

28 Février 2022

IMPLÉMENTEZ UN MODÈLE DE SCORING

Projet n° 7 Openclassrooms

OPENCLASSROOMS

Présenté par Louisy-Louis Matthieu

Sommaire

- I. Présentation de la problématique
- II. Présentation des données
- III. Modélisation
- IV. Présentation du Dashboard
- V. Questions/Réponses
- VI. Bilan



The background of the slide features a blue ballpoint pen resting on a document. The document has a grid pattern and a bar chart with several blue bars of varying heights. The overall color scheme is a muted blue and grey.

PRÉSENTATION DE LA PROBLÉMATIQUE

Description du projet et des objectifs

Création d'un outil de scoring crédit



- **Développer un algorithme de classification**
 - A partir de données diverses, classe un client en fonction de sa solvabilité (crédit accordé/refusé)
- **Développer un Dashboard interactif**
 - Simplifier l'utilisation du modèle pour une personne non initiée.
 - Permettre au chargé de clientèle de pouvoir expliquer de façon plus transparente les décisions d'octroi de crédit.

Outils utilisés

- Anaconda
 - *Pandas*
 - *ScikitLearn*
 - *Seaborn*
 - *LightGBM*
 - *Pickle*
 - *Shap*
- Git/Github
- Plotly
- Heroku



ANACONDA®



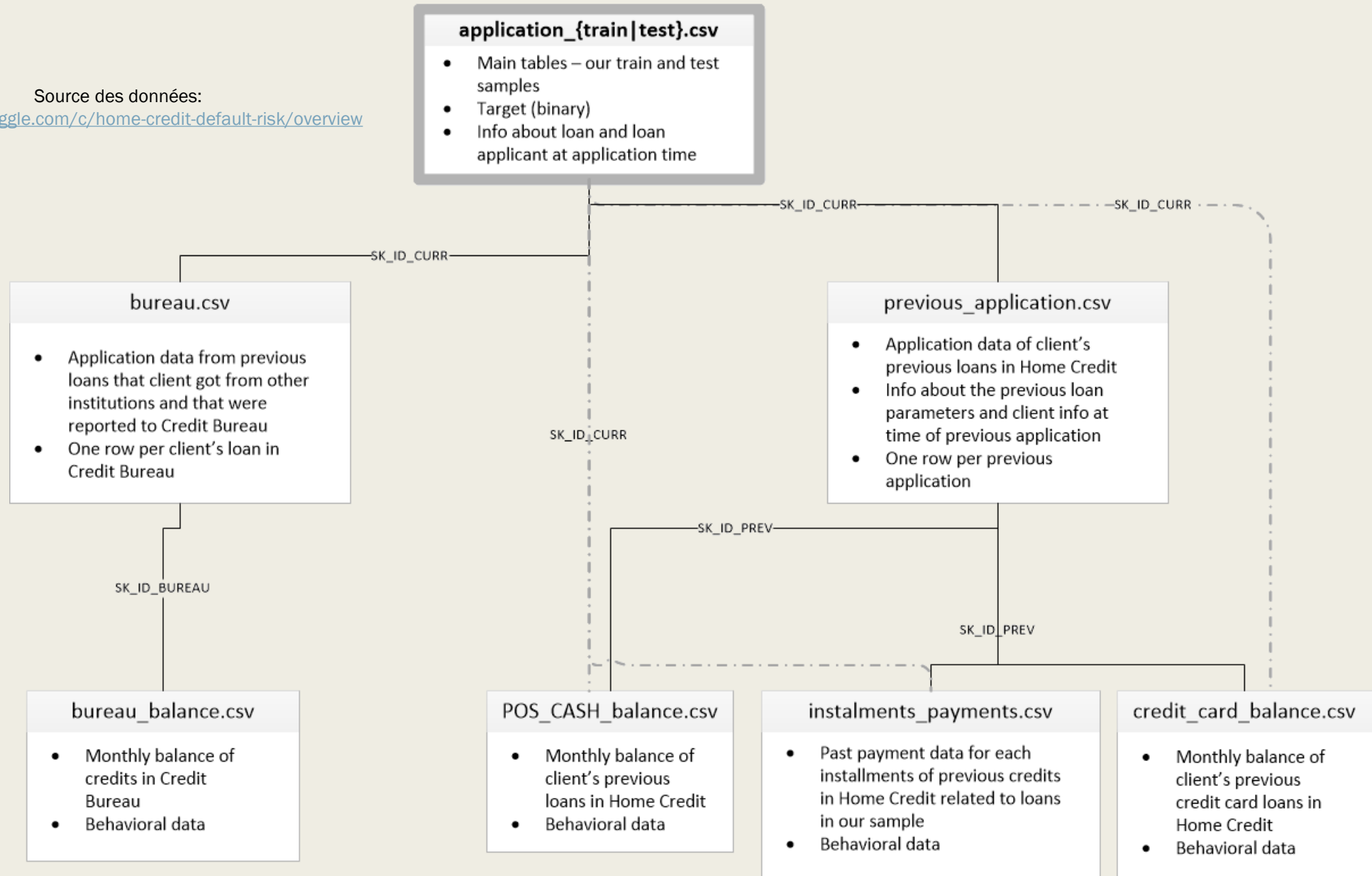


PRÉSENTATION DES DONNÉES

Structure des données

Source des données:

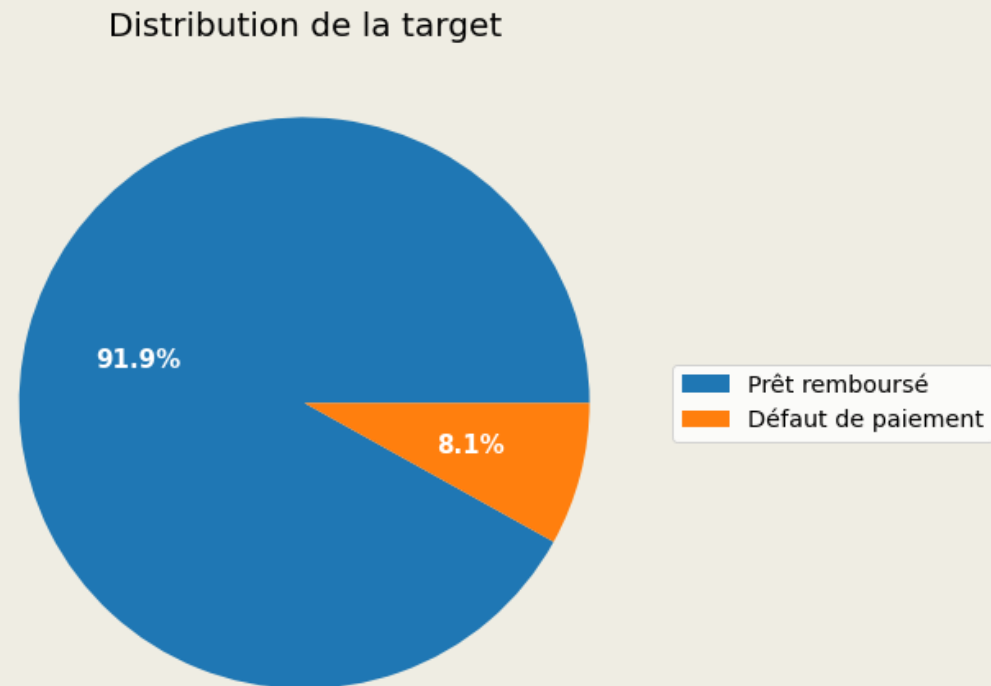
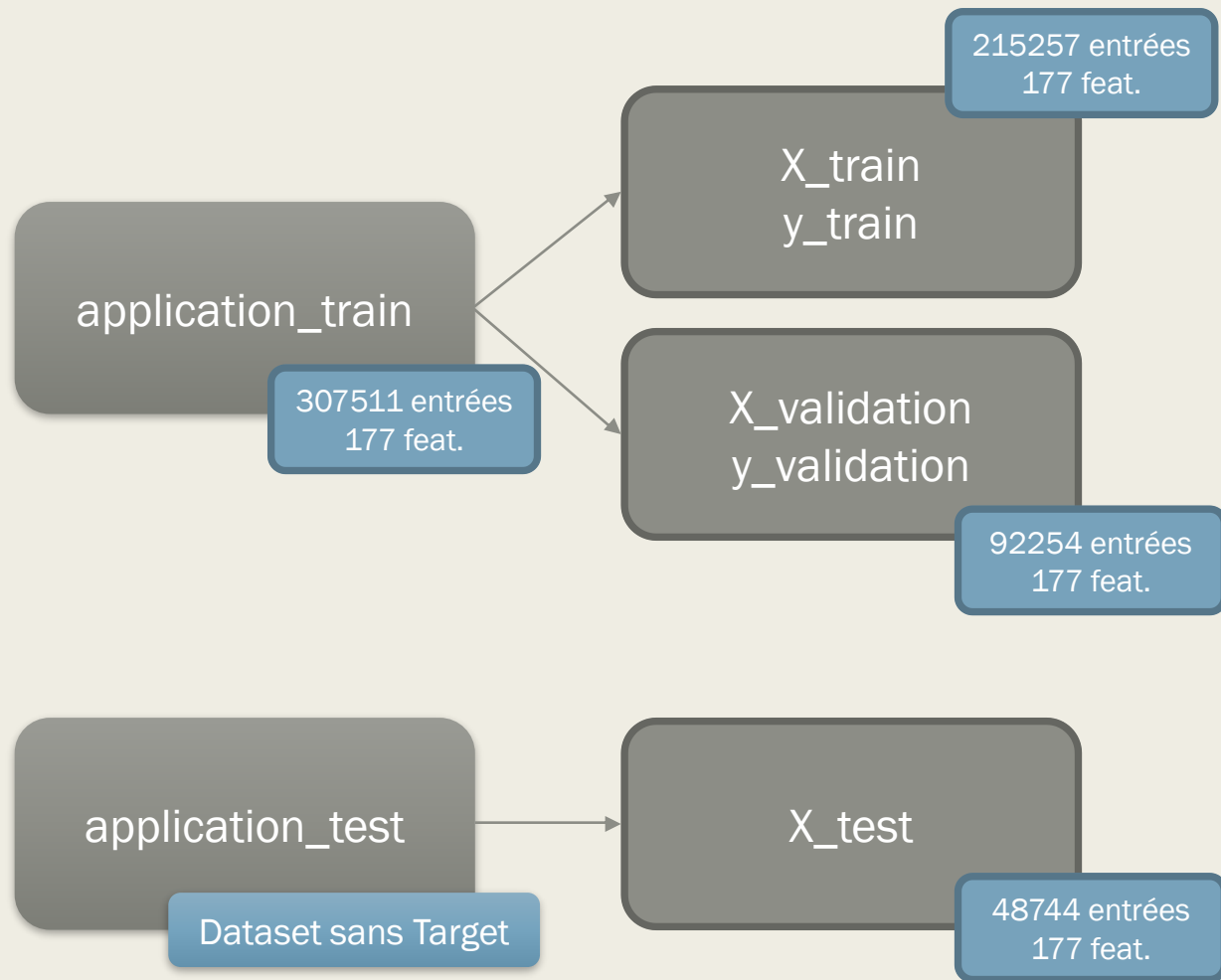
<https://www.kaggle.com/c/home-credit-default-risk/overview>



Feature engineering

- Agrégation sur les données du « bureau »
- Ajout de nouvelles features (par ex. Taux d'endettement = $\text{Anuité} / \text{Income} * 100$)
- Suppression des features trop corrélées avec la target
- Réduction des catégories (Organization_type)

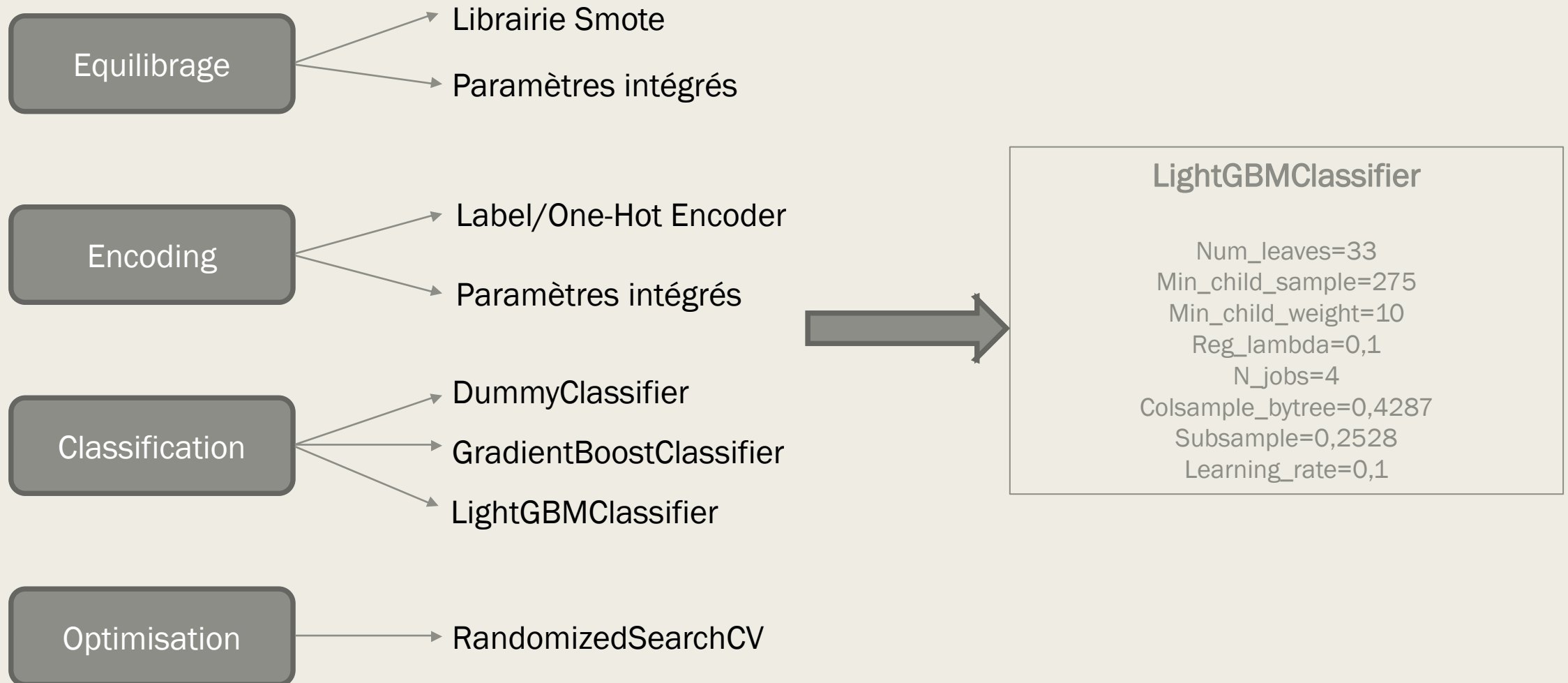
Split des données et équilibrage





MODÉLISATION

Choix du modèle

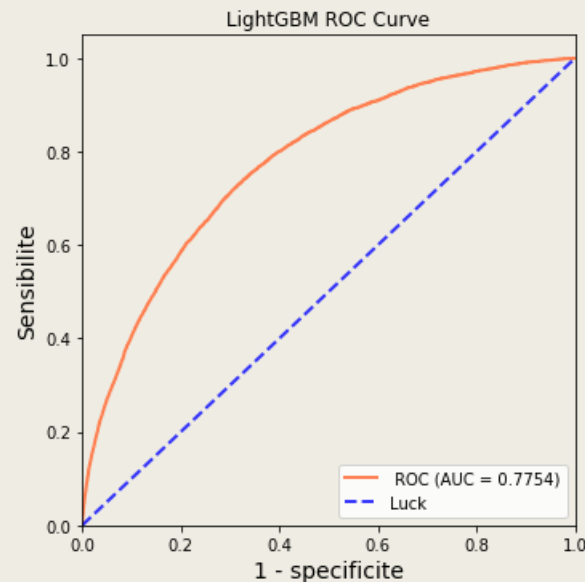


Métriques

Métriques simple pour classifieur

- **Accuracy:** Précision du modèle
- **Precision:** Proportion de prédictions correctes parmi la classe 1
- **Recall:** Proportion de classe 1 correctement prédite
- **F1-Score:** Moyenne harmonique de *precision* et *recall*

Un classifieur naïf qui ne prédirait que la classe 0 obtiendrait une accuracy de 92%

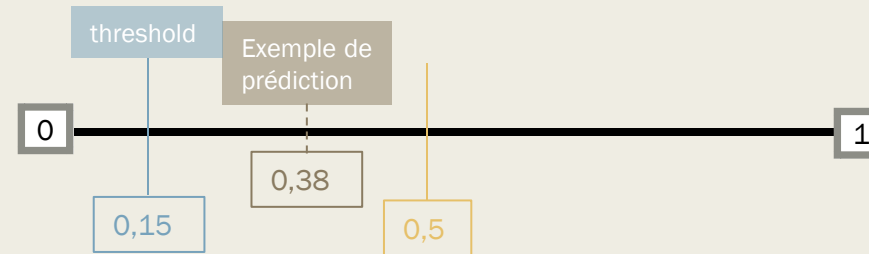


Utilisation de la courbe ROC (Receiver Operation Characteristic) et du score AUC (Area Under Curve) pour l'évaluation du modèle.

Problématique métier

Utilisation de la fonction predict pour le classifieur \longrightarrow seuil de 0,5
Utilisation de la fonction predict_proba pour le classifieur \longrightarrow seuil ajustable

Threshold: Valeur choisie pour déterminer si une valeur prédite appartient à la classe 0 (solvable) ou la classe 1 (non-solvable)



Un client non-solvable est une perte sèche pour l'entreprise

Importance de bien détecté les clients non-solvable (classe 1)

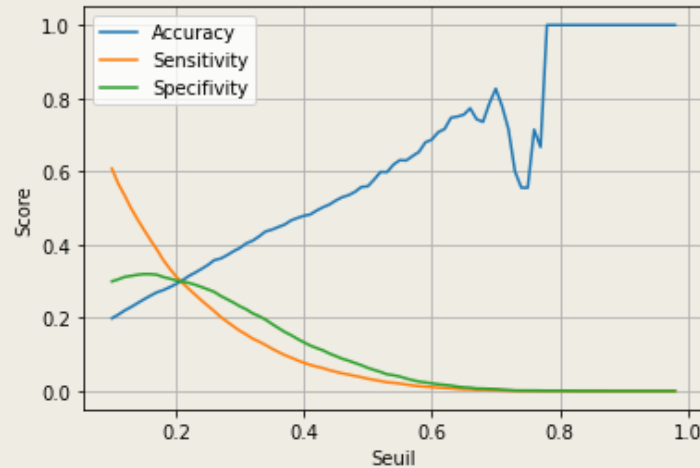
Limitation du nombre de Faux Négatif

Un client peu à risque qui n'obtient pas de prêt est une perte de profit potentiel

Eviter de classer inutilement des clients peu à risque (classe 0) comme étant non-solvable

Limitation du nombre de Faux Positif

Problématique métier



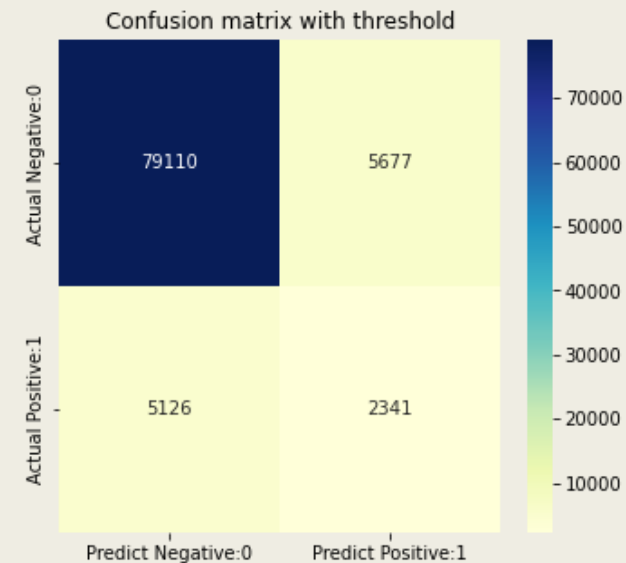
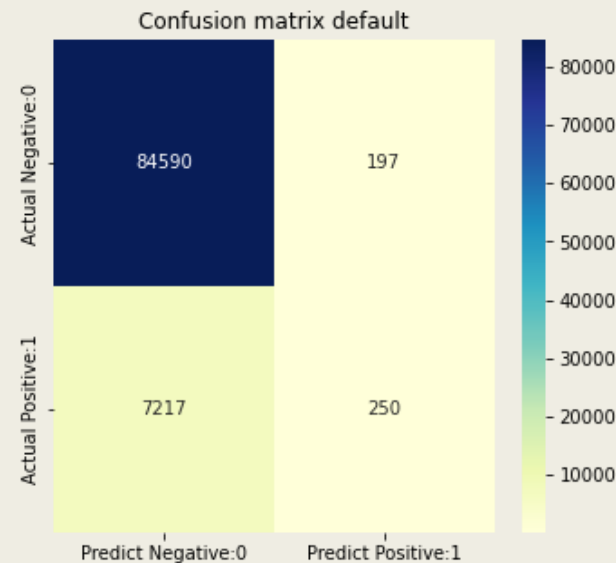
$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{FP+TN}$$

		Classe réelle	
		-	+
Classe prédite	-	True Negatives (vrais négatifs)	False Negatives (faux négatifs)
	+	False Positives (faux positifs)	True Positives (vrais positifs)

Seuil fixé à 0,2

Seuil = 0,5 →



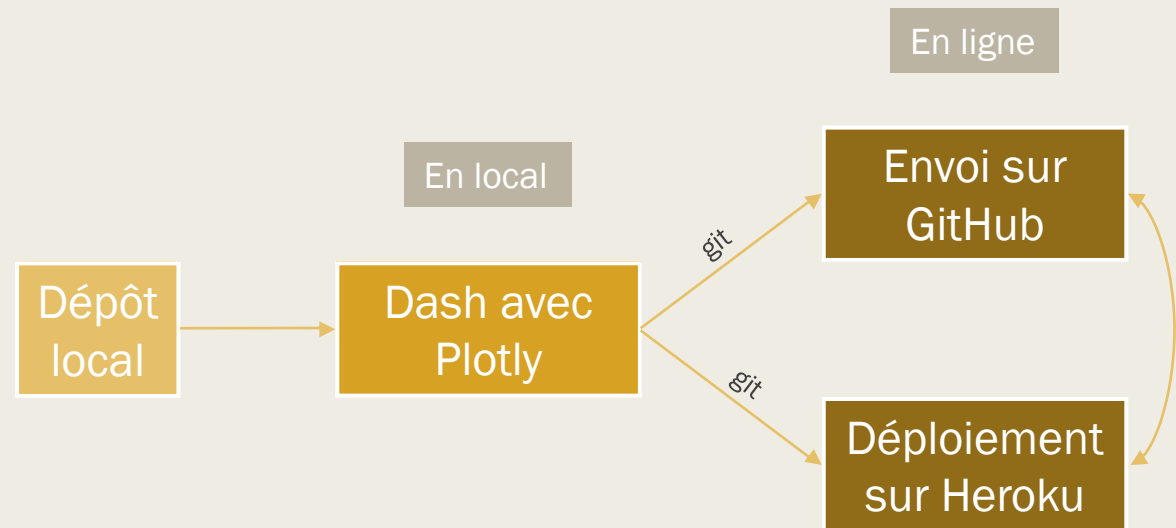
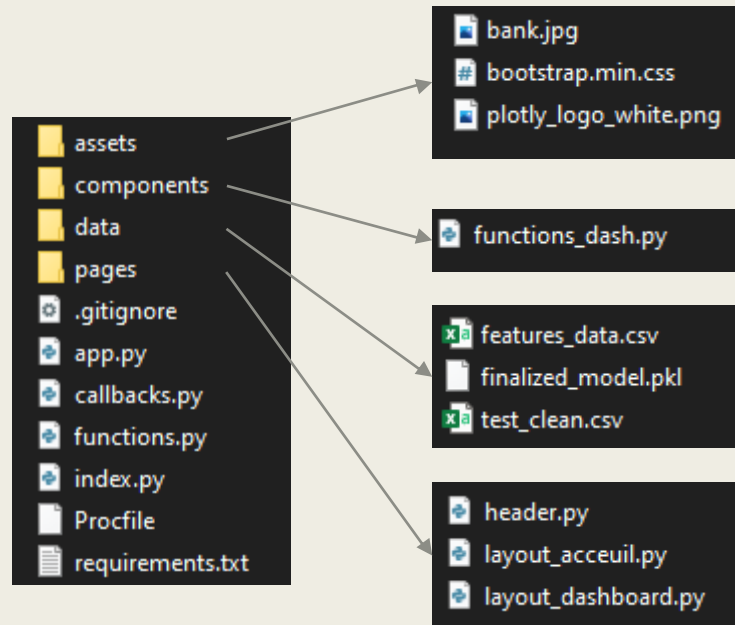
← Seuil = 0,2

PRÉSENTATION DU DASHBOARD

The background is a blurred image of a dashboard. It features several data visualization elements: a line chart with multiple data series in the upper left, a large line chart with a y-axis ranging from 110 to 170 and an x-axis with days of the week (Sun, Mon, Tue, Wed, Thu, Fri, Sat) in the lower left, a bar chart with blue and orange bars at the bottom, and a 3D pie chart in the bottom right. A table with numerical data is visible on the right side. The overall color scheme is dark with blue and orange accents.

Structure

Arborescence du dashboard



Lien GitHub: https://github.com/MatthieuLL/P7_OC_Scoring_Credit



Lien Heroku: <https://p07-oc-scoring-credit.herokuapp.com/>



Démonstration





QUESTIONS/RÉPONSES

Jeu de rôle – Mise en situation





BILAN

Améliorations possibles

- Le prétraitement des données n'a pas fait l'objet d'une attention particulière dans ce projet.
- Le modèle pourrait être affiné: la solution utilisée a pour but d'être simple d'utilisation et rapide en exécution.
- L'utilisation de la version gratuite d'Heroku limite les performances de l'application, qui pourrait être encore plus complète.

Classification
LightGBM relativement efficace



Réalisation du dashboard avec
les principales informations



FIN DE LA SOUTENANCE