

Risque de Crédit – Projet à rendre au plus tard le 28 Février 2023
A soumettre à : hdiallo01@gmail.com
I. Problématique

Une banque souhaite mettre en place une nouvelle grille de probabilité de défaut (PD) pour noter sa base de clients. La Direction Risque utilise l'information sur un échantillon de clients représentatifs afin de développer un modèle permettant de prévoir avec quelle probabilité le client tombera en défaut.

Ce modèle permettra de cibler les clients à risque élevé/faible. La Direction Risque a extrait des informations de 150 000 clients pour déceler les défauts. Est-ce que le client est tombé en défaut ? Deux réponses sont possibles à cette question : 1 – oui ou 0 – non.

Le but de cette étude est de construire un modèle basé sur cet échantillon.

Le rapport : (voir la structure indicative ci-dessous) devrait être clair, précis et concis. Les fautes d'orthographe et/ou de grammaire à bannir.

II. Variables explicatives
1. Variables explicatives

- Nettoyage du jeu de données
- Statistiques descriptives pour décrire le jeu de données

2. Profil du défaut

- Etablir un profil type de défaut à travers notamment les tests de corrélation entre les variables

3. Méthodes de traitement de l'information et résultats
III. Modélisation

Itérer jusqu'à obtenir le modèle le plus performant à vos yeux et interprétation.

IV. Utilisation du score (note) obtenu

Utiliser les notes obtenues (scores) pour construire des classes homogènes de risque (même nombre de CHR que la variable « grade ») et calcul des probabilités de défaut par CHR.

V. Conclusion
VI. Annexe (vos programmes SAS et autres éléments)
VII. La grille de notation ci-dessous sera appliquée

Topic	Sub-topic	Nb max de points
Forme	Présentation globale du rapport	1
	Qualité rédactionnelle (clarté, précision, fautes d'orthographe/grammaire à bannir)	2
Fond	Qualité des graphiques Data Visualization	2
	Interprétations des analyses et résultats	3
	Comparaisons des modèles	2
	Cohérence des CHR	3
Prise de risque et analyses optionnelles	Variables croisées et autres analyses	3
	Autres méthodes de traitement de l'information (arbres de décision, réseaux de neurones, etc...)	2
Code SAS/R/Python	Cohérence entre le code SAS ou R ou Python et les résultats	1
	Auditabilité des codes	1
Total		20