

Sélection de variables pour la classification multiclassé : Comparaisons et applicaion à la discrimination des banques selon le taux des prêts non performants

Asma Feki, Anis Ben Ishak, Saber Feki

► To cite this version:

Asma Feki, Anis Ben Ishak, Saber Feki. Sélection de variables pour la classification multiclassé : Comparaisons et applicaion à la discrimination des banques selon le taux des prêts non performants. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494721>

HAL Id: inria-00494721

<https://hal.inria.fr/inria-00494721>

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Sélection de variables pour la classification multiclassées :
Comparaisons et application à la
discrimination des banques selon le taux des prêts non performants*

Asma Feki, Anis Ben Ishak & Saber Feki

Unité de Recherche Gouvernance, ESC Sfax, Tunisie

Laboratoire BESTMOD, ISG Tunis, Tunisie

Department of Computer science, University of Houston, USA

Résumé

Le volume élevé des Prêts Non Performants (PNP) reste toujours le problème le plus important dans le secteur bancaire tunisien. La classification des banques selon leur niveau des PNP semble alors cruciale pour les autorités de supervision locales afin de prendre les actions correctives adéquates.

Dans cette étude, nous comparons trois méthodes de classification différentes, à savoir: les réseaux de neurones, les machines à vecteurs supports multiclassées et le classifieur bayésien sous l'hypothèse de normalité. Nous utilisons des données spécifiques aux dix banques commerciales tunisiennes collectées sur une période de sept ans, de 2000 à 2006. Il s'agit de vingt huit facteurs de risque appartenant au rating CAMELS. Ces facteurs sont ordonnés selon leur importance et sélectionnés en utilisant des algorithmes de sélection de variables.

La comparaison des trois méthodes utilisées montre une légère suprématie des machines à vecteurs supports multiclassées.

Abstract

One of the major problems in the Tunisian banking sector is the high rate of Nonperforming loans. Bank classification according to the level of NPL is then a crucial task for local authorities' supervision in order to be able to take the according corrective actions.

We compared in this study three different classification techniques, namely the back-propagation neural network, multi-class Support Vector Machines and a Bayesian classifier under a Gaussian distribution assumption. We used data from ten different commercial Tunisian banks during a period of seven years, from 2000 to 2006. The data consists of twenty eight selected and ordered variables belonging to rating CAMELS using feature selection algorithm.

The comparison between the three methods applied to our selected data shows that the best performance is obtained using the multiclass Support Vector Machines classifier.

Mots-clés: Apprentissage et classification, Sélection de variables, classification des banques, Modèle bayésien, Réseaux de neurones, Machines à vecteurs supports.

1 Introduction

En automne 2008, le système bancaire international a vécu des perturbations majeures débutées par une crise des sub-primes, en été 2007, incitée par des booms de crédits à des emprunteurs non performants. Le mauvais contrôle du risque de crédit représente donc un problème significatif auquel les banques doivent faire face.

Le risque de crédit peut être défini comme étant la perte probable supportée par la banque résultant de la non-performance des emprunteurs. On peut donc dire que le risque de crédit d'une banque se réalise lorsque l'un de ses débiteurs se trouve incapable d'honorer ses engagements contractuels ou lorsque la dégradation de la situation financière de l'emprunteur diminue la valeur des créances à l'actif du bilan de la banque. Ceci se manifeste essentiellement à travers l'accroissement du taux des prêts non performants (Non performing loans) qui représente la part des créances classées (2, 3 et 4) dans le total des engagements.

Le taux des prêts non performants représente, sans doute, le problème majeur des banques tunisiennes et qui reste, toujours, très élevé, en moyenne 20%. La classification des banques selon leur niveau de risque devrait donc faciliter la tâche des autorités de supervision locales de mettre en œuvre les mécanismes de contrôle efficaces.

L'objet de cette étude est d'appliquer trois approches supervisées de classification des banques Tunisiennes selon leur niveau des prêts non performants et de comparer la performance moyenne, sur les mêmes 50 partages aléatoires stratifiés, des différentes techniques de classification. On fait appel aussi bien, à l'approche paramétrique qu'à l'approche non paramétrique. Dans le premier cas, c'est la méthode bayésienne sous hypothèse gaussienne qui est appliquée du fait qu'elle pose moins d'hypothèses restrictives que les autres méthodes paramétriques. Au niveau de l'approche non paramétrique, on teste deux modèles, le premier utilise les réseaux de neurones et le deuxième utilise les machines à vecteurs supports (SVM).

Les données collectées appartiennent à la catégorie taille et au rating CAMELS. On sélectionne parmi les vingt huit variables celles les plus discriminantes en utilisant différentes méthodes de feature selection pour réduire le nombre de variables par rapport au nombre d'observations (problème de dimensionnalité) d'une part et pour sélectionner le meilleur paquet de facteurs de risque d'autre part.

Enfin, la comparaison entre les trois modèles montre que la meilleure performance est obtenue en utilisant des SVM multiclassés.

L'organisation du reste de l'article est comme suit : La section 2 s'intéresse à la description des principales caractéristiques méthodologiques de l'étude en termes de données, variables, sélection de variables et les techniques de classification utilisées. La section 3 sera consacrée aux principaux résultats ainsi que leurs interprétations. Enfin, La section 4 conclut ce travail.

2. Méthodologie

2.1. Données et variables utilisées

Après collecte et traitement des données de dix banques Tunisiennes pour une période de sept ans s'étalant entre 2000 et 2006, vingt huit variables ont été retenues et serviront d'inputs pour les différents modèles. Ces variables représentent les différents facteurs de risque appartenant à la catégorie taille (I) V13 et V14 et au Rating CAMELS : (C) adéquation du capital V1 et V17, (A) qualité des actifs V12, V18, et V26, (M) qualité de la gestion V2, V3, V4, V5, V11, V16 et V21, (E) rentabilité V6, V7, V15, V19, V20, V22 et V24 et (L) liquidité de la banque V8, V9, V10, V23 et V25 (S) sensibilité au marché V27 et V28.

D'autre part, on a calculé le taux des créances classées des banques en recourant à des guides et des documents de recherches de Tunisie Valeurs.

On a donc retenue la part des prêts non performants dans le portefeuille $\frac{\text{CRCL}}{\text{TENG}}$ (Créances classées / Total engagements), comme un critère de classification des banques, transformée en un output composé de trois classes de risque.

{	Classe 1 si $\frac{\text{CRCL}}{\text{TENG}} < 10\%$	Banque à faible risque
	Classe 2 si $10\% \leq \frac{\text{CRCL}}{\text{TENG}} < 20\%$	Banque à moyen risque
	Classe 3 si $\frac{\text{CRCL}}{\text{TENG}} \geq 20\%$	Banque à haut risque

Le taux des prêts non performants est jugé un bon critère de classification des banques Tunisiennes et il permet de dresser les deux seuils essentiels pour la classification : 20% la moyenne des taux des prêts non performants des banques Tunisiennes, un taux jugé élevé, et en dessous de 10% l'objectif fixé par la Banque Centrale de Tunisie (BCT) pour l'échéance (2009-2011), un taux jugé faible.

L'output est donc formé de trois classes de risque de crédit croissant. Cette sortie désirée est utile pour la comparer avec l'output calculé durant la phase d'apprentissage et ce pour la détermination de l'erreur. En adoptant ce critère de classification, on peut classifier les banques pour chaque année durant la période (2000-2006).

2.2. Sélection de variables et les modèles testés

La démarche empirique retenue pour cette recherche comporte deux étapes pour chacune des trois techniques de classification testées :

- Sélection du meilleur paquet de variables discriminantes
- Evaluation de la performance du modèle retenu.

Nous allons donc comparer trois méthodes de classification. Un modèle supervisé paramétrique (modèle bayésien sous hypothèse gaussienne) et deux autres supervisés non paramétriques (SVM multiclassés et réseaux de neurones : perceptron multicouches).

2.2.1 Le modèle bayésien sous hypothèse gaussienne

La méthode de Bayes sous hypothèse gaussienne est une approche paramétrique qui exige l'hypothèse de normalité des variables. Pour cela, on teste dans cette partie la normalité de toutes les variables au seuil de 5% en utilisant la statistique de Jarque-Bera.

La sélection de variables

L'approche bayésienne paramétrique correspond à une connaissance a priori sur les lois de probabilité dont il faut estimer les paramètres.

Pour la classe i on possède j points d'apprentissage notés x . On suppose que la distribution de probabilité de x , pour une dimension d , est gaussienne :

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right]$$

D'abord, on estime les paramètres des lois de probabilité par maximum de vraisemblance. Il s'agit des estimateurs de la moyenne μ et de la matrice de covariance Σ .

$$\mu = \varepsilon(x) = \int_{-\infty}^{+\infty} xp(x)dx$$

$$\Sigma = \varepsilon[(x - \mu)(x - \mu)'] = \int_{-\infty}^{+\infty} (x - \mu)(x - \mu)' p(x)dx$$

Enfin, les figures des densités marginales des 16 variables pour chaque classe permettent de sélectionner les variables discriminantes.

Pour l'élimination des variables redondantes, on s'est basé d'une part sur les figures des densités marginales estimées et d'autre part, sur la matrice de corrélation. En fait, on élimine tout d'abord les variables non discriminantes, c'est-à-dire celles dont les densités marginales pour chaque classe sont très avoisinantes, puis, on recourt à la matrice de corrélation en rejetant parmi les deux variables corrélées celle qui est la moins discriminante que l'autre.

L'évaluation du modèle

La décision bayésienne consiste à minimiser la probabilité d'erreur en associant à chaque réalisation x , la classe qui maximise sa densité de probabilité.

Les variables sélectionnées vont être introduites dans le modèle bayésien paramétrique. Le taux d'erreur du modèle retenu est calculé sur 50 partages aléatoires stratifiés.

2.2.2 Le modèle neuronal

Avant la phase d'apprentissage nécessaire pour l'élaboration du modèle neuronal optimal, il est indispensable de réduire le nombre de variables afin de résoudre le problème de dimensionnalité (un grand nombre de variables contre un nombre réduit d'observations) qui risque de biaiser les

La sélection de variables

L'élimination des variables redondantes repose sur un test de sensibilité permettant de sélectionner et hiérarchiser les meilleures variables. En répétant ce test 100 fois, on peut construire 100 paquets optimaux de variables permettant d'élaborer une meilleure hiérarchie de m variables $< n$ (où n désigne les 28 inputs). Les m variables sélectionnées sont celles qui ont les meilleures classes pour les 100 modèles. Ce même test sera appliqué sur le sous-ensemble de variables choisies (les m variables) et il sera répété 50 autres fois afin de choisir une hiérarchie de p variables tel que $p < m$. On réinitialise plusieurs fois le modèle avec les p variables en ajoutant ou en éliminant une variable afin de dégager le meilleur paquet. Les variables du paquet optimal seront les neurones d'entrées de 50 modèles neuronaux établis sur 50 partages aléatoires stratifiés.

L'évaluation du modèle

Le modèle non paramétrique supervisé utilisé dans cette étude correspond à un réseau feed-forward dans lequel il n'y a pas de retour en arrière de l'information. L'architecture de ce réseau est de type perceptron multicouches. L'algorithme choisi en matière d'apprentissage est l'algorithme de rétro propagation du gradient. On retient comme fonction de performance l'erreur quadratique moyenne (MSE). Enfin, on construit un modèle neuronal sur l'échantillon d'apprentissage et on retient le taux d'erreur sur l'échantillon test. Cette procédure est répétée sur 50 partages aléatoires stratifiés.

2.2.3 Le modèle SVM multiclassés

Plusieurs méthodes basées sur les machines à vecteurs supports (SVM) reposent principalement sur un algorithme de recherche et un critère d'évaluation pour mesurer la performance des sous-ensembles de variables.

La sélection des variables

Dans cette étude, la procédure de sélection de variables est basée sur des scores qui seront utilisés pour évaluer le degré d'importance d'une variable. Ces scores d'importance sont dérivés des SVM aux cas multiclassés et chaque score fournit une hiérarchie des variables explicatives.

Trois critères permettant d'établir un ordre d'importance sur les variables ont été introduits partiellement par Guyon et al. et Rakotomamonjy puis complétés par Ben Ishak et al. Une variable est d'autant plus importante que sa contribution au critère est forte. Nous utilisons ici les scores d'importance calculés par dérivation partielle (voir Ben Ishak 2007).

Dans cette étude, on va tester huit scores d'ordre un :

$\sum_{ovr} \delta W$, $\sum_{ovr} \delta RW$ et $\sum_{ovr} \delta Spb$ pour les scores obtenus en faisant la somme sur tous les sous problèmes relatifs à la décomposition une-contre-reste.

$\sum_{ovo} \delta W$, $\sum_{ovo} \delta RW$ et $\sum_{ovo} \delta Spb$ pour les scores obtenus en faisant la somme sur tous les sous problèmes relatifs à la décomposition une-contre-une.

$\sum_{wei} \delta W$ et $\sum_{wei} \delta RW$ pour les scores obtenus sur la base de la marge généralisée et ils ne sont calculés que pour le schéma de décomposition une contre-une. (Ben Ishak, 2007)

L'évaluation des modèles

Les modèles SVM multiclassés sont diverses et elles peuvent être réparties en deux catégories : directes et indirectes. Trois approches indirectes sont utilisées dans cette étude empirique : RADAG (Graphe orienté acyclique adaptatif réordonné), décodage basé sur la distance de Hamming et décodage basé sur la fonction de perte. L'extension des scores d'importance est obtenue en sommant sur tous les sous-problèmes binaires issus de la décomposition adoptée. Le taux d'erreur est estimé sur 50 partages aléatoires stratifiés.

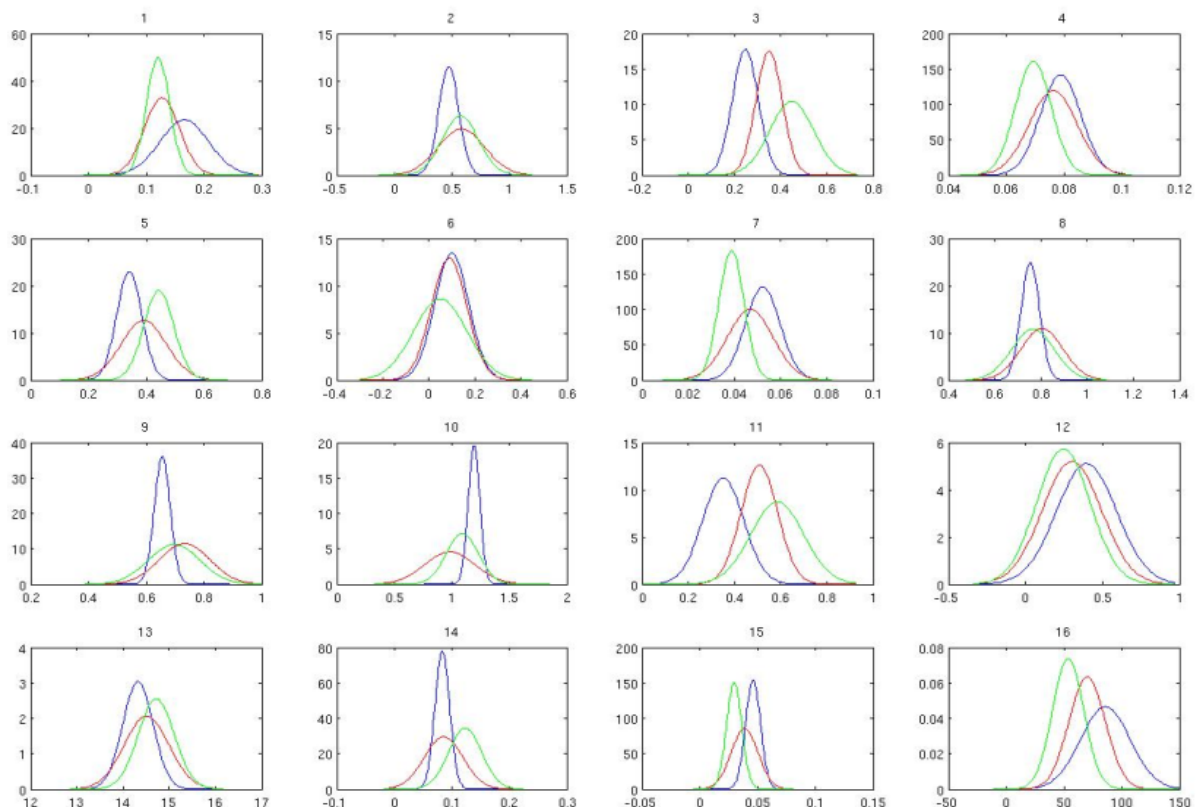
3. Analyse des résultats

On s'intéresse dans cette section à présenter les différents résultats des trois modèles étudiés, ainsi que leurs interprétations.

3.1 Le modèle Bayésien sous hypothèse gaussienne

En étudiant la statistique de Jarque Bera, seulement seize variables dont les distributions sont gaussiennes. Les paramètres sont estimés par maximum de vraisemblance. Les densités marginales pour chaque classe des 16 variables sont tracées dans les figures suivantes.

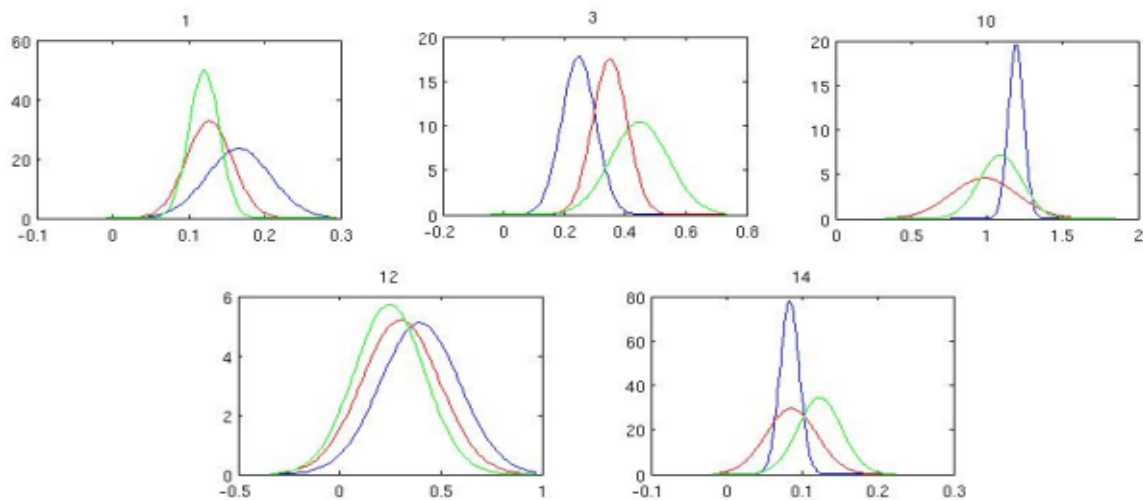
Figure 1 : Les densités marginales pour chacune des trois classes estimées par maximum de vraisemblance des seize variables.



D'après les figures, une première sélection consiste à éliminer les variables non discriminantes dont les densités marginales pour chaque classe sont trop attenantes. On a donc éliminé les variables suivantes : V2, V4, V6, V8, et V9. En recourant en deuxième étape à la matrice de corrélation, on a rejeté parmi les deux variables corrélées, la variable dont les densités marginales pour chaque classe sont plus adjacentes que l'autre.

Enfin, d'après les figures et la matrice de corrélation on a pu sélectionner les 5 variables suivantes : V1, V3, V10, V12 et V14.

Figure 2 : Les densités marginales pour chaque classe estimées par maximum de vraisemblance des cinq variables sélectionnées.



On peut donc construire le modèle Bayésien permettant de classifier les banques Tunisiennes. Les inputs du modèle sont donc les cinq variables sélectionnées représentant la taille, la capitalisation C, la gestion M, la rentabilité E et la liquidité L.

Après avoir testé les 50 modèles bayésiens sur 50 partages aléatoires, on a retenu le taux d'erreur moyen sur 50 partages aléatoires stratifiés qui est égal à 38.31%.

3.2 Le modèle neuronal

En introduisant comme inputs les 28 variables ($n = 28$), on a initialisé le modèle neuronal 100 fois. Le test de sensibilité a été établi pour les 100 modèles neuronaux permettant ainsi de construire 100 hiérarchisations de variables. Les différentes hiérarchies ont permis de choisir les 11 (m) meilleurs variables ($m < n$). Les 11 variables retenues sont R2, R3, R4, R5, R10, R13, R15, R16, R25, R26 et R27.

En deuxième lieu, on a réinitialisé 50 fois le modèle par les 11 variables sélectionnées et on a appliqué le test de sensibilité pour les 50 modèles neuronaux obtenus et on a choisi un meilleur paquet de 6 (p) variables discriminantes ($p < m$).

Enfin, on a réinitialisé 10 fois le modèle par les 6 variables sélectionnées et on a retenu le modèle le plus performant. Il s'agit d'un réseau à une seule couche caché dont les neurones sont au

nombre de 14. La réinitialisation du modèle 10 fois par l'ajout d'une variable et encore 10 fois en éliminant une variable n'améliore pas la performance du modèle.

On a donc retenu comme variables discriminantes : R2, R3, R4, R5, R15 et R25. Il s'agit de quatre variables de gestion M, d'une variable de rentabilité E et une variable de liquidité L. Ces résultats montrent que la qualité de gestion permet en grande partie de discriminer entre les différentes classes de risque des banques Tunisiennes.

Dans la phase de classification, on a introduit les 6 variables sélectionnées (les neurones d'entrées) dans le modèle du réseau de neurones. Il s'agit d'un réseau à une seule couche cachée. En modifiant l'architecture du réseau, on retient à chaque fois le taux d'erreur moyen et le taux d'apprentissage sur les 50 partages aléatoires stratifiés et on a abouti aux résultats suivants.

Tableau 1 : Les principaux résultats des modèles neuronaux à une seule couche cachée.

	Premier modèle	Deuxième modèle	Troisième modèle
Nombre des neurones cachés	14	10	5
Taux d'erreur moyen sur l'échantillon test	0.697895	0.692632	0.671579
Taux d'apprentissage moyen	0.937895	0.926316	0.866316

On remarque que le taux d'erreur moyen sur l'échantillon test est très élevé quelle que soit l'architecture du réseau. Ce taux est de plus en plus élevé que la structure du réseau est complexe. En conclusion, bien que le taux d'apprentissage moyen soit toujours acceptable, les réseaux de neurones ne sont pas capables de bien généraliser.

3.3 Le modèle SVM multiclassés

Dans cette partie, on a utilisé la méthode SVM multiclassés tout en combinant la sélection de variables et l'estimation du modèle en même tâche. Huit scores sont utilisés pour sélectionner et classer les plus importantes variables. En utilisant trois approches indirectes de SVM : RADAG (Graphe orienté acyclique adaptatif réordonné), Décodage basé sur la distance de Hamming et Décodage basé sur la fonction de perte, on a enfin abouti à 21 paquets optimaux de variables discriminantes. Pour choisir la meilleure hiérarchie de variables, on a retenu le paquet ayant le plus faible taux d'erreur moyen sur 50 partages aléatoires stratifiés.

Le paquet optimal qui a le plus faible taux d'erreur moyen a été sélectionné sur la base de deux scores différents. Il s'agit d'un taux d'erreur moyen égal à 23.053% pour un paquet optimal de 11 variables discriminantes. Les variables sélectionnées par ordre d'importance croissant sont : V26 (A), V13 (I), V20 (E), V22 (E), V21 (A), V3 (M), V19 (E), V15 (E), V25 (L), V5 (M) et V24 (E).

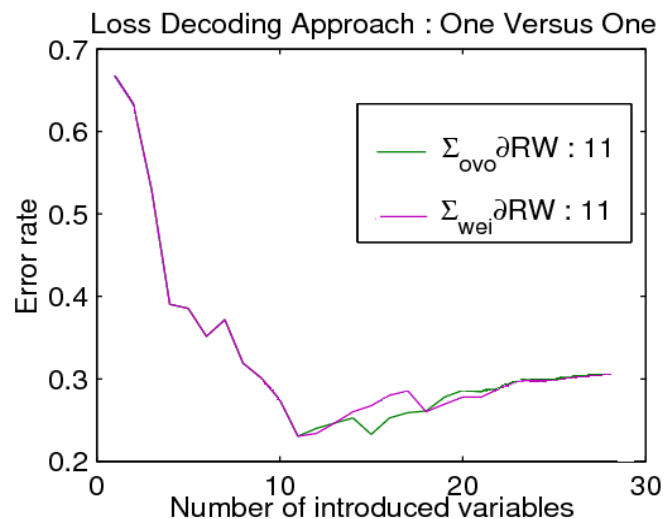
Les résultats montrent que la plus importante variable V26 appartienne à la catégorie qualité d'actifs. En fait, la détérioration de la qualité des prêts est le plus grave problème auquel les banques tunisiennes doivent faire face en assurant un montant minimum de provisions et en améliorant leur contrôle de gestion.

On remarque tout d'abord l'absence d'une variable qui s'inscrive à la catégorie « adéquation du capital » C. Ceci affirme que les banques Tunisiennes n'utilisent pas le capital pour ajuster leur niveau des créances douteuses.

D'autre part, on remarque la dominance des variables de rentabilités qui représentent la performance de l'ensemble des activités bancaires : V20 (E), V22 (E), V19 (E), V15 (E), V24 (E). Ceci montre que la mauvaise qualité du portefeuille sera reflétée dans les différents indicateurs de rentabilité.

Toutes ces variables peuvent être d'une utilité limitée s'elles sont aperçues d'une manière séparée. En effet, en introduisant, à chaque fois, au modèle une variable par importance, on voit que le taux d'erreur moyen diminue fortement et il atteint son minimum à la onzième variable.

Figure 3 : L'évolution du taux d'erreur en introduisant les variables séquentiellement par ordre d'importance.



4. Conclusion

La comparaison des trois modèles montre que les machines à vecteurs supports ont supplanté les réseaux de neurones et le modèle bayésien. En effet, le taux d'erreur moyen sur les mêmes 50 partages aléatoires stratifiés obtenu par le modèle SVM est plus faible que celui obtenu par le modèle « bayésien paramétrique » et par les « réseaux de neurones ».

Cette étude met en évidence l'efficacité des algorithmes de sélection de variables dans le contexte du risque bancaire. D'autre part, le modèle neuronal rend un taux d'erreur très élevé. Ce mauvais résultat semble être dû au problème de dimensionnalité. En effet, les 70 observations restent toujours un nombre réduit par rapport au 6 variables sélectionnées et l'application du modèle neuronale sur 50 partages aléatoires augmente la perturbation des résultats.

Ces résultats empiriques dévoilent la suprématie de la méthode SVM multiclassées. Ces machines ne permettent pas seulement de détecter les informations mutuelles entre les variables, mais aussi de considérer l'interaction existante entre la variable expliquée et les différents attributs. Il est donc crucial d'étudier et d'interpréter ces pondérations de point de vue analyse financière en utilisant les différentes approches des SVM multiclassées.

Références

- Altman, E.I. (1968), 'Financial Ratios, Discriminant analysis and the prediction of corporate bankruptcy', *Journal of finance*, **23**:589-609.
- Altman, E.I. (1977), 'Predicting performance in the savings and loan association industry', *Journal of monetary economics*, **3**:443-466.
- Altman, E.I., Marco, G., Varetto, F. (1994), 'Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks: the Italian experience', *Journal of banking and finance*, **18**:505-529.
- Avery, R., & Hanweck, G. (1984), 'A Dynamic Analysis of Bank Failures', Proceedings of the Conference on Bank Structure and Competition, *Federal Reserve Bank of Chicago*, 380-395.
- Barr, R.S., Seiford, L.M., & Siems, T.F. (1994), 'Forecasting Bank Failure: A Non-Parametric Frontier Estimation Approach', *Recherches Economiques de Louvain*, **60**:417-429.
- Barth, J., Brumbaugh, D., Sauerhaft, D., & Wang, G. (1985), 'Thrift Institutions Failures: Causes and Policy Issues', Proceedings of the Conference on Bank Structure and Competition, *Federal Reserve Bank of Chicago*, 184-216.
- Beaver, W.H (1966), 'Financial ratios as predictors of failure', *Journal of accounting research*, **4**:71-102.
- Ben Ishak A ET Ghattas B (2005), 'An efficient method for variable selection using svm-based criteria', Pré-publication de l'Institut de Mathématiques de Luminy, Marseille, France.
- Ben Ishak A. (2007), « Sélection de variables par les machines à vecteurs supports pour la discrimination binaire et multiclassée en grande dimension », Thèse de doctorat en cotutelle : "Informatique et mathématiques", Faculté des sciences de Luminy, "Gestion (option : modélisation) ", Institut supérieur de gestion de Tunis.
- Benston, G.J. (1985), 'An Analysis of the Causes of Savings and Loan Association Failures', *Monograph Series in Finance and Economics: New York University*.
- Demirgüç-Kunt A. (1989b), 'Modeling Large Commercial Bank Failures: A Simultaneous-Equation Analysis', Working Paper 8905, Federal Reserve Bank of Cleveland.
- failure predictions', *Management science*, **38**: 926-947.
- Gajewski G.R. (1988), 'Bank Risk, Regulator Behavior, and Bank Closure in the Mid- 1980s: A Two-Step Logit Model', Ph.D. Dissertation, the George Washington University.
- Godlewski C.J. (2003), 'Excès de risque et défaillance bancaire : une application des modèles de prévision de défaut aux banques des pays émergents', Mimeo, LARGE, Université Robert Schuman.
- Guyon I, Weston J, Barnhill S et Vapnik V.N. (2002), 'Gene selection for cancer classification using support vector machines', *Machine Learning*, 46(1-3), 389-422.
- Martin, D. (1977), 'Early Warning of Bank Failure: A Logit Regression Approach', *Journal of Banking and Finance*, **1**:249-276.
- Rakotomamonjy A. (2003), 'Variable selection using SVM-based criteria', *Journal of Machine Learning Research*, **3**, 1357-1370.
- Sinkey, J. (1975), 'A Multivariate Statistical Analysis of the Characteristics of Problem Banks', *Journal of Finance*, **30**:21-36.
- Tam, K.Y. (1991), 'Neural network models and the prediction of bank bankruptcy', *Omega*, **19**, 429-445.
- Tam, K.Y., Kiang, M.Y. (1992), 'Managerial application of neural networks: the case of bank.