

PLATEFORME D'ÉVALUATION DU RISQUE DE CRÉDIT

PARC

CREDIT RISK ASSESSMENT PLATFORM

Au sein du Département Innovation de Match Profiler, PARC est une plateforme destinée aux entreprises du secteur financier qui peuvent être liées à l'octroi de crédit et qui vise à détecter les manquements, en tenant compte des caractéristiques des prêts et de ceux qui les demandent.

DÉVELOPPEMENT DE LA PLATEFORME

Sélection de Datasets

Sélection de Datasets représentatifs du problème étudié. Données anonymes, préalablement classées.

Exploratory Data Analysis (EDA)

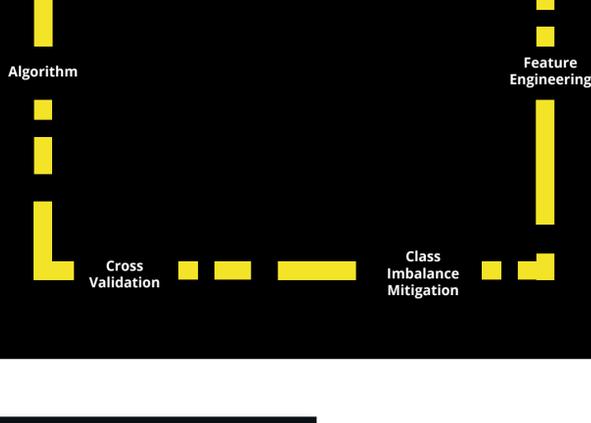
Traitement des données, sélection et transformation des variables : analyse des corrélations, prévention de data leakage, traitement des outliers et missing values, traitement des variables catégorielles.

Benchmarking des algorithmes de classification

Parallèlement à l'EDA, le choix de l'algorithme résulte d'un processus itératif à plusieurs niveaux afin de maximiser les paramètres du modèle. Les caractéristiques des datasets, avec un fort déséquilibre (class imbalance), exigent un soin particulier dans le choix et l'analyse des mesures de la performance du modèle.

BENCHMARKING DES ALGORITHMES

- **Algorithmes considérés:**
SVM, Random Forest, Gradient Boosting, LightGBM, CatBoost, XGBoost
- **Optimisation des paramètres :**
Grid Search, Bayesian Optimization, Genetic Algorithms
- **Feature engineering:**
Génération de nouvelles variables grâce à des transformations et des agrégations
- **Class imbalance:**
Atténuation du problème des données déséquilibrées (minorité de manquement) grâce au SMOTE et à l'oversampling
- **Cross validation:**
Validation en considérant comme paramètres les plus importants l'aire de la courbe ROC et le score F1



WEB SERVICE

- Le cœur des services est un modèle de classification optimisé pour une performance maximale grâce à la compilation en C.
- Le modèle de classification est disponible sous forme d'API REST dans les WebServices.
- La plateforme permet de classer les prêts sous forme de listes de prêts individuelles ou par lots.
- Compte tenu de l'analyse de sensibilité, il existe un endpoint permettant d'obtenir plusieurs niveaux de score correspondant à des fourchettes de montant et de durée.

SERVICE WEB

DE SCORE DE CRÉDIT - INUP

Le service est invoqué par la saisie d'un message composé de 21 variables numériques et qualitatives.

- **Variables spécifiques aux prêts :**
AMontant, durée, objet, garant, codemandeur, taux d'effort.
- **Variables spécifiques du demandeur:**
Antécédents de crédit, situation professionnelle, qualifications, état civil, âge, sexe, nombre de personnes à charge, nombre de crédits contractés auprès de l'institution, autres crédits, travailleur étranger, résidence, épargne, existence de biens mobiliers ou immobiliers, existence de contrats de fourniture de services (eau, électricité, télécommunications, etc.).T

SERVICE WEB

DE SCORE DE CRÉDIT - OUTPUT

Le résultat de l'invocation du service est composé d'une classification de crédit, ainsi que des paramètres de la version actuelle du modèle qui a fait la prévision.

- **Classification (manquement ou non)**
- **Paramètres du modèle:**
 - (AUROC) - Zone correspondant à la probabilité que le modèle classe correctement un cas de manquement réel. Une zone de 0,5 correspond à un modèle aléatoire.
 - Précision $P = \frac{Tp}{Tp + Fp}$: dans l'univers des cas classés comme positifs (manquement) par le modèle, quelle est la proportion de cas corrects. Un P élevé correspond à un faible nombre de faux positifs.
 - Rappel $R = \frac{Tp}{Tp + Fn}$: proportion de cas positifs (manquement) correctement identifiés par rapport au nombre de cas de manquement réels. Un R élevé correspond à un faible nombre de faux négatifs.
 - Score F1 = $2 \frac{P \cdot R}{P + R}$: moyenne harmonique de Précision et de Rappel.

ENDPOINTS

L'API dispose d'un ensemble de services de scoring spécifiques et est destinée à l'audit trail, au logging et à la facturation. Les services de même, sont constitués d'endpoints qui permettent une interaction au cas par cas ou bien de services capables de traiter des listes de prêts. La plateforme offre également la possibilité de mener des études paramétriques, permettant ainsi l'étude et l'évaluation de multiples scénarios de risques.

Endpoint unique :

Il correspond à la classification d'un prêt spécifique. Il consiste en la saisie d'un message au format JSON dans l'input.

Endpoint batch :

Il permet de traiter de multiples cas de prêts. Le message d'input est composé de N cas. La réponse contient la liste des résultats composée de codes d'erreur ou de réussite et de leurs classifications respectives.

Endpoint simulation :

Le message d'input comprend 2 fourchettes de variation pour le montant et la durée et le nombre de points à prendre en compte dans chaque fourchette. La réponse est une carte de score en fonction du montant et de la durée.

CARACTÉRISTIQUES TECHNIQUES DE LA SOLUTION

L'API a été développée dans une architecture sans serveur selon les normes de référence pour les micro-services.

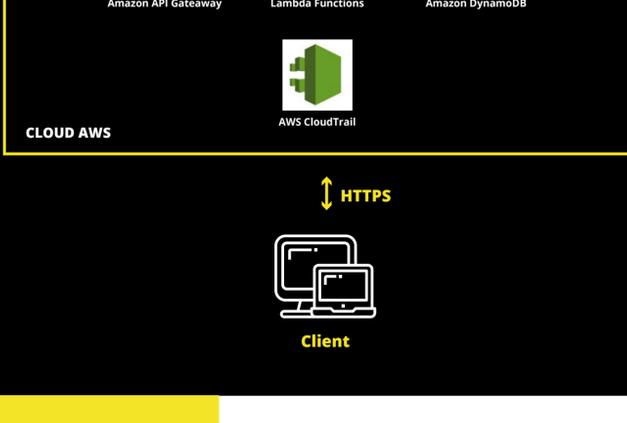
L'accès est contrôlé par les micro-services JWT ou Amazon Cognito.

La communication repose sur des mécanismes de cryptage en transit et sur la persistance, en d'autres termes sur la base de données.

Les invocations de l'API sont restreintes (throttled) conformément à ce qui est établi dans les contrats de service. Les endpoints renvoient les codes appropriés pour les mises en œuvre de back-off par les clients.

Solution sans serveur :

- Évolutif
- Résilient
- Haute disponibilité



FICHE TECHNIQUE

- Algorithme XGBoost
- Traitement des données avec Oversampling et Deep Feature Synthesis
- Modèle compilé avec Python Treelite
- Architecture AWS sans serveur
- API REST

Cofinancé par:

