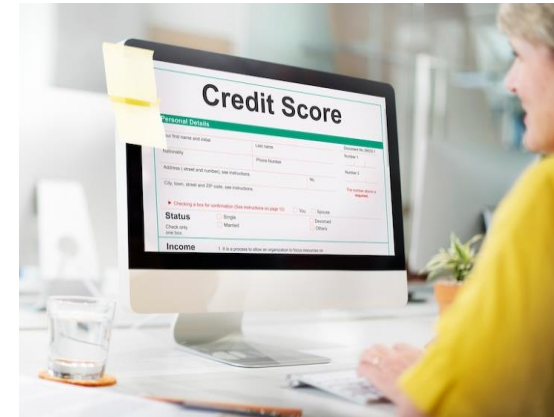




# Réalisation d'un **tableau de bord** de « scoring de crédit » et **d'une veille technique**



Projet : Data Scientist #8

Oumou Faye

Mentor : Medina Hadjem

# PLAN DE PRÉSENTATION

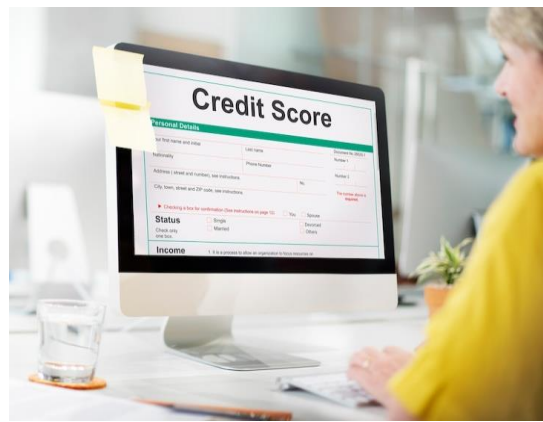
INTRODUCTION  
-  
MISSION



DASHBOARD :  
PRÉSENTATION  
+  
DÉMONSTRATION



PRÉSENTATION DE  
L' APPROCHE TECHNIQUE



# INTRODUCTION - MISSION

# INTRODUCTION - MISSION



## Objectif

Mettre à disposition un tableau de bord pour les chargés de relation client



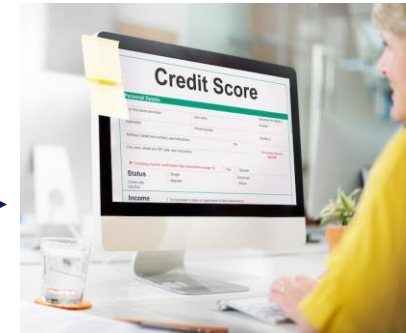
Permettre aux chargés de relation client d'accorder ou de refuser la demande de crédit



Motiver la décision d'octroi de crédit au regard des critères de solvabilité du client

## Mission

Mise en place d'un « dashboard de scoring credit »



Évaluer la solvabilité des clients demandeurs

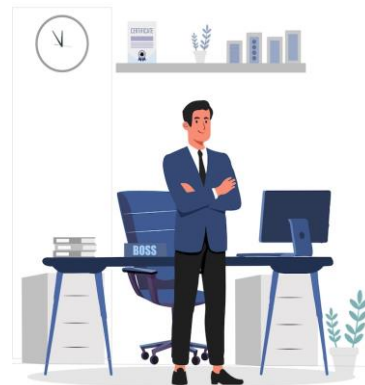
Déployer l'interface utilisateur pour les chargés de relation client

## Fonction

Data Scientist



Manager



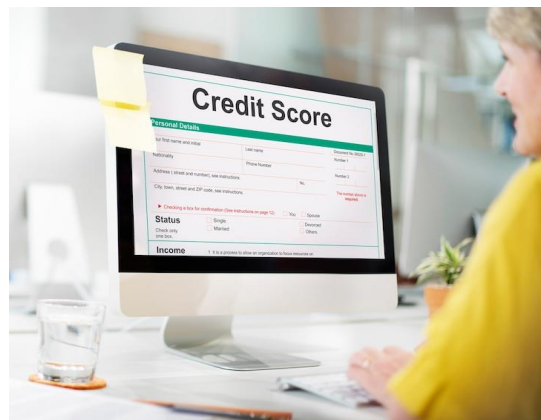
Mickaël

## Entreprise



Activité

Proposition de crédit à la consommation  
(clients à historique limité)



# PRÉSENTATION ET DÉMONSTRATION DU TABLEAU DE BORD





Développement du tableau de bord sous Python



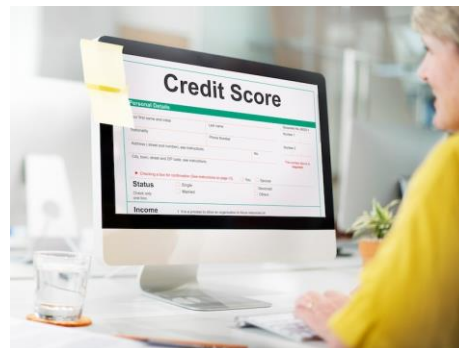
*Réalisation de tests du tableau de bord en local*



Dépôt des fichiers requis :

- Tableau de bord (fichier Python (interface UI))
- Modèle de prédiction au format .pkl
- Fichier des données de test
- Fichier des dépendances requises (requirements.txt)

Mise en production du tableau de bord  
Interface (UI) ·





Évaluation de la solvabilité du client



Les champs de saisie des  
caractéristiques du client

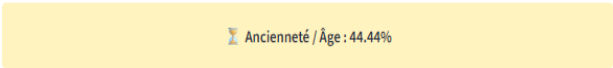
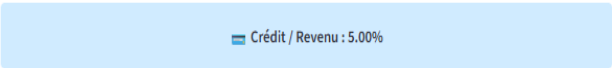
Critères du client et son score de solvabilité

Comparaison du client avec population étudiée

Influence du profil client sur la décision finale de prêt

Saisir les caractéristiques du client

|  |   |  |                                      |  |
|--|---|--|--------------------------------------|--|
| Revenu annuel (€)                      | Montant du crédit (€)                         | Mensualité du crédit (€)                         |                                      |  |
| 100.000                                | 5.000   | 100 - +  |                                      |  |
| Âge (années)                           | Durée de remboursement (mois)                 | Année d'ancienneté (emploi)                      |                                      |  |
| 45 - +                                 | 24 - +  | 20 - +   |                                      |  |
| Score de solvabilité externe n°1       | Score de solvabilité externe n°2              | Score de solvabilité externe n°3                 |                                      |  |
| 0,95 - +                               | 0,96 - +                                      | 0,97 - +   |                                      |  |
| Nombre d'enfants à charge              | Note de la région (ville 1: faible, 3: bonne) | Note de la région (générale 1: faible, 3: bonne) |                                      |  |
| 0 - +                                  | 3   | 3  |                                      |  |
| Genre                                  | Travaille-t-il dans une autre ville ?         | Propriétaire d'un bien immobilier ?              | Travailleur manuel ?                 | Type de revenu                           |
| <input checked="" type="radio"/> Homme | <input type="radio"/> Oui                     | <input checked="" type="radio"/> Oui             | <input type="radio"/> Oui            | <input checked="" type="radio"/> Travail |
| <input type="radio"/> Femme            | <input checked="" type="radio"/> Non          | <input type="radio"/> Non                        | <input checked="" type="radio"/> Non | <input type="radio"/> Autre revenu       |



Évaluer la Solvabilité du client

Réinitialiser le formulaire client

# PRÉSENTATION DES GRAPHIQUES DU DASHBOARD



## Évaluation de la solvabilité du client



Critères du client et son score de solvabilité

Comparaison du client avec population étudiée

Influence du profil client sur la décision finale de prêt

### Saisir les caractéristiques du client

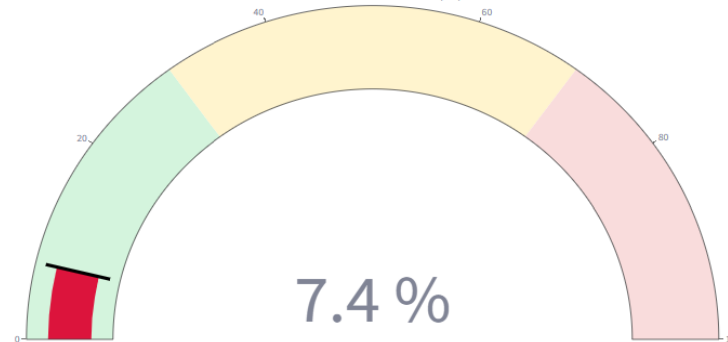
|  |  |  |
|--|--|--|
| Revenu annuel (€)  | Montant du crédit (€)  | Mensualité du crédit (€)   |
| 100.000  | 5.000  | 100 - +  |
| Âge (années)   | Durée de remboursement (mois)  | Année d'ancienneté (emploi)  |
| 45 - +   | 24 - +   | 20 - +   |
| Score de solvabilité externe n°1   | Score de solvabilité externe n°2   | Score de solvabilité externe n°3   |
| 0,95 - +   | 0,96 - +   | 0,97 - +   |
| Nombre d'enfants à charge  | Note de la région (ville 1: faible, 3: bonne)  | Note de la région (générale 1: faible, 3: bonne)   |
| 0 - +  | 3  | 3  |
| Genre<br><input checked="" type="radio"/> Homme<br><input type="radio"/> Femme | Travaille-t-il dans une autre ville ?<br><input type="radio"/> Oui<br><input checked="" type="radio"/> Non | Propriétaire d'un bien immobilier ?<br><input checked="" type="radio"/> Oui<br><input type="radio"/> Non |
|  | Travailleur manuel ?<br><input type="radio"/> Oui<br><input checked="" type="radio"/> Non                  | Type de revenu<br><input checked="" type="radio"/> Travail<br><input type="radio"/> Autre revenu         |

Évaluer la Solvabilité du client

Réinitialiser le formulaire client

Décision d'octroi de crédit : accepté

Probabilité de défaut (%)



Bouton d'activation du calcul de la probabilité de défaut de remboursement du demandeur

Bouton de réinitialisation du formulaire de saisie des informations

**i** Interprétation du score de solvabilité : Le score de solvabilité est un indicateur numérique. On l'utilise pour évaluer la solvabilité ou le risque d'un client à partir d'informations externes, comme des données de bureau de crédit ou d'autres bases de données financières. Plus le score est élevé (proche de 1), plus la probabilité de remboursement est forte.



# PRÉSENTATION DES GRAPHIQUES DU DASHBOARD



## Top 10 des critères qui influencent le plus l'accord de crédit

*Selon l'ensemble des clients étudiés*

### i Signification des valeurs: critères client ?

- Genre : 1 = Homme, 0 = Femme
- Travail : 1 = Oui, 0 = Non
- Travailleur manuel : 1 = Oui, 0 = Non
- Autre ville : 1 = Oui, 0 = Non
- Propriétaire : 1 = Oui, 0 = Non
- Type de revenu : 1 = Travail salarié, 0 = Autre (pension, aide...)
- Scores de solvabilité : proche de 1 = client plus fiable
- Ratios (ex. crédit / revenu) : plus c'est bas, mieux c'est

Menu déroulant affichant : la  
signification des valeurs des critères  
clients (encodés)

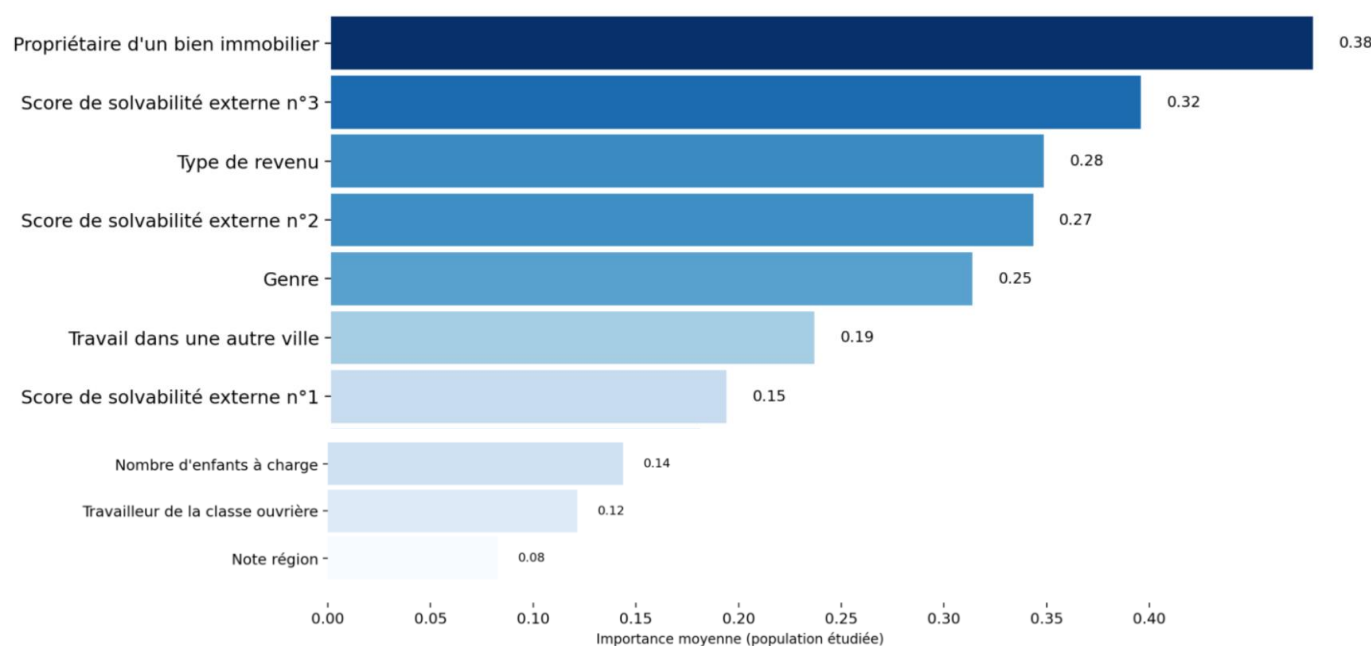
### Shapley Additive exPlanations

#### i Interprétation globale :

Ce graphique présente l'impact moyen de chaque variable sur la décision de crédit, en analysant l'ensemble des clients.

Les variables les plus influentes sont affichées en haut.

Plus la valeur absolue est grande, plus l'effet est déterminant sur la décision.



## Évaluation de la solvabilité du client



Critères du client et son score de solvabilité

Comparaison du client avec population étudiée

Influence du profil client sur la décision finale de prêt

Position du client face à la population étudiée selon la valeur du critère

Sélectionnez le critère client à comparer avec la population de référence :

Détails sur les valeurs des caractéristiques ⓘ

Durée de remboursement (mois) ▼



Client

24 mois

## Évaluation de la solvabilité du client



Option qui permet de comparer deux critères du client.

Critères du client et son score de solvabilité

Comparaison du client avec population étudiée

Influence du profil client sur la décision finale de prêt

Visualisez deux critères du dossier client à comparer :

Position du client face à la population étudiée selon la valeur des critères

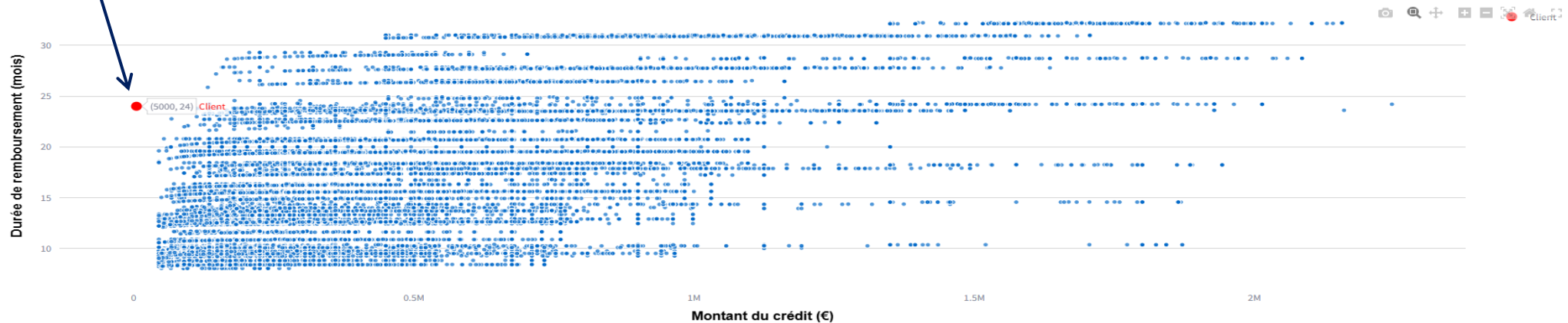
Détails sur les valeurs des caractéristiques ⓘ

Critère n°1

Montant du crédit (€)

Critère n°2

Durée de remboursement (mois)



## Évaluation de la solvabilité du client



Critères du client et son score de solvabilité

Comparaison du client avec population étudiée

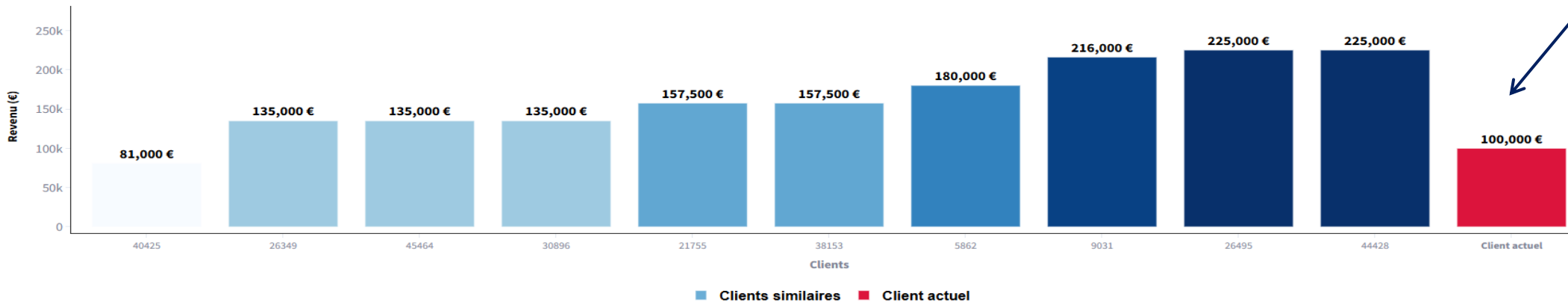
Influence du profil client sur la décision finale de prêt

## Profil du client comparé à des clients similaires

Sélectionnez un critère à comparer ⓘ

Revenu (€) ▼

Critère sélectionné : Revenu (€)



Valeur du client face à des clients similaires

ⓘ Les principaux critères de comparaison: Revenu du client / Montant du crédit demandé / Mensualité / Les scores de solvabilité (1,2 et 3).



Évaluation de la solvabilité du client



Critères du client et son score de solvabilité      Comparaison du client avec population étudiée      Influence du profil client sur la décision finale de prêt


Influence du profil client sur la décision finale

Cette section permet d'analyser l'impact de chaque caractéristique sur la décision de crédit.

Synthèse des caractéristiques du demandeur

Détails sur les valeurs des caractéristiques ⓘ

| Critère d'influence               | Valeur | Facteurs déterminants dans la décision de crédit du client |
|-----------------------------------|--------|--|
| Score de solvabilité externe n°3  | 0.97   | -0.64  |
| Score de solvabilité externe n°2  | 0.96   | -0.56  |
| Travail dans une autre ville      | 0      | -0.20  |
| Score de solvabilité externe n°1  | 0.95   | -0.20  |
| Type de revenu                    | 1      | -0.18  |
| Note région                       | 3      | -0.16  |
| Genre                             | 1      | -0.15  |
| Travailleur de la classe ouvrière | 0      | -0.14  |
| Nombre d'enfants à charge         | 0      | -0.13  |
| Propriétaire d'un bien immobilier | 1      | -0.12  |

 Télécharger les données au format CSV

- Genre : 1 = Homme, 0 = Femme
- Travail : 1 = Oui, 0 = Non
- Travailleur manuel : 1 = Oui, 0 = Non
- Autre ville : 1 = Oui, 0 = Non
- Propriétaire : 1 = Oui, 0 = Non
- Type de revenu : 1 = Travail salarié, 0 = Autre (pension, aide...)
- Scores de solvabilité : proche de 1 = client plus fiable
- Ratios (ex. crédit / revenu) : plus c'est bas, mieux c'est



## Évaluation de la solvabilité du client



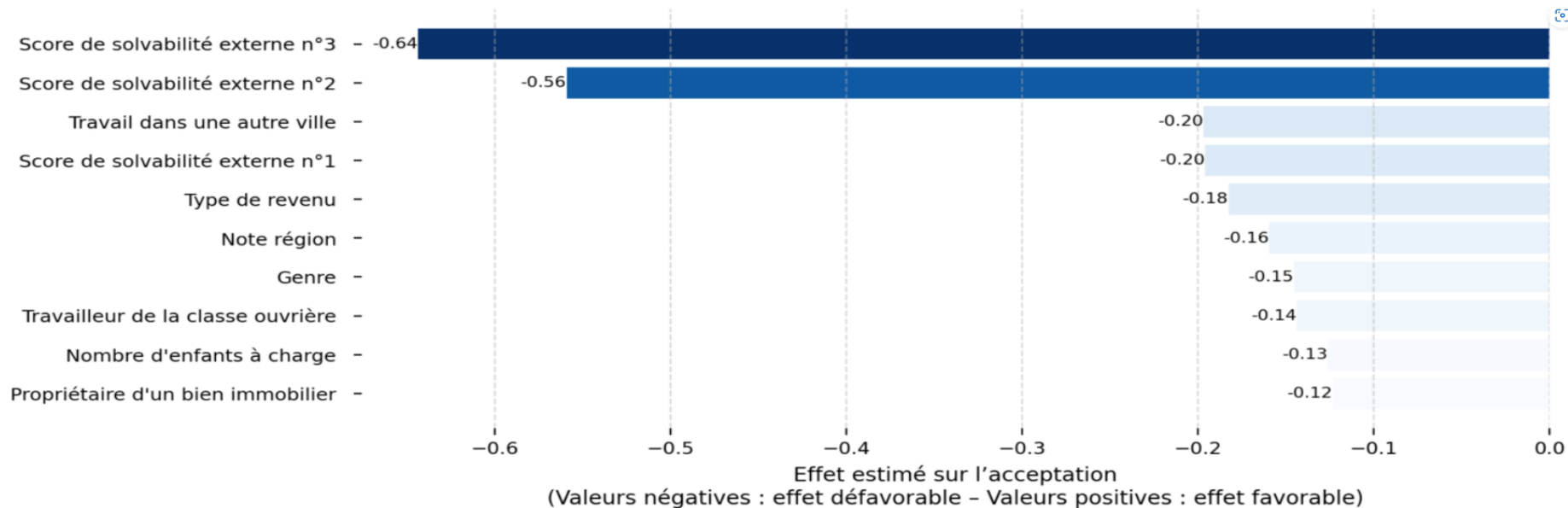
Critères du client et son score de solvabilité

Comparaison du client avec population étudiée

Influence du profil client sur la décision finale de prêt

Influence des critères du client sur la décision d'octroi de crédit

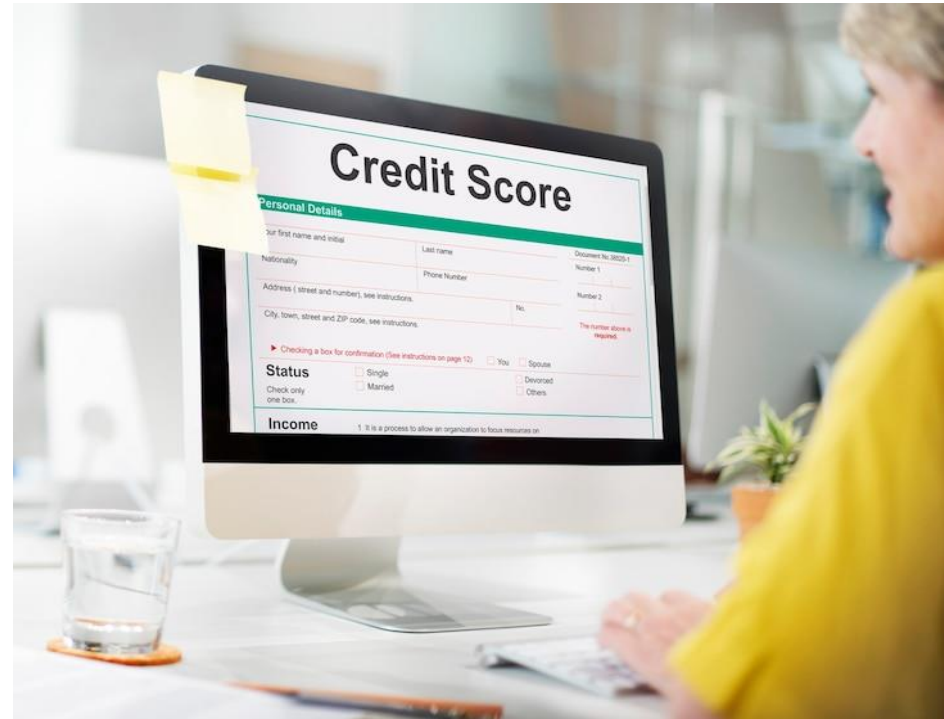
### Les critères ayant influencé la décision

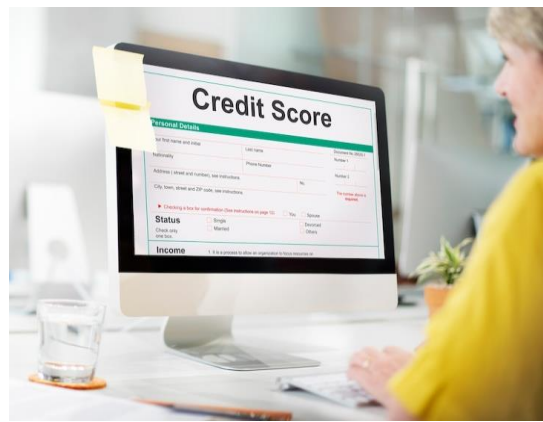


**i Interprétation :** Ce graphique montre les éléments qui ont le plus influencé la décision pour ce client. Même si certains indicateurs paraissent bons (comme une bonne solvabilité), d'autres critères peuvent venir contrebalancer cette impression.

Les barres vers la gauche indiquent ce qui a réduit les chances d'acceptation, et celles vers la droite ce qui les a renforcées. C'est l'ensemble du profil qui compte dans la décision finale.

# DEMONSTRATION DE L'APPLICATION DE SCORING CREDIT





# APPROCHE TECHNIQUE

# CONTEXTE DE LA VEILLE TECHNIQUE

## CONTEXTE

Site internet de vente d'articles  
(biens de consommations)  
en ligne



### Place de Marché

Transaction

Consumer to  
Consumer (C2C)

Vendeurs

Acheteurs



Les vendeurs proposent des articles  
à des acheteurs



## PROCESSUS DE VENTE

1

Le vendeur publie une photo de  
l'article ( données visuelles )

&

Le vendeur fournit une  
description de l'article  
( données textuelles )

2

Le vendeur attribue la catégorie du  
produit à vendre  
(Sélection d'une catégorie à l'aide  
d'une « drop down list »)

3

La catégorie attribuée par le  
vendeur est souvent peu fiable

## POSITION



DATA SCIENTIST

### MISSION

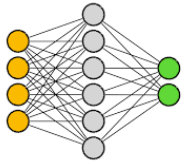


Attribuer une catégorie  
pour chaque image grâce à  
la mise en place  
d'algorithme de  
classification par l'image

# PRESENTATION DES ALGORITHMES CNN et ViT



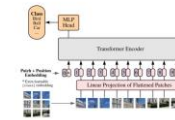
Yann LeCun  
1993



**CNN**

L'objectif de la veille technique présente est de comparer les performances des deux approches de classification d'image (Computer vision).

Vision Transformer - ViT



Google  
Research 2021

**ViT**

Un algorithme basé sur des filtres de convolution permettant d'**extraire des descripteurs visuels hiérarchiques d'une image.**



## Algorithme CNN classique (VGG16)

- 13 couches convolutionnelle et 3 couches entièrement connectées
- Utilisation des filtres pour l'extraction des motifs visuels
- Nécessite une grande quantité et paramètre de mémoire
- Performe correctement avec des images petites
- Algorithme sensible à l'overfitting si peu de données

Un modèle **qui transforme une image en séquence de patches et utilise des mécanismes d'attention** pour analyser la structure globale de l'image.



## Transformer (ViT)

- Découpage de l'image en patch (token) ex 16x16 traduits en mots
- Utilisation de l'attention pour capter les relations entre les patches des images éloignées
- Meilleure performance sur des grands jeux de données
- Interprétation facile avec des architectures pré entraînés
- Interprétable également grâce aux cartes d'attention





## [Collecte des images]

- 1. Identification
- 2. Chargement

## [Prétraitement des images]

- Les étapes 1 et 2 concernent uniquement CNN
- 1. Conversion aux niveaux de gris
  - 2. Filtrage de bruit avec filtre gaussien
  - 3. Augmentation de contraste

## [Extraction des Features]

- 1. CNN: réseau de neurones convolutifs
- 2. ViT : Transformers visuels (fine-tuning)

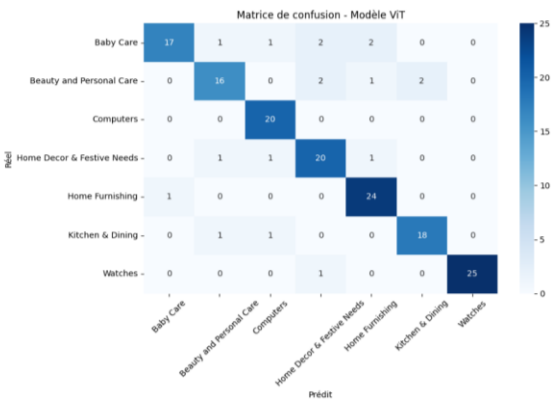
## Evaluation de la performance des modèles

### [Rapport de classification]

Rapport de classification - ViT :

|                            | precision | recall | f1-score | support |
|----------------------------|-----------|--------|----------|---------|
| Baby Care                  | 0.94      | 0.74   | 0.83     | 23      |
| Beauty and Personal Care   | 0.84      | 0.76   | 0.80     | 21      |
| Computers                  | 0.87      | 1.00   | 0.93     | 20      |
| Home Decor & Festive Needs | 0.80      | 0.87   | 0.83     | 23      |
| Home Furnishing            | 0.86      | 0.96   | 0.91     | 25      |
| Kitchen & Dining           | 0.90      | 0.90   | 0.90     | 20      |
| Watches                    | 1.00      | 0.96   | 0.98     | 26      |
| accuracy                   |           |        | 0.89     | 158     |
| macro avg                  | 0.89      | 0.88   | 0.88     | 158     |
| weighted avg               | 0.89      | 0.89   | 0.88     | 158     |

### [Matrice de Confusion]



# Répartition du dataset Train / Validation / Test



Répartition homogène des catégories (TRAIN) - TRAIN contient 787 images :

| Catégorie | label_name                 | Nombre d'images | Nombre total de labels |
|-----------|----------------------------|-----------------|------------------------|
| 1         | Beauty and Personal Care   | 113             | 113                    |
| 5         | Kitchen & Dining           | 113             | 113                    |
| 0         | Baby Care                  | 113             | 113                    |
| 4         | Home Furnishing            | 112             | 112                    |
| 6         | Watches                    | 112             | 112                    |
| 2         | Computers                  | 112             | 112                    |
| 3         | Home Decor & Festive Needs | 112             | 112                    |

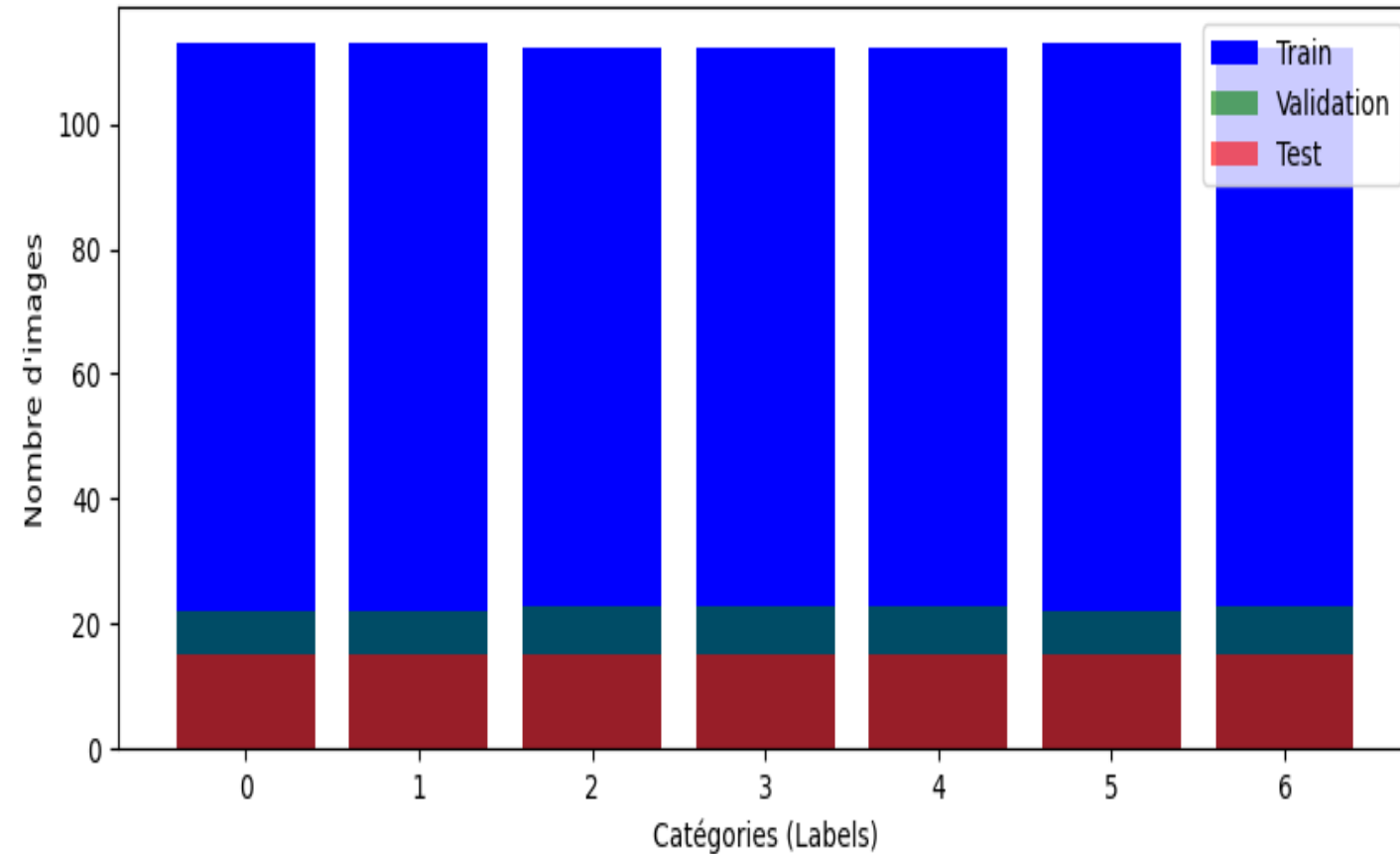
Répartition homogène des catégories (VAL) - VAL contient 158 images :

| Catégorie | label_name                 | Nombre d'images | Nombre total de labels |
|-----------|----------------------------|-----------------|------------------------|
| 3         | Home Decor & Festive Needs | 23              | 23                     |
| 6         | Watches                    | 23              | 23                     |
| 2         | Computers                  | 23              | 23                     |
| 4         | Home Furnishing            | 23              | 23                     |
| 5         | Kitchen & Dining           | 22              | 22                     |
| 1         | Beauty and Personal Care   | 22              | 22                     |
| 0         | Baby Care                  | 22              | 22                     |

Répartition homogène des catégories (TEST) - TEST contient 105 images :

| Catégorie | label_name                 | Nombre d'images | Nombre total de labels |
|-----------|----------------------------|-----------------|------------------------|
| 1         | Beauty and Personal Care   | 15              | 15                     |
| 0         | Baby Care                  | 15              | 15                     |
| 6         | Watches                    | 15              | 15                     |
| 5         | Kitchen & Dining           | 15              | 15                     |
| 3         | Home Decor & Festive Needs | 15              | 15                     |
| 4         | Home Furnishing            | 15              | 15                     |
| 2         | Computers                  | 15              | 15                     |

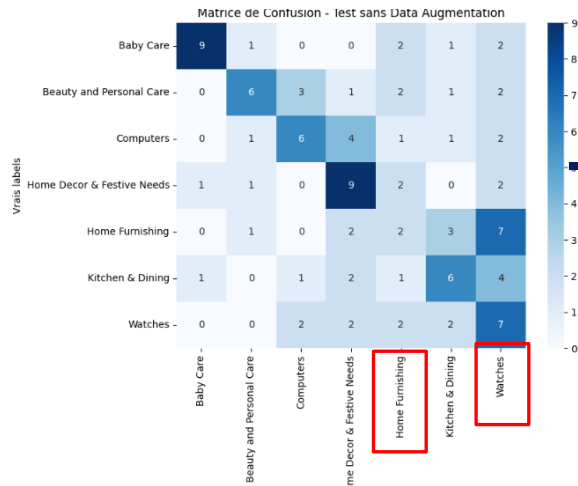
Distribution des labels dans Train 75% / Validation 15% / Test 10%



# Synthèse comparative des résultats



## CNN



Faible distinction entre les classes : beaucoup de confusions, notamment entre "Home Furnishing" et "Watches".

## ViT Visual Transformer



Prédictions nettes et précises : diagonale dominante, erreurs rares et bien réparties.

Rapport de classification - Test :

|                            | precision | recall | f1-score | support |
|----------------------------|-----------|--------|----------|---------|
| Baby Care                  | 0.82      | 0.60   | 0.69     | 15      |
| Beauty and Personal Care   | 0.60      | 0.40   | 0.48     | 15      |
| Computers                  | 0.50      | 0.40   | 0.44     | 15      |
| Home Decor & Festive Needs | 0.45      | 0.60   | 0.51     | 15      |
| Home Furnishing            | 0.17      | 0.13   | 0.15     | 15      |
| Kitchen & Dining           | 0.43      | 0.40   | 0.41     | 15      |
| Watches                    | 0.27      | 0.47   | 0.34     | 15      |
| accuracy                   |           |        | 0.43     | 105     |
| macro avg                  | 0.46      | 0.43   | 0.43     | 105     |
| weighted avg               | 0.46      | 0.43   | 0.43     | 105     |

F1-score faible et déséquilibré entre les classes. Prédictions peu robustes.

Rapport de classification - ViT :

|                            | precision | recall | f1-score | support |
|----------------------------|-----------|--------|----------|---------|
| Baby Care                  | 0.94      | 0.74   | 0.83     | 23      |
| Beauty and Personal Care   | 0.84      | 0.76   | 0.80     | 21      |
| Computers                  | 0.87      | 1.00   | 0.93     | 20      |
| Home Decor & Festive Needs | 0.80      | 0.87   | 0.83     | 23      |
| Home Furnishing            | 0.86      | 0.96   | 0.91     | 25      |
| Kitchen & Dining           | 0.90      | 0.90   | 0.90     | 20      |
| Watches                    | 1.00      | 0.96   | 0.98     | 26      |
| accuracy                   |           |        | 0.89     | 158     |
| macro avg                  | 0.89      | 0.88   | 0.88     | 158     |
| weighted avg               | 0.89      | 0.89   | 0.88     | 158     |

Excellente performance globale. Toutes les classes atteignent un f1-score  $\geq 0.80$

# CONCLUSION

---



## Le dashboard de Scoring Crédit :

1. Le dashboard qui à été développé et mis en production est accessible à toute personne disposant de son lien.
2. Il permettra aux chargés de relation client d'appuyer leur décision d'octroi de crédit en s'appuyant sur une application qui est fondée sur une population finement étudiée, grâce à la mise en place du modèle prédictif développé précédemment
3. Le dashboard permet enfin de mieux gérer les risques de perte en capital pour l'entreprise Prêt à dépenser

---

## La note de veille technique :

1. La veille technique a permis de démontrer l'intérêt des architectures récentes de type Vision Transformer (ViT) pour la classification d'images.
2. En comparaison avec une approche CNN classique, ViT offre des performances nettement supérieures et une meilleure interprétabilité grâce aux mécanismes d'attention.
3. Cette exploration confirme l'importance d'intégrer des modèles d'état de l'art dans les projets de data science pour gagner en efficacité et transparence.

# Prêt **à** dépendre



*Merci de votre attention*  
–  
*Question(s) – Réponse(s)*