2008-10-01

Neural nets versus conventional techniques in credit scoring in Egyptian banking

Abdou, H

http://hdl.handle.net/10026.1/8370

10.1016/j.eswa.2007.08.030
Expert Systems with Applications

Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 35 (2008) 1275–1292
Expert Systems with Applications
www.elsevier.com/locate/eswa

Neural nets versus conventional techniques in credit scoring in Egyptian banking

Hussein Abdou *, John Pointon, Ahmed El-Masry

Plymouth Business School, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

Abstract

Neural nets have become one of the most important tools using in credit scoring. Credit scoring is regarded as a core appraised tool of commercial banks during the last few decades. The purpose of this paper is to investigate the ability of neural nets, such as probabilistic neural nets and multi-layer feed-forward nets, and conventional techniques such as, discriminant analysis, probit analysis and logistic regression, in evaluating credit risk in Egyptian banks applying credit scoring models. The credit scoring task is performed on one bank's personal loans' data-set. The results so far revealed that the neural nets-models gave a better average correct classification rate than the other techniques. A one-way analysis of variance and other tests have been applied, demonstrating that there are some significant differences amongst the means of the correct classification rates, pertaining to different techniques.

JEL Classification: G21; G32

Keywords: Neural nets; Conventional techniques; Banking; Credit scoring

1. Introduction

The process of credit risk evaluation has the interest of many researchers nowadays. The role of credit risks has changed dramatically over the last 10 decades, from passive automation to a strategic device. Recently, bankers have come to realise that banking operations affect and are affected by the natural environment and that consequently the banks might have an important role to play in helping to raise environmental requirements. Although the environment presents significant risks to banks, in particular, environmental credit risk, it also perhaps presents profitable opportunities (Casu, Girardone, & Molyneux, 2006; Thompson, 1998).

Decision-making of accepting or rejecting a client's credit can be supported by judgemental techniques and/or credit scoring models. The judgemental techniques rely on the knowledge and both past and present experience of credit analysts, who evaluate the required requisites, such as the personal reputation of a client, the ability to repay credit, guarantees and client's character (Sarlija, Bensic, & Bohacek, 2004). Due to the rapid increase in fund-size invested through credit granted by Egyptian banks, and the need for quantifying credit risk, financial institutions including banks have started to apply credit scoring models.

The structure of the banking system varies from country to country. In the Egyptian environment the structure includes:[1] First, public sector banks (7 banks). Second,

* Corresponding author. Tel.: +44 1752 238654; fax: +44 1752 232249.
E-mail address: hussein.abdou@plymouth.ac.uk (H. Abdou).

[1] Before the 16th October 2006 the Egyptian banking structure were consists of: commercial banks (28 banks), comprising public sector banks (4 banks) and private and joint venture banks (24 banks); and secondly, business and investment banks (31 banks), comprising private and joint venture banks (11 banks) and branches of foreign banks – off-shore banks – (20 banks). In addition, there are also specialised banks (3 banks), which are the Egyptian Industrial Development Bank, the Arab Egyptian Real Estate Bank and Principal Bank for Development and Agriculture Credit. Egyptian banks abroad are not included, also two banks established under private laws and are not registered with Central Bank of Egypt; namely, Arab International Bank, and Nasser Social Bank (Central Bank of Egypt, 2003/2004).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2007.08.030

private and joint venture banks (28 banks).Third, branches of foreign banks (7 banks). Fourth, branches ceased its operations (9 banks)[2] (see: http://www.cbe.org.eg/links.htm for more details).

Since most banks in Egypt are currently using judgemental techniques, it is important to review judgemental techniques versus credit scoring techniques. Sullivan (1981) and Bailey (2004) argue that, in a judgemental technique evaluation, each credit application including information contained with it is evaluated individually by a decision maker "creditor". The success of a judgemental process depends on the experience and the common sense of the credit analyst. As a result, judgemental techniques are: incongruity, lack of motivation, control and risk quantification.

Otherwise, in a credit scoring model, analysts usually used their historical experience with debtors to derive a quantitative model for the segregation of acceptable and unacceptable credit applications. Using a credit scoring system, a credit application is self-operating processed and consistently all credit decisions are made. The scoring system is based on the addition or subtraction of a statistically extracted number of points relating to the applicant's score given to the predictor variables, such as time on a job or the number of credit sources used. As a result, it can be said that credit scoring enables advancers to assess the credit worthiness quickly. Also, provides moderate scale to adjust the accepted quality by the lenders, and of course provides statistical techniques which enable lenders to measure it. Moreover, credit scoring give a chance to the advancers to improve the customer services process to avoid any estimated future decline. By using a statistically extracted cut-off credit score, an analyst can of course separate the acceptable from the unacceptable credit applicants. On the other hand, credit scoring has been criticized because statistical problems with the data used to evolve the model assumptions of the statistical technique used to derive the point scores. Besides, variables used in a credit scoring system may have the effect of social discrimination. By analysing clients' characteristics to who were once granted credit, the scoring system may provide a bias results because of the different circumstances when those clients or new clients applying for credits. Despite the criticism of credit scoring models, these models can be regarded as one of the most successful models used in the field of business and finance (Bailey, 2004; Sullivan, 1981).

Credit scoring is a quantitative evaluation technique employed by financial institutions "banks" to assess the creditworthiness for both individuals and firms that applies for loans. On other words, the set of decision models that provide lenders in the granting of consumer credit. These techniques assess, and therefore help to decide, who will get credit, how much credit they should get, and what operational strategies will sustain the profitability of the borrowers to the lenders (Long, 1973; Thomas, Edelman, & Crook, 2002).

Recently neural nets have emerged as a practical technology, with successful applications in many fields in financial institutions in general and banks in particular. Concerning with many problems such as pattern recognition, and make use of feed-forward nets architecture such as the multi-layer feed-forward nets and probabilistic neural nets, are the majority of these applications (Bishop, 1995; Masters, 1995).

Linear regression and discriminant analysis are widely-used statistical techniques, as evidenced in the literature follows. The other methods are: logistic regression, probit analysis, mathematical programming, non-parametric smoothing methods, Markov chain models, expert systems, neural networks, genetic algorithms and others (Hand & Henley, 1997). For such a new banking environment, it would see appropriate, as a first step, to investigate neural nets versus some of the conventional techniques.

Indeed, discriminant analysis and logistic regression are still used in building and developing credit scoring models (Caouette, Altman, & Narayanan, 1998; Desai, Crook, & Overstreet, 1996; Hand & Henley, 1997; Sarlija et al., 2004). Generally, the best technique for all data sets does not exist. Therefore, the main thrust of this paper is to investigate the ability of neural nets such as probabilistic neural nets and multi-layer feed-forward nets, and conventional techniques such as discriminant analysis, probilt analysis and logistic regression in evaluating credit risk in Egyptian banks using credit scoring models, in terms of a case study. Discussion with banking officials would suggest that most banks in Egypt are using judgemental techniques in their evaluation process, except a limited number of banks using scoring sheets and/or semi-scoring systems in their evaluation process. We are examining integrated models for the evaluation of consumer credit risks in the banking sector in Egypt; especially since credit scoring models have undergone a noticeable success in different environments in Europe and the US, taking into account all requirements for the proposed models according to the nature of the Egyptian environment.

1.1. Neural nets versus traditional statistical methods

Neural nets provide an alternative to conventional statistical techniques. Such as Linear Regression, a function approximation is used. Otherwise, for the classification purposes, discriminant analysis, probit analysis and logistic regression are used. The point of using neural nets is that its capability of modelling extremely complex functions, and of course, this stands in contrast with the traditional linear techniques, such as, linear regression and linear discriminant analysis. Probabilistic neural nets usually trains presented cases faster than multi-layer feed-forward nets, and classifies like or better than multi-layer feed-forward

[2] The board of the CBE agreed to cancel two banks, Jammal Trust Bank and Rafidain Bank, from its record.

nets, taking into account that multi-layer feed-forward nets have been shown as excellent classifiers. However, a range of sophisticated algorithms for neural nets training, making them an attractive alternative to the more conventional techniques has been known (Masters, 1995; Palisade, 2005).

Our empirical results reveal an 86.75% average correct classification rate using discriminant analysis. With a stepwise discriminant approach, nine significant predictor variables were selected in the final model and we found an 86.92% average correct classification rate. For probit analysis an 87.78% average correct classification rate was found. Moreover, an 87.26% average correct classification rate was observed after excluding the insignificant variables. Using logistic regression, it was found that the average correct classification rate was 88.30%, and 87.95% after excluding the insignificant variables. The above conventional techniques were employed using a 0.50 cut-off point. A 96.21% average correct classification rate was found using probabilistic neural nets. A 94.84% average correct classification rate for multi-layer feed-forward nets with five nodes and 93.98% average correct classification rate using multi-layer feed-forward nets with four nodes. In general, all models gave better correct classification rates than the currently used system (74.5% of all accepted loans which did not lead to default, i.e., 433/581). Misclassification costs are also investigated in this paper; since the cost associated with type I errors differ from those associated with type II errors.

This paper is organized as follows: Section 2 discusses the literature review. Section 3 details the research methodology and data collection. Section 4 explains the results. Finally, Section 5 concludes the study results and suggests the area for the future researches.

2. Literature review

Credit scoring was one of the earliest financial tools developed use by US retailers and mail-order institutions in the 1950s to be used in the risk assessment process, which is coetaneous with the early applications of portfolio analysis to manage and diversify the risk inherent in investment portfolios. In addition, credit scoring aims to estimate risk of the clients in their loans, not to explain it. (Mays, 2004; Thomas et al., 2002). The objective of credit scoring models is to assign loan customers to either good credit or bad credit (Lee, Chiu, Lu, & Chen, 2002), or predict the bad creditors (Lim & Sohn, 2007). Therefore, scoring problems are related to classification analysis (Anderson, 2003; Hand, 1981; Lee et al., 2002). Classification models for credit scoring are used to categorize new applicants as either accepted or rejected with respect to their characteristics, such as, marital status, age, and income (Chen & Huang, 2003). At the same time, this suits the Egyptian environment, with perhaps the addition of other variables, such as corporate guarantee, monthly salary and education.

The credit scoring model is one of the most successful applications of research modelling in finance and banking, and the number of scoring analysts in the industry is constantly increasing. Yet because credit scoring does not have the same luster as the pricing of exotic financial derivatives or portfolio analysis, the literature on the subject is very limited. However, credit scoring has been vital in allowing the phenomenal growth in consumer credit over the last few decades. Without an accurate and automatically operated risk assessment tool, lenders of consumer credit could not have expanded their loan books in the way they have (Bailey, 2001; Bluhm, Overbeck, & Wagner, 2003; Lewis, 1992; Mays, 2001; Siddiqi, 2006; Thomas et al., 2002).

Possibly the earliest use of applying multiple discriminant analysis to credit scoring is the work by Durand (1941), who examined car loan applications. A well-known application in corporate bankruptcy prediction is one by Altman (1968), who developed the first operational scoring model based on five financial ratios, taken from eight variables from corporate financial statements. He produced a Z-Score, which is a linear combination of the financial ratios.

The evaluation of new consumer loans is one of the most important applications of credit scoring models and it has attracted some attention in the last few decades (Malhotra & Malhotra, 2003; Sarlija et al., 2004; Steenackers & Goovaerts, 1989). Some researchers have focused on existing consumer loans rather than new loan applications (Kim & Sohn, 2004; Orgler, 1971). Statistical techniques, such as discriminant analysis, regression analysis, probit analysis and logistic regression, used in building the scoring models have been examined (Banasik, Crook, & Thomas, 2001; Boyes, Hoffman, & Low, 1989; Greene, 1998; Orgler, 1971; Sarlija et al., 2004; Steenackers & Goovaerts, 1989). There have also been case studies of building credit scoring models (Banasik et al., 2001; Lee & Chen, 2005; Lee et al., 2002; Leonard, 1995).

A few credit scoring models using probabilistic neural nets have been investigated, (Masters, 1995; Zekic-Susac, Sarlija, & Bensic, 2004). Correspondingly, of course many scoring models applying multi-layer feed-forward nets have been used (Bishop, 1995; Desai et al., 1996; Dimla & Lister, 2000; Reed & Marks, 1999; Trippi & Turban, 1993; West, 2000). The neural network models have the highest average correct classification rate when compared with discriminant analysis and logistic regression, although results are very close.

Hybrid models, as well as neural networks and advanced statistical techniques have been used in building scoring models (Kim & Sohn, 2004; Lee et al., 2002; Blochlinger & Leippold, 2006; Lee & Chen, 2005; Seow & Thomas, 2006). Meanwhile, comparisons between traditional and advanced statistical techniques have been investigated too (Lee & Chen, 2005; Lee et al., 2002; Zekic-Susac et al., 2004; Malhotra & Malhotra, 2003; Ong, Huang, & Tzeng, 2005). Comparisons have also been extended to include feed-forward nets and back-propagation nets (Arminger, Enache, & Bonne, 1997; Malhotra & Malhotra, 2003). Statistical association measures showed that the neural network models are better representations of data than logistic regression and CART, (Zekic-Susac et al., 2004), while discriminant analysis, in general, has a better classification ability but worse prediction ability, whereas logistic regression has a relatively better prediction capability (Liang, 2003).

On the one hand, the use of only two groups of customer credit, either "good" or "bad" as it has been used in this paper, is appropriate with in such a new environment, such as the Egyptian banking sector, to credit scoring models, and is still one of the most important assortments in credit scoring applications (Banasik et al., 2001; Boyes et al., 1989; Kim & Sohn, 2004; Lee et al., 2002; Orgler, 1971). On the other hand, the use of three groups of consumer credit became one of the approaches for classification in credit scoring models. Some have used "good" or "bad" or "refused" (Steenackers & Goovaerts, 1989), whilst others have used "good" or "poor" or "bad" (Sarlija et al., 2004). Otherwise, the probit analysis (Banasik, Crook, & Thomas, 2003; Greene, 1998; Guillen & Artis, 1992) has been used in building credit scoring models beside other statistical techniques.

It is important for new users to apply the most appropriate technique(s) for the array of methods available, bearing in mind comparisons between different methods (Baesens et al., 2003; Bailey, 2004; Chen & Huang, 2003; Guillen & Artis, 1992; Hand & Henley, 1997; Ong et al., 2005), and the emphasis on a dichotomous variable of "good" and "bad" (Banasik et al., 2003; Chen & Huang, 2003; Desai et al., 1996; Guillen & Artis, 1992; Hand & Henley, 1997; Yang, Wang, Bai, & Zhang, 2004), in building the scoring models, especially for the new users to credit scoring models. Lim and Sohn (2007) argue that using existing models is quite troublesome to discriminate the creditability of borrowers with high default risks in the middle of the repayment term. However, with the cluster-based dynamic scoring models, the lender can identify the individual credibility at earlier stage of loan period without loosing its accuracy.

Generally, there is no overall best statistical technique/method used in building credit scoring models, for what is best depends on the details of the problem, the data structure, the characteristics used, the extent to which it is possible to segregate the classes by using those characteristics, and the objective of the classification (Hand & Henley, 1997). Most studies that made a comparison between different techniques found that, first, most recent/advanced statistical techniques such as neural networks and fuzzy algorithms are better than the traditional ones; second, there is no apparent difference between different statistical techniques in terms of the percentage of average correct classification rate. This sometimes depends on the original group that is used to compute the correct classification, depending on "bad" or "good and bad" together (Desai et al., 1996; Blochlinger & Leippold, 2006; Hoffmann, Baesens, Mues, Gestel, & Vanthienen, 2007). However, the more simple classification techniques, such as linear discriminant analysis and logistic regression, also have a very good performance in this context, which is in majority of the cases not statistically different from other techniques (Baesens et al., 2003).

The chosen environment will be the Egyptian banking sector, in which no study (in the best of our knowledge) has investigated the use of sophisticated statistical appraisal techniques in credit scoring. Indeed, from the review of literature to date, no studies were found in Egypt in covering credit scoring techniques. Therefore, we intend to cover this gap, which was found in the Egyptian banking sector.

3. Methodology and data collection

In this Section, four statistical techniques used in building credit scoring are described first. The first model is the discriminant analysis model (DA), which was first proposed by Fisher (1936) as a discrimination and classification technique. The second model is the probit analysis model (PA), which is also usually used with other statistical techniques for the purpose of comparing the results. Then, the logistic regression model (LR), unlike other conventional statistical techniques, can suit different kinds of distribution functions and is more suitable for credit scoring problems. In recent years, neural nets (NNs), one of the best statistical techniques used in building the scoring models, is regarded as a practical technology, with successful applications in many fields in financial institutions especially banks (Bishop, 1995; Masters, 1995). Here two different nets, probabilistic neural nets (PNNs) and multi-layer feed-forward nets (MLFNs) with four nodes were utilized in this research and the best net search (BNS), from multi-layer feed-forward net with two to six nodes and from probabilistic neural net, was an option selected in the current package. Later, the data collection method and the identification of variables will be discussed. Data cases in both hold-out and training samples were automatically selected by the Neural Tools software, applying 20% as a hold-out sample and 80% as a training sample.

3.1. Credit scoring models

3.1.1. Discriminant analysis

However, as to the statistical assumptions implicit in implementation, DA requires the data to be independent and normally distributed. Consequently, the general formula of DA is as follows:

$$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n,$$

where

Z represents the discriminant (zed) score, α is the intercept term, and β_i represents the respective coefficient in the linear combination of explanatory variables, X_i, for $i = 1$–n (Lee et al., 2002).

Specifically, the DA model assumes that (Desai et al., 1996):

- The independent variables are measured on an interval scale.
- There is equality of covariance matrices of the independent variables.
- The independent variables are multivariate-normal.

3.1.2. Probit analysis

PA is a technique that finds coefficient values, such that this is a probability of a unit value of a binary coefficient. As such Probit means "probability unit". Under a probit model, a linear combination of the independent variables is transformed into its cumulative probability value from a normal distribution. The method requires finding value for the coefficients in this linear combination, such that this cumulative probability equals the actual probability that the binary outcome is one, thus:

$$\text{Prob}(y = 1) = \Phi(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n),$$

where

y is the zero-one binary outcome for a given set of value. Φ is the value from the cumulative normal distribution function. α is the intercept term, and β_i represents the respective coefficient in the linear combination of explanatory variables, X_i, for $i = 1–n$.

PA is used as an alternative to LR. Early in the 1930s the term "Probit" has been developed which stands for probability unit (Maddala, 2001; Pindyck & Rubinfeld, 1997).

3.1.3. Logistic regression

LR is a widely used statistical modelling technique in which the probability of a dichotomous outcome (zero or one) is related to a set of potential predictor variables in the form:

$$\log[p/(1-p)] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n,$$

where

p is the probability of the outcome of interest, α is the intercept term, and β_i represents the respective coefficient in the linear combination of explanatory variables, X_i, for $i = 1–n$. The dependent variable is the logarithm of the odds, $\{\text{Log}\,[p/(1-p)]\}$, which is the logarithm of the ratio of two probabilities of the outcome of interest (Lee et al., 2002).

Given the set of explanatory variables, the probability of a value of one for the dichotomous outcome is (Desai et al., 1996):

$$Z = \frac{1}{1 + e^{-Z}}$$

where

Z = the probability that the dichotomous outcome is one; and

$$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

Thus, the objective of a logistic regression model in credit scoring is to determine the conditional probability of a specific observation belonging to a class, given the values of the independent variables of that credit applicant (Lee & Chen, 2005).

PA tends to be used as alternative to LR, although LR is more suited to dichotomous testing. Comparing LR with DA, the LR does not necessarily require the assumptions of DA. One advantage of DA is that the ordinary least square estimation procedure can be implemented to estimate the coefficient of the linear discriminant function, whereas the maximum likelihood method is required for the estimation of logistic regression models. Another advantage of DA over logistic regression is that prior probabilities and misclassification costs can easily be incorporated into the DA approach (Desai et al., 1996). Moreover, both DA and LR have been widely used in business, finance, science, and customer behaviour (Lee et al., 2002).

3.1.4. Neural nets

3.1.4.1. Neural net fundamentals. A system that takes numeric inputs, and outputs one or more numeric values, executing calculations on these inputs, is a neural net. Neural nets are an attempt to create nets that work in a very similar way to the human brain by setting up these nets using components that behave like the human brain. Hence, the idea of neural nets comes from the structure of the brain. In the human brain, electronic signals are carried to a neuron by a huge numbers of dendrites, and then takes place a conversion of the signals to pulses of electricity sending an axon to a number of synapses, which transfer ideas or information to the dendrites of other neurons. Therefore, a neuron may send/receive a signal to/from other neurons. As a result, neural nets consist of elements, each of which receive a number of inputs, and generate a single output. This is like the human brain (Palisade, 2005; Picton, 2000; Thomas et al., 2002).

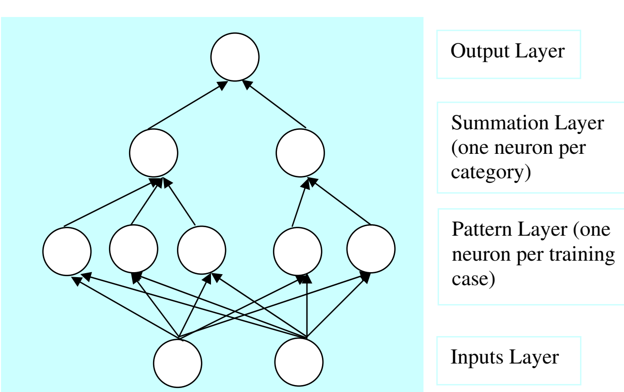
3.1.4.2. The structure of a neural net. Neural nets composed of a number of simple "nodes" or "neurons" elements, which are connected together from either a single layer or multiple layers. The basic neuron elements employed in neural nets are differing in terms of the type of net used. Each neuron executes a portion of the calculations inside the net, and then the neuron takes some numbers as inputs, performs a relatively simple computation on these inputs, and returns an output. The output value of a neuron is passed on as one of the inputs for another neuron, except for neurons that generate the final output values of the entire system (Irwin, Warwick, & Hunt, 1995; Palisade, 2005).

Neurons are arranged in layers. The input layer nodes receive the inputs for the previous calculations. These values are passed to the nodes in the first hidden (intermediate) layer, which perform computations on their inputs and pass their outputs to the next intermediate (hidden) layer, which could be another hidden layer, if there is one. The outputs from the nodes in the last intermediate layer are passed to the node or nodes that create the final outputs of the net (Irwin et al., 1995; Palisade, 2005; Trippi & Turban, 1993).

3.1.4.3. Neural nets types. Three different types of neural nets offered in the package used in this paper, probabilistic neural nets and generalized regression neural nets; which they are point-blank related, with the former used for category prediction, and the latter used for numeric prediction. Because of the categorical nature of the dependent prediction variable, the probabilistic neural nets, is only used in this research. And multi-layer feed-forward nets, basically four nodes are provided with multi-layer feed-forward nets. Besides, a range from two to six nodes are available with multi-layer feed-forward nets and probabilistic neural nets when the best net search, an option provided in the current package, is selected.

The advantage of selecting the best net search, current package tests all checked net configurations, including probabilistic neural nets and multi-layer feed-forward nets with node counts in the entered minimum-maximum range, from two to six nodes, which means more alternative models in the training and testing process.

3.1.4.3.1. Probabilistic neural nets. As an example of probabilistic neural nets structure, which assumes there are two independent numeric variables, two dependent categories, and five training cases including two cases in one category and three in the other one, is given below:



PNN structure (source: Palisade, 2005. p. 82)

An implementation of statistical techniques, called kernel discriminant analysis, in which the processes are structured into a multilayered feed-forward net with four layers, is a probabilistic neural net.

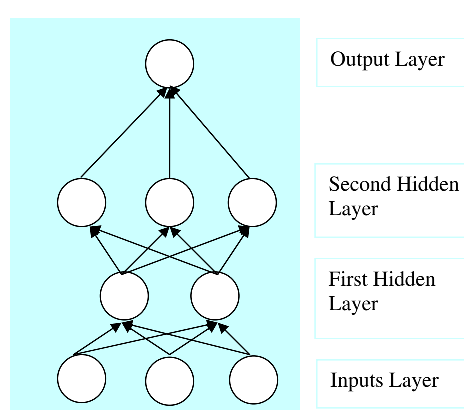
Therefore, a probabilistic neural net is predominantly a classifier mapping inputs to a number of classifications, and then might be imposed into more general function.

By introducing a case to the probabilistic neural net, each node in the first layer "pattern layer" calculate the distance between the input case and the training case reintroduced by the node. And then, the value pass to the second layer "summation layer" node, which is a function in the distance in the same time smoothing factors, taking into account that each input has its own smoothing factor. One node per dependant category/variable is in the second layer, each node sums up the output values for the nodes corresponding to the training cases in that category. The second layer output values can be interpreted as probability function predicts for each class. Finally, the category with the highest probability function value selected by the output node as the estimated category.

Probabilistic neural net training consists of two parts, optical smoothing factor and Conjugate Gradient method. Bishop (1995, p. 275–276), explains, that in finding a minimum line a search procedure, if search directions are always based on negative gradients, the search process may be very slow; indeed there is a problem, 'in which the search point (may oscillate) on successive steps'. Instead, non-interfering on conjugate directions can be chosen. A conjugate gradient algorithm can be usually employed, dressing in the work by Hestenes and Stiefel (1952), for example. The conjugate gradient algorithm provides a minimization technique, which requires only the evaluation of the error function and its derivatives and which, for a quadratic error function is guaranteed to find a certain number of steps (Bishop, 1995, p. 282) .

3.1.4.3.2. Multi-layer feed-forward nets. In situations of complex relationships between variables, it may be advisable to model a system using multi-layer feed-forward nets (multi-layer perceptron networks).

An example of multilayer feed-forward architecture is given below for three inputs, which are classified as numeric, and one output, between which these is a first hidden layer of two nodes and a second hidden layer of three nodes.



MLF nets architecture (source: Palisade, 2005. p. 73)

Palisade (2005) explains the behaviour of the net which is determined by:

The structure of the net in terms of numbers of nodes and hidden layers; parameters associated with connections and neurons; and conversion functions for each neuron, which map inputs to outputs.

The output at a given level (layer) may be expressed as a connection-weighted summation of outputs from a previous level (layer) plus a neuron-bias. A sigmoid function, which is also employed in a logistic regression, is sometimes used in neural nets. However, in the Neural Tools software the sigmoid function is not utilized. The reason is to avoid a restriction on outputs values, to create a superb model for training purposes (Palisade, 2005).

Particular attributes of multi-layer feed-forward nets include reliability out side the training data range, compactness in size, an excellent classifier, and with a capability to generalize results from small training data. By contrast, probabilistic neural nets are particularly fast, they do not require a number of hidden layers and nodes, they have a parallel structure, and they classify and return probabilities for different dependent categories, and guarantee convergence to the optimal case (Masters, 1995; Palisade, 2005).

3.2. Data collection and proposed variables

In order to build the proposed five credit scoring models, a personal loans data-set was provided by one of the commercial banks in Egypt. This consists of 581 personal loans with 433 good loans and 148 bad loans. It should be emphasized that this dataset is pertinent because of the large number of bad loans (25.5%) with good loans (74.5%). Each bank customer in this data-set is linked to 20 independent variables (see Appendix A for details), in addition to the dependent variable, which is loan quality explained by two values, good/paid = 1 and bad/defaulted = 0. Some variables had identical values for all cases and hence were excluded, e.g., loan duration was four years in all cases, and all customers had a credit card.

Selected variables for the proposed models were reduced to 12 variables, as shown in Appendix A. In addition, all clients must have an investigation report from the Central Bank of Egypt, which provides a comprehensive history of the clients' dealings with all banks in Egypt.

4. Results

In order to run the proposed models, STATGRAPHICS Plus 5.1, SPSS 14.00 and Neural Tools[3] software were used in this paper. The detailed credit scoring results using the above-mentioned five modelling techniques can be summarized as follows. Because of the high correlation between the loan amount and monthly salary, 0.963, an Orthogonalisation test has been used to keep the effect of both in the proposed models because of their potential importance. The revised correlation, after running the test, was 0.269; all other variables had correlations within an acceptable range.

4.1. Discriminant analysis

DA credit scoring models were designed to develop a set of discriminating functions, which can help predict the dependent variable. All the 12 predicted variables were entered. The one discriminating function with a *P*-value of 0.0000 was statistically significant at the 95% confidence level.

From the results revealed in Table 1, it can be observed that the average correct classification rate is 86.75%, depending on 0.5 prior probabilities for groups. Again a stepwise discriminant approach (Johnson & Wichern, 2002; Lee et al., 2002; Neter, Kutner, Wasserman, & Nachtsheim, 1996) was adopted in building the DA scoring model (which we call DA_1). The stepwise approach was run on a forward basis, entering at each step the variable that minimizes the overall Wilks' lambda. The minimum partial F to enter was 3.84, and the minimum partial F to remove was 2.71. Prior probabilities were used treating all groups equally, and the covariance matrix was applied 'within groups'. Nine significant predictor variables are selected in the final model (discriminant function), LOAN AMO, COR GUAR, TELE, LFOB, AGE, MAR STA, EDU, HOR, and SALA. From Table 1, 86.92% was observed as the average correct classification rate.

4.2. Probit analysis

PA credit scoring models were developed to describe the relationship between the dependent variable (LOAN QUA) and 12 independent variables. Because the *P*-value for the model in the analysis of deviance table (Appendix B) is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level. In addition, the *P*-value for the residuals is greater than or equal to 0.10, indicating that the model is not significantly worse

Table 1
Classification results using the DA and DA_1

Observed group	Predicted group			
DA	Good	Bad	Total	Overall%
Good	372	61	433	85.91
Bad	16	132	148	89.19
Total	388	193	581	86.75
DA_1				
Good	372	61	433	85.91
Bad	15	133	148	89.86
Total	387	194	581	86.92

Cut-off point 0.50.

[3] Neural Tools Professional, provided by Palisade Europe Corporation UK.

Table 2
Classification results using the PA and PA_1

Observed group	Predicted group			
PA	Good	Bad	Total	Overall%
Good	407	26	433	94.00
Bad	45	103	148	69.59
Total	452	129	581	87.78
PA_1				
Good	403	30	433	93.07
Bad	44	104	148	70.27
Total	447	134	581	87.26

Cut-off point 0.50.

Table 3
Classification results using the LR and LR_1

Observed group	Predicted group			
LR	Good	Bad	Total	Overall%
Good	407	26	433	94.00
Bad	42	106	148	71.62
Total	449	132	581	88.30
LR_1				
Good	406	27	433	93.76
Bad	43	105	148	70.95
Total	449	132	581	87.95

Cut-off point 0.50.

than the best possible model for this data at the 90% or higher confidence level, as it is shown in Appendix B.

All selected variables were significant at the 95% confidence level except three variables: ADD INC, SEX, and COMP.[4] But because of their potential importance we kept them in the model. Table 2 reveals an 87.78% average correct classification rate for this model using a 50% cut-off point. Nevertheless, the highest correct classification per cent was found using a 65% cut-off point, which is 89.33%.

Hence, we ran the model again, without ADD INC, SEX and COMP (calling this the PA_1 model). All included variables were significant, and an 87.26% average correct classification rate was observed with a cut-off of 50% as it is shown in Table 2. The highest average correct classification rate at 88.81%, using a 60% cut-off point, was found.

4.3. Logistic regression

Table 3 summarizes the results of the LR credit scoring model, using the original 12 predictor variables. It can be observed that the average correct classification rate was 88.30% with a 0.5 cut-off point. Because the *P*-value, (see Appendix B), for the model is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level. In addition, the *P*-value for the residuals is greater than or equal to 0.10, indicating that the model is not significantly worse than the best possible model for this data at the 90% or higher confidence level. The highest correct classification rate was 89.85% with a 0.60 cut-off point.

Actually, three variables were not significant at the 95% confidence level: ADD INC, SEX, and COMP.[5] The model was run again (which we called model LR_1) without ADD INC, SEX and COMP; all predictor variables were significant at the 95% confidence level. The average correct classification rate as it is shown in Table 3 was 87.95% with a 0.50 cut-off point, and 89.16% with a 0.60 cut-off point. Appendix C summarizes the PA, PA_1, LR, LR_1 different cut-offs, and their average correct classification rates (this option was not available using discriminant analysis, the standard cut-off being 0.50 only in SPSS 14.0 and STATGRAPHICS Plus 5.1).

4.4. Neural nets

In this paper, we apply a simple validation technique by dividing the data-set into training sample (80%, 465 cases) and a hold-out sample (20%, 116 cases) that tests the predictive effectiveness of the fitted model. To study the overall predictive capability of the classification models, we used the whole data-set as a test set. The experiment was repeated 20 times with a different hold-out (testing) subsample each time and the remaining data-set was the training sample. The reason for repeating the process was to investigate whether different results, in terms of average correct classification rate, was being achieved because of the random selection procedure as part of the software design. Actually, in our analysis we found significant differences between the various neural nets models, which we describe below.

4.4.1. Probabilistic neural nets

Table 4 summarizes the classification results of the PNN credit scoring models for the hold-out (testing), training and overall samples. A 96.21% average correct classification rate has been found with PNN_6. Besides, four different models which have the highest average correct classification rates have also been selected for comparison with the other NN models. The highest average correct classification rate in the hold-out (testing) sample was 90.52% with PNN_6 and PNN_1. Meanwhile the highest average correct classification rate in the training sample was 98.49% with PNN_4 with an overall average correct classification rate of 94.49%.

It can be observed from Table 4 that all PNNs predict the good credit much better than the bad credit in all samples (hold-out, training, and overall). Also, the highest bad predictor in the overall sample was 90.54% and 95.83% in

[4] In addition to HOR with a *P*-value of 0.1002, but after excluding the three variables became significant with a *P*-value of 0.0179.

[5] In addition to HOR with a *P*-value of 0.1695 but after excluding just the ADD INC, the *P*-value of HOR became 0.0429 and 0.0275 after excluding the ADD INC, SEX and COMP.

Table 4
Classification results for the 20 probabilistic neural nets

PNN trial	Hold-out sample (testing sample)			Training sample			Overall sample		
	Good%	Bad%	Overall%	Good%	Bad%	Overall%	Good%	Bad%	Overall%
PNN	91.58	76.19	88.79	97.04	92.13	95.70	95.84	89.86	94.32
PNN_1*	94.25	79.31	90.52	97.98	89.92	95.91	97.23	87.84	94.84
PNN_2	91.21	84.00	89.66	97.37	85.37	94.19	96.07	85.14	93.29
PNN_3	87.10	82.61	86.21	97.94	91.20	96.13	95.61	89.86	94.15
PNN_4	84.71	61.29	78.45	99.14	96.58	98.49	96.30	89.19	94.49
PNN_5	90.24	73.53	85.34	99.15	85.96	95.91	97.46	83.11	93.80
PNN_6*	93.67	83.78	90.52	99.72	90.99	97.63	98.61	89.19	96.21
PNN_7	88.64	67.86	83.62	97.97	90.83	96.13	96.07	86.49	93.63
PNN_8	94.19	60.00	85.34	98.56	88.98	96.13	97.69	83.11	93.98
PNN_9	86.59	76.47	83.62	99.43	89.47	96.99	97.00	86.49	94.32
PNN_{10}*	94.19	76.67	89.66	99.42	88.98	96.77	98.38	86.49	95.35
PNN_{11}	89.77	71.43	85.34	97.68	85.00	94.41	96.07	82.43	92.60
PNN_{12}	93.83	74.29	87.93	98.86	87.61	96.13	97.92	84.46	94.49
PNN_{13}*	94.05	62.50	85.34	99.71	91.38	97.63	98.61	85.14	95.18
PNN_{14}*	89.77	67.86	84.48	98.55	95.83	97.85	96.77	90.54	95.18
PNN_{15}	90.59	74.19	86.21	98.28	83.76	94.62	96.77	81.76	92.94
PNN_{16}	95.12	67.65	87.07	96.87	87.72	94.62	96.54	83.11	93.12
PNN_{17}	89.66	58.62	81.90	100.00	91.60	97.85	97.92	85.14	94.66
PNN_{18}	94.87	55.26	81.90	98.59	89.09	96.34	97.92	80.41	93.46
PNN_{19}	90.59	67.74	84.48	98.56	88.89	96.13	97.00	84.46	93.80

* Best five PNN.

the training sample for PNN_{14}; whilst it was 84.00% in the testing sample for PNN_2.

4.4.2. Multi-layer feed-forward nets

Following the same methodology which is used in PNNs, MLFN models have been run 20 times to investigate the expected difference between the proposed models in terms of average correct classification rates. Table 5 shows the classifications results for the hold-out (testing), training and overall samples of the MLFN with only four nodes. A 93.98% average correct classification rate has been found with $MLFN_8$. Again we select the best five models to compare with the other NN models. It can be observed that the highest average correct classification rate in the testing sample was 86.21% with $MLFN_2$, $MLFN_6$ and $MLFN_8$. Otherwise, a 95.91% average correct classification rate was the

Table 5
Classification results for the 20 multi-layer feed-forward nets

MLFN trial	Hold-out sample (testing sample)			Training sample			Overall sample		
	Good%	Bad%	Overall%	Good%	Bad%	Overall%	Good%	Bad%	Overall%
MLFN	88.51	65.52	82.76	96.53	91.60	95.27	94.92	86.49	92.77
$MLFN_1$*	82.95	78.57	81.90	97.10	92.50	95.91	94.23	89.86	93.12
$MLFN_2$	91.21	68.00	86.21	97.37	86.99	94.62	96.07	83.78	92.94
$MLFN_3$*	95.06	60.00	84.48	98.86	85.84	95.70	98.15	79.73	93.46
$MLFN_4$	79.52	75.76	78.45	94.00	94.78	94.19	91.22	90.54	91.05
$MLFN_5$	87.78	73.08	84.48	93.59	95.08	93.98	92.38	91.22	92.08
$MLFN_6$	93.18	64.29	86.21	99.13	82.50	94.84	97.92	79.05	93.12
$MLFN_7$	81.71	70.59	78.45	97.15	92.11	95.91	94.23	87.16	92.43
$MLFN_8$*	92.94	67.74	86.21	98.28	88.89	95.91	97.23	84.46	93.98
$MLFN_9$	90.12	54.29	79.31	98.86	86.73	95.91	97.23	79.05	92.60
$MLFN_{10}$	86.21	72.41	82.76	95.38	91.60	94.41	93.53	87.84	92.08
$MLFN_{11}$*	88.75	77.78	85.34	96.88	90.18	95.27	95.38	87.16	93.29
$MLFN_{12}$	86.90	81.25	85.34	94.84	93.97	94.62	93.30	91.22	92.77
$MLFN_{13}$	90.00	63.89	81.90	98.87	86.61	95.91	97.23	81.08	93.12
$MLFN_{14}$	87.95	78.79	85.34	97.43	86.96	94.84	95.61	85.14	92.94
$MLFN_{15}$	83.33	61.54	78.45	94.75	94.26	94.62	92.38	88.51	91.39
$MLFN_{16}$	97.44	57.89	84.48	99.15	77.27	93.98	98.85	72.30	92.08
$MLFN_{17}$	90.11	68.00	85.34	96.49	86.99	93.98	95.15	83.78	92.25
$MLFN_{18}$	88.75	63.89	81.03	97.73	89.29	95.70	96.07	83.11	92.77
$MLFN_{19}$*	87.80	76.47	84.48	98.01	86.84	95.27	96.07	84.46	93.12

* Best five MLFN.

Table 6
Classification results for the 20 best net searches

BNS trial	Hold-out sample (testing sample)			Training sample			Overall sample		
	Good%	Bad%	Overall%	Good%	Bad%	Overall%	Good%	Bad%	Overall%
BNS-PNN*	88.24	74.19	84.48	98.85	93.16	97.42	96.77	89.19	94.84
BNS_1–MLFN–5N	89.53	73.33	85.34	97.41	92.37	96.13	95.84	88.51	93.98
BNS_2–MLFN–6N	91.95	82.76	89.66	95.09	90.76	93.98	94.46	89.19	93.12
BNS_3–MLFN–4N	93.67	70.27	86.21	95.48	93.69	95.05	95.15	87.84	93.29
BNS_4–MLFN–3N	93.02	73.33	87.93	96.54	92.37	95.48	95.84	88.51	93.98
BNS_5–MLFN–2N	88.17	69.57	84.48	95.29	87.20	93.12	93.76	84.46	91.39
BNS_6–MLFN–6N	89.13	75.00	86.21	97.36	94.35	96.56	95.61	91.22	94.49
BNS_7–PNN*	91.76	80.65	88.79	98.85	89.74	96.56	97.46	87.84	95.01
BNS_8–PNN	96.20	62.16	85.34	97.46	88.29	95.27	97.23	81.76	93.29
BNS_9–PNN	91.95	75.86	87.93	97.69	84.03	94.19	96.54	82.43	92.94
BNS_{10}–MLFN–6N	90.00	80.56	87.07	96.60	94.64	96.13	95.38	91.22	94.32
BNS_{11}–MLFN–2N	96.47	67.74	88.79	95.69	79.49	91.61	95.84	77.03	91.05
BNS_{12}–MLFN–2N	91.21	72.00	87.07	96.49	84.55	93.33	95.38	82.43	92.08
BNS_{13}–MLFN–3N	89.77	89.29	89.66	95.07	93.33	94.62	94.00	92.57	93.63
BNS_{14}–PNN	96.51	73.33	90.52	97.41	88.98	95.27	97.23	85.81	94.32
BNS_{15}–MLFN–5N*	88.64	67.86	83.62	97.97	96.67	97.63	96.07	91.22	94.84
BNS_{16}–PNN	93.41	68.00	87.93	96.20	86.18	93.55	95.61	83.11	92.43
BNS_{17}–MLFN–5N*	95.45	71.43	89.66	97.97	90.00	95.91	97.46	86.49	94.66
BNS_{18}–MLFN–5N*	89.04	79.07	85.34	96.94	98.10	97.20	95.61	92.57	94.84
BNS_{19}–MLFN–3N	96.77	69.57	91.38	97.94	77.60	92.47	97.69	76.35	92.25

* Best five BNS.

highest in the training sample with $MLFN_1$, $MLFN_7$, $MLFN_8$, $MLFN_9$ and $MLFN_{13}$.

We see from Table 5 that all the MLFNs predict the good credit better than the bad credit in all samples, except two models in the training sample, $MLFN_4$ and $MLFN_5$. Regarding these exceptions, the bad creditor using $MLFN_4$ was 94.78%, whilst the good creditor was 94.00%; and it was 95.08% as a bad predictor in $MLFN_5$, whilst the good predictor was 93.59%. Also, the highest bad predictor in the overall sample was 91.22% in both $MLFN_5$ and $MLFN_{12}$ and 95.08%, 94.78% for $MLFN_5$ and $MLFN_4$, respectively, in the training sample. Correspondingly, the highest bad predictor in the testing sample was 81.25% for $MLFN_{12}$.

Furthermore, there were seven models producing high average correct classification rate, but we omitted two ($MLFN_6$, and $MLFN_{13}$) which had the same average correct classification rate as $MLFN_1$ and $MLFN_{19}$, but worse bad predictor rates.

4.4.3. Best net search

MLFN using two to six nodes was an option, under the best net search, which we investigated. So, with PNN as well, we had six models, from which the software selected the best one. Classification results for the 20 BNSs are shown in Table 6. It can be observed that the average correct classification rate was 95.01% with BNS_7–PNN and a 94.84% with both BNS_{15}–MLFN–5N[6] and BNS_{18}–MLFN–5N. A 91.38% average correct classification rate was found in the testing sample with BNS_{19}–MLFN–3N, while a 97.63% average correct classification rate was observed in the training sample with BNS_{15}–MLFN–5N.

We see from Table 6 that all the BNS models predict the good credit better than the bad credit, as well, except only one model in the training sample, which is BNS_{18}–MLFN–5N. In this case, the bad credit was 98.10%, while the good credit was 96.94%. Besides, this was the highest bad predictor in the training sample. Moreover, the highest bad predictor was 92.57% for both BNS_{18}–MLFN–5N and BNS_{13}–MLFN–3N in the overall sample; whilst it was 89.29% in the testing sample for BNS_{13}–MLFN–3N.

4.5. Comparison of results of different credit scoring models[7]

Since the average correct classification rate became an important criterion/tool in evaluating the classification capability of the scoring models, it was important to compare the different models' results. The classification results for all proposed models are compared in order to evaluate these models. Table 7 summarizes the average correct classification rate results for conventional techniques (DA, DA_1, PA, PA_1, LR and LR_1), and the best 5 models from PNN, the best 5 models from MLFN and the best 5 models from BNS.

It can be concluded from Table 7 that LR has the highest average correct classification rates, which is 88.30%, amongst the conventional techniques. Meanwhile PNN_6 has the highest average correct classification rate, which

[6] BNS_{15}–MLFN–5N means trial number 15 in best net search with multi-layer feed-forward net selecting 5 nodes as a best net.

[7] The conventional models compared in this section depend on the observed results, using a 0.50 cut-off point only.

Table 7
Comparing classification results for different techniques

Scoring model	Classification results (overall sample)		
	Good%	Bad%	Overall%
DA	85.91	89.19	86.75
DA_1	85.91	89.86	86.92
PA	94.00	69.59	87.78
PA_1	93.07	70.27	87.26
LR*	94.00	71.62	88.30
LR_1	93.76	70.95	87.95
PNN_1	97.23	87.84	94.84
PNN_6****	98.61	89.19	96.21
PNN_{10}	98.38	86.49	95.35
PNN_{13}	98.61	85.14	95.18
PNN_{14}	96.77	90.54	95.18
$MLFN_1$	94.23	89.86	93.12
$MLFN_3$	98.15	79.73	93.46
$MLFN_8$**	97.23	84.46	93.98
$MLFN_{11}$	95.38	87.16	93.29
$MLFN_{19}$	96.07	84.46	93.12
BNS-PNN	96.77	89.19	94.84
BNS_7-PNN***	97.46	87.84	95.01
BNS_{15}–MLFN–5N	96.07	91.22	94.84
BNS_{17}-MLFN–5N	97.46	86.49	94.66
BNS_{18}–MLFN–5N	95.61	92.57	94.84

* Best conventional technique.
** Best MLFN.
*** Best BNS with PNN.
**** Best PNN and best of all techniques.

is 96.21%, amongst all techniques. All models predict the good credit better than the bad credit, except only two models namely, DA and DA_1. In addition, the highest bad predictor was 92.57% for BNS_{18}–MLFN–5N, whilst the highest good predictor was 98.61% for both PNN_6 and PNN_{13}.

As shown in Table 7, on average the overall performance of the NNs is much better than the average performance of the conventional techniques.

For the purpose of comparing results of all models developed in this paper, and in order to evaluate the overall credit scoring capability and effectiveness, the misclassification costs have been taken into account, beside the average correct classification rate, in order to find the minimum expected misclassification cost in a credit scoring model (West, 2000).

The following equation is used in computing the estimated misclassification cost:

$$\text{Estimated cost} = C(B/G) \times P(B/G) \times \pi_1 + C(G/B) \times P(G/B) \times \pi_0$$

where,

$C(B/G)$, i.e., C (predicted bad/actually good) and $C(G/B)$, i.e., C (predicted good/actually bad), are the corresponding misclassification costs of both type I and type II errors. $P(B/G)$ and $P(G/B)$ measure the probabilities of type I and type II errors. π_1 and π_0, are the prior probabilities of good and bad, respectively (West, 2000).

It is a complicated and challenging task to provide reliable estimates of the misclassification costs, therefore valid prediction might not be available, especially in an environment such as the Egyptian banking sector. However, it is generally believed in a credit scoring application that the costs associated with both type I and type II errors are significantly different. Generally, the misclassification cost associated with a type II error is much higher than the misclassification cost associated with a type I error (Lee & Chen, 2005).

West (2000) noted that Dr Hofmann, who compiled his German credit data, reported that the ratio of misclassification costs associated with type II and type I is 5:1.

In this paper, this relative cost ratio will be used to calculate the estimated misclassification cost for the proposed models.[8] The prior probabilities of good and bad credit are set as 74.5% and 25.5%, respectively, using the ratio of good and bad credit in the Egyptian data-set.

Table 8 concludes the type I,[9] type II[10] errors and the estimated misclassification costs for all proposed models. In general, the misclassification error associated with type II are higher than those associated with type I, which is also true in other case studies based on credit card and housing loans datasets (Lee & Chen, 2005; Lee et al., 2002).

On the one hand, comparing conventional techniques, our results are consistent with the above analysis using probit and logistic models namely, PA, PA_1, LR and LR_1, while the discriminant models did not agree with them. The discriminant models, DA and DA_1, predicted bad credits much better than the other models did. The reason is that the type I errors in the discriminant models are higher than the type II errors. By contrast, PA, PA_1, LR and LR_1 predicted good credits much better than the DA and DA_1. Accordingly, the type I errors in the last four conventional models are lower than the type II errors.

Furthermore, where the type I error rate exceeds the type II error rate, as in the case of DA and DA_1, the lower misclassification cost at 0.2343 is for DA_1. Also, we know that the average correct classification rate criterion led to selecting DA_1 at 86.92%, (see Table 7). Correspondingly, where the type II error rate exceeds the type I error rate, as for PA, PA_1, LR and LR_1, the lowest misclassification cost at 0.4065 is for LR. This is also the chosen model between PA, PA_1, LR and LR_1, for LR has the highest correct classification rate at 88.30% (see Table 7).

[8] Misclassification costs have been calculated for all models including all trials. We suggest at this stage that the lowest misclassification cost might be found in a model that does not have the highest average correct classification rate.
[9] Good credit is misclassified as bad credit.
[10] Bad credit is misclassified as good credit.

Table 8
Errors and estimated misclassification costs for all the proposed models

Credit scoring model	Error results		Estimated misclassification cost	Credit scoring model	Error results		Estimated misclassification cost	Credit scoring model	Error results		Estimated misclassification cost
	Type I	Type II			Type I	Type II			Type I	Type II	
DA	0.1409	0.1081	0.2428	PA	0.0600	0.3041	0.4324	LR	0.0600	0.2838	0.4065
DA$_{1}$*	0.1409	0.1014	0.2343	PA$_{1}$	0.0693	0.2973	0.4307	LR$_{1}$	0.0624	0.2905	0.4169
PNN	0.0416	0.1014	0.1603	MLFN	0.0508	0.1351	0.2101	BNS–PNN	0.0323	0.1081	0.1619
PNN$_{1}$	0.0277	0.1216	0.1757	MLFN$_{1}$	0.0577	0.1014	0.1723	BNS$_{1}$–MLFN–5N	0.0416	0.1149	0.1775
PNN$_{2}$	0.0393	0.1486	0.2187	MLFN$_{2}$	0.0393	0.1622	0.2361	BNS$_{2}$–MLFN–6N	0.0554	0.1081	0.1791
PNN$_{3}$	0.0439	0.1014	0.1620	MLFN$_{3}$	0.0185	0.2027	0.2722	BNS$_{3}$–MLFN–4N	0.0485	0.1216	0.1912
PNN$_{4}$	0.0370	0.1081	0.1654	MLFN$_{4}$	0.0878	0.0946	0.1860	BNS$_{4}$–MLFN–3N	0.0416	0.1149	0.1775
PNN$_{5}$	0.0254	0.1689	0.2343	MLFN$_{5}$	0.0762	0.0878	0.1687	BNS$_{5}$–MLFN–2N	0.0624	0.1554	0.2446
PNN$_{6}$	0.0139	0.1081	0.1482	MLFN$_{6}$	0.0208	0.2095	0.2826	BNS$_{6}$–MLFN–6N	0.0439	0.0878	0.1447
PNN$_{7}$	0.0393	0.1351	0.2015	MLFN$_{7}$	0.0577	0.1284	0.2067	BNS$_{7}$–PNN	0.0254	0.1216	0.1740
PNN$_{8}$	0.0231	0.1689	0.2326	MLFN$_{8}$	0.0277	0.1554	0.2188	BNS$_{8}$–PNN	0.0277	0.1824	0.2532
PNN$_{9}$	0.0300	0.1351	0.1946	MLFN$_{9}$	0.0277	0.2095	0.2877	BNS$_{9}$–PNN	0.0346	0.1757	0.2498
PNN$_{10}$	0.0162	0.1351	0.1843	MLFN$_{10}$	0.0647	0.1216	0.2032	BNS$_{10}$–MLFN–6N	0.0462	0.0878	0.1464
PNN$_{11}$	0.0393	0.1757	0.2533	MLFN$_{11}$	0.0462	0.1284	0.1981	BNS$_{11}$–MLFN–2N	0.0416	0.2297	0.3239
PNN$_{12}$	0.0208	0.1554	0.2136	MLFN$_{12}$*	0.0670	0.0878	0.1619	BNS$_{12}$–MLFN–2N	0.0462	0.1757	0.2584
PNN$_{13}$	0.0139	0.1486	0.1998	MLFN$_{13}$	0.0277	0.1892	0.2619	BNS$_{13}$–MLFN–3N	0.0600	0.0743	0.1394
PNN$_{14}$*	0.0323	0.0946	0.1447	MLFN$_{14}$	0.0439	0.1486	0.2222	BNS$_{14}$–PNN	0.0277	0.1419	0.2016
PNN$_{15}$	0.0323	0.1824	0.2566	MLFN$_{15}$	0.0762	0.1149	0.2033	BNS$_{15}$–MLFN–5N	0.0393	0.0878	0.1412
PNN$_{16}$	0.0346	0.1689	0.2411	MLFN$_{16}$	0.0115	0.2770	0.3617	BNS$_{16}$–PNN	0.0439	0.1689	0.2481
PNN$_{17}$	0.0208	0.1486	0.2050	MLFN$_{17}$	0.0485	0.1622	0.2429	BNS$_{17}$–MLFN–5N	0.0254	0.1351	0.1912
PNN$_{18}$	0.0208	0.1959	0.2653	MLFN$_{18}$	0.0393	0.1689	0.2446	BNS$_{18}$–MLFN–5N*	0.0439	0.0743	0.1274
PNN$_{19}$	0.0300	0.1554	0.2205	MLFN$_{19}$	0.0393	0.1554	0.2274	BNS$_{19}$–MLFN–3N	0.0231	0.2365	0.3187

* Models associated with the lowest estimated misclassification costs for each technique.

On the other hand, all the neural nets models' type II errors were higher than type I errors. The lowest misclassification cost at 0.1447 is for PNN_{14} amongst all the PNN models. That was not the chosen model, according to the average correct classification rate, which is PNN_6 at 96.21% average correct classification rate (see Table 7). While the misclassification cost is 0.1619 for $MLFN_{12}$. Again that was not the chosen model according to the average correct classification rate. As to the MLFNs the chosen model was $MLFN_8$ at 93.98% average correct classification rate (see Table 7). Finally, the lowest misclassification cost using the BNS is 0.1274 for BNS_{18}–MLFN–5N, gives a 94.84% average correct classification rate, but this was not the highest average correct classification rate amongst all BNS models. The highest average correct classification rate in this case was for BNS_7-PNN at 95.01% (see Table 7).

Comparing all the techniques, the lowest misclassification cost criterion leads to selecting BNS_{18}–MLFN–5N, which is the best net search selecting multi-layer feed-forward net with 5 nodes, with a minimum cost of 0.1274. However, this does not provide the highest average correct classification rate, which was 96.21% for PNN_6. Correspondingly, we do suggest that the average correct classification rate is more reliable, while the misclassification costs calculated in this paper is more subjective.

There is evidence of significant differences between the neural nets models in Group 1, and between the neural nets and the conventional techniques in Group 2, which is an overlapping, group encapsulating Group 1. As it shown in Table 9, the ANOVA F-ratio was 12.73 and 83.18 for Group 1 models and Group 2 models, respectively. These were significant at 99% confidence level. Besides, all the neural net models namely, PNN, MLFN, BNS and conventional techniques (CON. TE) namely, DA, PA and LR are significantly different at 95% confidence level as revealed by Fisher's least significant difference test. However, there were statistically significant differences in variances between those within Group 1, and also between those within Group 2 according to the Cochran's C/Bartlett's/Levene's tests. Moreover, the Kruskal–Wallis Median Test Statistic shows statistically significant differences at 99% confidence level for Group 1 and Group 2 with test Statistics 19.8774 and 32.5968, respectively, which means that the average correct classification rates are significantly different in each proposed technique.

For more statistical details relating to Group 1 and 2, the reader is referred to Figs. 1 and 2, respectively.

5. Conclusion and area of future research

There has been enormous interest over the recent decades in the use of credit scoring for evaluating credit risk in the banking sector. Within a competitive environment for financial institutions, including banks, credit scoring

Table 9
A comparative statistical evaluation of the proposed models

	Group 1: NN models only				(Overlapping) Group 2: NN + conventional techniques				
	PNN(0)	MLFN(1)	BNS(2)	Overall	PNN(0)	MLFN(1)	BNS(2)	CON. TE.(3)	Overall
Count	20	20	20	60	20	20	20	6	66
Average (Mean)	94.1905	92.668	93.5375	93.4653	94.1905	92.668	93.5375	87.4933	92.9224
Standard deviation	0.902782	0.697979	1.20332	1.13189	0.902782	0.697979	1.20332	0.612721	2.04564
ANOVA F-ratio	–	–	–	12.73***	–	–	–	–	83.18***
Fisher's least significant difference test:									
PNN-MLFN	–	–	–	1.5225**	–	–	–	–	1.5225**
PNN-BNS	–	–	–	0.6530**	–	–	–	–	0.6530**
MLFN-BNS	–	–	–	−0.8695**	–	–	–	–	−0.8695**
PNN-CON.TE	–	–	–	–	–	–	–	–	6.69717**
MLFN-CON.TE	–	–	–	–	–	–	–	–	5.17467**
BNS-CON.TE	–	–	–	–	–	–	–	–	6.04417**
Cochran's C test:	–	–	–	0.526505*	–	–	–	–	0.463265**
Bartlett's test:	–	–	–	1.10247*	–	–	–	–	1.11951*
Levene's test:	–	–	–	3.42494**	–	–	–	–	2.74414*
Kruskal–Wallis median test statistic									
Average rank	41.575	17.275	32.65	–	47.575	23.275	38.650	3.50	–
Test statistic	–	–	–	19.8774***	–	–	–	–	32.5968***

*, **, and *** denotes a statistically significant difference at 10, 5 and 1% level, respectively.

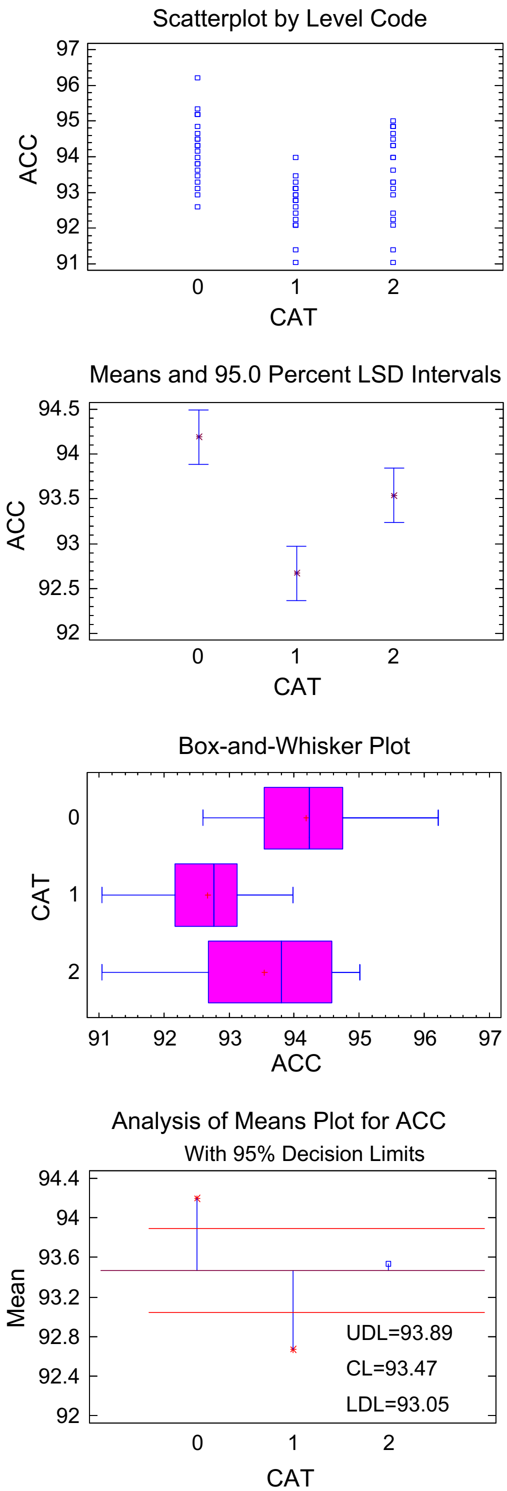


Fig. 1. Scatterplot by level code; means and 95.0% LSD intervals; Box-and-whisker plot and analysis of means plot for ACC with 95% decision limits. *Notation:* ACC = average correct classification; CAT = category.

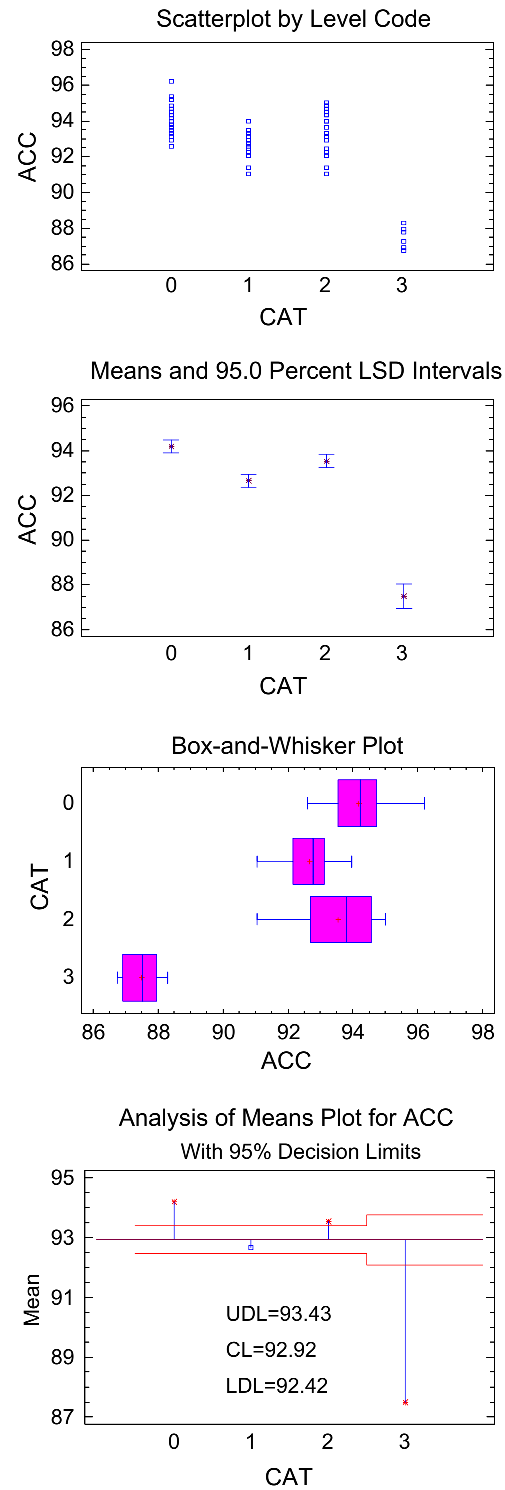


Fig. 2. Scatterplot by level code; means and 95.0% LSD intervals; Box-and-whisker plot and analysis of means plot for ACC with 95% decision limits. *Notation:* ACC = average correct classification; CAT = category.

techniques have become one of the most important tools currently used in the credit risk evaluation of loans. Besides, credit scoring is regarded as one of the basic applications of misclassification problems that have attracted more and more attention during the past decades. This paper presents an evaluation of personal loans to help strengthen the credit risk evaluation process in the Egyptian banking sector using four credit scoring statistical techniques: DA, PA, LR and NNs.

The ranking of the models varied according to the decision criterion. Using the highest average correct classification rate, PNN_6 is preferred, whereas using the lowest estimated misclassification cost, BNS_{18}–MLFN–5N is the best model. The final choice depends on the bank's decision maker's viewpoint.

The motivation behind this paper was to evaluate the relative performance of particular neural nets, such as PNNs and MLFNs, versus conventional techniques, such as DA, PA and LR, in the Egyptian banking sector. In our evaluation process, we utilized ANOVA for testing differences in mean correct classification rates of groups with and without the inclusion of conventional techniques. There were strong significant differences at the 99% confidence level. Fisher's least significant difference tests also revealed all the NNs were different from each other at the 95% confidence level. There were also significant differences in the variances of the classification rates, and in the medians, following the Kruskal–Wallis test.

Some of the predictor variables have not normally been used in published studies of credit scoring models, for example: corporate guarantee and loans from other banks. They are particularly appropriate within the Egyptian environment.

Future studies should aim to use other advanced statistical scoring techniques, such as genetic algorithms, besides the neural nets and traditional scoring models which were used in the current paper, and perhaps integrated with other techniques, such as fuzzy discriminant analysis. In addition to this, the plan is to collect more data and employ more variables that might increase the accuracies of the scoring models. Finally, future research would use more than one bank's data-set.

Appendix A. List of variables used in building the proposed credit scoring models

Variable/description	Code	Unit	Comment
X_1 Loan amount*	LOAN AMO	No.	–
X_2 Loan duration	–	–	Loan duration is 4 years in all cases in this sample.
X_3 Company*	COMP	10, 01, 00	10 = Public sector, 01 = Local private sector, 00 = Multinational company.
X_4 Branch	–	–	The bank has a branch to serve and collect instalments (i.e., clients work or live in a very remote area that there is no branch in the city).
X_5 Sex*	SEX	0, 1	0 = Male, 1 = Female
X_6 Marital status*	MAR STA	0, 1	0 = Married, 1 = Single
X_7 Age*	AGE	Years	Clients ages from 25 to 59 years.
X_8 Monthly salary*	SALA	No.	–
X_9 Additional income*	ADD INC	0, 1	0 = N/A, 1 = Suitable
X_{10} House owned or rented*	HOR	0, 1	0 = Rented, 1 = Owned
X_{11} House rent > loan tenure	–	–	The client must have a rent contract for 4 years or higher to be greater than loan tenure (4 years).
X_{12} Home telephone*	TELE	0, 1	0 = N/A, 1 = Ok confirmed (land line).
X_{13} Utility bill	–	–	Clients must have a utility bill not less than 6 months.
X_{14} Title/position	–	–	It means the occupation of customers: workers is less grade than white collar, workers are not accepted.
X_{15} Education level*	EDU	0, 1	0 = University, 1 = Higher education 100% university or higher, it is a must.
X_{16} Loans from other banks*	LFOB	0, 1	0 = N/A, 1 = Nil
X_{17} Relation with other banks	–	–	Through an investigation report from the central bank of Egypt (provides the client's history).
X_{18} Credit card status	–	–	All clients have valid credit card(s).
X_{19} Corporate guarantee*	COR GUAR	0, 1	0 = No, 1 = Ok from creditable company. There is no such a default with a client has a corporate guarantee.
X_{20} Other guarantors	–	–	If required.
Y Loan quality* (dependent variable)	LOAN QUA	0, 1	0 = Default/bad credit, 1 = Paid/good credit

* Variables finally selected in the credit scoring models.

Appendix B. Statistical analysis for conventional models

Discriminating function for DA model:

Functions derived	Wilks Lambda	Chi-Square	DF	*P*-Value
1	0.543615	349.2512	12	0.0000

Discriminating function for DA_1 model:

Functions derived	Wilks Lambda	Chi-Square	DF	*P*-Value
1	0.5438	349.9703	9	0.0000

Analysis of deviance and likelihood ratio tests for PA model:

Analysis of deviance

Source	Deviance	Df	*P*-Value
Model	374.5	13	0.0000
Residual	284.906	567	1.0000
Total (corr.)	659.407	580	

Likelihood ratio tests

Factor	Chi-square	Df	*P*-Value
ADD INC	0.00152616	1	0.9688
AGE	12.0717	1	0.0005
COR GUAR	72.313	1	0.0000
EDU	11.6285	1	0.0006
HOR	2.70153	1	0.1002
LFOB	72.0333	1	0.0000
LOAN AMO	78.0624	1	0.0000
MAR STA	5.04102	1	0.0248
SALA	5.69163	1	0.0170
SEX	0.53373	1	0.4650
TELE	61.4374	1	0.0000
COMP	3.49304	2	0.1744

Analysis of deviance and likelihood ratio tests for PA_1 model:

Analysis of deviance

Source	Deviance	Df	*P*-Value
Model	370.674	9	0.0000
Residual	288.732	571	1.0000
Total (corr.)	659.407	580	

Likelihood ratio tests

Factor	Chi-square	Df	*P*-Value
AGE	10.8605	1	0.0010
COR GUAR	72.5957	1	0.0000
EDU	10.7326	1	0.0011
HOR	5.60935	1	0.0179
LFOB	69.6341	1	0.0000
LOAN AMO	99.0516	1	0.0000
MAR STA	6.08719	1	0.0136
SALA	5.84293	1	0.0156
TELE	61.5081	1	0.0000

Analysis of deviance and likelihood ratio tests for LR model:

Analysis of deviance

Source	Deviance	Df	*P*-Value
Model	374.661	13	0.0000
Residual	284.746	567	1.0000
Total (corr.)	659.407	580	

Likelihood ratio tests

Factor	Chi-square	Df	*P*-Value
ADD INC	0.0171689	1	0.8958
AGE	13.4555	1	0.0002
COR GUAR	74.1195	1	0.0000
EDU	10.5227	1	0.0012
HOR	1.88777	1	0.1695
LFOB	71.4812	1	0.0000
LOAN AMO	78.5665	1	0.0000
MAR STA	4.72988	1	0.0296
SALA	5.23704	1	0.0221
SEX	0.716841	1	0.3972
TELE	62.0235	1	0.0000
COMP	3.80231	2	0.1494

Analysis of deviance and likelihood ratio tests for LR_1 mode:

Analysis of deviance

Source	Deviance	Df	*P*-Value
Model	370.372	9	0.0000
Residual	289.035	571	1.0000
Total (corr.)	659.407	580	

Likelihood ratio tests

Factor	Chi-Square	Df	*P*-Value
AGE	12.3538	1	0.0004
COR GUAR	73.6767	1	0.0000
EDU	9.75523	1	0.0018
HOR	4.86088	1	0.0275
LFOB	68.7425	1	0.0000
LOAN AMO	99.7909	1	0.0000
MAR STA	6.12316	1	0.0133
SALA	5.35199	1	0.0207
TELE	61.8505	1	0.0000

Appendix C. Average correct classification rates for PA, PA_1, LR and LR_1

Cut-off	0.35 (%)	0.40 (%)	0.45 (%)	0.50 (%)	0.55 (%)	0.60 (%)	0.65 (%)	0.70 (%)	0.75 (%)
PA	85.89	86.75	87.44	87.78*	87.95	88.98	89.33*	87.44	86.40
PA_1	85.89	86.75	87.61	87.26*	87.95	88.81*	87.95	86.75	86.23
LR	86.75	87.09	87.61	88.30*	88.81	89.85*	89.16	87.61	86.23
LR_1	86.75	87.61	87.44	87.95*	88.47	89.16*	87.61	86.75	86.40

Numbers in cells refer to the average correct classification rates under the different cut-offs.
* The 0.50 standard cut-off rates and the highest rates per model are indicated by asterisks.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, XXIII*(4), 589–609.

Anderson, T. W. (2003). An introduction to multivariate statistical analysis. New York: Wiley-Interscience.

Arminger, G., Enache, D., & Bonne, T. (1997). Analyzing credit risk data: A comparison of logistic discriminant, classification tree analysis, and feedforward networks. *Computational Statistics, 12*(2), 293–310.

Baesens, B., Gestel, T. V., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 54*(6), 627–635.

Bailey, M. (2001). Credit scoring: The principles and practicalities. Kingswood, Bristol: White Box Publishing.

Bailey, M. (2004). Consumer credit quality: Underwriting, scoring, fraud prevention and collections. Kingswood, Bristol: White Box Publishing.

Banasik, J., Crook, J., & Thomas, L. (2001). Scoring by usage. *Journal of the Operational Research Society, 52*(9), 997–1006.

Banasik, J., Crook, J., & Thomas, L. (2003). Sample selection bias in credit scoring models. *Journal of the Operational Research Society, 54*(8), 822–832.

Bishop, C. M. (1995). Neural networks for pattern recognition. New York: Oxford University Press Inc.

Blochlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance, 30*(3), 851–873.

Bluhm, C., Overbeck, L., & Wagner, C. (2003). An introduction to credit risk modeling. London: Chapman & Hall/CRC.

Boyes, W. J., Hoffman, D. L., & Low, S. A. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics, 40*(1), 3–14.

Caouette, J. B., Altman, E. I., & Narayanan, P. (1998). Managing credit risk: The next great financial challenge. New York: John Wiley & Sons Inc.

Casu, B., Girardone, C., & Molyneux, P. (2006). Introduction to banking. London: Prentice Hall.

Chen, M., & Huang, S. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications, 24*(4), 433–441.

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research, 95*(1), 24–37.

Dimla, D. E., & Lister, P. M. (2000). On-line metal cutting tool condition monitoring. II: Tool-state classification using multi-layer perceptron neural networks. *International Journal of Machine Tools & Manufacture, 40*(5), 769–781.

Durand, D. (1941). Risk elements in consumer instalment financing, studies in consumer instalment financing. New York: National Bureau of Economic Research.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188.

Greene, W. (1998). Sample selection in credit-scoring models. *Japan and the World Economy, 10*(3), 299–316.

Guillen, M. & Artis, M. (1992). Count data models for a credit scoring system. In *The European Conference Series in Quantitative Economics and Econometrics on Econometrics of Duration, Count and Transition Models. Paris.*

Hand, D. J. (1981). Discrimination and classification. New York: John Wiley & Sons Inc.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160*(3), 523–541.

Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standard, 49*(6), 409–436.

Hoffmann, F., Baesens, B., Mues, C., Gestel, T. V., & Vanthienen, J. (2007). Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. *European Journal of Operational Research, 177*(1), 540–555.

Irwin, G. W., Warwick, K., & Hunt, K. J. (1995). Neural networks applications in control. London: The Institution of Electronic Engineers.

Johnson, R. A., & Wichern, D. W. (2002). Applied multivariate statistical analysis. Prentice Hall.

Kim, Y. S., & Sohn, S. Y. (2004). Managing loan customers using misclassification patterns of credit scoring model. *Expert Systems with Applications, 26*(4), 567–573.

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

Lee, T., Chiu, C., Lu, C., & Chen, I. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications, 23*(3), 245–254.

Leonard, K. J. (1995). The development of credit scoring quality measures for consumer credit application. *International Journal of Quality & Reliability Management, 12*(4), 79–85.

Lewis, E. M. (1992). An introduction to credit scoring. California: Fair, Isaac & Co., Inc.

Liang, Q. (2003). Corporate financial distress diagnosis in China: Empirical analysis using credit scoring models. *Hitotsubashi Journal of Commerce and Management, 38*(1), 13–28.

Lim, M. K., & Sohn, S. Y. (2007). Cluster-based dynamic scoring model. *Expert Systems with Applications, 32*(2), 427–431.

Long, M. S. (1973). *Credit scoring development for optimal credit extension and management control. College on Industrial Management, Georgia Institute of Technology*. Atlanta Georgia: Purdue University.

Maddala, G. S. (2001). Introduction to econometrics. Chichester: John Wiley & Sons Inc.

Malhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *Omega the International Journal of Management Science, 31*(2), 83–96.

Masters, T. (1995). Advanced algorithms for neural networks: AC++ sourcebook. New York: John Wiley & Sons, Inc.

Mays, E. (2001). Handbook of credit scoring. Chicago: Glenlake Publishing Company, Ltd.

Mays, E. (2004). The rule of credit scores in consumer lending. In E. Mays (Ed.), *Credit scoring for risk managers: The handbook for lenders* (pp. 3–12). Australia: Thomson South-Western.

Neter, J., Kutner, M. H., Wasserman, W., & Nachtsheim, C. J. (1996). Applied linear statistical models. Chicago: McGraw-Hill/Irwin.

Ong, C., Huang, J., & Tzeng, G. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications, 29*(1), 41–47.

Orgler, Y. E. (1971). Evaluation of bank consumer loans with credit scoring models. *Journal of Bank Research, 2*(1), 31–37.

Palisade Corporation. (2005). Neural Tools: Neural Networks Add-In for Microsoft Excel. Version 1.0. New York: Palisade Corporation.

Picton, P. (2000). Neural networks. Chippenham, Wilts: Palgrave, Antony Rowe Ltd.

Pindyck, R. S., & Rubinfeld, D. L. (1997). Econometric models and economic forecasts. McGraw-Hill/Irwin.

Reed, R. D., & Marks, R. J. (1999). Neural smithing: Supervised learning in feedforward artificial neural networks. London: The MIT Press.

Sarlija, N., Bensic, M., & Bohacek, Z. (2004). *Multinomial model in consumer credit scoring. 10th International Conference on Operational Research*. Trogir: Croatia.

Seow, H., & Thomas, L. C. (2006). Using adaptive learning in credit scoring to estimate take-up probability distribution. *European Journal of Operational Research, 173*(3), 880–892.

Siddiqi, N. (2006). *Credit risk scorecards: Developing and implementing intelligent credit scoring*. New Jersey: John Wiley & Sons, Inc.

Steenackers, A., & Goovaerts, M. J. (1989). A credit scoring model for personal loans. *Insurance: Mathematics and Economics, 8*(8), 31–34.

Sullivan, A. C. (1981). Consumer finance. In E. I. Altman (Ed.), *Financial handbook* (pp. 9.3–9.27). New York: John Wiley & Sons.

Thomas, L. C., Edelman, D. B., & Crook, L. N. (2002). *Credit scoring and its applications*. Philadelphia: Society for Industrial and Applied Mathematics.

Thompson, P. (1998). Bank lending and the environment: Policies and opportunities. *International Journal of Bank Marketing, 16*(6), 243–252.

Trippi, R. R., & Turban, E. (1993). *Neural networks in finance and investing: Using artificial intelligence to improve real-world performance*. Chicago: IRWIN.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research, 27*(11-12), 1131–1152.

Yang, Z., Wang, Y., Bai, Y., & Zhang, X. (2004). Measuring scorecard performance. *Computational Science, 3039*, 900–906.

Zekic-Susac, M., Sarlija, N., & Bensic, M. (2004). Small Business Credit Scoring: A Comparison of Logistic Regression, Neural Networks, and Decision Tree Models. In 26th International Conference on Information Technology Interfaces. Croatia.