texte Images





MERCI

Questions et Réponses