

Prévision de la défaillance des entreprises tunisiennes à l'aide des méthodes non paramétriques : Réseaux de neurones et SVM

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 — La prévision de la défaillance des entreprises est un champ d'investigation, relevant du domaine économique et de gestion, qui attire l'attention des chercheurs et des praticiens depuis longtemps. Avec le développement des technologies de l'information, ce domaine d'étude passe de l'utilisation des techniques paramétriques vers l'utilisation des techniques non paramétriques basées sur l'utilisation d'algorithmes d'apprentissage automatique supervisé et non supervisé pour créer des modèles de prévision de la défaillance à partir de l'analyse des états financiers des entreprises. Au niveau de ce papier, nous présentons dans la première partie en revue les méthodes d'apprentissage automatique classique telle que l'analyse discriminante multivariée (AD) et la régression logistique (RL), ainsi que les nouvelles méthodes non paramétriques basées sur l'intelligence artificielle en particulier les réseaux de neurones (RN) et les machines à vecteurs de support (SVM).

Nous avons testé la performance de ces deux méthodes en termes de prévision de la défaillance. A la suite, les résultats sont comparés pour tester leurs degrés de précision tout en se basant sur le critère de la matrice de confusion. Nous constatons que les modèles basés sur la méthode des RN et des SVM sont plus performants que les modèles classiques basés sur l'analyse discriminante.

Abstract — Predicting business failure is a field of investigation, within the domain of economics and management, which has attracted the attention of researchers and practitioners for a long time. With the development of information technology, this field of study shifts from the use of parametric techniques to the use of non-parametric techniques based on the use of supervised and unsupervised machine learning algorithms to create failure prediction models based on the analysis of companies' financial statements. At the level of this paper, in the first part we present a review of classical machine learning methods such as multivariate discriminant analysis (AD) and logistic regression (RL), as well as new nonparametric methods based on the artificial intelligence in particular neural networks (RN) and support vector machines (SVM).

We tested the performance of these two methods in terms of failure prediction. Subsequently, the results are compared to test their degrees of precision while

being based on the criterion of the confusion matrix. We find that models based on the RN and SVM method perform better than classical models based on discriminant analysis.

Mots clés: Prévision, défaillance des entreprises, réseau de neurones, SVM, méthode non paramétrique.

I. INTRODUCTION

L'entreprise vit dans un environnement de plus en plus instable et turbulent qui peut affecter ses choix, sa performance et même sa survie. L'objectif de toute entreprise est le maintien de son activité et de faire face aux risques potentiels qui affectent son existence, d'où revêt l'importance cruciale des méthodes statistiques de prévision de la défaillance des entreprises.

En effet, l'évaluation de la santé financière d'une entreprise et l'appréciation de son risque de faillite sont d'un grand intérêt à la fois pour l'entreprise et pour les parties prenantes telles que les actionnaires, les créanciers, les gestionnaires et le gouvernement. Une prévision efficace et bien précise d'une faillite potentielle est très utile pour les acteurs du marché et présente un caractère informationnel important pour la gestion des risques et pour l'aide à la décision pour l'entreprise. En effet, la détection précoce de la défaillance revêt une importance cruciale du fait qu'elle permette de mener en temps opportun des actions préventives de restructuration ou de réorganisation à fin d'éliminer ou moins réduire le risque de faillite des entreprises. Plus précisément, la prévision des défaillances d'entreprise est cruciale pour l'analyse des flux de trésorerie, des budgets d'investissement, des marchés financiers, du développement de l'industrie et de la planification économique nationale.

En terme de méthodologie de prévision, les années 1960 ont marqué le développement d'un nombre considérable de travaux de recherche traitant ce domaine et proposant une panoplie de méthodes et de démarches aux service des chefs d'entreprises et des analystes financiers afin de se

prémunir contre le risque de faillite. La régression logistique (logit) et l'analyse discriminante multivariée (MDA) ont été largement appliquées à la défaillance d'entreprise à partir de cette date.

Depuis que les techniques d'intelligence artificielle ont commencé à être appliquées pour prédire les défaillances d'entreprises dans les années 1990, les techniques sont devenues plus robuste, plus sophistiquées et plus raffinées. Les approches du réseau de neurones (RN), la machine à vecteurs de support (SVM), et l'arbre de décision (DT) ont été largement appliqués à partir de cette date caractérisées comme nouvelles techniques non paramétriques de prévision.

L'objet de notre travail s'inscrit dans une perspective déductive qui vise à: (1) développer un modèle de prévision de défaillance d'entreprise performant en utilisant des données financières relatives aux entreprises tunisiennes durant la période 2018-2019 et (2) évaluer de manière exhaustive les variables financières qui sont susceptible d'expliquer la défaillance. Aussi, ce travail essaie de présenter la méthodologie employée lors de la modélisation de prévision et de montrer en quoi l'utilisation des méthodes non paramétrique adaptées aux réseaux de neurones et aux SVM offre une alternative utile à celles traditionnelles employées par les études expérimentales antérieures et cela en comparant les résultats en matière de détection issue des méthodes classique comme l'analyse discriminante avec les résultats obtenu par les réseaux de neurones et par les SVM.

Le reste du document est divisé en cinq parties. La partie 2 fournit un cadre théorique pour la recherche. La partie 3 décrit la présentation des données et la méthodologie utilisée dans l'analyse. La partie 4 présente les principaux résultats des modèles de prédiction des défaillances d'entreprise obtenue avec la méthode de l'analyse discriminante. Les parties 5 et 6 présentent les principaux résultats des modèles et les implications de ces résultats, tout en se basant successivement sur la technique du réseau de neurones et des SVM. Enfin, la dernière partie présente les conclusions retenues et les suggestions proposées pour de futures recherches.

II. REVUE DE LA LITTÉRATURE

Les répercussions dramatiques engendré suite à la faillite d'une entreprise sur l'environnement interne et externe de l'entreprise a suscité l'attention des économistes, des praticiens et des chercheurs depuis plusieurs années et a justifié le recours aux méthodes statistiques et techniques dans le but d'élaborer des modèles de prévision permettant la détection précoce de la défaillance des entreprises afin de limiter ses conséquences et prendre les mesures nécessaires au temps opportun. La prévision de la défaillance est un champ d'investigation qui remonte aux années 30 avec l'étude pionnier de [1], depuis cette date on a assisté à plusieurs générations de méthodes et de modèles permettant l'anticipation d'une faillite potentielle. Même s'il n'y a pas un grand changement au niveau du choix du thème et de la

sélection des variables utilisées, c'est l'évolution de la méthodologie et des outils statistiques qui ont fait de ce domaine d'étude un thème toujours d'actualité. En effet, la prévision de la faillite d'une entreprise peut être appréciée comme un problème de classification, qui essaie d'affecter un ensemble d'entreprises, sur la base d'un échantillon d'apprentissage, en deux groupes: groupe d'entreprises saines et groupes d'entreprises défaillantes. Deux méthodologies statistiques ont marqué l'histoire de ce type de modèle de prévision, les méthodes paramétriques et celles non paramétriques.

Les méthodes paramétriques de classification statistique essayent de créer une relation fonctionnelle entre les variables explicatives et la variable à expliquer. A partir des années 1960, la littérature a été caractérisée par la naissance de plusieurs méthodes, ayant pour but la modélisation du risque de faillite de l'entreprise, l'une des plus célèbres étant l'analyse discriminante. En effet l'approche unidimensionnelle de [2] a été parmi les premiers travaux visant à dissocier entre les entreprises saines et les entreprises défaillantes par l'étude des ratios financiers, l'objectif de Beaver était d'affecter les entreprises entre les deux groupes sur la base du ratio le plus discriminant. Pour cela, un seuil statistique a été déterminé de manière à séparer au mieux le groupe des entreprises saines à celui des défaillantes. La méthode de Beaver fournit donc un indicateur à la fois simple et efficace. En revanche le manque de robustesse lié à l'unicité du ratio utilisé explique la rareté de l'utilisation de cette méthode par la suite. De nombreuses études ont eu par la suite comme l'analyse discriminante multidimensionnelle qui fournit une explication plus riche de la situation de l'entreprise tout en se basant sur plusieurs variables explicatives, cette technique aboutit à la formulation d'une fonction score qui présente la combinaison linéaire d'un certain nombre de ratios jugés les plus discriminant. La famille des scores Z développée par [3] aux Etats Unis a connu un grand succès. Le premier de ces scores est formé à partir d'un échantillon composé de 66 entreprises étudiées durant la période de 1946 à 1965. Cette fonction score est formé de 5 ratios jugés les plus significatifs et les plus inter-corrélés. A la suite, une autre fonction Z' a été développée, cette fonction introduit à la fois 4 ratios jugés les plus discriminant et deux seuil critiques. Ce score a été suivi d'un score $Zeta$ qui introduit des critères de stabilité des résultats, de service de la dette et de liquidité du bilan. En France, de nombreuses études ont été développées à partir des années 70, par ([4], [5], [6]) utilisant cette méthode. Des travaux successifs ont également été élaborés au sein de la banque de France ([7], [8]). D'autres techniques faisant partie des méthodes paramétrique de classification sont apparues pendant les années 80 à côté de l'analyse discriminante, ces techniques sont les modèles recourant aux régressions sur variables qualitatives : Logit et Probit. Plusieurs recherches ont été menées pour comparer ces techniques par apport à l'analyse discriminante: l'étude réalisée par [9], au niveau de laquelle, elle a essayé de comparer la discrimination

linéaire aux modèles Probit à partir des mêmes données, elle a montré la supériorité de la première grâce à sa plus grande robustesse, des études successives aboutissent à la conclusion inverse. Dans ce cadre, d'autres études ([10], [11]) ont obtenu de meilleurs taux de classification respectivement par un modèle Logit et Probit que par un modèle d'analyse discriminante linéaire.

A partir des années 90, les modèles recourant aux techniques dotées de l'intelligence artificielle ont attiré l'attention des économistes et des chercheurs dans le domaine de prévision de la défaillance des entreprises. La popularité et l'importance de ces méthodes au niveau de ce domaine d'étude se reflètent dans un grand nombre d'articles résumés dans plusieurs revues récentes ([12], [13], [14], [15]). Malgré la variété toujours croissante des méthodes dotées de l'intelligence artificielle, il existe trois techniques prépondérantes qui sont abordées dans les travaux de recherche déjà mentionnés: réseau de neurones, les machines à vecteurs de support (SVM) et les arbres de décision.

La méthode la plus utilisée qui a connu un grand succès depuis les années 1990, est celle des réseaux de neurones. L'étude de synthèse réalisée par [15] au niveau de laquelle il répertorié 38 articles ou les auteurs ont utilisé cette technique. L'utilisation de plusieurs architectures des réseaux a été appliquée dans le domaine de prévision de la faillite (par exemple, les perceptrons multicouches [16], réseaux de neurones à rétro propagation [17] et les réseaux de neurones probabilistes [18]). La méthode des réseaux de neurones à convolution (feed-forward) fait partie des premières méthodes d'apprentissage automatique utilisées pour prédire la faillite, elle sert donc de référence pour les méthodes employées à la suite. Dans le contexte Tunisien, plusieurs travaux de recherches comme ([19], [20], [21], [22]) ont essayé de tester la performance des réseaux de neurones en matière de prévision.

Les réseaux de neurones ont réalisé de bons résultats même dans des domaines très compliqués ([23], [24]). Bien que les modèles des réseaux de neurones soient très précis, ils sont souvent décrits comme "boîte noire" [25]. Cette question a été abordée dans une étude [26], dans laquelle les auteurs soutiennent qu'il est possible de découvrir le processus sous-jacents cachés dans un tel modèle.

La méthode SVM a réalisé un grand succès dans le domaine de la prédiction de faillite à partir de la fin des années 2000 [27]. Cette méthode a été caractérisée comme la technique la moins sensible aux données déséquilibrées [28]. De même cette méthode a été qualifiée par sa robustesse et par sa qualité de précision en termes de prévision de la faillite [29]. Aussi, cette technique peut rendre séparables des ensembles de données inséparables en les mappant dans un espace de dimension supérieure à l'aide d'une fonction de noyau [30]. Plusieurs travaux de recherche ont essayé de tester la performance des méthodes SVM en terme de prévision ([31], [32], [33], [34], [35], [36]). La limite adressée à la technique SVM réside dans le

choix de la meilleure fonction noyau parmi les quatre formes présentés par cette méthode [37].

Dans les méthodes citées ci-dessus, il est difficile de dire que la précision d'un modèle obtenu avec une telle méthode est systématiquement meilleure que celle d'un autre modèle dans toutes les circonstances. Dans la plupart des situations, les performances de ces modèles dépendent du problème analysé [38].

III. LE CHOIX METHODOLOGIQUE

A. Présentation de la base de données

Grâce à la collaboration de L'Agence de Promotion de l'Industrie et de l'Innovation, de quelques experts comptables et de certains commissaires aux comptes nous avons pu obtenir une base de données qui comprend 340 entreprises saines et défaillantes pour les années 2018 et 2019. Pour déterminer l'échantillon définitif, nous avons procédé à un appariement en faisant correspondre à chaque entreprise défaillante une entreprise saine de même taille. Après l'élimination de certains 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, dont les données comptables sont disponibles en 2018 et 2019. Les ratios financiers issus de ces données comptables constituent notre base de données (Voir Tableau. I).

Dans le cas présent, nous justifions le choix de nos ratios par:

- Leur récurrence dans la littérature ([39], [40], [41], [42]).
- Leur pertinence par rapport à l'analyse financière, en incorporant les ratios de base existants dans la plupart des modèles de prévision de la faillite: ratios de liquidité, de rentabilité de gestion, de productivité et de structure financière.

Le choix des ratios financier qui sont capable d'analyser la situation des entreprises est une étape primordiale et importante dans l'élaboration d'un modèle de prévision de la défaillance des entreprises. Conformément à l'enseignement de l'analyse financière, la rentabilité économique ou financière de l'entreprise, la structure de son bilan, le niveau de sa liquidité, et sa capacité de remboursement sont les éléments les plus corrélés à la défaillance.

TABLEAU I : LES RATIOS FINANCIERS CHOISIS

| RATIOS | Code |
|-------------------------------|------|
| Rentabilité financière | R01 |
| Rentabilité économique | R02 |
| Rentabilité d'exploitation | R03 |
| Rendement du capital investi | R04 |
| Marge de profit brut | R05 |
| Rotation de l'actif | R06 |
| Rotation des immobilisations | R07 |
| Rotation des capitaux propres | R08 |
| Rotation des stocks | R09 |

| | |
|--|------------|
| Délai de recouvrement des créances | R10 |
| Délai de règlement des fournisseurs | R11 |
| Taux de rentabilité | R12 |
| Taux de croissance du chiffre d'affaires | R13 |
| Liquidité générale | R14 |
| Liquidité réduite | R15 |
| Liquidité immédiate | R16 |
| Liquidité de l'actif | R17 |
| Taux d'endettement | R18 |
| Endettement à MLT | R19 |
| Endettement à court terme | R20 |
| Autonomie Financière | R21 |
| Equilibre financier | R22 |
| Indépendance financière | R23 |
| Couverture des charges financières | R24 |
| Capacité de remboursement des dettes | R25 |
| Solvabilité générale | R26 |
| Ratio des capitaux propres | R27 |
| Ratio des immobilisations | R28 |
| Indicateur de taille | R29 |

B. Retraitement des données

La base de données est divisée en deux sous échantillons pour pouvoir tester la performance des méthodes 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'évaluation des techniques utilisées. Pour cette raison, nous avons effectué 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.

C. Présentation des logiciels employés

Pour le traitement des données et la modélisation de la défaillance, nous avons eu recours au logiciel SPSS pour l'analyse et le traitement des données, à Matlab pour la configuration du réseau de neurones et à R pour le paramétrage des SVM. Le logiciel SPSS a servi pour les traitements statistiques de base (certains recodages de variables, statistiques descriptives, tests d'hypothèses...), L'analyse discriminante, la sélection des variables à l'aide du critère Lambda de Wilks et du test de Fisher et l'estimation des paramètres des fonctions discriminantes. Tous les autres traitements ont été développés par nos soins : les procédures (Framework) du réseau de neurones employé pour effectuer les classifications (Perceptron multicouches) et celles permettant de paramétrer R pour l'application des SVM.

IV. PROCESSUS DE MODELISATION DE LA DEFAILLANCE A L'AIDE DE L'ANALYSE DISCRIMINANTE

A. Analyse de l'ensemble des ratios financiers

Dans une première étape, une batterie de vingt-neuf ratios financiers, codés de R01 à R29, a été constituée. Le choix de ces ratios est basé sur la sélection de ceux qui sont couramment utilisés dans le thème de prévision de la défaillance ou ceux qui présentent un contenu informationnel significatif dans l'analyse de la situation des entreprises. Les thèmes adoptés pour le choix de ces ratios sont: la rentabilité, la productivité, la gestion, la liquidité et le financement et enfin la structure financière.

La performance du processus de modélisation est dans l'identification et la sélection des ratios financiers qui affectent de manière significative la probabilité de défaillance [43].

TABLEAU III : TEST D'EGALITE DES MOYENNES

| Ratios | Lambda de wilks | F | Signification |
|------------|-----------------|---------------|---------------|
| R01 | ,894 | 35,280 | ,000 |
| R02 | ,973 | 8,256 | ,004 |
| R03 | 1,000 | ,146 | ,703 |
| R04 | 1,000 | ,128 | ,721 |
| R05 | 1,000 | ,052 | ,820 |
| R06 | ,995 | 1,626 | ,203 |
| R07 | ,988 | 3,467 | ,064 |
| R08 | 1,000 | ,147 | ,702 |
| R09 | ,997 | ,886 | ,347 |
| R10 | ,961 | 11,958 | ,001 |
| R11 | ,989 | 3,256 | ,072 |
| R12 | ,766 | 90,847 | ,000 |
| R13 | ,999 | ,312 | ,577 |
| R14 | ,969 | 9,585 | ,002 |
| R15 | ,976 | 7,443 | ,007 |
| R16 | ,985 | 4,588 | ,033 |
| R17 | ,996 | 1,060 | ,304 |
| R18 | ,805 | 72,257 | ,000 |
| R19 | ,898 | 33,794 | ,000 |
| R20 | ,899 | 33,389 | ,000 |
| R21 | ,958 | 13,161 | ,000 |
| R22 | 1,000 | ,045 | ,831 |
| R23 | ,956 | 13,810 | ,000 |
| R24 | ,990 | 3,017 | ,083 |
| R25 | ,992 | 2,392 | ,123 |
| R26 | ,953 | 14,742 | ,000 |
| R27 | ,995 | 1,555 | ,213 |
| R28 | ,994 | 1,772 | ,184 |
| R29 | ,975 | 7,699 | ,006 |

A partir du test d'égalité des moyens réalisé au niveau du tableau ci-dessus, l'examen du test de Fisher montre la présence d'un certain nombre de ratios significatifs et d'autres qui ne le sont pas au seuil de 5%. Les ratios significatifs qui peuvent discriminer au mieux entre les deux groupes d'entreprises appartiennent au quatre types de ratios utilisés dans cette analyse. On retrouve quelques ratios de rentabilité (R01, R02), le deuxième groupe d'indicateurs se rapporte à des ratios de gestion (R10, R12).

Les ratios traduisant la liquidité de l'entreprise sont significatifs. Il s'agit des ratios 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 grandes entreprises sont moins vulnérables à la défaillance.

En examinant les résultats obtenu, nous avons remarqué que les ratios R01 et R02 sont significatifs ce qui confirme que les entreprises saines sont les plus rentables. La rentabilité d'une entreprise peut être appréciée par sa capacité à tirer des profits. Les entreprises rentables dégagent des profits et disposent d'une liquidité qui leurs permet d'honorer leurs engagements. En effet, ce constat peut être visualisé sous une optique d'exploitation économique ou financière, ce qui permet d'avoir une idée sur la rentabilité générale et de la performance de l'entreprise.

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 très endettées à court terme par rapport à leur activité. D'après l'analyse du ratio R18, on constate que les entreprises défaillantes se caractérisent par un taux d'endettement moyennant élevé que celui des entreprises saines. Plus ce taux est élevé, plus l'entreprise est endettée et dépendante, ce qui peut l'amener à une situation de cessation des paiements. Le ratio R20 montre que les entreprises tunisiennes défaillantes, sont plus endettées à court terme en comparaison avec celles qui sont saines. Cet endettement élevé de la part de ces entreprises est due essentiellement à un manque de liquidité, ce constat est prouvé en mettant l'accent sur le niveau de liquidité de ces derniers, à travers les ratios R14, R15 et R16 représentant les ratios de liquidité générale et réduite. Ces ratios montrent que les entreprises défaillantes sont caractérisées par un niveau de liquidité moins élevé que celui des entreprises saines. Cela peut paraître évident, dans la mesure où les entreprises défaillante souffrent d'un état de cessation de paiement et ne sont plus capables d'honorer leurs engagements vu leur contrainte de liquidité. Lorsque la solvabilité de l'entreprise devient douteuse, le rôle des établissements financiers s'accroît et ils sont amenés à prendre des mesures de resserrement des conditions de crédit. De ce fait, l'importance des charges financières subites par les entreprises défaillantes associées à des

niveaux d'endettement élevé rend ces entreprises moins solvables (R26).

L'analyse concernant l'autonomie financière (ratio R21) des entreprises, confirme l'idée précédente puisque les entreprises saines disposent d'une autonomie financière plus élevée que les entreprises défaillantes. Ce résultat a été aussi confirmé par le ratio R23 qui a pour but essentiel de mesurer l'indépendance de l'entreprise vis-à-vis de ses créanciers. En effet un niveau élevé de ce ratio observé au sein des entreprises défaillantes montre que ces dernières sont plus dépendantes vis à vis de leurs prêteurs. Ce ratio est calculé suivant le rapport des capitaux propres sur endettement total. Il représente l'état d'endettement de l'entreprise par rapport à ses fonds propres.

La solvabilité de l'entreprise se mesure à travers la capacité de l'entreprise de faire face à ses dettes exigibles à partir de ses actifs les plus liquides (R14). Cependant l'entreprise, n'est pas toujours capable de liquider ses actifs pour répondre à ses engagements, d'où le rôle du fonds de roulement. La couverture des engagements exigibles est appréciée à travers le fonds de roulement, qui permet à l'entreprise de faire face à toutes ses dépenses exigibles en attendant les encaissements R26. En effet, nous constatons que le ratio de solvabilité R26 est moins élevé au sein des entreprises défaillantes que chez les entreprises saines. Les entreprises saines sont capables de rembourser leurs dettes au moment de l'échéance ainsi que de payer les intérêts au moment où ils deviennent exigibles.

À l'issue de cette analyse, nous constatons que la baisse notable de la rentabilité ainsi que le besoin de la liquidité constaté chez les entreprises défaillante affectent leur solvabilité. Aussi, l'analyse des deux groupes d'entreprises confirme que les entreprises défaillantes se caractérisaient par des performances économiques et financière médiocres.

B. Validation de la méthode

A partir de l'examen du test de Fisher, on peut constater la présence de plusieurs variables explicatives pertinente et d'autres qui ne le sont pas ayant des taux de significativité supérieurs à 5%. Pour établir la fonction score à partir de l'utilisation de l'analyse discriminante, on a procédé à une purification des données (combinaison linéaire des variables différenciant au mieux les deux groupes d'entreprises) en gardant que les variables significatives.

Pour vérifier la validité de l'analyse discriminante menée, on a fait recours à certains indicateurs.

☉ Le Test de BOX

TABLEAU III: RESULTAT DU TEST MULTI VARIE DE BOX

| M de Box | 630,538 | |
|-------------------|---------------|------------|
| Approximativement | 5,712 | |
| F | ddl1 | 105 |
| | ddl2 | 276659,610 |
| | Signification | ,000 |

Le Test de Box est un test statistique qui permet d'examiner l'hypothèse nulle d'indépendance des données.

Ce test suppose que les matrices de variance covariance sont identiques dans les sous populations. Il est possible d'éprouver cette hypothèse à l'aide de la version multidimensionnelle du test de Bartlett. La statistique M de Box basée sur le logarithme népérien du déterminant de chaque matrice de variance-covariance locale permet de construire un test multi varié pour la comparaison des matrices de variances- covariance.

Comme son homologue uni-varié, le test multi-varié de Box est très sensible à l'hypothèse de normalité de distributions. Cette normalité n'étant pas acquise, en raison du caractère asymétrique des distributions des différents ratios, il convient de considérer le résultat du test comme une preuve de l'hétéroscédasticité de la distribution des ratios relativement à ces deux groupes.

La valeur obtenue de M de Box par ce test a une valeur égale à 630,538. La valeur obtenue est élevée et la signification du test de F tend vers 0. Les résultats obtenus montrent la pertinence de ce test.

☞ La corrélation globale

TABLEAU IV: VALEUR PROPRE ET COEFFICIENT DE CORRELATION CANONIQUE ASSOCIE

| Fonction | Valeur propre | % de la variance | % cumulé | Corrélation canonique |
|----------|---------------|------------------|----------|-----------------------|
| 1 | ,925 | 100 | 100 | ,693 |

Une autre mesure d'évaluation de l'analyse discriminante est le coefficient de corrélation canonique entre le sous espace engendré par les variables indicatrices des deux groupes et le sous espace des combinaisons linéaire des ratios financiers. Plus ce coefficient est proche de 1, plus la fonction discriminante est meilleure. La valeur obtenue s'élève à 0,693 d'où s'avère la pertinence de la méthode utilisée.

C. Formulation de la fonction score

La fonction score est une combinaison linéaire de tous les ratios utilisés, ayant le pouvoir discriminant le plus élevé. Elle est donnée par les coefficients présentés dans le tableau V.

A la suite de la vérification de la validité de l'analyse discriminante par les indicateurs analysés ci-dessus, on procède à l'élaboration de la fonction discriminante linéaire, fonction score.

TABLEAU VI: COEFFICIENTS STANDARDISES DE LA FONCTION DISCRIMINANTE

| Fonction | |
|----------|-------|
| R01 | -,321 |
| R02 | -,179 |
| R10 | ,177 |
| R12 | ,545 |
| R14 | -,174 |
| R15 | -,190 |

| | |
|-----|-------|
| R16 | -,126 |
| R18 | ,513 |
| R19 | ,356 |
| R20 | ,308 |
| R21 | -,210 |
| R23 | -,213 |
| R26 | ,244 |
| R29 | ,191 |

La fonction score établie s'écrit :

$$Z = -0,321R01 - 0,179R02 + 0,177R10 + 0,545R12 - 0,174R14 - 0,190R15 - 0,126R16 + 0,513R18 + 0,356R19 + 0,308R20 - 0,21R21 - 0,213R23 + 0,244R26 + 0,191R29$$

☞ Evaluation de la qualité de classification

L'objectif de la fonction discriminante établie concerne son utilisation pour classer de nouvelles entreprises aux groupes prédéfinis à partir des données fournies dans leurs états financiers. Pour s'assurer de la performance de la méthode, le pouvoir discriminant du score est contrôlé. Une des mesures de ce pouvoir discriminant est la matrice de confusion présentant les taux de bon classement. Si le score d'une entreprise est positif, celle-ci est dite bien classée par la fonction score si elle est affectée dans le groupe des entreprises saines. De même si le score d'une entreprise est négatif, celle-ci est dite bien classée par la fonction score si elle est affectée au groupe des entreprises défectives (Voir tableau VII).

TABLEAU VIII: VALIDATION DU MODELE

| | | Classe d'affectation prévue par le modèle | | Total | |
|----------|----------|---|----------------------|-------|------------------|
| | | Groupe d'entreprise | entreprise défective | | entreprise saine |
| Original | Effectif | entreprise défective | 133 | 17 | 150 |
| | | entreprise saine | 30 | 120 | 150 |
| | % | entreprise défective | 88,7 | 11,3 | 100,0 |
| | | entreprise saine | 20,0 | 80,0 | 100,0 |

Nous constatons que le modèle a pu classer 133 entreprises défectives parmi les 150 introduites initialement, ce qui donne un taux de bon classement des entreprises défectives de 88,7%. De même ce modèle a pu identifier 120 entreprises saines parmi les 150 introduites initialement, ce qui donne un taux de bon classement des entreprises saines de 80%. En contrepartie, le taux d'erreur de première catégorie est de l'ordre de 11,3%, alors que celui de la seconde est de 20%. Nous concluons donc, que l'application de ce modèle à notre échantillon, un an avant la défectivité, nous a permis de classer correctement 253 entreprises parmi les 300, soit un taux de bon classement de 84,3%.

La même fonction score appliquée à l'échantillon test (celui des données relatives à l'année 2018) a permis de donner des résultats qui sont présentées dans le tableau ci-dessous :

TABLEAU IXI: QUALITE DE REPRESENTATION DE LA FONCTION SCORE

| | | Classe d'affectation prévue par le modèle | | | Total |
|----------|----------|---|------------------------|------------------|-------|
| | | Groupe d'entreprise | entreprise défaillante | entreprise saine | |
| Original | Effectif | entreprise défaillante | 108 | 44 | 150 |
| | | entreprise saine | 30 | 120 | 150 |
| | % | entreprise défaillante | 72 | 28 | 100,0 |
| | | entreprise saine | 20 | 80 | 100,0 |

Cette fonction score a permis de classer 108 entreprises défaillantes parmi les 150 présentes dans l'échantillon test ce qui donne un taux de bon classement des entreprises défaillantes de 72 %. De même cette fonction a permis de classer 120 entreprises saines parmi les 150 présentes dans l'échantillon test, soit un taux de 80% de bon classement. Pour conclure, ce modèle appliqué 2 ans avant la défaillance permet d'aboutir à un taux de bon classement des entreprises qui s'élève à 76%. Le modèle établi permet un reclassement des entreprises avec des taux satisfaisants, en effet il convient de souligner que plus l'événement de défaillance se rapproche, plus l'exactitude de prévision s'améliore.

V. PROCESSUS DE MODELISATION DE LA DEFAILLANCE A L'AIDE DE LA METHODE NEURONALE

Le recours à la technique des réseaux de neurones artificiels a débuté avec le travail pionnier de [44] lors de l'application de l'algorithme de rétropropagation des erreurs, qui a fait naître la possibilité de pouvoir utiliser ou tirer conclusion des phénomènes présents pour la prévision des phénomènes futures. Cet algorithme a été développé par la suite par [45] et a abouti à deux phases: une phase de forward propagation et une phase de backward propagation.

La procédure de paramétrage du réseau de neurones correspond à une méthode de calcul des poids optimaux. L'algorithme couramment appliqué à cette optique est l'algorithme de rétro-propagation (Back propagation), qui peut être utilisé pour l'apprentissage surveillé.

A. Procédure d'apprentissage : L'algorithme de rétro propagation

Le réseau de neurone artificiel est organisé en plusieurs couches. Chacune de ces couches est composée de plusieurs neurones. La couche d'entrée est composée de K neurones au nombre des ratios sélectionnés qui sont censées d'expliquer le phénomène de 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} . La phase « Backward » détermine le degré d'ajustement des pondérations aux connexions individuelles. En effet, le réseau est appliqué à toutes les entreprises de l'échantillon d'apprentissage, puis l'erreur de classement est calculée et les poids synaptiques sont modifiés de manière à réduire cette erreur, c'est en fait avec la variation progressive de ces pondérations jusqu'à la classification optimale (c.à.d. la minimisation de l'erreur de classification) que le réseau se paramètre pendant la phase d'apprentissage. L'algorithme de rétro-propagation peut être appliqué pour n'importe quel type de fonction d'erreur. La fonction d'erreur dépend de tous les poids du réseau et doit être minimisée. A cause de la non-linéarité du réseau, on ne peut pas trouver un minimum global par une méthode analytique. C'est pour cette raison qu'on cherche un minimum local par une méthode itérative par une descente de gradient récurrente. L'algorithme d'optimisation le plus utilisé est celui de rétro-propagation.

C'est grâce à l'algorithme de rétro-propagation du gradient de l'erreur que l'utilisation des réseaux de neurones a connu un essor considérable, en effet la règle de minimisation d'erreur adoptée 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 . Les ajustements des poids sont déterminés dans le chemin de propagation à chaque niveau, tout en évaluant les dérivées partielles de la fonction E par rapport aux poids synaptiques. Après l'interaction de larges nombres de cycles, l'erreur est réduite au niveau acceptable, et le processus s'arrête.

On a appliqué 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 technique exige de l'utilisateur d'effectuer plusieurs optimisations afin d'identifier l'architecture neuronale optimale qui améliore le pouvoir prédictif du modèle dans l'identification de la situation des entreprises [46].

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 [47]. Les données d'apprentissages du réseau de neurones comprennent ainsi les 14 ratios jugés initialement les plus significatifs, ces ratios forment les neurones de la couche d'entrée. L'output désiré est formé de valeurs binaires : 1 pour les entreprises saines et 0 pour les entreprises défaillantes (couche de sortie), cet output forme le neurone de la couche de sortie. La fonction de transfert retenue est celle sigmoïde puisque la variable à expliquer est binaire.

Et enfin, nous avons retenu la moyenne des erreurs quadratiques (MSE) comme fonction de performance pour juger la qualité prédictive du modèle.

B. L'identification de la meilleure architecture

Pour construire le réseau de neurones, nous avons utilisé le 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 afférent afin d'aboutir à l'architecture optimale qui permet d'avoir l'erreur moyenne la plus faible avec le taux de bon classement le plus performant relatif à l'échantillon test. Le choix d'une seule couche cachée a été motivé par le travail de [48] qui a montré qu'un réseau établi à partir d'une seule couche cachée et par un nombre optimal de neurones peut aboutir à des résultats performants (voir tableau X).

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

| Nombre de | Performance | Nombre |
|-----------|---------------|-----------|
| 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 |

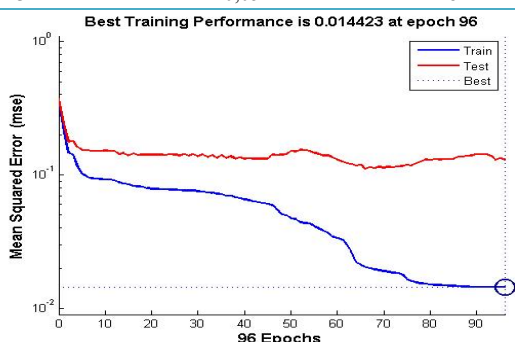


FIGURE 1: COURBE D'APPRENTISSAGE ET DE TEST DU RESEAU

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.

C. Evaluation de la performance du modèle

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

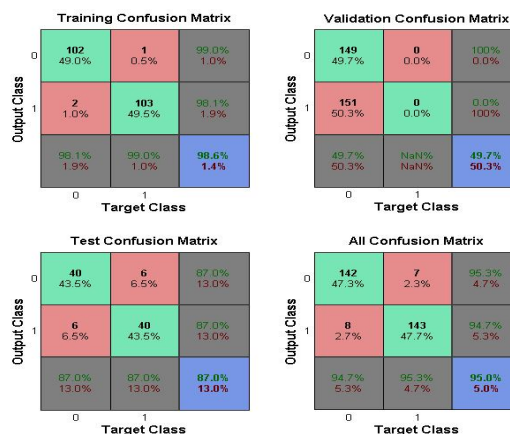


FIGURE 2: MATRICE DE CONFUSION DES RNA

TABLEAU XII: 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 | 40 | 6 | 46 |
| | | entreprise saine | 6 | 40 | 46 |
| % | | entreprise défaillante | 87 | 13 | 100 |
| | | entreprise saine | 13 | 87 | 100 |

TABLEAU XIII: 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 | 102 | 2 | 104 |
| | | entreprise saine | 1 | 103 | 104 |
| % | | entreprise défaillante | 98.1 | 1.9 | 100 |
| | | entreprise saine | 1 | 99 | 100 |

Ces tableaux montrent les résultats de classification des RNA. Le taux de bon classement global pour le sous-ensemble d'apprentissage est de 98,60% et pour le sous-ensemble de test est de 87,00%.

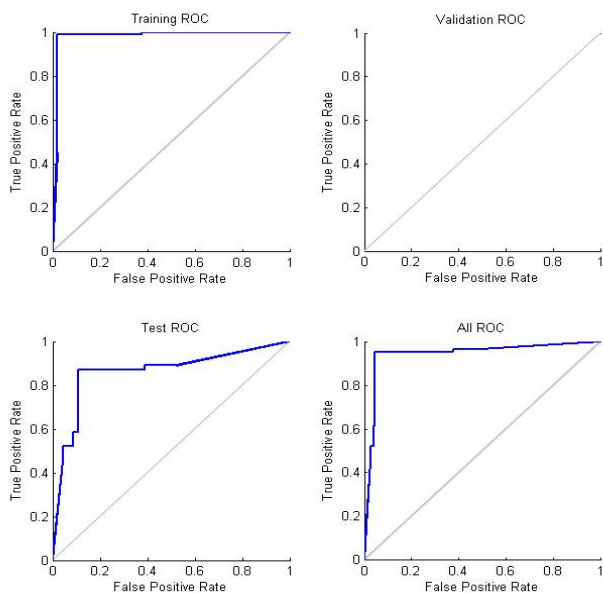


FIGURE 3: COURBE ROC OBTENU PAR RNA

Il est possible de caractériser numériquement la courbe ROC en calculant la surface située sous la courbe. C'est le critère d'aire sous la courbe (AUC, pour Area Under Curve) qui 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 de laquelle 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 d'un bon pouvoir discriminant performant.

VI. VI. PROCESUS DE MODELISATION DE LA DEFAILLANCE A L'AIDE DE LA METHODE SVM

A. Présentation de la technique

Vapnik[49] a développé la machine de vecteur de support comme nouvelle méthode d'apprentissage automatique de classification. Cette méthode basée sur la transformation de l'espace de représentation des données d'entrées en un espace de plus grande dimension dans lequel il est probable qu'il existe une séparation linéaire.

La SVM construit un hyperplan de séparation optimal pour classifier les données. Plusieurs études ont conclu que les SMV sont des modèles très prometteurs au sujet de la prévision, et surtout en termes du degré de précision et de capacité de généralisation.

Dans cette étude, nous appliquons SVM pour classifier les sociétés dans notre ensemble de données dans deux classes saines et défaillantes.

B. Procédure de fonctionnement

La méthode SVM est utilisée pour la résolution des problèmes de classifications. 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 ramener le problème de discrimination à celui de la recherche d'un hyperplan optimal (hyperplan séparateur). Cette technique construit deux hyperplans parallèles de chaque côté à l'hyperplan de séparation, pour calculer la marge. 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 de généralisation du classificateur la plus faible.

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 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.

La technique SVM nous permet d'essayer quatre modèles basés respectivement sur un noyau : linéaire, polynomial, à base radiale et sigmoïde. La bonne optimisation des paramètres des différents noyaux permet d'améliorer les performances des SVM en termes de classification. 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 choix du modèle efficace sera sur la base de la meilleure qualité de prévision obtenue tout en essayant les différents types des fonctions noyaux.

C. Modèle SVM

➤ Modèle SVM à base radiale

L'application de la méthode SVM à base radiale sur l'échantillon des variables choisis initialement nous a permis d'avoir une bonne performance (Voir tableau XVI).

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

| | | Classe d'affectation prévue par le modèle pour l'échantillon Test | | | Total |
|----------|----------|---|------------------------|------------------|-------|
| | | Groupe d'entreprise | entreprise défaillante | entreprise saine | |
| Original | Effectif | entreprise défaillante | 44 | 2 | 46 |
| | | entreprise saine | 10 | 36 | 46 |
| | % | entreprise défaillante | 95,65 | 4,35 | 100 |
| | | entreprise saine | 21,74 | 78,26 | 100 |

Ce tableau indique que le taux de bon classement global, obtenu par le modèle s'élève à 86,95%. Le taux de bon classement des entreprises saines s'élève à 78,26%, alors que celui des défaillantes s'élève à 95,66%. L'erreur de type I (classé une entreprise défaillante comme saine), s'élève à 4,35% alors que celui de type II est de 21,74%.

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

| | | Classe d'affectation prévue par le modèle pour l'échantillon apprentissage | | | Total |
|----------|----------|--|------------------------|------------------|-------|
| | | Groupe d'entreprise | entreprise défaillante | entreprise saine | |
| Original | Effectif | entreprise défaillante | 99 | 5 | 104 |
| | | entreprise saine | 6 | 98 | 104 |
| | % | entreprise défaillante | 95,20 | 4,80 | 100 |
| | | entreprise saine | 5,77 | 94,23 | 100 |

L'application du modèle avec l'échantillon d'apprentissage, nous a permis d'avoir un taux de bon classement de l'ordre de 94,71%.

➤ *Modèle SVM à base polynômiale*

Une bonne optimisation des paramètres de ce noyau permet de donner des résultats satisfaisants.

TABLEAU XVIII: 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 XIX: 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, obtenu par le modèle SVM à base polynômiale, s'élève à 80,43% pour l'échantillon test et 85,58% pour l'échantillon d'apprentissage. 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.

➤ *Modèle SVM à base du noyau sigmoïde*

L'application de la méthode SVM à l'aide du noyau sigmoïde nous a permis d'avoir une bonne performance.

TABLEAU XX: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON TEST A L'AIDE DES SVM A BASE SIGMOÏDE

| | | Classe d'affectation prévue par le modèle pour l'échantillon Test | | | Total |
|----------|----------|---|------------------------|------------------|-------|
| | | Groupe d'entreprise | entreprise défaillante | entreprise saine | |
| Original | Effectif | entreprise défaillante | 36 | 10 | 46 |
| | | entreprise saine | 10 | 36 | 46 |
| | % | entreprise défaillante | 78,26 | 21,74 | 100 |
| | | entreprise saine | 21,74 | 78,26 | 100 |

TABLEAU XXI: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON APPRENTISSAGE A L'AIDE DES SVM A BASE SIGMOÏDE

| | | Classe d'affectation prévue par le modèle pour l'échantillon apprentissage | | | Total |
|----------|----------|--|------------------------|------------------|-------|
| | | Groupe d'entreprise | entreprise défaillante | entreprise saine | |
| Original | Effectif | entreprise défaillante | 91 | 13 | 104 |
| | | entreprise saine | 23 | 81 | 104 |
| | % | entreprise défaillante | 87,5 | 12,5 | 100 |
| | | entreprise saine | 22,12 | 77,88 | 100 |

L'application de la technique SVM en utilisant le noyau sigmoïde nous a permis d'avoir un taux de bon classement satisfaisant qui s'élève à 78,26% pour le test et 82,69% pour l'apprentissage.

➤ *Modèle SVM à base linéaire*

Une bonne optimisation des paramètres de ce noyau permet de donner les résultats suivants (Voir tableaux XVII et XXI)

TABLEAU XVII: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON TEST A L'AIDE DES SVM A BASE LINEAIRE

| | | Classe d'affectation prévue par le modèle pour l'échantillon Test | | | Total |
|-----------------|----------|---|------------------------|------------------|-------|
| | | Groupe d'entreprise | entreprise défaillante | entreprise saine | |
| Original | Effectif | entreprise défaillante | 32 | 14 | 46 |
| | | entreprise saine | 5 | 41 | 46 |
| % | | entreprise défaillante | 69,56 | 30,43 | 100 |
| | | entreprise saine | 10,87 | 89,13 | 100 |

TABLEAU XXII: RESULTAT DE CLASSIFICATION DE L'ECHANTILLON APPRENTISSAGE A L'AIDE DES SVM A BASE LINEAIRE

| | | Classe d'affectation prévue par le modèle pour l'échantillon apprentissage | | | Total |
|-----------------|----------|--|------------------------|------------------|-------|
| | | Groupe d'entreprise | entreprise défaillante | entreprise saine | |
| Original | Effectif | entreprise défaillante | 84 | 20 | 104 |
| | | entreprise saine | 12 | 92 | 104 |
| % | | entreprise défaillante | 80,77 | 19,23 | 100 |
| | | entreprise saine | 11,54 | 88,46 | 100 |

Le taux de bon classement, obtenu par le modèle SVM à base linéaire, s'élève à 79,34% pour l'échantillon test et 84,61% pour l'échantillon d'apprentissage. Le taux de bon classement des entreprises défaillantes s'élève à 69,56% pour le test et 80,77% pour l'apprentissage alors que celui des saines s'élève à 89,13% pour le test et 88,46% pour l'apprentissage.

➤ *Le choix du modèle SVM optimal*

TABLEAU XXIII : CHOIX DU NOYAU OPTIMAL

| T. Bon Classement | N. radial | | N. polynomial | | N. sigmoïde | | N. linéaire | |
|-------------------|-----------|-------|---------------|-------|-------------|-------|-------------|-------|
| | T | % | T | % | T | % | T | % |
| | AP | 94,71 | AP | 85,58 | AP | 82,69 | AP | 84,61 |
| | T | 86,95 | T | 80,43 | T | 78,26 | T | 79,43 |

Les tests des taux de bon classement réalisés à l'aide de toutes les fonctions des noyaux montrent que le modèle SVM à noyau radial a le taux de bon classement le plus performant, d'où le choix de ce modèle comme technique de prévision de la défaillance.

VII. CONCLUSION

En guise de conclusion, il faut préciser que la prévision de la défaillance des entreprises a été étudiée en détail et de manière récurrente dans la littérature économique et financière. L'intérêt de notre recherche est de permettre une meilleure compréhension des risques et une meilleure identification des signes précurseurs de défaillance des entreprises tunisiennes, ainsi que de montrer l'apport des techniques dotées de l'intelligence artificielle en termes de prévision. Notre recherche a d'abord porté sur l'identification des facteurs financiers annonciateurs de la défaillance des entreprises. Il ressort de l'analyse effectuée par l'analyse discriminante que les PME tunisiennes défaillantes se caractérisaient par une faible liquidité et un fort niveau d'endettement témoignant leur dépendance financière. Aussi, l'application de l'analyse discriminante, de la technique des réseaux de neurones et de la technique SVM à deux échantillons d'entreprises saines et défaillantes, nous a permis d'obtenir des résultats significatifs et de proposer des modèles de prévision pertinents. Au moyen de ceci, nous avons constaté une supériorité des techniques non paramétriques par rapport à l'analyse discriminante traditionnelle en termes de degré de précision de la défaillance. Les performances des modèles obtenues avec la méthode des réseaux de neurones et des SVM sont presque équivalentes et s'élèvent respectivement à 87 % et 86,95% pour l'échantillon test et 98,6% et 94,71% pour l'échantillon d'apprentissage, alors que la performance obtenue avec la méthode de l'analyse discriminante s'élève à 76% pour l'échantillon test et 84,3% pour l'échantillon d'apprentissage. L'explication de ces résultats est liée en fait à des causes plutôt statistiques qu'économiques, puisque les techniques non paramétriques n'exigent pas des hypothèses restrictives comme c'est le cas pour la technique paramétrique.

Cette recherche présente une contribution dans l'appréciation du risque de défaillance à tous les meneurs de décision au sein ou en dehors de l'entreprise, elle leur permette de distinguer les ratios financiers qui sont susceptibles de présenter les signes précurseurs d'une faillite potentielle. De même, elle essaie de proposer une vision claire des techniques de prévision utilisées dans ce domaine et de proposer des modèles permettant la détection précoce de la défaillance des entreprises. Toutefois cette recherche présente certaines limites qui se situent aux niveaux de l'indisponibilité des données sur un horizon assez large qui dépasse les deux années et sur l'utilisation d'un échantillon multisectoriel. A cet égard, il faut noter que les modèles ont été appliqués sur des échantillons multisectoriels, en effet l'appartenance d'une entreprise donnée à un tel secteur exerce une certaine influence sur le

degré de discrimination des ratios au sein du modèle. Pour cela, il s'avère utile d'investiguer les secteurs les plus touchés par la défaillance et de proposer un modèle spécifique à chacun d'eux.

Aussi, le présent travail peut être étendu par la prise en compte d'autres variables de type qualitatif qui influencent profondément la situation des entreprises.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Fitzpatrick P.I. «A comparison of the successful industrial enterprise with those of failed companies», The Accountants Publishing Company(1932)
- [2] Beaver W., “Financial ratios as predictors of failure, Empirical Research in Accounting: Selected studies”, Journal of Accounting Research, 6, 165-182
- [3] Altman E., “Financial Ratio, Discriminant Analysis and Prediction of Corporate Bankruptcy”, The Journal Of Finance, September 1968, pp. 589-609.
- [4] Altman E.I., Hadelman R.G., Narayanan P. (1977), Zeta analysis, a new model to identify bankruptcy risk of corporations, Journal of Banking and Finance, 1,29-51.
- [5] Collongues Y., « Ratios financiers et prévision des faillites des petites et moyennes entreprises », Revue Banque, n° 365, septembre (1977), p. 963-970
- [6] Conan, J., & Holder, N. “Variables explicatives de performances et contrôle de gestion dans les PMI ». Paris: Université de Paris IX(1979).
- [7] Bardos M., « Les scores de la Banque de France : leur développement, leurs applications, leur maintenance ». Bulletin de la Banque de France (2005), n°144.
- [8] Bardos M., « Scoring sur données d’entreprises : instrument de diagnostic individuel et outil d’analyse de portefeuille d’une clientèle », Revue MODULA (2008), n° 38, pp159 – 177
- [9] Bardos M. « Trois méthodes d’analyse discriminante, Cahiers Economiques et Monétaires, Banque de France », 2^Ème trimestre, n°33. (1989).
- [10] Kira D.S., D. Doreen et D. Nguyen “An application of artificial neural networks and statistical methods in qualitative evaluation of small business loans”, Colloque de l'ASMDA, Capri 1997
- [11] Lennox, C. “Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches.” Journal of Economics and Business.(1999). 51: 347–364.
- [12] Kumar, PR, Ravi, V., “Bankruptcy prediction in banks and firms via statistical and intelligent techniques”. European Journal of Operational Research, 2007, vol. 180, issue 1, 1-28.
- [13] Lin, Y. W. Hu and Tsai C. , "Machine Learning in Financial Crisis Prediction: A Survey" Applications and Reviews, vol. 42, no. 4, pp. 421-436, July 2012,
- [14] Sun, J., Li, H., Huang, Q.-H., He, K.-Y., “Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches”. Knowledge-Based Systems Volume 57, February 2014, Pages 41-56
- [15] Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Kumar, V., Ajayi, S.O., Akinade, O.O., Bilal, M., 2018. « Systematic review of bankruptcy prediction models: towards a framework for tool selection”. Expert Syst. Appl. 94, 164–184
- [16] 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)
- [17] 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.
- [18] 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
- [19] Khoufi W. et Feki R. « Détermination et analyse des caractéristiques des entreprises défaillantes à l'aide des cartes auto organisatrices de Kohonen (Cas des PMI Tunisiennes) ». La Revue Comptable et Financière, RECHERCHES EN COMPTABILITE ET FINANCE N° 1(2008), pp. 97-110
- [20] Ben Jabeur .S et Fahmi . Y, « Prévision de la défaillance des entreprises : une approche de classification par les méthodes de Data-Mining ». Gestion 2000.(2013)/4 (Volume 30), pages 31 à 45
- [21] 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
- [22] ZAMMEL M. & KHOUFI W «The use of Bayesian Networks in the prediction of bankruptcy», Revue Française d’Economie et de Gestion 2020 «Volume 1: Numéro 1» pp: 60-88
- [23] Ghatasheh, N., Faris, H., AlTaharwa, I., Harb, Y., & Harb, “A.. Business analytics in telemarketing: Cost-sensitive analysis of bank campaigns using artificial neural networks”. Applied Sciences, (2020), 10 (7), 2581
- [24] Ozkan-Gunay, E. N., & Ozkan, M. (2007). Prediction of bank failures in emerging financial markets: An ANN approach. The Journal of Risk Finance, 8(5), 465–480.
- [25] Hung, C., & Chen, J.. “A selective ensemble based on expected probabilities for bankruptcy prediction”. Expert Systems With Applications, (2009).vol 36(3), pp 5297–5303
- [26] 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.
- [27] 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,
- [28] Veganzones, D., et Severin, e., 2018. An investigation of bankruptcy prediction in imbalanced datasets. Decision Support Systems 112

- [29] 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.
- [30] Kittler, J.. "A framework for classifier fusion: Is it still needed?". (2000), *Lecture Notes in Computer Science*, 1876, 45–56.
- [31] Chaudhuri, A., De, K., « Fuzzy support vector machine for bankruptcy prediction ». *Appl. Soft Comput.* (2011) , 2472–2486.
- [32] 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.
- [33] Karaa, A., Krichène, A., *Credit–Risk Assessment Using Support Vectors Machine and Multilayer Neural Network Models: A Comparative Study Case of a Tunisian Bank*, *Journal of Accounting and Management Information Systems*, (2012) Vol. 11, No. 4, pp. 587-620
- [34] 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
- [35] 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
- [36] 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
- [37] Kim, S., Y., and Upneja, A., "Majority voting ensemble with a decision trees for business failure prediction during economic downturns". *Journal of Innovation & Knowledge* 6 (2021). pp : 112–123
- [38] Jones, S., and Wang, T.,"Predicting private company failure: A multi-class analysis" *J. Int. Financ. Markets Inst. Money* 61 (2019) 161–188
- [39] Bardos M. (1998a). Detecting the risk of company failure at the Banque de France, *Journal of Banking and Finance*, vol. 22, pp. 1405-1419.
- [40] Ooghe H. et Waeyaert N., « Oorzaken van faling en falingspaden: Literatuur Overzicht en conceptueel verklarings model Economisch en Sociaal Tijdschrift », 57 (4) (2004), pp. 367-393
- [41] 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.
- [42] 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.
- [43] Petersen, C., & Plenborg, T. (2012). « Financial statement analysis: valuation, credit analysis and executive compensation". England: Pearson Education Limited
- [44] McCulloch & Walter Pitts. « A logical calculus of the ideas immanent in nervous activity ». *The bulletin of mathematical biophysics* volume 5, pages115–133(1943)
- [45] Rumelhart D., Hinton G. & Williams R. ,"Learning representations by back-propagating errors", *Nature* volume 323, pages533–536 (1986)
- [46] 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.
- [47]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.
- [48] Hornik, K. [1993]. "Some new results on neural network approximation". *Neural Networks*, 6(8), 1069–1072.
- [49] Vapnik, V.N., 1995. « The Nature of Statistical Learning Theory ». Springer, Berlin.