

Rapport

Credit-risque-project

1. Quelles sont les variables les plus influentes pour prédire le risque de crédit ?

L'analyse de l'importance des variables selon le modèle Random Forest montre que plusieurs facteurs jouent un rôle déterminant dans la prédiction du risque de crédit.

Les 5 variables les plus influentes

1. **loan_percent_income**
→ Rapport entre montant du prêt et revenu du client.
Plus il est élevé, plus le risque de défaut augmente.
2. **loan_int_rate**
→ Les prêts à taux élevés sont plus souvent associés à des clients risqués.
3. **person_income**
→ Un revenu plus faible augmente la probabilité de défaut.
4. **cb_person_default_on_file**
→ Historique de défaut : très fort indicateur d'un risque futur.
5. **loan_amnt**
→ Montant du prêt demandé : plus il est important, plus l'exposition au risque est élevée.

Ces variables sont cohérentes avec la logique bancaire : le poids du prêt dans le revenu, l'historique de crédit et les conditions du prêt expliquent fortement le comportement de remboursement.

2. Quel modèle est le plus performant et pourquoi ?

Deux modèles ont été entraînés :

Régression Logistique

Random Forest

- Le **Random Forest** est le modèle le plus performant, car :
- il réduit mieux les erreurs de classification,
 - il équilibre les classes grâce au paramètre *class_weight='balanced'*,
 - il capture des patterns complexes que la régression logistique ne peut pas modéliser.

Interprétation de la matrice de confusion :

La matrice de confusion obtenue pour le modèle Random Forest indique :

- **5059 vrais négatifs** : bons clients correctement identifiés
- **1017 vrais positifs** : clients risqués correctement identifiés
- **36 faux positifs** : la banque refuse un bon client
- **405 faux négatifs** : la banque accepte un client risqué

3. Quel type d'erreur est le plus coûteux ?

Les **faux négatifs** sont les erreurs les plus dangereuses pour une banque :

Le modèle prédit que le client n'est pas risqué alors qu'il ne remboursera pas.

→ Cela entraîne **une perte financière directe**, car l'argent prêté risque de ne jamais être récupéré.

Les faux positifs sont moins graves :

→ on refuse un bon client, ce qui génère un **manque à gagner**, mais pas une perte.

4. Comment une banque peut-elle utiliser ce modèle ?

Un modèle IA de scoring de crédit peut être utilisé pour :

- **Automatiser les décisions de crédit**
- **Personnaliser les taux d'intérêt**
- **Réduire les impayés**
- **Aider les analystes**

5. Risques liés au modèle

Même performant, un modèle comporte des risques :

- **Les erreurs de classification**
- **Biais dans les données**
- **Sur confiance dans le modèle**