

# Réalisez un Dashboard & Assurez une Veille Technique

OPENCLASSROOMS





# Réaliser un Dashboard

# Introduction au Dashboard

## Contexte du Projet

- **Entreprise** : Prêt à dépenser
- **Objectif** : Concevoir un dashboard interactif de scoring crédit
- **Problématique** : Besoin de transparence sur l'octroi des crédits

## Objectifs du Projet

- Visualisation du score de crédit et de sa probabilité
- Affichage des informations clés du client
- Comparaison avec les autres clients
- Accessibilité du dashboard
- Déploiement sur une plateforme Cloud

# Deploiement API :

## Choix Technologiques :



**Framework** : Streamlit pour une interface intuitive

**API de scoring** : Intégration de l'API développée précédemment :

<https://my-scoring-app-546acd78d8fa.herokuapp.com>

Heroku : héberger

## Documents Necessaires :



- **Procfile** : Fichier de configuration pour Heroku



- **dashboard.py** : Code streamlit de l'API

- **feature\_eng.py** : Code permettant le feature engineering



- **requirements.txt** : Liste des dépendances Python

- **runtime.txt** : Version de Python pour Heroku

# Dashboard Interactif : Présentation

The screenshot displays the 'Dashboard de Crédit Scoring' interface, which is divided into two main sections. The left section, titled 'Modification des informations client', contains five input fields for modifying client data: 'Modifier ACTIVE\_AMT\_ANNUITY\_MAX' (28359,00), 'Modifier ACTIVE\_AMT\_ANNUITY\_MEAN' (9453,00), 'Modifier AMT\_REQ\_CREDIT\_BUREAU\_DAY' (0,00), 'Modifier APPROVED\_AMT\_GOODS\_PRICE\_MEAN' (76655,25), and 'Modifier PREV\_NAME\_CONTRACT\_STATUS\_Cancelled\_MEAN' (0,00). Below these fields are three buttons: 'Mettre à jour les informations client', 'Voir les analyses avec les données de base', and 'Voir les analyses avec les données mises à jour'. The right section, titled 'Analyse du Client', contains three sub-sections: 'Analyse du Client' (with a dropdown for client ID '204020' and an 'Analyser le client' button), 'Analyse des Features' (with 'Analyse Globale & Locale des Features' and two buttons to display global and local feature importance), and 'Analyses Croisées' (with two dropdowns for feature selection, both set to 'ACTIVE\_AMT\_ANNUITY\_MAX', and three buttons to display feature distribution and bi-variate analysis). Annotations with arrows point to specific elements: 'Sélection & Analyse du Client' points to the client ID dropdown; 'Analyse des Informations Capitales du Client' points to the 'Analyse Globale & Locale des Features' section; 'Sélection & Analyses Individuelle & Croisée des Informations sur l'ensemble du Portefeuille Client' points to the feature selection dropdowns; and 'Modification & Mise à Jour des Informations Client' points to the 'Mettre à jour les informations client' button.

**Dashboard de Crédit Scoring**

**Modification des informations client**

- Modifier ACTIVE\_AMT\_ANNUITY\_MAX: 28359,00
- Modifier ACTIVE\_AMT\_ANNUITY\_MEAN: 9453,00
- Modifier AMT\_REQ\_CREDIT\_BUREAU\_DAY: 0,00
- Modifier APPROVED\_AMT\_GOODS\_PRICE\_MEAN: 76655,25
- Modifier PREV\_NAME\_CONTRACT\_STATUS\_Cancelled\_MEAN: 0,00

Mettre à jour les informations client

Voir les analyses avec les données de base

Voir les analyses avec les données mises à jour

**Analyse du Client**

Choisissez l'ID du client : 204020

Analyser le client

**Analyse des Features**

**Analyse Globale & Locale des Features**

Afficher l'importance globale des features

Afficher l'importance locale des features

**Analyses Croisées**

Choisissez une première feature : ACTIVE\_AMT\_ANNUITY\_MAX

Choisissez une deuxième feature : ACTIVE\_AMT\_ANNUITY\_MAX

Afficher la distribution des features

Afficher l'analyse bi-variée

Sélection & Analyse du Client

Analyse des Informations Capitales du Client

Sélection & Analyses Individuelle & Croisée des Informations sur l'ensemble du Portefeuille Client

Modification & Mise à Jour des Informations Client

# Fonctionnalités : Visualisation du Score de Crédit



Texte indiquant le niveau de risque de défaut de paiement.

Jauge de Probabilité de Remboursement & Seuil d'Acceptabilité (52%) :

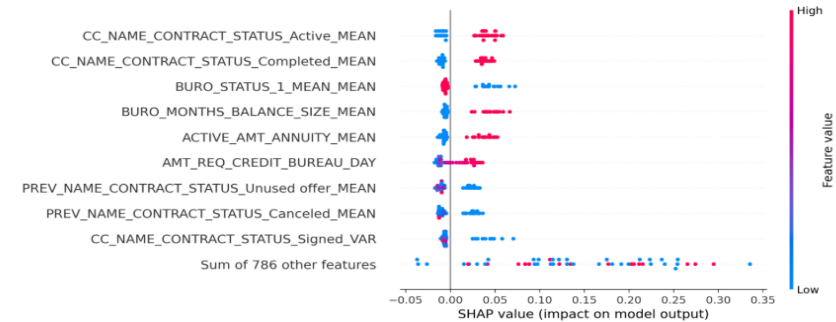
- Très Elevé
- Elevé
- Légèrement Elevé
- Moyen
- Légèrement Faible
- Faible
- Très Faible

Informations & Données du Client Sélectionné

# Features Importances Globales et Locales

Afficher l'importance globale des features

## Feature Importance Globale (SHAP)

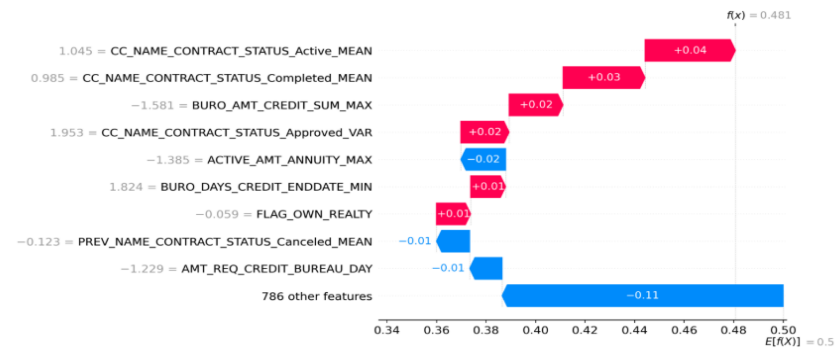


Données utilisées pour SHAP :

	CC_NAME_CONTRACT_STATUS_Active_MEAN	CC_NAME_CONTRACT_STATUS_Completed_MEAN	BU
13	None	None	

Afficher l'importance locale des features

## Feature Importance Locale (SHAP)



Données du client utilisées pour SHAP :

	CC_NAME_CONTRACT_STATUS_Active_MEAN	CC_NAME_CONTRACT_STATUS_Completed_MEAN	BU
16	None	None	

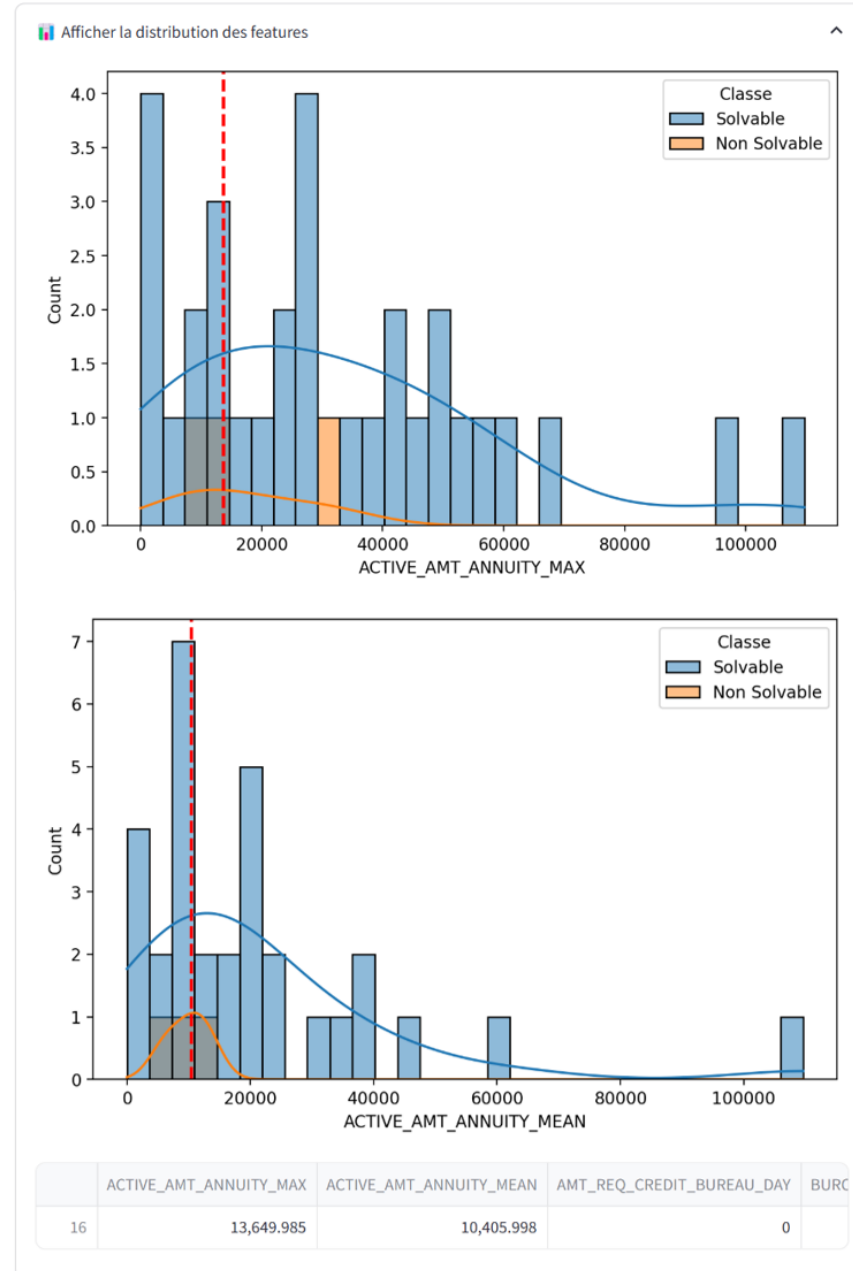
**Importance Globale :** Identification des variables les plus influentes dans le modèle global

**Importance Locale :** Explication de la décision pour un client donné

**Comparaison entre les tendances globales et les cas particuliers**

**Identification des variables clés expliquant les décisions individuelles**

# Analyse des Variables Clés & Solvabilité



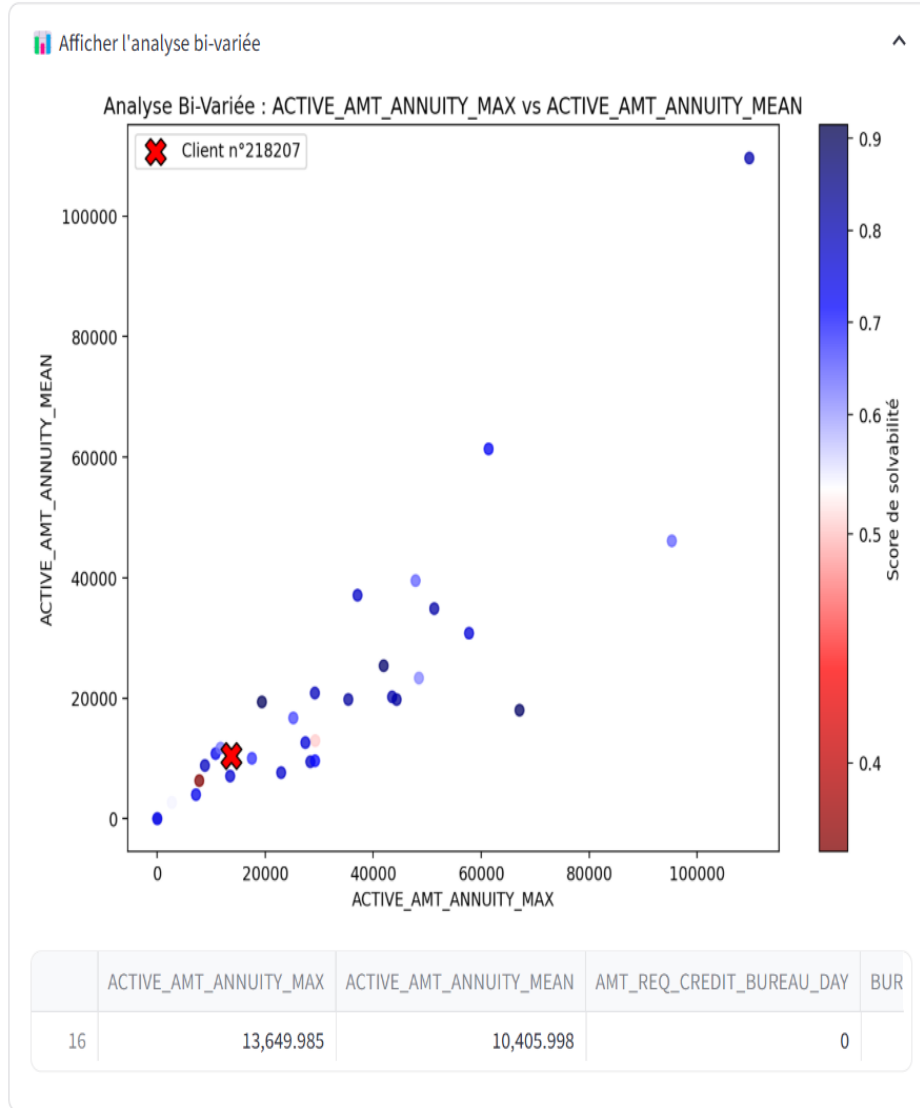
Identification des tendances entre clients à risques et clients solvables :

- Mettre en évidence des seuils critiques ou des zones à risque.
- Comprendre l'impact de chaque variable sur la solvabilité.
- Identifier les indicateurs clés dans la décision d'octroi de crédit & leurs importances.

Expliquer les décisions & identifier des pistes d'ajustement potentielles.



# Analyses Croisées et Multi-Variées



- Observation des tendances entre les 2 features sélectionnées & leur impact sur la décision de crédit.
- Détection d'éventuelles zones critiques où la probabilité de non-remboursement est plus élevée.
- Comprendre si certaines combinaisons de valeurs sont fortement corrélées avec le score.
- Affiner les analyses & explications visuelles de l'impact des critères sur la décision finale.

# Modification & Mise à Jour des Informations Client

The screenshot shows a mobile application interface titled "Modification des informations client". It features a list of six input fields, each with a "Modifier" icon and a label: "Modifier ACTIVE\_AMT\_ANNUITY\_MAX" (value: 28359,00), "Modifier ACTIVE\_AMT\_ANNUITY\_MEAN" (value: 9453,00), "Modifier AMT\_REQ\_CREDIT\_BUREAU\_DAY" (value: 0,00), "Modifier APPROVED\_AMT\_GOODS\_PRICE\_MEAN" (value: 76655,25), "Modifier PREV\_NAME\_CONTRACT\_STATUS\_Canceled\_MEAN" (value: 0,00), and "Modifier PREV\_NAME\_CONTRACT\_STATUS\_Unused offer\_MEAN" (value: 0,00). Each field has minus and plus icons for adjustment. Below the fields is a button labeled "Mettre à jour les informations client". At the bottom, there are two buttons with eye icons: "Voir les analyses avec les données de base" and "Voir les analyses avec les données mises à jour".

## Mise à jour dynamique des données client :

- Modification des informations via un formulaire interactif situé sur une barre latérale
- Soumission et validation des modifications
- Recalcul des scores & des explications SHAP
- Mise à jour du score de crédit
- Comparaison des résultats & analyses avant/après modification

## Impact & Bénéfices :

- Expérience utilisateur améliorée : interaction fluide & intuitive
- Prise de décision optimisée : impact immédiat des modifications visibles
- Exploration dynamique : évaluation rapide de scénarios alternatifs

**Approche interactive & transparente pour tester différentes hypothèses & affiner l'analyse du scoring crédit**

# Synthèse Dashboard

## **Dashboard interactif, fonctionnel & déployé**

### **Synthèse des Travaux :**

- Visualisation intuitive & explicative des scores crédit
  - Comparaison dynamique
  - Modification dynamique des informations client
- Mise à jour instantanée du score

### **Perspectives d'Amélioration :**

- Feature Engineering & Données volumineuses
- Interprétation des Variables
- Déséquilibre des Scores
- Tests utilisateurs & retour d'expérience pour améliorer l'UI/UX

## **Dashboard fonctionnel & explicable, meilleure transparence**

→ prise de décision facilitée

# **Assurer une Veille Technique**

# Introduction à la Veille Technique

## Objectif:

- Comparer une approche classique et un modèle d'état de l'art en NLP
- Tester leur efficacité, performance et robustesse sur un jeu de données e-commerce

Classique :  
**bert-base-uncased**



Etat de l'Art :  
**all-MiniLM-L6-v2**

## Jeu de données utilisé :

**Source :** Flipkart (plateforme e-commerce)

### Attributs clés :

- product\_name (Nom du produit)
- description (Description du produit)
- product\_category\_tree (Hiérarchie des catégories)
- product\_category (Catégorie principale)

# Présentation de MiniLM

MiniLM = version allégée de BERT/RoBERTa, obtenue par distillation de connaissance.

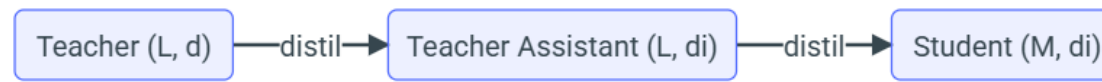
2x plus rapide que les modèles classiques

99% précision.

## Comment ça fonctionne ?

Apprentissage basé sur un transfert d'attention depuis un modèle plus grand

Seule la dernière couche du Transformer est distillée pour optimiser l'apprentissage.

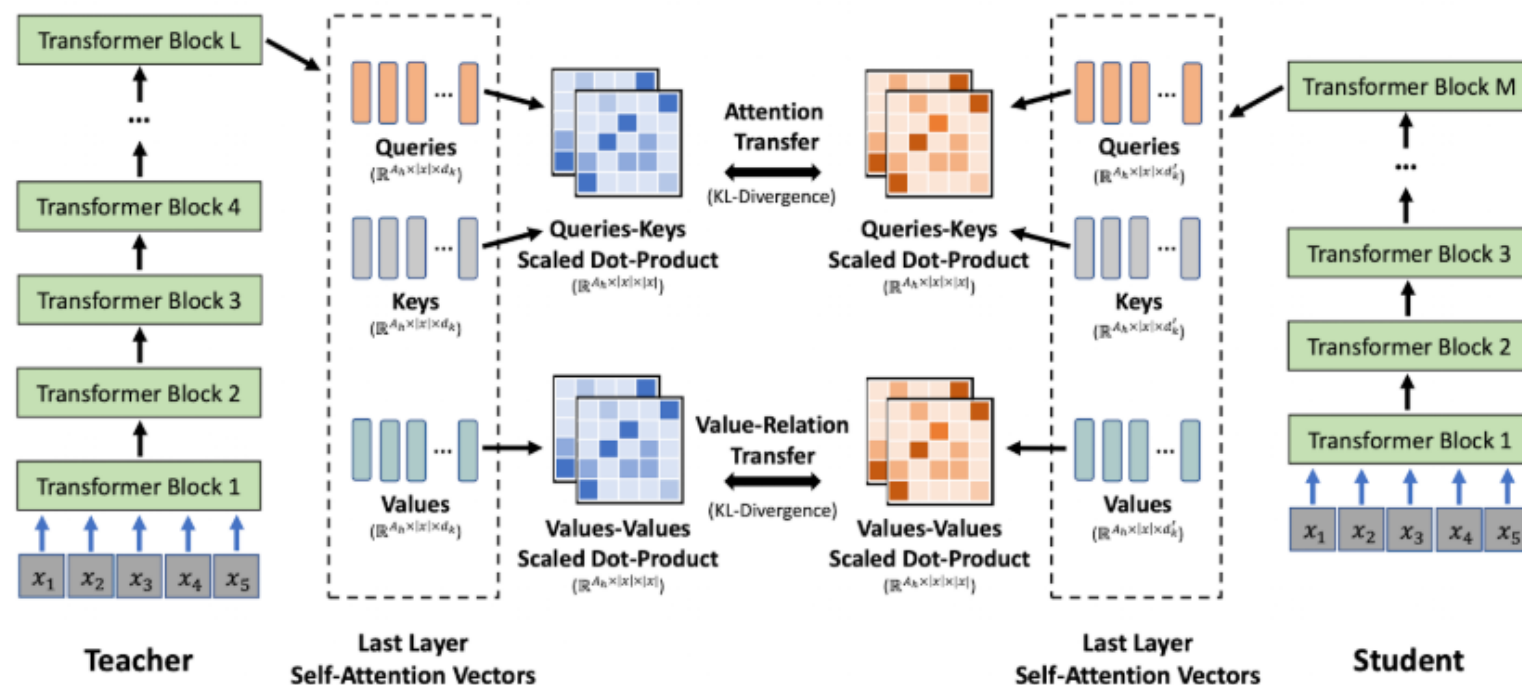


Dual Attention Transfer : transfert de relations entre les Questions-Clés (Q-K) et les Valeurs-Valeurs (V-V).

MiniLM 3x plus petit que BERT-Base

Distillation efficace → rivalise avec des modèles plus volumineux en classification et en compréhension du langage.

# Distillation Profonde de l'Auto-Attention & Comparatif de performances



Approach	Teacher Model	Distilled Knowledge	Layer-to-Layer Distillation	Requirements on the number of layers of students	Requirements on the hidden size of students
DistillBERT	BERT <sub>BASE</sub>	Soft target probabilities Embedding outputs			✓
TinyBERT	BERT <sub>BASE</sub>	Embedding outputs Hidden states Self-Attention distributions	✓		
MOBILEBERT	IB-BERT <sub>LARGE</sub>	Soft target probabilities Hidden states Self-Attention distributions	✓	✓	✓
MINILM	BERT <sub>BASE</sub>	Self-Attention distributions Self-Attention value relation			

# Méthodologie : Classification Report (sklearn.metrics)

Catégorie	Precision	Recall	F1-Score	Support
Baby Care	0,88	0,78	0,82	27
Beauty and Personal Care	1,00	1,00	1,00	21
Computers	0,97	0,97	0,97	38
Home Decor & Festive Needs	0,85	0,93	0,89	30
Home Furnishing	0,91	0,89	0,90	35
Kitchen & Dining	0,93	0,96	0,94	26
Watches	1,00	1,00	1,00	33
Accuracy			0,93	210
Macro Avg.	0,93	0,93	0,93	210
Weighter Avg.	0,93	0,93	0,93	210

**bert-base-uncased**



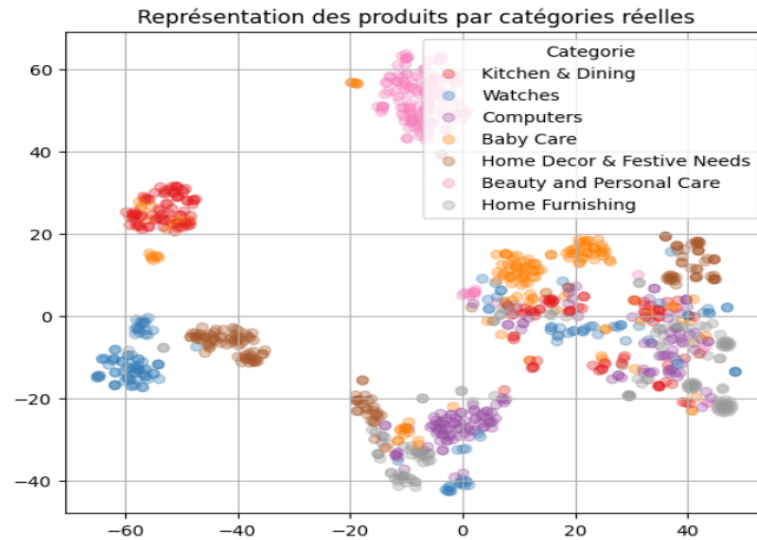
Catégorie	Precision	Recall	F1-Score	Support
Baby Care	1,00	0,70	0,83	27
Beauty and Personal Care	0,88	1,00	0,93	21
Computers	0,97	1,00	0,99	38
Home Decor & Festive Needs	0,85	0,97	0,91	30
Home Furnishing	0,94	0,97	0,96	35
Kitchen & Dining	1,00	0,96	0,98	26
Watches	1,00	1,00	1,00	33
Accuracy			0,95	210
Macro Avg.	0,95	0,94	0,94	210
Weighter Avg.	0,95	0,95	0,95	210

**all-MiniLM-L6-v2**

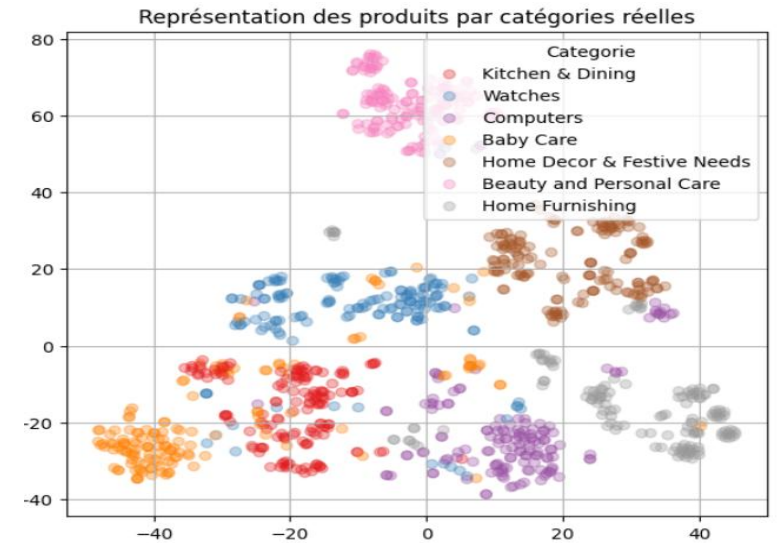
- all-MiniLM-L6-v2 offre une légère amélioration des scores globaux
- Meilleure précision pour certaines catégories comme Kitchen & Dining
- Baisse de précision pour Beauty & Personnel Care
- Catégorie Baby Care montre une baisse de recall pour all-MiniLM-L6-v2



# Méthodologie : Visualisation des embeddings avec t-SNE



**bert-base-uncased :**  
**ARI Score = 0,3095**



**all-MiniLM-L6-v2 :**  
**ARI Score = 0.7119**

- all-MiniLM-L6-v2 montre une bien meilleure séparation des clusters
- Les embeddings sont plus discriminants et alignés avec les catégories réelles
- bert-base-uncased produit des clusters plus mélangés, entraînant une moins bonne séparation

**all-MiniLM-L6-v2 capture mieux les structures sous-jacentes des catégories**

# Synthèse Veille Technique

## Résumé des performances :

- all-MiniLM-L6-v2 offre une meilleure classification et une séparation plus nette des clusters
- Amélioration du score ARI (0.71 vs 0.30), indiquant une meilleure robustesse
- Performances accrues, avec une précision légèrement supérieure à bert-base-uncased

## Perspectives :

- Intégration de all-MiniLM-L6-v2 dans le pipeline de classification e-commerce
- Exploration d'autres modèles légers comme DistilBERT pour optimiser rapidité et coût
- Utilisation de méthodes avancées de réduction de dimension (UMAP, PCA)

# Conclusion Générale

## Bilan du Projet

- ✓ Dashboard de Scoring Crédit : Développement d'une interface interactive permettant une analyse transparente et détaillée du score de crédit des clients.
- ✓ Veille Technique en NLP : Comparaison entre BERT-base-uncased et all-MiniLM-L6-v2 sur un jeu de données e-commerce pour évaluer leur efficacité, performance et robustesse.

## Apports & Enseignements

- ◆ Explicabilité et Interprétation : impact des variables sur le score & analyse des embeddings NLP via t-SNE et clustering KMeans.
- ◆ Contraintes et Optimisation : Gestion de données volumineuses pour le dashboard et enjeux liés à l'équilibre des classes en scoring crédit et en NLP.
- ◆ Performances Modélisation : Amélioration des scores avec all-MiniLM-L6-v2 par rapport à BERT-base-uncased, notamment en classification et structuration des embeddings.

## Perspectives Futures

- ✂ Automatisation complète du feature engineering pour une intégration cloud du dashboard.
- ✂ Optimisation du scoring face à l'équilibre des classes pour affiner les analyses prédictives.
- ✂ Exploration régulière de modèles NLP plus avancés pour améliorer la structuration et la classification des produits.

🌟 **Un projet mêlant analyse de données, modélisation et interprétabilité pour une meilleure prise de décision et une veille sur les dernières avancées en IA !**



**Merci pour votre  
attention**