BUSINESS FAILURE PREDICTION WITH SUPPORT VECTOR MACHINES AND NEURAL NETWORKS: A COMPARATIVE STUDY*

Jae H. Min
College of Business Administration, Sogang University, jaemin@ccs.sogang.ac.kr

Young-Chan Lee
College of Business Administration, Sogang University, chanlee@family.sogang.ac.kr

ABSTRACT

Bankruptcy prediction has attracted a lot of research interests in previous literature, and recent studies have shown that artificial neural networks (ANN) method achieved better performance than traditional statistical ones. ANN approaches have, however, suffered from difficulties with generalization, producing models that can overfit the data. This paper employs a relatively new machine learning technique, support vector machines (SVM), to the bankruptcy prediction problem in an attempt to provide a model with better explanatory power. To evaluate the prediction accuracy of SVM, we compare its performance with three-layer fully connected backpropagation neural networks (BNN). The experiment results show that SVM outperforms the BNN.

KEYWORDS: support vector machine, artificial neural networks, bankruptcy prediction

1. INTRODUCTION

Over the past decade, several financial crises were observed in some emerging markets enjoying a recent financial liberalization experience, and showed that debt financing built on capital inflow may result in large and sudden capital outflows, thereby causing a domestic "credit crunch". Experience with these recent crises forced banking authorities, i.e. the Bank of International Settlements (BIS), to draw a number of lessons. Among them, they all encourage commercial banks to develop internal models to better quantify financial risks (Basel Committee on Banking Supervision, 1999). Decision-making problems in credit evaluation are very important and difficult tasks for financial institutions due to the high level of risk from wrong decisions. Business failures, for example, affect shareholders, managers, lenders, suppliers, clients, financial community, government, competitors, and regulatory bodies. In order to effectively manage the credit risk exposure of financial institutions, there is a strong need for sophisticated decision support systems backed by analytical tools to measure, monitor, manage, and control financial and operational risks as well as inefficiencies (Park and Han, 2002; Emel et al., 2003).

A potential client's credit risk is often evaluated by financial institution's internal credit scoring models, which aim to determine whether an applicant has the capacity to repay by evaluating the risk of his loan application. Such models offer banks a means for evaluating the risk of their credit portfolio, in a timely manner, by centralizing global-exposures data and by analyzing marginal as well as absolute contributions to risk components. And quantitative risk-management systems provide the bank early warnings for predicting potential business failures (Emel et al., 2003; Lee et al., 2002; Lopez and Saidenberg, 2000; West, 2000).

So far, linear probability and multivariate conditional probability models, the recursive partitioning algorithm, artificial intelligence approaches, multicriteria decision making (MCDM), mathematical programming approaches have been proposed to support the credit decision (Bryant, 1997; Buta, 1994; Cielen and Vanhoof, 1999; Coakley and Brown, 2000; Davis et al., 1992; Dimitras et al., 1996, 1999; Emel et al., 2003; Falbo, 1991; Frydman et al., 1985; Martin, 1997; Reichert et al., 1983; Roy, 1991; Tam and Kiang, 1992). In particular, artificial neural networks (ANN) method is most frequently used in previous literature since the power of prediction is known to be better than the others. Due to the difficulty of interpreting the neural network models, however, most studies applying neural networks focused on prediction accuracy. Few efforts of using ANN models to provide better understanding of the bankruptcy prediction process have been reported in the literature (Huang et al., 2003). In addition, ANN suffers from difficulties with generalization because of overfitting, and fully depends on

* This work was supported by the Brain Korea 21 Project in 2004.

researchers' experience or knowledge for preprocessing of selecting a large number of control parameters that include relevant input variables, hidden layer size, learning rate, and momentum (Gemen et al., 1992; Lawrence et al., 1997; Moody, 1992; Sarle, 1995; Smith, 1993; Weigend, 1994).

This paper applies support vector machines (SVM) to bankruptcy prediction problem. The foundation of SVM has been developed by Vapnik (1998), and it is gaining popularity due to many attractive features, and excellent generalization performance on a wide range of problems. The formulation embodies the structural risk minimization (SRM) principle, which has been shown to be superior to traditional empirical risk minimization (ERM) principle employed by conventional neural networks (Gunn et al., 1997). SRM minimizes an upper bound of generalization error, as opposed to ERM that minimizes the error on training data. In addition, the solution of SVM may be global optimum while other neural network models tend to fall into a local optimal solution. Thus, overfitting is unlikely to occur with SVM (Cristianini and Shawe-Taylor, 2000; Gunn, 1998; Kim, 2003, Vapnik, 1998), and we expect to improve prediction accuracy by adopting this new algorithm.

This paper consists of five sections. Following Section 1, Section 2 provides the basic concept of SVM and its applications in financial decision problems with a short review of traditional bankruptcy prediction models. Section 3 describes research design and experiment steps, where we propose SVM and ANN models to the bankruptcy prediction. In Section 4, the empirical results of both methods are summarized and compared. Finally, Section 5 draws a conclusion of this study, and provides future research area.

2. LITERATURE REVIEW

2.1 Statistical Method

In the late 1960s, discriminant analysis (DA) was introduced to create a composite empirical indicator of financial ratios. Using financial ratios, Beaver (1966) developed an indicator that best differentiated between failed and non-failed firms using univariate analysis techniques. The univariate approach was later improved and extended to multivariate analysis by Altman (1968). During the years that followed, many researchers attempted to increase the success of MDA in predicting business failure (Dimitras et al., 1996). Among these are Eisenbeis (1978), Peel et al. (1986), and Falbo (1991). Linear probability and multivariate conditional probability models (logit and probit) were introduced to the business failure prediction literature in late 1970s. The contribution of these methods was in estimating the probability of a firm's failure. The linear probability model is a special case of ordinary least-squares regression with a dichotomous dependent variable (Dimitras et al., 1996). In the 1980s, the recursive partitioning algorithm (RPA) based on a binary classification tree rationale was applied to this business failure prediction problem by Frydman et al. (1985) and Srinivasan and Kim (1988).

2.2 Artificial Intelligence Method

From the late 1980s, artificial intelligence (AI) techniques, particularly rule-based expert systems, case-based reasoning systems and machine learning techniques such as artificial neural networks (ANN) have been successfully applied to bankruptcy prediction (Elmer and Borowski, 1988; Patuwo et al., 1993; Srinivasan and Ruparel, 1990; Srinivasan and Kim, 1988; Tam and Kiang, 1992). Recently, a hybrid approach that integrates ANN, statistical methods, and other machine learning techniques has been suggested in the literature as well as comparative studies between ANN and other techniques to improve the prediction accuracy (Curram and Mingers, 1994; Desai et al., 1996; Desai et al., 1997; Jensen, 1992; Lee et al., 1996; Lee et al., 1997; Lee et al., 1999; Lee et al., 2002; Malhotra and Malhotra, 2002; Markham and Ragsdale, 1995; Piramuthu, 1999; West, 2000; Zhang, 2000; Zhang et al., 1999). The results of comparative studies indicate that ANN shows better prediction accuracy (Barniv et al., 1997; Bell et al., 1990; Coates and Fant, 1993; Fanning and Cogger, 1994; Fletcher and Goss, 1993; Jo and Han, 1996; Lee et al., 1997; Odom and Sharda, 1990; Tam and Kiang, 1992; Wilson and Sharda, 1994; Zhang et al., 1999); however, ANN has a difficulty in explaining the causes of prediction result due to the lack of explanatory power and suffers from difficulties with generalization because of overfitting. In addition, it needs too much times and efforts to construct a best architecture (Lawrence et al., 1997; Sarle, 1995).

2.3 Support Vector Machine (SVM)

SVM uses linear model to implement nonlinear class boundaries through some nonlinear mapping the input vectors **x** into the high-dimensional feature space. A linear model constructed in the new space can represent a nonlinear decision boundary in the original space. In the new space, an optimal separating hyperplane is constructed. Thus, SVM is known as the algorithm that finds a special kind of linear model, the *maximum margin hyperplane*. The maximum margin hyperplane gives the maximum separation between the decision classes. The training examples that are closest to the maximum margin

hyperplane are called *support vectors*. All other training examples are irrelevant for defining the binary class boundaries (Cristianini and Shawe-Taylor, 2000; Gunn, 1998; Hearst et al., 1998, Kim, 2003).
SVM is simple enough to be analyzed mathematically since it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space. In this sense, support vector machines may serve as a good candidate for combining the strengths of conventional statistical methods that are more theory-driven and easy to be analyzed, and more data-driven, distribution free and robust machine learning methods. The SVM approach has been introduced in several financial applications recently, mainly in the area of time series prediction and classification (Huang et al., 2003; Kim, 2003; Tay and Cao, 2001; Viaene et al., 2002). A recent study (Fan and Palaniswami, 2000) closely related to this work investigated the use of SVM to select bankruptcy predictors, and reported that SVM was comparable to and even outperformed other classifiers (including ANN and linear discriminant classifier) in terms of generalization performance. In this study, we are interested in evaluating the performance of SVM in the domain of bankruptcy prediction in comparison with that of three-layer fully connected backpropagation neural networks. A simple description of the SVM algorithm is provided as follows. For more details, please refer to Cristianini and Shawe-Taylor (2000) and Gunn (1998).
Given a training set $D = \{x_i, t_i\}_{i=1}^{N}$ with input vectors, $x_i = (x_i^{(1)}, \ldots, x_i^{(n)})^T \in \mathrm{R}^n$ and target labels $t_i \in \{-1, +1\}$, the support vector machine (SVM) classifier, according to Vapnik's original formulation, satisfies the following conditions:

$$\begin{cases} \omega^T \varphi(x_i) + b \geq +1, & \text{if } t_i = +1 \\ \omega^T \varphi(x_i) + b \leq -1, & \text{if } t_i = -1 \end{cases} \tag{1}$$

Which is equivalent to:

$$t_i \left[\omega^T \varphi(x_i) + b \right] \geq 1, \quad i = 1, \ldots, N \tag{2}$$

Where ω represents the weight vector and b the bias. The nonlinear function $\varphi(\cdot) : \mathrm{R}^n \to \mathrm{R}^{n_k}$ maps the input or measurement space to a high-dimensional, and possibly infinite-dimensional, feature space. Equation (2) then comes down to the construction of two parallel bounding hyperplanes at opposite sides of a separating hyperplane $\omega^T \varphi(x) + b = 0$ in the feature space, with the margin width between both hyperplanes equal to $\frac{2}{\|\omega\|^2}$. In primal weight space, the classifier then takes the form:

$$y^t(x) = \mathrm{sgn}(\omega^T \varphi(x) + b) \tag{3}$$

But, on the other hand, it is never evaluated in this form.
One defines the optimization problem as:

$$\underset{\omega, b, \xi}{\mathrm{Min}}\, J(\omega, \xi) = \frac{1}{2} \omega^T \omega + c \sum_{i=1}^{N} \xi_i \tag{4}$$

subject to:

$$\begin{cases} t_i (\omega^T \varphi(x_i) + b) \geq 1 - \xi_i, & i = 1, \ldots, N \\ \xi_i \geq 0, & i = 1, \ldots, N \end{cases} \tag{5}$$

The variables ξ_i are slacks needed to allow misclassifications in the set of inequalities. $c \in \mathrm{R}^+$ is a tuning hyperparameter, weighting the importance of classification errors vis-à-vis the margin width. The solution of the optimization problem is obtained after constructing the Lagrangian. From the conditions of optimality, one obtains a quadratic programming (QP) problem in the Lagrange multipliers α_i. A multiplier α_i exists for each training data instance. Data instances corresponding to non-zero α_i are

called support vectors.
As is typical for SVMs, we never calculate ω or $\varphi(x)$. This is made possible due to Mercer's condition, which relates the mapping function $\varphi(x)$ to a kernel function $K(\cdot,\cdot)$ as follows.

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \tag{6}$$

For the kernel function $K(\cdot,\cdot)$, one typically has several design choices, such as $K(x_i, x_j) = x_i^T x_j$ (linear kernel), $K(x_i, x_j) = (x_i^T x_j + 1)^d$ (polynomial kernel of degree d), and $K(x_i, x_j) = \exp\left\{-\gamma \| x_i - x_j \|^2\right\}$ (radial basis function (RBF) kernel), where $d \in \mathrm{N}$ and $\sigma \in \mathrm{R}^+$ are constants. Then construct the SVM classifier as:

$$y^t(x) = \mathrm{sgn}\left(\sum_i^N \alpha_i t_i K(x, x_i) + b) \right) \tag{7}$$

Optimization is discussed in Cristianini and Shawe-Taylor (2000), Gunn (1998), and Vapnik (1995, 1998).

3. RESEARCH DESIGN

3.1 Data collection and preprocessing

Bankruptcy prediction is an ill-structured decision problem, which involves the analysis of a complex array of a firm's historical data (Park and Han, 2002). The database used in this study was obtained from the Korea's largest credit guarantee organization that has the public policy role of promoting growth among small and medium-sized enterprises in the country. All of the failure cases are from medium-sized firms (total assets range is from 1 to 7 billion dollars) in IT industry, which went bankrupt between 1999 and 2002. There are also a small number of bankruptcy cases that can be compared to non-bankruptcy cases. Therefore, it is possible to select a sound case, which is included in the same industry and of similar size as that of a bankruptcy case. For the purposes of this study, the experimental sets consisted of the same number of bankruptcy and non-bankruptcy cases. The total sample of 516 companies includes 258 bankruptcy and 258 non-bankruptcy cases placed in random order.
The original data are scaled into the range of [-1, 1]. The goal of linear scaling is to independently normalize each feature component to the specified range. It ensures the larger value input attributes do not overwhelm smaller value inputs; hence helps to reduce prediction errors (Sarle, 1995)
In choosing financial ratios, we apply statistical methods. Most studies that have been performed by using statistical methods such as discriminant analysis and logistic regression, have selected the independent variables employing stepwise approach, where financial ratios are initially selected for the evaluation model by factor analysis and t-test. We also reduce the number of financial variables into a manageable set of 10 using two different variable selection methods: stepwise and the t-test.
In case of SVM, each data set is split into two subsets: a training set of 80% and a test set of 20% of the total data, respectively. The holdout data is used to test results with the data that is not utilized to develop the model. In case of ANN, each data set is split into three subsets: a training set, a test set and a validation set of 60%, 20%, and 20% of the data, respectively. The validation data are used to test the results with the data that were not used to develop the system.

3.2 Analysis steps

This study is conducted according to the analysis steps in Figure 1.
In Step 1, we reduce the number of multi-dimensional financial ratios to two factors through the principle component analysis (PCA).[2] In Step 2, we calculate factor scores of all companies using factor loadings equal to or greater than 0.5 and compare the training performance between ANN and SVM graphically using the factor scores. In Step 3, we choose final financial ratios used in the bankruptcy prediction model via stepwise logistic regression and discriminant analysis. Lastly in Step 4, we compare the prediction accuracy between ANN and SVM using the final financial ratios and conduct a sensitivity

[1] We use statistically significant financial ratios via t-test for the principle component analysis.

analysis according to parameter adjustment.

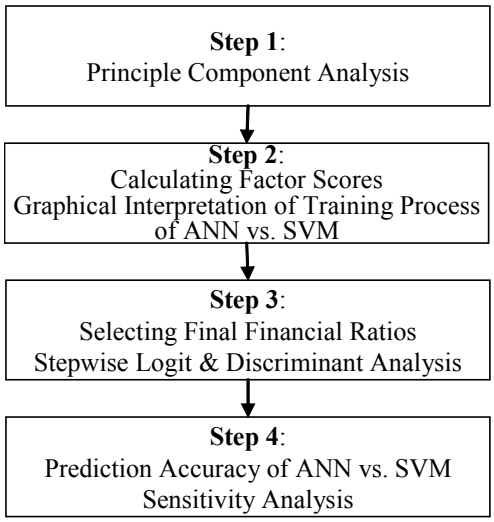


Figure 1. Analysis steps

3.3 Support Vector Machine

In this study, the Gaussian radial basis function (RBF) is used as the kernel function of SVM. There are two parameters associated with RBF kernels: C and γ. The upper bound C and the kernel parameter γ play an important role in the performance of SVMs (Tay and Cao, 2001; Hsu et al., 2004). Therefore, improper selection of these two parameters can cause the overfitting or the underfitting problems. Nevertheless, there is little general guidance to determine the parameters of SVM. Recently, Hsu et al. (2004) suggested a practical guideline to SVM using cross-validation and grid search, and this study will utilize it.

The goal is to identify good (C, γ) so that the classifier can accurately predict unknown data (i.e., testing data). Note that it may not be useful to achieve high training accuracy (i.e., classifiers accurately predict training data whose class labels are indeed known). Therefore, a common way is to separate training data into two parts, of which one is considered unknown in training the classifier. Then the prediction accuracy on this set can more precisely reflect the performance on classifying unknown data. An improved version of this procedure is cross-validation.

In v-fold cross-validation, we first divide the training set into v subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining v-1 subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data that are correctly classified.

The cross-validation procedure can prevent the overfitting problem. This study uses a "grid-search" on C and γ using cross-validation. Basically, pairs of (C, γ) are tried and the one with the best cross-validation accuracy is selected. We found that trying exponentially growing sequences of C and γ is a practical method to identify good parameters (for example, $C = 2^{-5}, 2^{-3}, \ldots, 2^{15}$, $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$).

The reasons why we use the grid-search is as follows. One is that psychologically we may not feel safe to use methods that avoid doing an exhaustive parameter search by approximations or heuristics. The other reason is that the computational time to find good parameters by grid-search is not much more than that by advanced methods since there are only two parameters. Furthermore, the grid-search can be easily parallelized because each (C, γ) is independent (Hsu et al., 2004).

This study uses *LIBSVM software system* (Chang and Lin, 2004) to perform SVM experiments.

3.4. Artificial Neural Networks

In this study, a three-layer back-propagation (BP) network is used as benchmarks. This study varies the number of nodes in the hidden layer and stopping criteria for training. In particular, 8, 12, 16, 24, 32 hidden nodes for each stopping criteria because the BP network does not have a general rule for determining the optimal number of hidden nodes (Kim, 2003). For the stopping criteria of BP, this study allows 50, 100, 200, 300 learning epochs per one training example since there is little general knowledge for selecting the number of epochs. The learning rate is 0.1 and the momentum term is 0.7. The hidden nodes use the hyperbolic tangent transfer function and the output node uses the same transfer

function. This study allows 10 input nodes because 10 input variables are employed.
This study uses *NeuroSolutions 4.32* to perform the ANN experiments.

4. EMPIRICAL ANALYSIS

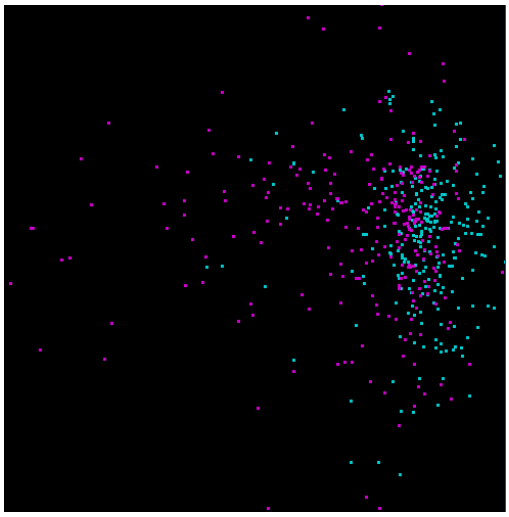
4.1 Principle Component Analysis

This study conducted the principle component analysis (Varimax rotation) to 50 financial ratios with significant t-value. The list of financial ratios and the factor loadings (≥ 0.5) is summarized in Table 1.

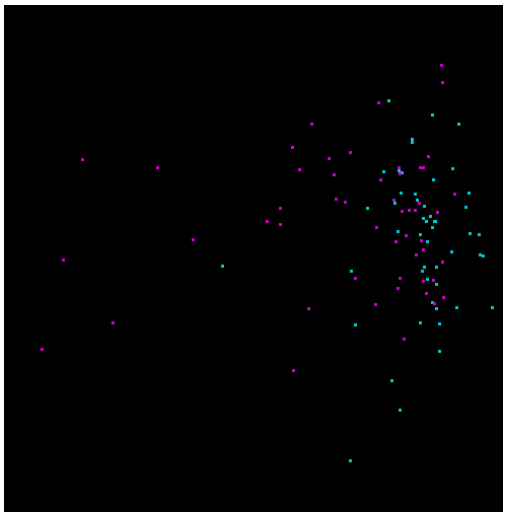
Table 1. The result of Principle Component Analysis

Variables	Factor1	Factor2
Growth rate of total assets		
Growth rate of current assets		
Ordinary income to total assets	0.868	
Net income to total assets	0.860	
Ordinary income to stockholders' equity		
Net income to stockholders' equity	0.557	
Ordinary income to sales	0.905	
Net income to sales	0.906	
Operating income to sales	0.893	
R&D costs to sales		
EBIT to sales	0.896	
EBITDA to sales	0.887	
Depreciation ratio		
Interest expenses to total expenses		
Interest expenses to sales	-0.799	
Net interest expenses to sales	-0.781	
Interest coverage ratio	0.700	
Stockholders' equity to total assets		0.510
Current ratio		0.787
Quick ratio		0.770
Cash flow to short term loan		
Cash flow to total loan		
Cash flow to total debt		
Cash flow to interest expenses		
Fixed assets to stockholders' equity and long-term liabilities		-0.605
Debt ratio		
Current liabilities ratio		
Total borrowings and bonds payable to total assets		
Total borrowings and bonds payable to sales	-0.763	
Net working capital to total assets		0.735
Total assets turnover	0.603	
Capital stock turnover		
Operating assets turnover	0.550	
Receivables turnover		
Gross value added to total assets and productivity of capital	0.758	
Gross value added to property, plant, and equipment	0.540	
Gross value added to sales	0.827	
Solvency ratio		
Ordinary income to ordinary expenses	0.681	
Earnings to interest ratio	0.808	

The scatter plots of two factor scores calculated by using the factor loadings in Table 1 is represented in Figure 2. In Figure 2, two different colors denote two classes of the training and the holdout examples. Green and purple bullets represent the two classes.



(a) The scatter plot of training set



(b) The scatter plot of test set

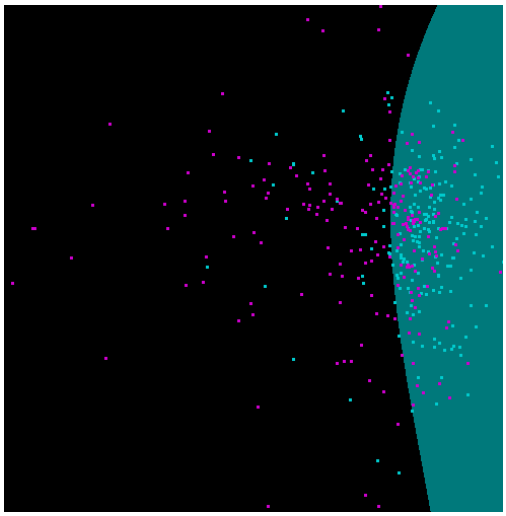
Figure 2. The scatter plot of factor scores: (a) training set and (b) test set

This study compared the training performance between three-layer BP ANN and SVM using Gaussian RBF kernel function to solve the problem whose split boundary is complex like Figure 2.

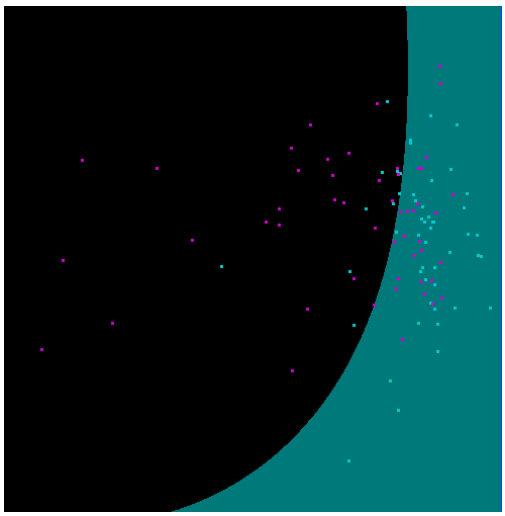
4.2 Graphical interpretation of the result of SVM and ANN

The following figures show a comparison of the SVM using Gaussian RBF kernel function and the three-layer BP ANN in attempting to classify two sets of factor score inputs.

First, the result of implementing SVM according to training data and test data is represented in Figure 3, where target variable is bankruptcy or non-bankruptcy (0 or 1), and input variables are two factor scores.



(a) After SVM is implemented (training data)

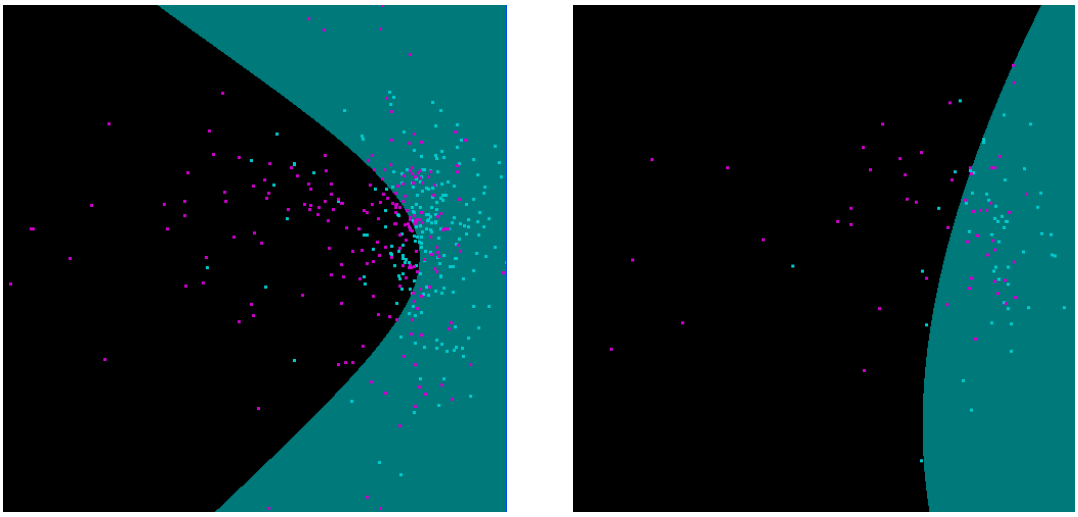


(b) After SVM is implemented (test data)

Figure 3. Graphical interpretation of the results of SVM: (a) training data and (b) test data

SVM trains the data corresponding to each Gaussian in order to determine the support vectors. These support vectors allow the network to rapidly converge on the data boundaries and consequently classify the inputs. As you see in Figure 3, SVM (with just default parameter options) classifies the two groups appropriately, where the prediction accuracy is 67.48% (training) and 66.35% (test data) respectively.

Second, the training result using the ANN is represented in Figure 4. Like SVM, target variable is bankruptcy or non-bankruptcy (0 or 1), and input variables are two factor scores.



(a) After ANN is implemented (training data) (b) After ANN is implemented (test data)

Figure 4. Graphical interpretation of the results of ANN: (a) training data and (b) test data

As you see in Fig. 4, three-layer BP ANN (with 100 epochs, 1 hidden-layer, and 10 hidden-node[3]) tries to solve the problem and its prediction accuracy is similar to that of SVM (training data: 66.51%, test data: 66.35%).

4.3 Selecting final financial ratios

In section 4.2, we interpreted the training process of ANN and SVM using the principle component of 50 financial ratios graphically and could find out the performance of SVM is attractive. Based on these results, we conduct a comparative analysis of the prediction accuracy between ANN and SVM. To serve this purpose, we select the final financial ratios used in the analysis through stepwise logistic regression and discriminant analysis. The final financial ratios are summarized in Table 2.

Table 2. Final financial ratios

Variables	Formula
Growth rate of total assets	(Total assets_t – total assets_{t-1}) ÷ total assets_t
R&D costs to sales	R&D costs ÷ sales
Interest expenses to total expenses	Interest expenses ÷ total expenses
Stockholders' equity to total assets	Stockholders' equity ÷ total assets
Cash flow to total debt	Cash flow after operating activity ÷ total debt
Total assets turnover	2* Sales ÷ (total assets_t + total assets_{t-1})
Operating assets turnover	Sales ÷ operating assets
Receivables turnover	2* Sales ÷ (receivables_t + receivables_{t-1})
Gross value added to total assets and productivity of capital	200 * Value added ÷ (total assets_t + total assets_{t-1})
Ordinary income to ordinary expenses	Ordinary income ÷ ordinary expenses

The performance of bankruptcy prediction fully depends on the input variables. In the literature, there are many cases that the economic interpretation of them is very difficult but they are significant to classify business units into bankruptcy or non-bankruptcy. We select the final financial ratios used in the corporate business analysis of the Bank of Korea. Therefore, the financial ratios selected in this study are both economically interpretable and useful to figure out the financial credibility.

[2] There is no universally accepted guideline for the number of hidden-layers, hidden-nodes, and epochs in ANN architecture. This is the default setting of ANN in general, where the number of input variables is 10 and the problem is a bankruptcy prediction.

4.4 Prediction accuracy of SVM and ANN

SVM

In SVM, each data set is split into two subsets: a training set of 80% and a test set of 20% of the total data (516), respectively.

One of the advantages of linear SVM is that there are no parameters to tune except for the constant C. But the upper bound C on the coefficient α_i affects prediction performance for the cases where the training data is not separable by a linear SVM (Drucker and Vapnik, 1999). For the nonlinear SVM, there is an additional parameter, the kernel parameter, to tune. There are two kernel functions including the Gaussian radial basis function and the polynomial function for the nonlinear SVM. The polynomial function, however, takes a longer time in the training of SVM, and provides worse results than the Gaussian radial basis function in preliminary test. Thus, this study uses the Gaussian radial basis function as the kernel function of SVMs (Kim, 2003).

There are two parameters associated with RBF kernels: C and γ. It is not known beforehand which C and γ are the best for one problem; consequently, some kind of model selection (parameter search) must be done (Hsu et al., 2004). We conduct a "grid-search" to find a best C and γ using cross-validation. Pairs of (C, γ) are tried and the one with the best cross-validation accuracy is picked. We first used a coarse grid (see Figure 5) and found that the best (C, γ) was (2^8, 2^{-10}) with the cross-validation rate of 81.5534%. Next we conducted a finer grid search on the neighborhood of (2^8, 2^{-10}) and obtained a better cross-validation rate of 81.7961% at (2^9, $2^{-6.94221}$). After the best (C, γ) was found, the whole training set was trained again to generate the final classifier.

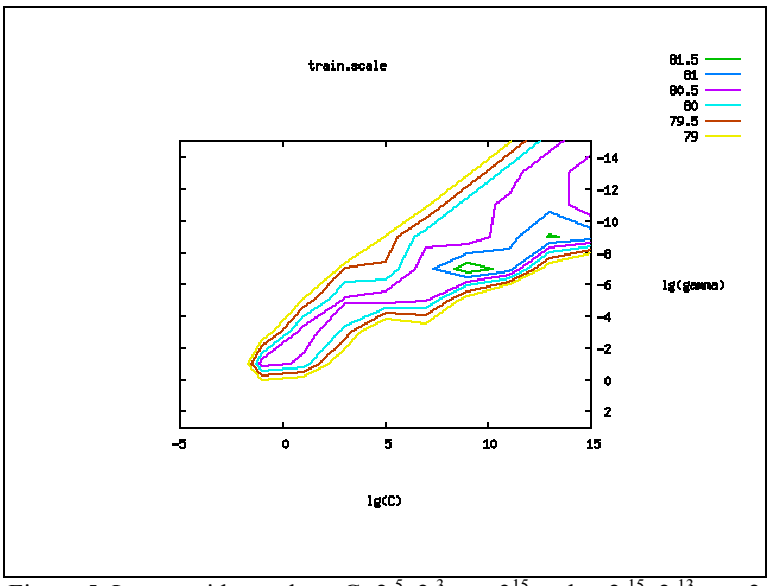


Figure 5. Loose grid search on C=2^{-5}, 2^{-3}, …, 2^{15} and γ=2^{-15}, 2^{-13}, …, 2

Finally, we used the best C and γ for training data and testing the prediction accuracy. The overall prediction accuracy of SVM using the grid-search turned out to be 76.92% (bankruptcy: 78.46%, non-bankruptcy: 75.00%).

Three-layer BP ANN

In ANN, each data set is split into three subsets: a training set, a test set and a validation set of 60%, 20%, and 20% of the data, respectively. Table 3 shows the results of three-layer BP ANN. As the training epoch increases, the prediction accuracy of training set becomes higher. The best prediction accuracy of the training set was found when epoch was 300 and the number of hidden-nodes was 24, and the prediction accuracy of the test data set turned out to be 75.96% (bankruptcy: 78.85%, non-bankruptcy: 73.08%). The results of prediction accuracy according to parameter adjustment are summarized in Table 3.

Table 3. The results of ANN

Learning epoch	Hidden nodes	Prediction accuracy (%)	
		Training data	Test data
50	8	83.77	75.00
	12	84.09	72.12
	16	84.74	73.08
	24	85.39	75.00
	32	85.39	74.04
100	8	85.39	74.04
	12	86.36	74.04
	16	87.99	73.08
	24	87.01	74.04
	32	87.66	74.04
200	8	87.01	74.04
	12	88.64	74.04
	16	88.96	74.04
	24	88.96	**75.96**
	32	90.26	73.08
300	8	89.29	73.08
	12	91.23	75.00
	16	92.21	74.04
	24	**92.21**	75.00
	32	91.88	73.08

As you see in Table 3, the best prediction accuracy of ANN (75.96%) to the test data is similar to that of SVM (76.92%). ANN, however, shows a serious overfitting problem when learning epoch is 300. As the learning epoch increases, the prediction accuracy of training data improves while the prediction accuracy of test data rather decreases.
From the results of the empirical experiment, we can conclude that SVM shows better performance in bankruptcy prediction problem avoiding overfitting and exhaustive parameter search.

5. CONCLUSION

This study employed SVM to predict bankruptcy of the medium-sized IT industry companies. In this study, the effects of the values of the upper bound C and the kernel parameter γ in SVM were investigated. Because the prediction performance of SVM is sensitive to the values of these parameters, we used "grid-search" to find the optimal values of C and γ using the training data.
In addition, this study compared SVM with standard three-layer BP ANN in terms of prediction accuracy. The results of empirical analysis showed that SVM outperformed BP ANN. The results may be attributable to the fact that SVM implements the structural risk minimization principle and this leads to better generalization than conventional techniques. With these results of this study, we claim that SVM can serve as a promising alternative for the bankruptcy prediction.
Although a grid-search in this study reduce the effort to find the best C and γ, the prediction performance may be different or improved if the other values of parameters of SVM are selected; and this remains a very interesting topic for further study. Also, the generalization of SVM should be tested in various industries and companies in the future.

REFERENCES

Altman, E.I., "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, Vol. 23, No. 4, 1968, pp. 589-609.

Altman, E.I., G. Marco, and F. Varetto (1994), "Corporate distress diagonis comparisons using linear discriminant analysis and neural networks," *Journal of Banking and Finance*, Vol. 18, No. 3, pp. 505-529.

Barniv, R., A. Agarwal, and R. Leach, "Predicting the outcome following bankruptcy filing: a three-state classification using neural networks," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 6, No. 3, 1997, pp. 177-194.

Basel Committee on Banking Supervision, *Credit Risk Modeling: Current Practices and Applications*, Basel Committee Publications, 1999.

Beaver, W., "Financial ratios as predictors of failure," *Journal of Accounting Research*, Vol. 5, 1966, pp. 71-111.

Bell, T., G. Ribar, and J. Verchio, "Neural nets vs. logistic regression: a comparison of each model's ability to predict commercial bank failures," *Proceedings of the 1990 Deloitte & Touche/University of Kansas Symposium on Auditing Problems*, 1990, pp. 29-58.

Bryant, S.M., "A Case-based Reasoning Approach to Bankruptcy Prediction Modeling," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 6, No. 3, 1997, pp. 195-214.

Buta, P., "Mining for Financial Knowledge with CBR," *AI Expert*, Vol. 9, No. 2, 1994, pp. 34-41.

Chang, C.-C. and C.-J. Lin, "LIBSVM: a library for support vector machines," Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, 2001, Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Charnes, A., W.W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, Vol. 2, 1978, pp. 429-444.

Chen, M-C. and S.-H. Huang, "Credit Scoring and Rejected Instances Reassigning through Evolutionary Computation Techniques," *Expert Systems with Applications*, Vol. 24, 2003, pp. 433-441.

Cielen, A. and K. Vanhoof , *Bankruptcy prediction using a data envelopment analysis*, Manuscript, Limburg University, Diebenpeek, 1999.

Coakley, J.R. and C.E. Brown, "Artificial Neural Networks in Accounting and Finance: Modeling Issues," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 9, No. 2, 2000, pp. 119-144.

Coates, P., and L. Fant, "Recognizing financial distress patterns using a neural network tool," Financial Management, Vol. 22, No.3, 1993, pp. 142-155.

Cristianini, N., and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge, England: Cambridge University Press, 2000.

Curram, S.P. and J. Mingers, "Neural Networks, Decision Tree Induction and Discriminant Analysis: An Empirical Comparison," *Journal of Operational Research Society*, Vol. 45, No. 4, 1994, pp. 440-450.

Davis, R.H., D.B. Edelman, and A.J. Gammerman, "Machine Learning Algorithms for Credit-Card Applications," *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 4, pp. 43-51, 1992.

Desai, V.S., J.N. Conway, and G.A. Overstreet Jr., "Credit Scoring Models in the Credit Union Environment Using Neural Networks and Genetic Algorithms," *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 8, 1997, pp. 324-346.

Desai, V.S., J.N. Crook, and G.A. Overstreet Jr., "A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment," *European Journal of Operational Research*, Vol. 95, 1996, pp. 24-37.

Dimitras, A.I., S.H. Zanakis, and C. Zopounidis, "A Survey of Business Failure with an Emphasis on Prediction Methods and Industrial Applications," *European Journal of Operational Research*, Vol. 90, No. 3, 1996, pp. 487-513.

Drucker, H., D. Wu, and V.N. Vapnik, "Support vector machines for spam categorization," *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, 1999, pp. 1048-1054.

Eisenbeis, R.A., "Problems in applying discriminant analysis in credit scoring models," *Journal of Banking and Finance*, Vol. 2, pp. 1978, pp. 205-219.

Elmer, P.J. and D.M. Borowski, "An expert system approach to financial analysis: the case of S&L bankruptcy," *Financial Management Autumn*, Vol. 17, No. 3, 1988, pp. 66-76.

Emel, A.B., M. Oral, A. Reisman, and R. Yolalan, "A Credit Scoring Approach for the Commercial Banking Sector," *Socio-Economic Planning Sciences*, Vol. 37, 2003, pp. 103-123.

Falbo, P., "Credit-scoring by enlarged discriminant models," *Omega*, Vol. 19, No. 4, 1991, pp. 275-289.

Fan, A., and M. Palaniswami, "Selecting bankruptcy predictors using a support vector machine approach," *Proceedings of the International Joint Conference on Neural Networks*, 2000.

Fanning, K., and K. Cogger, "A comparative analysis of artificial neural networks using financial distress prediction," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 3, No. 3, 1994, pp. 241-252.

Fletcher, D., and E. Goss, "Forecasting with neural networks and application using bankruptcy data," *Information and Management*, Vol. 24, 1993, pp. 159-167.

Frydman H.E., E.I. Altman, and D. Kao, "Introducing Recursive Partitioning for Financial Classification: the case of Financial Distress," *Journal of Finance*, Vol. 40, No. 1, 1985, pp. 269-291.

Geman, S., E. Bienenstock, and R. Doursat, "Neural Networks and the Bias/Variance Dilemma," *Neural Computation*, Vol. 4, 1992, pp. 1-58.

Gestel, T. V., J.A.K. Suykens, D.-E. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor, and J. Vandewalle, "Financial time series prediction using least squares support vector machines within the evidence framework," *IEEE Transactions on Neural Networks*, Vol. 12, No. 4, 2001, pp. 809– 821.

Gunn, S.R., *Support Vector Machines for Classification and Regression*, Technical Report, University of Southampton, 1998.

Gunn, S.R., M. Brown and K.M. Bossley, "Network Performance Assessment for Neurofuzzy Data Modelling," Lecture Notes in Computer Science, Vol. 1280, 1997, pp. 313-323.

Hand, D.J., *Discrimination and Classification*, New York, NY: Wiley, 1981.

Hearst, M.A., S.T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent System*, Vol. 13, No. 4, 1998, pp. 18-28.

Hornik, K., "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, Vol. 4, 1991, pp. 251-257.

Hsu, C.-W., C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Huang, Z., H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, "Credit rating analysis with support vector machine and neural networks: a market comparative study," *Decision Support Systems*, Vol. 37, 2004, pp. 543-558.

Jensen, H.L., "Using Neural Networks for Credit Scoring," *Managerial Finance*, Vol. 18, 1992, pp. 15-26.

Jo, H., and I. Han, "Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction," *Expert Systems with Applications*, Vol. 11, 1996, pp. 415-422.

Joachims, T., "Text categorization with support vector machines," *Proceedings of the European Conference on Machine Learning (ECML), 10th European Conference on Machine Learning*, 1998, pp. 137-142.

Kim, K.J., "Financial time series forecasting using support vector machines," *Neurocomputing*, Vol. 55, 2003, pp. 307-319.

Lawrence, S., C.L. Giles, and A.-C. Tsoi, "Lessons in Neural Network Training: Overfitting May be harder than Expected," *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, AAAl-97, AAAl Press, Mento Park, California, 1997, pp.540-545.

Lee, G., T.K. Sung, and N. Chang, "Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction," *Journal of Management Information Systems*, Vol. 16, 1999, pp. 63-85.

Lee, H., H. Jo, and I. Han, "Bankruptcy Prediction Using Case-based Reasoning, Neural Networks, and Discriminant Analysis," *Expert Systems With Applications*, Vol. 13, 1997, pp. 97-108.

Lee, K., I. Han, and Y. Kwon, "Hybrid neural networks for bankruptcy predictions," *Decision Support Systems*, Vol. 18, 1996, pp. 63-72.

Lee, T.-S., C.-C. Chiu, C.-J. Lu, and I.-F. Chen, "Credit Scoring Using Hybrid Neural Discriminant Technique," *Expert Systems with Applications*, Vol. 23, 2002, pp. 245-254.

Lopez, J.A. and M.R. Saidenberg, "Evaluating credit risk models," *Journal of Banking and Finance*, Vol. 24, No. 1-2, 2000, pp. 151-165.

Malhotra, R. and D.K. Malhotra, "Differentiating Between Good Credits and Bad Credits Using Neuro-fuzzy Systems," *European Journal of Operational Research*, Vol. 136, No. 2, 2002, pp. 190-211.

Markham, I.S. and C.T. Ragsdale, "Combining Neural Networks and Statistical Predictions to Solve the Classification Problem in Discriminant Analysis," *Decision Sciences*, Vol. 26, No. 2, 1995, pp. 229-242.

Martin, D., "Early Warning of Bank Failure: A Logit Regression Approach," *Journal of Banking and Finance*, Vol. 1, 1997, pp. 249-276.

Moody, J.E., "The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems," *NIPS*, Vol. 4, 1992, pp. 847-854.

Odom, M., and R. Sharda, "A neural network model for bankruptcy prediction," *Proceedings of the International Joint Conference on Neural Networks*, 1990, pp. II-163-II-168.

Ohlson, J.A., "Financial ratios and probabilistic prediction of bankruptcy," *Journal of Accounting Research*, Vol. 18, No.1, 1980, pp. 109-131.

Osuna, E., R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," *Proceedings of Computer Vision and Pattern Recognition*, 1997, pp. 130–136.

Park, C.-S. and I. Han, "A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction," *Expert Systems with Applications*, Vol. 23, No. 1, 2002, pp. 255-264.

Patuwo, E., M.H. Hu, and M.S. Hung, "Two-group classification using neural networks," *Decisions Science*, Vol. 24, No. 4, 1993, pp. 825-845.

Peel, M.J., D.A. Peel, and P.F. Pope, "Predicting corporate failure-some results for the UK corporate sector," *Omega*, Vol. 14, No. 1, 1986, pp. 5-12.

Piramuthu, S., "Financial Credit-risk Evaluation with Neural and Neurofuzzy Systems," *European Journal of Operational Research*, Vol. 112, 1999, pp. 310-321.

Reichert, A.K., C.C. Cho, and G.M. Wagner, "An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models," *Journal of Business and Economic Statistics*, Vol. 1, 1983, pp. 101-114.

Roy, B., "The outranking approach and the foundations of ELECTRE methods," *Theory and Decision*, Vol. 31, 1991, pp. 49-73.

Sarle, W.S., "Stopped Training and Other Remedies for Overfitting," *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, 1995, pp.352-360.

Smith, M., *Neural Networks for Statistical Modeling*, NY: Van Nostrand Reinhold, 1993.

Srinivasan, V. and B. Ruparel, "CGX: an expert support system for credit granting," *European Journal of Operational Research*, Vol. 45, 1990, pp. 293-308.

Srinivasan, V. and Y.H. Kim, "Designing expert financial systems: a case study of corporate credit management," *Financial Management*, Vol. 5, 1988, pp. 32-43.

Tam, K.Y and M.Y. Kiang, "Managerial Applications of Neural Networks: the Case of Bank Failure Predictions," *Management Science*, Vol. 38, No. 7, 1992, pp. 926-947.

Tay, F.E.H., and L. Cao (2001), "Application of support vector machines in financial time series forecasting," *Omega*, Vol. 29, pp. 309-317.

Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

Vapnik, V., *Statistical Learning Theory*, Springer, New York, 1998.

Vapnik, V., S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, 1997, pp. 281–287, Cambridge, MA, MIT Press.

Viaene, S., R.A. Derrig, B. Baesens, and G. Dedene, "A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection," *The Journal of Risk and Insurance*, Vol. 69, No. 3, 2002, pp. 373-421.

Weigend, A., "On overfitting and the effective number of hidden units," *Proceedings of the 1993 Connectionist Models Summer School*, 1994, pp. 335-342.

West, D., "Neural Network Credit Scoring Models," *Computers & Operations Research*, Vol. 27, 2000, pp. 1131-1152.

Wilson, R., and R. Sharda, "Bankruptcy prediction using neural networks," *Decision Support Systems*, Vol. 11, 1994, pp. 545-557.

Zhang, G.P., "Neural Networks for Classification: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 30, No. 4, 2000, pp. 451-462.

Zhang, G.P., M.Y. Hu, B.E. Patuwo, and D.C. Indro, "Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis," *European Journal of Operational Research*, Vol. 116, 1999, pp. 16-32.