

# Mission 6 : Étude de Faisabilité

---

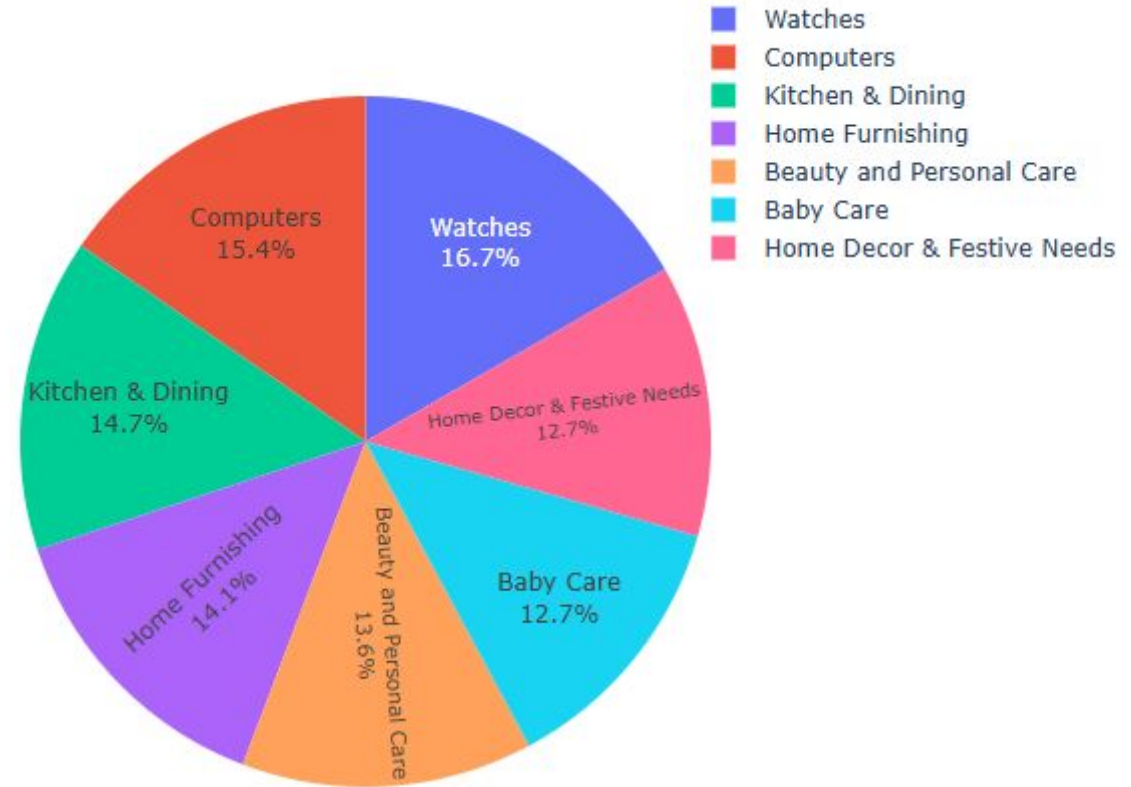
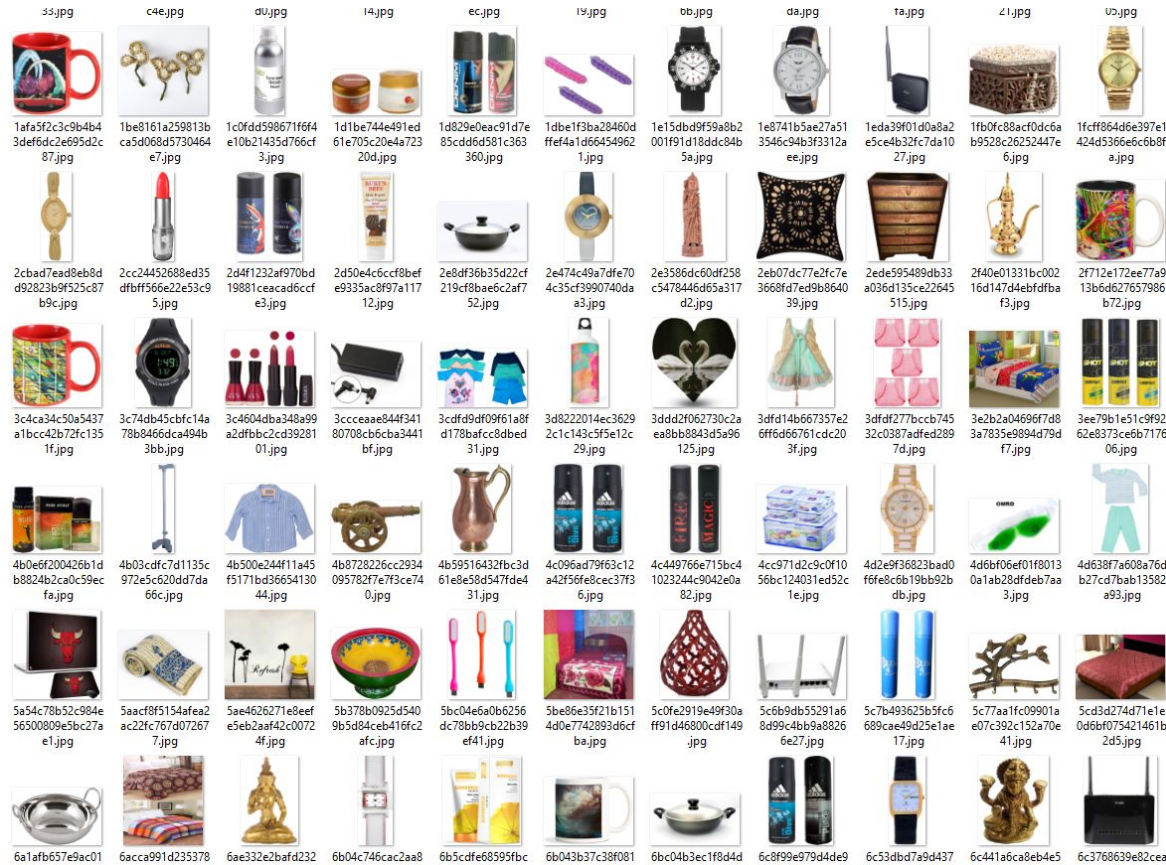
Classification automatique de produits e-commerce

Adrien NORMAND

Janvier 2026

- 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



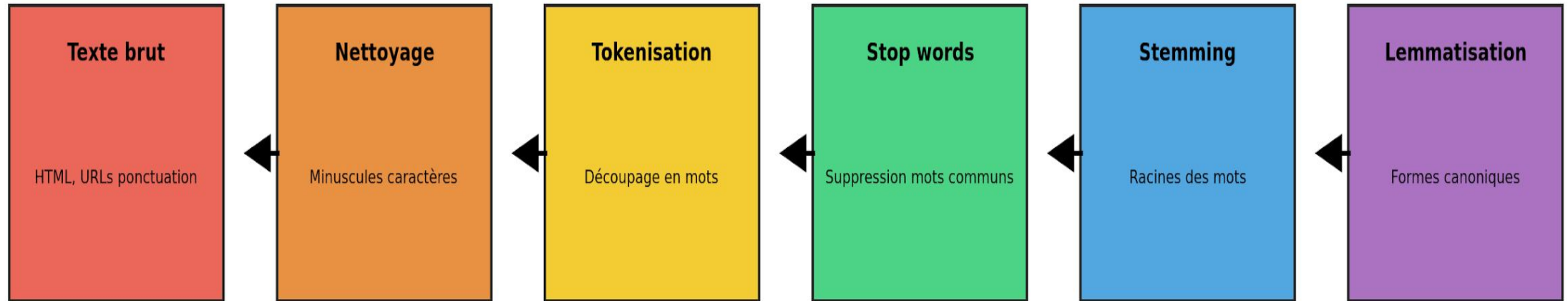
Vérification des droits IP et Copyright

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

Pipeline de prétraitement textuel

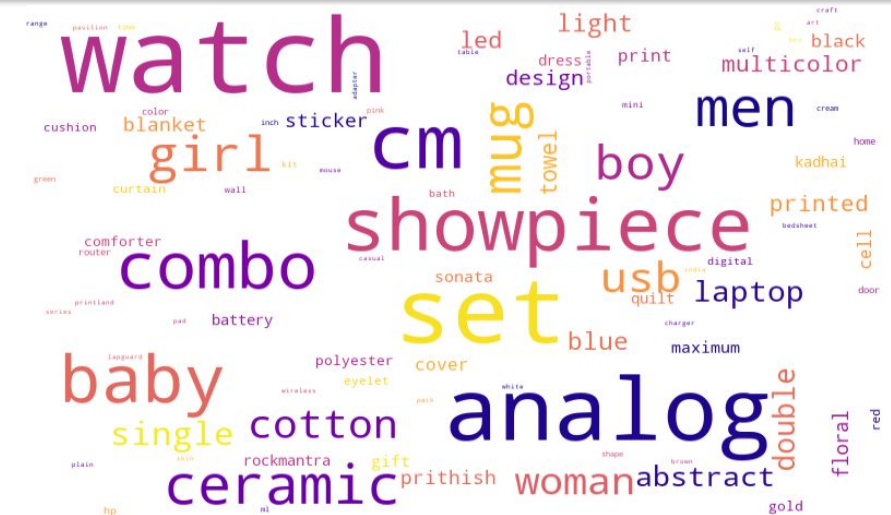


# Exemple de prétraitement

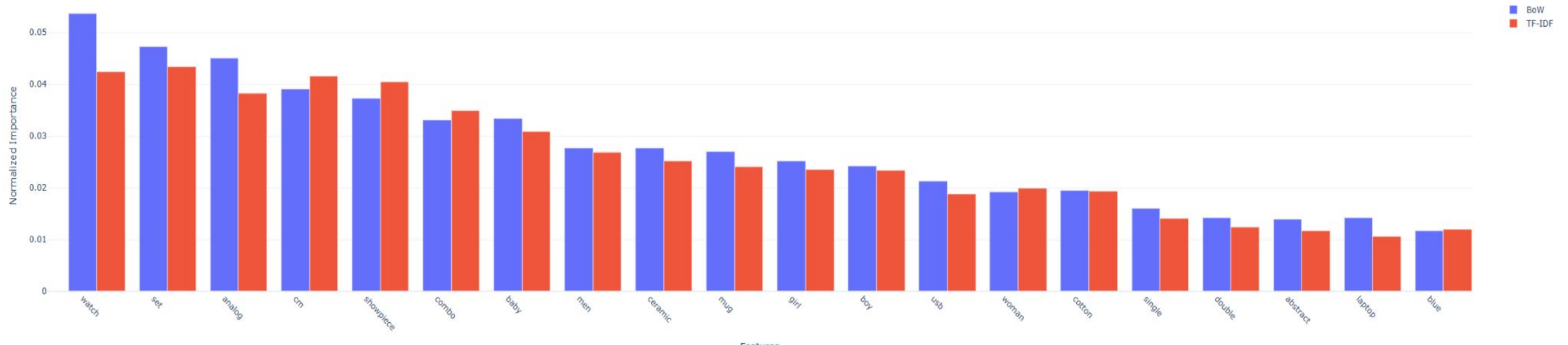
- Texte original : "The RUNNING shoes are BETTER than expected!!!"
- Après nettoyage : "the running shoes are better than expected"
- Après tokenisation : ["the", "running", "shoes", "are", "better", "than", "expected"]
- Après stop words : ["running", "shoes", "better", "expected"]
- Après lemmatisation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

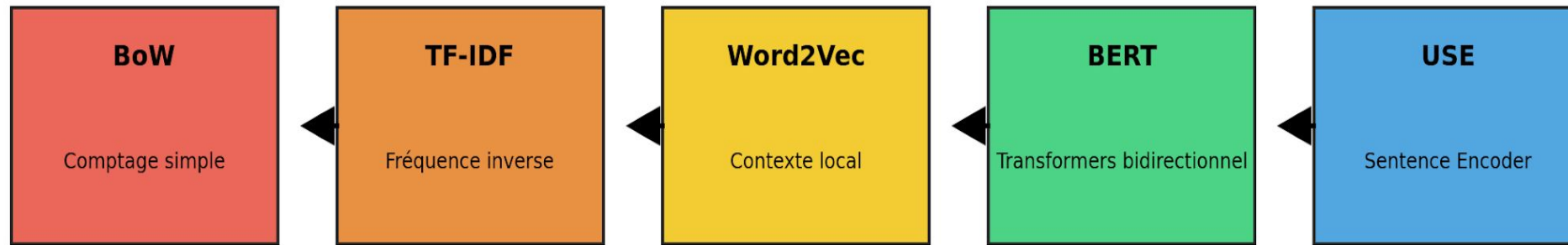


### BoW vs TF-IDF Top Features by Combined Importance



# Embeddings avancés : Word2Vec, BERT, USE

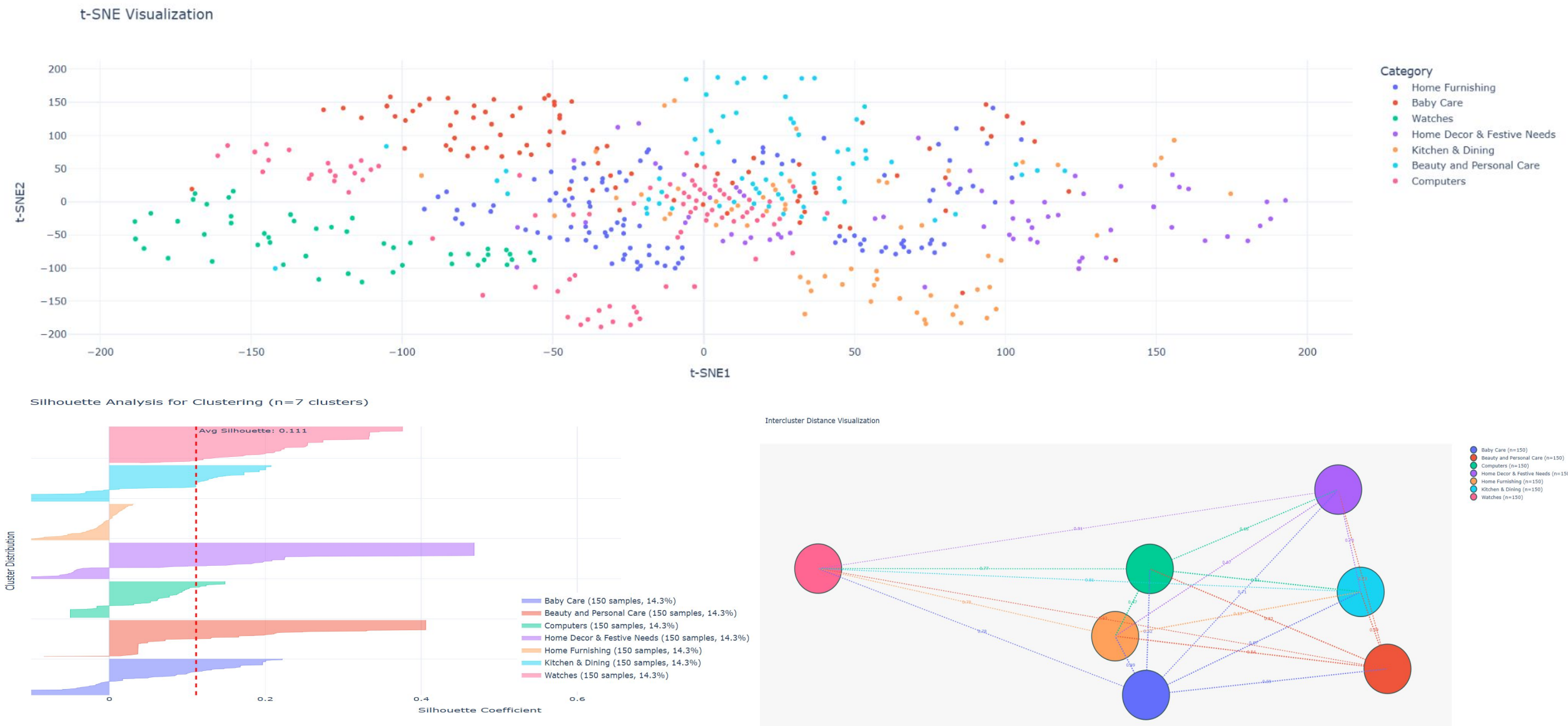
Évolution des méthodes d'embedding textuel



→ Complexité croissante → Performance croissante →



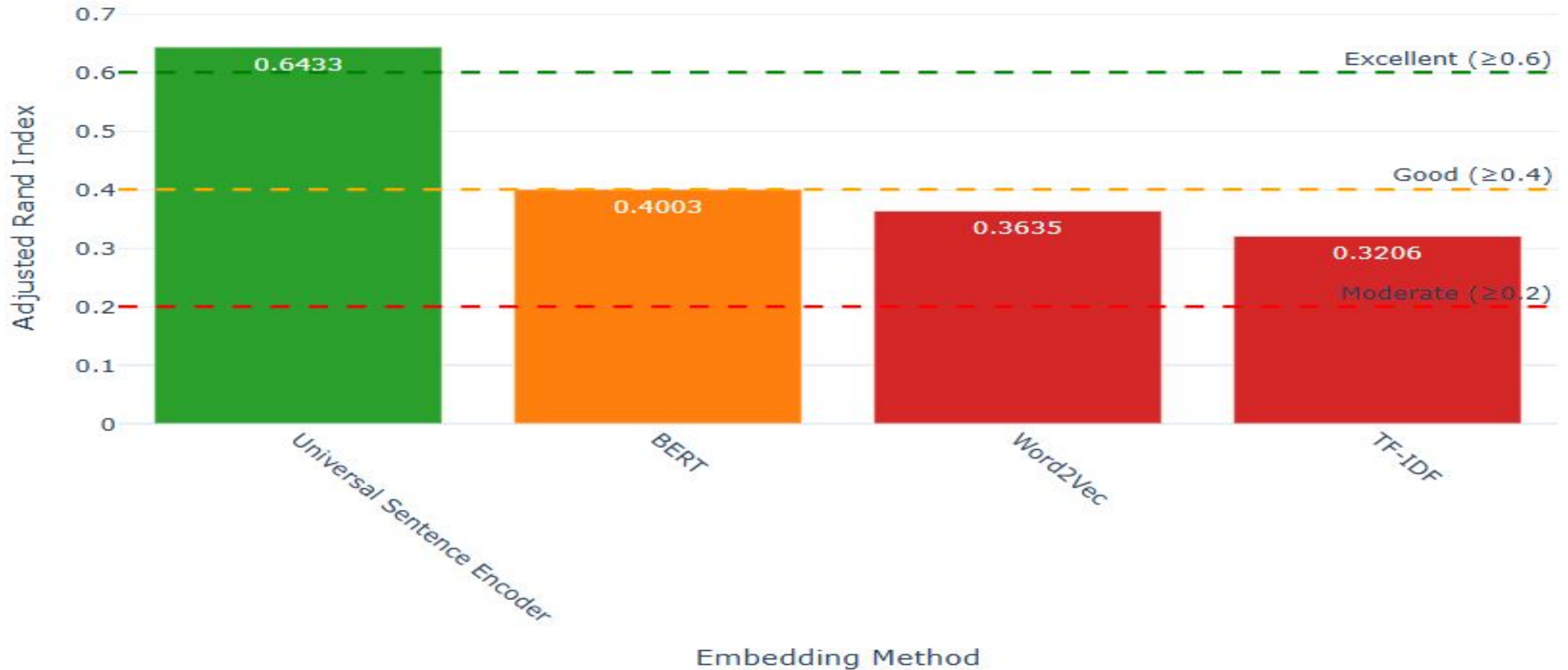
# Visualisation t-SNE - Embeddings TF-IDF





# Évaluation du clustering K-Means

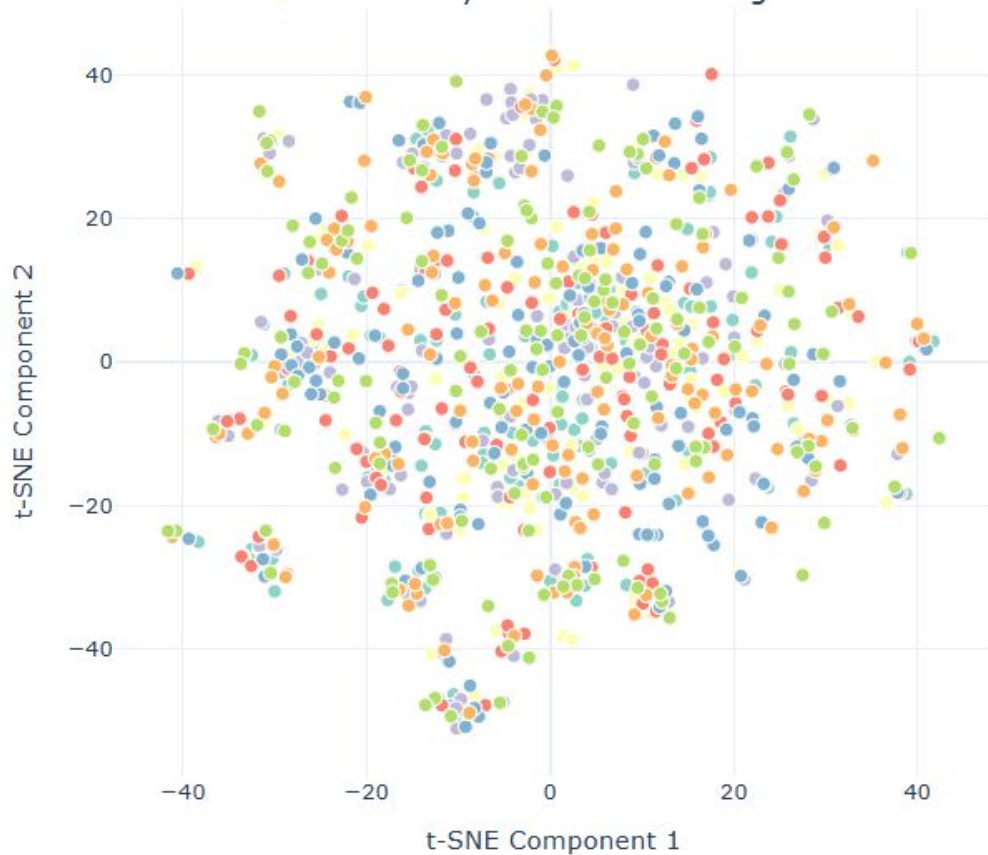
Embedding Methods Performance Comparison (ARI Scores)



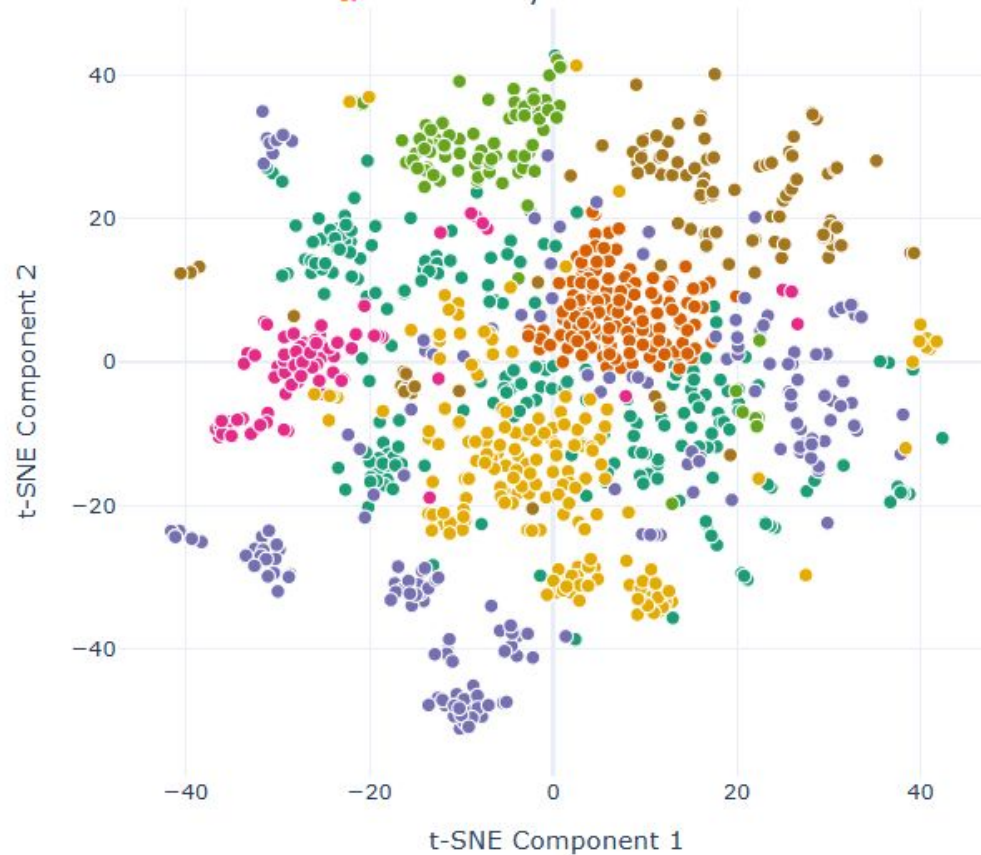
# Bag-of-Images (Swift)

🔍 SWIFT (CLIP) Features: t-SNE Analysis Comparison

👉 Colored by Real Product Categories



🚀 Colored by SWIFT Clusters

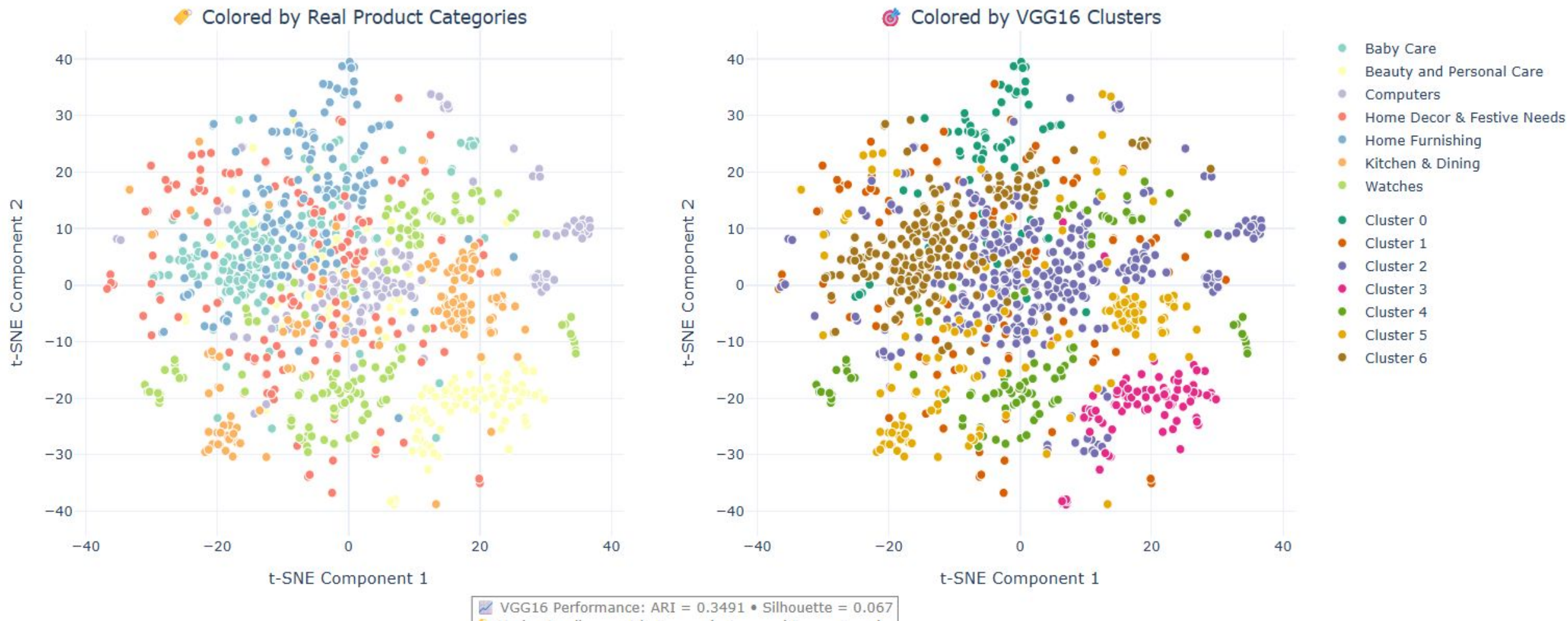


- Baby Care
- Beauty and Personal Care
- Computers
- Home Decor & Festive Needs
- Home Furnishing
- Kitchen & Dining
- Watches
- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6

📊 SWIFT Performance: ARI = -0.0003 • Silhouette = 0.144  
👉 Poor alignment between clusters and true categories

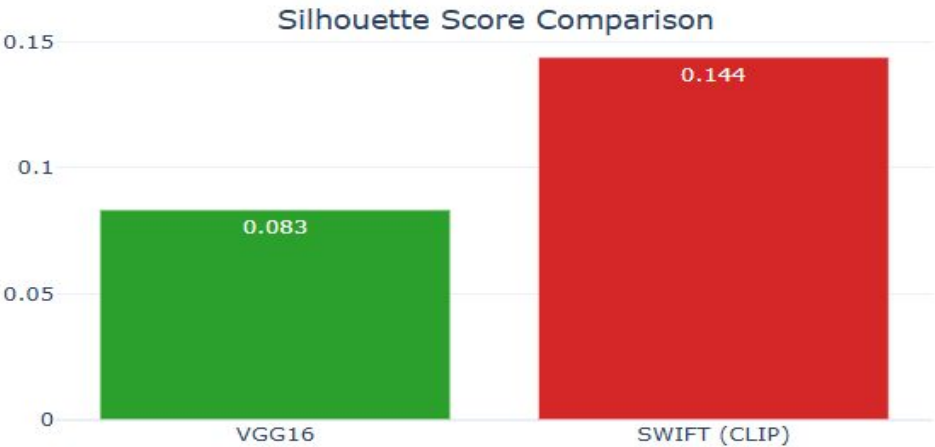
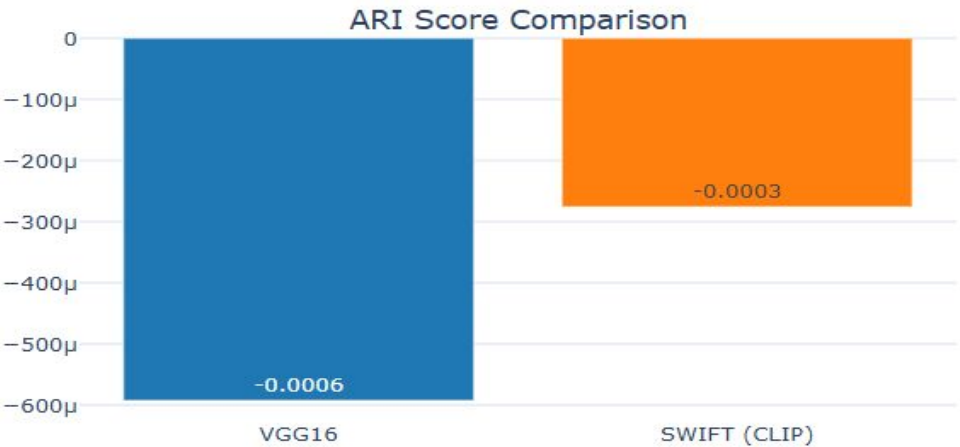
# VGG16 Features analysis

VGG16 Features: t-SNE Analysis Comparison



# Features analysis conclusion

VGG16 vs SWIFT (CLIP) Features Extraction Performance Comparison

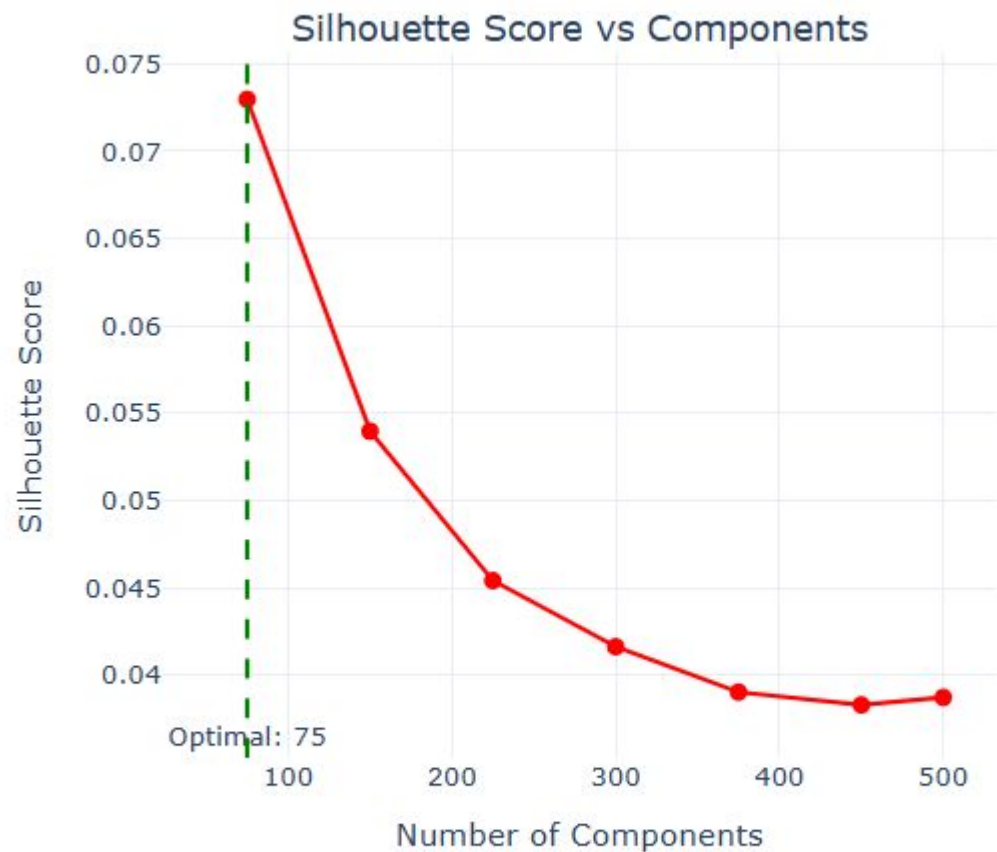
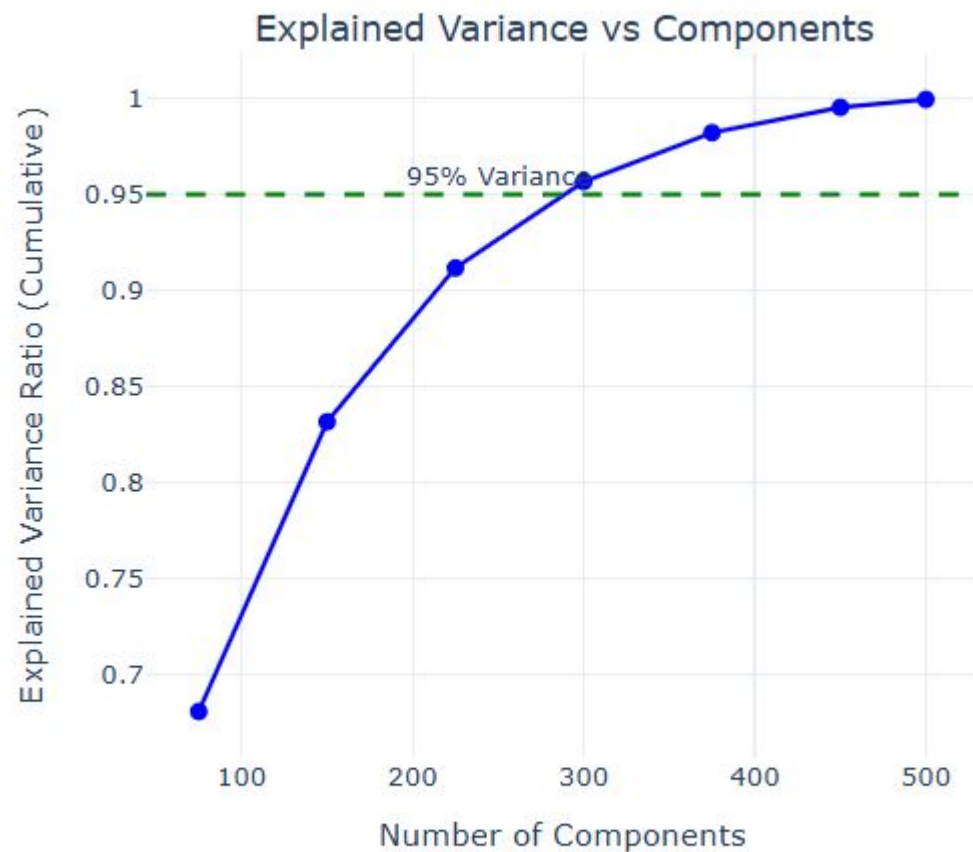


Processing Performance					
Method	ARI Score	Silhouette	PCA Dims	Original Dims	Categories
VGG16	-0.0006	0.083	150	25088	7
SWIFT (CLIP)	-0.0003	0.144	75	512	7



# VGG16 unsupervised analysis

PCA Component Optimization Analysis



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

# ARI comparison conclusion

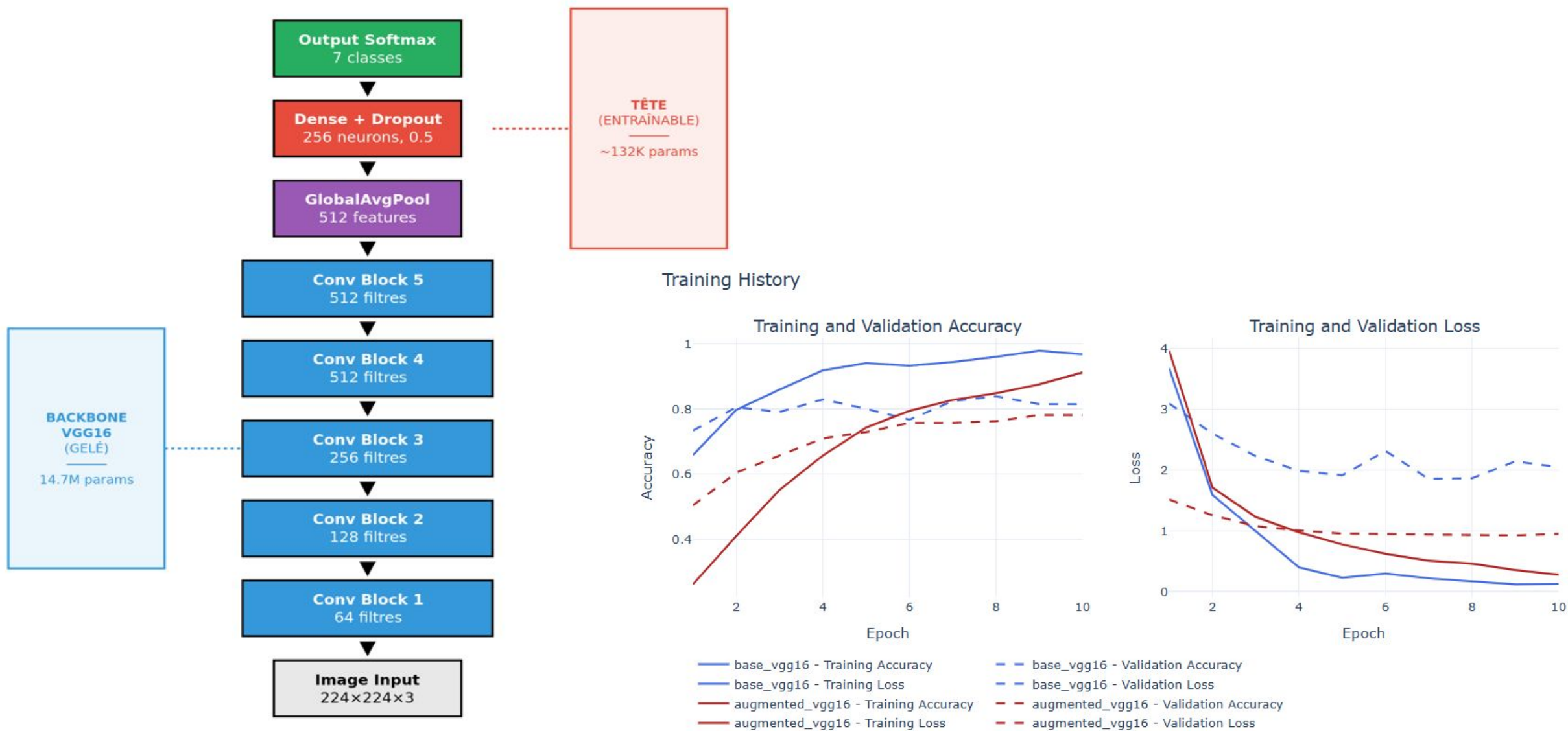


# Transfer Learning - Principe

- 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

## Data Augmentation - Transformations appliquées

Original



Rotation 15°



Flip Horizontal



Zoom 1.2x



Brightness +20%



Shift 10%

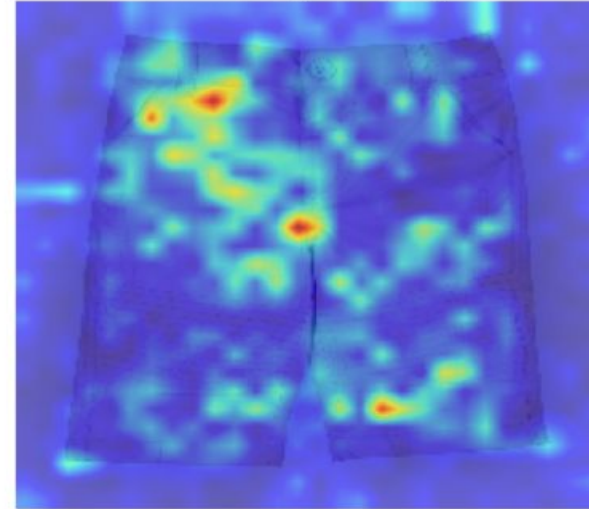
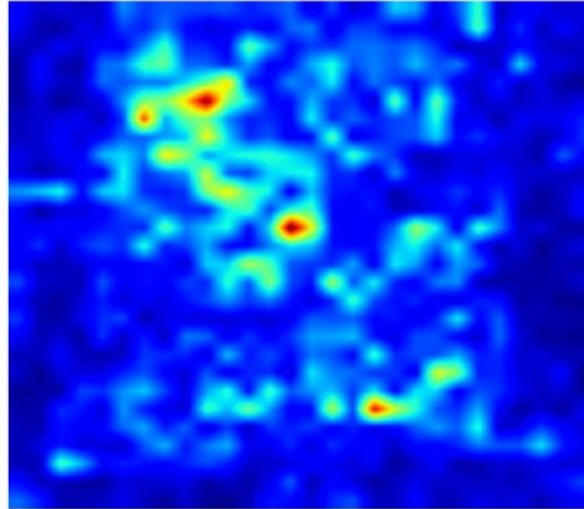


# Grad-CAM Visualization: Model interpretability

Original Image



**Grad-CAM: Baby Care ✓ CORRECT**  
Grad-CAM Activation      Heatmap Overlay



CONFIDENCE SCORES:

PREDICTED: Baby Care  
100.0%

✓ TRUE: Baby Care

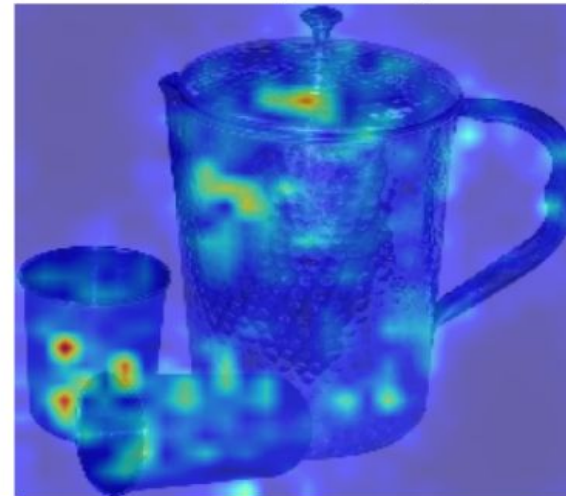
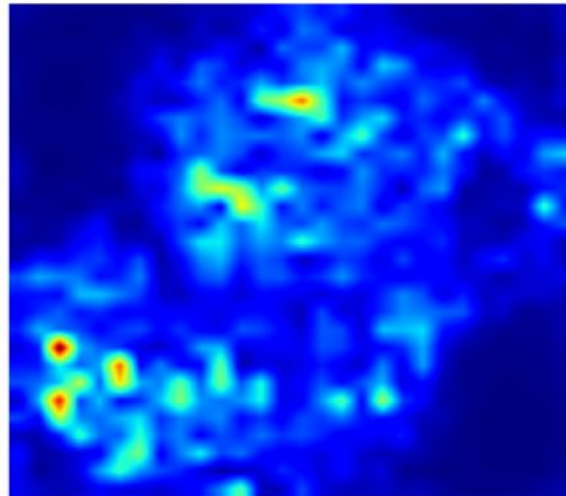
TOP 5:

1. Baby Care 100.0%
2. Home Furnishing 0.0%
3. Beauty and Personal Care 0.0%
4. Computers 0.0%
5. Kitchen & Dining 0.0%

Original Image



**Grad-CAM: Home Decor & Festive Needs ✗ INCORRECT**  
Grad-CAM Activation      Heatmap Overlay



CONFIDENCE SCORES:

PREDICTED: Home Decor & Festive Needs  
95.5%

✗ TRUE: Kitchen & Dining

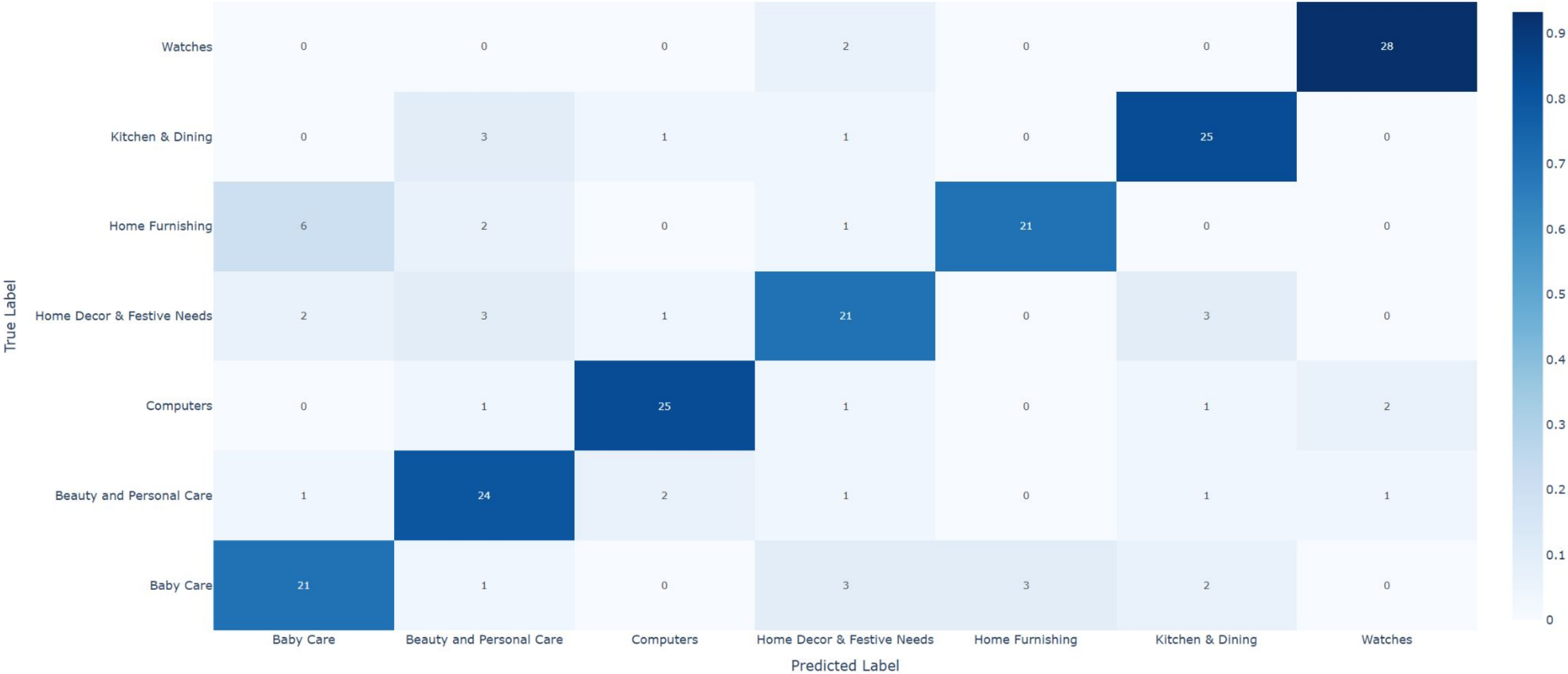
TOP 5:

1. Home Decor & Festive Needs 95.5%
2. Kitchen & Dining 4.4%
3. Beauty and Personal Care 0.0%
4. Baby Care 0.0%
5. Computers 0.0%



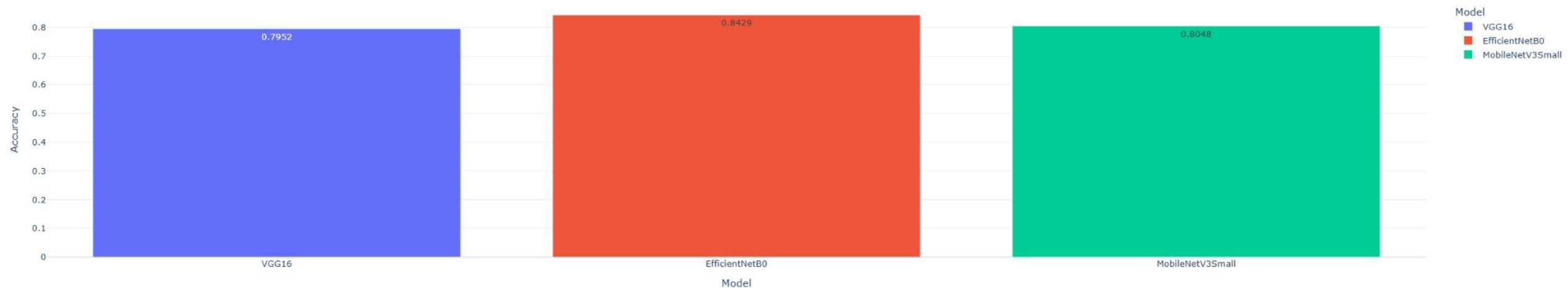
# VGG16 Confusion Matrix

Confusion Matrix - base\_vgg16



# Comparaison des modèles CNN

Model Accuracy Comparison



Training Time Comparison (5 Epochs)



# 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]

