Parcours AI Engineer

SOUTENANCE PROJET 9

“Développer une Preuve de Concept”

Stéphanie Duhem - Février 2025

00. CONTEXTE & DÉROULÉ DU PROJET

LE CONTEXTE

Marketplace "Place de Marché" :

- Utilisation de la classification automatique des annonces via les images des produits.
- Modèle DenseNet121 efficace.

Veille technologique révélant une nouvelle approche :

- Nouvelle méthode potentiellement plus performante.
- Association des images et des textes des descriptions produits.
- Évaluation de cette nouvelle approche.

LES GRANDES ÉTAPES DU PROJET

- La documentation ayant servie de point de départ
- L'analyse exploratoire du dataset
- La comparaison des modèles (baseline et nouvel algorithme)
- Le dashboard de présentation des résultats
- Les limites et les améliorations possibles

01. DOCUMENTATION

Article principal développant l'approche CLIP

- CLIP ("Contrastive Language-Image Pre-training") approche multimodale Texte+Image proposé par Open.Ai - 5 janvier 2021 : « *Learning Transferable Visual Models From Natural Language Supervision* »

Article sur l'explicabilité du modèle CLIP (textes et images)

- Sepideh Mamooler - 2021 : https://github.com/sMamooler/CLIP_Explainability/blob/main/CLIP_Explainability.pdf

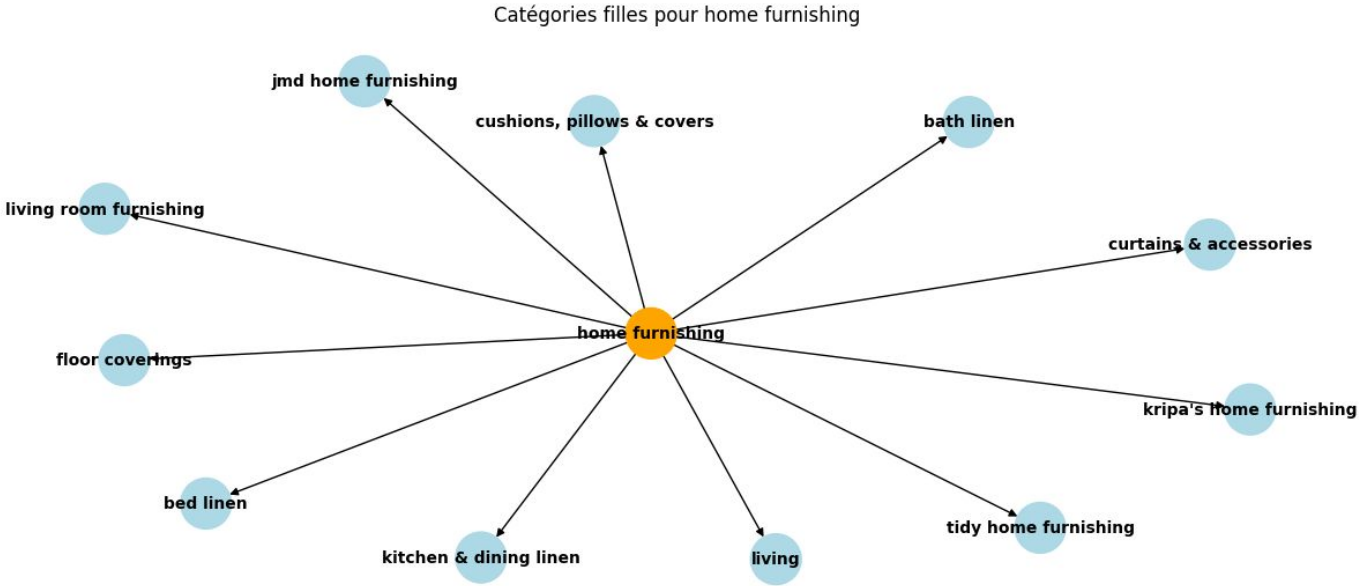
Sources GIT pour l'explicabilité de CLIP

- Sepideh Mamooler - 2021 : https://github.com/sMamooler/CLIP_Explainability
- Shashwat Trivedi : https://github.com/shashwattrivedi/Attention_visualizer?tab=readme-ov-file#readme
- Hila Chefer : https://github.com/hila-chefer/Transformer-MM-Explainability/tree/main

02-A. Analyse exploratoire globale

LES DONNÉES

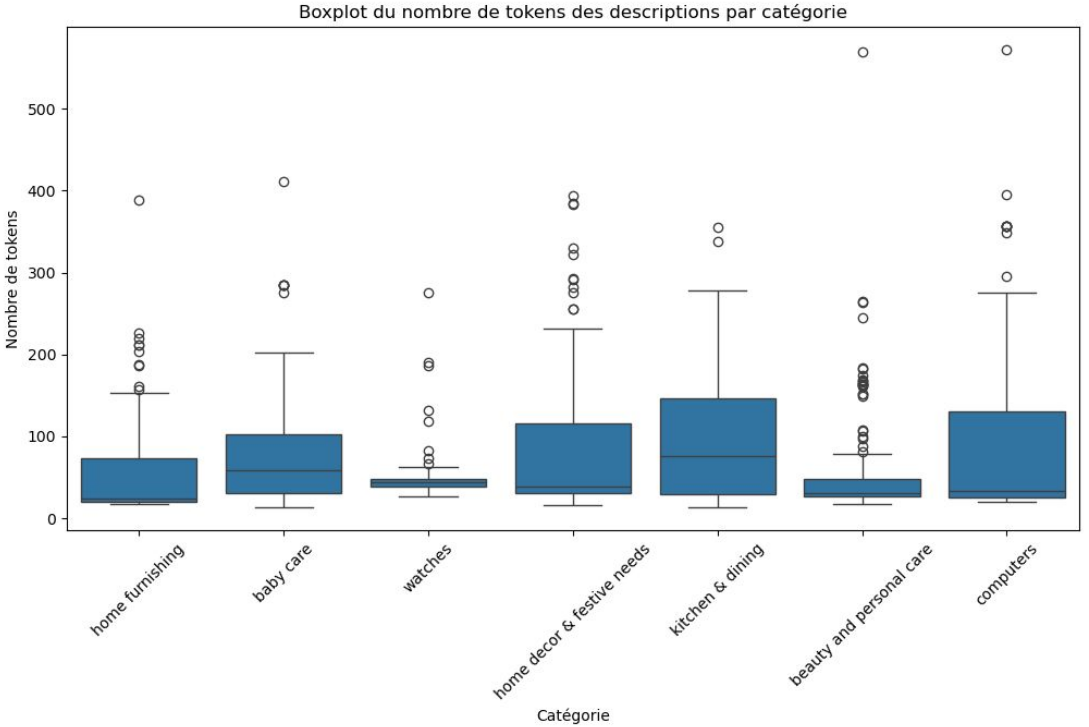
- 1050 articles classés par catégories avec leurs descriptions et leurs images
- Il y a 7 catégories principales (150 articles par catégories) qui ont chacune beaucoup de sous-catégories de produits
  - certaines catégories mères ont jusqu'à 5 niveaux de catégories filles
  - pour la classification, utilisation des 7 catégories principales



02-B. Analyse exploratoire des TEXTES

LES DESCRIPTIONS

- certaines catégories ont des descriptions plus variées avec un champ lexical bien plus importants que d'autres
- beaucoup de mots en commun entre certaines catégories



02-C. Analyse exploratoire des IMAGES

LES DONNÉES

- 1050 articles classés par catégories avec leurs images
- Dans l'ensemble plutôt bonne qualité d'images (majorité d'articles détourés sur fond blanc, image nette)
- Dimensions d'images très variables même au sein d'une même catégories.

LE PRÉ-TRAITEMENT

=› Rajout de bande de pixel sur les côtés pour obtenir des images carrées (pour les CNN)

Home furnishing

Watches

Baby cares

Home decor & festive needs

Kitchen & dining

Beauty and personal care

Computers

03-A. Comparaison des modèles - Méthodologie

Préparation des données

- **Images :** Mise au carré pour éviter la déformation lors du redimensionnement.
- **Textes (pour CLIP) :**
  - Option 1 : Textes bruts sans modification.
  - Option 2 : Tokenisation avec RegexTokenizer de NLTK.

Modélisation

- DenseNet121 : Utilisation de 4 GPU.
- CLIP : Modélisé uniquement avec le CPU (support CUDA non fonctionnel).
- Sélection de l'encodeur Vision Transformer (ViT-B/32) pour CLIP.

Comparaison des modèles

- **Baseline :** DenseNet121 uniquement sur les images.
- **CLIP 'model A' :** Images et textes sans transformation.
- **CLIP 'model B' :** Images et textes tokenisés avant l'envoi au modèle.

Évaluation / Dashboard sur Streamlit

- Taux de classification.
- Accuracy.
- Temps de calcul nécessaire.
- Analyse des erreurs de classification.

03-B. Comparaison des modèles - Résultats

	Scenario	Description	Accuracy_Train	Accuracy_Test	Durée_totale_computation
0	Baseline_DenseNet121	AVEC poids imagenet & SANS data-augmentation	97.533631	88.607597	2 min 54 sec
1	TEST_CLIP_model_A	SANS tokenisation NLTK et troncature avec CLIP...	96.547619	96.666667	2 min 42 sec
2	TEST_CLIP_model_B	AVEC tokenisation NLTK et troncature avec CLIP	95.833333	95.714286	2 min 45 sec

CONCLUSIONS

Le meilleur modèle est le CLIP model-A, sans préparation préalable des textes.

Malgré l'absence d'utilisation de GPU pour les approches CLIP elles sont moins coûteuses que celle du DenseNet121 avec support CUDA.

DASHBOARD

share.streamlit.io

04. Limites et améliorations possibles

Exploration des erreurs de CLIP

Prévoir une étude avec un **opérateur humain,** sur les erreurs de CLIP :

- en cas de **doute** sur la catégorie :
  - travail de révision des catégories
  - dans l'interface de dépot d'annonce, courtes recommandations aux vendeurs sur certains mots-clés dans les descirption pour les catégories principales
- en cas de **certitude** :
  - réentraînement du modèle (paramètres d'entraînements)
  - niveau de gris sur les images pur concentrer l'information sur la forme
  - ajouter du poids sur la présence de certains mots-clés pour augmenter la probabilité d'appartenance à une catégorie

Merci de votre attention