



P6- CLASSIFIEZ AUTOMATIQUEMENT DES BIENS DE CONSOMMATION

Parcours Data Scientist – Soutenance du 26 avril 2024

Charlotte Dubus

SOMMAIRE

- I. Missions
- II. Prétraitement des données
- III. Extraction de features
- IV. Conclusion de l'étude de faisabilité
- V. Classification supervisée
- VI. Test de l'API



place de marché

MISSIONS

Problématiques

- Catégorisation manuelle peu fiable
- Volume d'articles faibles : élargir la gamme produit

Objectifs

- Automatiser la classification d'objets
- Faciliter la mise en ligne et la recherche sur le site

Réalisations

- Etude de faisabilité basée sur la description et l'image
- Classification supervisée et collecte de données via API

PRESENTATION DES DONNEES

Jeu de données

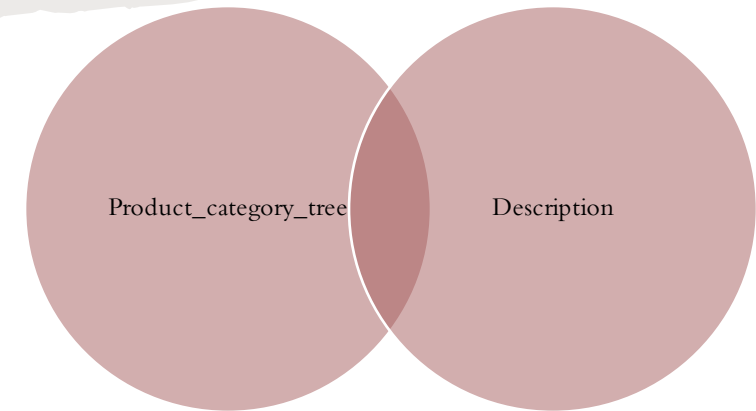
- 1050 lignes et 15 colonnes
- 3 catégories d'informations: identifications, descriptives et financières
- Aucun doublon dans les identifiants uniques

Images

- Dossier avec 1050 représentations de produits

Respect RGPD

- Nos données et images sont exemptes de contraintes de propriété intellectuelle



PRE-TRAITEMENT TEXTUEL : METHODOLOGIE

01

Identification et
sélection des
caractéristiques
essentiels au projet

02

Développement de
fonctions modulaires
pour assurer la
réutilisabilité du
code

03

Nettoyage des
données adapté à
chaque colonne

04

Évaluation
qualitative et
quantitative des
données nettoyées.

05

Transformation du
texte: Tokenisation,
Stematisation,
Lemmatisation

PRE-TRAITEMENT TEXTUEL: PRODUCT_CATEGORY_TREE

home furnishing	150
baby care	150
watches	150
home decor festive needs	150
kitchen dining	150
beauty and personal care	150
computers	150

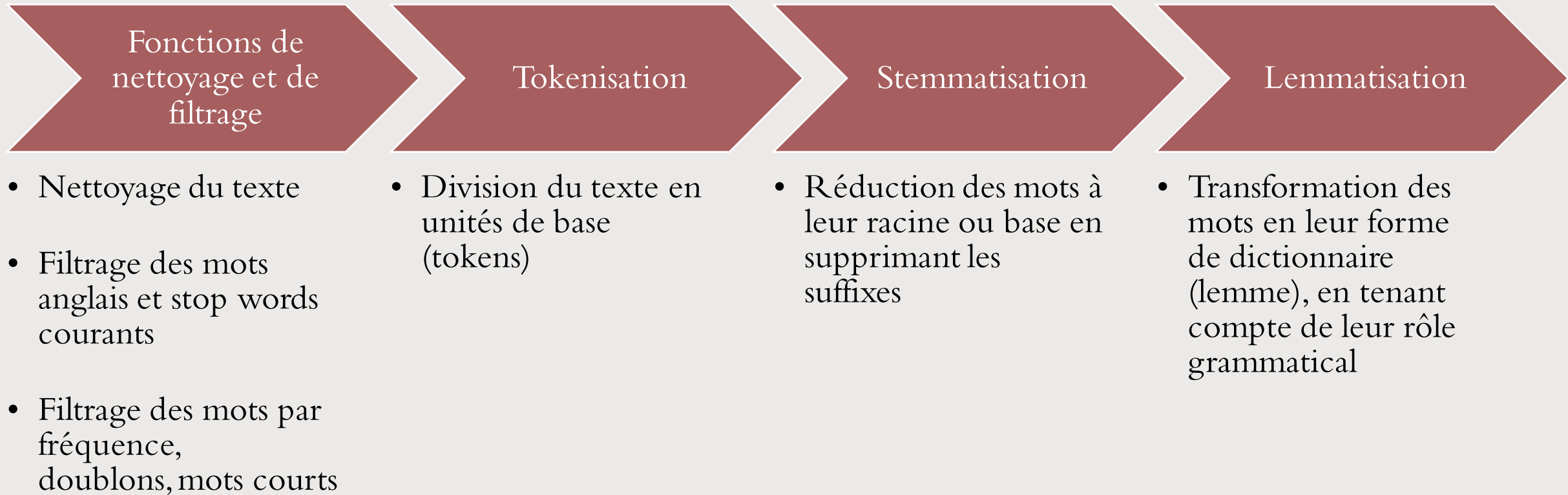
Segmentation
des catégories de
produits

Nettoyage textuel

Traitement itératif

Analyse
comparative des
catégories

PRE-TRAITEMENT TEXTUEL: DESCRIPTION

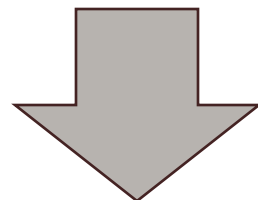


Description:

Nombre de mots dans la description originale:

230

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



Nombre de mots dans la description stemmatisée:

29

Description stemmatisée:

polyest eyelet door curtain floral stitch ring environ ant wrinkl anti shrinkag heart slide smoothli draw apart thing morn welcom sun want wish even creat given sooth amount sunlight

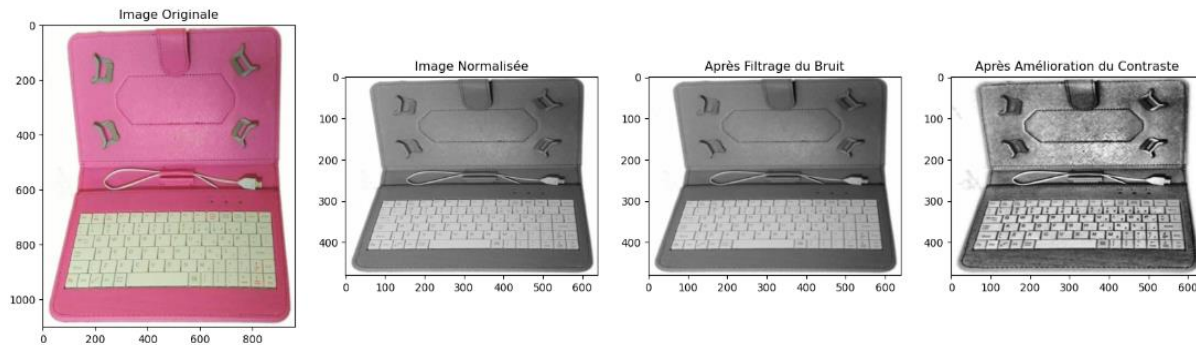
Nombre de mots dans la description stemmatisée:

29

Description lemmatisée:

polyester eyelet door curtain floral stitch ring environment ant wrinkle anti shrinkage heart slide smoothly draw apart thing morning welcome sun want wish evening create give soothe amount sunlight

PRE-TRAITEMENT IMAGE



Création des dossiers

Traitements d'image

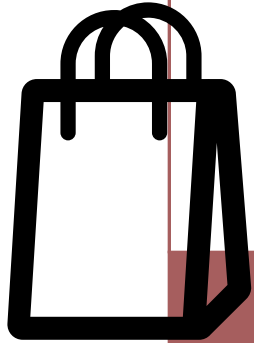
- Conversion en gris et redimensionnement
- Flou gaussien (réduction des bruits)
- Amélioration du contraste avec CLAHE

Visualisation

Traitement de masse

Sauvegarde

EXTRACTION DE FEATURES - TEXTE



Définition : Méthode qui transforme le texte en nombres

Fonctionnement : Chaque mot devient un chiffre et la fréquence du mot est sa valeur.

Bag of Words (BoW)

absorbency	absorbent	accent	access	accessory	accident	accidental	accord
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0

Définition : Mesure l'importance d'un mot dans un texte.

Fonctionnement : Chaque mot devient un chiffre et la fréquence du mot est sa valeur.

TF*IDF

TF*IDF = TERM FREQUENCY * INVERSE DOCUMENT FREQUENCY

Term Frequency – Inverse Document Frequency (TF-IDF)

absorbency	absorbent	accent	access	accessory	accident	accidental	accord
0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.295956	0.0	0.0	0.0	0.0	0.0	0.0	0.0

EXTRACTION DE FEATURES - TEXTE

Word2Vec

- **Définition:** Un modèle qui représente des mots par des vecteurs.
- **Fonctionnement :** Les mots utilisés dans des contextes similaires ont des vecteurs similaires et apprend de la position des mots dans les textes.

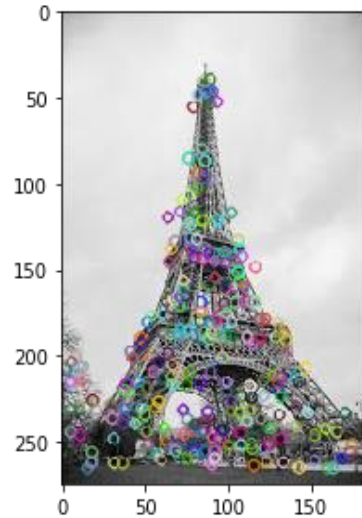
BERT : Bidirectional Encoder Representations from Transformers

- **Définition :** Un modèle avancé qui saisit le contexte des mots dans toutes les directions
- **Fonctionnement :** Analyse les mots en fonction de leurs voisins à gauche et à droite et utilise les transformateurs pour traiter le texte

USE : Universal Sentence Encoder

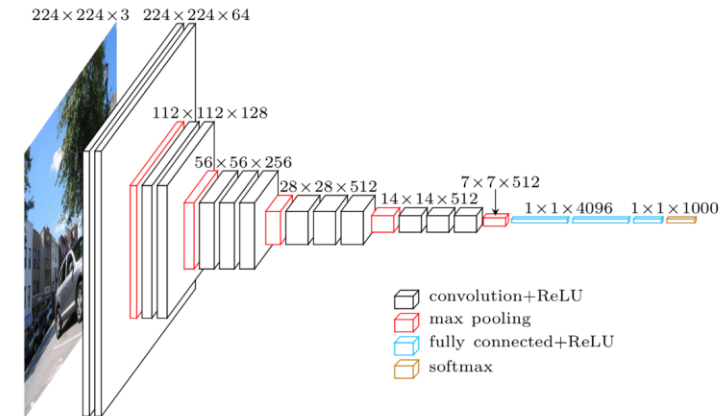
- **Définition :** Un modèle qui transforme les phrases en vecteurs denses
- **Fonctionnement :** Capte le sens global des phrases, pas seulement des mots isolés et utilise un apprentissage profond pour l'encodage de texte.

EXTRACTION DE FEATURES - IMAGES



SIFT

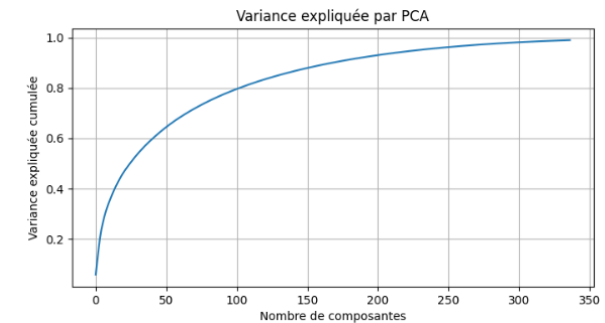
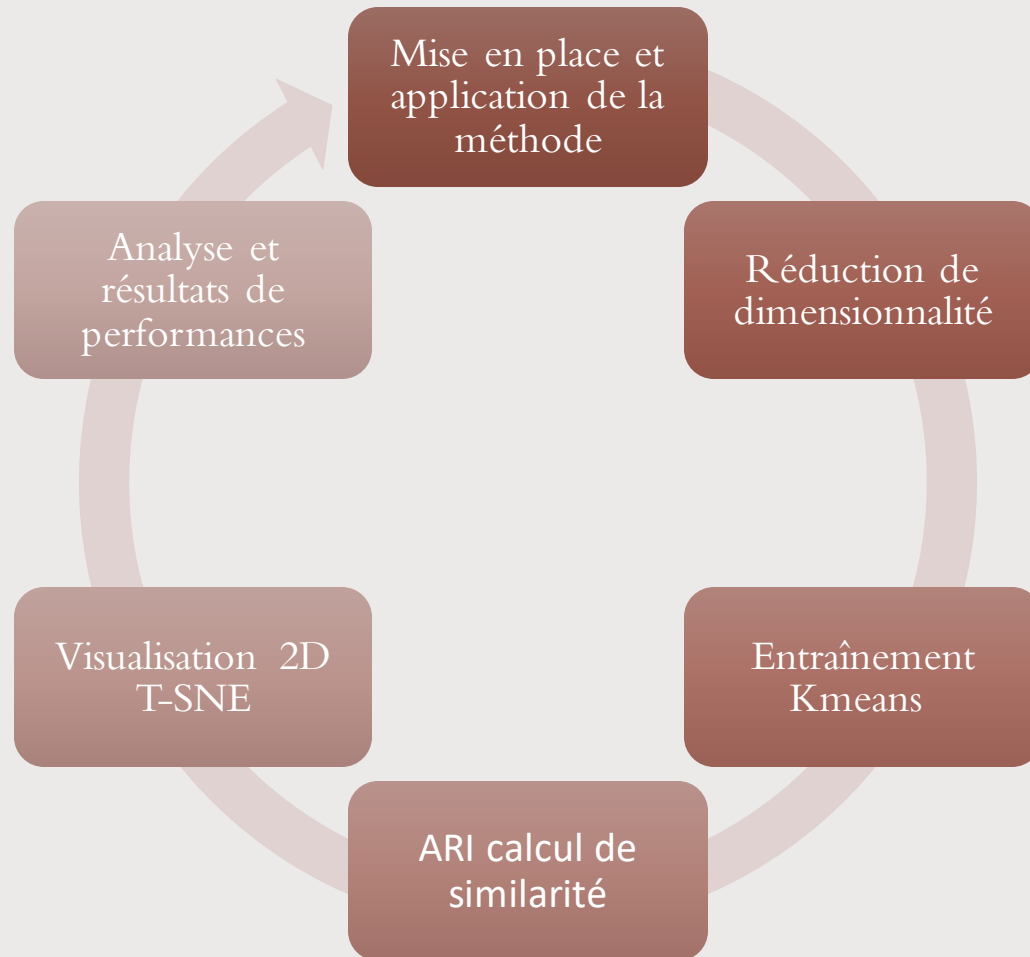
- Définition : Technique de détection et description de points d'intérêt dans les images.
- Fonctionnement : Détection multi-échelle, localisation précise, orientation et descripteurs des keypoints.
- Utilisation : Comparaison de points entre images pour reconnaissance d'objets et correspondance d'images.



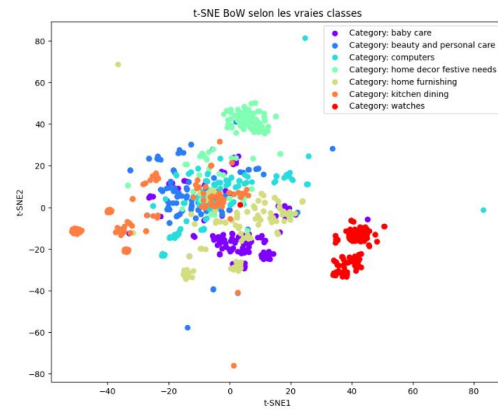
VGG16

- Définition : Modèle de réseau de neurones convolutif profond pour la classification d'images
- Fonctionnement : Couches de convolution, Couches de pooling, Couches entièrement connectées, Softmax
- Utilisation: Pré-entraîné sur ImageNet pour le transfert d'apprentissage sur de nouvelles tâches de classification d'images.

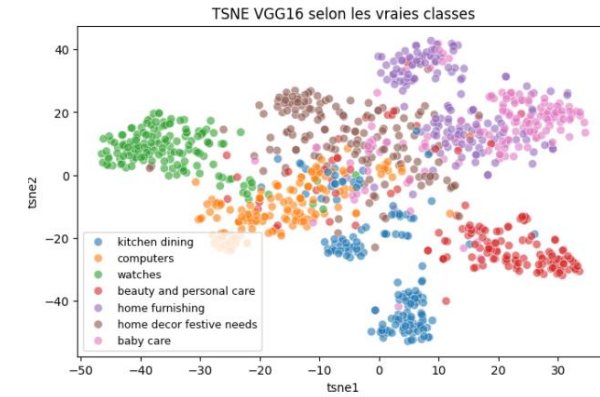
CONCLUSION ETUDE DE FAISABILITE



CONCLUSION ETUDE DE FAISABILITÉ

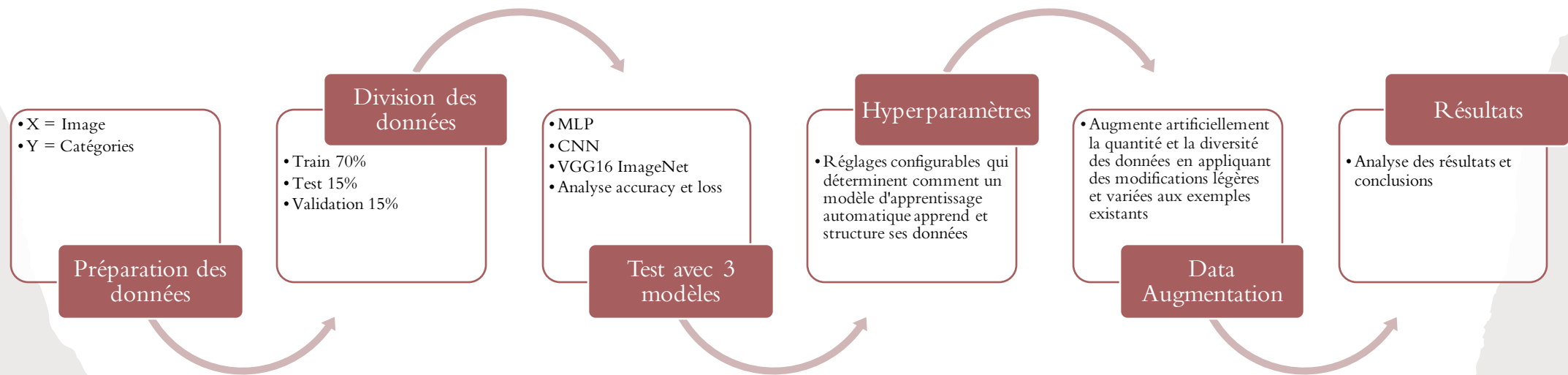


Méthodes avec PCA		Score ARI
0	BoW	0.382604
1	TF-IDF	0.193265
2	W2v	0.194719
3	BERT	0.144197
4	USE	0.280026
5	SIFT	0.063957
6	VGG16	0.453247



- **Résultats Graphiques :** L'interprétation visuelle des graphiques confirme la possibilité de catégoriser les images et les descriptions en fonction de leurs classes attribuées.
- **Prochaines Étapes :** Forts de ces résultats prometteurs, nous orienterons nos efforts vers la mise en œuvre de la classification supervisée, en utilisant les images et en explorant des architectures avancées de CNN.

CLASSIFICATION SUPERVISEE



CLASSIFICATION SUPERVISEE

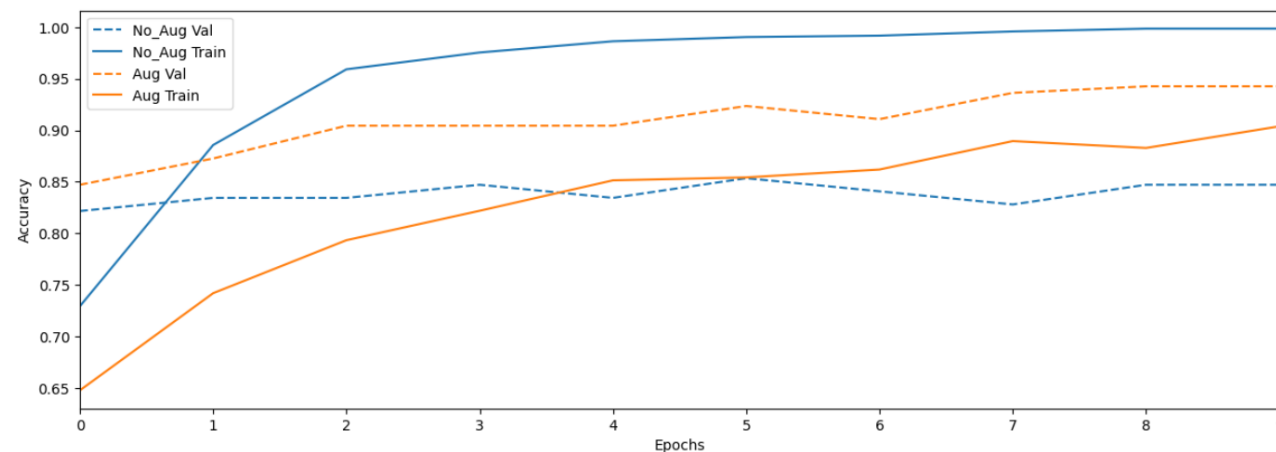
	Modèles	Test Loss	Test Accuracy
0	MLP	1.673886	0.481013
1	CNN	1.422901	0.556962
2	VGG 16 ImageNet	0.650294	0.810127
3	VGG 16 ImageNet + Data Aug	1.034525	0.911392

Meilleure
performance:

VGG16 +
Data
Augmentation

Courbe :

Stabilité et
amélioration
avec la Data
Augmentation



TEST API



Code Snippets

Example Responses

Results

(Python) Requests ▾

Copy Code

```
import requests
```

```
url = "https://edamam-food-and-grocery-database.p.rapidapi.com/api/food-database/v2/parser"
```

```
querystring = {"nutrition-type": "cooking", "category[0]": "generic-foods", "health[0]": "alcohol-free"}
```

```
headers = {
```

```
    "X-RapidAPI-Key": "9336f841e6msh32a18267260f9f6p123bb2jsnb3dc1a54f02c",
```

```
    "X-RapidAPI-Host": "edamam-food-and-grocery-database.p.rapidapi.com"
```

Aperçu des 10 premières données filtrées du dataframe

df.head(10)

	foodId	label	category	foodContentsLabel	image
0	food_a656mk2a5dmqb2adlamu6beihduu	Champagne	Generic foods	NaN	https://www.edamam.com/food-img/a71/a718cf3c52...
1	food_b753ithamdb8psbt0w2k9aquo06c	Champagne Vinaigrette, Champagne	Packaged foods	OLIVE OIL; BALSAMIC VINEGAR; CHAMPAGNE VINEGAR...	NaN
2	food_b3dyababjo54xobm6r8jzbghjgqe	Champagne Vinaigrette, Champagne	Packaged foods	INGREDIENTS: WATER; CANOLA OIL; CHAMPAGNE VINE...	https://www.edamam.com/food-img/d88/d88b64d973...
3	food_a9e0ghsamvoc45bwa2ybsa3gken9	Champagne Vinaigrette, Champagne	Packaged foods	CANOLA AND SOYBEAN OIL; WHITE WINE (CONTAINS S...	NaN
4	food_an4jueaucpus2a3u1ni8auhe7q9	Champagne Vinaigrette, Champagne	Packaged foods	WATER; CANOLA AND SOYBEAN OIL; WHITE WINE (CON...	NaN
5	food_bmu5dmkazwvupaa5prh1daa8jxs0	Champagne Dressing, Champagne	Packaged foods	SOYBEAN OIL; WHITE WINE (PRESERVED WITH SULFIT...	https://www.edamam.com/food-img/ab2/ab2459fc2a...
6	food_alpl44taoyv11ra0lic1qa8xculi	Champagne Buttercream	Generic meals	sugar; butter; shortening; vanilla; champagne;...	NaN
7	food_byap67hab8evc3a0f9w1oag3s0qf	Champagne Sorbet	Generic meals	Sugar; Lemon juice; brandy; Champagne; Peach	NaN
8	food_am5egz6aq3fjla8xpkdbc2asis	Champagne Truffles	Generic meals	butter; cocoa; sweetened condensed milk; vanil...	NaN
9	food_bcz8rhiajk1fuva0vldmeakbouc0	Champagne Vinaigrette	Generic meals	champagne vinegar; olive oil; Dijon mustard; s...	NaN



MERCI POUR VOTRE ATTENTION