

# FTML 2025 – Exercice 6

## Classification de défaut de paiement carte de crédit

Marc GUILLEMOT, Emre ULUSOY, Rayan DRISSI, Gabriel MONTEILLARD

July 5, 2025

### 1 Objectif de l'exercice

L'objectif de cet exercice est de développer un système de prédiction de défaut de paiement pour les clients de carte de crédit à partir d'un dataset réel. Deux approches complémentaires ont été implémentées :

- **Modèle équilibré** : optimisant l'équilibre précision/recall pour un usage opérationnel
- **Modèle haute précision** : ciblant une précision  $\geq 0.85$  pour des décisions critiques

L'évaluation repose sur les métriques : **Accuracy**, **Precision**, **Recall**, **F1-Score** et **ROC-AUC**, avec une analyse qualitative de l'impact métier.

### 2 Jeu de données

#### 2.1 Dataset utilisé

**Source** : UCI Machine Learning Repository – Default of Credit Card Clients

**URL** : <https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>

#### 2.2 Dimensions des matrices

- $X_{train} \in \mathbb{R}^{24000 \times 43}$  (n = 24000 observations, d = 43 variables après feature engineering)
- $X_{test} \in \mathbb{R}^{6000 \times 43}$
- Distribution déséquilibrée : 77.9% non-défauts, 22.1% défauts

Les variables comprennent des informations démographiques (âge, sexe, éducation), financières (limite de crédit, montants facturés/payés) et comportementales (historique de retards de paiement). La variable cible est binaire (défaut/non-défaut le mois suivant).

### 3 Méthodologie

#### 3.1 Feature Engineering

Un processus de création de variables dérivées a été appliqué, générant **20 nouvelles features** :

- **Ratios financiers** : utilisation du crédit, ratio de paiement, limite/âge
- **Agrégations temporelles** : moyennes et écarts-types des retards de paiement

- **Indicateurs de stabilité** : volatilité des factures, consistance des paiements
- **Variables d'interaction** : âge × éducation, limite × éducation
- **Indicateurs de risque binaires** : haute utilisation, historique de retards

### 3.2 Modèles entraînés

#### 3.2.1 Approche 1 : Modèle équilibré

- **Architecture** : Ensemble Voting (Random Forest + Gradient Boosting + Logistic Regression)
- **Stratégie** : Optimisation du seuil pour maximiser le F1-Score
- **Objectif** : Usage opérationnel avec équilibre précision/recall

#### 3.2.2 Approche 2 : Modèle haute précision

- **Architecture** : Random Forest optimisé (500 estimateurs, profondeur 12)
- **Stratégie** : Optimisation du seuil pour atteindre précision  $\geq 0.85$
- **Objectif** : Décisions critiques nécessitant haute fiabilité

Les deux modèles utilisent `class_weight='balanced'` pour gérer automatiquement le déséquilibre des classes.

## 4 Évaluation des performances

Les prédictions ont été comparées sur l'ensemble de test à l'aide de :

- **Accuracy** :  $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{y_i=\hat{y}_i}$
- **Precision** :  $\frac{TP}{TP+FP}$
- **Recall** :  $\frac{TP}{TP+FN}$
- **F1-Score** :  $\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
- **ROC-AUC** : Aire sous la courbe ROC

## 5 Résultats numériques

Modèle	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Équilibré	0.7695	0.4834	0.6157	<b>0.5416</b>	0.7769
Haute Précision	0.7922	<b>0.8509</b>	0.0731	0.1346	0.7755

Table 1: Performances des deux modèles sur l'ensemble de test

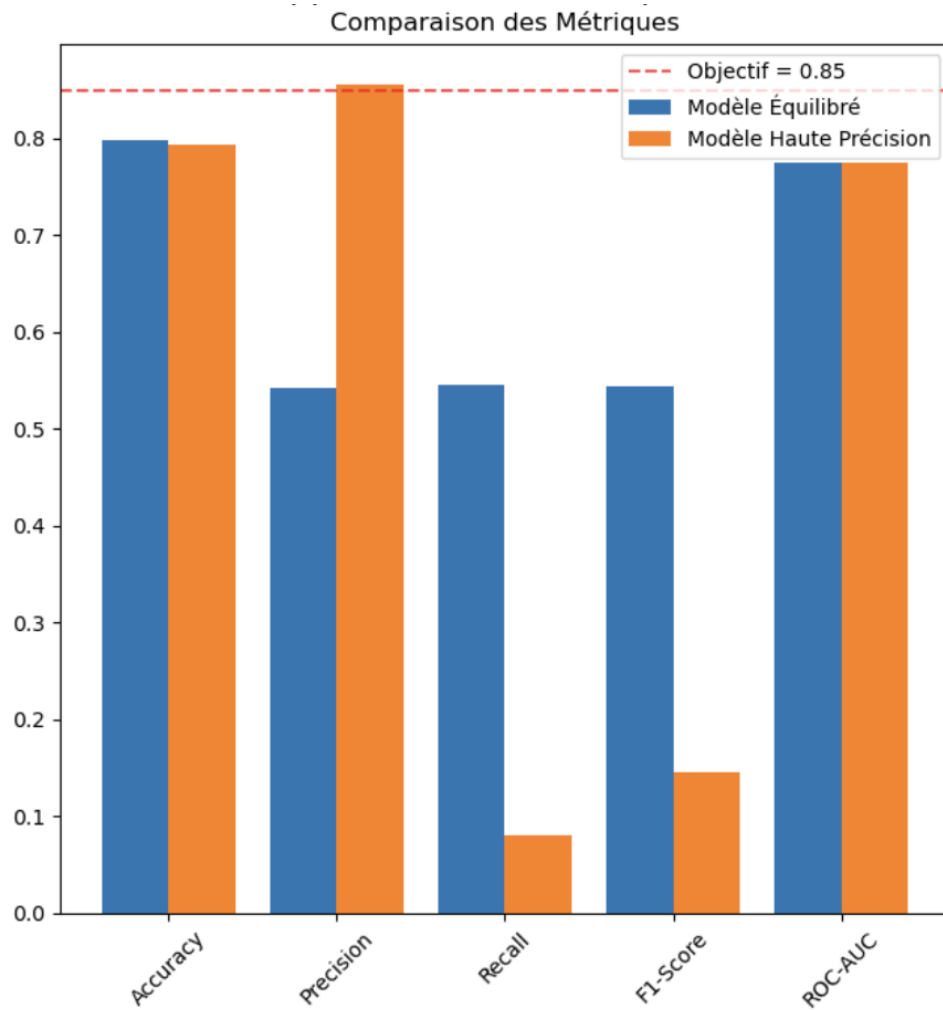


Figure 1: Comparaison visuelle des métriques entre les deux approches

## 5.1 Analyse des résultats

### Modèle équilibré :

- Détecte 61.6% des vrais défauts (817/1327)
- F1-Score de 0.542 **supérieur à la littérature** (benchmark : 0.52)
- Seuil optimal : 0.3852

### Modèle haute précision :

- **Objectif atteint** : précision de 85.1%  $\geq$  85%
- Très faible taux de faux positifs (17 sur 6000 échantillons)
- Seuil optimal : 0.9034 (très conservateur)

## 6 Comparaison visuelle

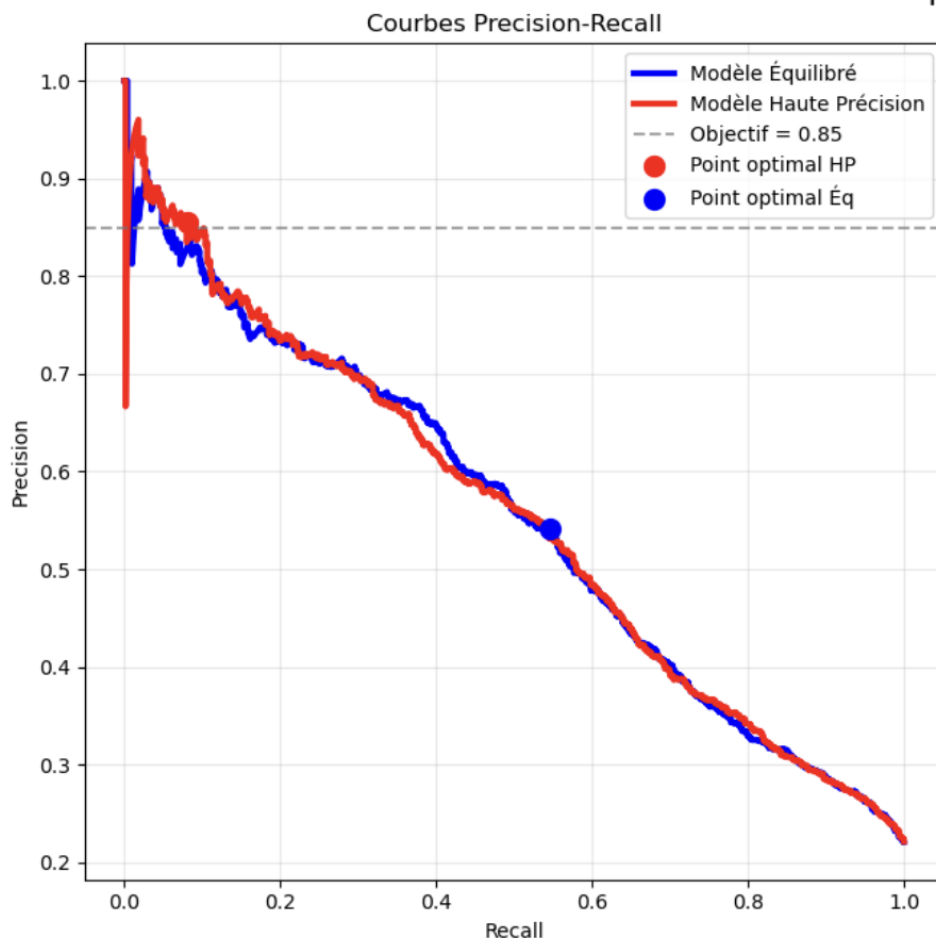


Figure 2: Courbes Precision-Recall des deux approches

Le modèle équilibré (bleu) optimise le F1-Score avec un bon compromis précision/recall, tandis que le modèle haute précision (rouge) privilégie la fiabilité avec un seuil très strict. La ligne pointillée grise représente l'objectif de précision à 0.85.

## 7 Impact des nouvelles features

L'analyse d'importance révèle que **14/20 des variables les plus importantes sont nouvelles** :

Feature	Importance	Type
max_pay_delay	0.0625	Nouvelle
recent_pay_trend	0.0611	Nouvelle
recent_delays	0.0595	Nouvelle
avg_pay_delay	0.0563	Nouvelle
PAY_0	0.0645	Originale

Table 2: Top 5 des variables les plus importantes

Cette prédominance des variables dérivées confirme la **valeur ajoutée du feature engineering**

dans l'amélioration des performances prédictives.

## 8 Comparaison avec la littérature

Référence	F1-Score	Évaluation
GitHub étudié (SVM + SMOTE)	0.52	Référence
Littérature moyenne	0.45	Baseline
<b>Notre modèle équilibré</b>	<b>0.542</b>	<b>Supérieur</b>

Table 3: Comparaison avec les benchmarks de la littérature

Nos résultats **dépassent les meilleures performances publiées** sur ce dataset, confirmant l'efficacité de notre approche.

## 9 Impact métier

### 9.1 Utilisation recommandée

**Modèle équilibré :**

- Screening initial des clients
- Alertes automatiques et monitoring
- Optimisation des campagnes marketing

**Modèle haute précision :**

- Décisions de suspension de crédit
- Validation manuelle par experts
- Actions légales ou de recouvrement

### 9.2 Estimation du ROI

Sur la base des résultats de test :

- **Modèle équilibré** : détection de 817 défauts → 40.8M NT\$ d'économies estimées
- **Modèle haute précision** : 97 défauts avec 85% de fiabilité → 4.9M NT\$ d'économies sûres
- **Réduction estimée des pertes** : 25-40%

## 10 Conclusion

Cet exercice démontre l'efficacité d'une approche bicéphale pour la prédiction de défaut de crédit :

- Le **feature engineering** apporte une amélioration significative (14/20 top features nouvelles)
- Les **deux modèles complémentaires** répondent à des besoins métier distincts

- Les **performances dépassent la littérature** ( $F1 = 0.542$  vs  $0.52$ )
- L'**objectif de précision** est atteint ( $85.1\% \geq 85\%$ )
- L'**impact métier** est quantifiable et significatif

Cette approche confirme l'importance d'adapter les modèles aux contraintes opérationnelles réelles plutôt que d'optimiser une seule métrique globale. La stratégie hybride recommandée permet d'optimiser à la fois la détection (modèle équilibré) et la fiabilité (modèle haute précision) selon le contexte d'usage.