Mission 6 : Étude de Faisabilité

Classification automatique de produits e-commerce

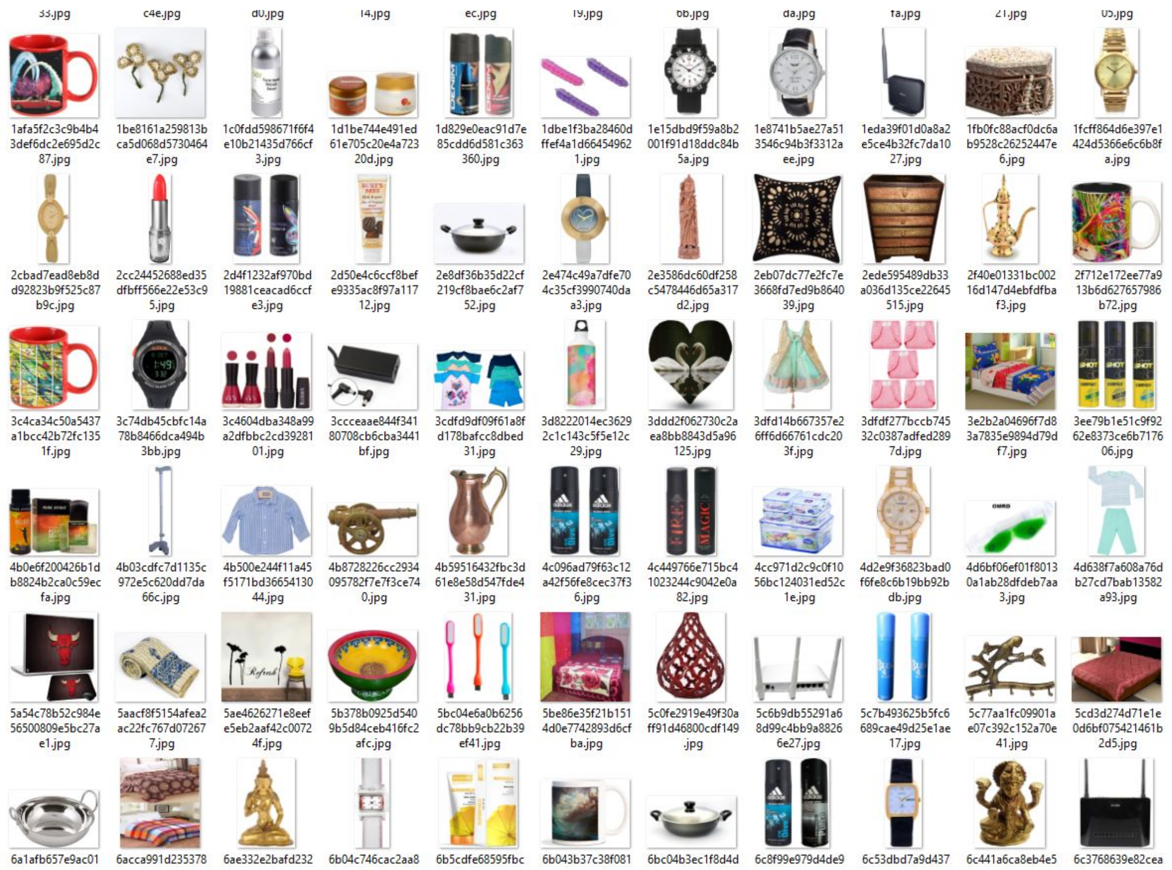
Adrien NORMAND

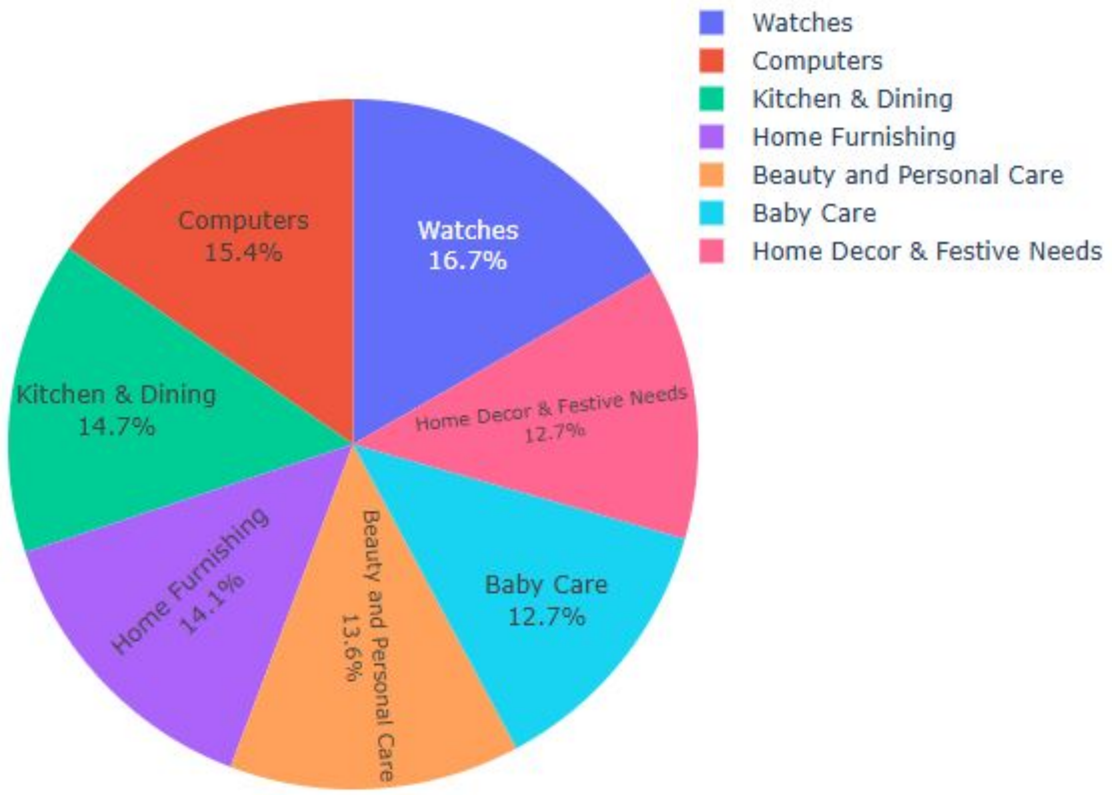
Janvier 2026

Contexte & Objectifs

- Client : Place de marché e-commerce en croissance
- Problème : Classification manuelle ⟶ non scalable
- Objectif : Automatiser la catégorisation via IA
- Approche : Étude de faisabilité sur données réelles
- Données : Descriptions textuelles + Images produits

Dataset Flipkart



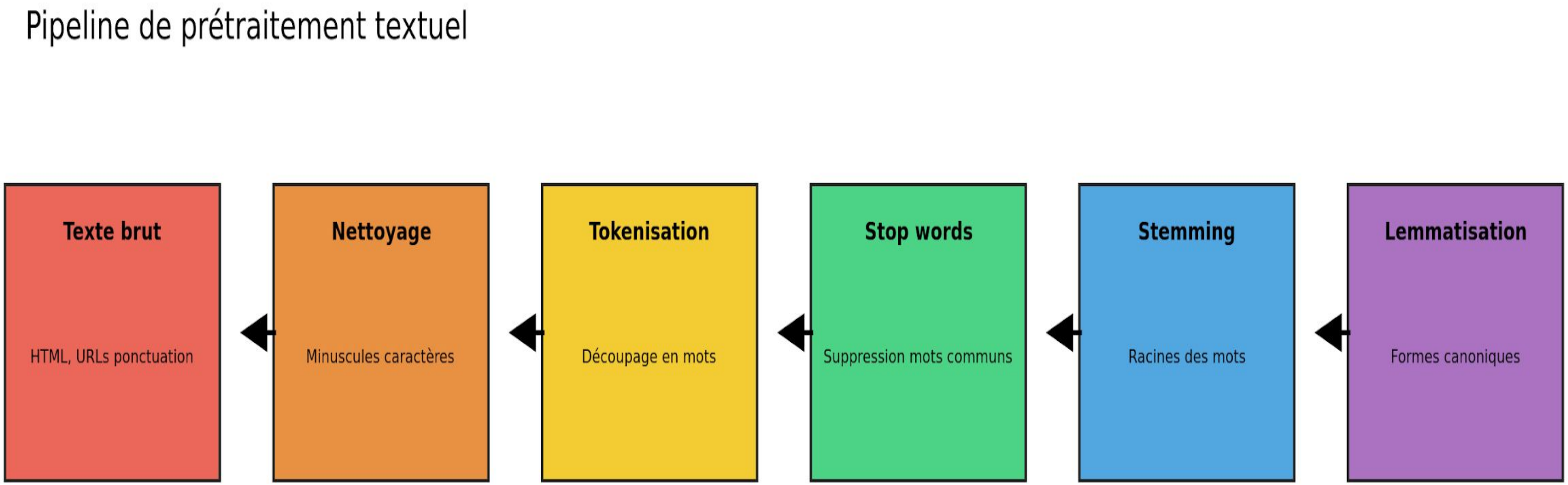


Vérification des droits IP et Copyright

Source des données : Dataset Flipkart (Kaggle)

- **Licence :** Usage éducatif et de recherche autorisé
- **Contexte :** Images provenant d'un site e-commerce public
- **Usage prévu :** Étude de faisabilité académique uniquement
- **Précautions :** Aucune redistribution commerciale / Attribution de la source (Flipkart via Kaggle) / Utilisation limitée au cadre du projet Mission 6

Pipeline de prétraitement textuel

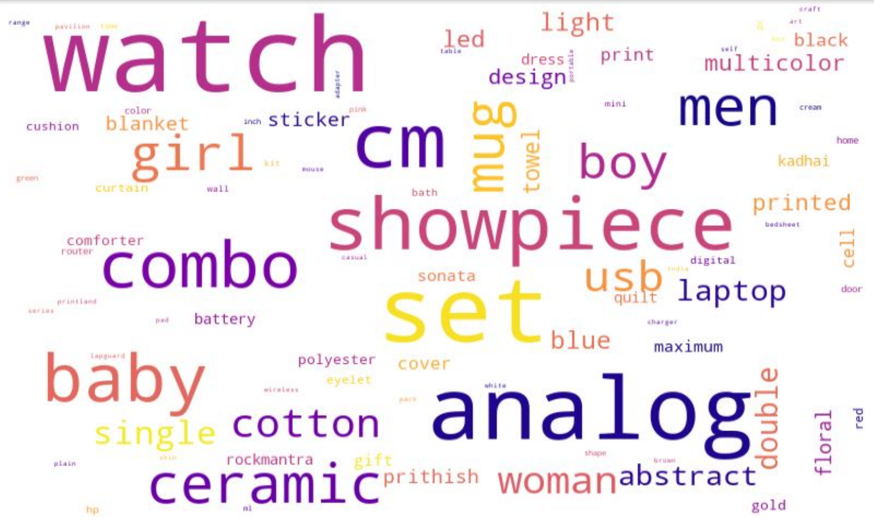


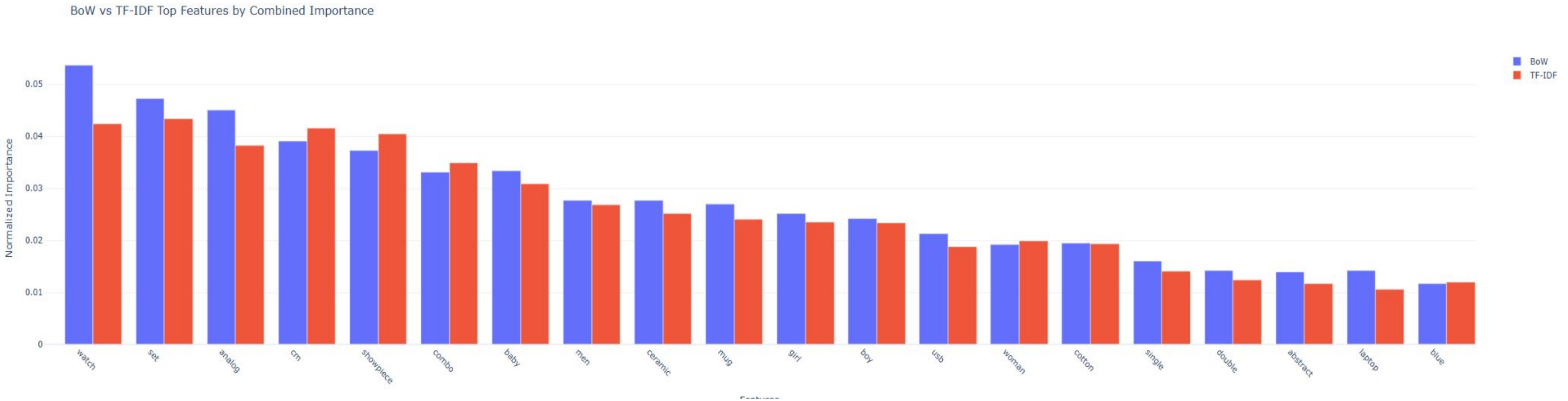
Exemple de prétraitement

- Texte original : "The RUNNING shoes are BETTER than expected!!!"
- Après nettoyage : "the running shoes are better than expected"
- Après tokenisation : ["the", "running", "shoes", "are", "better", "than", "expected"]
- Après stop words : ["running", "shoes", "better", "expected"]
- Après lemmatisation : ["run", "shoe", "good", "expect"]

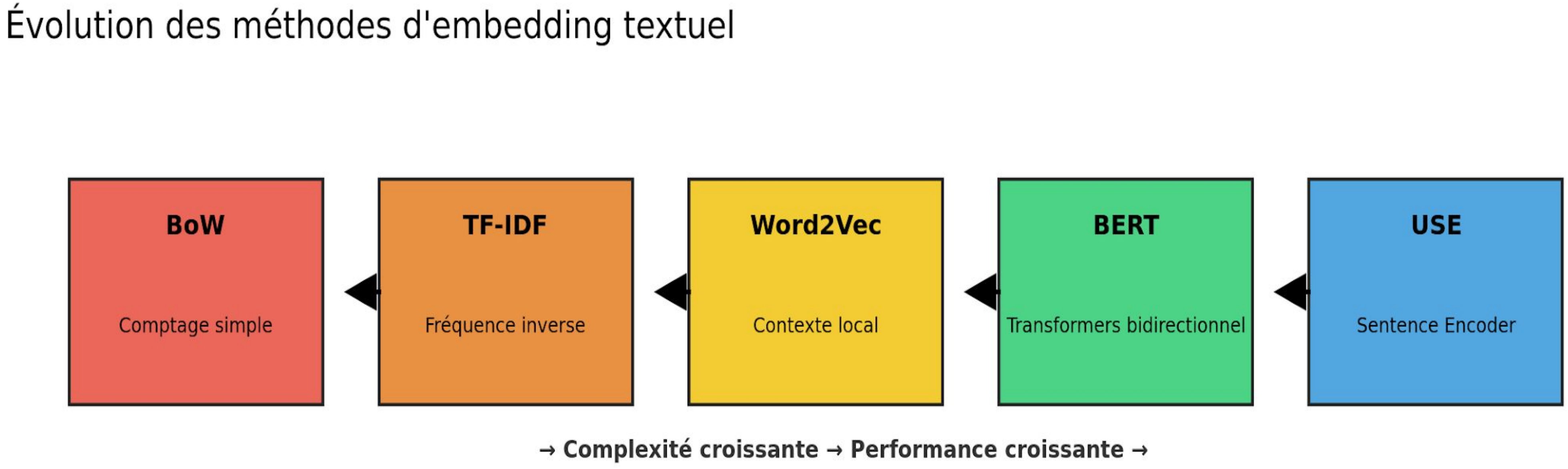
Extraction de features textuelles (BoW & TF-IDF)

- Bag of Words (BoW) : comptage simple des mots
- TF-IDF : pondération par fréquence inverse du document
- Paramètres : min_df=5, max_df=0.95 (seuil de fréquence)
- Vocabulaire final : ~2000 features après filtrage

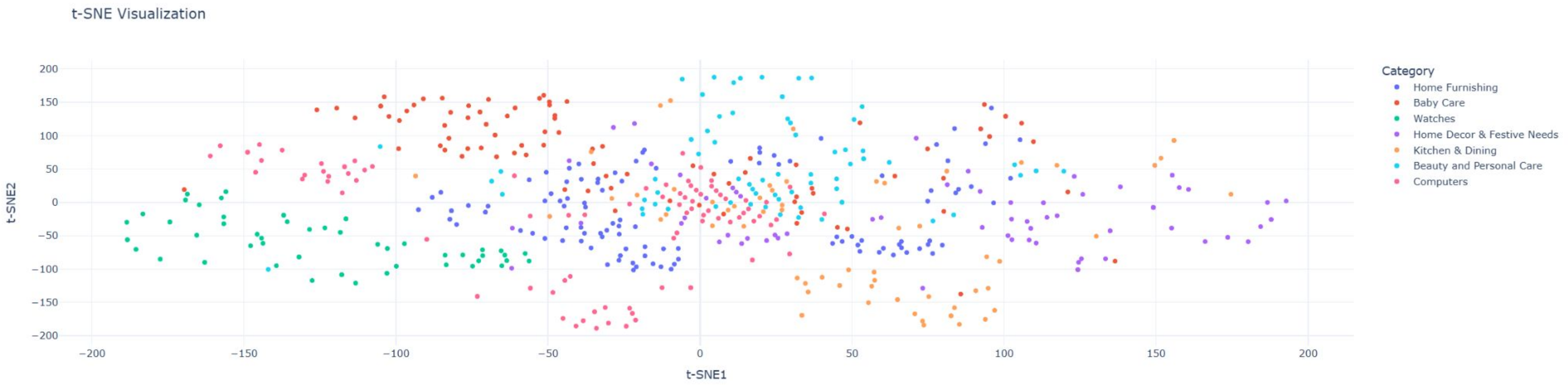


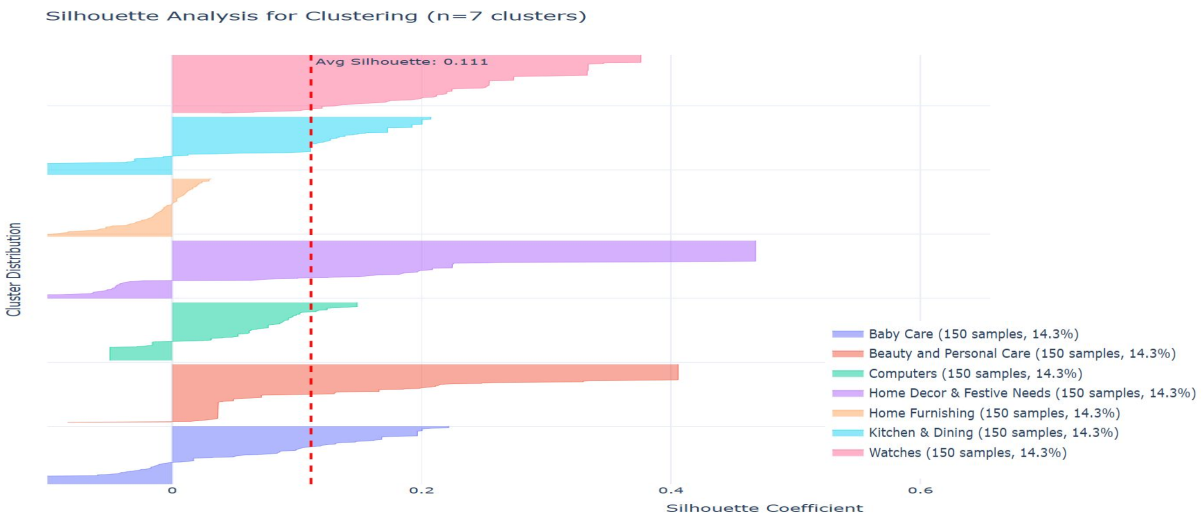


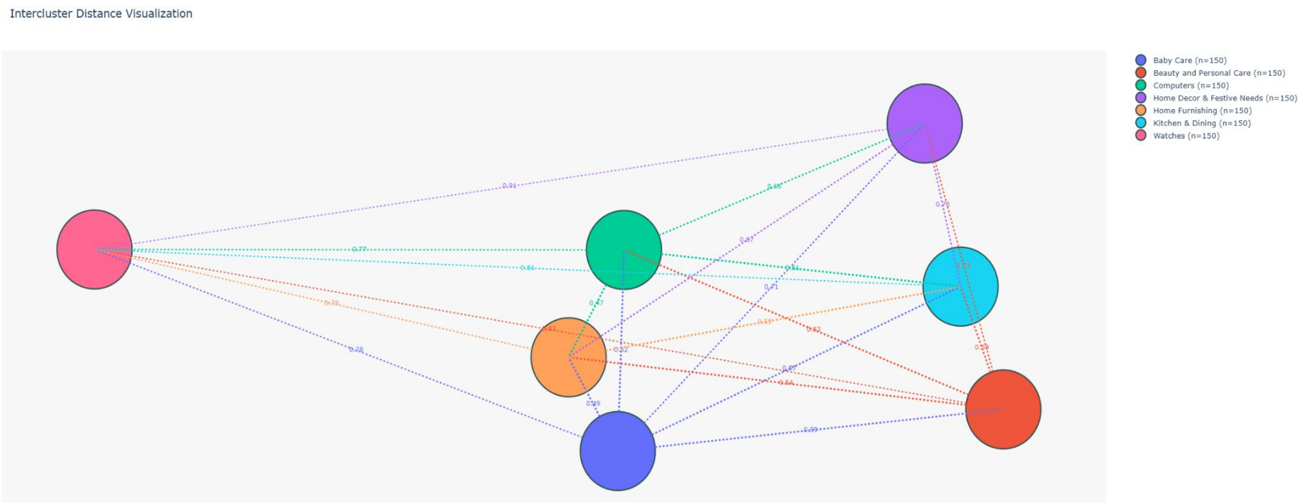
Embeddings avancés : Word2Vec, BERT, USE



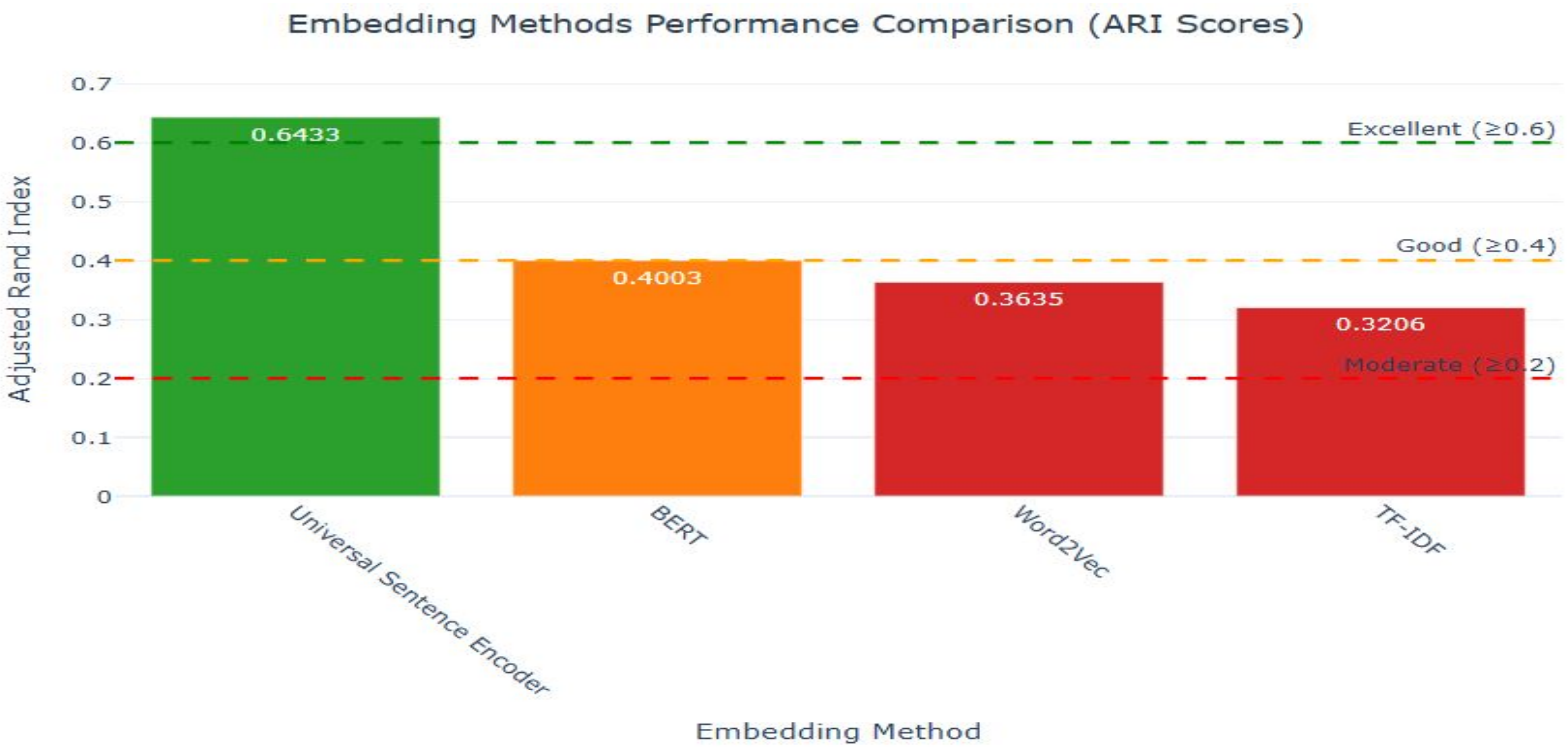
Visualisation t-SNE - Embeddings TF-IDF



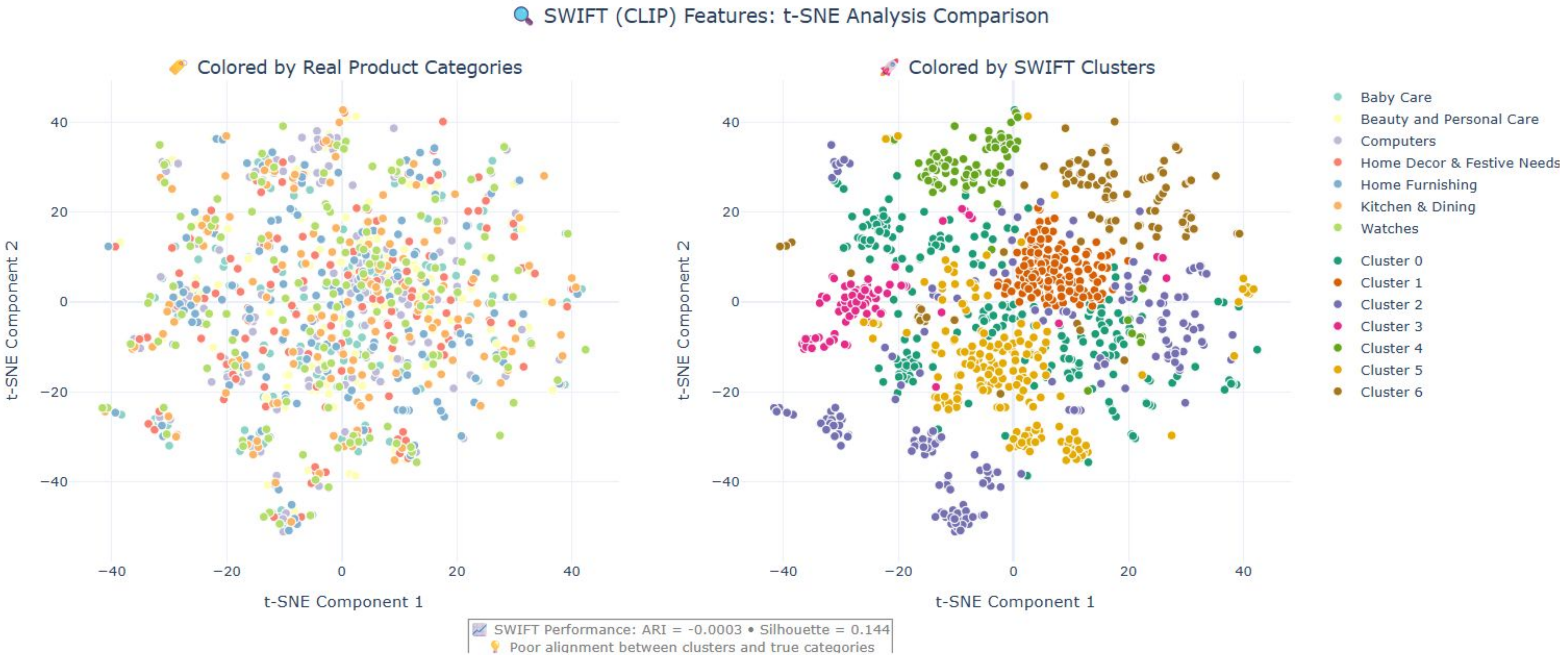




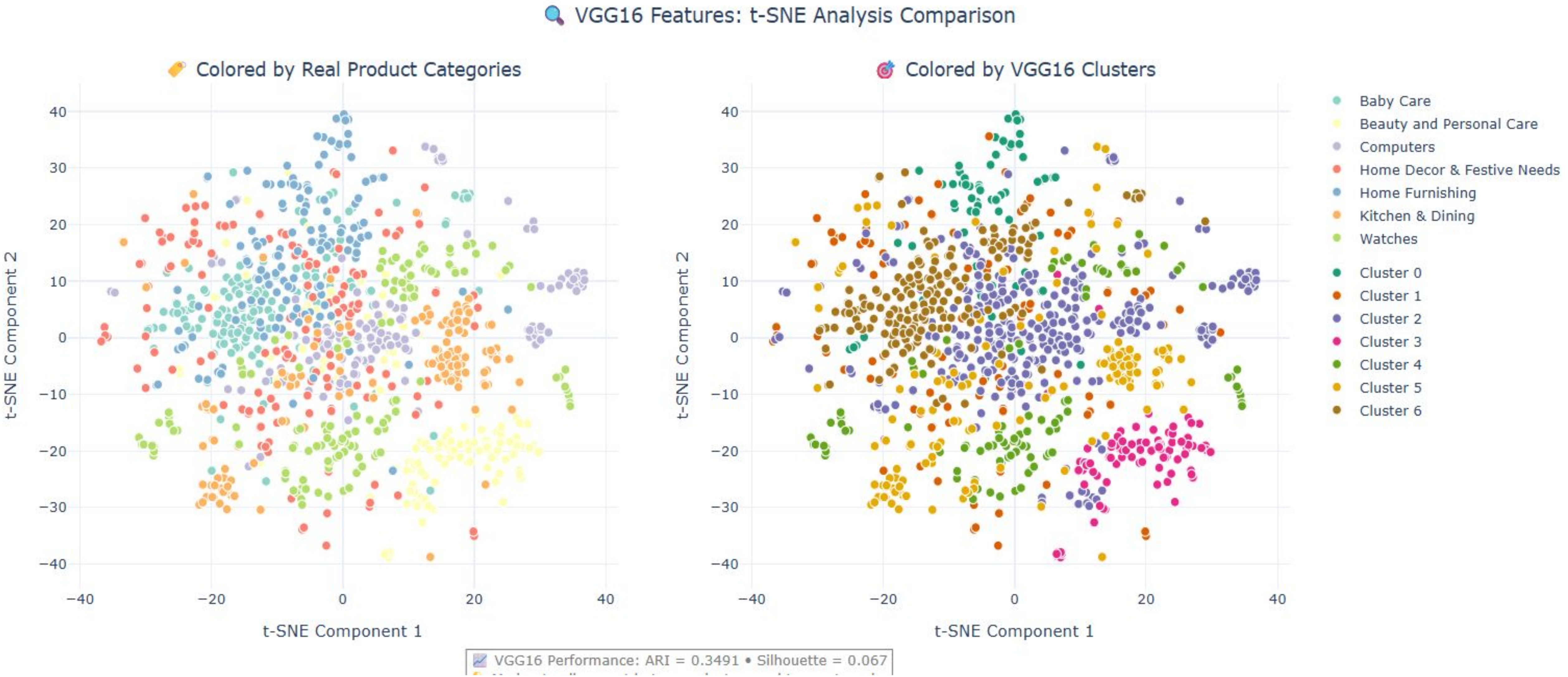
Évaluation du clustering K-Means



Bag-of-Images (Swift)

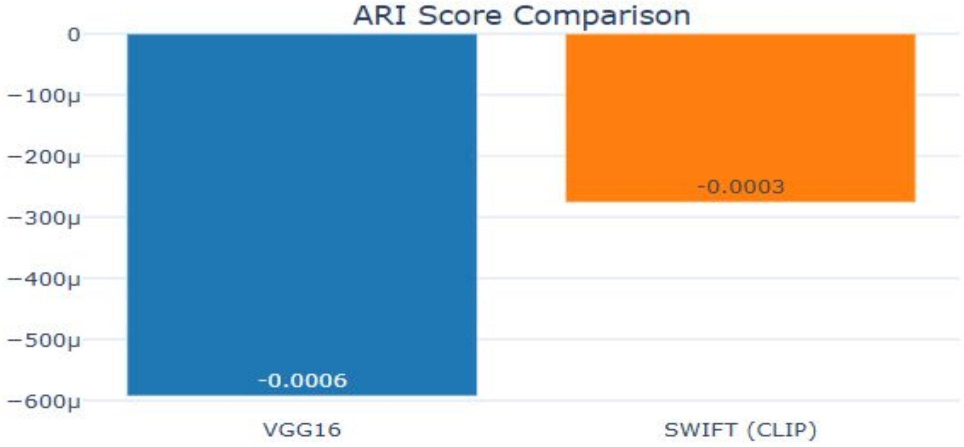


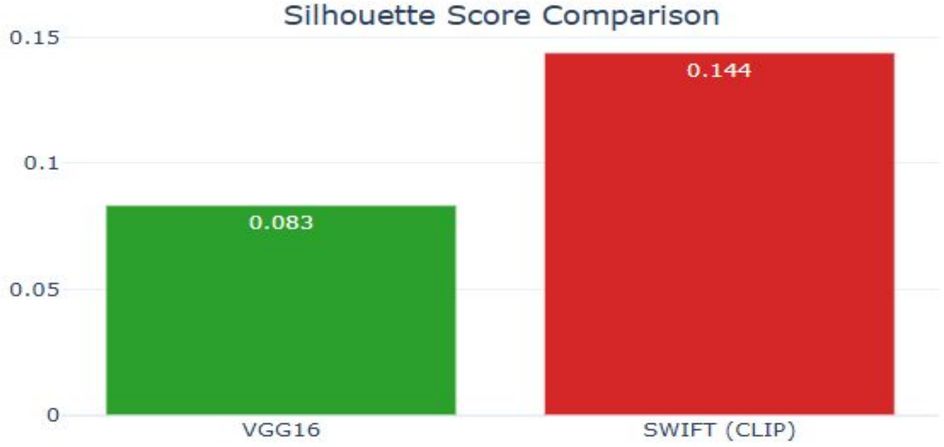
VGG16 Features analysis



Features analysis conclusion

🔍 VGG16 vs SWIFT (CLIP) Features Extraction Performance Comparison

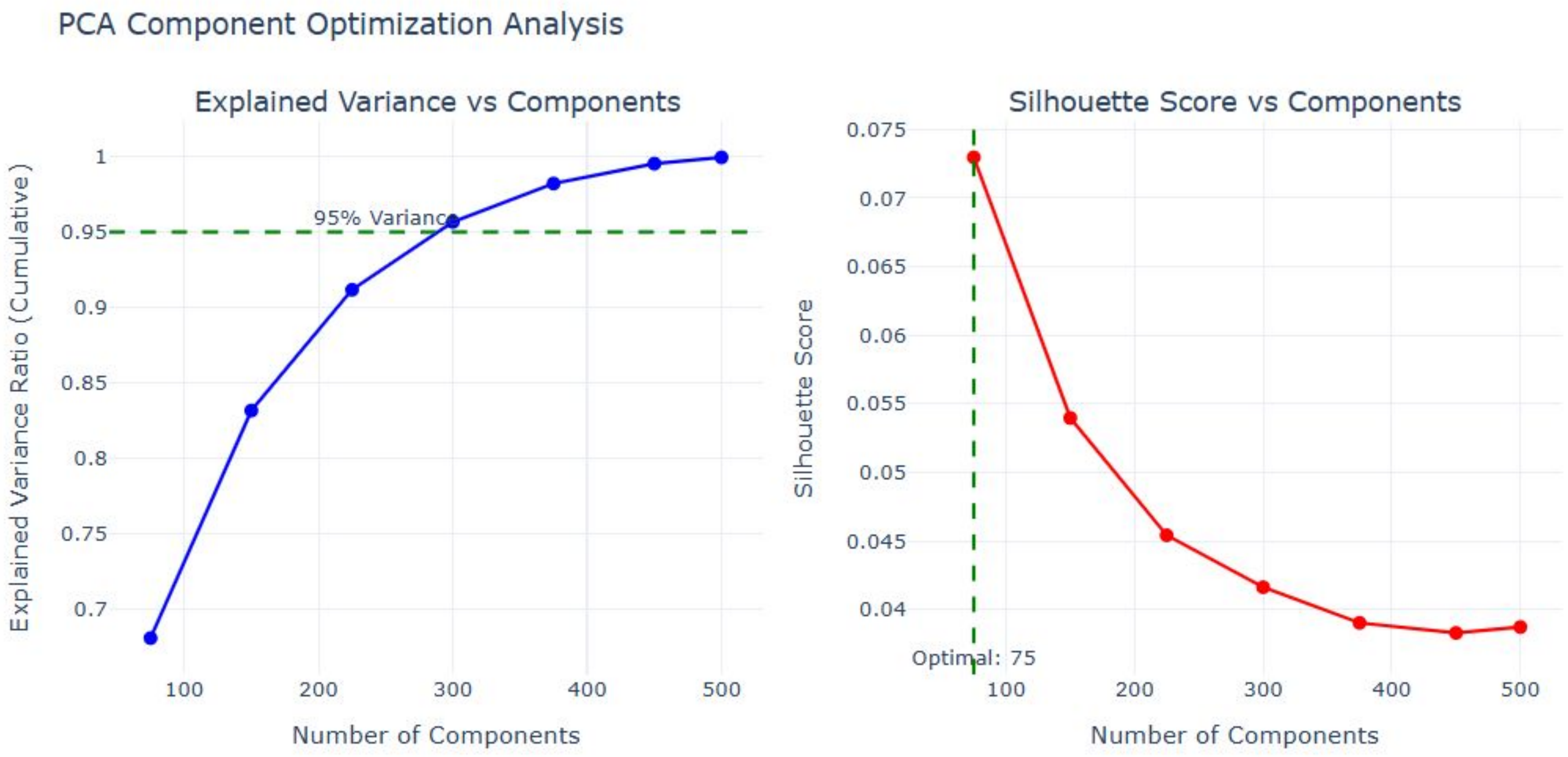




Processing Performance

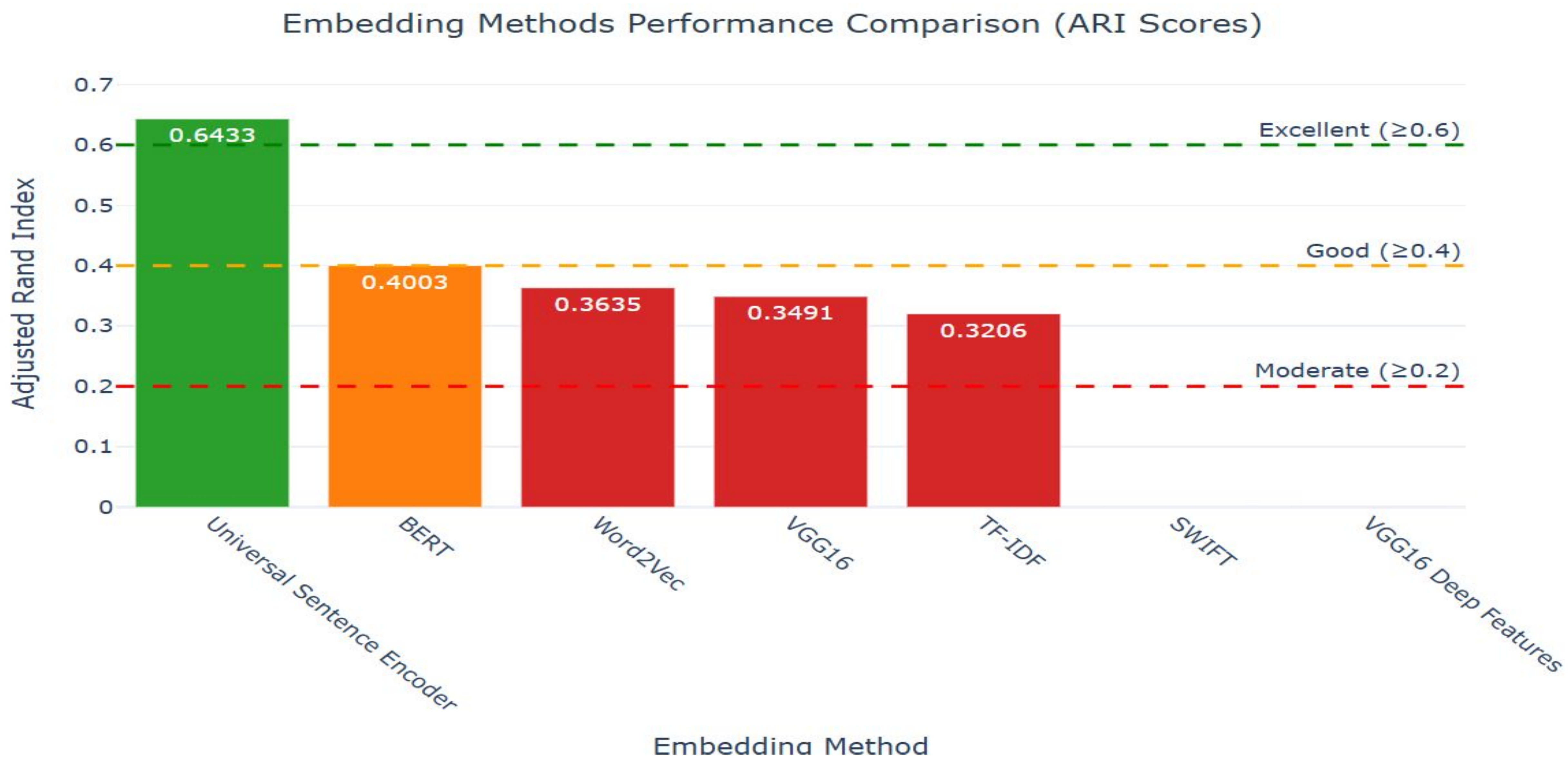
Method	ARI Score	Silhouette	PCA Dims	Original Dims	Categories
VGG16	-0.0006	0.083	150	25088	7
SWIFT (CLIP)	-0.0003	0.144	75	512	7

VGG16 unsupervised analysis




```
Applying PCA to reduce dimensions from 512 to 75...
PCA completed: 68.11% of variance preserved
```

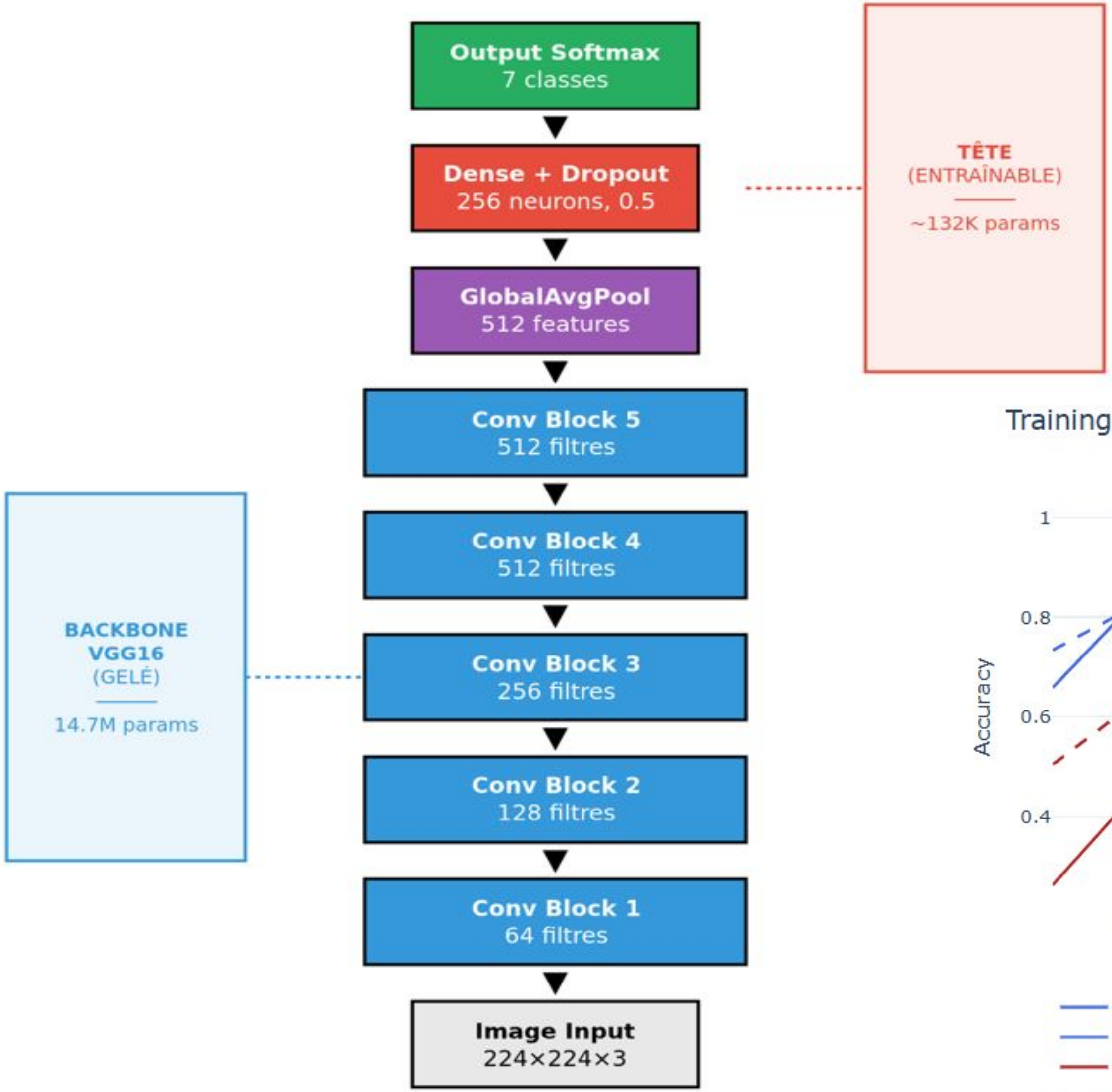
ARI comparison conclusion

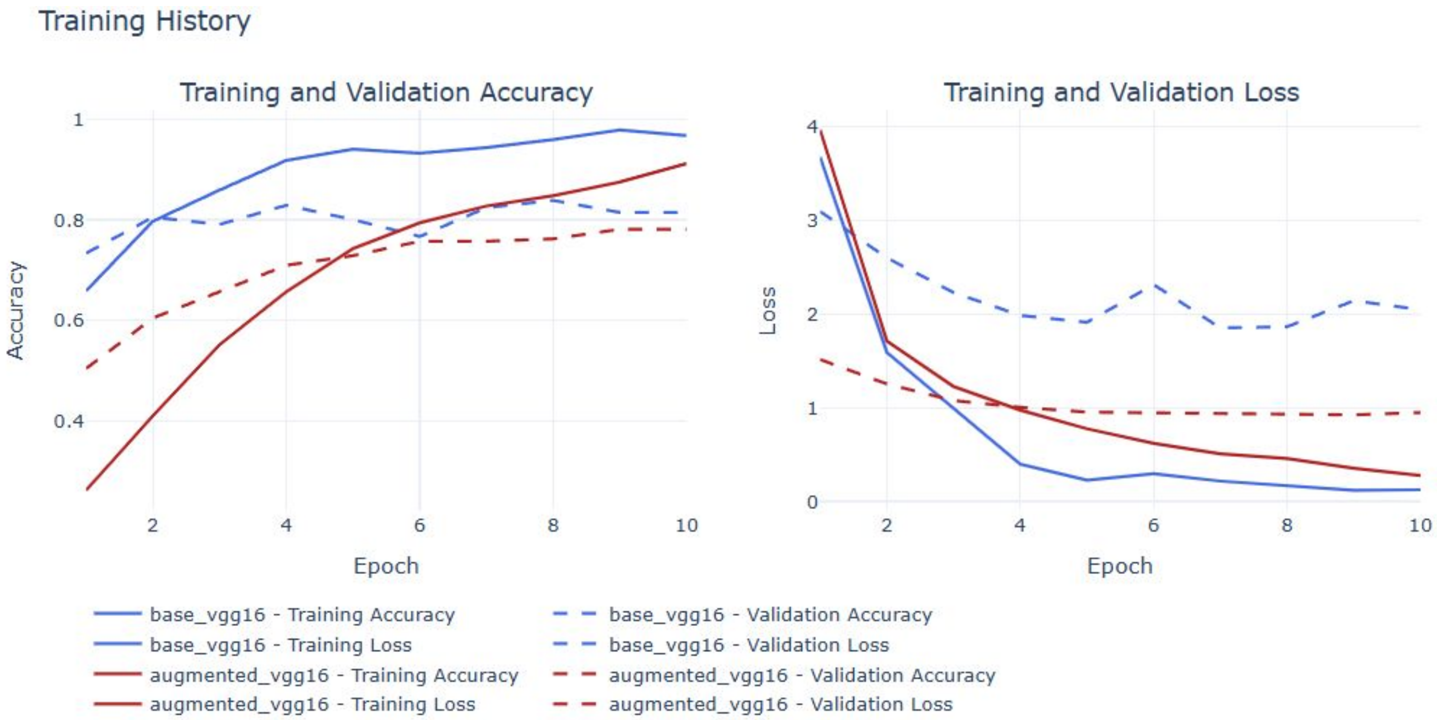


Transfer Learning - Principe

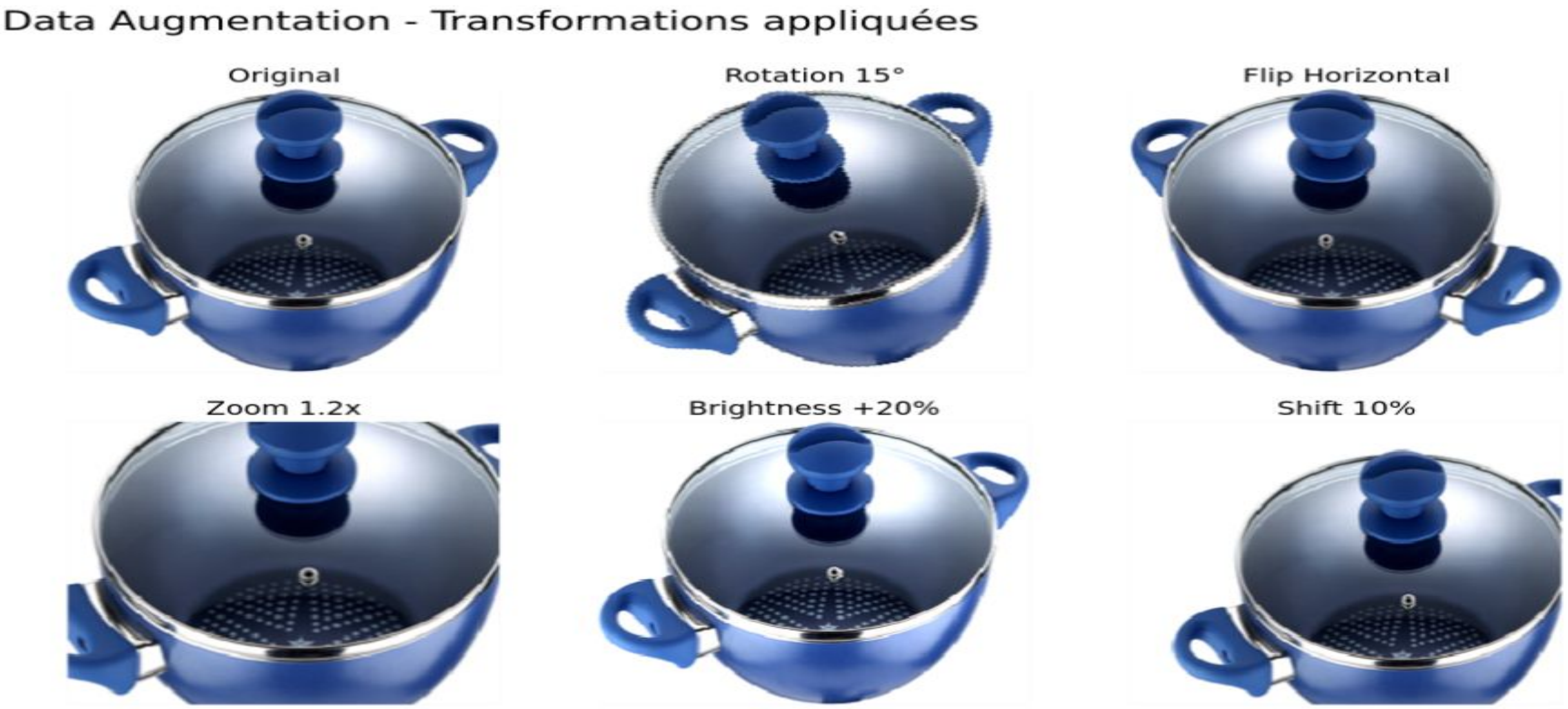
- Problème : 1050 images → insuffisant pour CNN from scratch
- Solution : poids pré-entraînés sur ImageNet (14M images)
- Backbone gelé : features génériques (edges, textures, formes)
- Head entraînable : classification spécifique à notre tâche
- Avantage : convergence rapide, meilleure généralisation

Architecture VGG16 - Transfer Learning

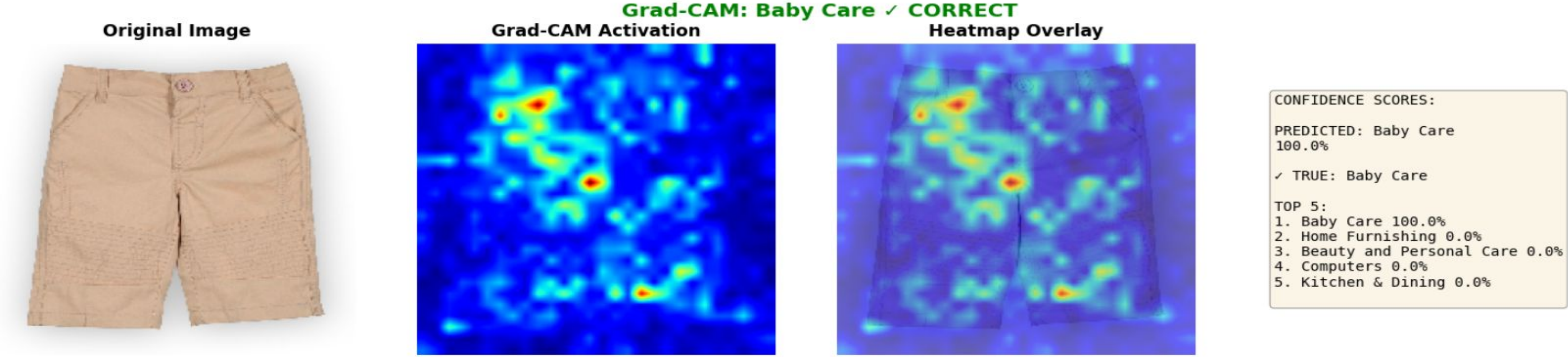


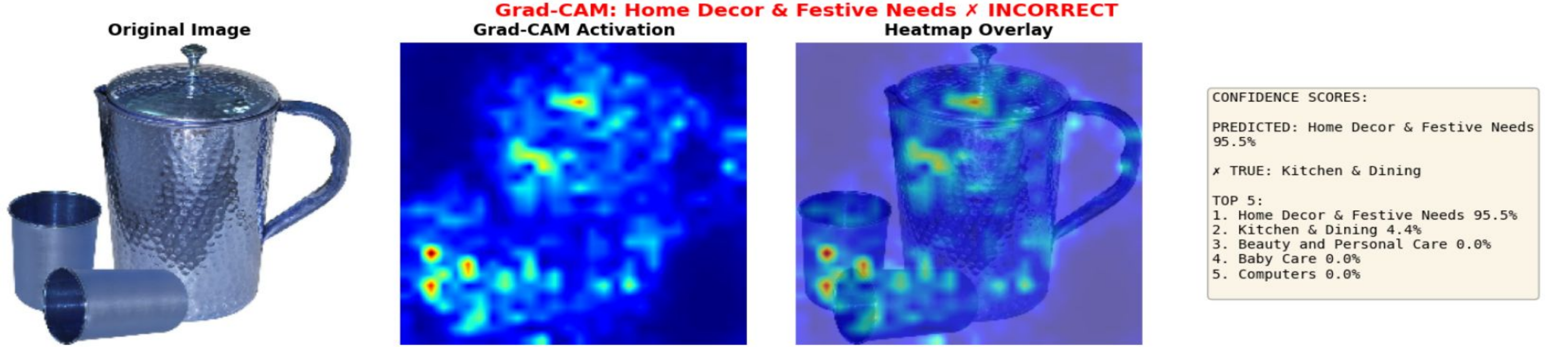


Data Augmentation - Vraie image transformée

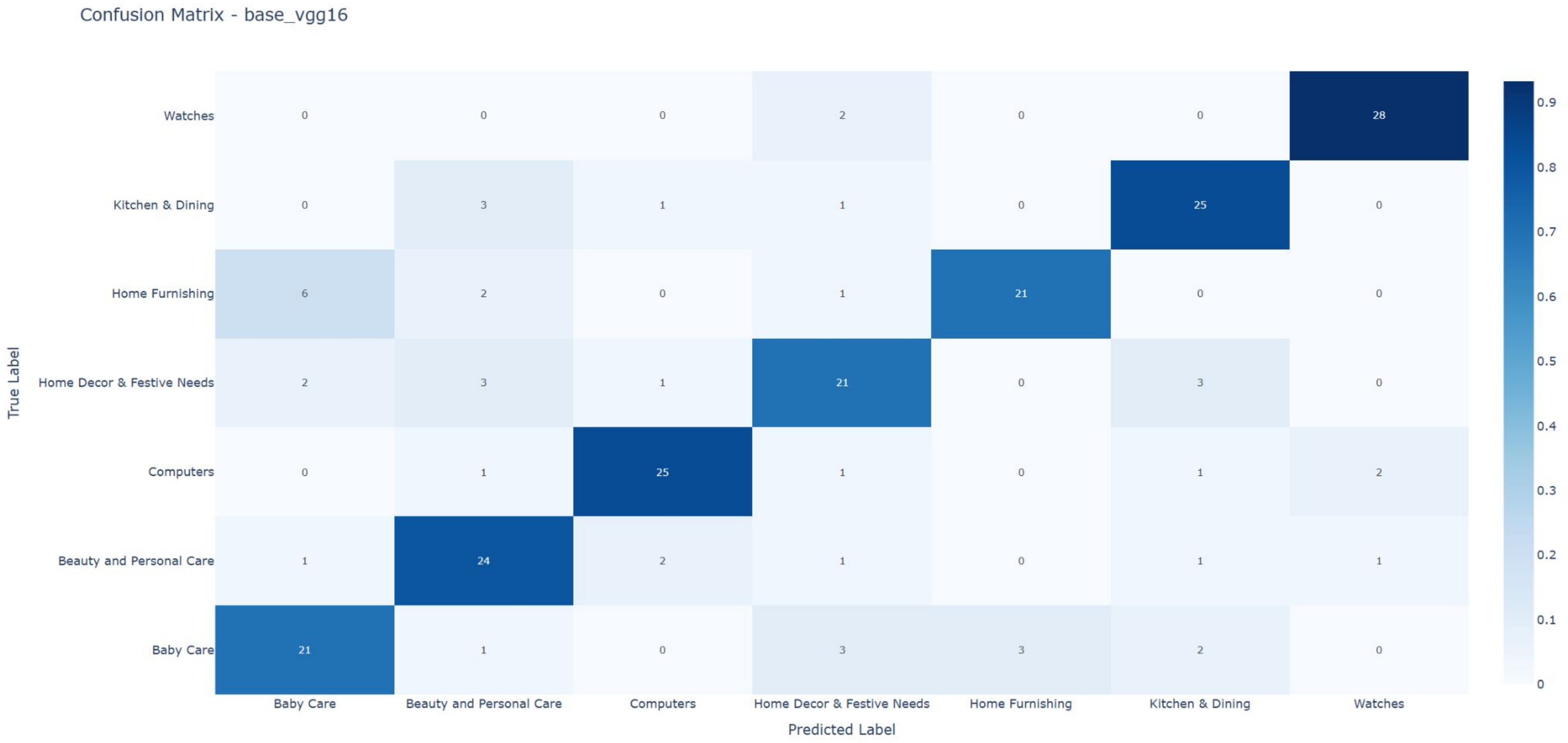


Grad-CAM Visualization: Model interpretability

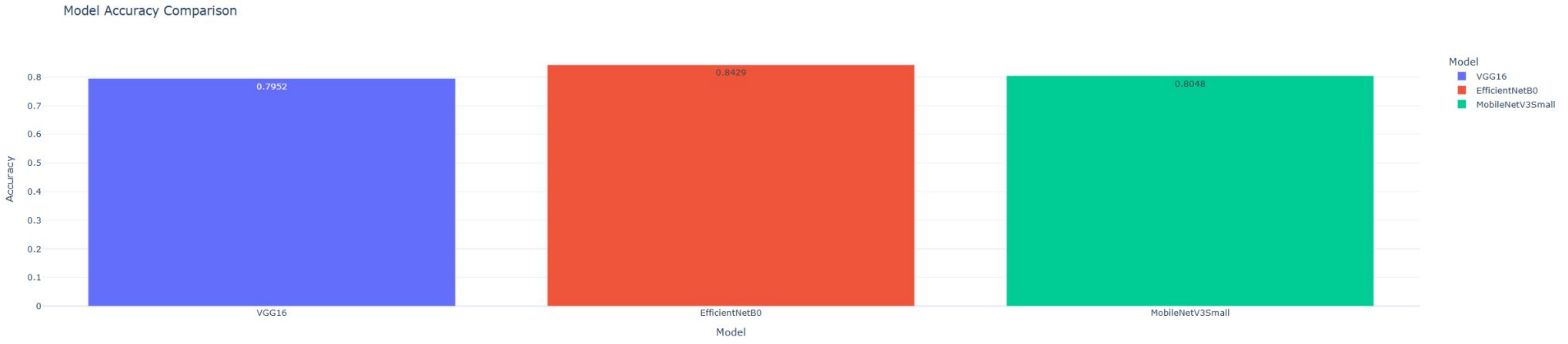




VGG16 Confusion Matrix



Comparaison des modèles CNN



Conclusion

In this project, we explored various techniques for classifying e-commerce products based on their images and text descriptions.

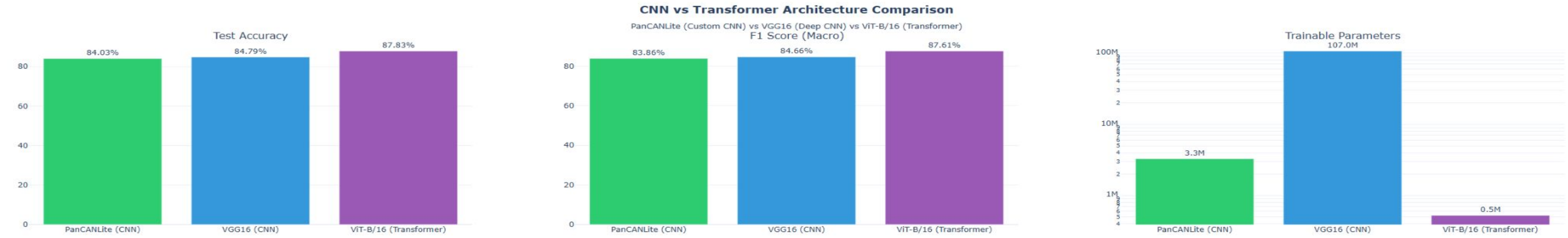
Key Findings:

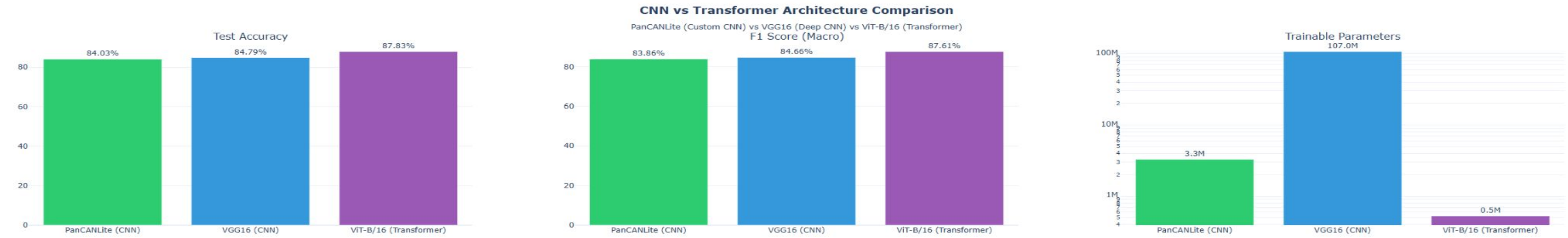
1. **Visual Analysis**:
    - **SIFT/ORB**: Traditional feature descriptors provided a baseline but struggled with semantic understanding.
    - **CNN (VGG16)**: Deep learning features significantly outperformed traditional methods, capturing high-level semantic concepts.
    - **Architecture Comparison**:
        - **VGG16** provided a strong baseline.
        - **EfficientNetB0** demonstrated superior efficiency, achieving competitive accuracy with fewer parameters.
        - **MobileNetV3** offered the fastest training times, suitable for resource-constrained environments.
2. **Text Analysis**:
    - **Bag of Words / TF-IDF**: Effective for keyword matching but lost semantic context.
    - **Word Embeddings (USE/BERT)**: Captured semantic meaning, allowing for better clustering of similar products even with different wording.
3. **Multimodal Fusion**:
    - Combining visual and textual features yielded the best results. The complementary nature of images (visual appearance) and text (specifications, usage) allowed the model to disambiguate difficult cases.

Future Work:

- **Fine-tuning**: Unfreezing the top layers of the pre-trained models could further improve accuracy.
- **Data Augmentation**: Increasing the dataset size with augmentations would help reduce overfitting.
- **Deployment**: The MobileNetV3 model is a strong candidate for deployment on edge devices or a mobile app for real-time product classification.

Teasing mission8….





Model	Test Accuracy	F1-Score	Key Reference
VGG16 (baseline)	84.79%	84.66%	-
PanCANLite	84.79%	84.68%	[Jiu et al., 2025]
ViT-B/16	86.69%	86.54%	[Wang et al., 2025]
Ensemble	88.21%	87.95%	[Abulfaraj & Binzagr, 2025]
Multimodal Fusion	**92.40%**	**92.15%**	[Dao et al., 2025], [Willis & Bakos, 2025]