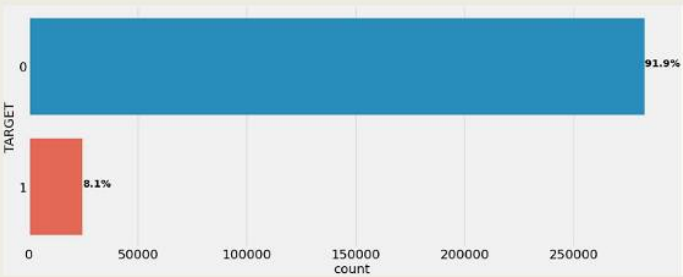


Scoring pour l'octroi de crédit

Application pour déterminer le risque d'octroyer un crédit à un client

Nous disposons d'une base de données de 500000 clients ayant contracté un prêt auprès d'une banque, certains (une minorité) ont connu des incidents de paiement, d'autres non. La base contient 240 données pour chaque client (âge, revenu, temps de travail, montant du prêt, situation de famille, nombre de personnes à charge, ...)

L'objectif est de développer une application de scoring en ligne (et une API hébergée chez Heroku) permettant si on peut accorder un prêt à une personne en limitant le risque pour la banque.



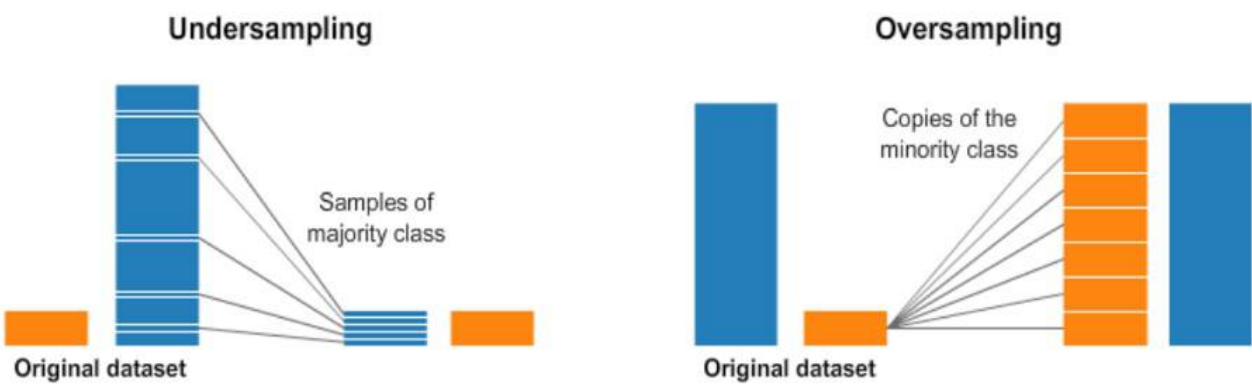
Fort déséquilibre :

91,9% de clients 0

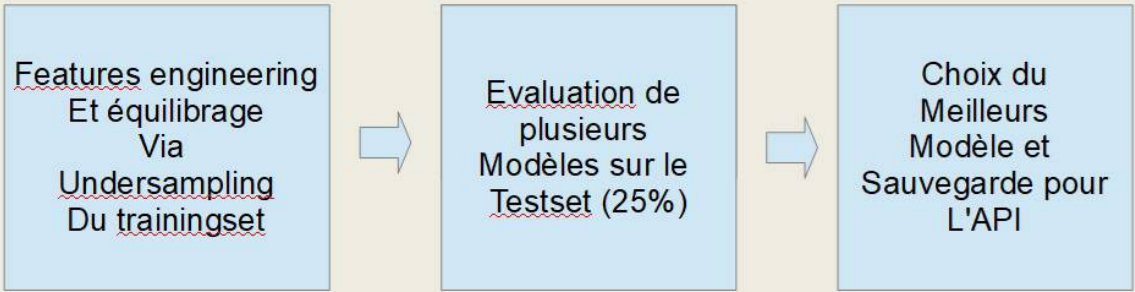
8,1% de clients 1

Il faut rééquilibrer le jeu de données :

- undersampling
- oversampling
- smote



Démarche générale



Modèles étudiés :

- Regression Logistique (base line)
- Random Forrest
- XGBoost

Résultats :

Over sampling

	Accuracy	Precision	Recall	F1_score
Logistic regression	0.69	0.16	0.68	0.25
Random Forest	0.72	0.18	0.74	0.29
XGBoost	0.76	0.22	0.79	0.34

Under sampling

	Accuracy	Precision	Recall	F1_score
Logistic regression	0.69	0.16	0.68	0.25
Random Forest	0.68	0.17	0.75	0.27
XGBoost	0.71	0.19	0.81	0.3

