

# **Projet 8 : Réalisez un dashboard et assurez une veille technique**

**25/11/2024**

**Soukaina GUAOUA ELJADDI**

**Parcours Data Scientist  
OpenClassrooms**

# Plan:

- ❑ Rappel de la problématique
- ❑ Présentation du dashboard
  - Présentation des graphiques du dashboard
  - Démonstration de l'application dashboard, déployée sur le Cloud
- ❑ Présentation du travail de veille
  - Description nouvelle approche ou technique
  - Synthèse comparative des résultats

# Problématique

**Contexte** : Société financière "**Prêt à dépenser**" qui propose des crédits à la consommation.

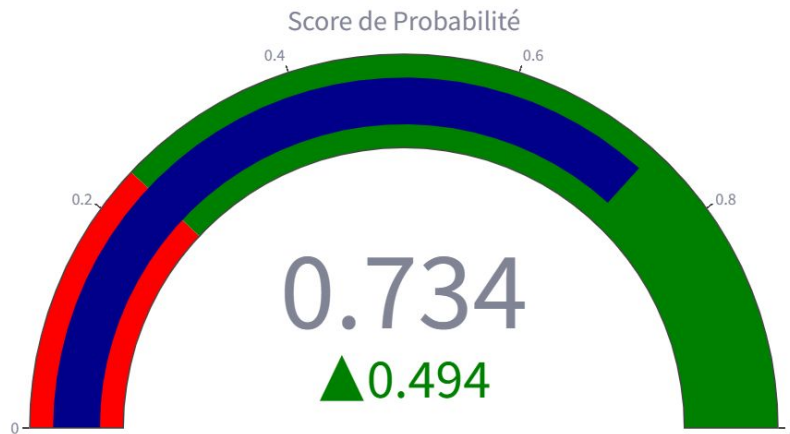
- “**scoring crédit**” pour calculer **la probabilité** qu’un client rembourse son **crédit**.
- un **algorithme de classification** (crédit accordé ou refusé)

## **Objectifs et Missions :**

- Concevez un dashboard de credit scoring.
- Réalisez une veille technique.

# Présentation du dashboard

Présentation des graphiques du dashboard :



## Critères WCAG :

1. Contenu non textuel
2. Utilisation de la couleur
3. Contraste (minimum)
4. Redimensionnement du texte
5. Titre de page

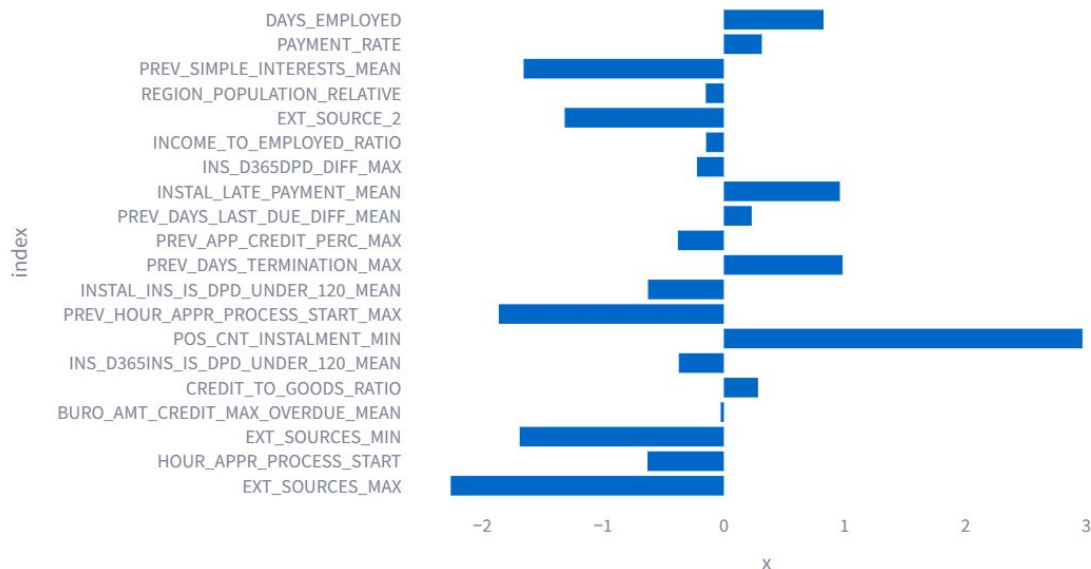
Prédiction : Crédit Accordé

Probabilité de paiement : 0.73

# Présentation du dashboard

Présentation des graphiques du dashboard :

## Importance Locale des Caractéristiques



# Présentation du dashboard

## Présentation des graphiques du dashboard :

### Analyse Univariée des Caractéristiques

Sélectionnez la première caractéristique

DAYS\_EMPLOYED |



Sélectionnez la deuxième caractéristique

POS\_CNT\_INSTALLMENT\_MIN



### Définition de la caractéristique sélectionnée

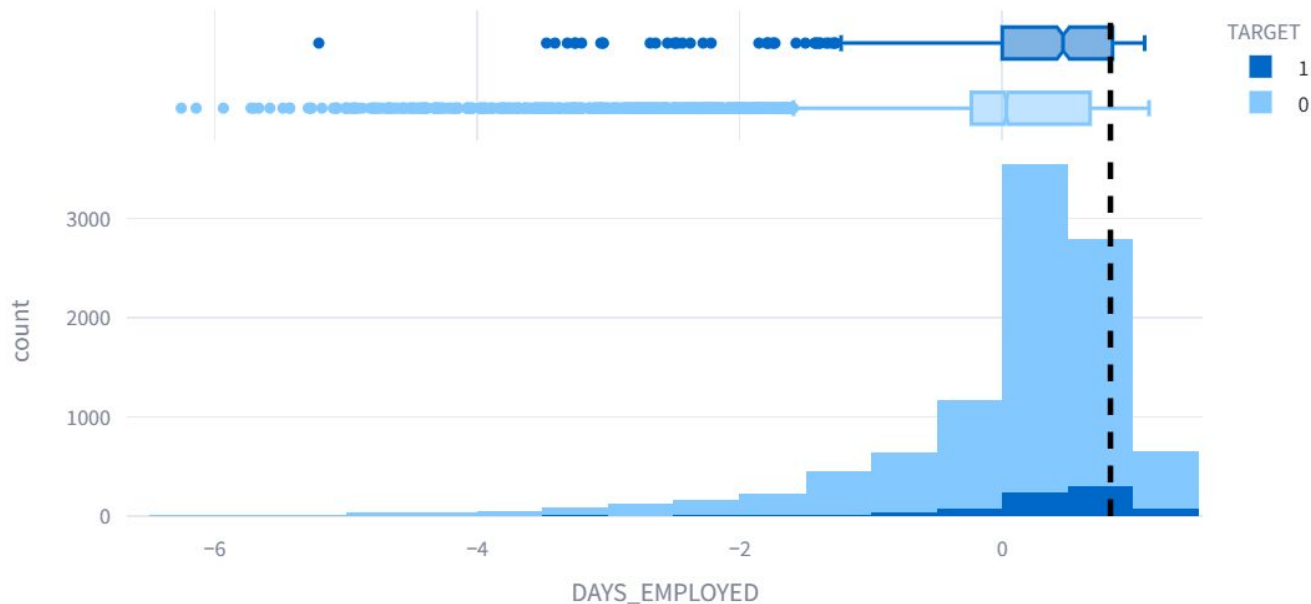
**DAYS\_EMPLOYED:** Nombre de jours que le client a été employé.

**POS\_CNT\_INSTALLMENT\_MIN:** Le nombre minimal de paiements par versement.

# Présentation du dashboard

Présentation des graphiques du dashboard :

Distribution de DAYS\_EMPLOYED



# Présentation du dashboard

## Présentation des graphiques du dashboard :

### Analyse Bi-Variée des Caractéristiques

Relation entre DAYS\_EMPLOYED et POS\_CNT\_INSTALLMENT\_MIN

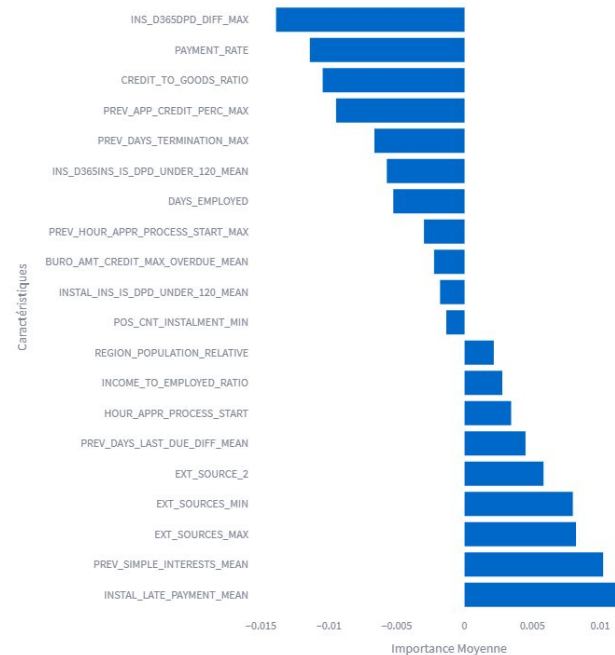




# Présentation du dashboard

Présentation des graphiques du dashboard : Importance globale des caractéristiques

Top 20 des Caractéristiques les Plus Importantes



# Présentation du dashboard

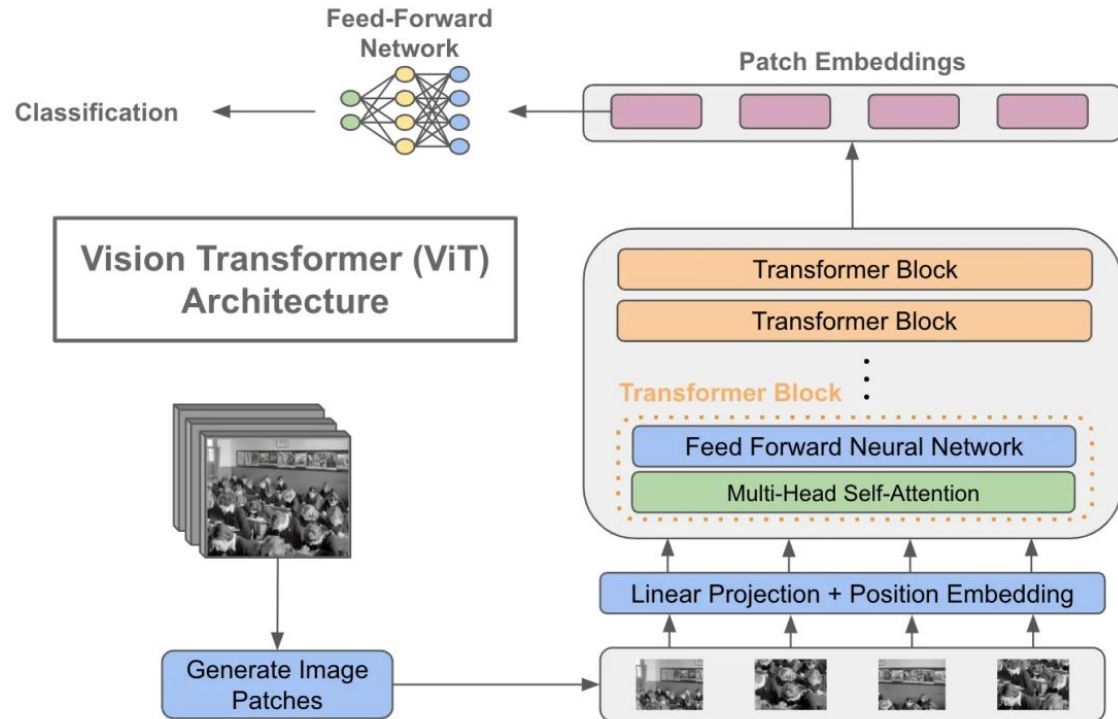
Démonstration de l'application dashboard, déployée sur le Cloud :

<https://p7apid-ploiement-e26adgmbbclbb5rzd4ujfd.streamlit.app/>

# Présentation du travail de veille

## Description nouvelle approche ou technique : ViT (Vision Transformer)

**ViT (Vision Transformer) :**  
Modèle basé sur les Transformers, adapté à la vision par ordinateur pour capturer des relations globales.



# Présentation du travail de veille

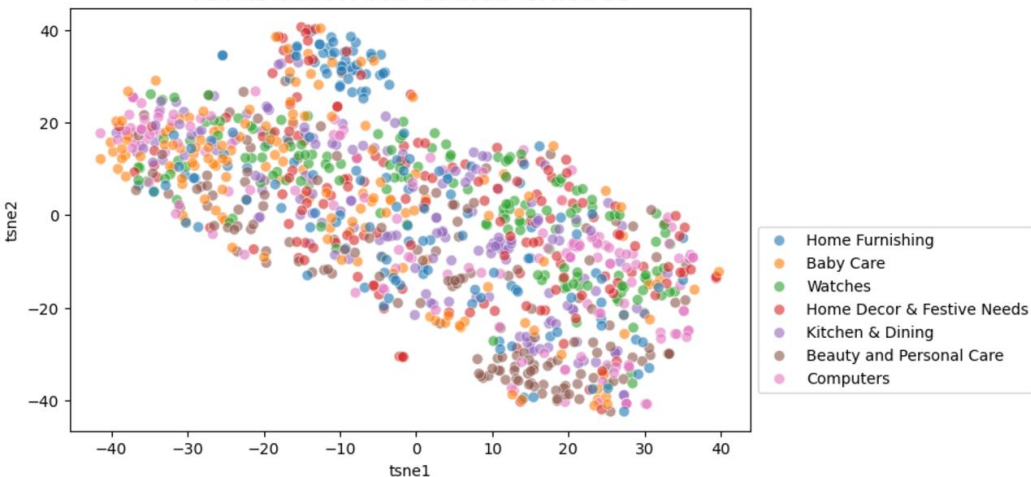
## Démarche de modélisation :

1. Préparation des données
2. Extraction des caractéristiques avec Vision Transformer (ViT)
3. Réduction de dimension (PCA)
4. Clustering avec KMeans
5. Évaluation et Matrice de confusion

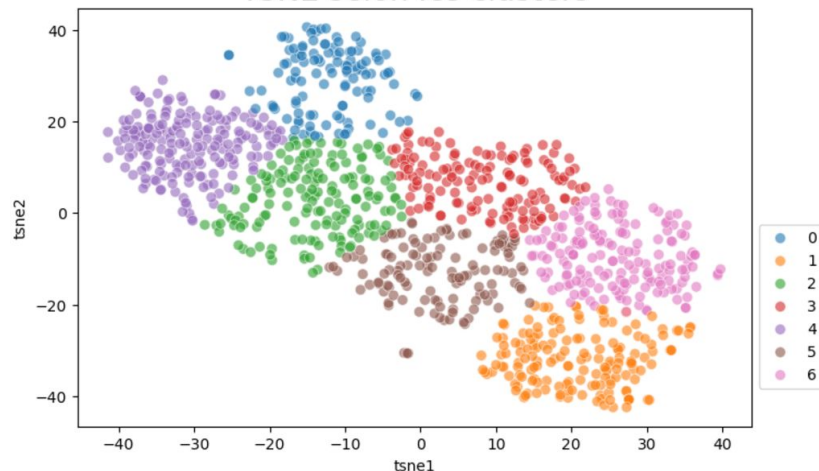
# Présentation du travail de veille

## Résultats : Visualisation des clusters formés par ViT

TSNE selon les vraies classes

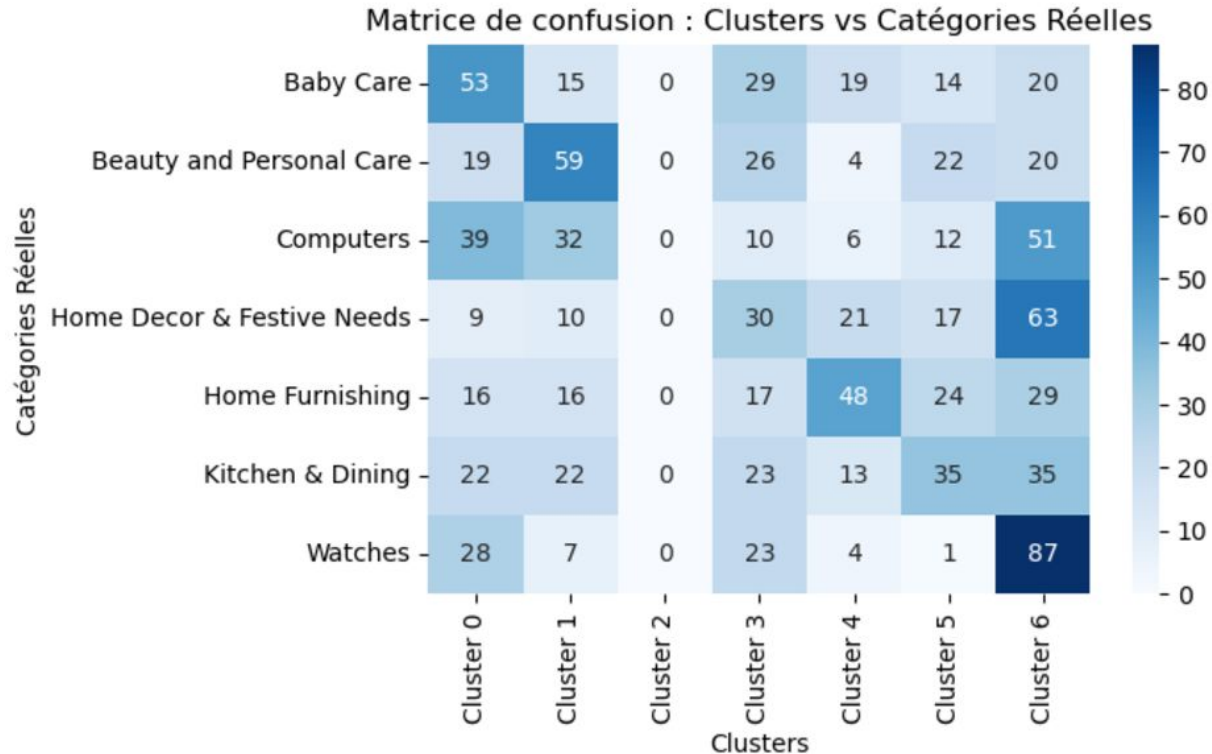


TSNE selon les clusters



# Présentation du travail de veille

## Résultats : Matrice de confusion



# Présentation du travail de veille

## Comparaison avec d'autres approches :

**SIFT (Scale-Invariant Feature Transform)** : Méthode basée sur l'extraction de points d'intérêt et de descripteurs locaux.

**VGG16** : Réseau convolutif profond pré-entraîné, capable d'extraire des caractéristiques globales après fine-tuning.

## Métriques utilisés :

**ARI** : Mesure la correspondance entre deux partitions (clusters vs classes réelles), allant de -1 (aucune correspondance) à 1 (parfaite correspondance).

**accuracy** : Pourcentage de prédictions correctes sur le total des échantillons.

**F1-score** : Moyenne harmonique de la précision et du rappel

# Présentation du travail de veille

## Synthèse comparative des résultats

	<b>SIFT</b>	<b>VGG16</b>	<b>Vit</b>
<b>ARI</b>	0.0019	0.439	0.054
<b>Accuracy</b>	0.14	0.65	0.30
<b>F1-score moyen</b>	0.14	0.66	0.27



# Présentation du travail de veille

## Limites et améliorations ViT:

### Limites identifiées :

1. Données d'Entraînement Limitées.
2. Complexité Computationnelle.
3. Confusions Inter-Catégories.
4. Faible Interprétabilité.
5. Sensibilité aux Hyperparamètres.

# Présentation du travail de veille

## Améliorations possibles :

1. Augmentation des Données.
2. Fine-Tuning du Modèle Pré-Entraîné.
3. Réduction de la Résolution et des Patches.
4. Exploration de ViT Hybrides.
5. Meilleure Sélection des Hyperparamètres.
6. Optimisation Matériel et Logiciel.

# Conclusion

- Les Performances limitées de ViT sont dues à l'absence de fine-tuning, qui est crucial pour adapter le modèle aux spécificités de l'ensemble de données.
- ViT est prometteur mais nécessite de grandes quantités de données pour un apprentissage optimal.