

**OpenClassrooms**

**Projet P7:**

**Implémentez un modèle de scoring**

N. Alves

# Implémentez un modèle de scoring

## Sommaire

- Contexte
- Présentation des informations disponibles
- Modélisation
- Tableau de bord (dashboard)
- Conclusion et perspectives

# Implémentez un modèle de scoring

- Contexte
- Présentation des informations disponibles
- Modélisation
- Tableau de bord
- Conclusion et perspectives

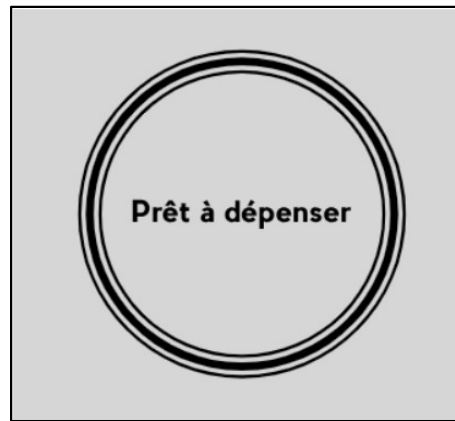
# Implémentez un modèle de scoring

## Contexte

La société financière « Prêt à dépenser » propose des crédits à la consommation. Elle souhaite évaluer le risque crédit : la probabilité qu'un client ne rembourse pas son crédit.

⇒ Conception d'une modélisation du risque client,

⇒ Développement d'un Tableau de bord



Prêt à dépenser

# Implémentez un modèle de scoring

- Contexte
- **Présentation des informations disponibles**
- Modélisation
- Tableau de bord
- Conclusion et perspectives

# Implémentez un modèle de scoring

## Présentation des informations disponibles

Les informations disponibles sont issues de la plateforme Kaggle:

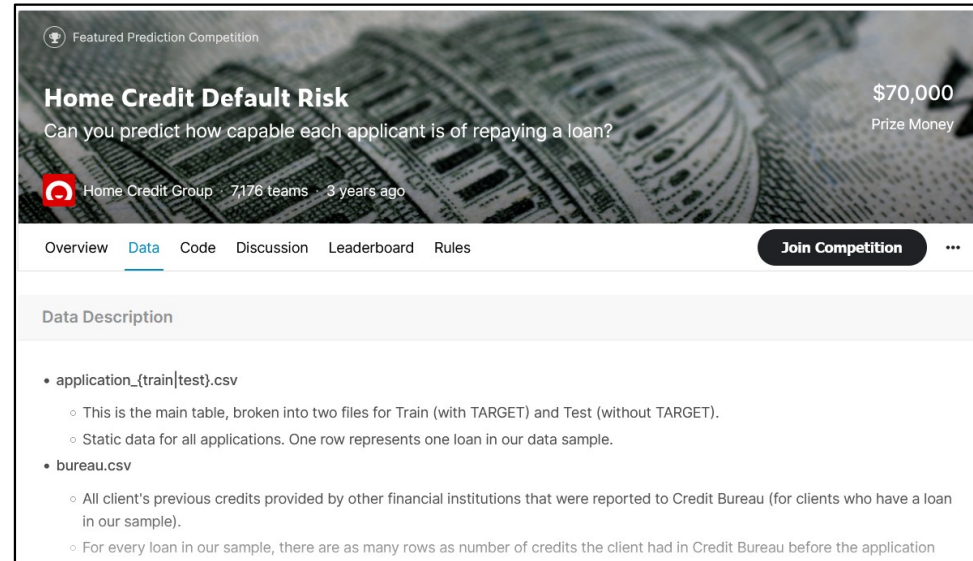
- Données : fichiers au format csv,
- Notebooks de compétiteurs.

Ligne → un client

Colonne → caractéristiques clients :

- Identifiant client,
- montant du crédit,
- ressources du client, etc.

⇒ ~300 000 clients, 250 variables.



La plateforme kaggle

# Implémentez un modèle de scoring

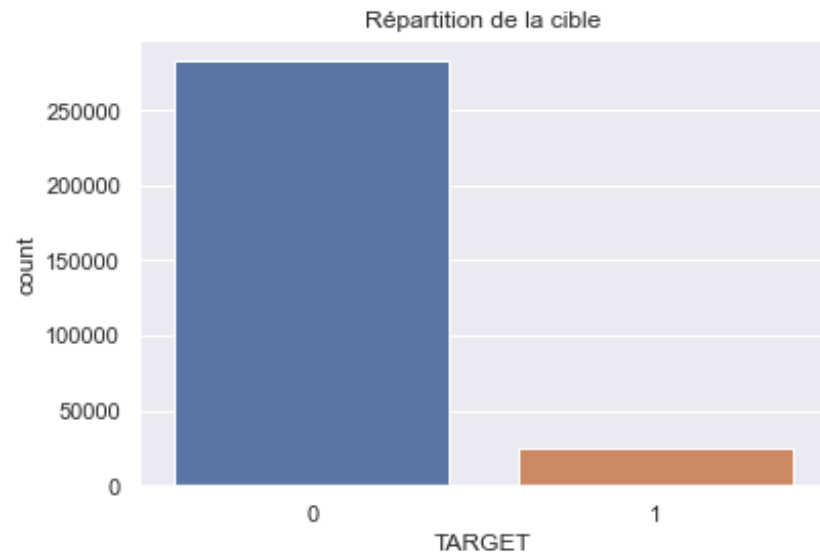
- Contexte
- Présentation des informations disponibles
- **Modélisation**
- Tableau de bord
- Conclusion et perspectives

# Implémentez un modèle de scoring

## Modélisation

### L'entraînement :

- Jeu d'entraînement et de test (80% / 20%)
- Distribution déséquilibré (92 % / 8 %):  
⇒ utilisation de la librairie smote
- Optimisation d'hyperparamètres:  
⇒ nombre d'estimateurs, de feuilles, de profondeur  
et paramètre pour l'équilibrage de données.





# Implémentez un modèle de scoring

## Modélisation

### Le modèle:

- Problème de classification binaire : probabilité de défaut client
- Algorithme ensembliste: Lightgbm
- Métrique : matrice de confusion , accuracy, précision, rappel.

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

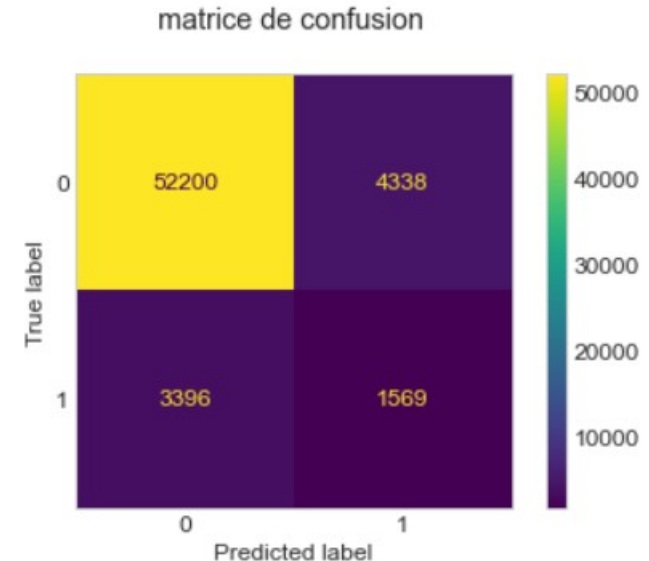
Matrice de confusion

# Implémentez un modèle de scoring

## Modélisation

Le modèle, résultats:

- accuracy  $\sim 0,75$
- précision  $\sim 30\%$
- rappel (sensibilité)  $\sim 30\%$

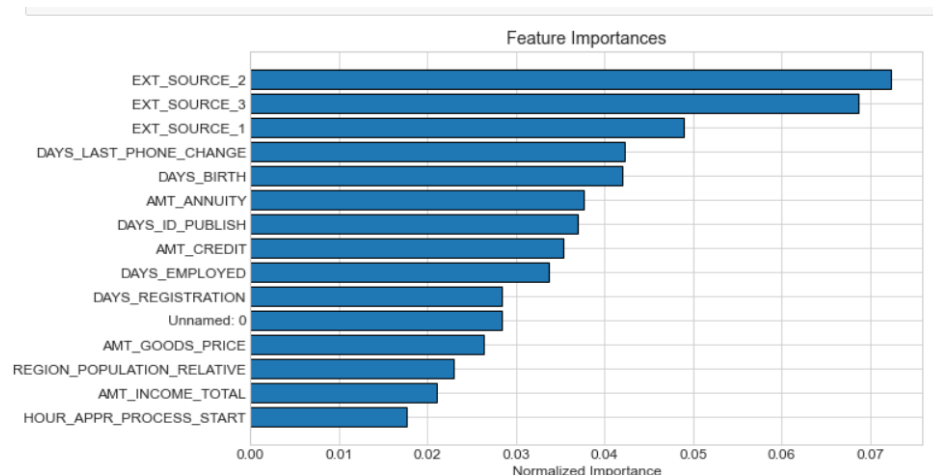


# Implémentez un modèle de scoring

## Modélisation

### Les variables les plus contributrices retenues:

- score provenant d'un calcul externe (EXT\_SOURCE)
- annuité (AMT\_ANNUITY)
- montant du crédit (AMT\_CREDIT)
- nombre de jours en emploi (DAYS\_EMPLOYED)
- montant du bien (AMT\_GOODS\_PRICE)
- ressources du client (AMT\_INCOME\_TOTAL)



# Implémentez un modèle de scoring

- Contexte
- Présentation des informations disponibles
- Modélisation
- **Tableau de bord**
- Conclusion et perspectives

# Implémentez un modèle de scoring

## Tableau de bord



### API

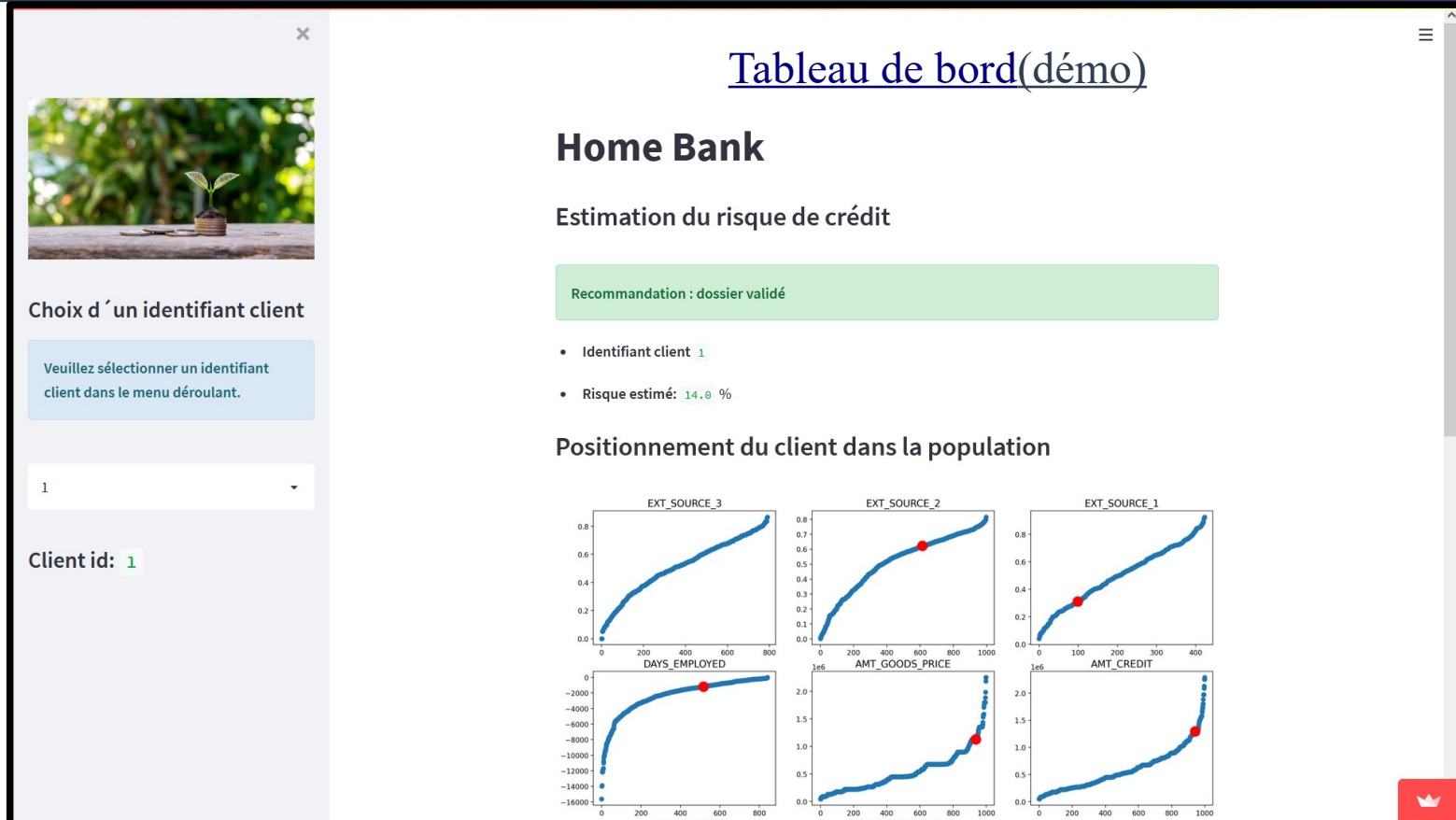
- Modèle encapsulé dans une API.
- Langages : Python, Flask
- Requête avec une méthode 'get' à une url, retour des données au format json.
- Stockage sur la plateforme Heroku



### Tableau de bord :

- Langage/Outil:Python, Streamlit, Matplotlib
- Sélection d'un client, requête de l'api, affichage des résultats
- Stockage sur la plateforme Streamlit

# Implémentez un modèle de scoring



# Implémentez un modèle de scoring

- Contexte
- Présentation des informations disponibles
- Modélisation
- Tableau de bord
- Conclusion et perspectives

# Implémentez un modèle de scoring

## Conclusion et perspectives

**Tableau de bord** : Maquette Heroku et Streamlit  $\Rightarrow$  BI, python/dash ou html, css, javascript, d3js, php/sql etc.

### **Modélisation :**

- Rappel de 30 % et 75 % d'accuracy
- Difficulté : le feature engineering
- Pistes d'améliorations :
  - $\Rightarrow$  notebook kaggle
  - $\Rightarrow$  modèle ensembliste ou réseaux de neurones
  - $\Rightarrow$  retour métiers (technique et éthique)





