Expert Systems with Applications 39 (2012) 2650–2661

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model

Bo-Wen Chi, Chiun-Chieh Hsu *

Department of Information Management, National Taiwan University of Science and Technology, Taipei, Taiwan

ARTICLE INFO

Keywords:
Dual scoring model
Mortgage behavioral scoring model
Credit bureau scoring model
Genetic algorithm
Logistic regression

ABSTRACT

Credit scoring model is an important tool for assessing risks in financial industry, consequently the majority of financial institutions actively develops credit scoring model on the credit approval assessment of new customers and the credit risk management of existing customers. Nonetheless, most past researches used the one-dimensional credit scoring model to measure customer risk. In this study, we select important variables by genetic algorithm (GA) to combine the bank's internal behavioral scoring model with the external credit bureau scoring model to construct the dual scoring model for credit risk management of mortgage accounts. It undergoes more accurate risk judgment and segmentation to further discover the parts which are required to be enhanced in management or control from mortgage portfolio. The results show that the predictive ability of the dual scoring model outperforms both one-dimensional behavioral scoring model and credit bureau scoring model. Moreover, this study proposes credit strategies such as on-lending retaining and collection actions for corresponding customers in order to contribute benefits to the practice of banking credit.

1. Introduction

Credit scoring model is the most successful example of statistical model applied in financial institutions (Thomas, 2000). To achieve the purpose of risk control, application scoring models help banks to decide whether to grant credit to new applicants based on customers' characteristics such as age, education, marital status, and so on (Chen & Huang, 2003). In a similar vein, behavioral scoring models help banks to make use of analyzing existing customers' consumption behavior to evaluate future risk of delinquency and loss (Setiono, Thong, & Yap, 1998). Therefore, financial institutions try very hard to develop credit scoring models to separate customers' risk accurately, while concurrently effectively control risks to enhance the flexibility in funding use.

The causes of the US subprime mortgage crisis in 2007 were due to some reasons. First, US house prices began their steep decline after the peak in mid-2006, and the adjustable-rate mortgages began to reset at higher rates resulted in mortgage delinquencies soar. Second, financial institutions offered loans to higher-risk borrowers without appropriate review and documentation. The warning of subprime mortgage crisis made Taiwan's Financial Supervisory Commission to pay attention to see whether the banks' lending policy was too loose, and request banks to set up a reasonable risk-based pricing policy. Therefore, it is essential to build up a mortgage scoring system to strengthen risk management mechanism.

The common approaches had been used to construct credit scoring models included linear discriminant analysis (e.g., Altman, Marco, & Varetto, 1994; Jo, Han, & Lee, 1997; Kim, Kim, Kim, Ye, & Lee, 2000), neural networks (e.g., Desai, Crook, & Overstreet, 1996; Mahlhotra & Malhotra, 2003; Tsai & Wu, 2008; West, 2000; Zhang, Hu, Patuwo, & Indro, 1999), classification and regression tree (e.g., David, Edelman, & Gammerman, 1992; Feldman & Gross, 2005; Lee, Chiu, Chou, & Lu, 2006; Li, Ying, Tuo, Li, & Liu, 2004), and logistic regression (e.g., Dinh & Kleimeier, 2007; Laitinen & Laitinen, 2000; Steenackers & Goovaerts, 1989; Westgaard & Wijst, 2001). Altman (1968) applies multiple discriminant analysis to predict corporate bankruptcy. Altman's study has a sample size of 66, which are comprised of 33 manufacturing companies that have filed for bankruptcy from 1946 to 1965 and 33 non-bankrupt companies matched on a stratified random basis by both industry and asset size. Altman selects the five most significant variables from 22 financial ratio variables to construct a bankruptcy prediction model. These variables are working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to book value of total liabilities and sales to total assets. This model has an accuracy rate of 95% one year prior to bankruptcy, and an accuracy rate of 72% two years prior bankruptcy. Atiya (2001) applies neural networks (NNs) to bankruptcy prediction for credit risk. He indicates that the use of

* Corresponding author. Address: No. 43, Sec. 4, Keelung Road, Taipei 106, Taiwan. Tel.: +886 2 27376766; fax: +886 2 27376777.

E-mail address: cchsu@cs.ntust.edu.tw (C.-C. Hsu).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.08.120

equity-based indicators in addition to financial ratio indicators provides significant improvements in the prediction accuracy. Feldman and Gross (2005) apply classification and regression trees to analyze Israeli mortgage default data. They find that if the costs of accepting bad customers are more than rejecting good ones, borrowers' features are the powerful predictors of default rather than mortgage contract features. However, if the costs are equal, mortgage contract features are powerful predictors as well. The higher (lower) the ratio of the misclassification costs of bad risks compared to good risks, the lower (higher) are the resulting misclassification rates of bad risks, and the higher (lower) are the misclassification rates of the good risks. Steenackers and Goovaerts (1989) apply stepwise logistic regression model to find out the criteria for the selection of new credits. They use personal loans collected from a Belgian credit company during the period from 1984 to 1986. They list 19 characteristics of applicants and indicate the selecting criteria including age, with a telephone or not, duration of residence in current place, duration of employment period, standard of living area, occupation, civil servant or not, monthly income, ownership of the house, number of past loans, and length of the loan.

The above-mentioned methods use a one-dimensional approach to construct the credit scoring model. However, more recent studies have proposed to use a combined model or a combined score approach to enhance the predictability of the one-dimensional credit scoring model.

Altman et al. (1994) compare NNs with linear discriminant analysis (LDA) for Italian industrial firms using a large sample size. The results show that LDA outperforms NNs. However, they suggest integrating NNs and LDA due to the performance of NNs is subject to enhancement by integration with statistical approaches. Lee and Jung (2000) compare the forecasting ability of logistic regression (LR) with that of NNs to identify creditworthiness of urban and rural customers. The results show that LR has better prediction for urban customers and NNs have better prediction for rural customers. They also mention a combination of two models to enhance the forecasting accuracy. Koh, Tan, and Goh (2006) combine individual credit scoring models to build a combined model which performs better than the individual models. Nonetheless this study suggests that limitation is the possible difficulty in interpreting rules generated by the combined model. Furthermore, the combined model is unlikely to make sense because it is built on individual models that are likely to generate different rules.

Zhu, Beling, and Overstreet (2001) investigate the performance of a regression-based combination of application credit score and bureau credit score. The results show that the combined score outperforms each set of scores from its basis; however, this research also suggests that the dependence of the two individual scores should be taken into account.

Due to the difficulty in interpreting the rules generated by the combined model and the high collinearity between two consumer credit scores, we combine the bank's internal mortgage behavioral scoring model with the external credit bureau scoring model to construct the dual scoring model. In this way, we are able to make more accurate risk judgment and segmentation to further discover the parts which are required to be enhanced in management or control from mortgage portfolio. The results show that the predictive and distinguish ability of the dual scoring model outperforms both of the one-dimensional behavioral scoring model and credit bureau scoring model. Consequently, this study recommends applying the dual scoring model in bank's credit strategies.

The remainder of this paper proceeds as follows. Section 2 discusses research design and empirical methodology. Section 3 presents the sampling procedure, data, and empirical results. Section 4 describes the application of bank's credit strategies based on the dual scoring model. A brief conclusion follows.

2. Analysis methodology

The literatures mostly emphasize on one-dimensional perspective to construct credit scoring model in order to measure customer risk. In this study, we use the bank's internal data to construct the mortgage behavioral scoring model. In addition, we utilize customers' credit bureau data from the Public Credit Registers (PCR) to construct the credit bureau scoring model, and combine the mortgage behavioral scoring model with the credit bureau scoring model to yield a two-dimensional dual scoring model. This dual scoring model is capable of more accurately segment customer risk and improve credit assessment. The development process of dual scoring model in this study is shown in Fig. 1:

Step 1: Data preprocessing.
In this step, we collect bank's internal data and credit bureau data from the PCR on existing mortgage customers, and engage in data cleaning, data integration, data transformation, data reduction, and generation of feature selection from candidate models for the above two types of data.
Step 2: Segmentation analysis.
In this step, we apply segmentation analysis to classify customers into homogeneous risk groups based on the customer characteristics, and thereby enhance the predictability of the model.
Step 3: One-dimensional credit scoring model building.
We build a one-dimensional credit scoring model respectively for segmented groups and calibrate several scoring models to derive a consistent score-to-odds relationship. The behavioral score and credit bureau score are divided into BHS1, BHS2, BHS3, BHS4, BHS5 and CBS1, CBS2, CBS3, CBS4, CBS5 risk ranks respectively in accordance with the degree of risk score.
Step 4: Dual scoring model building.
The five risk ranks of the mortgage behavioral scoring model and credit bureau scoring model are combined into a 5×5 risk matrix, and the Kolmogorov-Smirnov (K–S) statistic and Receiver Operating Characteristic (ROC) curve are used to measure the predictability of the credit scoring model.
Step 5: Credit strategy.
Based on the 5×5 risk matrix, we design corresponding credit strategies such as on-lending retaining and collection actions.

2.1. Data preprocessing

Data preprocessing is required to ensure data field consistency and determine the relative importance variables in credit scoring model building. In this study, we collect bank's internal data and credit bureau data from the PCR on existing mortgage customers. The bank's internal data include borrower characteristics, collateral characteristics and payment characteristics. The credit bureau data include mortgage customers' credit-related information of amounts owed, payment history, new credit, length of credit history, types of credit used, and account information. The data preprocessing will be proceeded as follows:

1. *Data cleaning*: Using the attribute mean to fill in the missing value, smooth out noisy data, identify or remove outliers, and resolve inconsistencies.
2. *Data integration*: Using a customer identifier to join different data sets.
3. *Data transformation*: Scaling attribute values to fall within a specified range (normalization), move up in the concept hierarchy on the numeric attributes (aggregation), move up in the

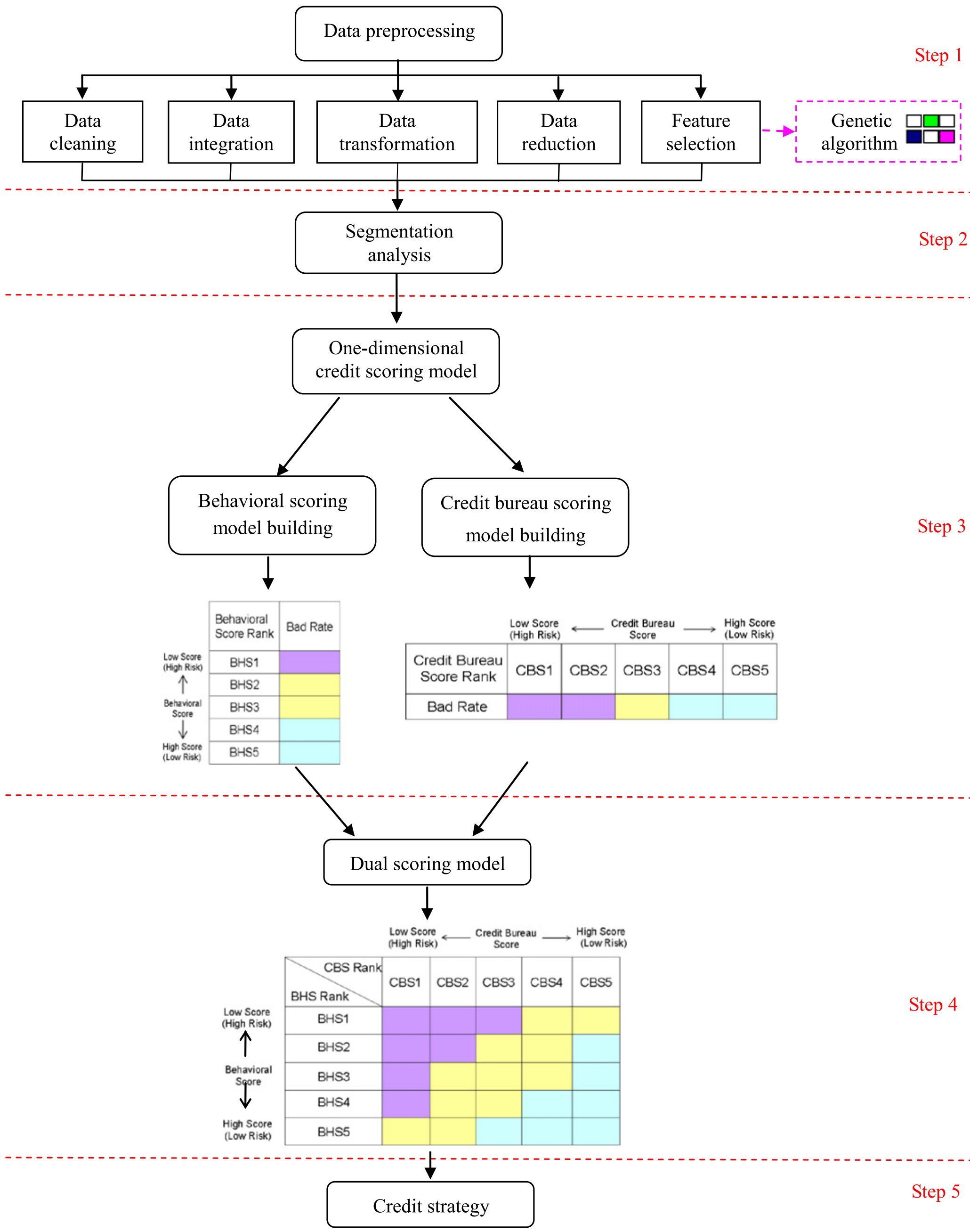

Fig. 1. Research framework.

concept hierarchy on nominal attributes (generalization), and replace or add new attributes inferred by existing attributes (attribute construction).

4. *Data reduction*: Removing irrelevant attributes from data and reducing the number of attributes by grouping them into intervals (binning).
5. *Feature selection*: Genetic algorithm (GA) was proposed by Holland in 1975 to be used as a heuristic combinatorial optimization search technique. Compared to traditional statistical approach, GA is not bounded by the form of functions, which is advantage of GA. This study takes advantage of the nature of the GA fitness function to analyze the input variables influencing the mortgage payment status for feature information, and then converts the rules of the hidden features of the various variables and transforms them into important numbers. The relative important numbers range from 0 to 1, and they are normalized so that all inputs add up to approximately 1. A variable with greater number means that it is more capable of predicting results. The use of GA as a technique for ranking the importance of variables enables us to systematically identify the usefulness of variables and objectively rank their importance, which is very helpful in eliminating the ineffective inputs and saving the useful ones when selecting model inputs (Chi & Tang, 2007).

2.2. Segmentation analysis

Segmentation analysis is used to segment customers into homogeneous groups according to the behavior and characteristics of the customers as well as the nature of the products. Customers in the same group possess similar risk characteristics or primary risk drivers. Building an individual credit scoring model for each subpopulation will allow us to separate the good accounts (Goods) from the bad accounts (Bads) more accurately than if one credit scoring model is built to handle the whole population. The development of the credit scoring model is often asked for a sample of at least 1000 Goods and 500 Bads. In general, the more of both Goods and Bads available, the more accurate the resulting score (see Mays, 2001).

In an investigation into the usefulness of traditional LDA and artificial NNs, Altman et al. (1994) work with financial data from over 1000 Italian industrial firms from 1982 to 1992 and find that NNs have similar level of accuracy as credit scoring model. Furthermore, they indicate that the results are part of great efforts involving separate models for industrial, retailing/trading and construction firms.

2.3. One-dimensional credit scoring model

2.3.1. Credit scoring model building

NNs and LR have been widely used in building the credit scoring models due to higher predictability. Desai et al. (1996) employ the personal loan information of three certain credit unions in US to investigate the predictive power of NNs, LDA, and LR for scoring credit decision. They conclude that NNs outperform LDA in classifying the most difficult group, namely poor loans, and LR is comparable to NNs. West (2000) examines the credit scoring accuracy of five neural network architectures and compares them to traditional statistical methods. The neural architectures and traditional models constructed using multilayer perceptron, mixture-of-experts, radial basis function, learning vector quantization, and fuzzy adaptive resonance; and LDA, LR, k nearest neighbor, kernel density estimation, and decision trees, respectively. The result from two real world data sets and testing the models using 10-fold cross validation from German and Australian credit scoring shows that LR is the most accurate traditional model, whereas, the mixture-of-experts and radial basis function perform best among neural architectures. West goes further to suggest that LR is a good alternative to NNs.

Using NNs for establishing credit scoring model becomes increasingly popular because of its high predictability from empirical studies and the rise of artificial intelligence (Tam & Kiang, 1992). However, recent studies have pointed out the drawbacks of neural networks. West (2000) indicates that NNs are commonly regarded as a black-box technique without logic or rule-based explanations for the input-output approximation. The main drawback of utilizing NNs for credit scoring model is the difficulty in explaining the underlying principle for the decision to reject credit applications. In a similar vein, Piramuthu (1999) shows that the NNs fail to interpret the rationale and conceptual framework behind its credit granting or denial decision. Meanwhile, NNs have been criticized for its poor performance and sensitivity when existing irrelevant attributes or small data sets (Castillo, Marshall, Green, & Kordon, 2003; Feraud & Cleror, 2002).

For predicting dichotomous outcomes, the LR has been concluded as one of the most appropriate techniques (Jo et al., 1997). A number of explorations of LR model for credit scoring applications have been reported in literature. The LR has been explored by Wiginton (1980), Steenackers and Goovaerts (1989), Joanes (1993), and Laitinen (1999) in building credit scoring models for personal loans and business loans. In addition, because the G/B odds ratio in the logistic regression is easy to calculate and interpret, the logistic regression is wildly applied in finance practice. For this season, we apply logistic regression to build the credit scoring model for the subpopulation following the segmentation analysis in this study.

2.3.2. Calibration

Credit scores are commonly scaled linearly to take more integer points and to conform to industry or company standards. This study scales the points such that a total credit score of 200 points corresponds to G/B odds of 1 to 1, and that an increase of the credit score of 20 points corresponds to a doubling of G/B odds. Derivation of the scaling rule that transforms the credit scores of each attribute is seen in Eqs. (1) and (2).

$$\begin{aligned} score &= \ln(odds) * factor + offset \\ &= \left(-\sum_{j,i=1}^{k,n}(woe_j * \beta_i) + a\right) * factor + offset \\ &= \left(-\sum_{j,i=1}^{k,n}\left(woe_j * \beta_i + \frac{a}{n}\right)\right) * factor + offset \\ &= \sum_{j,i=1}^{k,n}\left(-\left(woe_j * \beta_i + \frac{a}{n}\right) * factor + \frac{offset}{n}\right) \end{aligned} \tag{1}$$

$$\begin{aligned} &200 = \ln(1) * factor + offset \\ &220 = \ln(2) * factor + offset \\ &factor = 20/\ln(2) \\ &offset = 200 - factor^{*}\ln(1) \end{aligned} \tag{2}$$

where woe is the weight of evidence for each grouped attribute, β is the regression coefficient for each variable, a is the intercept term from LR, n is the number of variables, and k is the number of attributes in each variable.

Raw scores from scoring models are roughly related to the predicted probability of good credit but the relationship is not exact and the scores cannot be compared to scores produced from another scoring model. The calibration formula can standardize the relationship between the score and G/B odds, and the credit scoring model's scores in different segments can be compared directly.

2.3.3. Risk rank

To facilitate the use of credit strategies in collection actions, onlending retaining and marketing, the credit score is commonly divided into different risk ranks in accordance with the degree of risk score. For example, score below 251 is risk rank 1, score between 251 and 350 is risk rank 2, score between 351 and 450 is risk rank 3, and so forth.

2.4. Dual scoring model

The behavioral scoring model in past literature generally only uses bank's internal customer data, such as customers' application data, transaction data and repayment data. However the internal customer data collected independently by bank has its deficiencies in terms of incompleteness. For example, it lacks data on bank's internal customer interactions with other financial institutions, that is to say, the proprietary information resulted from business competition. The study additionally collects mortgage customers' credit-related information from the PCR to build the credit bureau scoring model, and combines this with the behavioral scoring model to yield a two-dimensional dual scoring model. The purpose is to present the possibility of customer payment by using objective and specific data. This dual scoring model can differentiate the customer risk more precisely and go further to design corresponding credit strategies. In order to ensure the effectiveness of the credit scoring model, numerous measures and charts are then applied to investigate the discriminatory power of the credit scoring models. These measures evaluate the model's ability to separate the Goods from the Bads based on their credit scores. Measures include

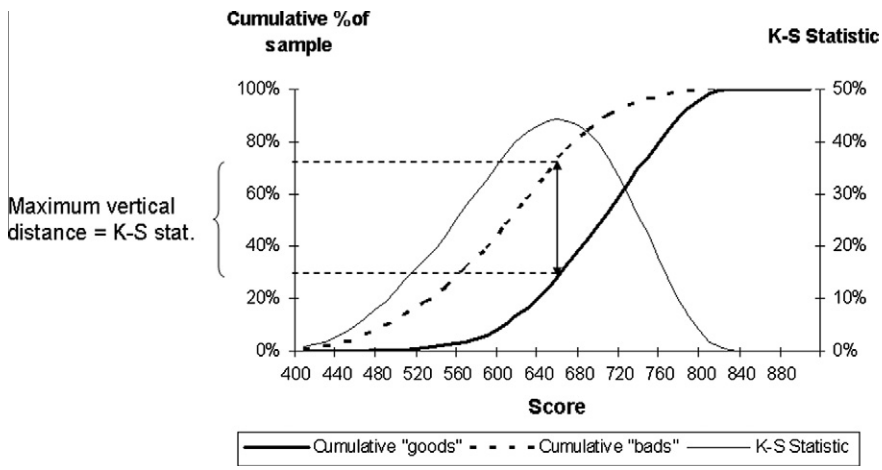

Fig. 2. The K–S statistic.

the Kolmorogov–Smirnoff (K–S) statistic and the Receiver Operating Character (ROC) curve analysis described as follows:

2.4.1. The K–S statistic

The K–S statistic measures the maximum difference between the plotted cumulative distribution functions of two discriminations (i.e., Goods and Bads), where each discrimination scores is transformed to lie between 0 and 1. The score that produces the greatest separability between the functions is considered the cutoff value for accepting or denying a credit application. The credit score model generating the greatest amount of separability between the two distributions is regarded as the better model. The equation is mentioned as below:

$$\text{K--S} = \max_s |F(s|G) - F(s|B)| \quad (3)$$

where $F(s|G)$ and $F(s|B)$ are cumulative distribution functions of Goods and Bads, and s is the corresponding score for the individual loan.Fig. 2 shows Bads rapidly accumulate at low scores, while Goods accumulate more quickly at high scores. In addition, the Goods' cumulative distribution function curve lies to the right of the Bads' cumulative distribution function curve.

2.4.2. The ROC curve analysis

The ROC curve, such as the one depicted in Fig. 3, is a graphical representation of the trade-off between the percentage of false alarms (e.g. 1 – specificity) of a credit scoring model on the x-axis against the percentage of hits (e.g. sensitivity) on the y-axis for all possible classification thresholds. If high scores are defined to present a low default probability, then x-values represent the error rate that Goods are classified as Bads by a credit scoring model (e.g. Type II error) and y-values represent one minus the error rate that Bads are classified as Goods by a credit scoring model (e.g. Type I error). Thus, the ROC curve is also a complete representation of Type I and Type II errors.

The area under the ROC curve (AUC) is an important index to measure the discriminatory power of a credit scoring model, which can be interpreted as the probability that the Goods receive better scores than the Bads. It is equivalent to the Gini coefficient and also to the Mann–Whitney–Wilcoxon two-independent sample non-parametric test statistic.

The AUC value ranges from 0.5 to 1. The larger the AUC value, the more accurate the credit scoring model. In the case of a perfect model, the area beneath will equal 1, while an area of 0.5 reflects a random model. When the AUC value is greater than 0.7, it means the model has good discriminating capacity (Cholongitas et al., 2006).

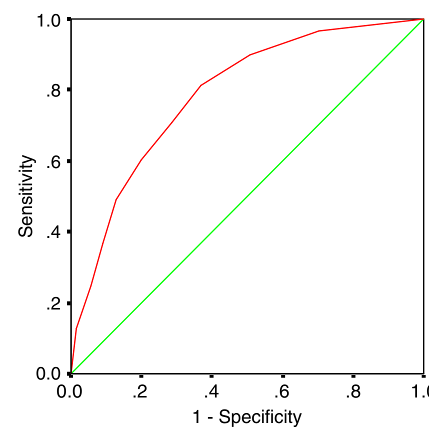

Fig. 3. ROC curve.

2.5. Credit strategy applications

The banks can design and implement the related credit strategies based on their behavioral scoring model of existing mortgage customers, e.g. collection actions, risk pricing, and on-lending retaining strategies. But if they can incorporate the credit bureau scoring model into the behavioral scoring model, they will undertake more refined risk segmentation. In this study, the dual scoring model is applied in collection actions and on-lending retaining strategies. The credit strategy applications are described in Section 4.

3. Empirical analysis

3.1. Behavioral scoring model

3.1.1. Description of data and development period

The mortgage data set which we analyze in this study came from one of the major banks in Taipei, Taiwan. There are totally 86,628 active accounts in the data set after excluding 7219

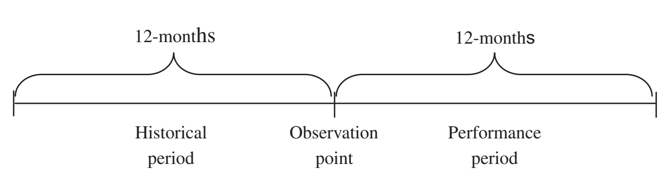

Fig. 4. Development period of the credit scoring model.

Table 1
Results of the GA for input feature selection of behavioral scoring model.

Characteristics		Variables	Useful	Importance of input
I	Borrower characteristics	Borrower's gender: male, female	No	0.003
		Borrower's age	Yes	0.053
		Borrower's material: single, married, divorced, and others	No	0.012
		Borrower's education: elementary, high school, college/university, graduate school, and others	Yes	0.051
		Borrower's occupation: white collar, professional person, military/public official/teacher, blue collar, freelancer, and others	Yes	0.062
		Borrower's job status: employed, self-employed, unemployed, and others	No	0.048
		Number of borrower's children	No	0.001
		Borrower's work seniority	No	0.042
		Borrower's income	No	0.022
		Number of guarantors	No	0.005
		Purpose of the mortgage: self living, investment, repair, and others	Yes	0.059
II	Collateral characteristics	Collateral area: north, middle, south, and east of the country	Yes	0.060
		Original value of the property	No	0.002
		Mortgage's loan amount	No	0.002
		Original loan-to-value ratio: the approval amount of mortgage divided by the appraised value of property	Yes	0.055
		Floor space of the property	No	0.002
III	Payment characteristics	Debt-to-income ratio	Yes	0.089
		Number of 30+ days delinquent in last 12 months	Yes	0.067
		Number of 60+ days delinquent in last 12 months	Yes	0.075
		Number of 90+ days delinquent in last 12 months	No	0.038
		Worst delinquency status in last three months	No	0.028
		Worst delinquency status in last 6 months	No	0.028
		Worst delinquency status in last 12 months	No	0.032
		Term to maturity of the mortgage	Yes	0.066
		Current loan-to-value ratio: the outstanding of the mortgage divided by the appraised value of property	Yes	0.098

Table 2
Results of behavioral scoring model.

Variables	Attributes	G/B odds	Attribute points
Age	20–25	3.00	60
	26–30	2.80	56
	31–40	1.95	39
	41–50	1.15	23
	Above 50 (reference)	1.00	20
Borrower's occupation	White collar	2.36	59
	Professional person	2.84	71
	Military/public official/teacher	3.00	75
	Blue collar	1.76	44
	Freelancer	1.28	32
	Others (reference)	1.00	25
Debt-to-income ratio	0	3.04	85
	0.1–0.3	2.29	64
	0.4–0.5	1.71	48
	0.6–0.7	1.54	43
	0.8–0.9	1.25	35
	Above 0.9 (reference)	1.00	28
Purpose of the mortgage	Self living	1.81	65
	Investment	0.58	21
	Repair	1.56	56
	Others (reference)	1.00	36
Collateral area	North	2.92	70
	Middle	2.29	55
	South	1.42	34
	East (reference)	1.00	24
Number of 60+ days delinquent in last 12 months	0	2.93	79
	1	1.96	53
	Above 1 (reference)	1.00	27
Term to maturity of the mortgage	13–24	2.92	76
	25–60	2.19	57
	61–108	2.04	53
	109–180	1.81	47
	Above 180 (reference)	1.00	26
Current loan-to-value ratio	Below 10%	3.10	90
	10–40%	2.93	85
	41–70%	1.97	57
	71–90%	1.48	43
	Above 90% (reference)	1.00	29

Table 3
Credit scoring results of the three constructed models.

	Training set		Testing set	
	K–S	AUC	K–S	AUC
Behavioral scoring model	0.446	0.787	0.415	0.733
Credit bureau scoring model	0.510	0.809	0.505	0.795
Dual scoring model	0.593	0.877	0.546	0.841

accounts opened for less than 12 months, 369 accounts who have 2 or more installments in arrears, 71 collection accounts, 221 write-off accounts. In this study, the development period of the behavioral scoring model is composed of the observation point, the historical period, and performance period. Fig. 4 shows the observation point, the historical period, and performance period. The observation point is December 31, 2008, and all active accounts are observed at this point in time. The 12-month period prior to the observation point is called the historical period (i.e., 2008/1/1–2008/12/31), and the mortgage payment and transaction records will be served as the independent variables of the behavioral scoring model during this time. In addition, the 12-month period following the observation point is called the performance period (i.e., 2009/1/1–2009/12/31), and the customers are classified as either Goods or Bads based on their payment performance during this time. Bads mean that the customers have ever 2 or more installments in arrears during the performance period. By contrast, Goods mean that the customers have never 2 or more installments in arrears during the performance period. For the 86,628 total accounts, 85,870 of which are Goods and 758 ones are Bads. Therefore, the G/B odds ratio is 113:1. To avoid overfitting the constructed model, the G/B odds ratio of 3:1 is used in this study (see Chuang & Chen, 2006). In other words, we randomly select 2242 Goods and combine with 758 Bads to form the development sample. To validate the stability and accuracy of the behavioral scoring model, the data set of 3000 accounts is split into two groups for training and testing data sets by the ratio of 8:2 (see Lee et al., 2006).

3.1.2. Variables

The independent variables of mortgage behavioral scoring model consisted of three major dimensions, which are borrower characteristics, collateral characteristics, and payment characteristics. In this study, we use GA to select 11 variables as input variables for constructing the mortgage behavioral scoring model by the rule

Table 4
Results of the GA for input feature selection of credit bureau scoring model.

	Characteristics	Variables	Useful	Importance of input
I	Amounts owed	Credit card balance	No	0.002
		Credit card revolving balance	No	0.001
		Secured balance	No	0.001
		Secured limit	No	0.001
		Unsecured balance	No	0.002
		Unsecured limit	No	0.001
		Total revolving balance	Yes	0.051
		Total revolving limit	No	0.002
		Outstanding amount of cash cards	Yes	0.082
II	Payment history	Number of 1+ days delinquent in last 6 months	No	0.015
		Number of 1+ days delinquent in last 12 months	No	0.007
		Number of 30+ days delinquent in last 6 months	Yes	0.078
		Number of 30+ days delinquent in last 12 months	Yes	0.055
		Number of 60+ days delinquent in last 6 months	No	0.013
		Number of 60+ days delinquent in last 12 months	No	0.005
		Number of 90+ days delinquent in last 6 months	No	0.008
		Number of 90+ days delinquent in last 12 months	No	0.003
		Number of loan 1+ days delinquent in last 6 months	No	0.002
		Number of loan 1+ days delinquent in last 12 months	No	0.002
		Worst current status	Yes	0.055
		Worst delinquency in last 6 months	No	0.027
		Worst delinquency in last 12 months	No	0.026
III	New credit	Number of total inquires	Yes	0.053
		Number of banks with inquiries for new business in last one month	No	0.019
		Number of banks with inquiries for new business in last two months	No	0.022
		Number of banks with inquiries for new business in last three months	Yes	0.051
		Date of most recent enquire	No	0.001
IV	Length of credit history	Months since first opened for credit card	Yes	0.059
V	Types of credit used	Bureau forced closed record	No	0.002
		Bureau bounced check record/refusal of check facilities	No	0.002
		Bureau abnormal credit record	Yes	0.052
		Remark for repayment by relatives	No	0.000
VI	Account information	Average utilization ratio of credit card in last 1 month	Yes	0.062
		Average utilization ratio of credit card in last 3 months	No	0.032
		Average utilization ratio of credit card in last 6 months	Yes	0.053
		Average utilization ratio of credit card in last 12 months	Yes	0.072
		Difference between 6-month average utilization ratio and 12-month average utilization ratio of credit card	No	0.001
		Number of total trade lines	No	0.001
		Number of trade lines opened in last 3 months	No	0.001
		Number of trade lines opened in last 6 months	No	0.028
		Number of trade lines opened in last 12 months	Yes	0.050
		Number of trade lines with balance > 0	No	0.000
		Number of revolving trade lines	No	0.000

Table 5
Results of delinquent credit bureau scoring model.

Variables	Attributes	G/B odds	Attribute points
Outstanding amount of cash cards	Below NT$ 10,000	3.00	75
	NT$ 10,000–NT$ 30,000	2.36	59
	NT$ 30,001–NT$ 50,000	1.76	44
	NT$ 50,001–NT$ 100,000	1.28	32
	Above NT$ 100,000 (reference)	1.00	25
Number of 30+ days delinquent in last 6 months	0	3.08	74
	1	1.75	42
	2–3	1.33	32
	Above 3 (reference)	1.00	24
Worst current status	Non-delinquent	3.38	71
	1–29 days delinquent	2.81	59
	30–59 days delinquent	2.10	44
	60–89 days delinquent	1.52	32
	Above 90 days delinquent (reference)	1.00	21
Number of total inquires	0	3.50	70
	1	3.15	63
	2–3	1.65	33
	4–5	1.25	25
	Above 5 (reference)	1.00	20
Bureau abnormal credit record	Yes (reference)	1.00	19
	No	3.63	69
Average utilization ratio of credit card in last one month	Below 10%	3.17	73
	10–30%	2.70	62
	31–70%	2.04	47
	71–90%	1.43	33
	Above 90% (reference)	1.00	23
Number of trade lines opened in last 12 months	0	3.78	68
	1	2.83	51
	2	1.61	29
	Above 3 (reference)	1.00	18

Table 6
Results of non-delinquent credit bureau scoring model.

Variables	Attributes	G/B odds	Attribute points
Total revolving balance	Below NT$ 50,000	2.85	77
	NT$ 50,000–NT$ 100,000	2.15	58
	NT$ 100,001–NT$ 500,000	1.74	47
	NT$ 500,001–NT$ 1,000,000	1.56	42
	Above NT$ 1,000,000 (reference)	1.00	27
Outstanding amount of cash cards	Below NT$ 10,000	2.52	83
	NT$ 10,000–NT$ 15,000	1.73	57
	NT$ 15,001–NT$ 20,000	1.30	43
	Above NT$ 200,000 (reference)	1.00	33
Number of total inquires	0	2.79	78
	1	2.00	56
	Above 1 (reference)	1.00	28
Number of banks with inquiries for new business in last 3 months	0	2.92	76
	1	2.27	59
	Above 1 (reference)	1.00	26
Months since first opened for credit card	Below 12 (reference)	1.00	30
	12–24	1.87	56
	25–60	1.60	48
	61–120	2.13	64
	Above 120	2.67	80
Average utilization ratio of credit card in last 12 months	Below 10%	2.61	81
	10–20%	2.29	71
	21–50%	1.90	59
	51–70%	1.42	44
	Above 70% (reference)	1.00	31
Number of trade lines opened in last 12 months	0	3.00	75
	1	2.20	55
	2	1.36	34
	Above 2 (reference)	1.00	25

of important number > 0.05 (see Wang, Feng, & Chang, 2010). A total of 25 variables are listed in Table 1 as follows:

3.1.3. Results of the analysis

In this study, we use LR to construct the behavioral scoring model. Table 2 indicates that the eight significant variables of the behavioral scoring model and the attributes, G/B odds, and attribute points of each variable. These variables are age, occupation, debt-to-income ratio, purpose of mortgage, collateral area, number of 60+ days delinquent in last 12 months, terms to maturity of the mortgage, and original loan-to-value ratio.

The empirical results in Table 3 show that K–S and AUC values of the behavioral scoring model are 44.6% and 78.7%, respectively. To facilitate the credit strategy applications, we classify the customers into five risk ranks from BHS1 to BHS5 in accordance with the degree of risk score. Among which BHS1 is the rank with the highest risk and BHS5 is the rank with the lowest risk.

3.2. Credit bureau scoring model

3.2.1. Description of data and development period

The development period of the credit bureau scoring model is consistent with the behavioral scoring model. However, the data used in this model comprises those for all consumer finance-related products, including credit cards, personal loans, car loans, cash cards, and so on, which are obtained from the PCR.

3.2.2. Variables

The independent variables of the credit bureau scoring model consisted of six major dimensions, which are amounts owed, payment history, new credit, length of credit history, types of credit used, and account information. In this study, we use GA to select 13 variables as input variables for constructing the credit bureau scoring model by the rule of important number > 0.05. A total of 43 variables are listed in Table 4 as follows:

3.2.3. Results of the analysis

The study first use segmentation analysis to classify the customers into two groups according to whether their loans have became delinquent or not. We then apply LR to construct the delinquent credit bureau scoring model and non-delinquent credit bureau scoring model for two groups, respectively. Table 5 indicates that the seven significant variables of the delinquent credit bureau scoring model and the attributes, G/B odds, and attribute points of each variable. The seven variables are outstanding amount of cash cards, number of 30+ days delinquent in last 6 months, worst current status, number of total inquiries, bureau abnormal credit record, average utilization ratio of credit card in last one month and number of trade lines opened in last 12 months. Table 6 indicates that the seven significant variables of the non-delinquent credit bureau scoring model and the attributes, G/B odds, and attribute points of each variable. The seven variables are total revolving balance, outstanding amount of cash cards, number of total inquiries, number of banks with inquiries for new business in last three months, months since first opened for credit card, average utilization ratio of credit card in last 12 months and number of trade lines opened in last 12 months. We calibrate two credit bureau scoring models to derive a consistent score-to-odds relationship. The K–S and AUC values of the calibrated credit bureau scoring model are 51.0% and 80.9%, respectively, as shown in Table 3. In addition, we classify the customers into five risk ranks from CBS1 to CBS5 in accordance with the degree of risk score. Among which CBS1 is the rank with the highest risk and CBS5 is the rank with the lowest risk.

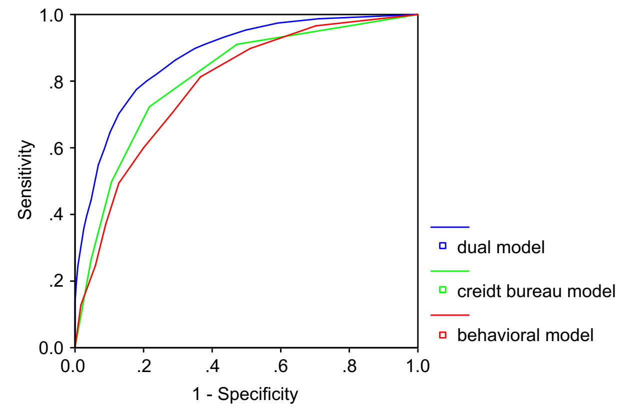

Fig. 5. ROC curves *f* or model results (training set).

3.3. Dual scoring model

3.3.1. Results of the analysis

The five risk ranks of the mortgage behavioral scoring model and the credit bureau scoring model are combined to construct a 5×5 dual scoring model. From Table 3 we can see that the K–S and AUC values of the dual scoring model are 59.3% and 87.7%, respectively. With regards to the K–S value, the dual scoring model increases the accuracy rate of prediction compared to both behavioral scoring model and credit bureau scoring model by 14.7% and 8.3%, respectively. With regards to the AUC value, the dual scoring model increases the accuracy rate of prediction

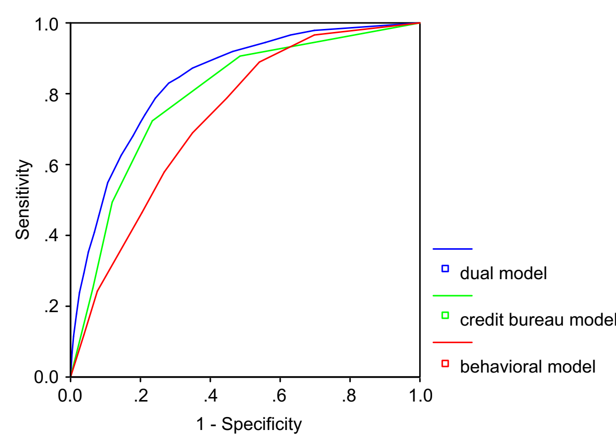

Fig. 6. ROC curves *f* or model results (testing set).

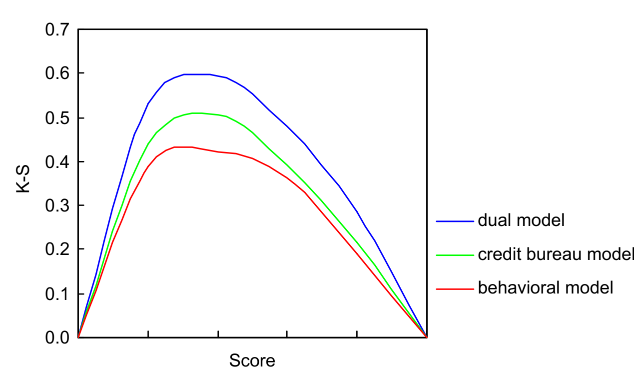

Fig. 7. K–S for model results (training set).

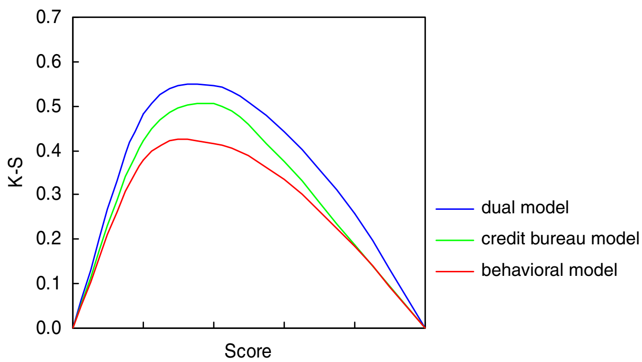

Fig. 8. K–S for model results (testing set).

compared to both behavioral scoring model and credit bureau scoring model by 9.0% and 6.8%, respectively. The empirical results show that the predictive ability of the dual scoring model outperforms both one-dimensional behavioral scoring model and credit bureau scoring model. In this way, we can undertake more refined risk segmentation.

3.3.2. Comparison of model accuracy for training set and testing set

From Figs. 5–8 we can see that the dual scoring model has higher K–S and AUC values in both training set and testing set than those of the behavioral scoring model and the credit bureau scoring model. In addition, the empirical results in Table 3 show that the accuracy of the testing set is slightly lower than that of the training set, indicating that the dual scoring model is stable.

4. Credit strategy application

From the results of the analysis we discover that the K–S and AUC values of the dual scoring model are significantly higher than those of the behavioral scoring model and the credit bureau scoring model. By using the dual scoring model in mortgage portfolio management, we classify 86,628 mortgage accounts into the 25 cells, so that we can more accurately segment customer risk. Furthermore, the cell can be grouped into three risk groups based on the degree of each cell's bad rate. The details are shown in Table 7, where the cell's bad rate exceeds 3% is categorized as high-risk group, the cell's bad rate falls between 1% and 3% is categorized as medium-risk group, and if the cell's bad rate is lower than 1%, it will be categorized as low-risk group. The purple zone is the

Table 7
Dual scoring model.

			Credit bureau score					
			200 or less (CBS1)	201-300 (CBS2)	301-450 (CBS3)	451-500 (CBS4)	501 and above (CBS5)	Total
Behavioral score	250 or less (BHS1)	Bad	224	68	36	11	6	345
		Total	2,623	1,101	977	455	309	5,465
		Bad rate	8.55%	6.16%	3.66%	2.44%	1.99%	6.31%
	251-350 (BHS2)	Bad	75	67	23	10	5	180
		Total	1,187	1,621	1,498	6,409	4,248	14,963
		Bad rate	6.33%	4.11%	1.56%	0.15%	0.12%	1.20%
	351-450 (BHS3)	Bad	47	28	15	10	6	106
		Total	791	1,014	1,140	8,066	5,618	16,629
		Bad rate	5.92%	2.79%	1.30%	0.12%	0.11%	0.64%
	451-550 (BHS4)	Bad	30	27	12	9	4	82
		Total	804	986	9,578	8,642	5,041	25,051
		Bad rate	3.68%	2.75%	0.13%	0.10%	0.07%	0.33%
	551 and above (BHS5)	Bad	17	15	7	6	0	45
		Total	576	899	9,434	7,921	5,689	24,520
		Bad rate	3.00%	1.65%	0.08%	0.08%	0.00%	0.18%
	Total	Bad	393	205	93	46	21	758
		Total	5,981	5,621	22,628	31,493	20,905	86,628
		Bad rate	6.57%	3.65%	0.41%	0.15%	0.10%	0.88%

Table 8
On-lending retaining strategy.

Group	Bad rate	Retention strategies	Interest rate
Low-risk group	0.10%	Accept	Lower interest rate for customer
Medium-risk group	2.00%	Accept	Maintain interest rate bounded by contract
High-risk group	5.83%	Reject	–

Table 9
Collection strategy.

Group	Bad rate	Collection strategies
Low-risk group	0.10%	Send a short message service to remind customer of payment
Medium-risk group	2.00%	1. Collection by junior collection personnel 2. Granting grace period (interest only) 3. Granting repayment program agreement
High-risk group	5.83%	1. Collection by senior collection personnel 2. Foreclosure auction

high-risk group with an average bad rate of 5.83%, the yellow zone is the mid-risk group with an average bad rate of 2.00%, and the green zone is the low-risk group with average bad rate of 0.10%. We can see from these three risk groups that there is significant risk segmentation. Hence, this study will use three risk groups generated from the dual scoring model for credit strategy applications.

4.1. On-lending retaining strategy

When the customers discover mortgage interest rate is higher than other banks, the customers will request loan transfer actively. Therefore, the bank can adopt the on-lending retaining strategy for different risk groups as shown in Table 8. If the customers belong to low-risk group, the banks can lower the mortgage interest rate to retain the customers and the level of decrease interest rates will be determined by the profit margin. If the customers belong to medium-risk group, the banks can remind the customers of bank transfer fees and recommend them to keep accounts with the current interest rate. If the customers belong to high-risk group, the banks can agree loan transfer to decrease the write-off loss of the bank.

4.2. Collection action strategy

When the customers are unable to pay the principal and interest within the payment period, then the bank will commence with collection procedures. At this time the bank can adopt the collection action strategy for different risk groups as shown in Table 9 due to cost considerations. For low-risk group, the customers may just occasionally forget to pay, the collection system can automatically send a short message service to remind customers of payment. For medium-risk group, the junior collection personnel will remind customers of payment in early stage collections. If the customer does not have sufficient ability to pay his debts, the bank can grant the customer a grace period or repayment program agreement for increasing the willingness of the customer to repay the debt. Finally, the high-risk group will be collected by the senior collection personnel or go into foreclosure auction to decrease the write-off loss of the bank.

5. Conclusion

Past researches mostly use the one-dimensional credit scoring model to segment customer credit risk. In this study, we combine the bank's internal behavioral scoring model with the external credit bureau scoring model to construct the dual scoring model for credit risk management of mortgage accounts. In addition, we classify the mortgage portfolio into three risk groups based on the degree of customer risk. By focusing attention on different risk groups, we are able to design corresponding collection actions and on-lending retaining strategies. For model validation, we apply the K–S statistic and ROC curve to measure the predictability of the credit scoring model. With regards to the K–S value, the dual scoring model increases the accuracy rate of prediction compared to both behavioral scoring model and credit bureau scoring model by 14.7% and 8.3%, respectively. With regards to the AUC value, the dual scoring model increases the accuracy rate of prediction compared to both behavioral scoring model and credit bureau scoring model by 9.0% and 6.8%, respectively. Overall, using the dual scoring model can more precisely and efficiently strengthen risk identification, assessment and management capability, which is indispensable risk avoidance tool for financial institutions in risk management.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609.
Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks. *Journal of Banking and Finance, 18*(3), 505–529.
Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks, 12*(4), 929–935.
Castillo, F., Marshall, K., Green, J., & Kordon, A. (2003). A methodology for combining symbolic regression and design of experiments to improve empirical model building. In *Genetic and evolutionary computation conference* (pp. 1975–1985).
Chen, M. C., & Huang, S. H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications, 24*(4), 433–441.
Chi, L. C., & Tang, T. C. (2007). Impact of reorganization announcements on distressed-stock returns. *Economic Modelling, 24*(5), 749–767.
Cholongitas, E., Senzolo, M., Patch, D., Shaw, S., Hui, C., & Burroughs, A. K. (2006). Review article: Scoring systems for assessing prognosis in critically ill adult cirrhotics. *Alimentary Pharmacology and Therapeutics, 24*(3), 453–464.
Chuang, R. J., & Chen, M. J. (2006). The building of credit scoring system on the residential mortgage finance. *Journal of Housing Studies, 15*(2), 65–90.
David, R. H., Edelman, D. B., & Gammerman, A. J. (1992). Machine learning algorithms for credit-card applications. *IMA Journal of Mathematics Applied in Business and Industry, 4*(1), 43–51.
Desai, V. S., Crook, J. N., & Overstreet, G. A. Jr., (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research, 95*(1), 24–37.
Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis, 16*(5), 471–495.
Feldman, D., & Gross, S. (2005). Mortgage default: Classification trees analysis. *The Journal of Real Estate Finance and Economics, 30*(4), 369–396.
Feraud, R., & Cleror, F. (2002). A methodology to explain neural network classification. *Neural Network, 15*(2), 237–246.

Holland, J. (1975). *Adaptation in natural and artificial systems.* The University of Michigan Press. Reissued by The MIT Press.
Jo, H., Han, I., & Lee, H. (1997). Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Systems with Applications, 13*(2), 97–108.
Joanes, D. N. (1993). Rejecting inference applied to logistic regression for credit scoring. *IMA Journal of Mathematics Applied in Business and Industry, 5*(4), 35–43.
Kim, J. C., Kim, D. H., Kim, J. J., Ye, J. S., & Lee, H. S. (2000). Segmenting the Korean housing market using multiple discriminant analysis. *Construction Management and Economics, 18*, 45–54.
Koh, H. C., Tan, W. C., & Goh, C. P. (2006). A two-step method to construct credit scoring models with data mining techniques. *International Journal of Business and Information, 1*(1), 96–118.
Laitinen, E. K. (1999). Predicting a corporate credit analyst's risk estimate by logistic and linear models. *International Review of Financial Analysis, 8*(2), 97–121.
Laitinen, E. K., & Laitinen, T. (2000). Bankruptcy prediction: Application of the Taylor's expansion in logistic regression. *International Review of Financial Analysis, 9*(4), 327–349.
Lee, T. H., & Jung, S. (2000). Forecasting creditworthiness: Logistic vs. artificial neural net. *The Journal of Business Forecasting Methods and Systems, 18*(4), 28–30.
Lee, T. S., Chiu, C. C., Chou, Y. C., & Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data Analysis, 50*(4), 1113–1130.
Li, X., Ying, W., Tuo, J., Li, B., & Liu, W. (2004). Applications of classification trees to consumer credit scoring methods in commercial banks. *IEEE International Conference on Systems, Man and Cybernetics, 5*, 4112–4117.
Mahlhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *OMEGA: The International Journal of Management Science, 31*(2), 83–96.
Mays, E. (2001). Handbook of credit scoring. *Global Professional Publishing, 23*, 56.
Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research, 112*(2), 310–321.
Setiono, R., Thong, J. Y. L., & Yap, C. S. (1998). Symbolic rule extraction from neural networks – An application to identifying organizations adopting IT. *Information and Management, 34*(2), 91–101.
Steenackers, A., & Goovaerts, M. J. (1989). A credit scoring model for personal loans. *Insurance: Mathematics and Economics, 8*(1), 31–34.
Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of the neural networks: The case of bank failure predictions. *Management Science, 38*(7), 926–947.
Thomas, L. C. (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting, 16*(2), 149–172.
Tsai, C. F., & Wu, J. W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 34*(4), 2639–2649.
Wang, M. L., Feng, Z. Y., & Chang, Y. H. (2010). A preliminary application of genetic algorithm on slope stability analysis. *Journal of Soil and Water Conservation, 42*(1), 65–82.
West, D. (2000). Neural network credit scoring models. *Computers and Operations Research, 27*(11/12), 1131–1152.
Westgaard, S., & Wijst, N. V. D. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research, 135*(2), 338–349.
Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis, 15*, 757–770.
Zhang, G., Hu, Y. M., Patuwo, E. B., & Indro, C. D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research, 116*, 16–32.
Zhu, H., Beling, P. A., & Overstreet, G. A. (2001). A study in the combination of two consumer credit scores. *Journal of the Operational Research Society, 52*(9), 974–980.