P7 - IMPLÉMENTEZ UN MODÈLE DE SCORING

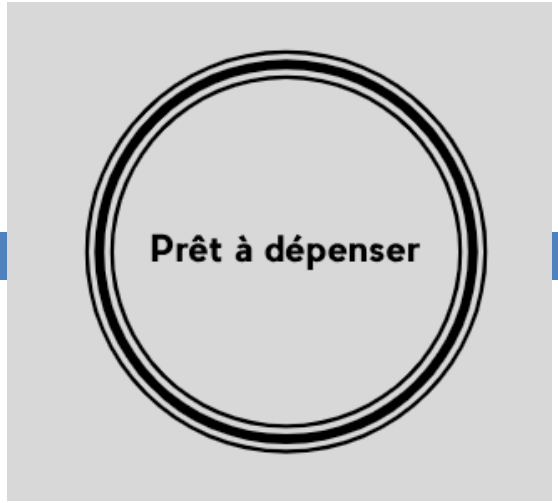
08/10/2021

Etudiant : Luc Rogers
Mentor : Etienne Sanchez

Sommaire

Problématique

Problématique



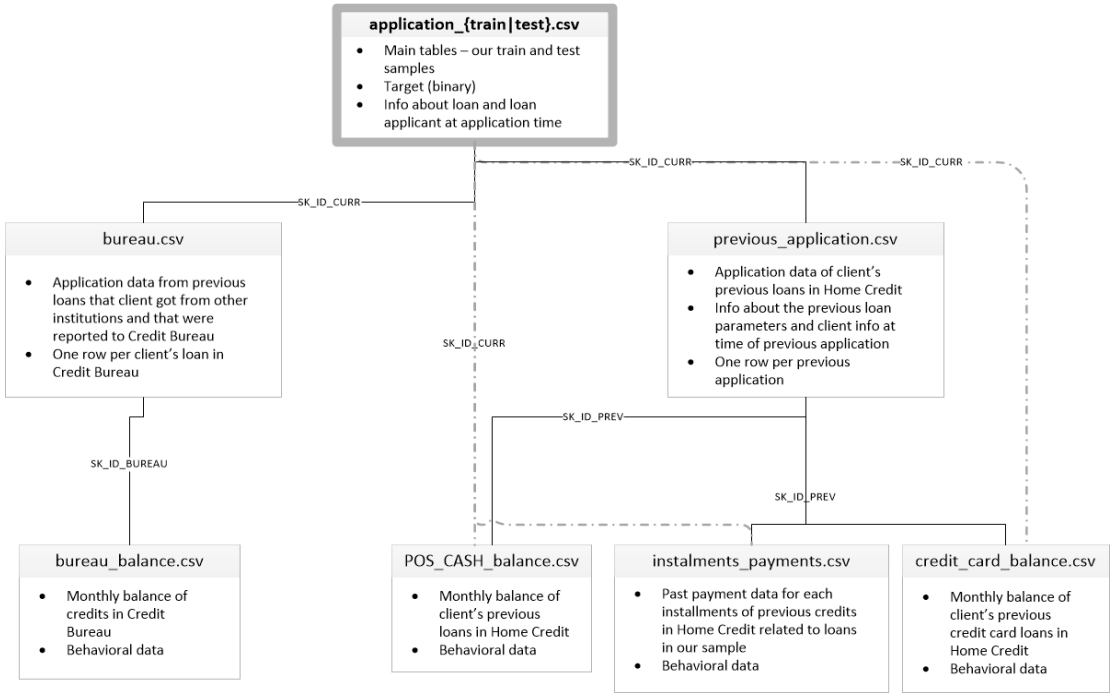
- Contexte: société de crédits à la consommation « Prêt à dépenser »

- Objectifs:
  - Développer un modèle de scoring de la probabilité de défaut de paiement client
  - Déployer un dashboard interactif en ligne répondant à des critères de transparence sur l'explicabilité du modèle

- Liens du projet:
  - Données : https://www.kaggle.com/c/home-credit-default-risk/data
  - Livrables : https://github.com/lucrogers/OpenClassrooms-Projet-7-Data-Scientist-Implementez-un-modele-de-scoring
  - Dashboard interactif : https://loan-management-dash.herokuapp.com

Problématique

- Base de données:
  - ~ 300 000 clients provenant de diverses sources
  - 121 features : salaire, métier, nombre d'enfants, âge, informations relatives aux crédits antérieurs, …
  - Etiquette cible:
    - 0 pour un client sans incident de paiement
    - 1 pour un client ayant connu des difficultés de paiement



Modélisation

Modélisation

- Reprise d'un kernel Kaggle
- Pre-processing:
  - Valeurs aberrantes supprimées
  - Valeurs manquantes remplacées par la moyenne ou par le mode selon leur nature quantitative ou catégorielle
  - Les variables catégorielles sont ensuite encodées avec un encodage de type one hot encoding
  - Création de features métiers
  - Aggrégation des différents datasets par client

➔ Obtention d'un jeu de données de 657 features

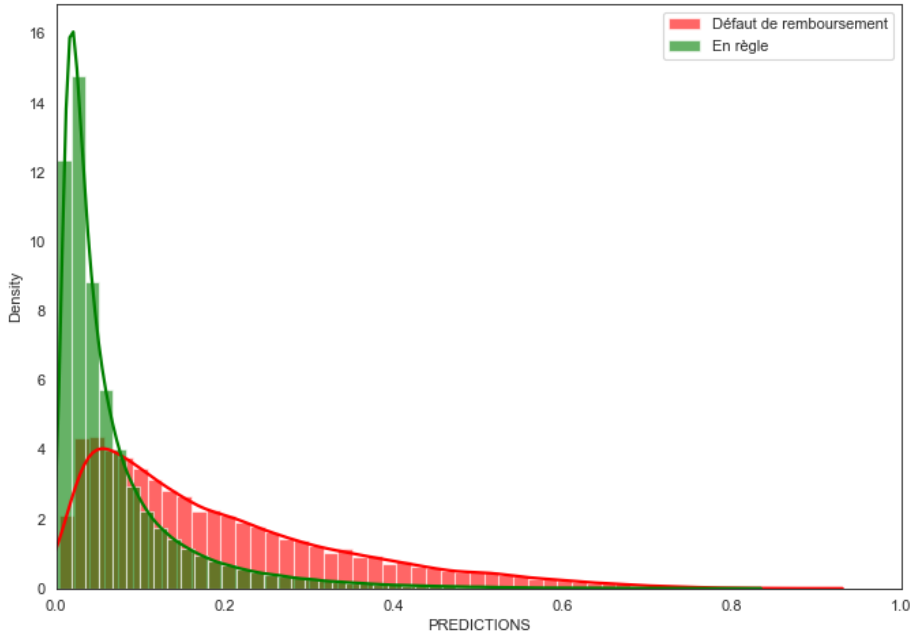
Modélisation

- Méthodologie type:
  - Essais de plusieurs modèles de classification
  - Recherche des hyperparamètres optimaux
  - Validation croisée
  - Choix d'une métrique appropriée: score AUC
- Modèle retenu:
  - Classifieur LightGBM
  - Score AUC: 0,80

Scoring et fonction coût

Scoring et fonction coût

- Score client = probabilité de défaut de paiement (entre 0 et 1)



➔ Une distribution des défauts de paiement légèrement plus étalée vers la droite

➔ Quel seuil choisir ? Compromis à trouver en fonction des coûts et profits

Scoring et fonction coût

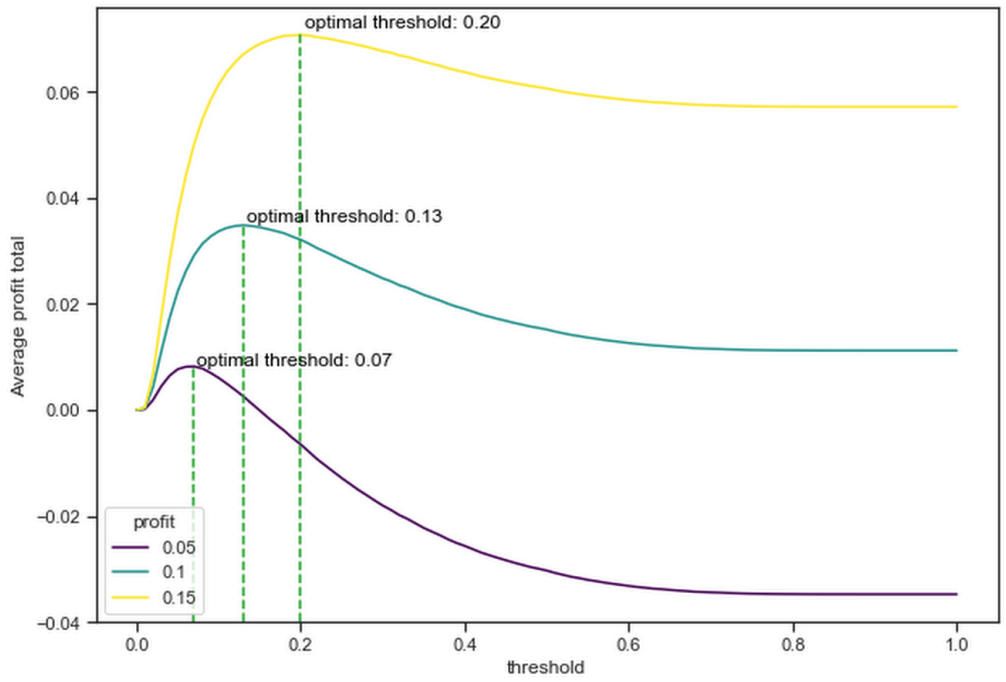
- Fonction coût = bénéfice réalisé par l'entreprise:

$$Bénéfice = [\,(1 - p(seuil)\,) * profit + p(seuil) * coût\,] * n(seuil)$$

Avec:

- p(seuil) : la proportion de clients en défaut de paiement
- profit : le profit moyen réalisé par un prêt en règle (en %)
- coût : le coût moyen associé à un défaut de paiement (en %)
- n(seuil) : le nombre de prêts accordés

Scoring et fonction coût



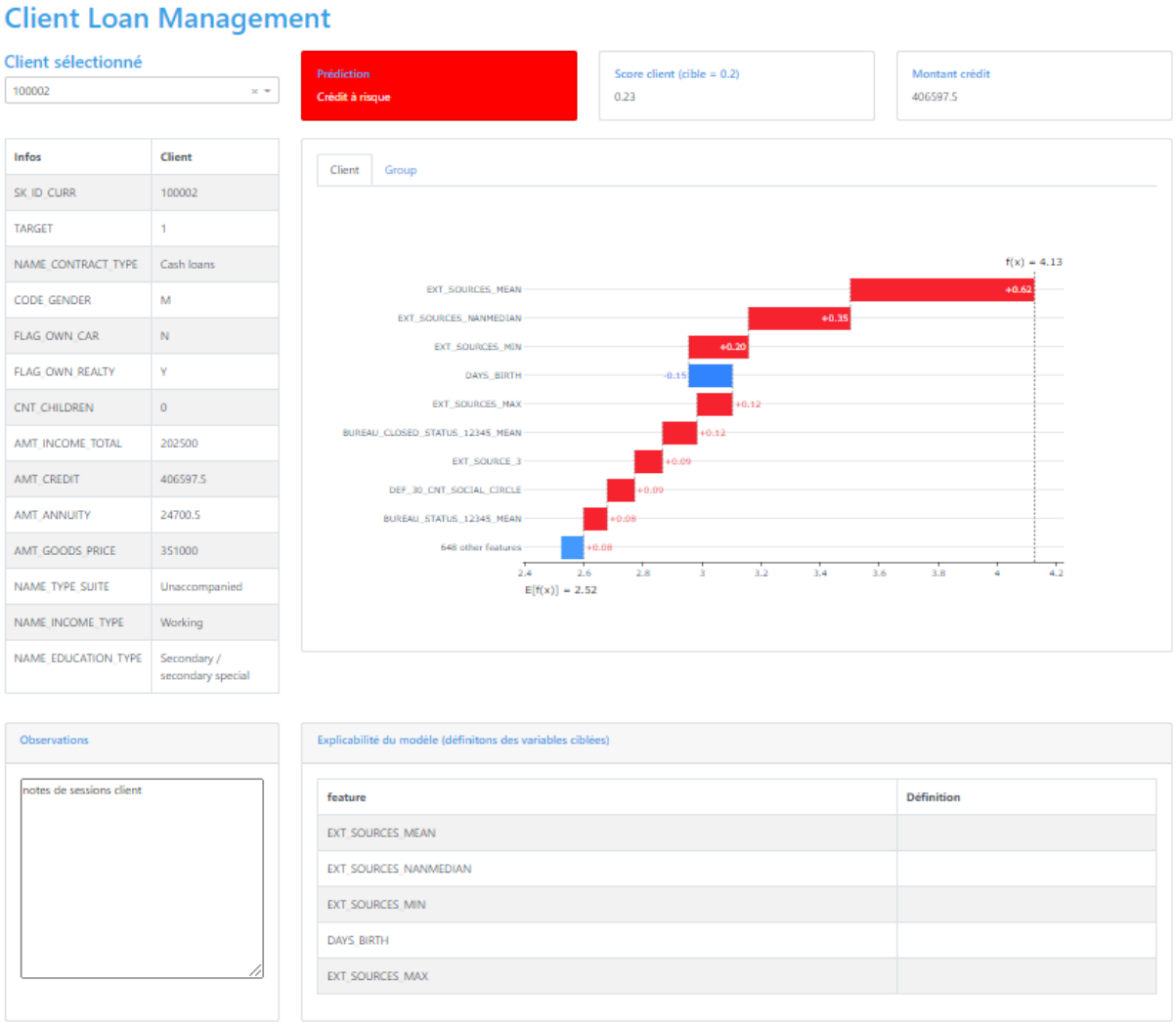
➔ Seuil optimal trouvé en fonction des coûts et profits générés par les pertes et le gains

Présentation du dashboard

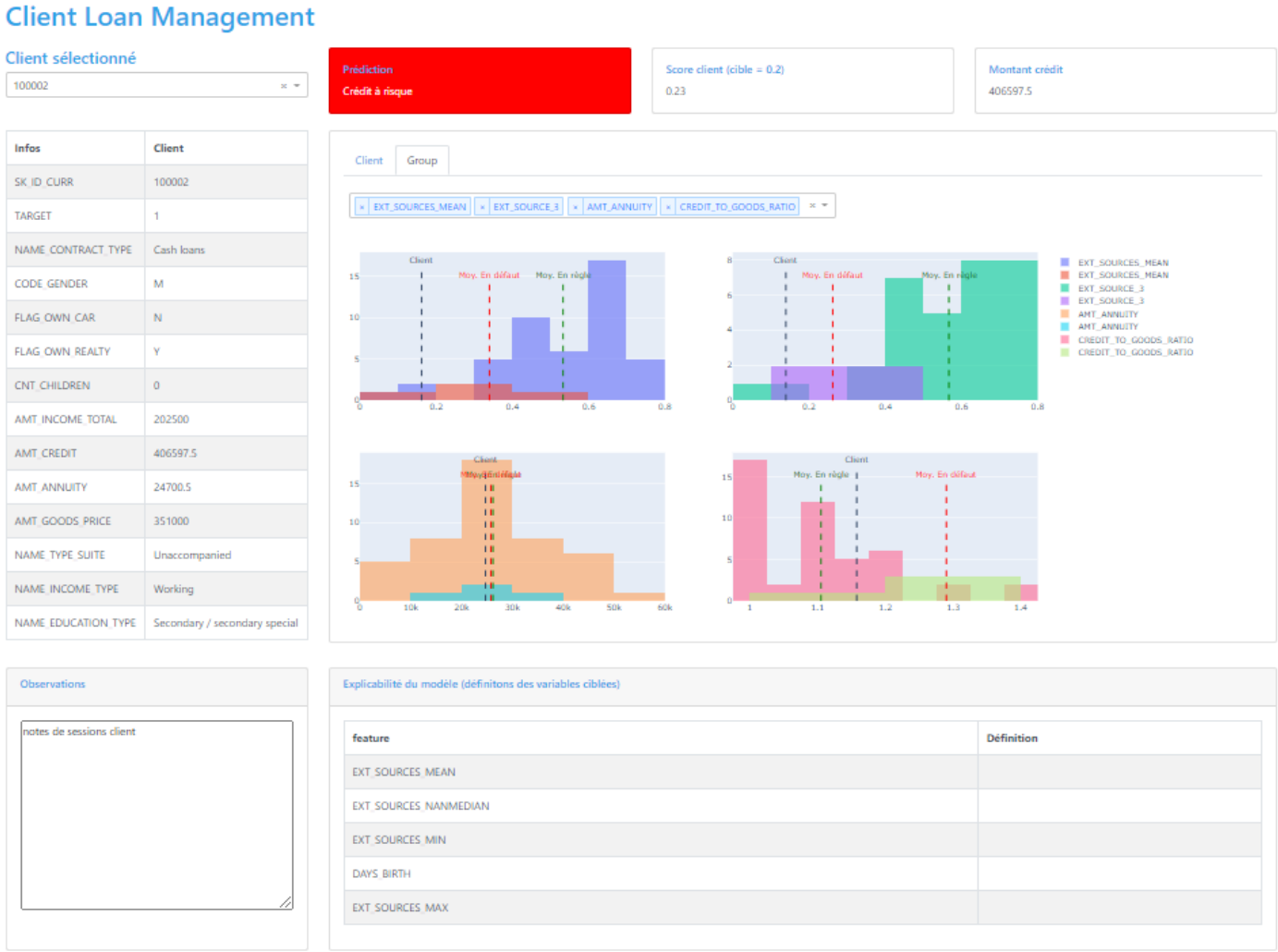
Présentation du dashboard

- Dashboard interactif à l'attention des gestionnaires clients
- Cahier des charges:
  - Permettre de visualiser le score et l'interprétation de ce score pour chaque client de façon intelligible pour une personne non experte en data science.
  - Permettre de visualiser des informations descriptives relatives à un client (via un système de filtre).
  - Permettre de comparer les informations descriptives relatives à un client à l'ensemble des clients ou à un groupe de clients similaires.
  - Déploiement en ligne.
- Lien du dashboard: https://loan-management-dash.herokuapp.com/
  - Réalisé avec Dash
  - Déployé sur Heroku

Présentation du dashboard



Présentation du dashboard



Limites du modèle et améliorations

Limites du modèle et améliorations

- Limites du modèle:
  - Population des demandeurs de prêts ≠ population d'entraînement du modèle ➔ modèle qui ne peut se suffire à lui-même
  - Données provenant de différentes sources ➔ règles d'acceptation de crédit différentes
  - Fonction coût qui repose sur l'observation empirique du nombre d'individus en difficulté de paiement dans notre jeu de données
  - Fonction coût calculée sur l'approximation d'un montant moyen par client
  - Fonction coût = rentabilité uniquement
- Amélioration du dashboard:
  - Prévoir espace de simulation qui permettrait de jouer avec les features client pour modifier dynamiquement le score

Merci de votre attention

Annexe