

SCORING CLIENT



**App &
dashboard**



PROJET

Votre client souhaite mettre à disposition un tableau de bord à ses conseillers financiers indiquant diverses informations et indicateurs.

Le but de ce projet est de montrer une première itération complète (ML + déploiement).

Le projet se découpe en 2 parties :

Partie 1 : Construction du modèle de scoring

Une démarche détaillée (les étapes, modèles préssentis à tester, les paramètres de performance utilisés) devra être présentée avant la réalisation pour validation.

Partie 2 : Mise à disposition via dashboard

DATA

Les données

Nous avons décidé de travailler avec le fichier `application_train.csv` et `application_test.csv`

Observations

- Colonne cible : TARGET
(1 > client avec difficultés de paiement / 0 > autre cas)
- Dataset déséquilibré : 0 > 282 686 (92%) / 1 > 24825 (8%)
- 0 duplicate
- Shape : 307 711 lignes x 122 colonnes
- Null value : 9 152 465
- Missing value : 9 152 465

Preprocessing

- Nous gardons uniquement les colonnes dont on a besoin pour application en ligne : 'TARGET', 'CODE_GENDER', 'NAME_FAMILY_STATUS', 'NAME_EDUCATION_TYPE', 'DAYS_BIRTH', 'CNT_CHILDREN', 'AMT_INCOME_TOTAL', 'NAME_CONTRACT_TYPE', 'OCCUPATION_TYPE', 'AMT_CREDIT'
- Suppression des NaN
- Transformation de la colonne 'DAYS_BIRTH' pour obtenir l'âge d'un client en année (et non en nombres de jours)
- Données finales : 10 colonnes et environ 200 000 lignes

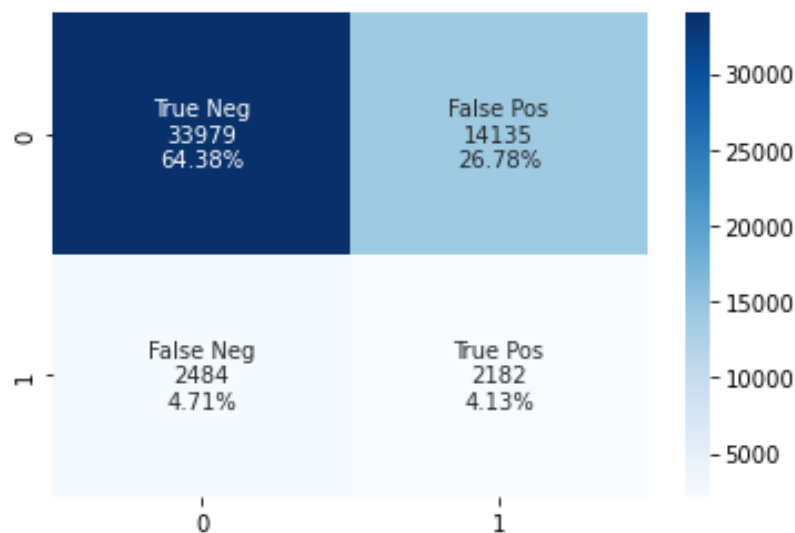


Pipeline & modèle

- Encodage (OneHotEncoder) des colonnes catégoriques
- Scaling (MinMaxScaler) des colonnes numériques
- Rééquilibrage (SMOTE) des données
- Modèle : Random Forest Classifier (avec des paramètres qui ont été testé auparavant grâce à un Grid Search)

Résultats

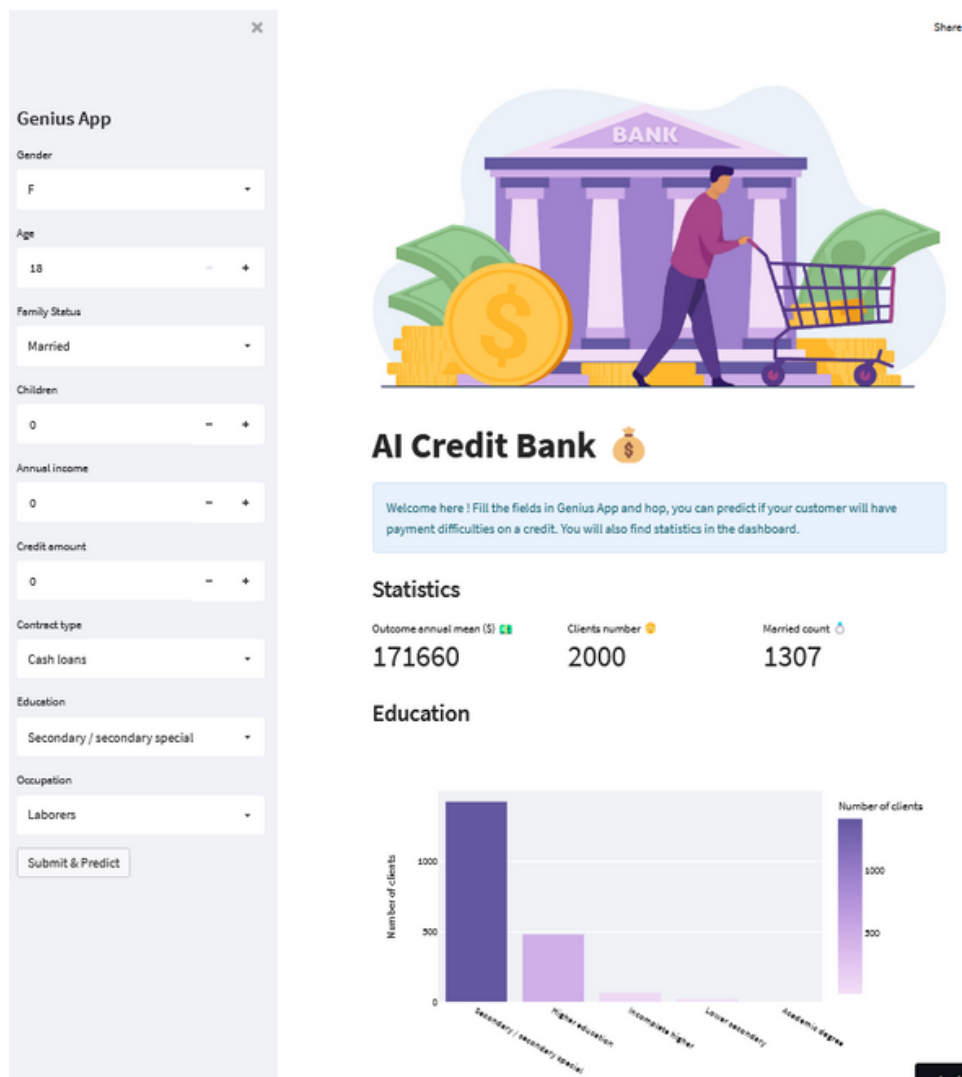
	precision	recall	f1-score	support
0	0.93	0.71	0.80	48114
1	0.13	0.47	0.21	4666
accuracy			0.69	52780
macro avg	0.53	0.59	0.51	52780
weighted avg	0.86	0.69	0.75	52780



WEB APP

Disponible sur

https://share.streamlit.io/bonjourcerise/simplon_scoringclientapp/app.py



WEB APP

Technologies

- Application permettant à l'utilisateur de savoir, selon les attributs d'un profil, si un crédit pourra être accordé ou non.
- Dashboard (visualisation et KPI) du fichier allégé (2000 lignes) application_test.csv

Technologies

- Environnement Python 3.9.7
- Langage : Python
- Librairie open source : Streamlit
- Visualisation : Plotly
- Déploiement test (et crash) : Heroku et Image docker sur Azure
- Déploiement final : Streamlit avec Github