

L'apport des variables qualitatives dans l'amélioration de la prévision du risque de défaillance des entreprises

Mohamed Sofien NOURI^{1*}

URECA, Faculté des sciences économiques et de gestion de Sfax, Tunisie

¹ *Docteur en sciences économiques,*

*medsofiennouri@gmail.com

Younes BOUJELBENE^{2**}

² *Professeur et Doyen de la FSEG Sfax*

**younes.boujelbene@gmail.com

Résumé—Les travaux se rapportant à la prévision de la défaillance des entreprises sont fortement intensifiés ces dernières années utilisant surtout les données financières des entreprises et les outils avancés en matière de classification telle que les techniques non paramétriques. Portant du constat que la majorité de ces études sont intéressés en général à l'analyse des ratios financiers des entreprises, qui ne présentent réellement que les symptômes initiaux du processus du déclin, notre étude vise à explorer le processus de défaillance dans ses différentes étapes. La volonté de mieux comprendre ce processus nous pousse à élargir le champ d'expertise par l'intégration des aspects d'ordre qualitatif, relevant d'avantage des modes d'organisation de l'entreprise en relation avec son environnement interne et externe.

Le principal objectif des auteurs dans ce papier est d'améliorer le degré de précision en termes de détection précoce de la défaillance et d'identifier les déterminants de la défaillance des PME tunisiennes. Afin d'atteindre ce but le champ d'expertise est élargi par à travers l'utilisation des variables qualitatives à côtés des variables quantitatives tout en déployant des nouvelles méthodes statistiques dotées de l'intelligence artificielle comme les réseaux de neurones et les méthodes SVM. Cette démarche permet une meilleure explication des risques de défaillance et une meilleure précision dans l'identification des signes précurseurs de la faillite.

En termes opérationnels, cette recherche a pour but d'apporter une contribution à l'intention des chefs d'entreprise, des investisseurs et à tous les meneurs de décision au sein ou en dehors de l'entreprise. Il s'agit, en fait d'identifier les causes de défaillance, de proposer une vision claire des techniques de prévision et de montrer l'apport des variables qualitatives en termes de prévision de la défaillance des entreprises.

Abstract— Work relating to the prediction of business failure has been intensified in recent years, mainly using business financial data and advanced

classification tools such as non-parametric techniques. Noting that the majority of these studies are generally interested in the analysis of the financial ratios of companies, which really only present the initial symptoms of the process of decline, our study aims to explore the process of failure in its various stages. The desire to better understand this process prompts us to broaden the field of expertise by integrating qualitative aspects, relating more to organizational methods of the company in relation to its internal and external environment.

The main objective of the authors in this paper is to improve the degree of precision in terms of early detection of failure and to identify the determinants of failure of Tunisian SMEs. In order to achieve this goal the field of expertise is widened through the use of qualitative variables alongside quantitative variables while deploying new statistical methods equipped with artificial intelligence such as neural networks and SVM methods. This approach allows for a better explanation of the risks of default and better precision in identifying the warning signs of bankruptcy.

In operational terms, this research aims to make a contribution to the benefit of business leaders, investors and all decision makers within or outside the company. It is, in fact, to identify the causes of failure, to offer a clear vision of forecasting techniques and to show the contribution of qualitative variables in terms of forecasting business failure.

Mots clés: Prévision, défaillance des entreprises, variables qualitatives, Analyse en composante principale, réseau de neurones, SVM.

I. INTRODUCTION

L'environnement économique actuel est caractérisé par une grande turbulence et une forte incertitude. En effet, la libéralisation des échanges, la mondialisation et les progrès

technologiques réalisés ont accentué la concurrence à l'échelle nationale et internationale et ont contraint les entreprises à déployer plus d'efforts pour puisse s'adapter et survivre. Devant cette concurrence intense et ces mutations accrues au niveau de l'environnement, les entreprises sont soumises à des difficultés continues dès leur création jusqu'à leur arrêt. Celles qui parviennent à anticiper aussi tôt les difficultés potentielles s'en sortent généralement sans grand dégâts, en revanche celles qui n'apprécient pas suffisamment tôt se trouvent dans des situations de faillite. L'augmentation du nombre des entreprises défaillantes, confirme la nécessité de chercher les causes et les outils permettant de détecter aussitôt ce phénomène afin de prendre les mesures nécessaires au temps opportun et limiter ses conséquences. En effet, Il est nécessaire de veiller à la protection des intérêts des créanciers, à la pérennité de l'entreprise, en détectant les difficultés que les entreprises peuvent rencontrer, ce qui exige une estimation du risque de défaillance et éventuellement une amélioration des méthodes d'évaluation.

La prévision de la défaillance de l'entreprise, est un thème de recherche qui a pris l'attention des chercheurs dans les domaines économiques, comptables et financiers depuis plusieurs décennies [1,2], et a connu une renaissance depuis la crise financière mondiale en 2008 (crise des subprimes)[3]. Le concept de prévision de la faillite d'entreprise est largement utilisé dans divers contextes : comptables, financiers et réglementaires, tels que le contrôle de la solvabilité par les institutions financières, l'évaluation de la santé de l'entreprise par les auditeurs, l'évaluation des emprunts accordés aux entreprises, l'évaluation de la fiabilité des entreprises envers ses engagements et l'analyse de portefeuille ([4,5,6]).

La santé de l'entreprise revêt un caractère très important pour ses créanciers, ses investisseurs, ses actionnaires, ses partenaires et même pour ses clients et ses fournisseurs. Conscients de cet état, les chercheurs et les praticiens se sont concentrés au développement des méthodes et des techniques qui leurs permettent d'avoir plus de performance et de précision dans la prévision de la faillite des entreprises.

Partant du constat que les recherches précédentes traitant la prévision de la défaillance des entreprises ont été focalisées essentiellement sur des variables de type quantitatif malgré l'importance qui revêt les variables de type qualitatif dans le déclenchement de la défaillance, notre objectif dans ce papier vise à élargir le débat par l'identification des déterminants de la défaillance des PME tunisiennes à travers l'utilisation des variables qualitatives à côtés des variables quantitatives pour la prévision de la défaillance des entreprises, en déployant des méthodes statistiques non paramétriques comme les réseaux de neurones et les méthodes SVM. Cette démarche permet une meilleure identification des signes précurseurs de la défaillance des entreprises et une meilleure estimation des risques de faillite.

II. REVUE DE LA LITTERATURE

Il existe une diversité significative, au niveau des études précédentes traitant le domaine de la prévision de la faillite des entreprises, en ce qui concerne les définitions de la faillite, la taille des échantillons, les types de modèles statistiques utilisés et le choix des variables explicatives.

A. LA NOTION DE FAILLITE

L'une des questions les plus traitées dans la littérature était la définition de la faillite d'une entreprise, un large éventail de définitions était utilisé dans divers études, notamment: liquidation ou liquidée par un tribunal [7]; cessation de paiement ou cessation des activités commerciales [8], la détresse financière [9], échec ou défaut [10], liquidation et inactivité [11]. En général, toute entreprise peut cesser ses activités pour de nombreuses raisons, parmi lesquelles: éviter de nouvelles pertes si elle n'arrive pas à couvrir ses charges variables, la retraite, les pénuries de liquidités, la cessation de paiement et divers autres raisons. Les incohérences dans la définition de la faillite d'une étude à l'autre peuvent limiter l'interprétation et la généralisation des résultats empiriques. Pour les besoins de cette étude, le critère adopté relatif à la défaillance est l'adhésion de l'entreprise aux régimes de sauvetages ou l'ouverture d'une procédure judiciaire à son encontre.

B. ECHANTILLON

La plupart des études traitant le sujet de prévision de la défaillance des entreprises s'appuient essentiellement sur les données financières comme principales variables d'intérêt (voir par exemple [12], [13], [14]). D'autres facteurs non financiers examinés de manière ponctuelle comprennent l'âge, la taille, l'industrie et la région [15], notations de crédit internes [16], la structure managériale, l'insuffisance du système d'information comptable et les retards d'élaboration des rapports d'audit [17], stratégie, relations avec les banques, tarification, marketing, caractéristiques des propriétaires et qualité de la main-d'œuvre [18]. Cependant, ces facteurs qualitatifs sont difficiles à mesurer et les données sont difficiles à rassembler. Au niveau de cette étude, on a opté à la diversification des sources de données. Les variables qualitatives utilisés ont été sélectionnés sur la base d'un questionnaire adressé aux responsables des entreprises. (Voir questionnaire en annexe)

C. LES MODELES STATISTIQUES

Le thème de prévision de la faillite des entreprises a débuté avec l'étude pionnier de [19]. Depuis cette recherche, on a assisté au développement de plusieurs outils statistiques essayant à modéliser la défaillance. Ce thème peut être apprécié comme un problème de classification entre deux groupes: groupe d'entreprises saines et groupe d'entreprises défaillantes. Deux types de méthodes ont marqué l'histoire des modèles de prévision, les méthodes paramétriques et celles non paramétrique. En ce qui concerne les techniques paramétriques, la majorité des études se sont appuyées sur

des modèles de choix discrets standard, tels que l'analyse discriminante multiple (MDA) et les modèles logit /probit [20]. Pour les méthodes non paramétriques de classification traitant ce domaine, trois techniques prépondérantes ont été couramment utilisées: les réseaux de neurones, les arbres de décision, et les machines à vecteurs de support (SVM).

La méthode des réseaux de neurones a marqué le champ d'étude de ce domaine à partir des années 90. Cette méthode est l'une des méthodes les plus populaires en apprentissage automatique et qui s'appuie sur l'inspiration de méthodes de calcul inspirée du fonctionnement neuronal dans le cerveau humain [21]. Plusieurs types de réseaux présentant des architectures différentes ont été appliqués dans le domaine de prévision de la faillite (par exemple, les perceptrons multicouches [22], réseaux de neurones à rétro-propagation [23] et les réseaux de neurones probabilistes [24]. Plusieurs études comme ([25] [26], [27], [28], [29], [30]) ont essayé de tester la performance des réseaux de neurones en matière de prévision. Cependant malgré le grand succès et le degré de précision des modèles des réseaux de neurones, ils sont souvent critiqués comme "boîte noire" [31].

A partir de la fin des années 2000, la méthode SVM fait son entrée et réalise un grand succès dans le domaine de la prévision de la défaillance [32]. Cette méthode a été caractérisée par plusieurs avantages : elle est marquée comme la technique la moins sensible aux données déséquilibrées [33]. De même cette technique est qualifiée par sa robustesse et sa performance en termes de prévision de la faillite [34]. Plusieurs travaux de recherche ont essayé de tester le degré de précision des méthodes SVM en terme de prévision ([35], [36], [37], [38], [39], [40]).

III. LE CHOIX METHODOLOGIQUE

Nous sommes amenés à construire un dispositif méthodologique basé à la fois sur une approche quantitative et qualitative. Pour l'approche quantitative, notre démarche comporte plusieurs étapes: construction de la base de données, la sélection des entreprises et le choix des indicateurs de défaillance. Quant à l'approche qualitative, un questionnaire a été adressé aux responsables des entreprises, au niveau duquel, on a donné aux dirigeants des entreprises la possibilité d'exprimer leurs opinions concernant les problèmes micro et macroéconomiques rencontrés par l'entreprise pendant son parcours (voir questionnaire en annexe).

A. Présentation de la base de données

La base de données est composée par des états financiers de 340 entreprises saines et défaillantes qui se rapportent aux années 2018 et 2019, ces états sont obtenus suite à la collaboration de l'Agence de Promotion de l'Industrie et de l'Innovation, de certains commissaires aux comptes et de quelques experts comptables. A la suite, un appariement des données était effectué afin de construire l'échantillon final composé des variables quantitatives en

faisant correspondre à chaque entreprise saine une entreprise défaillante de même taille. Après le retranchement de quelques états financiers contenant des données aberrantes, la structure finale de l'échantillon se compose de 300 entreprises, qui se répartissent en part égale entre saines et défaillantes. Les ratios financiers issus des documents comptables constituent notre base de données relative aux variables quantitative, qui ont été choisis sur la base de leurs récurrence dans la littérature ([41], [42], [43], [44]). (Voir tableau I)

TABLEAU I : LES VARIABLES QUANTITATIVES CHOISIES

RATIOS	Code
La Rentabilité financière	R01
La Rentabilité économique	R02
La Rentabilité d'exploitation	R03
Le Rendement du capital investi	R04
La Marge de profit brut	R05
La Rotation de l'actif	R06
La Rotation des immobilisations	R07
La Rotation des capitaux propres	R08
La Rotation des stocks	R09
Le Délai de recouvrement des créances	R10
Le Délai de règlement des fournisseurs	R11
Le Taux de rentabilité	R12
Le Taux de croissance du chiffre d'affaires	R13
La Liquidité générale	R14
La Liquidité réduite	R15
La Liquidité immédiate	R16
La Liquidité de l'actif	R17
Le Taux d'endettement	R18
L'Endettement à MLT	R19
L'Endettement à court terme	R20
L'Autonomie Financière	R21
L'Equilibre financier	R22
L'Indépendance financière	R23
La Couverture des charges financières	R24
La Capacité de remboursement des dettes	R25
La Solvabilité générale	R26
Le Ratio des capitaux propres	R27
Le Ratio des immobilisations	R28
L'Indicateur de taille	R29

Notre base de données présentant les variables qualitatives a été construite sur la base d'un questionnaire adressé aux différentes entreprises. Cette base présente des variables caractéristiques de 200 entreprises: 120 entreprises saines et 80 défaillantes. Ce choix s'est imposé de lui-même, en raison des contraintes liées à la collecte d'informations et la non collaboration de certains dirigeants. Nous avons administré 300 questionnaires aux dirigeants, auditeurs et experts comptables chargés du dossier des entreprises (Voir questionnaire en annexe).

Nous avons reçu des réponses concernant 80 entreprises défaillantes parmi les 150 ayant constitué notre base de données initiales et 120 réponses émanant des entreprises saines. Toutes les variables retenues sont présentées dans le tableau ci-dessous avec leur codification (Voir tableau II).

TABLEAU III : LES VARIABLES QUALITATIVES RETENUES

Variables	Code
Mauvaise gestion et organisation	MGESTOR
Incapacité d'anticipation et d'adaptation	INCANTIA
Déficiences Management	DEFMANA
Allongement des délais de paiement clients	ALLODLP
Disfonctionnement dans l'organisation interne	DISFORG
Absence d'analyse de la concurrence	ABSANCO
Incapacité de l'entreprise à évaluer correctement le marché sur lequel elle exerce	INCEVAL
Baisse du chiffre d'affaire	BAISSECA
Difficulté financière due à une insuffisance des fonds propres	DIFFINAN
Difficulté de recouvrement des créances clients	DIFFIREC
Alourdissement de la dette commerciale	ALLOUDT
Absence de positionnement du produit	ABSPOSPR
Mauvais climat social	MAUVCLI
Manque de responsabilisation du personnel	MQRESPER
Perte d'un client important	DEFACTL

Notre base de données est divisée en deux sous échantillons pour pouvoir tester la performance des techniques utilisées : échantillon d'apprentissage et échantillon test. Nous disposons au départ d'un échantillon dit d'apprentissage dont le classement est connu. Cet échantillon est utilisé pour paramétrer les divers modèles et servir à l'apprentissage des règles de classement d'une entreprise selon ses caractéristiques. Le deuxième échantillon test est nécessaire à l'étude de la fiabilité des techniques utilisées. Pour cette raison, nous allons effectuer un échantillonnage aléatoire sur la variable défaillance. Nous choisissons environ 70% de données pour l'apprentissage et 30% de données pour tester nos modèles de prévision.

B. Présentation des logiciels employés

Dans le but de prévenir et modéliser une défaillance potentielle des entreprises, nous avons utilisé le logiciel SPSS pour le traitement des données, à Matlab pour la programmation des réseaux de neurones et à R pour le paramétrage des SVM. Le logiciel SPSS a servi pour les traitements et les analyses statistiques de base (certains recodages de variables, statistiques descriptives, tests d'hypothèses...), l'analyse en composante principale, la sélection des variables discriminantes à l'aide du test de Fisher. Tous les autres traitements ont été développés par nos soins. Ainsi, les procédures de programmations (bootstrap), relatives au réseau de neurones utilisés pour

effectuer les classifications (Perceptron multicouches) et celles relatives au paramétrage du logiciel R pour l'application de la méthode SVM.

C. Procédure de fonctionnement des techniques choisies

✓ La technique des réseaux de neurones

Les réseaux de neurones artificiels sont des programmes informatiques d'inspiration biologique conçus pour simuler la façon dont le cerveau humain traite l'information. Ils rassemblent leurs informations sur les caractéristiques des données et les relations qui les lient par l'expérience pendant la phase d'apprentissage. Le succès croissant de réseaux de neurones peut s'attribuer à leur puissance et à leur polyvalence de traitement. L'avantage de cette technique est sa capacité à modéliser des relations entre des données ou des fonctions particulièrement complexes.

En effet, le réseau est formé par plusieurs neurones artificiels ou d'éléments de traitement, liés par des coefficients (poids synaptique), qui constituent la structure neuronale et qui est organisé en plusieurs couches. La couche d'entrée est composée de N neurones au nombre des variables discriminantes initialement sélectionnées qui sont censées expliquer le phénomène de la défaillance. La couche de sortie est composée d'un seul neurone qui présente la variable dichotomique défaillante ou saine. La couche cachée est composée de plusieurs neurones cachés qui traitent l'information entre les neurones de la couche d'entrée et celle de la couche de sortie. Ces neurones sont reliés entre eux par l'intermédiaire de poids synaptiques notés w_{ij} . On applique dans cette étude les réseaux à couches multiples (Multi Layer Perceptron : MLP) en association avec la méthode d'apprentissage de rétro-propagation du gradient de l'erreur appelée « Back-Propagation ». Cette méthode exige du manipulateur de réaliser plusieurs optimisations afin de déterminer l'architecture neuronale optimale qui améliore le pouvoir prédictif du modèle dans l'identification de la situation des entreprises [45]. L'algorithme d'optimisation le plus utilisé est celui de rétro-propagation du gradient de l'erreur. C'est grâce à cet algorithme que l'utilisation des réseaux s'est intensifié.

La règle de minimisation d'erreur adoptée par cet algorithme est la suivante :

$$E(w) = \frac{1}{N} \sum_{i=1}^n (y_{i,j} - y_{i,j}^*)^2$$

Où N représente le nombre de couples (x_i, y_i) utilisés pour l'apprentissage, n est le nombre de neurones dans la couche de sortie et $y_{i,j}^*$ est la réponse du réseau pour l'entrée x_i .

La fonction d'erreur dépend de tous les poids du réseau et doit être minimisée. Le principe de l'optimisation individuelle implique que nous devons continuellement rechercher des classificateurs parfaits tout en optimisant les paramètres du modèle ou la structure du modèle [46].

La capacité prédictive du réseau dépend du degré de connexion des neurones, mieux qu'elles soient

interconnectées, meilleure sera la prédiction en sortie. Aussi, l'activité d'un réseau de neurones est déterminée par les fonctions de transfert de ses neurones, par la règle d'apprentissage et par l'architecture elle-même. La fonction de transfert retenue est celle sigmoïde logistique (exponentielle) puisque la variable à expliquer est binaire. L'obtention de bons résultats dépend de la minimisation de l'erreur de prédiction par l'optimisation des connexions inter unités pendant la formation. Les ajustements des poids sont déterminés dans le chemin de propagation à chaque niveau et ce en évaluant les dérivées partielles de la fonction E par rapport aux poids synaptiques. Après l'interaction de large nombre de cycles, l'erreur est réduite à un niveau acceptable, et le processus s'arrête. Une fois que le réseau est formé avec une erreur de prédiction minimale et testé, il peut être utilisé avec de nouvelles informations d'entrée pour prédire la sortie.

✓ Technique des SVM

Les machines à vecteurs de support (SVM) sont des algorithmes d'apprentissage automatique, utilisés pour résoudre des problèmes de classification. Cette technique a connu un grand essor au cours de la dernière décennie. Le succès de cette méthode repose dans sa fonction noyau, qui consiste à transformer les données d'origine en données de grande dimension, afin de rendre les groupes séparables. Ce principe fait de cette technique un outil efficace et puissant dans le domaine de classification.

Il s'agit donc d'une technique d'apprentissage artificielle basée à la fois sur l'hypothèse de linéarité dans un espace multidimensionnel et sur la théorie de l'optimisation. Le principe de base de la technique des SVM consiste à maximiser la marge d'un hyperplan séparant les données en deux classes, en leur conférant une grande capacité de généralisation. L'hyperplan avec la plus grande distance aux points de repères des deux classes des données, indique une bonne séparation, puisqu'en général la marge la plus grande réalise l'erreur la plus faible du classificateur.

La détermination des surfaces séparatrices est réalisée à l'aide de l'insertion d'une fonction noyau (kernel) dans le produit scalaire induisant implicitement une transformation non linéaire des données vers un espace intermédiaire (feature space) de plus grande dimension. La fonction noyau utilisée dans SVM doit respecter le principe mathématique connu sous le nom de théorème de Mercer. Ce principe garantit que la fonction noyau peut être exprimée comme un produit scalaire entre les deux vecteurs d'entrée dans un espace de grandes dimensions. La performance de la technique SVM en matière de classification peut se définir en termes d'une recherche d'un ensemble de paramètres optimaux. Dans ce cadre, nous essayons à déterminer dans ce qui suit une fonction noyau efficace, un paramètre de régularisation C qui permet de minimiser l'erreur de mauvaise classification et une combinaison appropriée des paramètres de la fonction noyau.

Le paramétrage de ces noyaux est réalisé à l'aide de la fonction « tune.svm » sous R. Cette fonction permet de tester plusieurs valeurs du paramètre C en estimant la performance de prédiction pour un noyau donné. Le noyau utilisé dans cette recherche est celui polynomial, qui est largement utilisé dans la littérature et qui a abouti à des résultats performants dans le domaine de classification.

De nombreux outils de mesure sont proposés pour évaluer la capacité prédictive du modèle de prévision. Nous optons dans ce qui suit pour le choix de la matrice de confusion qui est l'une des critères les plus utilisés dans le domaine de l'économie et de la finance.

IV. ANALYSE FACTORIELLE DES VARIABLES QUALITATIVES

En matière d'analyse des données, il peut exister un grand nombre de variables, souvent corrélées. Il convient alors de les synthétiser afin de pouvoir en tirer une information utilisable. Dans ce but, l'analyse factorielle sera adoptée pour la sélection et la synthèse des données utilisées dans cette recherche.

A. Pertinence de l'analyse factorielle

Le test de sphéricité de Bartlett peut être utilisé pour tester l'hypothèse nulle que les variables ne sont pas corrélées dans cette étude. Ce test s'appuie sur une transformation du khi-deux du déterminant de la matrice des corrélations. Une valeur élevée sera favorable au rejet de l'hypothèse nulle. Dans le cas contraire, la pertinence de l'analyse factorielle devra être remise en question. Un autre test statistique utile est la mesure de l'indice d'adéquation de Kaiser-Meyer-Olkin (KMO), qui compare les grandeurs des coefficients de corrélation observés aux grandeurs des coefficients de corrélation partielle. De petites valeurs de cet indice indiquent que les corrélations entre paires de variables ne peuvent pas être expliquées par d'autres variables et que l'analyse factorielle peut ne pas être pertinente. En général, il est préférable d'avoir une valeur supérieure à 0,5.

TABLEAU III: INDICE KMO ET TEST DE BARTLETT

Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin.		,831
Test de sphéricité de Bartlett	Khi-deux approximé	1766,418
	Ddl	105
	Signification de Bartlett	,000

Les résultats du tableau III montrent que l'hypothèse nulle (la matrice des corrélations serait une matrice identité) est rejetée par le test de sphéricité de Bartlett. La statistique de khi-deux est de 1766,418 avec 105 degrés de liberté, ce qui est significatif à 5%. La valeur du test

statistique KMO (0,831) est également élevée (>0,5). L'analyse factorielle est alors appropriée pour synthétiser l'information contenue dans toutes les variables initiales. On peut constater d'après les statistiques obtenues que l'analyse Factorielle s'avère pertinente pour notre étude.

B. Extraction des axes

Dans l'analyse en composantes principales, la variance totale des données est considérée. Cette méthode est recommandée quand on cherche avant tout à déterminer le nombre minimal de facteurs, appelés composantes principales, correspondant à la variance maximale des données, afin de les utiliser ultérieurement dans une analyse multi variée. (Voir tableau IV)

TABLEAU IV: VARIANCE TOTALE EXPLIQUÉE

Composante	valeurs propres initiales			Extraction carrés des retenus	Sommes des facteurs	
	Total	% de la variance	% cumulés		Total	% de la variance
MGESTO	6,744	44,95	44,959	6,744	44,959	44,95
INCANTI	3,127	20,84	65,807	3,127	20,847	65,80
DEFMAN	1,644	10,96	76,769	1,644	10,962	76,76
ALLODLP	1,275	8,499	85,268	1,275	8,499	85,26
DISFORG	,635	4,237	89,505			
ABSANCO	,428	2,850	92,355			
INCEVAL	,338	2,251	94,606			
BAISSEC	,301	2,009	96,615			
DIFFINAN	,197	1,314	97,929			
DIFFIREC	,107	,715	98,645			
ALLOUD	,085	,568	99,212			
ABSPOSP	,077	,517	99,729			
MAUVCLI	,023	,157	99,887			
MQRESP	,010	,072	99,959			
DEFACT	,061	,041	100,00			
Méthode d'extraction : Analyse en composantes principales.						

L'analyse en composantes principales fournit quatre axes riches en termes d'informations. Ces axes factoriels sélectionnés peuvent être interprétés par l'analyse des corrélations les plus fortes avec chacun des axes.

L'axe n° 1 fournit 44,959 % de l'information totale. Il correspond à l'information la plus importante. Les axes 2,3 et 4 fournissent respectivement 20,847% ; 10,962% et 8,49% de l'information totale.

C. Interprétation des axes retenus

Pour faciliter l'interprétation des axes factoriels retenus, on a fait recours à des algorithmes de rotation. L'objectif de la rotation est que les facteurs aient un coefficient non nul ou significatif pour quelques variables. Pour interpréter les facteurs, il est nécessaire de revenir sur les variables initiales et déterminer le poids de chaque variable dans la formation de chaque facteur. Il est donc nécessaire de revenir à la matrice des composantes principales et d'étudier les coefficients de corrélation entre les nouvelles variables appelées facteur principal et les anciennes variables. (Voir Tableau V)

TABLEAU V: MATRICE DES COMPOSANTES

	Composante			
	1	2	3	4
MGESTORG	,948	-,172	,188	,082
INCANTIA	,948	-,181	,174	,084
DEFMANAG	,946	-,161	,207	,073
ALLODLP	-,930	,155	-,125	-,062
DISFORG	,892	,124	,156	,175
ABSANCON	-,828	,131	-,186	,366
INCEVAL	-,731	-,314	-,055	,410
BAISSECA	,207	-,851	,138	-,135
DIFFINAN	,091	,786	,371	,026
DIFFIREC	-,312	,779	-,087	-,313
ALLOUDTC	,151	-,756	,169	,096
ABSPOSPR	,393	,628	,282	,198
MAUVCLIM	,236	-,008	,954	,016
MQRESPER	,240	,001	,953	,004
DEFACTL	-,046	,005	-,028	-,903
Méthode d'extraction : Analyse en composantes principales.				
Méthode de rotation : Varimax avec normalisation de Kaiser.				
La rotation a convergé en 5 itérations.				

Les variables actives les mieux corrélées au **facteur1** sont: Mauvaise gestion et organisation des différents maillons de la chaîne de valeur interne, Incapacité d'anticipation et d'adaptation, Disfonctionnement dans l'organisation interne, Absence d'analyse de la concurrence, Incapacité de l'entreprise à évaluer correctement le marché sur lequel elle exerce. Par ailleurs,

nous pouvons interpréter l'axe F1 comme celui des erreurs de gestions rencontrées au sein de l'entreprise. Ces erreurs sont parmi les causes essentielles qui amènent les entreprises à des situations de difficultés et qui sont dues essentiellement à des incompétences managériales. Ces résultats ont été confirmés par les études antérieures. En effet, Kaplan [47] est arrivé dans ses recherches à montrer que la première cause de l'échec d'une entreprise résulte des insuffisances de son dirigeant. Dans ce contexte, une autre étude [48] s'est intéressé aux faiblesses inhérentes au management comme le savoir-faire ou l'expérience managériale. Aussi, les études réalisées en France par CEPME [49] et par l'INSEE [50] convergent à montrer que les erreurs de gestion, qui sont des causes endogènes, sont les origines de difficultés des entreprises et des dépôts de bilan. Dans ce contexte, l'étude de [51] montre qu'un mode de gestion avertit et rigoureux, peut réduire les risques de difficulté potentielle.

Les variables actives les mieux corrélées au facteur 2 sont : Baisse du chiffre d'affaire, Difficulté financière due à l'insuffisance des fonds propres, Difficulté de recouvrement des créances clients, Alourdissement de la dette commerciale, Absence de positionnement du produit. L'axe F2 peut être interpréter comme l'axe présentant les erreurs stratégiques commises au sein de l'entreprise. En effet, les entreprises défaillantes sont situées dans la plupart des cas sur des marchés saturés caractérisés par une concurrence intense. Aussi, l'absence de positionnement du produit et l'absence d'une vision stratégique prive les entreprises d'avantages concurrentielles (en termes de coût de revient ou en terme d'innovation, image de marque ...) et engendre la perte des parts de marché, ce qui aggrave d'avantage leurs situation difficile. Pour faire face au problème de liquidité immédiate ces entreprises se trouvent d'avantage soumises à la pression importante de leurs clients en ce qui concerne les remises demandées et les délais du règlement. L'entreprise est alors prise dans une spirale défavorable qui peut s'accélérer jusqu'à devenir incontrôlable. Ces résultats, ont été confirmés par [52] qui ont souligné que la défaillance des entreprises est essentiellement dû à des causes internes tels que: absence de veille concurrentielle ou stratégique, mauvais choix stratégiques et une mauvaise anticipation du marché.

Dans le même cadre, [53] a souligné dans son étude visant à limiter les déterminants de la faillite des entreprises, que les causes peuvent résulter de plusieurs éléments interne et externe à l'entreprise parmi lesquels le positionnement sur des marchés moins porteurs ainsi que l'absence des atouts stratégiques pour se démarquer des concurrents.

Les variables actives les mieux corrélées au facteur 3 sont : Mauvais climat social, Manque de responsabilisation du personnel. L'axe 3 présente la responsabilité sociale des entreprises. En effet, l'absence de sensibilisation, de contrôle et de motivation du personnel affecte l'environnement social au sein de l'entreprise et engendre automatiquement une déresponsabilisation de la part du

personnel qui peut affecter la survie de l'entreprise. L'absence de motivation au travail a été portée par [52] comme cause principale de la faillite des entreprises. En effet, selon ces derniers, l'absence de motivation (climat social décevant, problèmes entre les personnels, etc) au milieu de travail peut aboutir à la faillite de l'entreprise dans le moyen ou long terme. Dans le même ordre d'idée, [54], dans une enquête réalisée par les dirigeants et les cadres de PME, montrent qu'il existe des stratégies à mener pour protéger la survie des entreprises. En effet, ils signalent que des efforts particuliers doivent être consacrés à la création d'un cadre de travail agréable, convivial et souple ainsi qu'une valorisation du travail rendu par le personnel de l'entreprise. L'étude de [55] signale que la réussite des PME est due essentiellement au facteur humain.

Les variables actives les mieux corrélées au facteur 4 sont : Perte ou défaillance d'un client important. L'axe F4 présente la dépendance de l'entreprise par rapport à un facteur externe. En effet, la perte d'un client important peut affecter la survie de l'entreprise. Dans ce contexte, [53] affirme que lorsque l'entreprise est fortement dépendante d'un seul marché, d'un client ou d'un fournisseur, les risques de défaillances sont élevés. La diversification du portefeuille de la clientèle met l'entreprise à l'abri des difficultés potentielle de trésorerie. Dans ce contexte, [56] montre que les causes économiques de défaillances sont multiples: facteurs conjoncturels, facteurs interne à l'entreprises et facteurs externes comme la faillite d'un client important.

V. MODELISATION DE LA DEFAILLANCE EN UTILISANT LES VARIABLES QUANTITATIVES

A. Choix des variables

Une batterie de vingt-neuf ratios financiers, codés de R01 à R29, a été sélectionnée. Le choix de ces ratios a été motivé par leurs récurrence dans les études traitant le sujet de prévision de la défaillance des entreprises ou ceux qui présentent un caractère informationnel significatif dans l'évaluation de la santé financière des entreprises. Les thèmes choisis pour la sélection de ces ratios sont: la rentabilité, la productivité, la gestion, la liquidité et le financement et enfin la structure financière.

La modélisation et la qualité de prévision dépendent énormément du choix des ratios financiers qui apprécient de manière significative la probabilité de défaillance [57].

Etant donnée le nombre élevé des ratios financier initialement fixés, notre sélection s'est basé sur le choix des ratios qui contribuent au mieux à la discrimination entre les deux types d'entreprises (saines et celles défaillantes). Dans Afin d'atteindre cet objectif, nous avons appliqué le test de Fisher pour déterminer les ratios financiers les plus discriminants qui différencient au mieux entre les deux groupes d'entreprises (Voir tableau III).

TABLEAU IV: TEST D'EGALITE DES MOYENNES

Ratios	F	Signification
R01	35,280	,000
R02	8,256	,004
R03	,146	,703
R04	,128	,721
R05	,052	,820
R06	1,626	,203
R07	3,467	,064
R08	,147	,702
R09	,886	,347
R10	11,958	,001
R11	3,256	,072
R12	90,847	,000
R13	,312	,577
R14	9,585	,002
R15	7,443	,007
R16	4,588	,033
R17	1,060	,304
R18	72,257	,000
R19	33,794	,000
R20	33,389	,000
R21	13,161	,000
R22	,045	,831
R23	13,810	,000
R24	3,017	,083
R25	2,392	,123
R26	14,742	,000
R27	1,555	,213
R28	1,772	,184
R29	7,699	,006

L'examen des résultats obtenus à l'aide du test de Fisher dans le tableau ci-dessus, au seuil de 5%, montre la présence d'un certain nombre de ratios significatifs et d'autres qui ne le sont pas. Les ratios significatifs qui discriminent au mieux entre les deux groupes d'entreprises appartiennent au quatre types de ratios utilisés dans cette analyse. Le premier groupe traduisant les ratios de rentabilité est formé par les ratios R01 et R02.

Le deuxième groupe d'indicateurs se rapporte à des ratios de gestion : R10 et R12.

Le troisième groupe présentant la liquidité de l'entreprise est formé par les ratios significatifs : R14, R15, et R16.

Le quatrième groupe d'indicateurs reflète des ratios de financement et de structure. Les ratios significatifs sont R18, R19, R20, R21, R23, R26.

Enfin le ratio R29 qui reflète la taille de l'entreprise montre que les petites entreprises sont plus vulnérables au risque de faillite.

Pour conclure, l'analyse des deux groupes d'entreprises confirme que les entreprises défaillantes se caractérisent par

des performances économiques et financière médiocres. Le test de Fisher, qui permet de tester la pertinence d'une variable à différencier entre les deux groupes d'entreprises, montre que les entreprises défaillantes sont lourdement endettées à court terme et sont caractérisés des rentabilités faibles.

B. Application de la technique des réseaux de neurones

➤ L'identification de la meilleure architecture

Afin de paramétrer le réseau de neurones, nous avons eu recours au logiciel Matlab (2011) qui permet la modélisation des neurones artificiels à l'aide de l'application « neural network toolbox ». Nous avons utilisé une seule couche cachée, et nous avons varié à chaque fois le nombre des neurones y afférente afin d'aboutir à l'architecture optimale qui permet d'avoir l'erreur moyenne la plus faible avec le taux de prévision le plus performant relatif à l'échantillon test. Le choix d'une seule couche cachée a été motivé par le travail de [58] qui a montré qu'un réseau de neurones bien paramétré, composé d'une seule couche cachée et par un nombre optimal de neurones abouti à des résultats performants dans le domaine de classification.

TABLEAU V: ERREUR DE PREVISION EN FONCTION DE LA VARIATION DU NOMBRE DE NEURONES

Nombre de neurones	Performance (MSE)	Nombre d'itération
3	0,024	39
4	0,035	146
5	0,022	42
6	0,038	82
7	0,029	53
8	0,026	65
9	0,015	113
10	0,0144	96
15	0,052	48

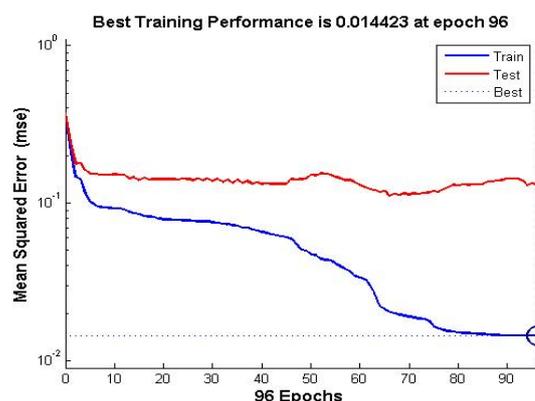


FIGURE 1: COURBE D'APPRENTISSAGE ET DE TEST DU RESEAU

L'architecture optimale du réseau est celle obtenue avec 10 neurones. Ce réseau nous a permis d'avoir une erreur

moyenne quadratique la plus faible qui s'élève à 0,0144 avec 96 itérations.

➤ *Evaluation de la performance du modèle*

La performance du réseau est mesurée par la matrice de confusion et par l'aire sous la courbe ROC (AUC).

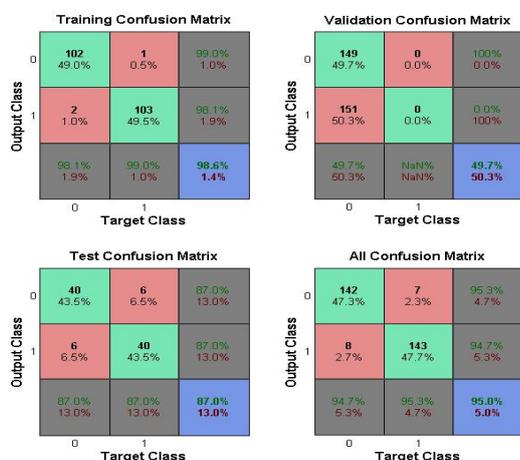


FIGURE 2: MATRICE DE CONFUSION DES RNA

TABLEAU VI: RESULTAT DE CLASSIFICATION DES RNA POUR L'ECHANTILLON TEST

		Classe d'affectation prévue par le modèle pour l'échantillon Test		Total
		entreprise défaillante	entreprise saine	
Original	Effectif	40	6	46
	%	87	13	100
Original	Effectif	6	40	46
	%	13	87	100

TABLEAU VII: RESULTAT DE CLASSIFICATION DES RNA POUR L'ECHANTILLON D'APPRENTISSAGE

		Classe d'affectation prévue par le modèle pour l'échantillon d'apprentissage		Total
		entreprise défaillante	entreprise saine	
Original	Effectif	102	2	104
	%	98.1	1.9	100
Original	Effectif	1	103	104
	%	1	99	100

Les tableaux ci-dessus montrent les résultats de classification obtenue par la méthode des RNA. Pour le sous ensemble d'apprentissage, le taux de bon classement global du modèle s'élève à 98,6%. Pour le sous-ensemble de test le modèle n'a pu classer correctement que 87,00% de l'ensemble des entreprises introduites.

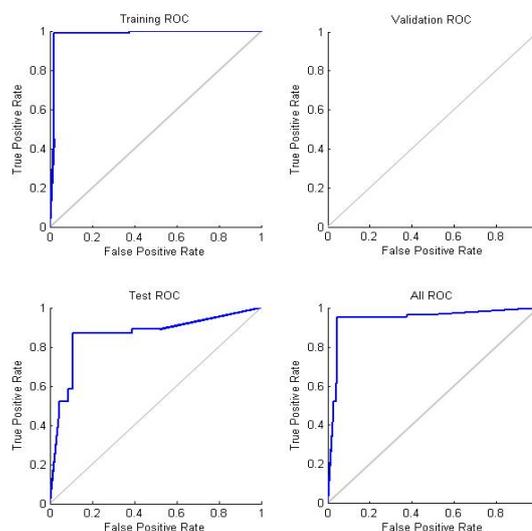


FIGURE 3: COURBE ROC OBTENU PAR RNA

Un autre critère d'évaluation de la performance du réseau est celui du calcul de l'aire sous la courbe ROC. En effet, il est possible de caractériser numériquement la surface située sous la courbe ROC désigné par le critère d'AUC (AUC, pour Area Under Curve). Il mesure la probabilité de placer une entreprise saine devant une défaillante. Ainsi, dans le cas d'une discrimination parfaite $AUC = 1$, les entreprises saines sont sûrs d'être placés devant les entreprises défaillantes. Au contraire, si $AUC = 0.5$ le classificateur affecte des scores au hasard, de ce fait il y a autant de chances d'affecter une entreprise saine devant une défaillante que l'inverse. La situation de référence correspond à la situation au niveau du quelle la courbe ROC se confond avec la première bissectrice.

L'aire sous la courbe ROC mesure la qualité de discrimination du modèle et traduit la probabilité qu'une entreprise saine ait un score supérieur au score d'une entreprise défaillante, ceux-ci étant tirés au hasard. L'aire sous la courbe ROC de réseaux de neurones est égale à 0.73682 ; cette valeur est supérieure à 0,5.

La valeur de l'AUC est égale à 0,97119 pour l'échantillon d'apprentissage et 0,73682 pour l'échantillon test. Notre modèle dispose donc d'un pouvoir discriminant performant.

C. *Application de la technique des SVM*

Nous avons eu recours au logiciel R pour le paramétrage des SVM. De plus, la sélection du noyau affecte grandement les performances de la classification. Comme le noyau linéaire n'offre pas une grande

prévisibilité dans les jeux de données non séparables, le noyau RBF et les autres noyaux non linéaires sont difficiles à analyser mais ils offrent des taux de prédiction satisfaisants dans le cas de données non séparables [2]. Nous avons opté pour le choix du noyau polynomial caractérisé par sa robustesse et sa performance en matière de classification. Une bonne optimisation des paramètres de ce noyau permet de donner des résultats satisfaisants. Contrairement au réseau de neurones, la méthode SVM intercepte les caractéristiques géométriques de l'espace de l'ensemble des données d'apprentissage, sans dériver les poids des réseaux, ce qui la rend capable d'extraire la solution optimale avec la petite taille de l'ensemble d'apprentissage.

L'application de la méthode SVM à base polynomiale sur l'échantillon de variables choisi nous a permis d'avoir une bonne performance. Le taux d'erreur est inférieur à 0,2.

TABLEAU VIII: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON TEST A L'AIDE DES SVM A BASE POLYNOMIALE

		Classe d'affectation prévue par le modèle pour l'échantillon test			Total
Groupe d'entreprise		entreprise défaillante	entreprise saine		
Original	Effectif	Entreprise défaillante	33	13	46
		entreprise saine	5	41	46
	%	entreprise défaillante	71,74	28,26	100
		entreprise saine	10,87	89,13	100

TABLEAU IX: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON APPRENTISSAGE A L'AIDE DES SVM A BASE POLYNOMIALE

		Classe d'affectation prévue par le modèle pour l'échantillon apprentissage			Total
Groupe d'entreprise		entreprise défaillante	entreprise saine		
Original	Effectif	entreprise défaillante	86	18	104
		entreprise saine	12	92	104
	%	entreprise défaillante	82,7	17,3	100
		entreprise saine	11,54	88,46	100

Le taux de bon classement des entreprises défaillantes s'élève à 71,74 % pour le test et 82,7% pour l'apprentissage, alors que celui des saines s'élève à 89,13% pour le test et 88,46 % pour l'apprentissage. En définitif, le taux de bon classement, obtenu par le modèle SVM à base polynomiale, s'élève à 80,43% pour l'échantillon test et 85,58% pour l'échantillon d'apprentissage.

VI. PROCESSUS DE MODELISATION DE LA DEFAILLANCE EN UTILISANT LES VARIABLES QUANTITATIVES ET QUALITATIVES

Les analyses effectuées en général, reposent principalement sur l'établissement de modèles quantitatifs construits à partir des données financières, et mettant en lumière les différents symptômes de la défaillance. La volonté de mieux comprendre le processus de la défaillance oblige à élargir le champ d'investigation de l'analyse et intégrer des aspects d'ordre qualitatif à côté des variables quantitatives initialement utilisées. Les variables qualitatives sélectionnées à partir du questionnaire adressé aux dirigeants des entreprises, auxquelles on a pratiqué une analyse en composantes principales pour les résumer en 4 facteurs, vont être utilisées dans cette analyse.

A. Choix des variables

Le test d'égalité des moyennes sera menée sur l'ensemble des variables quantitatives et les axes factoriels représentant les variables de types qualitatives à fin de déterminer celles ou ceux qui ont le pouvoir discriminant le plus élevé et qui sont capables d'enrichir le processus en terme d'informations utile à la prise de décision afin de dissocier les entreprises saines des défaillantes.

TABLEAU X: TEST D'EGALITE DES MOYENNES DES GROUPES

Ratios	Lambda de	F	Signification
R01	,925	12,748	,000
R02	,956	7,237	,008
R10	,946	8,999	,003
R12	,901	17,344	,000
R14	,974	4,284	,040
R15	,959	6,736	,010
R18	,950	8,248	,005
R19	,892	19,219	,000
R20	,931	11,790	,001
R21	,911	15,504	,000
R23	,971	4,672	,032
R26	,962	6,210	,014
R27	,961	6,468	,012
R29	,943	9,550	,002
F1	,820	34,608	,000
F2	,830	32,376	,000
F3	,942	9,748	,002
F4	,949	8,500	,004

L'examen du test de Fisher dans le tableau ci-dessus, nous indique qu'au seuil de 5%, toutes les variables sont significatives. De même, les quatre axes factoriels sont significatifs et sont dotés d'un fort pouvoir discriminant.

En effet, il est intéressant d'approfondir cette recherche par l'introduction de variables qualitatives à côtés de celles

quantitatives pour tester la capacité prédictive des modèles tout en utilisant la méthode des réseaux de neurones et celle des SVM.

B. L'identification de la meilleure architecture

La même démarche que celle employée avec l'introduction des variables quantitatives sera employée.

➤ L'identification de la meilleure architecture

TABLEAU XI: ERREUR DE PREVISION EN FONCTION DE LA VARIATION DU NOMBRE DE NEURONES

Nombre de neurones	Performance (MSE)	Nombre d'itération
2	0,044	26
4	0,065	142
5	0,0022	42
6	0,015	18
7	0,0012	20
8	0,0026	26
9	0,005	113
10	0,000048	12
15	0,012	48

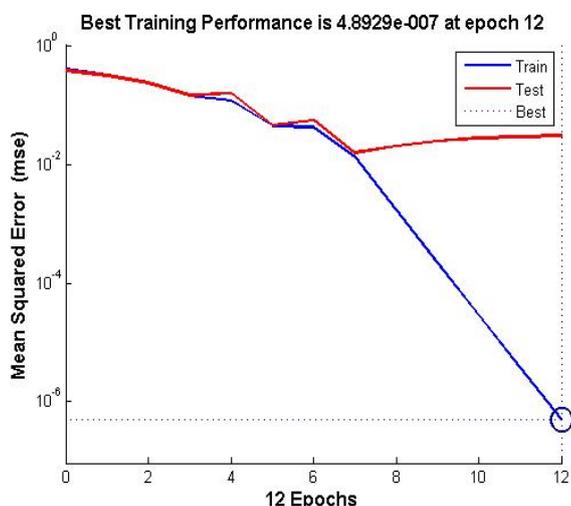


FIGURE 4 : COURBE D'APPRENTISSAGE ET DE TEST DU RESEAU

L'architecture optimale du réseau est obtenue avec une couche d'entrée, une couche de sortie qui correspond au résultat selon lequel l'entreprise est saine ou défaillante dans la classification du réseau et une couche intermédiaire avec 10 neurones. Ce réseau nous a permis d'avoir une erreur moyenne quadratique la plus faible qui s'élève à $4,8 \cdot 10^{-5}$.

➤ Evaluation de la performance du modèle

La performance du réseau est mesurée par la matrice de confusion.

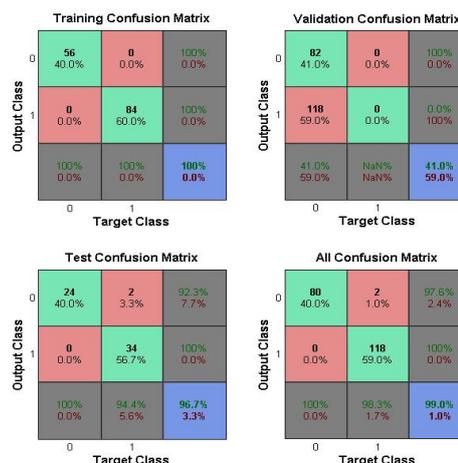


FIGURE 5: MATRICE DE CONFUSION DES RNA

TABLEAU XIV: RESULTAT DE CLASSIFICATION DES RNA POUR L'ECHANTILLON TEST

		Classe d'affectation prévue par le modèle pour l'échantillon Test			Total
		Groupe d'entreprise	entreprise défaillante	entreprise saine	
Original	Effectif	entreprise défaillante	24	0	24
		entreprise saine	2	34	36
	%	entreprise défaillante	100	0	100
		entreprise saine	5,6	94,4	100

TABLEAU XV: RESULTAT DE CLASSIFICATION DES RNA POUR L'ECHANTILLON D'APPRENTISSAGE

		Classe d'affectation prévue par le modèle pour l'échantillon d'apprentissage			Total
		Groupe d'entreprise	entreprise défaillante	entreprise saine	
Original	Effectif	entreprise défaillante	56	0	56
		entreprise saine	0	84	84
	%	entreprise défaillante	100	0	100
		entreprise saine	0	100	100

Ces tableaux montrent les résultats de classification des RNA avec l'association des variables quantitatives et qualitatives. Le taux de bon classement global pour le sous-ensemble d'apprentissage est de 100% et pour le sous-ensemble test est de 97,2%.

La technique des réseaux de neurones est une technique de prévision très performante. Elle permet la détection de

défaillance avec des taux élevés de reconnaissance. L'association des variables qualitatives au variables quantitatives dans le réseau nous a permis d'avoir un pourcentage de prévision hyper satisfaisant qui s'élève à 97,2% , supérieur au taux de prévision déterminé seulement à partir des données quantitatives qui s'élève à 87%.

D'après les résultats mentionnés ci-dessus, on constate l'apport appréciable des variables qualitatives en termes de prévision de la défaillance des entreprises.

C. Application de la technique des SVM

La technique SVM construit un hyperplan de séparation optimal pour classifier les données. La même démarche que celle employée avec l'introduction des variables quantitatives sera employée dans cette partie. Une bonne optimisation des paramètres du noyau polynomial permet de donner des résultats satisfaisants.

TABLEAU XVI: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON TEST A L'AIDE DES SVM A BASE POLYNOMIALE

Original	Effectif	Groupe d'entreprise	Classe d'affectation prévue par le modèle pour l'échantillon test		Total
			entreprise défaillante	entreprise saine	
		Entreprise défaillante	24	0	24
		entreprise saine	2	34	36
	%	entreprise défaillante	100	0	100
		entreprise saine	5,55	94,44	100

TABLEAU XVII: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON APPRENTISSAGE A L'AIDE DES SVM A BASE POLYNOMIALE

Original	Effectif	Groupe d'entreprise	Classe d'affectation prévue par le modèle pour l'échantillon apprentissage		Total
			entreprise défaillante	entreprise saine	
		entreprise défaillante	54	2	56
		entreprise saine	1	83	84
	%	entreprise défaillante	96,43	3,57	100
		entreprise saine	1,20	98,8	100

Le taux de bon classement des entreprises défaillantes s'élève à 100 % pour le test et 96,43% pour l'apprentissage, alors que celui des saines s'élève à 94,44% pour le test et 98,8% pour l'apprentissage. En définitif, le taux de bon classement, obtenu par le modèle SVM à base polynomiale, s'élève à 97,22% pour l'échantillon test et 97,61% pour l'échantillon d'apprentissage.

La méthode SVM est un outil très performant en matière de détection précoce de la faillite des entreprises.

Elle permet l'appréciation du risque de faillite avec des taux de reconnaissance satisfaisants. L'usage de cette méthode avec l'association des variables qualitatives et variables quantitatives nous a permis d'avoir un taux de prévision hyper satisfaisant qui s'élève à 97,61%. Ce taux est supérieur à celui obtenu seulement avec des données financières qui s'élevait à 80,43%. D'après ces résultats, on peut constater l'apport appréciable des variables qualitatives en termes de prévision de la défaillance des entreprises.

VII. Conclusion

Cet article a d'abord porté sur l'identification des ratios financiers qui sont capable d'expliquer la faillite des entreprises. Hors ceux-ci ne représentent que les symptômes initiaux annonciateurs d'une défaillance potentielle. Pour mieux identifier les causes réelles, l'application de l'analyse en composante principale sur un ensemble de variables qualitatives, collectées auprès des responsables des entreprises à l'aide d'un questionnaire, nous a permis de dégager quatre axes factoriels qui présentent les facteurs déterminants du processus de défaillance des entreprises tunisiennes, à savoir des facteurs internes en matière de qualité de management et d'orientation stratégique, la pression concurrentielle et l'état de conjoncture, le climat social ainsi que l'effet domino.

Les insuffisances et les contraintes liées à la modélisation paramétrique, nous ont poussés vers l'utilisation des outils plus puissants dans le domaine de classification qui sont les réseaux de neurones et la technique SVM. On a testé la performance de ces deux méthodes en matière de prévision tout en utilisant des variables quantitatives, dans une première étape, et en utilisant des variables qualitatives à côté des variables quantitatives, dans une deuxième étape.

Les résultats obtenus ont montré l'apport bénéfique et précieux des variables qualitatives en matière de prévision de la défaillance. La performance des réseaux s'est améliorée chaque fois qu'on associe des variables qualitatives à côté des variables quantitatives. En utilisant l'approche neuronale, le taux des entreprises biens classés passent de 87% à 97,2%. L'amélioration sera plus meilleure avec la technique SVM dans la mesure où le taux des biens classés est passé de 80,43% à 97,61%.

Au total, il ressort des résultats obtenus dans cette recherche que l'association des variables qualitatives à coté de celles financières est très prometteuse dans le domaine d'appréciation du risque de défaillance. Ces résultats offrent de nouvelles perspectives et inspirations pour la recherche dans le domaine de la prévision de la faillite.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Qu Yi, Quan Pei, Lei, M., Shi, Y., "Information technology and quantitative management based on Artificial Intelligence". *Procedia Computer Science* Volume 162, 2019, pp 895-899.
- [2] Altman, E.I. "Bankruptcy, Credit Risk and High Yield Junk Bonds". (2002). New York: Blackwell Publishers
- [3] Hensher, D.A, and Jones. S., "Forecasting Corporate Bankruptcy: Optimizing the Performance of the Mixed Logit Model", Volume 43, Issue 3, pp. 241-427, September 2007
- [4] Shumway, T., 2001. "Forecasting bankruptcy more accurately: A simple hazard model". *The Journal of Business*, 2001, vol. 74, issue 1, 101-24
- [5] Duffie, D., Singleton, K.J., "Credit risk, pricing, measurements, and management". *ASTIN Bulletin*, 2004, vol. 34, issue 1, 264-265.
- [6] Jones, S., and Wang, T., "Predicting private company failure: A multi-class analysis," *Journal of International Financial Markets, Institutions and Money*, 2019, vol. 61(C), pages 161-188.
- [7] Mc Namara, R.P., Cocks, N.J., Hamilton, D.F., 1988. Predicting private company failure. *Accounting and Finance*. 28 (2), 53-64.
- [8] Hall, G., "Factors distinguishing survivors from failures amongst small firms in the UK construction sector". *J. Manage. Stud.* (5). 1994. pp 737-760
- [9] Franks, J.R., "Predicting financial distress in farm businesses. *Eur. Rev. Agric. Econ.*" 1998., 25 (1), 30-52.
- [10] Mitchell, J., Roy, R.V., "Failure prediction models: performance, disagreements, and internal rating systems", Working Paper Research 123, National Bank of Belgium 2007.
- [11] Altman, E.I., Iwanicz-Drozowska, M., Laitinen, L., Suvas, A., "Financial distress prediction in an international context: a review and empirical analysis of Altman's Z-Score model". *J. Int. Financial Manage. Acc.* 28 (2) .2017. pp131-171.
- [12] Altman E., "Financial Ratio, Discriminant Analysis and Prediction of Corporate Bankruptcy". (1968) .*The Journal Of Finance*, September 1968, pp. 589-609.
- [13] Edmister, RO., "Un test empirique de l'analyse des ratios financiers pour la prédiction des défaillances des petites entreprises ».1972. *J. Financ. Quant. Anal.* 7 (2), 1477-1493.
- [14] Bhimani, A., Gulamhussen, M.A., Lopes, S.D. "Owner liability and financial reporting information as predictors of firm default in bank loans". 2014. *Rev. Acc. Stud.* 19 (2),pp.,769-804
- [15] Bhimani, A., Gulamhussen, M.A., Lopes, S.D., "Accounting and non-accounting determinants of default: An analysis of privately-held firms". (2010).*J. Acc. Publ.Policy.* 29 (6), 517-532.
- [16] Grunert, J., Norden, L., Weber, M., "The role of non-financial factors in internal credit ratings". (2005). *J. Bank. Financ.* 29 (2), pp.509-531.
- [17] Keasey, K., Watson, R., "The non-submission of accounts and small company financial failure prediction". (1988). *Acc. Busin. Res.* 19 (73), 47-54.
- [18] Hall, G., "Factors distinguishing survivors from failures amongst small firms in the UK construction sector". (1994). *J. Manage. Stud.* 31 (5),pp 737-760.
- [19] Fitzpatrick P.I. «A comparison of the successful industrial enterprise with those of failed companies», *The Accountants Publishing Company*(1932)
- [20] Jones, T., and Wang, T., "Predicting private company failure: A multi-class analysis". *J. Int. Financ. Markets Inst. Money* 61 (2019). pp161-188
- [21] Barboza F, Kimura H, Altman E. "Machine learning models and bankruptcy prediction". *Expert Systems with Applications*, 2017, vol183, pp 405-417
- [22] Iturriaga, F.L., et Sanz, I.P., "Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks". *Expert Systems with Applications*.2015. 42(6)
- [23] Lee, S., Choi, W.S., 2013. "A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis". *Expert Syst. Appl.* 40, 2941-2946.
- [24]YAJNIK M., MOON S., KUROSE J. & TOWSLEY D. "Measurement and modeling of the temporal dependence in packet loss". In *Proceedings of IEEE INFOCOM 1999*, p. 345-352
- [25] Almaskati N., Bird R., Yeung D., Lu Y., "A horse race of models and estimation methods for predicting bankruptcy" *Advances in Accounting*, Volume 52, March 2021
- [26] Perbo, P., and Arabnezhad, E., "A Machine Learning-based DSS for mid and long-term company crisis prediction". *Expert Systems with Applications*. Volume 174, 15 July 2021, 114758
- [27] Marsoa, S., and El Merouania, M., "Predicting financial distress using hybrid feedforward neural network with cuckoo search algorithm". *Procedia Computer Science*, Volume 170, 2020, Pages 1134-1140
- [28] Khemakhem S. et Boujelbene Y., "Credit Risk Prediction: A Comparative Study between Discriminant Analysis and the Neural Network" Approach. *Journal of Accounting and Management Information Systems*, 2015, vol. 14, issue 1, 60-78
- [29] ZAMMEL M. & KHOUIFI W «The use of Bayesian Networks in the prediction of bankruptcy», *Revue Française d'Economie et de Gestion* (2020).Volume 1. pp: 60-88
- [30] Zhao Z, Xu S, Kang B H, et al. Investigation and improvement of multi-layer perceptron neural networks for credit scoring[J]. *Expert Systems with Applications*, 2015, 42(7): 3508-3516

- [31] Olden, J.D., Jackson, D.A., "Illuminating the black box a randomization approach for understanding variable contributions in artificial neural networks". *Ecol. Model.* 154, (2002). 135–150.
- [32] Lin, Y. W. Hu and Tsai C. , "Machine Learning in Financial Crisis Prediction: A Survey," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421-436, July 2012,
- [33] Veganzones, D., et Severin, e., 2018. An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems* 112
- [34] Young Kim S.Y , Upneja A (2021). "Majority voting ensemble with a decision trees for business failure prediction during economic downturns". *Journal of Innovation & Knowledge*, Volume 6, Issue 2, April–June 2021, Pages 112-123.
- [35] Wang, G., Ma, J., « A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine”. *Expert Syst. Appl.* (2012), 39, 5325–5331.
- [36] Yi Qu., Pei Quan., Minglong Lei., and Yong Shi., “Review of bankruptcy prediction using machine learning and deep learning techniques” *Procedia Computer Science* 162, (2019). pp 895–899
- [37] Feng, Y.; Zhang, P., Yang, M., Li, Q., and Zhang, A., “Short Term Load Forecasting of Offshore Oil Field Microgrids Based on DA-SVM”. *Energy Procedia*, Volume 158, February 2019, Pages 2448-2455.
- [38]Li, X., Wang, F., Chen, X., “Support vector machine ensemble based on Choquet integral for financial distress prediction”. *International Journal of Pattern Recognition and Artificial Intelligence.*(2015), Vol. 29, No. 04
- [39] Khemakhem S.et Boujelbene Y.,“Support vector machines for credit risk assessment with imbalanced datasets”. *Int. J. Data Min. Model. Manag.* .(2018) , 10(2): 171-187
- [40] Martín-Torres, S., Carvelo, A., Casado, A., Rodríguez, L.,. “Authentication of the geographical origin and the botanical variety of avocados using liquid chromatography fingerprinting and deep learning methods”.*Chemometrics and Intelligent Laboratory Systems*. Volume 199, 15 April 2020, 103960
- [41] Bardos M. (1998a). Detecting the risk of company failure at the Banque de France, *Journal of Banking and Finance*, vol. 22, pp. 1405-1419.
- [42] Ooghe H. et Waeyaert N., « Oorzaken van falen en falingspaden: Literatuur Overzicht en conceptueel verklarings model Economisch en Sociaal Tijdschrift », 57 (4) (2004), pp. 367-393
- [43] Refait C. « La prévision de la faillite fondée sur l’analyse financière de l’entreprise: un état des lieux », *Économie et Prévision*, 2004, 162,129-147.
- [44] Bardos M., « Scoring sur données d’entreprises : instrument de diagnostic individuel et outil d’analyse de portefeuille d’une clientèle », *Revue MODULA*, (2008),38, 159 – 177.
- [45] Tam Y. , Lei Y. , Zheng J. and Wang W. , "ASR error detection using recurrent neural network language model and complementary ASR," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 2312-2316.
- [46]Young Kim S.Y , Upneja A., “Majority voting ensemble with a decision trees for business failure prediction during economic downturns”. *Journal of Innovation & Knowledge*, Volume 6, Issue 2, April–June 2021, Pages 112-123.
- [47] Kaplan,A.D.H., “Small Business: its place and problems” New York McGraw Hill Book Co.1948.
- [48]SCHINDLER M. & EPPLER M.J., “Harvesting Project Knowledge: A Review of Project Learning Methods and Success Factors” , *International Journal of Project Management*, (2003). Vol. 21, n° 3, pp. 219-228.
- [49] CEPME, « Crédit d’équipement des PME ». (1986), Les causes de défaillance des entreprises industrielles, Rapport de synthèse
- [50] Caisse Nationale des Marchés de l’Etat (CNME), "L'origine des difficultés des entreprises françaises tombées en règlement judiciaire", *Bulletin d'Information Economique*, (1976), n°73, 4ème trimestre, pp 41-70.
- [51] Gresse C. (1994), Les entreprises en difficulté, *Economica* 1994
- [52] OOGHE, H. ; WAEYAERT, N. (2004): “Oorzaken van falen en falingspaden: Literatuur overzicht en conceptueel verklaringsmodel”, *Economisch en Sociaal Tijdschrift*, Vol. 57, n°4, pp. 367-393
- [53] Lelogeais, L., “The role of qualitative variables in the early detection of default risk”, *Banque de France*(2004), *Business Management, Business Monitor* 44-1356
- [54] Géniaux, I., et Bonnardel , S.,M., « La gestion des connaissances et des compétences dans les petites entreprises technologiques : une approche empirique », *Revue internationale P.M.E* , Volume 14, Number 2, 2001
- [55] Bachelet, C., « La situation de la communication interne en PME : ébauche d'une typologie », *Revue internationale P.M.E*. Volume 11, numéro 2-3, 1998
- [56] Refait, C., « Soutien financier ou mise en faillite de l’entreprise ?Comprendre la décision de la banque », *Revue Finance Contrôle Stratégie – Volume 8, N° 1, mars 2005*, p. 131 – 157
- [57] Petersen, C., & Plenborg, T. (2012). “Financial statement analysis: valuation, credit analysis and executive compensation”. England: Pearson Education Limited
- [58] Hornik, K. [1993]. “Some new results on neural network approximation”. *Neural Networks*, 6(8), 1069–1072.