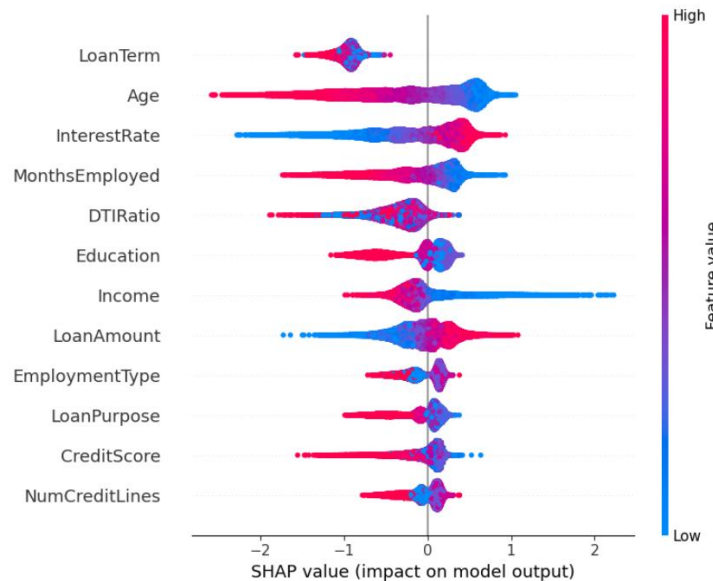


# ***Analyse Prédicative des Défaits de Paiement***



# Sommaire

- ① Problématique
- ② Données
- ③ Modélisation
- ④ Interprétation
- ⑤ Conclusions

# Sommaire

① Problématique

② Données

③ Modélisation

④ Interprétation

⑤ Conclusions

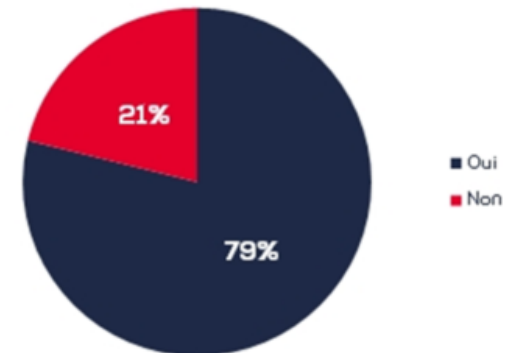
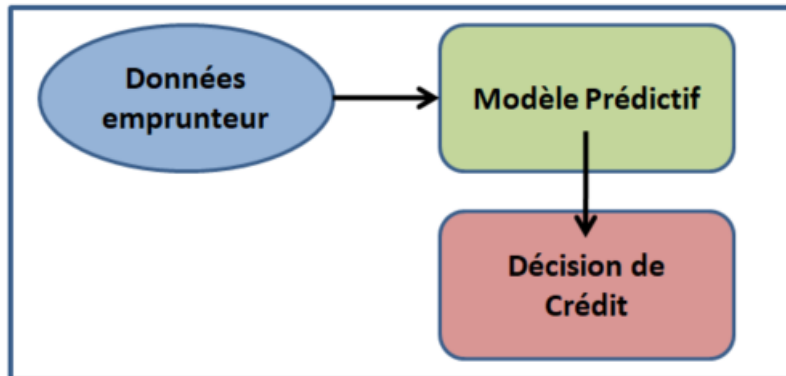
# Problématique: Introduction

## Contexte:

- Les institutions financières font face à un risque croissant de défauts de paiement.
- Importance de prédire ces risques pour minimiser les pertes.

## Objectifs:

- Développer un modèle prédictif pour identifier les emprunteurs à risque.
- Améliorer la gestion des risques de crédit grâce à l'analyse des données.



# Problématique: Etapes du projet

# EDA

- Importation jeu de données
- Types de données, nan, outliers, doublons, n-unique
- Analyses descriptives variables
- Corrélation des variables

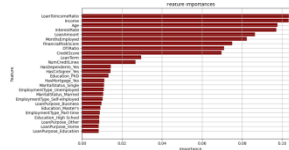
[illegible]

# PRÉ PROCESSING

- Séparation des variables,
- Encodage,
- Split du jeux de données,
- Standardisation

## FEATURES SELECTION

- Première modélisation
- Features importances
- Selection des variables les plus importantes



## MODELISATION

- Choix des modèles,
- Rééquilibrage des classes
- Choix des métriques
- Optimisation des modèles

## INTERPRETATION RESULTATS

- Tableau des métriques
- Matrices de confusion
- Analyse des performances des modèles

# Modèle Final

# Sommaire

① Problématique

**② Données**

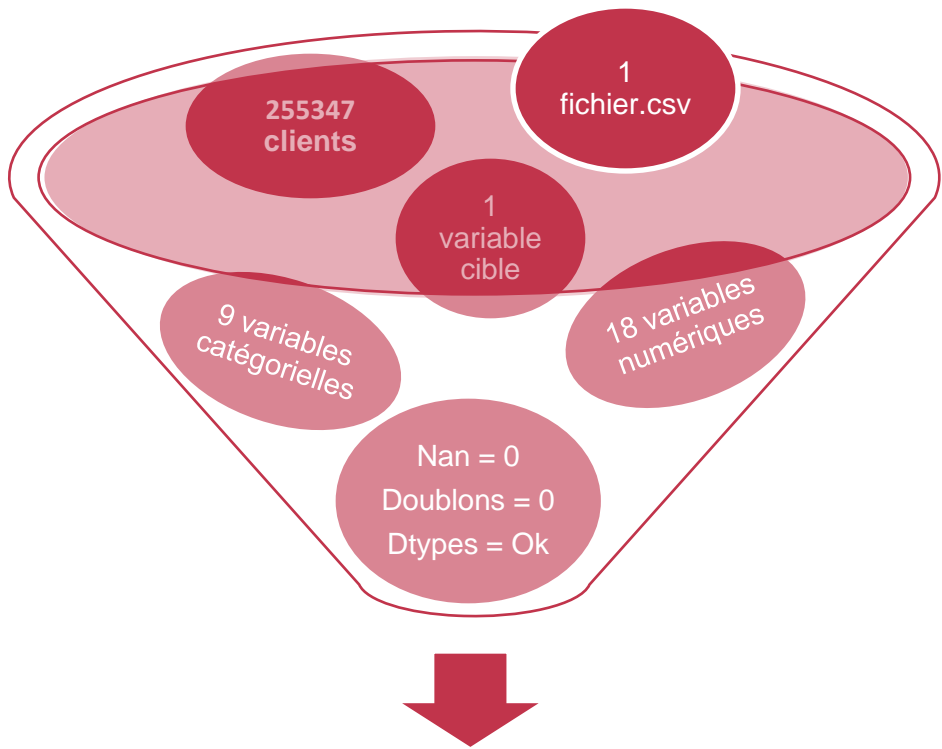
③ Modélisation

④ Interprétation

⑤ Conclusions

# Données

## Jeu de données:



	Colonne	Description
0	LoanID	Id unique client
1	Age	âge
2	Income	revenu annuel
3	LoanAmount	montant du crédit
4	CreditScore	score de crédit qui indique la capacité de rem...
5	MonthsEmployed	nombre de mois de travail
6	NumCreditLines	nombre de crédits déjà en cours
7	InterestRate	taux d'intérêt du crédit
8	LoanTerm	nombre de mois de remboursement du crédit
9	DTIRatio	ratio dette/revenu
10	Education	niveau d'études
11	EmploymentType	type d'emploi
12	MaritalStatus	statut marital
13	HasMortgage	hypothèque
14	HasDependents	personnes à charge
15	LoanPurpose	objet du prêt
16	HasCoSigner	co-signataire
17	Default	défauts de paiement

LoanID	Age	Income	LoanAmount	CreditScore	MonthsEmployed	NumCreditLines	InterestRate	LoanTerm	DTIRatio	Education	EmploymentType
I38PQUQS96	56	85994	50587	520	80	4	15.23	36	0.44	Bachelor's	Full-time
HPSK72WA7R	69	50432	124440	458	15	1	4.81	60	0.68	Master's	Full-time
C1OZ6DPJ8Y	46	84208	129188	451	26	3	21.17	24	0.31	Master's	Unemployed
V2KKSFM3UN	32	31713	44799	743	0	3	7.07	24	0.23	High School	Full-time
EY08JDHTZP	60	20437	9139	633	8	4	6.51	48	0.73	Bachelor's	Unemployed

# Données: EDA

## Analyse descriptive: variables numériques

	Age	Income	LoanAmount	CreditScore	MonthsEmployed	NumCreditLines	InterestRate	LoanTerm	DTIRatio
count	255347.00	255347.00	255347.00	255347.00	255347.00	255347.00	255347.00	255347.00	255347.00
mean	43.50	82499.30	127578.87	574.26	59.54	2.50	13.49	36.03	0.50
std	14.99	38963.01	70840.71	158.90	34.64	1.12	6.64	16.97	0.23
min	18.00	15000.00	5000.00	300.00	0.00	1.00	2.00	12.00	0.10
25%	31.00	48825.50	66156.00	437.00	30.00	2.00	7.77	24.00	0.30
50%	43.00	82466.00	127556.00	574.00	60.00	2.00	13.46	36.00	0.50
75%	56.00	116219.00	188985.00	712.00	90.00	3.00	19.25	48.00	0.70
max	69.00	149999.00	249999.00	849.00	119.00	4.00	25.00	60.00	0.90

### (Min, Max)

- Âge emprunteurs: 18 – 69 ans
- Montants de prêt: 5000 – 249999 €
- Taux d'intérêt: 2 – 25 %
- DTI Ratio: 0.1 – 0.9

### Tendances centrales (mean)

- Âge moyen: 43.5 ans
- Revenu moyen: 82 499€
- Montant moyen de prêt: 127 578€
- Taux d'intérêt moyen: 13.49%



**Score de crédit minimum = 300** → emprunteurs avec un profil de crédit à risque élevé  
**Ratio DTI = 0.1 – 0.9** → certains emprunteurs très endettés par rapport à leurs revenus



# Données: EDA

## Analyse descriptive: variables catégorielles

	Education	EmploymentType	MaritalStatus	HasMortgage	HasDependents	LoanPurpose	HasCoSigner	ScoreSegment
count	255347	255347	255347	255347	255347	255347	255347	254863
unique	4	4	3	2	2	5	2	5
top	Bachelor's	Part-time	Married	Yes	Yes	Business	Yes	Poor
freq	64366	64161	85302	127677	127742	51298	127701	129739

- La majorité des emprunteurs ont un **Bachelor** et travaillent à temps partiel (**Part-time**)
- Les emprunteurs sont majoritairement **mariés** (environ 50% sont **mariés**)
- La majorité des prêts sont accordés pour des raisons **Business**
- Plus de la moitié des emprunteurs ont une **hypothèque** et des **personnes à charge**
- La catégorie **Poor** est la plus courante dans les segments de score de crédit

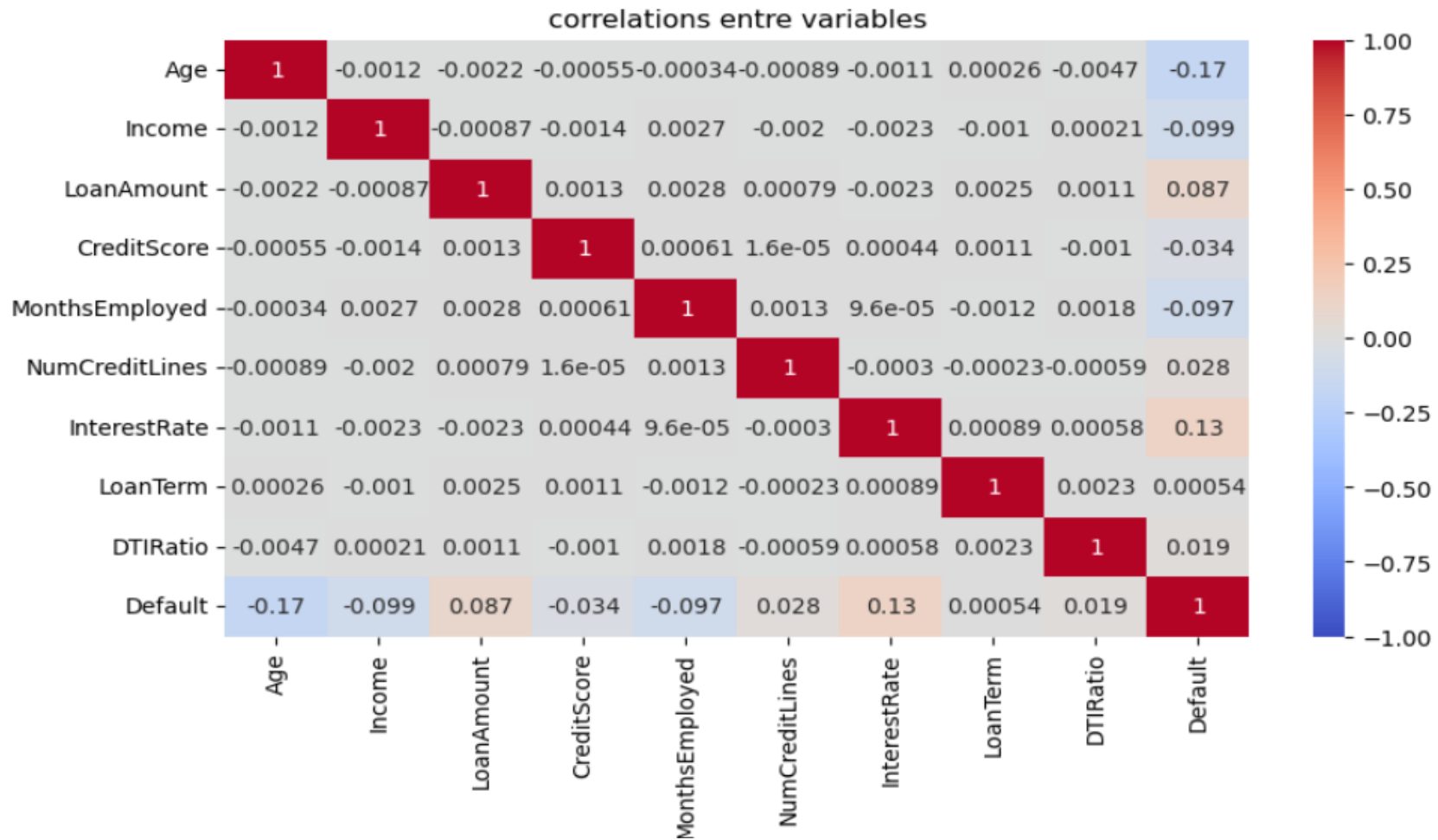
**Statut matrimonial:** majorité des emprunteurs **mariés** ≈ stabilité financière

**Segments de score de crédit:** majoritairement faible (**Poor**) → prise en compte dans les stratégies de modélisation, car risque élevé

**Concentration finalités de prêt :** majorité de prêts accordés pour des raisons commerciales → pertinent de cibler des stratégies spécifiques à ce segment

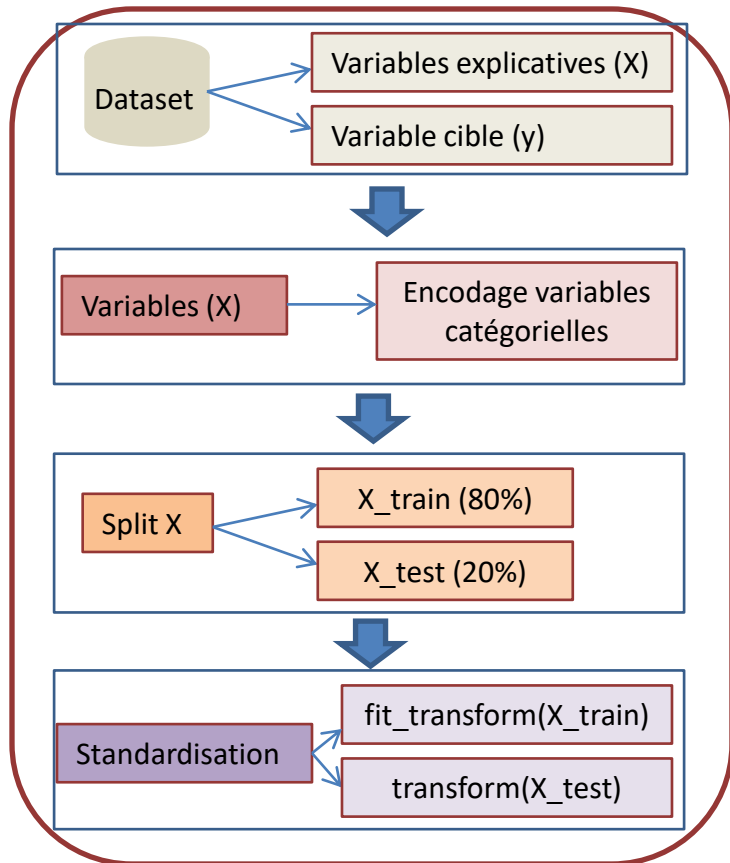
# Données: Corrélation

## Corrélations des variables

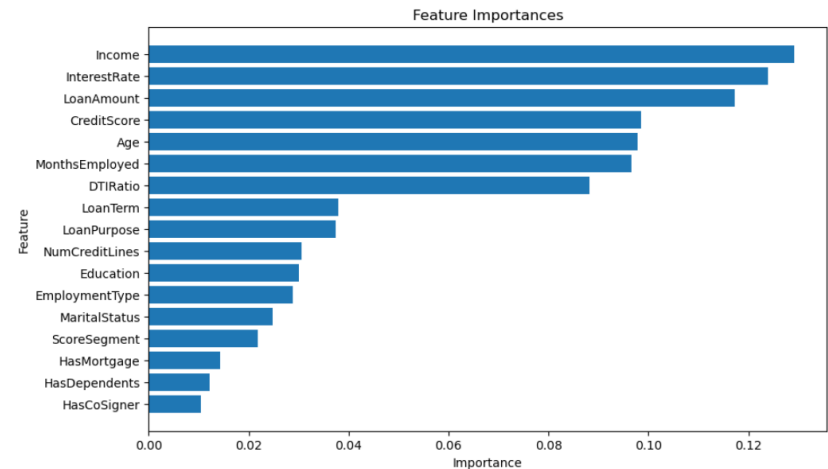


Variables très peu corrélées entre elles → variables quasi indépendantes

# Données: Pré processing et feature engineering



## feature importances



**Suppression variables  
moins importantes  
(feat\_imp < 0.3)**

# Sommaire

- ① Problématique
- ② Données
- ③ Modélisation**
- ④ Interprétation
- ⑤ Conclusions

# Modélisation: Variable cible

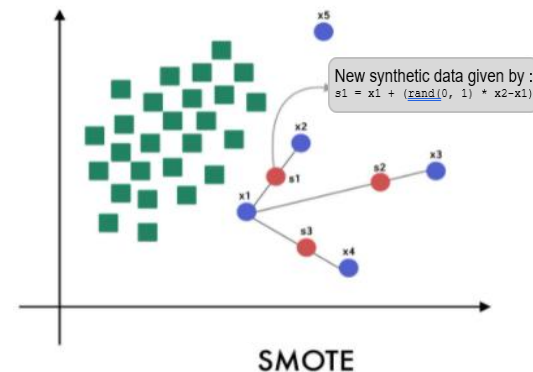
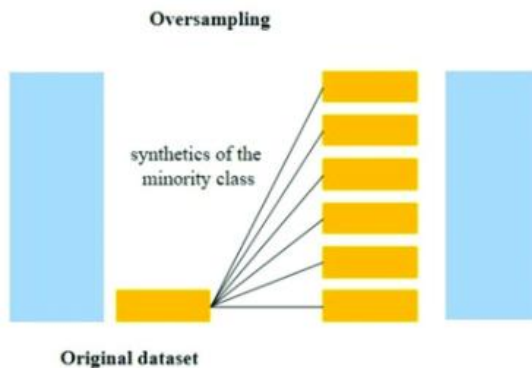
## Variable cible binaire

Classe 0	Classe 1
Clients <b>non défaillants</b>	Clients <b>défaillants</b>
88.6 % majoritaire	11.4% minoritaire



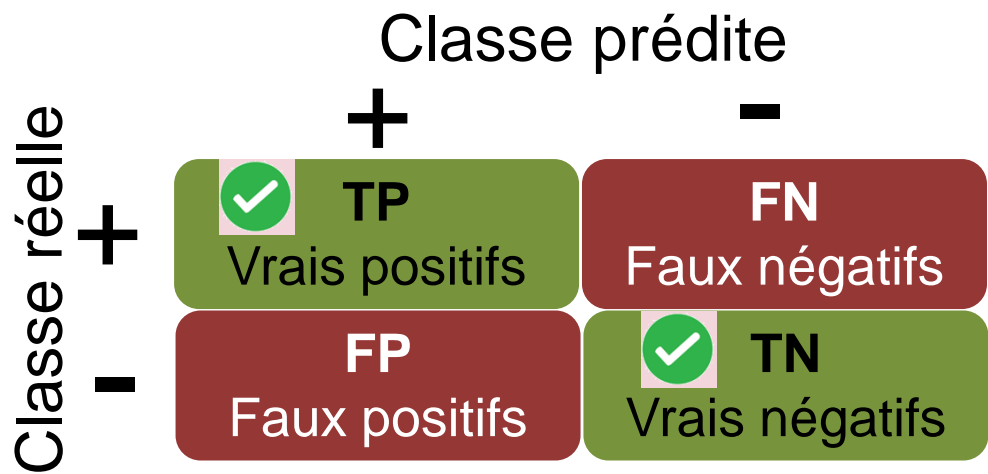
classes **déséquilibrées**

## Rééquilibrage des classes avec SMOTE



# Modélisation: Choix des métriques

## Matrice de confusion



$$\text{Précision} = \frac{TP}{TP+FP}$$

$$\text{Rappel} = \frac{TP}{TP+FN}$$

$$\text{F1-score} = \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Métrique	Description
Précision	proportion de prédictions positives correctes parmi toutes les prédictions positives
Rappel	proportion d'exemples positifs correctement identifiés par le modèle
F1-score	moyenne harmonique de la précision et du rappel
Accuracy	pourcentage de prédictions correctes sur l'ensemble des exemples
AUC-ROC	L'aire sous la courbe ROC, qui évalue la capacité du modèle à distinguer les classes positives des classes négatives
MSE	L'erreur quadratique moyenne

# Modélisation: Bilan des 3 modèles testés

## Tableau des métriques

Modèle	Précision	Rappel	F1-score	AUC-ROC	Accuracy	MSE
Decision Tree	0.175154	0.306799	0.222997	0.558281	0.751243	0.248757
Random Forest	0.246105	0.223325	0.234163	0.685368	0.830037	0.169963
XGBoost	0.316901	0.174184	0.224805	0.700935	0.860231	0.139769

**Meilleur modèle: XGBoost**

- ✓ Meilleure précision, AUC-ROC, accuracy, et MSE
- ✗ Rappel soit le plus faible

Modèle offrant un bon équilibre entre la précision et la capacité à distinguer correctement les classes, ce qui est souvent crucial dans les problèmes de classification binaire comme la prédiction de la capacité de remboursement d'un emprunteur.

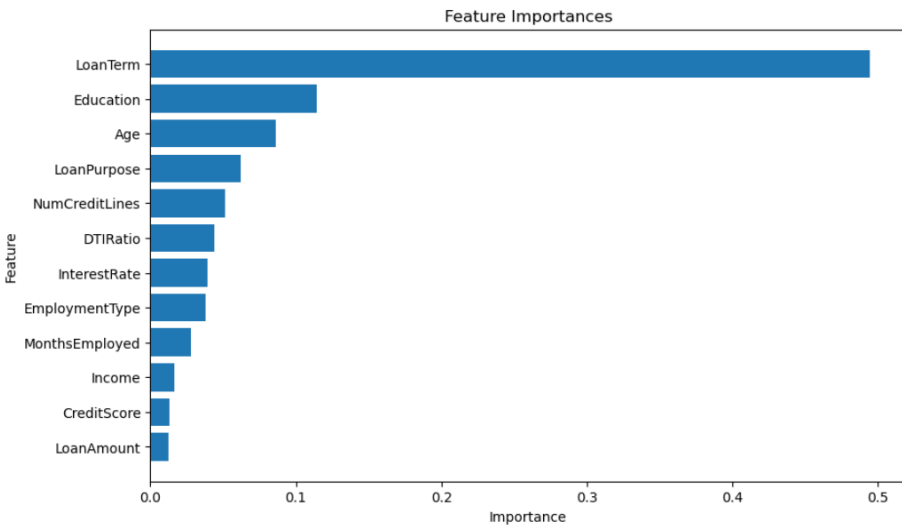
# Sommaire

- ① Problématique
- ② Données
- ③ Modélisation
- ④ Interprétation**
- ⑤ Conclusions

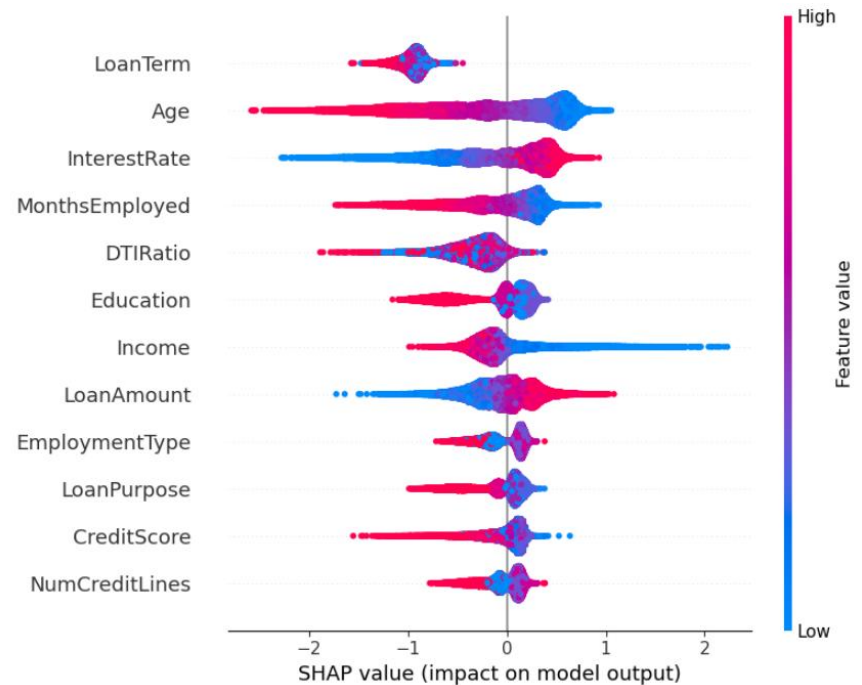


# Modélisation: Interprétation XGBoost

## Feature importances



## Shap values



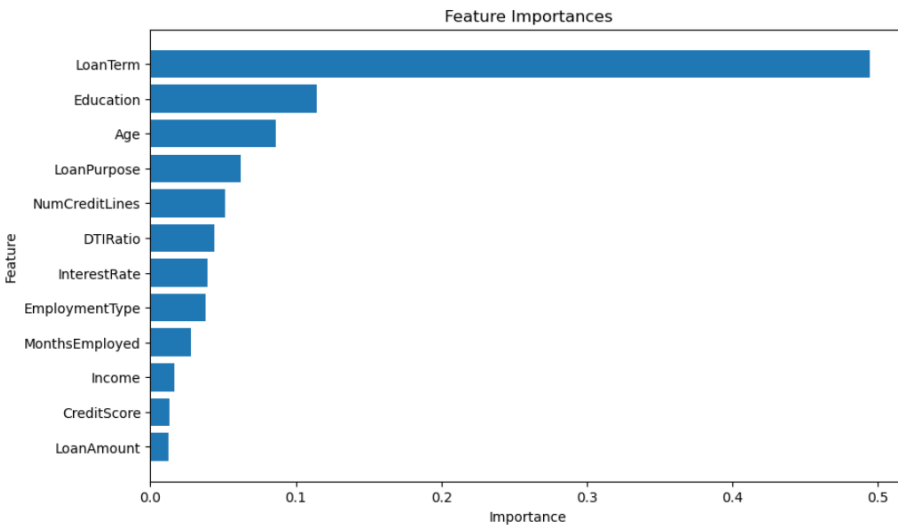
Le graphique de la SHAP value ci-dessus aide à comprendre non seulement quelles caractéristiques sont importantes pour les prédictions de notre modèle, mais aussi comment ces caractéristiques influencent la décision du modèle.

## Observations:

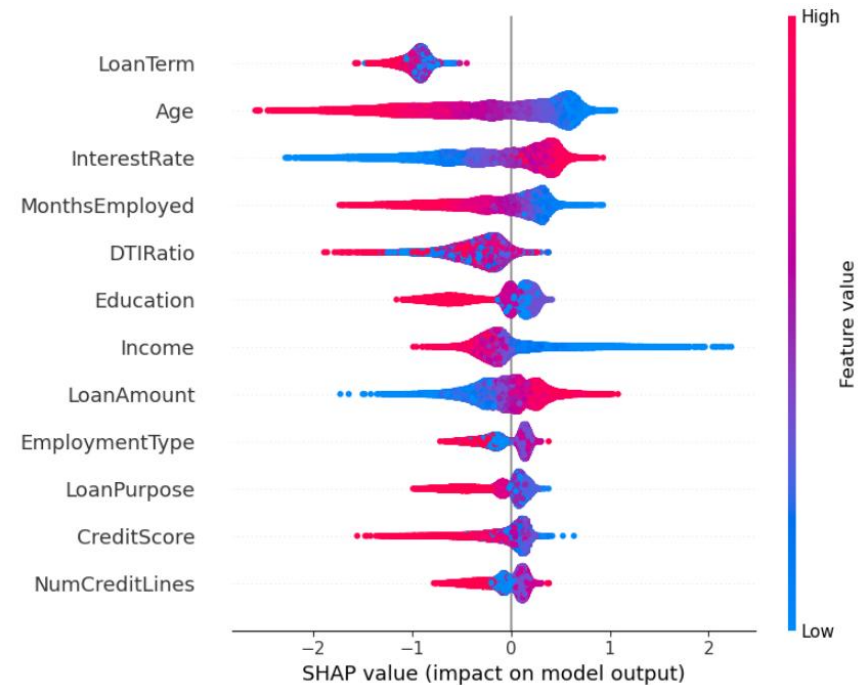
- Plus le terme de remboursement est court, plus l'emprunteur est susceptible d'être en défaut de paiement
- Plus l'emprunteur est jeune, plus la probabilité d'être en défaut de paiement est élevée
- Plus le taux d'intérêt est élevé, plus la probabilité de remboursement est faible
- Plus le montant du crédit est élevé, plus la probabilité de défaut de paiement est élevé
- Un revenu annuel faible augmente le risque de défaut de paiement

# Modélisation: Interprétation XGBoost

## Feature importances



## Shap values



## Les 5 variables avec le plus d'impact sur la prédiction :

**LoanTerm** (durée de remboursement), **Age**, **InterestRate** (taux d'intérêt), **MonthsEmployed**(nombre de mois en fonction), et **DTIRation** (ratio dette/revenu)

## Recommandations:

- Evaluation de la stabilité professionnelle et des revenus annuels
- Evaluation du ratio dette/revenu
- Adaptation du taux d'intérêt et du terme de remboursement en fonction du revenu
- Accent sur l'âge de l'emprunteur: fixer des montants bas pour le plus jeunes

# Sommaire

- ① Problématique
- ② Données
- ③ Modélisation
- ④ Interprétation
- ⑤ Conclusions**

# Conclusions et perspectives

**Problématique de classification binaire avec classes déséquilibrées**

**Modèle retenu: XGBoost en raison de sa robustesse et de ses performances**

**Améliorations potentielles:**

- Ajustement des hyperparamètres pour améliorer le rappel
- Exploration de modèles hybrides ou d'ensemble
- Collecte de données additionnelles pour enrichir le modèle