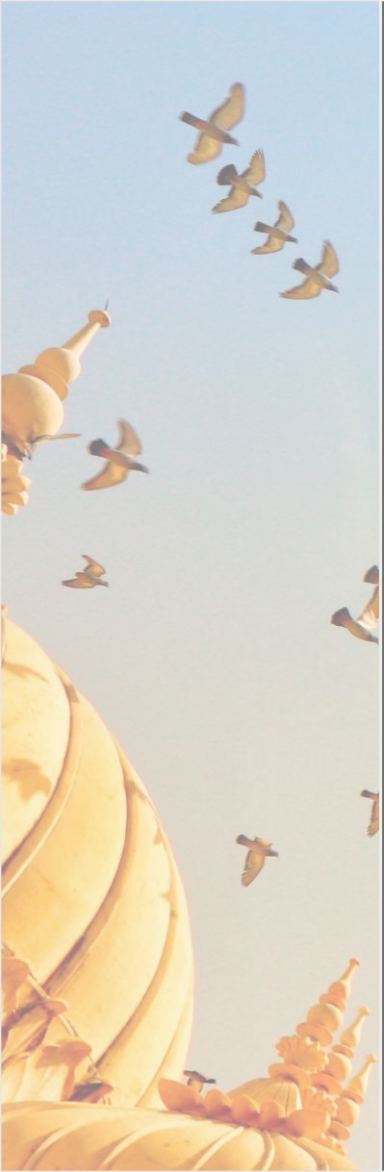




Place de marché

Projet 6 - Raphaël Koudache--Louvet



Place de marché

Objectif : étudier la faisabilité d'un **moteur de classification**

Nota Bene : faisabilité de regrouper automatiquement des produits de même catégorie => **clustering**



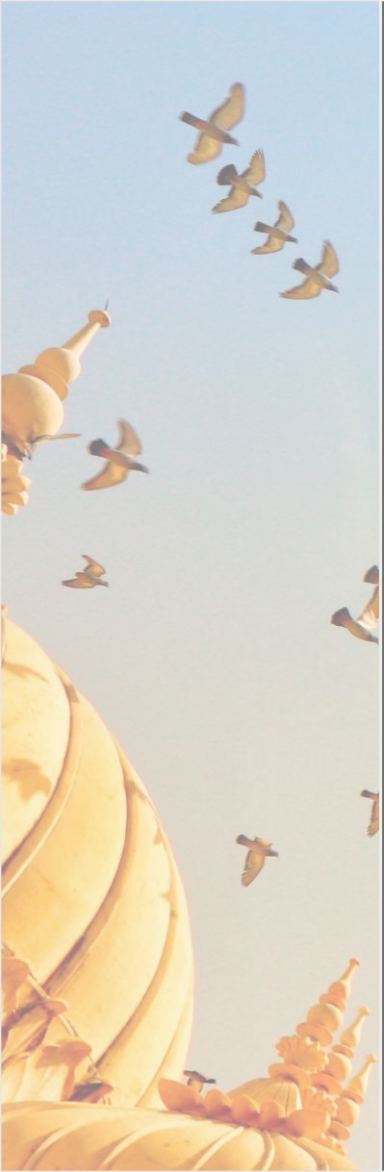
I - Analyse Exploratoire des données (EDA)

II - Feature Extraction

III - Modélisation

I - Analyse Exploratoire des données (EDA)



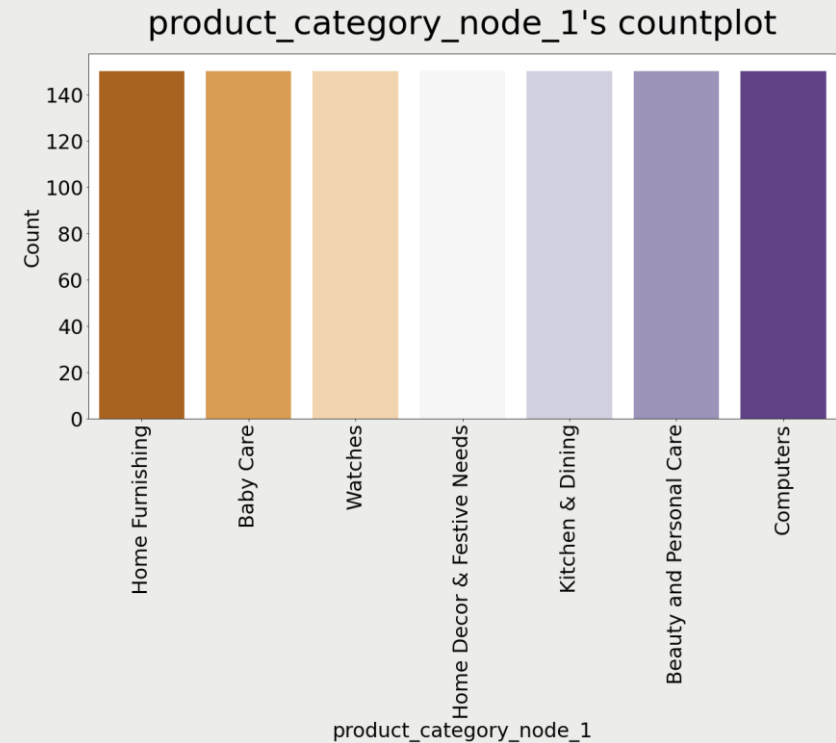


Dataframe

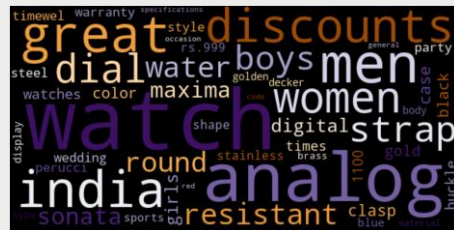
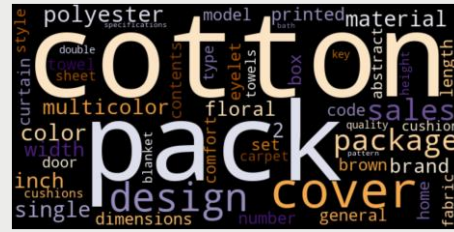
1050 produits indiens

Données
tabulaires +
données non
tabulaires (Texte
& Image)

Labels en nœuds



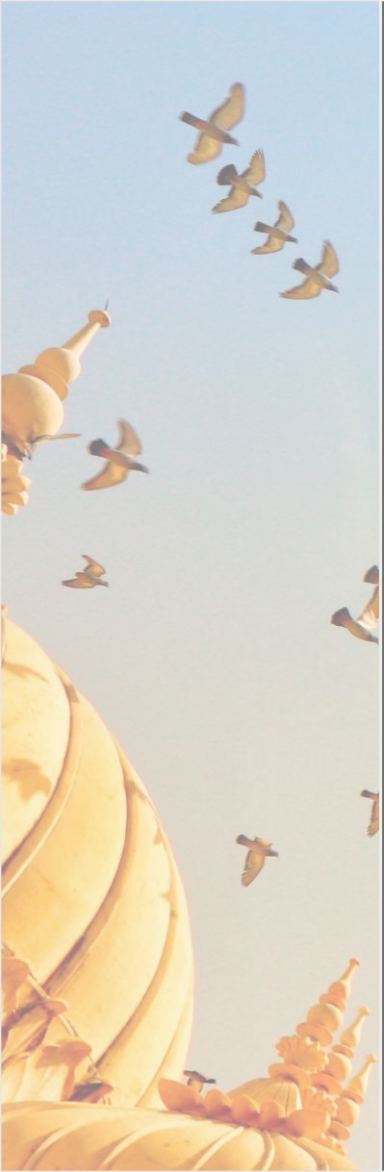
Catégories Nœud 1



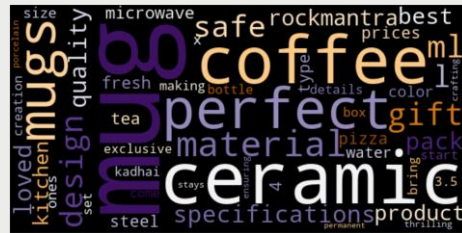
Home Furnishing

Baby Care

Watches



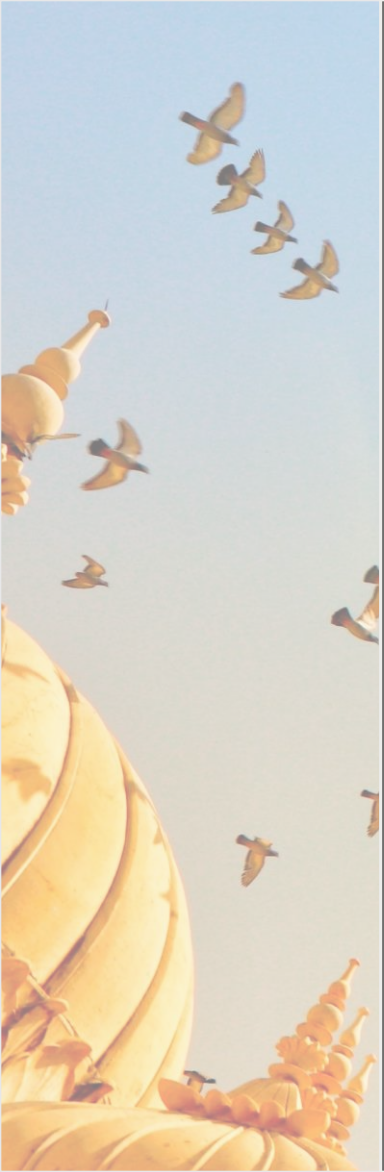
Catégories Nœud 1



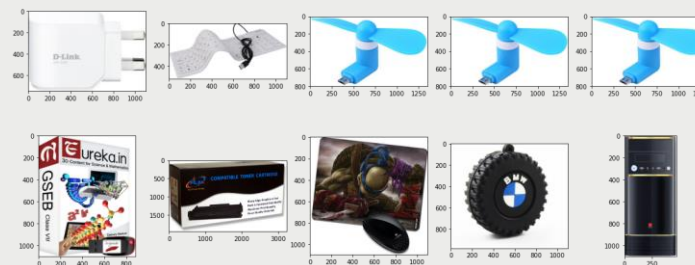
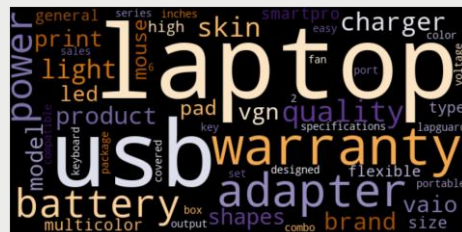
Home Decor & Festive Needs

Kitchen & Dining

Beauty and Personal Care



Catégories Nœud 1



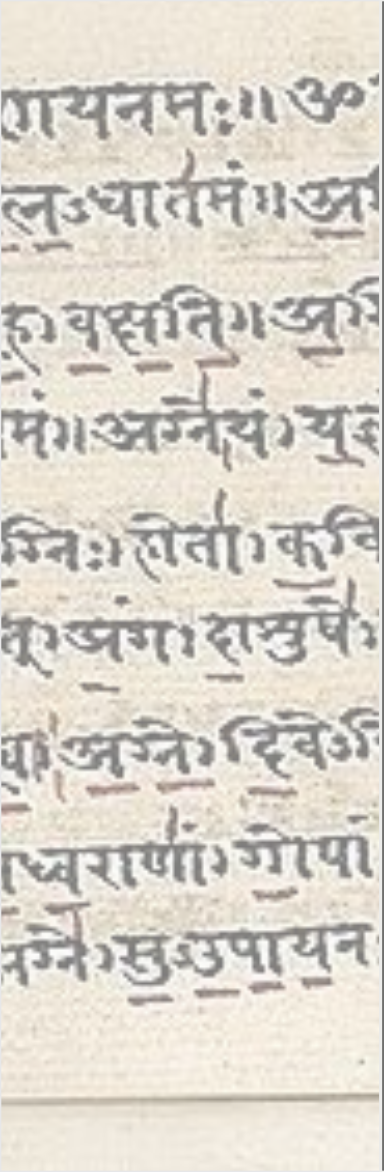
Computers

II - Feature Extraction



श्रीगणेशायनमः॥ ॐ नमः॥ अग्निं॥ ईळे॥ पुरः॥ र्हितं॥ यज्ञस्य॥ देवं॥ ऋत्विजं
 ॥ होतारं॥ रत्न॥ धातमं॥ अग्निः॥ पूर्वभिः॥ ऋषिभिः॥ इत्यः॥ नूतनैः॥ उत॥ सः॥ देवा
 न॥ आ॥ इह॥ वसति॥ अग्निना॥ रुचिं॥ अन्नवत्॥ पोषं॥ एव॥ दिवे॥ दिवे॥ यज्ञसं॥
 वीरवत्॥ तमं॥ अग्नेयं॥ युतं॥ अक्षरं॥ विश्वतः॥ परिभूः॥ अस्ति॥ सः॥ इत्॥ देवेषु॥ ग
 ष्ठति॥ अग्निः॥ होता॥ तमः॥ अन्नस्य॥ तमः॥ देवः॥ देवे॥ आ॥ गमु
 त्॥ ३॥ यत्॥ अंग॥ दाम्बुर्षः॥ तं॥ अग्ने॥ भद्रं॥ कुरिष्यसि॥ तव॥ इत्॥ तव॥ सप्तं॥ अग्नि
 रः॥ उर्षः॥ त्वा॥ अग्ने॥ दिवे॥ दिवे॥ शोषा॥ वस्तः॥ धिया॥ वयं॥ नमः॥ भरतः॥ आ॥ इमं॥ सि
 राजंतं॥ अध्वराणां॥ गोपां॥ ऋतस्य॥ दीदिविं॥ वर्धमानं॥ स्व॥ हमै॥ सः॥ नः॥ पिता॥ इवा
 सुनवे॥ अग्ने॥ सु॥ उपा॥ यनः॥ भव॥ सर्वस्वा॥ नः॥ स्वस्तये॥ २॥ गायो॥ इति॥ आ॥ युहि॥

Natural Language Processing



Prétraitement du texte

Tokenization

Casse

Stop Words

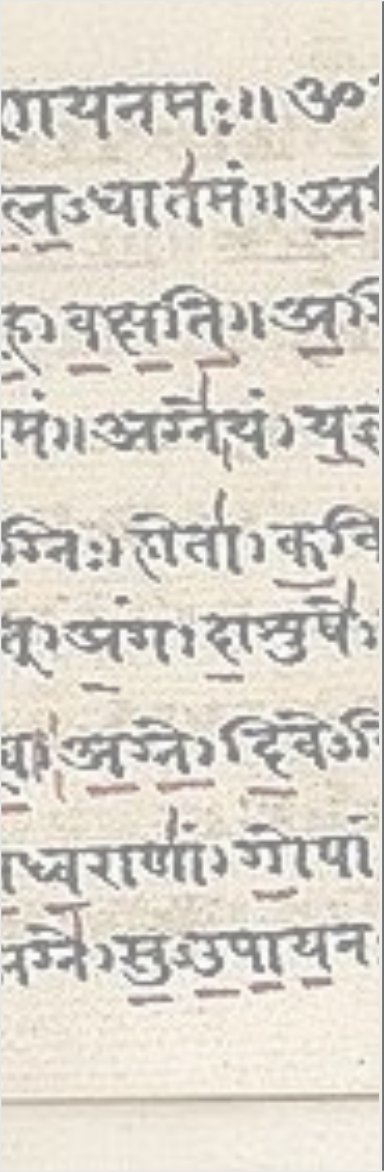
Mots Rares et Trop Fréquents

Punctuation

Part of Speech

(Out of Vocabulary)

Lemmatisation



Extraction des features

Bag of words:

Compte, TF-IDF
(1050, 3249)

Modèles Préentraînés:

1. Sans contextes :

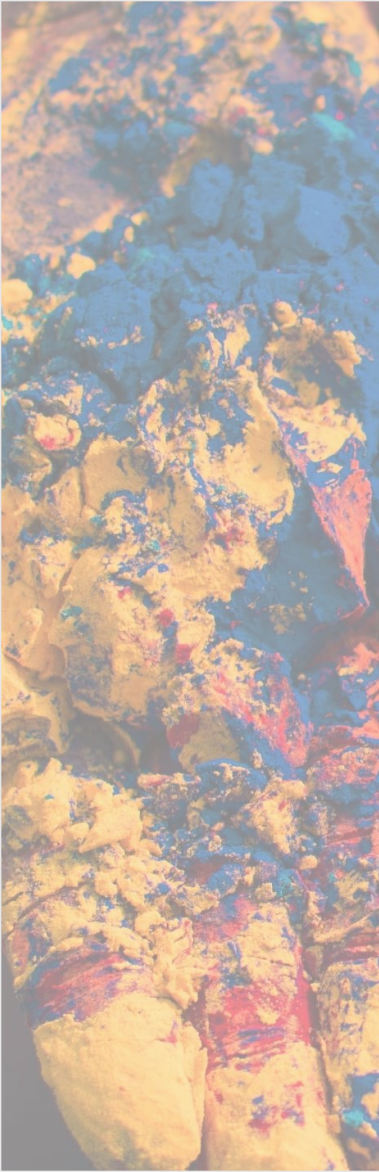
Word2VEC, Glove, FastText, spacy's
(1050, 300)

2. Avec contextes :

BERT (1050, 768)
USE (1050, 512)

A close-up photograph of two hands, palms facing each other, covered in thick, textured layers of blue, yellow, and red paint. The paint is applied in a way that creates a complex, almost abstract pattern across the skin. The background is a soft, out-of-focus blur of similar colors, creating a cohesive visual theme.

Computer Vision



Extraction des features

Bag of Visual Words:

SIFT

(1050, 719)

ORB

(1050, 721)

SIFT 500 first keypoints on a watch



ORB 500 first keypoints on a watch



Convolution neural NN (Flatten Layer):

VGG-16

(1050, 25088)

ResNet50V2

(1050, 100352)

EfficientNetB0

(1050, 62720)



III - Modélisation



राधे

कृष्ण

Non supervisé

Modélisation et Dimensionnalité

Modélisation :

KMEANS 7 Composants

Réduction de dimensionnalité :

PCA (% de variance)

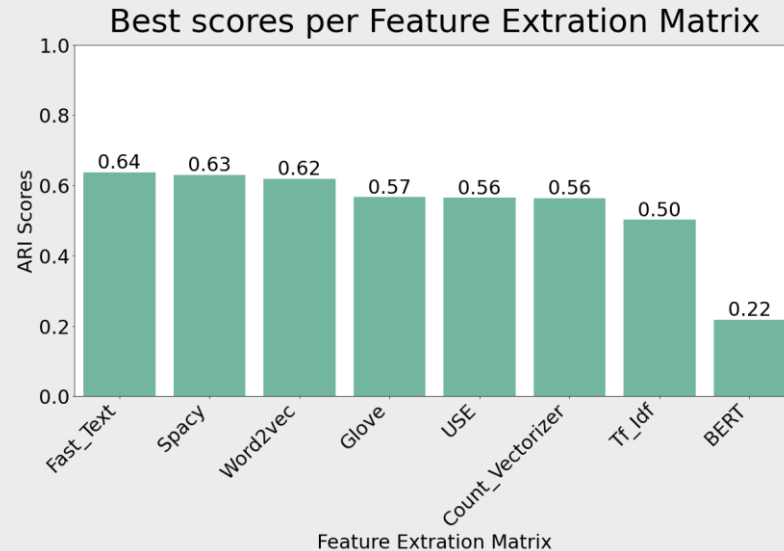
PCA puis TSNE

Latent Dirichet Allocation

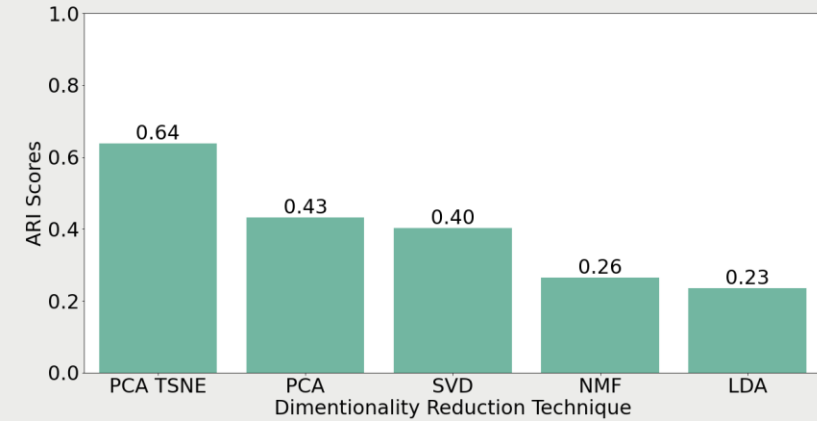
Singular Value Decomposition

Non negative Matrix Factorization

Catégorie : premier nœud (7 catégories) NLP

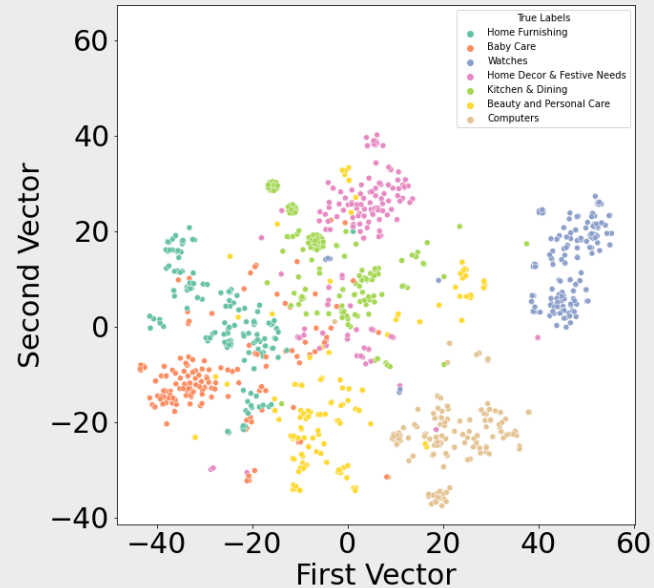


Best scores per Feature Dimentionality Reduction Technique

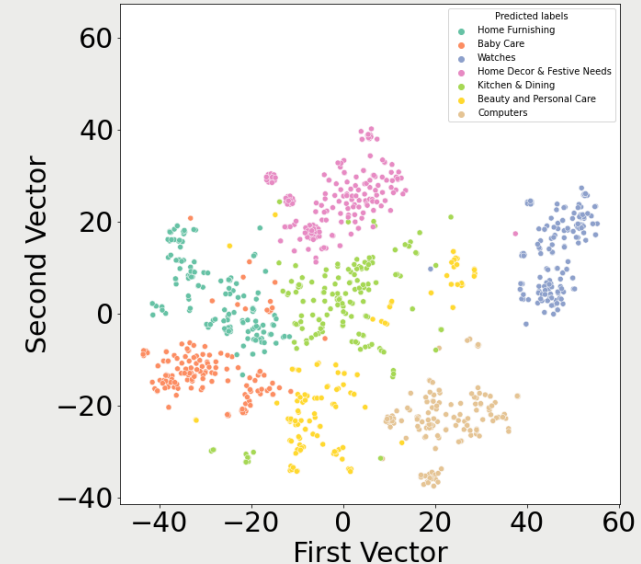


Catégorie : premier nœud (7 catégories) - Spacy's Embedding - ARI : 0,64

PCA t-SNE nlp_matrix True Labels



PCA t-SNE nlp_matrix Predictions

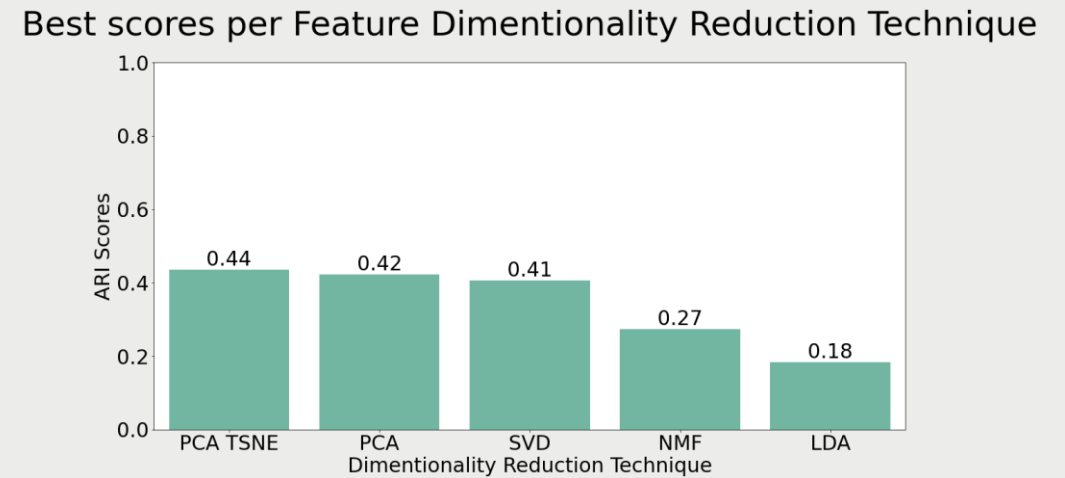
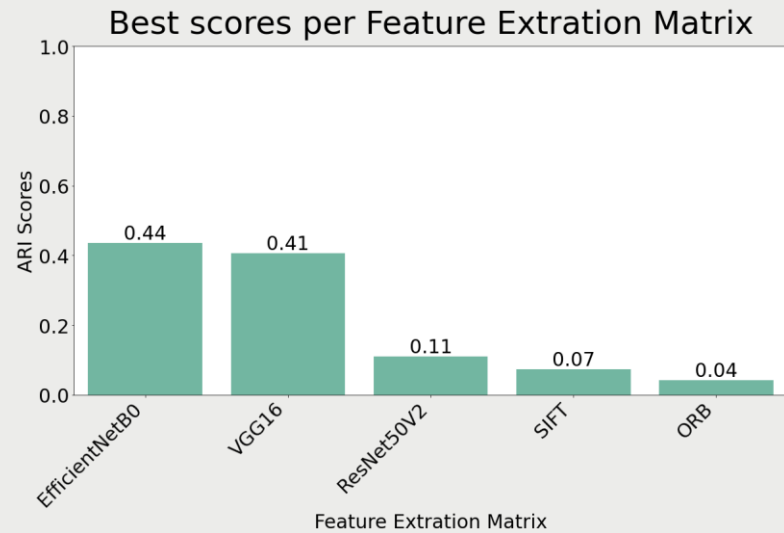


Catégorie : premier nœud (7 catégories) - NLP MATRIX - ARI : 0,63

nlp_matrix's Labels Confusion Matrix

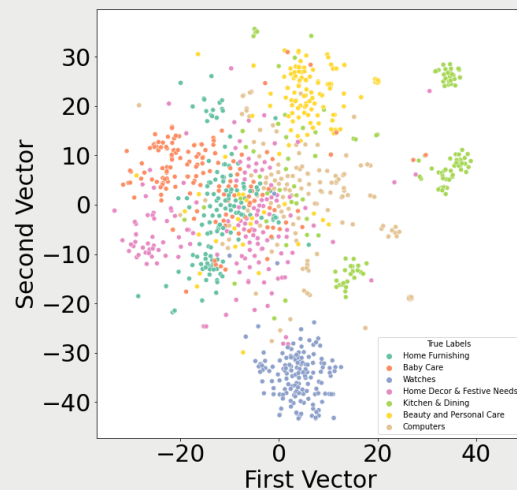
	Home Furnishing	Baby Care	Watches	Home Decor & Festive Needs	Kitchen & Dining	Beauty and Personal Care	Computers
Home Furnishing	123	26	0	0	1	0	0
Baby Care	16	106	0	4	21	3	0
Watches	0	0	145	3	2	0	0
Home Decor & Festive Needs	2	0	2	99	44	2	1
Kitchen & Dining	2	1	0	63	84	0	0
Beauty and Personal Care	0	9	0	8	5	126	2
Computers	0	0	0	0	5	1	144
	Home Furnishing	Baby Care	Watches	Home Decor & Festive Needs	Kitchen & Dining	Beauty and Personal Care	Computers

Catégorie : premier nœud (7 catégories) Computer Vision

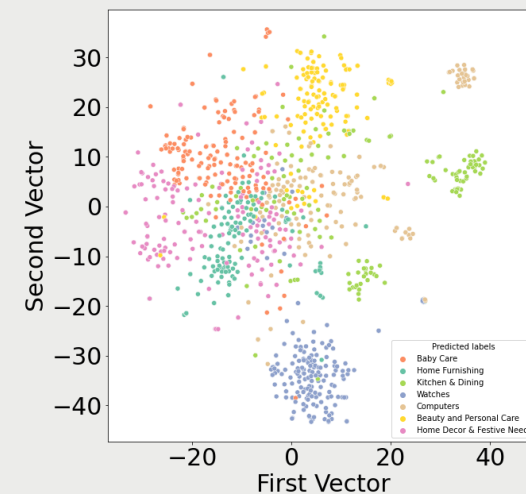


Catégorie : premier nœud (7 catégories) - EfficientNetB0 - ARI : 0,44

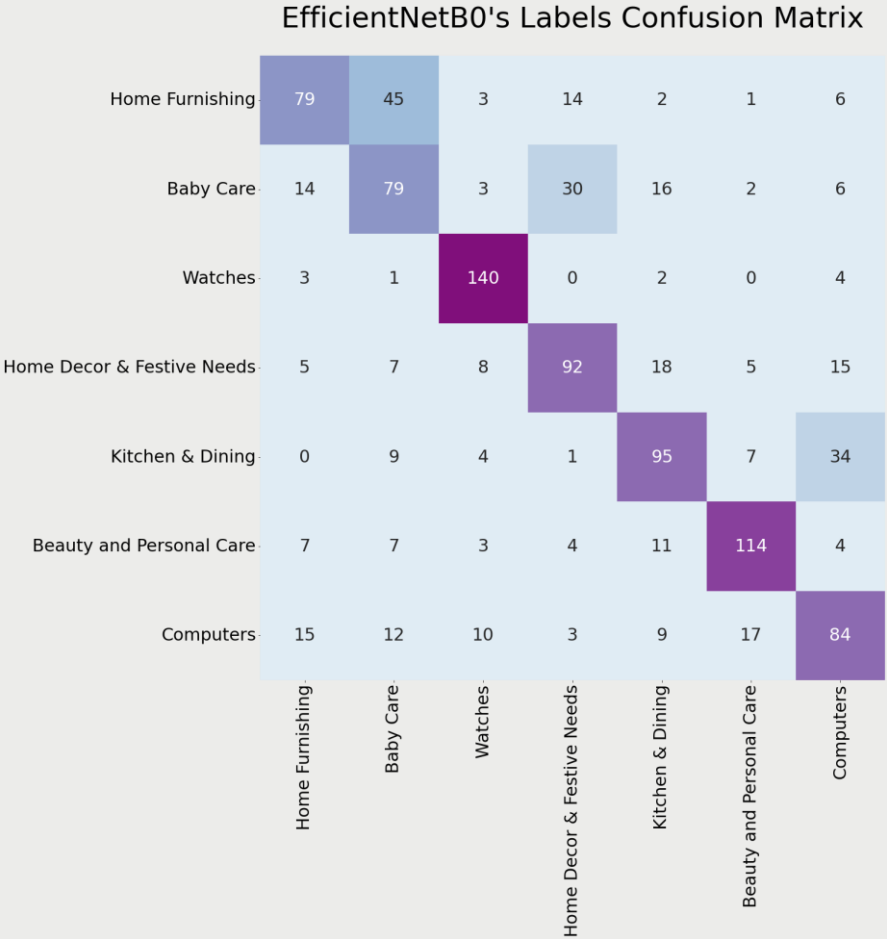
PCA t-SNE EfficientNetB0 True Labels



PCA t-SNE EfficientNetB0 Predictions



Catégorie : premier nœud (7 catégories) - NLP MATRIX - ARI : 0,43



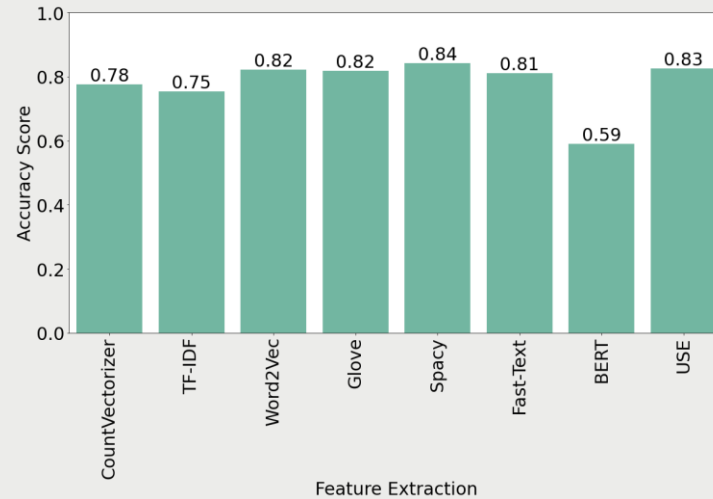
राधे

कृष्ण

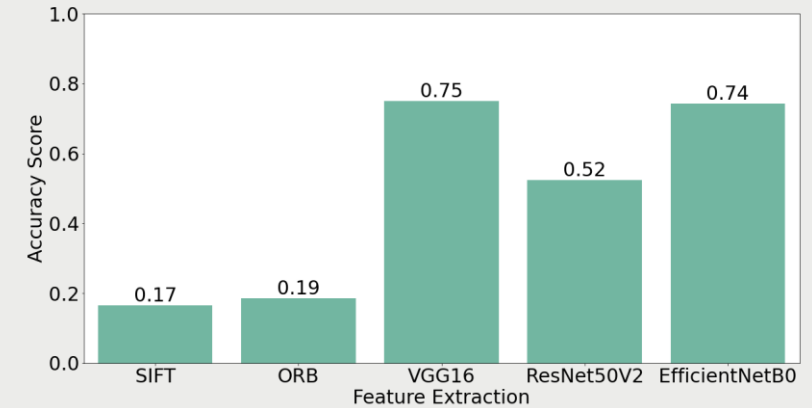
Supervisé

Catégorie : premier nœud (7 catégories) Supervisé

Mean test accuracy score of feature extraction techniques (CV:5)



Mean test accuracy score of feature extraction techniques (CV:5)

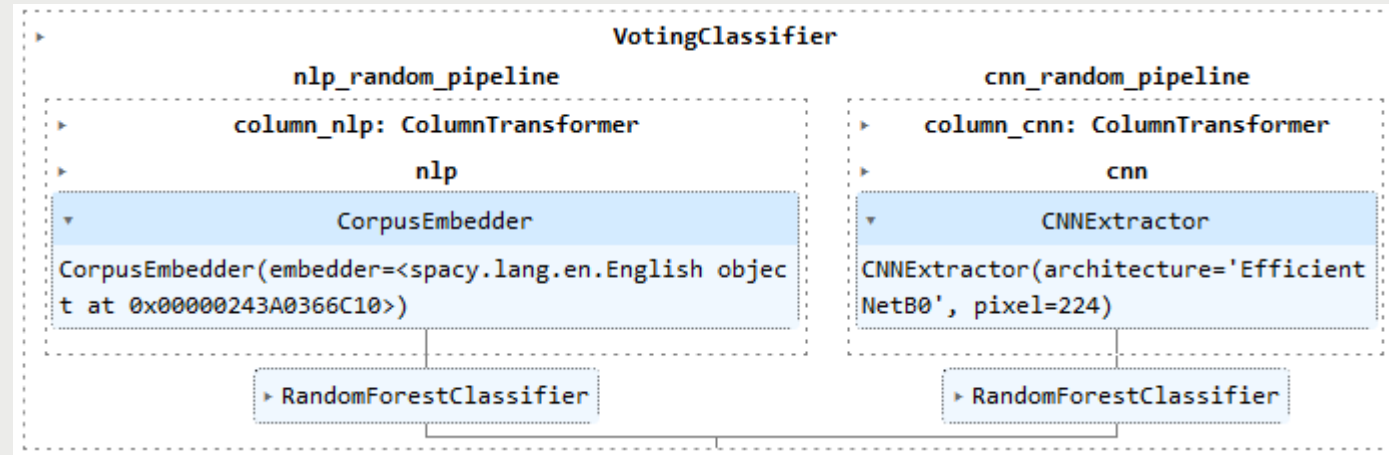


राधे

कृष्ण

Concaténation

Catégorie : premier nœud (7 catégories) Supervisé



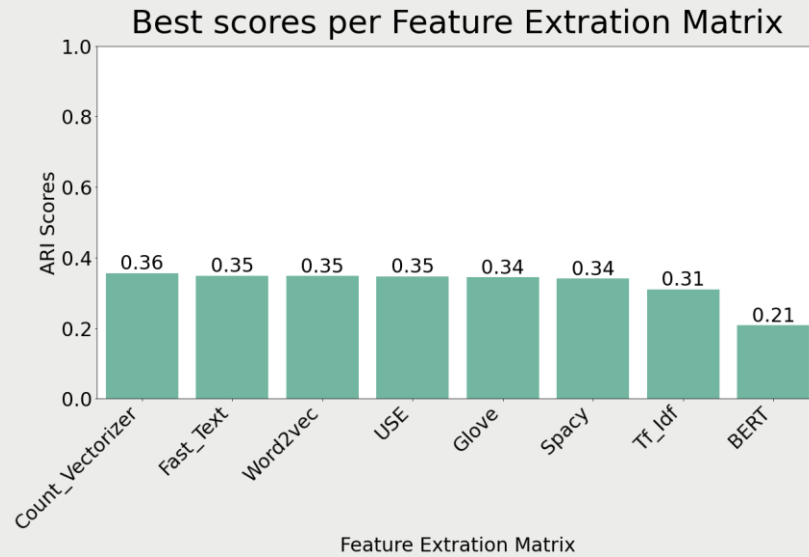
Accuracy : 0,86

राधे

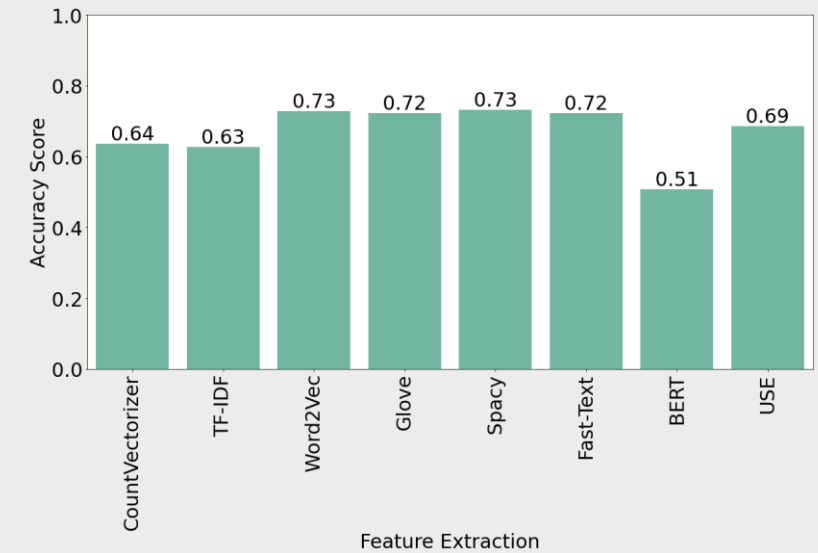
कृष्ण

2^{ème} nœud de catégories

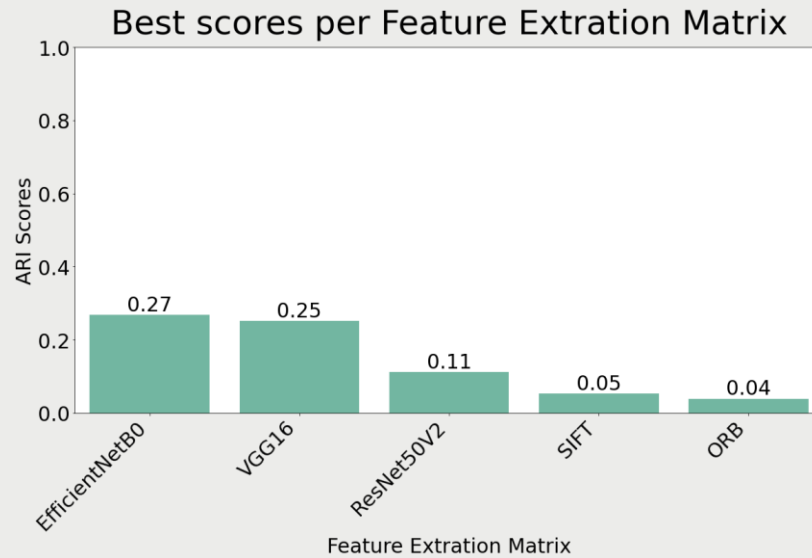
Catégorie : second nœud (62 catégories) - NLP



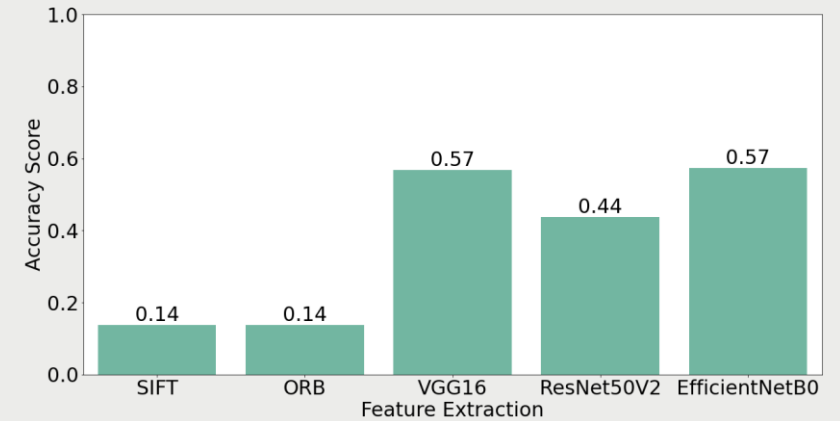
Mean test accuracy score of feature extraction techniques (CV:5)



Catégorie : second nœud (62 catégories) - Computer vision



Mean test accuracy score of feature extraction techniques (CV:5)



Conclusion


Faisable !

1^{er} Nœud de Catégories

Natural Language Processing performe mieux (en supervisé sur le 2^{ème} nœud également !)

Spacy's Embedding => ARI : 0,64 (Kmeans) et Accuracy : 0,84 (Random Forest)



A low-angle shot looking up at several large, golden, ribbed domes of a building, likely a mosque or cathedral. The sky is a pale, hazy blue with soft, wispy white clouds. Numerous birds, possibly pigeons, are captured in flight throughout the sky, their wings spread in various directions. The lighting is bright and warm, suggesting a sunny day. The text "Merci pour votre attention" is centered in the middle of the image.

Merci pour votre attention