



Soutenance du Projet 7
Parcours : Data Scientist
Open Classroom, Septembre 2022

Implémentation d'un modèle de scoring

Réalisé par **Habibatou TOGOLA**

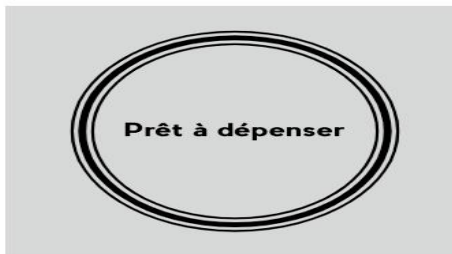


Sommaire



Contexte

Entreprise

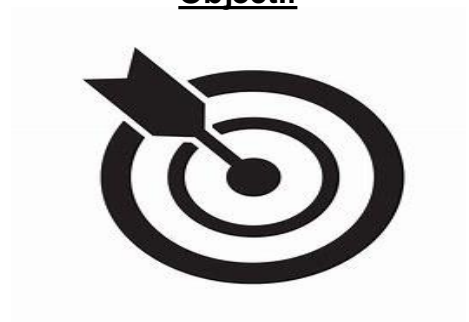


Société financière proposant des crédits à la consommation

Problématique spécifique

- Problème de classification binaire avec des **classes déséquilibrées**
- Limiter **les pertes** engendrées par les défauts de paiement

Objectif



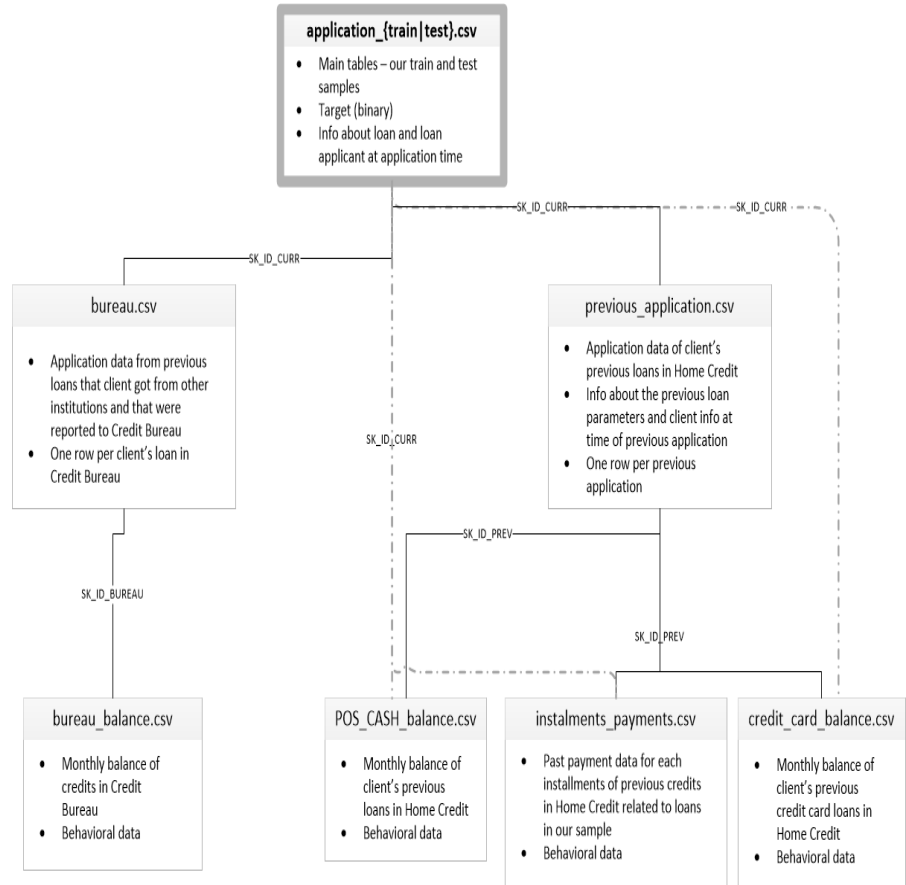
- Développer un modèle de scoring de la probabilité de défaut de paiement.
- Etayer la décision d'accorder ou non un prêt à un client potentiel.
- Construire un Dashboard interactif à destination des gestionnaire de la relation client.

Travail Réalisé : Les données

Les données sont issu d'un dataset sur kaggle:

[Home Credit Default Risk | Kaggle](#)

- 7 sources de données fournissant des informations relative aux clients et à la société
- 307511 clients
- Variable cible: **Target** (0: credit accordé, 1: credit rejeté)



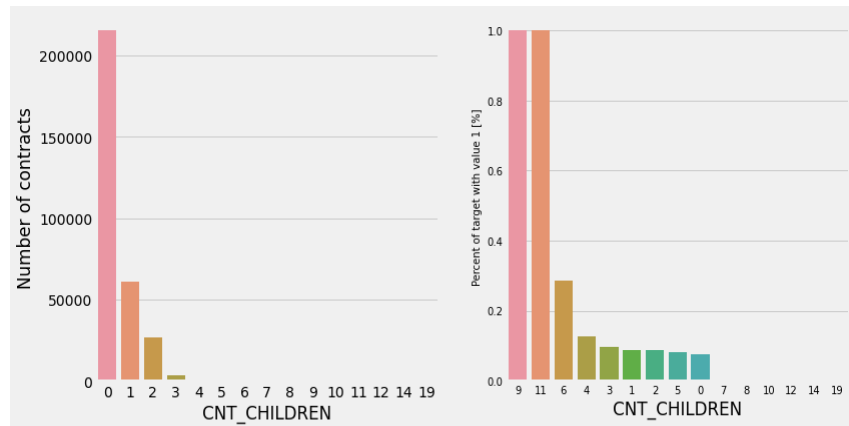
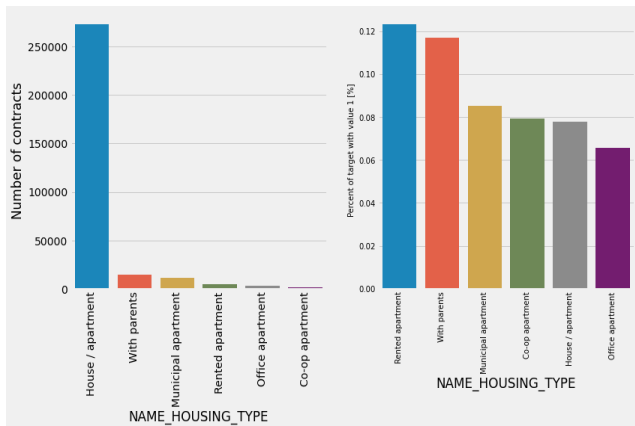
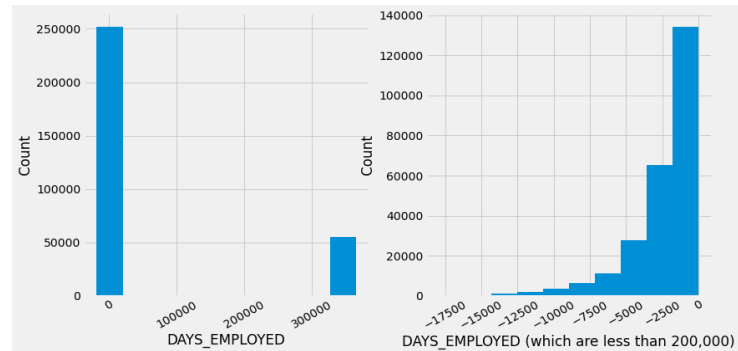
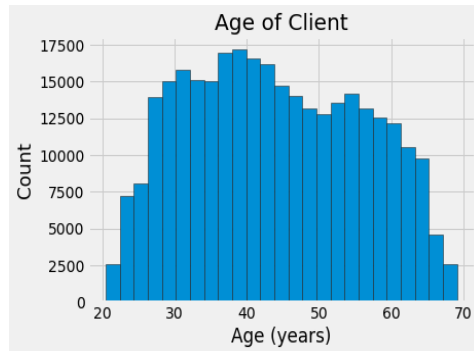
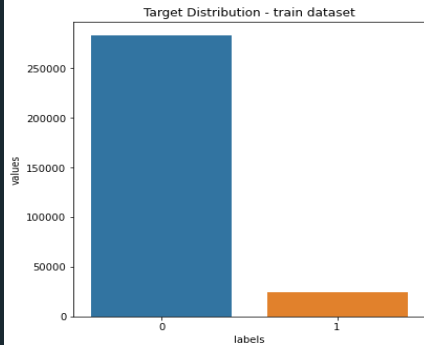
Travail Réalisé : Traitement des données et Feature engineering

- Traitement des valeurs aberrantes
- One Hot Encoder pour les variables catégorielles
- Construction de variables agrégées (min, max, moyenne, somme,...)
- Création de nouvelles variables interprétables
- Sélection de variables pertinentes à l'analyse

Les étapes de traitement et de feature engineering ont été réalisé à partir de ce kernel kaggle:

[LightGBM with Simple Features | Kaggle](#)

Travail Réalisé : Analyse exploratoire



Travail Réalisé : Modélisation

Preprocessing

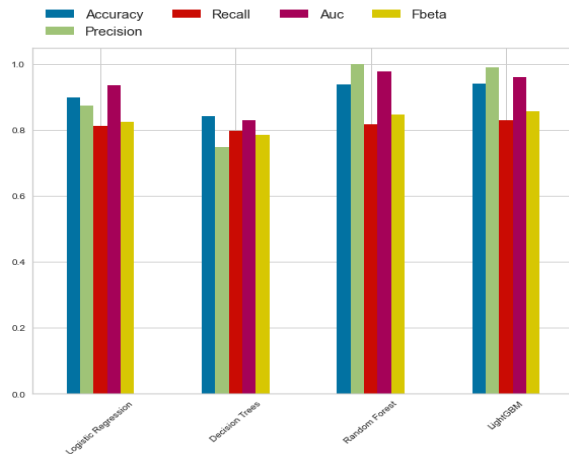
- Traitement des nan et valeurs infinis
- Séparation des données en train et test
- Gestion du déséquilibre: sur-échantillonnage, sous-échantillonnage
- Choix de la méthode de sous-échantillonnage: SMOTE

Modélisation

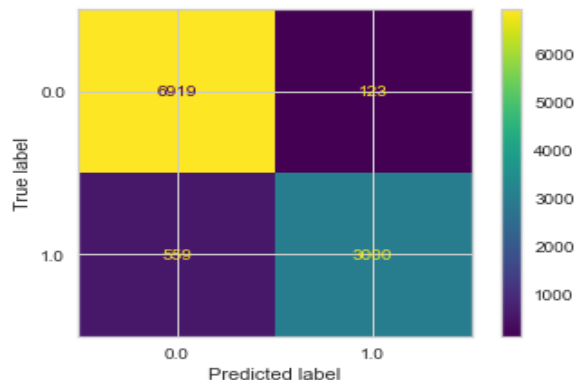
- Essai de 4 modèles avec leur paramètres par défauts
- Evaluation des métriques suivantes: Accuracy, Precision, Recall, Auc
- Choix du meilleur modèle
- Optimisation des hyperparamètres du modèle final
- Evaluation du modèle avec les hyperparamètres trouvés



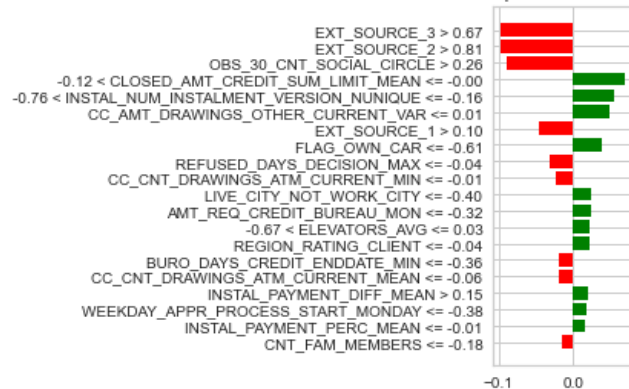
	Accuracy	Precision	Recall	Auc	Fbeta
Logistic Regression	0.897651	0.873490	0.812869	0.935272	0.824310
Decision Trees	0.841053	0.747099	0.796010	0.829914	0.785722
Random Forest	0.938119	0.999312	0.816241	0.975656	0.847285
LightGBM	0.939345	0.988279	0.829165	0.959710	0.856753



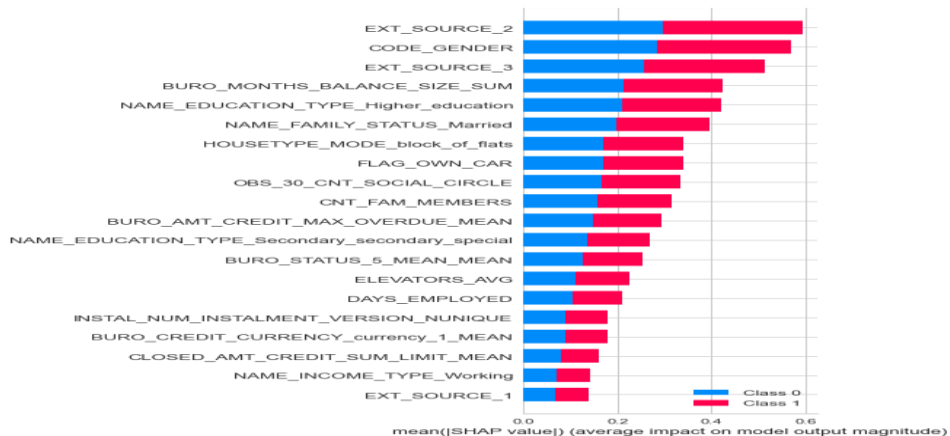
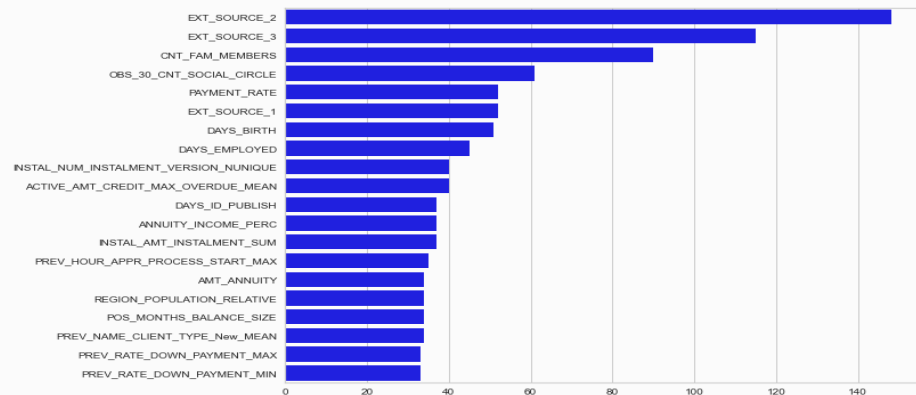
Travail Réalisé : Interprétabilité



Local explanation for class Bad



Feature Importance: LightGBM Stroke Prediction



Travail Réalisé: API et Dashboard

Lien Github : [habibatoutogola/p7scoring3108 \(github.com\)](https://github.com/habibatoutogola/p7scoring3108)

Lien Heroku : <https://p7api.herokuapp.com>

Lien Streamlit : [Streamlit \(streamlitapp.com\)](https://streamlitapp.com)

Bilan

❑ Bilan du projet:

- Construire un modèle de scoring
- Construire un Dashboard interactif

❑ Alternatives envisageables:

- Choix du métrique d'évaluation avec les équipes métier
- Sélection de variable et feature engineering en collaboration avec les équipes métier
- Enrichir les informations descriptives des clients en fonction des retours des utilisateurs du Dashboard

❑ Bilan personnel

- Savoir utiliser Git/Github
- Savoir Déployer un API sur Heroku
- Savoir Créer un Dashboard sur Streamlit