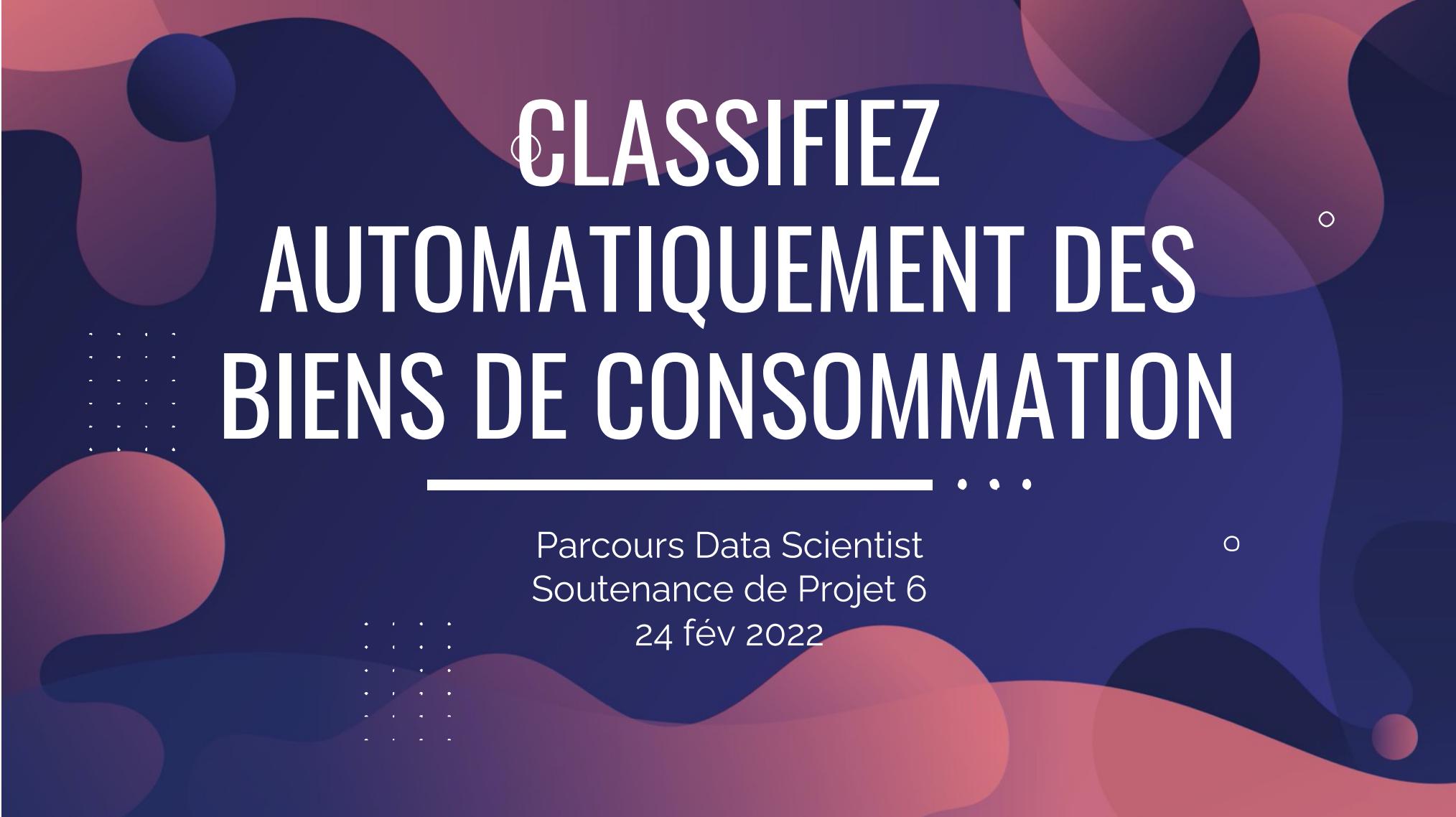


CLASSIFIEZ AUTOMATIQUEMENT DES BIENS DE CONSOMMATION



Parcours Data Scientist
Soutenance de Projet 6

24 fév 2022

• TABLE OF CONTENTS

01

Contexte

02

Données

03

Extraction Features NLP

04

Extraction Features Image

05

Extraction Features Combinées

06

Conclusion



| 01

• CONTEXTE



PLACE DE MARCHÉ

Automatiser l'attribution de la catégorie d'un article pour rendre l'expérience utilisateur des vendeur la plus fluide possible et dans l'optique d'un passage à l'échelle.

-
-
-

02

DONNÉES

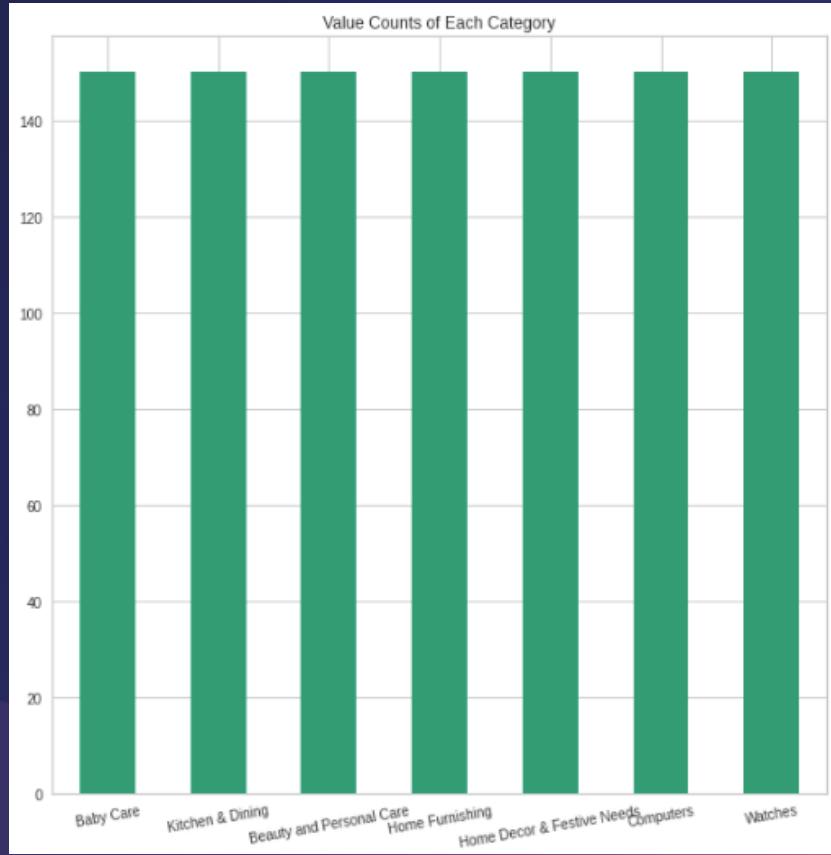




DONNÉES

- Images
- Description de produits
- Catégories
- hiérarchie de produits

CATÉGORIES DE PRODUITS



03

Extraction Features NLP



TEXTES ORIGINALES

'Key Features of Elegance Polyester Multicolor Abstract Eyelet Door 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 anti-wrinkle and anti shrinkage and have 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'

O

TEXTES – TRAITEMENT DE DONNÉES

LEMMATIZATION

key features polyester multicolor abstract eyelet door curtain floral curtain . polyester multicolor abstract eyelet door curtain height . pack price . curtain enhances look interior . curtain made high quality polyester fabric . feature eyelet style stitch metal ring . make room environment romantic loving . curtain ant . wrinkle anti shrinkage elegant appearance . 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 evening . create special moment joyous beauty given soothing print . bring home elegant curtain softly filter light room get right amount sunlight . specifications polyester multicolor abstract eyelet door curtain height . pack general brand designed door type eyelet model name abstract polyester door curtain set model color multicolor dimensions length box number contents sales package pack sales package curtains body design material polyester

○

STEMMING

key featur polyest multicolor abstract eyelet door curtain floral curtain . polyest multicolor abstract eyelet door curtain height . pack price . curtain enhanc look interior . curtain made high qualiti polyest fabric . featur eyelet style stitch metal ring . make room environ romant love . curtain ant . wrinkl anti shrinkag eleg appar . give home bright modernist appeal design . surreal attent sure steal heart . contemporari eyelet valanc curtain slide smooth draw a part first thing morn welcom bright sun ray want wish good morn whole world draw close even . creat special moment joyous beauti given sooth print . bring ho me eleg curtain soft filter light room get right amount sunlight . specif polyest multicolor abstract eyelet door curtain height . pack general brand design door type eyelet model name abstract polyest door curtain set model color multicolor dimens length box number content sale packag pack sale packag curtain bdi design materi polyest

O

O

EXTRACTION FEATURES NLP



WORD2VEC



TFIDF



TOPIC MODELING

TEXTES – WORD2VEC

LEMMATIZATION & STEMMING

Nombre de tokens avant

3534

Nombre de tokens après avoir créé '<UNK>'

2593

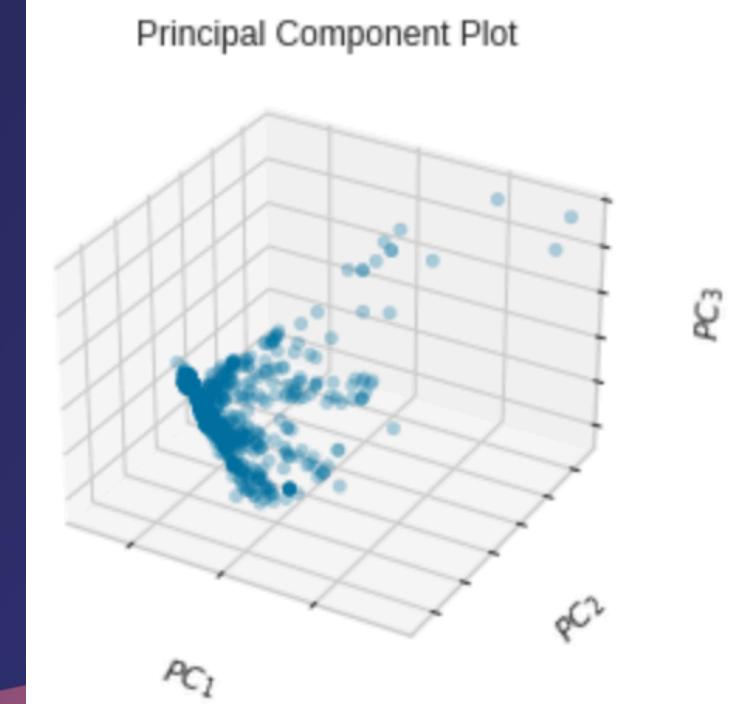
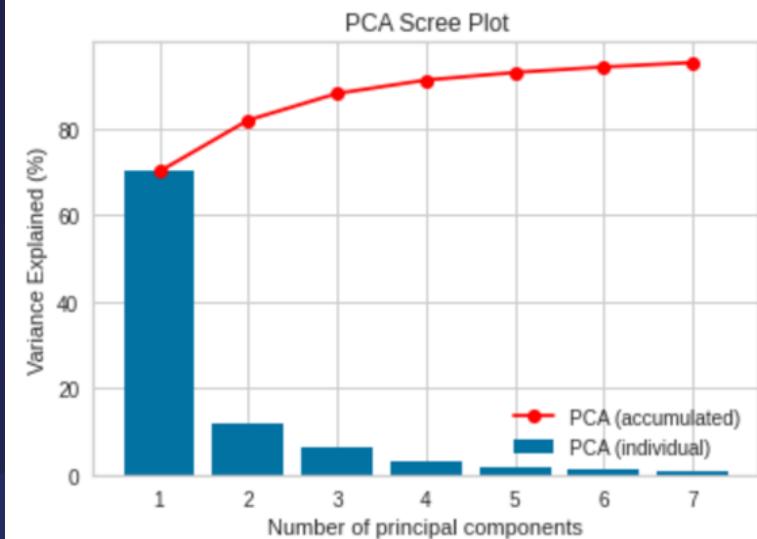
La taille de Bag of words réduite par

26.63%

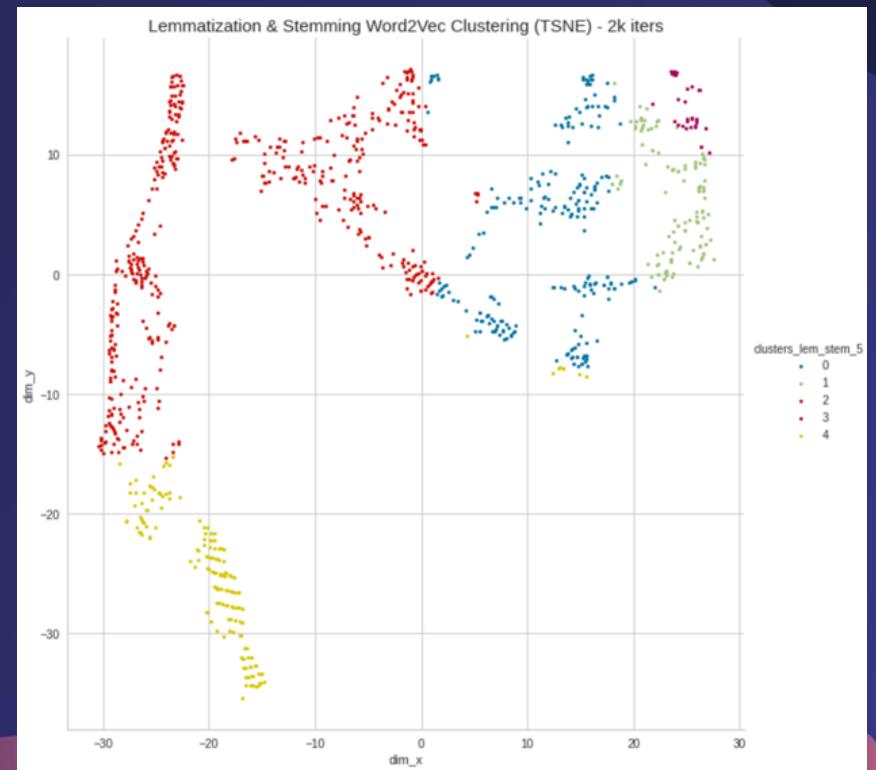
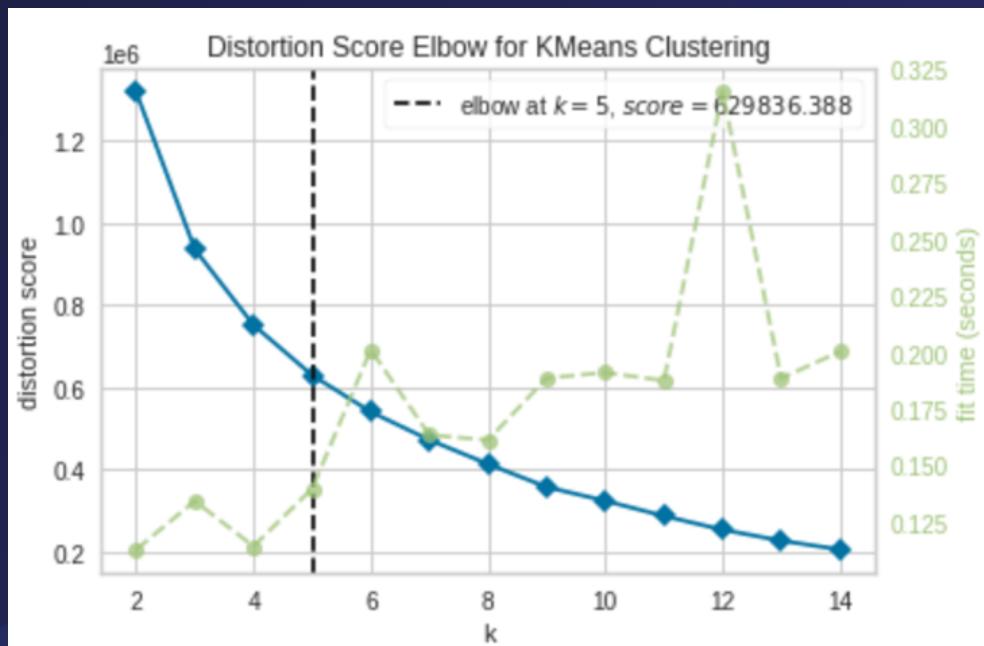
. <UNK> aapno aari abil abl abroad absorb abstract accent access accessori accid accident accord ach acid acn across acryl act activ act ptor add addit adhes adi adida adjust admir ador adorn ads1 advanc advic aero affect afford afternoon age agrasen aid air airtex airtight ala lli allin allow alloy allur almond almost alo aloevera along also alter altern aluminium alway amaz ambienc america amit ammonia among amount log andr andriod android angel anim anis anna anniversari anod anodis anoth ant antenna anthil anti antibacteri antiqu antra anymor anywhere a appeal appear appl appliq appreci approach approx arab arabian arch archi area arecacea ari armani aroma around arrang array ar act artifici artisan artist artpiec artwork ash asia ask aspect assembl assort assur aster asthma astonish astring asus atm attach attent att act audio australia australian authent auto automat avail avenu avocado avoid awaken away awesom azan babi babyoy babywear back background ba bacteria bag bain bakelit balanc ball balm balsam band bangl bank bankura bapa bar barbi barcod barrier base basic batch bath bathroom batter beach bead bean bear bearberri beat beauti beautifi beautiful becom bed bedroom bedsheets bee beig bela believ bell bella belong belt bendabl erlin best better beverag beyond bhagavan bib big bigger bird birthday bit black blackl blade bland blanket blast blazer bleach blemish ck blond blue bluecensor bluetooth blush bmw board bob bodi bodysuit bold bolster bomber bone bong bonsai book booti border born bors botan b ht bourn bow bowl box boy bpa branch brand brass bread break breast breathless breez brick bright brighten brillianc brilliant bring bristl b roken bronchiti broom brought brow brown brownish brush bsf bubbl bucket buckl buddha buddhism buddhist built bulb bull burn burner burp burt teri button buy buyer cabl cactus calf calib calibro california call calm camera camerii candl cannon canon canva cap capac capri captiv car rd care cargo caron carpentri carpet carri carrier cartier cartoon carv case cash casserol cast casual catarrh catchi categori cater caus ced center centr centuri ceram ceremoni certain certif certifi chair chakra chandan chang changer chappin charact characterist charg charger char cheeki cheer chef chemic cherri chest chhote chicken child children china chines chinmay chip chlorin choic chomper choos chopra chowkidar ch ph circuit circul citi civil claim clasp class classi classic classif clay clean cleaner cleans clear click clip clock close closur cloth clu cnl coast coaster coat cocoa cocoon cocovey code coffe coin cold collagen collar collect color colour column colur com combin combo come comf

TEXTES – RÉDUCTION DE DIMENSION (WORD2VEC)

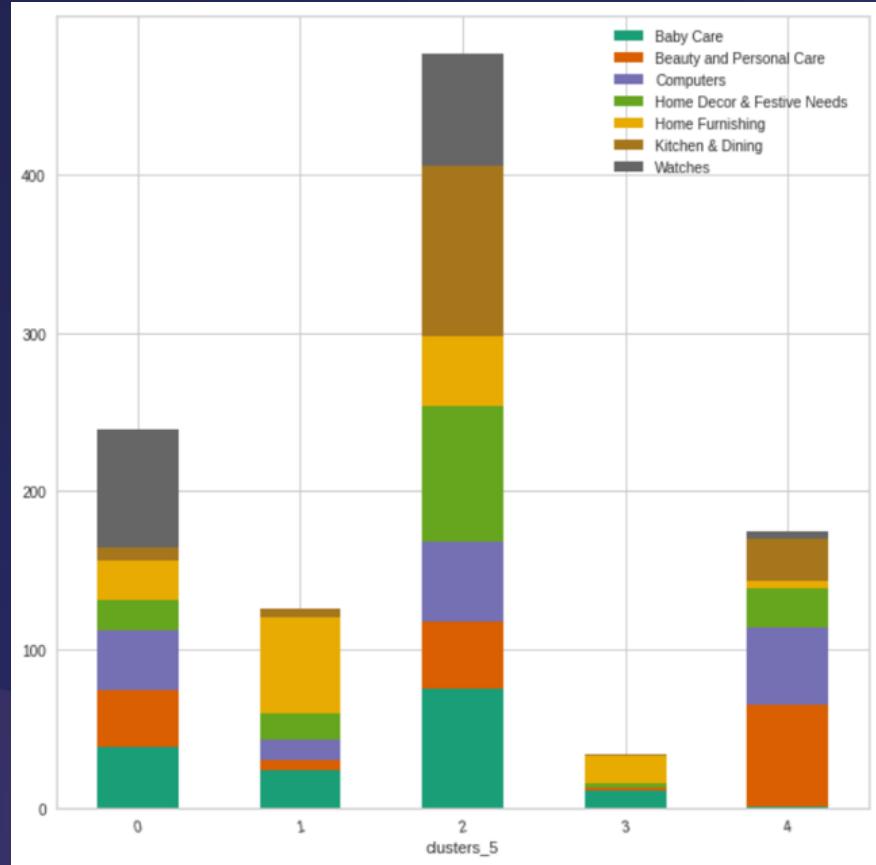
Original number of features: 2593
No. of features after dimensionality reduction: 7
Features reduced by: 99.73%



TEXTES – K-MEANS (WORD2VEC)



TEXTES – CLUSTERS VS. CAT. ORIGINALES (WORD2VEC)



EXTRACTION FEATURES NLP



WORD2VEC



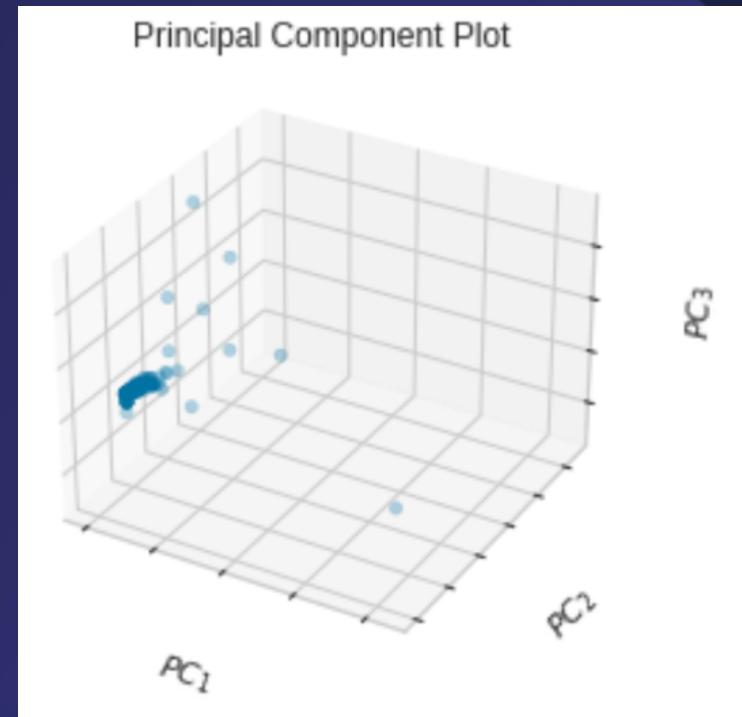
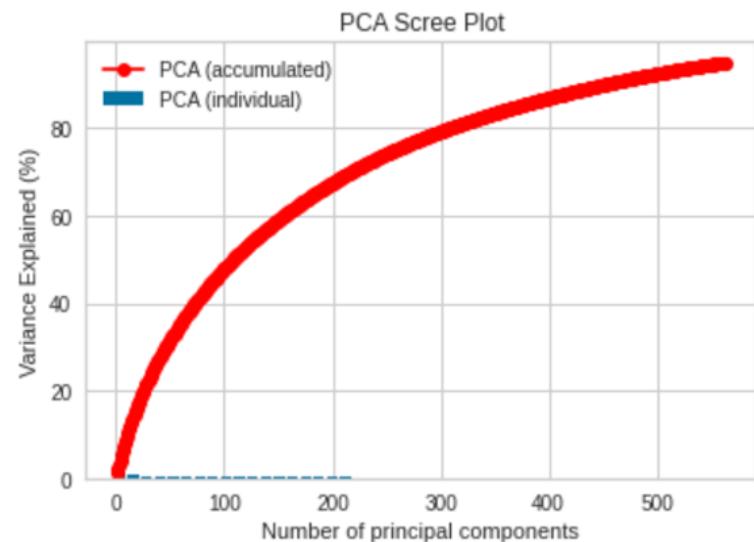
TFIDF



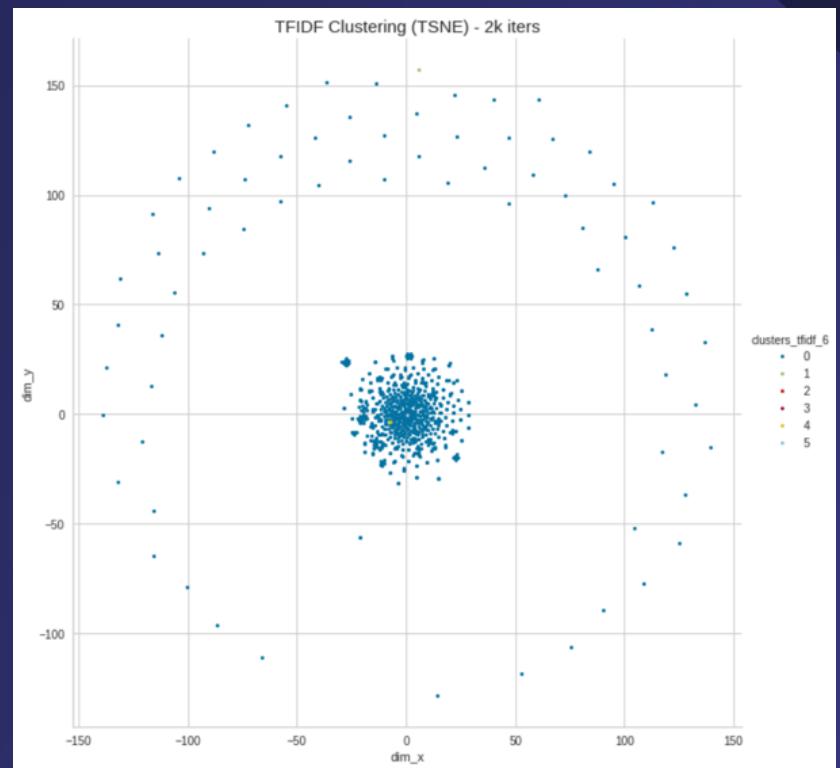
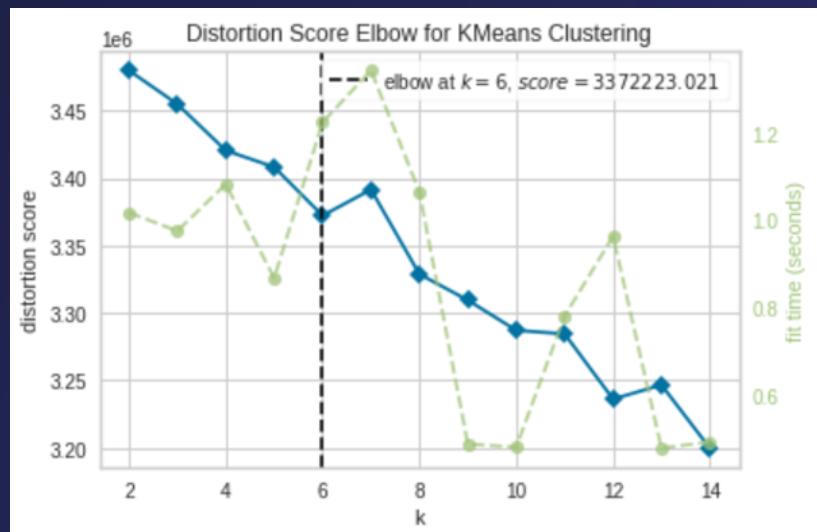
TOPIC MODELING

TEXTES – RÉDUCTION DE DIMENSION (TF-IDF)

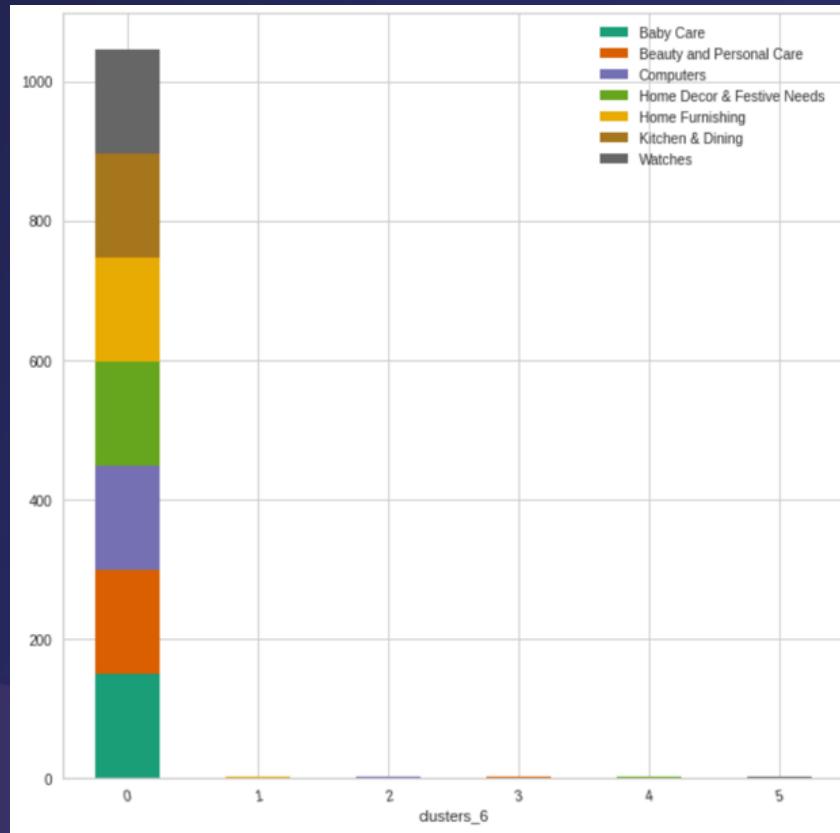
Original number of features: 3534
No. of features after dimensionality reduction: 563
Features reduced by: 84.07%



TEXTES – K-MEANS (TF-IDF)



TEXTES – CLUSTERS VS. CATÉGORIES ORIGINALES (TF-IDF)



EXTRACTION FEATURES NLP



WORD2VEC



TFIDF



TOPIC MODELING

TEXTES – TOPIC MODELING (LDA)

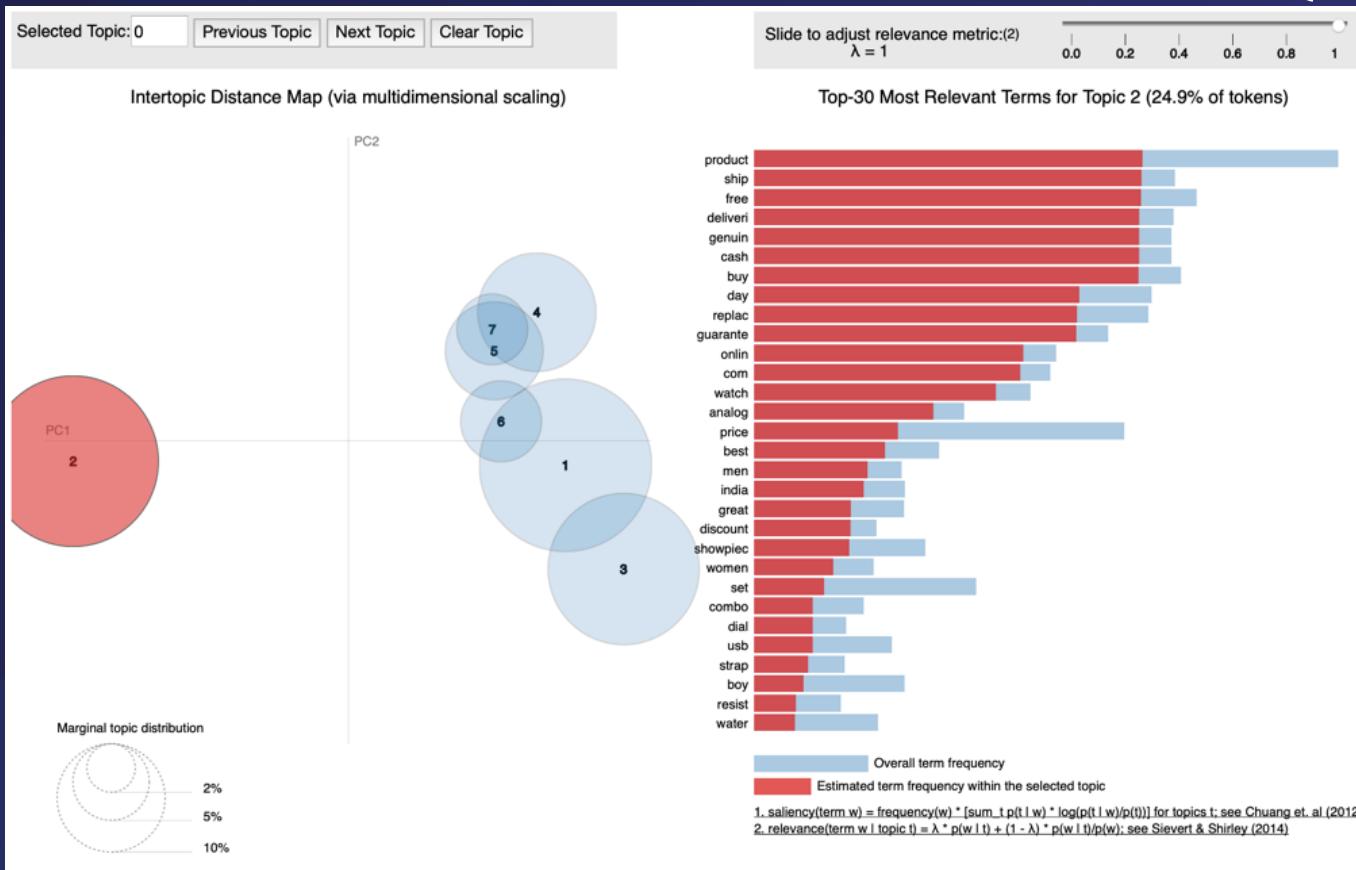
MODÈLE DE RÉFÉRENCE

| | |
|-----------------|-----------|
| Coherence Score | 0.583669 |
| Perplexity | -5.440031 |

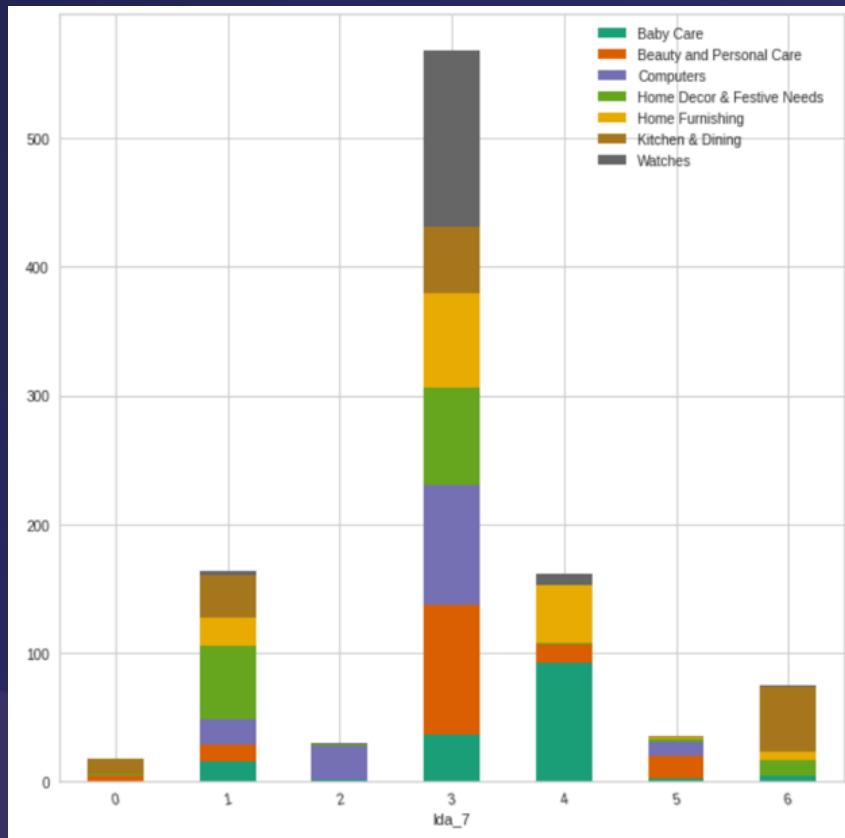
RÉSULTATS DE GRID SEARCH

| | topics | alpha | eta | coherence |
|------|--------|---------------------|--------------------|-----------|
| 52 | 4 | 0.44 | 451.35 | 0.677503 |
| 51 | 4 | 0.44 | 407.1 | 0.669235 |
| 1806 | 7 | symmetric | 8.85 | 0.664352 |
| 54 | 4 | 0.44 | 539.85000000000001 | 0.657874 |
| 962 | 6 | 0.36000000000000004 | 761.1000000000001 | 0.656461 |
| 53 | 4 | 0.44 | 495.6 | 0.653364 |
| 525 | 5 | 0.8000000000000002 | 8.85 | 0.648377 |
| 964 | 6 | 0.36000000000000004 | 849.6 | 0.647571 |
| 437 | 4 | symmetric | 761.1000000000001 | 0.646416 |
| 436 | 4 | symmetric | 716.85 | 0.646416 |

TEXTES – CLUSTERS VS. CATÉGORIES ORIGINALES (LDA)



TEXTES – CLUSTERS VS. CATÉGORIES ORIGINALES (LDA)



04 |

Extraction Features Image :

IMAGES ORIGINALES



CLASSIQUE

- Algorithmes classiques:
- SIFT, ORB ou SURF
-
-

DEEP LEARNING

- Transfer Learning d'un algorithme de CNN
-
-

CLASSIQUE

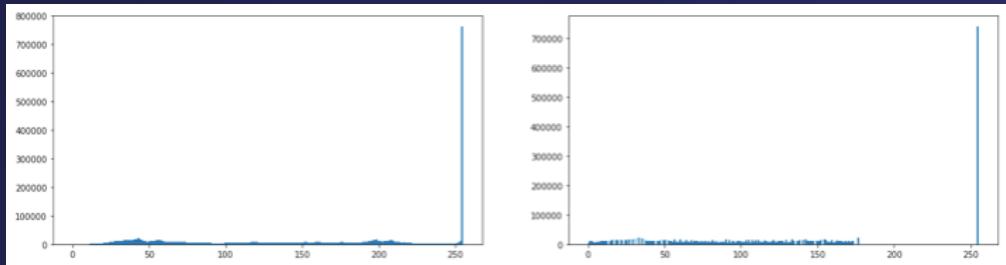
- Algorithmes classiques:
- SIFT, ORB ou SURF
-
-

DEEP LEARNING

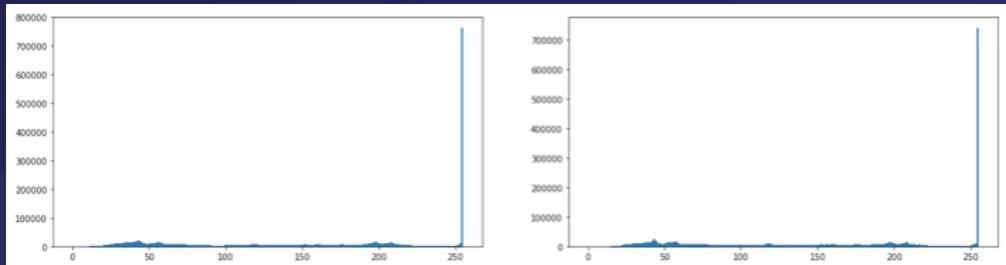
- Transfer Learning d'un algorithme de Classification
-
-
-

IMAGES – TRAITEMENT DE DONNÉES

EQUALISATION

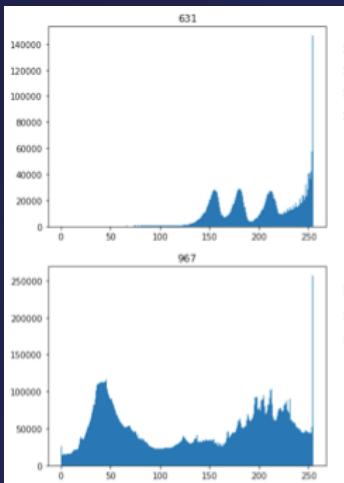


ELIMINATION DU BRUIT

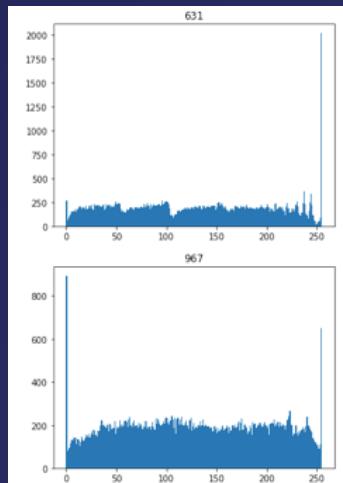


IMAGES – TRAITEMENT DE DONNÉES

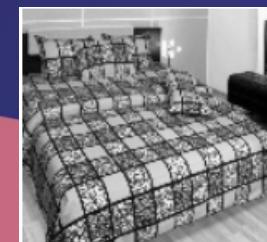
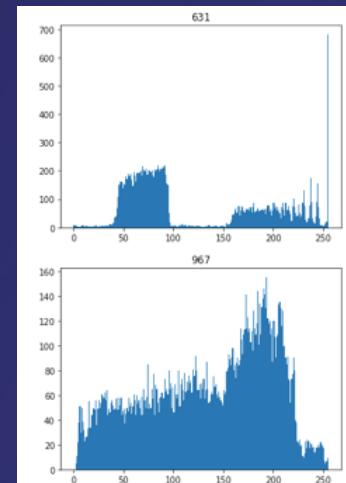
AVANT



APRÈS - COULEUR

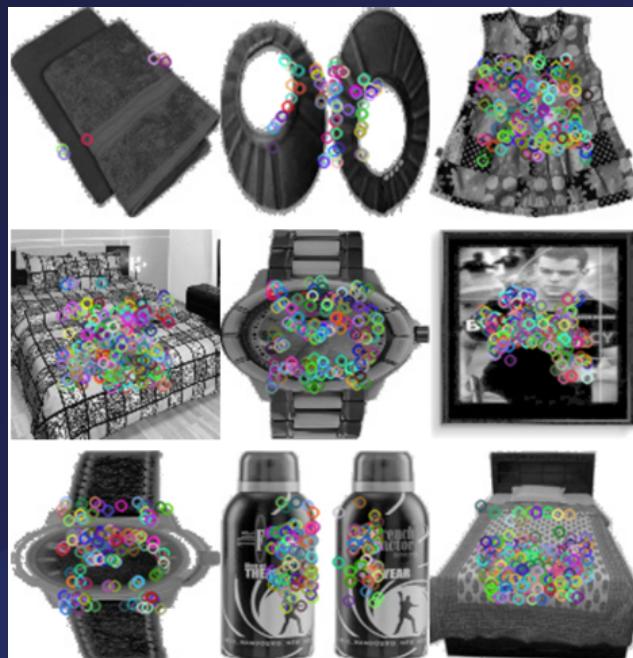


APRÈS - GREY

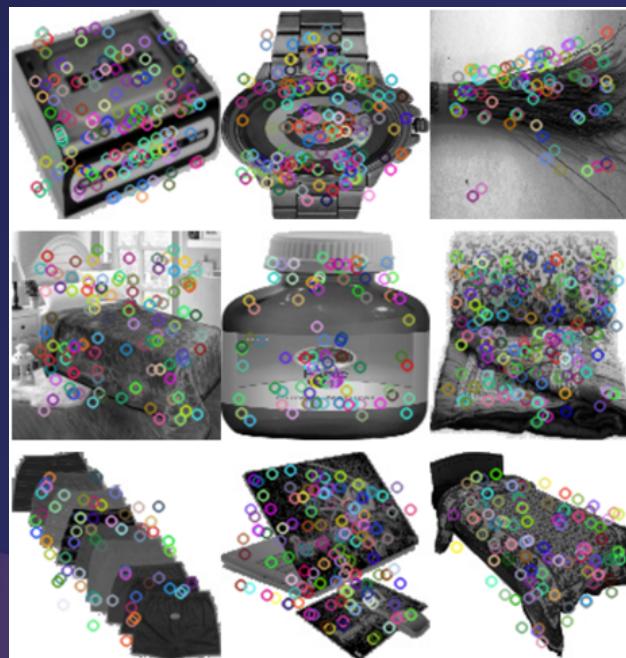


IMAGES – EXTRACTION DES CHARACTÉRISTIQUES

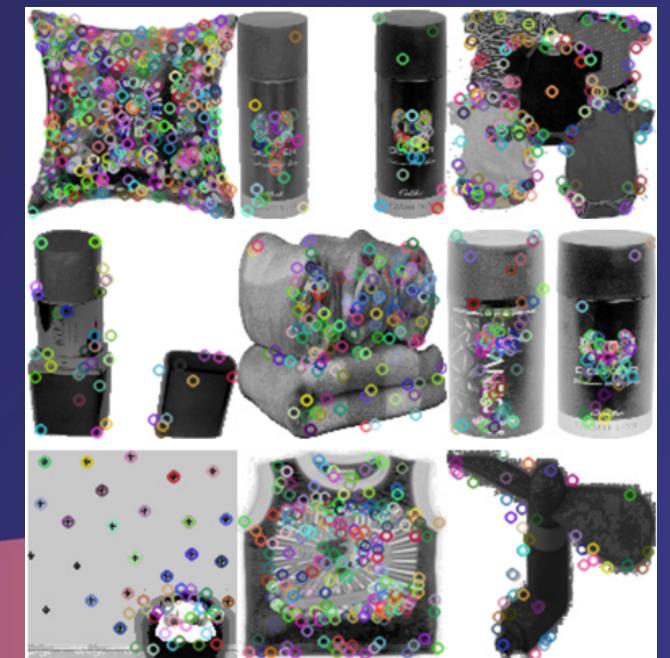
ORB



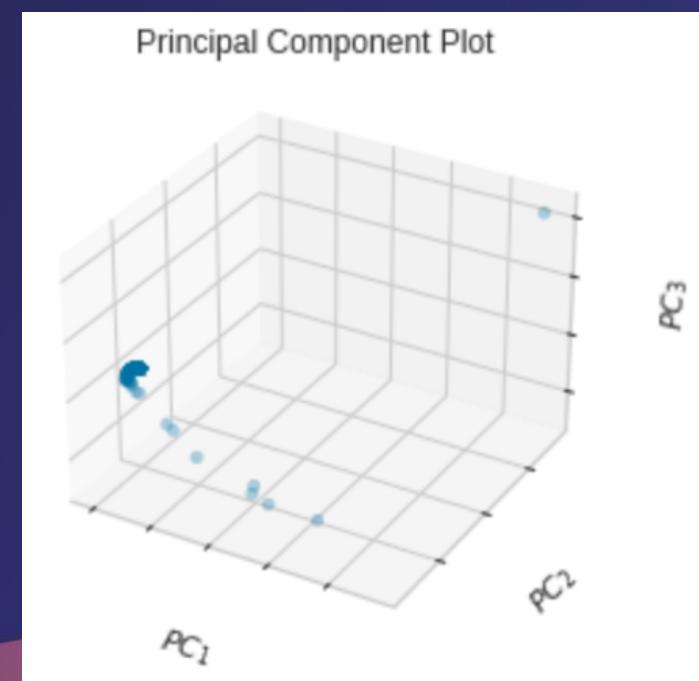
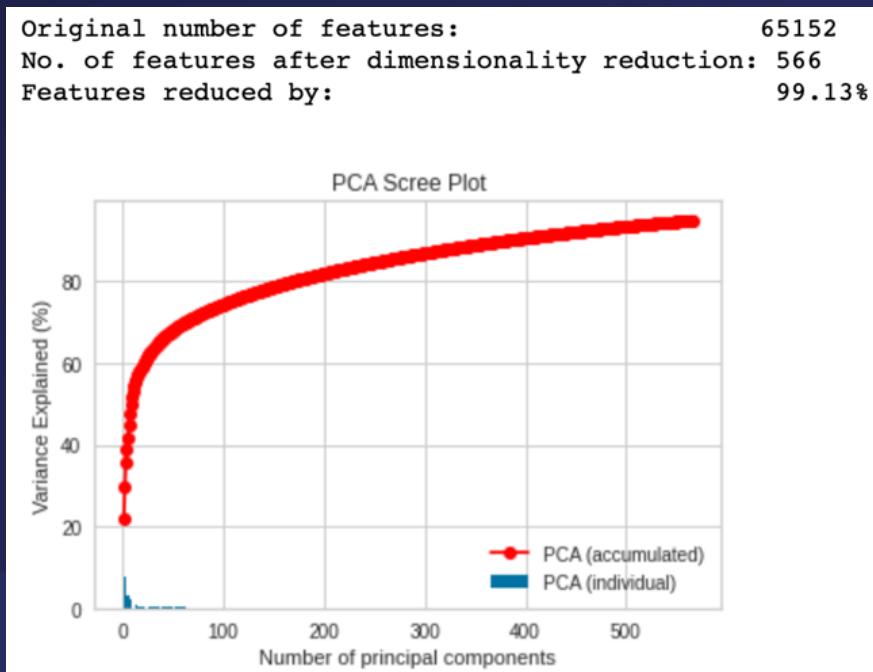
SUFT



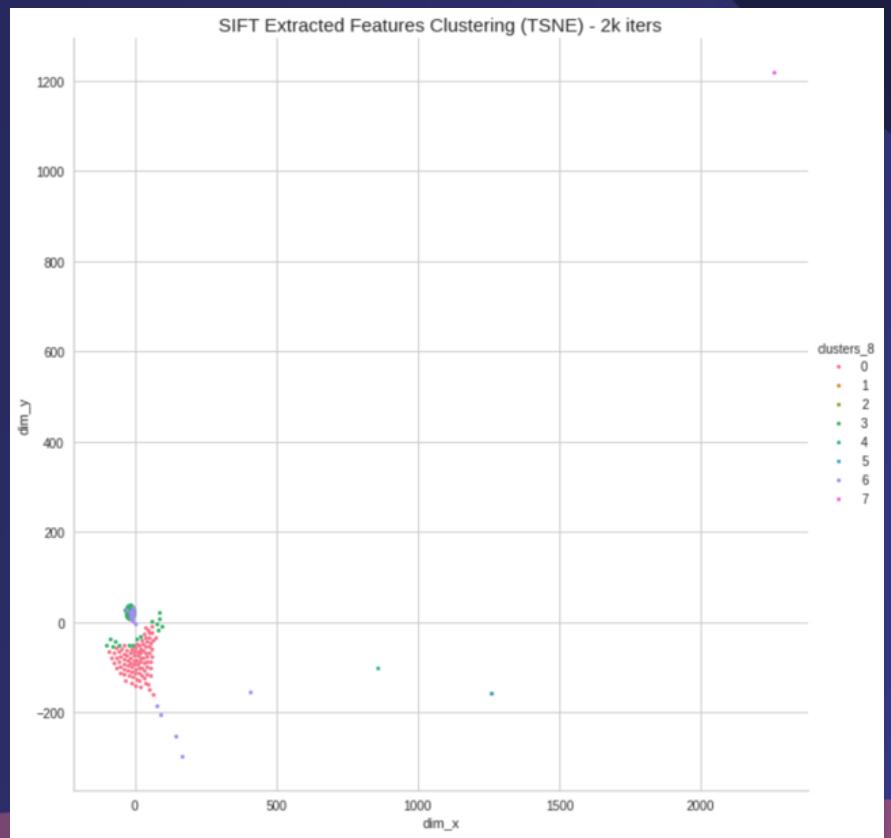
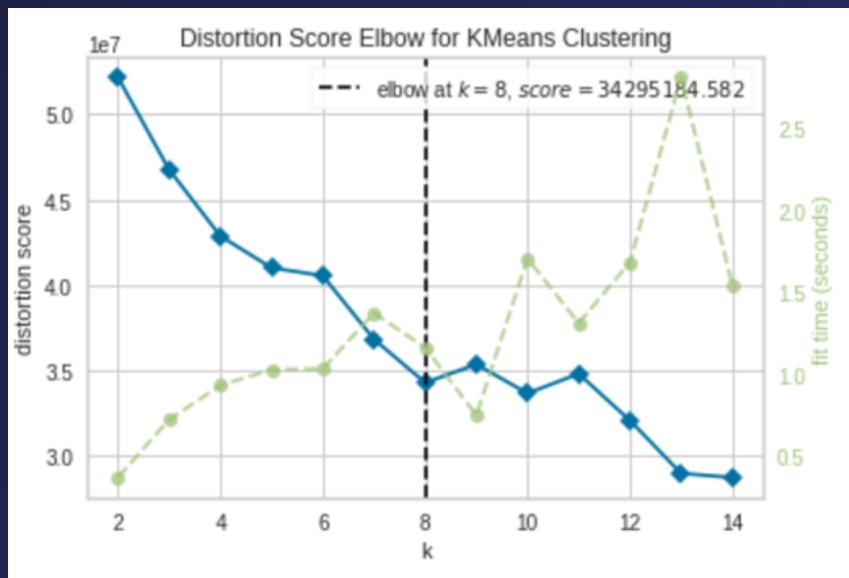
SIFT



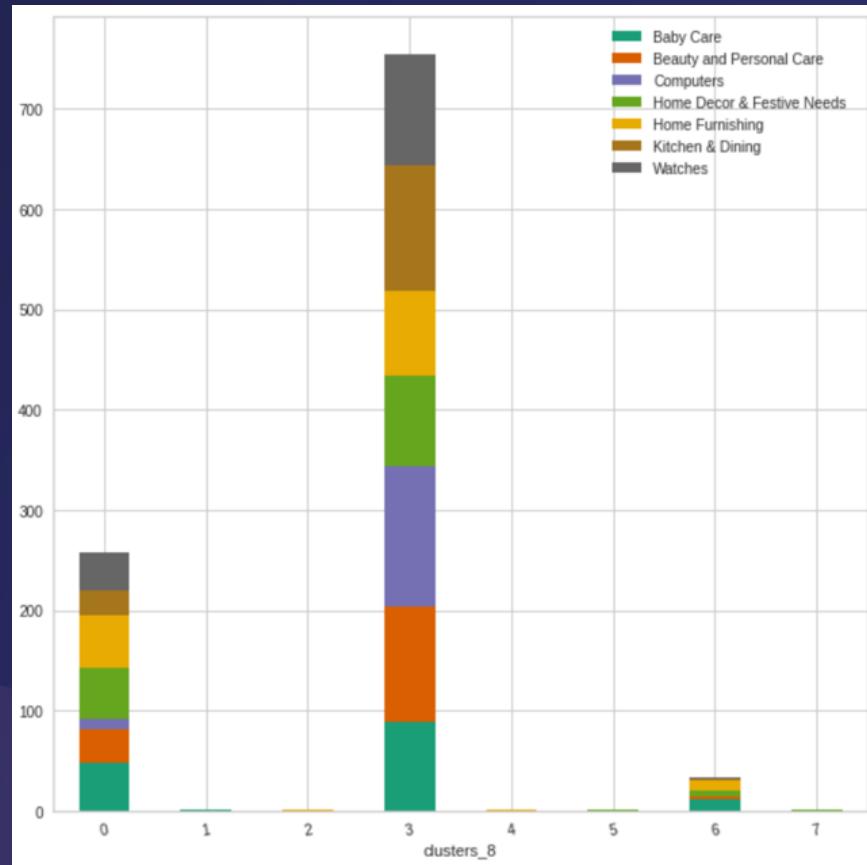
IMAGES – RÉDUCTION DE DIMENSION (SIFT)



IMAGES – K-MEANS (SIFT)



IMAGES – CLUSTERS VS. CATÉGORIES ORIGINALES (SIFT)



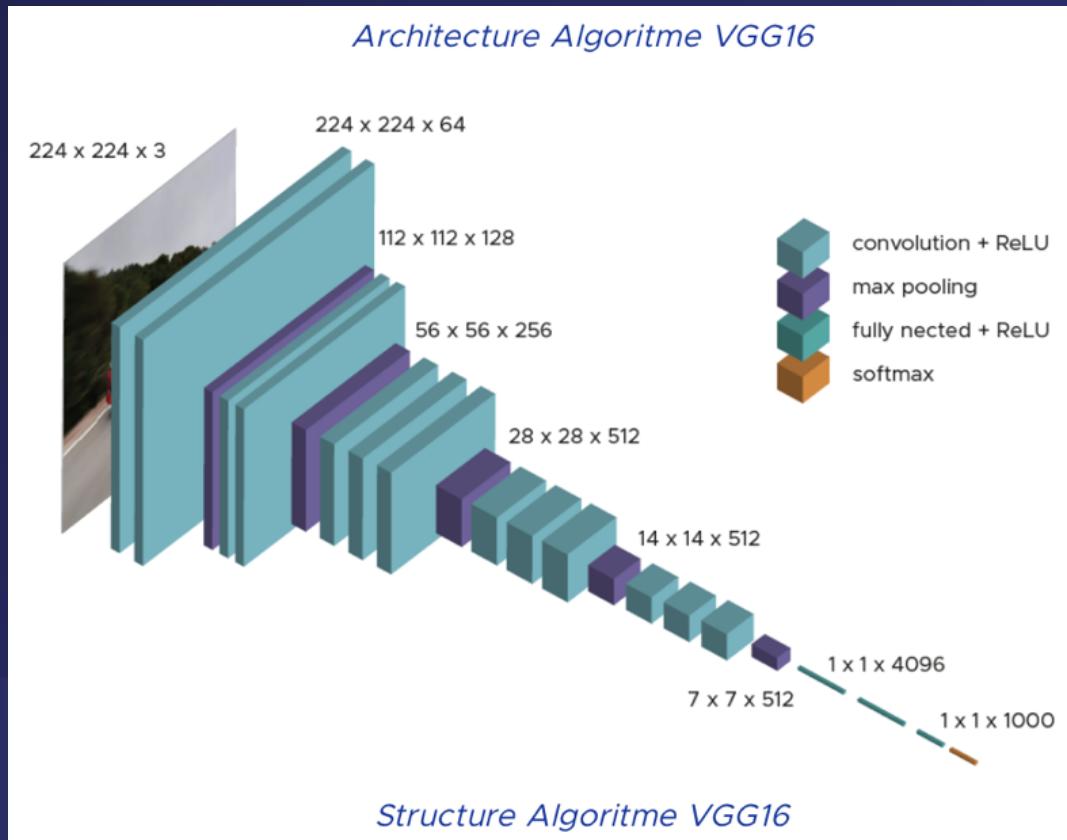
CLASSIQUE

Algorithmes classiques:
SIFT, ORB ou SURF

DEEP LEARNING

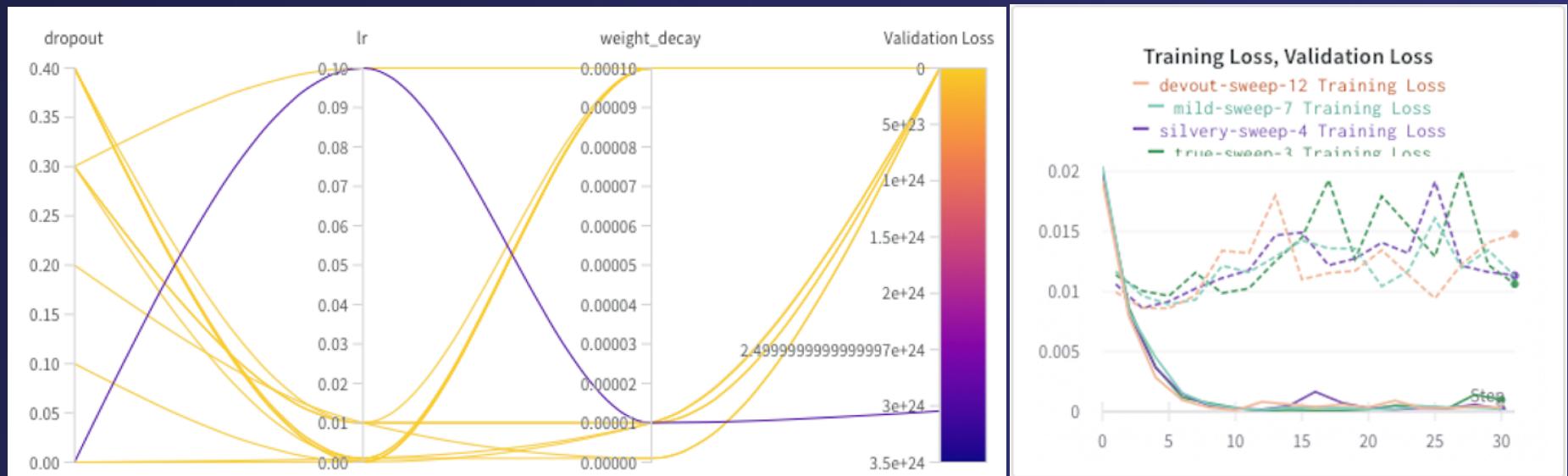
Transfer Learning d'un
algorithme de CNN

IMAGES – VGG16

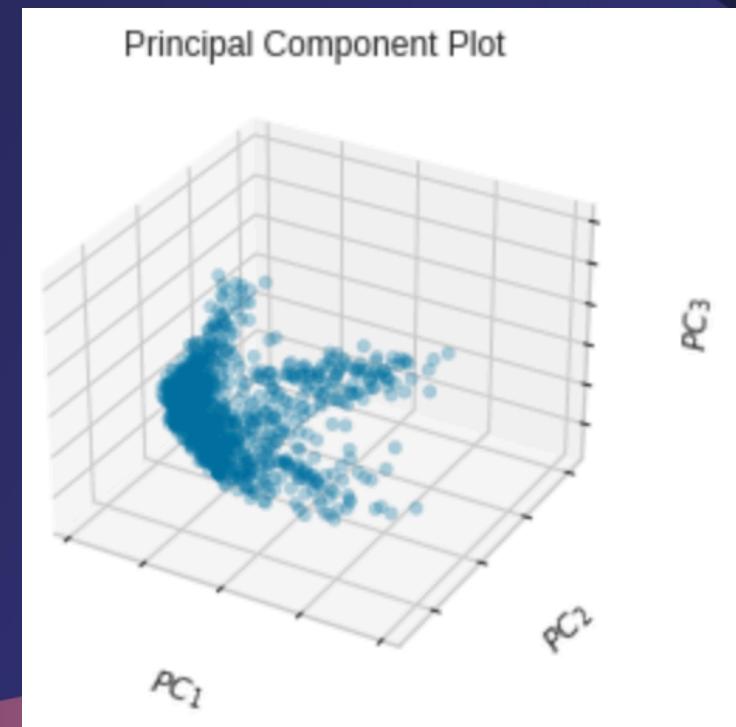
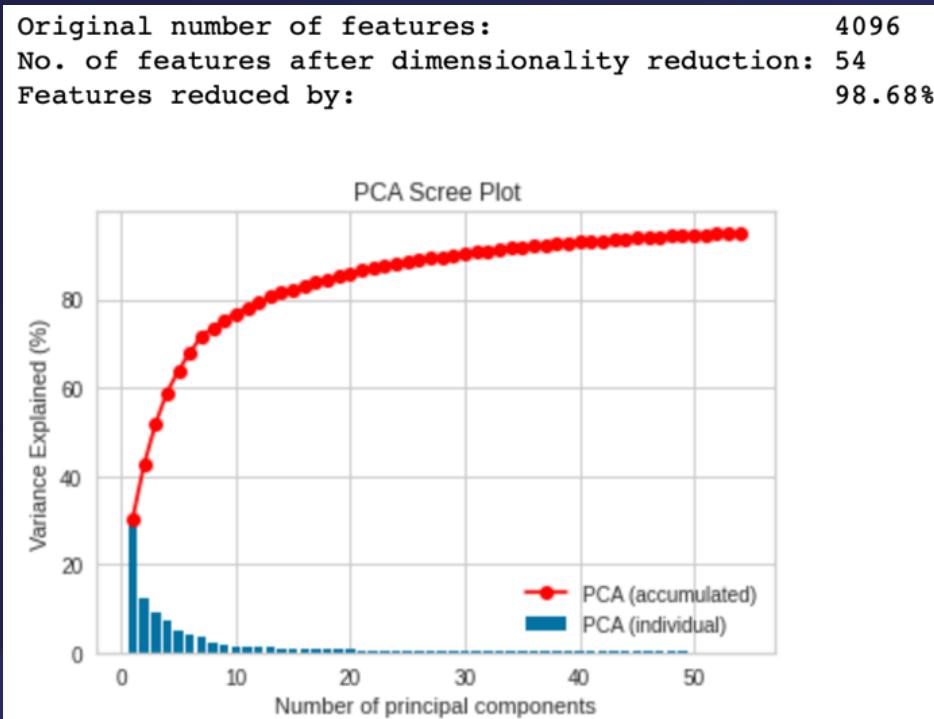


SOURCE: <https://datascientest.com/quest-ce-que-le-modele-vgg>

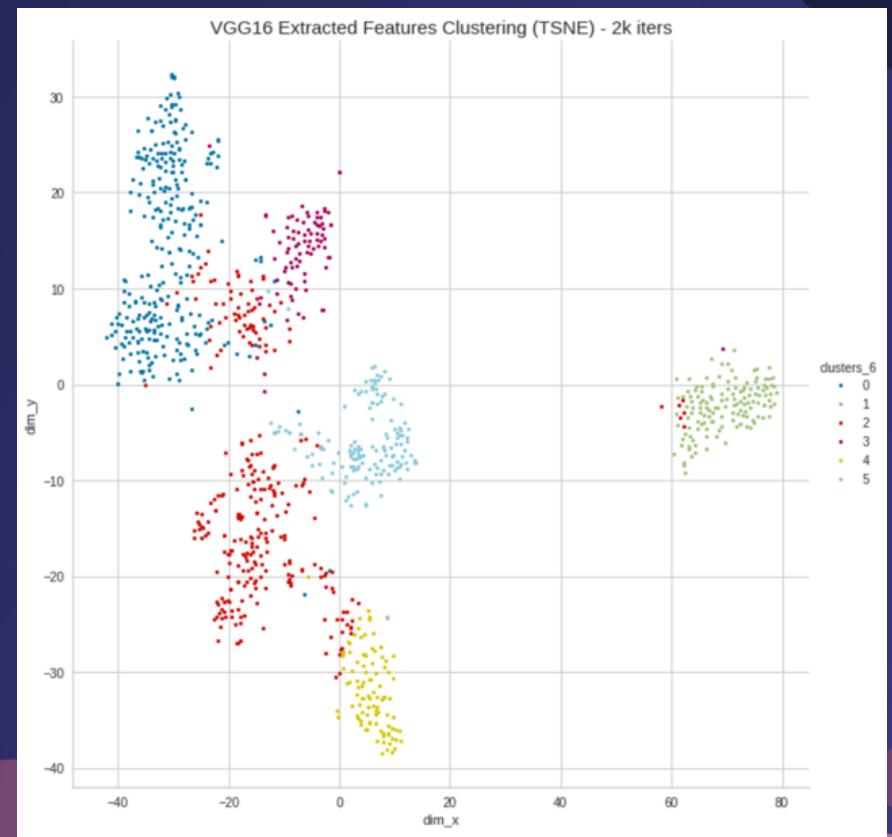
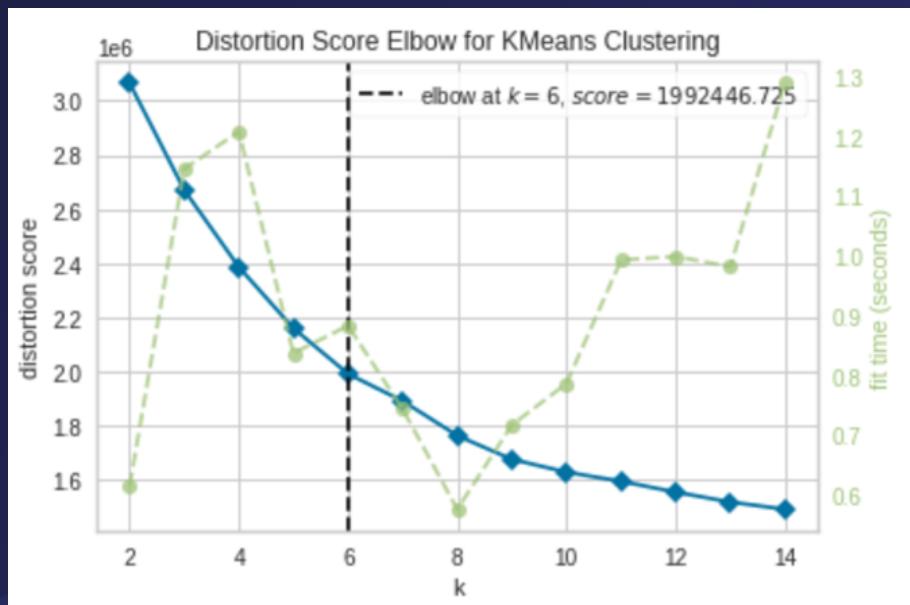
IMAGES – HYPERPARAMETERS TUNNING (VGG16)



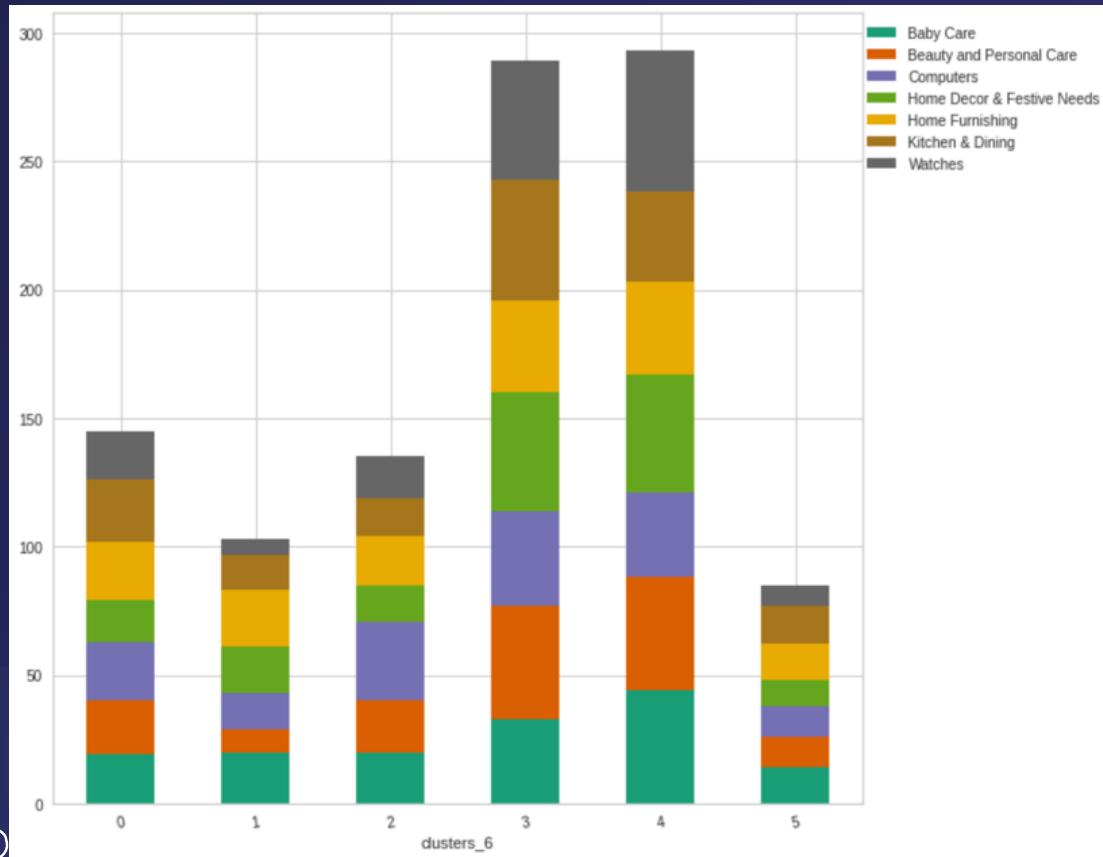
IMAGES – RÉDUCTION DE DIMENSION (VGG16)



IMAGES – K-MEANS (VGG16)



IMAGES – CLUSTERS VS. CATÉGORIES ORIGINALES (VGG16)



Extraction Features Combinées

05 !

COMBINÉES – K-MEANS

Shape of image SIFT descriptor vector

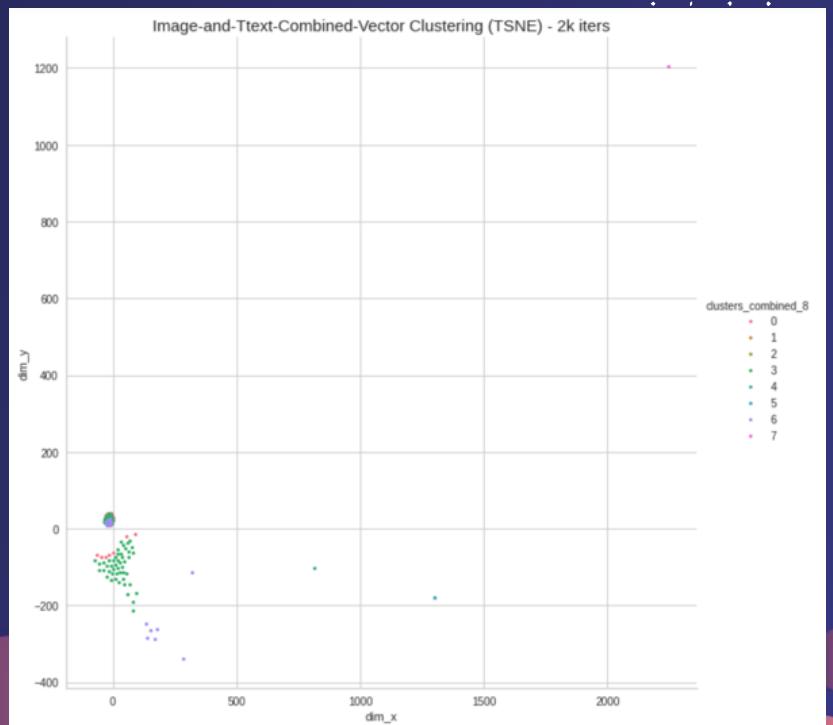
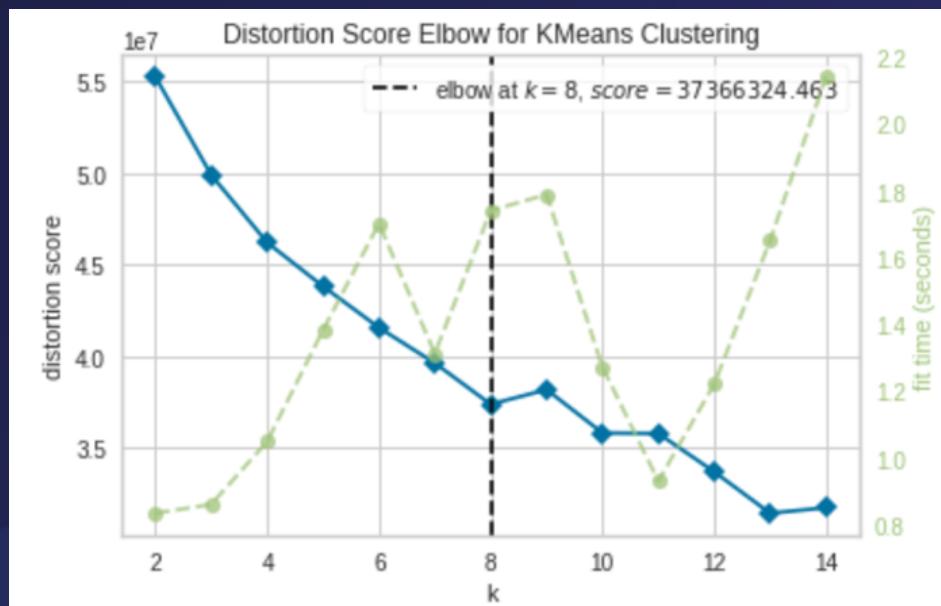
(1050, 566)

Shape of word2vec vector

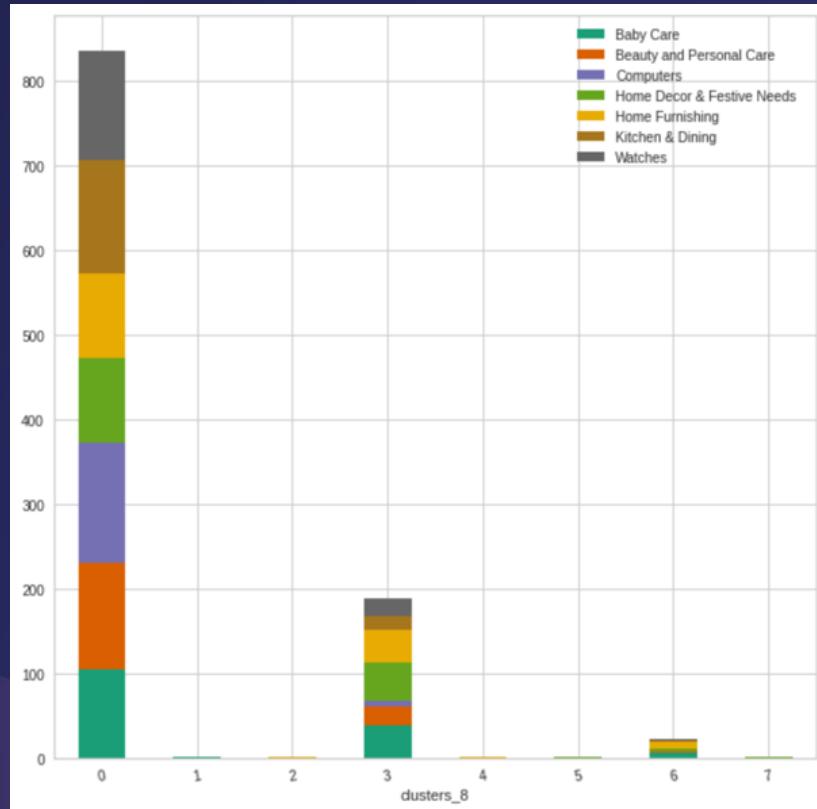
(1050, 7)

Shape of the combined vector

(1050, 574)



COMBINÉES – CLUSTERS VS. CATÉGORIES ORIGINALES

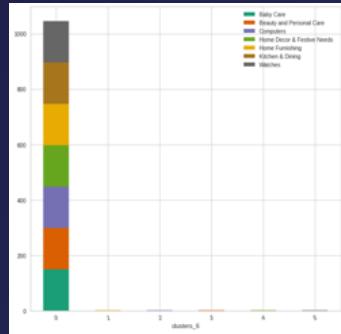


CONCLUSION

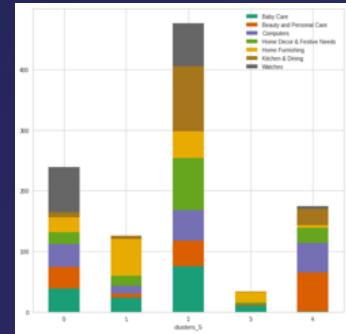
06

COMBINÉES – CLUSTERS VS. CATÉGORIES ORIGINALES

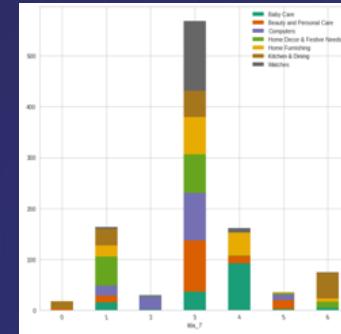
TEXTES – TF-IDF



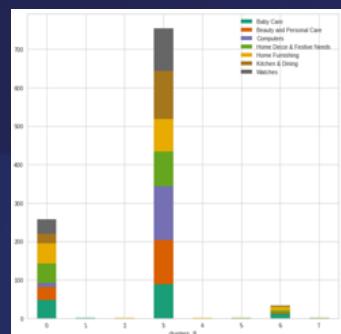
TEXTES – WORD2VEC



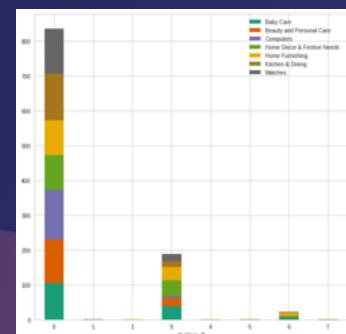
TEXTES – LDA



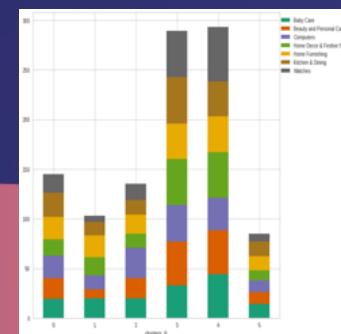
IMAGES – SIFT



COMBINÉES



IMAGES – VGG16



THANKS!

[linked.com/in/channingp](https://www.linkedin.com/in/channingp) 

Do you have any questions?

channing.platevoet@gmail.com
+33 6 9906 9777

CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), infographics & images by [Freepik](#)