Working Paper Cnr-Ceris, N. 04/2011

AN ARTIFICIAL NEURAL NETWORK APPROACH FOR ASSIGNING RATING JUDGEMENTS TO ITALIAN SMALL FIRMS

Greta Falavigna

Working Paper

Consiglio Nazionale delle Ricerche

CERIS Istituto di Ricerche sull'Impresa e Lo Sviluppo

An artificial neural network approach for assigning rating judgements to Italian small firms

Greta Falavigna

Cnr-Ceris researcher
Ph.D. in Economics and Technology Management

Cnr-Ceris
Via Real Collegio, 30
10024 Moncalieri (To)
Tel. 011-6824941
e-mail: g.falavigna@ceris.cnr.it

ABSTRACT: Based on new regulations of Basel II Accord in 2004, banks and financial institutions have now the possibility to develop internal rating systems with the aim of correctly judging financial health status of firms. This study analyses the situation of Italian small firms that are difficult to judge because their economic and financial data are often not available. The intend of this work is to propose a simulation framework to give a rating judgements to firms presenting poor financial information.

The model assigns a rating judgement that is a simulated counterpart of that done by Bureau van Dijk-K Finance (BvD). Assigning rating score to small firms with problem of poor availability of financial data is really problematic. Nevertheless, in Italy the majority of firms are small and there is not a law that requires to firms to deposit balance-sheet in a detailed form. For this reason the model proposed in this work is a three-layer framework that allows us to assign rating judgements to small enterprises using simple balance-sheet data.

KEYWORDS: rating judgements, artificial neural networks, feature selection

JEL CODES: C15; C45; G24

CONTENTS

.

INTRODUCTION

Bankruptcy prediction of firms has been largely studied by banks and researchers since late 1960 (Altman, 1968). Ravi Kumar and Ravi (2007) underlines that banks or firm health in competitive industries is dependent upon "(i) how financially solvent it is at the inception; (ii) its ability, relative flexibility and efficiency in creating cash from its continuous operations; (iii)its access to capital markets and (iv) its financial capacity and staying power when faced with unplanned cash short-falls". The concept of default probability has become very relevant after Basel II (2004) because the New Basel Accord (Basel Committee on Banking Supervision, 2004) provided the possibility for banks and financial institutions to define a model for assigning rating judgements to their customers. For this reason many statistical and intelligent techniques have been used for forecasting default probability and economic-financial performances of firms[1].

Rating judgements are known as evaluations of a potential borrower's ability to repay debt, prepared by a credit bureau at the request of the lender. Typically, a credit rating tells a lender or investor the probability of the subject being able to pay back a loan. However,

in recent years, credit ratings have also been used to adjust insurance premiums, determine employment eligibility, and establish the amount of a utility or leasing deposit.

A poor credit rating indicates a high risk of defaulting on a loan, and thus leads to high interest rate, or the refusal of a loan by the creditor. Standard and Poor's, Moody's and Fitch Rating are the biggest rating agencies but with the New Accord all financial institutions can build a system for assigning ratings. In this study judgements used have been extracted from Bureau van Dijk (BvD)[2] that defined 8 classes, from KR7 (the best performance) to D (default firms).

Obviously, the problem of determining rating scores or default probability is very close to the variables introduced in the model. Indeed big rating agencies assess performance of firms on a base of balance-sheet accounts, market information (i.e. share value) but also confidential data (economic situation of owner or shareholders or managers). Nevertheless, the problem in this study is not feature selection because the issue faced is to build a model able to determine rating score of firms with scarce balance-sheet data. Indeed rating judgement of Bureau van Dijk and then also that reproduced in this context is called "technical" because

---

[1] For a deeper illustration of techniques see Ravi Kumar and Ravi (2007).

[2] Bureau van Dijk is a company born in 1991. Its product range includes databases of company information and business intelligence for individual countries, regions and the world. Its global database, ORBIS, combines information from around 100 sources and covers approaching 65 million companies.

it does not draw from personal or informal information about firm.

The first studies on financial prediction model for bank were made by Beaver (1966) and Altman (1968) that have used univariate analysis and discriminant one. The last work has been revised in Zeta analysis by Altman *et al.* (1977) that have introduced comprehensive inputs in the previous analysis. In the following years many researchers used different models as logistic regression but also data envelopment analysis (DEA), for predicting failure probability of bank (Martin, 1977; Cielen *et al.*, 2004; Kao and Liu, 2004). All these researches use statistical methodologies suffering from the assumption of multivariate-normality for independent variables. If this hypothesis is violated these methods could be theoretically invalid for finite sample (Berrt and Linoff, 1997; Huang *et al.*, 2004). Data mining techniques and intelligent systems are good tools for finding out the potential and significant information from large data sets (Lin *et al.*, 2009). Indeed many authors displayed that decision tree methodology or artificial neural networks perform better than the discriminant analysis or logit regression in predicting bankruptcy event (Olmeda and Fernandez, 1997; de Andres *et al.*, 2005). A second flow of studies focuses on hybrid frameworks that include strengths of different models. In this direction is the work of Ahn *et al.* (2000) that combines a rough set approach and a neural network for forecasting failure of firms based on past financial performance data. More recent study is made by Ryu and Yue (2005) that introduce the so called "isotonic separation method" for bankruptcy prediction. The evidence emerging from these studies is that hybrid methodologies are the best performers because statistical methodologies, as logit, probit, ANOVA, discriminant analysis, are very effective in selecting determinant variables; whereas intelligent models are the best in providing previsions (Lee *et al.*, 1996; Lin and McClean, 2001). Other researches use hybrid intelligent models for selecting input variables and forecasting default event as fuzzy clustering, self-organizing map or support vector machines (Alam *et al.*, 2000; Huang *et al.*, 2007; Lin *et al.*, 2009).

The same considerations and results have been found on credit scoring field. Indeed creditors construct classification rules (rating or credit scoring models) based on data of applicants. Hybrid methodologies are the most numerous in the literature, in particular the methodologies combining statistical and intelligent techniques: clustering and neural network (Hsieh, 2005), artificial neural networks and multivariate adaptive regression splines (Lee and Chen, 2005), discriminant analysis and neural network (Lee *et al.*, 2002), genetic algorithm and neural network (Chen and Huang, 2003).

The research presented in this paper uses

a unique technique that is the feed-forward artificial neural network with back-propagation algorithm. It will show the ability of this model to predict insolvency of a firm and to simulate credit ratings. On the computation side, an algorithm for calculating an optimal threshold that separates good firms from bad ones is illustrated. Moreover, a three-layer neural network framework for the rating simulation is explained.

1. RESEARCH GOAL

In this study three subjects are considered. The first two have a computational soul and the third one represents an application of model to rating assessment problem.

The inspiring idea was born from the evidence that in Italy the majority of firms have micro or small size and laws do not require them to present completed balance-sheets. It became very difficult to assess these firms from the only information collected in these poor documents and also big analyst companies, as Bureau van Dijk (BvD), were not able to judge a lot of them. Ceris-CNR institute analyses the economic and financial situation of Piedmont firms applying descriptive statistics but the data lack can often produce distorted images of reality or weak considerations.

On a wake of these remarks in the first presented model we discuss the artificial neural network validity in forecasting default probability of firms and its classification power in the situation of weak disposability of balance-sheet data.

A second more complex framework will be shown for proving that it is possible to simulate rating judgements also with scarce data. Indeed Ceris-CNR buy from Bureau van Dijk rating judgements but for the above mentioned reason not all firms have their rating scores.

A third result shows a model customization built for improving outcomes just obtained from the previous framework.

1.1. Models and data

As explained in the previous section Italian laws do not require that all firms compile detailed balance-sheets. This depends on the nature of society and then, substantially, on the size of firms.

Rating agencies and analyser societies even have not disposability of data outside of balance-sheets for assessing financial and economic performances of firms.

AIDA[3] database has been the source of our Italian data and according to the mainstream analysis of Ceris-CNR and

---

[3] AIDA is a database edited by Bureau van Dijk that contains comprehensive information on listed companies in Italy, with up to five years of history. Provided information are about accessible accounts following the scheme of the 4th Directive CEE, probabilities of default and rating assessments, over to provide information about identification number, address and trade description of collected firms.

according to the unbiased method of BvD, in the model only balance-sheet accounts are used, so that the judgement for each firm is a technical rating.

2. FIRST MODEL

The first step of the analysis was to define a model for forecasting default probability of firms starting from a database of failed and best performance firms. The methodology chosen is a feed-forward artificial neural network with back-propagation algorithm as performance function.

An Artificial Neural Network (ANN) is a network made up of several simple processors, units or neurons, each one possibly having a local memory. The units are connected by unidirectional communication channels, connections, which carry numeric, as opposed to symbolic, data. The units operate only on their local data and on the inputs they receive via the connections.

A neural network is a processing device, either an algorithm or actual hardware, whose design was inspired by the design and functioning of animal brains and components thereof.

Most neural networks have some sort of training rule whereby, the weights of connections are adjusted on the basis of presented patterns. In other words, neural networks learn from examples, just like children who, for example, learn to recognise cats from examples of cats, and exhibit some structural capability for generalisation.

Neurons are often elementary non-linear signal processors, or they can be simple threshold discriminators. Another feature of ANNs, which distinguishes them from other computing devices, is a high degree of interconnection, which allows a high degree of parallelism. Furthermore, there is no idle memory containing data and programs, but rather each neuron is pre-programmed and continuously active.

The framework of an artificial neural network is represented as a group of nodes joined among them through links. In the Multi-Layer Perceptron (MLP) network the neurons are organised in layers, as shown in figure 1.

Literature suggests that an artificial network with a only one hidden layer is the best structure for facing classification problem (Dillon *et al.*, 1979; Hornik et al., 1989; Hornik, 1993; Olmeda and Fernandez, 1997; Min and Lee, 2005).

The feed-forward connections show that the information travels in the direction from the input layer to the output one, but it is possible to build different frameworks as the recurrent networks that provides different link among nodes.

The input layer has a number of neurons equal to the variables analysed, then the input matrix X is done by i rows and n columns (X_{in}). In this study subscript i represents balance-sheet indexes and n are firms.
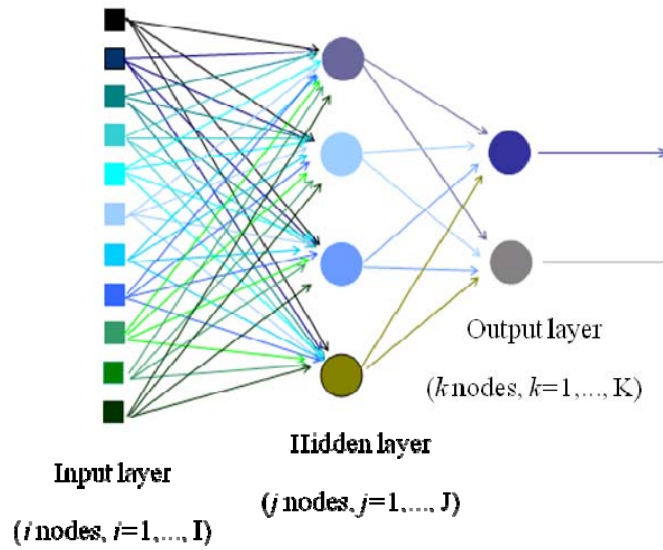
Figure 1: *Feed-forward neural network*

The optimal number of hidden nodes must be found empirically (Kim, 2003; Min and Lee, 2005) but many researchers have shown that there is a relation between the hidden node number and the input one. Patuwo *et al.* (1993), Nath *et al.* (1997) and Chauhan *et al.* (2009) suggests to use the formula *(2i+1)* where $i = 1, ..., I$ represents the number of considered variables and this is the method chosen in the present study. Salchenberger *et al.* (1992) and Olmeda and Fernandez (1997) propose the proportion *0.75i* that is more performing in time computing compared to that offered by Chauhan *et al.* (2009).

The output layer is formed by the same number of neurons as the answers required to the model (*k* rows and *n* columns). In this paper, only one neuron is the solution adopted, then $k = 1$ and *n* is the number of firms (Y_{kn}).

Arrows linking layers represent connection weights (the biological "synapses") that are tools collecting rules about inputs. This information is mathematically represented by weights: W_{ij} from input to hidden layer and W_{jk} from hidden to output one.

For computing the weight matrixes, that are the containers of signals between subsequent layers, are determined by mathematical transformation (called "transfer functions").

The transfer or activation function, ϕ ($\cdot$), from the hidden layer to the output one varies on the basis of a considered problem. This function defines the output of a neuron in terms of the induced local field. In this case, we use a logsigmoid function. The transfer function from the

input layer to the hidden one is tansigmoid.

The only requirement of the activation function is to satisfy the differentiability property. Logistic and hyperbolic tangent functions are continuously differentiable non-linear activation functions.

In this study the activation function between the input and the hidden layer is the hyperbolic tangent sigmoid function:

$$f(x) = \frac{2}{1 + e^{-2x}} - 1$$

Where $f(x) \in [-1,1]$ and $x \in (-\infty, +\infty)$. Similar function, logarithmic sigmoid function, is used from the hidden to the output layer for having a score included between 0 and 1, indeed:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Where $f(x) \in [0,1]$ and $x \in (-\infty, +\infty)$.

This allows to obtain a probability for each element (firm) on a base of variables introduced and analysed by network neurons.

Figure 2 represents how network works. The initial dataset is subdivided into two groups: the training set (X_{in}) and the validation one (V_{ih} where $h = 1, ..., H$).

Generally the second one is a little part of the first one, usually about 10%. Neural network runs twice: in a first phase the model analysed the relationships among the data because input variables (X_{in}) but also outputs (targets, T_{kn}) are introduced in the model. The network learns and collects this information into the weight matrixes (W_{ij} and W_{jk}). When the network is well trained to recognize inputs on a base of an error measure, the information saved is applied in the second phase on inputs of the validation set (V_{ih}). The main and fundamental difference between these two phases is that in the first one the unknown variables are the weight matrixes, whereas in the second phase the outputs are the real expected result. In this study, during the training phase of the model are introduced the firms with their variables (X_{in}) and also the healthy state (T_{kn}): the network learns and when the error is quite little, the same framework and weight matrixes are applied to the validation set (V_{ih}) for determining the healthy state of these firms. The criterion employed to evaluate the accuracy of the model is the percentage of correct classifications (Olmeda and Fernandez, 1997).
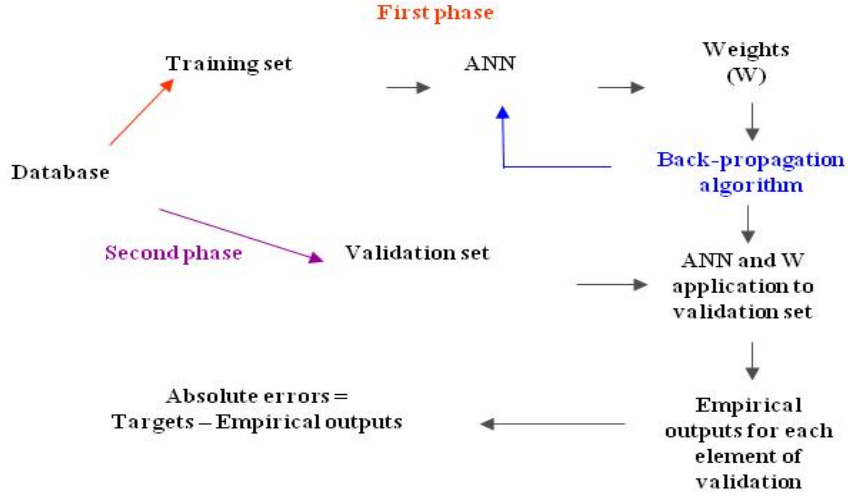
Figure 2: ANN process: the training and the validation phase

The percentage error is done comparing the number of errors made by the network to the number of elements to classify.

The back-propagation algorithm works on errors generated from the network in terms of empirical outputs (y_{kn}) too different from the expected ones (t_{kn}).

The error signal at the output of the k neuron at the m iteration (i.e. presentation of the nth training example) is defined by:

$$\varepsilon_n(m) = y_n(m) - t_n(m)$$

Neuron k is the output node and the instantaneous value of the error energy for neuron k is defined as:

$$\frac{1}{2}\varepsilon_k^2(m)$$

Correspondingly, the instantaneous value $\mathcal{E}(m)$ of the total error energy is obtained by summing the previous equation on all neurons in the output layer: these are the only "visible" neurons for which error signals can be calculated directly. The following may be written:

$$\mathcal{E}(m) = \frac{1}{2}\sum_{k \in C}\varepsilon_k^2(m)$$

Where set C includes all the neurons in the output layer of the network. This algorithm allows to minimize the absolute error both in training and validation phase. For improving the network generalization power in this study the Bayesian regulation is used. This algorithm updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well (MacKay, 1992).

In addition, a mean squared error with regularization performance function (*Mreg*) has been implemented in the

model. In this way network performance is measured as the weight sum of two factors: the mean squared error and the mean squared weight and bias values:

$$Mreg = \gamma \cdot \left(\frac{1}{N} \sum_{n=1}^{N} (t_n - y_n)^2 \right) + (1 + \gamma) \cdot \frac{1}{N} \sum_{n=1}^{N} w_n^2$$

Where γ is the performance index.

At the end of this complex process, the explained tool allows to obtain outputs defining for each firm its probability of default. Indeed the function used from the hidden layer to the output one is a logsig that has codomain comprised between 0 (the firm has best performances, its default probability is equal to 0%) and 1 (the firm is failed; its default probability is 100%).

The performances of the network are very good and we can provide a proof on Italian database of healthy/failed firms.

Balance-sheet data have been extracted from the AIDA database and they refer to Italian public societies failed in 2008. The considered accounts date to 2006 because it is necessary to forecast the earliest possible the default event and because from the analyses of BvD and K-Finance[4] the time period necessary for assessing correctly the bankruptcy is two years. Notice that in this paper failed firms are those judged by the court. Firms in bankruptcy procedures are not considered

in this study.

Database used to test the ability of network to classify firms into healthy and failed has been extracted by AIDA and it is formed by micro and small Italian firms of manufacturing industry. The initial sample was done by 631 companies whose economic-financial position in 2008 we know. On a wake of suggestion of K-Finance 2006 balance-sheet variables have been extracted. The problem of many works is that detailed variables are required for being able to forecast firm performance. The strength of artificial neural networks is that they can learn the relationships among data from an initial dataset (training phase) and after, the acquired rules are applied to a validation sample. This propriety allows to simulate results, also starting from different information regarding the initial ones.

The analysis of weight matrices built during the training of network allows to deduce the relevance of considered features. Also if the goal of this work is not the feature selection, because firms studied have poor balance-sheet data, it is necessary to highlight this issue.

Literature studied largely the problem of feature selection because when a model is build, its performance is often determined by inputs introduced.

Researchers proposed many methods for choosing model variables: starting from simpler methodologies, as the ANOVA (Lin and McClean, 2001; Huang *et al.*,

---

[4] K Finance is a consultancy firm that offers advisory on: mergers & acquisitions, private equity deals, planning & control, fairness opinions, structured finance deals, listings.

12

2004) for extending the analysis of problem to more complex models as combined methodologies of Support Vector Machines and F-score or genetic algorithms (Huang *et al.*, 2007; Ravi and Pramodh, 2008; Huang, 2009).

Moreover, artificial neural networks have the interesting advantage that don't require that $X_{in} \sim N(\mu, \sigma^2)$, as instead required by econometric models but too many information could be redundant and make difficult to perform model. Nevertheless, a variable that alone is completely useless can provide a significant performance improvement when taken with others (Guyon and Elisseeff, 2003). For these considerations variable selection is a very decisive topic for providing good results. The framework of a model can be perfect but if variables are unfit it could be very difficult to have performing outcomes.

In this study variables to introduce in the model are specified before but it could be interesting catches the relevance of each variable to the default problem. In this sense many researchers (Nath *et al.*, 1997; Mak and Blanning, 1998; Huang *et al.*, 2004; Huang *et al.*, 2007; Chauhan *et al*, 2009) suggest the Garson index (Garson, 1991) as a tool able to represent the importance of each feature introduced in the model.

Let i input variables (where $i = 1, ..., I$), j hidden nodes (where $j = 1, ..., J$) and k output neurons (where $k = 1, ..., K$), matrices are signed with capital letters

and their elements with minuscule ones, Garson index is a vector made by:

$$G_{ik} = \frac{W_{ik}^*}{Z_k} \cdot 100$$

where $W_{ik}^* = \frac{|W_{ij}|}{S_j} \cdot |W_{jk}|$

W_{ij} is the weight matrix between input and hidden layer and W_{jk} is the weight matrix between hidden and output one.

$$S_j = \sum_{i=1}^{I} |w_{ij}| \qquad j = 1, ..., J$$

$$Z_k = \sum_{i=1}^{I} w_{ik}^* \qquad k = 1, ..., K$$

The indexes used are only eight, built starting on very simple balance-sheet information: Receivables due from shareholders (called Share capital), Fixed assets, Total current assets, Equity, Total provisions for risks and charges, Total payables, Value of production, Costs of production and Financial interests. In table 1 balance-sheet variables and a short summary of database subdivided on a base of healthy state of firms is done.

Figure 3 shows the percentage weight of each variable in forecasting the default probability. The result is not surprising because the ratio between Total payables and Liabilities represents a measure of financial dependence of a firm from external creditors. From the analysis of statistics (table 1) healthy companies are less dependent than the failed ones, that is

that they are less leveraged and with less interests to pay.

The sample for training the network was formed by 588 firms, whereas the validation set was of 43 elements.

The results on validation sample are shown in figure 4 where blue stars (*) are the outputs of network, that is probabili-

ties of default, and red plus (+) represent the correct classification of a firm (targets). Vertical red lines display shifts between empirical outputs and expected ones. The relevance of this model does not only show good ability of network to classify elements but a new cue is done by the green horizontal line in the figure

Table 1*: Sample description for the default probability model*
- sample size and mean values of input variables –

Type of firms	Sample size	Total payables on Liabilities	Total payables on Value of production	Return on Sales (ROS)	Return on Asset (ROA)	EBIT on Total payables	Fixed assets on Value of production	Financial interests on EBIT	Fixed assets on Equity
Healthy	525	0.422	0.565	0.126	0.119	0.301	1.413	0.021	1.255
Failed	106	0.916	0.997	0.036	0.035	0.039	3.532	0.864	0.136
Total	**631**	**0.505**	**0.638**	**0.110**	**0.105**	**0.257**	**1.571**	**0.025**	**0.527**

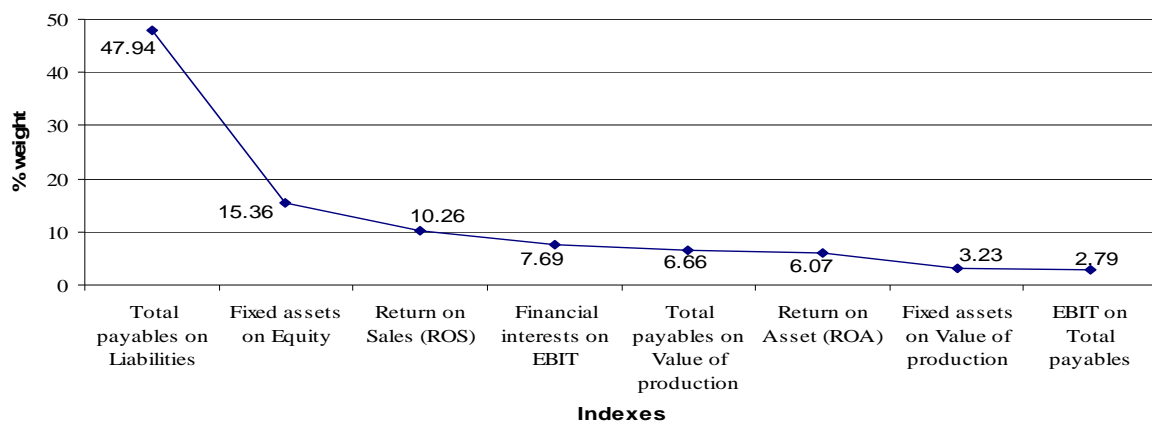Figure 3*: Garson indexes for default probability model*

Indeed this represents a threshold subdividing firms into healthy (network output < threshold) and failed (network output > threshold).

Many authors established 0.5 as the cut-off value (Olmeda and Fernandez, 1997; Huang, 2007) of a model analysing similar problem but in this paper the optimal value of threshold is not defined by the user but it is calculated in the training phase so that performing the results.

Let s a vector of thresholds where $s \in [0, 1]$ with a step of 0.001, y_n a vector of network outputs (where $n = 1, \ldots, N$) and t_n a vector of targets, it is possible to define a vector y_n^* as:

$$y_n^* = \begin{cases} 0 & if\ (s - y_n) > 0 \\ 1 & if\ (s - y_n) < 0 \end{cases}$$

(for each value of s)

In this manner for each threshold value a vector of approximated network outputs has been computed to compare with targets $|y_n^* - t_n|$.

At the end of the process we will obtain the number of correct classification of n elements for each s. At this point the algorithm computes the optimal threshold as the mean of cut-off values with the lowest number of misclassifications. This method of threshold selection takes into account the size of absolute errors (y_n-t_n) made by network because the threshold is compared directly with the empirical output of network not approximated yet.

Notice that due to the fact that the threshold is fixed to 4 digit we will obtain a threshold measure that is an approximation of the optimal cut-off value.

At the end the percentage error (E) is:

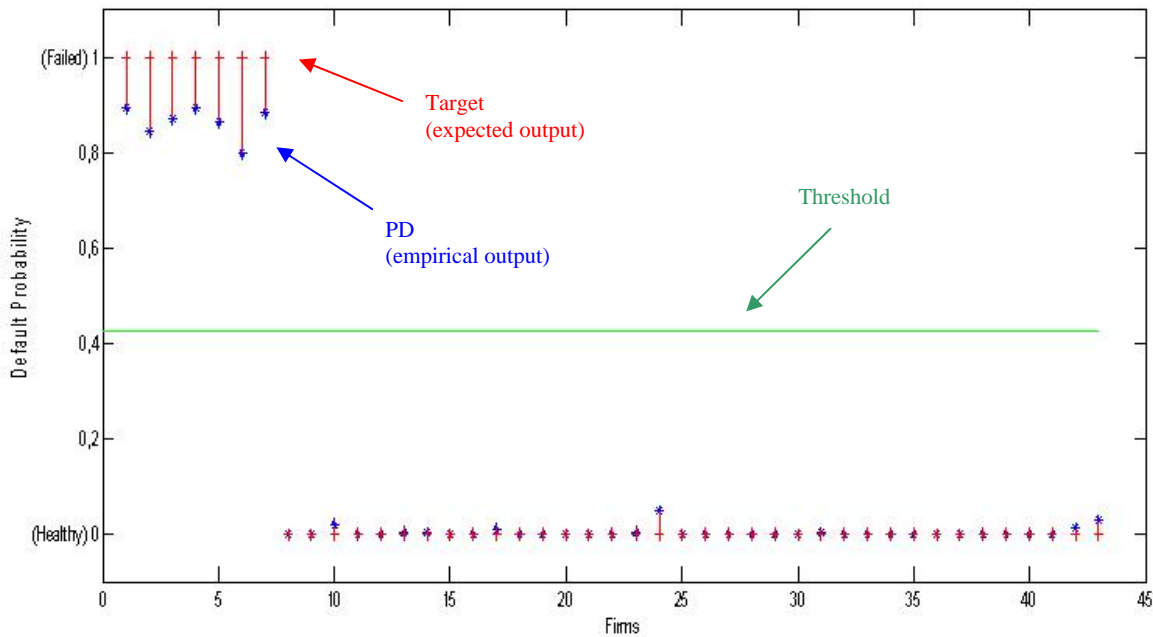$$E = \frac{\sum_{n=1}^{N} |y_n^* - t_n|}{N} \cdot 100$$

Figure 4: *Feed-forward neural network – Results on default probability*

3. SECOND MODEL

After validating the feed-forward artificial neural network with back-propagation algorithm and having computed a performing algorithm for classifying firms into healthy and failed ones, this study shows as the previous tested network can be used for simulating rating values of firms.

The database used is extracted by AIDA for the year 2008 and for each firm K-Finance with Bureau van Dijk have provided the rating judgements.

For the prior considerations rating judgements are not available for all firms and we propose a model able to simulate them and to upload missing values in the database.

Ratings are judgements defined by accredited agencies that analysing economic and financial statements of firms issue a score representing the economic safety of companies. The rating judgements are expressed as score, through an acronym. In particular, for K-Finance the best performance rating class is the best performance KR7 and the worst KR1. Default firms are assigned to D class whereas the firms with intermediate economic-financial performance are of KR6, KR5, KR4, KR3 and KR2.

The framework of the model proposed in this research is a three-layers structure: in each structure at least one neural network runs. In the figure 5 there is a scheme of the three-layer model.
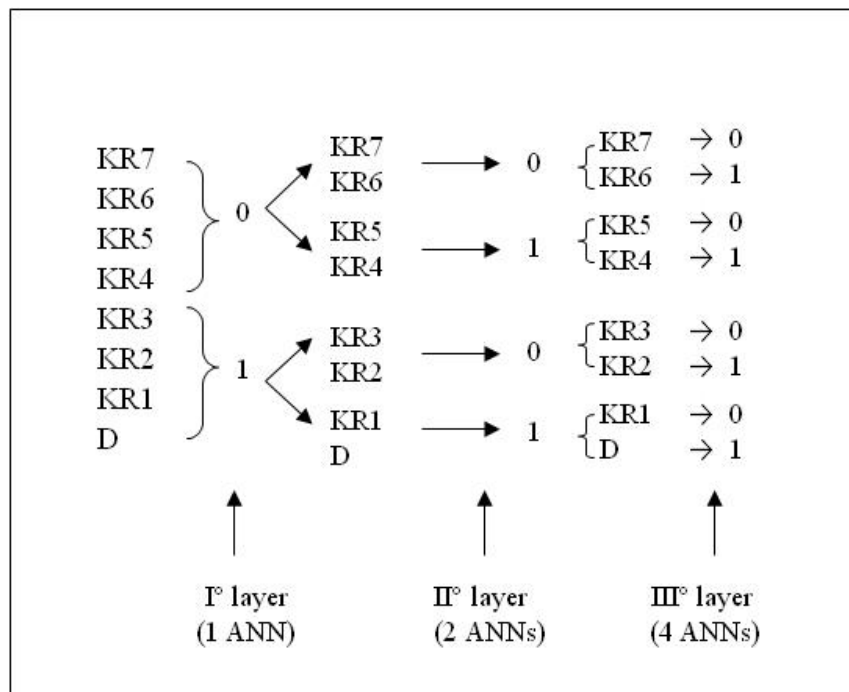

Figure 5: *Three-layer framework for simulating rating judgments*

As shown in this figure, in the first layer the training sample is subdivided into two groups: 0 for KR7, KR6, KR5, KR4 firms and 1 for KR3, KR2, KR1, D companies. Only one neural network runs, able to classify firms in correct group. In the second layer, two artificial neural networks run. The first one groups KR7 and KR6 as 0 and KR5 and KR4 as 1; the second network selects KR3 and KR2 firms as 0 and KR1 and D as 1. At the end, in the third layer, 4 neural networks run. Each model subdivides firms into two groups (0 an 1) and each firm will be classified in its correct class of rating.

The proposed model exploits the properties of artificial neural network and minimizes the potential error that should incur using the network with non dichotomous output.

The results are very surprising because they confirm the ability of a model to simulate correctly the rating also for firms with poor data.

The database used for testing the model is formed by all Italian manufacturing firms. After introducing data into the model a pre-processing methodology has been adopted. Input variables are the same considered in the previous model.

The steps of firm selection are the following three:

1) For each *n*-th firm the norm of indexes (N_n) introduced in the model has been calculated:

$$N_n = \sqrt{\sum_{i=1}^{I} |x_{in}|^2} \qquad \text{where } n = 1, ..., \text{N}$$

2) The mean and the variance of norms have been computed;

3) All firms with a norm value higher than the mean of norms plus once the variance are deleted from the database.

Nevertheless this procedure has been run not on the whole database but after a clustering on a base of firm size and rating judgement. In this manner firms have been subdivided into 5 classes of size[5] and inside these groups they are subdivided into the 8 rating classes. This means that the procedure for deleting outliers has been run 40 times.

In this manner the spikes are deleted taking into account both the size effect and the economic-financial performance aspect and the outliers are dropped from the database. In figure 6a there is the plot of the norms for the group of big KR7 firms with spikes in red circles.

In fig. 6b the histogram of norms, the mean (red vertical line) and the median (yellow vertical line) are shown. Green vertical lines in figure 6b represent the mean of norm plus respectively once and twice the variance.

After the firm selection process and the elimination of firms with missing value in necessary variables, the training database

---

[5] The firm size has been defined on a base of sales (*S*):

$$Firm = \begin{cases} if\ S \le €2mln & Micro \\ if\ €2mln < S \le €10mln & Small \\ if\ €10mln < S \le €50mln & Mean \\ if\ €50mln < S \le €290mln & Big \\ if\ S > €290mln & Large \end{cases}$$

is formed by 35,978 units. For improving results, three models are performed, on a base of size of firm (Micro, Small and Other). Table 2 represents statistics on the training sample and validation one, considering that elements of validation are not passed by pre-processing phase.

For simulating rating of presented firms the feed-forward artificial neural network with back-propagation tested in the previous section has been introduced in the proposed three-layer model including the threshold search algorithm in training phase.

This framework should minimize the number of big errors (more than one class) that could be very destructive for economic-financial evaluation of firms. One-class errors are considered less incident because the difference in judgments between sequential classes are not very big. A pre-test has been made on the whole database, not subdivided on a base of size and the results were good if two-or-more class errors are considered (0.41%) but one-class errors, also if less serious, are many still (37.63%).
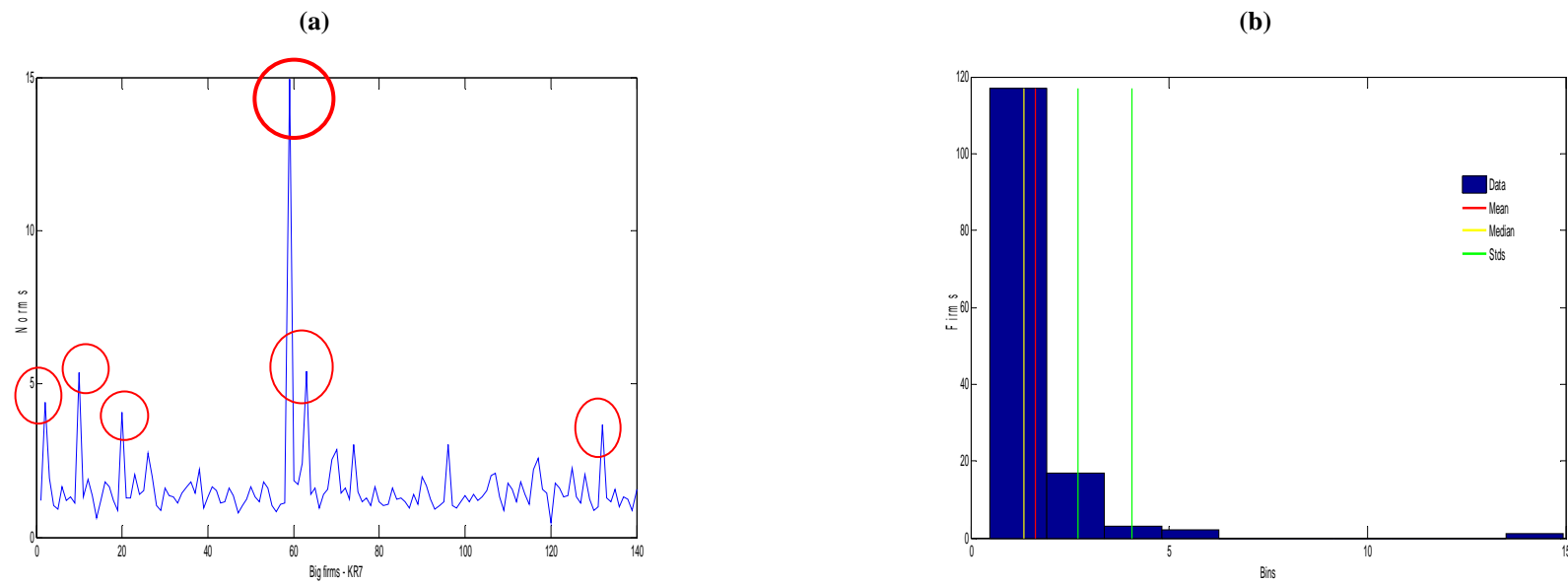
(a)

(b)

Figure 6: *(a) Plot of norms and spikes – (b) Histogram of norms, mean and variance*
[Big KR7 firms]

Table 2*: Descriptive statistics on training and validation sets*

Rating Classes	Training				Validation			
	Size				Size			
	Micro	Small	Other	Total	Micro	Small	Other	Total
D	521	255	79	855	67	7	8	82
KR1	744	154	34	932	80	17	3	100
KR2	2,267	871	139	3,277	248	96	15	359
KR3	2,863	1,623	509	4,995	284	180	56	520
KR4	3,872	3,002	1,757	8,631	321	330	195	846
KR5	2,495	2,043	1,502	6,040	235	220	166	621
KR6	3,947	2,944	2,034	8,925	405	320	226	951
KR7	1,058	719	546	2,323	120	80	60	260
Total	*17,767*	*11,611*	*6,600*	*35,978*	*1,760*	*1,250*	*729*	*3,739*

Nevertheless another test has been made on the same data. Subdividing the data on a base of three-size (Micro, Small and Other), for each dimensional class a specific model ran, providing more persuasive results, presented in figures 7, 8 and 9. In these figures empirical outputs (called "Network Outputs") are depicted with black five-pointed stars (☆) whereas red square (☐) are expected results (called "BvD Outputs"). Yellow lines represent the errors made by models.

Micro firms (figure 7) are the largest and heterogeneous group and for this reason the number of errors is big but, the model does not commit two or more class mis-take (0%), whereas 492 are firms not correctly classified (28%).

The second group of firms is small sized. This sample is both less numerous and less heterogeneous than that of Micro units and the effect on the network performance is that the number of one-class errors diminishes to 2.23% (29 non-correct classification, figure 8) whereas two-or-more class errors are always 0.

The last sample is formed by firms of non-micro and non-small size (so that "Other"). This is the smallest group of elements and the results are very good as figure 9 shows (19 one-class error, 2.61% and 0 two-or-more class mistakes).
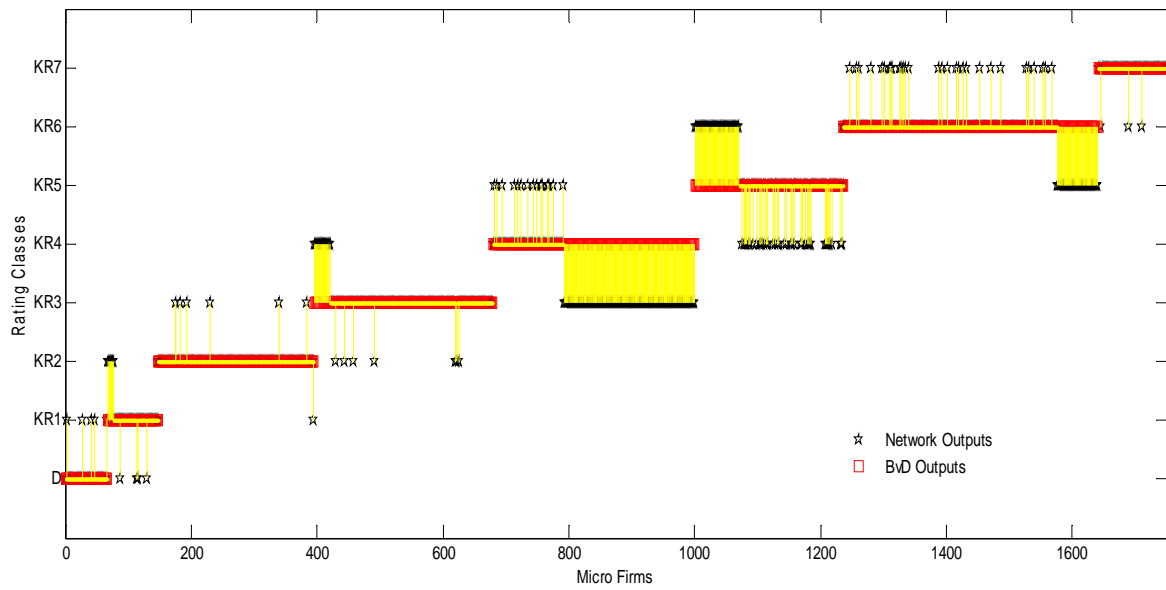
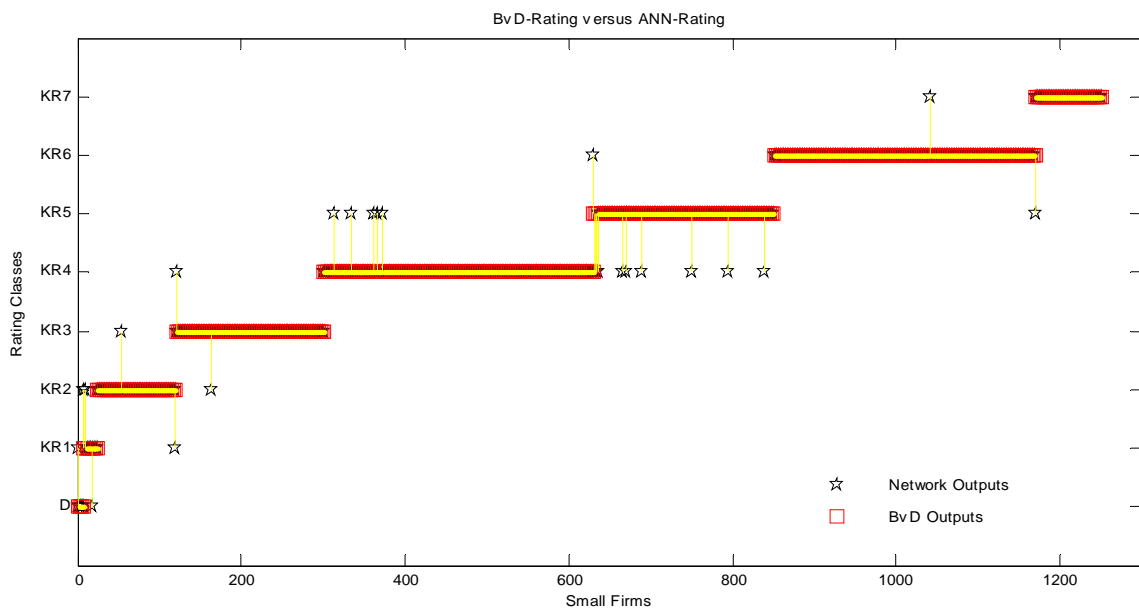Figure 7: *Three-layer Artificial Neural Network for Micro firms*

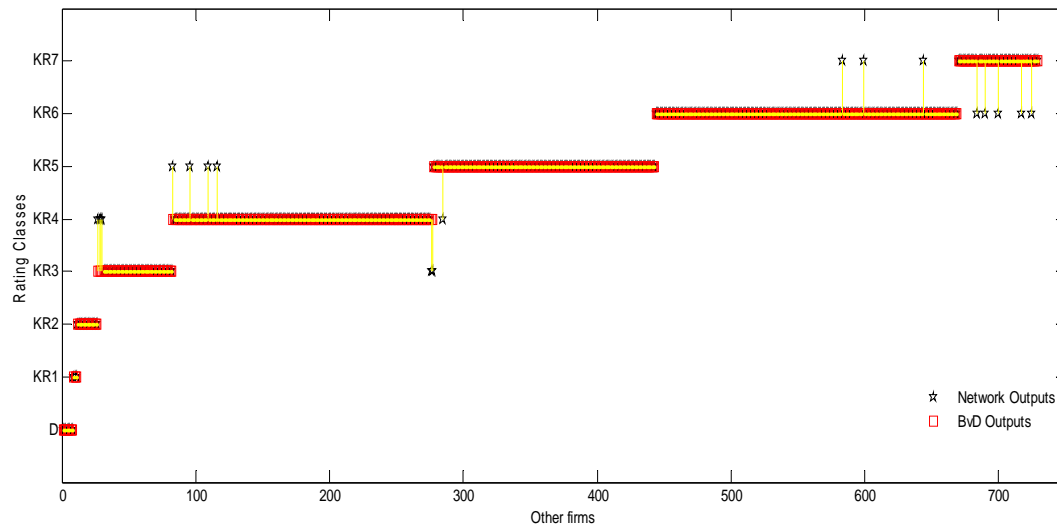Figure 8: *Three-layer Artificial Neural Network for Small firms*

Figure 9*: Three-layer Artificial Neural Network for other firms*

4 CONCLUSIONS

The results we obtained confirm that artificial neural networks are very good tools in generalization and classification. These proved characteristics allow to use neural network for solving problem in many different fields.

This paper proposes the well known methodology of artificial neural networks applied to financial problem of default of firms and rating judgements. In particular, generalization property is exploited so that to provide good solution to financial problem.

The study concurs to economic literature proposing a first tool able to forecast the default event two years before the bankruptcy. This model is used for validating the ability of network in generalization and in forecasting. Moreover, a threshold search algorithm for minimising errors of network has been built.

A second model, the three-layer neural network framework, uses the previous tested model for simulating rating judgements of firms with scarce data and covering missing values that often pose a threat to significance of analyses. This problem is frequent in Italy because firms are micro or small sized and it is difficult to study them.

The results are very good for both the models presented. From these issues stems the possibility to propose the tool to banks, as internal rating system, or to micro and small firms for having a previous evaluation before asking for a loan.

Nevertheless, it is necessary to perform better the results because the test has been made only on manufacturing firms. Moreover, it could be interesting to use cluster analysis in a previous phase and to see if results are different.

REFERENCES

Ahn B.A., Cho S.S. and Kim C.Y., 2000, "The integrated methodology of rough set theory and artificial neural network for business failure prediction", Expert Systems with Applications, 18, pp. 65-74.

Alam P., Booth D., Lee K. and Thordarson T., 2000, "The use of fuzzy clustering algorithm and self-organization neural networks for identifying potentially failing banks: An experimental study", Expert Systems with Applications, 18, pp. 185-199.

Altman E.I., 1968, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", Journal of Finance, 23, pp. 589-529.

Altman E.I., Haldeman R.G. and Narayanan P., 1977, "ZETATM analysis, a new model to identify bankruptcy risk of corporations", Journal of Banking and Finance, 1, pp. 29-54.

Basel Committee on Banking Supervision, 2004, "International Convergence of Capital Measurement and Capital Standards. A Revised Framework", Bank for International Settlements, ISBN web: 92-9197-669-5, http://www.bis.org/publ/bcbs107.pdf?noframes=1

Beaver W.H., 1966, "Financial ratios as predictors of failure", Journal of Accounting Research, 4, pp. 71-111.

Berrt M.J.A. and Linoff G., 1997, Data mining techniques: For marketing sales, and customer support, Canada: John Wiley & Sons.

Chauhan N., Ravi V. and Chandra D.K., 2009, "Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks", Expert Systems with Applications, 36, pp. 7659-7665.

Chen M.C. and Huang S.H., 2003, "Credit scoring and rejected instances reassigning through evolutionary computation techniques", Expert Systems with Applications, 24(4), pp. 433-441.

Cielen A., Peeters L. and Vanhoof K., 2004, "Bankruptcy prediction using data envelopment analysis", European Journal of Operational Research, 154, pp. 526-532.

De Andres J., Landajo M. and Lorca P., 2005, "Forecasting business profitability by using classification techniques: A comparative analysis based on a Spanish case", European Journal of Operational Research, 167, pp. 518-542.

Dillon W.R., Calantone R.J. and Worthing P., 1979, "The new product problem: an approach for investigating product failures", Management Science, 25, pp. 1184-1196.

Garson G.D., 1991, "Interpreting neural-network connection weights", AI Expert, 6 (4), pp. 47-51.

Guyon I. and Elisseeff A., 2003, "An introduction to Variable and Feature Selection", Journal of Machine Learning Research, 3, pp. 1157-1182.

Hornik K., 1993, "Some new results on neural network approximation", Neural Networks, 6, 1069-1072.

Hornik K., Stinchcombe M. and White H., 1989, "Multilayer feedforward networks are universal approximators", Neural Networks, 2, 359-366.

Hsieh N.-C., 2005, "Hybrid mining approach in the design of credit scoring models", Expert Systems with Applications, 28(4), pp. 655-665.

Huang C.-L., Chen M.-C. and Wang C.-J., 2007, "Credit scoring with a data mining approach based on support vector machines", Expert Systems with Applications, 33, pp. 847-856.

Huang S.-C., 2009, "Integrating nonlinear graph based dimensionality reduction schemes with SVMs for credit rating forecasting", Expert Systems with Applications, 36, pp. 7515-7518.

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

Kao C. and Liu S.T., 2004, "Prediction bank performance with financial forecast: A case of Taiwan commercial banks", Journal of Banking and Finance, 28, pp. 2353-2368.

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

Lee K.C., Han L. and Kwon Y., 1996, "Hybrid neural network models for bankruptcy prediction", Decision Support Systems, 18, pp. 63-72.

Lee T.-S. and Chen I.-F., 2005, "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines", Expert Systems with Applications, 28(4), p. 743-752.

Lee T.-S., Chiu C.-C., Lu C.-J. and Chen I.-F., 2002, "Credit scoring using the hybrid neural discriminant technique", Expert Systems with Applications, 23(3), pp. 245-254.

Lin F.-Y. and McClean S., 2001, "A data mining approach to the prediction of corporate failure", Knowledge-Based Systems, 14, pp. 189-195.

Lin S.W., Shiue Y.Ren., Chen S.C. and Cheng H.M. 2009, "Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks, Expert Systems with Applications", 36(9), pp. 11543-11551.

MacKay, 1992, "Bayesian Interpolation", Neural Computation, 4(3), pp. 415-447.

Mak B., Blanning R.W., 1998, "An empirical measure of element contribution in neural networks", IEEE transactions on systems, man and cybernetics – part C: applications and reviews, 28 (4), pp. 561-564.

Martin D., 1977, "Early warning of bank failure. A logit regression approach", Journal of Banking and Finance, 1, pp. 249-276.

Min J.H. and Lee Y.-C., 2005, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters", Expert Systems with Applications, 28, pp. 603-614.

Nath R., Rajagopalan B. and Ryker R., 1997, "Determining the saliency of input variables in neural network classifiers", Computers Operational Research, 24 (8), pp. 767-773.

Olmeda I. and Fernandez E., 1997, "Hybrid classifier for financial multicriteria decision making: the case of bankruptcy prediction", Computational Economics, 10, pp. 317-335.

Patuwo E., Hu M.Y. and Hung M.S., 1993, "Two-group classification using neural networks", Decision Sciences, 24(4), pp. 825-845.

Ravi Kumar P. and Ravi V., 2007, "Bankruptcy Prediction in Banks and Firms Via Statistical and Intelligent Techniques – A Review", European Journal of Operational Research, 180, pp. 1-28.

Ravi V., Pramodh C., 2008, "Threshold accepting trained principal component neural network and feature subset selection: Application to bankruptcy prediction in banks", Applied Soft Computing, 8, pp. 1539-1548.

Ryu Y.U. and Yue W.T., 2005, "Firm bankruptcy prediction: Experimental comparison of isotonic separation and other classification approaches", IEEE transactions on Systems, Man and Cybernetics – Part A: Systems and Humans, 35, pp. 727-737.

Salchenberger L.M., Cinar M.E. and Lash N.A., 1992, "A neural networks: a new tool for predicting thrift failures", Decision Sciences, 23, pp. 899-916.

Consiglio Nazionale delle Ricerche

CERIS

Working Paper Cnr-Ceris

ISSN (*print*): 1591-0709   ISSN (*on line*): 2036-8216

Download

↓

http://www.ceris.cnr.it/index.php?option=com_content&task=section&id=4&Itemid=64