

Application des réseaux de neurones artificiels pour la classification : cas des défaillances d'entreprises

Abdelkrim AAZZAB¹, Mimoun BENZAOUAGH²

et Ahmed ABRIANE³

¹Docteur en sciences économiques, EMAA Business School ; a.azzab@gmail.com

²Professeur Chercheur, FSJES, Agadir ; a.abriane@uiz.ac.ma

³Professeur Chercheur, EST, Agadir ; m.benzaouagh@uiz.ac.ma

Résumé :

Les techniques de classification statistique peuvent être divisées en techniques paramétriques et non-paramétriques dont on trouve les réseaux de neurones artificiels (RNA). Les RNA, contrairement aux techniques classiques de classification, ne nécessitent aucune hypothèse sur les variables et ils sont tout à fait adaptés pour des problèmes complexes non structurés (Parizeau, 2004). Cette robustesse a permis à ces techniques un engouement tout particulier dans divers domaines de recherche et principalement dans les défaillances des entreprises. En effet, les premières expériences de la modélisation de la défaillance des entreprises en appliquant les réseaux de neurones artificiels débutent en 1990 avec Bell et al.. Ensuite, l'utilisation de cette technique s'est intensifiée avec les travaux de Tam (1991), Keasey et Watson (1991), Dimitras et al. (1996), Altman et Narayanan (1997), Wong et al. (1997), Zhang et al. (1998), Coakley et Brown (2000), Aziz et Dar (2004), Balcaen et Ooghe (2004, 2006), Ravi Kumar et Ravi (2007) et Lin (2009).

L'objectif de cet article est de présenter et d'appliquer la méthodologie des réseaux de neurones à couches comme outil de classification pour la problématique de la défaillance des entreprises en se basant sur des données comptables d'un échantillon d'entreprises marocaines.

Mots clés : Techniques de classification, réseaux de neurones artificiels, défaillances d'entreprises.

Abstract :

The Statistical classification techniques can be divided into parametric and non-parametric techniques including artificial neural networks (ANN). The ANN, unlike conventional classification techniques, do not require any hypothesis about variables and they are well suited for complex unstructured problems (Parizeau, 2004). This robustness has allowed these techniques a particular craze in various fields of research and mainly in the business failures of companies. Indeed, the first experiences of modeling business failure by applying artificial neural networks begin in 1990 with Bell et al. Then, the use of this technique intensified with the work of Tam (1991), Keasey & Watson (1991), Dimitras & al. (1996), Altman & Narayanan (1997), Wong & al. (1997), Zhang & al. (1998), Coakley & Brown (2000), Aziz & Dar (2004), Balcaen & Ooghe (2004, 2006), Ravi Kumar & Ravi (2007) and Lin (2009).

The aim of this article is to present and apply the methodology of Multi-Layer Neural Networks as a classification tool for the problem of business failure based on accounting data from a sample of Moroccan companies.

Key words: Classification techniques, artificial neural networks, business failures.

Introduction

Les réseaux de neurones artificiels (RNA), inspirés des réseaux de neurones biologiques, constituent un domaine de recherche permettant d'aborder les problèmes de perception, de mémoire, d'apprentissage et de raisonnement. Comme techniques de classification statistique, contrairement aux techniques paramétriques, les RNA sont robustes aux erreurs de spécification et apparaissent comme étant des approximatrices universels particulièrement parcimonieux (Cybenko, 1989 ; Funahashi, 1989 ; Barron, 1993 ; Hornik et al., 1994 et Dreyfus et al., 2008).

L'utilisation des réseaux de neurones pour la prévision des défaillances d'entreprises débute véritablement dans les années 1990 avec les travaux d'Odom et Sharda (1990)¹. Cette méthode, qui s'inspire du traitement de l'information effectué par le cerveau humain², consiste à développer un algorithme d'apprentissage qui traite un ensemble d'informations pour obtenir un résultat. De multiples études et travaux de recherches sur la défaillance des entreprises ont pratiqué cette technique dont on trouve Bell et al. (1990), Keasey et Watson (1991), Dimitras et al. (1996), Altman et Narayanan (1997), Wong et al. (1997), Zhang et al. (1998), Coakley et Brown (2000), Aziz et Dar (2004), Balcaen et Ooghe (2004, 2006), Ravi Kumar et Ravi (2007) et Lin (2009).

Dans notre étude, en s'inspirant des travaux empirique, nous avons sélectionné d'abord deux classes d'entreprises selon le critère défaillant non-défaillant, ensuite, nous avons choisi un ensemble de variables dites explicatives et enfin on a cherché d'établir une relation statistique entre ces variables et l'état dichotomique d'être ou de ne pas être défaillant. La qualité du modèle élaboré dépend du taux de bon classement des entreprises dans la classe correspondante.

I. Présentation des réseaux de neurones artificiels

Un réseau de neurone artificiel est un processus composé d'unités de traitement simples, parallèlement distribuées, destinées à accumuler la connaissance expérimentale et la rendre opérationnelle³ (Aleksander et Morton, 1990). Il existe plusieurs types d'architectures des réseaux de neurones artificiels. Cependant l'une des architectures les plus connues et les plus

¹ BORITZ J.E. et D. B. KENNEDY D.B., [1995], "Effectiveness of Neural Network Types for Prediction of Business Failure", *Expert Systems with Applications*, vol. 9, pp. 503-512.

² C'est le célèbre psychologue américain William James (1890) qui a introduit le concept de mémoire associative qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de loi de Hebb.

³ ALEKSANDER I. et MORTON H., [1990], "An introduction to neural computing", London, Chapman et Hall.

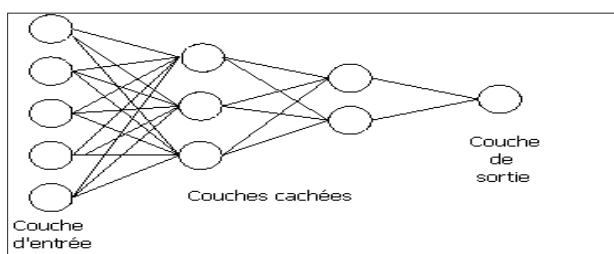
appliquées dans le domaine économique et financier est celle des réseaux de neurones multicouches (MLP).

1.1. Architecture des réseaux de neurones à couches

Un réseau de neurones à couches est un ensemble de neurones formels associés en couches (d'entrée, cachées et de sortie) et fonctionnant en parallèle. Chaque neurone formel est un ensemble d'opérateurs mathématiques simples. C'est une fonction algébrique non linéaire, paramétrée et à valeurs bornées.

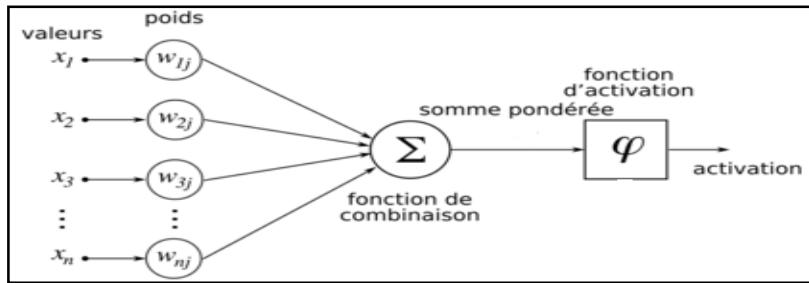
L'architecture d'un réseau de neurones est la façon dont les neurones sont ordonnées et connectés entre eux au sein d'un même réseau. C'est alors un maillage de plusieurs neurones, généralement organisé en couches : couche d'entrée, couches cachées et couche de sortie :

Figure 1 : Exemple simple de réseau de neurones



Un neurone artificiel ou formel est considéré alors comme un dispositif qui reçoit, à partir de d'autres neurones ou de l'extérieur, des stimulations par des entrées (inputs), au nombre de n , et les pondère grâce à des valeurs réelles appelées coefficients synaptiques ou poids synaptiques. Ces poids peuvent être positifs, on parle alors des synapses excitatrices, ou négatifs et dans ce cas on dit synapses inhibitrices.

Figure 2 : Structure d'un neurone artificiel (Haykin, S. 1994: Neural Networks⁴)



Un neurone j calcule ainsi, un potentiel P_j , égal à la somme de ses entrées (inputs) (x_1, x_2, \dots, x_n) pondérées, par les coefficients synaptiques respectifs (w_1, w_2, \dots, w_n), à laquelle on ajoute un terme constant : le biais b_j . La valeur du potentiel P_j est donnée par l'équation suivante :

$$P_j = \sum_{i=1}^n w_{ij} x_i + b_j$$

Équation 1 : Calcul du potentiel du neurone

A ce potentiel, le neurone applique une fonction d'activation Φ , de manière à ce que la sortie y_j , calculée par le neurone, soit égale à $\Phi(P_j)$, tel que :

$$y_j = \Phi(P_j) = \phi\left(\sum_{i=1}^n w_{ij} x_i + b_j\right)$$

Équation 2 : Calcul de la sortie du neurone en appliquant une fonction d'activation

L'utilisation du biais est de nature à appliquer une transformation affine au potentiel. En fait, le biais est un paramètre externe du neurone j , il peut être intégré dans l'équation du potentiel, comme étant le signal x_0 qui prend la valeur 1, pondéré par le poids w_{0j} , dont la valeur est égale au biais b_j .

La valeur de sortie y_j (output) est émise par le neurone vers d'autres neurones ou vers l'extérieur. Ainsi, un neurone est caractérisé par trois concepts : son état interne c'est-à-dire son potentiel, ses connexions avec d'autres neurones et sa fonction de transfert⁵.

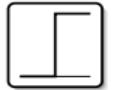
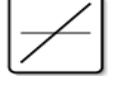
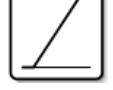
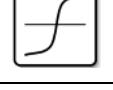
⁴ HAYKIN S. [1994], « Neural Networks. A Comprehensive Foundation », Macmillan, New York.

⁵ DAHMANI H. et ZAOUI A. [2005], « Application des réseaux de neurones artificiels pour la détection et la localisation des défauts de la machine asynchrone », Thèse d'ingénierat, université de M'sila.

1.2. Fonction de transfert

La fonction de transfert, appelée également fonction d'activation, est la fonction employée pour propager l'information de couche en couche. Les fonctions les plus courantes et les plus citées dans la littérature sont recensées dans le tableau suivant⁶ :

Tableau 1 : Liste des fonctions de transfert $a = f(n)$

Nom de la fonction	Relation Entrée/Sortie	Description de la fonction	Icône
Seuil	$\begin{cases} a = 0 & \text{si } n < 0 \\ a = 1 & \text{si } n \geq 0 \end{cases}$	<ul style="list-style-type: none"> - Une entrée négative ne passe pas le seuil ; - Une entrée positive ou nulle dépasse le seuil, et la fonction retourne 1 ; 	
Seuil symétrique	$\begin{cases} a = -1 & \text{si } n < 0 \\ a = 1 & \text{si } n \geq 0 \end{cases}$	<ul style="list-style-type: none"> - La fonction retourne -1 si une entrée est négative et retourne 1 si une entrée est positive ou nulle 	
Linéaire	$a = n$	<ul style="list-style-type: none"> - Affectation directe de l'entrée à la sortie 	
Linéaire saturée	$\begin{cases} a = 0 & \text{si } n < 0 \\ a = n & \text{si } 0 \leq n \leq 1 \\ a = 1 & \text{si } n > 1 \end{cases}$	<ul style="list-style-type: none"> - Une entrée négative ne passe pas le seuil ; - Une entrée nulle ou supérieure ou égale au seuil, la fonction affecte directement l'entrée à la sortie ; - Une entrée strictement supérieure au seuil, la fonction retourne 1 ; 	
Linéaire saturée symétrique	$\begin{cases} a = -1 & \text{si } n < -1 \\ a = n & \text{si } -1 \leq n \leq 1 \\ a = 1 & \text{si } n > 1 \end{cases}$	<ul style="list-style-type: none"> - Une entrée strictement inférieure à -1 (supérieure à 1), la fonction retourne -1 (1) ; - Autrement, la fonction affecte directement l'entrée à la sortie ; 	
Linéaire positive	$\begin{cases} a = 0 & \text{si } n < 0 \\ a = n & \text{si } n \geq 0 \end{cases}$	<ul style="list-style-type: none"> - Une entrée négative ne passe pas le seuil ; - Autrement, la fonction affecte directement l'entrée à la sortie ; 	
Sigmoïde	$a = \frac{1}{1 + \exp^{-n}}$	<ul style="list-style-type: none"> - Un arrangement entre la fonction seuil et la fonction linéaire 	
Tangente hyperbolique	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$	<ul style="list-style-type: none"> - La fonction affecte l'entrée à la sortie dont les asymptotes ont pour équation $y = -1$ et $y = 1$ 	
Compétitive	$\begin{cases} a = 1 & \text{si } n \text{ maximum} \\ a = 0 & \text{autrement} \end{cases}$	<ul style="list-style-type: none"> - La fonction retourne 1 pour une entrée maximum ; - Autrement la fonction retourne 0 	

1.3. Algorithme d'apprentissage

Les réseaux de neurones sont généralement optimisés par l'apprentissage qui représente la propriété fondamentale des réseaux neuronaux. Il s'agit d'une phase où le comportement du

⁶ PARIZEAU M., [2004], « Réseaux de neurones », Université de laval, 117 p.

réseau se modifie jusqu'à l'acquisition du résultat voulu. Dans le domaine de prévision des défaillances, l'algorithme de rétro-propagation du gradient est le plus connu.

On appelle souvent technique de rétro-propagation du gradient, l'algorithme permettant de corriger les erreurs en se basant sur le calcul du gradient et à l'aide de la rétro-propagation. L'apprentissage par l'algorithme de rétro-propagation du gradient est une application d'une méthode statistique connue sous le nom « approximation stochastique » qui a été proposée par Robbins et Monro (1951). Cet algorithme a été développé par plusieurs équipes de chercheurs (Le Cun 1986, Rumelhart *et al.* 1986, Parker 1985) et son principe de base consiste à minimiser une fonction dépendante de l'erreur. Cela consiste, à chaque configuration des poids correspondant à un coût, à rechercher un coût minimal sur une surface de coût. Cependant, la minimisation du gradient risque d'avoir des minima locaux au lieu du minimum global.

Avant de commencer l'apprentissage, les poids du réseau sont d'abord initialisés avec des valeurs aléatoires. Ensuite, on considère un échantillon d'apprentissage qui va être utilisé pour la phase d'apprentissage. Enfin, le réseau de neurones est appelé alors, à terme, de prévoir et donc de générer, à partir de cet échantillon, des valeurs similaires que celles désirées.

L'algorithme se présente comme ceci⁷ :

- Soit un échantillon \vec{x} qui représente l'entrée du réseau et \vec{t} la sortie correspondante désirée ;
- Le signal se propage en avant de la couche d'entrée vers la couche de sortie en passant par la ou les couche(s) cachée(s) : $x_k^{(n-1)} \alpha x_j^n$;
- Le calcul de la propagation vers l'avant se fait par une fonction d'activation g , par une fonction d'agrégation h ⁸ et à l'aide des poids synaptiques w_{jk}^{\rightarrow} entre le neurone $x_k^{(n-1)}$ et le neurone x_j^n (w_{jk}^{\rightarrow} indique bien un poids de k vers j).

$$x_j^{(n)} = g^{(n)}(h_j^{(n)}) = g^{(n)}\left(\sum_k w_{jk}^{(n)} x_k^{(n-1)}\right)$$

Équation 3 : La propagation vers l'avant

- Ensuite, après la propagation, le résultat \vec{y} est calculé à la sortie du réseau.

⁷ Nous avons adopté ici la notation adoptée par Byrd A. dans son ouvrage intitulé « Auto-calibration d'un réseau de capteurs de pollution », haute école d'ingénierie et de gestion de Canton de Vaud, 2006, pp. 21-22.

⁸ La fonction d'agrégation est souvent un produit scalaire entre les poids et les entrées du neurone.

- L'erreur de sortie, qui représente la différence entre \vec{y} et \vec{t} , est ensuite calculée pour chaque neurone de la couche de sortie :

$$e_i^{sortie} = g'(h_i^{sortie})[t_i - y_i]$$

Équation 4 : Erreur de sortie

- Cette erreur est propagée ensuite vers l'arrière $e_i^{(n)}$ ou $e_j^{(n-1)}$ grâce à la formule suivante :

$$e_j^{(n-1)} = g'^{(n-1)}(h^{(n-1)}) \sum_i w_{ij} e_i^{(n)}$$

Équation 5 : Propagation de l'erreur vers l'arrière

- Après, une mise à jour des poids dans toutes les couches se fait par la formule suivante :

$$\Delta w_{ij}^{(n)} = \lambda e_i^{(n)} x_j^{(n-1)}$$

Équation 6 : Mise à jour des poids des connexions

Où λ indique le taux d'apprentissage (inférieur à 1)

L'algorithme de rétro-propagation du gradient comporte donc deux phases :

- la première phase « *Forward* » : le signal fonction se propage vers l'avant du réseau, d'un neurone à un autre, à partir des neurones d'entrée. Pendant cette phase, les poids synaptiques restent inchangés. Pour chaque neurone j , le signal fonction est $y_j(n) = \varphi(P_j(n))$, où $P_j(n)$ est le potentiel à l'entrée du neurone j , qui est la somme pondérée par les poids synaptiques respectifs de tous les neurones de la couche précédente.
- La deuxième phase : « *backward* », dans ce cas c'est le signal erreur qui se propage, dans le sens contraire, couche par couche, à partir de la couche de sortie. Pour chaque neurone j , on calcule le gradient local $\psi_j(n)$ qui permet de calculer la correction des poids synaptiques, liaison par liaison, selon une règle dite Delta⁹ (aussi connue sous le nom least-mean-squares rule (LMS), delta rule, ou Widrow-Hoff rule). Dans le cas d'un neurone de sortie, le gradient local est égal au signal d'erreur multiplié par la dérivée de la fonction d'activation appliquée au potentiel à l'entrée de ce neurone :

⁹ Cette méthode développée par WIDROW B. et HOFF M.E. (1960) sous l'acronyme d'ADALINE (ADaptive LInear NEuron). C'est une règle d'apprentissage qui minimise l'erreur de sortie en utilisant une descente de gradient de l'erreur approximée. Après chaque itération, la correction s'applique au poids proportionnellement à l'erreur. La correction est calculée avant le seuillage (fonction d'activation). Pour plus de détail, voir DAVALO E. et NAIM P. [1990], « Des réseaux de neurones », Eyrolles, Paris, p. 232.

$$\psi_j(n) = \varphi_j(p_j(n)) \sum_{k=1}^{N_Q} \omega_{kj}(n) \psi_k(n)$$

Équation 7 : Calcul du gradient local

1.4. La mise en œuvre d'un réseau de neurones

Les réseaux de neurones peuvent approcher toute fonction de régression si l'on trouve un réseau qui rend minimum la fonction de coût mesuré par la somme des carrés des différences entre les valeurs calculées par le réseau de neurones et les valeurs souhaitées. La mise en œuvre d'un réseau de neurones nécessite les éléments suivants :

- Déterminer les grandeurs d'entrées pertinentes, c'est-à-dire les variables qui ont une influence significative sur le phénomène que l'on cherche à modéliser ;
- Collecter les données et construire un échantillon nécessaires à l'apprentissage et un échantillon test pour l'évaluation des performances du réseau de neurones ;
- Trouver le nombre de neurones cachés nécessaires pour obtenir une meilleure approximation ;
- Estimer les valeurs des paramètres qui correspondent à un minimum de la fonction de coût ;
- Évaluer les performances du réseau de neurones ;
- Effectuer plusieurs itérations, autant de fois que nécessaire, de tout ou partie de cette procédure, en fonction des résultats obtenus ;

Pour la prévision de la défaillance, la constitution d'un réseau avec une structure adaptée est la conception d'un modèle consistant à déterminer un ensemble de paramètres avant toute expérimentation du modèle. Il s'agit d'abord de choisir une architecture à travers la détermination du nombre de couches qui sont généralement au nombre de trois : la couche d'entrée, cachée et de sortie. Ensuite on définit le nombre de neurones pour chaque couche afin d'avoir une meilleure optimisation qui a pour objectif de diminuer dans la mesure du possible le nombre de synapses (poids) et de neurones afin de réduire la complexité du réseau et d'améliorer les temps de calcul, ainsi que de maintenir les capacités de généralisation. Enfin, on fixe les paramètres relatifs à l'apprentissage à savoir : l'échantillon d'apprentissage, le choix des entreprises et la période d'analyse, la sélection des variables d'entrées, les prétraitements des données, le mode d'apprentissage, le test du réseau et la validation du modèle.

II. Données et méthodologie de recherche

D'abord, nous avons entamés à la sélection des variables explicatives parmi un ensemble de variables candidates choisies sur la base des travaux empiriques précédents. Ainsi, la procédure de la sélection des variables est basée sur 1000 échantillons *bootstraps* et les variables retenues sont celles ayant le plus fort pouvoir de discrimination calculé en classant les variables par ordre croissant selon la statistique de Fisher et selon de la fréquence de sélection¹⁰. Ensuite, dans un deuxième point, nous avons développé des modèles prédictifs de la défaillance en se basant sur la méthode des réseaux de neurones artificiels.

2.1.Choix des entreprises

Notre démarche de collecte des données comporte trois étapes : le choix de la base de données, la sélection des entreprises et la le choix des indicateurs de défaillance. En effet, pour constituer notre échantillon, en se basant sur une source d'information officielle, nous avons acheté les documents de synthèses auprès de l'OMPIC¹¹. Il s'agit au départ de 160 entreprises dont 50% représente des entreprises défaillantes. Ainsi, pour délimiter notre champ d'investigation et pour assurer un maximum possible d'homogénéité de l'échantillon composé, nous avons sélectionné les entreprises opérantes dans le secteur de l'industrie et qui sont de petites et moyennes tailles. Le choix de ce secteur revient d'abord au nombre important des entreprises défaillantes qui y opèrent et à la possibilité de calculer l'ensemble des ratios financiers décrits par la théorie, une chose qui n'est pas possible pour les entreprises de services, par exemple, qui ne disposent pas de certains indicateurs.

Ainsi, le critère taille concerne les entreprises qui ont réalisé, au cours de trois exercices qu'on a retenus pour l'analyse, un chiffre d'affaires annuel qui ne dépasse pas 75MDhs ou un total du bilan qui n'excède pas 50MDhs. Notre échantillon final est composé de 132 entreprises dont la moitié est défaillante¹². Cet équilibre entre les deux types d'entreprises revient aux considérations empiriques qui montrent qu'un déséquilibre entre les classes a un effet négatif sur le taux de classement correct de chaque groupe et sur le taux global de bon classement (Altman, 1968 ; Berg, 2005 ; Westgaard et Van Der Wijst, 2001).

¹⁰ Selon SAUERBREI et SCHUMACHER (1992), les variables à retenir sont celles qui sont apparues dans au moins 70 % des cas.

¹¹ L'OMPIC (Office Marocain de la Propriété Industrielle et Commerciale) offre, à travers le service Directinfo, un accès aux états financiers des entreprises avec un prix de 60dhs/état. Cinq types de documents sont disponibles : le Bilan (BL), le Compte de Produits et Charges (CPC), l'Etat des Soldes de Gestion (ESG), le Tableau de Financement (TF) et l'Etat des Informations Complémentaires (ETIC)

¹² Une entreprise défaillante, dans notre cas, est celle soit définitivement radiée du registre de commerce, soit celle qui a vécu une procédure de redressement ou de liquidation judiciaire et les données concerne dans ce cas la période couvrant trois ans avant le déclanchement de cette procédure.

Pour les entreprises en bonne santé, nous avons entamé un choix au hasard sans aucune autre hypothèse. Alors que pour les entreprises défaillantes, il s'agit d'abord de les identifier auprès des tribunaux de commerce avant de demander leurs états de synthèses. Les tribunaux de commerce choisis sont ceux d'Agadir, de Marrakech et de Casablanca. Notre choix ici est motivé par la facilité d'accès aux informations et par la proximité.

Ainsi, pour chaque entreprise défaillante, nous avons demandé les documents de synthèses d'un exercice comptable avant la date de déclaration de défaut. Pour les non-défaillantes, il s'agit également d'un exercice tiré aléatoirement. Aussi, notre échantillon couvre une période de cinq ans allant de 2006 à 2010. Le choix de cette période est dû principalement aux difficultés liées à l'identification des entreprises défaillantes sur une période plus courte. Le tableau suivant résume la description des entreprises qui composent notre base de données selon la nature et selon les régions.

Tableau 2 : Répartition des entreprises selon les régions

	Agadir	Marrakech	Casablanca	Total
Non défaillante	18	22	26	66
Défaillante	13	17	36	66
Total	31	39	62	132

2.2.Choix des variables

Notre base de données est composée de 18 ratios financiers. Ces ratios sont calculés à partir des documents collectés afin de constituer une batterie pertinente et crédible susceptible de répondre à notre question relative aux facteurs explicatifs de la défaillance des entreprises. La justification de choix de ces ratios est basée principalement sur la littérature théorique et empirique (Altman, 1968, 1984, Bardos 1995, Conan et Holder 1979, Rose et Giroux, 1984, Refait, 2004).

Ainsi, la variable à expliquer est dichotomique, elle prend la valeur 1 si l'entreprise est défaillante et la valeur 0 si l'entreprise est non-défaillante. Pour les variables explicatives, et dans la mesure où il n'existe pas de théorie unificatrice définissant la défaillance des entreprises, notre travail s'inscrit également dans la même voie que la plupart des modèles empiriques qui commencent avec un nombre élevé de facteurs et les réduisent afin de ne garder que quelques-uns jugés comme les plus explicatifs du risque de défaillance. Alors, nous sommes limités donc à une batterie de base composée de 18 ratios selon leur popularité et leur performance dans les

études antérieures. L'annexe 1 synthétise les ratios de notre étude qui représentent l'ensemble des indicateurs financiers choisis.

2.3. Méthodes de sélection de variables

Nous avons opté pour les méthodes automatiques de sélection de variables en comparant entre deux méthodes pour enfin sélectionner celle qui présente plus de précision. En effet, nous avons procédé à la sélection des variables à l'aide de la méthode dite « *Stepwise Discriminant analysis SDA* ». Alors, nous avons comparé entre l'approche FORWARD et celle dite BACKWARD. Pour ces méthodes de sélection de variables et afin d'évaluer le rôle significatif d'une variable, nous utilisons la statistique F de Fisher. Il suffirait donc de comparer la p-value calculée pour la variable à évaluer et la comparer avec le seuil de signification choisi¹³. Ainsi, le lambda de Wilks, qui varie entre 0 et 1, représente l'indicateur privilégié pour l'évaluation statistique du modèle¹⁴. Il indique dans quelle mesure les centres de classes sont distincts les uns des autres dans l'espace de représentation. Tant qu'il tend vers 0 le modèle sera bon parce que les nuages sont bien distincts.

2.4. Construction du modèle neuronal

Le réseau de neurone développé est de type « Perceptrons multi-couches » avec la descente de gradient simple basée sur l'algorithme de rétro-propagation de l'erreur¹⁵ comme technique d'optimisation. Ainsi, nous avons retenu la fonction tangente hyperbolique comme fonction d'activation et l'erreur des moindres carrés comme fonction de coût. En outre, pour la modification des poids du réseau, nous avons opté pour un terme de moment et chaque couche est entamée d'un biais et un terme de régularisation. Enfin, nous avons retenu la somme des erreurs quadratiques (SSE) comme fonction de mesure de performance.

Pour le paramétrage du réseau, nous avons adopté l'apprentissage supervisé, à couches, non bouclé, complètement connecté, à une couche cachée et à une sortie linéaire. Pour la couche d'entrée (Input), il s'agit du vecteur des variables candidates sélectionnées pour l'apprentissage. Pour le nombre de neurones à introduire dans la couche cachée, il s'agit de tester les différentes configurations ayant conduit à un niveau d'apprentissage élevé. Pour la couche de sortie (output), la variable à expliquer est dichotomique. Il s'agit d'un vecteur qui prend la valeur 1 si

¹³ Il n'y a pas de repères pour définir une règle d'arrêt, cependant, nous avons choisi la valeur 3, 84 proposée par défaut par de nombreux logiciels et qui ressemble à la valeur critique d'un test à 5% lorsque nous travaillons sur un échantillon de quelques milliers d'individus.

¹⁴ BARDOS M., [2001], « Analyse Discriminante - Application au risque et scoring financier », Dunod.

¹⁵ RUMELHART et al., [1986], Op. cit.

l'entreprise est défaillante et la valeur 0 si l'entreprise est non-défaillante. Ainsi, pour assurer un meilleur apprentissage et pour stabiliser le processus de sélection de variables, nous avons employé des techniques de ré-échantillonnage de type *bootstrap*¹⁶.

Ainsi, nous avons recouru à la définition d'un générateur aléatoire par création d'une variable de partitionnement afin de recréer exactement les échantillons utilisés dans les analyses. Il s'agit d'une variable de Bernoulli générée aléatoirement avec un paramètre de probabilité de 0,7, modifiées de manière à prendre la valeur 1 ou -1, au lieu de 1 ou 0 (défaillante ou non-défaillante). Alors, les observations contenant des valeurs positives sur la variable de partitionnement sont affectées à l'échantillon d'apprentissage, celles avec des valeurs négatives sont affectées à l'échantillon de validation et celles avec une valeur égale à 0 sont affectées à l'échantillon de test. Ce dernier est formé pour éviter les problèmes de sur-apprentissage¹⁷ afin d'aider le réseau à rester « sur la bonne voie ». Pour les autres paramètres du réseau (le pas d'apprentissage, le terme de moment et les termes de régularisation des poids), les valeurs sont fixées sur la base des travaux empiriques relevées dans la littérature. Ainsi, le nombre d'itérations à retenir est celui pour lequel l'erreur ne varie quasiment plus au-delà de ce nombre. Enfin, dans le but de supprimer tout ce qui est modélisable afin de réduire la complexité du réseau et d'accélérer sa convergence, nous avons effectué des prétraitements concernant principalement la normalisation des données en se basant sur la méthode Min-Max.

III. Analyse des résultats

3.1. Performances de classification et sélection de variables

3.1.1. Matrices de confusion

Les matrices de confusion des deux méthodes de sélection de variables (Tableau 3) indiquent un taux de mauvais classement de 0,0909 pour la SDA de type Forward et un taux de 0,0985 pour la SDAB. Les taux d'erreur calculés sur les données d'apprentissage sont alors très optimistes et l'estimateur de l'erreur *bootstrap* donne l'avantage à la SDA de type Forward qui a une valeur de 0,1221 au lieu de 0,1279 pour la SDA de type Backward.

¹⁶Un échantillon *bootstrap* correspond à un échantillon de taille identique à l'échantillon initial et constitué à partir de ce dernier par tirage aléatoire avec remise. Cette méthode de ré-échantillonnage est recommandée par plusieurs auteurs (GUYON et ELISSEEFF, 2003 ; FERGUSON et al., 2003 ; ZELLNER, 2004). Par cette méthode, la sélection serait répétée sur différentes répliques de l'échantillon initial de façon à lisser les perturbations pouvant affecter la procédure.

¹⁷C'est-à-dire que le réseau recherche les modèles faux apparaissant dans les données d'apprentissage par variation aléatoire.

Tableau 3 : Performances de classification des méthodes de sélection de variables

SDA (FORWARD)				SDA (BACKWARD)			
Error rate				Error rate			
Bootstrap error estimation				Bootstrap error estimation			
Confusion matrix							
	D	ND	Sum		D	ND	Sum
D	61	5	66	D	59	7	66
ND	7	59	66	ND	9	57	66
Sum	68	64	132	Sum	68	64	132

La première méthode indique alors que 61 entreprises défaillantes ont été bien reclassées et 5 l'ont mal été. De même, pour les entreprises non défaillantes, 7 d'entre elles ont été mal reclassées et 59 sont bien reclassés. Au total, c'est donc 120 entreprises (60 + 57) qui ont été correctement reclassées soit un taux de bon classement de 90,90%.

3.1.2. Test de MANOVA

L'analyse de la variance multivariée (

Tableau 4) montre que c'est la méthode SDA de type Forward qui présente de bons résultats. En effet, elle a la statistique de lambda de Wilks la plus faible (0,37). Ce résultat est confirmé par les transformations de Bartlett ou de Rao qui statuent sur la significativité des écarts et qui aboutissent à la même conclusion au seuil d'erreur de 5%, on rejette alors l'hypothèse selon laquelle les centres de classes sont confondus (p-value=0).

Tableau 4 : Analyse de la variance multivariée

SDA (FORWARD)			SDA (BACKWARD)		
Stat	Value	p-value	Stat	Value	p-value
Wilks' Lambda	0,3789	-	Wilks' Lambda	0,3909	-
Bartlett C(18)	123,74	0,00	Bartlett C(18)	120,22	0,00
Rao F(18, 113)	41,31	0,00	Rao F(18, 113)	49,47	0,00

En mariant alors le résultat du test de MANOVA avec celui de la matrice de confusion, nous comprenons que la bonne tenue du modèle tient surtout à l'application de la méthode SDA de type Forward.

3.1.3. Sélection de variables

L'évaluation individuelle des variables prédictives montre que cinq variables qui contribuent à l'explication de la défaillance pour la SDA de type Forward et quatre variables pour la SDA de type Backward. Ainsi, le Tableau 5 qui montre ces résultats indique également que les variables communes entre les deux méthodes sont au nombre de quatre (R3, R5, R7 et R16).

Tableau 5 : Evaluation individuelle des variables prédictives (N)

SDA (FORWARD)	SDA (BACKWARD)
R7, R16, R5, R3 et R15	R3, R5, R7 et R16

L'ensemble de ces variables sont sélectionnées sur la base de la statistique F qui est significativement différente de zéro, car la p-value est inférieur à 5%.

3.2. Réseau de neurones

3.2.1. Architecture des modèles neuronaux

D'après le tableau 5, nous remarquons que, par l'emploi de l'ensemble des variables candidates, la meilleure architecture est celle utilisant une couche cachée avec un seul neurone (Net1_1 [18 1 1]). En effet, c'est l'architecture pour laquelle la somme des erreurs quadratiques est minimale pour l'échantillon d'apprentissage (7,68). La SSE pour l'échantillon test est de 2,89, ce n'est pas donc la valeur minimale mais elle correspond comme même à une faible valeur si on la compare avec les autres. Cette architecture a permis d'enregistrer un taux de bon classement de 87,3% pour l'échantillon d'apprentissage et un taux de 85% pour l'échantillon test (tableau 6).

Cependant, l'architecture optimale correspondante à l'emploi des variables sélectionnées par la méthode SDAF est celle composée d'une couche cachée avec 9 neurones (Net2_6 [6 9 1]). Quant à elle, cette architecture a permis d'enregistrer la valeur la plus faible de la somme des erreurs quadratiques pour l'échantillon d'apprentissage avec 7,86% et un taux d'erreur de 2,98% pour l'échantillon test. Pour ce réseau, le taux de bon classement n'est que de 84,8% pour l'échantillon d'apprentissage et de 85% pour l'échantillon test (tableau 6).

Tableau 6 : Récapitulation des tests des architectures des réseaux¹⁸

Emploi des 18 variables candidates			Emploi de 5 variables de sélection (SDAF)		
Architecture	SSE ¹⁹ de l'échantillon d'apprentissage	SSE ²⁰ de l'échantillon test	Architecture	SSE de l'échantillon d'apprentissage	SSE de l'échantillon test
Net1_1 [18 1 1]	7,68	2,89	Net2_1 [5 1 1]	10,83	2,78
Net1_2 [18 3 1]	10,15	2,39	Net2_2 [5 3 1]	13,79	2,69
Net1_3 [18 5 1]	9,61	2,82	Net2_3 [5 5 1]	9,83	3,08
Net1_4 [18 7 1]	9,96	2,61	Net2_4 [5 7 1]	10,89	3,22
Net1_5 [18 8 1]	11,42	2,83	Net2_5 [5 8 1]	9,75	3,11
Net1_6 [18 9 1]	11,30	2,74	Net2_6 [5 9 1]	7,86	2,98
Net1_7 [18 10 1]	9,58	2,91	Net2_7 [5 10 1]	11,17	2,91

Tableau 7 : Matrices de confusion des modèles neuronaux

Modèle_1: Emploi de 18 variables candidates				Modèle_2: Emploi de 5 variables sélectionnées					
Echantillon		Prévisions			Echantillon		Prévisions		
		D	ND	% correct			D	ND	%
Apprentissage	D	34	6	85,0%	Apprentissage	D	34	6	85,0%
	ND	4	35	89,7%		ND	6	33	84,6%
	% global	48,1%	51,9%	87,3%		% global	50,6%	49,4%	84,8%
Test	D	7	1	87,5%	Test	D	7	1	87,5%
	ND	2	10	83,3%		ND	2	10	83,3%
	% global	45,0%	55,0%	85,0%		% global	45,0%	55,0%	85,0%
Validation	D	15	3	83,3%	Validation	D	16	2	88,9%
	ND	1	13	92,9%		ND	1	13	92,9%
	% global	50,0%	50,0%	87,5%		% global	53,1%	46,9%	90,6%

3.2.2. Validation des modèles neuronaux

Quant à l'échantillon de validation, qui permet d'évaluer le réseau de neurones final et donc de valider le modèle, le pourcentage de bon classement montre que 88,9% des entreprises

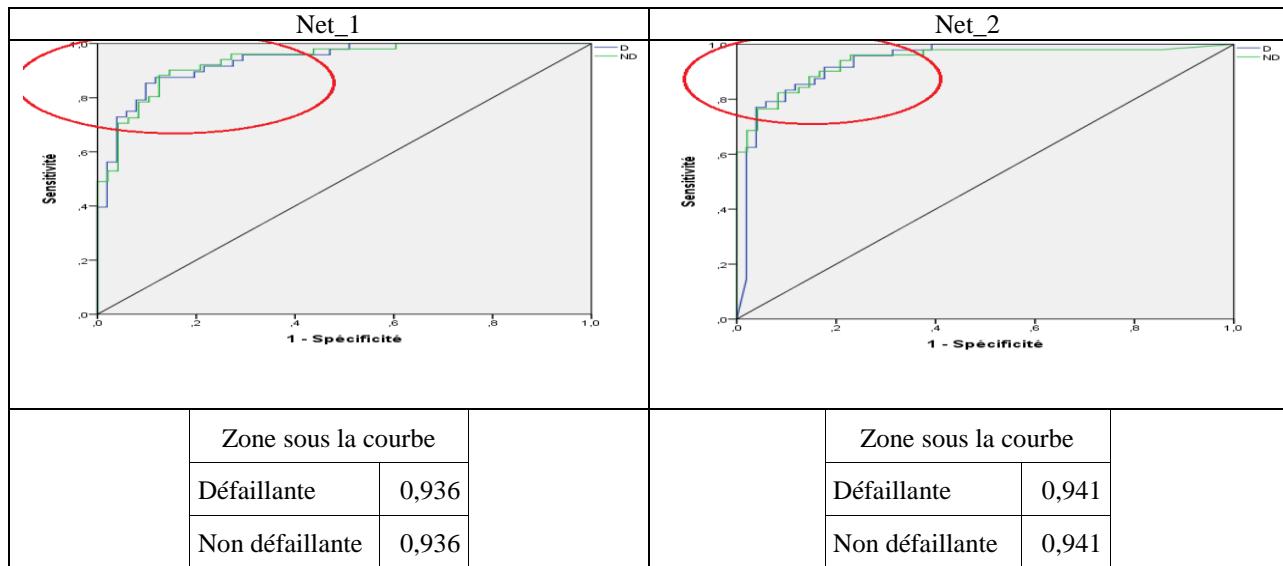
¹⁸ Il s'agit de signaler ici que ce tableau compte seulement quelques exemples d'architectures testées y compris celles qui enregistrent les faibles taux d'erreur. En fait, le nombre total des architectures de réseaux testées est de 220 (110 pour chaque niveau d'emploi des variables).

¹⁹ La somme des erreurs quadratiques (Sum squared error SSE) commise lors du classement des entreprises de l'échantillon d'apprentissage.

²⁰ La somme des erreurs quadratiques (Sum squared error) commise lors du classement des entreprises de l'échantillon test.

défaillantes son bien classées par le réseau Net_2 alors que le réseau Net_1 arrive seulement à bien classer 83,3% de ces entreprises. Pour les entreprises non-défaillantes, les deux réseaux ont classé correctement 92,9%. Par conséquent, le taux global de bon classement du Net_2 affiché est de 90,6% et celui de Net_1 n'est que de 87,5%. La validation des modèles neuronaux peut être renforcée par l'analyse des courbes ROC²¹.

Graphique 1 : Les courbes ROC des deux réseaux de neurones



Nous constatons pour les deux réseaux que les courbes ROC sont un peu proches du coin haut-gauche. Alors, les performances des facteurs de discrimination sont acceptables. Cela signifie que la probabilité pour que la fonction SCORE, développée par le modèle neuronal, place une entreprise défaillante devant une entreprise non défaillante est presque proche de 1 pour les deux modèles neuronaux. Ainsi, pour le réseau de neurones Net_2, pour un choix aléatoire d'une entreprise défaillante et d'une entreprise non-défaillante, il existe une probabilité de 94,1% que la pseudo-probabilité de manquement prévue par le modèle soit plus élevée pour entreprise défaillante. C'est-à-dire que la probabilité pour que le réseau place une entreprise défaillante devant une entreprise non-défaillante est de 94,1%. Ce taux est de 93,6% pour le réseau Net_1.

Il est apparu que, à partir de l'analyse de l'ensemble de ces éléments de validation, les deux modèles neuronaux sont valides et enregistrent de bons résultats. Ainsi, les taux enregistrés par les deux modèles sont très optimistes que ce soit pour l'échantillon d'apprentissage, de test ou de validation. Or, pour trancher entre eux, nous pouvons dire que, sans doute, le modèle

²¹ Une courbe ROC (Receiver Operating Characteristic) affiche les modalités pour chaque variable dépendante qualitative. Elle présente un affichage visuel de *sensibilité* et *spécificité* pour toutes les césures possibles dans un diagramme unique, ce qui constitue un outil plus clair et plus puissant qu'une série de tableaux.

neuronal basé sur 6 variables est le plus performant sur tous les niveaux. En effet, à l'exception de l'échantillon d'apprentissage, ce réseau a enregistré les taux les plus élevés pour l'échantillon test et pour celui de validation. De plus avec un nombre réduit de variables, le deuxième modèle a permis d'enregistrer des résultats plus marquants par rapport au premier modèle basé sur 18 variables. Alors, notre modèle des réseaux de neurones choisi est celui basé sur 6 variables explicatives avec une architecture composée d'une couche cachée composée de 9 neurones.

Conclusion

Au terme de cet article, il semble que les RNA sont des outils puissants et donnent de bons résultats en matière de classification. En effet, les pourcentages de bon classement de l'échantillon d'apprentissage, du test et de validation sont atteints des niveaux très optimistes. Ce résultat confirme ceux déjà établis par des études empiriques sur le sujet qui ont montré le succès de ces méthodes non paramétriques dans la prédiction de la défaillance d'entreprises (Kerling et Poddig, 1994 ; Oden et Sharada, 1990; Balcean et Ooghe, 2004 ; Abdou et al., 2008).

Cependant, les réseaux de neurones sont critiqués par leur opacité. En effet, ils sont incapables d'expliquer les résultats obtenus, car les règles de fonctionnement internes sont inconnues, d'où le nom de la boîte noire. Cependant, grâce à l'analyse de sensibilité, nous pouvons mesurer l'influence relative de chaque variable sur le résultat de sortie.

Notez enfin que cette étude comporte certaines limites dans le cadre où les modèles développés sont basés sur un nombre réduit d'observations. Ainsi, les tests de multicolinéarité, de multi-normalité et d'homoscédasticité ne sont pas vérifiés.

Bibliographie

- ALEKSANDER I. et MORTON H., [1995], "An introduction to neural computing", London, Chapman et Hall, 284 p.
- ALTMAN E. I., MARCO G. et VARETTO F., [1994], "Corporate Distress Diagnosis : Comparisons Using Linear Discriminant Analysis and Neural Network – The Italian Experience", *Journal of Banking and Finance*, vol. 18, n° 3, pp. 505-529.
- ANANDARAJAN M., LEE P. et ANANDARAJAN A., [2001], "Bankruptcy Prediction of Financially Stressed firms: An Examination of the Predictive Accuracy of Artificial Neural Networks", *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 10, n° 2 pp 69-81.
- ATIYA A. F., [2001], "Bankruptcy Prediction for Credit Risk Using Neural Networks : A Survey and New Results", *IEEE Transactions on Neural Networks*, vol. 12, n° 4, pp. 929-935.
- BARDOS M. et ZHU W. H., [1997], « Comparaison de l'analyse discriminante et des réseaux de neurones. Application à la détection des défaillances d'entreprises », *Revue de statistique appliquée*, tome 45, n°4, pp. 65-92.
- BERTRAND L., [2008], « Cours de Datamining - 8 : Modélisation, réseaux de neurones et de Kohonen », 270 p.
- BLAZY R. et COMBIER J., [1997], « La défaillance d'entreprise : Causes économiques, traitement judiciaire et impact financier », INSEE Méthodes n° 72-73, *Économica*, Paris ;
- BLAZY R., [2000], « La faillite : Éléments d'analyse économique », Paris: *Economica*.
- BLAZY R., [2008], « Une analyse comparative des procédures de faillite : France, Allemagne, Royaume-Uni », Paris, observatoire Oséo, la documentation française.
- BONE R., ASSELIN de BEAUVILLE J. P., ZOLLINGER M., [996], « Les réseaux de neurones artificiels : un apport potentiel aux études marketing », *Recherches et Application en Marketing*, vol.11, n°2, pp. 63-80.
- BOSE I. et PAL R., [2006], "Predicting the Survival or Failure of Click-and-Mortar Corporations : A Knowledge Discovery Approach", *European Journal of Operational Research*, vol. 174, n° 2, pp. 959-982.
- CLAUDE T., [1992], « Les Réseaux de Neurones Artificiels, Introduction au connexionnisme » EC2, 160 p.
- CONAN J. et M. HOLDER [1979], « Analyse des causes de faillite des PMI », Thèse d'Etat, Université ParisDauphine, CEREG, cités par DAIGNE, 1986.
- CRUCIFIX F. et DERNI A. [1993], « Symptômes de défaillance et stratégies de redressement de l'entreprise », Maxima, Laurent du Mesnil, collection Comprendre pour agir.

- DAHMANI H. et ZAOUI A., [2005], « Application des réseaux de neurones artificiels pour la détection et la localisation des défauts de la machine asynchrone », Thèse d'ingénierat, université de M'sila.
- DAVALO E. et NAIM P., [1990], « Des réseaux de neurones », *Eyrolles*, Paris, 232 p.
- DESMET P., [1996], « Comparaison de la prédictivité d'un réseau de neurone à rétropropagation avec celles des méthodes de régression linéaire, logistique et AID pour le calcul des scores en marketing direct», *Recherches et Application en Marketing*, vol.11, n°2, pp.17-27.
- DOMENS J. [2007], « Les défaillances d'entreprises entre 1993 et 2004 », PME/TPE, en bref, n° 23, Ministère de l'économie, des finances et de l'emploi – DCASPL.
- DU JARDIN P., [2007], « Prévision de la défaillance et réseaux de neurones : l'apport des méthodes numériques de sélection de variables », Thèse de doctorat, Université de Nice Sophia-Antipolis, 88 p.
- ETHERIDGE H. L. SRIRAM R. S., [1997], “A Comparison of the Relative Costs of Financial Distress Models : Artificial Neural Networks, Logit and Multivariate Discriminant Analysis”, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 6, n° 3, pp. 235-248.
- HERAULT L., [2002], « Réseaux de neurones: méthodologies et applications», Editions Eyrolles.
- KOUAM A., BADRAN F. et THIRIA S., [1992], « Approche méthodologique pour l'étude de la prévision à l'aide des réseaux de neurones », Actes du congrès Neuro-Nîmes 92 : Neural Networks and their Applications, Nîmes, 2-6 Novembre, pp. 117-127.
- MALECOT J.F. [1991], « Analyse historique des défaillances d'entreprises : une revue de la littérature », *Revue d'économie financière*, vol. 19, p. 205-226.
- MASCRET H. F. [1871], « Tableau synoptique des faillites », dans *Dictionnaire des faillites pour année*, Paris.
- PARIZEAU M., [2004], « Réseaux de neurones », Université de laval, 117 p.
- RIVALS I., PERSONNAZ L., DREYFUS G. et PLOIX J.L., [1995], « Modélisation, classification et commande par réseaux de neurones : principes fondamentaux, méthodologie de conception et applications industrielles », J.P. Corriou éd., Lavoisier Tec et Doc.
- TOUZET C., [1992], « Les réseaux de neurones : Introduction au connexionnisme », EC2, 160 p.
- WIRA P., [2009], « Réseaux de neurones artificiels: architectures et application », Université de Haute Alsace, Laboratoire MIPS.

Annexe

Liste des 18 principaux ratios de l'étude

Dimensions	Ratio	Formule	Code
STRUCTURE FINANCIERE	Capacité de remboursement	$\frac{Dettes \ de \ financement}{CAF}$	R1
	Autonomie financière	$\frac{Dettes \ de \ financement}{Financement \ permanent}$	R2
	Complément de capacité de remboursement	$\frac{Frais \ financier}{Chiffre \ d'affaires(HT)}$	R3
ACTIVITE	Taux de la valeur ajoutée	$\frac{Valeur \ Ajoutée}{Chiffre \ d'affaires(HT)}$	R4
	Taux de Marge Brut	$\frac{Marge \ Brute}{Chiffre \ d'affaires(HT)}$	R5
	Croissance de l'activité	$\frac{VA_t - VA_{t-1}}{VA_{t-1}}$	R6
LIQUIDITE	Liquidité générale	$\frac{Actif}{Dettes \ à \ court \ terme}$	R7
	Liquidité réduite	$\frac{Créances + trésorerie}{Dettes \ à \ court \ terme}$	R8
	Liquidité immédiate	$\frac{Trésorerie \ Actif}{Dettes \ à \ court \ terme}$	R9
ENDETTEMENT	Endettement global	$\frac{Dettes \ totales}{Passif}$	R10
	Structure d'endettement	$\frac{Dettes \ à \ court \ terme}{Dettes \ totales}$	R11
RENTABILITE	Taux de rentabilité financière	$\frac{Résultat \ Net}{Capitaux \ propres}$	R12
	Rentabilité économique	$\frac{EBE}{immob. \ brutes + BFRE}$	R13
RENDEMENT	Coefficient du capital	$\frac{immob. \ brutes + BFRE}{Valeur \ ajoutée}$	R14
GESTION	Taux d'écoulement des stocks	$\frac{Stock \ moyen \ de \ Mses + BFRE}{Achats \ de \ Mses(HT)} \times 360$	R15
	Taux d'écoulement des créances	$\frac{Clients \ et \ comptes \ rattachés}{Chiffre \ d' \ affaires(TTC)} \times 360$	R16
	Taux d'écoulement des crédits fournisseurs	$\frac{Fournisseurs}{Achats \ de \ Mses(TTC)} \times 360$	R17
L'EXCE. DE TRES-ORERIE D'EXP.	Facteur travail	$\frac{Frais \ de \ personnels}{Valeur \ ajoutée}$	R18