

Alexandre ROGUES - Data Scientist

PROJET

8

Réalisez un dashboard et assurez une veille technique

Prêt à
dépenser





INTRODUCTION



Prêt à dépenser

Objectif : Délivrer des crédits à des clients qui les remboursent

Projet 7: mise en oeuvre d'un outil de scoring, et détermination d'un seuil selon le coût pour l'entreprise

Besoin désormais : **Dashboard interactif** à destination du conseiller client

=> pouvoir expliquer la décision au client

PROBLÉMATIQUE

- 01** Visualiser le score
- 02** Visualiser les informations du client
- 03** Comparer les informations

On souhaite offrir au conseiller client un outil lui permettant :

- 04** De prendre en compte son besoin en cas de handicap
- 05** D'y accéder dans un simple navigateur



DASHBOARD INTERACTIF

1ère étape

Renseigner l'identifiant client

Sélectionnez la taille du texte

Moyen

Calculateur de risque de remboursement

Entrez le numéro de client

DASHBOARD INTERACTIF

Un parcours simple

Le conseiller n'a plus qu'à dérouler le menu

Sélectionnez la taille du texte

Moyen

Calculateur de risque de remboursement

Entrez le numéro de client

368274

Afficher infos client

Afficher la décision de l'algorithme

Afficher facteurs de décision

Afficher les valeurs de la base de donnée

DASHBOARD INTERACTIF

Des informations facilement interprétable

Sélectionnez la taille du texte

Moyen

Calculateur de risque de remboursement

Entrez le numéro de client

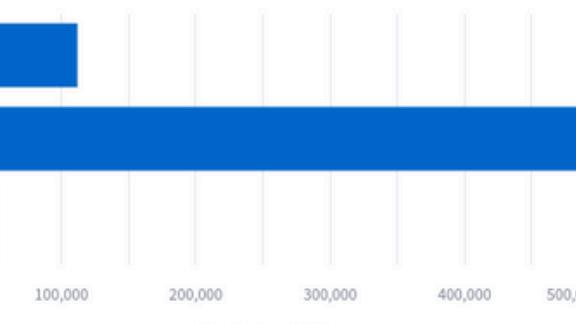
368274

Afficher infos client

Le client #368274 :

- ⌚ est une femme
- 🏃 ne possède pas de voiture
- ❌ n'est pas propriétaire
- ❌ n'a pas d'enfants

Informations financières



Informations financières	Montant en USD
Salaire annuel	~100,000
Montant du crédit	~500,000
Montant de l'ann...	~20,000

Sélectionnez la taille du texte

Moyen

Calculateur de risque de remboursement

Entrez le numéro de client

368274

Afficher infos client

Afficher la décision de l'algorithme

Le client a peu de chance d'avoir des soucis de remboursement.



Le score de crédit prédit pour le client 368274 est : 0.49%

Afficher facteurs de décision

Afficher les valeurs de la base de donnée

DASHBOARD INTERACTIF

The screenshot shows a web-based dashboard titled "Dashboard Client". At the top, there's a text size selector set to "Moyen" and a URL bar showing "https://projet8.rogues.fr/site/projet8/". Below the header, there's a section for "Calculateur de risque de remboursement" where a client number "368274" is entered. There are several checkboxes: "Afficher infos client", "Afficher la décision de l'algorithme", and "Afficher facteurs de décision" (which is checked). A "Importances des caractéristiques" section is present, divided into "Importances locales" and "Importances globales". Under "Importances locales", there's a note explaining positive and negative local importances, followed by two tables: "Importances locales positives" and "Importances locales négatives", each listing features like gender, age, and occupation with their respective importances and descriptions. The "Importances globales" section also has a note and a table listing global importances for features such as EXT_SOURCE_3, EXT_SOURCE_2, DAYS_REGISTRATION, etc.

des notes explicatives aux moments critiques

Importances locales

Note explicative :

- Les **importances locales positives** indiquent les caractéristiques qui augmentent le score de risque de crédit pour ce client.
- Les **importances locales négatives** indiquent les caractéristiques qui diminuent le score de risque de crédit pour ce client.
- Les valeurs de ces importances montrent l'impact de chaque caractéristique sur la prédiction du modèle.

Note explicative :
Les importances globales indiquent les caractéristiques les plus influentes sur l'ensemble des prédictions du modèle. Caractéristiques Clés : Portez une attention particulière aux caractéristiques avec des importances globales élevées, car elles ont le plus grand impact sur les prédictions du modèle. Comparaison : Utilisez les importances globales pour comparer et comprendre comment le modèle prend ses décisions de manière générale.

Importances globales

Note explicative :

- Les **importances globales** indiquent les caractéristiques les plus influentes sur l'ensemble des prédictions du modèle.
- **Caractéristiques Clés** : Portez une attention particulière aux caractéristiques avec des importances globales élevées, car elles ont le plus grand impact sur les prédictions du modèle.
- Les **valeurs de ces importances** indiquent la contribution d'une caractéristique sur la prédiction.

formation spécifique à fournir au conseiller

DASHBOARD INTERACTIF

Afficher les valeurs de la base de donnée

Sélectionnez la caractéristique à afficher

EXT_SOURCE_3

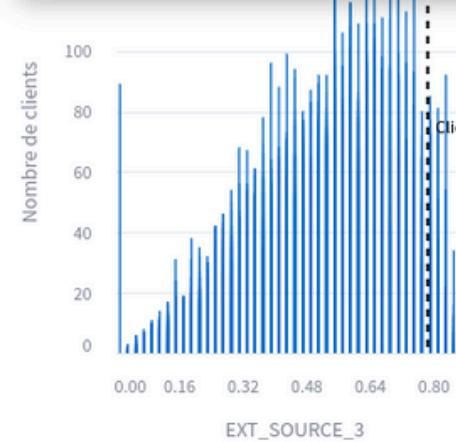
EXT_SOURCE_3

EXT_SOURCE_2

DAYS_REGISTRATION

DAYS_EMPLOYED

EXT_SOURCE_1



Possibilité de comparer les valeurs du client à celles des autres clients dans leur globalité

Ou selon un critère

Afficher les valeurs de la base de donnée

Sélectionnez la caractéristique à afficher

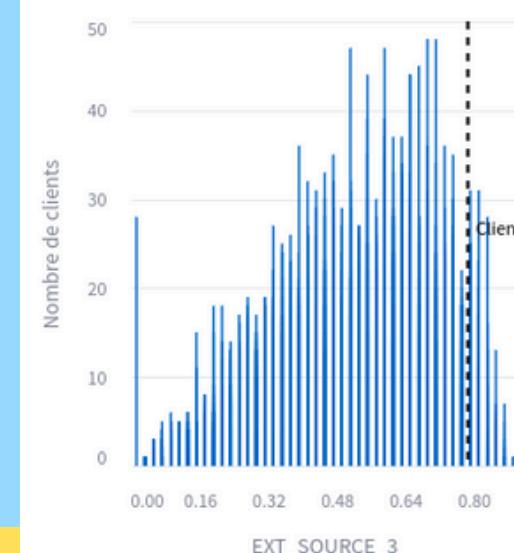
EXT_SOURCE_3

Sélectionnez l'option d'affichage

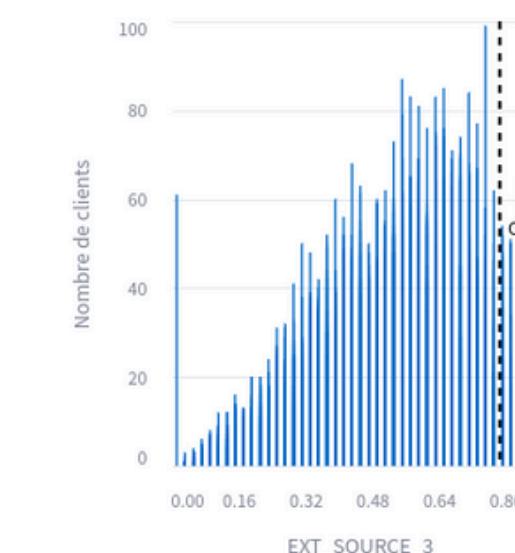
Tout le monde

Comparer homme/femme

Hommes

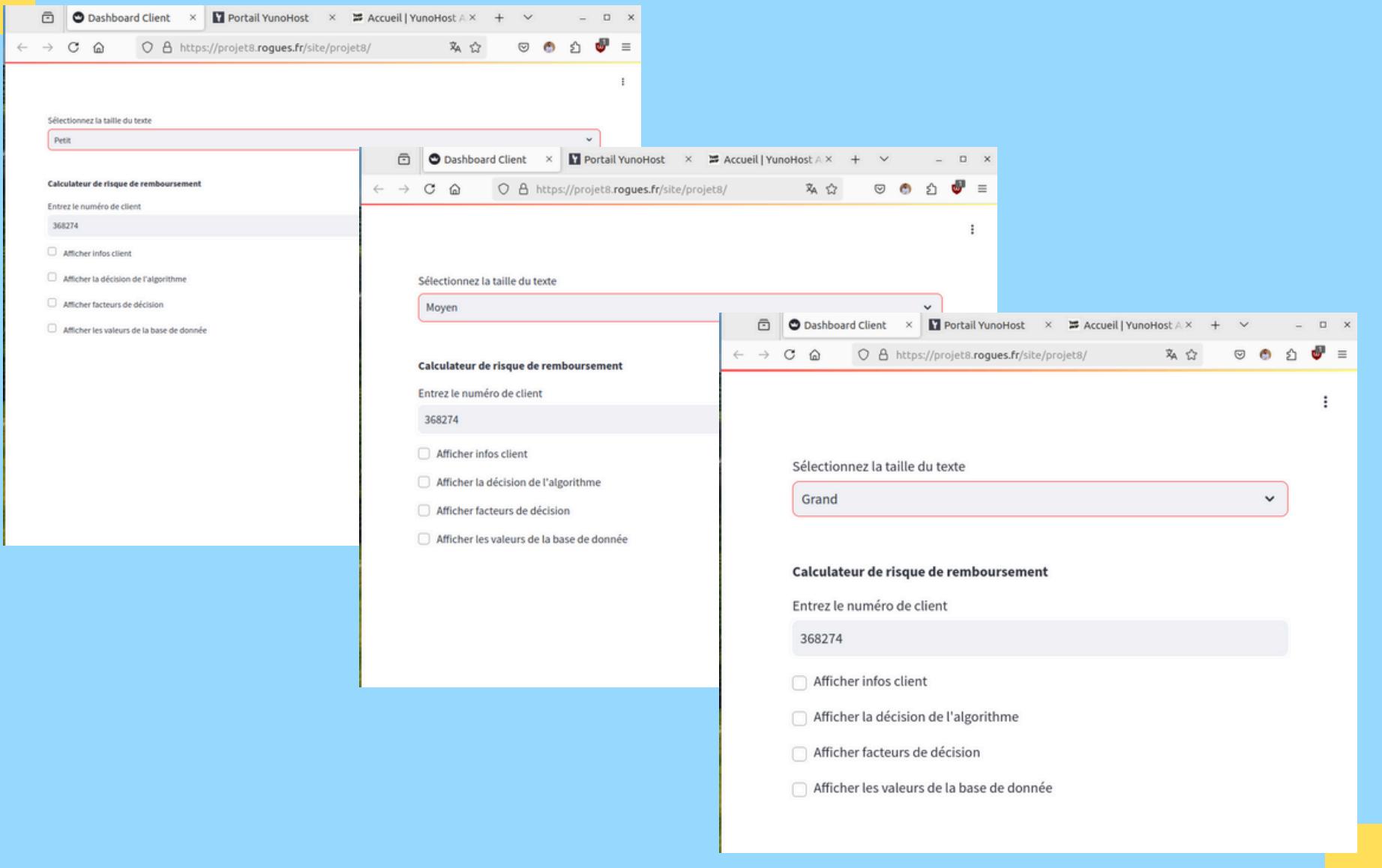


Femmes



DASHBOARD INTERACTIF

respect des critères d'accessibilité



Option de changement de taille de police

Afficher infos client

Le client #368274 :

- 🚹 est une femme
- 🚻 ne possède pas de voiture
- ✗ n'est pas propriétaire
- ✗ n'a pas d'enfants

images

Afficher la décision de l'algorithme

Le client a peu de chance d'avoir des soucis de remboursement.

Le score de crédit prédit pour le client 368274 est : 0.49%

jauge colorée

HEBERGEMENT

- autohebergement sur VPS perso chez OVH
- OS Debian/YunoHost (solution d'autohébergement libre)
- création sous -domaine projet8.rogues.fr, sécurisation SSL
- création d'un service streamlit.service pour assurer la disponibilité
- modification paramètres nginx pour redirection via proxy
- installation appli via script deploy GITHUB Actions
- redémarrage du service streamlit manuel via connection SSH

Stabadev/Projet8

Réalisez un dashboard et assurez une veille technique



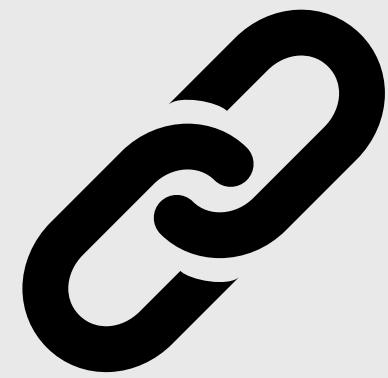
1 Contributor 0 Issues 0 Stars 0 Forks

Stabadev/Projet8: Réalisez un dashboard et assurez une veille technique

Réalisez un dashboard et assurez une veille technique - Stabadev/Projet8

 GitHub

DEMO



<https://projet8.rogues.fr/site/projet8/>

#368274 #230664

#141653 #225599

id client

VEILLE TECHNIQUE

Processus de surveillance et d'analyse des nouvelles technologies, des tendances de l'industrie, des pratiques concurrentielles et des normes.

Arxiv : Archive de recherches scientifiques prépubliées

Blogs reconnus, éventuellement chaînes youtube



1050 items

CONTEXTE

Rappel Projet 6 - Classifiez automatiquement des biens de consommation

*Etude de la faisabilité de labellisation automatique
d'items en vente sur internet, à partir d'info textuelles et
images.*

Objectif : faciliter

- mise en ligne par les vendeurs
- recherche par clients

7 catégories balancées :

- Home Furnishing
- Baby Care
- Watches
- Home Decor & Festive Needs
- Kitchen & Dining
- Beauty and Personal Care
- Computers

15 features/items

'uniq_id', 'crawl_timestamp',
'product_url', 'product_name',
'product_category_tree',
'pid', 'retail_price', 'discounted_price',
'image', 'is_FK_Advantage_product',
'description',
'product_rating', 'overall_rating',
'brand', 'product_specifications'

MODEL PHI3-MINI-4K

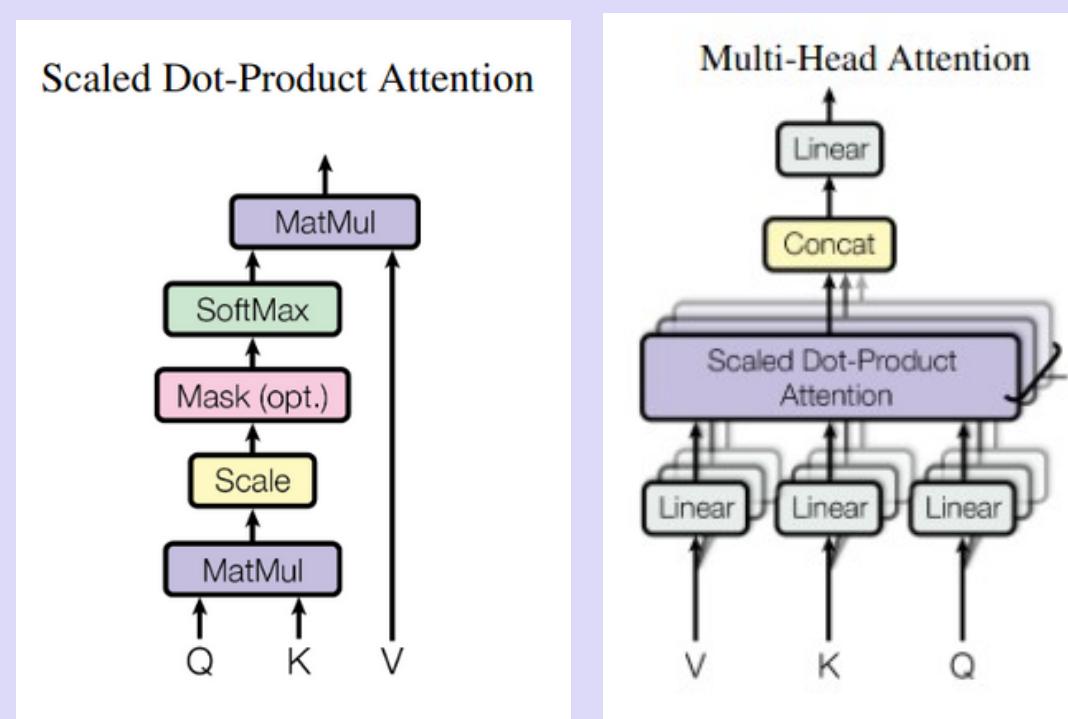
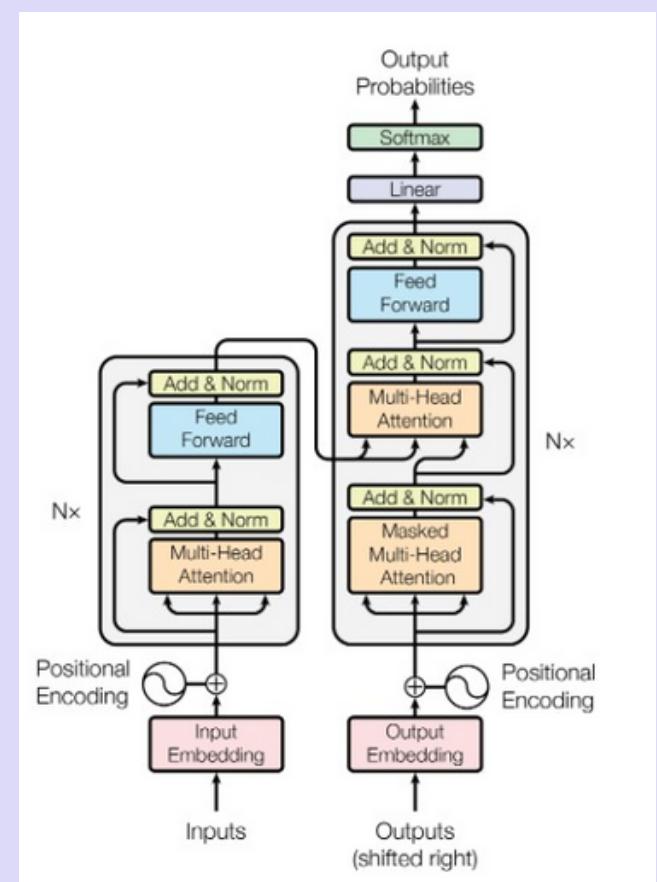
Modèle génératif basé sur les transformers

Encoders:

Transforme les séquences en représentations vectorielles denses et informatives sur les éléments de contexte et de sémantique

Decoders:

Génère les séquences de texte de sortie



Principe de l'attention:

introduit par la publication "Attention Is All You Need"

Self-Attention:

Permet à chaque mot d'une séquence de se concentrer sur tous les autres mots de la séquence, indépendamment de leur position, améliorant ainsi la compréhension contextuelle.

Multi-Head Attention:

Utilise plusieurs "têtes" d'attention pour capturer différentes nuances contextuelles simultanément, offrant une compréhension plus riche et variée du contexte.



MODEL PHI3-MINI-4K

Des spécificités propres au modèle de Microsoft

Compact

Entraîné sur 3,3 trillions de tokens, l'ensemble de données d'entraînement comprend des données web filtrées et des données synthétiques

Performant

Des performances comparables à des modèles beaucoup plus grands (GPT, Mixtral)

Différentes versions

- mini
- small
- medium
- vision (multimodal)

Filtrage des Données

- Données web filtrées et synthétiques pour un entraînement de haute qualité.

Techniques Avancées

- Hyperparamètres ajustés (muP et activation GEGLU) pour de meilleures performances.
- Modules d'attention blocksparse pour optimiser vitesse et performance sur les longues séquences.

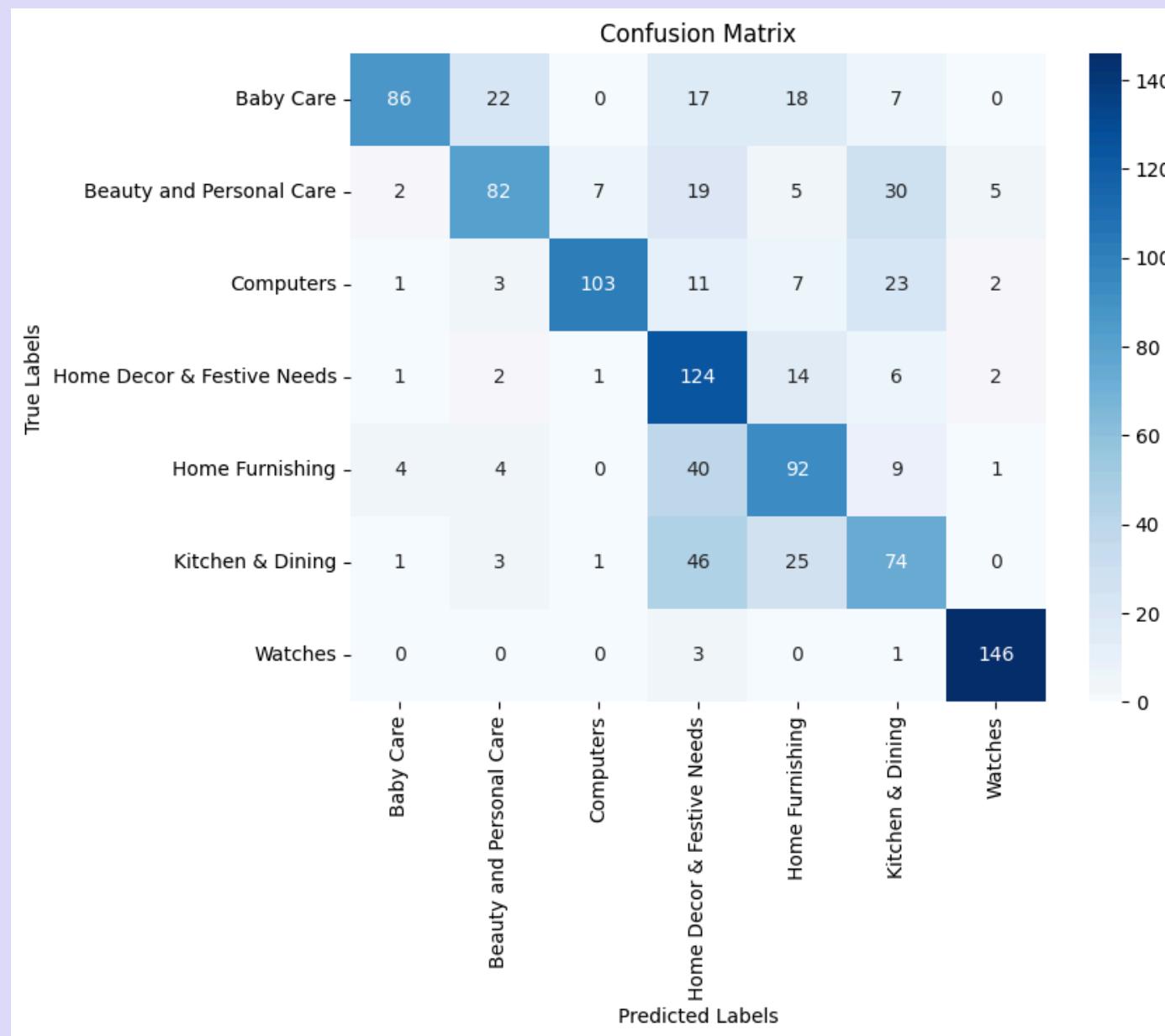
Sécurité

- Post-formation axée sur la sécurité avec fine-tuning supervisé.
- Utilisation de données Responsable AI pour évaluer et améliorer la sécurité.
- Équipe indépendante de "red-teaming" pour identifier et corriger les risques.

MÉTHODE 1

Compléction

- Soumission description brute + Question “choisi parmi 7 catégories”
- Nettoyage réponse
- Matrice de confusion => accuracy

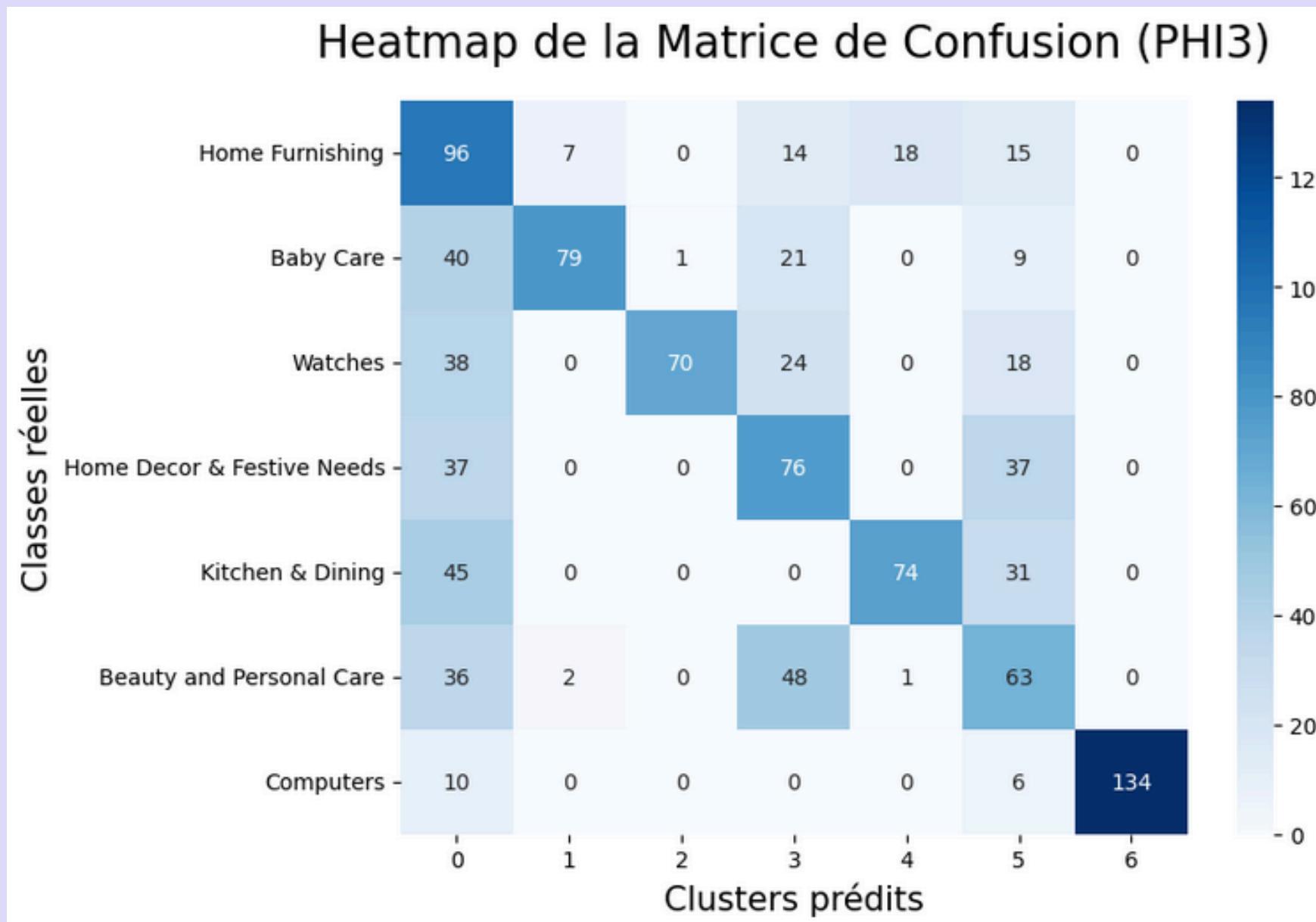


		precision	recall	f1-score	support
	Baby Care	0.91	0.57	0.70	150
	Beauty and Personal Care	0.71	0.55	0.62	150
	Computers	0.92	0.69	0.79	150
	Home Decor & Festive Needs	0.48	0.83	0.60	150
	Home Furnishing	0.57	0.61	0.59	150
	Kitchen & Dining	0.49	0.49	0.49	150
	Watches	0.94	0.97	0.95	150
		accuracy		0.67	1050
		macro avg	0.72	0.67	1050
		weighted avg	0.72	0.67	0.68
					1050

MÉTHODE 2

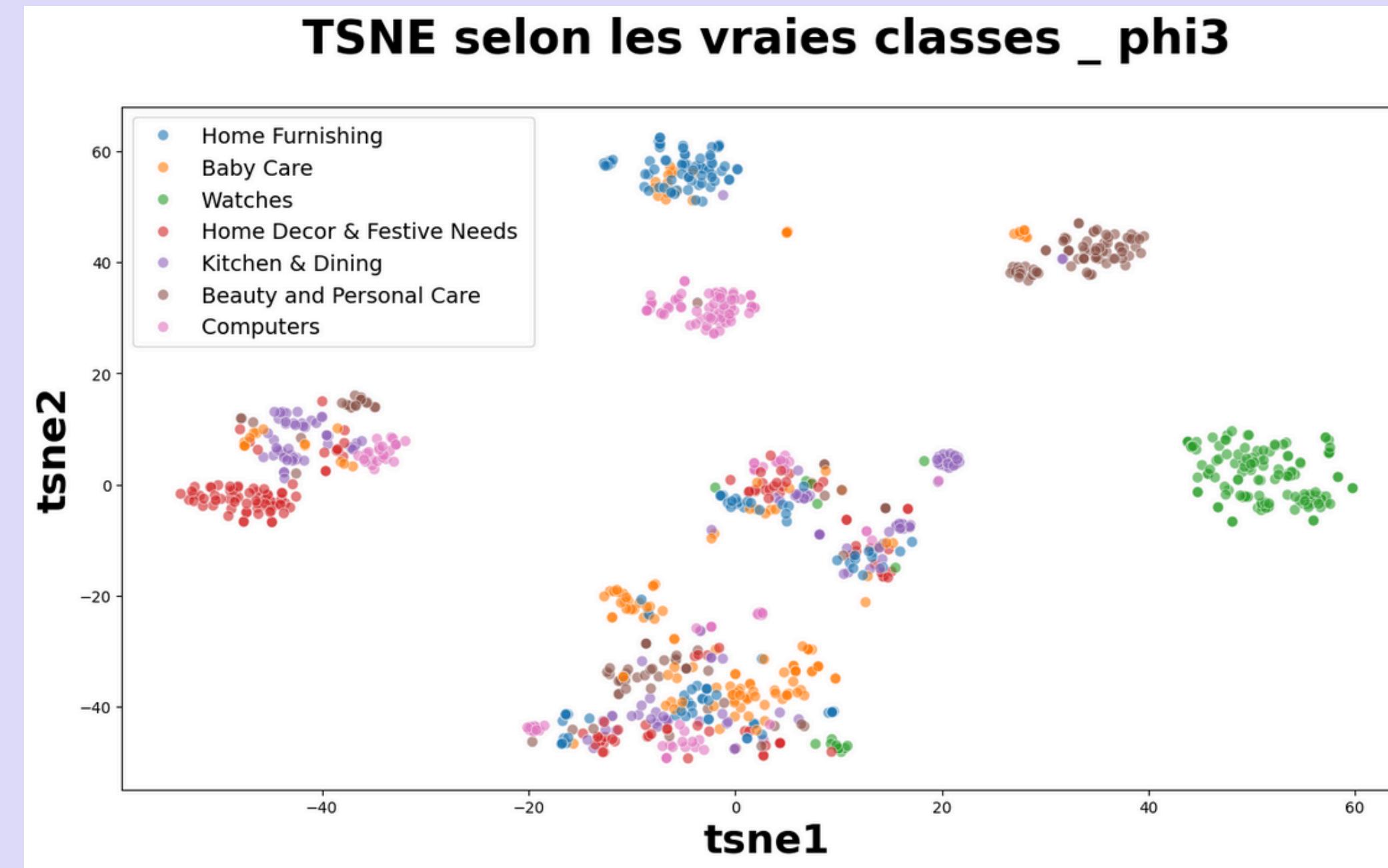
Vectorisation

- Extraction des embeddings
- PCA / T-SNE
- Clustering (kmeans)
- ARI
- Matrice de confusion

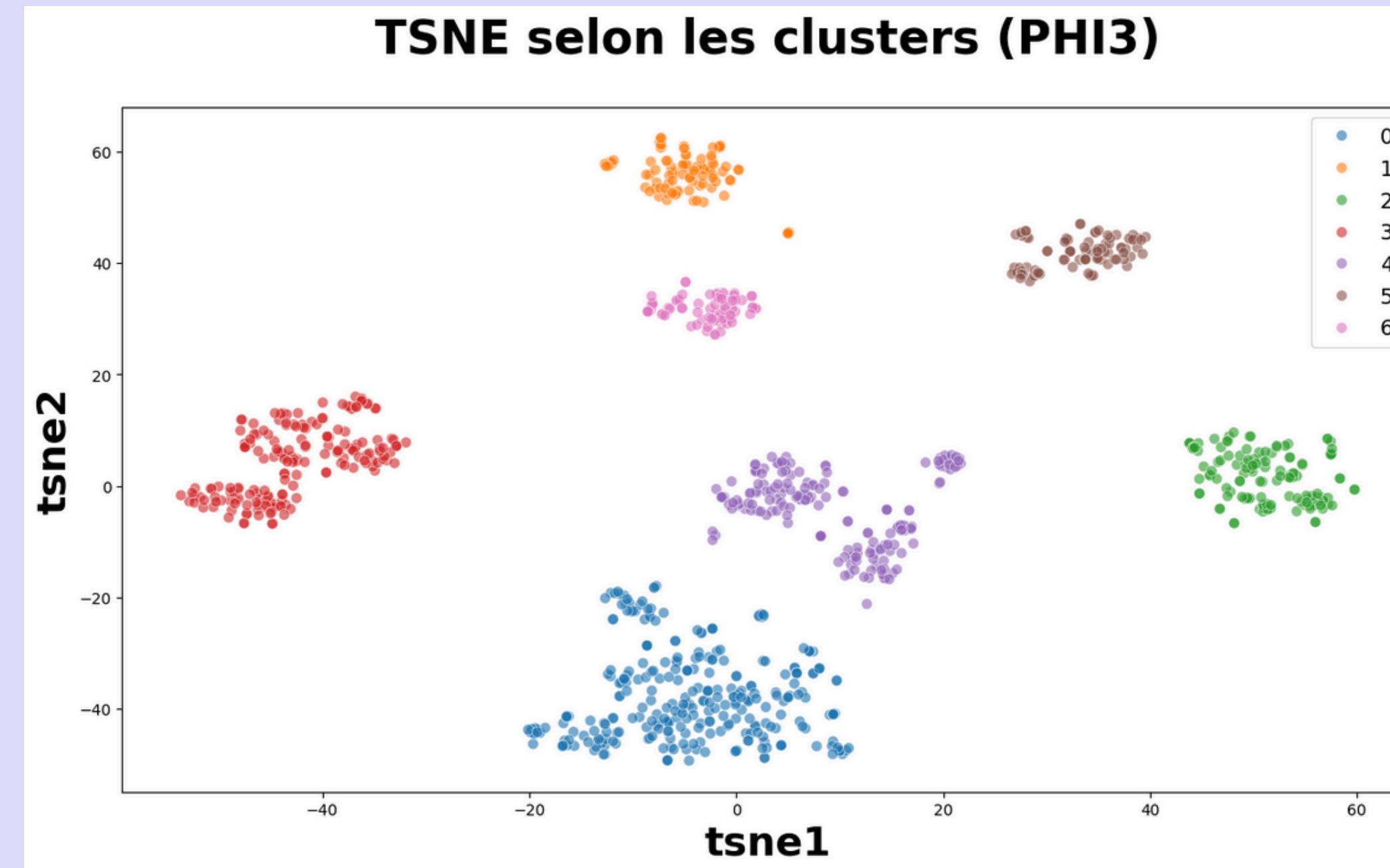


	precision	recall	f1-score	support
0	0.32	0.64	0.42	150
1	0.90	0.53	0.66	150
2	0.99	0.47	0.63	150
3	0.42	0.51	0.46	150
4	0.80	0.49	0.61	150
5	0.35	0.42	0.38	150
6	1.00	0.89	0.94	150
accuracy				0.56
macro avg	0.68	0.56	0.59	1050
weighted avg	0.68	0.56	0.59	1050

MÉTHODE 2



MÉTHODE 2



ARI (PHI3) : 0.27

RÉSULTATS



	ACCURACY	ARI
W2V	0.48	0.218
BERT	0.55	0.303
USE	0.66	0.434
PHI3 compléction	0.72	N/A
PHI3 embeddings	0.68	0.267

LIMITES & AMÉLIORATIONS POSSIBLES

Dépendance aux Descriptions Textuelles :

Le modèle repose uniquement sur la qualité et la consistance des descriptions textuelles, limitant la richesse des informations pour la classification.

Complexité des Transformers :

Les transformers, comme Phi-3 mini 4K, sont considérés comme des boîtes noires, rendant les décisions du modèle difficiles à expliquer.

Optimisation et Multimodalité :

Amélioration possible via l'optimisation des hyper-paramètres. Intégration de données multimodales (images, prix, métadonnées) pour une classification plus précise et robuste.

Interprétabilité :

Utilisation d'outils comme LIME ou SHAP pour expliquer les prédictions et rendre les résultats plus transparents.



Références

Abdin, M., Huynh, J., Saarikivi, O., Ade, S., Jacobs, M. A., Iter, D., Saied, A., et al. (2024). *Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone*. arXiv.
<https://arxiv.org/pdf/2404.14219.pdf>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. arXiv.
<https://arxiv.org/pdf/1706.03762.pdf>





Merci