

## Projet IA & ML

Prédire les clients qui vont résilier leur abonnement

**A DEPOSER ICI :** <https://forms.gle/jcFzMj42Wd9WudDq7>

**Date de fin de rendu : 05/05/2025**

-3 points par jour de retard.

**PROJET A FAIRE SEUL OU EN BINOME**

**Contexte :**

Tu travailles pour une entreprise qui propose un service par abonnement (comme Netflix, une salle de sport, ou une appli mobile). Le service marketing a remarqué que certains clients **résilient** leur abonnement, et aimerait pouvoir **anticiper ces départs**.

**Ta mission** : Créer un modèle de Machine Learning capable de **prédire si un client risque de résilier**.

#### Variables du dataset à utiliser

Colonne	Description
Age	Âge du client
Revenu	Revenu mensuel
Sexe	Homme / Femme
Anciennete	Ancienneté en années
Frequence_utilisation	Combien de fois il utilise le service par mois
Support_contacte	A-t-il contacté le support (Oui/Non)
Score_satisfaction	Score de satisfaction de 1 à 10
Résilié	A-t-il résilié ? (0 = Non, 1 = Oui) → <b>c'est la colonne à prédire !</b>

## 1. Comprendre le problème

- Quelles informations peuvent influencer la décision de résilier ?
- Pourquoi est-ce utile pour l'entreprise ?
- Que peut-on faire avec une bonne prédiction ? (ex : envoyer des offres personnalisées, relancer les clients...)

## 2. Explorer et préparer les données

### 1. Télécharge le fichier

### 2. Regarde à quoi ressemblent les données

- Combien de clients ?
- Quelles sont les colonnes ?

### 3. Nettoie les données

- Y a-t-il des valeurs manquantes ?
- Faut-il normaliser certaines colonnes ?
- Doit-on encoder des variables catégorielles ? (ex : "Oui/Non" → 1/0)

### 4. Fais quelques visualisations :

- Histogrammes des âges, du revenu
- Corrélation entre satisfaction et résiliation ?

## Créer un premier modèle

### 1. Sépare les données :

- **80% pour l'entraînement, 20% pour le test**

### 2. Teste plusieurs modèles :

- Régression logistique
- Arbre de décision
- K-plus proches voisins (KNN)

### 3. Évalue les modèles avec :

- Précision
- Rappel

- F1-score
- Matrice de confusion

Ton objectif : trouver un **modèle équilibré** qui fonctionne bien sans sur-apprendre.

## 4. Améliorer ton modèle

1. Utilise la **validation croisée**
2. Teste différents **paramètres (hyperparamètres)** : profondeur des arbres, nombre de voisins pour KNN...
3. Supprime ou ajoute des colonnes, vois l'impact (feature engineering)
4. Si besoin : applique une réduction de dimensionnalité (PCA)

## 5. Interpréter les résultats

Tu dois **comprendre pourquoi le modèle a pris certaines décisions** :

1. Affiche l'importance des variables :
  - Quelles colonnes influencent le plus la prédiction ?
2. Utilise **SHAP** (ou LIME si tu préfères) pour un client donné :
  - Pourquoi le modèle pense-t-il qu'il va résilier ?

## 6. Présente ton travail

Imagine que tu présentes ton projet à un manager marketing

1. Prépare une **slide ou un rapport simple** :
  - Résume ce que tu as fait
  - Montre des visuels
  - Explique ce que le modèle apprend
  - Donne une **recommandation** : "on pourrait cibler les clients ayant moins de 2 ans d'ancienneté + une note de satisfaction < 5"

Option bonus : crée une **mini appli** avec Streamlit pour prédire et expliquer automatiquement une situation.

# Récapitulatif

## À la fin tu auras appris à :

- Comprendre et formuler un problème de ML réel
- Nettoyer et préparer des données
- Tester plusieurs modèles
- Éviter le sur-apprentissage
- Expliquer une prédiction (interprétabilité)
- Communiquer comme une pro les résultats 🗨️

## Barème de notation (sur 20 points)

Critère	Détail	Points
<b>Compréhension du problème</b>	Reformulation claire, enjeux, impact	2 pts
<b>Préparation des données</b>	Nettoyage, encodage, exploration visuelle pertinente	4 pts
<b>Modélisation</b>	Choix de plusieurs modèles, évaluation avec les bonnes métriques	4 pts
<b>Amélioration &amp; optimisation</b>	Validation croisée, tuning d'hyperparamètres, essais justifiés	3 pts
<b>Interprétation</b>	Importance des variables + SHAP/LIME sur un cas concret	3 pts
<b>Présentation finale</b>	Clarté des résultats, capacité à vulgariser, recommandations	4 pts

Bonus (jusqu'à +4 pts) : mise en forme originale (Streamlit), dashboard interactif, idée innovante