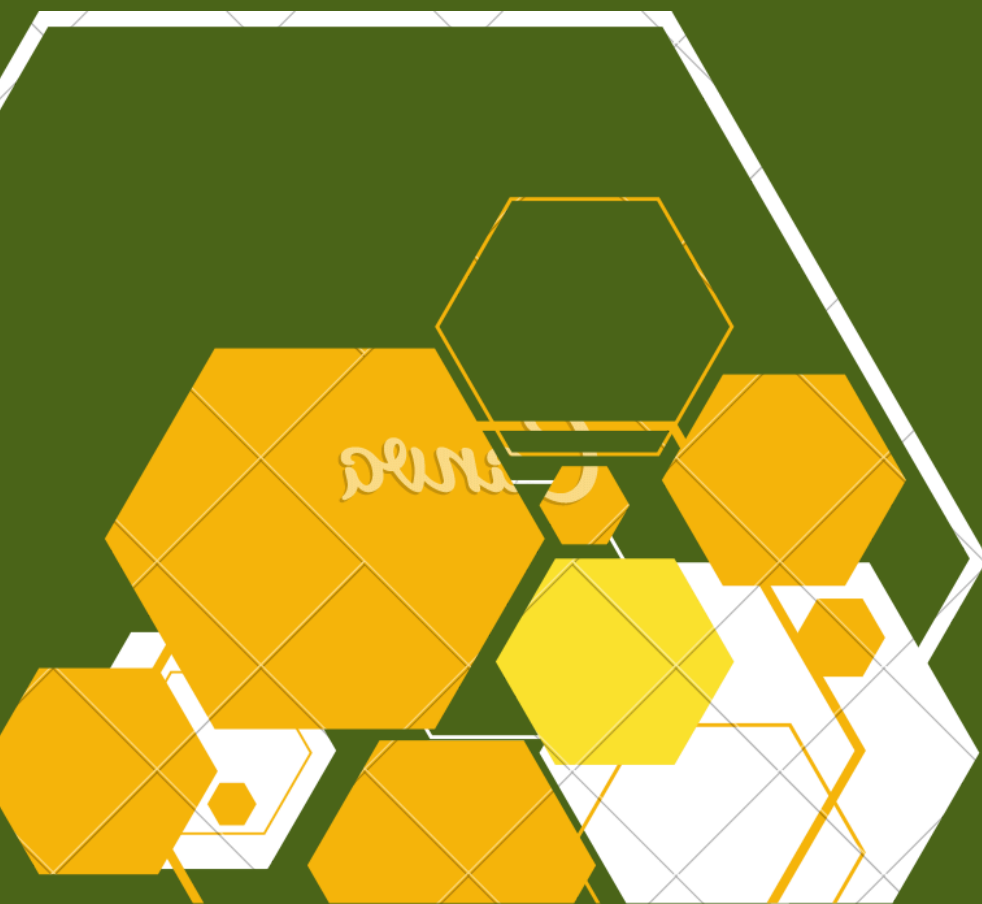


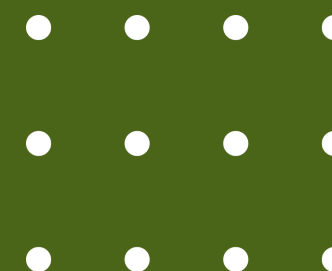
Prêt à dépenser

Réalisez un dashboard et
assurez une veille technique

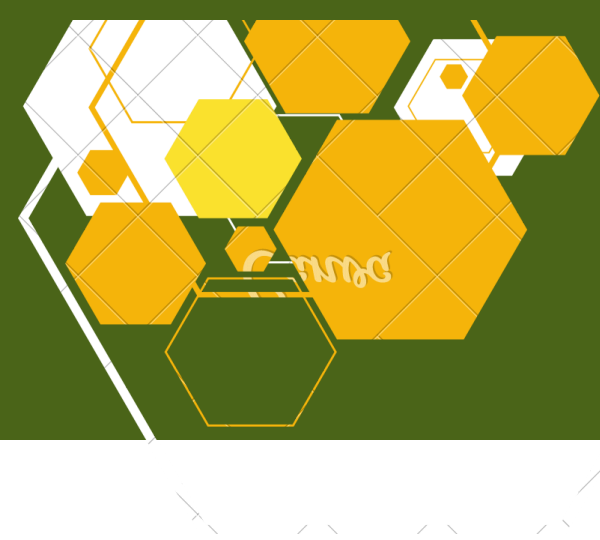


https://github.com/tkappe/Scoring_dashboard.git

Présenté par Thierry KAPPE



Plan



01 Contexte et problématique

02 Présentation du dashboard

01

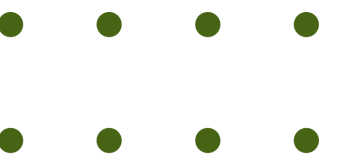
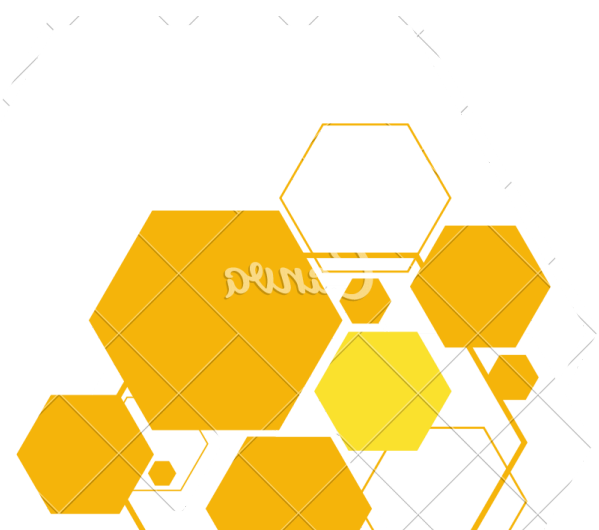
Contexte et problématique

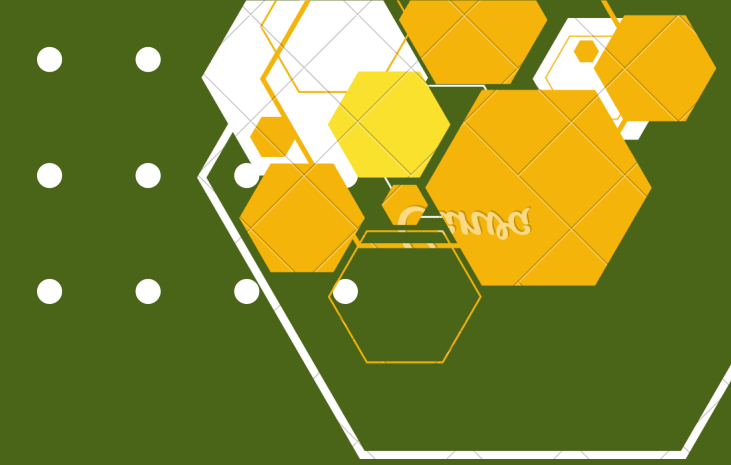
02

Présentation du dashboard

03

Présentation de la veille technique





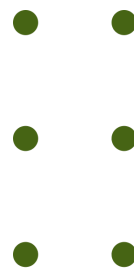
La société financière Prêt à dépenser propose des crédits à la consommation pour des personnes ayant peu ou pas du tout d'historique de prêt.

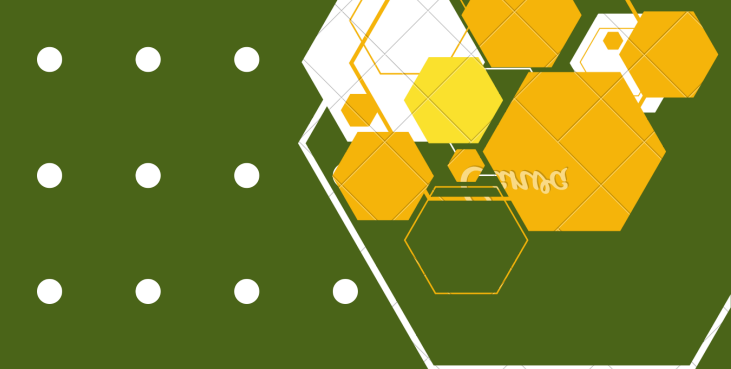
L'entreprise a mis en place un outil de scoring crédit qui calcule la probabilité qu'un client rembourse son crédit, puis classifie la demande en crédit accordé ou refusé.

Il est maintenant question **d'élaborer un dashboard interactif qui aidera les chargés de relation client à expliquer avec la plus grande transparence possible aux clients les décisions d'octroi de crédit.**

Le dashboard devrait permettre de :

- Visualiser du score, de la probabilité et l'interprétation de ce score pour chaque client,
- Visualiser les principales informations descriptives du client,
- Obtenir des comparaisons des principales informations du client avec d'autres clients,
- Obtenir un score et probabilités rafraichis suite à la modification d'informations,
- Un déploiement sur le cloud pour une accessibilité totale.



**Ecran d'accueil**

Lien pour la démo : [ici](#)

**Menu de navigation**

Aller à

- ☒ Accueil
- ☐ Résultats
- ☐ Compréhension
- ☐ Analyses Univariées
- ☐ Analyses Bivariées
- ☐ Projections

Sélectionnez l'ID du client

208550



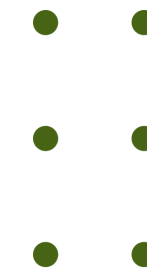
Soumettre

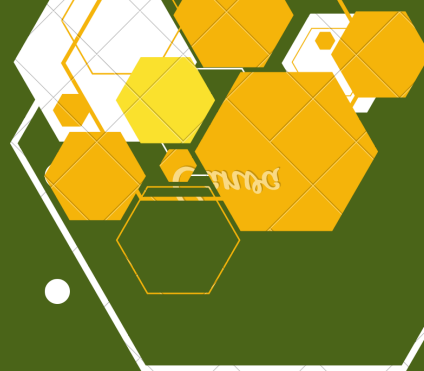
Dashboard de Prédiction de Scoring

Bienvenue sur le dashboard de prédiction de scoring.

Ce tableau de bord vous permet de comprendre et d'analyser les risques de défaut de crédit pour les clients. Vous pouvez explorer les résultats de la prédiction, comprendre les facteurs qui influencent les décisions, et effectuer des analyses univariées et bivariées.

L'objectif est de fournir une explication transparente des prédictions et d'explorer les impacts possibles de la modification de certaines caractéristiques. Améliorer sans cesse notre service est une priorité absolue !





Ecran Résultats

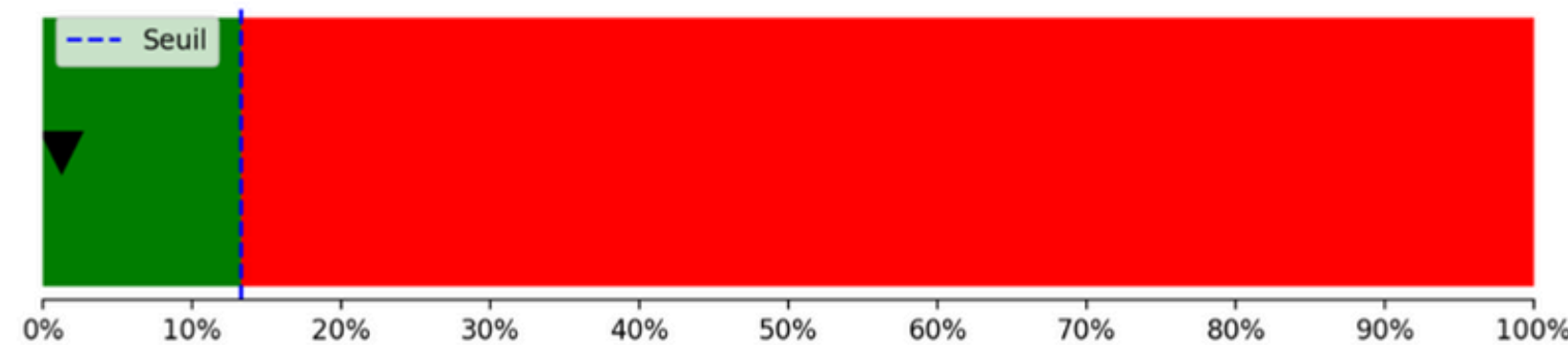
Prédiction du risque de défaut pour le client 208550

Résultat de la demande

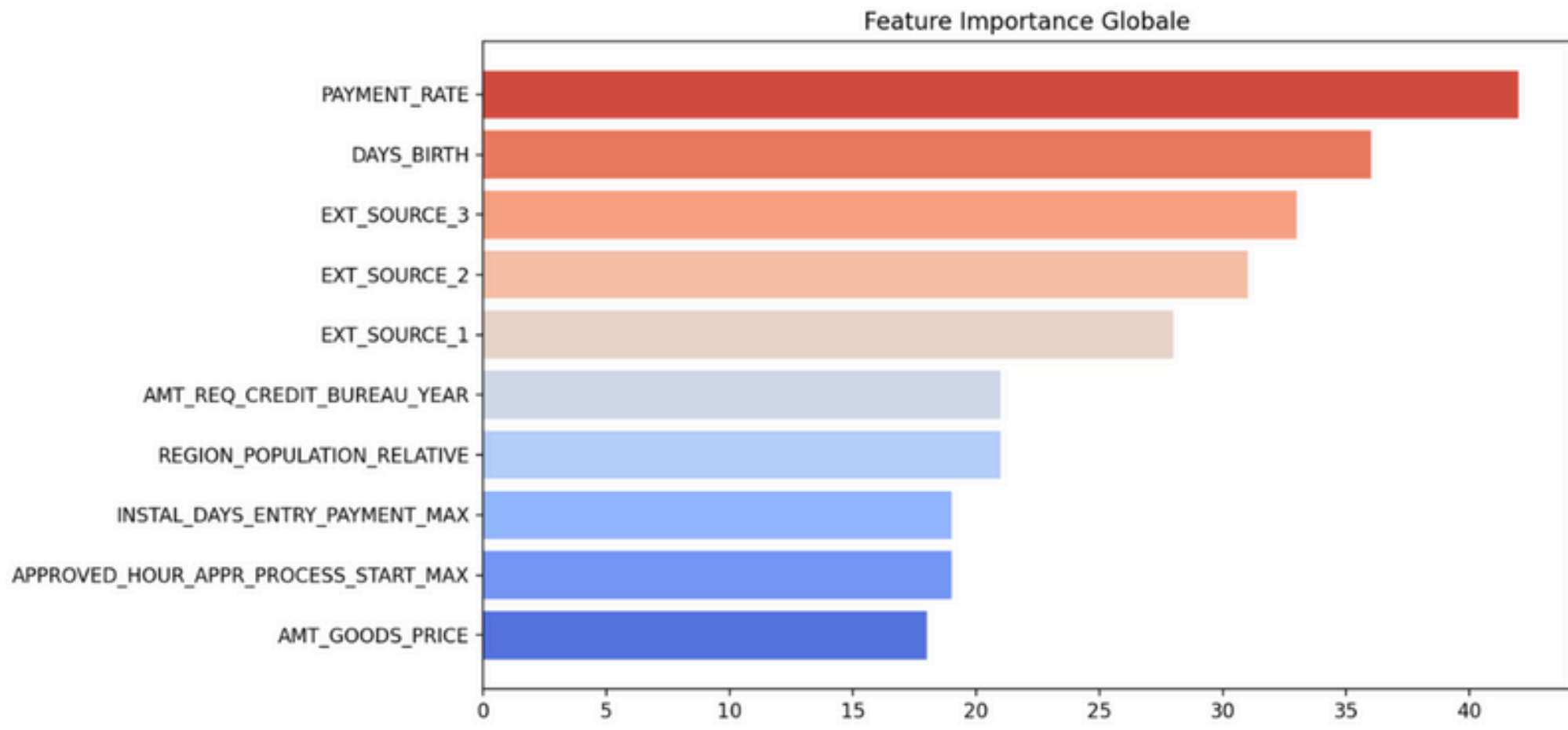
Probabilité de défaut : 1.20%

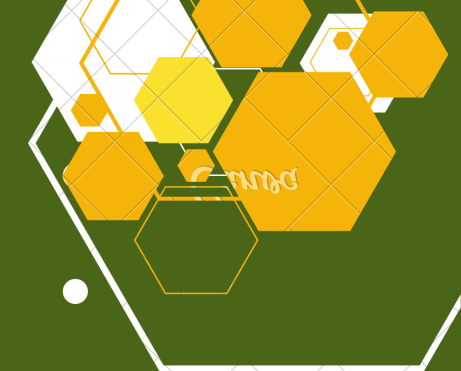
Décision : Prêt accepté

Positionnement du client par rapport au seuil



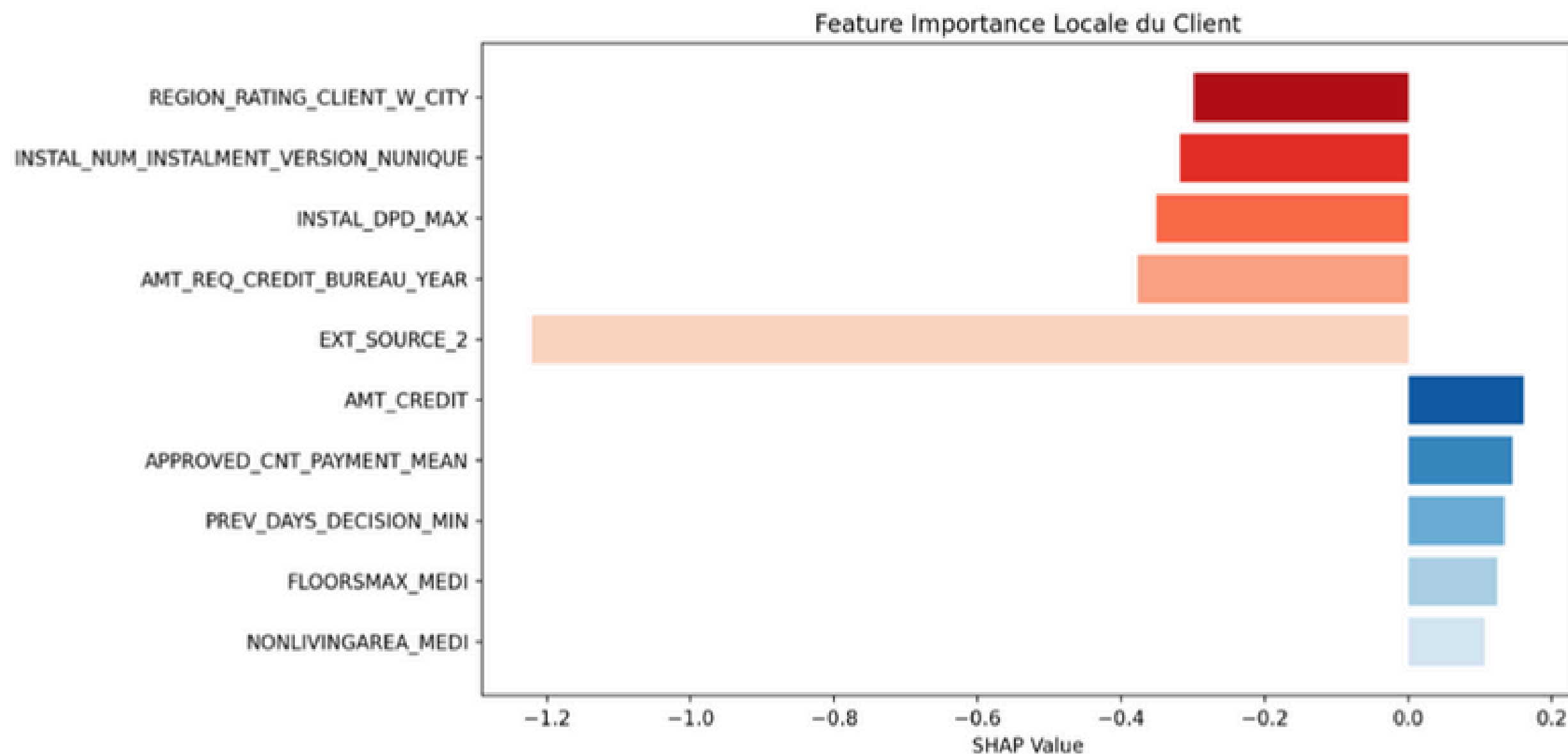
Importance Globale des Features

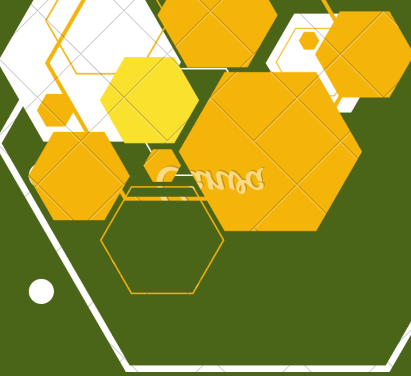





Ecran Compréhension de la prédiction sur 10 features clés

Compréhension de la Prédiction pour le Client 208550





Analyses univariées et position du client



Menu de navigation

Aller à

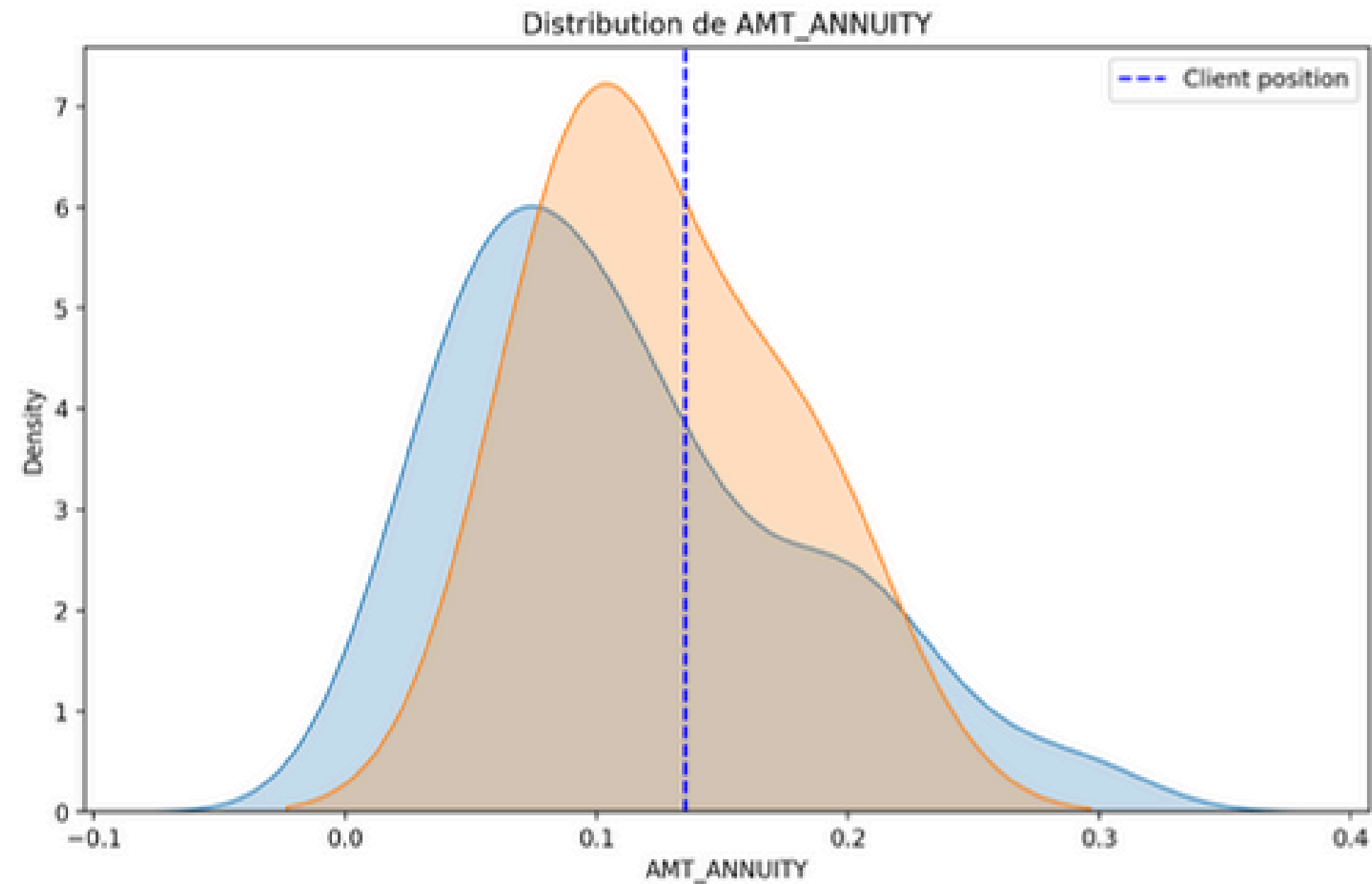
- ☐ Accueil
- ☐ Résultats
- ☐ Compréhension
- ☒ Analyses Univariées
- ☐ Analyses Bivariées
- ☐ Projections

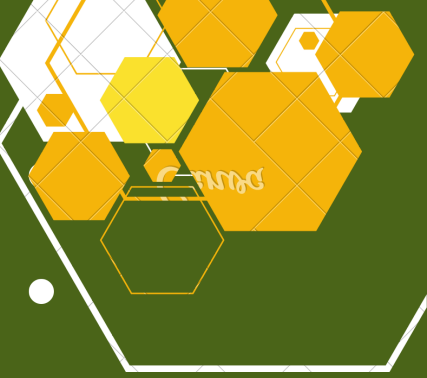
Sélectionnez l'ID du client

Soumettre

Sélectionnez une feature

AMT_ANNUIITY





Analyses bivariées et position du client

Prêt à dépenser

Menu de navigation

Aller à

- ☐ Accueil
- ☐ Résultats
- ☐ Compréhension
- ☐ Analyses Univariées
- ☒ Analyses Bivariées
- ☐ Projections

Sélectionnez l'ID du client

208550

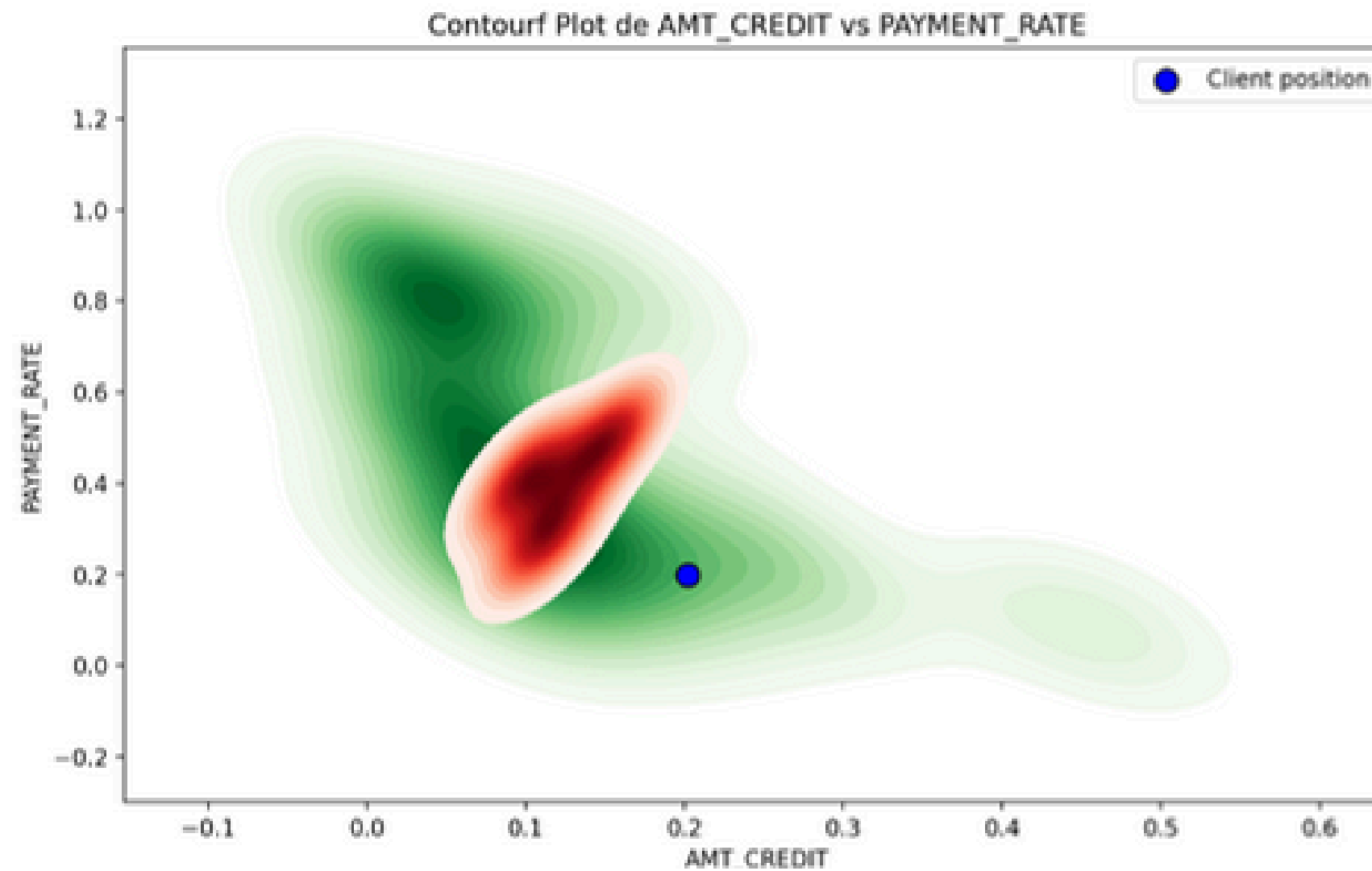
Soumettre

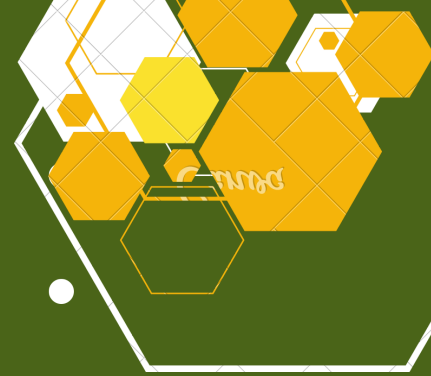
Sélectionnez la première fonctionnalité

AMT_CREDIT

Sélectionnez la deuxième feature

PAYMENT_RATE





Projections et actualisations



Menu de navigation

Aller à

- ☐ Accueil
- ☐ Résultats
- ☐ Compréhension
- ☐ Analyses Univariées
- ☐ Analyses Bivariées
- ☒ Projections

Sélectionnez l'ID du client

208550



Soumettre

Modifier les valeurs des features les plus impactantes négativement :

EXT_SOURCE_2

0,74

- +

AMT_REQ_CREDIT_BUREAU_YEAR

0,07

- +

INSTAL_DPD_MAX

0,00

- +

INSTAL_NUM_INSTALLMENT_VERSION_UNIQUE

0,02

- +

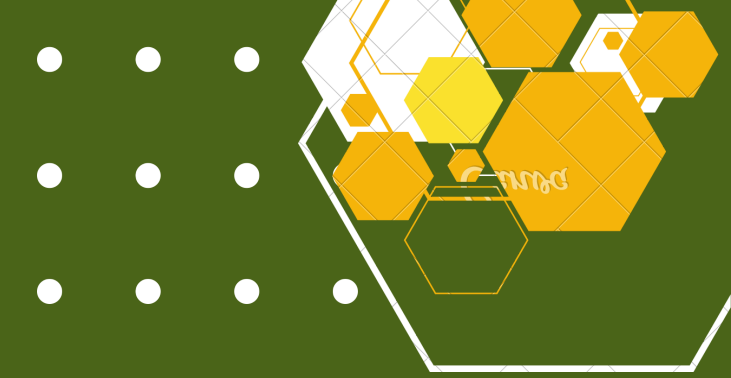
REGION_RATING_CLIENT_W_CITY

0,50

- +

Nouvelle probabilité de défaut : 23.55%

Nouvelle décision : Prêt rejeté



Limites et améliorations

- *Rendre disponible non seulement le code des variables tels que fournis à présent dans le dashboard mais également leurs descriptions littérales pour le chargé de relation client, ceci rendra les échanges bien plus fluides.*
- *Les analyses comparatives telles que fournies actuellement dans le dashboard comparent la position du client concerné par la prédiction analysée avec l'ensemble des prédictions du modèles. Il aurait fallu que cette position soit plutôt comparée avec les réalités terrain c'est à dire avec les labels réels d'un historique de clients.*
- *Pour les analyses bivariées, il aurait été judicieux de faire une distinction des différents variables à expliquer afin d'adapter le graphique d'analyse au type de variables pour un visuel plus adapté.*
- *Il aurait été préférable dans la partie projections du dashboard, d'avoir en visuel les données du client avant transformation par le modèle. Ceci aurait été bien plus simple à expliquer et à comprendre pour un chargé de relation client et bien plus concret pour le client.*

• Présentation des données utilisées

Le jeu de données utilisé ici est celui du projet 6 "Classifiez automatiquement des biens de consommation.

Ce dataset contient un ensemble de 1050 images et descriptifs classés en 7 catégories de 150 images chacune.

Labellisation automatique des objets via une image et une description.



Home Furnishing



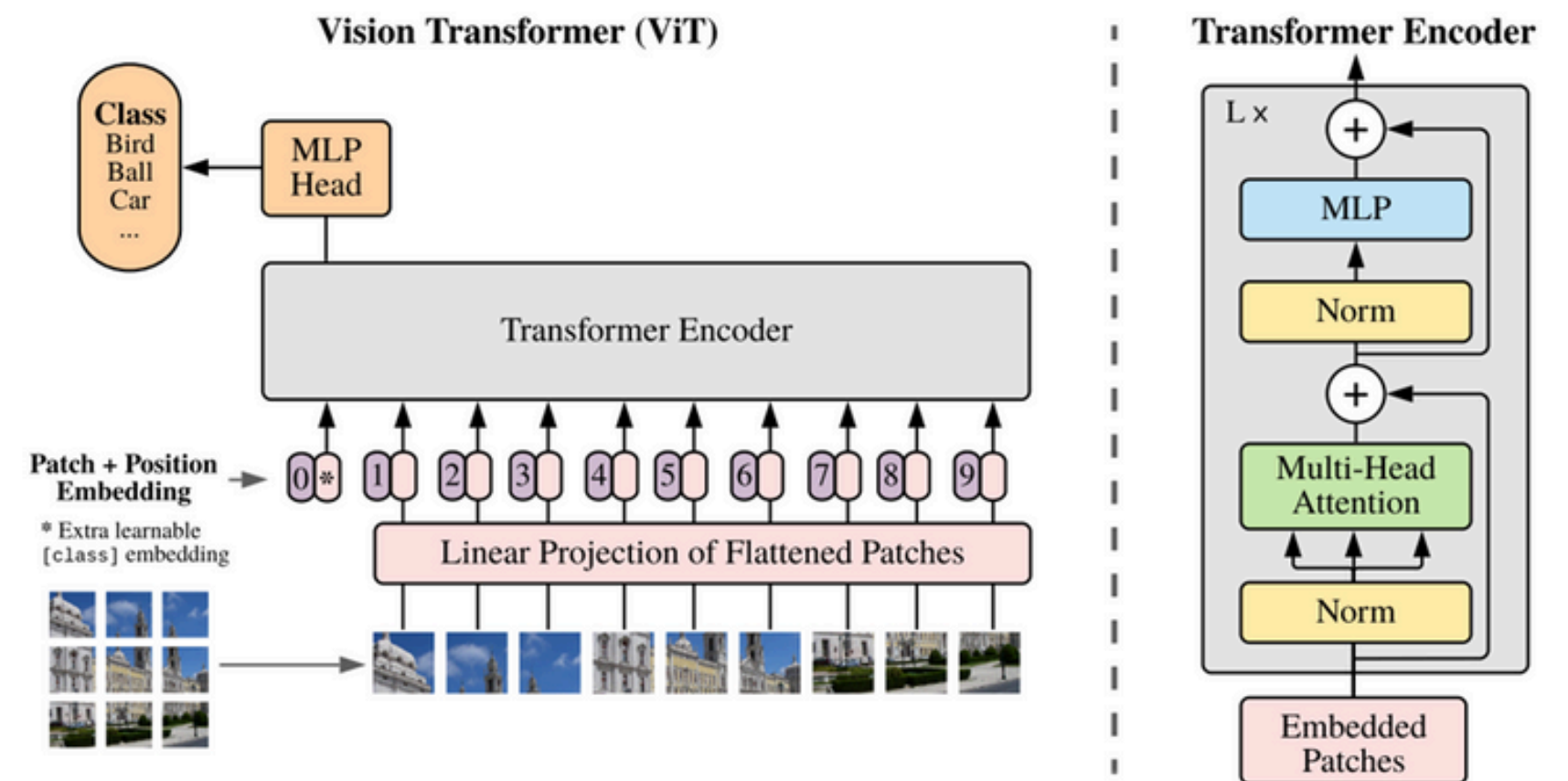
Baby Care

• Le concept de l'algorithme récent

Le jeu de données utilisé ici est celui du projet 6 "Classifiez automatiquement des biens de consommation.

Ce dataset contient un ensemble de 1050 images et descriptifs classés en 7 catégories de 150 images chacune.

Schéma synthétique des traitements

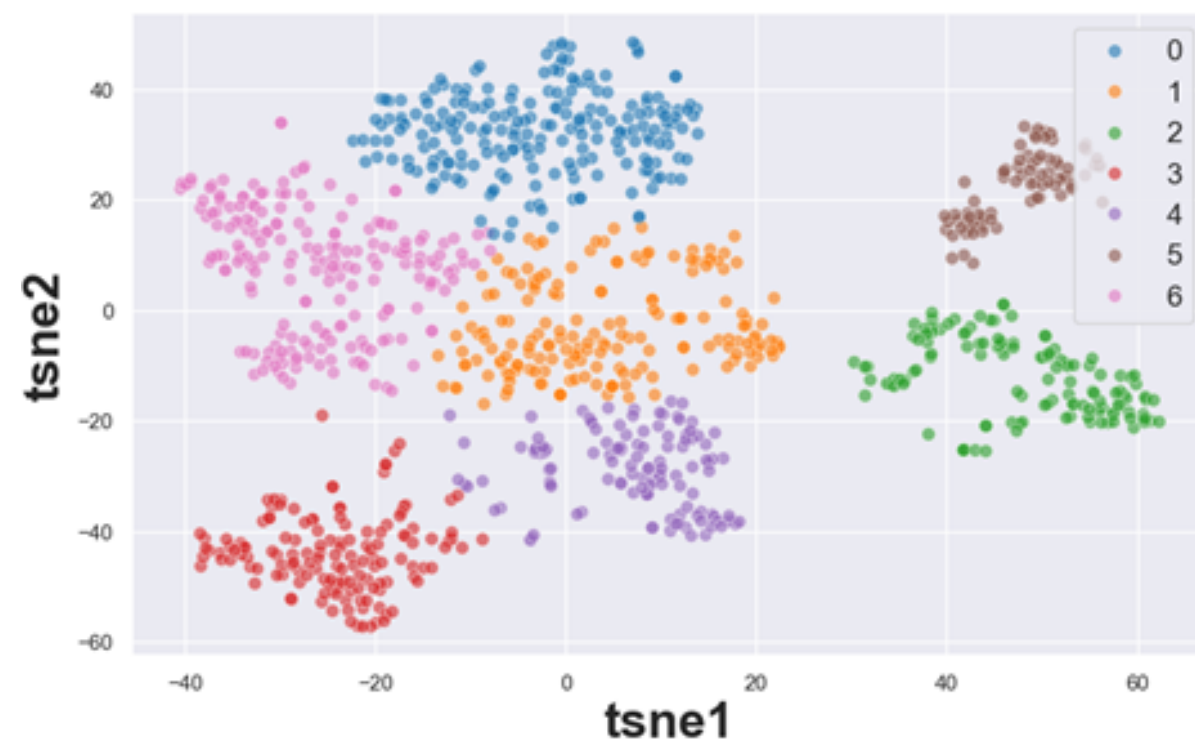


• Synthèse des résultats

	VGG16	ViT
Temps de traitement	8 minutes	13 minutes
Longueur embedding	(1050, 4096)	(1050, 768)
Après réduction de dimension via PCA	(1050, 793)	(1050, 531)
Après réduction de dimension via t-SNE	(1050, 2)	(1050, 2)
Score ARI après clustering	0.46	0.57

• Clusters après extraction VGG16

TSNE selon les clusters

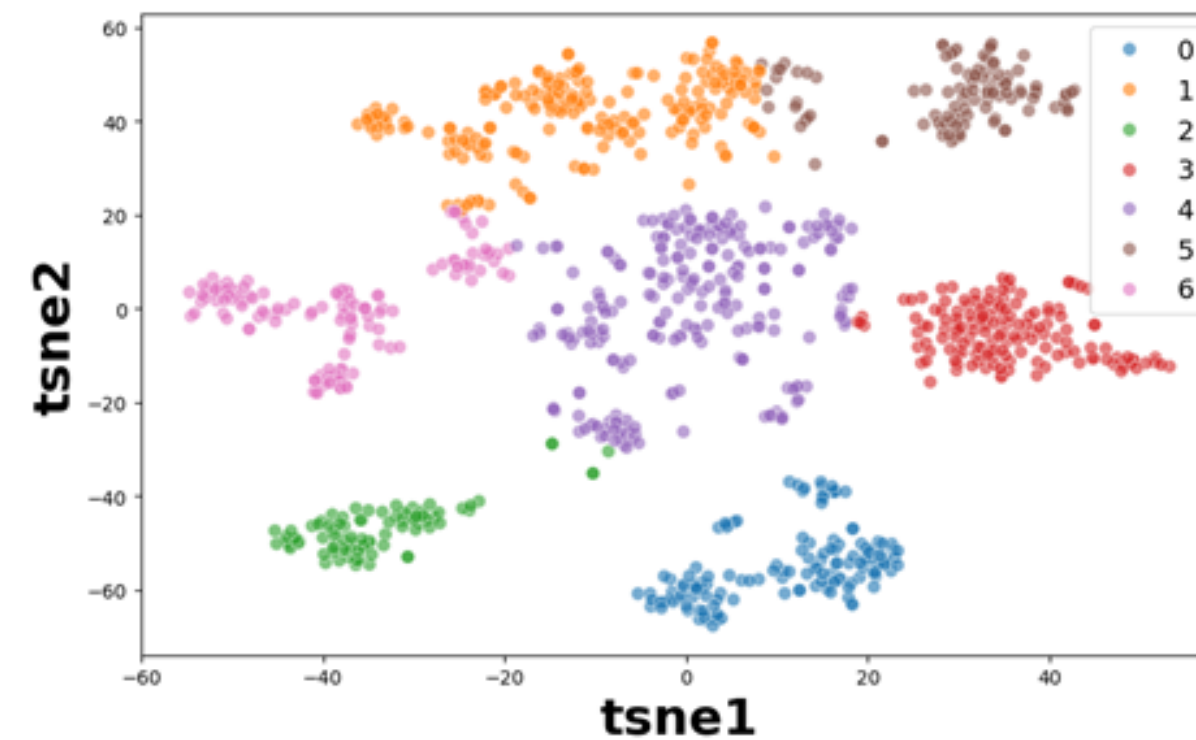


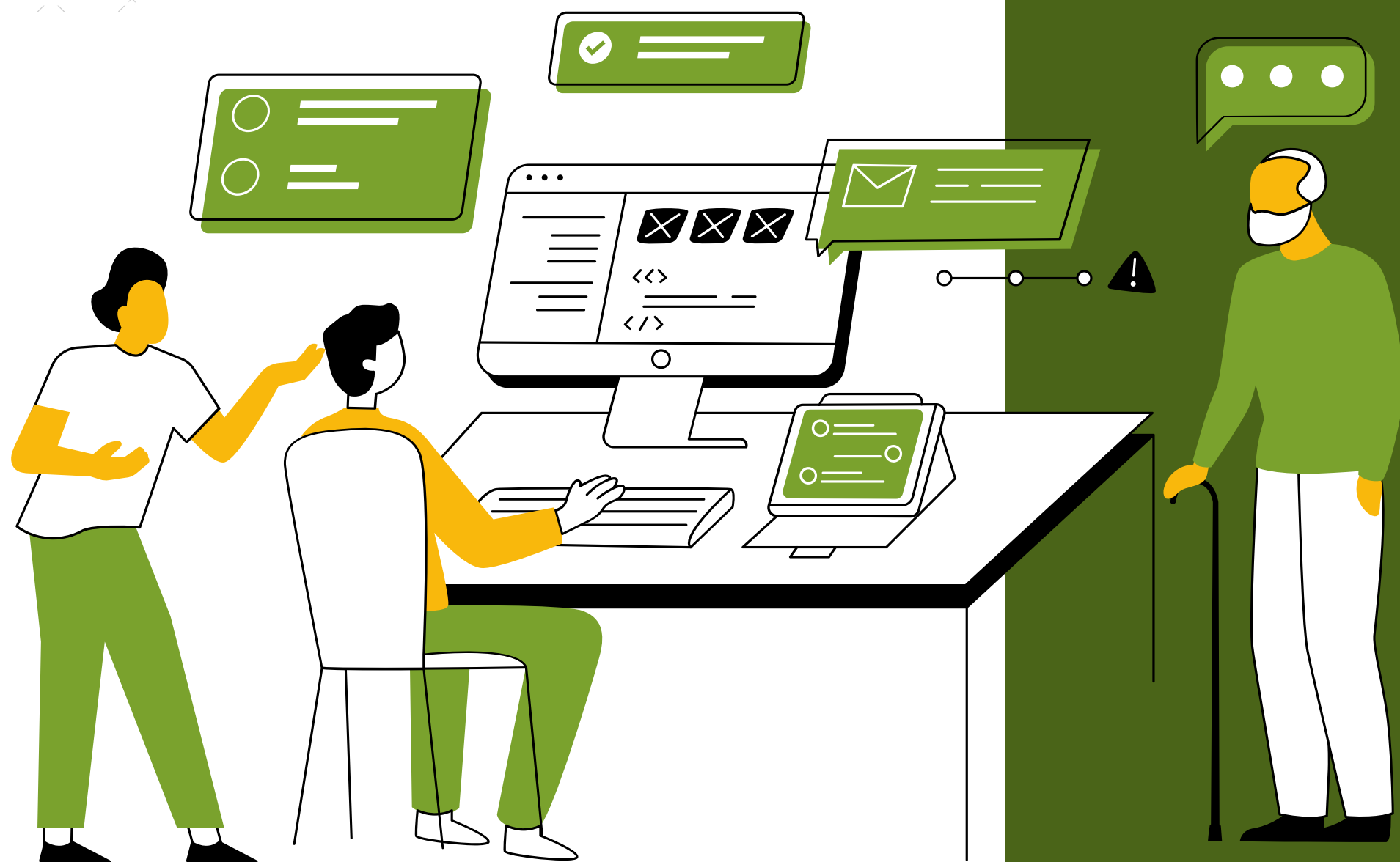
• Conclusion

Avec l'extracteur de ViT, les clusters sont mieux déterminés et mieux séparés comme nous pouvons l'observer sur les visualisations graphiques. Le score ARI de 0.57 nettement supérieur vient bien confirmer cette analyse. Nous pouvons ainsi dire que pour ce jeu de données le modèle ViT est plus adapté que celui de VGG16.

• Clusters après extraction ViT

TSNE selon les clusters





Merci

