Rapport de projet

Modélisation en Risque de Crédit et Segmentation en Classes Homogènes de Risque

Roland DUTAUZIET
Lina RAGALA
Maeva N'GUESSAN

6 janvier 2026

Table des matières

Introduction

Ce projet vise à développer un modèle de notation de crédit (*Credit Scoring*) robuste et interprétable, permettant d'estimer la Probabilité de Défaut (PD) d'un emprunteur. La démarche suit une méthodologie rigoureuse, conforme aux standards bancaires (Bâle II/III), allant du nettoyage des données à la segmentation finale en classes de risques homogènes et l'estimation de la probabilité de défaut.

1 Compréhension et Préparation des Données

L'objectif de cette première phase est de garantir la qualité et la fiabilité des données avant toute analyse statistique. Le dataset de base provient de Kaggle. On travaille sur le jeu de données **credit_risk_dataset.csv**. On va éffectuer une analyse univariée des variables explicatives. Cela permet d'identifier pour chacune des variables le pourcentage de valeurs manquantes/aberrantes, non-applicables ou à exclure. Dans notre cas, on a peu de variables et l'essentiel des transformations ont été faites mais c'est quand même important de bien maîtriser ses données.

1.1 Traitement des Valeurs Manquantes

Une analyse exploratoire a permis d'identifier les taux de remplissage par variable.
— **Variables numériques :** Imputation par la médiane pour limiter l'impact des valeurs extrêmes.
— **Variables catégorielles :** Imputation par le mode (valeur la plus fréquente) ou création d'une modalité spécifique "Missing" si le taux de vide est significatif.

1.2 Gestion des Valeurs Aberrantes (Outliers)

Pour éviter de biaiser le modèle, les valeurs extrêmes ont été traitées selon deux approches :
— **Filtres Métiers :** Application de seuils logiques (ex : âge maximum de 100 ans, exclusion des revenus incohérents).
— **Méthode IQR :** Détection statistique des outliers via l'intervalle interquartile pour les variables continues restantes.

2 Discrétisation optimisée des variables continues

La discrétisation est une étape clé en scoring. Elle transforme les variables continues en classes (bins), permettant de capturer les non-linéarités et de rendre le modèle plus stable.

2.1 Stratégie Mixte

Nous avons adopté une approche hybride adaptée à la nature de chaque variable :

1. **ChiMerge (Basé sur le Chi-2) :** Utilisé pour les variables ayant une relation complexe avec le risque (ex : *loan_int_rate*). L'algorithme fusionne les classes adjacentes tant qu'elles présentent une distribution de défaut statistiquement similaire.
2. **Monotone :** Appliqué aux variables où une relation d'ordre logique est attendue (ex : *person_income*). Cette méthode force le taux de défaut à être strictement croissant ou décroissant, garantissant l'interprétabilité économique.
3. **K-Means (Clustering) :** Utilisé pour les variables socio-démographiques (ex : *person_age*) afin de créer des groupes naturellement homogènes.

2.2 Ajustements Manuels

Des regroupements post-algorithmiques ont été effectués pour éviter les classes à trop faible effectif (sur-apprentissage) et assurer la robustesse statistique.

3 Analyse des Corrélations

Avant la modélisation, il est impératif de réduire la multi-colinéarité entre les variables explicatives pour ne pas instabiliser la régression logistique.

— **V de Cramer :** Calculé pour mesurer l'intensité de la liaison entre deux variables qualitatives discrétisées.

— **Test du Chi-2 :** Réalisé pour vérifier la dépendance significative entre chaque variable explicative et la variable cible (*loan_status*).

TABLE 1 – Tests d'indépendance du χ^2 - Couples de variables dépendantes au seuil 5%

Variable 1	Variable 2	Chi-2	ddl	p_value	dependance_significative	chi2_valide
person_age	cb_person_cred_hist_length	39883.623	6	0.000	True	True
loan_int_rate	cb_person_default_on_file	7440.067	6	0.000	True	True
loan_amnt	loan_percent_income	6966.772	9	0.000	True	True
person_income	loan_amnt	4318.562	15	0.000	True	True
person_income	person_home_ownership	3885.711	10	0.000	True	True
person_income	loan_percent_income	3786.505	15	0.000	True	True
person_emp_length	cb_person_cred_hist_length	3253.798	4	0.000	True	True
person_age	person_emp_length	3170.324	6	0.000	True	True
person_home_ownership	person_emp_length	1764.583	4	0.000	True	True
person_income	person_emp_length	1485.153	10	0.000	True	True
loan_amnt	loan_int_rate	931.191	18	0.000	True	True
person_age	loan_intent	756.689	15	0.000	True	True
person_home_ownership	loan_intent	754.581	10	0.000	True	True
person_home_ownership	loan_int_rate	717.725	12	0.000	True	True
person_home_ownership	loan_percent_income	630.453	6	0.000	True	True
loan_int_rate	loan_percent_income	507.686	18	0.000	True	True
person_age	person_income	486.127	15	0.000	True	True
person_home_ownership	loan_amnt	434.633	6	0.000	True	True
loan_intent	cb_person_cred_hist_length	400.087	10	0.000	True	True
person_income	loan_intent	361.025	25	0.000	True	True
person_emp_length	loan_amnt	297.355	6	0.000	True	True
person_income	cb_person_cred_hist_length	296.791	10	0.000	True	True
person_emp_length	loan_int_rate	146.878	12	0.000	True	True
person_income	loan_int_rate	136.887	30	0.000	True	True
person_home_ownership	cb_person_default_on_file	135.215	2	0.000	True	True
loan_amnt	cb_person_default_on_file	105.607	3	0.000	True	True
person_emp_length	loan_percent_income	98.512	6	0.000	True	True
person_age	loan_amnt	96.395	9	0.000	True	True
person_emp_length	loan_intent	91.809	10	0.000	True	True
person_age	person_home_ownership	90.666	6	0.000	True	True
loan_intent	loan_amnt	87.729	15	0.000	True	True
loan_percent_income	cb_person_default_on_file	59.730	3	0.000	True	True
person_home_ownership	cb_person_cred_hist_length	59.238	4	0.000	True	True
loan_amnt	cb_person_cred_hist_length	58.695	6	0.000	True	True
person_age	loan_percent_income	51.660	9	0.000	True	True
loan_intent	loan_int_rate	49.184	30	0.015	True	True
person_income	cb_person_default_on_file	44.847	5	0.000	True	True
person_emp_length	cb_person_default_on_file	34.109	2	0.000	True	True
loan_percent_income	cb_person_cred_hist_length	33.101	6	0.000	True	True
loan_intent	cb_person_default_on_file	9.376	5	0.095	True	True

4 Sélection de Variables (Stepwise)

Afin de construire un modèle parcimonieux et performant, une procédure de sélection automatique a été mise en œuvre.

4.1 Méthodologie

- **Encodage :** Les variables discrétisées ont été transformées en variables binaires (*One-Hot Encoding*).
- **Algorithme Stepwise :** Une approche combinée (Forward et Backward) a été utilisée.
- **Critères de sélection :**
  - Minimisation du critère **BIC** (Akaike Information Criterion).
  - Validation de la significativité des coefficients via le **Test de Wald** (p-value < 0.05).

Formule mathématique du logit

Soit

$$p = P(Y = 1 \mid X)$$

la probabilité de défaut.

Le **logit** est défini par :

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

Modèle de régression logistique

Dans un modèle logit, le logit est exprimé comme une combinaison linéaire des variables explicatives :

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

où :

- β_0 est l'intercept,
- β_i sont les coefficients estimés,
- X_i sont les variables explicatives (ou les modalités après discrétisation).

Formule inverse (probabilité de défaut)

La probabilité de défaut s'obtient par l'inverse du logit :

$$p = \frac{1}{1 + \exp\left(-\left(\beta_0 + \sum_{i=1}^{k} \beta_i X_i\right)\right)}$$

Interprétation en Credit Scoring

- Un coefficient $\beta_i > 0$ augmente la probabilité de défaut.
- Un coefficient $\beta_i < 0$ diminue la probabilité de défaut.
- Le caractère additif du logit permet la construction d'une grille de score par somme de contributions.

Variable	Modalité	Coefficient	p-value	Significativité
person_age	person_age_0	0.1081	0.2504	_
person_age	person_age_1	0.0027	0.9699	_
person_age	person_age_2 (Référence)	0.0000	_	_
person_income	person_income_0	1.5305	3.62e-95	***
person_income	person_income_1	1.0821	4.87e-52	***
person_income	person_income_2	0.3298	3.81e-10	***
person_income	person_income_3 (Référence)	0.0000	_	_
person_home_ownership	person_home_ownership_0 (Référence)	0.0000	_	_
person_home_ownership	person_home_ownership_1	1.6490	5.60e-50	***
person_home_ownership	person_home_ownership_2	2.5540	8.69e-126	***
person_emp_length	person_emp_length_0	0.0112	0.8850	_
person_emp_length	person_emp_length_1	-0.1225	0.1310	_
person_emp_length	person_emp_length_2 (Référence)	0.0000	_	_
loan_intent	loan_intent_0	0.9705	8.90e-47	***
loan_intent	loan_intent_1	0.9088	1.25e-49	***
loan_intent	loan_intent_2	0.2913	1.94e-06	***
loan_intent	loan_intent_3 (Référence)	0.0000	_	_
loan_amnt	loan_amnt_0 (Référence)	0.0000	_	_
loan_amnt	loan_amnt_1	-0.1015	0.2786	_
loan_amnt	loan_amnt_2	0.1422	0.0523	_
loan_amnt	loan_amnt_3	0.4400	2.86e-08	***
loan_percent_income	loan_percent_income_0 (Référence)	0.0000	_	_
loan_percent_income	loan_percent_income_1	0.2667	2.47e-08	***
loan_percent_income	loan_percent_income_2	0.5052	1.73e-12	***
loan_percent_income	loan_percent_income_3	2.7524	1.85e-300	***
cb_person_default_on_file	cb_person_default_on_file_0 (Référence)	0.0000	_	_
cb_person_default_on_file	cb_person_default_on_file_1	1.1616	1.54e-156	***
cb_person_cred_hist_length	cb_person_cred_hist_length_0	-0.0452	0.4728	_
cb_person_cred_hist_length	cb_person_cred_hist_length_1 (Référence)	0.0000	_	_

Table 2 – Modèle Logit Naïf

Variable	Modalité	Coefficient	p-value	Significativité
loan_percent_income	loan_percent_income_0 (Référence)	0.0000	_	_
loan_percent_income	loan_percent_income_1	0.3276	3.03e-13	***
loan_percent_income	loan_percent_income_2	0.6189	7.75e-21	***
loan_percent_income	loan_percent_income_3	2.9469	0.0000	***
person_home_ownership	person_home_ownership_0 (Référence)	0.0000	_	_
person_home_ownership	person_home_ownership_1	1.6898	6.60e-52	***
person_home_ownership	person_home_ownership_2	2.5957	8.09e-128	***
cb_person_default_on_file	cb_person_default_on_file_0 (Référence)	0.0000	_	_
cb_person_default_on_file	cb_person_default_on_file_1	1.1757	1.92e-161	***
person_income	person_income_0	1.3959	3.71e-108	***
person_income	person_income_1	0.9319	3.75e-51	***
person_income	person_income_2	0.2178	3.15e-06	***
person_income	person_income_3 (Référence)	0.0000	_	_
loan_intent	loan_intent_0	0.9734	4.55e-47	***
loan_intent	loan_intent_1	0.9126	3.73e-50	***
loan_intent	loan_intent_2	0.2954	1.40e-06	***
loan_intent	loan_intent_3 (Référence)	0.0000	_	_

Table 3 – Modèle Logit Stepwise

Variable	Modalité	Coefficient	p-value	Significativité
loan_percent_income	loan_percent_income_0 (Référence)	0.0000	_	_
loan_percent_income	loan_percent_income_1	0.3252	8.34e-14	***
loan_percent_income	loan_percent_income_2	0.6431	8.76e-24	***
loan_percent_income	loan_percent_income_3	2.7861	0.0000	***
person_home_ownership	person_home_ownership_0 (Référence)	0.0000	_	_
person_home_ownership	person_home_ownership_1	1.7302	2.80e-57	***
person_home_ownership	person_home_ownership_2	2.6547	1.29e-140	***
person_income	person_income_0	1.3252	2.06e-104	***
person_income	person_income_1	0.9117	2.72e-52	***
person_income	person_income_2	0.1922	2.31e-05	***
person_income	person_income_3 (Référence)	0.0000	_	_

TABLE 4 – Modèle Logit Stepwise Strict

5 Construction de la Grille de Score

Le modèle prédictif final est une **Régression Logistique**, dont les résultats sont traduits en une grille de score. Cette transformation permet d'attribuer une note sur 1000 aux clients, facilitant l'interprétation opérationnelle du risque.

La grille de score ainsi créée regroupe les informations suivantes :

— Les variables explicatives retenues ;
— Les classes (modalités) des variables explicatives ;
— La *p-value* associée au test de Wald pour chaque classe ;
— La note normalisée attribuée à chaque classe ;
— La contribution relative de la variable au score total ;
— Le taux de défaut observé pour chaque classe ;
— L'effectif de chaque classe.

Règles d'acceptation Pour garantir la robustesse du modèle, les règles suivantes ont été appliquées :

— **P-value associée à chaque classe :** Toutes les classes doivent être statistiquement significatives au seuil de 5%. *Note : Le non-respect de cette règle induirait l'existence d'une corrélation résiduelle entre les classes et/ou les variables.*
— **Cohérence des signes :** Le signe des coefficients doit être économiquement cohérent (respect de la monotonicité du risque par rapport à la variable).

5.1 Calcul des Notes (N_i^j)

Les coefficients β issus de la régression sont transformés en points selon une échelle normalisée (de 0 à 1000) :

Calcul des notes par modalité

La note associée à la modalité j de la variable i est définie par :

$$N_i^j = \frac{\left|\min\left(\beta_i^1, \ldots, \beta_i^p\right) - \beta_i^j\right|}{\sum_{i=1}^{k}\left[\max\left(\beta_i^1, \ldots, \beta_i^p\right) - \min\left(\beta_i^1, \ldots, \beta_i^p\right)\right]} \times 1000$$

où :

— N_i^j est la note associée à la modalité j de la variable i,
— β_i^j est le coefficient estimé de la modalité j de la variable i dans le modèle logit,
— $\min(\beta_i)$ et $\max(\beta_i)$ représentent respectivement les coefficients minimum et maximum parmi les modalités de la variable i,

— p est le nombre de modalités de la variable i,
— k est le nombre total de variables du modèle.
Cette transformation garantit une échelle de score comprise entre 0 et 1000, où :
— une note élevée correspond à un profil plus risqué,
— une note faible correspond à un profil moins risqué.

5.2 Contribution des Variables

L'importance relative de chaque variable (c_i) dans le score final a été calculée en fonction de la dispersion des notes pondérée par les effectifs, mettant en lumière les facteurs de risque prédominants.

$$c_i = \frac{\sqrt{\sum_{j=1}^{p} r_j \left(N_i^j - \overline{N_i}\right)^2}}{\sum_{i=1}^{k} \sqrt{\sum_{j=1}^{p} r_j \left(N_i^j - \overline{N_i}\right)^2}}$$

où :
— c_i est la contribution de la variable i au score global,
— N_i^j est la note associée à la modalité j de la variable i,
— r_j est la proportion d'individus appartenant à la modalité j de la variable i,
— p est le nombre de modalités de la variable i,
— k est le nombre total de variables du modèle.
La note moyenne pondérée $\overline{N_i}$ est calculée comme :

$$\overline{N_i} = \sum_{j=1}^{p} r_j N_i^j$$

Cette contribution mesure la capacité discriminante de chaque variable dans la grille de score :
— une contribution élevée indique une variable fortement discriminante,
— une contribution faible signale une variable peu informative pour la différenciation du risque.

Variable	Modalité	Bornes modalité	Effectif	Coefficient	Note	Contribution (%)	Taux de défaut observé (%)	p-value	Significativité
loan_percent_income	loan_percent_income_0 (Référence)	]-∞, 0.15]	17185	0.0000	0	31.41	12.15	_	_
loan_percent_income	loan_percent_income_1	]0.15, 0.25]	9110	0.3276	36	31.41	18.62	3.03e-13	***
loan_percent_income	loan_percent_income_2	]0.25, 0.30]	2444	0.6189	68	31.41	25.70	7.75e-21	***
loan_percent_income	loan_percent_income_3	]0.30, +∞[	3833	2.9469	324	31.41	70.31	0	***
person_home_ownership	person_home_ownership_0 (Référence)	Modalité 0	2584	0.0000	0	25.04	7.47	_	_
person_home_ownership	person_home_ownership_1	Modalité 1	13441	1.6898	186	25.04	12.57	6.60e-52	***
person_home_ownership	person_home_ownership_2	Modalité 2	16547	2.5957	286	25.04	31.57	8.09e-128	***
cb_person_default_on_file	cb_person_default_on_file_0 (Référence)	Modalité 0	26829	0.0000	0	15.31	18.40	_	_
cb_person_default_on_file	cb_person_default_on_file_1	Modalité 1	5743	1.1757	129	15.31	37.80	1.92e-161	***
person_income	person_income_0	]-∞, 28590.00]	3259	1.3959	154	15.26	47.41	3.71e-108	***
person_income	person_income_1	]28590.00, 35000.00]	3370	0.9319	103	15.26	39.26	3.75e-51	***
person_income	person_income_2	]35000.00, 63000.00]	12942	0.2178	24	15.26	20.85	3.15e-06	***
person_income	person_income_3 (Référence)	]63000.00, +∞[	13001	0.0000	0	15.26	11.85	_	_
loan_intent	loan_intent_0	Modalité 0	5212	0.9734	107	12.98	28.59	4.55e-47	***
loan_intent	loan_intent_1	Modalité 1	9675	0.9126	100	12.98	26.48	3.73e-50	***
loan_intent	loan_intent_2	Modalité 2	11970	0.2954	32	12.98	18.45	1.40e-06	***
loan_intent	loan_intent_3 (Référence)	Modalité 3	5715	0.0000	0	12.98	14.82	_	_

TABLE 5 – Grille de Score

Variable	**Modalité**	**Bornes modalité**	**Effectif**	**Coefficient**	**Note**	**Contribution (%)**	**Taux défaut (%)**	**p-value**	**Signif.**
loan_percent_income	loan_percent_income_0 (Réf.)	$]-\infty, 0.15]$	17185	0.0000	0.00	42.45	12.15	_	_
loan_percent_income	loan_percent_income_1	$]0.15, 0.25]$	9110	0.3252	48.07	42.45	18.62	8.34×10^{-14}	***
loan_percent_income	loan_percent_income_2	$]0.25, 0.30]$	2444	0.6431	95.05	42.45	25.70	8.76×10^{-24}	***
loan_percent_income	loan_percent_income_3	$]0.30, +\infty[$	3833	2.7861	411.78	42.45	70.31	0	***
person_home_ownership	person_home_ownership_0 (Réf.)	Modalité 0	2584	0.0000	0.00	36.58	7.47	_	_
person_home_ownership	person_home_ownership_1	Modalité 1	13441	1.7302	255.72	36.58	12.57	2.80×10^{-57}	***
person_home_ownership	person_home_ownership_2	Modalité 2	16547	2.6547	392.36	36.58	31.57	1.29×10^{-140}	***
person_income	person_income_0	$]-\infty, 28590.00]$	3259	1.3252	195.86	20.97	47.41	2.06×10^{-104}	***
person_income	person_income_1	$]28590.00, 35000.00]$	3370	0.9117	134.75	20.97	39.26	2.72×10^{-52}	***
person_income	person_income_2	$]35000.00, 63000.00]$	12942	0.1922	28.41	20.97	20.85	2.31×10^{-05}	***
person_income	person_income_3 (Référence)	$]63000.00, +\infty[$	13001	0.0000	0.00	20.97	11.85	_	_

TABLE 6 – Grille de Score Strict

6 Segmentation en Classes de Risque Homogènes

L'étape ultime du processus consiste à transformer le score continu en *Classes Homogènes de Risque* (CHR) afin d'estimer la *PD Long-Run Average* (PD_LRA). Une fois la grille de score construite, celle-ci est associée à une échelle de notation permettant de regrouper les individus présentant des profils de risque similaires au sein de classes distinctes.

Plusieurs algorithmes peuvent être mobilisés pour la construction des CHR, parmi lesquels :

— les arbres de décision,
— les algorithmes génétiques,
— la méthode de Jenks (*Natural Breaks Optimization*).

La segmentation doit respecter un ensemble de contraintes réglementaires et méthodologiques afin d'assurer sa robustesse et son exploitabilité opérationnelle :

— une forte homogénéité du risque au sein de chaque classe,
— une hétérogénéité marquée entre les classes,
— l'absence de concentration excessive dans une même classe (généralement limitée à 30% de la population),
— une augmentation monotone et régulière des taux de défaut d'une classe à la suivante.

6.1 Méthodologie de Segmentation

Un algorithme d'**Arbre de Décision** (*DecisionTreeClassifier*) a été utilisé pour identifier les seuils de score (*cut-offs*) optimaux. Cette méthode permet de maximiser l'écart de taux de défaut entre les classes.

L'arbre de décision est particulièrement adapté car son fonctionnement natif consiste à trouver le meilleur point de coupure dans une variable pour séparer au mieux les "Bons payeurs" (0) des "Mauvais payeurs" (1). L'algorithme cherche les seuils dans la note finale qui maximisent la pureté des groupes par rapport au défaut (y).

La configuration du modèle repose sur les contraintes suivantes :

— **Nombre de feuilles :** L'arbre est forcé de s'arrêter dès qu'il a créé 7 feuilles finales (correspondant aux 7 classes de risque).
— **Robustesse :** Une contrainte interdisant la création de classes contenant moins de 5% de la population totale a été ajoutée afin de garantir la représentativité statistique.
— **Critère d'entropie :** L'algorithme teste différents seuils (ex : Note $>$ 300 vs Note $>$ 305) et conserve celui qui minimise l'entropie le plus rapidement. Cela permet de créer des groupes ayant des taux de défaut très hétérogènes entre eux.

Une fois les seuils identifiés, les clients sont regroupés et les statistiques clés suivantes sont calculées pour chaque classe :

— **Min/Max Score :** Définit la fourchette de la classe (ex : Classe 2 = Score entre 220 et 280).
— **Effectif :** Le nombre de clients présents dans la classe.
— **Défauts :** Le nombre de clients ayant fait défaut.
— **PD_LRA (*Probability of Default - Long Run Average*) :** Il s'agit du taux de défaut moyen de la classe.
  — *Calcul :* $\frac{\text{Nombre de défauts}}{\text{Nombre total de clients}} \times 100$
  — C'est l'indicateur de référence utilisé par la banque pour le provisionnement des risques.

6.2 Respect des Contraintes

La segmentation obtenue valide les critères réglementaires suivants :

— **Monotonicité :** Le taux de défaut augmente strictement de la classe AAA (faible risque) à la classe CCC (haut risque).
— **Concentration :** La distribution des clients est équilibrée, évitant une concentration excessive ($>$30%) dans une seule classe.
— **Discrimination :** Les classes extrêmes sont bien différenciées (Taux de défaut $<$3% pour la meilleure classe vs $>$90% pour la pire).

Classe Risque	Notation Bâle	Min Score	Max Score	Effectif	Defauts	Part de population(%)	PD_LRA_chr(%)
0	AAA	0.0	221.98	4771	117	14.65	2.45
1	AA	222.15	289.8	4206	283	12.91	6.73
2	A	290.07	392.75	10045	1162	30.84	11.57
3	BBB	395.6	456.28	4600	879	14.12	19.11
4	BB	456.71	568.61	4418	1329	13.56	30.08
5	B	569.34	719.87	2644	1533	8.12	57.98
6	CCC	734.29	1000.0	1888	1804	5.8	95.55

TABLE 7 – Synthèse des Classes Homogènes de Risques (CHR)

Classe Risque	**Notation Bâle**	**Min Score**	**Max Score**	**Effectif**	**Défauts**	**Part Pop. (%)**	**PD LRA (%)**
0	AAA	0.00	229.80	1907	22	5.85	1.15
1	AA	243.93	255.72	5471	460	16.80	8.41
2	A	284.13	411.78	9665	1230	29.67	12.73
3	BBB	420.77	468.84	7251	1167	22.26	16.09
4	BB	485.52	546.53	1954	524	6.00	26.82
5	B	546.63	802.25	3884	1310	11.92	33.73
6	CCC	804.14	1000.00	2440	2394	7.49	98.11

TABLE 8 – Synthèse des Classes Homogènes de Risques (CHR) Strict

Conclusion

Le modèle développé permet une évaluation précise du risque de crédit. La transformation en grille de score assure une transparence totale de la décision, tandis que la segmentation optimisée fournit un outil de pilotage du risque conforme aux exigences prudentielles.

Annexes et Liens utiles

Cette section regroupe les ressources externes associées au projet, permettant d'en assurer la reproductibilité, la transparence méthodologique et l'accès aux données utilisées.

Code source du projet

— Repository GitHub (code, notebooks, résultats et documentation) :
https://github.com/Shawndarm/Credit-Risk-Project-PD-Estimation

Données

— Dataset Kaggle – Credit Risk Dataset :
https://www.kaggle.com (jeu de données utilisé pour la modélisation du risque de défaut)

Ressources méthodologiques et réglementaires

— Comité de Bâle – Principes pour la gestion du risque de crédit :
https://www.bis.org/bcbs
— IFRS 9 – Instruments financiers :
https://www.ifrs.org
— Documentation `statsmodels` (régression logistique) :
https://www.statsmodels.org