



Marie-Pierre Jacquemin

DIGI 4

Semaine 2 - RUSH 4

Profil client et solvabilité financière

Groupe de 2/3 personnes

Durée : 2 jours

*Pas de soutenance, seulement un rapport + PYTHON (apprécié et fortement suggéré).
(Introduction à l'algorithmique...).*

Attentes :

- ☞ Diverses régressions et corrélations
- ☞ Pour celles/ceux qui souhaitent aller plus loin sur la réflexion mathématique et statistiques : je suis là
- ☞ Être vigilant quant à la corrélation qui ne justifie pas nécessairement une notion de causalité...

Pour celles et ceux qui ne savent pas coder, ou pour qui ceci est compliqué/dur/long, il est au moins important de comprendre les concepts de régression linéaire, logit, et les KNN → WORKSHOP

Liberté au niveau des outils/langage de programmation utilisés : C++/R/Python/PowerBI

Objectifs :

**Analyser les données d'une banque sur les crédits à la consommation.
Faire de l'analyse exploratoire une étape préliminaire à la compréhension du marché.**

Soit en clair :

- ☞ Dégager des KPI
- ☞ Analyser les données d'une banque sur les crédits à la consommation.
- ☞ Faire une analyse exploratoire, qui est une étape préliminaire à la compréhension du marché.
- ☞ Analyser les données recueillies, identifier les patterns, interroger la corrélation et la causalité.
- ☞ Introduire des notions algorithmiques de capacité prédictive des modèles.



Marie-Pierre Jacquemin

Finalement parvenir à prédire les Caractéristiques du client : Bon ou Mauvais client

Qu'est ce qu'un bon client ?

Un bon client est celui qui pourra honorer ses engagements avec l'entité financière, soit ici, qui sera en mesure de rembourser son crédit.

Contexte : 1723 observations sur le risque crédit d'une banque.

Vous êtes Consultant Business Intelligence pour une banque.

La mission qui vous est confiée est d'optimiser le processus de tri des clients entre "bons" et "mauvais" clients.

Pour cela vous devrez développer une solution algorithmique justifiant de votre compréhension des réalités data.

Livrables attendus:

- ☞ Un rapport écrit de 5 à 8 pages (hors annexes) justifiant les choix analytiques de l'exploration data et les variables explicatives du risque crédit
- ☞ Un code en PYTHON exprimant votre capacité à prédire les données de la banque par une méthode de régression ou de KNN.

REMARQUE :

L'analyse doit mettre en évidence les écueils de la caractérisation par la data.

Ce n'est pas parce qu'on emploie des méthodes analytiques complexes qu'on a les meilleurs résultats. Il faut prendre conscience de la qualité de l'information à disposition.

Ici, le dataset est assez logiquement très déséquilibré entre les bons et les mauvais clients.

Même la partie la plus technique demandée, ie un algo de prédiction type méthode simple de ML, peut s'avérer "incompétent" s'il est entraîné sur un dataset "défaillant".

En effet, la base contenant 88% de bons clients, on risque d'avoir un taux de prédictions faussé, ne parvenant finalement pas à identifier les nouveaux mauvais clients.....Et ce au-delà de l'aspect technique pur.