

# IMPLÉMENTEZ UN MODÈLE DE SCORING

Projet 7 – Parcours Data Scientist

**OPENCLASSROOMS**

# sommaire

INTRODUCTION

PRÉSENTATION DU JEU DE DONNÉES

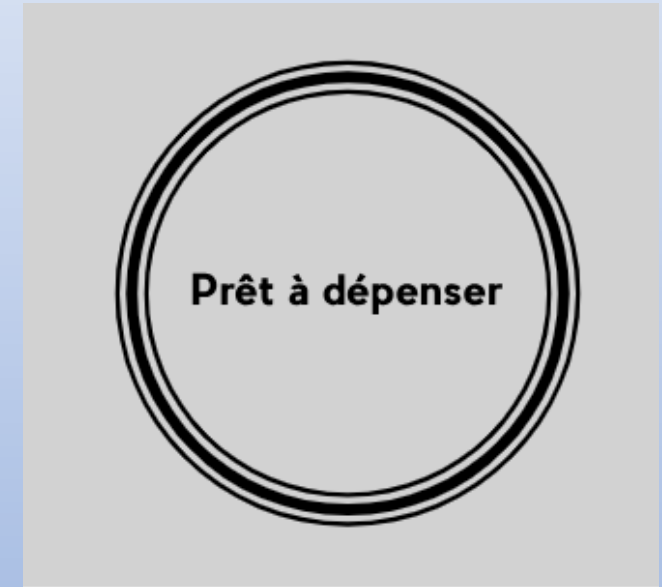
ANALYSE EXPLORATOIRE

MODÉLISATION

DASHBOARD

# INTRODUCTION

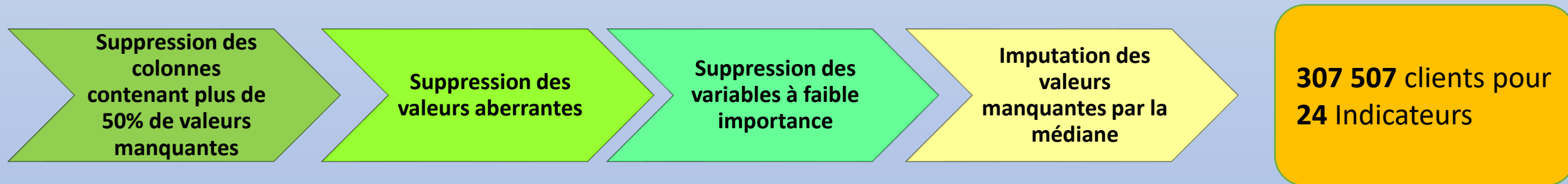
- Société financière 'Prêt à dépenser' propose des crédits à la consommation
- Mettre en œuvre un outil de scoring
- Développer un Dashboard interactif et qui serait transparent pour le client



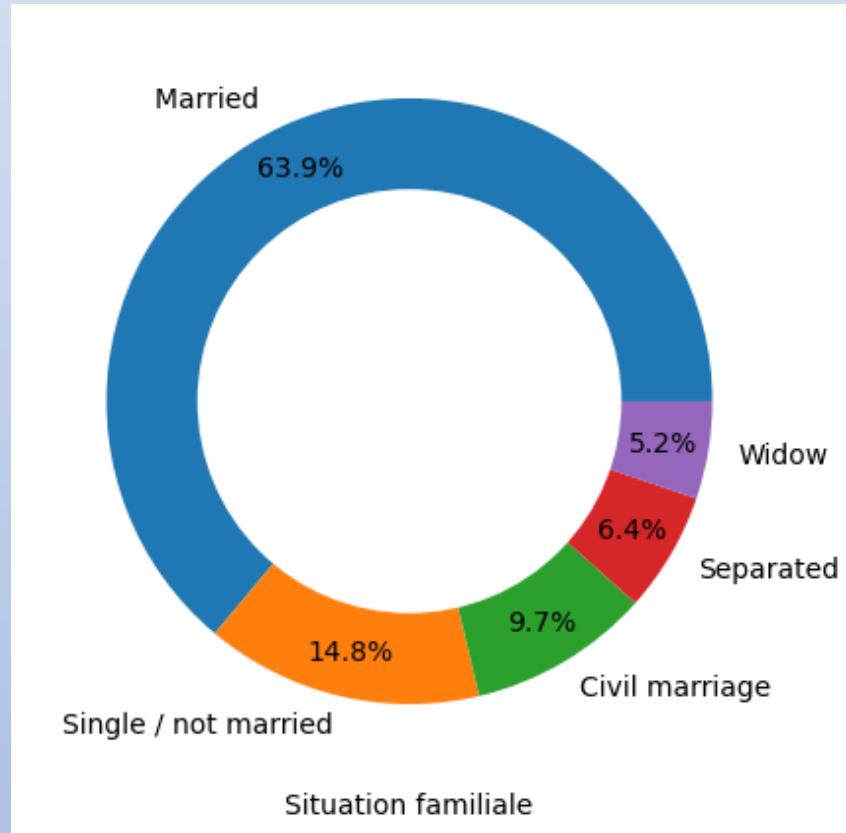
# PRESENTATION DU JEU DE DONNÉES

Les données originales sont téléchargeables sur [kaggle](#)

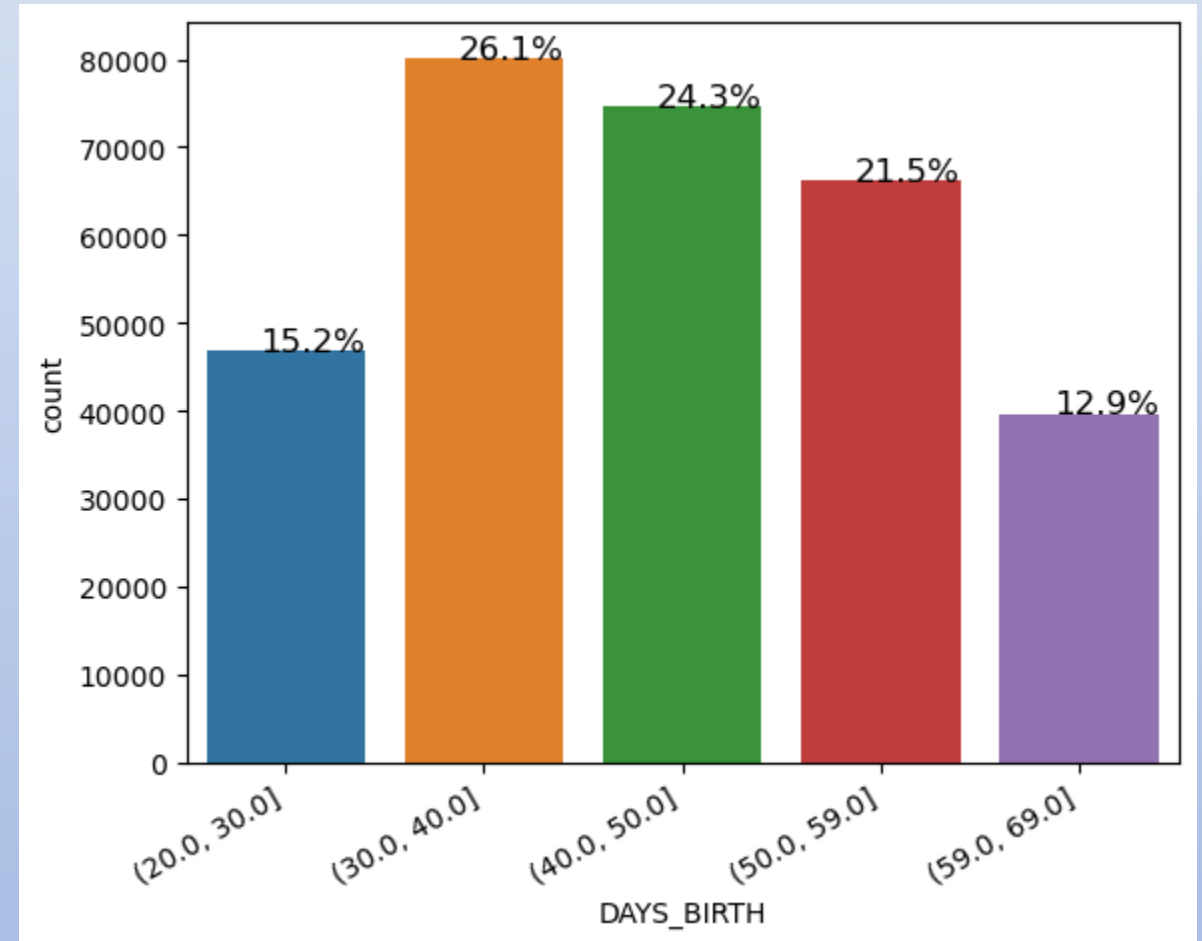
307 511 clients pour 122 Indicateurs



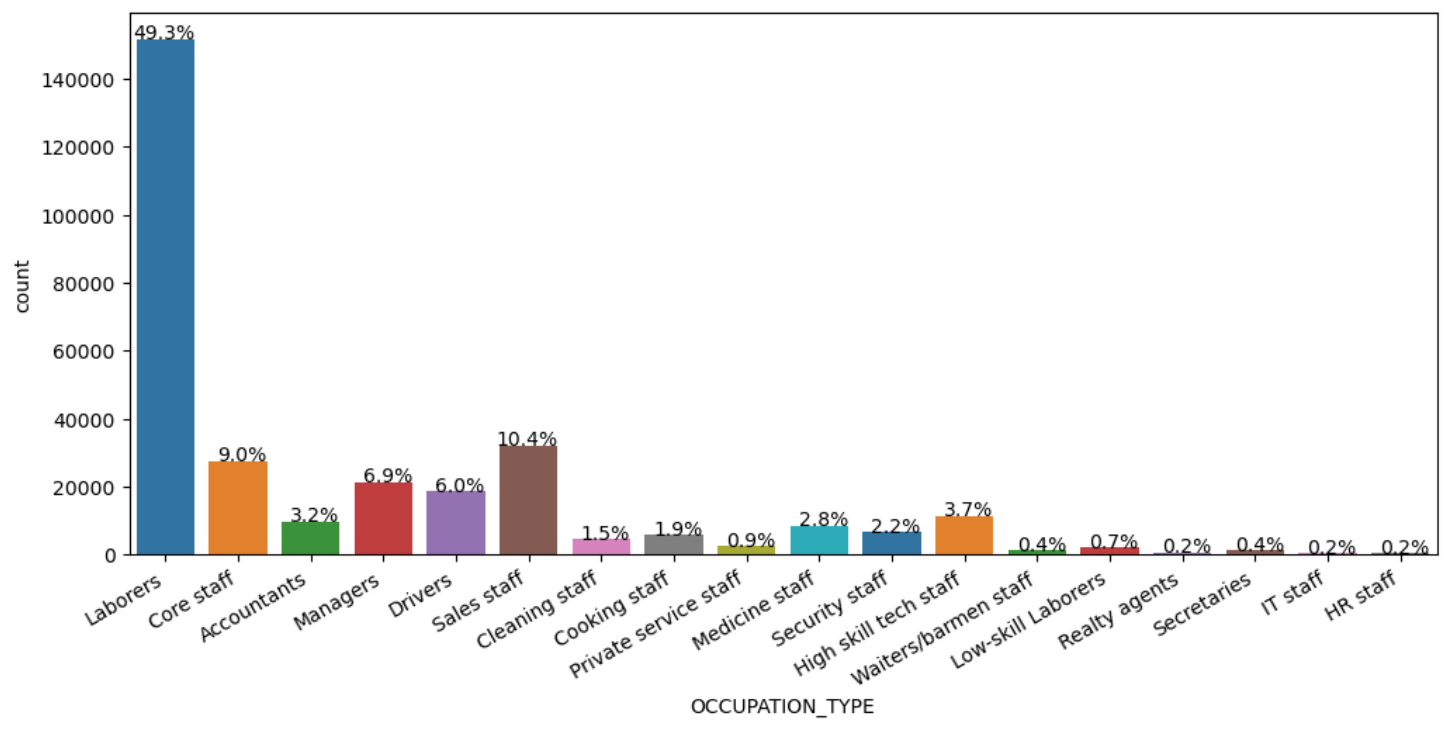
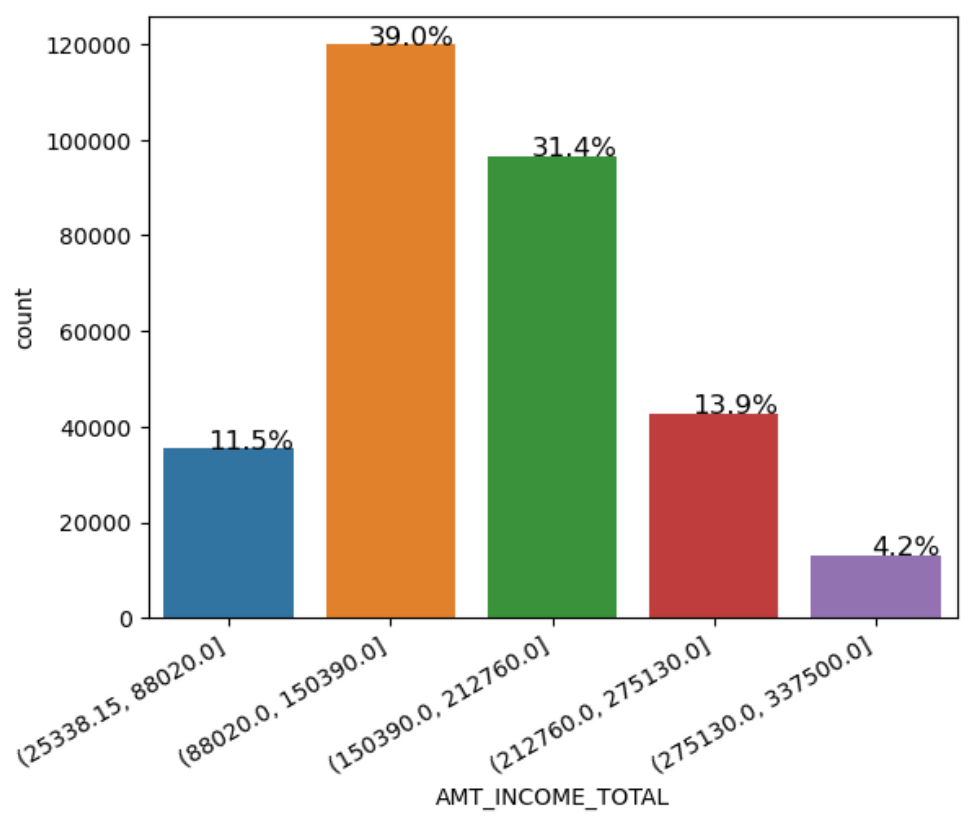
# ANALYSE EXPLORATOIRE



73,5% des clients sont mariés.  
72% des clients ont entre 30 et 59 ans.

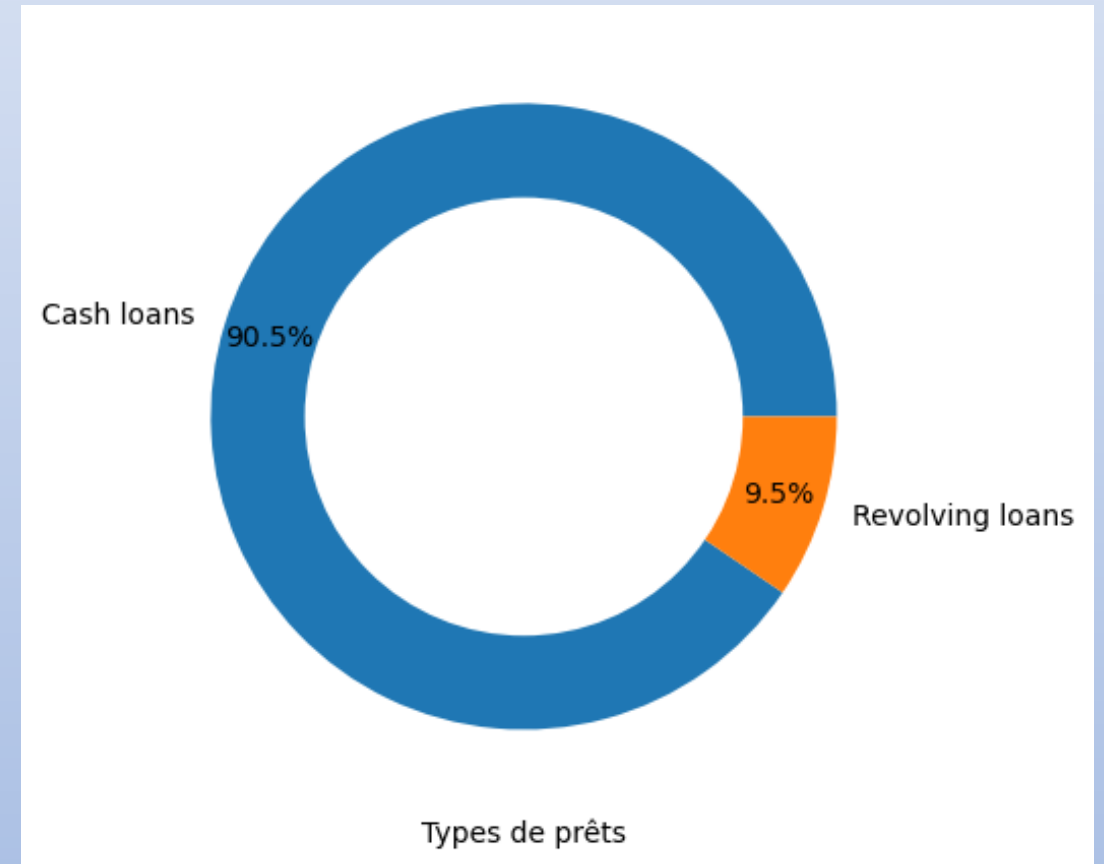


Analyse exploratoire - suite



On peut constater que 49% des clients occupent des emplois d'ouvriers, et que 70% des clients présentent des revenus compris entre : 88000 et 212700

La grande majorité des prêts contractés par les clients sont des prêts de trésorerie, représentant 90,5 % de l'ensemble des prêts.



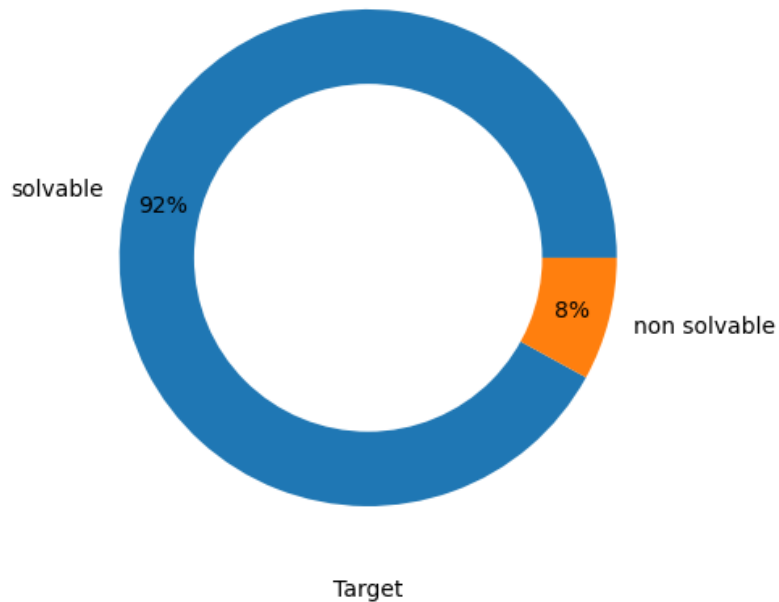


# MODÉLISATION

**Encodage des variables  
catégoriques**

**Standardisation des  
données**

**Séparation du dataset  
en train et test**




Jeu de donnée déséquilibré



- ❖ Client solvable (Target:0) : 92 %
- ❖ Client à risque (Target:1) : 8 %

**Smote:** rééquilibré les  
données


## Sélection du modèle de Machine Learning

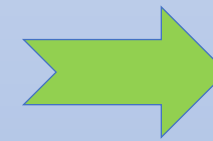
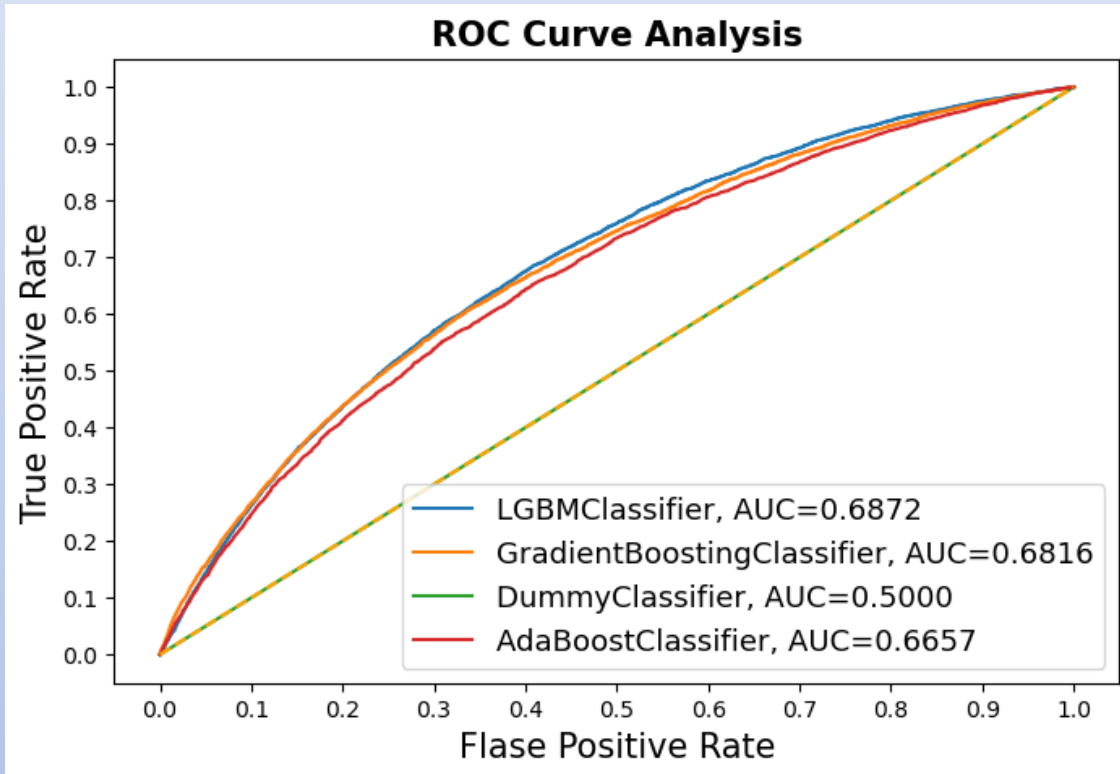


Sélectionner le meilleur  
algorithme adapté à notre  
problématique



Optimiser ses  
hyperparamètres

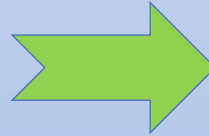
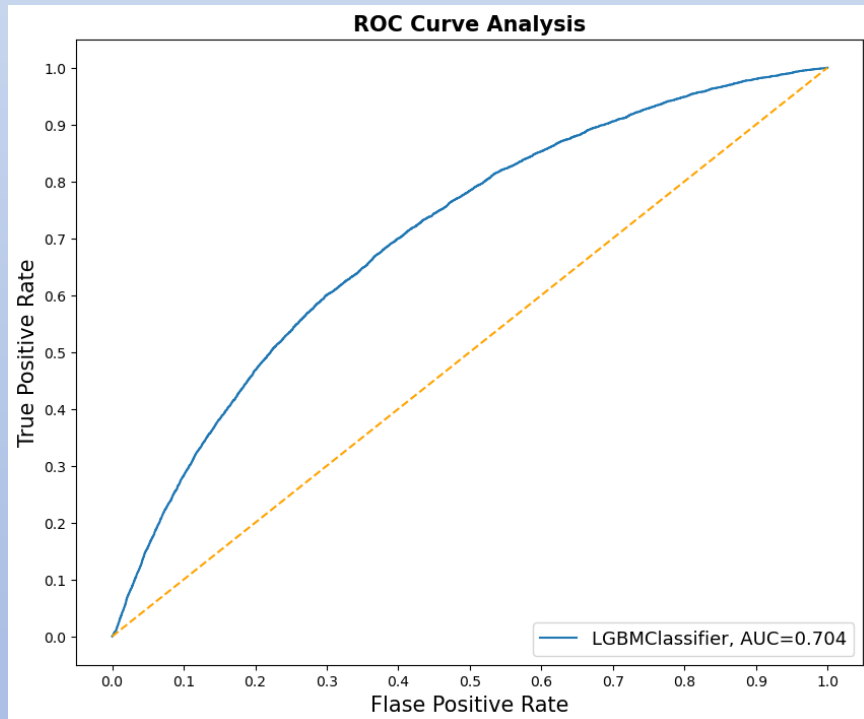




Le meilleur modèle dans  
notre cas est  
**LGBMClassifier**

Algorithme	Accuracy	AUC
Light Gradient Boosting	0,90	0,687
Dummy Classifier	0,92	0,5
AdaBoost Classifier	0,77	0,666
Gradient Boosting Classifier	0,86	0,682

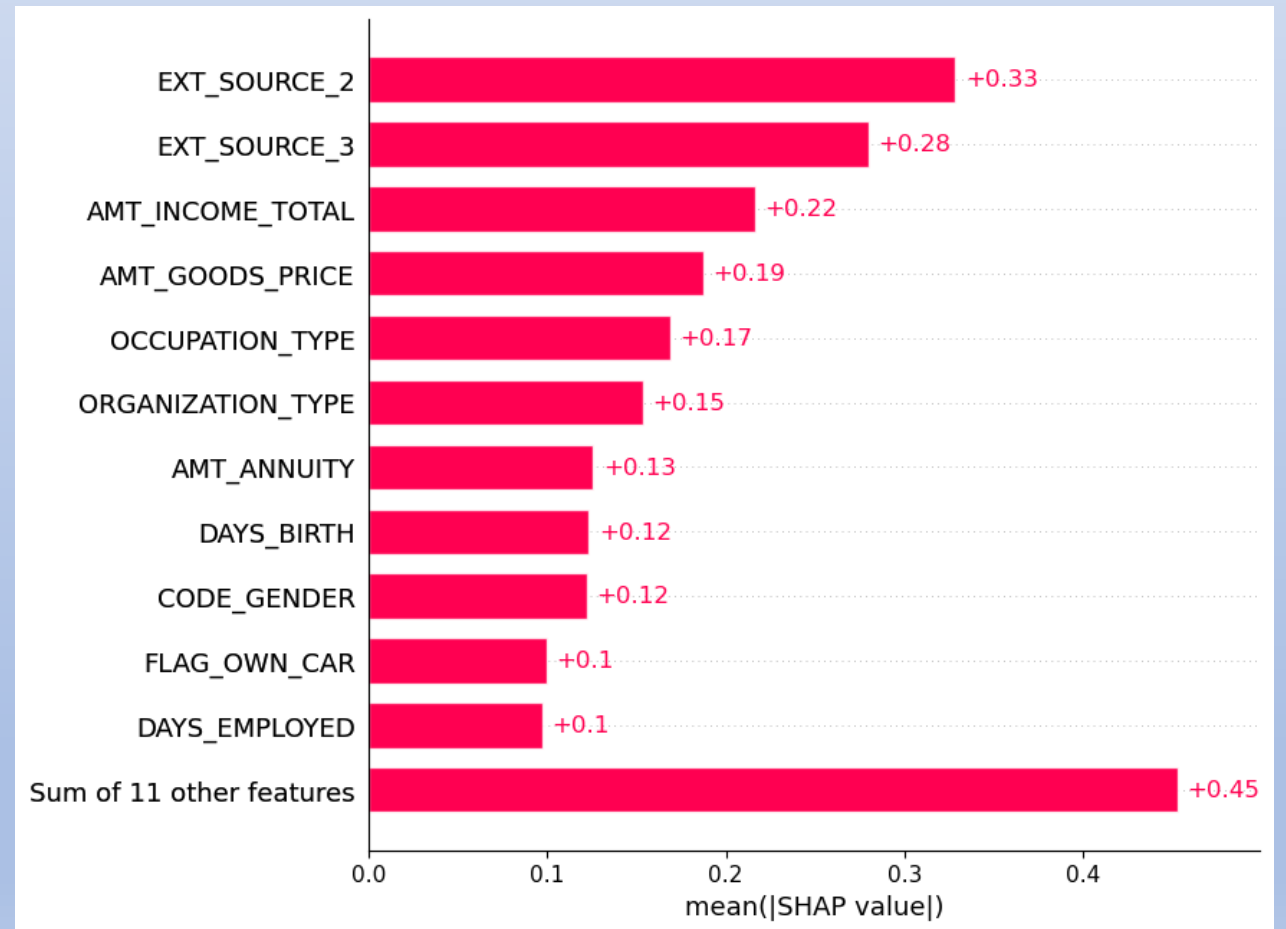
### Performances du modèle après optimisation des hyperparamètres



L'AUC a augmenté de 0,686 à 0,704, tandis que la précision est passée de 0,90 à 0,91.

### Interprétation des résultats du modèle

Les variables sont classées par ordre d'influence sur la prédiction du modèle. Pour cet exemple, « Ext\_source\_2 » est la caractéristique la plus importante, suivie de " Ext\_source\_3 ", " Amt\_Income\_Total ", " Amt\_Good\_Price " ...



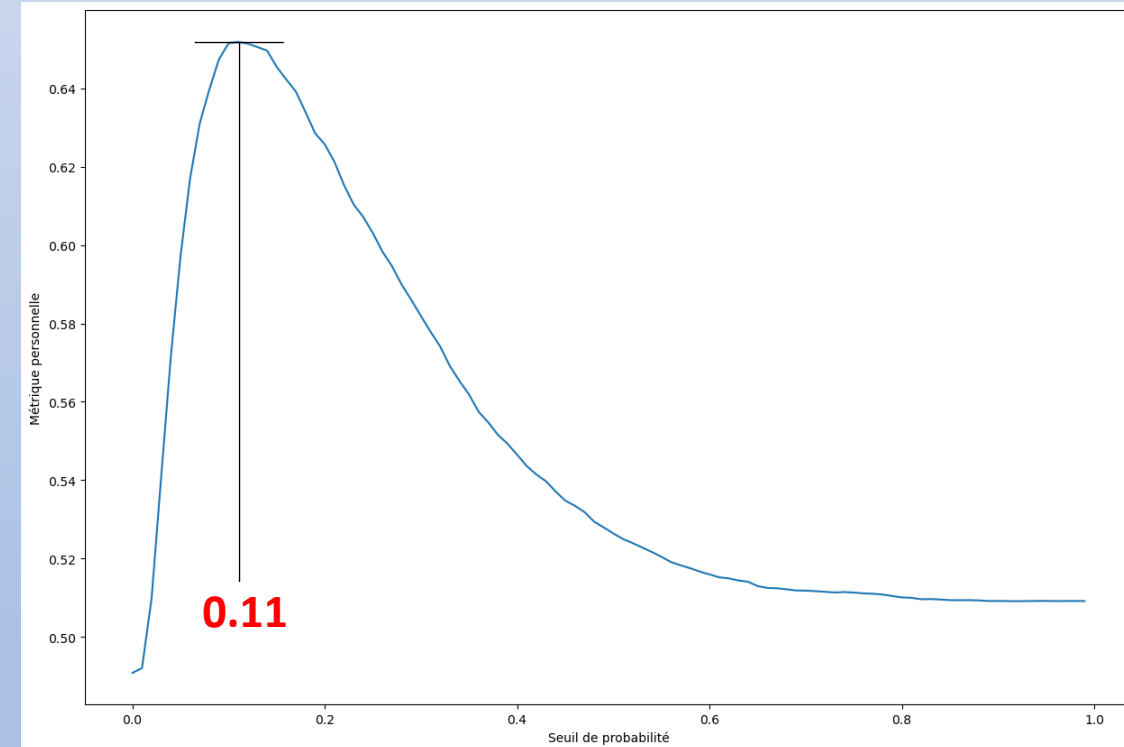
### Définition du Seuil dans notre problématique

Un prêt non remboursé coûte bien plus cher à la banque que ce qu'il aurait potentiellement rapporté.

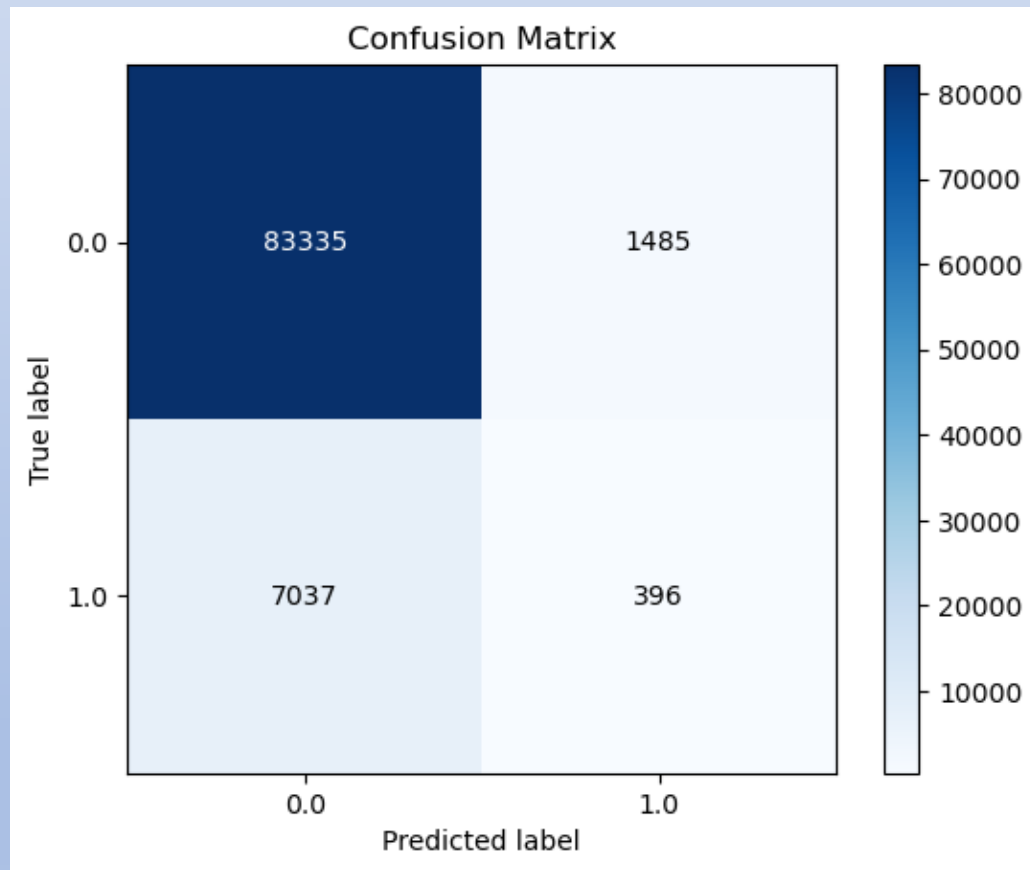
Nous modélisons ça de la manière suivante :

- 0 pour les vrais négatifs (TN)
- 1 pour les faux positifs (FP)
- +1 pour les vrais positifs (TP)
- 10 pour les faux négatifs (FN)

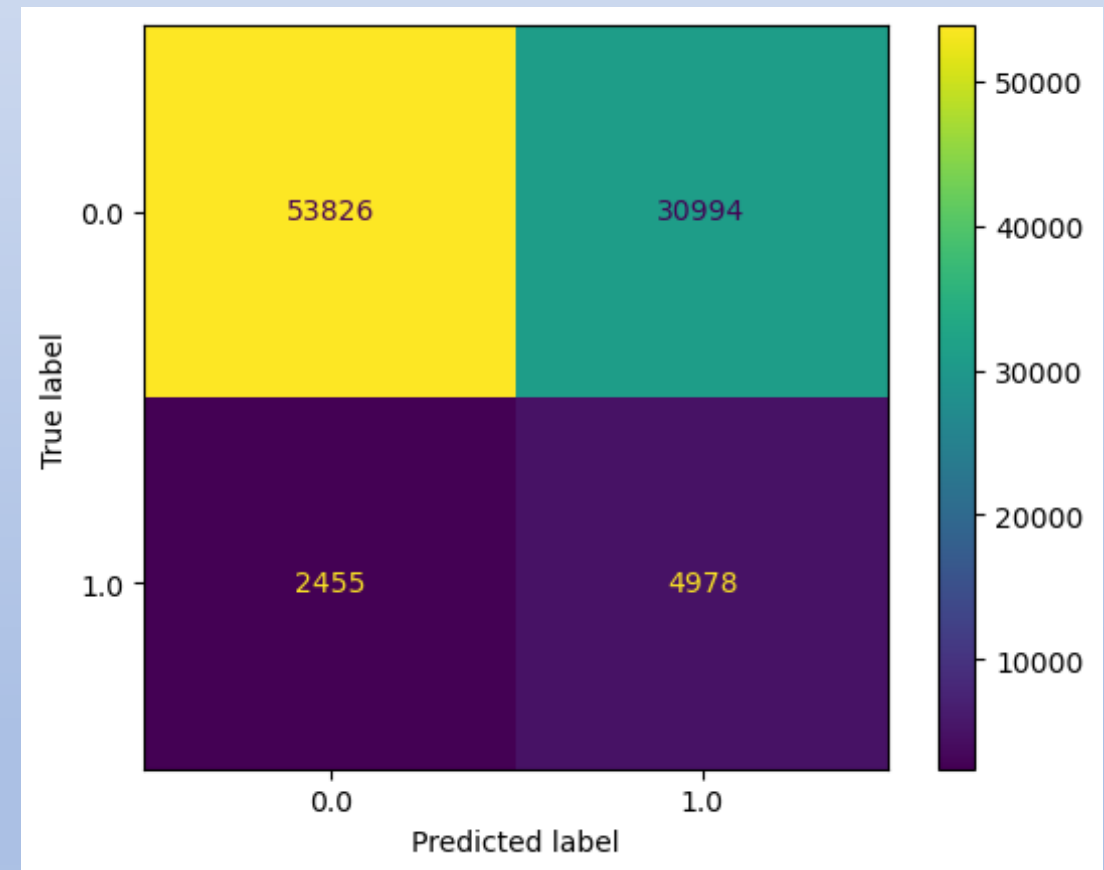
Le seuil qui optimise la fonction se situe à 0,11, ce qui signifie qu'un crédit sera accordé si la probabilité de défaut d'un client est inférieure à 11%.



**Avec le seuil par défaut à 0.5**



**Avec le seuil optimal à 0.11**

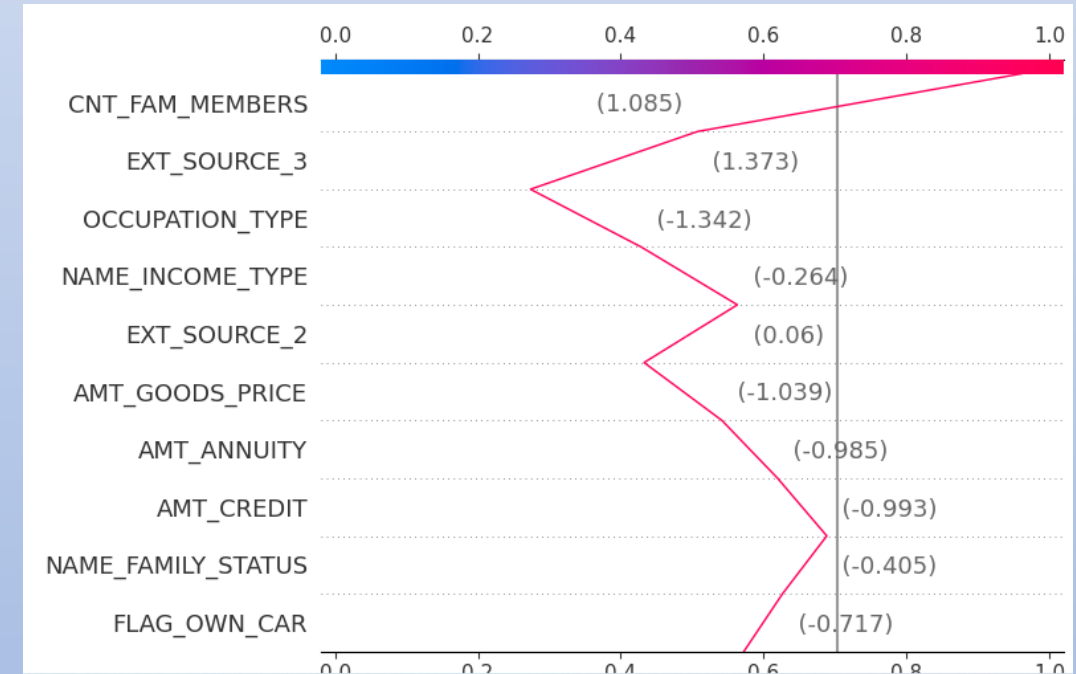




# DASHBOARD

<https://github.com/zhmidi78>

<https://zh-pretprediction.streamlit.app/>



**MERCI**