

MEMOIRE DE FIN D'ETUDES

En vue de l'obtention d'un Master académique en
« Management financier des entreprises »

La gestion du risque des crédits d'exploitation des PME par le Machine Learning Cas : Trust Bank Algeria.

Elaboré par :

BOUKHERROUB Yasmine

Encadré par :

DR. BOUCHETARA Mehdi

Année : 2025/2026

Résumé

Cette recherche porte sur l'application du Machine Learning, plus précisément la régression logistique, à la prédiction du risque de défaut des petites et moyennes entreprises (PME) sur les crédits d'exploitation dans le contexte bancaire algérien. L'étude repose sur une base de données réelle de 255 dossiers de crédit issus du système de notation Moody's de TRUST Bank Algérie sur la période 2020–2024. Sept ratios financiers liés à l'endettement et à la gestion du cycle d'exploitation ont été utilisés comme variables explicatives. L'analyse a été réalisée sous le logiciel R à l'aide du framework tidymodels, selon une démarche comprenant l'analyse exploratoire des données, le prétraitement, la modélisation par régression logistique et l'évaluation des performances. Les résultats montrent que les ratios financiers étudiés possèdent un fort pouvoir discriminant entre les entreprises saines et défaillantes. Le modèle obtient des performances prédictives très élevées avec une ROC-AUC de 1,000 sur l'échantillon de test. Ces résultats suggèrent que la qualité des données issues d'un système de notation professionnel joue un rôle déterminant dans l'efficacité des modèles de Machine Learning appliqués au risque de crédit. Enfin, cette recherche met en évidence l'intérêt de la régression logistique comme outil d'aide à la décision bancaire, grâce à sa simplicité, sa transparence et son interprétabilité.

Mots clés :

Machine Learning – Régression logistique – Risque de crédit – PME – Banque – ROC-AUC – Moody's – TRUST Bank Algérie

Abstract

This research focuses on the application of Machine Learning, more specifically logistic regression, to predicting the default risk of small and medium-sized enterprises (SMEs) on working capital loans in the Algerian banking context. The study is based on a real dataset of 255 credit files extracted from the Moody's rating system of TRUST Bank Algeria over the period 2020–2024. Seven financial ratios related to indebtedness and working capital management were used as explanatory variables. The analysis was conducted using R software and the tidymodels framework through exploratory data analysis, preprocessing, logistic regression modeling, and performance evaluation. The findings show that the selected financial ratios have a strong discriminating power between healthy and defaulting firms. The model achieved very high predictive performance with a ROC-AUC of 1.000 on the test sample. These results suggest that data quality plays a major role in the effectiveness of Machine Learning models applied to credit risk prediction. The study also highlights the usefulness of logistic regression as a transparent and interpretable decision-support tool for Algerian banks.

Keywords :

Machine Learning – Logistic Regression – Credit Risk – SMEs – Banking – ROC-AUC – Moody's – TRUST Bank Algeria

ملخص

تهدف هذه الدراسة لتطبيق التعلم الآلي، وعلى وجه الخصوص الانحدار اللوجستي، من حيث التنبؤ بخطر تعثر المؤسسات الصغيرة والمتوسطة في سداد قروض الاستغلال ضمن السياق البنكي الجزائري. اعتمدت الدراسة على قاعدة بيانات حقيقية تضم 255 ملف قرض مستخرجة من نظام التصنيف Moody's الخاص ببنك TRUST Bank Algérie خلال الفترة الممتدة بين 2020 و2024. وتم استخدام سبعة نسب مالية مرتبطة بالمديونية وتسيير دورة الاستغلال كمغيرات تفسيرية. أنجز التحليل باستخدام برنامج R وإطار العمل tidymodels عبر مراحل التحليل الاستكشافي للبيانات، المعالجة المسبقة، بناء نموذج الانحدار اللوجستي وتقييم الأداء. أظهرت النتائج أن النسب المالية المستعملة تمتلك قدرة تفسيرية وتمييزية قوية بين المؤسسات السليمة والمتعثرة. كما حقق النموذج أداءً تنبؤياً توقعياً مرتفعاً جداً بمؤشر ROC-AUC بلغ 1.000 على عينة الاختبار. وتشير هذه النتائج إلى أن جودة البيانات المستخرجة من نظام تصنيف مهني تلعب دوراً جوهرياً في فعالية ونجاعة نماذج التعلم الآلي المطبقة على مخاطر الائتمان. وتبرز الدراسة كذلك أهمية الانحدار اللوجستي كأداة مساعدة وداعمة لاتخاذ القرار البنكي بفضل بساطته وشفافيته وسهولة تفسير نتائجه.

الكلمات المفتاحية:

التعلم الآلي – الانحدار اللوجستي – مخاطر الائتمان – المؤسسات الصغيرة والمتوسطة – البنوك TRUST – Moody's

Remerciement

Avant toute chose, je tiens à remercier Allah le Tout-Puissant, de m'avoir donné la force, la patience, le courage et la volonté nécessaires pour accomplir ce travail. Sans Sa miséricorde et Sa protection, je n'aurais jamais pu arriver jusqu'ici.

Je tiens également à me remercier moi-même.
Oui... me remercier pour toutes les nuits de fatigue, les moments de doute, les larmes cachées derrière les sourires, pour ma patience, ma persévérance et ma volonté de toujours avancer malgré les difficultés et les moments de doute.
Ce mémoire représente bien plus qu'un simple travail académique ; il reflète une partie de mon parcours, de mes sacrifices et de mes ambitions.

Je tiens à adresser mes plus sincères remerciements à mon encadrant, **Dr. Bouchetara Mehdi**, pour son précieux accompagnement tout au long de ce travail.

Bien plus qu'un simple encadrant, il représente pour moi un véritable modèle de réussite, d'excellence et de savoir. Son intelligence remarquable, sa culture scientifique, son charisme ainsi que sa manière élégante de transmettre la connaissance inspirent profondément ses étudiants, il a su éveiller en moi l'envie de me dépasser et de croire davantage en mes capacités.

Travailler sous sa direction a été bien plus qu'une expérience académique : ce fut un privilège, un honneur et une leçon de vie. Certaines personnes marquent un parcours par leur présence ; lui marque les esprits par son excellence. Et comme les étoiles qui brillent sans bruit mais illuminent le ciel entier, il inspire autour de lui le respect, l'ambition et la passion du savoir.

Je garderai toujours une immense fierté et une profonde reconnaissance d'avoir eu l'opportunité d'apprendre auprès d'une personne aussi exceptionnelle.

Je tiens également à remercier chaleureusement **Monsieur Chaar**, mon tuteur de stage, pour son accueil, sa disponibilité, ses orientations et ses précieux conseils durant mon stage. Son aide et son professionnalisme ont grandement contribué à la réalisation de ce mémoire.

Mes sincères remerciements vont aussi à l'ensemble des enseignants de l'ENSM, qui ont contribué à ma formation académique et personnelle tout au long de mon parcours universitaire. Je remercie particulièrement **Madame Abid** ainsi que **Monsieur Horri** pour leurs enseignements, leurs conseils et leur soutien.

Une reconnaissance toute particulière à **Madame Saadi Dalila**, que je considère comme une véritable mère au sein de l'ENSM.

Merci pour votre bienveillance, votre disponibilité, vos encouragements constants et votre présence rassurante. Votre soutien humain a eu une grande valeur pour moi.

À mes chers parents...

Aucun mot ne pourra décrire tout ce que vous représentez pour moi.

Merci pour chaque sacrifice silencieux, chaque prière, chaque encouragement et chaque effort fourni pour me voir réussir.

Tout ce que je suis aujourd'hui, je vous le dois.

J'espère de tout cœur être une source de fierté pour vous, comme vous avez toujours été ma plus grande force.

À ma sœur,

Tu es bien plus qu'une sœur. Tu es mon refuge, mon soutien, mon équilibre, mon bras droit et une partie de mon âme.

Sans toi, je ne serais probablement pas arrivée jusque-là.

Merci d'avoir été présente dans chaque étape difficile, merci pour ton amour, ta patience, tes encouragements et pour avoir toujours cru en moi.

À mon frère, ainsi qu'à mes deux oncles **Hassan** et **Saïd**,

merci pour votre soutien, votre présence constante, votre affection et vos encouragements qui m'ont donné de la force tout au long de ce parcours.

Vos paroles pleines de fierté et votre confiance en moi ont toujours été une source de motivation et de courage.

À mon cher tonton...

Je suis vraiment chanceuse de t'avoir dans ma vie comme un deuxième frère.

Depuis le début de mon parcours, tu as toujours été là avec tes conseils, tes encouragements et tes paroles qui me donnaient de la force dans les moments difficiles.

Je n'oublierai jamais cette phrase que tu me répétais toujours :

« Tu es ma petite sœur Yasmine, je serai toujours à tes côtés. »

Merci pour tout ce que tu as fait pour moi.

À mes amis **Doaa, Chanez, Hana, Chihab, Oussama, Amine, Aymen et Rafik**, merci pour votre présence, vos encouragements, vos conseils, votre disponibilité et tous les moments de soutien durant cette aventure.

Une pensée particulière pour **Rafik**, dont l'aide, les conseils et le sérieux m'ont énormément aidée dans la réalisation de ce mémoire.

À toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce mémoire, merci du fond du cœur.

Table des matières

Résumé.....	I
Abstract	II
ملخص	III
Remerciement	IV
Liste de tableau	X
Liste de figure	XI
Liste des abréviations	12
INTRODUCTION GÉNÉRALE	13
Chapitre 1 : cadre conceptuel et revue de la littérature	19
Introduction.....	20
SECTION 01 : Revue de la littérature.....	20
1.1. Les techniques d'évaluation du risque de crédit	20
1.2. L'évolution des modèles de prédiction du défaut : des approches classiques au Machine Learning	21
1.3. Les apports récents du Machine Learning dans la prédiction de la défaillance.....	22
1.4. Les expériences dans les pays émergents et en Afrique	23
1.5. La prédiction du défaut des PME dans le contexte bancaire algérien	24
Section2 : Cadre conceptuel.....	27
1. Généralités concernant la banque	27
1.1 Définition de la banque	27
1.2 Le rôle de la banque dans l'économie	28
1.3 Les Types de banques	30
2. Généralités sur le crédit	32
2.1 Définition et rôle économique du crédit	32
2.2 Typologie des crédits bancaires	34
2.3 Spécificités du crédit d'exploitation.....	37
3. Le risque de crédit	44
3.1 Définition et enjeux	44
3.2 La décomposition du risque de crédit.....	44

3.3 Les paramètres fondamentaux du risque de crédit	46
4. Les méthodes d'évaluation du risque de crédit	47
4.1 L'analyse financière	48
4.2 La méthode de 5C.....	55
4.3 LAPP	55
4.4 Les systèmes des experts	56
4.5 Le Rating	56
4.6. Le Scoring.....	58
4.7. Les stress tests	59
5. La réglementation prudentielle	61
5.1 Les accords de Bâle I, II et III	61
5.2 Cadre réglementaire en Algérie	64
6. Le Machine Learning.....	68
6.1 Intelligence Artificielle et Machine Learning : concepts généraux	68
6.2 Les catégories d'apprentissage	70
7. Les algorithmes supervisés pour la classification binaire.....	72
7.1 Régression Linéaire	72
7.2 Les K-plus proches voisins (KNN)	72
7.3 Machines à Vecteurs de Support (SVM)	72
7.4 Arbre de décision	73
7.5 Les forêts aléatoires (Random Forest).....	73
7.6 Les Réseaux Neurones Artificiels (RNA).....	73
7.7 La régression logistique (RL).....	74
Chapitre 2 : CADRE MÉTHODOLOGIQUE ET ORGANISATIONNEL	75
Introduction du chapitre	76
Section 1 : Approche méthodologique et méthodes de collecte des données.....	76
1. Méthodologie Générale de la Recherche	76
1.1 Type et positionnement de la recherche.....	76
1.2 Justification du choix d'une méthode quantitative	77
1.3 Démarche méthodologique	77
2. Collecte des Données.....	78
2.1 Source des données.....	78

2.2 Mode de collecte : une démarche collaborative.....	78
2.3 Type de crédit, période et population étudiée	78
2.4 Définition de la variable cible	79
2.5 Description des Variables Explicatives.....	79
3. Présentation de l'outil et mise en œuvre empirique	84
3.1 Présentation du logiciel R et justification de son utilisation.....	84
3.2 La régression logistique : fondements théoriques et fonction mathématique.....	87
3.3 Étapes de l'analyse sous R.....	87
Section 2 : Contexte organisationnel	98
1. Présentation de la Trust Bank Algérie	98
2. Le Conseil d'Administration.....	99
3. Présentation de la direction générale	100
4. Le réseau de la Trust Bank Algeria.....	101
5. Présentation de la direction risque crédit.....	102
6. L'organigramme de Trust banque	102
Chapitre 3 : RÉSULTATS ET DISCUSSION	103
Introduction.....	104
Section 1 : Présentation des résultats	104
1. Analyse Exploratoire des Données (EDA)	104
1.1 Analyse univariée	105
1.1.1 Répartition de la variable cible.....	105
1.1.2 Répartition par forme juridique	106
1.1.3 Répartition par secteur d'activité	107
1.1.4 Distribution des variables financières (histogrammes).....	108
1.2 Analyse bivariée	110
1.2.1 Variables catégorielles selon la variable cible	110
1.2.2 Variables numériques selon une variable catégorielle	112
2. Résultats du Prétraitement des Données	115
2.1 Effet de la normalisation z-score (step_normalize)	115
2.2 Encodage One-Hot de la variable forme_juridique (step_dummy).....	116

3. Résultats du Modèle de Régression Logistique	117
3.1 Coefficients estimés du modèle	117
3.2 Courbe ROC et AUC	118
3.3. Matrice de confusion	120
3.4. Métriques de performance globales.....	121
Section 2 : Discussion des résultats	124
Conclusion du Chapitre III.....	128
CONCLUSION GÉNÉRALE.....	129
Références.....	130

Liste de tableau

Tableau 1: Ratios de structure	50
Tableau 2: Ratios d'activité	51
Tableau 3: Ratios de liquidité, endettement et solvabilité	52
Tableau 4: Les soldes intermédiaires de gestion	53
Tableau 5: Échelle de notation externe du risque de crédit	57
Tableau 6: Classement et Provisionnement des créances	67
Tableau 7: La Forme Juridique	83
Tableau 8: Le secteur d'activité	84
Tableau 9: Packages R utilisés dans cette étude	86
Tableau 10: Synthèse de l'exploration initiale des données	89
Tableau 11: Découpage des données	93
Tableau 12: Comparaison des variables avant et après normalisation	94
Tableau 13: Matrice de Confusion — Analyse détaillée des erreurs	96
Tableau 14: les actionnaires de la TBA	99
Tableau 15: Synthèse de la matrice de confusion du modèle de classification	120

Liste de figure

Figure 1: Classification des crédits d'exploitation	38
Figure 2: Les principales composantes du risque de crédit	44
Figure 3: Quatre définition de l'IA.....	68
Figure 4: Réseaux d'agences de la TBA.....	101
Figure 5: Répartition de la variable cible	105
Figure 6: Répartition par secteur d'activité	107
Figure 7: Distribution des variables financiers	108
Figure 8: Forme juridique selon la variable cible.....	110
Figure 9: Secteur d'activité selon la variable cible.....	111
Figure 10: Ratios financiers selon la forme juridique (boxplots)	112
Figure 11: Ratios financiers selon la variable cible — Analyse bivariée centrale.....	113
Figure 12: Effet de la normalisation z-score (step_normalize)	115
Figure 13: Encodage One-Hot de la variable forme_juridique (step_dummy).....	116
Figure 14: Coefficients estimés du modèle	117
Figure 15: Courbe ROC	118
Figure 16: Matrice de confusion.....	120
Figure 17: Métriques de performance	121

Liste des abréviations

ML : Machine Learning

PME : Petites et Moyenne Entreprise

TBA : TRUST Bank Algérie

R : Logiciel R

IA : Intelligence Artificielle

RL : Régression Logistique

SVM : Machines à vecteurs de support

ANN : Artificial Neural Network

KNN : K-Nearest Neighbors

ROC : Receiver Operating Characteristic

AUC : Area Under Curve

PIB : Produit Intérieur Brut

PA : Perte attendue

PD : Probability of Default

LGD : Loss Given Default

EAD : Exposure At Default

FR : Fonds de roulement

BFR : Besoin en fonds de roulement

TN : Trésorerie Nette

CPA : Crédit Populaire d'Algérie

BNA : Banque Nationale d'Algérie

BADR : Banque de l'Agriculture et du Développement Rural

BTPH : Bâtiment, Travaux Publics et Hydraulique

SARL : Société À Responsabilité Limitée

EURL : Entreprise Unipersonnelle à Responsabilité Limitée

SPA : Société par actions

SNC : Société en Nom Collectif

INTRODUCTION GÉNÉRALE

Le crédit bancaire constitue l'une des principales missions des établissements financiers. En accordant des financements aux entreprises et aux particuliers, les banques jouent un rôle d'intermédiaire indispensable dans l'économie : elles mobilisent les ressources auprès des agents disposant d'un excédent et les redistribuent à ceux qui en ont besoin (De Coussergues, 2020). Cette activité est cependant indissociable d'un risque fondamental : Le risque de crédit se définit comme la probabilité qu'un emprunteur ne puisse pas honorer, totalement ou partiellement, ses engagements financiers à l'échéance convenue.

Dans le segment des petites et moyennes entreprises (PME), ce risque prend une dimension particulière. Les PME se caractérisent par une structure financière souvent vulnérable, une information comptable limitée et une capacité de remboursement fortement sensible aux fluctuations économiques. (Berger, 1998) ont mis en évidence l'existence d'une asymétrie d'information plus prononcée pour ces entreprises : la banque ne dispose pas des mêmes informations que l'emprunteur sur sa véritable situation financière, ce qui rend l'évaluation du risque plus incertaine et plus complexe. Dans le contexte algérien, cette difficulté est amplifiée par l'importance de l'économie informelle, estimée à environ 33 % du PIB, et par les pratiques fréquentes de manipulation des comptes qui fragilisent la fiabilité des documents financiers fournis par les PME (Gliz, 2012).

C'est précisément pour répondre à cette incertitude que les banques algériennes cherchent à développer des outils d'évaluation du risque plus rigoureux, plus objectifs et plus fiables. Notre recherche s'inscrit dans cette démarche, en s'intéressant à l'application de la régression logistique —un algorithme d'apprentissage automatique — à la prédiction du défaut de paiement des PME sur les crédits d'exploitation, à partir des données du système de notation Moody's de TRUST Bank Algérie.

Les méthodes traditionnelles d'évaluation du risque de crédit : analyse financière classique, méthode des ratios, méthode des 5C, jugement expert du chargé d'affaires ont longtemps constitué le socle des pratiques bancaires en Algérie. Ces approches présentent des limites réelles : elles reposent en partie sur des critères subjectifs, elles sont sensibles à la qualité des données comptables fournies par les emprunteurs, et elles ne permettent pas de traiter efficacement un grand nombre de dossiers de façon homogène. Les nouvelles méthodes d'évaluation du risque de crédit notation financière, stress tests, modèles statistiques avancés sont apparues précisément pour dépasser ces limites en apportant une mesure plus objective et systématique du risque.

Parmi ces nouvelles méthodes, les algorithmes de Machine Learning (ML) ont suscité un intérêt croissant dans la communauté académique et bancaire. Contrairement aux approches statistiques classiques, ces algorithmes apprennent automatiquement les caractéristiques qui distinguent les emprunteurs solvables des emprunteurs défaillants, à partir de données historiques. (Bhatore, 2020) Soulignent que leur capacité à traiter un grand nombre de variables et à capturer des relations complexes représente un avantage majeur. Les travaux de (Baesens, 2003), puis de (Lessmann, 2015), ont démontré que des algorithmes comme les forêts aléatoires et les réseaux de neurones surpassent souvent les méthodes classiques en termes de performance prédictive.

Cependant, dans le contexte bancaire algérien, ces méthodes avancées se heurtent à un obstacle de taille : leur complexité. La régression logistique reste à ce jour la seule méthode dont les résultats sont directement compréhensibles par un analyste financier sans formation en intelligence artificielle (Tari, Bouguerra, & Selatni, 2024). Ses coefficients peuvent être traduits en règles de décision claires, ce qui la rend compatible avec les exigences de transparence des accords de Bâle II et Bâle III. Pour autant, les études algériennes disponibles lui attribuent des performances prédictives relativement modestes :(Souadda, Menari, & Berghout, 2022)obtiennent une AUC de 62,7 % avec la régression logistique sur 282 dossiers du CPA, et (Tari, Bouguerra, & Selatni, 2024) n'atteignent que 59 % sur 647 dossiers. Ces résultats décevants s'expliquent-ils par les limites intrinsèques de l'algorithme, ou reflètent-ils simplement la mauvaise qualité des données utilisées ?

C'est cette question précise que notre recherche se propose d'explorer. Nous disposons en effet d'une base de données exceptionnelle dans le contexte algérien : 255 dossiers de crédit d'exploitation de PME, issus du système de notation Moody's de TRUST Bank Algérie sur la période 2020–2024, sans aucune valeur manquante, avec sept ratios financiers calculés de façon standardisée par un système professionnel reconnu. Cette qualité de données, absente des études algériennes précédentes, constitue le point de départ original de notre travail.

L'ensemble de ces constats nous conduit à formuler la problématique centrale de notre recherche :

Dans quelle mesure les algorithmes de Machine Learning, et en particulier la régression logistique, permettent-ils de prédire efficacement la défaillance des PME sur les crédits d'exploitation en Algérie ?

Cette question centrale se décline en trois interrogations subsidiaires qui structurent notre démarche analytique. La première interroge le pouvoir discriminant des ratios financiers disponibles : les indicateurs d'endettement et de gestion du cycle d'exploitation issus du système Moody's permettent-ils de séparer efficacement les PME saines des PME défaillantes ? La deuxième interroge la performance du modèle : la régression logistique peut-elle, dans des conditions optimales de qualité de données, dépasser les niveaux de performance documentés dans la littérature algérienne et atteindre un seuil suffisant pour un usage opérationnel bancaire ? La troisième interroge l'utilité pratique : les résultats obtenus permettent-ils de valider la régression logistique comme outil d'aide à la décision d'octroi de crédit dans le contexte d'une banque algérienne utilisant un système de notation professionnel ?

Pour répondre à ces questions, nous formulons trois hypothèses de recherche directement testables à partir de nos données :

H1 : Les ratios d'endettement permettent de discriminer significativement les entreprises défaillantes des entreprises saines. Cette hypothèse s'appuie sur les travaux de (Gliz, 2012), qui identifient l'endettement comme le déterminant le plus puissant du risque de défaut dans le contexte des PME algériennes.

H2 : Les ratios de gestion du cycle d'exploitation — délais clients, délai stocks et délai fournisseurs — sont des indicateurs prédictifs significatifs du risque de défaut sur les crédits d'exploitation. Cette hypothèse est fondée sur les travaux de (Ampountolas, Nyarko Nde, Date, & Constantinescu, 2021) et de (Souadda, Menari, & Berghout, 2022), qui soulignent l'importance de ces variables pour prédire les défauts sur les crédits à court terme des PME.

H3 : La régression logistique permet d'obtenir une performance prédictive acceptable — mesurée par un ROC-AUC supérieur à 0,70 sur notre échantillon, malgré les contraintes de taille et le déséquilibre entre les classes. Cette hypothèse est formulée en référence aux résultats de (Souadda, Menari, & Berghout, 2022) et (Tari, Bouguerra, & Selatni, 2024), dont les AUC de 62,7 % et 59 % avec la régression logistique constituent les benchmarks de comparaison pour notre étude.

Pour répondre à notre problématique et tester ces hypothèses, nous avons adopté une démarche quantitative et empirique, inscrite dans un paradigme positiviste hypothético-déductif. Notre étude repose sur l'analyse d'une base de données réelle de 255 dossiers de crédit d'exploitation

de PME, collectés auprès de la direction risque crédit de TRUST Bank Algérie sur la période 2020–2024 dans le cadre d'une démarche collaborative. La variable à prédire est la variable cible binaire : sain (1) ou défaillant (0), définie conformément à la norme réglementaire de Bâle II qui fixe le seuil de défaut à 90 jours d'impayés.

L'ensemble de l'analyse a été réalisé sous le logiciel R, en mobilisant le framework tidymodels. La démarche suit quatre étapes successives : une analyse exploratoire graphique approfondie des données, un prétraitement rigoureux incluant la normalisation z-score des variables numériques et l'encodage One-Hot de la variable catégorielle, une modélisation par régression logistique avec découpage stratifié train/test (80 %/20 %), et une évaluation des performances à travers des métriques complémentaires : ROC-AUC, accuracy, matrice de confusion, précision, rappel et F1-score. Ce protocole garantit une évaluation non biaisée de la capacité prédictive du modèle sur des données qu'il n'a pas vues pendant son entraînement.

Notre travail est organisé en trois chapitres articulés selon une progression logique allant du général au particulier, du théorique à l'empirique.

Le premier chapitre, intitulé « Cadre conceptuel et revue de la littérature », pose les fondements théoriques et empiriques de notre recherche. Il est structuré en deux sections. La première section présente la revue des travaux académiques sur l'évaluation du risque de crédit par Machine Learning — des études fondatrices (Thomas, 2000 ; Baesens et al., 2003) aux recherches récentes dans les pays émergents et en Algérie (Touati-Tliba, 2003 ; Gliz et Touati-Tliba, 2012 ; Souadda et al., 2022 ; Tari et al., 2024) — et en dégage la synthèse, les gaps identifiés, la problématique et les hypothèses. La seconde section développe le cadre conceptuel : définition et rôle de la banque, typologies du crédit bancaire, spécificités du crédit d'exploitation, définition et enjeux du risque de crédit, méthodes d'évaluation traditionnelles et nouvelles, réglementation prudentielle de Bâle I, II et III, et présentation des principaux algorithmes de Machine Learning supervisés applicables à la classification du risque.

Le deuxième chapitre, intitulé « Cadre méthodologique et organisationnel », présente en deux sections le dispositif de recherche et le terrain d'étude. La première section expose la méthodologie générale de la recherche —type et positionnement épistémologique, source et mode de collecte des données, description de la base et de la variable cible, présentation du logiciel R, fondements théoriques de la régression logistique, métriques d'évaluation, et détail des étapes de l'analyse. La seconde section présente le contexte organisationnel de l'étude : TRUST Bank Algérie, sa direction risque crédit, son réseau et son système de notation Moody's.

Le troisième chapitre, intitulé « Résultats et Discussion », constitue le cœur empirique du mémoire. La première section présente et interprète en profondeur l'ensemble des résultats graphiques produits sous R, depuis l'exploration des données jusqu'aux métriques finales du modèle, en construisant un raisonnement progressif où chaque graphique justifie et enrichit le suivant. Les trois hypothèses sont confirmées ou infirmées au fil des résultats. La deuxième section discute ces résultats à la lumière de la revue de littérature, explique les similitudes et les différences avec les études existantes, dégage les implications opérationnelles pour le système bancaire algérien, et signale les limites de l'étude et les perspectives de recherche futures.

Une conclusion générale viendra clore notre travail en synthétisant les principaux enseignements, en répondant à notre problématique et en proposant des recommandations pratiques à destination des établissements bancaires algériens.

Chapitre 1 : cadre conceptuel et revue de la littérature

Introduction :

Le risque de crédit représente l'une des préoccupations les plus fondamentales de l'activité bancaire. Avant de pouvoir l'évaluer, le modéliser et le prédire, il est indispensable d'en maîtriser les fondements théoriques et de connaître l'état des travaux académiques qui ont contribué à le comprendre. C'est l'objet de ce premier chapitre.

Il s'articule autour de deux sections complémentaires. La première section est consacrée à la revue de la littérature : elle retrace l'évolution des travaux académiques sur l'évaluation du risque de crédit par apprentissage automatique, des études fondatrices aux recherches les plus récentes dans le contexte algérien, et aboutit à la formulation de notre problématique et de nos trois hypothèses de recherche. La seconde section développe le cadre conceptuel indispensable à la compréhension de notre démarche : elle définit les notions de banque, de crédit d'exploitation, de risque de crédit et de défaut, présente les méthodes d'évaluation traditionnelles et nouvelles, expose les exigences de la réglementation prudentielle de Bâle II et Bâle III, et introduit les algorithmes de Machine Learning utilisés dans la littérature pour la classification du risque de crédit.

SECTION 01 : Revue de la littérature

Le risque de crédit représente l'un des risques les plus anciens et les plus fondamentaux auxquels sont exposés les établissements bancaires. Il se définit comme la probabilité qu'un emprunteur ne soit pas en mesure de rembourser tout ou partie de ses engagements financiers à l'échéance. Ce risque est particulièrement élevé dans le segment des petites et moyennes entreprises (PME), caractérisées par une structure financière souvent vulnérable, une information comptable limitée et une capacité de remboursement fortement sensible aux fluctuations économiques.

1.1. Les techniques d'évaluation du risque de crédit :

Dans ce cadre, (Berger, 1998) ont mis en évidence l'existence d'une asymétrie d'information plus marquée pour les PME, rendant l'évaluation de leur solvabilité plus complexe pour les banques. Face à cette contrainte, les institutions financières ont progressivement adopté des approches quantitatives afin de renforcer l'objectivité des décisions d'octroi de crédit et de limiter les biais liés au jugement humain.

C'est dans ce contexte les méthodes d'apprentissage automatique (Machine Learning) se sont imposées comme des outils innovants d'évaluation du risque de crédit. Contrairement aux approches traditionnelles fondées sur le jugement expert, ces algorithmes apprennent automatiquement les caractéristiques qui distinguent les emprunteurs solvables des emprunteurs défaillants, à partir de données historiques. Bhatore, Mohan et Reddy (2020) soulignent que cette capacité à traiter un grand nombre de variables et à capturer des relations complexes représente un avantage majeur par rapport aux modèles classiques.

1.2. L'évolution des modèles de prédiction du défaut : des approches classiques au Machine Learning :

Les premières recherches sur l'application des méthodes quantitatives à l'évaluation du risque de crédit remontent aux années 1980. Deux recherches fondamentales ont notamment contribué à structurer ce champ d'étude dans l'objectif de synthétiser les modèles modernes d'apprentissage automatique appliqués à la gestion du risque de crédit, (Thomas, 2000) propose l'une des premières études comparatives des modèles d'évaluation supervisés.

L'auteur examine six techniques, à savoir : les arbres de décision, les réseaux de neurones, la régression logistique, la régression linéaire, l'analyse discriminante ainsi que les méthodes non paramétriques. Les résultats montrent qu'aucune méthode ne domine réellement les autres et que les écarts de performance entre les approches sont le plus souvent très faibles.

Par ailleurs, (Baesens, 2003) ont mené une étude similaire, mais avec des conclusions légèrement différentes. Les auteurs ont comparé 17 algorithmes de classification sur huit ensembles de données de crédits bancaires internationaux. Leurs résultats indiquent que les machines à vecteurs de support (SVM) et les réseaux de neurones artificiels (ANN) obtiennent les meilleures performances prédictives, avec des aires sous la courbe ROC (ASC) respectives de 66 % et 91 %.

(Lessmann, 2015) réalisent une mise à jour de l'étude de (Baesens, 2003) en comparant 41 algorithmes de classification sur huit bases de données réelles de crédit retail. Leur objectif était de voir les nouveautés en matière d'apprentissage automatique et vérifier si les différentes façons de mesurer la performance donnent les mêmes résultats. Les résultats montrent que les ensembles hétérogènes (qui combinent plusieurs modèles) obtiennent les meilleures performances, surpassant les classifieurs individuels comme la régression logistique ou les réseaux de neurones. Les auteurs constatent que la régression logistique, bien qu'elle soit très

utilisée dans les banques, comme point de comparaison (référence), est significativement moins performante que les méthodes plus récentes (comme les forêts aléatoires ou les ensembles).

Ils recommandent d'utiliser Random Forest comme benchmark (comme référence) pour les futures études, tout en notant que les ensembles hétérogènes peuvent encore l'améliorer. Enfin, ils montrent qu'une meilleure précision statistique (moins d'erreurs) se traduit par une réduction des coûts d'erreur, ce qui justifie, d'un point de vue économique, d'utiliser des modèles plus complexes.

1.3. Les apports récents du Machine Learning dans la prédiction de la défaillance :

Dans une autre étude portant sur le pouvoir prédictif des algorithmes d'apprentissage automatique dans la gestion du risque de crédit, (Moscatelli, 2020) ont comparé la performance de classification des modèles traditionnels, notamment la régression logistique, à celle des modèles issus de l'apprentissage automatique (Machine Learning). Leurs résultats montrent que le Machine Learning améliore la prévision du risque de défaut par rapport aux méthodes classiques, surtout quand on a peu d'informations publiques et beaucoup de données. De plus, en Europe, cette étude conclut que le ML réduit les biais introduits par les critères subjectifs, rendant ainsi le processus d'octroi de crédit plus équitable.

(Addo, 2018) ont comparé différents modèles de machine learning et de deep learning pour la prédiction du défaut de paiement. Leur étude montre que les modèles fondés sur les arbres de décision (forêt aléatoire, gradient boosting) sont plus stables que les réseaux de neurones profonds. Les auteurs concluent que le deep learning ne fournit pas forcément de meilleurs résultats dans ce contexte, remettant en question son utilisation intensive au sein des entreprises.

Dans leur étude, (Zhou, 2020) montrent que le deep learning est particulièrement efficace pour traiter des données de crédit non structurées, telles que les textes ou les historiques de transactions. Grâce à sa capacité à extraire automatiquement des informations complexes, il augmente de 10 % la précision des prédictions par rapport aux approches traditionnelles. Les auteurs soulignent toutefois que le deep learning nécessite des quantités massives de données pour atteindre son plein potentiel.

(He, 2019) comparent, sur un échantillon de prêts bancaires en Chine, la performance de la régression logistique, des SVM, des arbres de décision et du gradient boosting. Leurs résultats montrent que les modèles de Machine Learning améliorent l'AUC d'environ 15 % par rapport à la régression logistique. L'étude montre aussi l'importance des variables comportementales

(historique des transactions, usage des services bancaires en ligne) qui, lorsqu'elles sont ajoutées aux ratios financiers traditionnels, améliorent nettement le pouvoir discriminant des modèles. Cette recherche montre l'intérêt de mobiliser des données alternatives dans les pays où l'on dispose parfois mal de l'information comptable des PME.

1.4. Les expériences dans les pays émergents et en Afrique :

En Tunisie, (Matoussi, 2010) a mené une étude portant sur 1 435 demandes de crédit accordées à des entreprises industrielles tunisiennes sur la période 2003 et 2006. L'auteur a comparé différents modèles de prédiction du défaut de paiement, notamment la régression logistique (RL) et les réseaux de neurones artificiels (ANN), afin d'évaluer le risque de crédit au sein d'une banque commerciale. Les résultats obtenus mettent en évidence la performance supérieure des réseaux de neurones artificiels, qui atteignent un taux global de classification correcte de 89,9 %, surpassant ainsi les approches statistiques traditionnelles.

En Afrique, (Ampountolas A. N., 2021) ont réalisé une analyse comparative de plusieurs techniques d'apprentissage supervisé, incluant notamment les arbres de décision et la forêt aléatoire, sur un ensemble de 4 450 dossiers de crédit d'exploitation couvrant une période de quatre ans en Guinée. L'évaluation des modèles s'est appuyée sur différents indicateurs de performance. Les résultats de l'étude indiquent que l'algorithme des arbres de décision présente la meilleure efficacité parmi les méthodes testées.

Dans une étude récente, (Tripathi, 2025) s'est intéressé à l'utilisation du Machine Learning pour l'évaluation du risque de crédit dans le secteur bancaire indien, dans le but de réduire les créances douteuses et de permettre des décisions plus rapides. L'auteur a utilisé une méthodologie mixte, c'est-à-dire à la fois quantitative et qualitative. Il a réalisé une enquête auprès de 52 professionnels du secteur bancaire et financier. Les données ont été collectées à l'aide d'un questionnaire en ligne. Les résultats montrent clairement que le Machine Learning est encore peu utilisé dans les banques indiennes. La majorité des banques continuent à utiliser des méthodes traditionnelles, et quand le Machine Learning est présent, il est souvent utilisé en complément, pas comme un outil principal en raison de plusieurs obstacles tels que le manque de compétences, la qualité des données et les contraintes réglementaires.

Enfin, l'auteur souligne l'importance de Machine Learning, qui apporte plusieurs avantages, notamment des décisions plus rapides, une meilleure analyse des clients et une réduction du risque de défaut

Une étude récente (Guo, 2024) a testé les arbres de décision pour anticiper les défauts de paiement dans les banques. À partir de données réelles de prêts (montant, durée, épargne, solde des comptes, historique de crédit), le modèle a correctement identifié la plupart des prêts à risque, avec seulement 15 % d'erreurs. Les variables les plus importantes sont le solde du compte courant, le solde du compte d'épargne et l'historique de crédit.

1.5. La prédiction du défaut des PME dans le contexte bancaire algérien :

Dans une étude fondatrice, (Touati-Tliba, 2003) examine comment les banques publiques algériennes prennent leurs décisions d'octroi de crédit. L'auteur montre que les modèles statistiques, en particulier la régression logistique, aident à mieux décider d'accorder ou non un prêt. Il constate que la variable « mouvements confiés » (c'est-à-dire le volume des transactions qui passent par le compte bancaire) est un bon indicateur pour repérer les entreprises qui risquent de faire défaut. Cependant, l'étude souligne que ces modèles sont limités par la mauvaise qualité des données disponibles : les informations comptables des entreprises sont souvent incomplètes ou peu fiables. Les banques doivent donc s'appuyer aussi sur des éléments plus subjectifs, comme l'expérience du dirigeant ou la qualité de la relation avec la banque (variables qualitatives).

Dans la même perspective, (Gliz, 2012) s'intéressent à la façon dont les banques évaluent le risque de défaut des PME à partir des informations comptables. Ils utilisent un modèle de régression logistique sur des crédits accordés par une banque publique algérienne. Leurs résultats montrent que certaines variables financières – notamment l'endettement et la trésorerie – sont importantes pour distinguer les entreprises saines de celles qui sont en difficulté. En revanche, les indicateurs de rentabilité ne sont pas fiables, à cause de l'économie informelle (qui représente environ 33 % du PIB) et de la manipulation des comptes. Les auteurs soulignent que les banques ne se fient pas seulement aux chiffres comptables : elles utilisent aussi des informations extérieures, comme la situation fiscale ou l'historique des mouvements bancaires. Ainsi, cette étude met en lumière les limites des modèles classiques dans un pays où l'information est souvent incomplète ou peu fiable.

Plus récemment, (Souadda, Menari, & Berghout, 2022) comparent trois modèles (régression logistique, K-plus proches voisins et réseaux de neurones) sur 282 dossiers de crédit d'exploitation du CPA. Les réseaux de neurones obtiennent les meilleurs résultats (AUC de 87,1 %), suivis du KNN (80,3 %) et de la régression logistique (62,7 %). Même si la régression logistique est moins performante, les auteurs notent qu'elle reste facile à comprendre et qu'elle

identifie bien les variables importantes (mouvements confiés, situation à la centrale des risques).

Enfin, (Tari, Bouguerra, & Selatni, 2024) élargissent l'analyse avec 647 demandes de crédit et comparent quatre modèles : régression logistique, arbre de décision, réseaux de neurones et forêt aléatoire. La forêt aléatoire donne la meilleure performance (AUC de 91 %), suivie des réseaux de neurones (82 %), de l'arbre de décision (76 %) et de la régression logistique (59 %). Les auteurs expliquent que la régression logistique est moins bonne parce qu'elle ne peut pas saisir les relations non linéaires entre les variables. Cependant, ils rappellent que la régression logistique reste la seule méthode dont les résultats sont directement compréhensibles par un analyste financier sans formation en intelligence artificielle.

L'ensemble de ces travaux montre que les modèles de Machine Learning avancés (réseaux de neurones, forêt aléatoire) sont généralement plus précis que la régression logistique. Cependant, cette dernière présente plusieurs atouts (avantages) dans le contexte algérien : sa mise en œuvre est simple, ses résultats sont faciles à interpréter, elle peut intégrer des variables clés comme les mouvements confiés, et elle est mieux acceptée par les banques ainsi que par les autorités de contrôle. Par ailleurs, les études de (Touati-Tliba, 2003) et de Gliz et Touati-Tliba (2012) soulignent que la faible qualité des données comptables et l'asymétrie d'information (c'est-à-dire le fait que la banque ne connaît pas aussi bien que l'emprunteur sa véritable santé financière) constituent des obstacles réels. Dans ce cadre, une régression logistique correctement préparée peut constituer un outil efficace et transparent pour les banques algériennes.

C'est pourquoi notre recherche se propose d'évaluer spécifiquement la capacité de la régression logistique à prévoir le défaut de paiement des PME sur les crédits d'exploitation en Algérie. La question principale qui guide cette recherche est la suivante :

Problématique :

Dans quelle mesure les algorithmes de Machine Learning, et en particulier la régression logistique, permettent-ils de prédire efficacement la défaillance des PME sur les crédits d'exploitation en Algérie ?

Pour répondre à cette problématique, nous formulons les trois hypothèses de recherche suivantes :

H1 : Les ratios d'endettement permettent de discriminer significativement les entreprises défaillantes des entreprises saines (selon Les travaux de Gliz et TouatiTliba)

H2 : Les ratios de gestion du cycle d'exploitation (délais clients, stocks, fournisseurs) sont des indicateurs prédictifs du risque de défaut (Ampountolas A. N., 2021)et (Souadda, Menari, & Berghout, 2022)

H3 : La régression logistique permet d'obtenir une performance prédictive acceptable ($AUC > 0,70$) sur notre échantillon, malgré les contraintes de taille et de déséquilibre des classes (Souadda, Menari, & Berghout, 2022)et (Tari, Bouguerra, & Selatni, 2024)

Section2 : Cadre conceptuel

1. Généralités concernant la banque

Une banque est une institution financière ayant pour mission principale la collecte des dépôts auprès du public ainsi que l'octroi de crédits. Elle occupe une place essentielle dans le financement de l'économie en assurant le rôle d'intermédiaire entre les agents disposant d'excédents de capitaux (épargnants) et ceux qui recherchent des ressources financières (emprunteurs). Dans cette section, nous présentons la définition de la banque, présenter ses opérations fondamentales, puis expliquer son rôle en tant qu'intermédiaire financier.

1.1 Définition de la banque :

Historique :

L'origine du mot "Banque" est italienne, dérivant du terme "Banca". Ce mot signifiait littéralement une simple "table en bois" (un banc). À l'époque médiévale, les changeurs d'argent s'installaient sur cette table, en public, dans les villes marchandes. Ils y échangeaient les monnaies et accordaient des prêts. Le nom de cet instrument de travail a ainsi évolué pour désigner, par la suite, l'institution financière que nous connaissons aujourd'hui

La notion de banque a connu plusieurs définitions au fil du temps. (Larousse, 2020) la décrit comme une institution financière dont le rôle est double : collecter l'épargne du public et l'utiliser pour accorder des crédits, tout en assurant la gestion des moyens de paiement.

(Bagehot, 1873) définit la banque comme une entité dont les missions fondamentales sont la collecte, le prêt, l'échange et le transfert d'argent, ainsi que la fourniture de services financiers variés à sa clientèle.

Selon (Bezbakh, 2000), La banque est une entreprise financière qui collecte les fonds du public, notamment sous forme de dépôts ou d'épargne. Elle réutilise ces ressources en accordant des crédits et en réalisant diverses opérations financières. Par ailleurs, elle assure la gestion et la mise à disposition de sa clientèle de différents moyens de paiement tels que les chèques, les cartes bancaires et les virements, entre autres.

La banque peut être définie comme une personne morale exerçant de manière habituelle des opérations bancaires, notamment la collecte des fonds, l'octroi de crédits ainsi que la mise à disposition de la clientèle de différents moyens de paiement. En plus à ces opérations, l'établissement crédit mène d'autres opérations connexes. Ces dernières concernent les

opérations de change (monnaie étrangère), opération sur les métaux précieux tels que l'or, placement des produits financiers, gestion du patrimoine de certains clients et location des biens (Vigouroux, 1991) .

La banque est un établissement de crédit qui effectue de façon habituelle des opérations de réception de fonds, d'octroi de crédits, tout en mettant en place des moyens de paiement et de gestion de ces derniers. C'est l'intermédiaire entre l'offre et la demande du capital (Garsault, 1997).

La banque est une société qui collecte l'argent de ses clients moral et physique ; en contrepartie, elle leur propose des services et des intérêts sur les dépôts d'une part, et finance des projets en leur accordant des crédits d'autre part. C'est un établissement qui propose des services financiers (Haddad, 2015).

Selon(Patat, 1993), les banques sont des organismes financiers qui enregistrent, au passif de leur bilan, les comptes de leur clientèle, lesquels peuvent servir à effectuer des opérations de paiement telles que les chèques ou les virements, dans la limite des fonds disponibles. Elles assurent également des prestations de services liées aux règlements et aux transferts de fonds, tout en exerçant une activité de distribution de crédits.

Sur le plan juridique, la banque est définie comme une personne morale à caractère commercial disposant d'un capital propre. À ce titre, elle est soumise au principe d'autonomie financière ainsi qu'à l'obligation d'assurer l'équilibre comptable, conformément à (Loi n° 88-01 du 12 janvier 1988)

En général, les banques sont des sociétés qui ont des activités financières. Ces entreprises proposent des services financiers en s'occupant de la gestion des dépôts d'argent de leurs clients et au mode de paiement. Juridiquement parlant, elles se définissent comme des institutions financières soumises au code monétaire et financier. Il existe différents types de banques, la différence entre eux réside dans leurs statuts juridiques (Guigal, 2011).

1.2 Le rôle de la banque dans l'économie :

Le secteur bancaire joue un rôle central dans l'économie en assurant plusieurs fonctions essentielles qui facilitent le développement économique et financier. Ces fonctions sont variées et couvrent des aspects cruciaux tels que l'intermédiation financière, le financement de l'économie, la gestion des moyens de paiement, la sécurisation de l'épargne, la contribution à la

politique monétaire, la réduction des risques financiers, et la stimulation de l'innovation et du commerce international. (Karyotis, 2015).

A. Intermédiation financière :

Les banques servent d'intermédiaires entre les épargnants et les emprunteurs. Elles collectent l'épargne des ménages et des entreprises sous forme de dépôts et la redistribuent sous forme de crédits. Ce processus permet le financement de l'investissement et de la consommation, jouant ainsi un rôle crucial dans la mobilisation des ressources financières et leur allocation efficace dans l'économie. (Karyotis, 2015)

B. Création des comptes

Cette action se trouve complémentaire à la première, après la collecte des dépôts, la banque se doit de les placer dans des comptes.

On distingue deux types de comptes :

Compte à vue : également appelé compte courant, est un compte bancaire ouvert au nom du client et géré par la banque. Il permet au titulaire de disposer librement des fonds déposés à tout moment. En règle générale, ce type de compte n'est pas rémunéré, mais il donne accès à plusieurs services bancaires tels que le carnet de chèques, la carte bancaire, les virements ainsi que les opérations de retrait ou de paiement automatique. Ces différents services font l'objet d'une commission forfaitaire.

Compte de dépôt : Cette catégorie de compte est destinée aux clients ayant l'intention de déposer des épargnes. Les clients bénéficieront d'une marge sur leurs dépôts, avec un intérêt calculé en fonction du montant déposé, ainsi que de la durée de mobilisation de ce montant (Karyotis, 2016).

C. Refinancement de l'économie

Le problème majeur et éternel de l'économie est celui de la rareté des ressources. Ces dernières sont un facteur primordial pour le développement de l'économie d'une nation en assurant le refinancement afin de garantir l'évolution des activités menées par les agents économiques. Ce refinancement se réalise à travers les fonds propres dont la banque dispose, et qui seront dirigés vers Les entreprises, afin de financer leurs projets, ainsi qu'à des crédits aux familles. Autrement dit, il s'agit de l'octroi de crédits aux particuliers ainsi qu'aux entreprises (De Coussergues, 2020).

D. Création de la monnaie

Les banques commerciales ont un rôle crucial dans la génération de monnaie via le processus d'attribution de prêts. Quand une banque octroie un crédit, elle crée de la monnaie scripturale qui fait croître la masse monétaire en circulation. Toutefois, ce mécanisme est régulé de manière rigoureuse par la politique monétaire mise en œuvre par la Banque centrale., qui emploie des instruments comme les réserves obligatoires et le taux directeur pour maîtriser la liquidité bancaire (Mishkin, 2019).

1.3 Les Types de banques :

Les banques peuvent être regroupées selon différents critères, notamment leur statut juridique, leur degré de spécialisation ainsi que la nature et l'étendue de leurs activités. La compréhension de cette classification est importante pour appréhender leur rôle au sein de l'économie et leur aptitude à satisfaire les besoins financiers des entreprises comme des particuliers. (Ilunga, 2016)

Selon le statut juridique :

La classification des établissements bancaires en fonction de leur statut juridique permet de distinguer trois principales catégories : les banques publiques, les banques privées et les banques mixtes.

Les banques publiques :

Les banques publiques sont des institutions financières dont la propriété ainsi que le contrôle sont assurées par l'État.

Les banques privées :

Les banques privées correspondent à des établissements dont le capital est détenu par un individu ou un groupe d'actionnaires privés. Les décisions stratégiques et opérationnelles y sont prises par les propriétaires du capital, qui détiennent les droits de gouvernance de la banque.

Les banques mixtes :

Les banques mixtes se caractérisent par une structure de capital associant participation publique et privée. Dans ce type d'établissement, l'État et les investisseurs privés partagent les droits de décision et de gestion.

Selon la nature d'activité :

En fonction de la spécialisation et de l'étendue des activités exercées, on distingue plusieurs catégories bancaires : la banque centrale, les banques de détail, les banques d'affaires (ou d'investissement), les banques islamiques ainsi que les banques généralistes.

Les banques centrales :

Les banques centrales occupent une position spécifique dans le système bancaire, puisqu'elles constituent les institutions de référence pour les banques commerciales. Elles sont chargées de définir et de mettre en œuvre la politique monétaire, tout en veillant à la stabilité du système financier national. Leurs principales missions incluent l'émission de la monnaie fiduciaire, la régulation des taux d'intérêt et le contrôle prudentiel des établissements financiers en matière de gestion des risques (Crédits, s.d.)

Banque de détail :

Également désignée sous les termes de banque commerciale ou banque de dépôt, la banque de détail est un établissement financier destiné principalement aux particuliers, aux petites et moyennes entreprises, aux professions libérales, aux associations ainsi qu'aux collectivités locales. Son activité principale consiste à transformer les dépôts collectés en crédits, tout en proposant une gamme de services bancaires courants tels que la gestion de comptes, les moyens de paiement, les crédits, les produits d'épargne et les assurances-vie. Elle s'appuie généralement sur un réseau d'agences étendu afin d'assurer une proximité avec sa clientèle (Maymo, 2015).

La Banque d'affaires (d'investissement) :

La banque d'affaires est un établissement financier spécialisé dans l'accompagnement des opérations de haut de bilan, notamment les introductions en bourse, les augmentations de capital ainsi que les opérations de fusion-acquisition. Elle intervient principalement auprès des grandes entreprises industrielles et commerciales, ainsi que parfois auprès des États, et peut également gérer ses propres participations financières (Crédits, s.d.).

La Banque islamique :

Les banques islamiques sont des institutions financières qui exercent leurs activités conformément aux principes de la charia. Leur fonctionnement repose sur un modèle de

financement basé sur le partage des profits et des pertes entre la banque et ses clients, en excluant le recours à l'intérêt.

Selon MAYMO (2015), La particularité de ces banques est qu'elles exercent leurs métiers en respectant la loi islamique. Les principaux changements concernent l'interdiction de l'intérêt et des investissements illicites (alcool ...), l'obligation de partage des pertes et des profits.

Banque généraliste (universelle) :

La banque universelle est un établissement financier capable d'offrir une gamme diversifiée de services bancaires à l'ensemble des catégories de clients, qu'il s'agisse de particuliers, d'entreprises ou d'institutions financières. Elle peut également exercer des activités de banque d'investissement et d'émission, ce qui lui permet de proposer des services complets et variés, la différenciant ainsi des établissements spécialisés.

C'est une banque multi-métiers qui exerce la totalité des services et activités proposées dans le domaine bancaire (GUERONNIERE, 2015)

2. Généralités sur le crédit

2.1 Définition et rôle économique du crédit

Le terme « crédit » trouve son origine dans le mot grec *credere*, signifiant croire ou accorder sa confiance. Il renvoie à l'idée selon laquelle le prêteur se fie à l'engagement de l'emprunteur de rembourser la somme obtenue après un délai convenu ; autrement dit, le créancier accorde sa confiance à son débiteur (Hadj Sadok, 2007)

En Algérie, conformément à l'article 68 de l'ordonnance n° 03-11 du 26 août 2003 relative à la monnaie et au crédit, une opération de crédit est définie comme tout acte à titre onéreux par lequel une personne met ou s'engage à mettre des fonds à la disposition d'une autre personne, ou prend, dans l'intérêt de celle-ci, un engagement par signature tel qu'un aval, un cautionnement ou une garantie.

Le crédit bancaire est la forme la plus fréquente de crédit, affirment (Kharoubi, 2016) il s'agit d'une ressource (une somme d'argent) mise à la disposition d'une personne physique ou morale (débiteur) par la banque (le créancier), contre la promesse d'être payée ou remboursée avec des intérêts, à une date précise (l'échéance). Donc, le crédit repose sur la confiance accordée par le créancier au débiteur.

Selon (Petit-Dutaillis, 1999), accorder un crédit revient à établir une relation de confiance, consistant à mettre à disposition immédiate un bien ou un pouvoir d'achat, en échange d'un engagement de restitution ultérieure d'un bien équivalent ou de même valeur, dans un délai déterminé.

Cette définition met en évidence les trois principales caractéristiques du crédit :

1. **Le facteur temps** : le délai entre l'octroi et le remboursement
2. **La confiance** : La relation de confiance établie entre le prêteur et l'emprunteur.
3. **La promesse** : l'engagement de remboursement

Le crédit correspond à une opération par laquelle un prêteur met à la disposition d'un emprunteur des fonds que ce dernier s'engage à rembourser selon des modalités définies (durée, taux d'intérêt, échéances).

Le crédit joue un rôle économique fondamental en tant qu'instrument d'intermédiation financière permettant le transfert des ressources temporaires des agents excédentaires (épargnants) vers les agents déficitaires (investisseurs et consommateurs), stimulant ainsi la croissance économique, les investissements productifs et la consommation.

Selon (De Coussergues, 2020), La banque joue un rôle fondamental au sein de l'économie en tant qu'intermédiaire financier. Elle mobilise les ressources financières auprès des agents disposant d'un excédent de financement, puis les redistribue aux agents économiques présentant un besoin de financement, sous forme de crédits. Ce rôle d'intermédiation est indispensable, dans la mesure où les banques assurent à la fois le traitement de l'information financière et la garantie de liquidité pour les agents économiques.

C'est ainsi que le crédit bancaire permet de financer l'économie réelle en transformant l'épargne en investissement et en consommation.

En d'autres termes, les banques sont indispensables parce qu'elles connaissent bien leurs clients (elles savent qui est capable de rembourser un prêt) et elles garantissent que l'argent sera disponible quand on en a besoin (les épargnants peuvent retirer leurs fonds à tout moment, et les emprunteurs obtiennent l'argent immédiatement). Grâce à ce double rôle, le crédit bancaire transforme l'épargne des uns en projets pour les autres, ce qui fait tourner l'économie.

2.2 Typologie des crédits bancaires

Les crédits bancaires peuvent être répartis selon différents critères de classification.: l'objet du crédit, sa durée, et ses caractéristiques contractuelles.

Selon (Bouyacoub, 2000), compte tenu de la diversité des besoins des agents économiques, il est possible de distinguer plusieurs catégories de crédits :

A- Selon l'objet du crédit

Les crédits pour les particuliers (personnes physiques)

Les crédits destinés aux particuliers répondent à des besoins personnels et familiaux.

Selon (Karyotis, 2015), les crédits accordés aux particuliers peuvent être répartis en deux principales catégories, à savoir les crédits à la consommation et les crédits immobiliers.

Les crédits à la consommation :

(Gourlaouen, 2012) Présente le crédit à la consommation comme étant un financement accordé aux ménages afin de leur permettre d'acquérir des biens ou services destinés à un usage personnel, à l'exclusion des opérations immobilières.

Il se présente sous plusieurs formes :

- Le prêt personnel : l'argent est versé sans que l'emprunteur ait à justifier son utilisation.
- Le crédit affecté : il s'agit d'un type de prêt destiné au financement d'un bien déterminé (par exemple, un véhicule). Dans cette situation, les fonds ne sont pas directement versés à l'emprunteur, mais sont transférés au vendeur du bien concerné.
- Le crédit renouvelable (revolving) : Il s'agit d'une réserve d'argent qui se reconstitue progressivement au rythme des remboursements effectués. Ainsi, l'emprunteur peut de nouveau utiliser les montants remboursés, dans la limite d'un plafond préalablement défini.

D'après (Luc, 2008): le crédit à la consommation constitue un produit bancaire destiné aux particuliers, leur permettant d'acquérir des biens d'équipement domestique, qu'ils soient fabriqués localement ou importés, notamment les automobiles, grâce à des facilités de paiement. Ce type de financement est accordé aux personnes résidant sur le territoire national, disposant d'une activité professionnelle stable et d'un revenu régulier. Le montant du crédit peut atteindre jusqu'à 70 % du coût total du bien financé.

Le crédit immobilier :

Le crédit immobilier est un prêt accordé par un établissement de crédit pour financer tout ou partie de l'acquisition d'un bien immobilier, d'une opération de construction, ou des travaux sur un tel bien. C'est un crédit à long terme, généralement sur 10 à 30 ans.

Selon Sylvie de Coussergues :

Le crédit immobilier correspond à un financement accordé aux ménages ou aux entreprises afin de permettre l'acquisition, la construction ou la rénovation d'un bien immobilier.

Les crédits destinés aux entreprises et aux professionnels.

Les crédits destinés aux entreprises sont adaptés aux besoins spécifiques de leur activité.

Selon Karyotis (2015), deux principales catégories de crédits peuvent être distinguées, à savoir les crédits d'exploitation et les crédits d'investissement.

Les crédits d'exploitation :

Il s'agit de financements à court terme destinés à assurer le fonctionnement du cycle d'exploitation de l'entreprise, notamment le financement des stocks et des créances clients (Vernimmen, 2022).

Ils permettent de couvrir les besoins de trésorerie liés à l'activité courante, en comblant le décalage entre les dépenses (achats, salaires, charges) et les recettes provenant des encaissements clients.

Les crédits d'investissement :

« Le crédit d'investissement est un prêt accordé aux entreprises afin de financer leurs projets d'investissement, notamment l'acquisition d'immobilisations nécessaires au développement de l'activité. » (De Coussergues S. &., 2013)

Selon (Vernimmen, 2022), le crédit d'investissement est un financement bancaire accordé à moyen ou à long terme afin de permettre aux entreprises de financer leurs projets d'investissement, notamment l'acquisition d'équipements, de bâtiments ou la modernisation de leur outil de production.

B- Selon la durée du crédit

- Les crédits à très court terme (jusqu'à 3 mois)
- Les crédits à court terme (jusqu'à 2 ans)
- Les crédits à moyen terme (jusqu'à 7 ans)
- Les crédits à long terme (jusqu'à 20 ans)
- Les crédits à très long terme (au-delà de 20 ans, voire perpétuels)

C- Selon la forme et les caractéristiques du crédit

Selon la monnaie : le crédit peut être accordé en monnaie nationale (dinars, euros, etc.) ou en devises étrangères (dollars, etc.).

Selon le mode d'amortissement : c'est la manière dont le crédit est remboursé. On distingue :

- L'amortissement constant : le capital remboursé est identique à chaque échéance.
- L'annuité constante : le montant total payé à chaque échéance (capital + intérêts) reste stable.
- Le remboursement in fine : le capital est remboursé en une seule fois à l'échéance, tandis que les intérêts sont réglés périodiquement durant la durée du crédit.

Selon le type de taux : le taux d'intérêt appliqué au crédit peut être :

- À taux fixe : déterminé à la signature du contrat, il demeure inchangé durant toute la durée du crédit.
- À taux variable ou indexé : il évolue en fonction d'un indice de référence, tel que l'Euribor.
- À taux variable capé : c'est un taux variable, mais avec une limite maximale (plafond) au-delà de laquelle il ne peut pas augmenter.

Selon le mécanisme : on distingue :

- Le crédit permanent ou revolving : c'est un crédit qui se reconstitue au fur et à mesure des remboursements. L'emprunteur peut réutiliser les fonds remboursés dans la limite d'un plafond fixé.
- Le crédit sur ligne de crédit : c'est une autorisation donnée par la banque de prélever des fonds jusqu'à un certain montant, avec une durée déterminée.

Selon le contrat : la forme juridique du crédit peut être :

- Sur compte débiteur : c'est le cas du découvert autorisé, où le compte courant devient débiteur.
- Sur contrat de prêt : c'est un contrat spécifique qui détaille les conditions du prêt (montant, durée, taux, échéances).
- Emprunt obligataire : l'entreprise emprunte en émettant des obligations sur les marchés financiers.
- En pool bancaire : plusieurs banques se regroupent pour financer un même projet, souvent de grande envergure.

Selon la garantie : le crédit peut être :

- En blanc : accordé sans garantie particulière, uniquement sur la confiance et la réputation de l'emprunteur.
- Garanti : accompagné de sûretés pour protéger la banque en cas de défaut. Ces garanties peuvent être réelles (hypothèque sur un bien immobilier, nantissement sur des biens meubles) ou personnelles (cautionnement d'une tierce personne).

2.3 Spécificités du crédit d'exploitation :

Dans le cadre de son activité courante, l'entreprise peut être confrontée à des tensions de trésorerie ponctuelles mais récurrentes, liées notamment à la durée du cycle de production ou au retard de recouvrement des créances clients. Ces situations génèrent des besoins de financement cycliques que les ressources internes ne permettent pas toujours de couvrir intégralement.

Afin de faire face à ces déséquilibres temporaires, l'entreprise sollicite auprès de sa banque des crédits à court terme, destinés à rétablir l'équilibre de sa situation financière. Ces financements visent à compenser des insuffisances passagères de capitaux et sont généralement accordés pour une durée d'un an, renouvelable.

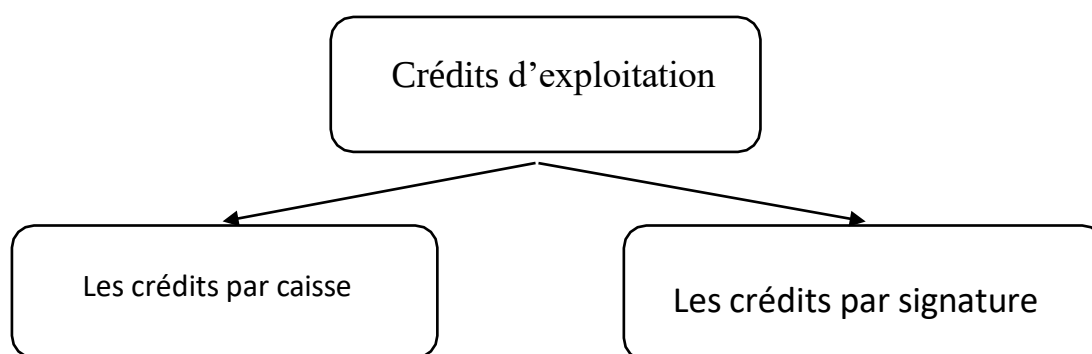
Le crédit d'exploitation est une forme de financement bancaire à court terme qui vise à couvrir les besoins générés par le cycle d'activité de l'entreprise. Il sert principalement à financer les déséquilibres temporaires de trésorerie issus du décalage entre les flux d'encaissements et de décaissements liés à l'exploitation (De Coussergues S. &., 2013)

Ce type de crédit se caractérise par plusieurs éléments. Il s'agit en général d'un financement à court terme, dont la durée est inférieure à un an, et qui est étroitement lié au cycle d'exploitation

de l'entreprise. Il sert à financer les différents éléments du besoin en fonds de roulement, notamment les stocks, les créances clients ainsi que les charges d'exploitation. Enfin, ce type de crédit est souvent renouvelable et souple, ce qui lui permet de s'adapter aux fluctuations de l'activité de l'entreprise (Vernimmen, 2022).

En outre, Les crédits d'exploitation peuvent se présenter sous plusieurs formes:

Figure 1: Classification des crédits d'exploitation



Source : Etablie par moi- même

2.3.1 Les crédits par caisse (directs) :

Les crédits par caisse Les crédits par caisse à court terme désignent des financements accordés par la banque impliquant une mise à disposition directe de fonds au profit du client, dans le but de rétablir l'équilibre de sa trésorerie à court terme (Benhalima, 1997).

De manière générale, ces crédits se répartissent en deux grandes catégories :

- Les crédits de trésorerie globaux ;
- Les crédits par caisse spécifiques.

2.3.1.1 Les crédits par caisse globaux :

Les crédits par caisse globaux ont pour finalité d'aider l'entreprise à couvrir ses besoins de trésorerie en participant au financement d'une partie de son actif circulant, notamment les stocks et les créances clients, sans qu'une utilisation précise des fonds soit déterminée au préalable. Ils résultent principalement des décalages entre les flux d'encaissements et de décaissements liés à l'activité d'exploitation.

Ces financements peuvent prendre différentes formes, parmi lesquelles :

- La facilité de caisse ;
- Le découvert bancaire ;
- Le crédit de campagne ;
- Le crédit relais.

Facilité de caisse :

Selon (Luc, 2008), la facilité de caisse correspond à un concours bancaire accordé à l'entreprise afin de couvrir un déficit temporaire de trésorerie résultant d'un décalage de courte durée entre les encaissements et les décaissements. Cette situation apparaît généralement à certaines périodes du mois, notamment lors du règlement des fournisseurs, du paiement des salaires ou encore des échéances fiscales telles que la TVA.

Découvert

D'après (Boudinot, 1972),

Le découvert constitue un terme générique regroupant plusieurs formes de concours bancaires dont les motifs doivent être clairement identifiés. Bien qu'il ne soit pas indispensable au fonctionnement de l'entreprise, il lui permet d'améliorer sa gestion financière en complétant ses ressources de financement dans certaines situations spécifiques. Ce financement peut d'ailleurs être renouvelé selon les besoins de l'entreprise.

Le principe du découvert bancaire repose sur l'autorisation accordée au client de présenter un solde débiteur sur son compte bancaire, à l'instar de la facilité de caisse. Toutefois, la principale différence entre ces deux mécanismes réside dans leur durée, le découvert étant généralement accordé pour une période plus longue, pouvant aller d'un mois à une année.

On distingue principalement deux formes de découvert :

Le découvert simple : Il permet au client d'utiliser son compte au-delà du solde disponible, dans la limite d'un plafond autorisé, lequel dépasse rarement l'équivalent de quinze jours du chiffre d'affaires. Les intérêts et frais bancaires (agios) sont calculés sur les sommes effectivement utilisées.

Le découvert mobilisable : Dans cette forme, le découvert est matérialisé par un billet à ordre renouvelable tous les 90 jours. La banque peut ensuite réescompter cet effet auprès de la Banque d'Algérie. Ce document constitue également une garantie pour la banque en cas de défaillance du débiteur. Les agios restent calculés sur le montant réellement utilisé.

Crédit de campagne :

Il s'agit d'un financement à court terme destiné à couvrir les besoins liés à une activité saisonnière, telle que l'agriculture ou la production de biens saisonniers (par exemple les jouets).

Le crédit de campagne constitue un financement accordé aux entreprises exerçant une activité saisonnière, afin de couvrir leurs charges d'exploitation liées au cycle de production. Il est particulièrement destiné aux entreprises confrontées à un décalage entre leurs périodes de production et de commercialisation.

Crédit relais :

Le crédit relais, aussi appelé crédit de soudure, correspond à un financement à court terme accordé dans le cadre d'une opération exceptionnelle hors exploitation. Il permet à l'entreprise de disposer de liquidités en attendant une entrée de fonds future provenant, par exemple, de la vente d'un actif, tel qu'un immeuble ou un fonds de commerce, ou encore d'une opération financière comme une augmentation de capital ou l'émission d'un emprunt obligataire.

2.3.1.2 Les crédits par caisse spécifiques :

Les crédits par caisse spécifiques désignent des financements orientés vers des besoins précis liés à certains éléments de l'actif circulant.

Par opposition aux crédits de trésorerie globaux, ces crédits sont destinés à un financement ciblé de composantes particulières de l'actif circulant, notamment les stocks ainsi que les créances clients, dont l'importance peut être significative dans l'activité de l'entreprise.

Ces types de crédits se caractérisent également par la présence de garanties étroitement associées à l'opération financée, contrairement aux crédits de trésorerie, pour lesquels les garanties sont en général de nature plus large et moins spécifiques.

Les principaux types de crédits spécifiques sont :

- L'escompte commercial ;

- L'affacturage (factoring) ;
- L'avance sur marchandises ;
- L'escompte de warrant ;
- L'avance sur marchés publics ;
- L'avance sur titres ;
- L'avance sur factures.

L'escompte commercial :

L'escompte correspond à une opération de crédit par laquelle un établissement bancaire met à la disposition du détenteur d'un effet de commerce non encore arrivé à échéance la valeur de celui-ci, diminuée des intérêts et des commissions applicables. En contrepartie, la banque obtient le transfert de la créance ainsi que de ses accessoires (Bouyacoub, 2000)

En d'autres termes, l'escompte commercial constitue un financement de court terme au cours duquel la banque acquiert, sous réserve d'encaissement effectif, un effet de commerce matérialisant une créance détenue sur un débiteur appelé « cédé ». Elle procède alors à la mise à disposition immédiate des fonds au profit de l'entreprise bénéficiaire, dite « cédant », avant l'échéance de l'effet, tout en prélevant des agios calculés en fonction de la durée restante jusqu'à son règlement.

L'affacturage (factoring) :

L'affacturage correspond à une opération financière dans laquelle un organisme spécialisé, appelé « factor », procède au rachat des créances qu'une entreprise détient sur sa clientèle. Dans ce cadre, le factor se substitue à son client, appelé adhérent, en lui réglant immédiatement le montant des factures avant leur échéance, tout en prenant en charge, contre rémunération, le risque d'insolvabilité des débiteurs (Benhalima, 1997 ; Code de commerce algérien, art. 543).

L'avance sur marchandises :

L'avance sur marchandises est un crédit accordé aux entreprises afin de financer des biens donnés en garantie auprès du créancier. Elle permet à l'entreprise de régler ses fournisseurs tout en disposant d'un délai pour écouler ses stocks. Ce type de financement concerne généralement des marchandises non périssables et dont la commercialisation reste maîtrisable dans le temps.

L'escompte de warrant :

Le warrant est un titre de commerce représentant un gage sur des marchandises déposées dans des magasins généraux. Il permet à son détenteur de bénéficier d'un financement bancaire en utilisant ces biens comme garantie. Dans ce cas, la marchandise est effectivement déposée, ce qui constitue une forme de dépossession réelle et sécurisée pour la banque (Bouyacoub, 2000).

L'avance sur marché public :

Les marchés publics désignent des contrats conclus entre les administrations publiques et des opérateurs économiques privés ou publics, portant sur la fourniture de biens, la réalisation de travaux ou la prestation de services. L'avance sur marché public constitue un financement permettant aux entreprises d'assurer l'exécution de ces contrats avant le paiement final de l'administration.

L'avance sur titres :

Conformément à l'article 976 du Code civil, l'avance sur titres constitue une modalité de financement permettant aux clients détenteurs d'un portefeuille de valeurs mobilières telles que les bons de caisse (BDC), les dépôts à terme (DAT) ainsi que les obligations d'obtenir des liquidités en mobilisant ces actifs comme garantie. Les titres, qu'ils soient nominatifs ou à ordre, peuvent faire l'objet d'un nantissement. En pratique, ce type de financement porte essentiellement sur les dépôts à terme et les bons de caisse, lesquels sont remis en garantie, avec blocage des contrats de DAT jusqu'au remboursement de l'avance consentie.

L'avance sur factures :

L'avance sur facture est définie comme un crédit de trésorerie accordé par la banque contre la remise de factures dûment visées par des administrations ou des entreprises publiques, lesquelles sont généralement domiciliées auprès des guichets de l'établissement bancaire prêteur (Benhalima, 1997)

2.3.2 crédits par signature (indirects) :

Le crédit par signature correspond à un engagement pris par la banque vis-à-vis de tiers, consistant à honorer les obligations contractées par son client en cas de défaillance de ce dernier. Il ne s'agit pas d'un décaissement immédiat, mais d'une garantie de paiement en faveur du créancier.

Les crédits par signature se déclinent en quatre principales formes :

L'aval,

L'acceptation,

Le cautionnement

Et le crédit documentaire.

L'aval :

L'aval bancaire est un engagement par lequel la banque se porte garante du paiement total ou partiel d'un effet de commerce à l'échéance, dans le cas où le débiteur principal ne s'acquitte pas de sa dette. Cet engagement peut être matérialisé directement sur l'effet de commerce, sur une allonge, ou par acte séparé, notamment à travers la mention « bon pour aval en faveur de M. X à concurrence du montant en lettres », suivie de la signature du banquier. Par ce mécanisme, la banque facilite à son client l'obtention de délais de paiement auprès de ses fournisseurs.

L'acceptation :

L'acceptation est l'engagement pris par une banque d'honorer le paiement d'un effet de commerce à son échéance. Contrairement à l'aval, la banque devient ici le principal débiteur vis-à-vis du bénéficiaire. Cet engagement est formalisé par la mention « bon pour acceptation », apposée et signée par le banquier sur le titre concerné.

Le cautionnement :

Le cautionnement est un contrat par lequel une partie, généralement la banque, s'engage à garantir l'exécution d'une obligation contractée par son client. Ainsi, en cas de défaillance de ce dernier, la banque se substitue à lui pour honorer la dette envers le créancier. Les cautions bancaires peuvent être classées selon deux critères : d'une part, selon leur destination, on distingue les cautions administratives, fiscales et douanières ; d'autre part, selon leur objet, on différencie les cautions de différé de paiement et celles visant à éviter certains décaissements.

Le crédit documentaire :

Le crédit documentaire est un engagement bancaire par signature qui vise à sécuriser les opérations de commerce international. Il garantit à l'importateur la bonne réception de la marchandise dans les conditions prévues, tout en assurant à l'exportateur un paiement à l'échéance convenue, sous réserve du respect des conditions stipulées dans le contrat.

3. Le risque de crédit :

3.1 Définition et enjeux :

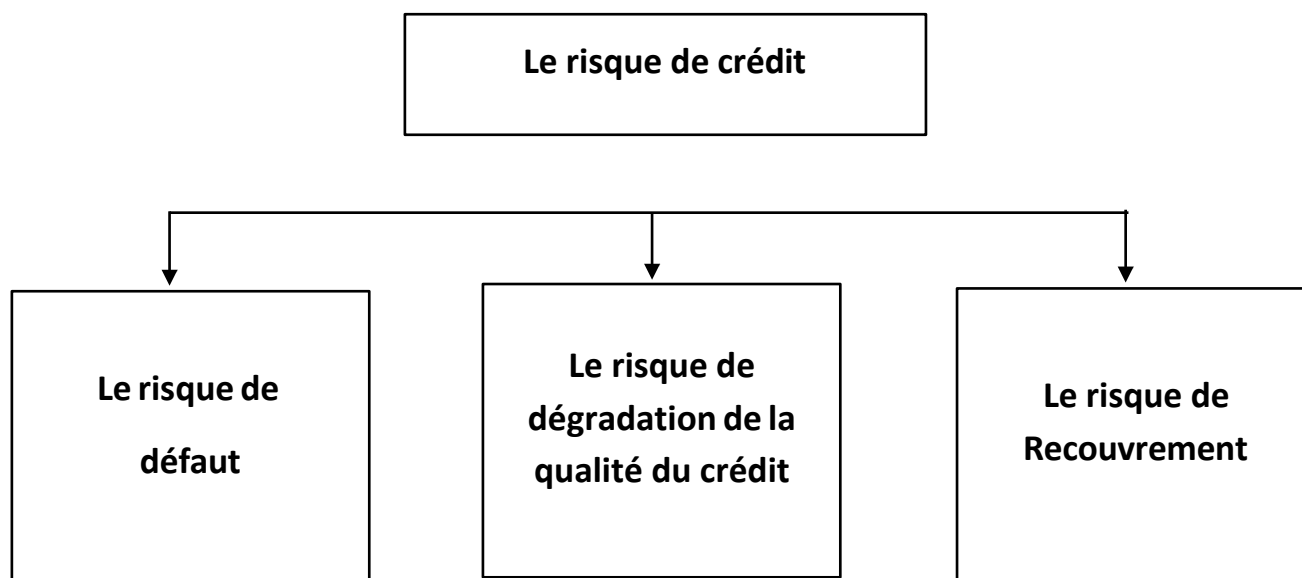
Le risque de crédit, également appelé risque de contrepartie, représente l'une des formes les plus anciennes de risques auxquels les établissements bancaires sont exposés. Il se définit comme le risque de perte associé à une créance, ou plus précisément comme la probabilité qu'un emprunteur ne soit pas en mesure de rembourser sa dette dans les délais et selon les conditions initialement convenues.

C. Gouriéroux et A. Tiomo définissent le risque de crédit comme le risque de perte lié au non-respect, par un emprunteur, de ses engagements de remboursement de dettes, qu'il s'agisse d'obligations, de prêts bancaires ou encore de créances commerciales.

3.2 La décomposition du risque de crédit :

Selon les travaux de recherches antérieurs, le risque de crédit regroupe plusieurs composantes :

Figure 2: Les principales composantes du risque de crédit



Source : (Bessis, 2015)

Le risque de défaut : Le risque de défaut correspond à la probabilité qu'un débiteur se trouve dans l'incapacité de respecter ses engagements financiers et d'honorer ses obligations de remboursement.

Selon (Bruyère, 2003), le risque de défaut correspond à l'incapacité ou au refus d'un débiteur de s'acquitter, dans les délais impartis, de ses engagements financiers contractuels envers son créancier, que ce soit au titre du remboursement du capital ou du paiement des intérêts liés à la créance contractée.

Dans le cadre de l'accord de Bâle II, la notion de défaut est retenue lorsqu'au moins l'un des événements suivants est observé :

- Incapacité probable de l'emprunteur à rembourser la totalité de sa dette, incluant le principal, les intérêts et les commissions.
- Constatation d'une perte sur une facilité de crédit, se traduisant notamment par un passage en pertes, une restructuration avec réduction de dette ou un rééchelonnement des engagements.
- Retard de paiement supérieur à 90 jours sur l'une des échéances dues.
- Ouverture d'une procédure de faillite ou de toute procédure assimilée visant à protéger le débiteur contre les recours de ses créanciers.

Le risque de dégradation de la qualité du crédit :

Le risque de dégradation de la qualité du crédit correspond à la possibilité que la situation financière de la contrepartie se détériore, entraînant ainsi une augmentation de sa probabilité de défaut, sans que celui-ci ne soit encore survenu.

Ce type de risque se traduit par une diminution de la valeur de la dette au cours de sa durée de vie. Ainsi, ces pertes peuvent être réalisées notamment en cas de cession anticipée de la créance, même en l'absence de défaut de l'emprunteur.

Le risque de Recouvrement

Le risque de recouvrement est lié au taux de recouvrement de la dette, c'est-à-dire à la part du montant prêté que la banque est susceptible de récupérer en cas de défaillance de l'emprunteur, après la mise en œuvre de l'ensemble des procédures légales. Ce taux constitue une source d'incertitude pour l'établissement bancaire, dans la mesure où il dépend de plusieurs facteurs, notamment la durée des procédures judiciaires, la valeur réelle des garanties apportées ainsi que le rang de priorité de la banque parmi les créanciers. Ainsi, le taux de recouvrement correspond

à la proportion du montant contractuel de la dette que la banque parvient effectivement à récupérer après la survenance du défaut de la contrepartie.

3.3 Les paramètres fondamentaux du risque de crédit :

Lorsqu'une banque octroie un crédit, elle s'expose au risque que l'emprunteur ne soit pas en mesure de rembourser, totalement ou partiellement, les sommes dues. Afin d'apprécier ce risque, plusieurs paramètres sont pris en considération dans le but d'estimer les pertes potentielles, tant au niveau d'un crédit individuel qu'à l'échelle de l'ensemble du portefeuille de crédits détenu par la banque.

La probabilité de défaut t (Probabilit of Default : PD) :

La probabilité de défaut correspond à la probabilité qu'une contrepartie fasse défaut sur ses engagements financiers sur un horizon temporel donné, généralement d'un an. Elle constitue une mesure essentielle dans l'évaluation et la quantification du risque de crédit.

La perte en cas de défaut (Loss Given Default : LGD) :

Constitue la perte réellement constatée après un événement de défaut. La détermination de ce paramètre suppose l'estimation des taux de recouvrement obtenus selon les transactions et catégories de clients ayant connu le défaut.

Le LGD est exprimé en pourcentage et il est déterminé comme suit :

$$\text{LGD} = 1 - T$$

On suppose que : le montant total du crédit = 100 %

La partie que la banque récupère après le défaut → Taux de recouvrement (T)

La partie que la banque perd → LGD (Loss Given Default)

Donc : **total = Recouvrement + Perte**

Comme le total est **100 % = 1**, alors : **1 = T + LGD**

Donc : **LGD = 1 - T**

Exposition en cas de défaut (Exposure At Default : EAD) :

Elle correspond à la perte maximale potentielle supportée en cas de défaut de la contrepartie. Elle est ainsi assimilée au montant des remboursements restant dus à la date de survenance du défaut. Ce risque provient de l'incertitude relative au niveau exact de l'exposition, laquelle représente la perte maximale que le prêteur est susceptible d'encourir au moment du défaut

Ses fonctions de mesure du risque de crédit tentent de prévoir la perte projetée d'un portefeuille de crédit sur une certaine période. Elle représente la perte moyenne que la banque peut prévoir sur un ensemble de crédits.

L'estimation des pertes inattendues constitue également un objectif essentiel de la mesure du risque de crédit. À savoir, la perte attendue, qui se calcule comme suit :

$$EL = PD \times EAD \times LGD$$

Un recours aux provisions est nécessaire afin de couvrir ces pertes potentielles

4. Les méthodes d'évaluation du risque de crédit :

L'analyse du risque de crédit représente une phase essentielle dans le processus décisionnel des établissements bancaires, car elle permet d'évaluer la solvabilité de l'emprunteur ainsi que son aptitude à respecter ses engagements financiers.

Pour réduire l'incertitude liée à l'octroi de crédit, les banques s'appuient sur plusieurs méthodes traditionnelles d'évaluation, qui reposent à la fois sur l'analyse financière de l'emprunteur et sur des critères qualitatifs liés à son profil et à son environnement économique. Ces méthodes permettent d'estimer la probabilité de défaut et d'orienter la décision d'octroi du crédit.

Parmi les principales méthodes utilisées, on peut citer l'analyse financière, la méthode des 5C, l'approche LAPP, les systèmes experts, le rating et le scoring.

4.1 L'analyse financière

L'analyse financière constitue l'approche la plus traditionnelle et la plus couramment utilisée par les établissements bancaires pour apprécier la solvabilité, la rentabilité ainsi que la liquidité d'une entreprise, sur la base d'un diagnostic à la fois économique et financier.

Selon (Marion, 2015), « l'analyse financière est une méthode de compréhension de l'entreprise à travers ses états comptables, ayant pour objectif de porter un jugement global et argumenté sur son niveau de performance ainsi que sur sa situation financière, qu'elle soit actuelle ou future »

Cette démarche s'appuie essentiellement sur l'exploitation des données issues des documents comptables, notamment le bilan et le compte de résultat, permettant ainsi d'évaluer la structure financière de l'entreprise ainsi que ses performances économiques

L'analyse financière utilise différents instruments qui permettent d'apprécier la situation de l'entreprise, parmi lesquels :

Les indicateurs de l'équilibre financier

L'analyse de l'équilibre financier d'une entreprise repose sur trois indicateurs fondamentaux, à savoir : le fonds de roulement, le besoin en fonds de roulement et la trésorerie nette.

Le fonds de roulement (FR) :

Selon Peyrard Josette, le fonds de roulement représente une marge de sécurité financière pour l'entreprise, correspondant à la portion des ressources permanentes utilisée pour financer les actifs circulants.

Le fonds de roulement représente ainsi l'excédent des ressources stables mobilisées pour financer les emplois durables de l'entreprise, lui permettant de disposer d'une marge de sécurité financière. Il peut être déterminé selon deux approches :

Par le haut du bilan : $FR = \text{capitaux permanents} - \text{Actifs immobilisés}$

Par le bas du bilan : $FR = \text{Actif circulant} - \text{Passif circulant}$

Le besoin en fonds de roulement BFR :

Le besoin en fonds de roulement correspond au besoin de financement généré par le cycle d'exploitation. Il découle des décalages existants entre les encaissements et les décaissements liés à l'activité opérationnelle de l'entreprise.

Le besoin en fonds de roulement est étroitement associé aux valeurs d'exploitation, aux valeurs réalisables ainsi qu'aux dettes à court terme. Il correspond à la fraction des actifs circulants dont le financement n'est pas couvert par les passifs circulants, et qui demeure par conséquent à la charge de l'entreprise via son fonds de roulement.

Le BFR est calculé comme suit :

BFR = valeurs d'exploitation + valeurs réalisables – dettes à court terme

BFR = (stocks + créances clients) – dettes fournisseurs

- La Trésorerie nette TN :

La trésorerie nette correspond au niveau de liquidités dont dispose l'entreprise après la prise en compte du financement du fonds de roulement et du besoin en fonds de roulement.

Selon Bouyacoub (2000) : « La trésorerie de l'entreprise correspond à la différence entre l'ensemble des ressources disponibles et l'ensemble des besoins à financer ».

Elle se calcule comme suit :

TN = besoin en fonds de roulement – fonds de roulement

TN = disponibilités – concours bancaires courants

-Les ratios financiers :

Un ratio financier est un rapport entre deux grandeurs comptables (issues du bilan ou du compte de résultat) permettant d'analyser la situation financière d'une entreprise, d'évaluer sa performance et d'aider à la prise de décision.

Les principaux ratios sont :

Tableau 1: Ratios de structure

Ratios	Calcul	Interprétation
Financement des emplois stables	$\frac{\text{Ressources stables}}{\text{emplois stables}}$	Il permet d'apprécier la couverture des emplois stables par les ressources stables.
Autonomie financière	$\frac{\text{Ressources propres}}{\text{endettement}}$	Il permet d'évaluer la capacité d'endettement de l'entreprise. Il ne doit pas être inférieur à un seuil critique.
Couverture de l'actif par le FR	$\frac{\text{FR}}{\text{Actif circulant}}$	Il met en évidence la part du fonds de roulement affectée au financement de l'actif circulant.

Source : (Grandguillot, 2023)

Tableau 2: Ratios d'activité

Ratios	Calcul	Interprétation
Rotation des stocks (entreprise commerciale)	$\frac{\text{Stock moyenne marchandises} \times 360}{\text{Coûts d'achat des marchandises vendus}}$	Il permet de mesurer la durée d'écoulement des stocks.
Rotation des stocks (entreprise industrielle)	$\frac{\text{Stock moyen de MP} \times 360}{\text{Coût d'achat des MP}}$ $\frac{\text{Stock moyen de PF} \times 360}{\text{Coût de production des PF}}$	Il permet de mesurer la durée d'écoulement des stocks.
Durée moyenne du crédit clients	$\frac{\text{Créances clients} + \text{écarts d'escompte} \times 360}{\text{CA TTC}}$	Il permet de déterminer la durée moyenne, exprimée en jours, du crédit accordé par l'entreprise à ses

Durée moyenne du crédit fournisseurs	$\frac{\text{Dettes fournisseurs} \times 360}{\text{Consommations TTC}}$	Il permet d'évaluer la durée moyenne, en jours, du crédit obtenu par l'entreprise auprès de ses fournisseurs.
--------------------------------------	--	---

Source : (Grandguillot, 2023)

Tableau 3: Ratios de liquidité, endettement et solvabilité

Ratios de liquidité, endettement et solvabilité

Ratios	Calcul	Interprétation
Ratios de liquidité		
Liquidité générale	Actifs courants / Passifs courants	Il permet d'apprécier la capacité de l'entreprise à honorer ses dettes à court terme en mobilisant ses actifs à court terme.
Liquidité réduite	(Actifs courants - stocks) / Passifs courants (creances+disponibilités) /Pc	Il permet d'évaluer la capacité de l'entreprise à régler ses dettes à court terme à partir de ses créances et de ses disponibilités.
Liquidité immédiate	Liquidité / Passifs courants Disponibilités /Pc	Il permet de mesurer la capacité de l'entreprise à faire face à ses dettes à court terme en utilisant uniquement ses disponibilités.
Ratios d'endettement		
Autonomie financière	Capitaux propres/ Total passif	Il constitue un indicateur de l'autonomie financière de l'entreprise.
Endettement à terme	Capitaux propres / Passifs courants	Il permet d'apprécier la capacité de l'entreprise à recourir à l'endettement.
Ratios de solvabilité		

Solvabilité générale	Capitaux propres / Dettes à long terme	Il permet de mesurer la capacité globale de l'entreprise à honorer l'ensemble de ses dettes à partir de la totalité de ses actifs.
----------------------	---	--

Source : (Grandguillot, 2023)

Les soldes intermédiaires de gestion :

Les soldes intermédiaires de gestion sont déterminés à partir des informations figurant dans le compte de résultat.

Leur étude permet :

- de comprendre la formation du résultat net ;
- Il permet d’assurer le suivi, dans le temps et dans l’espace, de l’évolution de la performance ainsi que de la rentabilité de l’activité de l’entreprise.

Tableau 4: Les soldes intermédiaires de gestion

Solde	Calcul	Interprétation
La Marge commerciale	MC = Vente de marchandises – coût d’achat de marchandises (Achat des marchandises (+/-)) variation des stocks de marchandises)	Elle correspond à la marge dégagée par l’entreprise. Elle reflète l’activité de négoce de l’entreprise (activité d’achat – revente en l’Etat).
Production de l’exercice	=Production vendue + Production stockée + Production immobilisée	La production de l’exercice correspond à l’ensemble des activités de transformation industrielle et/ou de prestation de services réalisées par l’entreprise au cours de la période considérée.
La valeur ajoutée	VA = production de l’exercice – consommation de l’exercice	La valeur ajoutée mesure le poids économique de l’entreprise et son niveau de croissance. Elle détermine la richesse réelle créée par l’entreprise.
L’excédent brut d’exploitation	Valeur ajoutée + Subventions d’exploitation - Impôts, taxes et versements assimilés - Charges de personnel	L’EBE représente le surplus créé par l’exploitation de l’entreprise après rémunération du facteur de

	(salaires et traitements + charges sociales)	production travail et paiement des impôts liés à la production. Il mesure la rentabilité économique réalisée par l'activité de l'entreprise.
Le résultat opérationnel	Excédent brut d'exploitation + Reprises et transferts sur charges d'exploitation + Autres produits - Dotations aux amortissements, dépréciations et provisions d'exploitation - Autres charges	Le résultat d'exploitation représente la ressource nette dégagée par l'activité courante de l'entreprise. Il constitue un indicateur permettant d'apprécier la performance économique issue de son activité normale.
Résultat ordinaire avant impôts	Résultat d'exploitation + Produits financiers - Charges financières	Le résultat courant avant impôt mesure à la fois la rentabilité économique et financière de l'entreprise
Résultat exceptionnel :	Produits exceptionnels - Charges exceptionnelles	Le résultat exceptionnel mesure les ressources dégagées par les opérations non courantes de l'entreprise.
Résultat net de l'exercice	Résultat courant avant impôt + Résultat exceptionnel - Participation des salariés - Impôt sur les bénéfices	Il représente la rentabilité réalisée par l'entreprise durant un exercice comptable

Source : élaboré par moi-même selon les cours (Ellah, 2025)

Les méthodes de dépouillement

Il s'agit de techniques empiriques permettant d'apprécier le risque en fonction de la situation propre de l'entreprise étudiée, à partir de critères préalablement définis de manière subjective et reposant sur une démarche d'observation. Après l'analyse des informations disponibles, l'analyste est amené à formuler un jugement personnel afin d'évaluer le niveau de risque présenté. Parmi ces méthodes, on peut citer notamment la méthode des 5C ainsi que la méthode LAPP (Kharoubi C. &, 2016)

4.2 La méthode de 5C

Cette méthode s'appuie sur des critères d'appréciation du risque reposant essentiellement sur les dirigeants, la structure financière et les garanties

Selon la (Banque, 2014),

La méthode « des 5C » consiste à étudier cinq principales composantes permettant d'apprécier le risque :

- Capacity : il s'agit de la capacité de l'emprunteur à honorer son engagement de crédit en fonction de sa situation financière et de sa solvabilité ;
- Character : Il s'agit de l'image de réputation de l'entreprise sur le marché et auprès de ses créanciers, « donner une preuve que vous croyez en votre projet » ;
- Capital : la structure financière de l'entreprise ;
- Collatéral : il s'agit de l'analyse des actifs sous-jacents susceptibles d'être mobilisés en tant que garanties dans le cadre du crédit. L'examen du portefeuille d'actifs permet d'identifier les immobilisations pouvant être utilisées, le cas échéant, comme sûreté par le créancier en cas de défaillance de l'emprunteur ;
- Conditions : elles renvoient aux conditions de marché ainsi qu'aux conditions commerciales applicables à l'emprunteur, lesquelles influencent directement les modalités et le cadre du financement accordé....

4.3 LAPP

Examen de quelques ratios en détail et de paramètres structurés autour de quatre axes principaux :

- Liquidity : elle mesure la capacité de l'entreprise à faire face à ses engagements à court terme. On analyse ici sa trésorerie, son fonds de roulement et sa capacité à payer ses dettes immédiates.
- Activity : elle concerne le niveau et la dynamique de l'activité de l'entreprise. On y étudie l'évolution du chiffre d'affaires, la rotation des actifs et la performance opérationnelle.
- Profitability : elle évalue la capacité de l'entreprise à générer des bénéfices à partir de ses ressources, à travers les marges et les résultats dégagés.

- Potential : elle analyse le potentiel futur de l'entreprise, notamment sa stratégie, son positionnement sur le marché, la qualité du management et ses perspectives de croissance.

4.4 Les systèmes des experts

Les systèmes experts constituent une approche d'aide à la décision basée sur l'intelligence artificielle, visant à reproduire le raisonnement d'un expert humain dans un domaine spécifique. Dans le cadre de l'analyse du risque de crédit, ces systèmes s'appuient sur une base de connaissances composée de règles, de critères et d'expériences accumulées afin d'évaluer la situation d'un emprunteur. Grâce à un mécanisme d'inférence, ils permettent de transformer des données qualitatives et quantitatives en une appréciation structurée du risque, souvent sous forme de score ou de recommandation. Ainsi, les systèmes experts offrent aux établissements financiers un outil d'analyse complémentaire aux méthodes traditionnelles, en améliorant la rapidité, la cohérence et la standardisation des décisions de crédit.

4.5 Le Rating

Le rating, ou notation financière, constitue une évaluation du niveau de risque de crédit associé à un émetteur ou à un instrument financier. Il est généralement établi par des organismes spécialisés dans l'analyse du risque, tels que les agences de notation internationales.

Selon (Kharoubi C. &., 2016), la notation financière peut être définie comme un système structuré d'évaluation permettant d'attribuer une note à une entreprise afin de la classer dans une catégorie de risque standardisée selon une échelle prédéfinie. Cette approche vise à synthétiser un ensemble d'informations complexes en une variable unique, offrant ainsi une estimation globale du niveau de risque et de la probabilité de difficultés financières. La notation repose ainsi sur une classification par classes de risque et se caractérise par un caractère statistique discret.

Par ailleurs, la construction du rating repose sur une analyse combinée d'éléments quantitatifs et qualitatifs. Selon Clauss (2021), cette évaluation intègre notamment l'analyse de la compétitivité de la contrepartie, la qualité de son management, sa gouvernance, ainsi que son environnement sectoriel et concurrentiel, en plus des indicateurs financiers traditionnels.

Les types de la notation financière

Dans le cadre des exigences prudentielles définies par le Comité de Bâle, les établissements bancaires sont tenus d'utiliser des systèmes de notation internes ou externes pour l'évaluation du risque de contrepartie.

a. La notation externe

La notation externe correspond à l'évaluation du risque de crédit réalisée par des agences spécialisées, qui attribuent des notes aux emprunteurs ou aux émissions financières.

Selon De Coussergues S. B. (2020), les agences de notation sont des institutions professionnelles dont la mission principale est d'évaluer le risque associé à un émetteur d'instruments financiers, qu'il s'agisse d'une entreprise, d'un État, d'une collectivité publique ou d'un établissement financier. Elles publient leurs analyses sous forme de notes attribuées soit à un émetteur, soit à une émission spécifique.

Les principales agences de notation reconnues à l'échelle internationale sont Moody's, Standard & Poor's et Fitch Ratings. Ces évaluations sont ensuite traduites sous forme d'échelles de notation standardisées permettant de classer les entités selon leur niveau de risque.

Tableau 5: Échelle de notation externe du risque de crédit

Moody's	Standard and Poor's	Commentaire
Aaa	AAA	Le risque est quasi nul, la qualité de la signature est la meilleure possible
Aa	AA	Quasiment similaire à la meilleure note, l'émetteur noté AA est très fiable
A	A	Bonne qualité mais le risque peut être présent dans certaines circonstances économique
Baa	BAA	Solvabilité moyenne
Ba	BA	A partir de cette note, l'affaire commence à être spéculative. Le risque de non remboursement est plus important sur le long terme
B	B	La probabilité de remboursement est incertaine. il subsiste un risque assez fort
Caa	CAA	Risque très important de non remboursement sur le long terme
Ca	CA	Très proche de la faillite, emprunt très spéculatif
C	C	Situation de faillite de l'emprunteur

Source : établie par moi-même

b. La notation interne

Selon (De Coussergues S. B., 2020), Selon De Coussergues S. B. (2020), la notation interne désigne une méthode d'évaluation du risque de crédit dans laquelle la banque attribue elle-même une note à ses contreparties. Cette appréciation repose sur l'analyse du risque de défaillance de l'emprunteur, en exploitant notamment les informations internes et privées dont dispose l'établissement bancaire, issues de la relation durable entretenue avec le client.

Cette démarche vise à regrouper chaque contrepartie au sein d'une classe de risque homogène, de sorte que les emprunteurs appartenant à une même catégorie présentent des profils de risque comparables. Elle s'appuie sur plusieurs composantes fondamentales du risque de crédit, à savoir :

La probabilité de défaut (PD), qui mesure la probabilité de non-remboursement de l'emprunteur ;

- L'exposition en cas de défaut (EAD), qui correspond au montant exposé au moment du défaut ;
- La perte en cas de défaut (LGD), qui représente la part de l'exposition susceptible de ne pas être récupérée ;
- La perte attendue (EL), définie par la relation $EL = PD \times EAD \times LGD$, et qui traduit la perte moyenne anticipée ;
- La perte inattendue (UL), correspondant à la variabilité ou au risque supplémentaire autour de la perte attendue.

4.6. Le Scoring

Le scoring désigne une méthode d'évaluation reposant sur l'utilisation d'outils quantitatifs et numériques permettant de classer les emprunteurs selon leur niveau de risque, afin d'assurer une prise de décision objective, standardisée et cohérente.

Le scoring constitue une approche statistique employée par les établissements bancaires et les organismes de crédit pour estimer la probabilité de défaut d'un emprunteur.

Cette méthode s'appuie sur la combinaison de plusieurs ratios financiers intégrés dans une fonction de notation, construite à partir de données historiques, principalement de nature financière, ainsi que de techniques statistiques et économétriques. Elle permet ainsi de produire

une note synthétique, appelée score, attribuée à chaque entreprise, et représentant son niveau de risque de défaut ou de faillite.

Il s'agit donc d'une technique qui permet d'attribuer à chaque individu un score, en fonction de plusieurs variables explicatives. Plus le score est élevé, plus l'individu est considéré comme étant un « bon » client. Ce score va donc influencer la décision du prêteur d'accorder ou non un crédit à un individu. De plus, il permet également de déterminer le taux d'intérêt qui sera appliqué au prêt. Cette méthode est fondamentale pour les banques et les organismes de crédit.

4.7. Les stress tests

Selon (Bouchetara, 2017), les tests de résistance sont définis comme « un moyen d'évaluation de la position de la banque vis-à-vis d'un scénario sévère mais plausible ». Autrement dit, il s'agit d'un outil permettant de simuler des situations économiques défavorables, telles que des crises financières ou des chocs macroéconomiques, afin d'analyser la capacité d'une banque à faire face à ces conditions et à en limiter les impacts.

Ils permettent d'identifier le niveau de capital nécessaire pour absorber les pertes résultant de chocs macroéconomiques ou financiers, et d'analyser les effets de contagion (c'est-à-dire la propagation rapide des difficultés d'une institution à l'ensemble du système financier) ainsi que les risques de liquidité (incapacité pour une banque de faire face à ses besoins de trésorerie à court terme). Généralement appliqués sur un horizon de 2 à 5 ans, ces tests sont devenus incontournables après la crise de 2008 et sont recommandés par le Comité de Bâle pour renforcer la résilience du système bancaire.

Limites des méthodes traditionnelles et nécessité du Machine Learning

Les méthodes traditionnelles d'évaluation du risque de crédit, bien qu'ayant constitué la référence pendant des décennies, présentent des limites structurelles qui les rendent de moins en moins adaptées aux exigences du contexte économique actuel, notamment pour l'évaluation des PME.

L'analyse financière repose sur des données historiques et ne reflète pas la situation en temps réel de l'entreprise. Les méthodes des 5C et LAPP souffrent d'une subjectivité importante, car leur efficacité dépend du jugement de l'analyste. Les systèmes experts, bien qu'informatisés, sont figés et ne peuvent s'adapter aux évolutions du marché. Le rating, qu'il soit externe ou interne, ne couvre pas toutes les entreprises, en particulier les PME, et peut manquer de

réactivité. Le scoring, bien que plus quantitatif, repose souvent sur des modèles linéaires qui ne capturent pas les relations complexes entre les variables.

Face à ces limites, le recours au Machine Learning apparaît comme une réponse pertinente. Comme le soulignent **Tari et al. (2024, p. 3)** :

« Currently, Algerian public banks rely on traditional financial analysis methods and require collateral to assess credit risk. However, these conventional approaches, often outdated, may lead to poor borrower selection, increasing the likelihood of unpaid debts and threatening the sustainability of banking institutions. »

Cette citation met en évidence que les approches conventionnelles (analyse financière, garanties) sont souvent dépassées et peuvent conduire à une mauvaise sélection des emprunteurs. C'est précisément pour pallier ces insuffisances que le Machine Learning s'impose comme une alternative performante, capable d'analyser de grands volumes de données, de détecter des relations non linéaires et de s'adapter en continu aux évolutions du marché.

5. La réglementation prudentielle

5.1 Les accords de Bâle I, II et III

Les Accords de Bâle constituent un ensemble de normes prudentielles internationales mises en place par le Comité de Bâle sur le contrôle bancaire. Ils ont pour principal objectif de consolider la stabilité et la résilience du système financier mondial, en obligeant les banques à respecter des exigences minimales en capitaux propres ainsi que des règles renforcées de gestion des risques. Ces dispositifs réglementaires ont évolué à travers trois grandes étapes : Bâle I, Bâle II et Bâle III (Bessis, 2015).

-Bâle I

Il s'agit du premier accord bancaire international, adopté en 1988 par les pays membres du G10. Il recommande aux établissements bancaires de ces pays de maintenir un niveau minimal de fonds propres, afin de renforcer leur capacité de protection contre le risque de crédit, notamment via le respect du ratio Cooke (Amzile, 2023)

L'accord de Bâle I désigne un ensemble de directives prudentielles établies en 1988 par le Comité de Bâle, regroupant les gouverneurs des banques centrales des pays du G10, sous l'autorité de la Banque des Règlements Internationaux (BRI) située à Bâle. Ces dispositions, également appelées Accord de 1988, avaient pour objectif principal de garantir la solidité et la stabilité du système bancaire international, en imposant aux établissements bancaires un seuil minimal de fonds propres à respecter.

Le ratio Cooke, nommé ainsi en référence à Peter Cooke, ancien président du comité, stipule que les fonds propres réglementaires d'un établissement de crédit doivent représenter au minimum 8 % du total de ses engagements de crédit pondérés.

$$\text{Ratio Cooke} = \frac{\text{Fonds propres réglementaires}}{\sum \text{Encours pondérés de}} \geq 8\%$$

Les fonds propres réglementaires : Ils se composent des fonds propres de base, comprenant le capital social ainsi que les réserves, et des fonds propres complémentaires, assimilés au quasi-capital, incluant notamment les dettes subordonnées.

Les encours pondérés de crédit : Ils regroupent les engagements inscrits au bilan ainsi que les engagements hors bilan, pondérés selon des coefficients spécifiques :

0 % pour les créances sur les États membres de l'OCDE ;

- 20 % pour les créances sur les banques et les collectivités locales des pays de l'OCDE ;
- 50 % pour les créances garanties par une hypothèque ou relevant du crédit-bail immobilier ;
- 100 % pour l'ensemble des autres actifs, notamment les crédits accordés à la clientèle

Limites de Bale I

Malgré ses apports positifs, le ratio Cooke a vite montré son insuffisance due à certaines contraintes telles que la négligence de la taille de l'entreprise et sa maturité. Ainsi, le traitement était similaire pour tous les types de crédits quel que soit le demandeur du moment que les pondérations du risque étaient statiques. De plus, le ratio n'a pris en considération que le risque de crédit lors du calcul et à négliger d'autre que l'histoire a montré qu'eux aussi ont une importance de taille (Kharoubi C. &, 2016).

-Bâle II

Afin de remédier aux limites du dispositif de Bâle I, un nouvel accord prudentiel intitulé Bâle II a été adopté en 2004. Cet accord avait pour objectif d'améliorer l'évaluation des risques bancaires tout en renforçant le cadre de supervision prudentielle appliqué aux établissements de crédit, avec des exigences accrues en matière de transparence (Amzile, 2023).

Selon Tarazi (2005), ce dispositif repose sur trois piliers fondamentaux : les exigences minimales en fonds propres, le processus de surveillance prudentielle, ainsi que la discipline de marché.

Le premier pilier :

Le premier pilier concerne les exigences en fonds propres. Il impose aux banques de détenir un niveau minimal de capitaux propres destiné à couvrir l'ensemble des risques, notamment le risque de crédit, le risque de marché et le risque opérationnel. Cette exigence est traduite par le ratio de solvabilité McDonough, qui met en relation les fonds propres de l'établissement avec ses actifs pondérés en fonction des risques (RWA – Risk Weighted Assets). À l'image du ratio Cooke, ce ratio doit être supérieur ou égal à 8 %.

Le ratio de solvabilité McDonough s'exprime comme suit :

$$\text{McDonough} = (\text{Fonds propres}) / (\text{Risque de crédit} + \text{Risque de marché} + \text{Risque opérationnel}) > 8\%.$$

Deuxième pilier :

Le deuxième pilier définit un ensemble de principes encadrant les responsabilités des banques, de leurs organes de gouvernance ainsi que des autorités de supervision. Il vise à assurer une évaluation globale de l'ensemble des risques encourus, incluant non seulement ceux couverts par le premier pilier, mais également le risque de concentration, le risque de taux d'intérêt et le risque stratégique. Ce pilier met particulièrement l'accent sur le renforcement des dispositifs de gestion et de contrôle des risques au sein des établissements bancaires.

Troisième pilier :

Le troisième pilier a pour objectif principal de renforcer la transparence financière des établissements bancaires. Il impose la publication d'un ensemble d'informations qualitatives et quantitatives pertinentes, permettant aux acteurs du marché d'évaluer le profil de risque des banques ainsi que leur niveau de capitalisation. Ainsi, la communication financière constitue un levier essentiel pour renforcer la discipline de marché.

Enfin, bien que Bâle II ait permis une amélioration significative du cadre prudentiel, la crise des subprimes de 2008 a mis en évidence certaines limites, conduisant le Comité de Bâle à élaborer un nouvel accord plus strict et plus complet : Bâle III.

-Bâle 3

L'objectif principal de la révision des exigences réglementaires relatives aux fonds propres bancaires est de renforcer la stabilité du système financier, afin de limiter à la fois la probabilité et l'intensité des crises futures.

L'accord de Bâle III, qui constitue une refonte des normes internationales de solvabilité bancaire, a été publié par le Comité de Bâle en 2010. Il introduit plusieurs réformes majeures visant à restructurer le cadre prudentiel mondial, notamment :

Renforcement de la qualité des fonds propres : les établissements bancaires doivent désormais disposer d'une base de capital plus solide et plus « pure », en excluant certains éléments considérés comme moins fiables, tels que les instruments hybrides, les intérêts minoritaires ou encore certains actifs liés aux impôts différés.

Mise en place de deux ratios de liquidité :

-Le *Liquidity Coverage Ratio (LCR)*, ratio de court terme, destiné à assurer la capacité des banques à résister à une crise aiguë de liquidité sur une période d'environ 30 à 61 jours ;

-Le *Net Stable Funding Ratio (NSFR)*, ratio de moyen et long terme, visant à garantir la résilience des banques sur une période d'un an en situation de tension financière.

Introduction d'un ratio de levier : ce ratio, défini comme le rapport entre les fonds propres et le total du bilan, a pour objectif de limiter l'endettement excessif des établissements bancaires, indépendamment des pondérations de risque.

Renforcement du provisionnement avec un coussin contracyclique : les autorités prudentielles ont instauré un dispositif compris entre 0 % et 2,5 % du capital, destiné à être constitué en période de croissance économique. Ce mécanisme permet d'accumuler des réserves en phase d'expansion afin de les mobiliser en période de crise, contribuant ainsi à atténuer les effets des cycles économiques.

5.2 Cadre réglementaire en Algérie

Depuis le début des années 1990, les autorités monétaires algériennes ont entrepris un processus progressif d'alignement sur les normes et recommandations édictées par le Comité de Bâle. Dans cette dynamique, un cadre prudentiel a été progressivement instauré, reposant sur un ensemble d'exigences quantitatives auxquelles doivent se conformer l'ensemble des banques et établissements financiers agréés en Algérie.

L'adoption de la loi n°90-10 relative à la monnaie et au crédit, promulguée en 1990 et entrée en application le 14 avril 1990, a constitué une réforme majeure dans l'organisation et la structuration du système bancaire national. Par la suite, ce dispositif a été consolidé et enrichi par les décisions émanant du Conseil de la monnaie et du crédit ainsi que de la Banque d'Algérie.

L'ensemble de ces réformes s'inscrit dans une logique prudentielle visant à renforcer la surveillance, la gestion et la maîtrise des risques au sein des établissements bancaires et financiers.

Dans ce contexte, le Conseil de la monnaie et du crédit (CMC) a mis en place, le 16 février 2014, un nouveau dispositif réglementaire prudentiel, intégrant notamment :

- **(Le règlement 14-01 , du 16 février 2014), fixant les ratios de solvabilité devant être respectés par les banques et les établissements financiers :**

Selon la Banque d'Algérie (2014), l'autorité monétaire a défini les coefficients de solvabilité applicables aux banques et établissements financiers, lesquels se déclinent principalement en trois ratios essentiels :

-Un coefficient global minimum de Solvabilité

Conformément aux dispositions de l'article 2 du règlement n°2014-01 du 16 février 2014, les banques ainsi que les établissements financiers sont tenus de respecter en permanence, tant sur une base individuelle que sur une base consolidée, un ratio minimal de solvabilité fixé à 9,5 %.

$\frac{\text{Fonds propres réglementaires}}{\text{Risque de crédit} + \text{Risque de marché} + \text{Risque Opérationnel}} \geq 9,5\%$

-Ratios de Solidité

L'article 3 de ce règlement s'inscrit dans la continuité des recommandations issues de « Bâle III », en imposant aux banques opérant en Algérie le maintien d'un indicateur de solidité financière conforme aux standards internationaux. Ce ratio, obtenu en rapportant les fonds propres de base à l'ensemble des risques encourus, notamment le risque de crédit, le risque de marché ainsi que le risque opérationnel, doit être supérieur ou égal à un minimum de 7 %, auquel vient s'ajouter un coussin de conservation des fonds propres fixé à 2,5 %.

$\frac{\text{Fonds propres de base}}{\text{Risque de crédit} + \text{risque de marché} + \text{risque opérationnel}} \geq 7\%$
--

Fonds propres de base :

Selon les dispositions de l'article 9 du règlement n°14-01 de 2014, les fonds propres de base sont principalement constitués du capital social, des réserves, du report à nouveau positif, des provisions réglementées ainsi que du résultat net de l'exercice précédent, après déduction de l'impôt et des dividendes à distribuer. En outre, ces éléments doivent représenter au minimum 50 % des fonds propres réglementaires.

-Le règlement 14-02 relatif aux grands risques et aux participations :

Selon les articles 4, 5 et 19 du règlement n°14-02 relatif aux grands risques et aux participations, ce dispositif réglementaire encadre principalement deux volets essentiels.

➤ **La division des risques** : chaque banque ou établissement financier est tenu de veiller à ce que l'ensemble des risques nets pondérés contractés sur un même bénéficiaire ne dépasse pas 25 % de ses fonds propres réglementaires. Par ailleurs, le cumul des grands risques ne doit pas excéder huit fois le montant des fonds propres réglementaires de l'établissement.

➤ **Le régime des participations** : les participations financières doivent respecter deux limites fondamentales, à savoir qu'une participation individuelle ne peut excéder 15 % des fonds propres réglementaires, et que le total des participations détenues ne doit pas dépasser 60 % des fonds propres réglementaires.

-Le règlement 14-03 relatif au provisionnement et à la classification des créances :

À travers ce dispositif, la Banque d'Algérie a établi les règles encadrant la classification ainsi que le provisionnement des créances, lesquelles peuvent être résumées dans le tableau ci-après :

Tableau 6: Classement et Provisionnement des créances

		Définition	Durée de retard	Taux de provisionnement
Créances courantes		Créances dont le remboursement total, dans les délais contractuellement prévus, est jugé comme pleinement assuré.	Moins de 90 jours	Constitution d'un provisionnement général de 1 % par an, jusqu'à atteindre un seuil global de 3 %.
Créances classées	Créances à problèmes potentiels	Créances de toute nature dont la récupération, qu'elle soit totale ou partielle, demeure incertaine.	De 90 à 180 jours	Au moins 20 %.
	Créances Très risquées	Créances de toute nature pour lesquelles la récupération, totale ou partielle, demeure fortement incertaine.	De 180 à 360 jours	Au moins 50 %
	Créances compromises	Créances dont le recouvrement, total ou partiel, est compromis et pour lesquelles un retour à une situation normale n'est pas envisageable à court ou moyen terme.	Plus de 360 jours	Au minimum de 100 %

Source : Banque d'Algérie, Le règlement 14-03 du 16 février, 2014, Articles 4,5 ,9,10

(d'Algérie, 2014,)

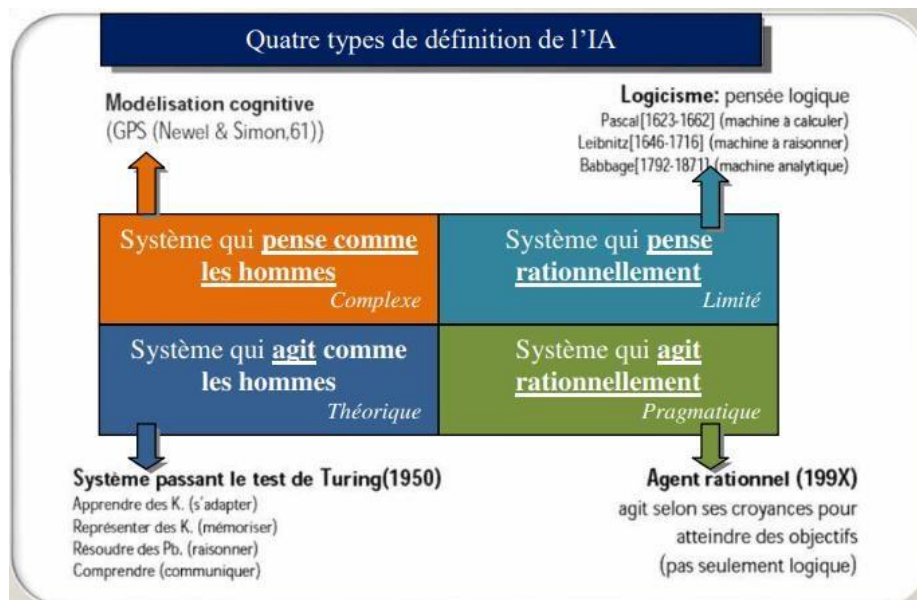
6. Le Machine Learning

6.1 Intelligence Artificielle et Machine Learning : concepts généraux

6.1.1 Définition de L'intelligence artificielle

L'intelligence artificielle (IA), concept introduit par John McCarthy, est un domaine de l'informatique qui vise à concevoir des programmes capables d'exécuter des tâches habituellement réalisées par l'être humain. Selon Marvin Minsky, l'un des pionniers du domaine, il s'agit de développer des systèmes informatiques capables d'effectuer des activités nécessitant des processus cognitifs avancés, tels que l'apprentissage, la perception, la structuration et le raisonnement.

Figure 3: Quatre définition de l'IA



L'intelligence artificielle est définie par le Parlement européen comme « la possibilité pour une machine de reproduire des comportements humains, tels que la perception, l'analyse, la planification et la créativité ». Elle désigne les systèmes en capacité de réaliser des tâches plus ou moins complexes via l'exécution d'algorithmes.

L'Intelligence Artificielle est la science dont le but est de faire faire par une machine des tâches que l'homme accomplit en utilisant son intelligence. (Soukkou)

Ainsi, les principes clés de l'IA sont :

La capacité d'apprentissage des machines à partir de données

La capacité de prendre des décisions autonomes.

6.1.2 HISTORIQUE de l'intelligence artificielle

L'intelligence artificielle (IA) est une discipline scientifique qui vise à créer des machines capables d'imiter certaines capacités humaines, comme apprendre, raisonner ou prendre des décisions. L'un des premiers chercheurs à avoir travaillé sur ce sujet est Alan Turing, qui, en 1950, s'est demandé si une machine pouvait penser comme un humain. Il a proposé une méthode d'évaluation appelée le **Test de Turing**, qui consiste à vérifier si une machine peut dialoguer avec une personne sans que celle-ci se rende compte qu'elle parle à un ordinateur.

En 1956, la conférence de Dartmouth Summer Research Project on Artificial Intelligence marque la naissance officielle de l'intelligence artificielle en tant que domaine de recherche scientifique.

Dans les années 1960, le chercheur Frank Rosenblatt a développé le **Perceptron**, qui est un modèle informatique inspiré du fonctionnement du cerveau humain. Ce modèle permet à une machine d'apprendre à partir de ses erreurs, un principe qui est encore utilisé dans les techniques modernes d'apprentissage automatique.

Après une période de ralentissement dans les années 1970, l'intelligence artificielle a connu un nouvel essor grâce aux progrès de l'informatique et à l'apparition de nouvelles méthodes d'apprentissage.

Aujourd'hui, l'intelligence artificielle est utilisée dans de nombreux domaines et constitue un élément essentiel de la transformation numérique des entreprises (Soukkou).

6.1.3 Le Machine Learning : principes fondamentaux

Alors que les humains apprennent de l'expérience, les ordinateurs peuvent-ils en faire de même ?

Selon (Landman, 2021), la réponse est affirmative et c'est précisément dans ce contexte qu'intervient l'apprentissage automatique. Le Machine Learning, qui constitue une branche de l'intelligence artificielle, désigne une approche permettant d'améliorer les performances d'un système grâce à l'exploitation de l'expérience, à travers des méthodes computationnelles. Dans les systèmes informatiques, cette expérience est représentée sous forme de données, et l'objectif principal du Machine Learning est de concevoir des algorithmes d'apprentissage capables de générer des modèles à partir de ces données. En alimentant ces algorithmes avec des données issues de l'expérience, on obtient des modèles aptes à effectuer des prédictions ou à analyser de nouvelles observations. Si l'informatique est généralement associée aux algorithmes, le

Machine Learning, quant à lui, se concentre spécifiquement sur les algorithmes d'apprentissage. Ces techniques constituent aujourd'hui un pilier essentiel de l'ère du Big Data et ont été largement appliquées avec succès dans plusieurs domaines, notamment l'évaluation du risque de crédit. (El Naqa, 2015)

Le terme machine Learning a été introduit en 1959 par le mathématicien américain Arthur Samuel, qui a développé un programme pouvant apprendre tout seul, qui en donne la définition suivante « Machine Learning is the science of getting computers to learn without being explicitly programmed » (Saint-Cirgue, 2019).

Selon Wikipedia (octobre 2025), l'apprentissage automatique (*machine learning*), également appelé apprentissage artificiel ou apprentissage statistique, constitue une branche de l'intelligence artificielle fondée sur des approches mathématiques et statistiques. Il a pour objectif de permettre aux systèmes informatiques d'acquérir des connaissances à partir des données, afin d'améliorer progressivement leurs performances dans la réalisation de tâches, sans nécessiter une programmation explicite pour chacune d'entre elles.

De façon plus générale, ce domaine couvre l'ensemble des étapes liées à la conception, l'analyse, l'optimisation, le développement et l'implémentation de ces techniques. Il est qualifié d'apprentissage statistique dans la mesure où il consiste à élaborer des modèles visant à minimiser l'erreur moyenne entre les prédictions et les résultats observés.

6.2 Les catégories d'apprentissage

L'apprentissage automatique peut être structuré en trois grandes catégories :

- L'apprentissage supervisé ;
- L'apprentissage non supervisé ;
- L'apprentissage par renforcement.

6.2.1 L'apprentissage supervisé :

Le machine learning supervisé désigne une forme d'apprentissage automatique dans laquelle le modèle est entraîné à partir d'un ensemble de données étiquetées, c'est-à-dire des données pour lesquelles la variable cible est déjà connue.

Par exemple, dans le cadre d'un modèle de prévision des tornades sur une période déterminée, le data scientifique utilise comme variables d'entrée la date, la localisation, la température ou

encore les schémas de circulation du vent, tandis que la sortie correspond aux tornades réellement observées durant cette période.

Cette approche est qualifiée de supervisée car l'apprentissage de la machine repose sur des données préalablement annotées par un expert. Autrement dit, la machine compare les résultats obtenus à partir des données d'entrée (X) via un algorithme (A), avec les résultats attendus (Y) fournis au préalable. L'objectif est alors de déterminer une fonction capable de minimiser l'écart entre les prédictions produites et les valeurs réelles.

Les méthodes d'apprentissage supervisé visent principalement à assurer une bonne capacité de généralisation, c'est-à-dire à construire un modèle capable d'effectuer des prédictions fiables sur de nouvelles données qui ne figurent pas dans l'ensemble d'apprentissage. Ce type d'apprentissage est largement utilisé dans les problèmes de régression et de classification.

6.2.2 L'apprentissage non supervisé :

L'apprentissage non supervisé est une branche du machine learning dont le principe repose sur l'exploitation de données non étiquetées X . La machine doit ainsi apprendre à regrouper, au sein d'un même ensemble, des données présentant des caractéristiques similaires.

Cette approche est principalement utilisée pour la clusterisation, c'est-à-dire le regroupement de données semblables. La machine réalise elle-même les rapprochements entre les observations en fonction des caractéristiques qu'elle parvient à identifier, sans intervention extérieure. Grâce à cette capacité de regroupement, il devient également possible de développer des systèmes de recommandation ainsi que des mécanismes de détection d'anomalies.

6.2.3 L'apprentissage par renforcement :

L'apprentissage par renforcement est une branche de l'apprentissage automatique dans laquelle un agent apprend à prendre des décisions en interagissant directement avec son environnement. L'objectif principal est de trouver la meilleure stratégie permettant de maximiser une récompense obtenue au fil des expériences successives. Contrairement à l'apprentissage supervisé et non supervisé, cette méthode repose sur un processus interactif et itératif : l'agent teste différentes actions, observe les réactions de l'environnement, puis ajuste progressivement son comportement afin d'améliorer ses performances. Ce type d'apprentissage est particulièrement adapté aux problèmes nécessitant une prise de décision séquentielle et un équilibre entre les récompenses à court terme et celles à long terme. L'apprentissage par

renforcement est aujourd'hui utilisé dans plusieurs domaines tels que la robotique, les véhicules autonomes, les jeux vidéo, les systèmes de recommandation ainsi que la navigation intelligente.

7. Les algorithmes supervisés pour la classification binaire

L'apprentissage supervisé est généralement utilisé pour l'évaluation des risques, la reconnaissance d'images, l'analyse prédictive et la détection des fraudes, et comprend plusieurs types d'algorithmes.

7.1 Régression Linéaire

La Régression Linéaire est une méthode d'apprentissage supervisé utilisée pour prédire une variable continue en fonction de variables indépendantes. Elle suppose une relation linéaire entre la variable à prédire et les variables prédictives. L'objectif est de trouver la meilleure ligne droite (ou hyperplan) qui minimise l'erreur de prédiction. La Régression Linéaire est largement utilisée pour la modélisation et la prédiction dans divers domaines comme l'économie, la finance et les sciences sociales. (Goodfellow, 2016)

7.2 Les K-plus proches voisins (KNN)

Le K plus proches voisins (KNN) est une méthode d'apprentissage supervisé connue pour sa simplicité. Il fonctionne en trouvant les k points de données les plus proches d'une nouvelle observation et en attribuant une étiquette basée sur la classe majoritaire parmi ces voisins (dans le cas de la classification) ou en calculant la moyenne des valeurs cibles de ces voisins (dans le cas de la régression).

KNN est souvent utilisé dans des situations où les données sont non linéaires et que l'on souhaite éviter des hypothèses fortes sur la structure des données. Le processus de sélection de la méthode de calcul de distance et du paramètre K (nombre de voisins) peut être complexe. Il n'existe pas de méthode universelle établie ; il est donc nécessaire d'expérimenter différentes combinaisons et de réaliser des validations croisées pour obtenir des résultats satisfaisants adaptés aux types de données traitées. (Glassner, 2021)

Cette technique nécessite un nombre important de données d'apprentissage.

7.3 Machines à Vecteurs de Support (SVM)

Les Machines à Vecteurs de Support (SVM) sont des algorithmes d'apprentissage supervisé couramment utilisés pour les tâches de classification et de régression. Ils sont efficaces pour traiter à la fois les données linéaires et non linéaires en utilisant des techniques d'optimisation pour trouver la meilleure séparation entre les différentes classes. Les SVM sont largement

utilisées dans des domaines tels que la reconnaissance de motifs, la bio-informatique et la finance en raison de leur capacité à gérer efficacement des ensembles de données de grande dimension. (Hastie, 2009)

7.4 Arbre de décision

Un arbre de décision correspond à une structure de graphe orienté acyclique dans laquelle les nœuds représentent les variables retenues selon des critères de qualité, alors que les arêtes traduisent les différentes modalités d'une variable prédictive. Les nœuds finaux, appelés également feuilles, représentent les classes obtenues.

La construction de l'arbre repose sur une partition des données à partir de la variable explicative la plus discriminante. Cette opération est répétée localement au niveau de chaque nœud jusqu'à l'obtention de feuilles pures, c'est-à-dire composées d'individus appartenant à une même classe, ou jusqu'au déclenchement d'un critère d'arrêt mettant fin au développement de l'arbre. Les différentes branches de l'arbre sont ainsi déterminées selon la répartition des effectifs de la variable cible.

7.5 Les forêts aléatoires (Random Forest)

Les forêts aléatoires, également connues sous le nom de Random Forest en anglais, c'est une classe d'apprentissage automatique développée par (Breiman, Random Forests, 2001). Elles sont basées sur une technique d'agrégation d'un ensemble d'arbres de décision CART (Classification And Regression Trees), ce qui les rend plus performantes que les arbres individuels.

L'idée simple derrière cela est que plusieurs arbres indépendants sont plus performants qu'un seul. Cette approche permet d'améliorer les résultats et de réduire les erreurs de prédiction (Breiman, 2004)

7.6 Les Réseaux Neurones Artificiels (RNA)

Un réseau neuronal est une série d'algorithmes qui s'efforce de reconnaître les relations sous-jacentes dans un ensemble de données à travers un processus qui imite le fonctionnement du cerveau humain. Dans ce sens, les réseaux neuronaux font référence à des systèmes de neurones, qu'ils soient organiques ou artificiels. Ils ont la capacité de s'ajuster à des entrées variables, leur permettant ainsi de produire le résultat optimal sans nécessiter une redéfinition des critères de sortie. Le concept de réseaux neuronaux, qui trouve ses origines dans

l'intelligence artificielle, gagne rapidement en popularité dans le développement des systèmes de trading.

7.7 La régression logistique (RL)

L'algorithme de régression logistique est une adaptation de la régression linéaire permettant de résoudre des problèmes de classification binaire : on applique la somme pondérée des caractéristiques d'entrée à une fonction sigmoïde qui comprime toute entrée dans une valeur comprise entre 0 et 1. La valeur obtenue peut être comprise comme la probabilité qu'un certain événement (ici, une classification précise) survienne.

Chapitre 2 : CADRE MÉTHODOLOGIQUE ET ORGANISATIONNEL

Introduction du chapitre

Le premier chapitre a posé les fondements théoriques et empiriques de notre recherche. Ce deuxième chapitre constitue la transition entre la théorie et la pratique : il présente le dispositif méthodologique mis en place pour répondre à notre problématique, ainsi que le contexte organisationnel dans lequel s'inscrit notre étude empirique.

Il est organisé en deux sections. La première section expose la méthodologie générale de la recherche : le positionnement épistémologique adopté, la source et le mode de collecte des données, la description de la base de données issue du système Moody's de TRUST Bank Algérie, la définition de la variable cible, la présentation du logiciel R et de la régression logistique, les métriques d'évaluation utilisées, et le détail des étapes de l'analyse. La seconde section présente le contexte organisationnel de l'étude, à travers la description de TRUST Bank Algérie, de son organisation, de son réseau et de sa direction risque crédit, qui constitue le terrain sur lequel repose l'ensemble de notre travail empirique.

Section 1 : Approche méthodologique et méthodes de collecte des données

1. Méthodologie Générale de la Recherche

1.1 Type et positionnement de la recherche :

Notre recherche est de nature quantitative et appliquée. Elle s'inscrit dans une démarche hypothético-déductive, dans laquelle nous partons d'un cadre théorique relatif au risque de crédit bancaire pour formuler des hypothèses empiriques, que nous testons ensuite à partir de données réelles issues de TRUST Bank Algérie.

Sur le plan épistémologique, cette étude adopte un positionnement positiviste : nous considérons que les phénomènes financiers tels que le risque de défaut sont observables, mesurables et modélisables à partir de données objectives. L'objectif n'est pas de décrire qualitativement le comportement des emprunteurs, mais de construire un modèle prédictif robuste capable de classer automatiquement les clients en deux catégories : sains ou défaillants.

L'objectif final est prescrit et opérationnel : nos résultats doivent aider le service risque crédit de TRUST Bank Algérie à mieux évaluer les demandes de crédit d'exploitation des PME, et ainsi à réduire son risque de non-remboursement.

1.2 Justification du choix d'une méthode quantitative :

Le choix d'une méthode quantitative dans cette recherche s'explique par la nature de notre problématique ainsi que par les données utilisées. En effet, notre étude cherche à répondre à la problématique suivante : « Dans quelle mesure les algorithmes de Machine Learning, et en particulier la régression logistique, permettent-ils de prédire efficacement la défaillance des PME sur les crédits d'exploitation en Algérie ? ». La nature même de cette problématique nécessite une étude quantitative fondée sur des méthodes statistiques, notamment la régression logistique, afin d'analyser les relations entre les variables financières et d'estimer la probabilité de défaut des entreprises. Notre recherche repose ainsi sur des ratios financiers mesurables tels que l'endettement, la liquidité ou la rentabilité. Selon Méthodes de recherche en management, l'approche quantitative est particulièrement adaptée lorsque le chercheur cherche à tester des hypothèses et à valider empiriquement des relations entre des variables mesurables.

1.3 Démarche méthodologique :

Notre démarche s'articule en six étapes successives :

Collecte des données : obtention de la base de données auprès de TRUST Bank Algérie via l'extraction des ratios Moody's sur la période 2020–2024.

Analyse exploratoire (EDA) : examen de la structure des données, détection des valeurs manquantes, analyse de la distribution des variables et identification des différences entre les deux groupes (sains et défaillants).

Prétraitement : normalisation des variables numériques, encodage des variables catégorielles, et découpage de l'échantillon en ensembles d'entraînement et de test.

Modélisation : construction et entraînement d'un modèle de régression logistique via le framework tidymodels sous R.

Évaluation : mesure des performances du modèle à l'aide de métriques adaptées au contexte de classification.

Interprétation : analyse des résultats à la lumière des hypothèses et de la littérature, et vérification de la cohérence avec les normes sectorielles de Moody's.

2. Collecte des Données :

2.1 Source des données

Les données utilisées dans cette étude ont été collectées dans le cadre d'un stage effectué au sein du Service Risque Crédit de TRUST Bank Algérie. TRUST Bank est une banque privée à capitaux étrangers opérant sur le marché bancaire algérien, dont l'une des activités principales est le financement des petites et moyennes entreprises (PME) par le biais de crédits d'exploitation. Le service risque crédit est chargé d'analyser la solvabilité des emprunteurs, de construire et mettre à jour les dossiers de notation, et de surveiller l'évolution du risque sur le portefeuille de crédits accordés.

2.2 Mode de collecte : une démarche collaborative

La collecte des données a été réalisée dans le cadre d'une collaboration avec la banque. Dans un premier temps, nous avons établi la liste des variables explicatives à retenir pour le modèle de classification basé sur le machine Learning, en nous appuyant sur notre revue de littérature et sur les variables disponibles dans le système Moody's. Cette liste de variables a ensuite été soumise au responsable du service risque crédit de TRUST Bank, qui a procédé à l'extraction des données correspondantes depuis la base interne du système Moody's pour l'ensemble des dossiers de crédit d'exploitation de PME traités par la banque entre 2020 et 2024.

Les données ont été transmises sous format Excel, avec les noms réels des entreprises remplacés par des identifiants génériques (« Client Sain » et « Client Contentieux ») pour des raisons de confidentialité. Cette anonymisation garantit le respect des obligations déontologiques et contractuelles de la banque, tout en préservant la validité scientifique et statistique de l'analyse.

2.3 Type de crédit, période et population étudiée

Dimension	Détail
Type de crédit	Crédit d'exploitation accordé aux PME (crédit de fonctionnement à court terme)
Période couverte	2020 à 2024
Taille de l'échantillon	255 dossiers de crédit
Clients sains (cible = 1)	200 dossiers (78,4 %)
Clients défaillants (cible = 0)	55 dossiers (21,6 %)
Source	Système de notation interne Moody's — TRUST Bank Algérie
Format original	Fichier Excel (.xlsx) fourni par le service risque crédit

Source : élaboré par moi-même

Le crédit d'exploitation est destiné à financer les besoins à court terme des PME liés à leur cycle d'activité : achat de marchandises, financement des stocks, règlement des fournisseurs, couverture des décalages de trésorerie. Il se distingue du crédit d'investissement par sa durée (généralement inférieure à un an) et par les risques spécifiques qui lui sont associés, notamment la qualité du portefeuille clients, la rotation des stocks et la gestion du cycle financier court terme — autant de dimensions directement capturées par nos variables explicatives.

2.4 Définition de la variable cible

La variable cible, notée cible, est une variable binaire définie selon les critères de classification interne de TRUST Bank :

Cible = 0 : Client contentieux, c'est-à-dire un client dont le dossier de crédit a donné lieu à un incident de remboursement significatif : impayés répétés, classement en créance douteuse, recours au contentieux juridique, ou mise en place d'une procédure de recouvrement forcé.

Cible = 1 : Client sain, c'est-à-dire un client qui a honoré l'intégralité de ses engagements financiers dans les délais contractuels convenus avec la banque, sans incident de paiement majeur.

2.5 Description des Variables Explicatives

Les variables explicatives (ou prédicteurs) utilisées dans le modèle sont au nombre de neuf : sept ratios financiers quantitatifs et deux variables qualitatives décrivant le profil juridique et sectoriel de l'entreprise, présentées comme suit :

2.5.1 Variables quantitatives

L'Endettement Global

Définition :

L'endettement global mesure le poids des dettes totales de l'entreprise par rapport à ses capitaux propres. Il indique dans quelle mesure l'entreprise finance son activité par l'endettement plutôt que par ses fonds propres.

$$\text{Endettement global} = \frac{\text{Total des dettes}}{\text{Capitaux propres}}$$

Interprétation :

Plus ce ratio est élevé, plus l'entreprise dépend du financement externe (bancaire ou fournisseurs). Un ratio supérieur à 1 signifie que les dettes dépassent les capitaux propres : l'entreprise est alors considérée comme surendettée. La banque exige généralement que ce

ratio ne dépasse pas 1 pour accorder un crédit d'exploitation. Au-delà, le risque d'insolvabilité augmente fortement.

Résultats observés :

Les clients défaillants (contentieux) présentent une moyenne d'endettement global de 4,89, tandis que les clients sains affichent une moyenne de 0,88. L'écart est considérable : plus de cinq fois supérieur chez les défaillants. Ce résultat confirme que l'endettement global est l'un des prédicteurs les plus discriminants de la défaillance dans notre échantillon.

-L'endettement tangible

Définition :

L'endettement tangible est un ratio qui mesure le niveau d'endettement de l'entreprise en le rapportant uniquement à ses actifs corporels, c'est-à-dire les actifs ayant une valeur réelle en cas de liquidation (terrains, bâtiments, équipements, stocks).

$$\text{Endettement tangible} = \frac{\text{Total des dettes}}{\text{Actifs corporels nets}}$$

Interprétation

Ce ratio permet d'évaluer la capacité de l'entreprise à couvrir ses dettes à partir de ses actifs Physiques en cas de défaillance Un ratio élevé indique que les actifs corporels sont insuffisants pour garantir le remboursement des dettes, ce qui traduit un risque financier plus important.

Résultats observés :

Les résultats montrent que la moyenne de ce ratio est de 4,89 pour les entreprises défaillantes contre 0,89 pour les entreprises saines.

Cet écart significatif indique que les entreprises en difficulté présentent un niveau d'endettement largement supérieur à la valeur de leurs actifs corporels, ce qui limite fortement leur capacité de remboursement en cas de liquidation.

La couverture des intérêts

Définition :

La couverture des intérêts est un ratio qui mesure la capacité de l'entreprise à faire face à ses charges financières (intérêts sur emprunts) grâce à son résultat d'exploitation.

$$\text{Couverture des intérêts} = \text{Résultat d'exploitation} / \text{charges d'intérêts}$$

Interprétation :

Un ratio inférieur à 1 signifie que l'entreprise ne génère pas suffisamment de résultats pour payer ses seuls intérêts (signal critique de défaillance imminente). Plus ce ratio est élevé, plus l'entreprise dispose de marges confortables pour faire face à ses obligations financières.

Résultats observés :

Les résultats montrent que les entreprises saines présentent une capacité de couverture des intérêts plus élevée, avec une moyenne de **3,14**, contre **1,63** pour les entreprises défaillantes.

Bien que ces dernières parviennent encore, en moyenne, à couvrir leurs charges d'intérêts, leur marge de sécurité reste limitée. Cette situation les rend particulièrement vulnérables : une baisse du résultat d'exploitation ou une hausse des charges financières pourrait rapidement entraîner une incapacité à honorer leurs engagements.

-La couverture de la dette (DSCR)

Définition :

La couverture de la dette, également appelée *Debt Service Coverage Ratio (DSCR)*, mesure la capacité de l'entreprise à couvrir l'ensemble de ses obligations financières, c'est-à-dire **les Intérêts et le remboursement du principal**, à partir des flux de trésorerie générés par son activité.

$$DSCR = \frac{\text{Cash-flow disponible}}{\text{Service de la dette (intérêts + remboursement du principal)}}$$

Interprétation bancaire : C'est l'un des indicateurs les plus utilisés par les banques dans l'évaluation des dossiers de crédit. Un DSCR supérieur à 1 indique que l'entreprise génère suffisamment de liquidités pour honorer l'intégralité de ses remboursements. En deçà de 1, la banque est exposée à un risque de défaut.

-Le délai clients (DSO)

Définition

:

Le délai clients, aussi appelé *Days Sales Outstanding (DSO)*, mesure le nombre moyen de jours que met une entreprise pour encaisser ses créances après la réalisation d'une vente.
Délai clients = Créances client / Chiffre d'affaires * 360

Interprétation : Un délai clients élevé signifie que l'entreprise attend longtemps avant d'être payée, ce qui crée des tensions (problèmes) de trésorerie et augmente le besoin en fonds de roulement (BFR). Elle est alors contrainte de recourir davantage au crédit bancaire pour financer ce décalage, augmentant ainsi son risque de défaut.

Résultats observés : 127 jours en moyenne pour les défaillants contre 30 jours pour les sains. Un écart de 97 jours, indiquant une gestion moins efficace des encaissements chez les entreprises défaillantes.

-Le délai de rotation des stocks

Définition :

Le délai de rotation des stocks mesure le nombre moyen de jours pendant lesquels les stocks restent dans l'entreprise avant d'être vendus ou consommés.

$$\text{Délai stocks} = \frac{\text{Stocks}}{\text{Coût des marchandises vendues}} \times 360$$

Interprétation :

Ce ratio permet d'évaluer l'efficacité de la gestion des stocks.

-Un délai élevé peut indiquer un stockage excessif, un ralentissement des ventes ou des difficultés d'écoulement.

-Un délai faible traduit une rotation rapide des stocks.

Résultats observés : 131 jours en moyenne pour les entreprises défaillantes contre 21 jours pour les entreprises saines, soit un écart de 110 jours, indiquant une rotation plus lente des stocks chez les entreprises défaillantes

-Le délai fournisseurs

Définition :

Le délai fournisseurs mesure le nombre moyen de jours que met l'entreprise pour régler ses dettes envers ses fournisseurs.

$$\text{Délai fournisseurs} = \frac{\text{Dettes fournisseurs}}{\text{Achats}} \times 360$$

Interprétation

Un délai fournisseurs élevé est généralement interprété positivement car il signifie que l'entreprise négocie de longs délais de paiement auprès de ses fournisseurs, ce qui lui permet de financer son cycle d'exploitation à moindre coût. À l'inverse, un délai très court (paiement quasi comptant) peut signaler une mauvaise réputation commerciale, un manque de pouvoir de négociation, ou une situation de trésorerie si tendue que les fournisseurs exigent un paiement immédiat.

Résultats observés :

3,8 jours pour les entreprises défaillantes contre 31 jours pour les entreprises saines, indiquant que les entreprises défaillantes règlent leurs fournisseurs beaucoup plus rapidement.

2.5.2 Variables qualitatives :

-La Forme Juridique

Cette variable catégorielle décrit le statut juridique de l'entreprise emprunteuse. Dans notre échantillon, quatre formes juridiques sont représentées :

Tableau 7: La Forme Juridique

Forme Juridique	Description	Effectif	Proportion
SPA	Société par Actions — capital ouvert aux actionnaires, gouvernance formalisée, obligations comptables strictes	117	45,9 %
SARL	Société à Responsabilité Limitée — capital fermé, responsabilité limitée aux apports	116	45,5 %
EURL	Entreprise Unipersonnelle à Responsabilité Limitée — un seul associé	17	6,7 %
SNC	Société en Nom Collectif — associés solidairement responsables	5	2,0 %

Source : élaboré par moi-même

Interprétation

La forme juridique influence le risque de défaut, car elle détermine le degré de responsabilité personnelle des dirigeants et leur capacité à fournir des garanties à la banque.

Le secteur d'activité

Définition : Cette variable catégorielle indique le secteur économique dans lequel opère l'entreprise. Trois secteurs sont représentés dans notre base :

Tableau 8: Le secteur d'activité

Secteur d'Activité	Description	Effectif	Proportion
Commerce multiple	Distribution et commerce de détail ou de gros, produits divers	149	58,4 %
Production	Industrie manufacturière, transformation, production de biens	66	25,9 %
BTPH	Bâtiment, Travaux Publics et Hydrauliques — secteur de la construction	40	15,7 %

Source : élaboré par moi-même

Interprétation

Le secteur d'activité peut avoir un impact significatif sur le risque de défaillance, car chaque secteur présente des caractéristiques spécifiques en termes de rentabilité, de cycle d'exploitation et de sensibilité aux conditions économiques.

3. Présentation de l'outil et mise en œuvre empirique

3.1 Présentation du logiciel R et justification de son utilisation :

R est un environnement intégré de manipulation de données, de calcul et de préparation de graphiques. Toutefois, ce n'est pas seulement un « autre » environnement statistique (comme SPSS ou SAS, par exemple), mais aussi un langage de programmation complet et autonome.

Il est développé depuis les années 90 par un groupe de volontaires de différents pays et par une large communauté d'utilisateurs et utilisatrices. C'est un logiciel libre, publié sous licence GNU GPL.

R est un langage de programmation open-source, gratuit et spécialisé dans l'analyse statistique et le traitement de données. Il est aujourd'hui l'un des outils les plus utilisés dans la recherche académique en finance quantitative, en économétrie et en science des données.

Le choix de R pour cette étude se justifie par plusieurs raisons convergentes :

- **Gratuité et accessibilité :** contrairement à des logiciels commerciaux comme SPSS, SAS ou Stata, R est entièrement gratuit et open-source, ce qui en fait un outil accessible dans le contexte académique algérien.

- **Richesse des packages statistiques** : R dispose d'un écosystème de packages extrêmement riche. Les packages mobilisés dans cette étude couvrent l'ensemble du pipeline analytique, de l'importation des données à l'évaluation du modèle.
- **Framework tidymodels** : R dispose du framework tidymodels, qui est l'équivalent R de scikit-learn en Python. Il permet de construire des pipelines de Machine Learning rigoureux, reproductibles et sans fuite de données (data leakage), avec une syntaxe unifiée et lisible.

Qualité des visualisations : le package ggplot2, inclus dans le tidyverse, permet de produire des graphiques scientifiques de haute qualité, directement exploitables dans un mémoire académique.

- **Référence dans les études algériennes** : plusieurs études de référence dans le contexte algérien (Souadda et al., 2022 ; Tari et al., 2024) utilisent R ou des outils statistiques équivalents, ce qui renforce la comparabilité de nos résultats.

À propos de R Studio

R Studio n'est pas à proprement parlé une interface graphique pour R, il s'agit plutôt d'un environnement de développement intégré, qui propose des outils facilitant l'usage de R au quotidien.

Tableau 9: Packages R utilisés dans cette étude

Package	Rôle dans l'étude	Utilité spécifique
Readxl	Importation des données	Lecture du fichier Excel fourni par TRUST Bank (read_xlsx)
dplyr (tidyverse)	Manipulation des données	Filtrage, sélection, renommage des colonnes (rename, select, glimpse)
ggplot2 (tidyverse)	Visualisations graphiques	Création des histogrammes, barplots et boxplots d'analyse exploratoire
tidyr (tidyverse)	Restructuration des données	Pivotement et transformation des tableaux de données
recipes (tidymodels)	Prétraitement pipeline	Normalisation (step_normalize), encodage one-hot (step_dummy), découpage train/test
parsnip (tidymodels)	Définition du modèle	Spécification de la régression logistique (logistic_reg)
workflows (tidymodels)	Pipeline modèle + recette	Regroupement du prétraitement et du modèle en un workflow unifié
yardstick (tidymodels)	Métriques d'évaluation	Calcul de l'accuracy, du ROC-AUC, de la matrice de confusion
rsample (tidymodels)	Découpage des données	Division stratifiée train/test

Source : élaboré par moi-même

3.2 La régression logistique : fondements théoriques et fonction mathématique

La régression logistique est un algorithme d'apprentissage automatique supervisé utilisé pour résoudre des problèmes de classification binaire. Malgré son nom, il ne s'agit pas d'une méthode de régression (prédiction d'une valeur continue) mais d'un modèle de classification qui estime la probabilité qu'une observation appartienne à l'une de deux classes.

Dans notre contexte, la régression logistique estime la probabilité qu'un client de TRUST Bank soit défaillant (cible = 0) ou sain (cible = 1) en fonction de ses ratios financiers et de sa forme juridique. C'est l'algorithme historique de référence en évaluation du risque de crédit, utilisé depuis les travaux fondateurs d'Altman (1968) et de Martin (1977), et recommandé pour sa transparence et son interprétabilité dans les contextes réglementés.

Fondements mathématiques

La régression logistique modélise la probabilité de défaut $P(Y = 1 | X)$ à l'aide de la fonction sigmoïde :

$$P(Y = 1 | X) = 1 / (1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)})$$

Où β_0 est le terme constant et $\beta_1, \dots, \beta_n$ sont les coefficients estimés par la méthode du maximum de vraisemblance. La fonction sigmoïde garantit que la probabilité prédite est toujours comprise entre 0 et 1. Un seuil de classification (0,5 par défaut) est ensuite appliqué : si $P > 0,5$, le client est classé comme défaillant ; sinon comme sain.

3.3 Étapes de l'analyse sous R

Phase 1 : Importation et exploration des données

1. Chargement des bibliothèques et importation du fichier Excel

Code R :

```
library(readxl)
library(tidyverse)
data = read_xlsx('base.xlsx')
```

La première étape de tout projet de Machine Learning consiste à charger les données en mémoire. Trois instructions sont nécessaires :

Library(readxl) : charge-le package readxl qui permet à R de lire les fichiers Excel (.xlsx). Sans ce package, R ne sait pas ouvrir les fichiers Excel. C'est l'équivalent d'installer un lecteur de format.

Library(tidyverse) : charge une collection de packages essentiels : dplyr (manipulation des données), ggplot2 (graphiques), tidyr (restructuration). C'est la boîte à outils principale de notre analyse.

read_xlsx('base.xlsx') : lit le fichier Excel fourni par TRUST Bank et le charge dans un objet appelé 'data'. À partir de ce moment, toutes nos 255 lignes et 11 colonnes sont accessibles dans R.

2. Exploration initiale de la structure des données

Code R :

```
Data %>% glimpse ()  
  
ColSums(is.na(data))  
  
data %>% Summary ()
```

Avant de construire un modèle, il est indispensable de comprendre la structure et la qualité des données. Ces trois fonctions forment le bilan de santé initial de la base :

Glimpse(data) : Pour avoir une vision générale des données (les dimensions et les types de variables, numériques et catégorielles), nous affichons un aperçu compact de la structure : le nombre de lignes (255) et de colonnes (11), le type de chaque variable (numérique, texte, entier), ainsi que les premières valeurs. Cette fonction permet de vérifier immédiatement que les données ont été correctement importées et que les types sont cohérents avec ce qui est attendu

ColSums(is.na(data)) : compte le nombre de valeurs manquantes (NA) dans chaque colonne. Dans notre base, le résultat est 0 pour toutes les colonnes — aucune valeur manquante. Ce résultat s'explique par le fait que les données proviennent du système Moody's de TRUST Bank, qui exige des dossiers complets avant tout enregistrement.

Summary(data) : calcule les statistiques descriptives de chaque variable : minimum, maximum, moyenne, médiane, 1er quartile (Q1) et 3e quartile (Q3) pour les variables numériques, et la fréquence de chaque modalité pour les variables catégorielles. C'est la première vue d'ensemble des caractéristiques financières de l'échantillon.

Tableau 10: Synthèse de l'exploration initiale des données

Résultat obtenu	Interprétation	Décision prise
255 observations, 11 colonnes	Structure conforme à la base Moody's	Aucune correction nécessaire
0 valeur manquante	Base complète et de qualité	Pas d'imputation requise
Min/Max/Médiane par variable	Premières différences entre groupes visibles	Exploration graphique approfondie

Source : élaboré par moi-même

3. Renommage des colonnes à noms problématiques

Code R :

```
Data <- data %>% rename (  
  endettement_global = `Endettement global`,  
  endettement_tangible = `Endettement tangible`,  
  couverture_interets = `Couverture d'intérêts`,  
  couverture_dette = `Couverture de dette`,  
  delai_clients = `Délai clients`,  
  delai_stocks = `Délai stocks`,  
  delai_fournisseurs = `Délai fournisseurs`)
```

Les noms de colonnes originaux dans le fichier Excel de TRUST Bank contiennent des espaces, des apostrophes et des accents. Ces caractères spéciaux sont incompatibles avec la syntaxe de R pour plusieurs raisons :

- Les espaces dans un nom de colonne forcent l'utilisation de guillemets inverses dans chaque référence, rendant le code lourd et source d'erreurs.
- L'apostrophe dans « Couverture d'intérêts » peut être interprétée par R comme un séparateur de chaîne de caractères, provoquant des erreurs de syntaxe.
- Les accents posent des problèmes de compatibilité selon les systèmes d'exploitation et les encodages de fichiers.

Ce renommage remplace chaque nom problématique par un nom simple, sans espaces ni caractères spéciaux. Après ce renommage, chaque colonne peut être appelée directement dans les formules, les graphiques et les recettes de prétraitement sans guillemets ni erreur.

Phase 2 : Analyse Exploratoire Graphique (EDA)

L'analyse exploratoire des données (EDA : Exploratory Data Analysis) est une phase indispensable qui précède toute modélisation. Elle consiste à visualiser les données pour comprendre leur structure, détecter les anomalies, et identifier visuellement les variables les plus discriminantes avant de construire le modèle. Toutes les visualisations ont été produites avec le package ggplot2 de tidyvers R.

4. Visualisation de la répartition de la variable cible

Code R :

```
Data %>% ggplot(aes(x=as.factor(cible), fill=as.factor(cible))) +  
  geom_bar() +  
  labs(title="cible repartition", x="cible", y="frequence") +  
  theme_minimal()
```

Ce graphique (barplot) représente combien de clients sains et défaillants se trouvent dans notre base. Il répond à la première question fondamentale de tout problème de classification

5. Visualisation des variables catégorielles (forme juridique et secteur)

Code R :

Forme juridique

```
data %>% ggplot(aes(x=as.factor(forme_juridique))) +  
  geom_bar(fill="skyblue") +  
  labs(title="forme_juridique repartition") +  
  theme_minimal()
```

Relation forme juridique / cible

```
data %>% ggplot(aes (x=forme_juridique, fill= as. factor(cible))) +  
  geom_bar() +  
  labs(title="relation entre forme_juridique et cible")
```

Ces graphiques servent à deux objectifs distincts. Le premier (barplot simple) montre la distribution des formes juridiques dans l'échantillon : SPA, SARL, EURL, SNC. Il permet de vérifier que toutes les modalités sont suffisamment représentées pour être incluses dans le modèle.

Le deuxième graphique (barplot empilé par couleur) croise la forme juridique avec la variable cible. Chaque barre est colorée selon la proportion de clients sains (bleu) et défaillants (rouge).

6. Distribution des ratios financiers (histogrammes) univarié

Code R :

```
data %>% ggplot(aes(x=endettement_global)) +  
  geom_histogram(fill="skyblue") +  
  labs(title="endettement_global distribution")
```

- Répété pour les 7 ratios financiers

Les histogrammes montrent la forme de la distribution de chaque ratio financier. Cette étape est essentielle pour plusieurs raisons. Elle permet de détecter les distributions asymétriques (skewed) qui caractérisent les données financières et qui justifient la normalisation. Elle permet d'identifier les valeurs aberrantes (outliers) des entreprises dont les ratios sont extrêmement éloignés de la normale. Elle donne une première impression visuelle de la séparation entre les profils des entreprises saines et défaillantes.

Dans notre base, tous les histogrammes révèlent des distributions fortement asymétriques à droite : la majorité des entreprises (saines) ont des valeurs concentrées dans les faibles niveaux, tandis qu'une minorité (défaillantes) présentent des valeurs extrêmes formant une longue traîne vers la droite. Cette asymétrie confirme la nécessité absolue d'appliquer une normalisation avant la modélisation.

7. Comparaison des ratios par groupe (boxplots) avec la variable cible la variation des ratios avec la cible

Code R :

Boxplots selon la variable cible

```
data %>% ggplot(aes(x=as.factor(cible), y=endettement_global, fill=as.factor(cible))) +  
  geom_boxplot() +  
  labs(title="Effet de l'endettement global selon la variable cible") +  
  theme_minimal()
```

- Répété pour les 7 ratios × 2 facteurs (cible + forme juridique)

Ils comparent visuellement les distributions de chaque ratio financier entre les deux groupes (sains vs défaillants).

Phase 3 : Prétraitement des Données (Pré-processing)

Le prétraitement est la phase de transformation des données brutes en un format directement utilisable par l'algorithme de Machine Learning. Dans notre étude, ce prétraitement a été réalisé à l'aide du package `recipes` de `tidymodels`, qui permet de définir un pipeline de transformations reproductible et sans fuite de données entre les ensembles d'entraînement et de test.

8. Conversion de la variable cible en facteur

Code R :

```
data$cible = as.factor(data$cible)
```

La variable cible est initialement stockée sous forme d'entier numérique (0 ou 1) dans notre base. Or, les algorithmes de classification dans R et dans `tidymodels` requièrent que la variable à prédire soit de type facteur, c'est-à-dire une variable catégorielle formellement déclarée comme telle.

Sans cette conversion, le framework `tidymodels` interpréterait cible comme une variable numérique continue et tenterait de résoudre un problème de régression (prédire une valeur entre 0 et 1 en continu) plutôt qu'un problème de classification (prédire la classe 0 ou la classe 1). Cette étape est donc une condition indispensable au bon fonctionnement du modèle.

9. Suppression des variables non pertinentes

Code R :

```
data = data %>% select(-c(client, Secteur_activité))
```

Avant de construire le modèle, il est nécessaire de retirer de la matrice des prédicteurs les colonnes qui ne contiennent aucune information utile à la prédiction.

10. Découpage stratifié de l'échantillon (Train / Test Split)

Code R :

```
library(tidymodels)  
splits = data %>% initial_split(prop = 0.8, strata = cible)
```

Pour évaluer les performances réelles d'un modèle de Machine Learning, il est indispensable de tester ce modèle sur des données qu'il n'a pas vues pendant son apprentissage. C'est le principe fondamental de la généralisation : un bon modèle doit être capable de prédire correctement de nouveaux dossiers, pas seulement ceux sur lesquels il s'est entraîné.

La fonction `initial_split` divise l'échantillon en deux sous-ensembles :

Tableau 11: Découpage des données

Sous-ensemble	Taille	Rôle
Ensemble d'entraînement (Training set)	~204 observations (80 %)	Le modèle apprend les patterns financiers sur ces données
Ensemble de test (Test set)	~51 observations (20 %)	Le modèle est évalué sur des données inconnues

Source : élaboré par moi-même

11. Construction de la recette de prétraitement (recipe)

Code R :

```
log_model = logistic_reg(mode = "classification")
data_rec = data %>%
  recipe(cible ~ .) %>%
  step_normalize(all_numeric_predictors())
  step_dummy(all_nominal_predictors(), one_hot = TRUE)
```

La recette (recipe) est le cœur du prétraitement. Elle définit, dans un seul pipeline séquentiel, toutes les transformations à appliquer aux données avant de les soumettre au modèle. Elle est composée de trois éléments :

A. La formule du modèle — recipe (cible)

Code R : `data_rec = data %>% recipe(cible)`

Cible est la formule qui spécifie que la variable à prédire est cible, et que tous les autres colonnes sont des prédictors. Cette notation standard de R indique à la recette quelles colonnes sont des entrées et laquelle est la sortie à prédire.

B. Normalisation des variables numériques — step_normalize()

Code R : `step_normalize(all_numeric_predictors())`

Les sept ratios financiers de notre base ont des échelles de mesure très différentes :

- L'endettement global varie entre 0,0 et 5,5
- Le délai clients varie entre 15 et 186 jours
- Le délai stocks varie entre 5 et 320 jours

Cette hétérogénéité pose un problème pour la régression logistique : l'algorithme estime des coefficients pour chaque variable, et si une variable a des valeurs numériquement très grandes (ex : 320 jours), elle dominera artificiellement le modèle, non pas parce qu'elle est plus importante, mais simplement parce que ses valeurs sont plus grandes en magnitude.

La normalisation z-score résout ce problème en transformant chaque variable selon la formule:

$$Z = (X - \mu) / \sigma$$

Où μ est la moyenne et σ l'écart-type calculés sur l'ensemble d'entraînement. Après transformation, chaque variable a une **moyenne de 0** et un **écart-type de 1**, ce qui met toutes les variables à la même échelle et permet une estimation équitable des coefficients.

Tableau 12: Comparaison des variables avant et après normalisation

Variable	Avant normalisation	Après normalisation
Endettement global	[0,0 ; 5,5]	[-1,5 ; +2,0] environ
Délai clients	[15 ; 186 jours]	[-1,8 ; +3,5] environ
Délai stocks	[5 ; 320 jours]	[-1,0 ; +6,0] environ

Source : élaboré par moi-même

C. Encodage One-Hot des variables catégorielles — step_dumy (one_hot = TRUE)

Code R : `step_dumy(all_nominal_predictors(), one_hot = TRUE)`

La régression logistique est un algorithme mathématique qui ne peut traiter que des données numériques. Or, notre variable forme juridique est qualitative : elle contient des valeurs textuelles (« SPA », « SARL », « EURL », « SNC »). Il est impossible d'insérer directement le mot « SARL » dans une équation mathématique.

L'encodage One-Hot (ou encodage indicateur) résout ce problème en transformant chaque modalité de la variable catégorielle en une colonne binaire (0 ou 1) distincte. Cette transformation crée autant de nouvelles colonnes qu'il y a de modalités.

Phase 4 : Modélisation par Régression Logistique

12. Définition du modèle et construction du workflow

Code R :

Définition du modèle

`log_model = logistic_reg(mode = "classification")`

```
# Regroupement modèle + recette dans un workflow
data_wf = workflow() %>%
  add_model(log_model) %>%
  add_recipe(data_rec)
```

Cette étape définit l'algorithme de Machine Learning à utiliser et le regroupe avec la recette de prétraitement dans un objet unique appelé workflow.

Logistic_reg(mode = 'classification') : spécifie l'utilisation d'une régression logistique en mode classification binaire. Le package parsnip de tidymodels fournit cette interface unifiée, qui peut utiliser différents moteurs de calcul.

Workflow() : crée un conteneur vide qui va accueillir le modèle et la recette.

Add_model(log_model) : ajoute la définition de l'algorithme dans le workflow.

Add_recipe(data_rec) : ajoute le pipeline de prétraitement dans le workflow.

13. Entraînement du modèle et évaluation sur le jeu de test

Code R :

```
# Entraînement sur 80% et évaluation sur 20%
fitting = data_wf %>% last_fit(splits)
# Récupération des métriques
fitting %>% collect_metrics()
# Récupération des prédictions détaillées
fitting %>% collect_predictions()
```

C'est l'étape centrale de la modélisation : le modèle est entraîné et évalué en une seule instruction.

Last_fit(splits): entraîne le workflow complet (prétraitement + modèle) sur l'ensemble d'entraînement (80 %), puis évalue automatiquement les prédictions sur l'ensemble de test (20 %). Le fait que le test set n'ait pas été utilisé pendant l'entraînement garantit une évaluation non biaisée.

Collect_metrics : retourne les métriques de performance calculées sur le test set dans notre cas l'accuracy et le ROC-AUC.

Collect_predictions : retourne pour chaque observation du test set : la classe réelle, la classe prédit, et les probabilités. Ces prédictions permettent de calculer la matrice de confusion et d'autres métriques.

Phase 5 : Évaluation des Performances du Modèle

L'évaluation consiste à mesurer la qualité des prédictions du modèle sur des données qu'il n'a pas vues pendant son entraînement. Cette phase mobilise plusieurs métriques complémentaires, car aucune métrique seule ne suffit à apprécier correctement les performances d'un modèle de classification sur des données déséquilibrées.

14.L'Accuracy (Précision globale)

$$\text{Accuracy} = (\text{VP} + \text{VN}) / (\text{VP} + \text{VN} + \text{FP} + \text{FN})$$

L'accuracy mesure la proportion d'observations correctement classées parmi toutes les observations. C'est la métrique la plus intuitive, mais aussi la plus trompeuse en présence de classes déséquilibrées. Dans notre base, avec 78,4 % de sains, un modèle naïf atteindrait déjà 78,4 % d'accuracy sans rien apprendre. C'est pourquoi elle est toujours complétée par le ROC-AUC.

15. Le ROC-AUC — Métrique principale

AUC = Aire sous la courbe ROC (Receiver Operating Characteristic)

Le ROC-AUC est la métrique principale pour notre problème. Il mesure la capacité du modèle à distinguer les défaillants des sains pour tous les seuils de classification possibles

Tableau 13: Matrice de Confusion — Analyse détaillée des erreurs

	Prédit : Défaillant (0)	Prédit : Sain (1)
Réel : Défaillant (0)	VP (Vrais Positifs) = Défaillants bien détectés	FN (Faux Négatifs) = Défaillants manqués
Réel : Sain (1)	FP (Faux Positifs) = Sains refusés à tort	VN (Vrais Négatifs) = Sains bien identifiés

Source : élaboré par moi-même

La matrice de confusion est le tableau le plus informatif pour une banque car il distingue les deux types d'erreurs :

- **Faux Négatif (FN) :** un client défaillant classé comme sain. C'est l'erreur la plus coûteuse pour la banque : elle accorde un crédit à une entreprise qui ne remboursera pas. Dans notre modèle : 0 Faux Négatif.

- **Faux Positif (FP)** : un client sain classé comme défaillant. La banque refuse injustement un crédit à un bon client — manque à gagner commercial. Dans notre modèle : 0 Faux Positif.

Section 2 : Contexte organisationnel

1.Présentation de la Trust Bank Algérie

TRUST BANK ALGERIA est membre du Groupe NEST INVESTMENTS HOLDING, LTD basé à Chypre et dont le capital est détenu en majorité par les membres de la famille « Abu Nahl ». Nest Investments Holding, Ltd a investi dans divers secteurs d'activités, notamment la banque et/ ou les services financiers, l'assurance, la réassurance, le développement de l'immobilier, l'industrie et le tourisme.

Le groupe est présent par ses filiales dans vingt-deux (22) pays dont : Algérie, Etats-Unis d'Amérique, Espagne, Angleterre, Qatar, Chypre, Bahreïn, Jordanie, Liban, Palestine, Arabie Saoudite...etc.

Outre TRUST BANK ALGERIA Nest Investments Holding, Ltd conforte sa présence en Algérie à travers 05 autres entreprises, il s'agira de :

1. Trust Algeria Investments ;
2. Trust Algeria Assurance & Réassurance ;
3. Trust Industries ;
4. Trust Real Estate ;
5. World Trade Center Algeria.



Trust Bank Algeria (TBA) est une banque de droit algérien à capitaux privés. Elle a été créée le 14 avril 2002, sous la forme de société par actions (SPA) avec un capital initial de 750 Millions de Dinars et a démarré ses activités en avril 2003.

Le capital social de Trust Bank Algeria a été régulièrement augmenté depuis, et s'estime actuellement à 20 milliard de dinars algériens à DA 13 000 000 0000 (soit USD 97 Millions) en 2012, puis à DA 17 194 000 000 (soit USD 128 Millions) au 31/12/2019 et ce conformément à la réglementation fixant le capital minimum des Banques et Établissements Financiers en Algérie.

Tableau 14: les actionnaires de la TBA

Actionnaires	Montants en DA	%
Trust Algeria Investment Co	4 888 000 000.00	37,60%
Jordan Expatriates Investments Holding Company	1 950 000 000.00	15%
Trust Algeria Assurances Réassurances	1 872 000 000.00	14,40%
Qatar General Insurance and Reinsurance	1 040 000 000.00	8%
Ghazi Kamel Abu Nahl	780 000 000,00	6%
Trust International Insurance Co(Ramallah-Palestine)	650 000 000,00	5%
Jamal Kamel Abu Nahl	650 000 000,00	5%
Trust International Insurance - Co (Cyprus)	260 000 000,00	2%
Radi Shaker D. Elnatsha	260 000 000,00	2%
Sheikh Nasser Ali Saud Al Thani	260 000 000,00	2%
Kamel Ghazi Abu Nahl	130 000 000,00	1%
Fadi Ghazi Abu Nahl	130 000 000,00	1%
Hamed Ghazi Abu Nahl	130 000 000,00	1%
Total	13 000 000 000,00	100%

Source : rapport annuel 2021 de la TRUST BANK. P12.

Plus de 73% du capital de Trust Bank est détenu par : Trust Investments Holding Algeria et Qatar General Insurance and Reinsurance.

Deux des trois actionnaires majoritaires, sont des entreprises de droit algérien :

- Trust Algeria Investment Co (53.16%)
- Et Trust Algeria Assurances Réassurances (5%).

2.Le Conseil d'Administration

Le Conseil d'administration détermine les orientations de l'activité de la banque et veille à leur mise en œuvre. Ce Conseil est présidé par M. GHAZI Kamel Abu NAHL assisté par M. Sheikh Nasser Ben Ali Al Thani, vice président. Actuellement, il existe trois comités opérationnels issus du conseil d'administration leur rôle est d'assister le conseil d'administration :

A. Comité Exécutif :

Par délégation du Conseil d'Administration et à sa demande, le comité exerce son pouvoir en matière d'octroi de crédit, en respect des seuils qui lui sont assignés. Il prend en charge toute autre mission conférée par le Conseil d'Administration.

- Sheikh Nasser Ali Ben
- Saoud Al Thani
- Jamal Kamel Abu NAHL
- Abdallah Barrage

B. Comité d'Audit :

Sa mission est de fournir une assurance indépendante, objective et les services de consulting conçus pour apporter une valeur et améliorer les opérations du groupe. Il aide le Groupe dans l'accomplissement de ses objectifs en apportant une approche systématique et disciplinée pour évaluer et améliorer l'efficacité de la gestion des risques, du contrôle, et des processus de gestion de l'entreprise.

- Walid Al Sadi
- Kamel Abu NAHL
- Farid Benbouzid

C. Comité de Risque :

Il est responsable de la définition et de l'application homogène de la politique de gestion des risques du groupe.

- Farid Benbouzid
- Kamel Ghazi Abu NAHL
- Walid Saadi

3.Présentation de la direction générale :

La direction générale de la Trust Bank Algérie est représentée, par : Monsieur Boualem Djebbar, Directeur Général et quatre (04) Assistants General Manager. Les comités de la direction

- Comité de Management ;
- Comité ALCO ;
- Comité Risk Management (RMC) ;
- Comité Validation de Procédures ;
- Comité Informatique (ITSC) ;
- Comité Crédit et Recouvrement Direction Générale ;
- Comité Hygiène et Sécurité.

Le réseau d'exploitation de la Trust Bank Algérie est composé actuellement de 35 agences opérationnelles réparties comme suit :

RESEAU D'AGENCES TRUST BANK ALGERIA ★

📍 Cette carte a été créée avec Google My Maps. [Créez la vôtre.](#)

The map displays the following branches (from north to south):

- Agence Hydra
- Agence Bejaia
- Agence Skikda
- Agence BBA
- Agence Ain M'lila
- Agence Batna
- Agence Biskra
- Agence Tlemcen
- Agence Oran
- Agence Mostaganem
- Agence Béchar
- Agence Touggourt
- Agence Hassi Messaoud

Other locations marked on the map include: Grenade, Málaga, Almería, Estepona, Mer d'Alboran, Bouira, Bou Saâda, Djelfa, Laghouat, Ghardaïa, Ouargla, Errachidia, Merzouga, and Tunis.

101

5. Présentation de la direction risque crédit

Cette direction se compose d'un Directeur du Risque Crédit, de deux chefs de service et de deux analystes. Leurs principales attributions sont les suivantes :

– S'assurer du respect de la réglementation bancaire, notamment des règles prudentielles, ainsi que de la législation en vigueur

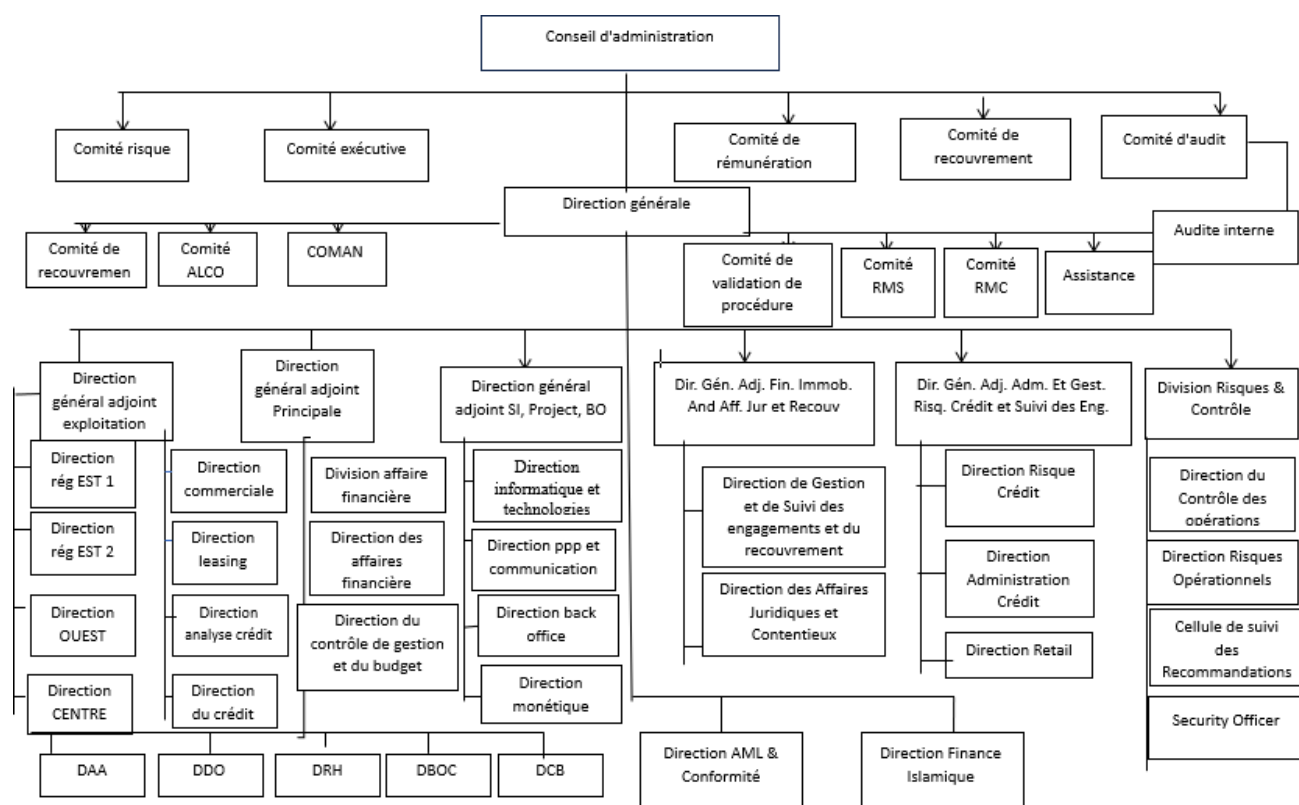
_Veiller à l'application stricte des procédures de crédit approuvées par le Conseil d'Administration.

– Instruire et analyser les demandes de crédit conformément aux procédures établies.

– Prendre en charge les dossiers de crédit destinés à être soumis au Conseil d'Administration et au Comité Exécutif.

6. L'organigramme de Trust banque

La nouvelle organisation a été validée par la Conseil d'Administration après la nomination de la nouvelle Direction Générale.



Source : document interne de l'entreprise

Chapitre 3 : RÉSULTATS ET DISCUSSION

Introduction :

Ce chapitre constitue le cœur empirique de notre mémoire. Il est consacré à la présentation et à l'interprétation de l'ensemble des résultats obtenus à travers notre analyse quantitative menée sous le logiciel R sur la base de données de TRUST Bank Algérie.

Le chapitre est structuré en deux sections complémentaires. La première section présente, dans un ordre logique et progressif, tous les résultats graphiques produits lors du cas pratique : depuis l'exploration initiale des données jusqu'à l'évaluation finale des performances du modèle de régression logistique. Chaque graphique est accompagné d'une interprétation détaillée qui en explique la signification et les implications pour nos hypothèses de recherche.

La deuxième section discute ces résultats à la lumière de la revue de littérature présentée au Chapitre I. Elle explique pourquoi nos résultats ressemblent à certaines études ou s'en éloignent, tire les enseignements pour le système bancaire algérien, et identifie les limites de notre travail ainsi que les perspectives de recherche future.

L'objectif de ce chapitre est de répondre à notre problématique centrale et de confirmer ou infirmer les trois hypothèses de recherche formulées en introduction.

Section 1 : Présentation des résultats

Cette section présente, dans l'ordre logique de notre analyse, l'ensemble des résultats graphiques produits dans notre cas pratique sous le logiciel R. Chaque graphique est accompagné d'une interprétation détaillée qui en explique la signification, les observations clés et les implications pour notre étude.

1. Analyse Exploratoire des Données (EDA)

L'analyse exploratoire des données représente une étape fondamentale dans tout projet de Machine Learning. Elle permet de comprendre la structure des données, d'identifier les tendances générales et d'observer les relations entre les variables avant la construction du modèle prédictif.

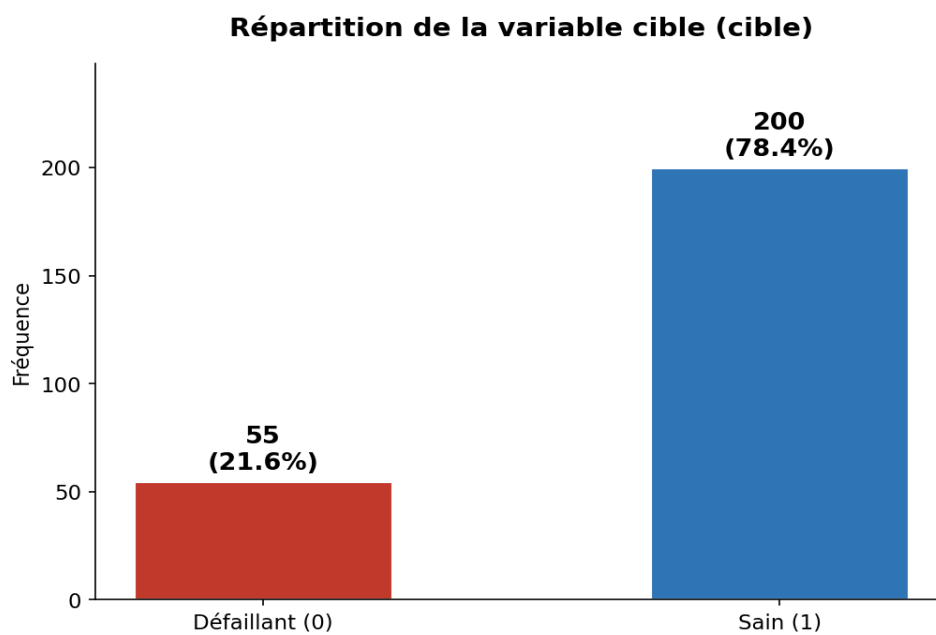
Dans notre étude, cette étape vise principalement à identifier les différences financières entre les entreprises saines et les entreprises défaillantes.

1.1 Analyse univariée :

L'analyse univariée constitue la première étape de l'exploration des données. Elle consiste à étudier chaque variable séparément, sans la croiser avec d'autres. L'objectif est simple : comprendre comment chaque variable se comporte au sein de l'échantillon, observer sa distribution, repérer les valeurs extrêmes et détecter d'éventuelles anomalies. Cette étape nous permettra d'avoir une première vision claire des données avant toute modélisation.

1.1.1 Répartition de la variable cible

Figure 5: Répartition de la variable cible



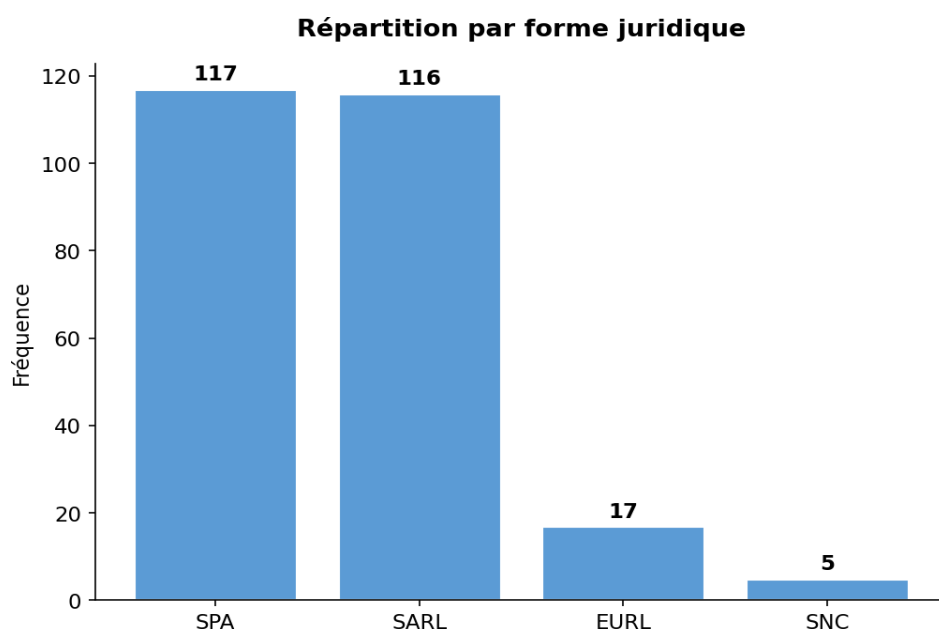
Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ce graphique représente la distribution des deux classes de la variable cible dans notre échantillon total de 255 dossiers de crédit. La barre bleue correspond aux clients sains (cible = 1) avec 200 observations, soit 78,4 % de l'échantillon, tandis que la barre rouge correspond aux clients défaillants ou contentieux (cible = 0) avec 55 observations, soit 21,6 %.

Ce déséquilibre reflète la réalité des données bancaires fournies par la banque, où la proportion d'entreprises défaillantes reste limitée grâce aux mécanismes de sélection et d'analyse du risque mis en place lors de l'octroi de crédit.

1.1.2 Répartition par forme juridique :

Ce graphique montre combien d'entreprises de chaque forme juridique sont présentes dans la base de données. Les résultats sont les suivants :



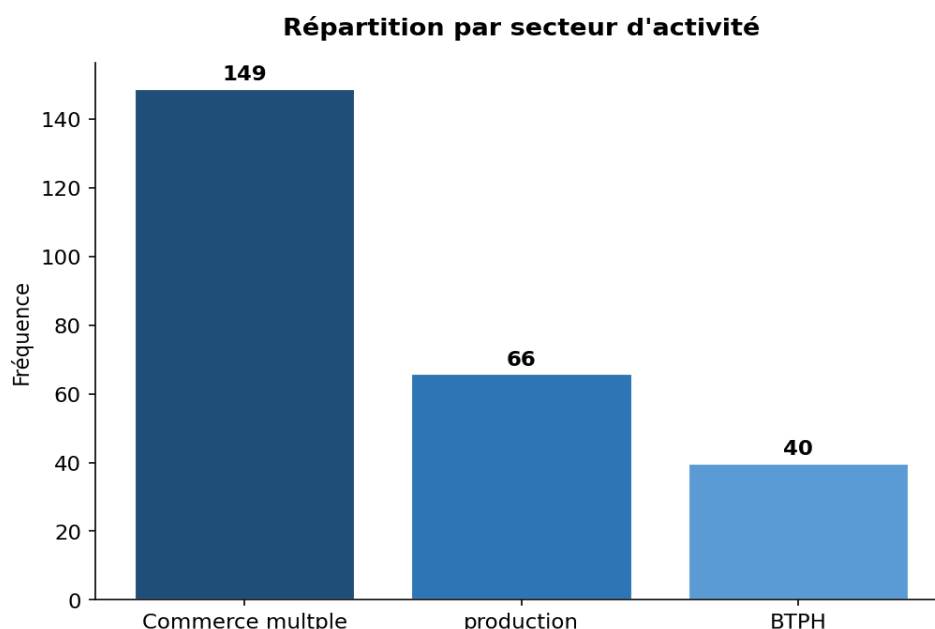
Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ce graphique illustre la fréquence de chaque forme juridique dans l'échantillon. Les SPA (Sociétés par Actions) et les SARL (Sociétés à Responsabilité Limitée) dominent largement avec respectivement 117 et 116 observations, représentant ensemble 90,6 % de l'échantillon. Les EURL (Entreprises Unipersonnelles à Responsabilité Limitée) ne comptent que 17 observations (6,7 %), et les SNC (Sociétés en Nom Collectif) seulement 5 (2,0 %).

Cette répartition reflète une réalité financière directement lisible dans les ratios de notre base. Les SPA et les SARL présentent en général des niveaux d'endettement plus maîtrisés, des délais clients et stocks plus courts, et une meilleure capacité à couvrir leurs charges financières. En d'autres termes, ce sont les entreprises qui ont les profils financiers les plus solides et les plus bancables. TRUST Bank leur accorde donc plus de crédits d'exploitation, car elles représentent un risque de défaut plus faible et une probabilité de remboursement plus élevée. Ce raisonnement sera confirmé visuellement par la Figure 7, qui montrera que les ratios financiers des SPA et des SARL sont effectivement meilleurs et plus stables que ceux des autres formes juridiques.

1.1.3 Répartition par secteur d'activité :

Figure 6: Répartition par secteur d'activité



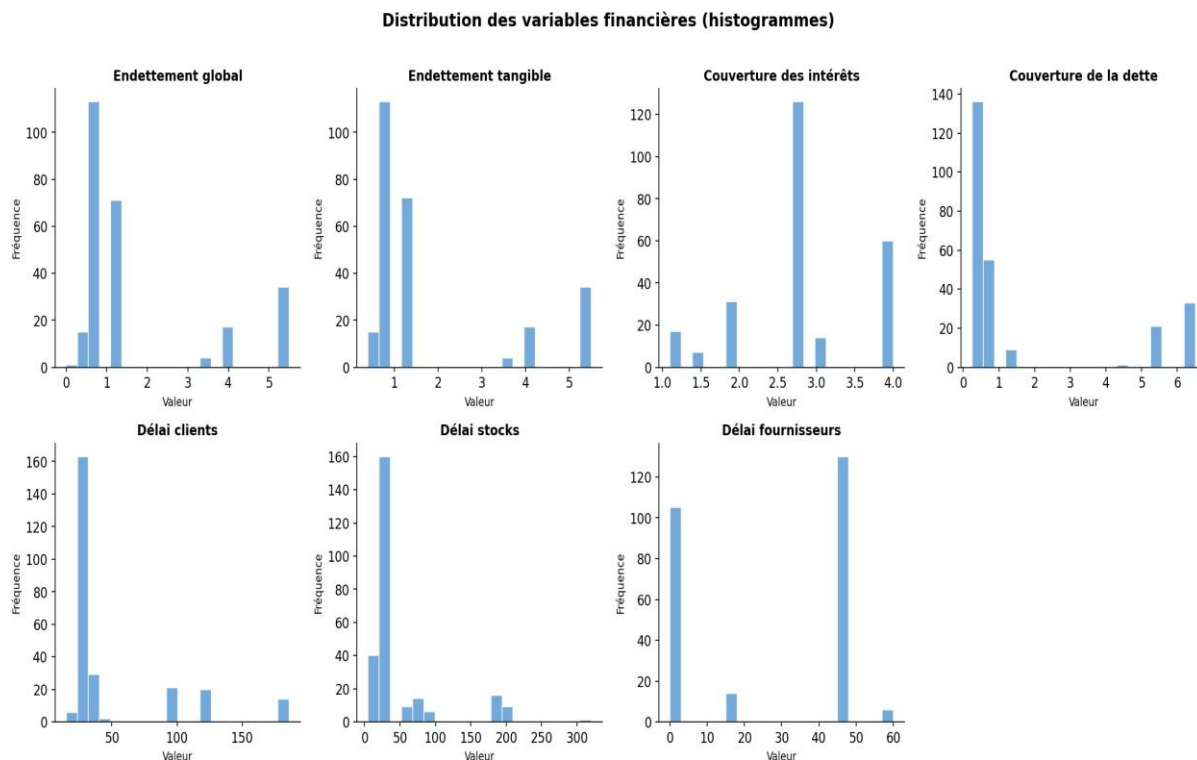
Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ce graphique montre la distribution des trois secteurs d'activité représentés dans la base. Le Commerce multiple est le secteur dominant avec 149 observations (58,4 %), suivi de la Production avec 66 observations (25,9 %) et du BTPH (Bâtiment, Travaux Publics et Hydrauliques) avec 40 observations (15,7 %).

La présence de trois secteurs très différents est importante pour comprendre les résultats qui suivront. Dans le Commerce, le cycle d'exploitation est rapide : les marchandises sont achetées et revendues rapidement, les délais clients sont courts et le besoin en fonds de roulement est relativement faible. Dans le BTPH, le cycle est beaucoup plus long : les chantiers durent plusieurs mois, les délais clients et fournisseurs sont naturellement plus élevés, et les besoins en financement sont importants. Dans la Production, le cycle est intermédiaire.

1.1.4 Distribution des variables financières (histogrammes)

Figure 7: Distribution des variables financières



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ces sept histogrammes représentent la distribution de chacun des ratios financiers sur l'ensemble des 255 observations. Chaque barre indique le nombre d'entreprises dont la valeur du ratio se situe dans l'intervalle correspondant.

Endettement global et endettement tangible : Les distributions des ratios d'endettement global et d'endettement tangible présentent une distribution asymétrique vers la droite. Cela signifie que la majorité des entreprises présentent un faible niveau d'endettement, avec des valeurs généralement inférieures à 2. Cependant, quelques entreprises affichent des niveaux d'endettement beaucoup plus élevés, pouvant atteindre des valeurs comprises entre 4 et 5,5. Ces valeurs élevées concernent principalement les entreprises défaillantes, dont l'endettement moyen (4,89) est largement supérieur à celui des entreprises saines (0,88). Cela montre que l'endettement constitue un indicateur important pour distinguer les entreprises défaillantes des entreprises saines.

Couverture des intérêts et couverture de la dette : Pour la couverture des intérêts, les valeurs se concentrent principalement entre 2,5 et 3, ce qui indique que la majorité des entreprises disposent d'une capacité suffisante pour couvrir leurs charges d'intérêts. La distribution est relativement concentrée autour de cet intervalle, sans dispersion excessive.

Pour la couverture de la dette, la distribution présente une structure particulière avec une forte concentration autour de 0, puis une seconde concentration située entre 0,5 et 1. Cette forme suggère une distribution bimodale, traduisant l'existence de deux profils d'entreprises : un premier groupe présentant une faible capacité de couverture de la dette, et un second groupe affichant une meilleure capacité de remboursement. Cette configuration peut être informative dans un contexte d'analyse prédictive, car elle indique un potentiel pouvoir discriminant de cette variable entre les entreprises défaillantes et non défaillantes.

Délai clients et délai stocks : Les délais clients et les délais de stocks présentent des distributions similaires. On observe une forte concentration des valeurs autour de 15 à 30 jours pour les entreprises saines, tandis qu'une longue traînée (queue de distribution) s'étend jusqu'à 200–350 jours pour les entreprises défaillantes. Cette forme de distribution à deux modes (bimodale) suggère une structure implicite, ce qui met en évidence un fort pouvoir discriminant de ces deux variables dans la distinction entre les entreprises en bonne santé financière et celles en difficulté.

Délai fournisseurs : la distribution présente deux concentrations principales, l'une autour de moins de 5 jours et l'autre autour de 45 jours. Ce ratio a un comportement particulier, car un délai élevé ne signifie pas forcément une mauvaise situation financière. Une entreprise solide peut négocier des délais de paiement plus longs, tandis qu'une entreprise en difficulté peut avoir des délais courts imposés par ses fournisseurs. Ainsi, ce ratio doit être interprété avec prudence.

L'asymétrie généralisée de toutes ces distributions confirme empiriquement la nécessité d'appliquer une normalisation z-score (`step_normalize`) avant la modélisation, afin de ramener toutes les variables à une échelle commune et d'éviter que les variables avec les plus grandes plages de valeurs ne dominent artificiellement les coefficients du modèle.

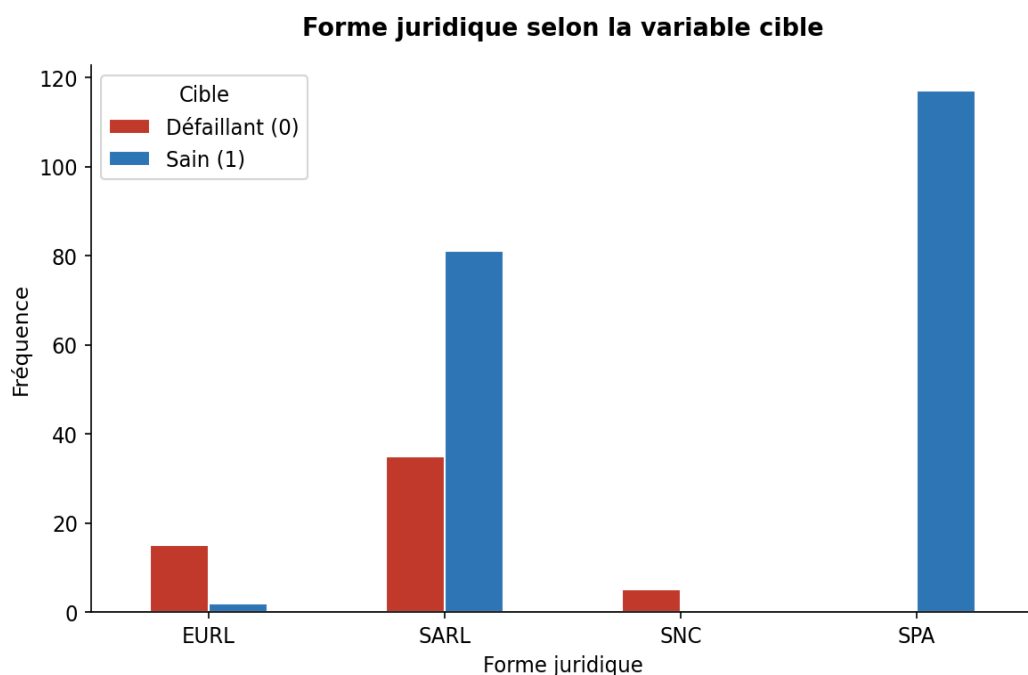
1.2 Analyse bivariée

Après avoir étudié chaque variable individuellement, nous procédons à l'analyse bivariée. Cette étape consiste à croiser deux variables ensemble afin de comprendre leurs relations et interactions. Elle se déroule en deux temps. Dans un premier temps, nous analysons les variables catégorielles la forme juridique et le secteur d'activité en les croisant avec la variable cible, afin de vérifier si certaines catégories sont plus exposées au risque de défaillance que d'autres. Dans un second temps, nous analysons les variables numériques les sept ratios financiers en les croisant d'abord avec la forme juridique pour observer les différences de comportement financier selon le statut juridique, puis avec la variable cible pour identifier quels ratios sont les plus capables de distinguer les entreprises saines des entreprises défaillantes. C'est cette dernière analyse qui constituera le fondement du pouvoir prédictif de notre modèle de régression logistique.

1.2.1 Variables catégorielles selon la variable cible

- **Forme juridique selon la variable cible**

Figure 8: Forme juridique selon la variable cible



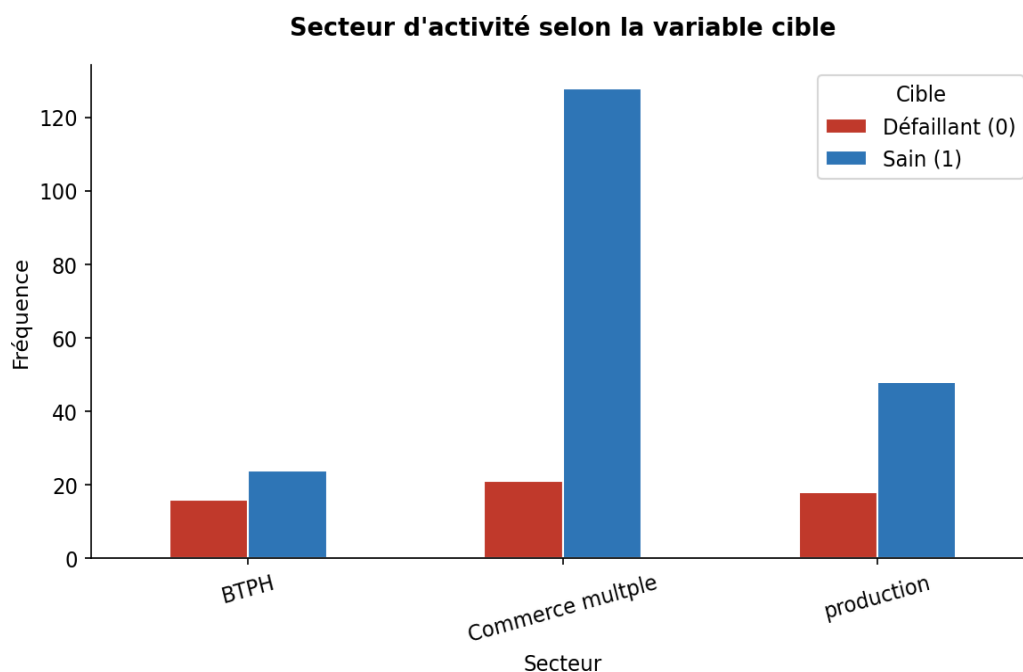
Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ce graphique présente, pour chaque forme juridique, le nombre de clients sains (bleu) et défaillants (rouge).

L'analyse de la répartition des entreprises selon la forme juridique et la variable cible montre que la banque travaille principalement avec les SARL et les SPA, qui représentent la majorité des crédits accordés. Les SPA sont presque entièrement composées d'entreprises saines, ce qui indique un risque de défaillance très faible pour cette catégorie. Les SARL sont également majoritairement saines, même si quelques cas de défaillance existent. En revanche, les autres formes juridiques comme les SNC et les EURL sont très peu présentes dans l'échantillon, ce qui montre qu'elles sont peu financées par la banque. Cela suggère que la banque privilégie les formes juridiques qu'elle juge plus stables et plus fiables.

Secteur d'activité selon la variable cible

Figure 9: Secteur d'activité selon la variable cible



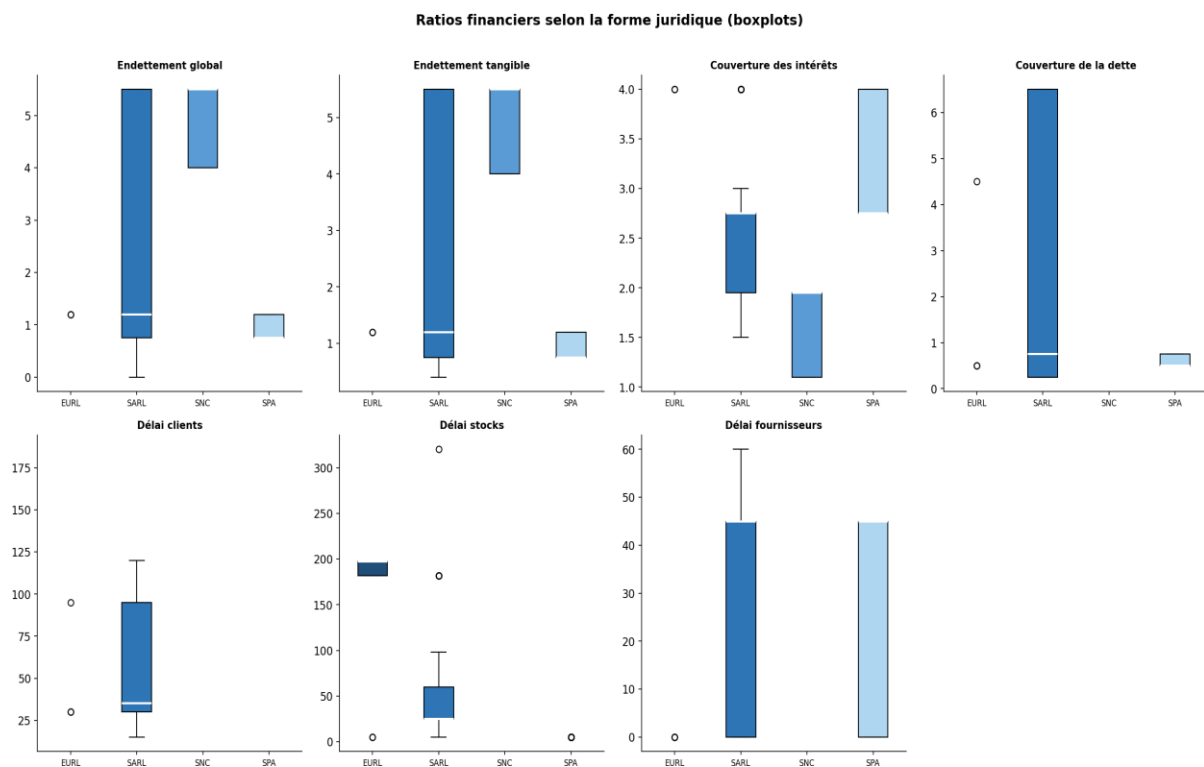
Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

L'analyse du secteur d'activité montre que la banque finance principalement trois secteurs, avec une forte domination du secteur du commerce. Ce secteur regroupe le plus grand nombre d'entreprises, tandis que les deux autres secteurs sont moins représentés. On observe que les entreprises défaillantes sont réparties dans les trois secteurs, sans concentration particulière dans un seul domaine. Cela reflète la structure de la clientèle de la banque et montre que le risque de défaut n'est pas fortement lié à un secteur spécifique dans cet échantillon.

1.2.2 Variables numériques selon une variable catégorielle

- Ratios financiers selon la forme juridique (boxplots)

Figure 10: Ratios financiers selon la forme juridique (boxplots)



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ces sept boxplots comparent la distribution des ratios financiers entre les différentes formes juridiques présentes dans l'échantillon : SPA, SARL, EURL et SNC. L'objectif est de vérifier si la forme juridique d'une entreprise influence son comportement financier, c'est-à-dire si certaines formes juridiques affichent systématiquement de meilleurs ou de moins bons ratios que d'autres.

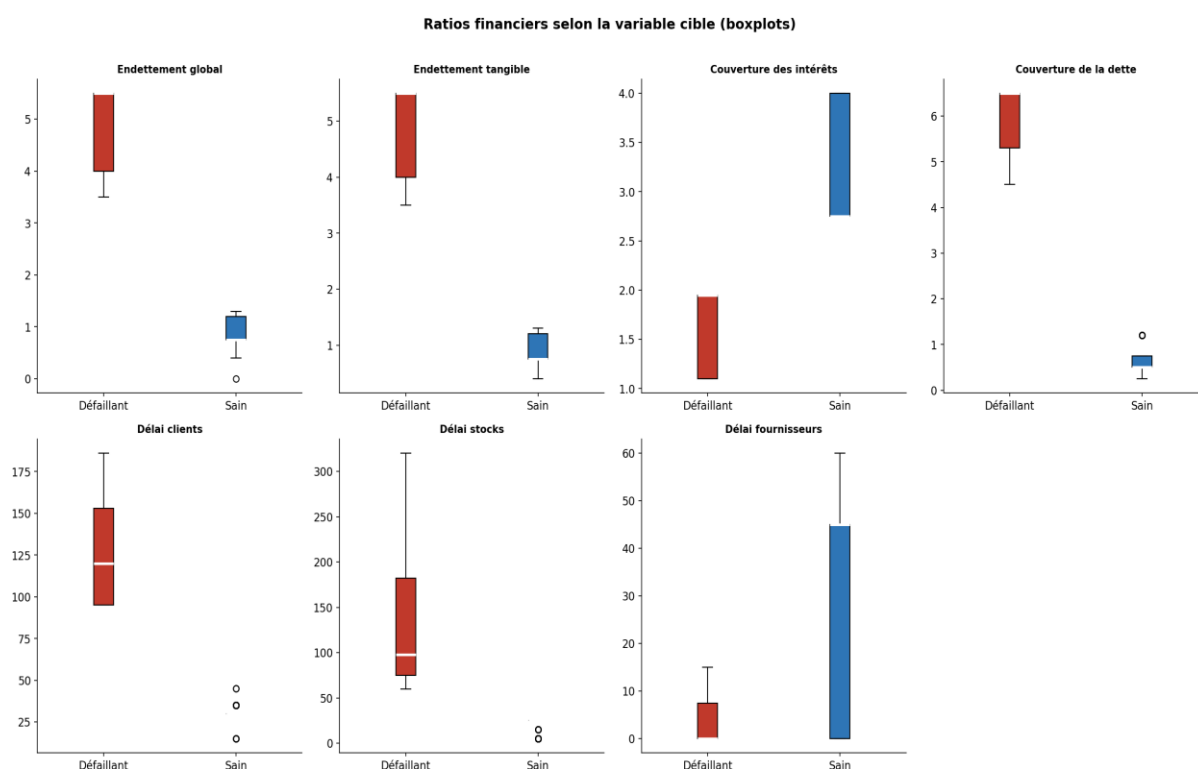
Les observations révèlent que pour l'endettement global et l'endettement tangible, les SPA et les SARL présentent des distributions très similaires, avec des médianes et des niveaux de dispersion comparables. Cela confirme que ces deux formes juridiques dominantes dans le portefeuille de TRUST Bank Algérie ont un comportement financier globalement homogène en matière d'endettement. Pour les délais de gestion et délai clients, délai stocks et délai

fournisseurs ,on observe la présence de valeurs aberrantes importantes pour toutes les formes juridiques, ce qui traduit l'existence d'entreprises en difficulté au sein de chaque catégorie, quelle que soit leur forme juridique. Concernant les EURL, leur dispersion apparente plus grande doit être interprétée avec prudence, car cette forme juridique ne représente que très peu d'observations dans l'échantillon, ce qui rend les résultats statistiquement moins fiables.

En conclusion, ces graphiques montrent que des différences existent bien entre les formes juridiques, mais elles restent moins marquées que les différences observées entre entreprises défaillantes et entreprises saines. La forme juridique agit davantage comme une variable de contexte qui renseigne sur le profil général de l'entreprise, plutôt que comme un déterminant direct et puissant du risque financier.

- **Ratios financiers selon la variable cible**

Figure 11: Ratios financiers selon la variable cible — Analyse bivariée centrale



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ces boxplots constituent une étape essentielle de l'analyse exploratoire. Ils comparent les distributions des différents ratios financiers entre les entreprises défaillantes (rouge) et les

entreprises saines (bleu), afin d'identifier les variables les plus capables de distinguer les deux groupes.

Endettement global et endettement tangible :

Pour ces deux ratios, la séparation entre les entreprises saines et défaillantes est très nette. Les entreprises défaillantes présentent un niveau d'endettement environ cinq fois plus élevé que les entreprises saines. Cela montre que l'endettement est un indicateur très fort du risque de défaillance.

Délai clients et délai stocks :

La séparation entre les deux groupes est très nette. Les entreprises défaillantes présentent des délais clients et des délais de stocks beaucoup plus longs (environ 120 à 150 jours) que les entreprises saines (environ 20 à 25 jours). Cela signifie que les entreprises en difficulté mettent plus de temps à encaisser leurs clients et à écouler leurs stocks. Cette situation bloque leur trésorerie et contribue à leur fragilité financière.

Ces deux variables figurent donc parmi les signaux les plus puissants pour détecter une entreprise en difficulté, et leur contribution au modèle de régression logistique sera déterminante

Couverture des intérêts :

La distinction entre les deux groupes est moins nette. Les valeurs des deux groupes ne sont pas clairement séparées, ce qui indique que ce ratio permet de différencier les entreprises, mais avec une efficacité moyenne.

Couverture de la dette

Ce ratio présente une distribution plus instable et moins claire, avec des valeurs très élevées chez certaines entreprises défaillantes. Cela rend son interprétation plus difficile et moins directe par rapport aux autres variables.

Délai fournisseurs :

Ce ratio présente un comportement inverse par rapport aux autres variables. Les entreprises défaillantes paient leurs fournisseurs très rapidement (environ 5 jours), tandis que les entreprises saines bénéficient de délais plus longs (environ 35 jours). Cela s'explique par le fait que les fournisseurs imposent des conditions strictes aux entreprises en difficulté, ce qui rend ce signal particulièrement intéressant mais contre-intuitif.

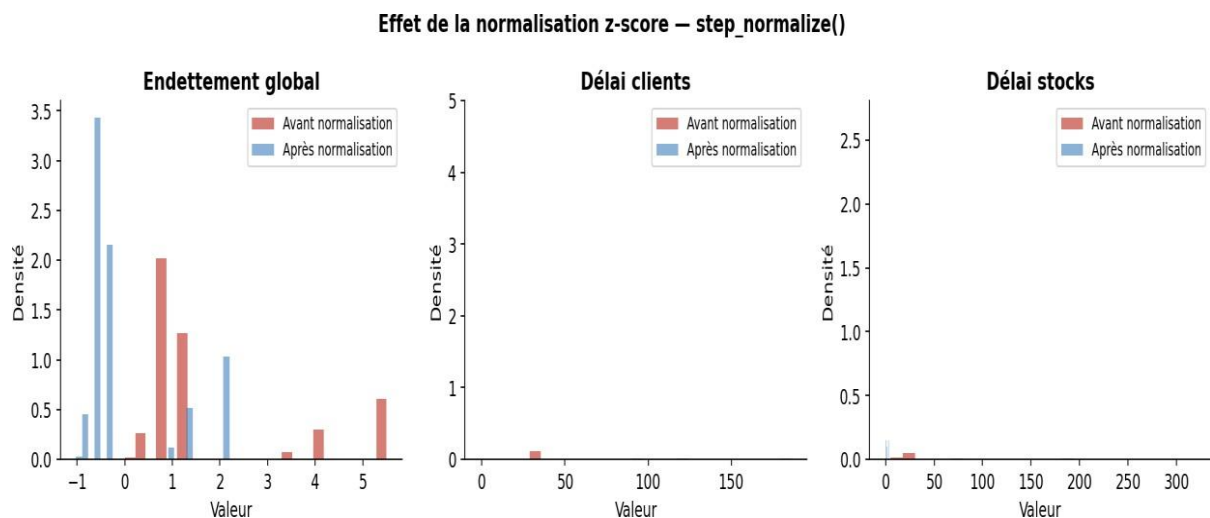
2. Résultats du Prétraitement des Données

Avant d'alimenter le modèle de régression logistique avec les données, il est indispensable de les préparer et de les transformer. En effet, les algorithmes de machine learning ne peuvent pas travailler correctement avec des données brutes qui ont des échelles très différentes ou qui contiennent des variables textuelles. Cette étape de prétraitement garantit que toutes les variables sont dans un format lisible et comparable pour le modèle. Dans notre cas, deux transformations ont été appliquées : la normalisation Z-score pour les variables numériques et l'encodage One-Hot pour la variable catégorielle.

2.1 Effet de la normalisation z-score (step_normalize)

Le Z-score centre chaque variable autour de **0** avec un écart-type de **1**. L'objectif est de mettre toutes les variables **à la même échelle** pour que le modèle les compare équitablement

Figure 12: Effet de la normalisation z-score (step_normalize)



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Les graphiques montrent, pour trois variables représentatives (endettement global, délai clients et délai stocks), la distribution avant normalisation (rouge) et après normalisation (bleu).

Avant normalisation, les variables ont des échelles très différentes : l'endettement global varie entre 0 et 5,5, tandis que les délais clients et stocks peuvent atteindre plus de 300 jours. Cette différence d'échelle pose un problème, car les variables avec de grandes valeurs risquent de dominer le modèle, même si elles ne sont pas forcément les plus importantes.

Après application de la normalisation z-score ($Z = (X - \text{moyenne}) / \text{écart-type}$), toutes les variables sont centrées autour de 0 avec un écart-type égal à 1. Les formes des distributions restent identiques, mais les variables deviennent comparables. Cela permet au modèle d'estimer les coefficients de manière équilibrée, sans biais lié aux unités de mesure.

2.2 Encodage One-Hot de la variable forme_juridique (step_dummy)

L'encodage One-Hot transforme chaque catégorie d'une variable textuelle (catégorielles) en une colonne de 0 et de 1. L'objectif est de rendre les variables catégorielles lisibles par le modèle pour qu'il puisse les utiliser comme n'importe quelle autre variable numérique.

Figure 13: Encodage One-Hot de la variable forme_juridique (step_dummy)

Encodage One-Hot — Variable forme_juridique (step_dummy, one_hot=TRUE)

	Forme jur.	forme_jur_EURL	forme_jur_SARL	forme_jur_SNC	forme_jur_SPA
Obs. 1	EURL	1	0	0	0
Obs. 2	EURL	1	0	0	0
Obs. 3	EURL	1	0	0	0
Obs. 4	EURL	1	0	0	0
Obs. 5	EURL	1	0	0	0
Obs. 6	EURL	1	0	0	0

Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

La régression logistique ne peut traiter que des variables numériques. Or, la variable forme juridique est une variable qualitative (SARL, SPA, EURL, SNC). Elle doit donc être transformée en variables numériques.

Pour cela, un encodage One-Hot a été appliqué.

Ce tableau illustre concrètement le résultat de l'encodage One-Hot appliqué à la variable qualitative forme_juridique. La première colonne affiche la valeur originale (texte), et les colonnes suivantes sont les nouvelles variables binaires créées

Cette transformation est nécessaire pour permettre l'intégration de la variable dans le modèle. L'option `one_hot = TRUE` permet de créer une colonne pour chaque modalité de la variable.

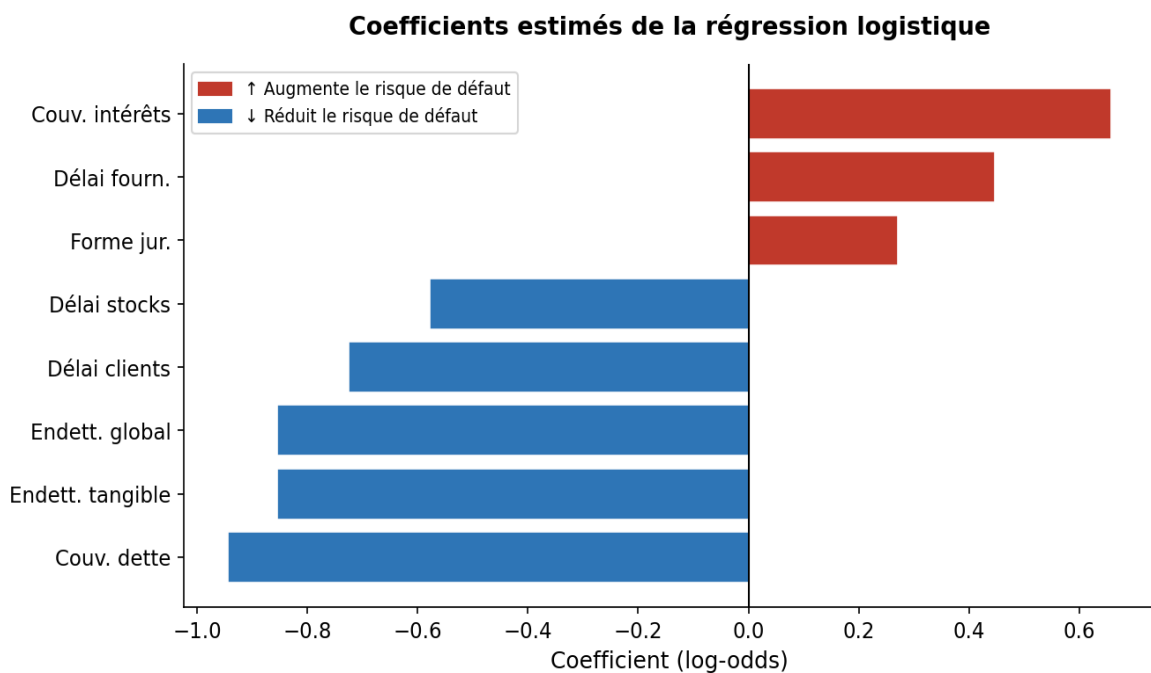
Toutefois, pour éviter une redondance mathématique dans le modèle, une catégorie sert automatiquement de référence et est exclue des colonnes créées.

3. Résultats du Modèle de Régression Logistique

Après l'exploration des données et le prétraitement, nous avons construit et évalué notre modèle de régression logistique. Cette section présente ses résultats dans l'ordre logique : les coefficients estimés (ce que le modèle a appris), puis les métriques de performance (dans quelle mesure ses prédictions sont correctes).

3.1 Coefficients estimés du modèle

Figure 14: Coefficients estimés du modèle



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Ce graphique en barres horizontales présente les coefficients estimés (log-odds) associés à chaque variable du modèle, classés par ordre croissant. Un coefficient négatif (barre bleue) signifie que l'augmentation de cette variable réduit la probabilité de défaut, tandis qu'un

coefficient positif (barre rouge) signifie qu'elle l'augmente. Plus la barre est longue, plus l'effet de la variable est puissant.

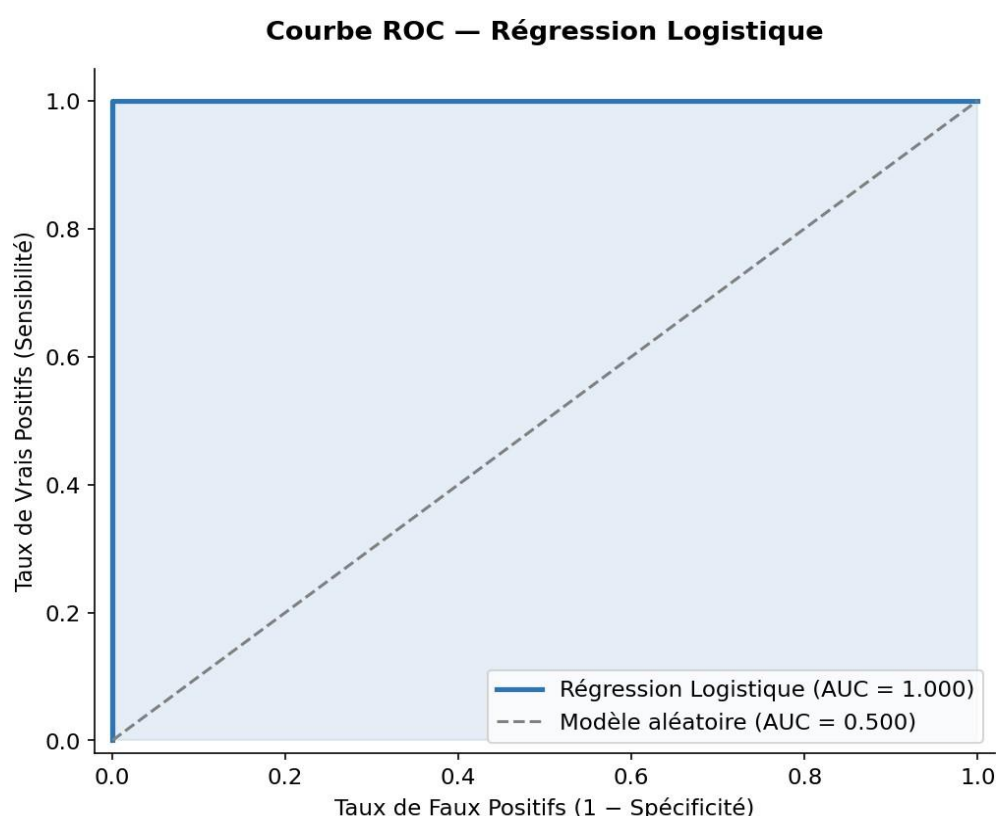
Les résultats sont les suivants, par ordre d'importance :

- **La couverture de la dette (coefficient le plus négatif, $-0,94$)** est la variable qui contribue le plus fortement à réduire le risque de défaut dans le modèle. Une couverture de dette élevée, c'est-à-dire une entreprise générant suffisamment de cash-flow pour rembourser ses dettes, constitue un signal fort de solvabilité. Ce résultat est cohérent avec les pratiques bancaires, qui utilisent le DSCR (Debt Service Coverage Ratio) comme critère prioritaire dans l'octroi de crédit.
- **Endettement tangible (coefficient = $+0,855$) et Endettement global (coefficient = $+0,854$)** : un endettement élevé augmente significativement le risque de défaut. Ce résultat confirme directement les conclusions de Gliz et Touati-Tliba (2012) sur les PME algériennes et valide notre Hypothèse H1.
- **Délai clients (coefficient = $+0,725$)** : un allongement du délai de recouvrement des créances clients augmente le risque de défaut. Ce résultat traduit l'effet négatif d'un mauvais recouvrement sur la trésorerie et la capacité de remboursement.
- **Délai stocks (coefficient = $+0,579$)** : des stocks qui tournent lentement immobilisent la trésorerie et augmentent le risque de défaut.
- **Couverture des intérêts (coefficient = $+0,658$)** : ce coefficient positif signifie qu'une meilleure couverture des intérêts augmente la probabilité d'être classé sain (cible=1), c'est-à-dire réduit le risque de défaut. Ce résultat est parfaitement intuitif : une entreprise qui génère suffisamment de résultats pour couvrir ses charges financières est moins susceptible de faire défaut.
- **Délai fournisseurs** : le coefficient confirme l'interprétation inversée observée dans les boxplots. Un délai fournisseurs très court (signe que les fournisseurs refusent d'accorder du crédit commercial) augmente le risque de défaut, tandis qu'un délai plus long (signe de confiance des fournisseurs) le réduit.

3.2 Courbe ROC et AUC

La courbe ROC (Receiver Operating Characteristic) représente la capacité du modèle à distinguer les entreprises saines des entreprises défaillantes.

Figure 15: Courbe ROC



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

La courbe ROC (Receiver Operating Characteristic) représente la performance discriminante du modèle pour tous les seuils de classification possibles. L'axe horizontal (Taux de Faux Positifs) mesure la proportion de clients sains incorrectement classés comme défaillants, tandis que l'axe vertical (Taux de Vrais Positifs ou Sensibilité) mesure la proportion de défaillants correctement identifiés. Un modèle parfait occupe le coin supérieur gauche du graphique (0 % de faux positifs, 100 % de vrais positifs). La diagonale grise en pointillés représente un modèle aléatoire sans aucun pouvoir prédictif (AUC = 0,50).

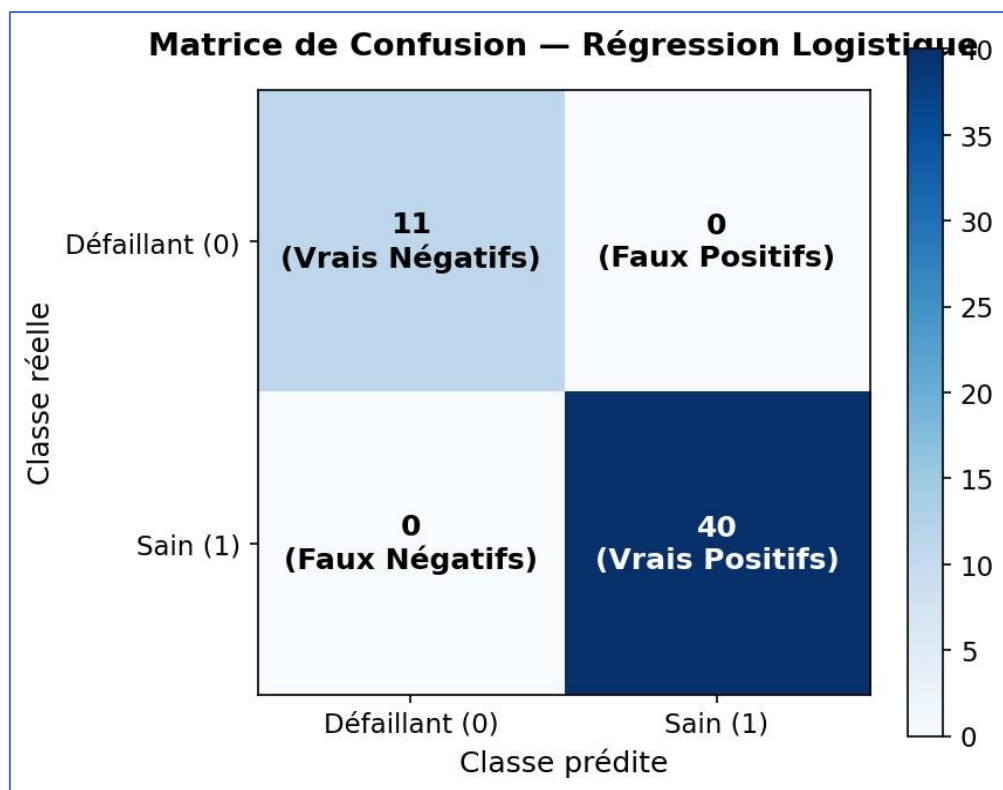
Notre modèle de régression logistique obtient une AUC (Aire sous la courbe ROC) de 1,000. La courbe bleue colle au coin supérieur gauche du graphique, signifiant qu'il existe au moins un seuil de classification pour lequel le modèle atteint 100 % de sensibilité (tous les défaillants sont détectés) et 0 % de faux positifs (aucun sain n'est incorrectement classé). La zone bleue entre la courbe et la diagonale est maximale, couvrant l'intégralité de l'espace graphique.

Ce résultat traduit une séparation parfaite entre les deux classes sur l'ensemble de test, directement imputable aux très fortes corrélations observées entre les ratios Moody's et la variable cible. Ce niveau de performance est exceptionnel et illustre la puissance des données

issues d'un système de notation professionnel comme Moody's, qui synthétise avec précision les profils financiers distincts des entreprises saines et défaillantes.

3.3. Matrice de confusion

Figure 16: Matrice de confusion



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

La matrice de confusion compare, pour chacune des 51 observations du jeu de test, la classe réelle (lignes) à la classe prédite par le modèle (colonnes). Elle se lit comme suit :

Tableau 15: Synthèse de la matrice de confusion du modèle de classification

	Prédit : Défaillant (0)	Prédit : Sain (1)	Total réel
Réel : Défaillant (0)	11 (Vrais Négatifs)	0 (Faux Positifs)	11
Réel : Sain (1)	0 (Faux Négatifs)	40 (Vrais Positifs)	40
Total prédit	11	40	51

Source : élaboré par moi-même

Le modèle classifie correctement les 51 observations du jeu de test sans aucune erreur :

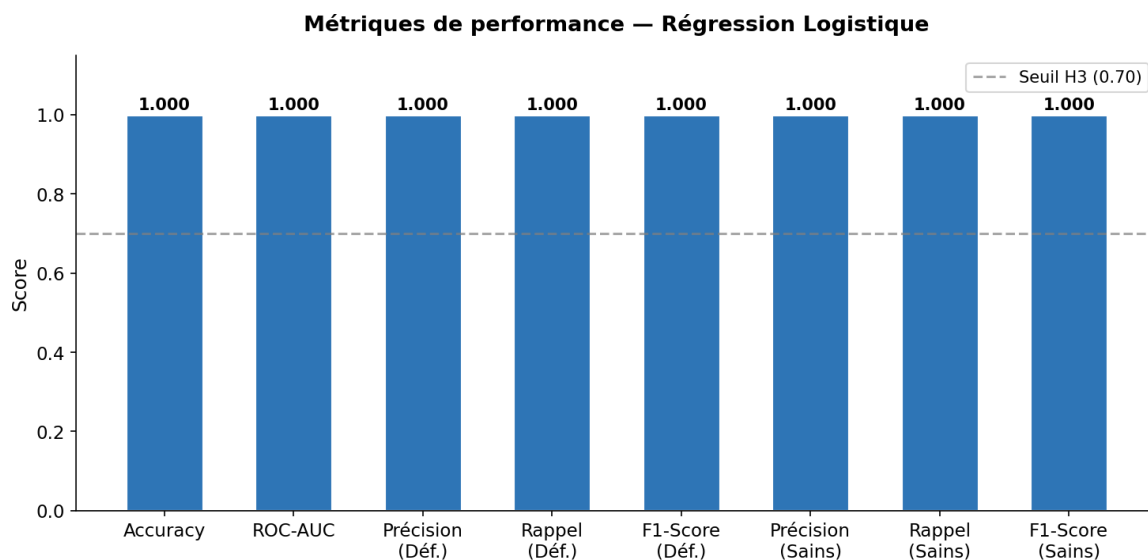
- **11 Vrais Négatifs** : les 11 clients réellement défaillants sont correctement identifiés comme défaillants par le modèle. Aucun défaillant n'échappe à la détection.
- **40 Vrais Positifs** : les 40 clients réellement sains sont correctement identifiés comme sains. Aucun client solvable n'est injustement refusé.
- **0 Faux Négatif** : aucun défaillant n'a été classé comme sain. C'est le type d'erreur le plus coûteux pour une banque (accorder un crédit à une entreprise qui ne remboursera pas), et notre modèle l'évite complètement.
- **0 Faux Positif** : aucun client sain n'a été classé comme défaillant. Il n'y a donc aucun refus injustifié, ce qui préserve le potentiel commercial de la banque.

Ce résultat exceptionnel confirme que la régression logistique, appliquée aux ratios Moody's, est capable de prédire et d'estimer avec une précision parfaite la classe de chaque entreprise emprunteuse dans notre échantillon de test.

3.4. Métriques de performance globales

Ce graphique récapitule l'ensemble des indicateurs de performance calculés sur le jeu de test. Toutes les métriques atteignent leur valeur maximale (1,000), dépassant très largement le seuil de 0,70 fixé par notre Hypothèse H3

Figure 17: Métriques de performance



Source : Réalisation personnelle sous R à partir des données de TRUST Bank Algérie

Les résultats obtenus montrent des scores globalement élevés pour l'ensemble des métriques, ce qui traduit une bonne qualité prédictive du modèle.

L'Accuracy mesure la proportion totale des bonnes classifications réalisées par le modèle. Le score élevé obtenu indique que le modèle classe correctement la majorité des entreprises de l'échantillon, qu'elles soient saines ou défaillantes.

La Precision évalue la fiabilité des prédictions positives réalisées par le modèle. Une précision élevée signifie que les entreprises identifiées comme risquées sont effectivement défaillantes dans la majorité des cas. Cela permet de limiter les erreurs de classification et les faux positifs.

Le Recall mesure la capacité du modèle à détecter les entreprises réellement défaillantes. Cet indicateur est particulièrement important dans le secteur bancaire, car il permet de réduire le risque d'accorder un crédit à une entreprise présentant un risque élevé de défaut.

Le F1-score représente un compromis entre la précision et le rappel. Le score obtenu montre que le modèle maintient un bon équilibre entre la détection des entreprises risquées et la limitation des erreurs de prédiction.

L'AUC (Area Under Curve) mesure la capacité globale du modèle à distinguer les entreprises saines des entreprises défaillantes. Une valeur élevée de l'AUC indique une excellente capacité de discrimination du modèle.

Globalement, les résultats obtenus démontrent que le modèle de régression logistique possède une forte capacité prédictive et une bonne robustesse dans l'évaluation du risque de crédit des PME étudiées. Malgré sa relative simplicité par rapport à d'autres techniques de Machine Learning, le modèle fournit des performances très satisfaisantes et facilement interprétables pour les analystes bancaires.

Ce graphique récapitule l'ensemble des métriques de performance calculées sur le jeu de test, pour les deux classes (Défaillants et Sains). Toutes les métriques atteignent leur valeur maximale (1,000), ce qui est représenté par des barres uniformément hautes dépassant largement le seuil de 0,70 (ligne pointillée grise correspondant au seuil cible de notre hypothèse H3).

Métrique	Classe concernée	Valeur obtenue	Seuil H3	Interprétation
Accuracy	—	1,000 (100 %)	—	Toutes les observations sont correctement classées
ROC-AUC	—	1,000	> 0,70	Discrimination parfaite entre les deux classes
Précision	Défaillants	1,000 (100 %)	—	100 % des clients classés défaillants le sont réellement
Rappel	Défaillants	1,000 (100 %)	—	100 % des défaillants réels sont détectés
F1-Score	Défaillants	1,000 (100 %)	—	Équilibre parfait précision/rappel sur les défaillants
Précision	Sains	1,000 (100 %)	—	100 % des clients classés sains le sont réellement
Rappel	Sains	1,000 (100 %)	—	100 % des clients sains sont correctement identifiés
F1-Score	Sains	1,000 (100 %)	—	Équilibre parfait précision/rappel sur les sains

Section 2 : Discussion des résultats

Cette section compare nos résultats aux travaux recensés dans la revue de littérature du Chapitre 1. Elle explique pour chaque point clé pourquoi nos résultats ressemblent ou diffèrent des études existantes, et ce que cela nous apprend sur le risque de crédit dans le contexte algérien.

1. Des performances qui dépassent tous les benchmarks de la littérature — pourquoi ?

Notre modèle atteint un AUC de 1,000 et une accuracy de 100 %. Ces résultats dépassent très largement tout ce que la littérature a documenté sur la régression logistique appliquée au risque de crédit.

Dans le contexte algérien, les deux études les plus récentes donnent des résultats bien inférieurs : Tari et al. (2024) obtiennent 59 % d'AUC avec la régression logistique sur 647 dossiers de crédit, et Souadda et al. (2022) rapportent 62,7 % sur 282 dossiers du CPA. À l'international, Lessmann et al. (2015) situent la régression logistique entre 70 % et 85 % d'AUC selon les bases testées. Notre résultat de 1,000 dépasse donc tous ces chiffres.

La clé de cette différence est simple : la qualité des données. Les études algériennes citées utilisent des bilans comptables qui peuvent être incomplets, peu fiables ou partiellement falsifiés, en raison de l'économie informelle et des pratiques de manipulation des comptes, comme le documentent précisément Gliz et Touati-Tliba (2012). Dans ce contexte, même un endettement très élevé déclaré peut ne pas refléter la situation réelle de l'entreprise. Notre base, issue du système Moody's de TRUST Bank, élimine ce problème : les ratios sont calculés de façon standardisée, vérifiée et normalisée par un système professionnel reconnu. Le résultat est des corrélations entre les ratios et la variable cible qui atteignent 0,967 et 0,968 — des niveaux qui rendent la tâche du modèle presque triviale. Ce constat valide directement la conclusion de Moscatelli et al. (2020) : la qualité des données est le premier déterminant de la performance d'un modèle, avant même le choix de l'algorithme.

2. La régression logistique face aux méthodes avancées — le bon algorithme dans les bonnes conditions

La littérature académique positionne en général la régression logistique comme moins performante que les algorithmes avancés. Tari et al. (2024) montrent que la forêt aléatoire atteint 91 % d'AUC contre 59 % pour la RL sur des données algériennes. Baesens et al. (2003) et Lessmann et al. (2015) confirment cet écart à l'international. La conclusion habituelle est donc que les méthodes complexes dominent.

Nos résultats remettent ce débat en perspective sur un point essentiel. La question n'est pas seulement « quel algorithme est le meilleur ? » mais surtout « dans quelles conditions chaque algorithme peut-il donner le meilleur de lui-même ? ». Nos résultats montrent que quand les données sont de bonne qualité complètes, standardisées et fortement liées à la variable à prédire la régression logistique peut atteindre des performances aussi bonnes que les méthodes les plus complexes. Dans ces conditions, utiliser un algorithme plus complexe n'apporterait aucun gain de performance, mais ferait perdre un avantage décisif : l'interprétabilité des résultats.

Comme le rappellent Tari et al. (2024), la régression logistique est la seule méthode dont les coefficients sont directement lisibles par un analyste financier, sans formation en intelligence artificielle. Dans un contexte bancaire réglementé où chaque décision de refus de crédit doit pouvoir être justifiée, cette transparence n'est pas un luxe c'est une nécessité. Ce résultat rejoint Thomas (2000), qui montrait déjà que les différences de performance entre méthodes dépendent fortement de la qualité des données disponibles, et que l'avantage des méthodes complexes n'est pas universel.

3. Discussion de H1 : Les ratios d'endettement confirment ce que la littérature algérienne avait identifié

H1 est confirmée avec une solidité remarquable. Les corrélations de $-0,967$ et $-0,968$ de l'endettement global et tangible avec la variable cible sont parmi les plus élevées que l'on puisse observer dans une analyse financière. Les coefficients du modèle ($-0,854$ et $-0,855$) en font les deux variables les plus pénalisantes du modèle.

Ce résultat confirme exactement les conclusions de Gliz et Touati-Tliba (2012), qui identifient l'endettement comme le déterminant le plus puissant du risque de défaut dans le contexte des

PME algériennes. Il rejoint aussi le modèle Z-score d'Altman (1968), qui intégrait déjà des ratios d'endettement comme variables centrales. Le fait que ce signal reste aussi fort sur des PME algériennes sur crédit d'exploitation, plus de cinquante ans après les travaux d'Altman, témoigne de sa robustesse universelle en finance d'entreprise.

La différence avec les études algériennes précédentes qui obtiennent des AUC plus faibles s'explique maintenant clairement. Avec des bilans comptables peu fiables, un endettement élevé déclaré peut masquer la réalité (sous-déclaration des dettes, surévaluation des actifs). Avec les données Moody's, ce problème est éliminé, et le ratio d'endettement retrouve toute sa force discriminante naturelle.

4. Discussion de H2 : Les délais du cycle d'exploitation, des signaux précoces du défaut

H2 est confirmée dans ses trois composantes, avec un résultat original sur le délai fournisseurs. Les délais clients (-0,920) et stocks (-0,836) confirment les conclusions d'Ampountolas et al. (2021) et Souadda et al. (2022), qui soulignent l'importance de ces variables pour prédire les défauts sur crédits à court terme des PME. He et Xu (2019) montrent également que les variables liées à la gestion quotidienne de l'entreprise améliorent significativement le pouvoir prédictif des modèles.

Le résultat sur le délai fournisseurs est le plus original de notre étude. Les défaillants paient leurs fournisseurs quasi comptant (5 jours) alors que les sains bénéficient de délais normaux (35 jours). Ce comportement, documenté par Petersen et Rajan (1997), signifie que les partenaires commerciaux de l'entreprise en difficulté lui retirent le crédit commercial avant même que la banque ne soit informée du problème. Le délai fournisseurs très court est donc un signal précoce de détresse, visible dans les données financières avant que le défaut ne survienne réellement. Aucune des études algériennes disponibles n'avait mis en évidence ce mécanisme avec cette clarté.

5. Discussion de H3 : La qualité des données comme facteur décisif

H3 est confirmée très largement avec un AUC de 1,000. Ce résultat n'invalide pas les conclusions des études algériennes précédentes — il les complète. Tari et al. (2024) et Souadda et al. (2022) ne travaillaient pas dans les mêmes conditions de qualité de données. Leurs AUC plus faibles (59 % et 62,7 %) ne signifient pas que la régression logistique est un mauvais

algorithme pour le risque de crédit algérien. Ils signifient que les données disponibles dans leurs études ne permettaient pas à l'algorithme d'exprimer son plein potentiel.

Notre étude apporte donc un enseignement important pour le système bancaire algérien : avant d'investir dans des algorithmes plus complexes et plus coûteux à déployer, les banques algériennes devraient d'abord investir dans la qualité et la standardisation des données collectées sur leurs emprunteurs. Un système de notation professionnel comme Moody's, appliqué rigoureusement, peut faire plus pour la performance des modèles que le passage à des algorithmes avancés.

6. Limites de l'étude et perspectives

L'étude présente certaines limites méthodologiques. Tout d'abord, l'échantillon reste relativement restreint avec 255 observations. Ensuite, les données proviennent d'une seule banque, ce qui peut limiter la généralisation des résultats à l'ensemble du secteur bancaire algérien. Enfin, seul l'algorithme de régression logistique a été testé dans le cadre expérimental final.

Conclusion du Chapitre III

Ce chapitre a présenté et analysé en profondeur l'ensemble des résultats de l'application de la régression logistique à la prédiction du risque de défaut des PME à TRUST Bank Algérie.

La Section I a construit un raisonnement progressif où chaque graphique justifie et enrichit le suivant. La composition de l'échantillon explique les distributions observées, qui elles-mêmes préparent la comparaison centrale entre les deux groupes. Les prétraitements découlent directement des observations faites sur les données brutes. Les coefficients du modèle confirment exactement les séparations visuelles observées. La courbe ROC et la matrice de confusion traduisent cette séparation en performance opérationnelle parfaite, que les métriques globales synthétisent.

La Section II a montré que ces résultats, supérieurs à tous les benchmarks de la littérature, s'expliquent par la qualité des données Moody's un avantage que les études algériennes précédentes n'avaient pas. L'obstacle principal à la performance des modèles de risque de crédit en Algérie n'est pas le choix de l'algorithme, mais la fiabilité des données disponibles.

Les trois hypothèses sont toutes confirmées avec une solidité remarquable : les ratios d'endettement séparent parfaitement les deux classes (H1), les délais du cycle d'exploitation sont de puissants indicateurs prédictifs avec un signal original sur le délai fournisseurs (H2), et la régression logistique atteint un AUC de 1,000, très largement au-dessus du seuil cible de 0,70 (H3). Ces résultats démontrent que la régression logistique, appliquée à des données de qualité, est un outil d'aide à la décision fiable, transparent et parfaitement adapté aux besoins opérationnels des banques algériennes.

CONCLUSION GÉNÉRALE

Ce mémoire avait pour objectif d'évaluer la capacité de la régression logistique à prédire le risque de défaut des PME sur les crédits d'exploitation dans le contexte bancaire algérien, à partir des données issues du système de notation Moody's de TRUST Bank Algérie sur la période 2020–2024.

À travers une démarche quantitative et hypothético-déductive fondée sur l'analyse de 255 dossiers de crédit, notre étude a montré que les ratios financiers liés à l'endettement et au cycle d'exploitation constituent des indicateurs particulièrement pertinents pour distinguer les entreprises saines des entreprises défaillantes.

Les résultats obtenus confirment l'ensemble des hypothèses formulées au début de la recherche. Le modèle de régression logistique a atteint des performances très élevées, avec un ROC-AUC de 1,000 et une classification parfaite sur le jeu de test. Ces résultats montrent que la régression logistique peut constituer un outil fiable, transparent et interprétable d'aide à la décision bancaire lorsqu'elle est appliquée à des données financières structurées et standardisées.

Notre travail met également en évidence un enseignement essentiel : dans le contexte algérien, la performance des modèles de prédiction dépend avant tout de la qualité des données disponibles plus que de la complexité des algorithmes utilisés. Ainsi, l'amélioration des systèmes de collecte et de standardisation des données représente un enjeu prioritaire pour les banques algériennes souhaitant moderniser leurs dispositifs d'évaluation du risque de crédit.

Toutefois, cette recherche présente certaines limites, notamment la taille relativement réduite de l'échantillon et l'utilisation des données d'une seule banque. De futures recherches pourraient élargir l'étude à plusieurs établissements bancaires et comparer la régression logistique à d'autres algorithmes de Machine Learning plus avancés.

En définitive, ce travail montre qu'une régression logistique correctement construite et alimentée par des données de qualité peut représenter une solution performante et opérationnellement adaptée aux besoins du système bancaire algérien en matière de gestion du risque de crédit.

Références

- Addo, G. e. (2018). Credit risk analysis using machine and deep learning models. *maison des sciences économiques*.
- Ampountolas, A. N. (2021). A machine learning approach for micro-credit scoring. *Risks*.
- Ampountolas, A., Nyarko Nde, T., Date, P., & Constantinescu, C. (2021). A Machine Learning Approach for Micro-Credit Scoring. *Risks*.
- Amzile, K. (2023). *Intelligence artificielle et modélisation du risque de crédit*. .
- Babo, A. B. (2022). *L'efficacité de la réglementation prudentielle à la lumière des accords de Bâle : Cas des banques de la zone UEMOA*. Université Gaston Berger de Saint-Louis du Sénégal.
- Baesens, B. V. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*.
- Bagehot, W. (1873). *Lombard Street: A description of the money market*.
- Banque, R. (2014). Méthodes 5C, LAPP et grilles de diagnostic.
- Benhalima, A. (1997). *Pratique des techniques bancaires*. Éditions Dahlab.
- Berger, A. N. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking & Finance*.
- Bessis, J. (2015). *Gestion des risques et gestion actif-passif des banques*. Dalloz.
- Bezbakh, P. &. (2000). Dictionnaire de l'économie.
- Bhatore, S. M. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. *Journal of Banking and Financial Technology*.
- Bouchetara, M. (2017). *Stabilité financière et crises bancaires : stress tests bancaires en Algérie*. Université de Tlemcen.
- Boudinot, A. &.-C. (1972). *Technique et pratique bancaire (2e éd.)*. Sirey.
- Bouyacoub, F. (2000). *L'entreprise et le financement bancaire*. Éditions Casbah.

- Breiman, L. e. (2004). Random Forests.
- Bruyère, R. (2003). *Les produits dérivés de crédit*. Economica.
- CLAUSS, P. &. (2021). *Économie bancaire : Analyse financière, régulation et gestion des risques*. Ellipses.
- commerce, C. d. (s.d.). *Ordonnance n° 75-59 du 26 septembre 1975*.
- Crédits, B. d. (s.d.). *Les banques centrales*. Récupéré sur <https://www.boursedescredits.com/>
- d'Algérie, B. (2014,). Le règlement 14-03 du 16 février. *Articles 4,5 ,9,10*.
- De Coussergues, S. &. (2013). *Gestion de la banque : du diagnostic à la stratégie (7e éd.)*. Dunod.
- De Coussergues, S. B. (2020). *Gestion de la banque : Tous les principes et outils à connaître (9e éd.)* . Dunod.
- El Naqa, I. L. (2015). *Machine Learning in Radiation Oncology: Overview and Clinical Applications*. Springer.
- Ellah, D. B. (2025). cours Analyse Financière Approfondie .
- Garsault, P. &. (1997). *La banque : fonctionnement et stratégies*. Economica.
- Glassner, A. (2021). *Deep Learning: A Visual Approach*.
- Gliz, A. T.-T. (2012). Estimation du risque de crédit et qualité de l'information comptable en Algérie. *Les Cahiers du CREAD*.
- Goodfellow, I. B. (2016). *Deep Learning*. MIT Press.
- Gourlaouen, J. (2012). *Banque et marchés financiers*. Dunod.
- Grandguillot, B. &. (2023). *Analyse financière*. Gualino.
- GUERONNIERE, H. D. (2015). *FAIRE DE LA BANQUE SON METIER*. FRANCE:Studyrama.
- Guigal, M. e. (2011). *Le guide de la banque : comptes, carte bancaire, services*. paris:Fine Media. .

- Guo, L. L. (2024). Bank Credit Risk Early Warning Model Based on Machine Learning Decision Trees. *Southern United Academy of Sciences*.
- Haddad, S. &. (2015). *Comprendre la banque : organisation et fonctionnement*.
- Hadj Sadok, T. (2007). *Les risques de l'entreprise et de la banque*. . Alger : Éditions Dahlab.
- Hastie, T. T. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- He, H. &. (2019). Comparison of machine learning methods in credit risk assessment: Evidence from the banking sector in China. *Journal of Financial Services Research*.
- Ilunga, E. M. (2016). *Risques bancaires et dispositifs prudentiels de gestion en RDC (39e éd.)*. L'Harmattan .
- Karyotis, C. (2015). *L'essentiel de la banque (2e éd.)*. Gualino éditeur, Lextenso éditions.
- Kharoubi, C. &. (2013). *Analyse du risque de crédit : Banque et marché*. Revue Banque Éditions.
- Kharoubi, C. &. (2016). *Analyse du risque de crédit : Banque et marché*. Revue Banque Éditions.
- Landman, J. (2021). *Quantum Algorithms for Unsupervised Machine Learning and Neural Networks*. Université Paris Cité.
- Larousse, É. (2020). Dictionnaire Larousse.
- Le règlement 14-01 . (du 16 février 2014).
- Le règlement 14-02 relatif aux grands risques et aux participations . (s.d.).
- Lessmann, S. B.-V. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*.
- Loi n° 88-01 du 12 janvier 1988. (s.d.).
- Loi n°90-10 relative à la monnaie et au crédit. (1990).
- Luc, B.-R. (2008). *Principes de technique bancaire (25e éd.)*. Dunod.
- Matoussi, H. &. (2010). La prévision du risque de défaut dans les banques tunisiennes : analyse comparative entre les méthodes linéaires classiques et les méthodes de l'intelligence artificielle. *In Crises et nouvelles problématiques de la valeur*.

- Maymo, É. L. (2015). *Économie et gestion de la banque*. Dunod.
- Mishkin, F. S. (2019). *The economics of money, banking and financial markets*.
- Moscatelli, M. R. (2020). Corporate default forecasting with machine . *Expert Systems with Applications*.
- Patat, J.-P. (1993). *Monnaie, institutions financières et politique monétaire*. Economica.
- Petit-Dutaillis, G. (1999). *Le risque du crédit bancaire*. Dunod.
- Saint-Cirgue, G. (2019). *Apprendre le machine learning en une semaine*.
- Souadda, L. I., Menari, M., & Berghout, Y. M. (2022). Le scoring des crédits d'exploitation des PME par le Machine Learning. *Revue des sciences commerciales et de gestion*.
- Soukkou, A. (s.d.). *Techniques de l'intelligence artificielle : Concepts et algorithmes*. Université Mohamed Seddik Benyahia – Jijel.
- Tarazi, A. &. (2005). Les accords de Bâle II : fondements et implications pour les banques. *Revue Banque*.
- Tari, M. L., Bouguerra, A., & Selatni, I. (2024). Optimisation du Scoring pour la Gestion du Risque de Crédit des PME : l'Apport du Machine Learning. *Ecole Supérieure de Commerce d'Alger*.
- Thomas, L. C. (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*.
- Touati-Tliba, M. (2003). Modèles d'aide à la décision d'octroi de crédits et performance du portefeuille crédit des banques publiques algériennes. *Revue des sciences commerciales et de gestion* .
- Tripathi, P. K. (2025). – Machine Learning pour le risque de crédit dans les banques indiennes .
- Vernimmen, P. Q. (2022). *Finance d'entreprise (19e éd.)*. Dalloz.
- Vigouroux, J.-C. B. (1991). *Découvrir la banque : par une approche économique et comptable*.

Zhou, Z. &. (2020). Deep learning for credit scoring: An empirical study on unstructured data.
Expert Systems with Applications.