



**PROJET INTER-PROMO 2021-22**



# Généralités

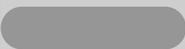
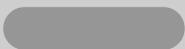
- Les 3 promotions impliquées ( $\approx$  100 étudiants)
- Deux thématiques complémentaires proposées par le Crédit Agricole (Toulouse et Languedoc)
- 9 groupes de 12 étudiants
  - Un enseignant référent par groupe
  - Un chef de groupe + un sous-chef
- Inscription à Moodle (EIMAB4B1) avec clé : 
- Signer l'accord de confidentialité avant de partir !!!
- Questions : à poser directement sur Discord

# Objectifs du premier sujet

- **Détection de clients sensibles à la fraude via phishing**
  - Scénario de fraude : Le fraudeur se fait passer pour le Crédit Agricole pour extirper des informations de connexion aux applications du Crédit Agricole. Une fois connecté sur le compte du client, le fraudeur peut effectuer des opérations via Carte ou Virement.
  - L'idée est de détecter ici les clients sensibles à ce type de fraude.

# Objectifs du premier sujet

- **Volumétrie :**

	Toulouse 31	Languedoc
Nombre clients total		
Nombre clients fraudés		

- **Données :**

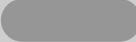
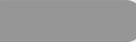
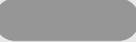
- Données signalétiques : ge, ancienneté, CSP, statut marital, code postal...
- Équipement bancaire
- Habitudes sur les outils digitaux
  - Environ 300 indicateurs au total

# Objectifs du deuxième sujet

- **Contribution des parcours digitaux sur la souscription d'un crédit consommation**
  - Intérêt : Le crédit conso est un axe stratégique majeur et le digital est un levier fort de développement. Grâce au dispositif de tracking de notre site web ainsi qu'aux outils et méthodes du « big data », le Crédit Agricole cherche à éclairer des zones d'ombre :
    - Contribution du digital à la production
    - Identification des leviers du parcours client générant de la production

# Objectifs du deuxième sujet

- **Volumétrie :**

	Toulouse 31	Languedoc
Nombre clients ayant souscrit un crédit conso sur les 6 derniers mois		
Nombre de pages consultées sur 45j (clients reconnus)		

- **Données :**

- La production de crédits conso sur les 6 derniers mois
- Les données signalétiques : ge, ancienneté, CSP, statut marital, code postal, score conso...
- Les données de webtracking sur 45 jours
- Les opportunités de contacts (Emails, SMS, Lead...)
  - Environ 300 indicateurs au total

# Organisation

- **Répartition :**

- Sujet 1 : Groupes 1 à 5
- Sujet 2 : Groupes 6 à 9

- **Objectifs essentiels du PIP 2021-22 pour tous les groupes**

- Interpréter et expliquer les modèles de machine learning utilisés
- Evaluer et comparer les modèles de façon rigoureuse et systématique
- Présenter des visualisations homogènes qui rendent compte de manière claire de tous les résultats (dashboard commun à tous les groupes)

# Groupe 1 : Détection d'anomalie supervisée

- **Composition :**

- Tuteur enseignant : Philippe Berthet
- Chef de groupe : Pauline Dupuy
- Sous-Chef / resp. visualisation : Jules Boutibou

- **Objectifs :**

- Détection des clients sensibles à la fraude par classification supervisée
- Mise en évidence des facteurs de risque de fraude

- **Données : jeu n°1**

# Groupe 2 : Détection d'anomalie non supervisée

- **Composition :**

- Tuteur enseignant : Edouard Pauwels

- Chef de groupe : 

- Sous-Chef / resp. visualisation : 

- **Objectifs :**

- Recherche de profils de clients atypiques selon plusieurs types de représentations

- Lien entre les profils clients atypiques et les fraudés

- **Données : jeu n°1**

# Groupe 3 : Clustering

- **Composition :**

- Tuteur enseignant : Mathieu Serrurier
- Chef de groupe : [REDACTED]
- Sous-Chef / resp. visualisation : [REDACTED]

- **Objectifs :**

- Recherche de groupes homogènes de clients selon plusieurs types de représentations
- Analyse de la sensibilité des groupes à la fraude

- **Données : jeu n°1**

# Groupe 4 : Intimité différentielle

- **Composition :**

- Tuteur enseignant : Adrien Mazoyer

- Chef de groupe : 

- Sous-Chef / resp. visualisation : 

- **Objectifs :**

- Est-il possible de déduire certaines variables à l'aide d'autres ?

- Quel est le nombre minimal de variables nécessaires pour détecter la fraude ?

- Peut-on bruite les données sans dégrader la détection de fraude ?

- **Données : jeu n°1**

# Groupe 5 : Génération de nouvelles observations

- **Composition :**

- Tuteur enseignant : Mathieu Serrurier

- Chef de groupe : [REDACTED]

- Sous-Chef / resp. visualisation : [REDACTED]

- **Objectifs :**

- Construction d'un générateur de données ayant les mêmes caractéristiques que le jeu de données initiales

- **Données : jeu n°1**

# Groupe 6 : Statistiques descriptives du webtracking et recherche de patterns

- **Composition :**

- Tuteur enseignant : Dominique Bontemps

- Chef de groupe : [REDACTED]

- Sous-Chef / resp. visualisation : [REDACTED]

- **Objectifs :**

- Qu'est-ce qui décrit le parcours d'un client de façon pertinente ? Comment le généraliser sur l'ensemble des données ?

- Comment lie-t-on les données clients aux données de son parcours ?

- **Données : jeu n°2**

# Groupe 7 : Représentation des parcours digitaux

- **Composition :**

- Tuteur enseignant : Emmanuelle Claeys

- Chef de groupe : 

- Sous-Chef / resp. visualisation : 

- **Objectifs :**

- Comment représenter le parcours digital des clients à partir de données de webtracking ?

- **Données : jeu n°2**

# Groupe 8 : Prédiction de souscription (sans données de webtracking)

- **Composition :**

- Tuteur enseignant : Cécile Chouquet

- Chef de groupe : 

- Sous-Chef / resp. visualisation : 

- **Objectifs :**

- Quelles sont les caractéristiques des souscripteurs ?

- A quel point peut-on prédire la probabilité de souscription à partir des ces informations ?

- **Données : jeu n°2**

# Groupe 9 : Prédiction de souscription (sur les données de webtracking)

- **Composition :**

- Tuteur enseignant : Emma Grandgirard
- Chef de groupe : 
- Sous-Chef / resp. visualisation : 

- **Objectifs :**

- Quel est le parcours type de la souscription ?
- Quelles sont les variables explicatives de la souscription ?
- Les données de webtracking permettent-elles d'améliorer la prédiction de la souscription ?
- Estimer la longueur d'historique pour prédire la souscription

- **Données : jeu n°2**

# Dates clés

- Phase préparatoire
  - Composition et description des groupes : le 17/11 à 20h30 sur Moodle
  - Formation gestion de projet (Catalyseur) :  
24/11 matin (M2), 30/11 matin (M2) et 10/12 après-midi (tous)
  - Travail préparatoire des M2 non-alternants : du 13/12 au 17/12
  - Travail préparatoire : du 17/11 au 17/12
- Période du projet : 03/01 au 14/01
- Restitution du projet : le 17/01 après-midi

# Travail préparatoire pour tous

- **Initié par les chefs de groupe**
- **Clairement spécifié dans les fiches de présentation**
  - Que faire pour monter en compétences ?
  - Que rendre avant les vacances de Noël ?
    - Travail avec le catalyseur
    - Documents élaborés pour la montée en compétence
- **Sensibilisation à la charte de codage**
- **Concertation pour le choix de la plateforme de dashboard**

# Travail préparatoire des M2 non-alternants

- **Charte de codage**
- **Préparation des fichiers de données**
  - Organisation et mise en forme du jeu de données n°2
  - Dans quelle mesure y a-t-il une intersection entre les 2 jeux de données ? Y a-t-il un intérêt à les concaténer ?
- **Mise en place d'une plateforme de dashboard choisie**

# Des questions ?

