



Mémoire de fin
d'étude

Thème

**Machine learning pour la mesure et la gestion
du risque de crédit**

Présenté et soutenu par

BOUHOUHOUH Heyam

Encadré par

Pr. CHICHTI Jameleddine

Etudiant (e) parrainé (e) par

Ministère des finances Algérie

Dédicaces

Je dédie ce travail à :

Mes chers parents qui ont tout fait pour me voir réussir

À toute ma famille

À tous mes amis

Heyam...

Remerciements

Au terme de ce travail, je tiens à exprimer mes remerciements les plus sincères, tout d'abord au bon dieu pour la patience et le courage qu'il m'a offert tout au long de mes études.

Mes sincères remerciements s'adressent en premier lieu à Pr. CHICHTI jameleddine mon encadrant, pour sa disponibilité, ses précieux conseils, pour son implication, son aide, ainsi que son soutien lors de la rédaction de ce travail de fin d'études.

Je tiens également à remercier tout le personnel de la direction des crédits aux industries et aux services du Crédit Populaire d'Algérie.

Je tiens à adresser mes vifs remerciements à tous les professeurs et les administrateurs de l'Institut de Financement du développement du Maghreb Arabe.

Enfin, Je tiens aussi à remercier les membres du Jury pour l'honneur qu'ils me font en acceptant d'évaluer ce travail.

Résumé

Les crises financières qu'a connu le monde, de plus de l'effondrement successif du secteur bancaire, et qui est la conséquence directe d'une mauvaise quantification et gestion des risques en particulier le risque de crédit qui constitue le risque le plus important pour une banque ou institution financière. Cette situation a renforcé l'idée que les banques sont dans l'obligation de les gérer, les risques bancaires sont devenus donc un sujet central des accords de Bâle, et des instances réglementaires internationales.

Pour ce faire, les établissements bancaires s'efforcent pour mettre une mesure et gestion appropriée, à travers l'implantation des méthodes sophistiquées, qui se basent sur différents critères et techniques. Parmi ces méthodes on trouve le crédit scoring, qui constitue la méthode la plus utilisée par les banques en matière de gestion du risque. Et avec le progrès de l'intelligence artificielle il découle d'une nouvelle vague de modèles de crédit scoring axés sur l'apprentissage automatique (ML).

Dans notre travail, nous avons été intéressés par l'implantation d'une application web pour la classification des clients potentiels, commençant par la construction des différents modèles de l'apprentissage automatique supervisé, puis nous avons utilisés le modèle le plus performant dans l'implantation de notre application.

Mots clés : risque de crédit, crédit scoring, machine learning, apprentissage supervisé, les modèles de classification.

Abstract

The financial crises that the world has experienced, in addition to the successive collapse of the banking sector, and which is the direct consequence of a poor quantification and management of risks especially the credit risk which constitutes the most important risk for a bank or financial institutions. This situation has reinforced the idea that banks are obliged to manage them, so banking risks have become a central issue of the Basel accords, and of international regulatory bodies.

For this reason, banking institutions strive to put an appropriate measurement and management, through the implementation of sophisticated methods, which are based on different criteria and technique. Among these methods is credit scoring, which is the most used method by banks in terms of risk management. And with the advancement of artificial intelligence, it stems from a new wave of credit scoring models focused on machine learning (ML).

In our work, we were interested in implementing a web application for the classification of potential customers, starting with the construction of the different models of supervised machine learning, then we used the most efficient model in the implementation of our application.

Key words: credit risk, credit scoring, machine learning, supervised learning, classification models.

Liste des abréviations :

ACT : Actif à Court Terme

AI : Artificial Intelligence

AUC : Area Under the Curve

CA : Chiffre d’Affaires

CAF : Capacité d’AutoFinancement

CMC : Consiel de la Monnaie et du Crédit

CPA : Crédit Populaire d’Algérie

DCT : Dette à Court Terme

EAD: Exposure At Default

EBE: Excédant Brute d’Exploitation

EL: Expected Loss

FMI : Fonds Monétaire International

FPR : False Positive Rate

IRB: Internal Ratings Based

KNN: k-nearest neighbor

LCR: Liquidity Coverage Ration

LGD: Loss Given Default

LMC : Lois sur la Monnaie et le Crédit

M-KMV : Moody’s – Kealhofer, Mcquown et Vasicek

ML: Machine Learning

NSFR: Net Stable Funding Ration

PD : Probabilité de Défaut

PME : Petit et Moyenne Entreprise

SVM: Support Vector Machine

TPR: True Positive Rate

UL: Unexpected Loss

VAR: Value At Risk

Liste des figures :

Figure 1: échelonnement de la notation	11
Figure 2: la qualité de crédit solen la notation	12
Figure 3: L'architecture d'un modèle de risque de crédit	20
Figure 4: le mécanisme de CDSs	47
Figure 5: le processus de scoring	53
Figure 6: le processus de l'apprentissage supervisé.....	57
Figure 7 : le processus de l'apprentissage non supervisé	58
Figure 8: l'apprentissage par renforcement	59
Figure 9: l'arbre de décision	63
Figure 10: la fonction sigmoïde.....	65
Figure 11 : les machines à vecteur support	66
Figure 12 : le modèle K-plus proche voisins.....	67
Figure 13: le modèle de la forêt aléatoire.....	68
Figure 14: répartition de l'échantillon	78
Figure 15: les valeurs aberrantes pour les ratios de trésorerie.....	79
Figure 16: les valeurs aberrantes pour les ratios de liquidité	80
Figure 17: les valeurs aberrantes pour les ratios de rentabilité	80
Figure 18: les valeurs aberrantes pour les ratios de structure financière.....	81
Figure 19: présentation des valeurs manquantes	81
Figure 20: rapports de classification avant et après l'optimisation de l'arbre de décision	87
Figure 21: les matrices de confusion de l'arbre de décision avant et arés l'optimisation.....	88
Figure 22: les courbes d'apprentissage avant et après l'optimisation de l'arbre de décision	89
Figure 23: les variables importantes pour l'arbre de décision	90
Figure 24: les rapports de classification avant et après l'optimisation de la régression logistique.....	91
Figure 25: les courbes d'apprentissage avant et après l'optimisation de la régression logistique.....	92
Figure 26: les matrices de confusion avant et après l'optimisation de la régression logistique	92
Figure 27: les rapports de classification après et avant l'optimisation du modèle SVM	95
Figure 28: les courbes d'apprentissage avant et après l'optimisation du modèle SVM	95
Figure 29: les matrice de confusion après et avant l'optimisation du modèle SVM	96
Figure 30: les rapports de classification avant et après l'optimisation du modèle KNN	97
Figure 31: les courbes d'apprentissage avant et après l'optimisation du modèle KNN	98
Figure 32: les matrices de confusion avant et après l'optimisation du modèle KNN	98
Figure 33: les rapports de classification avant et après l'optimisation du modèle Adaboost	99
Figure 34: les courbes d'apprentissage avant et après l'optimisation du modèle Adaboost	100
Figure 35: les matrices de confusion avant et après l'optimisation du modèle Adaboost	101
Figure 36: les variables importantes pour le modèle Adaboost	102
Figure 37: les rapports de classification avanies et après l'optimisation du modèle forêt aléatoire	103
Figure 38: les courbes d'apprentissage avant et après l'optimisation du modèle forêt aléatoire	104
Figure 39: les matrices de confusion avant et après l'optimisation du modèle forêt aléatoire	104
Figure 40: les variables importantes pour le modèle forêt aléatoire.....	105
Figure 41: Application Web.....	107

Liste des tableaux :

Tableau 1: listes des variables quantitatives et qualitatives	76
Tableau 2: le nombre des valeurs manquantes dans chaque variable	82
Tableau 3: liste des variables qualitatives après encodage	83
Tableau 4: le fractionnement de l'échantillon	84
Tableau 5: les hyperparamètres de l'arbre de décision.....	87
Tableau 6: les hyperparamètres de la régression logistique.....	90
Tableau 7: les coefficients des variables de la régression logistique	93
Tableau 8: les hyperparamètres avant et après l'optimisation	94
Tableau 9: les hyperparamètres du modèle KNN	97
Tableau 10: les hyperparamètres du modèle Adaboost.....	99
Tableau 11: les hyperparamètres du modèle forêt aléatoire.....	102
Tableau 12: comparaison des modèles.....	106

Liste des annexes :

Annexe 1: Matrice de corrélation	119
Annexe 2: Arbre de décision	120
Annexe 3: courbe ROC des modèles	121
Annexe 4: l'aire sous la courbe pour chaque modèle.....	122

Sommaire :

<i>Dédicaces</i>	i
<i>Remerciements</i>	ii
Résumé	iii
Liste des abréviations :	v
Liste des figures :	vi
Liste des tableaux :	vii
Liste des annexes :	vii
Introduction générale	A
<i>Partie théorique</i>	1
Chapitre I:Mesure du risque de crédit	2
Introduction	1
Section 1 : cadres conceptuels	2
Section 2 : La mesure du risque de crédit	10
Conclusion.....	21
Chapitre II:Gestion du risque de crédit	22
Introduction	23
Section1 : la gestion prudentielle.....	24
Section 2 : les techniques de gestion de risque de crédit	41
Conclusion	49
<i>Partie pratique</i>	50
Chapitre I:Méthodologie de construction d'un modèle de classification	51
Introduction	52
Section 1 : revue de littérature	53
Section 2 : méthodologie de traitement des données	59
Conclusion.....	72
Chapitre II: L'implantation des modèles de classification basés sur l'apprentissage supervisé	73
Introduction	74
Section 1 : préparation des données.....	75
Section 2 : La construction des modèles	86

Conclusion.....	108
Conclusion générale	109
Bibliographie :	112
Table des matières :.....	116
Annexes :	119

Introduction générale

Data science constitue un ensemble de disciplines complémentaires qui permettent de produire la connaissance en faisant parler les données. Plusieurs techniques se trouvent à l'intérieure de celle-ci, qui permettent l'analyse et l'exploitation des données, commençant par une simple visualisation graphique, allant jusqu'à l'intelligence artificielle.

L'intelligence artificielle a révolutionné de nombreux aspects de notre vie de tous les jours, composée de données, d'algorithmes, de matériel, et de connectivité, L'AI permet aux machines d'imiter certains aspects de l'intelligence humaines, comme la perception, la résolution de problèmes, l'interaction linguistique, ou même la créativité, ces technologies facilitent la vie par la mise en œuvre des programmes de développement durable.

A ce stade, on trouve le machine Learning ou l'apprentissage automatique, qui présente une discipline de l'AI, plus spécifique sur les algorithmes, et qui devient prépondérant dans bon nombre de domaines. La finance n'échappe pas de cette révolution.

La banque en tant qu'acteur de la finance, son activité est caractérisée par sa diversité, et avec un environnement très instable et très vulnérable face aux fluctuations de la sphère financière, le métier bancaire est inséparable du risque. Plusieurs types de risques peuvent affecter la survie de la banque. Le risque de crédit constitue un puissant facteur.

Historiquement, le risque de crédit est réputé le plus important pour les banques, et ceci est le résultat d'une mauvaise quantification et gestion de ce dernier, Cette situation a renforcé l'idée que les banques sont dans l'obligation de le gérer, le risque de crédit est devenu donc un sujet central des accords de Bâle, et des instances réglementaires internationales.

Pour ce faire, les établissements bancaires s'efforcent pour mettre une mesure et gestion approprié, à travers l'implantation des méthodes, qui se basent sur différents critères et techniques qui combinent le calcul et l'intuition. A ce titre plusieurs méthodes ont été proposés, à savoir l'analyse financière, qui est basée sur un diagnostic financier des informations comptables. Au fil du temps, cette méthode classique a connu des insuffisances.

En vue de combler ces lacunes, d'autres méthodes d'évaluation ont été proposées, qui dépendent des modèles statistiques sophistiqués. Parmi ces modèles on trouve le crédit scoring, qui constitue la méthode la plus utilisée par les banques en matière de gestion des risques.

L'idée est d'identifier certains facteurs clés qui déterminent la probabilité de défaut (par opposition au remboursement) et de les combiner ou de les pondérer dans un score quantitatif. Dans certains cas, le score peut être littéralement interprété comme une probabilité de défaut, dans d'autres, le score peut être utilisé comme système de classification. C'est-à-dire qu'il place un emprunteur potentiel dans un bon ou un mauvais groupe, en fonction d'un score et de certaines caractéristiques.

Plusieurs techniques sont utilisées pour la construction des modèles de scoring, surtout avec le progrès récent en matière d'intelligence artificielle est de plus en plus représenter et influent dans la finance, il découle d'une nouvelle vague de modèles de crédit scoring axés sur l'apprentissage automatique (ML) qui ont suscité une attention considérable.

Dans ce contexte, les banques algériennes utilisent des méthodes classiques pour l'évaluation du risque de crédit, donc l'adoption d'un modèle d'évaluation plus développé est une nécessité, où les algorithmes de machine learning peuvent être utilisés pour une meilleure performance des modèles.

Cela nous amène à demander : **Comment se fait l'évaluation du risque de crédit selon les algorithmes de machines learning ?**

De cette problématique déroule un certain nombre de sous questions qui doivent être posées :

- Comment le risque de crédit est-il perçu à l'égard de la réglementation prudentielle Internationale et algérienne ?
- Quelle est la méthodologie de construction d'un modèle du crédit scoring à l'égard des algorithmes de machines learning ?

Pour répondre à la problématique ci-dessus deux hypothèses guideront notre recherche et nous tenterons de les vérifier tout au long de ce travail à savoir :

- Les réglementations prudentielles internationales et nationales prennent au sérieux le risque du crédit et les pertes qu'il engendre. De plus, elles mettent l'accent sur la gestion de ces dernières.
- L'élaboration d'un modèle de scroing en utilisant les algorithmes de machine learning consiste à définir un critère de défaut ainsi que les variables explicatives et la modélisation se fait grâce à des modèles de classification.

La démarche méthodologique que nous avons suivie afin de répondre aux questions posées et de vérifier les hypothèses est une démarche analytique. Dans un premier temps, nous allons essayer de cerner les éléments théoriques relatifs à notre thème de recherche. Ensuite, nous recueillerons des données auprès de la direction des crédits aux industries et aux services de crédit populaire d'Algérie, pour répondre à notre problématique.

Ainsi, notre travail est structuré en deux parties, une première partie donnera dans un Premier chapitre une brève présentation du risque de crédit et les différents outils de mesure de ceci, un deuxième chapitre exposera les techniques de gestion du risque de crédit. La deuxième partie est consacré à notre étude de cas, elle sera également scindée en deux chapitres dans un premier chapitre, nous exposerons les différentes notions de la modélisation du crédit scoring à l'aide des algorithmes de machine learning et de même exposera notre méthodologie d'analyse, dans le deuxième chapitre nous présenterons les différents résultats ainsi que leurs interprétations chose qui nous permet de nous répondre à notre problématique.

L'objectif que nous poursuivons est d'analyser et de comprendre comment se fait la classification des clients selon l'apprentissage automatique, et de voir quelle est le modèle le plus performant qui prédit au mieux le défaut, ce modèle va être utilisé par la suite pour l'élaboration d'une application web qui va faciliter le classement des demandeurs des crédits aux niveau du CPA.

Cette étude comprenait le portefeuille des PME comme clientèle, vu leur importance dans l'économie algérienne, et leurs fortes concentrations.

Partie
théorique

Chapitre I

Mesure du risque de crédit

Introduction

L'activité principale des établissements de crédit est l'octroi de crédits. Les principales préoccupations du secteur bancaire sont les risques qui en résultent, en particulier le risque de crédit. Le travail des banquiers consiste à rechercher des opportunités passionnantes. Parallèlement, de nombreux universitaires et penseurs s'intéressent à la compréhension de la répartition du risque de crédit et à l'explication des facteurs qui influent sur la détermination de ce dernier et de ses composantes. En raison de cet intérêt, des théories et des modèles qui traitent des questions du risque de crédit ont été développés. Ces modèles ont évolué au fil du temps, et au fur et à mesure que le monde s'est développé, de nombreuses nouvelles versions sont également apparues. Ce chapitre est segmenté en deux sections, dans une première section nous avons présentés le cadre conceptuel du risque de crédit, et dans la deuxième section nous avons intéressé à la mesure du risque de crédit.

Section 1 : cadres conceptuels

Premièrement, il est important de définir l'aspect théorique du risque de crédit, de distinguer ces différentes notions de base ainsi que ses composantes, et les différents facteurs liés à ce-ci.

1. Définition du risque de crédit :

Toute opération de crédit entraîne l'anticipation de revenus futurs, ce qui comporte le risque que cette anticipation ne se matérialise pas ou que, en cas de défaillance de l'emprunteur, le remboursement à l'échéance ne soit que partiellement effectué. De même, chaque fois qu'un titre est acheté, le créancier financier assume le risque que l'émetteur du titre ne soit pas en mesure de payer le revenu qui lui est associé ou de le retourner à son échéance.

Selon C. KHAROUBI et P. THOMAS « Le risque de crédit est le risque particulier naissant d'une opération de prêt. Il correspond à la probabilité qu'un événement négatif affecte le service de la dette sur lequel le débiteur s'est engagé »¹.

C.Gouriéroux et A.Tiomo définissent le risque de crédit comme suit : « Le risque de crédit est le risque de perte inhérent au défaut d'un emprunteur par rapport au remboursement de ses dettes (obligations, prêts bancaires, créances commerciales) »².

Ainsi, lorsqu'il se produit dans le contexte d'une transaction de crédit, nous l'appelons risque de crédit. C'est le risque auquel un créancier est exposé si le débiteur n'est pas en mesure de tenir ses promesses, c'est-à-dire de rembourser son prêt ou de payer les intérêts de cette dette³.

Dans ce cas, le risque de crédit est lié à deux situations : soit le débiteur est temporairement frappé d'incapacité en raison de difficultés occasionnelles, soit le débiteur est frappé d'incapacité permanente en raison de problèmes structurels pouvant entraîner la disparition de l'emprunteur.

Le risque de crédit est composé des risques de défaut de crédit, des garants et des contreparties aux produits dérivés. Tous les secteurs du marché financier sont sensibles à ce risque, mais les

¹ C. KHAROUBI et P. THOMAS, **Analyse du risque de crédit**, 2^{ème} édition, revue banque édition, Paris, 2016, p15

² C.Gouriéroux et A.Tiomo, **risque de crédit une approche avancés**, les cahiers de CREF, avril 2007, p 11.

³ A.BEITONE, **Dictionnaire des sciences économiques**, 6^{ème} édition, dunod,,2019,p 553.

banques sont particulièrement exposées en raison du crédit et des activités hors bilan telles que les garanties¹.

La principale source du risque de crédit est les prêts, mais d'autres catégories d'actifs comme les obligations, les titres de créance à court terme et les produits dérivés peuvent également y être sujettes (ainsi que les obligations hors bilan comme les lignes ou limites de crédit non utilisées, les garanties et les crédits documentaires).

Le risque de crédit a donc de nombreuses appellations, y compris le risque de contrepartie dans les transactions sur les marchés financiers et le risque de défaut au sens propres dans les transactions sur le marché du crédit.

2. La décomposition du risque de crédit :

Pour plusieurs auteurs, ce risque est décomposé en :

2.1. Le risque de défaut :

Un « accident de crédit » se produit lorsqu'un emprunteur manque une échéance, et cette situation est connue sous le nom de défaut².

Dans l'accord de Bâle 2, un défaut intervient lorsque l'un des évènements suivants est constaté:

- Le débiteur n'est pas en mesure de payer ;
- Le report de paiement (non-paiement) est lié à une situation telle qu'une annulation de dette, une provision particulière ou une restructuration pendant une période difficile ;
- Il y a eu un retard de paiement de plus de 90 jours ;
- L'emprunteur est en faillite légale.

2.2. Le risque de dégradation de la qualité de crédit :

Si la dette est vendue de manière anticipée sans défaillance, les pertes associées à la dégradation de qualité de la contrepartie sont constatées. En effet, le risque de baisse se traduit par une dévaluation potentielle de la dette pendant sa durée de vie. Il convient de souligner que la probabilité de défaut et la dégradation de la qualité du crédit sont fortement associées dans la mesure où la détérioration de la qualité de crédit d'une contrepartie peut signaler un défaut

¹ E. SPUCHLAKOVA, **The Credit Risk and its Measurement, Hedging and Monitoring**, ICOAE, N°2, juillet, 2015, p 676.

² C. KHAROUBI et P. THOMAS, op-cite, p 25.

imminent et qu'un défaut entraîne inévitablement une dégradation de la qualité du crédit. Mais ce sont deux types de risques différents. La prime de risque offerte par les marchés financiers augmente parallèlement à une baisse perçue de la solvabilité de l'emprunteur, et la valeur de marché de ce prêt diminue.

2.3. Le taux de couverture en cas de défaut :

Ce risque est lié à la capacité de la banque à choisir un taux afin de compenser partiellement la perte subie du fait d'un prêt. Cependant, comme le défaut de paiement est un événement ultérieur provoqué par l'action de chaque emprunteur, il est difficile de calculer ce taux de recouvrement. Les banques sont conscientes du nombre typique de crédits qui peuvent manquer pendant une certaine période de temps. Cependant, elles sont moins conscientes du montant qu'elles peuvent finalement perdre en raison des crédits en défaut. En fin de compte, cette fluctuation des pertes par défaut peut s'avérer tout aussi préjudiciable que le changement du taux de défaut. La corrélation entre le taux de défaut et le taux de recouvrement en cas de défaut est également cruciale à comprendre.

3. Les paramètres du risque de crédit :

Le report de prêts au bilan de la banque s'accompagne d'une incertitude importante quant aux remboursements futurs, ce qui rend un portefeuille bancaire très risqué. Cette imprévisibilité provient principalement du risque inhérent à chaque client et à chaque prêt, qui ont souvent une très longue échéance, de sorte que si le statut d'un client se détériore, la banque en supporte de plein fouet les conséquences car la marge payée par le prêt fourni ne change pas. Cette asymétrie est typique du risque de crédit et peut être observée dans la loi des pertes du portefeuille de la banque. Dans le pire des cas, si l'entreprise n'est plus en mesure d'assurer le service de sa dette, la banque ne recevra qu'un pourcentage du montant prêté¹. Il est donc essentiel pour la banque de connaître :

3.1. La probabilité de défaut :

La probabilité de défaut (PD) représente la probabilité qu'une contrepartie fasse défaut dans un horizon temporel donné, souvent un an, Il s'agit d'une mesure fondamentale du risque dans l'analyse du risque de crédit.

¹ V.BRUNEL ; **gestion des risques et risque de crédit** ; 2009; P 7.

La PD d'un débiteur est déterminée à la fois par des variables de risque spécifiques à l'entreprise et macro-économiques. Selon des articles sur ce dernier facteur de risque, la PD a tendance à augmenter pendant les récessions et à baisser pendant les expansions. L'incidence des variables macroéconomiques sur la PD a donné lieu à deux estimations distinctes de la PD, la PD through the cycle (TTC) et la PD point in time (PIT). L'estimation de la PD de la TTC n'est essentiellement pas touchée par la conjoncture économique, tandis que l'estimation de la PD de PIT varie selon le cycle économique¹.

3.2. La perte en cas de défaut LGD :

La LGD calcule la différence entre le montant de l'exposition au moment de défaut et la perte subie en cas de défaillance d'un emprunteur.

En effet, la perte en cas de défaut quantifie en pourcentage combien la banque (ou le prêteur) perdra finalement sur les prêts accordés. La perte en cas de défaut est une variable importante dans le cadre prudentiel de Bâle II. En conséquence, en cas de défaillance d'une contrepartie, le recouvrement de la créance, qui prime sur la rémunération des actionnaires, est effectué sur les actifs de la société défaillante. Après déduction du montant de la garantie précédemment obtenue, le recouvrement couvrira le principal et les intérêts. Le taux de recouvrement (RR) est ainsi défini comme le pourcentage de la dette qui pourrait être remboursé en cas de défaut. Soit² :

$$LGD = 1 - RR$$

3.3. L'exposition en cas de défaut :

Cette mesure, qui a le même horizon temporel que la probabilité de défaut, représente la perte réelle si la contrepartie fait défaut sur un engagement spécifique. En réalité, il s'agit du capital encore dû à l'horizon en question plus les intérêts qui ne sont pas encore arrivés à échéance.

Ses fonctions de mesure du risque de crédit tentent de prévoir la perte projetée d'un portefeuille de crédit sur une certaine période. Il s'agit du montant moyen que la banque risque de perdre

¹ J.Hallblad ; **The Multi-year Through-the-cycle and Point-in-time Probability of Default** ; mémoire d'un master en economies; UMEA universitet; 2014; p1.

² B.Garnier ; **Mesure du risque de crédit d'un portefeuille obligataire** ; Mémoire présenté le : pour l'obtention du diplôme de Statisticien Mention Actuariat et l'admission à l'Institut des Actuariers ; Institut de statistique (ISUP) - Sorbonne Université ; 2013 ; p 32.

sur son portefeuille de créances. L'évaluation des pertes imprévues est également un objectif de la mesure du risque de crédit. A savoir :

- La perte attendue qui se calcule ainsi :

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

Un recours aux provisions est nécessaire pour couvrir ces pertes.

- La perte inattendue : est caractérisée par des pertes imprévues, sans lesquelles les pertes réelles peuvent dépasser considérablement les pertes attendues. La formule est la suivante¹ :

$$UL = VAR - EL$$

Donc il s'agit de l'écart entre la perte maximale et la perte attendue. Ces pertes extraordinaires nécessitent une certaine dotation en fonds propres : le capital économique. Après avoir modélisé l'incertitude des pertes potentielles et créé une distribution de probabilité des pertes totales au niveau du portefeuille, il est important d'évaluer les UL à un horizon temporel spécifique. Cela permet de spécifier le montant du capital économique nécessaire pour les couvrir (absorber) à un horizon particulier. Ensuite, il est encouragé pour une institution financière d'avoir suffisamment de capital pour couvrir ces UL. Il s'agit d'un quantile de la distribution de probabilité des pertes sur l'horizon temporel choisi dans ce cas, par exemple : 99% à un an².

4. Les facteurs de risque de crédit :

Un aspect externe relatif à l'insolvabilité de l'emprunteur et un aspect interne relatif à la manière dont la banque organise la fonction de distribution de crédit sont des composantes du risque de contrepartie. Les facteurs internes :

4.1. L'insolvabilité de l'emprunteur :

Le bénéficiaire d'un crédit, qu'il s'agisse d'une entreprise, d'un particulier, d'un établissement de crédit, d'une collectivité locale ou de l'État, n'est pas en mesure de rembourser les avances qui lui ont été faites, et le prêteur subit une perte totale ou partielle de dette ainsi que des revenus

¹ B.Garnier ; op-cite ; p 34.

² C. KHAROUBI et P. THOMAS, op-cite ; p 138.

y afférents en raison de l'insolvabilité de l'emprunteur. Il y a beaucoup de raisons différentes pour lesquelles les gens deviennent insolubles, mais il y en a généralement trois¹. A savoir :

➤ Le risque général :

L'insolvabilité de l'emprunteur résulte de causes extérieures provoquées par l'environnement politique ou économique dans lequel il opère. Outre les événements catastrophiques tels que les tremblements de terre ou les inondations, les crises politiques, mais surtout les crises économiques, augmentent le risque de crédit. Les crises économiques conduisent souvent à l'insolvabilité dans les économies modernes, comme l'a démontré la crise financière de 2007-2008².

➤ Le risque professionnel :

Selon l'état de l'économie, ce risque est associé à un secteur spécifique de l'économie. Les menaces qui pèsent sur les entreprises d'une industrie donnée peuvent prendre de nombreuses formes différentes, comme des changements dans le type ou l'intensité de la demande, des innovations qui modifient les processus de production, des changements de prix pour les intrants, une surproduction généralisée et une concurrence accrue qui affecte directement la capacité d'une organisation à demeurer solvable. La sidérurgie, l'immobilier et l'agriculture ne sont que quelques-unes des secteurs qui ont déjà connu des crises³.

➤ Le risque lié directement à l'emprunteur :

Le client ne peut pas tenir ses promesses pour diverses raisons. L'un des risques les plus fréquents dans l'activité de crédit est celui-ci. Sa difficulté d'identification rend sa gestion beaucoup plus difficile⁴.

➤ Le risque pays :

Le risque pays, également connu sous le nom de risque souverain, a considérablement augmenté depuis le début des années 1980 et touche désormais plus que les pays sous-développés ayant une dette extérieure importante, comme le montre la situation difficile de la Grèce. Il commence

¹ S.De GOUSSERGUES et autres, **Gestion de la banque**, 8^{ème} édition, dunod,2017, p250.

² Ibid p 250.

³ R.SUBLET ; **La gestion du risque de crédit bancaire sur les portefeuilles professionnels et particuliers** ; mémoire pour l'obtention du diplôme de master ; école de commerce de lion ; 2015 ; p18.

⁴Ibid ; p 18.

par couvrir les éléments de risque de contrepartie typiques, tels qu'une catastrophe naturelle, une crise politique ou économique et l'insolvabilité de l'emprunteur. Mais il comprend également un autre élément lié à la situation financière du pays où l'emprunteur est installé. Le bénéficiaire du crédit est solvable, mais parce que son pays est en faillite monétaire, la Banque centrale est incapable de transférer les fonds nécessaires pour assurer le service du prêt à l'étranger. Les organisations internationales, y compris le FMI, la Banque mondiale, les clubs de Londres et de Paris, ainsi que les banques touchées, sont profondément préoccupés par la question du risque pays. De nombreux rééchelonnements de la dette et programmes d'ajustement structurel en ont résulté¹.

4.2. Le risque de crédit et l'organisation de la banque :

La banque doit élaborer une politique de crédit et proposer des méthodes adaptées au traitement des dossiers de crédit puisque la distribution du crédit s'applique à un large éventail de situations complexes² :

➤ La politique de crédit :

La direction générale de la banque adopte les grands principes de la politique de crédit en précisant si elle s'appuiera sur un comité d'engagement ou de risque si nécessaire :

- Les objectifs recherchés en termes de clients, de types de crédit et d'emplacements géographiques sont conformes au plan stratégique ;
- Les taux d'intérêt du client qui doivent être facturés afin de produire suffisamment de profits pour payer les coûts associés aux prêts (ressources, gestion, risque et capitaux propres), ainsi que les garanties nécessaires ;
- Les délégations de pouvoir qui limitent le montant du crédit qu'un comité de crédit local ou un opérateur peut approuver avec sa seule signature font partie d'une décentralisation de la prise de décision. Selon le risque de crédit posé par le crédit, ces délégations peuvent également varier pour une même organisation.

¹ S.De GOUSSERGUES et autres ; op cite 251.

² Ibid ; p 253, p255, p256.

Les contrôleurs internes devraient vérifier soigneusement que toute octroi de crédit relève des pouvoirs du prêteur, car une délégation de pouvoir informelle ou ambiguë est une source importante de risque pour un établissement de crédit.

- Les procédures de traitement interne :

Il est nécessaire de créer des processus formalisés pour l'évaluation des demandes de crédit, le suivi des dossiers de crédit et les contrôles internes :

- L'étude de la demande de crédit : L'analyste de crédit est responsable de compiler toutes les informations essentielles sur l'emprunteur, qui sont ensuite consignées dans des dossiers conformément à une liste préalablement élaborée en fonction de la qualité de l'emprunteur, qu'il s'agisse d'une personne ou d'une entreprise. Le dossier est ensuite traité par le fonctionnaire compétent, qui rend un jugement d'acceptation ou de rejet. En cas d'acceptation, le montant, les conditions tarifaires, les modalités de remboursement et les garanties doivent être indiqués. Bien que la méthode de traitement d'un crédit varie, les normes de contrôle interne doivent être respectées. Il est essentiel de faire la distinction entre le devoir commercial de maintenir la relation client et la tâche d'examiner la demande de crédit confiée à un analyste de crédit. L'analyste de crédit examine le dossier et fait une recommandation de position. Il est essentiel que le décideur, l'exploitant ou le comité d'engagement reçoive l'autorité et la délégation nécessaires des niveaux hiérarchiques supérieurs, peu importe qui occupe ces rôles.
- Le suivi du dossier de crédit : Une fois le contrat de crédit accordé, la banque et l'emprunteur signent un contrat de prêt décrivant les conditions particulières des obligations de chaque partie, y compris les calendriers de remboursement et les restrictions tarifaires. L'étape suivante consiste à mettre en place l'encadrement du crédit jusqu'à son remboursement intégral et à définir les sanctions qui seront appliquées si l'emprunteur ne respecte pas ses obligations. Pour être en mesure d'identifier la faillite de la contrepartie et d'initier le traitement nécessaire dans les meilleurs délais, l'équipe de surveillance doit disposer d'une liste des actions à entreprendre et des délais à respecter. La surveillance des risques peut devenir difficile lorsque les employés en contact direct avec les clients sont les premiers à remarquer que la situation financière d'un client se détériore et prennent un certain temps pour alerter leurs supérieurs de peur d'être blâmés pour avoir fait une mauvaise évaluation des risques ou parce qu'ils

entretiennent une bonne relation de travail (ou personnelle) avec le client. Ces situations doivent être évitées par un contrôle interne.

- Le contrôle du risque de crédit : Le contrôle interne doit garantir la circulation de l'information sur les risques afin que la direction soit consciente de tous les dangers pour sa banque. Le contrôle tente également de confirmer le respect des réglementations bancaires suivantes : répartition des risques, tarification avec des marges suffisantes et provisionnement des prêts non performants.

Section 2 : La mesure du risque de crédit

Les normes prudentielles du Comité de Bâle, qui exigent que les banques maintiennent un système d'évaluation du risque de crédit, sont fondées sur la technique standard, ou la méthode de base IRB pour calculer la probabilité de défaut au niveau individuel ou la méthode avancée de IRB pour un portefeuille de crédit. Afin de présenter les nombreux modèles qui ont été créés dans le cadre des théories qui expliquent le risque de crédit, cette section sera consacrée à ce faire.

1. La mesure du risque de crédit au niveau individuel :

En mesurant le risque de crédit au niveau de chaque emprunteur, nous pouvons calculer le risque de recouvrement en établissant des critères de mesure pour chaque emprunteur et en recueillant des données à leur sujet. Ce risque est évalué sur la base de la collecte et de l'analyse de données concernant les nombreux aspects du risque de crédit pour chaque crédit d'un portefeuille, ce qui implique essentiellement d'évaluer le risque de crédit de chaque emprunteur. Il s'agit essentiellement de classer les crédits en groupes de risque et de calculer les probabilités de réussite en cas de défaut et de modification des notes. Étant donné que le montant des dommages en cas de défaut dépend du taux de recouvrement, l'évaluation de ce dernier taux fait également partie de la détermination du risque de crédit au niveau individuel.

1.1. Notation externe :

Ou l'approche standard, les agences de notation attribuent des notations externes basées sur des évaluations impartiales, fiables et transparentes. Cette vue qualitative est généralement exprimée à l'aide d'une ou de plusieurs lettres qui représentent la qualité de crédit de l'émetteur,

et la collecte de ces notations est connue sous le nom d'échelle de notation. Ces barèmes, qui s'appliquent à la fois à l'émetteur et à un produit de dette spécifique émis par ce dernier.

Dans le cas où l'échec est confirmé, les notes sont fournies par ordre décroissant du meilleur au terrible, comme indiqué dans la figure ci-dessous :

Figure 1: échelonnement de la notation

	S&P	Moody's	Fitch
Investment Grade	AAA	Aaa	AAA
	AA	Aa	AA
	A	A	A
	BBB	Baa	BBB
Speculative Grade	BB	Ba	BB
	B	B	B
	CCC	Caa	CCC
Default	D	D	D

Source : V.BRUNEL; op-cite ; p82.

Les entreprises ayant une forte capacité de remboursement et une vulnérabilité réduite aux risques économiques relèvent de la catégorie dite « Investment Grade ». Au contraire, les entreprises classées dans la catégorie « spéculative grade » offrent beaucoup moins de garanties, sont beaucoup plus vulnérables aux chocs économiques et sont plus risquées, « default » pour les entreprises qui l'ont déjà fait un défaut ou qui sont très susceptibles de le faire à l'avenir. Chaque notation se rapporte à une qualité de crédit particulière. Pour chaque notation, nous pouvons spécifier la qualité de crédit¹ :

¹ V.BRUNEL ; op-cite, p 18.

Figure 2: la qualité de crédit selon la notation

AAA	Meilleure qualité de crédit, excellente solidité
AA	Très bonne qualité de crédit, très solide
A	Bonne qualité de crédit, plus sensible aux conditions économiques
BBB	La plus basse qualité de crédit en Investment Grade
BB	Prudence requise, la meilleure qualité de crédit en Speculative Grade
B	Vulnérable, peut encore honorer ses engagements
CCC	Hautement vulnérable
D	Un défaut de paiement est déjà survenu

Source : *Ibid* ; p 18.

➤ Notation et probabilité de défaut :

Chaque émetteur reçoit une notation par un système de notation de crédit, qui est basé sur une échelle de notation qui hiérarchise le risque de défaut probable. Le score est intrinsèquement une indication synthétique du risque de défaut.

Étant donné que le risque de défaut ou le montant estimé de la perte exprimé dans la notation externe d'un émetteur, qui exprime une opinion sur la qualité de crédit de l'émetteur n'est pas une mesure aussi précise que la probabilité de défaut, il est nécessaire d'établir une relation solide entre les deux. Pour ce faire, les agences de notation publient régulièrement des tableaux des défauts observés par année, secteur d'activité, classe de notation et région géographique. Par rapport à l'univers total couvert par l'agence, ces tableaux fournissent une moyenne empirique des fréquences par défaut.

Il est possible d'établir un lien entre les notations et la probabilité de défaut grâce à ces faits. Alors que les classes de notation les moins bonnes permettent les plus grandes fréquences par défaut et donc une probabilité de défauts plus élevés, les meilleures classes de notation permettent les probabilités de défaut les plus faibles¹.

➤ La matrice de transition :

Les matrices de transition sont souvent calculées par les agences de notation. Ces tableaux en partie double donnent la probabilité qu'un émetteur ou un instrument appartenant à une classe de notation initiale passe à une autre classe au cours d'une durée pour un horizon particulier. L'idée est de définir un échantillon important de crédits sur une période historique

¹A. DE SERVIGNY, op-cite, p 86, 87

suffisamment longue pour être représentatif. A l'aide de ceci une statistique de migration est obtenue, et qui peut être utilisée pour déterminer tous les éléments de la matrice de transition¹.

1.2. Notation interne :

L'approche interne, également connue sous le nom de IRB (Internal Ratings Based), basée sur des évaluations internes des banques, tente de fournir une sensibilité supplémentaire à la mesure du risque de crédit contrairement à l'approche standard, qui ignore les impacts de la mutualisation des risques.

Afin d'imiter avec précision les outils qualitatifs développés par les agences de notation externe, les banques cherchent à établir des procédures en ajoutant des outils quantitatifs. Il y a évidemment une très forte motivation pour rendre ces systèmes efficaces en interne étant donné que le prêt est l'activité principale de la banque. Les systèmes internes se sont de plus en plus améliorés à partir d'une approche largement binaire, conduisant simplement à examiner si l'emprunteur était bon ou mauvais, vers des systèmes plus développés, en particulier avec l'impulsion du nouveau cadre réglementaire de Bâle II (Nous renvoyons le lecteur au chapitre 2 pour sa présentation complète.), ce dispositif encourage les banques à mettre en place un système interne plus sophistiqué pour évaluer la qualité de crédit de leurs clients et contreparties².

Selon le système de notation interne, la banque a le choix entre deux méthodes, l'une dite avancée et l'autre de base, Par conséquent, si la banque choisit la technique de base, elle n'aura qu'à évaluer la probabilité de défaut de la contrepartie, Le superviseur fixera les paramètres restants. Les quatre indicateurs de risque, à savoir la probabilité de défaut, l'exposition en cas de défaut, la perte en cas de défaut et l'échéance moyenne restante de l'engagement, doivent être déterminés si la banque choisit la méthode avancée³.

En conséquence, la banque décide des fonds propres à constituer en évaluant le risque de défaillance de la contrepartie en utilisant les informations confidentielles de l'emprunteur auxquelles elle a accès en raison de la relation à long terme⁴. Trois catégories de méthodes sont utilisées :

¹ Ibid p 87

²A. DE SERVIGNY., op-cite, p 95.

³ A.BENTHAMI, **Les limites de la notation interne appliqué par les banques aux petites et moyennes entreprise : cas du maroc**, revue techniques financière et développement, N°124, mars 2016,p142.

⁴ S.De GOUSSERGUES et autres, **Gestion de la banque**, 9^{ème} édition, dunod,2020, p181.

1.2.1. scoring :

Une technique de notation quantitative qui donne à chaque élément d'information qui reflète la solvabilité de l'emprunteur une pondération dans le but de synthétiser le risque de contrepartie sous forme de score. La décision d'approuver ou de refuser la demande de crédit peut être prise immédiatement en fonction de la somme des pondérations par rapport à un score limite préalablement défini¹.

L'objectif de cette valeur est de prévoir la probabilité d'un défaut de paiement de l'emprunteur. De ce fait, la création d'un système d'anticipation des défaillances de l'entreprise implique d'établir un lien entre les caractéristiques de l'entreprise et son défaut. Cette relation causale peut être exprimée quantitativement comme une fonction qui transforme ces valeurs en une note, un nombre dont la valeur représentera le risque d'échec².

Les scores utilisent donc des procédures quantitatives basées sur des informations comptables, financières et parfois non financières. Le Z-score créé par Altman (1968) est l'un des moyens les plus connus. Cette approche repose sur l'hypothèse que les données comptables offrent des informations prévisionnelles sur la probabilité de défaut d'une entreprise, des informations qui s'appliquent à la fois aux organisations saines et à celles déjà en défaut³.

1.2.2. L'approche structurelle :

Les modèles d'évaluation structurelle de l'entreprise (Merton 1974, MKMV Credit Monitor) ont permis d'estimer la probabilité de défaut des entreprises cotées en bourse à l'aide de diverses approximations empiriques.

La stratégie est basée sur la philosophie optionnelle de l'entreprise. La structure financière de l'entreprise est modélisée selon Merton comme un échange d'options sur la valeur de l'entreprise V . Une option d'achat détenue par l'actionnaire sur V a un prix d'exercice égal au montant nominal de la dette D . La valeur excédentaire de $V - D$ est restituée aux actionnaires sous forme de dividendes au moment de l'échéance de la dette, une fois les créanciers payés⁴.

¹ Ibid p 191.

² K. BOUAISS et autres, **Economie et gestion de la banque**, édition management et société, p 132.

³ A. DE SERVIGNY; op-cite, p97.

⁴ C. KHAROUBI et P. THOMAS; op-cite; p 107.

Pour décrire le risque de crédit, ce modèle utilise les principes de pricing des options de Black & Scholes (1973). Il est basé sur la valeur de la société emprunteuse et l'échange d'options qui a lieu entre le prêteur et l'actionnaire lorsqu'un prêt est accordé¹.

Plus tard, de nombreux modèles ont été proposés. Bien qu'elles soient beaucoup plus compliquées sur le plan mathématique, ces extensions ne donnent pas de résultats sensiblement différents. Les modèles appelés à frontières exogènes sont des modèles qui suppriment l'une des hypothèses irrationnelles de Merton : le défaut peut se développer à tout moment dès $V_r < K$, K étant un seuil exogène. Puisque le taux de guérison est déterminé de manière exogène, ils sont néanmoins moins « généralistes ». Ils produisent des résultats comparables. Le modèle de Longstaff et Schwartz (1995), est l'exemple original du genre. Ils fixent vraisemblablement un point en dessous duquel l'entreprise éprouve des difficultés financières, et donc, les dettes de chacun font toutes défaut en même temps. Une petite partie des prêts sont récupérés par les créanciers (taux de recouvrement connu a priori, ce qui élimine le problème d'ancienneté de la dette du modèle Merton). Zhou (1997) ajoute des processus de saut à la dynamique en cours qui déterminent la valeur des actifs de l'entreprise. Contrairement aux modèles à frontières exogènes, il définit des trajectoires qui ne sont plus tenues d'être continues, ce qui permet d'atteindre des spreads de crédit à court terme autres que zéro. Le modèle M-KMV est l'une des nombreuses extensions de modèle de Merton que les praticiens utilisent le plus fréquemment. M-KMV estime la distance de l'entreprise au défaut et détermine, pour un certain horizon, sa probabilité de défaut en fonction de la valeur comptable de la dette d'une entreprise et de l'évaluation de sa valeur marchande².

1.2.3. Les systèmes experts :

Dans les systèmes experts, la méthodologie est de nature qualitative et vise à reproduire de manière cohérente les critères de jugement des experts en crédit ou leur méthodologie d'évaluation des risques. En interrogeant les experts sur leurs méthodes, en confrontant leurs points de vue et en leur demandant d'approuver conjointement les règles de décision qui résultent de ces débats et désaccords, ces règles sont élaborées de manière entièrement empirique. Cet ensemble de règles pondérées est utilisé pour identifier les facteurs de risque de

¹ A. DE SERVIGNY ; op-cite ; p 100.

² C. KHAROUBI et P. THOMAS; op-cite; p 114 115.

l'emprunteur et pour évaluer l'emprunteur. Cette méthode est utilisée pour l'évaluation des grands clients¹.

2. La mesure du risque de crédit au niveau de portefeuille :

Après avoir examiné le risque de crédit en termes individuels, où le risque de chaque client est évalué isolément, il devient clair que le raisonnement sur un portefeuille va au-delà de l'accumulation du risque de crédit individuels. Une banque doit disposer d'un modèle pour évaluer le risque de crédit du portefeuille.

L'estimation de la distribution des pertes est la façon dont le risque de crédit est mesuré au niveau du portefeuille (fonction de distribution de probabilité ou PDF). L'analyse du portefeuille révèle deux catégories de pertes : les pertes attendues (EL), qui sont des pertes anticipées dans des conditions normales de marché, et les pertes inattendues (UL), qui sont des pertes extrêmes qui ne peuvent être observées que dans des conditions de marché exceptionnellement volatiles. Les institutions financières traitent ces deux catégories de pertes différemment. Des informations sur chaque exposition de crédit doivent être fournies afin d'utiliser cette fonction PDF et de construire son modèle au niveau du portefeuille. Les institutions financières peuvent utiliser ces informations pour calculer les provisions nécessaires (en fonction des pertes anticipées) et le capital économique (en fonction des pertes imprévues) afin d'assurer leur solvabilité et leur capacité à faire face à des catastrophes prévisibles et imprévues². Parmi ces modèles :

2.1. Credit Metrics :

Le modèle CreditMetrics est utilisé pour analyser et gérer le risque de crédit du portefeuille, le modèle CreditMetrics est constamment développé et amélioré depuis sa création pour devenir aussi flexible que possible pour répondre aux évolutions du marché et aux exigences réglementaires. Ce modèle appartient à la catégorie des modèles market to market, car il estime le défaut de l'émetteur sur la base des changements de notation³.

¹ M.DIETSCH et J. PETEY ; **Mesure et gestion du risque de crédit dans les institutions financières** ; édition revue banque ; 2ème édition ; 2008 ; p 36.

² Ibid 137.

³ K.BORIS, et autres, **Quantification of credit risk with the use of CreditMetrics**; procedia economics and finance N° 26, 2015, p 311.

Grâce à l'examen d'éléments dont les rendements individuels sont présumés indépendants et suivent une distribution normale, il permet de simuler l'évolution de la valeur du portefeuille. CreditMetrics fournit un total de six entrées. Les inputs de CreditMetrics sont au nombre de six, respectivement¹ :

- Les risques de défaut et de transition ;
- Les moyennes du taux de récupération et les écarts-types pour chaque industrie et niveau d'ancienneté ;
- La structure de corrélation factorielle et la sensibilité de chaque instrument à ces facteurs ;
- La structure de durée de la structure des taux sans risque ;
- Pour chaque classe de notation, les structures sont organisées par terme de spread de crédit.
- Chaque instrument a son propre profil d'exposition.

Le modèle est basé sur des matrices de transition (fournies, par exemple, par des agences de notation telles que Moody's ou Standard & Poor's) pour déterminer la valeur future d'un portefeuille. Plus précisément, on suppose que les probabilités de transition à l'avenir seront semblables à leur moyenne historique. Par conséquent, un tel système est incapable de tenir compte des chocs potentiels qui pourraient modifier la probabilité de transitions. Ces matrices ne tiennent pas compte non plus des corrélations de migration de divers titres. L'étape initiale de la méthode consiste à créer des matrices de transition qui tiennent compte des corrélations. Une fois produites, ces matrices de transition permettront le calcul de la VaR en permettant le processus itératif d'obtention d'une distribution de la valeur future du portefeuille après une phase d'évaluation².

2.2. credit risk+ :

CreditRisk+ est un modèle de risque par défaut qui considère uniquement la possibilité d'un défaut et le modélise comme une variable continue avec une distribution de probabilité. Le processus utilise ensuite des méthodes actuarielles pour simuler la distribution des pertes et ne tient pas compte du risque. En ce qui concerne les raisons de la faute, aucune hypothèse n'est faite dans cette modélisation. Les défauts individuels sont répartis indépendamment et

¹ A. DE SERVIGNY, op cite, P 184.

² C.Gouriéroux et A.Tiomo, op cite, p 355.

identiquement selon la loi de Bernoulli, sous réserve de variables de risque spécifiques. La probabilité individuelle de défaut, elle-même aléatoire, est décrite par un modèle factoriel. Les corrélations de défauts sont prises en compte via la dépendance des probabilités de défaut individuelles sur un ensemble de facteurs communs. Les facteurs introduits dans CreditRisks+ sont très simples. Le portefeuille est organisé en un certain nombre de sous-portefeuilles, dont chacun, se rapporte à une industrie spécifique. Le taux d'échec est considéré comme aléatoire basé sur la loi gamma pour chaque secteur. Cette loi gamma admet deux paramètres qui peuvent être calibrés à l'aide de moyennes historiques de taux de défaillance et de données de variabilité. Le modèle calcule les probabilités de défaut individuelles de chaque contrepartie à partir de la probabilité de défaut par classe de risque. En suivant la méthodologie à deux périodes, les probabilités de défaut pour un portefeuille donné sont dérivées. La répartition des pertes actuarielles prospectives est finalement estimée sur la base de ces probabilités et en connaissant l'exposition de chaque contrepartie, c'est-à-dire le montant des pertes en cas de défaut¹.

2.3. Portfolio manager :

KMV ont d'abord préconisé d'utiliser le modèle de valeur de la firme de Merton pour évaluer les différentiels de taux d'intérêt sur les obligations de sociétés. La société a étendu sa technique à d'autres types de crédit et de dérivés de crédit avec la mise en œuvre de la nouvelle règle. Notamment, à la suite de sa fusion avec Moody's, ils ont intégré des effets de notation dans leur modèle de base. Le cœur du modèle est actuellement une boîte noire, et il n'est pas clair comment la méthodologie de valeur de l'entreprise peut être rendue cohérente avec la procédure de calcul de la notation de Moody's.

La méthode de KMV évalue l'évolution des actifs et des obligations d'une entreprise, et considère qu'une entreprise fait défaut lorsque la valeur vénale de ses actifs est insuffisante pour rembourser ses dettes. La première étape consiste à analyser la valeur marchande non observable des actifs et leur volatilité. À cette fin, KMV propose d'utiliser la valeur de marché des actions et leur volatilité comme approximation de la valeur de marché des actifs et de leur volatilité sur la base de deux relations. Une fois la valeur de l'actif estimée, KMV calcule une distance par défaut (DD), qui est définie comme la distance entre la valeur attendue des actifs avec un horizon fixe (un an, par exemple) et le point de défaut (K), représentée en termes d'écart type de la valeur future des actifs σA^2 .

¹ Ibid p 357.

² Ibid, p 347.

2.4. VAR de crédit :

J.P Morgan a contribué à la formulation de la valeur à risque, qui tente de synthétiser l'ensemble du risque d'un portefeuille d'actifs financiers en un seul nombre. Cette mesure est depuis largement acceptée par les organisations financières. La mesure de la VaR est également utilisée par le Comité de Bâle pour calculer le capital requis par les banques pour le risque de crédit¹. La valeur à risque est définie comme la perte maximale possible qu'une organisation financière peut subir sur une période donnée (horizon de détention) et avec un niveau de probabilité spécifique (intervalle de confiance). Il peut être mesuré au niveau agrégé ou au niveau du portefeuille. En conséquence, il est facile d'identifier les positions qui présentent un risque. De plus, la sélection des caractéristiques permet de formuler une stratégie de gestion des risques. La prudence s'exprime par l'identification claire de trois paramètres : l'horizon de détention, le niveau de risque et la limite de la VaR. L'horizon de détention est déterminé par la composition du portefeuille. On s'attend à ce qu'il soit lié à l'ordre de marché de liquider le portefeuille. Le niveau de probabilité représente l'aversion au risque de l'entreprise. En conséquence, les pertes imprévues (UL) peuvent être calculées².

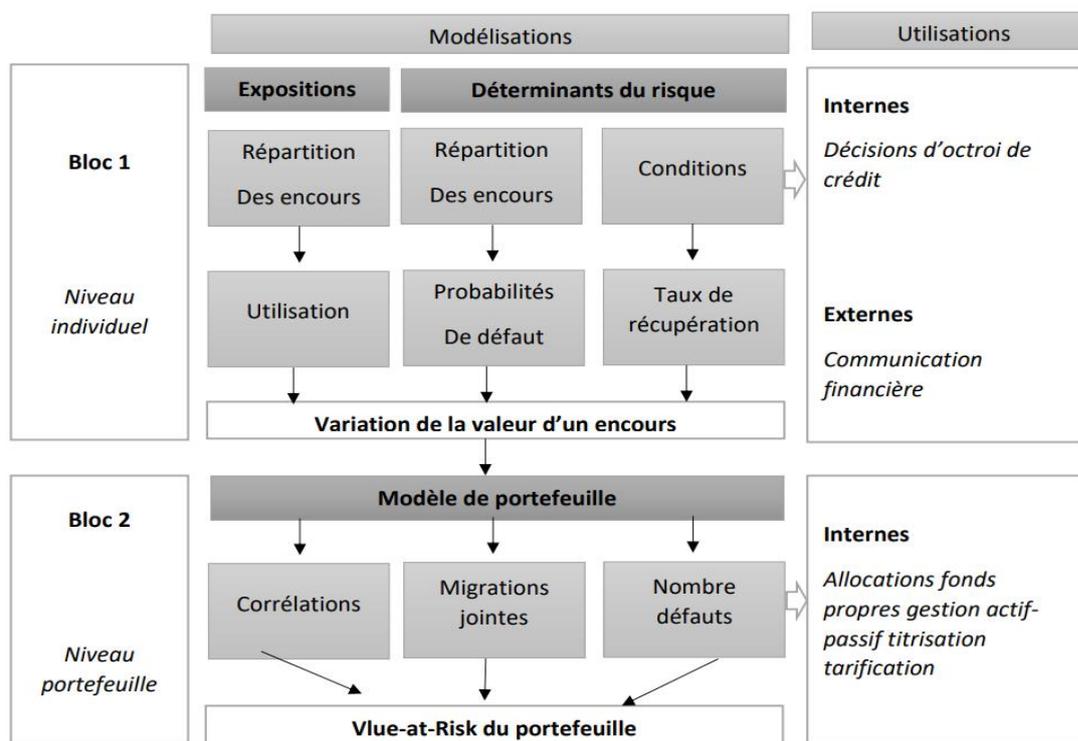
3. L'architecture d'un modèle du risque de crédit :

Comme nous avons vu précédemment l'analyse du risque de crédit est effectuée au niveau individuel et au niveau de portefeuille, selon une architecture suivante :

¹ J.HULL et autre ; **gestion des risques et institutions financières** ; 2^{ème} édition ; pearson education ; 2007 ; P 163.

² C. KHAROUBI et P. THOMAS ; opcite ; p 143.

Figure 3:L'architecture d'un modèle de risque de crédit



Source : M. DIETSCH et J. PETEY ; op cite ; p 27.

L'architecture à deux niveaux et les éléments clés d'un modèle de risque de crédit sont présentés dans le diagramme ci-joint, ainsi que les principaux extrants du modèle pour la mesure du risque de crédit.

Afin de calibrer les paramètres du modèle de portefeuille, le système de notation interne (bloc 1) offre les éléments d'information fondamentaux qui sont les probabilités de défaut, les expositions par défaut, et les taux de recouvrement par défaut (bloc 2).

Au premier niveau, les banques de détail ont été principalement impactées par la création de nouveaux instruments de mesure des risques au niveau individuel. Au deuxième niveau, la modélisation de la volatilité des probabilités de défaut et le calcul des corrélations entre sinistres constituent le cœur du modèle de portefeuille pour cette clientèle, et c'est à partir de ces composantes que nous construisons la fonction de densité future qui détermine la VaR du portefeuille et les sorties du modèle¹.

¹ M.DIETSCH et J.PETEY; op-cite; p 26.

Conclusion

Le crédit est le principal moteur de la croissance économique et la meilleure source de capitaux pour relancer les économies de toutes les nations. À cet égard, il est nécessaire d'adopter des réglementations souples et changeantes pour promouvoir la croissance économique et, dans toute la mesure du possible, prévenir les litiges. Ce chapitre vise à donner un aperçu général des aspects du risque de crédit et des facteurs qui nous aident à l'identifier, ainsi que de la façon dont les institutions financières réagissent à ce risque, planifient ses pertes prévisibles et évaluent la résilience de ces institutions face à des circonstances extrêmes (telles que des crises financières ou économiques) ou la réalisation de scénarios ou de pertes improbables. Nous avons également discuté des facteurs quantitatifs et des procédures nécessaires pour établir le modèle de crédit, ainsi que des nombreuses étapes de la création du modèle de crédit au niveau individuel et au niveau de portefeuille. La réglementation bancaire répond et encadre ces relations par le biais d'une législation souveraine, comme nous le verrons dans le chapitre suivant, afin d'organiser les relations des banques avec leurs clients, ainsi que les relations des banques entre elles et avec les organismes de surveillance.

Chapitre II

Gestion du risque de crédit

Introduction

Il est essentiel de garder à l'esprit que la gestion d'un risque ne signifie pas nécessairement le gérer. Cependant, la gestion d'un risque est impossible sans l'identifier et le mesurer. Les modèles, aussi sophistiqués soient-ils, ne peuvent remplacer une saine gestion.

Le risque de crédit lié à ses activités de prêt est de loin le plus grand risque pour les institutions de financement comme la Banque. Par conséquent, la viabilité à long terme de la Banque et sa capacité d'atteindre son objectif de réduction de la pauvreté et de développement durable dépendent essentiellement d'une saine gestion du risque de crédit.

Ce chapitre est segmenté en deux sections, première section consacrée à la gestion prudentielle, et une deuxième section pour décrire les techniques de gestion du risque de crédit.

Section1 : la gestion prudentielle

La gestion du risque de crédit et la réglementation prudentielle ne devraient pas être séparées. Cette partie décrira la réglementation prudentielle internationale et nationale, les normes décrites précédemment dans les théories du risque de crédit et les normes qui ont surgi à la suite du développement mondial et du début des crises.

1. La réglementation prudentielle baloise :

Un ratio minimum de fonds propres sur le total des actifs a été établi comme principale méthode de réglementation des fonds propres des banques dans plusieurs pays avant 1998. Cependant, les définitions des ratios et des fonds propres variaient d'un pays à l'autre. Comme les banques étaient en concurrence mondiale, la réglementation était également plus ou moins stricte selon les pays, donnant aux institutions opérant dans des pays où les lois bancaires étaient faibles un avantage concurrentiel sur leurs rivaux. En outre, les prêts importants accordés par les grandes banques internationales à des pays émergents comme le Mexique, le Brésil et l'Argentine, ainsi que la comptabilité inventive des banques, ont remis en question l'utilité de la règle actuelle sur le capital des banques. Le Comité de Bâle sur le contrôle bancaire a été créé à la suite de ces nombreuses questions par les régulateurs prudeniels¹.

Le Comité de Bâle a été créé en 1974 à la demande du Groupe des 10 (Allemagne, Belgique, Canada, États-Unis, France, Italie, Japon, Pays-Bas, Royaume-Uni, Suède et Suisse) à la suite d'une crise bancaire provoquée par la faillite d'un important établissement de crédit allemand, la Herstatt Bank, à la Banque des règlements internationaux à Bâle. Suisse. Son objectif est de formuler des recommandations que les régulateurs nationaux peuvent mettre en œuvre. Les nombreuses recommandations sont soumises aux pays membres, mais elles n'ont pas de force juridique².

Le Comité a publié divers rapports et suggestions entre 1974 et 1988. Ce sont les prémisses de l'accord de Bâle I, qui a pris la forme d'un document normatif approuvé par les différents membres du comité en juillet 1988, et est appelé Bâle I³.

¹ H.JOHN et autres ; op cite ; p 234, 235.

² R.HENNANI ; **De Bâle I à Bâle III: les principales avancées des accords prudeniels pour un système financier plus résilient** ; rapport technique, janvier 2015 ; p 2.

³ Ibid, p 7.

1.1. La réglementation de Bâle I :

La suffisance des fonds propres des banques par rapport aux risques qu'elles assument a été prise comme raisonnement. Le ratio Cooke (du nom du président du Comité) permet la construction d'un cadre réglementaire unifié qui s'applique à toutes les entreprises bancaires. Il encourage les banques multinationales à augmenter leurs fonds propres et minimise les distorsions de concurrence entre les banques dont la législation nationale était auparavant très variée. Il précise que les fonds propres réglementaires d'un établissement de crédit ne peuvent être inférieurs à 8 % du total de ses engagements de crédit pondérés¹:

$$\text{ration cook} = \frac{\text{fonds propres}}{\text{encours pondérés de crédit}} \geq 8\%$$

L'Accord de Bâle sur les fonds propres, signé en juillet 1988, était une réponse à la montée des risques et à la diminution des fonds propres des plus grandes banques du monde. Cet accord a été créé comme une fondation qui évoluera au fil du temps. Il a fait l'objet de multiples ajustements ultérieurs, chacun reflétant la nécessité d'améliorer le cadre prudentiel en réaction à des changements substantiels, voire à des crises ou à des perturbations, dont le plus important est la modification de 1996 visant à intégrer les risques de marché².

On peut dire donc que le ratio Cooke a souffert de nombreuses limites³ :

- Les pondérations des actions pondérées en fonction du crédit sont statiques, arbitraires et ne reflètent pas fidèlement le risque de crédit réel ;
- La contrainte est trop faible pour les petites entreprises (celles qui sont susceptibles de faire défaut) et trop élevée pour les grandes entreprises ;
- Quelle que soit la qualité de crédit de la contrepartie, la même pondération pour tous les types de crédit crée une certaine confusion entre l'adéquation des fonds propres et la tarification des prêts ;
- Quelle que soit la maturité de l'engagement, le règlement ne traite pas la structure par terme du risque de manière uniforme.

La taille de l'entreprise, sa notation, sa maturité, sa focalisation sur un seul émetteur, la corrélation entre émetteurs (et par conséquent les effets de la diversification du risque de crédit), la rémunération des positions longues et courtes sont autant de facteurs négligés.

¹ C.KHAROUBI et P.THOMAS ; op-cite ; p 139,140.

² A. DE SERVIGNY et I.ZELENKO, op-cite ; p 252.

³ C.KHAROUBI et P.THOMAS ; op-cite ; p 139,140

En réponse à ces plaintes, les organismes de réglementation des différents pays ont décidé d'accroître la réglementation prudentielle en tentant de mieux prendre en compte diverses catégories de risques, ce qui a mené à la création d'un nouvel accord connu sous le nom de Bâle II.

1.2. Les accords de Bâle II :

Le nouvel accord repose sur trois piliers interdépendants ; Prises dans leur ensemble, ces trois composantes devraient soutenir la stabilité et la sécurité du système financier. Le Comité souligne qu'il est important d'appliquer les trois piliers de manière rigoureuse et simultanée, et il prévoit de travailler en étroite collaboration avec les autorités de surveillance bancaire pour mettre en œuvre efficacement toutes les dispositions de l'accord. Toutefois, la principale garantie pour stabiliser et protéger le système financier demeure le montant des fonds propres détenu par la banque¹.

1.2.1. Exigences minimales en fonds propres :

Le pilier 1 spécifie la valeur des fonds propres minimum requis. Il s'efforce de garantir que les risques de crédit, de marché et opérationnel sont suffisamment capitalisés. Par rapport au ratio de solvabilité de Bâle I, le caractère unique de Bâle II réside non seulement dans l'éventail plus large de risques couverts (le risque opérationnel n'est pas inclus dans le ratio Cooke), mais aussi dans la capacité des banques à choisir d'autres niveaux de sophistication pour calculer les exigences de fonds propres. En conséquence, les entreprises pourront utiliser soit des pondérations forfaitaires graduées en fonction de la qualité de la contrepartie, soit des notations internes pour le risque de crédit et opérationnel².

La définition des fonds propres et le seuil minimal de 8 % restent inchangés dans le cadre actuel. Toutefois, le montant des fonds propres a été réparti différemment sur la base de pourcentages établis par le Comité en fonction des différents types et de l'ampleur des risques (75 % des fonds propres seront affectés au risque de crédit, ce qui équivaut à 6 %, 20 % au risque opérationnel et 5 % au risque de marché). Nous tombons donc sur le nouvel ratio McDonough³:

$$\text{ration MC Donough} = \frac{\text{totales des fonds propres}}{\text{risque de r dit} + \text{risque de march } + \text{risque op rationnel}} \geq 8\%$$

¹ H.FEKIR ; **La crise financi re et l'application des accords de B le : est-ce le meilleur rem de ?** ; Universit  Ibn Tofail, La Facult  des Sciences Juridiques, Economiques et Sociales de K nitra, Maroc ; may 2010 ; p 7.

² P.THORAVAL ; **Le dispositif de B le II : r le et mise en oeuvre du pilier 2** ; Revue de la stabilit  fi nanciere • N  9 • D cembre 2006 ; p126.

³H.FEKIR ; op cite ; p

1.2.2. Processus de surveillance prudentielle :

Le deuxième pilier consiste non seulement à veiller à ce que les banques disposent de fonds propres suffisants pour couvrir tous les risques associés à leurs opérations, mais aussi à encourager les banques à élaborer et à mettre en œuvre des pratiques améliorées de surveillance et de gestion des risques¹.

La procédure de surveillance a un objectif à deux niveaux ²:

- Au niveau de la banque : Encourager les banques à créer des stratégies pour réglementer leurs exigences de fonds propres et leur exposition au risque. Chaque banque peut surveiller efficacement ses risques et évaluer le calibre de ses actifs grâce à cette méthode interne de surveillance des risques. Donc elle peut pratiquer :
 - Le back testing : La banque doit démontrer la fiabilité de ses procédures statistiques sur des périodes relativement longues (5 à 7 ans) ;
 - Le stress testing : La banque doit démontrer la validité de ses fonds propres en cas de catastrophe économique lors de simulations de scénarios extrêmes.
- Au niveau du système réglementaire : Les autorités de réglementation devraient être capable à augmenter les exigences de capital réglementaire si nécessaire, en fonction des résultats de ces tests en général.

1.2.3. Discipline de marché :

Le troisième pilier, la discipline de marché, est destiné à compléter le premier pilier, les exigences minimales de fonds propres et le processus de surveillance (deuxième pilier). Le Comité a l'intention de renforcer la discipline de marché en adoptant un ensemble de critères d'information financière qui permettront aux participants au marché d'examiner les informations essentielles sur le périmètre de l'établissement, les fonds propres, les expositions aux risques, les procédures d'évaluation des risques et, par conséquent, la suffisance des fonds propres. Le Comité estime que cette information est particulièrement importante dans le contexte du cadre modifié, où l'utilisation de techniques internes offre aux banques une plus grande souplesse dans l'évaluation de leurs exigences de fonds propres³.

¹ M.MERHOUN et I. BENHALIMA ; Les **Banques Algériennes A La Lumière Des Règles De Bale I, II, III : Bilan Et Perspectives** ; revue des reformes économique et l'intégration dans l'économie mondiale, N°23 ; juin 2017 ; P 137.

² S.Chebaiki et KH.Slimani ; **Les Accords de Bale ou L'Évolution des Règles Prudentielles** ; revue Djadid El-iktissad ;N°09 ; décembre 2014 ; p37.

³ Comité de Bâle sur le contrôle bancaire ; **Convergence internationale de la mesure et des normes de fonds propres** ; banque des règlements internationales ; juin 2004 ; p 156.

Cécile Kharoubi & Philippe Thomas citent les insuffisances suivantes quant aux normes de Bale II :

- C'est un indicateur procyclique. En période d'exubérance financière, les risques pondérés diminuent (puisque'ils sont basés sur des pertes antérieures), les banques ont besoin de moins de fonds propres, et elles se contentent de détenir le strict minimum. Lorsque la situation s'aggrave, les banques doivent augmenter leurs fonds propres pour respecter les normes de solvabilité, les ressources devenant plus rares et plus chères, contribuant à l'asphyxie financière et à une réduction de la disponibilité du crédit (phénomènes de resserrement du crédit), ce qui accentue la récession économique.
- Les éléments les plus risqués et les plus complexes, ou les dangers du marché, sont sous-pondérés. En conséquence, les banques ont été en mesure d'évaluer correctement les risques qu'elles prenaient. Par conséquent, leur degré de trésorerie propre était disproportionné par rapport aux dangers réels pris.
- Les positions hors bilan sont difficiles à évaluer. Souvent l'analyse des risques pertinents est difficile, car les produits dérivés hors bilan sont souvent de taille assez importante.

1.3. Bale III :

Les limites et les défauts de Bâle II ont été mis en lumière par la crise financière de 2008 au point que certains analystes ont même accusé Bâle II de contribuer à l'aggravation des conséquences de la crise. Bale II est, en fait, principalement liée au fonds propres minimum et ignore certains risques tout aussi importants, y compris les risques de liquidité et les actifs risqués financés avec peu ou pas de fonds propres. Les banques sous-estiment les risques qu'elles prenaient et ont des niveaux de fonds propres déraisonnablement élevés. Avec les nouvelles recommandations connues sous le nom de Bâle III, le Comité de Bâle a une fois de plus sauvé la situation¹.

L'objectif principal de la mise à jour de la réglementation précisant les exigences minimales de fonds propres pour les banques est d'accroître la stabilité du système financier, ce qui réduira la possibilité et la gravité de crises futures. L'Accord de Bâle III, un ensemble révisé de règlements internationaux sur les exigences de fonds propres pour les banques, a été publié par le Comité de Bâle en 2010. Le groupe a dévoilé un certain nombre d'initiatives visant à réorganiser complètement le système réglementaire mondial, notamment²:

¹ S.Chebaiki et KH.Slimani ; op cite ; p 38, 39.

² F.Hache ; **Bâle 3 en 5 questions** ; finance watch ; 2012 ;p 3.

1.3.1. Augmentation des exigences en matière de fonds propres :

L'objectif est d'accroître la cohérence, la qualité et la transparence du capital. Les banques doivent disposer de plus de fonds propres de haute qualité pour faire face aux pertes imprévues. Le ratio minimum de fonds propres de catégorie 1 passe de 4% à 6%, avec au moins les trois quarts des titres de la plus haute qualité requis (actions ordinaires et bénéfiques non distribués). Des exigences de fonds propres supplémentaires s'appliquent aux banques d'importance systémique mondiale (EBISm)¹.

1.3.2. Étendre la couverture des risques :

L'un des enseignements les plus importants de la crise a été la nécessité d'améliorer la couverture des risques liés aux capitaux. Pendant la crise, un aspect déstabilisateur majeur a été l'incapacité à tenir compte de manière appropriée des risques graves sur les positions de bilan et hors bilan, ainsi que des expositions aux produits dérivés. Pour remédier à ces lacunes, le Comité a achevé de nombreux règlements essentiels de Bâle II en juillet 2009, renforçant ainsi les exigences de fonds propres².

En conséquence, un nouveau coussin contracyclique obligatoire de 0 % à 2,5 % est créé en plus du ratio décrit ci-dessus, chaque trimestre, chaque État membre doit évaluer si la création de crédit est excessive et risque de créer une bulle et décider si ses banques ont besoin de capitaux supplémentaires sur la base de cette évaluation. Cette mesure est basée sur l'hypothèse que le crédit est cyclique et sensible aux bulles. Les banques doivent donc constituer des fonds propres plus importants en période de prospérité, ralentir l'expansion du crédit pendant les périodes d'expansion et renforcer la solidité bancaire pendant les récessions³.

1.3.3. Compléter l'exigence de fonds propres fondée sur le risque par un ratio de levier :

L'accumulation d'un effet de levier insoutenable dans le secteur bancaire, tant dans les bilans qu'en hors bilan, est l'une des caractéristiques fondamentales de la crise. Ce trait était également répandu dans les crises financières antérieures, comme la période juste avant septembre 1998. Au plus fort de la crise, le marché a contraint le secteur bancaire à réduire son effet de levier d'une manière qui a accru la pression à la baisse sur les prix des actifs, exacerbant la spirale des

¹ Comité de Bâle sur le contrôle bancaire, **Finalisation de Bâle III**, banque des règlements internationaux, Décembre 2017, p2.

² Comité de Bâle sur le contrôle bancaire, **Bâle III : dispositif réglementaire mondial visant à renforcer la résilience des établissements et systèmes bancaires**, banque des règlements internationaux, Décembre 2010, p3.

³ F.Hache ; op cite ; p 3,4.

pertes, l'épuisement du capital des banques et la réduction de la disponibilité du crédit. En conséquence, le Comité ajoute un nouveau critère sous la forme d'un ratio de levier avec les objectifs suivants ¹:

- Limiter l'effet de levier du secteur bancaire afin de réduire la possibilité que son renversement ait de graves répercussions sur le système financier et sur l'économie ;
- L'ajout d'une mesure simple, transparente et sans risque en complément de la mesure fondée sur les risques aiderait à mieux protéger contre le risque du modèle et les erreurs de mesure.

Ce ratio impose un niveau minimum de fonds propres afin de garantir la stabilité financière des banques. Il devrait être supérieur ou égal à 3 % et devenir obligatoire à compter de 2018² :

$$\text{ration levier} = \frac{\text{fonds de propres de base}}{\text{total des actifs}} \geq 3\%$$

1.3.4. Instauration de normes mondiales de liquidité :

Pour que le secteur financier soit stable, des exigences strictes en matière de fonds propres sont nécessaires, mais elles sont insuffisantes. Il est tout aussi crucial de disposer de liquidités adéquates, soutenues par des exigences prudentielles strictes. Cependant, il n'y a pas eu d'harmonisation mondiale dans ce secteur à ce jour. En conséquence, le Comité de Bâle élabore des critères internationaux uniformes en matière de liquidité. À l'instar des normes mondiales en matière de fonds propres, les normes de liquidité favoriseraient l'égalité des chances et établiraient des normes minimales pour aider à prévenir un nivellement par le bas. En créant deux normes minimales pour le financement des liquidités, le Comité a encore renforcé son cadre de liquidité. Dans le cadre du dispositif de liquidité, un ensemble d'indicateurs contribuera également à une plus grande uniformité de la surveillance prudentielle transfrontalière. Ces critères ont deux objectifs différents mais complémentaires. Le premier objectif est d'améliorer le profil de risque de liquidité à court terme d'une banque, Le deuxième

¹ Comité de Bâle sur le contrôle bancaire, **Bâle III : dispositif réglementaire mondial visant à renforcer la résilience des établissements et systèmes bancaires, banque des règlements internationaux**, Décembre 2010, p5.

² S.Chebaiki et KH.Slimani ; op-cite ; p 40.

objectif est d'améliorer la résilience à long terme en incitant davantage les banques à financer leurs activités à partir de sources structurellement plus stables ¹.

➤ Ratio de liquidité à court terme :

L'objectif de ce critère est de s'assurer que la banque dispose d'un niveau adéquat d'actifs liquides non grevés de haute qualité pouvant être convertis en liquidités pour répondre à ses besoins sur une période de 30 jours en cas de graves difficultés de financement, sur la base d'un scénario défini par le superviseur. La banque devrait être en mesure de survivre jusqu'au 30^{ème} jour du scénario de crise, la direction et/ou les autorités de surveillance de l'établissement doivent décider des mesures correctives nécessaires et/ou que le problème de la banque peut être résolu de manière ordonnée². Le ratio se définit comme suit :

$$\frac{\text{encours d'actifs de haute qualité}}{\text{total des sorties nettes de trésorerie sur les 30 jours calendaires suivantes}} \geq 100\%$$

Lors de la détermination de l'exposition d'une banque à des situations pouvant entraîner des appels de liquidités, le LCR utilise des approches conventionnelles de « ratio de couverture » de liquidité. Le total des sorties nettes de trésorerie du scénario devrait être calculé sur une période de 30 jours. La valeur du ratio doit répondre à l'exigence minimale de 100 % de la norme, (c'est-à-dire que le stock d'actifs liquides de haute qualité doit être au moins égal à la sortie nette totale de trésorerie). Afin de faire face à un épisode potentiel de crise de liquidité extrême, les banques devraient toujours répondre à ces critères et maintenir une réserve d'actifs liquides de haute qualité et non grevés. Étant donné qu'il n'est pas clair quand les entrées et les sorties ont lieu, les banques et les responsables prudentiels devraient tenir compte de la probabilité d'asymétries entre eux au cours de la période de 30 jours et s'assurer qu'il y a suffisamment d'actifs liquides disponibles pour faire face à ces asymétries potentielles. La situation hypothétique qu'implique ce ratio implique un choc à la fois à l'échelle du marché et exclusif à certaines banques, avec les effets suivants ³ :

- Retrait partiel des dépôts des consommateurs ;
- Perte partielle de capacité de financement pour les prêts de gros non garantis ;

¹ Comité de Bâle sur le contrôle bancaire, **Bâle III : dispositif réglementaire mondial visant à renforcer la résilience des établissements et systèmes bancaires, banque des règlements internationaux**, Décembre 2010, p9, p 10.

²² Comité de Bâle sur le contrôle bancaire, **Bâle III : dispositif international de mesure, normalisation et surveillance du risque de liquidité**, Décembre 2010 ; p 3.

³ Ibid; p 4.

- L'arrêt partiel du financement à court terme de certaines contreparties garanti par des suretés spécifiques ;
 - D'autres résiliations contractuelles, y compris l'obligation de fournir une garantie, à la suite d'une réduction de trois crans de la notation de crédit de l'établissement ;
 - Une augmentation de la volatilité du marché qui aurait une incidence sur la qualité de la garantie ou sur l'exposition future potentielle des avoirs dérivés, nécessitant une décote plus importante ou la fourniture de garanties supplémentaires, ou d'autres exigences de liquidité ;
 - Les retraits imprévus des engagements de crédit et de liquidité confirmés des clients que la banque a pris mais qu'elle n'a pas encore utilisés ;
 - La possibilité que la banque doive racheter ses titres de créance ou remplir des obligations non contractuelles afin de protéger sa réputation.
- Ratio structurel de liquidité à long terme :

Le Comité a créé le ratio de financement net stable pour encourager les organismes financiers à accroître leur financement à moyen et à long terme de leurs actifs et activités (NSFR). Ce critère est un minimum raisonnable de financement constant en ligne avec les profils de liquidité annualisés de leurs actifs et activités. Il est destiné à être utilisé comme un mécanisme minimal et fonctionne conjointement avec le ratio de liquidité à court terme (LCR) et d'autres contrôles prudentiels. Il incite à modifier structurellement les profils de risque de liquidité des établissements. Ces modifications impliquent de passer des déficits de financement à court terme à un soutien plus stable et à long terme des actifs et des activités. Le NSFR, en particulier, doit être conçu de manière à ce que les actifs à long terme soient financés avec une quantité minimale de passifs stables correspondant à leur profil de risque de liquidité. Il tente également d'encourager une meilleure évaluation du risque de liquidité dans les actifs du bilan et hors bilan afin d'éviter une dépendance excessive à l'égard des emprunts de gros à court terme lorsque la liquidité du marché est abondante. Voici comment il est défini¹ :

$$\frac{\text{montant de financement stable disponible}}{\text{montant de financement stable exigé}} > 100\%$$

En période de tensions prolongées, le financement stable est la partie des différents types et montants de financement sous forme de capitaux propres ou d'autres passifs qui sont supposés représenter des ressources fiables au cours d'une année. Les besoins de financement d'une institution donnée sont déterminés par les caractéristiques de liquidité des diverses catégories

¹ Ibid ; p27.

d'actifs qu'elle détient, ses expositions hors bilan conditionnelles et/ou les activités qu'elle exerce¹.

1.4. Vers Bâle IV :

En fait, les régulateurs ont estimé qu'il était nécessaire et essentiel de compléter la réforme de Bâle III par une nouvelle vague réglementaire sous la forme de propositions et de recommandations pour le secteur financier et bancaire afin que les banques puissent résister aux chocs potentiels au niveau mondial et que le système financier soit résilient et immunisé. Grâce aux modifications et révisions importantes proposées, Bâle IV est une nouvelle réforme entrée en vigueur en 2017. Il devait initialement commencer en janvier 2022 et se terminer en 2027, mais en raison de l'épidémie de coronavirus à la fin de 2019, il a été retardé jusqu'en 2023²:

¹ Ibid ; p28.

² S.Haloui ; **Accord de Bâle IV : Quels apport et impact pour l'industrie financière et bancaire à l'ère du COVID-19?**; IJAFAME ; N°2 ; novembre 2021 ; p279,P280.

	Bâle III	Bâle IV
Extension de la couverture des risques	Les exigences de fonds propres au titre du risque de marché augmentent sensiblement. Elles sont calculées sur la base d'une période de 12 mois de tensions sur les marchés. Le risque d'ajustement de l'évaluation de crédit est désormais intégré dans le dispositif.	Les révisions des approches standards pour le calcul du risque de crédit, du risque de marché, du risque d'ajustement de l'évaluation de crédit et du risque opérationnel permettent une plus grande sensibilité au risque et une meilleure comparabilité. Les contraintes posées à l'usage des modèles internes visent à réduire la variabilité indésirable du calcul des RWA qui constituent une estimation du risque déterminant le niveau minimum de fonds propres réglementaires qu'une banque doit conserver pour faire face à des pertes imprévues. Par les banques. Un plancher (output floor), Le plancher minimal de capital ou le solde prévisionnel ou provisionnel minimum qu'un établissement bancaire est en obligation de préserver en dehors de tout risque de perte, Limite les avantages que les banques peuvent tirer de l'utilisation de modèles internes pour le calcul des exigences minimales de fonds propres.
Limitation du levier bancaire	Un ratio de levier limite l'accumulation de dette visant à financer les investissements et activités des banques (levier bancaire), réduisant le risque d'une spirale de désendettement en phase de retournement conjoncturel.	Les grandes banques d'importance systémique mondiale sont assujetties à des ratios de levier plus élevés.

2. La réglementation prudentielle au niveau nationale :

Depuis le début des années 1990, les autorités monétaires algériennes s'efforcent d'adopter aussi étroitement que possible ces lignes directrices afin de s'aligner sur les recommandations du Comité de Bâle. En établissant un cadre prudentiel avec un certain nombre d'exigences quantitatives que toutes les banques et autres institutions financières autorisées en Algérie sont tenues de respecter rigoureusement.

La loi n ° 90-10 sur la monnaie et le crédit (LMC), depuis sa promulgation le 14 avril 1990, le paysage bancaire algérien s'est formé. A ces mesures s'ajoutent les instructions du conseil de la monnaie et du crédit (CMC) et de la banque d'Algérie. Ces normes s'inscrivent dans le cadre des règles prudentielles de gestion imposées aux banques et établissements financiers.

On trouve le règlement N°91-09 DU 14 AOUT 1991 relative à la fixation des règles prudentielles de gestion des banques et établissement financiers, compléter par l'instruction N°34-91 DU 14 NOVEMBRE 1991 relative à la fixation des règles prudentielles de gestion des banques et établissement financiers, cette instruction a été modifier plus tard par l'instruction N°74-94 DU 29 NOVEMBRE 1994 relative à la fixation des règles prudentielles de gestion des banques et établissement financiers, cette instruction a été modifier et compléter par l'instruction N°09-2007 DU 25 OCTOBRE 2007 modifiant et complétant l'instruction N°74-94 DU 29 NOVEMBRE 1994 relative à la fixation des règles prudentielles de gestion des banques et établissement financiers. Avec la promulgation du règlement N°2014-01 DU 16 FEVRIER 2014 portant coefficients de solvabilité applicables aux banques et établissements financières et règlement N°2014-02 DU 16 FEVRIER 2014 relatif aux grands risques et aux participations, tout autres dispositions contraintes sont abrogées, parmi ces normes quantitatives :

2.1. Le ratio de solvabilité :

Conformément à l'article 2 du Règlement N° 2014-01 DU 16 FÉVRIER 2014, les banques et les établissements financiers doivent maintenir à tout moment un ratio de solvabilité minimum de 9,5% entre le total de leurs fonds propres réglementaires ou nets et la somme de leurs risques de crédit, risques opérationnels et risques du marché pondéré, que ce soit sur une base individuelle ou consolidée.

$$\frac{\text{fonds propres nets}}{\text{risques pondérés}} \geq 9,5\%$$

Dans son troisième article, cette législation est influencée par les exigences de « Bâle III », imposant aux banques algériennes de maintenir un ratio de solidité, ou les fonds propres de base par rapport aux risques de crédit, opérationnels et de marché, d'au moins 7 %, soit un Core Tier 1 de 4,5 % plus un coussin de sécurité de 2,5 %.

$$\frac{\text{fonds propres nets}}{\text{risques pondérés}} \geq 7\%$$

- Les risques pondérés sont la somme des expositions au risque de crédit, au risque opérationnel et au risque de marché, qui sont les suivantes :
 - Le montant pondéré du risque opérationnel est calculé en multipliant le capital requis pour ces risques par 12,5.
 - Le montant des risques de marché pondérés est calculé en multipliant par 12,5 l'exigence de capital pour ces risques.

- Les risques de crédit au bilan et hors bilan sont tous deux présents :

Les risques du bilan	
Pondération	Risques du bilan
0%	<ul style="list-style-type: none"> - Créances sur l'Etat et organismes assimilés ; - Dépôts et créances sur la Banque d'Algérie et les services financiers d'Algérie Poste ; - Créances sur les administrations centrales et locales.
20%	<ul style="list-style-type: none"> - Dépôts et concours aux banques et établissements financiers installés en Algérie ; - Titres de créances émis par les banques et établissements financiers installés en Algérie ; - Dépôts et concours aux banques et établissements financiers ou assimilés, installés à l'étranger, bénéficiant d'une notation au moins égale à AA- ou équivalent.
50%	<ul style="list-style-type: none"> - Dépôts et concours aux banques et établissements financiers ou assimilés, installés à l'étranger, bénéficiant d'une notation au moins égale à BBB- ou équivalent et inférieure à AA- ou équivalent.
100%	<ul style="list-style-type: none"> - Tous les crédits aux entreprises, particuliers et associations, y inclus les crédits-bails ;
Les éléments du hors bilan	
Facteur de conversion	Les risques de hors bilan
0%	<ul style="list-style-type: none"> - Facilités de découvert et engagements de prêter non utilisés qui peuvent être annulés sans condition à tout moment et sans préavis.
20%	<ul style="list-style-type: none"> - Crédits documentaires accordés ou confirmés lorsque les marchandises correspondantes constituent une garantie.
50%	<ul style="list-style-type: none"> - Engagements de payer résultant de crédits documentaires lorsque les marchandises correspondantes ne constituent pas une garantie ; - Cautionnements de marché public, garanties de bonne fin et engagements douaniers et fiscaux ; - Facilités irrévocables non utilisées, telles que découvert et engagement de prêter dont la durée initiale est supérieure à un (1) an.
100%	<ul style="list-style-type: none"> - Acceptations ; - Ouvertures de crédits irrévocables et cautionnements constituant des substituts de crédit ; - Garanties de crédits distribués ; - Autres engagements par signature données de manière irrévocable, non cités ci-dessus.

Source : Elaboré par nous même à partir du règlement N°2014-02 DU 16 FEVRIER 2014.

- Les fonds propres réglementaires ou nets, sont définis en ses articles 8 à 11. Les fonds propres règlementaires comprennent les fonds propres de base et les fonds propres complémentaire. Selon l'article 9 du même règlement les fonds propres de base sont constitués de la somme :
 - Du capital social ou de la dotation ;
 - Des primes liées au capital ;
 - Des réserves (hors écarts de réévaluation et d'évaluation) ;
 - Du report à nouveau créditeur ; - des provisions réglementées ;
 - Du résultat du dernier exercice clos, net d'impôts et de distribution de dividendes à prévoir.

Il convient de conclure de ces éléments :

- Actions propres rachetées ;
- Report au nouveau débiteur ;
- Résultats déficitaires en attente d'affectation ;
- Résultats déficitaires décidés semestriellement ;
- Immobilisations incorporelles nettes des dotations aux amortissements et provisions hors valeur ;
- 50% des participations et autres créances équivalentes détenues dans d'autres banques et institutions financières ;
- Dépassement des restrictions de participation ;
- Dispositions supplémentaires imposées par la commission bancaire.

Les fonds propres de base peuvent contenir des plus-values intermédiaires à condition qu'ils soient :

- Calculés après prise en compte de toutes les charges liées à la période, ainsi que des amortissements et provisions ;
- Calculés après soustraction des acomptes sur dividendes et de l'impôt sur les sociétés ;
- Approuvé par la Commission bancaire et certifié par les commissaires aux comptes.

Les fonds propres complémentaires sont définis à l'article 10 du Règlement n° 14-01 du 16 février 2014 comme :

- 50% du montant des écarts de réévaluation ;
- 50 % du montant des plus-values latentes provenant de l'évaluation à la juste valeur des actifs disponibles à la vente (à l'exclusion des titres de participation détenus dans les banques et les institutions financières) ;
- Les provisions pour risques bancaires généraux, constituées sur les créances courantes du bilan, dans la limite de 1,25% des actifs pondérés du risque de crédit ;
- Les titres participatifs et autres titres à durée indéterminée ;
- Les fonds provenant d'émission de titres ou d'emprunts, à condition que :
 - Ils ne soient remboursables qu'à l'initiative de l'emprunteur et avec l'accord préalable de la commission bancaire,
 - Ils donnent à l'emprunteur la possibilité de différer le paiement des intérêts dans le cas où son niveau de rentabilité l'en empêche,

- Le remboursement anticipé n'est pas autorisé avant cinq ans, à l'exception de la conversion de ce remboursement en fonds propres,
- Les créances du prêteur sur la banque ou l'institution financière sont subordonnées à celles de tous les autres créanciers et sont disponibles pour couvrir les pertes même s'il n'y a pas de cessation d'activité.

Sans satisfaire aux exigences susmentionnées, les fonds provenant de l'émission de titres ou de prêts subordonnés doivent également satisfaire aux exigences suivantes :

- Si le contrat spécifie une date d'échéance fixe, la durée initiale ne doit pas être inférieure à cinq (05) ans ;
- Si aucune échéance n'est fixée, la dette ne peut être remboursée qu'après un préavis de cinq (05) ans,
- Le contrat de prêt ne contient aucune disposition exigeant le remboursement anticipé de la dette dans des situations autres que la liquidation de la banque ou de l'institution financière visée par le prêt, et seulement après règlement de toutes les autres dettes dues à la date de la liquidation.

Déduire de ces fonds propres supplémentaires 50 % du montant des participations et de toute autre créance comparable aux fonds propres détenus dans d'autres banques et établissements financiers.

2.2. Le ratio de division des risques :

Les autorités monétaires algériennes disposent d'une mesure prudentielle quantitative supplémentaire après avoir exigé des banques et des établissements financiers autorisés en Algérie qu'ils disposent d'un niveau de fonds propres adapté aux risques associés tels que spécifiés par le Comité de Bâle. Rachid AMROUCHE interprète le ratio de répartition des risques dans son livre « Réglementation, risque et supervision bancaire », comme une stratégie qui tente d'éviter la concentration importante des risques sur un seul ou un groupe de bénéficiaires, ce qui, en cas de défaillance d'insolvabilité, menace d'entraîner la banque vers le bas.¹ La réglementation prudentielle algérienne distingue deux normes :

¹ ALIANE.S et AMRI H ; **la réglementation prudentielle en Algérie et son niveau de conformité avec les standards de Bâle 1 et 2** ; mémoire en vue de l'obtention de diplôme de Master ; Option : Monnaie, Banque et environnement International ; Faculté des Sciences Economiques, sciences de gestion et sciences commerciales ; Université Abderrahmane Mira ; Bejaia ; Algérie ; juin 2013.

- Le risque encourus sur un même bénéficiaire :

Selon l'article 4 du règlement N°2014-02 DU 16 FEVRIER 2014 Chaque banque ou institution financière doit respecter le rapport maximum entre tous les risques résultant de l'exploitation sous le même bénéficiaire et ses fonds propres. Ce maximum est fixé par à 25 % à compter du 1 janvier 1995 :

$$\frac{\text{Risque encourus sur un bénéficiaire}}{\text{les fonds propres nets}} \leq 25\%$$

- Risque encourus sur un ensemble des bénéficiaires :

Au même article précédent est stipulé également « Chaque banque et établissement financier est tenu de respecter un rapport maximum entre, d'une part, l'ensemble des risques qu'il encourt du fait de ses opérations avec des bénéficiaires ayant reçu pour chacun d'entre eux des concours supérieurs à une certaine proportion des fonds propres nets et, d'autre part, le montant des dits fonds propres nets »¹.

Ce rapport est fixé par le règlement N°2014-01 DU 16 FEVRIER 2014 le montant total des risques encourus sur les bénéficiaires dont les risques dépassent pour chacun d'entre eux 10 % des fonds propres nets des dits banques ou établissements financiers n'excède pas huit fois ces fonds propres nets.

$$\left(\sum \text{risques sur les bénéficiaires} > 10\% \text{ FPN} \right) \leq 8 * \text{FPN}$$

2.3. Coefficient minimum de liquidité :

L'article 3 du Règlement N° 2011-04 DU 24 MAI 2011 identifiant, mesurant, gérant et contrôlant le risque de liquidité stipule que les banques et les établissements financiers doivent respecter un rapport entre le total des engagements et engagements à court terme et à vue reçus des banques et le total des actifs à court terme et réalisables à court terme reçus. Les banques et autres institutions financières sont tenues de maintenir un ratio de liquidité minimum d'au moins 100% à tout moment.

Ce coefficient permet de suivre et de gérer la capacité d'une banque à rembourser ses dettes dans un avenir très proche (au maximum un et trois mois d'échéance). Il vise à s'assurer que

¹ L'article 2 de l'instruction 74-94.

l'institution dispose d'actifs liquides, d'actifs à courte échéance (un à trois mois au maximum) ou d'actifs pouvant être mobilisés rapidement pour payer au moins 100% des passifs à court terme (tels que les valeurs mobilières cotées sur un marché organisé).

$$\frac{\text{actifs liquides à court terme}}{\text{passifs exigible à court terme}} = 100\%$$

Les éléments d'actifs liquides à court terme et les quotités à hauteur desquelles ils doivent être pris en considération, sont précisé dans l'article 4 de l'instruction N°07-2011 DU 21 DECEMBRE 2011.

L'article 6 de même instruction précise les éléments du dénominateur du coefficient de liquidité et les quotités à hauteur desquelles ils doivent être pris en considération.

2.4. Le ratio de transformation :

Afin de maintenir un équilibre spécifique entre leurs emplois et les ressources à long terme en monnaie nationale, les banques et les institutions financières sont tenues de respecter un ratio connu sous le nom de « rapport entre la monnaie propre et les ressources permanentes ».

L'article 2 du même règlement précise les composantes permettant de calculer le numérateur de ce coefficient, et l'article 9 du Règlement N° 2004-04 DU 19 JUILLET 2004 stipule que ce rapport doit être au moins égal à 60%. Seules les ressources dont l'échéance résiduelle est supérieure à 5 ans sont incluses dans le numérateur, y compris les fonds propres et les actifs assimilés. Le dénominateur reflète les travaux de longue durée et les composantes utilisées pour le calculer sont décrites aux articles 7 et 8. Le coefficient ressemble à ceci :

$$\frac{\text{fonds propres} + \text{ressources permanents} > 5\text{ans}}{\text{emplois} > 5\text{ans}} = 60\%$$

3. Les instruments IFRS 9 :

La norme comptable IFRS 9 intitulée « Instruments financiers : comptabilisation et évaluation » vient remplacer la norme IAS 39 « Instruments financiers ». Les critiques de l'idée de juste valeur, qui évalue les actifs financiers à leur valeur marchande, sont l'un des principaux moteurs de cette réforme. En fait, cette méthode d'évaluation implique la comptabilisation dans les comptes de résultat et dans les capitaux propres des plus et moins-values latentes, ce qui devrait accroître la volatilité des bilans bancaires. En outre, étant donné que les pertes bancaires limitent le nouveau financement économique pendant une crise, l'évaluation des actifs au prix du

marché introduit une procyclicité qui retarde la résolution de la crise. En conséquence, l'IASB a publié la norme IFRS 9, qui s'est assigné l'obligation de fournir aux utilisateurs d'états récapitulatifs des informations fiables pour l'évaluation des montants, du calendrier et de l'incertitude des flux de trésorerie futurs de la Banque, en tenant compte de toutes les critiques et recommandations formulées à la suite du G20 de 2008. Cette nouvelle norme sur les instruments financiers porte principalement sur trois composantes¹ :

- « Classification et évaluation » : Cette partie offre de nouvelles lignes directrices distinctes de celles déjà en vigueur en vertu d'IAS 39 pour la comptabilisation et l'évaluation des instruments financiers ;
- « Dépréciation » : Cette deuxième section traite de la façon dont les instruments pour risque de crédit sont comptabilisés pour leur dépréciation. Cela signifie un véritable écart par rapport au modèle d'amortissement en vigueur ;
- Le dernier volet de la réforme d'IFRS 9, « Comptabilité de couverture », présente une méthodologie plus simple pour les états financiers de la comptabilité de couverture. L'objectif est de démontrer comment les activités de gestion des risques à l'aide d'instruments de couverture affectent le résultat net.

Section 2 : les techniques de gestion de risque de crédit

Cette section est consacrée pour la présentation de la littérature sur certaines techniques utilisées dans la gestion du risque de crédit, plusieurs techniques sont prises en considération.

1. Le processus de gestion des risques :

La théorie de la gestion des risques, dont le but est d'arrêter les pertes « accidentelles », peut être utilisée pour analyser le risque de crédit auquel les créanciers financiers sont exposés. Ce domaine, qui était à l'origine axé sur les entreprises industrielles et commerciales, offre une liste de lecture utile pour la gestion du risque de crédit. L'objectif principal de la gestion des risques est de protéger le capital économique d'une entreprise et de réduire la susceptibilité de son rendement à d'éventuelles pertes involontaires. En adoptant une approche en quatre phases des risques, elle devrait contribuer à réduire l'exposition. A savoir² :

¹ S. LOTFI et S. BENSAIDA ; **Mesure des impacts de la norme IFRS 9 sur le risque de crédit bancaire** ; Revue du Contrôle de la Comptabilité et de l'Audit ; N°5 ; Juin 2018 ; p 555.

² C. KHAROUBI et P. THOMAS ; op cite ; p 123, p126.

1.1. L'identification des risques :

« Prise de Conscience de l'existence du risque » est le nom de cette étape initiale. Elle cherche à identifier tout risque (menace de circonstance) pouvant entraîner des pertes financières pour une organisation. Concrètement, elle se compose de deux composantes distinctes :

- La cartographie des risques implique l'identification de tous les risques existants dans l'entité (connaissance des risques) ;
- Chaque acteur interne doit être conscient de la présence du risque, ce qui nécessite à la fois des informations et des preuves que de nombreux actes récents et spécifiques sont liés au risque pris (conscience du risque).

1.2. Evaluation des risques :

Il s'agit de chiffrer les conséquences éventuelles en cas de survenance du risque. Pour ceci, il convient de mesurer :

- La probabilité de survenance du risque : p
- Les conséquences, financières, en cas de survenance du risque (appelée précédemment Loss Given Default) : C .

On en déduit alors, pour chaque risque, une estimation du sinistre potentiel : S , qui correspond à la mesure du risque :

$$S = P \times C$$

La probabilité que l'événement se produise est représentée par la variable p , et les caractéristiques de cette distribution et son évolution révèlent des informations sur le profil de risque. Cela peut impliquer de calculer le risque le plus élevé (p et C maximum), ou le risque moyen (p et C moyens).

- Les dépenses associées aux conséquences de l'événement (C) comprennent toutes les dépenses engagées : répercussions financières immédiates : le coût des actions requises pour rétablir la situation où elle était avant que les pertes ne soient générées ;
- Les dépenses de réputation (non incluses dans les risques énumérés par le Comité de Bâle), la responsabilité, les effets sur d'autres activités (contagion), les réactions des parties prenantes, etc. sont des exemples de conséquences financières indirectes.

1.3. La réduction du risque :

Cela implique la mise en place de procédures de gestion interne pour réduire la probabilité que le risque se matérialise et la gravité des conséquences s'il se produit. Il existe deux types différents d'actions coordonnées dans cette phase opérationnelle.

- Réduction préventive : L'objectif de la prévention est de réduire la probabilité qu'un risque se produise, ou sa fréquence. L'objectif est de réduire le risque. Ceux-ci couvrent à la fois les processus internes et le suivi de la mesure des risques, en particulier pour garder un œil sur l'évolution des chances d'occurrence. Il est courant de définir des systèmes d'alerte dont le but est d'avertir les gens sur les risques croissants.
- Réduction curative : En prenant une série de mesures dans le cas où le risque se matérialise, la protection vise à minimiser les dommages dans toute la mesure du possible. Il s'agit d'une phase proactive au cours de laquelle un certain nombre de mesures sont prises selon un échéancier pour réduire les impacts immédiats du risque et prévenir un phénomène de contagion indirecte.

1.4. Traitement du risque résiduel :

Un risque a la propriété qu'il est impossible de minimiser sa probabilité d'occurrence. Par conséquent, des décisions doivent être prises sur le sort du risque résiduel, qui est égal à la probabilité (réduite) d'occurrence multipliée par l'ampleur des effets. C'est ainsi que le risque est traité économiquement.

Trois solutions sont possibles.

- Le risque est réduit. C'est de l'auto-couverture : l'entité prendra en compte les répercussions du risque (structure, réserves, provisions, etc.). Cette solution interne implique que le risque résiduel ne dépasse pas la capacité d'absorption de l'entité, Les scénarios dans lesquels l'entité peut disparaître sont exclus (seuil de vulnérabilité).
- L'assurance couvre le risque. L'organisme paie une prime et conclut un contrat d'assurance. L'assureur fixe les modalités d'indemnisation (délai, plafond, etc.). Si le risque se matérialise, l'assureur comble tout manque à gagner en payant la totalité ou une partie de la perte.
- Il est possible de transférer le risque à une contrepartie. Dans ce cas, l'entité veut une « protection ». Il paie une prime et, si un risque se matérialise, le vendeur de la protection

tiendra l'entité indemnisée conformément aux conditions énoncées dans le contrat. La contrepartie peut être située sur le marché financier ou de gré à gré.

2. Les méthodes de gestion des risques :

Étant donné que le risque de crédit ne peut pas être entièrement éliminé mais peut être réduit grâce à une gestion compétente du risque, la gestion du risque de crédit dans une banque de plus de la réglementation prudentielle comprend :

2.1. La politique des garanties :

Une banque a besoin de garanties pour se protéger en cas de défaut de paiement de l'emprunteur, et dans des conditions variées. A savoir ¹:

➤ Les garanties réelles :

Elles portent sur des biens, sous forme d'hypothèque lorsqu'un immeuble est transféré pour satisfaire une dette, de gage lorsque des biens meubles sont donnés en garantie, soit le débiteur est avec ou sans dépossession, ou sur des actifs incorporels ou encore de nantissements.

➤ Les garanties personnelles :

Ils proviennent d'un tiers à titre de cautionnement ou d'aval. Dans le cas des prêts aux PME, il est très courant que le banquier demande la caution solidaire du ou des administrateurs de la société afin d'éviter de s'opposer à la responsabilité financière limitée des apports et de motiver les dirigeants à gérer l'entreprise avec l'intention de rembourser les créanciers. Cela remonte aussi aux, signature par les entreprises mères de lettres d'intention ou de garanties à première demande pour soutenir leurs filiales.

☞ Les procédures de prise de garantie doivent être explicitement formalisées. Ces procédures comprennent des contrats qui garantissent la sécurité juridique de l'opération, une définition des circonstances dans lesquelles des garanties sont requises et des taux de couverture. En outre, il est reconnu que l'obtention de garanties n'élimine pas complètement le risque de non-remboursement, compte tenu en particulier de la difficulté des procédures d'insolvabilité.

¹ S.De GOUSSERGUES et autres ;8^{ème} édition ; p 279.

2.2. La politique de provisionnement :

Comme nous l'avons déjà indiqué, divers risques sont présents dans les activités quotidiennes des banques et d'autres organisations financières. La Banque gère le risque de crédit au moyen d'un système intégré de politiques de gestion responsable du crédit, conformément à sa philosophie globale de gestion du risque. Ces règlements traitent de choses comme l'admissibilité pour certains pays, les plafonds d'engagement, un montant adéquat de capital, la tarification et la sécurité, les arriérés et le recouvrement, et la provision pour pertes. L'une des principales raisons de la cote de crédit élevée de la Banque est sa structure conservatrice.

Pour les institutions financières, l'octroi de prêts est une procédure comptable courante. Les institutions reconnaissent ouvertement que les pertes sont un effet secondaire attendu de la conduite des affaires lors de l'établissement des provisions¹. Le calcul des provisions s'effectue selon la formule suivante :

$$\text{provision} = \text{proportion} \times (\text{engagement} - \text{garanties réelle} - \text{agios réservé})$$

La proportion dépend de la classe de créances ; la proportion est 0% pour les actifs courants, 20% pour les créances de classe 2, 50% pour les créances de classe 3, et 100% pour les créances de classes 4. La classification des créances s'effectue soit par :

➤ Classification des créances selon le nombre de jour de retard :

A ce stade, il faut définir les catégories de créances qui peuvent être rencontrés dans un portefeuille de crédit, selon le nombre de jours de retard, on trouve : (capture)

- Classe 0 (les actifs courants) : Les actifs circulants sont ceux dont la réalisation en temps opportun ou le recouvrement complet semble assuré.
- Classe 1 (les actifs à surveiller) : Il s'agit d'obligations détenues par des entreprises opérant dans des secteurs difficiles ou dont la situation financière se détériore, et dont l'exécution en temps opportun ou le rétablissement complet est néanmoins garanti. Les paiements d'intérêts ou de capital en souffrance ne peuvent pas être en retard de plus de 90 jours.
- Classe 2 (les engagements douteux) : Il s'agit des engagements dont la réalisation complète ou le recouvrement en temps opportun est douteux, et qui

¹ Banque africaine de développement ; **MISE À JOUR DE LA POLITIQUE DE PROVISIONNEMENT DES PRÊTS ET GARANTIES AU SECTEUR PUBLIC** ; p 1.

sont détenues par des entreprises connaissant des difficultés financières ou autres, qui peuvent mettre en doute leur légitimité et nécessitent la mise en œuvre de mesures correctives. Le délai de grâce pour retard de paiement d'intérêts ou de capital est de 90 à 180 jours.

- Classe 3 (les créances préoccupantes) : Il s'agit des engagements dont la réalisation ou le recouvrement est compromis et qui reposent sur des engagements qui présentent les caractéristiques de la classe 2 avec plus de gravité. Le délai de grâce pour le paiement tardif des intérêts ou du principal est de 180 à 360 jours.
- Classe 4 (les créances compromises) : Cette catégorie comprend les actifs impayés depuis plus de 360 jours, les réclamations en litige et les réclamations pour lesquelles les retards de paiement ont dépassé 360 jours.

➤ Classification par le gel du compte :

Sont pertinents pour le découvert et utilisés pour déterminer les classes 2, 3 et 4. Le découvert est généralement considéré comme gelé et doit être classé dans la catégorie 2 lorsqu'une période de 90 jours s'est écoulée après la suspension des intérêts sans que le compte n'enregistre de mouvements de revenus permettant d'équilibrer le montant total des charges d'intérêts et autres charges. La classe 3 fait référence aux retards qui se situent entre 180 et 360 jours, La classe 4 fait référence aux retards supérieurs à 360 jours.

3. Couverture des risques :

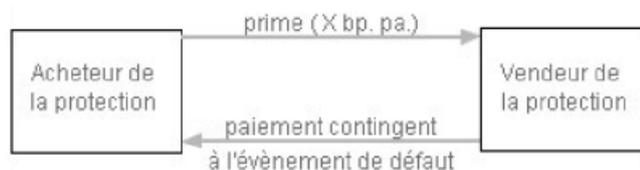
3.1. Couverture sur le marché financier :

Le risque le plus important pour les banques d'investissement ou commerciales est le risque de crédit, car l'une de ses principales responsabilités est d'accorder directement ou indirectement des prêts et des crédits. Ils ont contribué au développement de « dérivés de crédit » pour atténuer ce risque sur les marchés financiers afin de réduire leur exposition. Le volume de ces dérivés a considérablement augmenté au cours des 30 dernières années. Ils servent à la fois de classe d'actifs et d'outil de gestion du risque financier pour les investisseurs.

- Les Credit Default Swaps (CDSs): Dans un CDS, une partie (l'acheteur de la protection) accepte de payer une prime (exprimée en pourcentage du notionnel) sur une base périodique tandis que l'autre partie (le vendeur de protection) accepte de payer, en cas de défaut de l'emprunteur de référence, un montant qui couvrira la perte résultant du

défaut jusqu'à concurrence du montant notionnel. L'acheteur de la protection fait référence à un emprunteur bien défini dans le contrat financier. Le mécanisme CDS est illustré dans la figure ci-dessous¹ :

Figure 4: le mécanisme de CDSs



Source : A. DE SERVIGNY ; op cite ; p 220.

- Les Collateralized Debt Obligations (CDOs) : Les actifs adossés dans les CDOs peuvent inclure des obligations, des prêts des banques aux entreprises ou des dérivés de crédit qui sont eux-mêmes définis sur la base du même sous-jacent, comme un CDS qui protège contre le risque de défaut associé à la signature d'une entreprise. Par conséquent, les CDOs sont soutenus par divers portefeuilles d'actifs ou de produits dérivés qui nécessitent un petit nombre de signatures (environ 500 noms liquides sur le marché des CDS)².
- Titrisation : Il n'est pas toujours nécessaire que l'entreprise ou l'institution financière qui a accordé le prêt assume ce risque. Il peut offrir un portefeuille de prêts à une entreprise, qui le transformera par la suite en un produit financier pour les investisseurs. ce processus est connue sous le nom de titrisation. Ce dernier est depuis longtemps crucial pour les marchés financiers et s'est avéré très utile. Il prend en charge le modèle largement utilisé de originate-to-distribute, par les banques. La titrisation a été cruciale pour le développement de la bulle immobilière. En fait, la possibilité de revendre les prêts consentis par la titrisation a eu un impact sur le comportement des initiateurs³.

3.2. Autre technique de couverture :

De ces techniques on trouve⁴:

- La tarification : Il est important d'évaluer le crédit en fonction du risque qu'il comporte, L'écart facturé doit augmenter à mesure que la probabilité de défaut ou de faillite augmente.

¹ A. DE SERVIGNY ; op cite ; p 220.

² Ibid p 231.

³ J.Hull; op cite; p 348.

⁴ C. KHAROUBI et P. THOMAS ; op cite ; 128.

- La contractualisation : Des clauses restrictives qui diminuent le risque de la banque doivent être incluses dans le contrat de prêt : clauses restrictives (obligations de faire, interdictions, ratios à respecter, contraintes d'information, etc.).
- La « collatéralisation » : Le créancier peut exiger des actifs comme garantie du débiteur. Si l'emprunteur fait défaut, le créancier conserve les actifs collatéraux en paiement de la perte qui en résulte. Il s'agit souvent d'une sûreté réelle ou personnelle pour la banque commerciale, qui peut prendre la forme d'une garantie d'une société mère. Il s'agit de titres négociables ou d'espèces pour des opérations de marché. Le processus juridique de transfert de garantie est complexe et peut être envisagé de diverses manières (gage, cautionnement, fiducie de sécurité, etc.).

Conclusion

Tous les créanciers exposés au risque de crédit sont d'abord invités à le quantifier pour chacun de leurs emprunteurs afin de le gérer efficacement. Afin de minimiser les problèmes et de réduire la probabilité d'une crise systémique, les réglementations prudentielles appliquent également cette mesure, qui sert de base à la définition du montant minimum de fonds propres à détenir.

Afin d'atténuer les effets d'un tel risque de crédit, des approches de gestion sont apparues tant au niveau individuel qu'institutionnel de l'organisation financière.

*Partie
pratique*

Chapitre I

*Méthodologie de construction d'un
modèle de classification*

Introduction

Trouver des modèles efficaces qui fournissent l'estimation la plus précise est l'un des plus gros problèmes dans la prédiction des défauts.

Avec l'introduction de l'apprentissage automatique, les banques sont désormais en mesure de tirer parti et de créer de solides outils de traitement informatique qui permettent de construire des modèles prédictifs qui permettent aux banques de prévoir la possibilité de défauts des clients, ce qui leur permet une gestion efficace de cela.

Plusieurs modèles d'apprentissage automatique ont été mis en avant pour y parvenir, mais en ce qui concerne l'estimation des défauts, la question de savoir quel type d'apprentissage utiliser est toujours ouverte.

Ainsi dans ce chapitre nous abordons les différents types d'apprentissage, et les modèles utilisés dans notre travail, ce chapitre est segmenté en deux sections, dans un premier lieu on procédera à définir le machine learning et les types d'apprentissage, par la suite des différents modèles utilisés seront présentés avec une description du processus de prétraitement des données adaptés.

Section 1 : revue de littérature

Premièrement, il est important de définir l'aspect théorique des méthodes utilisées, de distinguer ces différentes notions ainsi que leurs utilisations.

1. Crédit scoring :

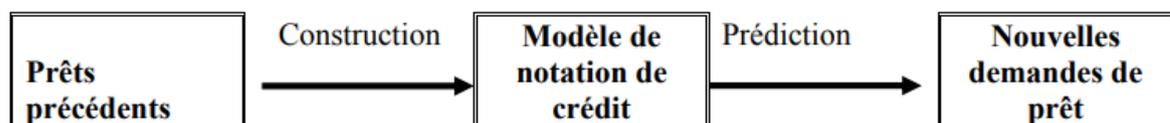
Une présentation des principes généraux de crédit scoring est nécessaire.

1.1. Définition :

En donnant un poids à chaque information indiquant la solvabilité d'un emprunteur, la score de crédit est une technique qui vise à condenser le risque de contrepartie. Par rapport aux scores de seuil précédemment définis, la pondération totale permet un jugement rapide pour approuver ou refuser une demande de crédit¹.

Pour élaborer un modèle de score, la performance historique des emprunteurs ou l'historique des prêts qui leur sont accordés, ainsi que des approches statistiques, sont souvent utilisées pour prédire la performance des prêts futurs présentant des caractéristiques comparables. Ces renseignements sont recueillis à partir des dossiers de crédit des clients ou d'autres sources. Pour ce faire, les modèles de score doivent d'abord être en mesure d'identifier les facteurs de risque les plus critiques, c'est-à-dire ceux qui influencent le plus la probabilité de défaut d'un emprunteur, puis d'évaluer la contribution relative de chaque élément au risque de défaut². Et donc le processus de scoring peut être résumé comme la figure ci-après :

Figure 5: le processus de scoring



Source : B. Diallo, UN MODELE DE "CREDIT SCORING" POUR UNE INSTITUTION DE MICRO-FINANCE AFRICAINE : LE CAS DE NYESIGISO AU MALI, Laboratoire d'Economie d'Orléans (LEO), Université d'Orléans, may 2006.

¹ S. DE COUSSERGUES, et autres, 9^{ème} édition, op-cite, P 191.

² M.DIETSCH et J.PETE, op-cite, p 50.

1.2. Les éléments requis pour la construction du modèle score :

Pour élaborer un modèle de score, il est important d'identifier un échantillon qui comprend deux types d'emprunteurs, les emprunteurs défaillants et les emprunteurs sains, et pour construire cet échantillon, divers éléments doivent être définis, à savoir ¹ :

➤ Le critère de défaut :

Il existe deux types de cet événement dans l'analyse du risque de crédit : il peut s'agir d'un défaut, ce qui signifie que le client est incapable de remplir ces obligations ou qu'il y a un retard de paiement de plus de 90 jours. Il peut également s'agir d'une faillite, ce qui signifie que le premier jugement est rendu dans le cadre d'une affaire de faillite légale.

➤ La construction de l'échantillon :

Comme déjà mentionné, afin d'assurer la qualité du score, l'échantillon doit comprendre deux sous-échantillons : l'un qui a connu le critère de défaut, et l'autre échantillon, n'a pas connu le critère de défauts. De plus, les données historiques collectés doivent couvrir l'ensemble d'un cycle économique.

➤ La définition d'un horizon temporel :

Il est nécessaire de définir un horizon pour le modèle, ce qui permettra aux données traitées de remonter à une époque historique antérieure à la faillite plus ou moins prolongée. L'horizon utilisé dépend par les modèles générés et la disponibilité des données.

➤ Le choix des variables explicatives :

Il est parfois difficile d'identifier ces variables, car elles dépendent principalement de l'information disponible, car les variables explicatives choisies doivent refléter divers aspects du critère du défaut. Pour la création de modèles de score, une variété de variables peut être utilisés. Il s'agit notamment de variables comptables sous forme de ratios financiers, de données comportementales et de variables qualitatives.

¹ C.KAROUBI, et P.THOMAS, op-cite, p 82.

- Le choix de la technique utilisée :

Les modèles optimaux de score de crédit doivent être trouvés en fonction des attributs et des traits des variables explicatives retenues. Pour ce faire, des approches économétriques, des techniques de classification et même des techniques d'intelligence artificielle peuvent être utilisées.

- La validation des modèles construites :

Pour garantir l'efficacité des modèles et leur capacité à identifier avec précision l'occurrence du défaut, les approches de validation convenues devraient être adoptées.

1.3. L'utilisation du modèle score :

Le modèle de score est principalement utilisé pour évaluer toute nouvelle demande de crédit, donc grâce à une évaluation du risque du demandeur, et selon la méthode utilisée, soit classer le client ou attribue un score au même client, ce score peut être comparé chronologiquement pour se renseigner sur l'évolution du risque.

Le score de crédit peut également être appliquée largement, par exemple, à un portefeuille d'entreprises, créant ainsi un tableau de bord pour une clientèle. L'application d'un score à une population importante permet un suivi macroéconomique, montrant l'évolution de la situation générale des entreprises et le risque économique¹.

1.4. L'apport de machine learning pour le crédit scoring :

L'utilisation de nouvelles données qui peuvent mieux tenir compte du risque de crédit est le principal avantage de l'apprentissage automatique. Les nouveaux comportements des clients, la digitalisation des interactions clients, ou encore la disponibilité de nouvelles sources d'information comme les réseaux sociaux contribuent à cette diversité de données. Et grâce aux algorithmes les techniques d'apprentissage automatique améliorent les performances de prédiction des modèles de score de crédit, tout comme la diversité des données².

Les techniques d'apprentissage automatique peuvent être utilisées d'innombrables façons en amont de la création d'un modèle, en plus des avantages en termes d'amélioration des performances et de diminution des erreurs d'estimation. Les acteurs sont encouragés à utiliser

¹ Ibid, p84

² C.Hurlin, et C.pregnon, **MACHINE LEARNING ET NOUVELLES SOURCES DE DONNÉES POUR LE SCORING DE CRÉDIT**, Revue d'économie financière, N°135, mars, 2019.

ces nouvelles techniques dans le but d'augmenter la rentabilité et de réduire les coûts en raison de la disponibilité accrue des données et de la puissance croissante des solutions technologiques¹.

2. Machine learning :

2.1. Définition de machine learning :

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 »².

L'apprentissage automatique est une branche de l'intelligence artificielle dont le but est de permettre aux ordinateurs d'apprendre par eux-mêmes. En termes simples, un algorithme d'apprentissage automatique peut reconnaître des modèles dans les données observées, générer des modèles qui décrivent le monde et prédire les choses sans avoir besoin de règles et de modèles préprogrammés explicites³.

2.2. Les types d'apprentissage en machine learning :

Nous employons des méthodes dites d'apprentissage, qui sont fortement influencées par la façon dont nous, les humains, apprenons à faire les choses, pour donner à la machine la capacité d'apprendre. Bien qu'il existe de nombreux types d'apprentissage automatique, ils peuvent être divisés en trois groupes:

➤ L'apprentissage supervisé :

C'est la méthode d'apprentissage automatique la plus largement utilisée, l'apprentissage supervisé, consiste à alimenter l'ordinateur d'un ensemble de facteurs appelés variables explicatives (features) ou prédicteurs et d'un autre ensemble de variables, généralement constitué d'une seule variable, appelée variables cibles (label) ou à prédire. En fait, l'objectif

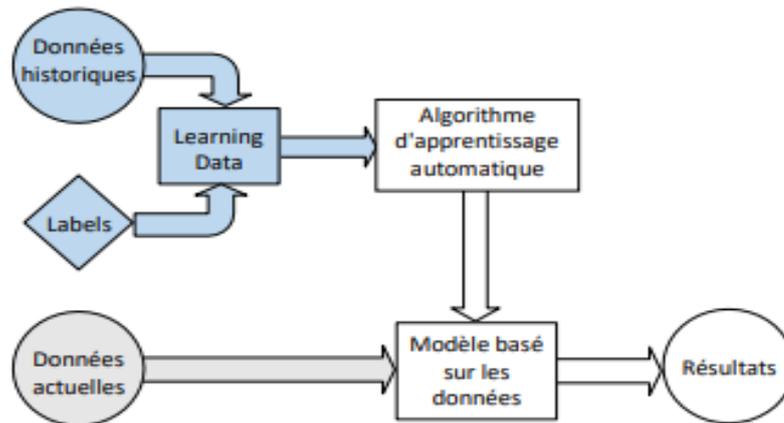
¹ N.Grennepois, et autres., **Machine learning and credit risk: a situable marriage ?** , risk advisory august 2019, p1

² G.saint-cirgue, **apprendre le machine learning en une semaine**, 2019, p11.

³ F. Francisco et A.Caixaero, **Infrastructure and Machine Learning for Credit Scoring**, memoire pour l'obtention d'un master en informatique et telecommunication, técnico lisboa, novembre 2018.

est de développer un modèle en utilisant des techniques d'optimisation pour comprendre la relation entre les variables explicatives et la variable à prédire¹.

Figure 6: le processus de l'apprentissage supervisé



Source : F. YALCOUE et A.ADNANE, *machine learning pour la maintenance prédictive*, juillet 2021, p15.

Deux techniques de prédiction sont réalisables du point de vue de l'apprentissage automatique supervisé : les techniques basées sur des algorithmes de régression et les techniques basées sur des algorithmes de classification.

- Les algorithmes de régression :

Ces algorithmes sont utilisés, lorsque la variable de sortie est une valeur continue, donc ici la machine doit prédire la valeur d'une variable quantitative.

- Les algorithmes de classification :

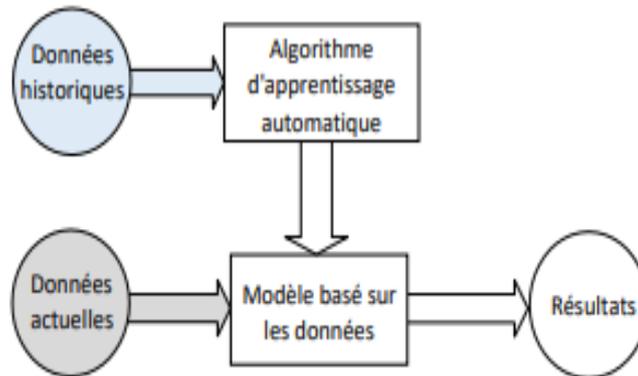
Ces méthodes sont utilisées pour déterminer à quel groupe de catégories appartient une nouvelle observation lorsque la variable à prévoir est une variable qualitative. Les techniques de classification des données peuvent être basées sur des recherches dans des espaces d'hypothèses, des idées de proximité ou des hypothèses probabilistes.

¹ A. SAMBA, **Science des données au service des réseaux d'opérateur**, thèse de doctorat, l'école nationale supérieure mines-télécom 2018, p 13.

➤ L'apprentissage non supervisé :

Dans ce type d'apprentissage, la valeur à prédire n'est pas disponible dans le jeu des données. A ce stade l'algorithme utilise un jeu de données non étiquetées c'est à dire sans la valeur cible.

Figure 7 : le processus de l'apprentissage non supervisé



Source : F. YALCOUE et A.ADNANE, op cite, p16.

Dans ce type d'apprentissage, les machines apprennent elles-mêmes à reconnaître les similitudes et les différences dans les données, ainsi qu'à regrouper ou à catégoriser des objets qui partagent des traits similaires, le tout sans supervision humaine. Les tâches non supervisées les plus courantes sont les suivantes :

- Clustering (classification non supervisé) :

Cette méthode regroupe les données non étiquetées en groupes présentant des caractéristiques connexes ; Le défi avec cette méthode est de décider quelles caractéristiques doivent être prises en compte lors du regroupement des données. Cette méthode est très utile, pour la segmentation des consommateurs dans le marketing.

- La détection d'anomalies :

Ici la machine analyse la structure des données, trouve les échantillons dont les caractéristiques sont très éloignées de celles des autres échantillons

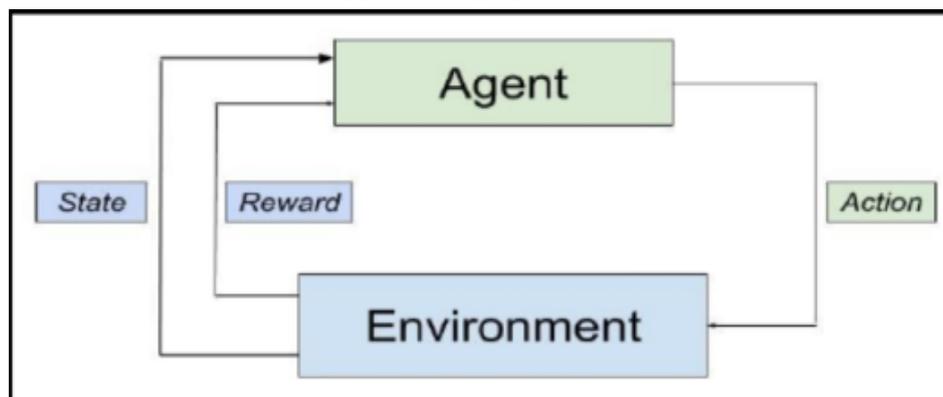
- La réduction de la dimensionnalité :

Afin d'exécuter ce processus, la machine apprend réellement à projeter nos données du mieux qu'elle peut dans des endroits plus petits. En étudiant les données, l'ordinateur est capable d'apprendre à simplifier cette structure tout en conservant les résultats importants.

➤ L'apprentissage par renforcement :

Au lieu de recevoir un ensemble de données étiquetées, comme dans le scénario d'apprentissage supervisé, la machine dans ce cas reçoit des données d'apprentissage directement de l'environnement. Il convient de noter que l'environnement dans ce cas peut être réel ou un simulateur. L'agent reçoit deux informations différentes concernant sa situation actuelle dans l'environnement dans ce cas, ainsi qu'une récompense pertinente pour la tâche à accomplir et son résultat escompté. L'agent veut choisir le plan d'action optimal afin de maximiser les avantages offerts par l'environnement. Récompenser un comportement ou retarder la punition est un élément crucial de l'apprentissage par renforcement. Le scénario général de l'apprentissage par renforcement est illustré par la figure suivante :

Figure 8: l'apprentissage par renforcement



Source : S.SADI, *Implantation et évaluation d'un modèle d'apprentissage automatique pour l'estimation de la valeur marchande*, 2020, p19.

Section 2 : méthodologie de traitement des données

Dans cette section les différents modèles utilisés seront exposés, avec une définition au préalable de la démarche de préparation de donnée.

1. Les bibliothèques clés de machine learning :

Pour l'implantation de nos modèles, nous avons utilisé comme langage de programmation python, puisqu'il est facile à apprendre, c'est le langage de programmation le plus utilisé de nos jours.

Python développé à l'origine par Guido van Rossum en 1991 (Pays Bas), est une licence open source, ce qui le rend librement utilisable et distribuable, il est utilisé avec succès dans des milliers d'applications commerciales réelles à travers le monde, la version python utilisée dans ce travail est la 3.9.

Ce langage de programmation dispose d'une variété de bibliothèques et modules qui facilite l'implantation des modèles, parmi les bibliothèques utilisées dans ce travail, on note :

➤ Numpy :

Numpy est l'abréviation de Numerical python, est la librairie qui permet de créer et manipuler des matrices simplement et avec efficacité. En Machine Learning, on insère le plus souvent notre base de données dans des matrices. Donc Numpy fournit à Python une vaste bibliothèque mathématique capable d'effectuer des calculs numériques de manière efficace et efficiente afin de pouvoir travailler avec des tableaux multidimensionnels et des structures de données matricielles.

➤ Pandas :

Pandas est l'abréviation de panel data, est une excellente librairie pour importer des tableaux dans Python dans le but de tirer des statistiques, Pandas intègre deux structures de données supplémentaires dans Python, à savoir Pandas Series et Pandas DataFrame. Ces structures de données nous permettent de travailler avec des données étiquetées et relationnelles de manière simple et intuitive. Elle aide aussi à la manipulation et l'analyse des données.

➤ Matplotlib :

Matplotlib est la librairie de traçage qui permet de visualiser les données, et les résultats sous forme de graphes, courbes et nuages de points.

➤ Seaborn :

Seaborn est une bibliothèque qui permet la visualisation des données, basé dans matplotlib, elle fournit une interface de haut niveau pour dessiner des graphiques statistiques attrayants et informatifs.

➤ Scikit-learn :

Scikit-learn est une bibliothèque clé construite à partir de NumP, SciPy et Matplotlib, et qui se concentre sur les outils d'apprentissage automatique, y compris les algorithmes mathématiques, statistiques et à usage général qui constituent la base de nombreuses technologies d'apprentissage automatique. En tant qu'outil gratuit, Scikit-learn est extrêmement important dans de nombreux types de développement d'algorithmes pour l'apprentissage automatique et les technologies connexes.

2. L'environnement de développement pour python :

De nombreux outils logiciels sont créés pour simplifier le processus difficile de l'analyse des données et pour fournir des environnements intégrés qui permettent l'utilisation de tous les algorithmes existants. Les environnements de programmation Python incluent pycharm, spider, jupyter notebook et google colab.

Comme environnement de développement pour notre travail, nous avons utilisé Google Colab Notebook, qui est basé sur Jupyter Notebook et nous permet de former des modèles d'apprentissage automatique directement dans le cloud sans avoir à installer quoi que ce soit sur notre ordinateur. Et nous avons utilisé pycharm pour le développement de notre application.

3. Préparation de la base données :

Sans aucun doute, la qualité des données utilisées dans un modèle d'apprentissage automatique est tout aussi importante que leur quantité en matière de performances. En règle générale, les bases de données utilisées pour stocker les données ne sont pas propres et contiennent souvent des anomalies et des erreurs susceptibles de ruiner l'apprentissage automatique. Par conséquent, il y a quelques tâches à terminer afin de préparer correctement la base de données :

- Nettoyage et préparation des données :

La correction des valeurs manquants, l'élimination des valeurs aberrantes et la conversion des variables catégorielles en valeurs numériques sont toutes importantes pour que l'apprentissage automatique soit plus rapide et plus efficace. Sans oublier de normaliser les données en mettant toutes les valeurs sur la même échelle.

- Sélection des variables :

Trouver les variables les plus utiles pour notre modèle d'apprentissage automatique parmi toutes les variables de notre base de données est nécessaire, car fournir un modèle avec trop de

variables, y compris des variables qui ne transportent pas l'information, peut affecter la qualité de son apprentissage. Pour nous aider à choisir les variables les plus pertinentes, nous avons utilisé des tests de dépendance :

- Test khi 2 :

Il est utile pour tester l'indépendance entre deux variables qualitatives, on a utilisé pour tester l'indépendance entre les variables qualitative et la variables cible, il repose sur les hypothèses suivantes :

H0 : variables indépendantes si $p\text{-value} > \alpha$

H1 : variables non indépendantes si $p\text{-value} < \alpha$

α Présente un seuille de prédiction, il est généralement égal à 5%.

- Le test Anova :

De la même logique que test Khi 2, nous avons utilisé le test Anova pour tester la dépendance entre les variables quantitatives et la variables cible. Les hypothèses de test Anova :

H0 : variables indépendantes si $p\text{-value} > \alpha$

H1 : variables non indépendantes si $p\text{-value} < \alpha$

- Matrice de corrélation :

Et à l'aide de la matrice de corrélation qui se présente sous d'un tableau où chaque cellule de tableau donne la corrélation entre toutes les variables de la base de données, nous avons éliminé l'une des variables fortement corrélées et qui a le pouvoir discriminant le moins élevés.

- Traitement de déséquilibre des classes :

Une distribution déséquilibrée des classes dans une base de données est typique ; Dans ce scénario, le biais du modèle de prédiction est évident, favorisant la catégorie majoritaire et classant mal les cas appartiennent aux classes minoritaires. Il existe de nombreuses méthodes de suréchantillonnage des classes minoritaires ou de sous-échantillonnage des classes majoritaire pour résoudre ce problème.

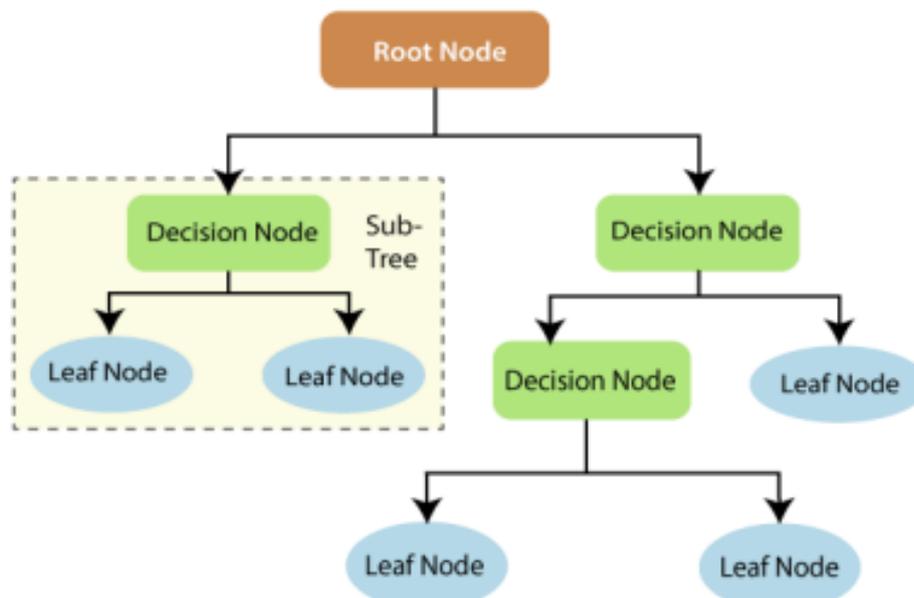
4. Les modèles prédictifs :

Parce que les données historiques sont accessibles, les approches d'apprentissage supervisées sont les plus couramment utilisées dans les domaines de la gestion du risque de crédit, et parce que nous nous concentrons sur le classement des clients, nous avons utilisé des modèles de classification.

4.1. L'arbre de décision :

L'une des techniques efficaces fréquemment utilisées dans divers domaines, notamment l'apprentissage automatique, le traitement d'images et la reconnaissance de modèles, sont les arbres de décision. Un ensemble de tests fondamentaux est combiné de manière efficace et fiable par des DT (decision tree), où chaque test compare une propriété numérique à une valeur de seuil. La méthode d'arbre de décision la plus populaire est CART (Classification and Regression Tree), qui a une structure similaire à un arbre avec des nœuds, des branches et des feuilles comme illustré dans l'image ci-dessous¹ :

Figure 9: l'arbre de décision



Donc l'arbre est constitué d'un :

- Un nœud racine (**Root Node**) : qui contient toutes les observations de notre base

¹ T.Bahzad, et A. Mohsin Abdulazeez , **Classification Based on Decision Tree Algorithm for Machine Learning**, journal of applied science and technology trends, vol2, N°1, mars 2021, p 21.

- Un nœud non terminal (**Decision Node**) : qui représente un test sous forme de questions sur nos caractéristiques qui les découpe en deux parties : les individus vérifiant la condition du nœud sont affectés à la branche gauche et les restants à la branche droite
- Un sous arbre (**sub-tree**) : qui est tout arbre obtenu en élaguant arable initiale obtenue en supprimant d'un nombre arbitraire de ses nœuds non terminaux
- Les feuilles : (**Leaf Node**) : qui représentent une classification ou une décision finale.

Le processus d'entraînement consiste à créer une collection de feuilles de calcul faibles qui sont ajoutées séquentiellement dans le but de segmenter l'espace des entités afin de catégoriser les données. Ensuite, cette collection de classeurs est représentée sous la forme d'une collection de nœuds interconnectés et organisés hiérarchiquement. Le résultat final est un arbre qui représente l'ordre des classeurs en insérant le nœud correspondant au classeur ayant la capacité de classification la plus élevée en haut et en descendant à partir de là, affichant d'autres classeurs faibles par ordre décroissant d'importance. Il existe de nombreux critères pour sélectionner la meilleure séparation d'un nœud, mais l'indice de Gini est le plus utilisé¹.

Les arbres de décision ont des modèles de résolution rapides et simples, les données ne nécessitent pas beaucoup de traitement, mais à mesure qu'elles s'approfondissent, elles peuvent parfois devenir très complexes.

4.2. La régression logistique :

C'est un classificateur probabiliste utilisé en apprentissage supervisé, la régression logistique est un modèle qui vise à prédire la probabilité d'un événement en fonction des valeurs de variables indépendantes, catégorielles ou numériques. Il peut être binaire ou multinomial.

Former un classificateur capable de porter un jugement binaire sur la classe d'une nouvelle observation d'entrée est le but de la régression logistique binaire. Ici, nous présentons le classificateur sigmoïde qui aide à prendre cette décision.

En apprenant un terme de biais et un vecteur de pondération à partir d'un ensemble d'apprentissage, la régression logistique est capable d'effectuer le travail de décision. Chaque pondération w_i est un entier réel qui est associé à une caractéristique particulière de l'entrée x_i qui affiche les variables explicatives. L'importance de cette caractéristique d'entrée pour le

¹ J.GONZALEZ, **EXPLORATION DES ARBRES DE DÉCISION ET DES SUPPORT VECTOR MACHINES EN VUE D'APPLICATIONS DANS L'ANALYSE DE TEXTE**, mémoire pour l'obtention de la maîtrise en mathématique et informatique appliqué, université du québec, juin 2016, p58.

choix de classification est indiquée par le poids w_i . Le terme biais, également connu sous le nom d'interception, fait référence à l'ajout d'un deuxième nombre réel aux entrées pondérées.

Après avoir appris les poids à l'entraînement, le classificateur multiplie chaque x_i par son poids w_i , résume les attributs pondérés et ajoute le terme biais b pour porter un jugement sur un échantillon de test. Le nombre unique z qui en résulte exprime la somme pondérée des preuves pour la classe¹.

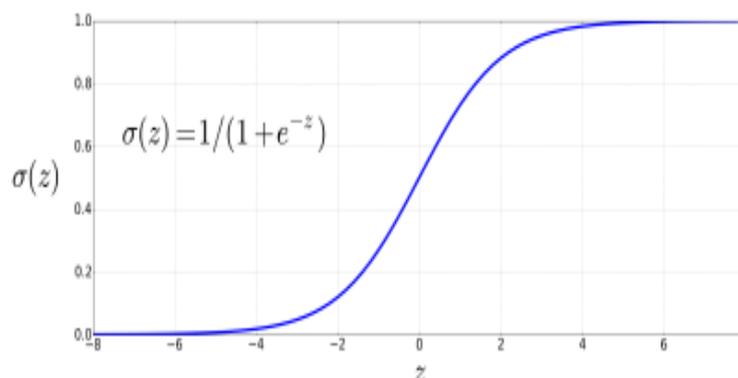
$$z = \left(\sum_{i=1}^n x_i w_i \right) + b$$

Ici la valeur z peut prendre une valeur entre $(-\infty, +\infty)$, puisque les pondérations sont des valeurs réelles, cette valeur z est transformée ensuite en une probabilité à l'aide de la fonction sigmoïde, pour avoir la fonction suivante :

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

La fonction sigmoïde prend une valeur réelle et la mappe à la plage $[0,1]$, comme le montre la figure suivante :

Figure 10: la fonction sigmoïde



D.JURAFSKY et H.MARTIN, « logistic regression », décembre 2021, p 3

La fonction sigmoïde aide à prendre les décisions, une valeur proche de 1, signifie, l'évènement à une grande chance d'être réalisé, une valeur proche de 0, signifie une faible probabilité que l'évènement sera réalisé.

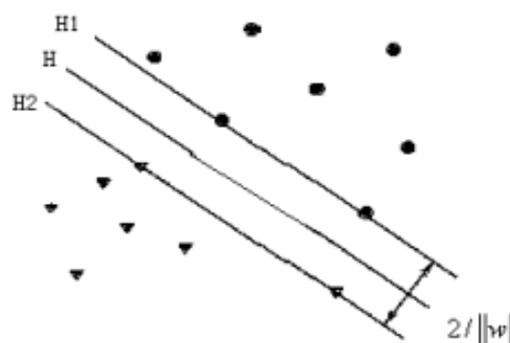
¹ D.JURAFSKY et H.MARTIN, « logistic regression », décembre 2021, p 3

Bien qu'elle s'agisse de l'une des techniques de classification les plus fiables, la régression logistique présente certaines limites, notamment la nécessité de nombreux ensembles de données prétraitées et de variables linéairement indépendantes. Elle est également simple à comprendre et à analyser.

4.3. Les machines à vecteur de support (SVM) :

L'algorithme SVM classe les données en localisant l'hyperplan optimal qui sépare tous les points de données d'une classe de ceux de l'autre. Les SVM classent les données en deux groupes à l'aide d'une approche géométrique. Ils créent un plan qui sépare les données dans chacune des catégories en modélisant les données sous forme de points (vecteurs) dans l'espace. Une fois la limite de décision établie, le SVM sera en mesure de catégoriser les nouvelles données en observant de quel côté de la frontière elles se situent¹.

Figure 11 : les machines à vecteur support



Le concept de base peut être utilisé ci-dessus, Ce modèle est une représentation bidimensionnelle de la surface optimale résultant du cas séparable linéairement. Le concept principal est représenté dans la figure. H sépare deux types sans erreur. H1, H2 et d'autres endroits passent le point le plus récent de H. L'intervalle de classe a été défini comme la séparation entre H1 et H2. En plus d'assurer une séparation sans erreur des deux types de données, la meilleure surface de séparation est celle qui optimise la distance entre le SVM et les points. En d'autres termes, le SVM sélectionne l'hyperplan qui a la distance maximale de tous les points parmi tous ceux qui pourraient éventuellement diviser les échantillons².

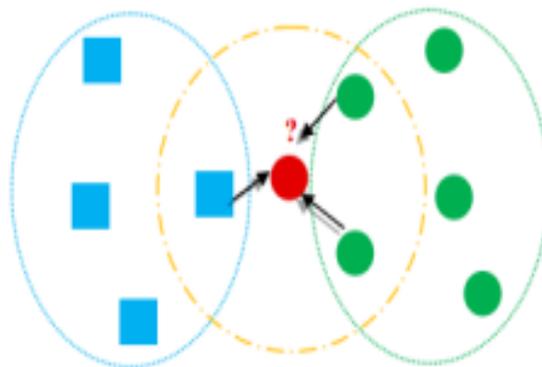
¹ J.GONZALEZ, op-cite, p 9.

² Y.ZHANG, **support vector classification algorithm and its application**, springer, 2012, p 180.

4.4. K-plus proche voisins :

L'une des méthodes d'apprentissage automatique les plus simples est l'algorithme K-Nearest Neighbor (KNN). En réalité, il s'agit d'une technique basée sur la mémoire avec une idée sous-jacente relativement simple qui stocke tous les exemples et évalue les nouveaux cas en fonction du degré de similitude. Avec une fonction de distance d , un entier k et un ensemble de données d'entraînement $D (x_1, x_2, \dots, x_n)$. La procédure recherche dans D les k points les plus proches d'un nouveau point de test x en fonction de la distance d et affecte à x la classe la plus répandue parmi ses voisins, comme le montre la figure ci-après¹ :

Figure 12 : le modèle K-plus proche voisins



Par conséquent, pour classer le cercle rouge, si nous définissons $k = 3$ dans ce cas, l'algorithme recherche les trois plus proches voisins du cercle rouge afin de le catégoriser soit dans la classe des cercles, soit dans la classe des carrés. Dans ce cas, un carré et deux autres cercles constituent les trois plus proches voisins du cercle rouge. En conséquence, l'algorithme inclura le cercle rouge dans la classe des cercles.

4.5. Adaboost :

L'algorithme Adaboost, fait référence à la technique d'ensemble d'apprentissage le Boosting. L'idée principale derrière le boosting est de construire une famille d'estimateurs qui sont ensuite agrégés par un vote majoritaire pour les modèles de classification après avoir abouti à de nombreux modèles raisonnablement fiables. Chaque fois qu'une estimation est construite,

¹B. FUEL CELLS et autres, **advanced in renewable energies and power technologies**, ELSEVIER, Netherlands, 2018, p445.

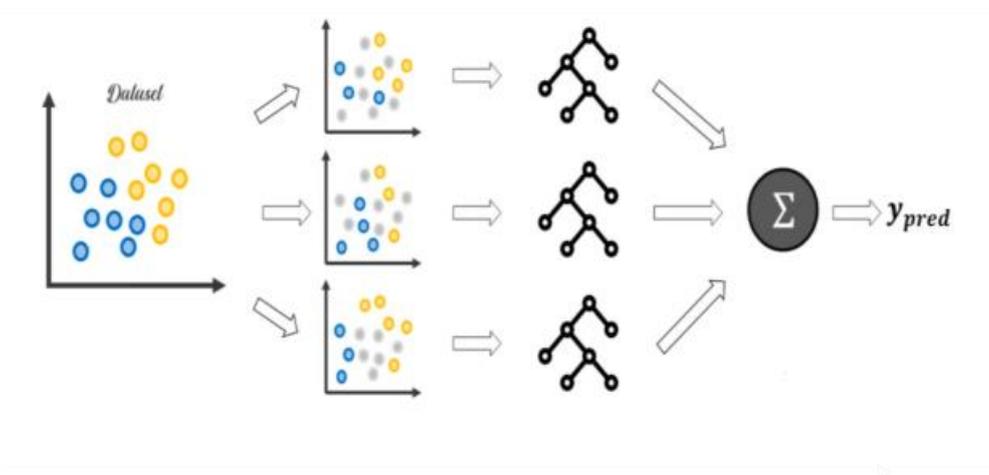
elle améliore celle qui la précède en donnant plus de poids aux observations qui ont été mal prédites ou ajustées. Par conséquent, l'estimateur créé à l'étape k se concentrera sur les données qui ont été mal ajustées par l'estimation de l'étape $k-1$ ¹.

L'idée de boosting a été affinée par Freund et Schapire, qui ont décrit l'algorithme original Adaboost (Adaptive Boosting), cet algorithme fut considéré comme la meilleure méthode puisqu'il ne nécessite pas un long prétraitement des données.

4.6. La forêt aléatoire :

Pour résoudre le problème de l'ajustement excessif des données dans les algorithmes d'arbre de décision, le modèle forêt aléatoire a été introduit, il s'agit d'une collection d'arbres de décision où chaque arbre est légèrement différent des autres. L'algorithme de forêt aléatoire est le plus couramment utilisé dans l'apprentissage des techniques d'ensemble, le Bagging ou le Bootstrap aggregating, qui consiste à créer plusieurs entités du même modèle, la figure ci-dessous montre la technique de bagging² :

Figure 13: le modèle de la forêt aléatoire



5. Les techniques d'évaluation :

L'évaluation des modèles est une étape importante après leur construction. L'étape de l'évaluation est essentielle pour mesurer la performance et la qualité de l'apprentissage de ces modèles, la robustesse et la fiabilité des résultats, et leur capacité à prévoir de nouvelles conditions réelles. Parmi les outils de l'évaluation nous trouvons :

¹ R.SCHAPIRE, *the boosting approach to machine learning an overview*, December, 2001, p3.

² L.ROUVIERE, *introduction aux méthodes d'agrégation : boosting, bagging, et forêts aléatoires*, Université Rennes 2, p 19.

5.1. La matrice de confusion :

Il s'agit d'une matrice de 2 lignes et 2 colonnes pour notre cas, elle aide à visualiser les résultats des modèles. Dans cette matrice on croise les classes cibles réelles avec les classes prédites obtenues. Ceci nous donne le nombre d'observation correctement classées et mal classées.

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

- Les vrais positifs (TP) : est le nombre d'observation positives correctement classées.
- Les faux positifs (FP) : est le nombre d'observation négatives et qui sont prédites comme positives.
- Les faux négatifs (FN) : est le nombre d'observation positives classées comme négatives.
- Les vrais négatifs (TN) : est le nombre d'observation négatives correctement classées.

5.2. Le rapport de classification :

Il s'agit de l'une des mesures d'évaluation des performances d'un modèle d'apprentissage automatique basé sur la classification. Il affiche la précision du modèle à travers quelques paramètres à savoir :

- Précision : c'est le nombre des positifs correctement prédites par rapport au nombre totale des positifs.

$$precision = \frac{VP}{VP + FP}$$

- Recall (sensibilité) : c'est le nombre d'observation correctement prédites, par rapport à l'ensemble des observations de la classes.

$$recall = \frac{VP}{VP + FN}$$

- F1_score : le f1 est la moyenne pondérée de précision et de recall. Ce score prend à la fois les faux positifs et les faux négatifs, Plus la valeur du score f1 est proche de 1, meilleures sont les performances attendues du modèle.

$$f1_{score} = \frac{2 \times (precision \times recall)}{(precision + recall)}$$

- Accuracy (taux de succès) : c'est la prédiction correcte par rapport au nombre total des prédictions.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

5.3. La courbe d'apprentissage :

Une courbe d'apprentissage est un graphique de la performance d'apprentissage du modèle au fil de l'expérience. Les courbes d'apprentissage sont un outil de diagnostic largement utilisé dans l'apprentissage automatique pour les algorithmes qui apprennent d'un jeu de données. Le modèle peut être évalué sur l'ensemble de données d'apprentissage et sur un jeu de données de validation après chaque mise à jour pendant l'entraînement des graphiques de la performance mesurée peuvent être créés pour montrer les courbes d'apprentissage. Dans la courbe d'apprentissage nous trouvons :

- Courbe d'apprentissage de l'entraînement (train learning curve) : courbe d'apprentissage calculée à partir de l'ensemble de données d'apprentissage qui donne une idée de la qualité de l'apprentissage du modèle.
- Courbe d'apprentissage de validation (validation learning curve) : courbe d'apprentissage calculée à partir d'un jeu de données de validation qui donne une idée de la façon dont le modèle se généralise.

5.6. La courbe ROC (Receiver Operating Characteristic) :

La courbe ROC est l'affichage graphique de la sensibilité (TPR) sur l'axe y et (1 – spécificité) (FPR) sur l'axe des x pour faire varier les seuils des valeurs de test.

L'aire sous la courbe (AUC) est une mesure efficace et combinée de la sensibilité et de la spécificité pour évaluer la validité inhérente d'un test diagnostique. AUC maximale = 1 et cela signifie que le modèle est parfait. En outre, l'AUC peut également être utilisée pour trouver la valeur seuil optimale pour un test spécifique, ainsi que Comparer les performances entre deux ou plusieurs tests alternatifs.

6. La validation des modèles :

La validation est une étape importante dans le processus d'apprentissage automatique, elle contribue à l'amélioration de la robustesse des modèles pour éviter des problèmes de généralisation, parmi les méthodes de validation, nous avons utilisés la validation croisée. Cette technique sert à diviser l'échantillon d'apprentissage en certain nombre de sous-ensembles k , et faire entraîner le modèle sur l'ensemble de $K-1$ sous-échantillons restants, puis tester la performance sur les sous_échantillons mis à part, et regrouper le score des échantillons tests pour avoir le score final, du modèle.

7. L'optimisation des hyperparamètres des modèles :

Il s'agit de modifier les paramètres d'ajustement des différents algorithmes, pour améliorer leurs performances, nous avons utilisé la recherche par quadrillage (Gridsearch CV), nous avons constitué une liste des valeurs possibles pour chaque hyperparamètres, donc ici le modèle va entraîner sur toutes les combinaisons, à travers divers essais. L'objectif de la recherche par quadrillage est de conserver le bon paramètre pour le modèle.

Conclusion

Dans ce chapitre, nous avons envisagé, l'aspect théoriques des différents modèles utilisés, avec l'éclaircissement du fonctionnement de chaque modèle et la description du fondement de machine Learning et son apport au crédit scoring.

De même nous avons présentés l'étape de préparation des données, et nous avons donné l'ensembles des techniques d'évaluation et de validation des modèles utilisées. Pour avoir le modèle le plus adéquat, pour prédire le mieux le défaut des clients.

Chapitre II

*L'implantation des modèles de
classification basés sur l'apprentissage
supervisé*

Introduction

Après avoir fait un tour d'horizon, à travers une investigation théorique expliquant les divers concepts liés, notamment à la gestion du risque de crédit et sa mesure à partir des modèles statistiques qu'elles sont souvent utilisés dans les méthodes de scoring crédit, et des modèles de machine learning.

Une approche sera menée, dans le cadre de ce chapitre, pour essayer de mettre en parallèle ces éclairages à la réalité du terrain. En effet, une démarche de l'implantation des différents modèles de l'apprentissage supervisés proposés afin de pouvoir réaliser une application web pour la prédiction des défauts. Une comparaison des différents modèles a été effectuée pour choisir le meilleur modèle.

Dans une première section le processus de prétraitement des données sera présenté. La deuxième section sera consacrée aux analyses des différents résultats obtenus, et au choix du meilleur modèle.

Section 1 : préparation des données

Cette section est consacrée pour présenter l'échantillon de l'étude et décrire le processus de préparation des données.

1. Choix et description des données de l'échantillon :

Nous avons centré pour la construction de notre échantillon décidé aléatoirement, sur des entreprises privées domiciliées auprès du crédit populaire d'Algérie, qui ont bénéficié de nouveaux financements ou de renouvellement au cours de la période 2017-2022.

Notre échantillon regroupe 387 entreprises, se rapportant à l'historique de paiement de crédits bancaires des PME, chaque entreprise est décrite à l'aide de 29 caractéristiques (variables). Nous avons pu recueillir des feuilles de dépouillement du bilan, afin de calculer les ratios et déterminer les informations qualitatives. Nous n'avons pas ciblé un secteur d'activité bien déterminé, les entreprises appartiennent à des secteurs différents (industriel, commercial, service ...)

Il y a quelques réserves au sujet de l'échantillonnage et de l'élaboration de la base de données. Tout d'abord, la taille de l'échantillon est très discutable pour un projet de machine learning et peut remettre en cause la généralisation des modèles construits. En effet, nous avons rencontré des problèmes de disponibilité des données, de contraintes de temps et d'efforts d'une part, parce que nous avons commencé à collecter ce que nous pouvions comme données à partir de fichiers physiques et numériques au niveau de la Direction des crédits industriels et des services. Nous avons éliminé, d'autre part, la période 2020 pour des raisons de la pandémie de covid pour éviter un classement biaisé, ces mêmes problèmes ne nous ont pas permis de constituer une population homogène au regard de la proportion d'entreprises saines et d'entreprises défaillantes.

1.1. Définition d'étiquette de classe :

Nous avons considéré comme mauvais ($y=1$) les dossiers de crédit financés ayant eu un incident de paiement, comme bons ($y=0$) les dossiers n'ayant pas eu d'impayés. Et nous avons tracé comme critère de défaillance un retard de remboursement supérieur ou égal à 90 jours, critère qui permet de répondre aux exigences du comité de Bâle.

*Chapitre II : L'implantation des modèles de classification basés sur
l'apprentissage supervisé*

1.2. Description des variables :

Tableau 1: listes des variables quantitatives et qualitatives

Les variables quantitatives	
Ratios de trésorerie	Valeur ajoutée/ CA
	Fonds de roulement en jour de CA
	Besoin en fonds de roulement en jour de CA
	Délai de règlement des clients
	Délai de règlement des fournisseurs
Ratios de liquidité	ACT-Stocks/DCT
	Disponibilité (net) / Actif
	Disponibilité (net) / DCT
	Charges financières / EBE
	DCT/Total dette
	Ration de rotation des stocks
Ratios de rentabilité	Résultat net après impôts / Total actif
	Résultat net / Fonds propres
	CAF / CA
	EBE / CA
	Résultat opérationnel / CA
	Résultat Financier / CA
	Résultat net après impôts / CA
Ratios de structure Financière	Total dettes / Total Actif
	Dette financière / Fonds propres
	Fonds propres / Passif non courant
	Fonds propres / Total actif
	Dette financière / Résultat net
	Dette à long terme / CAF
Les variables qualitatives	
La centrale des risques	Variables catégorielle binaire
Forme juridique	Variables catégorielles à 4 modalités

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Impayés confrères	Variables catégorielle binaire
Mouvement confiés	Variables catégorielle binaire
Secteur d'activités	Variables catégorielles à 8 modalités

Source : établis par nous même

Notre base est constituée de 29 variables dont 25 variables quantitatives, il s'agit des ratios financiers calculés à partir des bilans et tableaux de résultats, les ratios utilisés renseignent sur la trésorerie, la liquidité, la rentabilité, et la structure financières de chaque entreprise. Et 5 variables qualitatives, Ces ratios sont présentés dans le tableau ci-dessus.

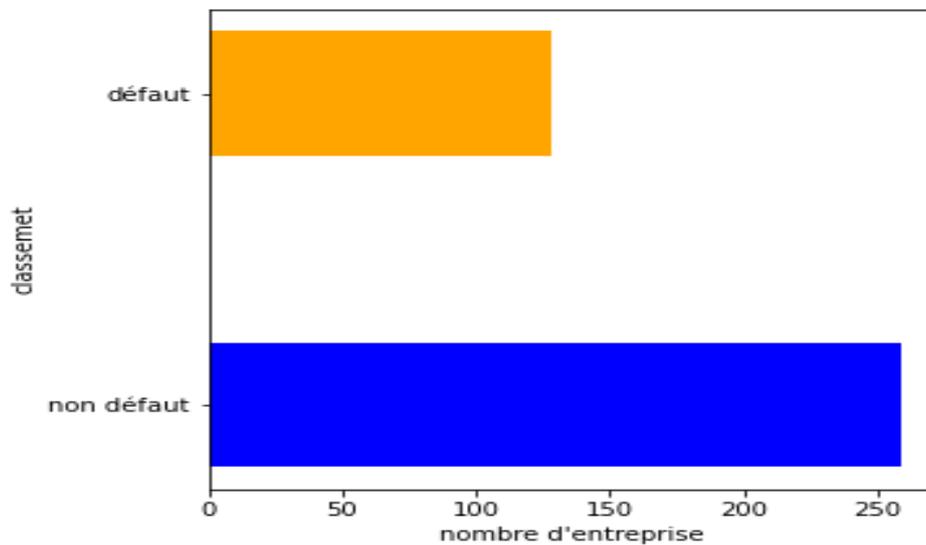
2. Les pré-traitements des données :

Le pré-traitement des données est l'une des étapes les plus importantes pour développer des modèles avec bonne performance. Les algorithmes de machines Learning apprennent à partir des données qui leur sont fournies, par conséquent si ces données sont de mauvaise qualité, erronées, incomplètes, ou redondantes, alors l'algorithme sera lui-même assez mauvais puisque il est censé de refléter ce qu'il voit dans les données, C'est pour cette raison, qu'il est impératif de préparer les données avant le passage à la machine, Il faut les nettoyer, les normaliser, les filtrer, et c'est cette étape quand appel le « **pre-processing** » ou le pré-traitement des données.

2.1. Visualisation de la valeurs cible :

Notre échantillon est constitué de 387 entreprises, il comporte 259 entreprises saines, et 128 entreprises défailtantes, soit donc 67% d'entreprises saines, et 33% des entreprises défailtantes, Comme le montre la figure suivante :

Figure 14: répartition de l'échantillon



Source : établis par nous même avec google colab

Il se doit de signaler que le déséquilibre entre les deux classes est dû au manque de données, notamment le manque d'entreprises défailtantes.

2.2. Le nettoyage de la base des données :

Le nettoyage des données consiste à éliminer toutes les informations que l'on ne souhaite pas conserver telles quelles. Ces informations peuvent être erronées, inexactes, induisant des erreurs ou simplement être sans intérêt pour la suite de l'analyse ou la modélisation, il s'agit de vérifier leurs qualités, Les données peuvent être : manquantes, aberrantes ou en double.

Les valeurs manquantes ou aberrantes sont présentées pratiquement dans toutes les bases de données des applications réelles. Elles peuvent correspondre aux erreurs de saisie ou l'indisponibilité de l'information. La mauvaise gestion de ces valeurs peut conduire à l'induction de modèles erronés et à des analyses fallacieuses. Le traitement des valeurs manquantes et aberrantes est souvent une tâche exigeante, tant du point de vue méthodologique qu'en termes de calcul.

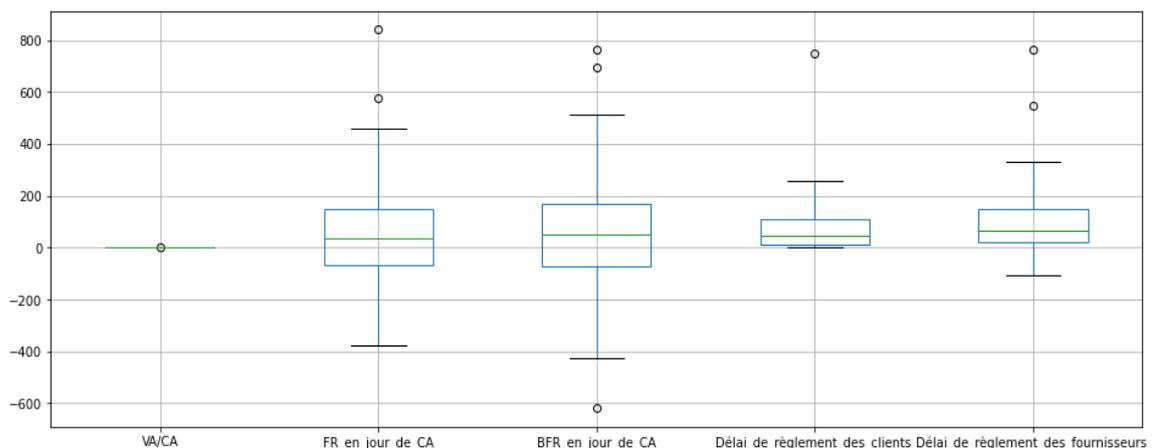
2.2.1. Traitement des valeurs Aberrantes :

Avant d'entreprendre le traitement des données manquantes, on doit tout d'abord traiter les valeurs aberrantes. Une valeur aberrante est une valeur qui diffère de façon significative de la

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

tendance globale des autres observations quand on observe un ensemble de données ayant des caractéristiques communes, ces valeurs pouvant provenir d'erreurs de saisie. Donc à l'aide de diagramme à moustaches, nous avons écrits un algorithme pour détecter les valeurs aberrantes, comme son nom l'indique, une boîte à moustaches, est composé d'une boîte et 2 moustaches, l'extrémité de la moustache bas nous informe sur la valeurs minimale de notre base, l'extrémité de la moustache en haut, nous informent sur la valeur maximale de la série, le bord en bas de la boîte nous indique la valeur du 1^{er} quartile, le bord en haut de la boîte, nous indique la valeur du 3^{ème} quartile de cette même série, donc le principe consiste à détecter toutes les valeurs en dehors des bornes inférieures et supérieurs, c'est-à-dire celles qui sont inférieures à la limites inférieure qui est égale à : $Q1 - 1.5 \times (Q3 - Q1)$, et celle qui sont supérieures à la limite supérieure qui est égale à $Q3 + 1.5 \times (Q3 - Q1)$. Les valeurs aberrantes se présentent comme suit :

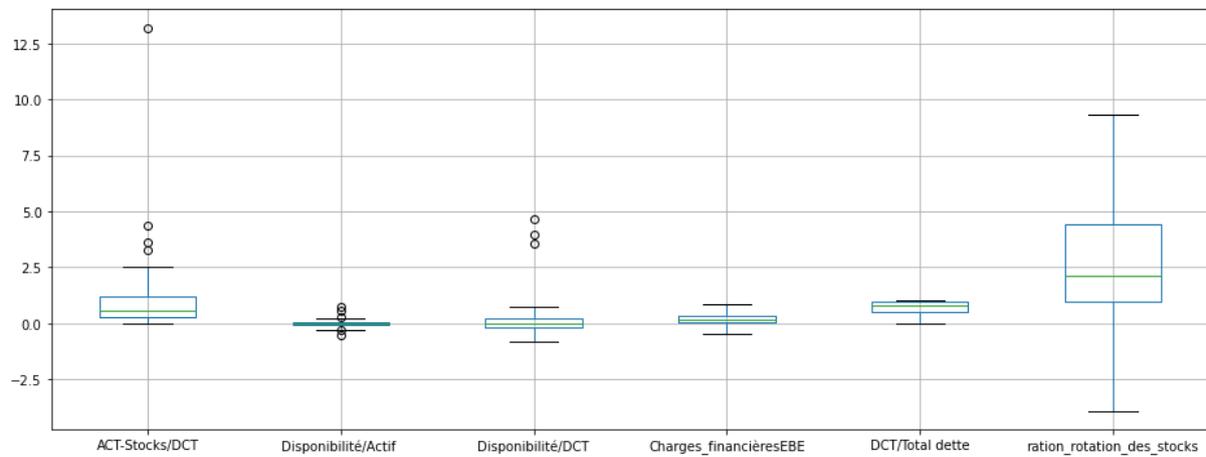
Figure 15: les valeurs aberrantes pour les ratios de trésorerie



Source : établis par nous même avec google colab

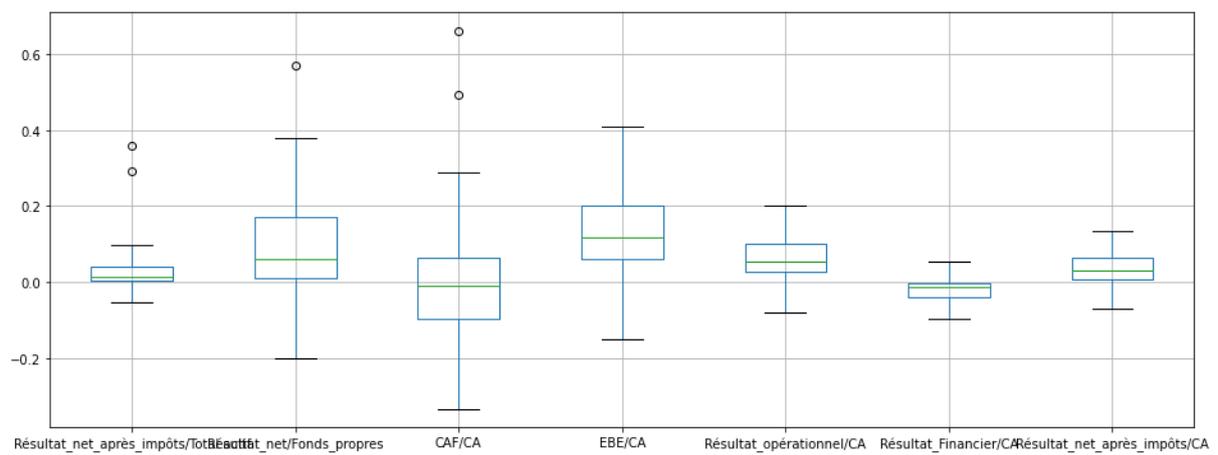
Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Figure 16: les valeurs aberrantes pour les ratios de liquidité



Source : établis par nous même avec google colab

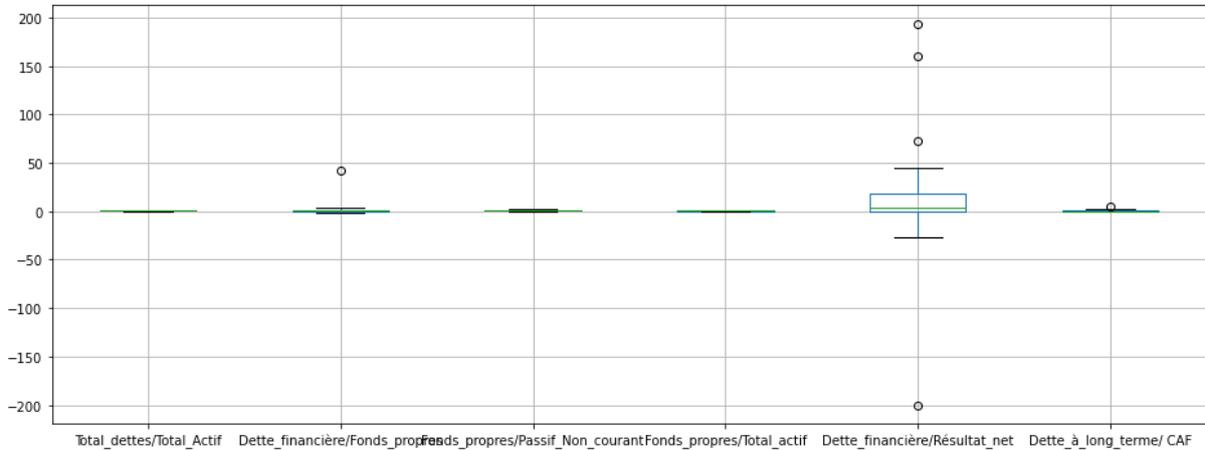
Figure 17: les valeurs aberrantes pour les ratios de rentabilité



Source : établis par nous même avec google colab

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Figure 18: les valeurs aberrantes pour les ratios de structure financière



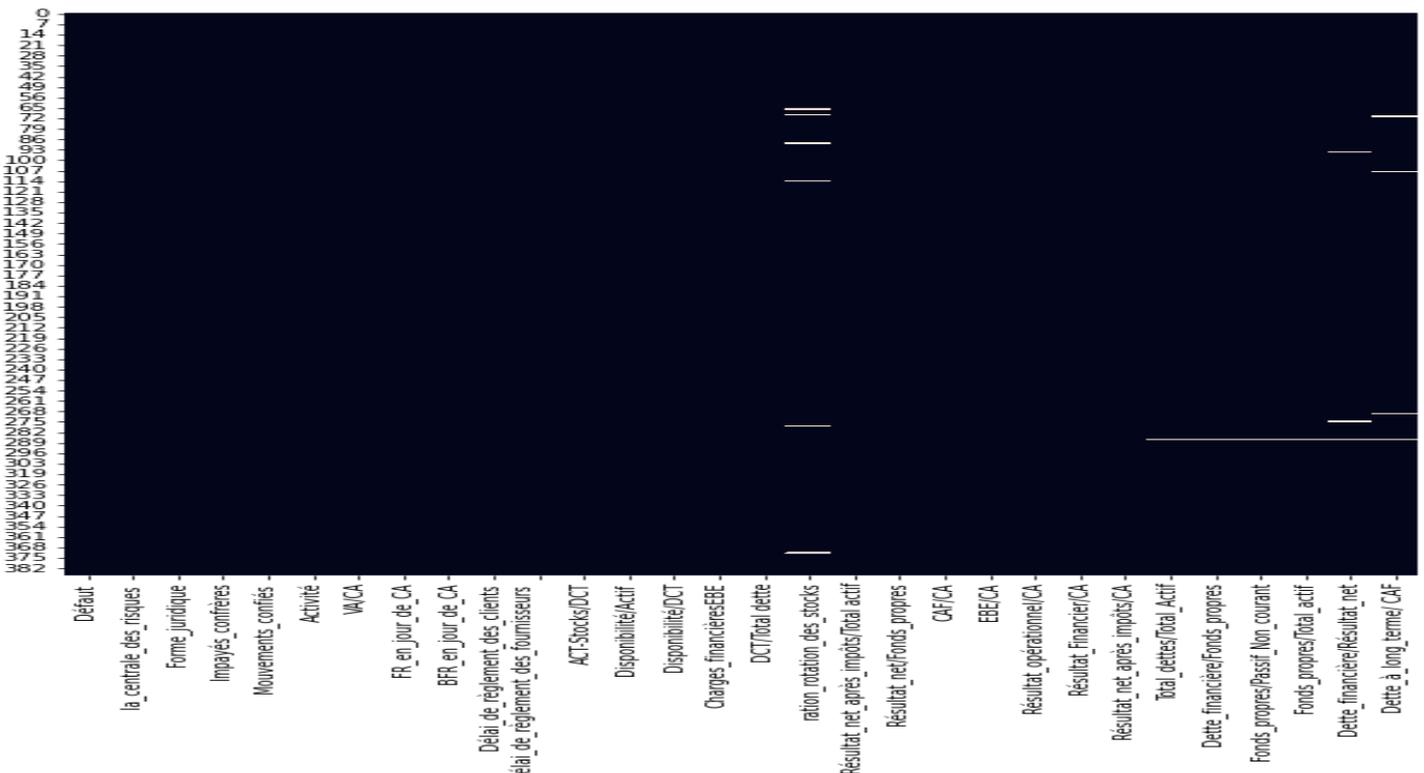
Source : établis par nous même avec google colab

Comme montre les graphes ci-dessous, les valeurs aberrantes sont présentées par des points, après la détection de ces valeurs, nous avons les éliminer de la base des données

2.2.2. Le traitement des valeurs manquantes :

A l'aide de la fonction **heatmap** de **seaborn**, nous avons affiché notre base en image, pour détecter les valeurs manquantes, Nous recourons à cette figure présentée ci-dessous :

Figure 19: présentation des valeurs manquantes



Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Donc ici, notre base de données est affichée avec toutes ses lignes, et ses colonnes, les lignes en blancs correspondent aux positions où il manque des valeurs, et en noire les positions où nous avons des valeurs.

Nous remarquons que notre base de données ne contient pas beaucoup de valeurs manquantes, et pour obtenir une base de données complète et fiable, nous avons étudié chaque variable au cas par cas.

Tableau 2: le nombre des valeurs manquantes dans chaque variable

Total_dettes/Total_Actif	1
Dette_financière/Fonds_propres	1
Fonds_propres/Passif_Non_courant	1
Fonds_propres/Total_actif	1
Dette_financière/Résultat_net	3
Dette_à_long_terme/ CAF	4
ration_rotation_des_stocks	6

Source : établis par nous même à partir des résultats de google colab

Nous remarquons que l'ensemble des valeurs manquantes sont de types quantitatifs. Pour certaines variables, comme le ratio de rotation des stocks, il s'agit bien des erreurs lors de la division par zéro qui correspond un à encours des stocks nul, et qui se traduit par une valeur manquante dans la visualisation de notre base. Pour le reste des variables, une partie des données est indisponible.

Pour ce faire, et à l'aide du transformer **fillna**, nous avons remplacé les valeurs vide dans le ratio de rotation des stocks par 0, partant du principe que l'entreprise adopte l'approche zéro stocks comme technique de gestion des stocks, pour le reste des variables, nous avons remplacé les valeurs manquantes de la manière suivantes :

- Les valeurs manquantes des ratios « fonds propres/passif non courant », et « fonds propres/total actif », par le mode.

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

- Les valeurs manquantes des ratios « total dettes/ total actif », « dettes financières/ fonds propres », « dettes financières / résultat net », et « dettes à long terme/ CAF », par la moyenne.

2.3. L'encodage des variables qualitatives :

Quand on développe un modèle de machine learning, il est nécessaire de présenter à la machine des valeurs numériques, avec lesquelles elle puisse faire des formules de calcul, par conséquent, si notre base de données contient des valeurs qualitatives, Il est indispensable de transformer les variables qualitatives en valeur numériques.

A partir des tranformers du module **sklearn.preprocessing**, nous avons utilisé **LabelEncoder**, qui encode chaque classe de la variables en une valeur, pour l'encodage des variables binaire.

Et pour les variables catégorielles à plusieurs modalités, nous avons utilisé la méthode **get_dummies** de **pandas**, avec cette technique chaque catégorie de classe est représentée de façon binaire dans une colonne qui lui propre, on décompose ici la variables initiale en plusieurs sous variables créant donc autant de colonnes que nous avons de catégorie ou de variables, de cette manière, les algorithmes de machine learning peuvent pas comparer les valeurs associe aux variables catégories sur un seul et même axe car ces catégorie est désormais séparer en plusieurs dimension. Est nos variables qualitatives deviennent :

Tableau 3: liste des variables qualitatives après encodage

Variable de base	Variable dummies
La centrale des risques	La centrale_des_risques_Néant
	La centrale_des_risques_Existant
Forme juridique	Forme_juridique_EURL
	Forme_juridique_SARL
	Forme_juridique_SNC
	Forme_juridique_SPA
Impayés confrères	Impayés_confrères_oui
	Impayés_confrères_non
Mouvement confiés	Mouvement_confiés_quasi_intégral

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

	Mouvement_confiés_partiel
	Acitivité_industries_manufacturières
	Acitivité_pharmacie
Activité	Acitivité_industries_agroalimentaire
	Acitivité_service
	Acitivité_Import/export
	Acitivité_BTPH
	Acitivité_hydrocarbures_energie_mine_et _service_liés
	Acitivité_commerce

Source : élaborés par nous même

2.4. Fractionnement et partitionnement des données :

En règles générale, dans les projets de machine learning, il est nécessaire de tester la performance du modèle sur des données qu'il n'a pas encore vues depuis la phase de formation, pour ce faire, nous avons créé un échantillon sur lequel , nous avons constitué le modèle, et un échantillon de test, sur lequel nous avons tester le modèle, à l'aide de la fonction **train_test_split** de **sklearn.model**, où nous avons prendre 80% des données pour l'échantillon d'entraînement et 20% des données pour l'échantillon de test, On a procédé à la définition de la variable cible à la machine. Cette variable cible (y) est égale à la variable défaut. Le reste des variables sont considérées comme des variables explicatives. Ce qui a permis d'obtenir la répartition suivante :

Tableau 4: le fractionnement de l'échantillon

	0	1
<i>Echantillon d'entraînement</i>	202	98
<i>Echantillon test</i>	48	28

Source : établis par nous même à partir des résultats de google colab

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

2.5. Traitement des déséquilibres des classes :

Le problème du déséquilibre des classes peut entraîner une rareté des exemples pour une classe spécifique, classe de défaut pour notre cas, ce qui présente des défis pour les algorithmes de classifications binaires supervisés qui tentent de découvrir des régularités dans de telles données, qui peut entraîner même des effets négatifs sur les performances, pour surmonter ce problème nous avons utilisé la technique de **SMOTE.NC** du module **imblearn.over_sampling**, pour notre échantillon. Cette technique n'a apporté aucune amélioration, puisque nous n'avons pas une grande différence entre les deux classes.

2.6. La normalisation des données :

En Data science, il est indispensable de normaliser nos données quantitatives c'est à dire les mettre toutes sur une même échelle, ce qui facilite considérablement l'apprentissage du modèle machine learning, puisque la différence d'échelle peut conduire à de moindres performances.

Pour surmonter de cette difficulté, nous avons utilisé le transformateur **StandardScaler** de module **sklearn.feature_selection**, qui consiste à transformer les données de telle sorte que chaque variable est une moyenne égale à 0 et écart-type égale à 1, pour ce faire, il y a lieu de soustraire chaque valeur à la moyenne initiale de notre variable et de diviser le tout par l'écart-type initiale de notre variable on obtient ainsi des données très simples à utiliser pour la plupart des modèles statistiques.

2.7. La sélection des variables :

Comme il existe plusieurs méthodes pour la sélection des variables dans le modèle. Dans cette étape, nous avons essayé des méthodes de sélection de variables pour déterminer la meilleure fonction à appliquer qui nous donne les meilleurs résultats. Nous avons essayé le **selectFwe** de **sklearn**. Au final, nous avons décidé d'utiliser les résultats de la méthode **selectkbest** pour la sélection des variables.

La sélection des variables pertinentes à l'aide du transformateur **selectKbest** de **sklearn** a été initiée, pour ce faire nous avons utilisé le test χ^2 pour sélectionner les variables qualitatives. Le f_{classif} (test Anova) a été pratiqué pour les variables quantitatives. Les

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

variables qui présentent le score du test de dépendance avec la variable cible le plus élevés ont été retenues.

Après la sélection des variables les plus pertinentes, nous avons testé la redondance entre ces variables, et nous avons éliminé ceux qui ont une forte corrélation entre eux. Cette approche a donné lieu à une matrice de corrélation (Annexe 1).

Section 2 : La construction des modèles

A ce stade, notre base est prête pour commencer la modélisation, nous avons entraîné des modèles différents, et nous avons retenu celui qui a la meilleure performance, et pour s'assurer que les modèles établis seront performants, nous avons utilisé la validation croisée.

La validation croisée consiste à entraîner puis valider notre modèle sur plusieurs découpe possible de l'échantillon d'apprentissage, dans notre cas nous avons choisi un nombre de splits (découpe) de 4, dans ici le modèle sera entraîné sur les 3 premières parties du trainset puis le valider sur la 4^{ème} partie, ensuite on va refaire tout ça pour toutes les configurations possibles au final on aura la moyenne des 4 scores que l'on obtient et qui présente le taux de succès du modèle et ainsi lorsque on voudra comparer deux modèles ou plus alors on sera sûr de prendre celui qui a au moyenne eu les meilleures performances. Nous avons tenté d'améliorer la performance de chaque modèle, par l'optimisation de ces hyperparamètres grâce au algorithme **GridSearch CV**, la recherche par grille sert à rechercher parmi un ensemble des paramètres prédéfini, la valeur des hyperparamètres qui permet d'obtenir les meilleurs résultats.

1. L'arbre de décision :

Après l'exécution de l'algorithme de l'arbre de décision, nous avons créé une liste d'hyperparamètres pour optimiser notre modèle, après l'optimisation, nous avons obtenu les hyperparamètres ci – après :

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Tableau 5: les hyperparamètres de l'arbre de décision

	Les hyperparamètres par défaut	Les hyperparamètres après optimisation
criterion	gini	entropy
Max_depth	None	2
Max_features	None	log2
Min_samples_leaf	1	1
Min_samples_split	2	6

Source : établis par nous même à partir des résultats de google colab

Criterion désigne la mesure qui permet de quantifier l'incertitude sur la règle de discrimination utilisé, plus l'entropie est élevé moins le modèle est performant, pour le **max_depth** qui indique la profondeur maximale de l'arbre où les nœuds sont étendus jusqu'à ce que toutes les feuilles contiennent un seul point de données la valeur optimale est de 2, **max_features**, qui indique le nombre des variables considéré pour avoir la meilleur classification est log2 après l'optimisation, pour le **min _sapmles_split** qui indique le nombre minimum d'échantillons requis pour diviser un nœud interne est de 1, et finalement le **min_samples_split** cela indique le nombre minimum d'échantillons requis pour fractionner un nœud, la valeur optimale est de 4.

1.1. Les rapports de classification :

Nous allons évaluer maintenant notre modèle à partir d'une analyse de rapport de classification, et pour faciliter l'interprétation des résultats nous allons présenter ci-dessous un état de comparaison entre les deux rapports, avant et après l'optimisation :

Figure 20: rapports de classification avant et après l'optimisation de l'arbre de décision

	precision	recall	f1-score	support
0	0.73	0.85	0.79	48
1	0.65	0.46	0.54	28
accuracy			0.71	76

	precision	recall	f1-score	support
0	0.74	0.83	0.78	48
1	0.64	0.50	0.56	28
accuracy			0.71	76

Source : établis par nous même avec google colab

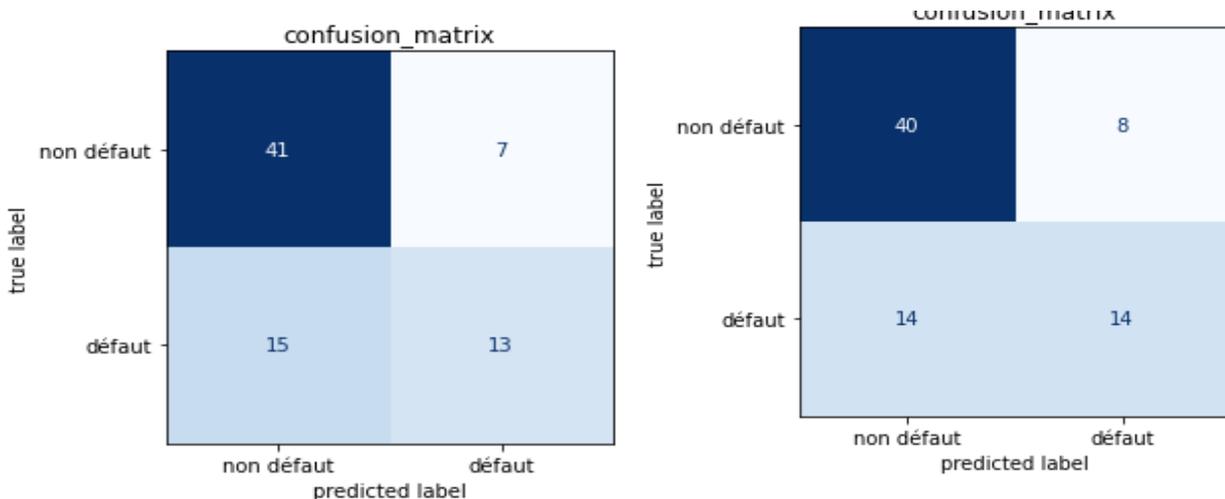
Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Le modèle nous donne un taux de succès de 71%, après l'optimisation le modèle est capable de prédire les classes défailtantes avec une précision de 64% ce qui nous donne un taux de succès de 71%, cette précision résume le score attribué aux classes défailtantes qui appartiennent réellement aux classes défailtantes, et avec un f1_score après optimisation de 56% pour les classes défailtantes, nous pouvons dire que notre modèle est légèrement plus performant après l'optimisation.

1.2. Les matrices de confusion :

Nous analysons les matrices de confusion ci-dessous :

Figure 21: les matrices de confusion de l'arbre de décision avant et après l'optimisation



Source : établis par nous même avec google colab

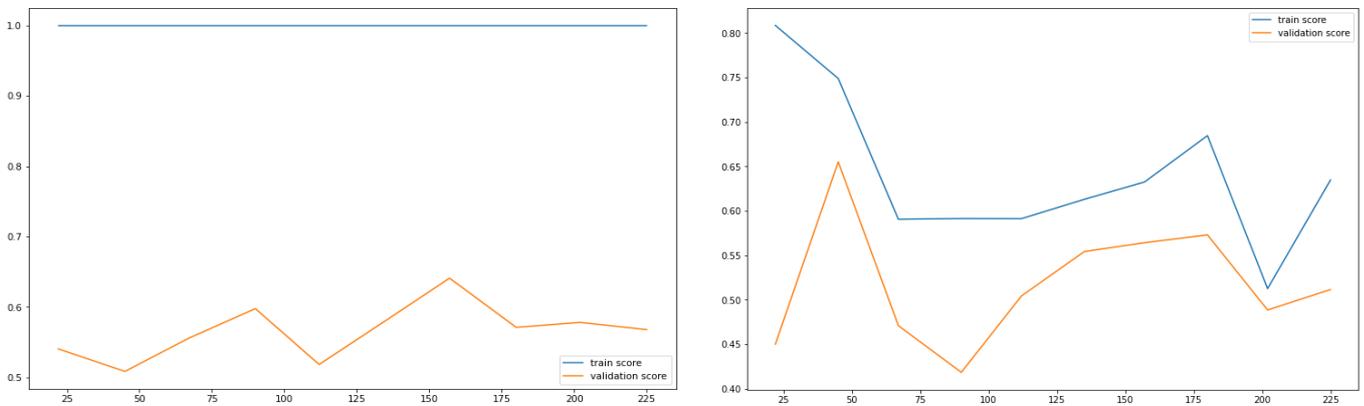
Sur 48 entreprises saines dans l'échantillon de validation, le modèle a identifié correctement 41 entreprises saines et il a rejeté 7, et sur 28 entreprises défailtantes, le modèle a identifié correctement 13 entreprises, après optimisation, le modèle est capable d'identifier mieux les entreprises en défauts, sur 28 entreprises, il a correctement identifié 14 entreprises.

1.3. Les courbes d'apprentissages :

Nous analysons maintenant les courbes d'apprentissage ci-dessous :

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Figure 22: les courbes d'apprentissage avant et après l'optimisation de l'arbre de décision



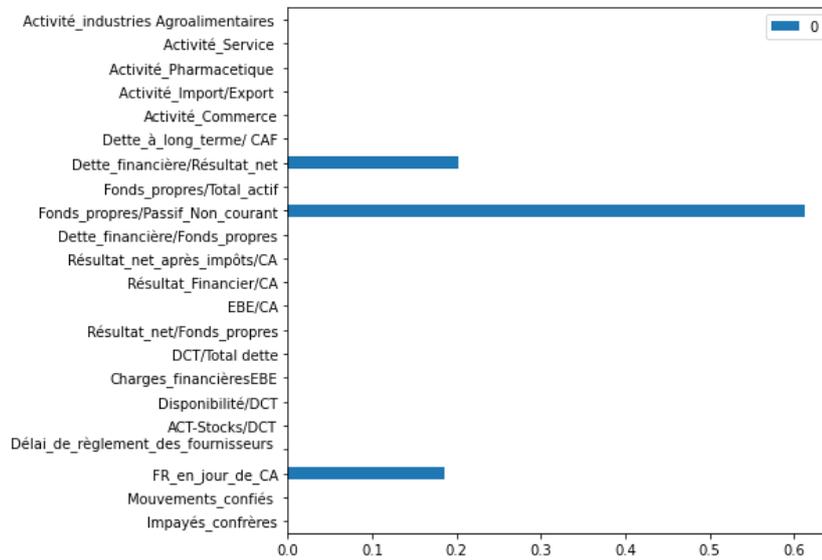
Source : établis par nous même avec google colab

Nous pouvons dire que notre modèle est en overfitting avant l'optimisation, puisque le modèle a un score de 100% sur l'échantillon d'apprentissage (courbe en bleu), par contre il a un problème de généralisation (courbe en orange). Après l'optimisation des hyperparamètres du modèle. Nous remarquons que la courbe de train score est en diminution, mais la courbe de validation score augmente, et commence à atteindre un pallier où les deux se rejoint, cela indique que notre modèle n'est pas en overfitting, puisque son score sur l'échantillon d'apprentissage est presque proche de son score sur l'échantillon de validation.

1.4. Décision :

Puisque l'optimisation des hyperparamètres, nous a donné une meilleure performance, donc nous avons choisis le modèle d'arbre de décision après optimisation, nous obtenons donc une présentation graphique de notre arbre (Annexe 2), et les variables qui explique mieux le défaut pour le modèle d'arbre de décision sont :

Figure 23: les variables importantes pour l'arbre de décision



Source : établis par nous même avec google colab

Après l’optimisation, le nombre de variables importantes pour le modèle arbre de décision à diminuer, « le ratio Fonds propres / passif non courant » constitue la variable la plus importantes avec taux d’importance de 61%, suivie par le ratio « Dette financières / résultat net » avec un taux de 20%. Le ratio « fonds de roulement en jour » présentent une importance faible par rapport au 2 ratios déjà cités, avec un taux d’importance de 13%. Après optimisation le modèle a seulement trois variables les plus importantes, deux variables désignent l’endettement, et une variable qui affecte la rentabilité.

2. La régression logistique :

Nous appliquons l’algorithme de la régression logistique, nous obtenons les hyperparamètres ci-dessous :

Tableau 6: les hyperparamètres de la régression logistique

	Les hyperparamètres par défaut	Les hyperparamètres après l’optimisation
C	1.0	0.03
solver	lbfgs	lbfgs

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Max_iter	100	100
-----------------	-----	-----

Source : établis par nous même à partir des résultats de google colab

Le paramètre **C** présente la force de régularisation, la régularisation minimise l'erreur et rend le modèle capable à généraliser, la valeur optimale est de 0.03, c'est-à-dire une régularisation inférieure et donc un modèle dans l'ensemble plus susceptible de se généraliser, le paramètre **solver** désigne le type d'algorithme utilisé dans l'optimisation, quant au **max_iter** désigne le nombre maximal d'itérations nécessaires pour que les solveurs convergent.

2.1. Les rapports de classification :

Une lecture des rapports de classification nous donne :

Figure 24: les rapports de classification avant et après l'optimisation de la régression logistique

	precision	recall	f1-score	support
0	0.77	0.90	0.83	48
1	0.75	0.54	0.63	28
accuracy			0.76	76

	precision	recall	f1-score	support
0	0.74	0.94	0.83	48
1	0.80	0.43	0.56	28
accuracy			0.75	76

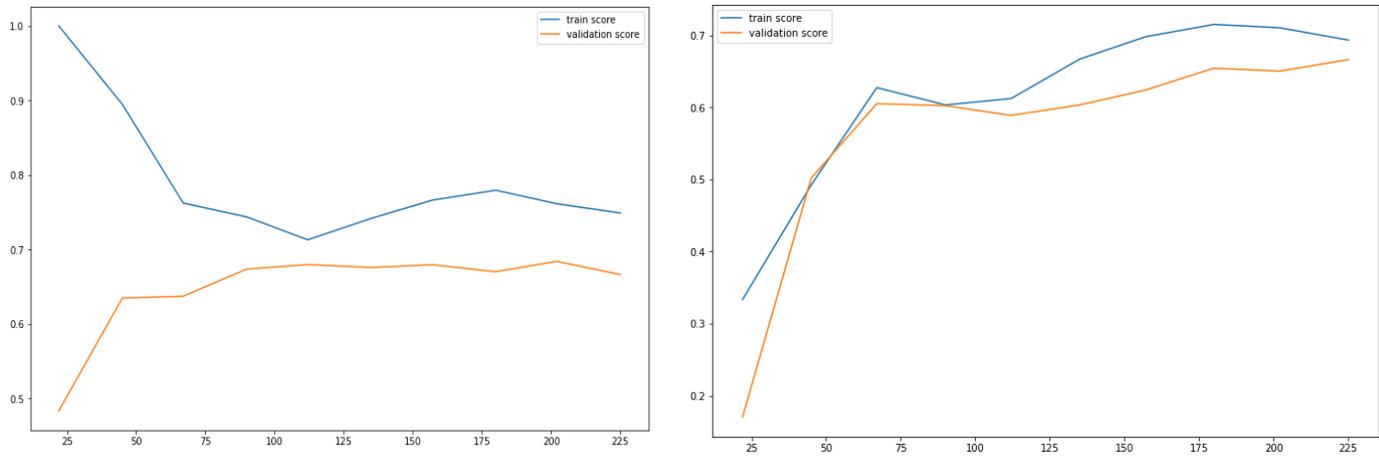
Source : établis par nous même avec google colab

Le modèle de régression logistique est plus performant avec les paramètres par défaut, avec un taux de succès globale de 76%, est un f1_score des classes défaillantes de 63%, nous pouvons dire donc que l'optimisation a réduit la performance du modèle.

2.2. Les courbes d'apprentissage :

Pour s'assurer de cette analyse, nous avons examiné les courbes d'apprentissage :

Figure 25: les courbes d'apprentissage avant et après l'optimisation de la régression logistique



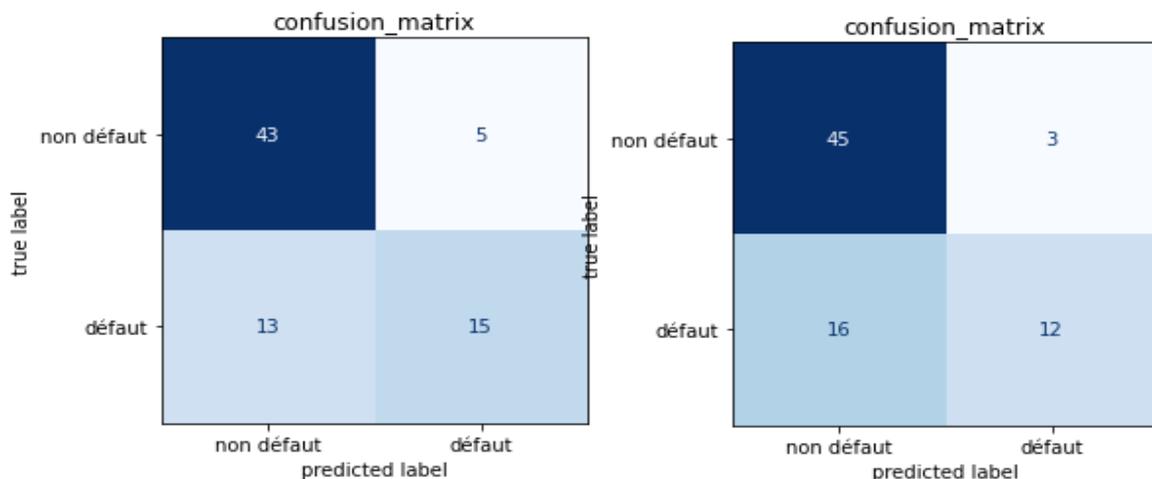
Source : établis par nous même avec google colab

La première chose à remarquer, consiste à savoir que l'optimisation a entraîné des modifications dans la courbe d'apprentissage, ici la courbe de construction et de validation ont une tendance haussière ce qui signifie que le modèle sera plus performant avec plus de données. Mais dans les deux situations, la courbe de construction et la courbe de validation se rapprochent ce qui indique que le modèle n'a pas un problème de généralisation.

2.3. Les matrices de confusion :

L'élaboration des matrices de confusion nous donne :

Figure 26: les matrices de confusion avant et après l'optimisation de la régression logistique



Source : établis par nous même avec google colab

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Le modèle par défaut a détecté correctement 15 entreprises défaillantes, et 43 entreprises saines, après l'optimisation le modèle a détecté 16 entreprises saines qui sont réellement des entreprises défaillantes.

2.4. Décision :

Donc ici, nous pouvons dire que le modèle de la régression logistique est plus performant avec ses paramètres par défauts, une fois nous avons obtenu les coefficients de chaque variable du modèle, nous avons déterminé les gènes qui sont les bons prédicteurs, et éliminé ce qui sont moins bons prédicteurs. Les variables qui expliquent mieux le défaut selon le modèle de la régression logistique sont :

Tableau 7: les coefficients des variables de la régression logistique

Variables	Béta
Impayé confrères	-0,119
Mouvement confiés	-0,453
FR en jour de CA	-0,125
Charges financières/EBE	0,158
EBE/CA	-0,177
Résultat financier / CA	-0,206
Résultat net après impôts / CA	-0,314
Fonds propres/ passif non courant	-0,369
Fonds propres/ total actif	-0,200
Dette à long terme / CAF	0,103

Source : établis par nous même à partir des résultats de google colab

Du point de vue financière, le modèle est satisfaisant puisque le signe attendu de tous les ratios est confirmé. En effet, on attend à ce que la rentabilité affecte négativement la probabilité de faillite, l'endettement affecte positivement cette dernière,

Le ratio « Charges financières / EBE », possède le coefficient le plus significatif parmi les ratios d'endettement. Cette importance réside dans la formation de ce ratio. En effet, ce ratio permet d'exprimer la capacité de l'entreprise à honorer la charge de la dette, plus ce ratio est élevé, plus élevée sera la proportion de l'EBE absorbée par le paiement des intérêts, et plus

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

importantes seront les chances de faire défaut. Le ratio « Dettes à long terme / CAF ». Le deuxième plus important ratio. Il présente le poids de l'endettement de l'entreprise par rapport à la capacité d'autofinancement. Plus la valeur de ce ratio est élevée, plus la capacité de l'entreprise de rembourser le principal de la dette est faible.

Les ratios de solvabilités, présentes les coefficients les plus élevés dans la régression logistique, ce ratio désigne la capacité des entreprises à faire face à leurs engagements.

Les deux variables qualitatives « mouvement confiés » et « impayés confères » sont désormais significatives.

3. Le sector vector machine :

Nous avons exécuté le modèle avec les paramètres par défaut, et nous avons l'optimiser par la modification des hyperparamètres, suivantes :

Tableau 8: les hyperparamètres avant et après l'optimisation

	Hyperparamètres par défaut	Hyperparamètre après l'optimisation
C	1	1
gamma	Scale	0.01
Kernel	rbf	rbf

Source : établis par nous même à partir des résultats de google colab

C, est un paramètre de contrôle du compromis entre une limite de décision fluide et la classification correcte des points d'entraînement, la valeur optimale est 1. **Le kernel** (noyau) spécifie le type de noyau à utiliser dans l'algorithme, il projette les données non séparables linéairement des dimensions inférieures vers des données séparables linéairement dans des dimensions supérieures de telle sorte que les points de données appartenant à différentes classes sont alloués à différentes dimensions, le modèle a pris la valeur par défaut, et pour **le gamma** qui définit le coefficient de noyau, en d'autres termes, avec un gamma faible, les points éloignés de la ligne de séparation sont présent en compte dans le calcul de la ligne de séparation, un gamma élevé signifie que les points proches de la droite qui sont uniquement présent en compte dans le calcul. Le modèle a pris une valeur de 0.01.

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

3.1. Les rapports de classification :

Par une analyse des rapports de classification ci-dessous :

Figure 27: les rapports de classification après et avant l'optimisation du modèle SVM

	precision	recall	f1-score	support
0	0.75	0.96	0.84	48
1	0.87	0.46	0.60	28
accuracy			0.78	76

	precision	recall	f1-score	support
0	0.72	0.98	0.83	48
1	0.91	0.36	0.51	28
accuracy			0.75	76

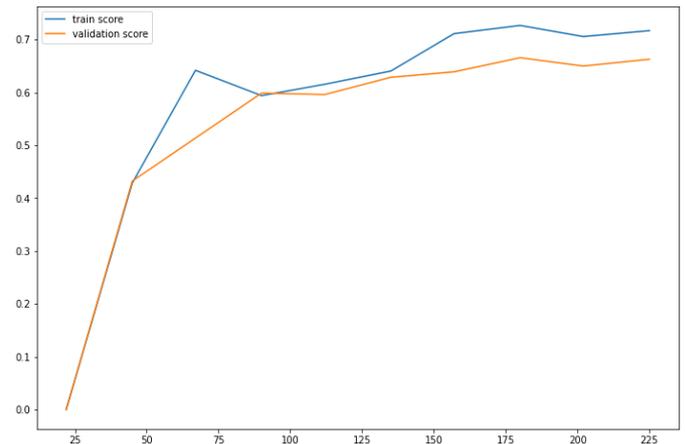
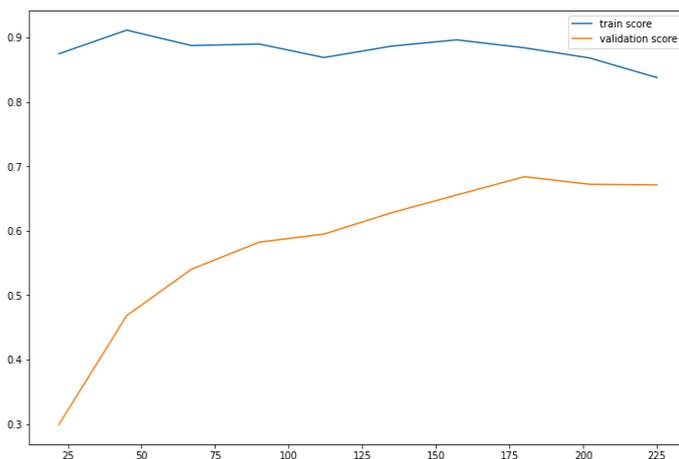
Source : établis par nous même avec google colab

Il est clair que le modèle après optimisation est moins performant, avec un taux de succès moins élevé, soit 75%, contre 78% avant l'optimisation. Avec un f1_score de 60%, le modèle avant optimisation est capable de détecter mieux les classes défailtantes.

3.2. Les courbes d'apprentissage :

Nous analysons les courbes d'apprentissage :

Figure 28: les courbes d'apprentissage avant et après l'optimisation du modèle SVM



Source : établis par nous même avec google colab

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

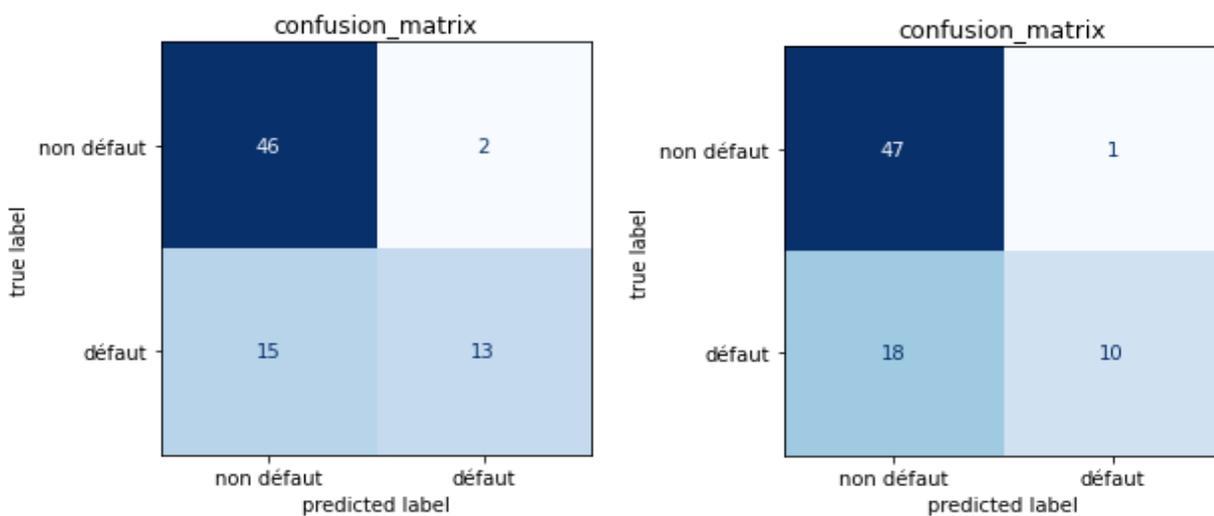
Similaire à la courbe de validation, la courbe de construction est en tendance haussière, avec le nombre de donnée fournis, est les deux courbes sont proche, la chose qui indique que notre modèle généralise beaucoup plus mieux qu'avant l'optimisation des hyperparamètres.

Mais même avant l'optimisation le modèle ne souffre pas d'un sur apprentissage, la courbe de construction diminue et la courbe de validation augmente en fonction des données, c'est-à-dire si nous fournissons plus de données à notre modèle, il sera plus performant.

3.3. Les matrices de confusion :

Nous analysons maintenant les matrices de confusion :

Figure 29: les matrice de confusion après et avant l'optimisation du modèle SVM



Source : établis par nous même avec google colab

Nous remarquons que le modèle dans les deux cas prédit presque le même nombre des entreprises saines, où le modèle a détecté la quasi intégralité des entreprises saines. Pour 28 entreprises défailtantes, il a correctement détecté 13, l'optimisation n'améliore pas la détection des entreprises défailtantes.

3.4. Décision :

Nous basons sur ces analyses, nous pouvons dire que le modèle est plus performant avant l'optimisation.

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

4. Le K-neighbor classifier :

Comme pour les modèles précédents, nous avons exécuté le modèle avec les paramètres par défauts, et après optimisation des hyperparamètres :

Tableau 9: les hyperparamètres du modèle KNN

	Hyperparamètres avant l'optimisation	Hyperparamètres après l'optimisation
N_neighbors	5	4
metric	minkowski	Euclidean

Source : établis par nous même à partir des résultats de google colab

N_neighbors, désigne le nombre de voisins pris par le paramètre k de l'algorithme, sa valeur optimale est 4, **metric**, désigne la distance, puisque le modèle KNN, a besoin d'une fonction de calcul de distance entre deux observations. Pour le calcul de cette distance il existe plusieurs fonctions parmi ces fonctions l'optimale pour notre modèle est distance euclidean

4.1. Les rapports de classification :

Nous analysons les rapports de classifications :

Figure 30: les rapports de classification avant et après l'optimisation du modèle KNN

	precision	recall	f1-score	support
0	0.75	0.85	0.80	48
1	0.67	0.50	0.57	28
accuracy			0.72	76

	precision	recall	f1-score	support
0	0.71	0.85	0.77	48
1	0.61	0.39	0.48	28
accuracy			0.68	76

Source : établis par nous même avec google colab

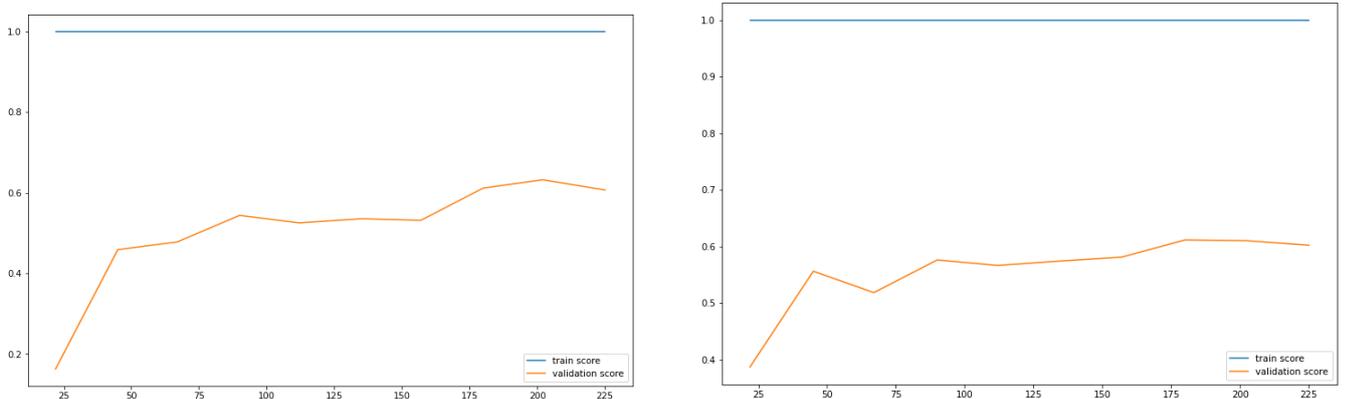
Nous observons que le modèle donne des meilleurs résultats avant l'optimisation, un taux de succès de 72%, et un f1_score des classes défailtantes de 57%, contre uniquement 48% à celui du modèle avec optimisation, le modèle après optimisation est moins performant dans la détection des entreprises saines.

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

4.2. Les courbes d'apprentissage :

Quant aux courbes d'apprentissage :

Figure 31: les courbes d'apprentissage avant et après l'optimisation du modèle KNN



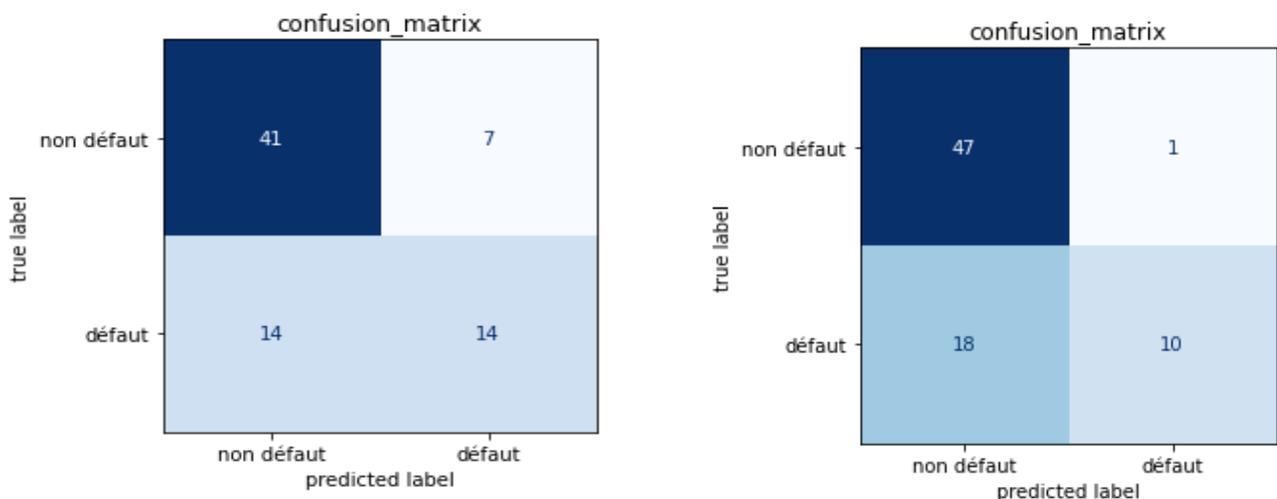
Source : établis par nous même avec google colab

Quel que soit avant ou après l'optimisation, le modèle souffre toujours d'un sur apprentissages, cela est traduit par une ligne droite de la courbe de construction, donc le modèle est en overfitting, c'est à dire, le modèle manque de capacité de généralisation.

4.3. Les matrices de confusion :

Et une analyse des matrices de confusion nous donne :

Figure 32: les matrices de confusion avant et après l'optimisation du modèle KNN



Source : établis par nous même avec google colab

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Les matrices de confusion confirment les analyses précédentes, après l'optimisation le modèle est capable de prédire seulement 10 entreprises défaillantes, alors sans optimisation le modèle a détecté 14 entreprises de l'ensemble des 28 entreprises défaillantes. Après optimisation le modèle a prédit correctement l'intégralité des entreprises saines.

4.4. Décision :

A ce stade, nous pouvons dire le modèle KNN est plus performant avec ses paramètres par défauts.

5. Adaboost :

De même pour le modèle Adaboost, nous évaluons sa performance en variant les hyperparamètres :

Tableau 10: les hyperparamètres du modèle Adaboost

	Hyperparamètres avant l'optimisation	Hyperparamètres après l'optimisation
Learning_rate	1	0.1
N_estimators	50	500

Source : établis par nous même à partir des résultats de google colab

Leaning rate est la contribution de chaque modèle aux pondérations, un taux d'apprentissage élevé augmente la contribution de chaque classificateur, sa valeur optimale ici est 0.1, quant à **n_estimators** désigne le nombre des estimateurs à utiliser dans l'algorithme.

5.1. Les rapports de classification :

A partir des rapports de classification de classification :

Figure 33: les rapports de classification avant et après l'optimisation du modèle Adaboost

	precision	recall	f1-score	support			precision	recall	f1-score	support
0	0.77	0.85	0.81	48		0	0.77	0.98	0.86	48
1	0.70	0.57	0.63	28		1	0.93	0.50	0.65	28
accuracy			0.75	76		accuracy			0.80	76

Source : établis par nous même avec google colab

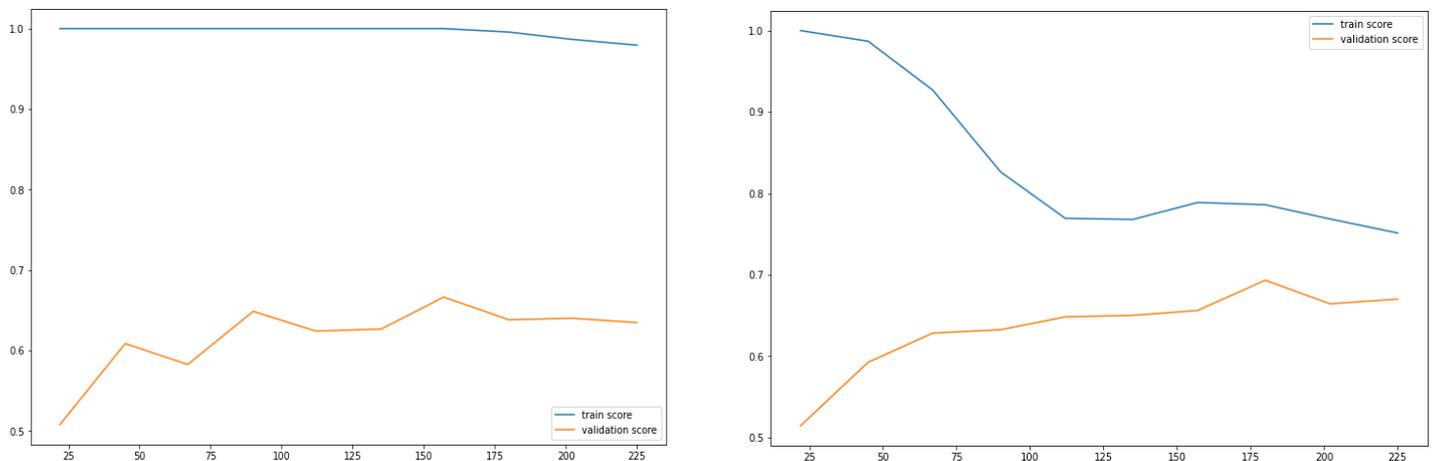
Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Nous remarquons le modèle sans optimisation, donne un taux de succès de 75%, soit un score moins élevé à celui de modèle après optimisation, le f1_score à l'instar du taux de succès est moins élevé avant l'optimisation, soit 63% contre 65% après l'optimisation, ce qui nous amène à dire que le modèle est plus performant avec optimisation.

5.2. Les courbes d'apprentissage :

Pour confirmer les résultats précédents, nous analysons ci- ainsi les courbes d'apprentissage :

Figure 34: les courbes d'apprentissage avant et après l'optimisation du modèle Adaboost



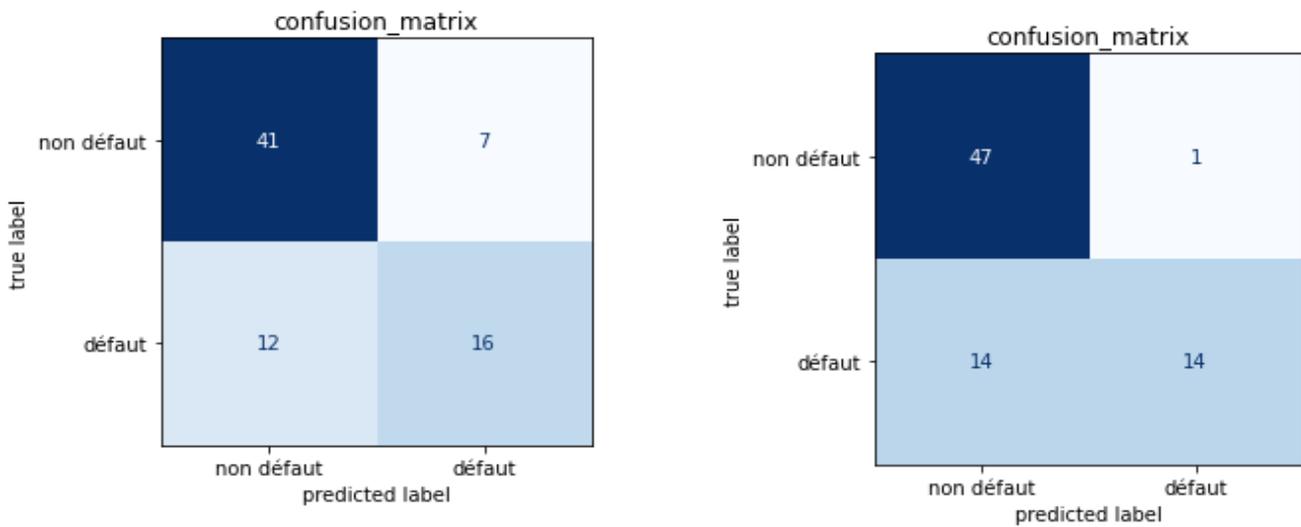
Source : établis par nous même avec google colab

Avant l'optimisation, le modèle a un problème de généralisation, ceci est expliqué par la ligne droite de la courbe de construction, après l'optimisation, les deux courbes commencent à se rejoindre, le modèle n'est plus en sur apprentissage, cela signifie que si nous fournissons plus de donnée au modèle, la performance va s'améliorer de plus.

5.3. Les matrices de confusion :

Une dernière analyse des matrices de confusion :

Figure 35: les matrices de confusion avant et après l'optimisation du modèle Adaboost



Source : établis par nous même avec google colab

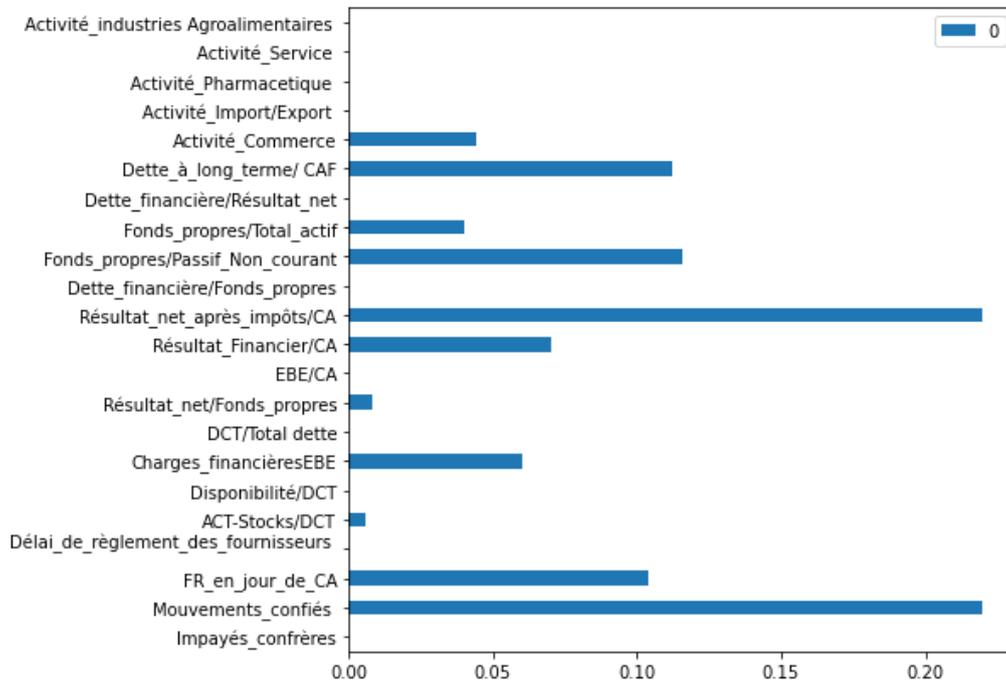
16 entreprises défailtantes sur 28 sont correctement détectées, avec le modèle sans optimisation, contre 14 entreprises après l'optimisation, l'optimisation a amélioré la capacité du modèle, en revanche, le modèle a détecté plus d'entreprises saines, soit une seule entreprise saine incorrectement détecter

5.4. Décision :

Il est clair que, le modèle après l'optimisation présente des paramètres de performance meilleur, il est capable de généraliser, ceci ne va pas impacter la capacité du modèle à prédire de nouvelles données ce qui va engendrer des faux résultats, donc nous avons choisis le modèle après optimisation, les variables les plus importantes dans le modèle d'adaboost sont :

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Figure 36: les variables importantes pour le modèle Adaboost



Source : établis par nous même avec google colab

« Le mouvement confié » et le ratio « résultat net après impôts/CA », présentent les variables les plus importantes, avec un taux d'importance de 22%, suivi par les ratios « dettes à long terme /CAF », « fonds propres / passif non courant », avec un taux d'importance de 11%, le ratio « charges financière /EBE », et le ratio « résultat financier / CA » ont un taux d'importances de 7%, le reste des variables présentent un taux d'importance faible inférieure à 5%.

6. Le forêt aléatoire (Random forest) :

Nous faisons varier les hyperparamètres de modèle :

Tableau 11: les hyperparamètres du modèle forêt aléatoire

	Hyperparamètres avant l'optimisation	Hyperparamètres après l'optimisation
N_estimators	100	56
Max_features	Auto	Auto

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Max_depth	None	4
Min_samples_split	2	1
Min_samples_leaf	1	2
bootstrap	True	True

Source : établis par nous même à partir des résultats de google colab

Le nombre d'arbre de décision utiliser dans la forêt désigné par **n_estimators** est de 25 après optimisation, le **max_features**, désigne le nombre de caractéristique à tester dans la construction de chaque nœud, il est par défaut auto, la profondeur maximale de la forêt (**max_depth**) est de 4 après l'optimisation, **min_samples_split** optimale est de 5, et le **min_samples_leaf** est de 2, le **bootstrap**, est qui désigne si les échantillons pris par la stratégie bootstrap sont employés en construisant des arbres ou non.

6.1. Les rapports de classification :

L'analyse des rapports de classification nous donne :

Figure 37:les rapports de classification avanies et après l'optimisation du modèle forêt aléatoire

	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.75	0.96	0.84	48		0.75	0.96	0.84	48
1	0.87	0.46	0.60	28		0.87	0.46	0.60	28
accuracy			0.78	76				0.78	76

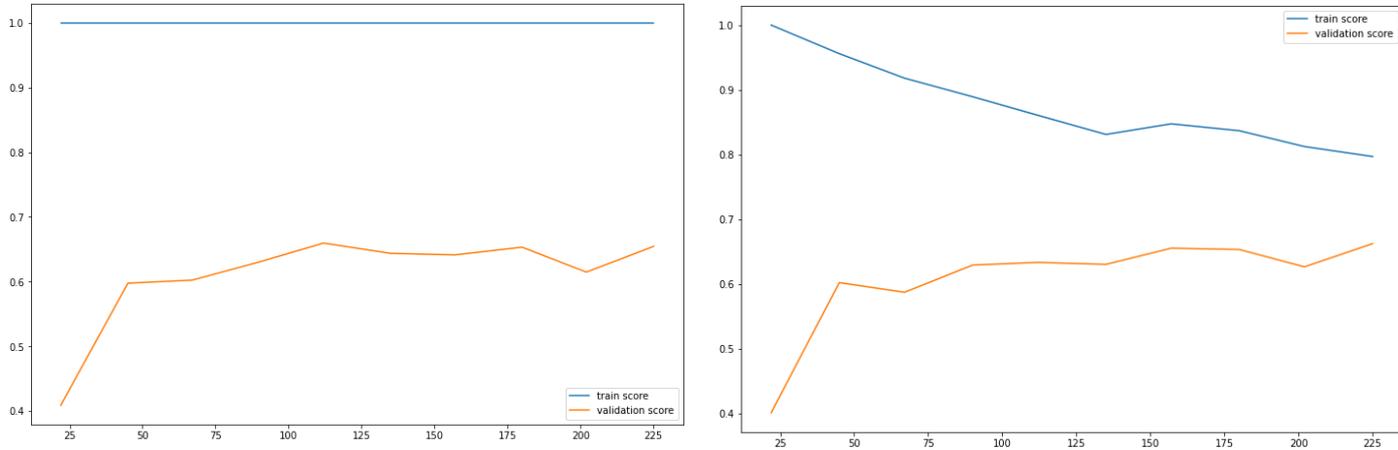
Source : établis par nous même avec google colab

Il est clair que l'optimisation des hyperparamètres n'apporte aucun changement dans la performance du modèle, le taux de succès est de 78%, le f1_score des classes défaillantes est de 60%, dans les deux situations.

6.2. Les courbes d'apprentissage :

Les courbes d'apprentissage confirment cette analyse :

Figure 38: les courbes d'apprentissage avant et après l'optimisation du modèle forêt aléatoire



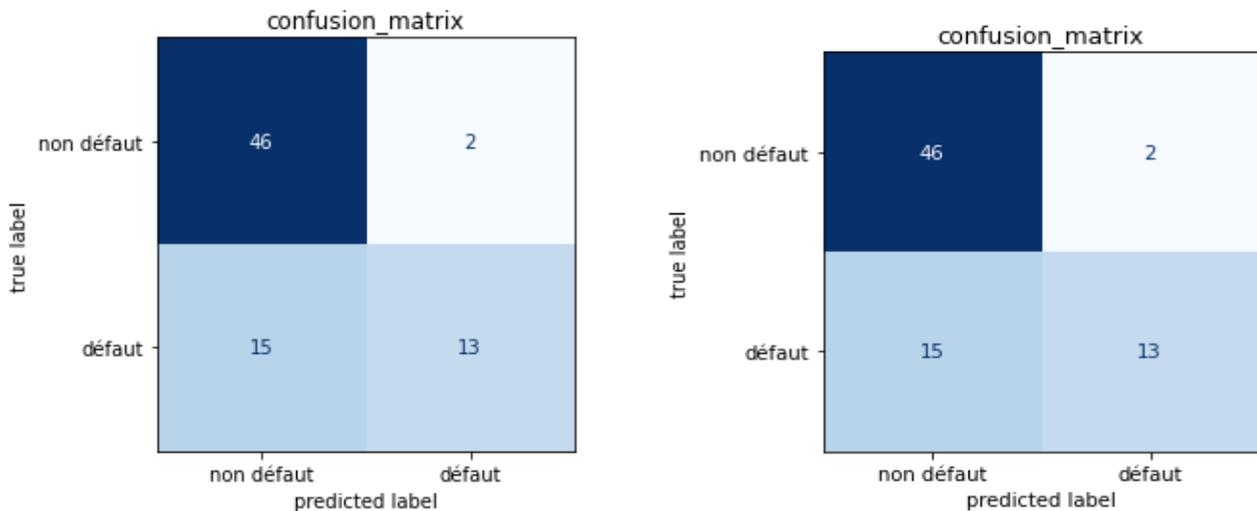
Source : établis par nous même avec google colab

Le modèle par défaut souffre d'un surapprentissage, la courbe de construction est une ligne droite, ce qui signifie que le modèle est bon dans les données de construction, il se base uniquement sur ces données initiales d'apprentissage. Il souffre alors d'un problème de généralisation et ne pourra pas reconnaître d'autres données, après l'optimisation, les courbes de construction et de validation commencent à se rejoindre, ce qui indique que l'optimisation des hyperparamètres a amélioré la robustesse du modèle.

6.3. Les matrices de confusion :

Les matrices de confusion montrent :

Figure 39: les matrices de confusion avant et après l'optimisation du modèle forêt aléatoire



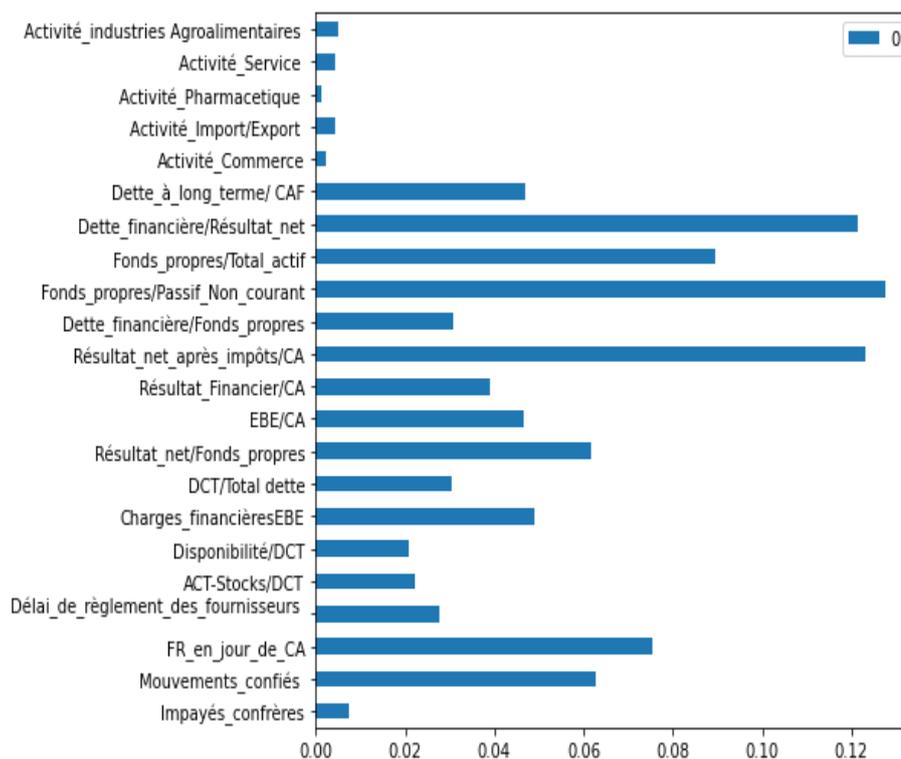
Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

Malgré l'amélioration de la robustesse, mais le modèle prédit toujours le même nombre d'entreprises défailtantes, 13 entreprises parmi les 28 défailtantes, et le même nombre des entreprises saines.

6.4. Decision :

A partir de ces analyses, nous pouvons dire que le modèle après optimisation est plus performant, les variables importantes pour ce modèle :

Figure 40: les variables importantes pour le modèle forêt aléatoire



Source : établis par nous même avec google colab

Le ratio « fonds propres/ passif non courant », constitue la variable la plus importantes, avec un taux d'importance de 13%, les deux ratios « dettes financière / résultat net » et « résultat après impôt / CA » présentent un taux d'importance de 12%, nous trouvons ensuite « les fonds propres /total actif », et la variable « mouvement confies » avec un taux d'importance de 9% et 6% consécutive. Le ratio « fonds propres / passif non courant » présente le même taux d'importance de 6%. Le reste des variables explicatives utilisés dans la construction de modèle, ont un taux d'importance qui ne dépasse pas les 5%.

Chapitre II : L'implantation des modèles de classification basés sur l'apprentissage supervisé

7. Le choix du meilleur modèle :

Il y a lieu, par ailleurs, de comparer la performance des modèles achevés pour la sélection du meilleur modèle. Où nous avons appliqué les mesures de performance numérique et graphique.

Le tableau ci-dessous regroupe l'ensemble des modèles déjà choisis :

Tableau 12: comparaison des modèles

Modèle	Classe	Précision	Sensibilité	F1_score	Taux de succès
L'arbre de décision	0	74%	83%	78%	71%
	1	64%	50%	56%	
Régression logistique	0	77%	90%	83%	76%
	1	75%	63%	63%	
SVM	0	75%	96%	84%	78%
	1	87%	46%	60%	
KNN	0	75%	85%	80%	72%
	1	67%	50%	57%	
Adaboost	0	77%	98%	86%	80%
	1	93%	65%	65%	
Forêt aléatoire	0	75%	96%	84%	78%
	1	87%	46%	60%	

Source : établis par nous même à partir des résultats de google colab

Pour rappel, nous avons pris le modèle arbre de décision, et le modèle Adaboost après l'optimisation des hyperparamètres, les modèles régression logistiques, SVM, et KNN sont prises par les hyperparamètres par défauts, quant au modèle du forêt aléatoire, l'optimisation n'a rien changé.

D'après le tableau, nous pouvons remarquer que le modèle Adaboost présente le modèle le plus performant, avec un taux de succès de 80%, soit le taux le plus élevé par rapport au reste des modèles, et de même il a le meilleur f1_score des classes défailites. En faisant recours à l'évaluation graphique à travers la courbe ROC de chaque modèle (Annexe 3), et avec comparaison du l'aire sous la courbe pour chaque modèle à l'aide de auc score (Annexe 4), nous trouvons que le modèle Adaboost est le plus performants avec une superficie sous la courbe de 80%.

8. Application Web :

Pour le rendu de notre travail, nous avons opté pour la mise en place d'une application web, développée avec **Streamlit**,

Figure 41: Application Web

The screenshot shows a web application interface. On the left, there is a sidebar with the logo of 'Crédit Populaire d'Algérie' and the text 'القرض الشعبي الجزائري'. Below the logo are several input fields with dropdown menus and numeric inputs. The main content area on the right has the title 'Prédiction de défaut' and 'Crédit Populaire d'Algérie'. Below the title is a button labeled 'Exécuter'.

Source : elabore par nous même à partir de logiciel paycharm

Nous avons utilisé le logiciel paycharm, et avec Streamlit, qui est une librairie python open source permet des créé les applications web, nous avons créé cette application.

L'application est accessible sur le lien :

<https://hayem-ifid-streamlitappscoring-main-qak0de.streamlit.app/>

Conclusion

Après avoir réalisé l'étape de prétraitement des données, celle-ci s'avère une étape importante, puisque c'est de la pertinence de son déroulement dont dépend les résultats que nous avons présentés. Nous avons implanté les différents modèles d'apprentissage supervisés.

Ainsi, l'analyse des résultats nous a permis de déterminer le meilleur modèle, à travers une comparaison par des techniques numériques et des techniques graphiques.

La comparaison des résultats a abouti à plusieurs constatations, dans un premier temps, certains modèles sont plus performants avec leurs hyperparamètres par défaut, pour certain modèle l'optimisation a amélioré la performance.

Parmi les modèles d'apprentissage de classification utilisés, le modèle Adaboost est le plus performant, les résultats de ce modèle ont été utilisés ensuite pour développer une application web à l'aide de streamlit

Conclusion générale

Vu l'importance de l'activité bancaire dans l'économie contemporaine, la mise en place et le renforcement des exigences est nécessaire pour garantir le bon fonctionnement de celle-ci. C'est dans ce sens que s'inscrit la réglementation bâloise, qui impose aux banques un ensemble des normes de gestion prudentes qui sert à mieux mesurer les risques bancaires pour une solidité et stabilité du secteur bancaire.

L'Algérie, a s'inspiré des travaux du comité de Bâle pour établir ses normes prudentielles, qui exige aux banques et établissement financier le respect des règles de gestion, néanmoins l'évaluation du risque de crédit a été toujours un enjeu essentiel pour tous les créanciers. Ces institutions sont incitées à le gérer d'une manière efficace. Pour se faire, les banque cherchent depuis longtemps des méthodes et techniques efficace afin de répondre à ce problème. Ces techniques ont évolué ce qui engendre le développement de nouveaux outils.

Cependant, les banques algériennes utilisent toujours les méthodes classiques qui se basent sur l'analyse financière, dans la mesure du risque de crédit, l'adoption de nouvelles techniques s'avère obligatoire. Dans ce cadre, ce travail s'est intéressé à la mise en place d'un outil plus développé pour l'évaluation du risque de crédit.

Au cours de notre travail nous avons essayé de cerner notre thématique dans un cadre théorique repartie en deux chapitres, l'exploration théorique du premier chapitre qui sur la mesure du risque de crédit, nous a permis de comprendre les différentes méthodes d'évaluation conformes aux exigences de régulateurs. Cette exploration s'appuie également sur les composantes d'un risque de crédit.

Notre investigation théorique nous a également dans le deuxième chapitre permis de savoir les procédures de gestion du risque de crédit. Cela permet d'approfondir nos connaissances concernant les outils de couverture contre ce dernier.

A cet effet, on a choisi un cas pratique émanant du crédit populaire d'Algérie. L'intérêt a porté sur le crédit scoring comme outil de mesure et gestion. Partant du développement qu'a connu le monde de l'analyse des données, nous avons utilisé le « machine Learning » comme techniques d'élaboration d'un modèle de score. Ce choix est soutenu par la contribution des

algorithmes de l'apprentissage automatique avec leur performance dans la mesure du risque de crédit, ce qui a donné lieu à un meilleur classement avec précision des clients.

La partie pratique que nous avons menée se divise en deux chapitres, un chapitre qui décrit la méthodologie de construction des modèles de classification et un chapitre pour interpréter l'ensemble des résultats trouvés. Pour la construction des modèles et vu notre incapacité de prendre un nombre important des clients vu le manque des données au niveau de lieu de stage nous nous sommes basés sur un échantillon de 387 PME, collecté auprès de la direction des crédits aux industries et aux services.

L'ensemble des données de notre échantillon a été prétraité avant la sélection finale des variables explicatives à l'aide de plusieurs méthodes classiques telles que le chi deux et le test Anova. Ensuite, les données ont été entraînées et testées avec un nombre multiple de modèles d'apprentissage automatique supervisés, tels que la régression logistique, l'arbre de décision, Support Vector Machine, la méthode des k plus proches voisins, la forêt aléatoire, et le modèle Adaboost. Une technique de validation croisée a été réalisée et les paramètres des algorithmes ont été correctement réglés pour un meilleur travail des modèles. La performance a été évaluée à l'aide de plusieurs mesures de performance telles que la matrice de confusion, le rappel, la précision, le F1 score, la courbe ROC, et le score AUC.

Les modèles ont été comparés selon plusieurs mesures et après Toutes les comparaisons nous avons sélectionnées le meilleur modèle. Le Adaboost a montré des meilleurs résultats, ce modèle a été utilisé ensuite pour le développement d'une application web en étant un outil d'aide à la classification des PME demandeurs de crédits.

La qualité de cette application n'est pas parfaite, et sa mise en application comme un outil d'aide à la décision pour le CPA est envisageable mais nécessite la prise en considération de certaines recommandations pour améliorer la qualité des modèles utilisée à savoir :

- Exiger aux PME demandeurs de crédit de fournir des documents financiers certifier pour éviter toutes informations fallacieuses et donc réduire les valeurs aberrantes ceci va améliorer la performance des modèles utilisés.
- Les banques algériennes sont tenues de travailler sur la construction des bases de données numériques qui regroupent l'ensemble des clients, chose qui va faciliter la disponibilité de l'information pour la modélisation de crédit.

- Il y a lieu, par ailleurs, de prendre en compte les dossiers de crédits refusés initialement pour optimiser le nombre des clients potentiels et donc atteindre une meilleure classification.

Nous tenons toutefois à préciser que notre étude peut encore être améliorée. En il y a encore de nombreuses perspectives :

- Il faut penser à d'autres méthodes d'évaluation pour l'utilisation d'un différent type d'apprentissage à savoir l'apprentissage non supervise, ou encore le deep learning.
- Il est intéressant de construire des nouvelles bases de données qui comporte en plus de variables quantitative qui décrivent la situation financière du client, des variables qualitatives qui expriment le comportement du client. Ce qui permettrait de concevoir des approches autres que le scoring d'octroi et développer le scoring de comportement de l'emprunteur, ce qui aide davantage les banques dans l'assurance de leurs suivie.

Bibliographie :

Ouvrages :

A. DE SERVIGNY ; **Risque de crédit**, 4^{ème} édition ; DUNOD.

B. FUEL CELLS et autres, **advanced in renewable energies and power technologies**, ELSEVIER, Netherlands, 2018.

C. KHAROUBI et P. THOMAS, **Analyse du risque de crédit**, 2^{ème} édition, revue banque édition, Paris, 2016.

C.Gouriéroux et A.Tiomo, **risque de crédit une approche avancés**, les cahiers de CREF, avril 2007.

G.saint-cirgue, **apprendre le machine learning en une semaine**, 2019.

J.HULL et autre ; **gestion des risques et institutions financières** ; 2^{ème} édition ; pearson education ; 2007.

K.BOUAISS et autres, **Economie et gestion de la banque**, édition management et société.

M.DIETSCH et J. PETEY ; **Mesure et gestion du risque de crédit dans les institutions financières** ; édition revue banque ; 2^{ème} édition ; 2008.

S.De GOUSSERGUES et autres, **Gestion de la banque**, 8^{ème} édition, dunod,2017.

S.De GOUSSERGUES et autres, **Gestion de la banque**, 9^{ème} édition, dunod,2020.

V.BRUNEL ; **Gestion des risques et de crédit** ; economica ; 2019.

Dictionnaire :

A.BEITONE, **Dictionnaire des sciences économiques**, 6^{ème} édition, dunod,,2019.

Articles:

A.BENTHAMI, **Les limites de la notation interne appliqué par les banques aux petites et moyennes entreprise : cas du maroc**, revue techniques financière et développement, N°124, mars 2016.

C.Hurlin, et C.pregnon, **MACHINE LEARNING ET NOUVELLES SOURCES DE DONNÉES POUR LE SCORING DE CRÉDIT** , Revue d'économie financière, N°135, mars, 2019.

Comité de Bâle sur le contrôle bancaire ; **Convergence internationale de la mesure et des normes de fonds propres**; banque des règlements internationales ; juin 2004.

E. SPUCHLAKOVA, The Credit Risk and its Measurement, Hedging and Monitoring, ICOAE,N°2, juillet, 2015,p 676.

H.FEKIR ; **La crise financière et l'application des accords de Bâle : est-ce le meilleur remède ?** ; Université Ibn Tofail, La Faculté des Sciences Juridiques, Economiques et Sociales de Kénitra, Maroc ; may 2010 ; p 7.

J.GONZALEZ, **EXPLORATION DES ARBRES DE DÉCISION ET DES SUPPORT VECTOR MACHINES EN VUE D'APPLICATIONS DANS L'ANALYSE DE TEXTE**, mémoire pour l'obtention de la maîtrise en mathématique et informatique appliqué, université du québec, juin 2016.

M.BELLALAH et S.STRTK, **réflexion sur les décisions de crédit, le risque de défaut, et l'évaluation de la perte en cas de défaut**, la revue des sciences de gestion, N°294, juin 2018.

M.MERHOUN et I. BENHALIMA ; **Les Banques Algériennes A La Lumière Des Règles De Bale I, Ii, Iii : Bilan Et Perspectives** ; revue des reformes économique et l'intégration dans l'économie mondiale, N°23 ; juin 2017.

P.THORAVALL ; **Le dispositif de Bâle II : rôle et mise en oeuvre du pilier 2** ; Revue de la stabilité financière ; N° 9 ; Décembre 2006.

S. LOTFI et S. BENSALIDA ; **Mesure des impacts de la norme IFRS 9 sur le risque de crédit bancaire** ; Revue du Contrôle de la Comptabilité et de l'Audit ; N°5 ; Juin 2018.

S.Chebaiki et KH.Slimani ; **Les Accords de Bale ou L'Évolution des Règles Prudentielles** ; revue Djadid El-iktissad ;N°09 ; décembre 2014.

S.Haloui ; **Accord de Bâle IV : Quels apport et impact pour l'industrie financière et bancaire à l'ère du COVID-19?**; IJAFAME ; N°2 ; novembre 2021.

T.Bahzad, et A. Mohsin Abdulazeez , **Classification Based on Decision Tree Algorithm for Machine Learning**, journal of applied science and technology trends, vol2, N°1, mars 2021

T.QUAMAR, **Les systèmes internes de notation : Architecture et enjeux sur la relation banque-PME marocaines**, Revue du Contrôle de la Comptabilité et de l'Audit, N°6, septembre 2018.

Thèses et mémoire :

ALIANE.S et AMRI H ; **la réglementation prudentielle en Algérie et son niveau de conformité avec les standards de Bâle 1 et 2** ; mémoire en vue de l'obtention de diplôme de Master ; Option : Monnaie, Banque et environnement International ; Faculté des Sciences Economiques, sciences de gestion et sciences commerciales ; Université Abderrahmane Mira ; Bejaia ; Algérie ; juin 2013.

A. SAMBA, **Science des données au service des réseaux d'opérateur**, thèse de doctorat, l'école nationale supérieure mines-télécom 2018.

B.Garnier ; **Mesure du risque de crédit d'un portefeuille** obligataire ; Mémoire présenté le : pour l'obtention du diplôme de Statisticien Mention Actuariat et l'admission à l'Institut des Actuaires ; Institut de statistique (ISUP) - Sorbonne Université ; 2013

F. Francisco et A.Caixeiro, **Infrastructure and Machine Learning for Credit Scoring**, memoire pour l'obtention d'un master en informatique et telecommunication, técnico lisboa, novembre 2018.

J.Hallblad ; **The Multi-year Through-the-cycle and Point-in-time Probability of Default** ; memoire d'un master en economies; UMEA universitet; 2014

R.SUBLET ; **La gestion du risque de crédit bancaire sur les portefeuilles professionnels et particuliers** ; mémoire pour l'obtention du diplôme de master ; école de commerce de lion ; 2015

W.AICHOUNI ; **La gestion du risque de liquidité bancaire : cas de la Banque d'Agriculture et de Développement Rural** ; mémoire pour l'obtention du diplôme de magister en sciences commerciales et financières ; option finance ; école supérieur de commerce ; Alger ; 2015.

Rapport :

R.HENNANI ; **De Bâle I à Bâle III: les principales avancées des accords prudentiels pour un système financier plus résilient** ; rapport technique, janvier 2015

B.Diallo, **UN MODELE DE "CREDIT SCORING" POUR UNE INSTITUTION DE MICRO-FINANCE AFRICAINE: LE CAS DE NYESIGISO AU MALI** , Laboratoire d'Economie d'Orléans (LEO), Université d'Orléans, may 2006.

Dispositifs :

Comité de Bâle sur le contrôle bancaire, **Finalisation de Bâle III**, banque des règlements internationaux, Décembre 2017.

Comité de Bâle sur le contrôle bancaire, **Bâle III : dispositif réglementaire mondial visant à renforcer la résilience des établissements et systèmes bancaires**, banque des règlements internationaux, Décembre 2010.

Table des matières :

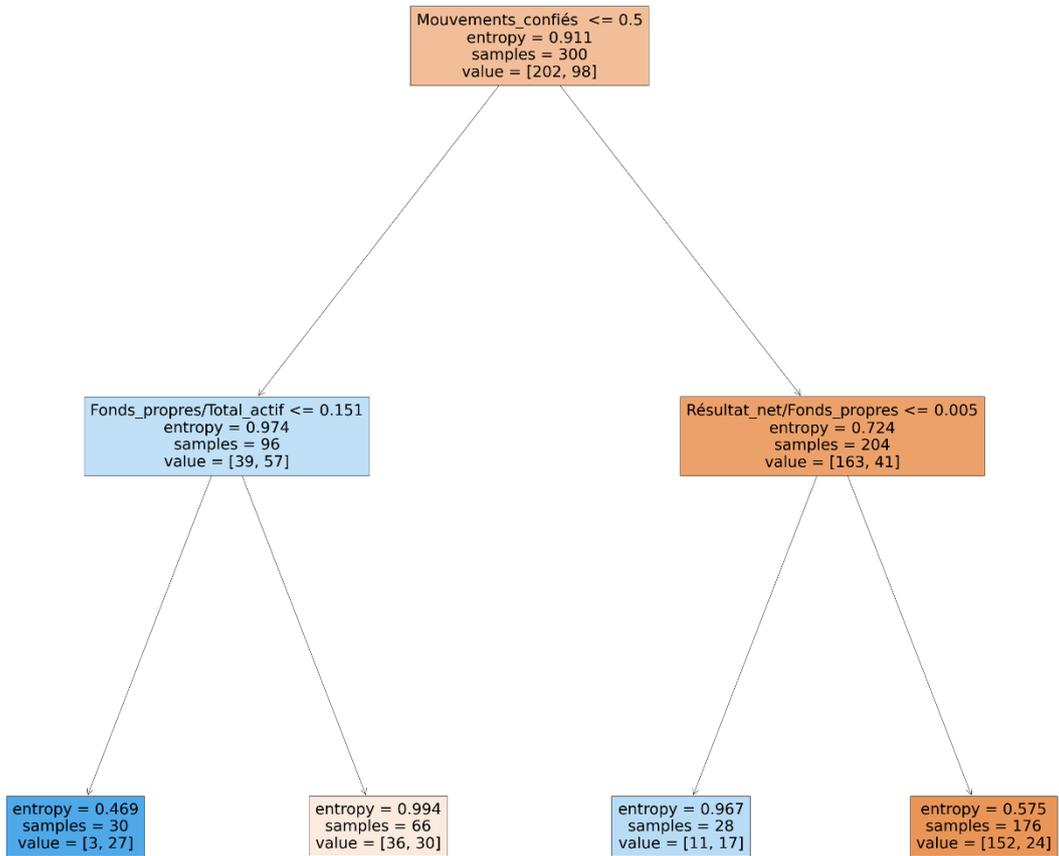
<i>Dédicaces</i>	i
<i>Remerciements</i>	ii
Résumé	iii
Liste des abréviations :	v
Liste des figures :	vi
Liste des tableaux :	vii
Liste des annexes :	vii
Introduction générale	A
<i>Partie théorique</i>	1
Chapitre I:Mesure du risque de crédit	2
Introduction	1
Section 1 : cadres conceptuels	2
1. Définition du risque de crédit :.....	2
2. La décomposition du risque de crédit :	3
3. Les paramètres du risque de crédit :	4
4. Les facteurs de risque de crédit :	6
Section 2 : La mesure du risque de crédit	10
1. La mesure du risque de crédit au niveau individuel :	10
2. La mesure du risque de crédit au niveau de portefeuille :.....	16
3. L'architecture d'un modèle du risque de crédit :.....	19
Conclusion	21
Chapitre II:Gestion du risque de crédit	22
Introduction	23
Section1 : la gestion prudentielle	24
1. La réglementation prudentielle baloise :.....	24
2. La réglementation prudentielle au niveau nationale :	34
3. Les instruments IFRS 9 :	40
Section 2 : les techniques de gestion de risque de crédit	41
1. Le processus de gestion des risques :	41
2. Les méthodes de gestion des risques :.....	44
3. Couverture des risques :	46

Conclusion	49
Partie pratique	50
Chapitre I:Méthodologie de construction d'un modèle de classification	51
Introduction	52
Section 1 : revue de littérature	53
1. Crédit scoring :	53
2. Machine learning :	56
Section 2 : méthodologie de traitement des données	59
1. Les bibliothèques clés de machine learning :	59
2. L'environnement de développement pour python :	61
3. Préparation de la base données :	61
4. Les modèles prédictifs :	63
5. Les techniques d'évaluation :	68
6. La validation des modèles :	71
Conclusion	72
Chapitre II:L'implantation des modèles de classification basés sur l'apprentissage supervisé	73
Introduction	74
Section 1 : préparation des données	75
1. Choix et description des données de l'échantillon :	75
2. Les pré-traitements des données :	77
Section 2 : La construction des modèles	86
1. L'arbre de décision :	86
2. La régression logistique :	90
3. Le sector vector machine :	94
4. Le K-neighbor classifier :	97
5. Adaboost :	99
6. Le forêt aléatoire (Random forest) :	102
7. Le choix du meilleur modèle :	106
8. Application Web :	107
Conclusion	108
Conclusion générale	109
Bibliographie :	112

Table des matières :..... 116

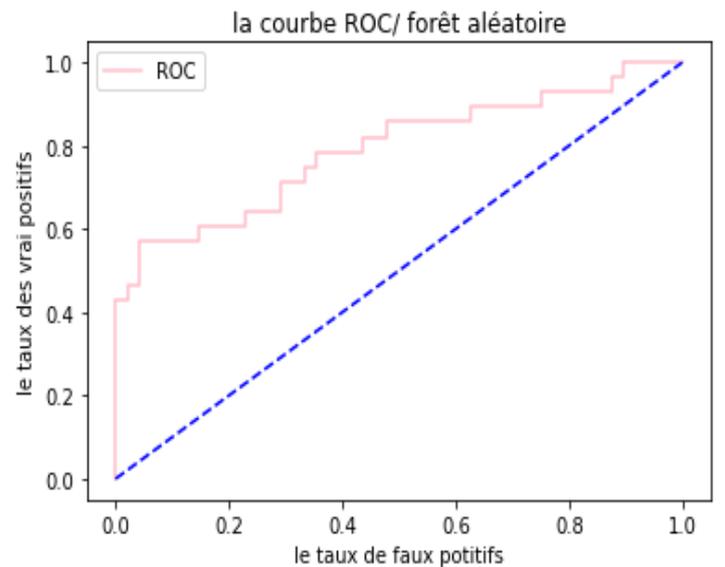
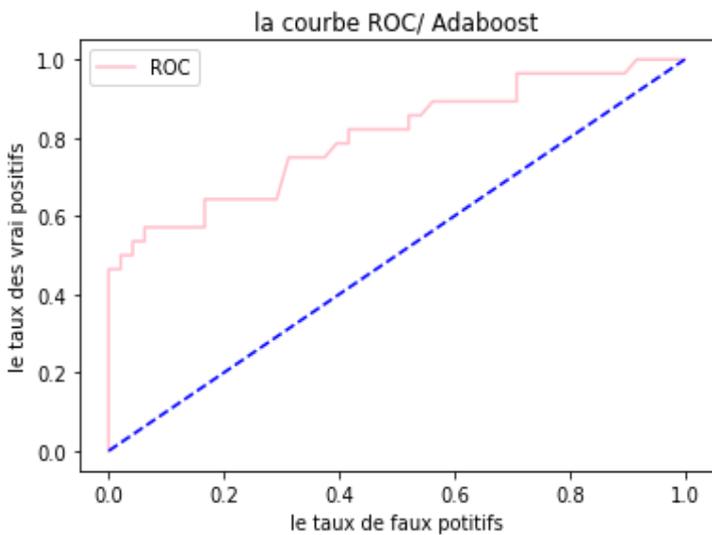
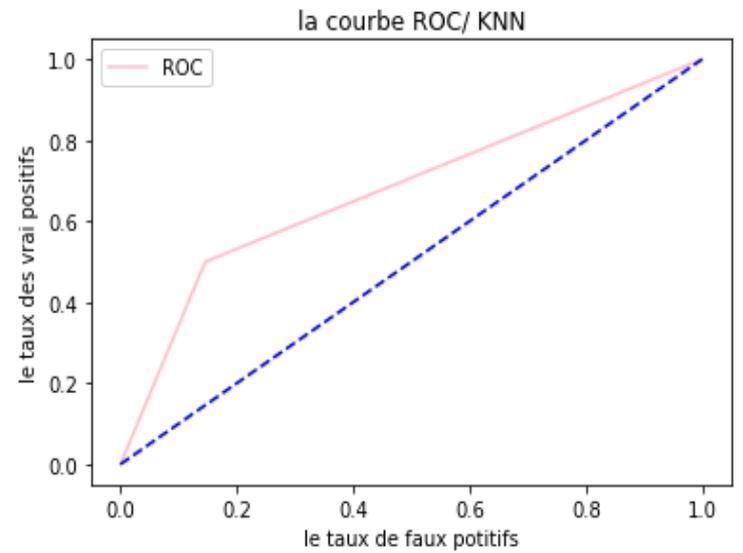
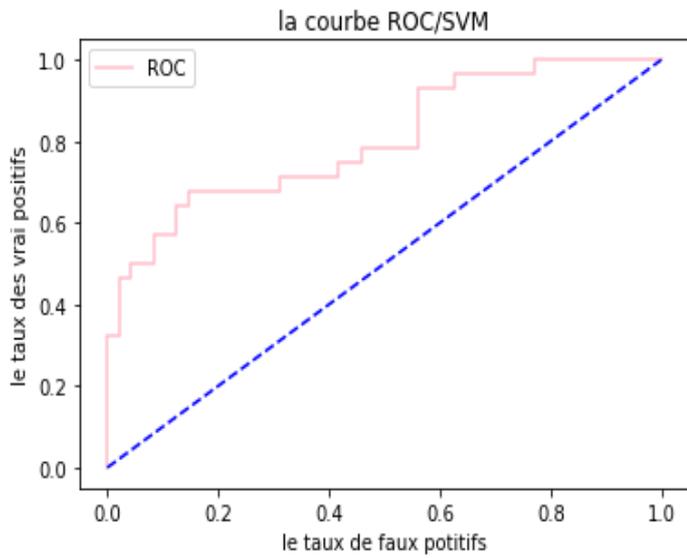
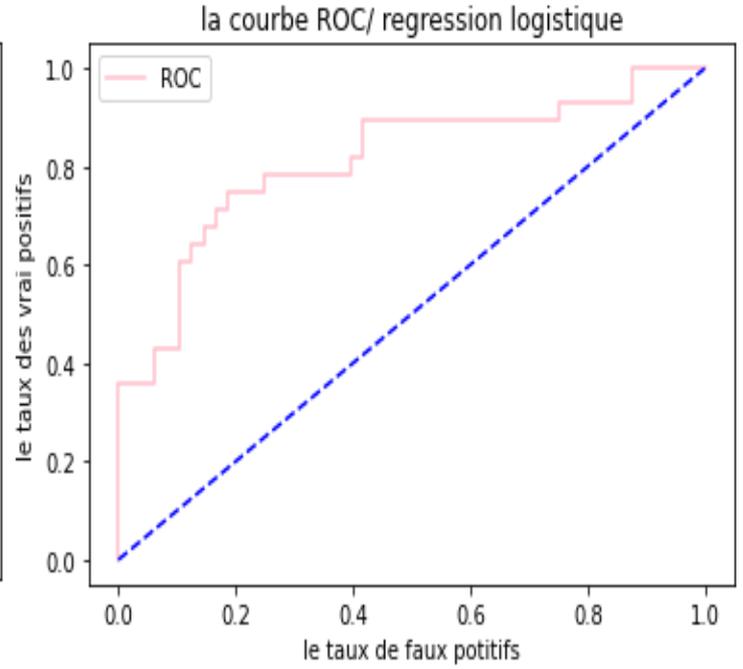
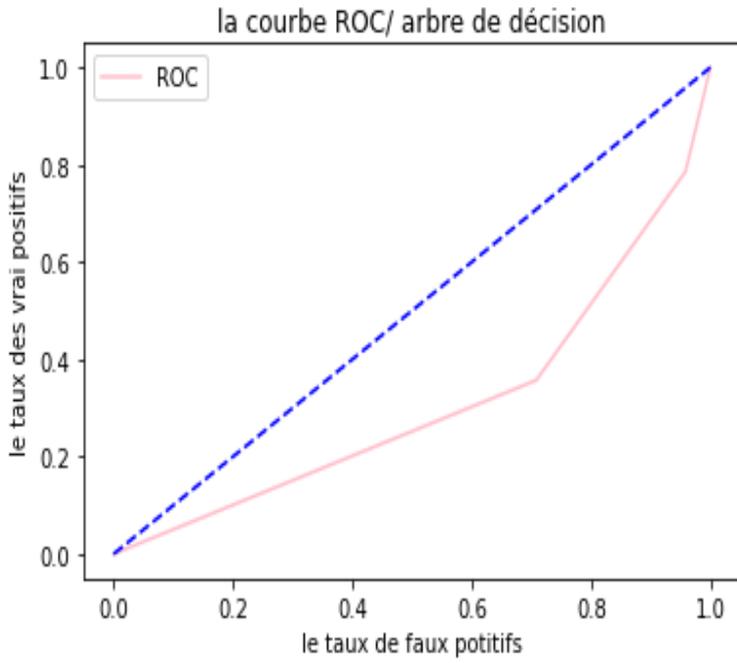
Annexes : 119

Annexe 2: Arbre de décision



Source : établis par nous même avec google colab

Annexe 3: courbe ROC des modèles



Annexe 4: l'aire sous la courbe pour chaque modèle

Modèle	AUC
L'arbre de décision	30%
La régression logistique	30%
Le SVM	80%
Le KNN	67%
Adaost	80%
L e forêt aléatoire	79%

Source : élaborer par nous même à partir des résultats de google colab