

Implémentation d'un modèle de scoring

API et Dashboard

Raquel Sanchez Pellicer
Projet 7 OpenClassrooms
15/08/2023

Table de matières

1. Objectif de l'API et du dashboard.....	3
2. Dépôt des sources.....	3
3. Accès.....	3
4. Présentation de l'API.....	3
5. Présentation et Mode d'emploi du dashboard.....	3

CAHIER DES CHARGES

Prêt à dépenser est une société financière d'offre de crédit à la consommation pour la clientèle ayant peu ou pas d'historique de prêt.

Notre mission est de **développer un modèle de scoring de la probabilité de défaut de paiement du client** pour étayer la décision d'accorder ou non un prêt à un client potentiel en s'appuyant sur des sources de données variées.

Le **développement d'un dashboard interactif** permettra aux chargés de clientèles d'expliquer avec transparence la décision d'octroi ou non du prêt et de mettre à disposition des clients l'exploration de leurs informations personnelles.

1. Objectif de l'API et du dashboard

La **transparence** est au cœur des valeurs de Prêt à dépenser. L'objectif du dashboard développé est de servir aux chargés de relation client d'outil pour **expliquer** de la façon la plus transparente possible les **décision** d'octroi de crédit. Il permet également d'**améliorer la connaissance** client des chargés de relation client, leur permettant d'explorer les informations personnelles des clients.

Les données relatives à la décision ainsi que les informations personnelles des clients sont disponibles via l'API.

2. Dépôt des sources

Link : https://github.com/Raquel-SP/OC_DS_P7_implementez_modele_scoring/

3. Accès

Accès API : <http://13.37.97.118/>

Accès dashboard : IP publique changeant à chaque connexion sur l'instance EC2

4. Présentation de l'API

L'API est structurée en quatre sections.

Dans la première section l'API charge les données nécessaires pour la prédiction, l'ensemble des données personnelles disponibles concernant les clients et le modèle entraîné. Les données qui serviront à la prédiction sont ceux issus de la phase de *feature engineering et sélection* des variables, ils sont standardisés. Les données qui permettront aux gestionnaires des relations avec la clientèle d'expliquer les décisions comprennent également des données personnelles et d'autres informations non standardisées, c'est pourquoi les données de la source sont également utilisés.

La deuxième et troisième sections permettent l'accès aux données nécessaires pour la prédiction (« */dataAPI* ») et aux données non standardisées (« */dataGeneral* ») respectivement.

Dans quatrième section (« *dataAPI/client/<id_client>* ») à partir d'un identifiant client on applique le modèle développe pour obtenir la prédiction, ainsi que les données nécessaires pour faciliter interopérabilité locale (valeurs de Shap des variables pour le client en question).

5. Présentation et Mode d'emploi du dashboard

Le dashboard (Fig. 1) comporte de différents sections permettant de sélectionner le client pour lequel on souhaite des informations, à travers son ID client, ainsi que d'afficher des informations concernant le score attribué ou lui comparer au reste des clients.

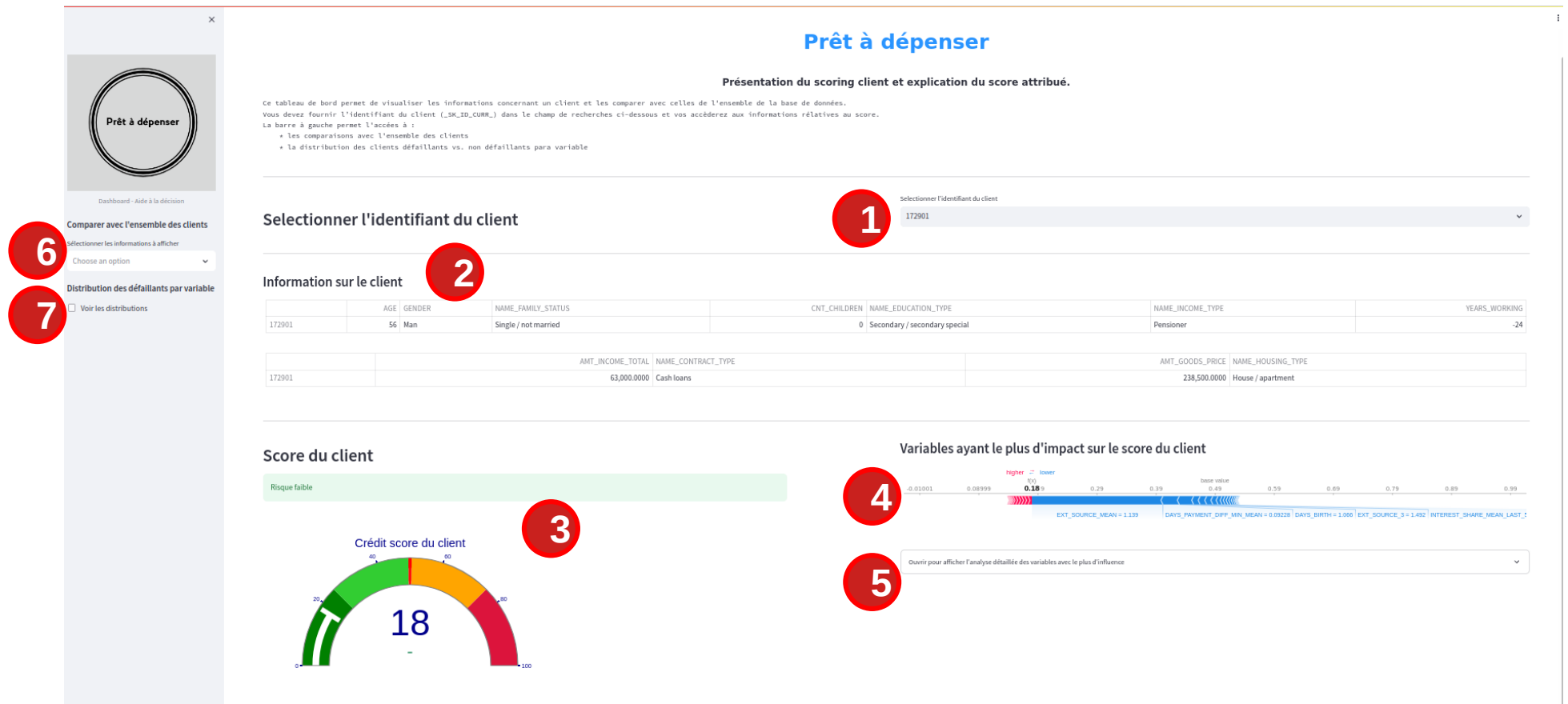


Figure 1: Copie d'écran du Dashboard

- 1** Liste de sélection d'un client pour décision d'octroi de prêt (jeu de données de validation à partir de application_train.csv pré traité de la même manière que le jeu de données ayant servi à entraîner le modèle, la variable cible a été utilisé pour créer un sous-ensemble ayant autant de défaillants que de non défaillants à fin de dynamiser la présentation).
- 2** Informations personnelles et bancaires minimales pour le client sélectionné.
- 3** Jauge permettant de visualiser rapidement le score du client.
- 4** Graphique, « force plot », illustrant les variables ayant le plus d'impact sur le score octroyé au client. Affiche les facteurs d'influence sur le modèle pour notre client (à gauche en rouge : les variables poussant vers la défaillance, à droite en bleue celles poussant vers la non défaillance).
- 5** Drop-down qui permet d'accéder à un autre graphique, "waterfall plot", qui représente la même information sous une autre représentation (Fig. 2). La lecture et interprétation de ce graphique est similaire au précédent. Bien qu'il n'apporte pas de nouvelles informations, il a été inclus car il peut faciliter la compréhension.

Variables ayant le plus d'impact sur le score du client

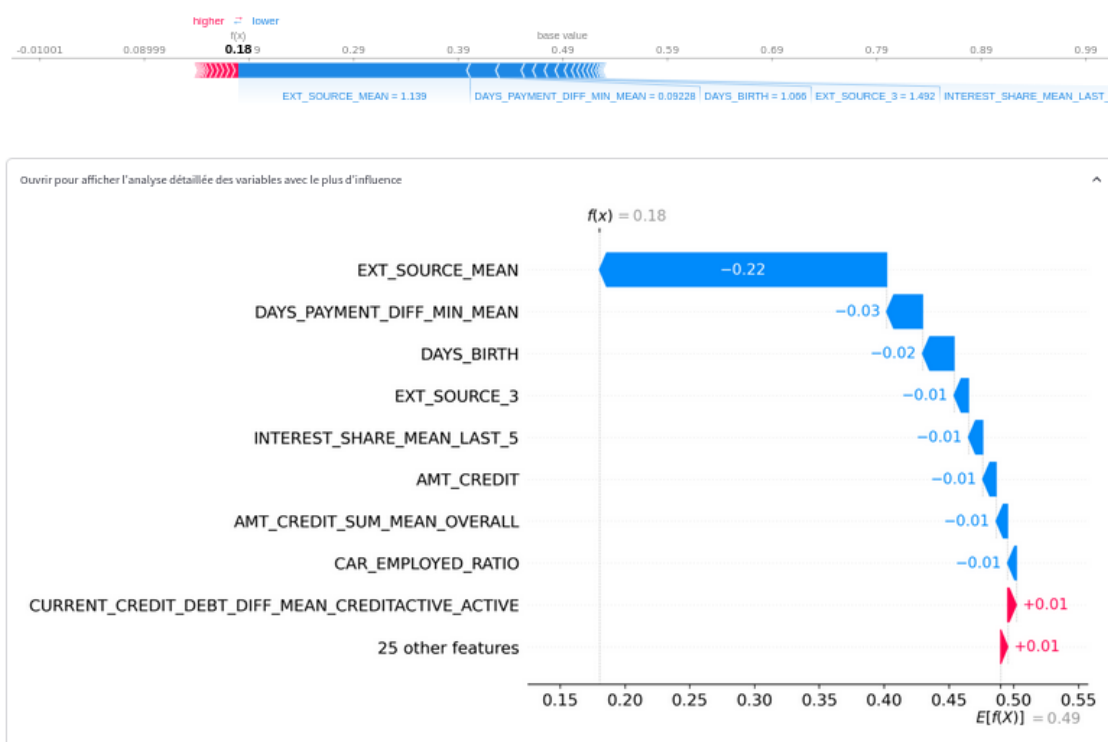


Figure 2: Graphique, waterfall plot, affiché dans le drop-down de la section informant sur le score du client.

- 6** Liste de sélection permettant de choisir les variables pour lesquelles on veut réaliser une comparaison avec l'ensemble des clients (Fig. 3). Pour les variables sélectionnées on connaîtra la valeur caractérisant le client ainsi que la distribution des valeurs sur l'ensemble des clients.



Figure 3: Section permettant de comparer le client d'intérêt à l'ensemble des clients.

- 7** Section permettant la sélection des variables utilisées lors de la modélisation pour visualiser la distribution des défaillants et non-défaillants par catégorie.

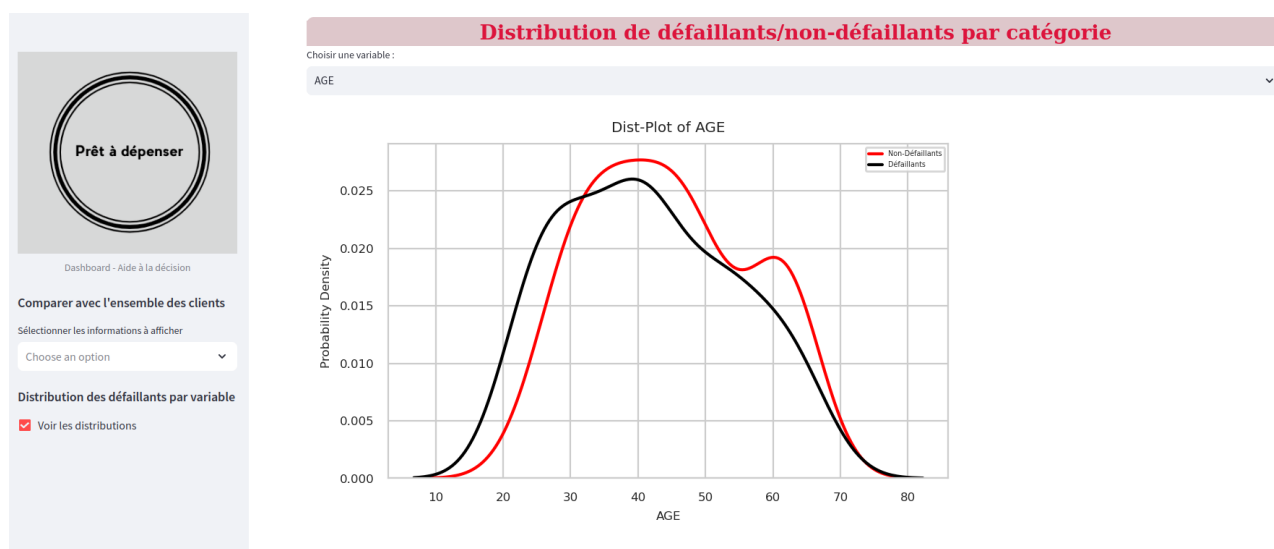


Figure 4: Section permettant d'illustrer la distribution des défaillants et non défaillants par variable.