Expert Systems with Applications 38 (2011) 13274–13283

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Using data mining to improve assessment of credit worthiness via credit scoring models

Bee Wah Yap [a,*], Seng Huat Ong [b,1], Nor Huselina Mohamed Husain [a,2]

[a] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[b] Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Keywords:
Data mining
Credit scoring
Logistic regression
Decision tree
Classification
Predictive modeling

ABSTRACT

Credit scoring model have been developed by banks and researchers to improve the process of assessing credit worthiness during the credit evaluation process. The objective of credit scoring models is to assign credit risk to either a "good risk" group that is likely to repay financial obligation or a "bad risk" group who has high possibility of defaulting on the financial obligation. Construction of credit scoring models requires data mining techniques. Using historical data on payments, demographic characteristics and statistical techniques, credit scoring models can help identify the important demographic characteristics related to credit risk and provide a score for each customer. This paper illustrates using data mining to improve assessment of credit worthiness using credit scoring models. Due to privacy concerns and unavailability of real financial data from banks this study applies the credit scoring techniques using data of payment history of members from a recreational club. The club has been facing a problem of rising number in defaulters in their monthly club subscription payments. The management would like to have a model which they can deploy to identify potential defaulters. The classification performance of credit scorecard model, logistic regression model and decision tree model were compared. The classification error rates for credit scorecard model, logistic regression and decision tree were 27.9%, 28.8% and 28.1%, respectively. Although no model outperforms the other, scorecards are relatively much easier to deploy in practical applications.

1. Introduction

Credit scoring models are very useful for many practical applications especially for banks and financial institutions. The decision-making process of accepting or rejecting a client's credit by banks is commonly executed via judgmental techniques and/or credit scoring models. Most banks and financial institutions use the judgmental approach which is based on the 3C's, 4C's or 5C's which are character, capital, collateral, capacity and condition. Credit scoring is a system creditors use to assign credit applicants to either a "good credit" one that is likely to repay financial obligation or a "bad credit" one who has a high possibility of defaulting on financial obligation. Generally, Linear Discriminant Analysis and logistic regression are two popular statistical tools to construct credit scoring models (Abdou, Pointon, & El-Masry, 2008; Desai, Crook, & Overstreet, 1996; Gao, Zhou, Gao, & Shi, 2006; Hand & Henley, 1997; Thomas, 2000; Vojtek & Kocenda, 2006). However, with the advance in information and computer technology new techniques are appearing under the name of data mining. Data mining software such as SAS® Enterprise Miner and SPSS PASW® 13 modeler provide not only the classical methods but new novel predictive modeling and classification techniques such as decision tree, neural networks, support vector machine (SVM), and *k*-nearest neighbors.

Although credit scoring methods are widely used for loan applications in financial and banking institutions, it can be used for other type of organizations such as insurance, real estate, telecommunication and recreational clubs for predicting late payments. For example, Gschwind (2007) showed a data mining application in real estate for predicting late payments by tenant. Due to privacy concerns and unavailability of data from banks, for this paper, historical payment of monthly subscription from members of a local recreational club was used. Payment of the monthly subscription fee is an obligation of the club members besides paying the permanent membership fee. The management faces the problem in the rising number of defaulters. So far, there has been no significant effort to improve cash flow by proactively predicting non-payments using quantitative methods, and taking corrective actions before a late payment happened. Discussion with the

* Corresponding author. Tel.: +60 03 55435461; fax: +60 03 55435501.
E-mail addresses: beewah@tmsk.uitm.edu.my, yapbeewah@salam.uitm.edu.my (B.W. Yap), ongsh@um.edu.my (S.H. Ong), huselina_11@yahoo.com (N.H.M. Husain).
[1] Tel.: +60 03 79674306; fax: +60 03 79674143.
[2] Tel.: +60 03 55435461; fax: +60 03 55435501.

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.04.147

management of the club revealed that they use judgmental techniques to determine the defaulters or non-defaulters and whether to terminate the membership of defaulters. The main source of income for most recreational clubs is the membership monthly payments. A large number of defaulters will result in cash flow problem and loss of income for the club. This will affect the financial planning of the club activities and the management faces the problem of ensuring that the club does not go bankrupt. The objective of this paper is to illustrate the use of data mining in assessing credit worthiness using credit scoring models and for prediction of an event such as default in payment so that early intervention can be done to prevent financial loss.

This paper is organized as follows. Section 2 provides a review of the applications of data mining and credit scoring models. Then, the conceptual framework is presented. The methodology for constructing the credit scoring models is covered in Section 3. The results are discussed in Section 4. Finally, the limitations of the data mining approach to the construction of credit scoring models are highlighted in the concluding section.

2. Literature review

2.1. Data mining

Data mining refers to the extraction of useful patterns or rules from a large database. The data mining process involves identifying the business problem and data mining goal, retrieving the database needed, and using data mining techniques to analyze the data with the final aim of achieving important results for making strategic decisions (Berry & Linoff, 2004). Data mining is an integral part of knowledge discovery in databases (KDD). Data mining involves techniques such as anomaly detection, association analysis, clustering, and predictive modeling (Berry & Linoff, 2004; Han & Kamber, 2001; Tan, Steinbach, & Kumar, 2006). Anomaly detection involves algorithm that can discover real anomalies such as detection of fraud, unusual disease, unusual weather conditions and ecosystem disturbances. Association analysis enables the discovery of group of objects that occur frequently together such as items that are often bought together or genes that are similar. Clustering involves segmentation of objects in a large database into clusters or segments with similar characteristics. Such segmentation is especially useful for target marketing. Predictive modeling involves using statistical models, machine-learning technique such as decision tree algorithm or artificial intelligence model to predict the outcome of a dependent variable based on several attributes (or independent variables). Data mining has been applied in many fields such as banking, finance, telecommunication, manufacturing, healthcare, insurance, real estate, education, marketing, customer relationship management and weather study such as avalanche forecasting (Abdou et al., 2008; Ang, Chua, & Bowling, 1979; Chien & Chen, 2008; Chien, Hsiao, & Wang, 2004; Chien, Wang, & Chen, 2005; Cho & Ngai, 2003; Davis, Elder, Howlett, & Bouzaglou, 1999; Gschwind, 2007; Kurt, Ture, & Kurum, 2008; Lee, Chiu, Chou, & Lu, 2006; Rygielski, Wang, & Yen, 2002).

2.2. Credit scoring models

Credit scoring was first introduced in the 1940s and over the years had evolved and developed significantly. In the 1960s, with the creation of credit cards, banks and other credit card issuers realized the advantages of credit scoring in the credit granting process. In the 1980s, credit scoring was used for other purposes such as aiding decision in approving personal loan applications. In recent years, credit scoring has been used for home loans, small business loans and insurance applications and renewals (Koh, Tan, & Goh, 2004; Thomas, 2000). A credit scoring model provides an estimate of a borrower's credit risk – i.e. the likelihood that the borrower will repay the loan as promised, based on a number of quantifiable borrower characteristics (Dinh & Kleimeier, 2007). Credit scoring is based on statistical or operational research methods. Historically, discriminant analysis and linear regression have been the most widely used techniques for building scorecards. Other techniques include logistic regression, probit analysis, nonparametric smoothing methods especially *k*-nearest neighbors, mathematical programming, Markov chain models, recursive partitioning, expert systems, genetic algorithms and neural networks (Hand & Henley, 1997).

Artificial neural networks (ANNs) have been critized for its 'black box' approach and interpretative difficulties. Multivariate adaptive regression splines (MARS), classification and regression tree (CART), case based reasoning (CBR), and support vector machine (SVM) are some recently developed techniques for building credit scoring models. Huang, Chen, Hsu, Chen, and Wu (2004) investigated the performance of the SVM approach in credit rating prediction in comparison with back propagation neural networks (BNN). However, only slight improvement of SVM over BNN was observed. Huang, Chen, and Wang (2007) reported that compared with neural networks, genetic programming and decision tree classifiers, the SVM classifier achieved identical classification accuracy with relatively few input variables.

Lee et al., 2006 demonstrated the effectiveness of credit scoring using CART and MARS. Their results revealed that, CART and MARS outperform traditional discriminant analysis, logistic regression, neural networks, and support vector machine (SVM) approaches in terms of credit scoring accuracy. Recently, with the development of data mining software the process involved in building credit scoring model is made much easier for credit analysts. Despite the development of new novel techniques, for practical applications the popular techniques for banking and business enterprises are credit scorecards, logistic regression and decision trees as it is relatively easy to identify the important input variable, interpret the results and deploy the model.

2.3. Conceptual framework

In building a scoring model, or "scorecard", historical data on the performance of previously made loans and borrowers characteristics are required. A good scoring model should give a higher percentage of high scores to 'good borrowers' and a higher percentage of low scores to those who are 'bad borrowers'. Ang et al. (1979) investigated the profiles of late-paying consumer loan borrowers using variables such as gross amount of loan, age, sex, marital status, number of dependents, years lived at residence, monthly take home pay, monthly take home pay of spouse, own or rent residence, other monthly income, total monthly payments on all debts, type of bank accounts, number of credit references listed, years on job, total family monthly income per month, debt to income ratio, total number of payments on the loan, and annual percentage interest on the loan. Koh et al. (2004) used age, annual income, gender, marital status, number of children, number of other credit cards held and whether the applicant has an outstanding mortgage loan to construct a credit scoring model to predict credit risk of credit card applicants as bad loss, bad profit and good risk. Abdou et al. (2008) used twenty variables some of which were loan amount, loan duration, sex, marital status, age, monthly salary, additional income, house owned or rent, and education level for building credit scoring models to evaluate credit risk (paid or default) for personal loan. Gschwind (2007) concluded that mining basic tenant data, accounts receivable data, and government-published data can generate predictions of late payments of rental. Mavri, Angelis, and Loannou (2008) used variables such as gender,

age, education, marital status and monthly income to estimate the risk level of credit card applicants. Vojtek and Kocenda (2006) provided a table of indicators that are typically important in retail credit scoring models. They classify the indicators as demographic, financial, employment and behavioral indicators.

The abovementioned studies serve as a guide in determining the conceptual framework of this study. The behavioral and the financial indicators data were not available. Hence, only the demographic and employment indicators were considered. The number of cars owned is used as a proxy of financial indicator. The demographic indicators consist of age, gender, marital status, district of address, race, and number of dependents. These variables typically are useful in capturing various regional, gender, and other relevant differences. For example, it is often found that older women are less risky than young men. In general, the risk of default decreases with age and is also higher for married applicants with dependants. Relations like this can help to better discriminate between good/bad applicants (Vojtek & Kocenda, 2006). The employment indicators consist of occupation and their working sector. The variables and their categories are shown in the conceptual framework in Fig. 1.

3. Methodology

This section explains the process of constructing credit scoring models.

3.1. Variable role identification

The credit scoring models were built using SAS® Enterprise Miner 5.3. In building credit scoring model, the variable role (input or independent variable and target which is the dependent variable) and measurement level must be stated. The target (dependent) variable of interest is payment status, a binary variable with two categories: defaulter or non-defaulter. Defaulters are those who failed to pay their monthly subscription fee for three months consecutively. Table 1 lists the variables, their role and measurement level.

3.2. Data preparation

First, the data given which was in text format was converted into Excel format. Next, we proceed with the data cleaning stage.

Table 1
List of variables in data set.

Variable description	Variable name	Role	Measurement level
Gender (M = Male, F = Female)	Gender	Input	Nominal
Age in years	Age	Input	Interval
District of address	DisAdd	Input	Nominal
Occupation	Occupation	Input	Nominal
Race	Race	Input	Nominal
Marital status	Mar_Status	Input	Nominal
Number of dependents	NoOfdepend	Input	Interval
Number of cars	NoOfcars	Input	Interval
Defaulters/non-defaulters	Status	Target	Binary
Work sector	Sector	Input	Binary

In general, data cleaning involves removing outliers, redundant data and imputation of missing values. There were a few missing values for age and gender and the values were determined from the member's Identification Card Number. The highest number of missing values were for numbers of cars (10.5%), followed by district of address (6.3%) and number of dependents (2.5%). The telephone number was used to identify the district of address, while missing values for number of cars and number of dependents was replaced using the median value. Once the data cleaning process was completed, the Excel data file was converted into a SAS file using SAS Enterprise Guide.

3.3. Data modeling

This section discusses the construction of the credit scorecard model, logistic regression model and decision tree model. The data consists of 977(35%) defaulters and 1788(65%) non-defaulters. The dependent variable, which is payment status is coded as Defaulter = 1 and Non Defaulter = 0. SAS® Enterprise Miner 5.3 software was used for building the credit scorecard models. This data mining software provides a graphical-user-interface (GUI) workspace whereby nodes (tool–icon) can be easily selected from a tools palette and placed into the diagram workspace. Nodes are then connected to form a process flow diagram that structure and document the flow of analytical activities. Fig. 2 shows the process flow diagram and the process begin with the sample data node. The sample data node is connected to the Data Partition node to split the data into training and validation sample. The sample data was partitioned at 70:30 that is 70% for training (used for model building) and 30% for validation sample. The three credit scoring

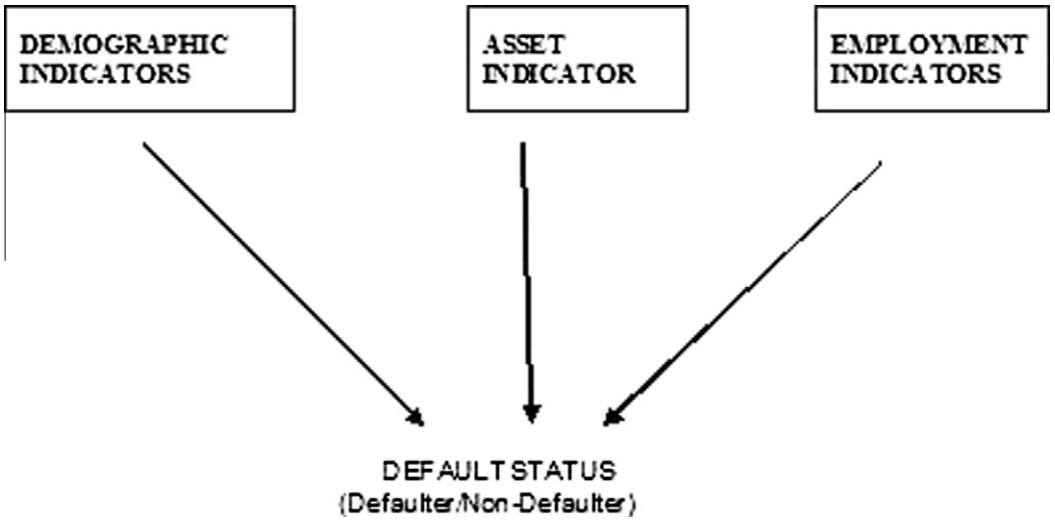

Fig. 1. Conceptual framework.

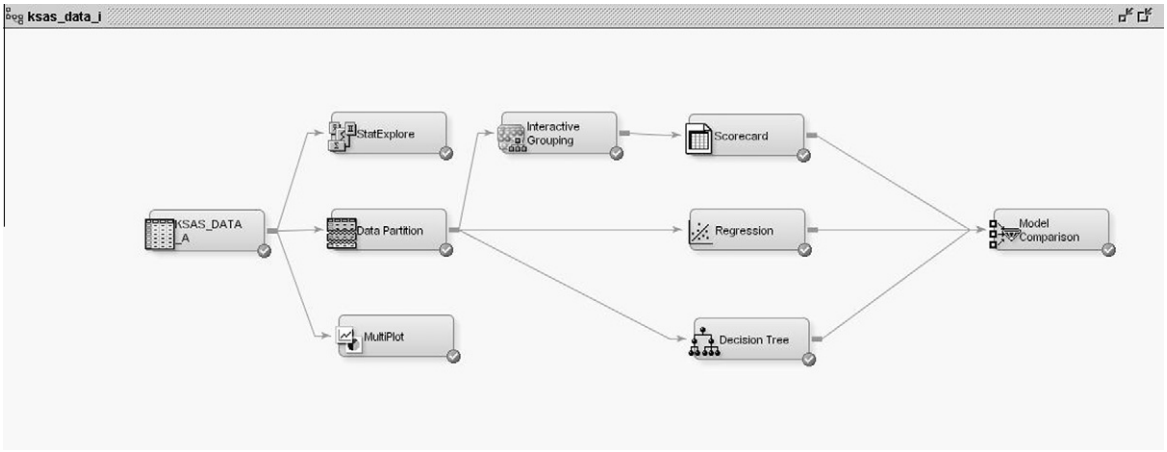

Fig. 2. Data mining process flow diagram.

models, the scorecard model (consisting of Interactive Grouping node and Scorecard node), Logistic Regression node and Decision Tree node were connected to the Data Partition node. The predictive performance of the three credit scoring models were then assessed and compared using the Model Comparison node. Meanwhile, the Multiplot and StatExplore node is a multipurpose tool used to examine variable distributions and statistics of the variables.

3.3.1. Credit scorecard model

SAS® Enterprise Miner 5.3 provides a Scorecard node that enables the creation of a credit-scorecard model. Classing in Enterprise Miner refers to the process of automatically and/or interactively binning interval and grouping nominal or ordinal input variables in order to manage the number of attributes per variable and to improve the predictive power of the variable. Classing in Enterprise Miner takes place in the Interactive Grouping node. This node has been specifically developed for credit scorecard applications. The Interactive Grouping node classes and selects variable automatically and interactively using weights of evidence (WOE) and information value (IV) measures. The Weight of Evidence (WOE) of an attribute is defined as the logarithm of the ratio of the proportion of 'goods' in the attribute over the proportion of 'bads' in the attribute. High negative values therefore correspond to high risk, high positives values correspond to low risk. The WOE of an attribute is computed as follows (SAS Institute Inc, 2009):

$$\text{Weight of Evidence}_{\text{attribute}} = \log\frac{p_{\text{goodattribute}}}{p_{\text{badattribute}}} \quad \text{where}$$
$$P_{\text{badattribute}} = \frac{\#\ \text{goods}_{\text{attribute}}}{\#\ \text{goods}} \quad \text{and}$$
$$P_{\text{badattribute}} = \frac{\#\ \text{bads}_{\text{attribute}}}{\#\ \text{bads}}. \tag{1}$$

After the classing stage, the variable predictive power, i.e., its ability to separate high risks from low risks was assessed with the information value (IV) measure. This aided the selection of variable for inclusion in the credit scorecard model. The information value is the weighted sum of the weights of evidence of the variable's attributes. The weights are the difference between proportion of 'goods' and the proportion of 'bads' in the respective attribute. The IV is computed as:

$$\text{Information Value} = \sum\left(p_{\text{goodattribute}} - p_{\text{badattribute}}\right) * \text{WeightofEvidence}. \tag{2}$$

The information value should be greater than 0.02 for a variable to be considered for inclusion in the credit scorecard model. The Scorecard node then makes use of the WOE and IV to generate the scorecard (SAS Institute Inc., 2009).

3.3.2. Logistic regression model

Logistic regression is a widely used statistical modeling technique in which the probability of a dichotomous outcome ($Y = 0$ or $Y = 1$) is related to a set of potential predictor variables. The logistic regression model is written as:

$$\log\left[\frac{P(Y=1)}{1-P(Y=1)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k, \tag{3}$$

where $P(Y = 1)$ is the probability of the outcome of interest.

Eq. (3) can be solved to obtain:

$$P(Y=1) = \frac{1}{1+e^{-z}}, \quad \text{where } z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k. \tag{4}$$

Thus, the objective of a logistic regression model in credit scoring is to determine the conditional probability of a specific applicant belonging to a class (defaulter or non-defaulter), given the values of the independent variables of that credit applicant. For this study, the logistic regression was used to model the event $Y = 1$ (defaulter).

3.3.3. Decision tree model

A decision tree model consists of a set of rules for dividing a large collection of observations into smaller homogeneous group with respect to a particular target variable. The target variable is usually categorical and the decision tree model is used either to calculate the probability that a given record belongs to each of the target category, or to classify the record by assigning it to the most likely category. Decision tree can also be used for continuous target variable although there are other techniques such as multiple regressions which are more suitable for such variable (Berry & Linoff, 2004). Given a target variable and a set of explanatory variables, decision algorithms automatically determine which

Table 2
Partitioned data for training and validation.

	Defaulter	Non-defaulter	Total
Training	683	1252	1935 (70%)
Validation	294	536	830 (30%)
Total	977 (35%)	1788 (65%)	2765 (100%)

Table 3
Grouping results using IGN.

Variable	Group	Group values	Event (default) count	Non-event (not default) count
Number of dependent	1	6, 7, 8, 9	14	82
	2	5	37	148
	3	2, 3, 4	359	824
	4	1	119	147
	5	0	154	51
Number of cars	1	3	184	239
	2	2	361	635
	3	1, 6	105	192
	4	4, 5, 7	33	186
District of address	1	Other	96	65
	2	KL	126	162
	3	Petaling, Jaya	387	845
	4	Klang	74	180
Age	1	Age < 32	154	187
	2	32 ⩽ Age < 37	226	338
	3	37 ⩽ Age < 46	223	481
	4	46 ⩽ Age < 53	61	167
	5	53 ⩽ Age	19	79

Table 4
Information value.

Variables	Information value	Information value ordering	Information value decision
Number of dependents	0.471	1	Input
Number of cars	0.141	2	Input
District of address	0.137	3	Input
Age	0.100	4	Input
Occupation	0.083	5	Rejected
Race	0.025	6	Rejected
Marital status	0.014	7	Rejected
Gender	0.007	8	Rejected
Sector	0.000	9	Rejected

variables are most important, and subsequently sort the observations into the correct output category (Olson & Yong, 2006). The common decision tree algorithms in data mining software are chi-square automatic interaction detector (CHAID), classification and regression tree (CART) and C5. CART uses gini as the splitting criteria while C5 uses entropy. Meanwhile, CHAID uses chi-square as the splitting criteria (Berry & Linoff, 2004).

4. Results

4.1. Background of sample

The target (dependent) variable is payment *status*, a binary variable with 2 categories: defaulter and non-defaulter which were coded using numerical values (1 and 0). Out of 2765, 35% was defaulters while 65% was non-defaulters. The majority of the members are male (80%) and more than half (74%) of the members are from non-government sector. A high majority of the members (92%) are married and almost half of the members (49%) are Chinese. More than half (64%) of the members reside in Petaling Jaya while 39% of the members are in the management team of their company. About 51% own two cars and more than half (61%) of the members are between 30 to 40 years old. Majority (74%) are working in non-government sectors while 26% are from government sector.

The number of defaulters for the training and validation samples as shown in Table 2 is 683 and 294 which is 35%, respectively.

4.2. Credit scorecard model

Table 3 shows the grouping for input variables selected by the Interactive Grouping node (IGN).

4.2.1. Information Value (IV)

The information value should be greater than 0.02 for a characteristic (or variable) to be considered for inclusion in the scorecard model. information values lower than 0.1 can be considered weak, smaller than 0.3 is medium and smaller than 0.5 is strong. If the information value is greater than 0.5, the characteristics may be over-predicting, meaning that it is in some form trivially related to the good/bad information (SAS Institute Inc., 2009). Based on the information values shown in Table 4, four variables with information value 0.1 and above were selected. The variables are number of dependents, number of cars, district of address, and age. Meanwhile occupation, race, marital status, gender and sector were rejected.

Table 5
WOE and score points.

Variable accepted	Group	Attributes	Weight of evidence	Coefficient of regression	Score points
Age	1	Age < 32	**−0.41185**	−0.66	95
	2	32 ⩽ Age < 37	**−0.20349**	−0.66	101
	3	37 ⩽ Age < 46	0.16269	−0.66	111
	4	46 ⩽ Age < 53	0.40112	−0.66	118
	S	53 ⩽ Age	0.81901	−0.66	130
Number of dependents	1	6, 7, 8, 9	1.16166	−0.90	l52
	2	5	0.78029	−0.90	137
	3	2, 3, 4	0.22485	−0.90	115
	4	1	**−0.39469**	−0.90	91
	5	0	**−1.71113**	−0.90	40
Number of cars	1	3	**−0.34447**	−0.75	95
	2	2	**−0.04126**	−0.75	105
	3	1, 6	**−0.00247**	−0.75	107
	4	4, 5.7	1.12324	−0.75	143
District of address	1	KL	**−0.35469**	−0.90	93
	2	Petaling Jaya	0.17491	−0.90	113
	3	Klang	0.28289	−0.90	118
	4	Other	**−0.99596**	−0.90	68

Bold Values is significant at $p < 0.01$ (obtained from Table 6)

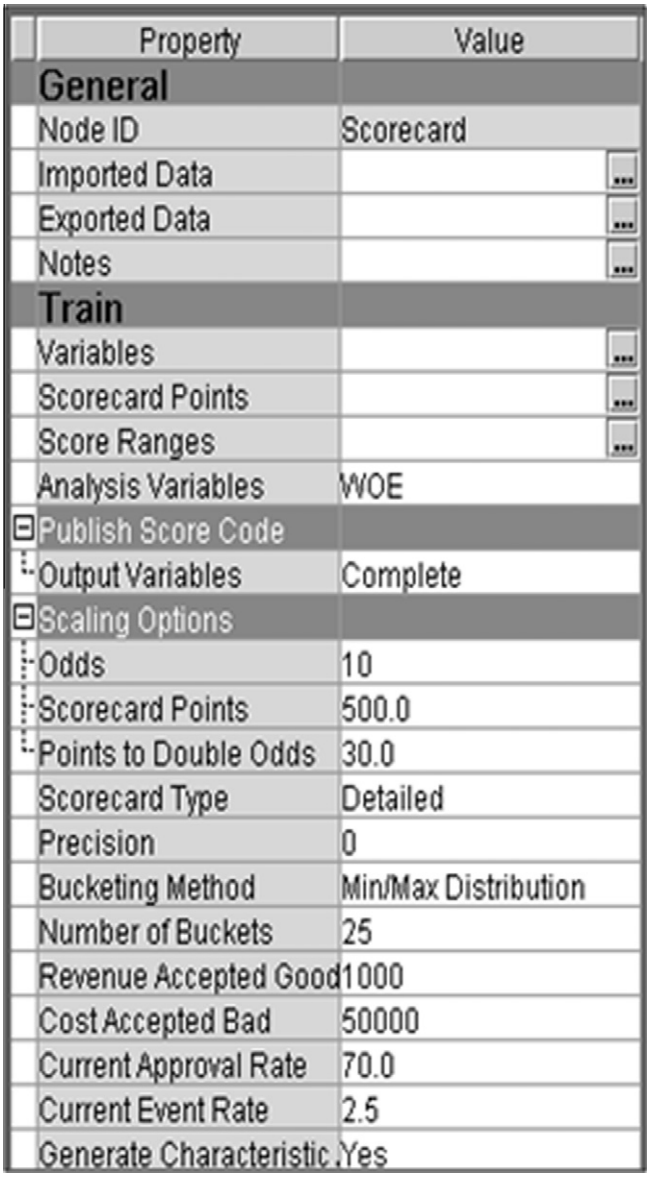

Fig. 3. Screenshot of scorecard node property.

Table 6
Logistic regression results from scorecard node.

Parameter	DF	Estimate	Standard error	Wald chi-square	Pr > chi-square
Intercept	1	−0.6062	0.0517	137.25	<.0001
WOE_DisAdd	1	−0.8970	0.1368	42.96	<.0001
WOE_NoOfcars	1	−0.7546	0.1534	24.19	<.0001
WOE_NoOfdepend	1	−0.8978	0.0785	130.65	<.0001
WOE_Age	1	−0.6550	0.1640	15.95	<.0001

4.2.2. Weight of evidence (WOE)

The *relative risk* of an attribute is determined by its Weight of Evidence (WOE). This value depends on the value of the binary target variable, which is either event or non-event. Here, the bad attribute is defaulter while non-defaulter is the good attribute. The Weight of Evidence of an attribute is the logarithm of the ratio of the proportion of "goods" in the attribute over the proportion of "bads" in the attribute. A high negative WOE value corresponds to high risk while high positive values correspond to low risk. The WOE for each of the variable, the attributes, the regression coefficients and the score points generated by the Scorecard node are as depicted in Table 5. WOE results indicate that for the age variable, the highest risk is those in Group 1 (Age < 32) and Group 2 (32–37) while members in Group 4 (53 or older) have lowest risk to default. Members who have none or only one dependent are at higher risk of defaulting in payment. Results also show that those who own many cars recorded a lower risk to default and those living in Kuala Lumpur (KL) and other places have higher risk to default.

Table 7
Scorecard results.

Variable	Group value	Group	Scorecard points	Weight of evidence
Age	Age < 32	1	95	−0.41
	32 ⩽ Age < 37	2	101	−1.20
	37 ⩽ Age < 46	3	111	0.56
	46 ⩽ Age < 53	4	118	0.40
	53 ⩽ Age	5	130	0.82
DisAdd	OTHER	1	68	−1.00
	KL	2	93	−0.35
	Petaling Jaya	3	113	0.17
	Klang	4	118	0.28
NoOfcars	3	1	95	−0.34
	2, Missing	2	105	−0.04
	1, 6	3	107	0.00
	4, 5, 7	4	143	1.12
NoOfdepend	6, 7, 8, 9	1	152	1.16
	5	2	137	0.78
	2, 3, 4	3	115	0.22
	1	4	91	−0.39
	0	5	40	−1.71

Table 8
Type 3 Analysis of Effects.

Parameter	DF	Wald chi square	Pr > chi-square
Age	1	19.9240	<0.0001
District of address	3	34.6386	<0.0001
Gender	1	12.5617	0.0004
Marital status	1	4.6713	0.0307
Number of cars	1	5.7742	0.0163
Number of dependents	1	130.2572	<0.0001
Occupation	4	24.1125	<0.0001
Race	3	15.8020	0.0012
Sector	1	5.5217	0.0188

Fig. 3 shows the Scorecard node property. The scorecard points were calculated using the following equations:

$$\text{score} = \log(\text{odds}) * \text{factor} + \text{offset},$$
$$\text{factor} = \text{points to double odds} / \log 2,$$
$$\text{offset} = \text{score} - \text{factor} * \log(\text{odds}).$$

Based on Fig. 3, the values of odds = 10, value of scorecard points = 500, value of points to double odds = 30. Therefore, factor = 43.28 and offset = 400.34.

As an example, the calculations of score points for Group 1 of the *Age variable* are as follows:

WOE_i is obtained from Table 5 while β_i and α figures are extracted from Table 6:

$$\begin{aligned}\text{Scorepoints} &= -\left(\text{woe}_i * \beta_i + \frac{\alpha}{n}\right) * \text{factor} + \frac{\text{offset}}{n} \\ &= -\left(-0.4118 * -0.655 + \frac{-0.6062}{4}\right) * 43.2808 \\ &\quad + \frac{400.3422}{4} = 94.9693 \approx 95.\end{aligned}$$

Table 9
Maximum likelihood estimates.

Parameter	DF	Estimate	Standard error	Waid chi-square	Pr > chi-square	Exp (Est)
Intercept	1	2.0341	0.3464	34.48	<.0001	7.645
Age	1	−0.0335	0.0075	19.92	<.0001	0.967
Dis Add KL	1	0.0668	0.1123	0.35	0.5520	1.069
Dis Add Klang	1	−0.3499	0.1250	7.83	0.0051	0.705
Dis Add Other	1	0.6413	0.1385	21.45	<.0001	1.899
Gender F	1	−0.2700	0.0762	12.56	0.0004	0.763
Mar_Status M	1	0.2355	0.1090	4.67	0.0307	1.266
NoOfcarS	1	−0.1272	0.0529	5.77	0.0163	0.881
NoOfdepend	1	−0.4201	0.0368	130.26	<.0001	0.657
Occupation						
Businessman/politician/pensioner	1	−0.4278	0.1537	7.75	0.0054	0.652
Education and finance services	1	0.4037	0.1366	8.73	0.0031	1.497
Management	1	0.2607	0.0912	8.17	0.0043	1.298
Other	1	−0.2195	0.1169	3.53	0.0604	0.803
Race A	1	0.0953	0.1222	0.61	0.4353	1.100
Race B	1	−0.2104	0.1193	3.11	0.0777	0.810
Race C	1	−0.5092	0.1605	10.07	0.0015	0.601
Sector G	1	0.1694	0.0721	5.52	0.0188	1.185

Hence, for attribute Age < 32, the score points is 95. The other score points are depicted in Table 7. A member score is obtained by summing the attributes scorecard points. Lower scores denote higher risk. For example, the total score for a member who is 55 years old, stays in Klang, has four cars and six dependents will be 543.

4.2.3. Determination of cut off score (threshold) value

According to Siddiqi (2005), the cut off score (threshold) can be determined by the value of Kolmogorov Smirnov statistic or K–S test for each bucket of score in the validation sample. The K–S test measures the distance between the distribution functions of the two classifications (e.g., good and bad credit risks). The score that generates the greatest separability (highest K–S test) between the good and bad is considered to be the threshold value. Based on the scorecard node results, the highest K–S test value is 0.3333 and the score is 415. Therefore, the score of 415 is the cut off score (threshold). In other words, 415 is the minimum acceptable level of risk. Hence, members with total score above 415 will be classified as non-defaulter.

4.3. Logistic regression model

There are three variable selection methods available in the Regression node of SAS® Enterprise Miner which are forward, backward and stepwise method. For this study, stepwise selection method was used. The SAS result of the logistic regression using the stepwise selection method is presented in Tables 8 and 9.

Based on Table 8, all nine variables were selected and are significant predictors. The odds-ratio_results in Table 10 indicate that older members are less likely to default. Female members are also less likely to default. The odds ratio for marital status show that married person are more likely to default compared to those who are single while members who own more cars (financial indicator) are less likely to default. Results also show that members who are in the education and finance services are more likely to default.

4.4. Decision tree model

A decision tree model using chi-square as the splitting criteria was built using SAS® Enterprise Miner 5.3. A decision tree is a structure that can be used to divide up a large collection of records into successfully smaller sets of records by applying a sequence of simple decision rules. The decision tree rules are shown in Table 10 while the decision tree is as depicted in Fig. 4.

Table 10
Decision tree classification rules.

```
IF Mar_Status EQUALSM
AND NoOfdepend EQUALS 0
THEN
  NODE : 5
  N : 98
  0 : 1.0%
  1 : 99.0%
IF DisAdd EQUALS OTHER
AND NoOfdepend IS ONE OF: 1 3 2 4 6 5 8 7 9
THEN
  NODE : 6
  N : 141
  0 : 41.1%
  1 : 58 9%
IF DisAdd IS ONE OF: KL KLANG FETALING
AND NoOfdepend IS ONE OF: 1 3 2 4 6 5 8 7 9
THEN
  NODE : 7
  N : 1589
  0 : 71.9%
  1 : 28.1%
IF NoOfcars IS ONE OF: 3 4 5 6 7 8 9 10
AND Mar_Status EQUALS S
AND NoOfdepend EQUALS 0
THEN
  NODE : 9
  N : 34
  0 : 2.9%
  1 : 97.1%
IF Gender EQUALS M
AND NoOfcars IS ONE OF: 1 2
AND Mar_Status EQUALS S
AND NoOfdepend EOUALS 0
THEN
  NODE : 14
  N : 16
  0 : 37.5%
  1 : 62.5%
IF Gender EQUALS F
AND NoOfcars IS ONE OF: 1 2
AND Mar_Status EQUALS S
AND NoOfdepend EQUALS 0
THEN
  NODE : 15
  N : 57
  0 : 75.4%
  1 : 24.6%
```

Note: 1 = defaulter; 0 = non-defaulter.

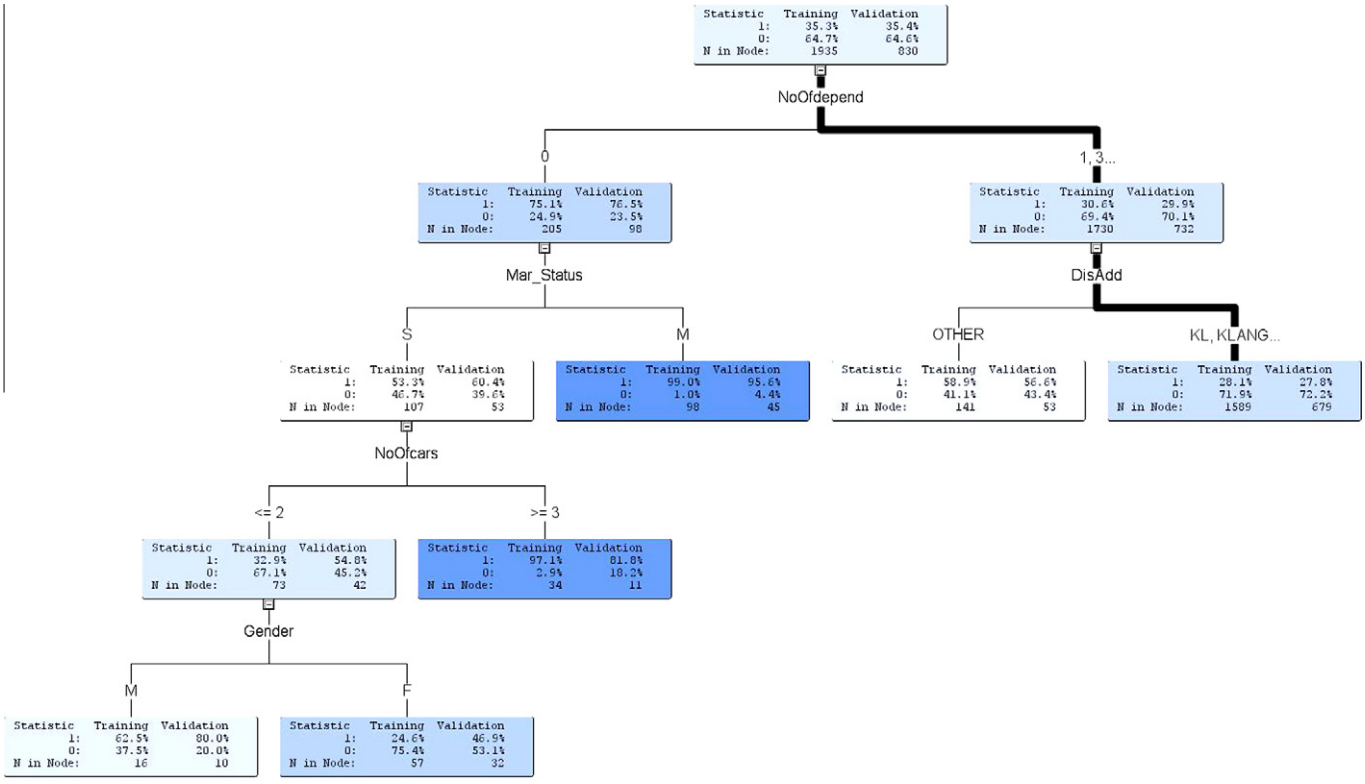

Fig. 4. Decision tree model.

Table 11
Event classification table.

Model	Model description	Data role	False negative	True negative	False positive	True positive
Reg	Regression	Train	398	1099	153	285
Reg	Regression	Validate	175	472	64	119
Tree	Decision tree	Train	460	1186	66	223
Tree	Decision tree	Validate	204	507	29	90
Scorecard	Credit scorecard	Train	440	1145	107	243
Scorecard	Credit scorecard	Validate	181	485	51	113

Table 12
Misclassification rate for training and validation sample.

Selected model	Model node	Model description	Training misclassification rate	Validation misclassification rate
Y	Scorecard	Credit scorecard	0.2847	0.2795
	Tree	Decision tree	0.2728	0.2807
	Reg	Logistic regression	0.2827	0.2879

Note: Y indicates the selected model.

4.4.1. Interpretation of the classification rules results

(1) If the members are married and number of dependents is 0, then they tend to default.
(2) If district of address is 'OTHER' and number of dependents is more than 0, then the members tend to default.
(3) If district of address is one of Kuala Lumpur, Klang or Petaling Jaya and number of dependents is more than 0 then the members tend not to default.
(4) If the members have more than 2 cars, single and number of dependents is 0, then they tend to default.
(5) If the members are male, have 1 or 2 cars, single and number of dependents is 0, then they tend to default.
(6) If the members are female, have 1 or 2 cars, single and number of dependents is 0, they are less likely to default.

From the above classification rules, the profile of defaulters are:

(i) Married with no children.
(ii) Single with more than 2 cars.
(iii) Male, single with 1 or 2 cars.
(iv) Members staying outside of Selangor.

4.5. Model comparisons

In this section comparison of the performance of the models is discussed. The models were compared based on validation misclassification rate, ROC chart and Type I and Type II errors. The misclassification rate is the percentage of misclassified observation by the model. Table 11 shows the event classification table for the

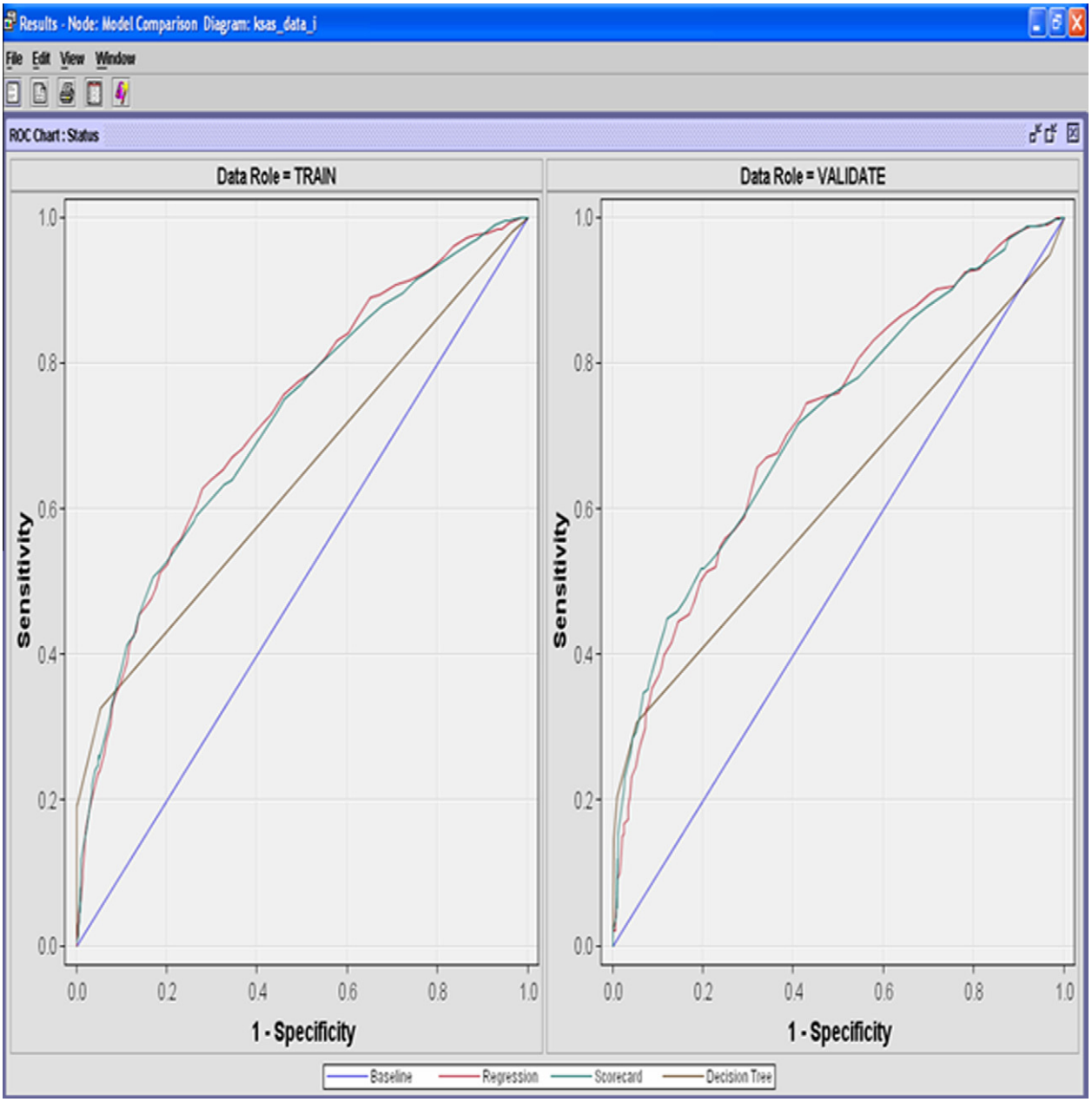

Fig. 5. ROC chart for training and validation sample.

Table 13
Sensitivity, specificity and misclassification rate.

Model	Sample	Sensitivity	Specificity	Misclassification rate	Type I error	Type II error
Logistic	Training	**0.4173**	0.8777	0.2848	12.2	**58.3**
Regression	Validation	**0.4048**	0.8806	0.2879	11.9	**59.5**
Decision	Training	0.3265	0.9473	0.2718	5.3	67.3
Tree	Validation	0.3061	0.9459	0.2807	5.4	69.4
Credit	Training	0.3558	0.9145	0.2827	8.5	64.4
Scorecand	Validation	0.3843	0.9049	0.2795	9.5	61.5

three credit scoring models for both the training and validation sample.

A summary of the misclassification rates and the selected model by the Model Comparison node are shown in Table 12. The three validation misclassification rates are quite similar and the credit scorecard model is the selected model as it has the lowest misclassification rate of 27.95%. Based on the validation misclassification rates, the accuracy rate for, credit scorecard model, decision tree and logistic regression is 72.0%, 71.2%, 71.9%, respectively.

The receiver operating characteristic (ROC) chart graphically displays sensitivity (percentage of the defaulters predicted correctly as defaulters) versus 1-specificity (percentage of the non-defaulters wrongly classified as defaulters), or the ratio of the true

positive rate versus the false positive rate. Based on Fig. 5, it can be see that there are four lines in the ROC graph. The straight line is the baseline (blue[3] color) and the logistic regression is represented by the red (highest) line. The second highest line (in green color) is for the scorecard model while the lowest line (brown line) is for decision tree. The baseline indicates that at each point of the line, the percentage of the defaulters predicted correctly as defaulters (true positive rate) is equivalent with the percentage of the non-defaulters wrongly classified as defaulters (false positive rate). The ROC charts show that the credit scorecard model and logistic regression model is quite comparable and outperforms the decision tree model. Beside ROC chart, the models also were assessed by their Type I and Type II error.

Table 13 displays the sensitivity, specificity and the misclassification rate for each model. The sensitivity rate is the true positive rate (the percentage of defaulters predicted correctly as defaulters) while specificity is the true negative rate (percentage of the non-defaulters predicted correctly as non-defaulters).

In comparing predictive models, we need to look at the Type I error (a good credit customer being misclassified as bad credit customer) and Type II error (a bad credit customer being misclassified as a good credit customer) of the models. The misclassification costs associated with Type II errors are much higher than those associated with Type I error (West, 2000). Based on Table 13, it is found that the Logistic regression model has the highest sensitivity and the lowest Type II error (a defaulter misclassified as non-defaulter). The decision tree is the worst model as it has the highest Type II error and the lowest sensitivity. However, it has to be noted that although TYPE I errors are low for all the three credit scoring models while the Type II errors are high.

5. Conclusion

For the past decade, the availability and high computing capability of data mining software enables business organizations to analyze and gain useful information from their large customer database. The main data mining techniques are predictive modeling, classification, cluster analysis and association (a.k.a market and basket) analysis. These techniques are highly useful for the purpose of credit scoring, target marketing, customer retention, customer profiling, marketing campaigns, fraud detection, churn modeling, customer segmentation, product-bundling, cross-selling and up-selling of products. Here we discuss some limitations to constructing credit scoring models. Two main limitations are the availability of data and sample selection issues. All too often a good credit scoring model cannot be obtained due to unavailability or poor quality (recording errors and high percentage of missing values) of available data. Moreover, credit scoring models built using historical data of past applicants who were accepted could lead to a biased sample when used to evaluate new applicants. To remedy this bias, SAS Enterprise Miner provide a Reject Inference node whereby the rejected applicants are scored (predicted as good or bad) using the model built based on accepted applicants. These scored data are then added to the accepted sample and the augmented sample serves as an input to a second modeling run (SAS Institute Inc., 2009). Next is the issue is which model is the best? According to results from past studies there is no overall 'best' model. The performance of credit scoring models depends on the data structure, data quality and the objective of the classification. Sophisticated techniques such as ANNs, MARS and SVM have shown only slight improvements in classification accuracy. In practical applications, classification methods which are easy to understand such as scorecards and decision trees are more appealing to users.

[3] For interpretation of color in Fig. 5, the reader is referred to the web version of this article.

References

Abdou, H., Pointon, J., & El-Masry, A. (2008). Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert System with Applications, 35*, 1275–1292.

Ang, J. S., Chua, J. H., & Bowling, C. H. (1979). The profiles of late paying consumer loan borrowers: An exploratory study. *Journal of Money, Credit and Banking, 11*(2).

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques: For marketing, sales, and customer support*. New York: John Wiley and Sons, Inc.

Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with Applications, 34*, 280–290.

Chien, C. F., Hsiao, A., & Wang, I. (2004). Constructing semiconductor manufacturing performance indexes and applying data mining for manufacturing data analysis. *Journal of the Chinese Institute of Industrial Engineers, 21*(4), 313–327.

Chien, C. F., Wang, I., & Chen, L. F. (2005). Using data mining to improve the quality of human resource management of operators in semiconductor manufactures. *Journal of Quality, 12*(1), 9–28.

Cho, V., & Ngai, E. (2003). Data mining for selection of insurance sales agents. *Expert Systems, 20*(3), 123–132.

Davis, R. E., Elder, K., Howlett, D., & Bouzaglou, E. (1999). Relating storm and weather factors to dry slab avalanche activity at Alta, Utah, and Mammoth Mountain, California, using classification and regression trees. *Cold Regions Science and Technology, 30*, 79–89.

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.

Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis, 16*, 471–495.

Gao, L., Zhou, C., Gao, H.-B., & Shi, R.-Y. (2006). Credit scoring model based on neural network with particle swarm optimization. In: *ICNC, Part I. LNCS* (Vol. 4221, pp. 76–79).

Gschwind, M. (2007). Predicting late payments: A study in tenant behavior using data mining techniques. *Journal of Real Estate Portfolio Management, 13*(3), 269–288.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco, CA: Morgan Kaufman.

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.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support System, 37*, 543–558.

Huang, C.-L., Chen, M.-C., & Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert System with Applications, 37*, 847–856.

Koh, H. C., Tan, W. C., & Goh, C. P. (2004). Credit scoring using data mining techniques. *Singapore Management Review, 26*(2), 25–47.

Kurt, I., Ture, M., & Kurum, A. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert System with Applications, 34*, 366–374.

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*, 1113–1130.

Mavri, M., Angelis, V., & Loannou, G. (2008). A two-stage dynamic credit scoring model based on customers profiles and time horizon. *Journal of Financial Services Marketing, 13*(1), 17–27.

Olson, D., & Yong, S. (2006). *Introduction to business data mining*. McGraw Hill International Edition.

Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in Society, 24*, 483–502.

SAS Institute Inc, (2009). *Building credit scorecards using credit scoring for SAS® Enterprise Miner™: A SAS best practices paper*. A SAS white paper. Cary, NC: SAS Institute, Inc.

Siddiqi, N. (2005). *Credit risk scorecards: Developing and implementing intelligent credit scoring*. Cary, NC: SAS Press.

Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Pearson Education Inc.

Thomas, L. C. (2000). A survey of credit and behavioral scoring: Forecasting financial risk of lending to customers. *International Journal of Forecasting, 16*, 149–172.

Vojtek, M., & Kocenda, E. (2006). Credit scoring methods. *Czech Journal of Economics and Finance, 56*(3–7), 152–167.

West, D. (2000). Neural network scoring models. *Computer and Operations Research, 27*, 1131–1152.