

# Analyse RGPD & Bloc de Compétence 2 — Étude de Cas Néo-Banque

## 🎯 Objectif pédagogique

Valider la **compétence C18** :

Contrôler le respect de la réglementation en matière de protection des données personnelles (RGPD) en établissant une politique de collecte conforme et en présentant les enjeux éthiques de l'IA.

## 🔍 Analyse des données utilisées

📁 Fichier source : `application_train.csv`

Provenant du challenge Home Credit Default Risk (Kaggle)

### 🔒 Analyse RGPD

Élément	Détail	Statut
Données personnelles directement identifiables	Aucune (pas de nom, email, adresse...)	☑ Conforme
Identifiant client ( <code>SK_ID_CURR</code> )	Supprimé dès le chargement dans le pipeline	☑ Conforme
Données sensibles (origine, santé...)	Absentes	☑ Conforme
Finalité du traitement	Scoring d'éligibilité au prêt bancaire	☑ Légitime
Prise de décision automatisée	Oui, mais pas définitive	⚠ À documenter
Explication du score	À mettre en œuvre (SHAP/LIME)	⚠ Recommandé
Minimisation des données	Colonnes inutiles ou vides supprimées	☑ Conforme
Conservation des données	Non définie dans l'étude	⚠ À définir (politique interne)

## ☑ Conformité RGPD mise en place

- 1. **Suppression des identifiants clients** (`SK_ID_CURR`)
- 2. **Nettoyage des variables inutiles ou à forte absence** (> 40%)
- 3. **Aucune variable sensible ou identifiable**
- 4. **Utilisation strictement limitée à la prédiction d'un score d'éligibilité**

## Recommandations RGPD complémentaires (à mettre en œuvre dans l'interface)

### 1. Transparence & Information client

Inclure un encart dans l'application : "Vos données sont analysées de façon pseudonymisée pour évaluer votre éligibilité à un prêt. Aucune décision n'est prise automatiquement."

### 2. Interprétabilité du score

Ajouter un visuel expliquant les facteurs qui influencent la prédiction (via SHAP)

### 3. Conservation & consentement

- Définir une politique claire de conservation : *ex. données conservées 12 mois maximum après décision*
- Optionnel : ajouter un champ de consentement dans l'interface si usage étendu prévu

---

## Justification finale pour le bloc C18

*Les données utilisées ne contiennent aucune information personnelle ou sensible. Les identifiants ont été supprimés, les données inutiles écartées, et la finalité du traitement est limitée à l'analyse d'éligibilité au crédit. Une politique d'explicabilité et de transparence est proposée afin d'être conforme au RGPD et de garantir un usage éthique de l'intelligence artificielle.*