A Two-step Method to Construct Credit Scoring Models with Data Mining Techniques

Hian Chye Koh
Associate Professor and Dean, School of Business
SIM University (Singapore)
hckoh@unisim.edu.sg

Wei Chin Tan
Statistical Reporting Manager
Singapore Telecommunications Ltd

Chwee Peng Goh
Assistant Vice President (Marketing)
Great Eastern Life Assurance Co Ltd

Abstract

Credit scoring can be defined as a technique that helps credit providers decide whether to grant credit to consumers or customers. Its increasing importance can be seen from the growing popularity and application of credit scoring in consumer credit. There are advantages not only to construct effective credit scoring models to help improve the bottom-line of credit providers, but also to combine models to yield a better performing combined model. This paper has two objectives. First, it illustrates the use of data mining techniques to construct credit scoring models. Second, it illustrates the combination of credit scoring models to give a superior final model. The paper also highlights the prerequisites and limitations of the data mining approach.

Keywords: Credit scoring; data mining techniques; construction of models; combination of model; meta-modeling.

International Journal of Business and Information
 A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

1. Introduction

The last two decades have seen a rapid growth in both the availability and the use of consumer credit. Until recently, the decision to grant credit was based on human judgment to assess the risk of default [Thomas, 2000]. The growth in the demand for credit, however, has led to a rise in the use of more formal and objective methods (generally known as credit scoring) to help credit providers decide whether to grant credit to an applicant [Akhavein, 2005]. This approach, first introduced in the 1940s, has evolved over the years and developed significantly [Rimmer, 2005]. In recent years, the progress in credit scoring was fueled by increased competition in the financial industry, advances in computer technology, and the exponential growth of large databases.

The 2003 Credit Research Foundation survey reported that 36% of the respondents were actively using credit scoring and that 72% of the respondents who were not using credit scoring intended to do so within the next two years [Cundiff, 2004]. Wendel and Harvey [2003] also reported that virtually all credit card loans and about 70% of home loans utilize credit scoring for the loan decisions. Further, in 2002, Alan Greenspan, then chairman of the U.S. Federal Reserve, commented that credit scoring had helped banks stave off the effects of the most recent economic downturn, provided a greater depth of risk management, and served as the foundation for the development of consumer and mortgage credit [PBI Media, 2002].

On another front, since the mid-1990s, three new

interrelated areas that emphasized obtaining more information from data have emerged strongly in information systems and information technology. They are data warehousing, knowledge management, and data mining, the last of which aims to identify valid, novel, potentially useful and understandable correlations and patterns in data [Chung and Gray, 1999]. Coupled with advances in both computer hardware and software, many data mining applications are now more accessible and affordable to businesses than before.

This paper discusses and illustrates the use of data mining techniques in the construction and combination of credit scoring models. The remainder of the paper is organized as follows. The first section provides a formal definition of credit scoring, describes its usefulness and advantages, and lists some of its applications. It also discusses the construction of credit scoring models by looking at the methodology and techniques that are commonly used. The second section identifies data mining techniques that can be used to construct credit scoring models and the methods of combining the different models. The third section illustrates the use of data mining techniques to construct credit scoring models and the various methods of combining them. Finally, the concluding section highlights the limitations of credit scoring as well as the prerequisites and limitations of the data mining approach to the construction and combination of credit scoring models.

2. Credit Scoring

Credit scoring can be formally defined as a statistical (or quantitative) method that is used to predict the probability that

100

International Journal of Business and Information
 A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

a loan applicant or existing borrower will default or become delinquent [Mester, 1997]. This helps to determine whether credit should be granted to a borrower [Morrison, 2004]. Credit scoring can also be defined as a systematic method for evaluating credit risk that provides a consistent analysis of the factors that have been determined to cause or affect the level of risk [Fensterstock, 2005]. The objective of credit scoring is to help credit providers quantify and manage the financial risk involved in providing credit so that they can make better lending decisions quickly and more objectively. In the United States, the Circuit Court has found considerable actuarial evidence that credit scores are a good predictor of risk of loss [Johnson-Speck, 2005]. Similarly, a recent actuarial study has concluded that credit scores are one of the most powerful predictors of risk; they are also the most accurate predictor of loss seen in a long time [Miller, 2003].

2.1 Benefits of Credit Scoring

Credit scoring has many benefits that accrue not only to the lenders but also to the borrowers. For example, credit scores help to reduce discrimination because credit scoring models provide an objective analysis of a consumer's creditworthiness. This enables credit providers to focus on only information that relates to credit risk and avoid the personal subjectivity of a credit analyst or an underwriter [Fensterstock, 2005]. In the United States, under the Equal Credit Opportunity Act, variables of overt discrimination such as race, sex, religion, and age cannot be included in the credit scoring models. Instead, only information that is non-discriminatory in nature and that has

been proven to be predictive of payment performance can be included in the models.

Credit scoring also helps to increase the speed and consistency of the loan application process and allows the automation of the lending process [Rimmer, 2005]. As such, it greatly reduces the need for human intervention on credit evaluation and the cost of delivering credit [Wendel and Harvey, 2003; Diana, 2005]. With the help of the credit scores, financial institutions are able to quantify the risks associated with granting credit to a particular applicant in a shorter time. Leonard's [1995] study of a Canadian bank found that the time for processing a consumer loan application was shortened from nine days to three days after credit scoring was used. The time saved in processing the loans can be used to address more complex issues. Banaslak and Kiely [2000] concluded that with the help of credit scores, financial institutions are able to make faster, better, and higher quality decisions. With this, credit scoring can also improve the allocation of resources toward the "first best equilibrium" [Jacobson and Roszback, 2003].

Further, credit scores can help financial institutions determine the interest rate that they should charge their consumers and to price portfolios [Avery et al., 2000]. Higher-risk consumers are charged a higher interest rate and vice versa. Based on the consumer's credit scores, the financial institutions are also able to determine the credit limits to be set for the consumers [Sandler et al., 2000; Park, 2004]. These help financial institutions to manage their accounts more effectively and profitably. As an extension, profit scoring can be used to maximize profits across a range of products [Thomas, 2000; Park, 2004].

International Journal of Business and Information
 A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

Related to the above, credit scoring models have enabled the development of the sub-prime lending industry where sub-prime consumers have poor credit records and fall short of credit acceptance and risk. They may not meet the requirements for traditional financing because of credit impairment, missing data in their credit histories, or difficulty in validating their income [Quittner, 2003]. One of the major factors in the progress of sub-prime lending has been automated underwriting, which allows sub-prime mortgage loans to be packaged and sold as investment securities. The initial success of specialized financial institutions in this market has driven more financial institutions to enter the sub-prime lending market, which is expected to grow as technology in credit scoring advances [Perin, 1998]. This success has also been observed in small business loans [Stanton, 2001; Akhavein et al., 2005].

Because of advances in technology, more intelligent credit scoring models are being developed. Consequently, credit card issuers are able to make use of the information generated from the models to formulate better collection strategies and hence use their resources more effectively [Cundiff, 2004]. Lucas [2000b], for example, had found recovery rates to average 15.9% in 1999, up from 12.1% in the previous year and 9.1% in 1997.

Finally, the insurance industry has used credit scoring to streamline the insurance application and renewal process. In particular, credit scores can help insurance companies to make a better prediction on claims and hence to control risk more effectively. They also make pricing more accurate. This enables

insurance companies to offer more insurance coverage to more consumers at a more equitable cost, react quickly to market changes and gain a competitive edge [Kellison and Brockett, 2003].

2.2 Credit Scoring Applications

In the early years, financial institutions used credit scoring mainly to make credit decisions for loan applications. Over the past 25 years, however, the application of credit scoring has grown from making credit decisions to making decisions related to housing, insurance, basic utility services, and even employment. However, not all these applications are equally widely used.

The most common use of credit scores is in making credit decisions for loan applications [Rimmer, 2005]. In addition to decisions on personal loan applications, financial institutions now make use of credit scores to help set credit limits, manage existing accounts, and forecast the profitability of consumers and customers [Lucas, 2000a]. For example, the Australia and New Zealand Banking Group uses credit scoring to help identify applicants who should receive credit, determine the amount of credit that the applicants should receive, and the steps that should be taken should there be a failure in the payment of loans (see http://www.sas.com/ success/anzcredit.html). Also, credit card issuers use credit scores as a decision support tool to identify their target market for credit cards [Lucas, 2000a]. In recent years, credit scores have also been used as part of the decision process for providing credit to small businesses [Rowland, 2003; Akhavein et al., 2005]. Previously, the Fleet Financial Group has used credit scoring for loans under US

104

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

$100,000 [Zuckerman, 1996].

Credit scoring models have also been used in the insurance industry (e.g., for mortgage and automobile insurance) to decide on the applications of new insurance policies and the renewals of existing polices. The premise is that there is a direct relationship between financial stability and risk. It has been argued that there is a strong relationship between credit rating and loss ratios in both automobile and mortgage insurance. Statistical evidence has shown that relative loss ratios (which are a function of both claim frequency and cost) decrease as credit rating improves [Schiff, 2003]. GE Capital Mortgage Corporation uses credit scoring to help screen mortgage insurance applications [Prakash, 1995]. Credit scores are also used as a basis to adjust premiums. Generally, consumers with bad credit scores have a higher chance of filing insurance claims, compared with customers with good credit scores. Therefore, the former are charged a higher premium. Further, credit information is used to assess a consumer's accountability and performance under the conditions of an insurance policy.

In addition to the above, other credit scoring applications have also been reported [Consumer Federation of America, 2002]. For example, landlords can make use of credit scores to determine whether potential tenants are likely to pay their rent on time. There is substantial use of credit scoring in the mortgage industry too [Wagner, 2005]. Also, some utility suppliers in the United States have used credit scores to determine whether to provide their services to consumers. Finally, some employers make use of credit history and credit scores to decide whether to hire a potential employee, especially

for posts where employees need to handle huge sums of money [Consumer Federation of America, 2002]. The implication is that employee trustworthiness and hence personal character can be assessed through their credit scores.

3. Construction of Credit Scoring Models

The methodology for constructing credit scoring models generally involves the following process. First, a sample of previous customers is selected and classified as "good" or "bad" depending on their repayment performance over a given period (for simplicity, only a dichotomy is used here). Next, data are compiled from loan applications, personal and/or business credit records, and various sources if available (e.g., credit bureau reports). Finally, statistical (or other quantitative) analysis is performed on the data to derive a credit scoring model. (A more elaborate methodology will be presented in the concluding section.) The model comprises weights to apply to the different variables (or attributes) in the data and a cut-off point. The sum of the weights applied to the variables for an individual consumer or customer constitutes the credit score. The cut-off point determines whether this consumer or customer should be classified as "good" or "bad." The probability associated with this classification can also be generated. Different models can be constructed for different segments of the data (e.g., for different products).

To date, several techniques have been used in the construction of credit scoring models. The most common techniques used are traditional statistical methods. For example, some of the earliest credit scoring models used

106

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

discriminant analysis. However, discriminant analysis requires rather restrictive statistical assumptions that are seldom satisfied in real life. Consequently, logistic regression (which is less restrictive) has been proposed as an alternative to discriminant analysis. Some of the techniques that have been previously used, but rather infrequently, to construct credit scoring models include genetic algorithm, k-nearest neighbor, linear programming, and expert systems [Thomas, 2000].

In recent years, new techniques have been increasingly used to construct credit scoring models. In particular, the decision tree approach has become a popular technique for developing credit scoring models because the resulting decision trees are easily interpretable and visualized. Further, neural networks are also commonly used. All the methods and techniques mentioned above can be considered as important data mining techniques for predictive modeling. Empirical studies on credit scoring models using data mining techniques include Lee and Jung [1999/2000] and West [2000].

4. Combination of Credit Scoring Models

As discussed above, several techniques can be used in the construction of credit scoring models. In such situations, the results of the different resulting models will be compared and the champion (or best) model chosen. The rest of the models will be discarded. However, each model may have its particular strengths and weaknesses. This may affect the characteristics of an individual model's performance. For example, two models may have comparable results in terms of accuracy rate and/or

other criteria, but one model may perform better in a particular segment of the data and the other model may perform better in another segment. Hence, it may be advantageous to combine the two models in order to make better use of the potential of the models.

Specifically, the purpose of combining credit scoring models is to produce a combined model that is better than the individual models in terms of accuracy rate and/or other criteria [Lee and Jung, 1999/2000]. In their study, Zhu et al. [2001] combined two sets of consumer credit scores. The results showed that the combined score outperformed each set of scores upon which the combined model was based.

Credit scoring can be combined in the several different ways. This paper illustrates four methods for combining credit scoring models.

The first method is to build a model based on the predictions or results of two or more models and the rest of the attributes. This is commonly known as meta-modeling. For the second method, the final prediction is a bad credit when any individual model predicts a bad credit – this is a conservative form of voting. In contrast, the third method is a liberal form of voting, where all individual models must predict a bad credit for the final prediction to be a bad credit. The second and third combination methods suggest that there can be variation in the degree of conservatism in voting. Hence, a fourth method is added where the final prediction is a bad credit when any two individual models predict a bad credit. The combined model will be considered superior to the individual models that it is based on only if the combined model performs better than the individual models in terms of accuracy rate and/or other criteria

in the training and scoring data sets. The four methods mentioned above are illustrated below.

5. An Illustration

The data set for the illustration is taken from the UCI Machine Learning Repository [Blake and Merz, 1998]. The data relate to a credit screening application in a German bank. There are 20 attributes (7 numerical and 13 categorical) and a binary outcome. Among the 1,000 observations, 700 (or 70.0%) are good credit risk and 300 (or 30.0%) are bad credit risk. The 20 attributes available for constructing credit scoring models include demographic characteristics (e.g., gender and age) and credit details (e.g., credit history and credit amount).

Suppose that the German bank is interested in developing a credit scoring model to predict the credit risk of loan applicants as bad or good risk. The bank intends to deploy the model at the time the loan applications are processed. Construction of the credit scoring model requires predictive modeling to be done. For this purpose, three data mining techniques are appropriate; namely, logistic regression, neural network, and decision tree. SPSS Clementine (a data mining software) is used in this illustration.

5.1 Modeling Approach

Before performing predictive modeling, 37 foreign workers from the sample of loan applicants are removed from subsequent modeling as it is noted that the profile of these foreign workers seems different from their local counterparts

and hence may confound the results. In addition, the sample data is partitioned into a construction/training sample (approximately 80.00%) and a validation/test sample (approximately 20.00%). In this illustration, the performance indicators of the respective prediction models are as follows: (1) overall accuracy rate; (2) stability of model; (3) proportion of responses (bad credit) captured; (4) accuracy rate for good credit (i.e., Type I accuracy rate); and (5) accuracy rate for bad credit (i.e., Type II accuracy rate). That is, the above criteria are used to assess each model and to compare across models.

The modeling approach can be summarized as follows:

(1) Variants within each modeling algorithm are tried and applied on the training sample by changing the parameters and model settings. The modeling algorithms used are the Logistic, Neural Net, C5.0 and C&R Tree nodes in Clementine.

(2) Each model is applied on the validation sample. The champion model within each modeling algorithm is identified based on the above performance indicators.

(3) The best three performing champion models are then combined using the four combination methods described in the previous section. For the first method (which uses meta-modeling to combine a model by taking into account the outputs from the various champion models), the approach is similar to that described in (1) and (2) above. In other words, the four modeling algorithms are applied on the training sample and the resulting models are scored on the validation sample. Each of the models constructed is then evaluated based on the

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

performance indicators as listed earlier.

The best three performing champion models are derived using logistic regression, neural network and decision tree. The data mining diagram is shown in Figure 1, which can be interpreted as follows. The left side of the data mining diagram (comprising two columns of nodes or icons) shows the reading-in of the data and the partitioning of data into the training sample and validation sample. For example, the "select" and "filter" nodes specifies the appropriate observations and variables to read-in and the "randnum" node partitions the data into the training and validation samples.

The middle top part of Figure 1 shows the construction of the individual and combined models, and the middle bottom part shows the validation of these models. In particular, the pentagon-shaped nodes at the middle top part show the construction of credit scoring models using decision trees (C5.0 and C&R Tree), neural network, and logistic regression. For the middle top and middle bottom parts, the diamond-shaped nodes show that the model outputs (i.e., credit scores) of the respective models are incorporated into the data for further analysis.

The right side of the data mining diagram shows the generation of results. For example, the rectangular and "analysis" nodes indicate the computation of accuracy rates. Finally, the combination of models is shown by nodes such as "Vote1" and "Vote3".*(More information on the nodes can be found in user guides issued by SPSS).*

To summarize, the appropriate sample data are first read-in and partitioned into training and validation samples

(approximately in the proportion of 80% to 20%, respectively). Next, credit scoring models are constructed on the training sample data (using logistic regression, neural networks and decision trees) and validated on the validation data. Based on the validation results (e.g., accuracy rates), the best individual models are selected for combination. Then, four combination methods are used to combine the individual models based on the training sample data. Finally, the resulting four combined models are validated (i.e., performance indicators computed) using the validation sample data.
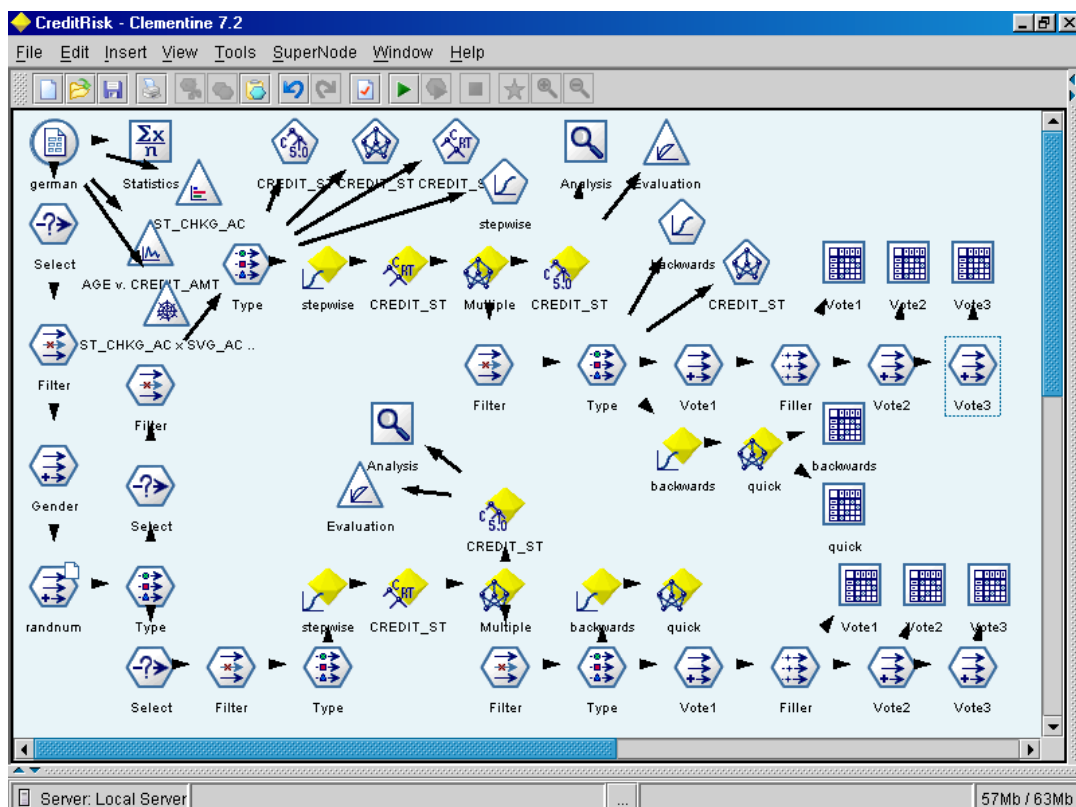

Figure 1. Data Mining Diagram

(More discussion can be found in Section 5.4 where a more

general algorithm is presented).

5.2 Predictive Modeling Results

The logistic regression results indicate that the model is statistically significant (at a 0.05 significance level). In addition, the following input variables are statistically significant in predicting credit risk: status of checking account, duration, credit history, purpose, other debtors/guarantors, installment rate in percentage of disposable income, savings account/bonds, other installment plan, and years of employment.   For the logistic regression model, the overall accuracy rate is 77.0%, proportion of responses captured 54.5%, Type I accuracy rate 66.8% and Type II accuracy rate 80.6%.

The neural network model has one input layer, two hidden layers, and one output layer. The top five most important input variables in descending order of importance are: status of checking account, credit history, savings account/bonds, purpose, and type of property. The overall accuracy rate of the neural network model is 73.4%, proportion of responses captured 38.9%, Type I accuracy rate 62.9% and Type II accuracy rate 75.9%.

Finally, the decision tree model has eight terminal nodes (predicting good and bad credit) and the following six important input variables: status of checking account, savings account/bonds, duration, other debtors/guarantors, other installment plan, and purpose. The overall accuracy for the decision tree model is 71.4%, proportion of responses captured 79.1%, Type I accuracy rate 53.3% and Type II accuracy rate 87.5%.

From the results presented above, the logistic regression model is the most accurate (based on overall accuracy rate). However, as the performance of the three models on the construction/training sample is upward-biased (since the same observations are used for model construction and model evaluation), it is important to assess the performance of the models on the validation/test sample. These results are summarized in Table 1, Panel A.

As shown, logistic regression is the most stable model. Its performance indicators fluctuate the least compared with the other two models. The decision tree model, on the other hand, performs the best in terms of the proportion of responses captured and Type II accuracy. Neural network has the best overall accuracy rate and Type I accuracy rate. Since each model has its own strengths, it is useful to explore the combination of the three models.

Table 1. Comparison of Models
A. Performance of Best Three Models

	Logistic Regression	Neural Network	Decision Tree (CART)
Overall Accuracy Rate (%)	75.6	76.7	68.9
Proportion of Responses Captured (%)	53.8	42.3	73.1
Type I Accuracy Rate (%)	54.9	59.5	45.2
Type II Accuracy Rate (%)	83.1	80.8	87.2

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

<u>B.</u> Performance of the Four Combined Methods

	Logistic Regression	VOTE1	VOTE 2	VOTE3
Overall Accuracy Rate (%)	78.8	68.4	75.6	77.2
Proportion of Responses Captured (%)	55.8	76.9	57.7	34.6
Type I Accuracy Rate (%)	61.7	44.9	54.5	64.3
Type II Accuracy Rate (%)	84.2	88.5	84.1	79.4

<u>C.</u> Comparative Results on German Credit Database

	Kaburlasos and Kazarlis (2002)	Cho and With-rich (2002)	Merkevicius and Garsva (2004)
Overall Accuracy Rate (%)	71.6	61.6	83.7
Type I Accuracy Rate (%)	–	–	95.3
Type II Accuracy Rate (%)	–	–	18.1

5.3 Combination of Models

The four methods mentioned in the previous section are used to combine the best model in each family (i.e., logistic regression, neural network, and decision tree). Using the first combination method, the variables that are used for building the combined model include the variables used to build the individual models and the predictions that are generated from

the three best models. The combined model is built using the training sample and evaluated using the validation sample. The same performance indicators are used to evaluate the combined model; namely, overall accuracy rate, stability of model, proportion of responses (bad credit) captured, accuracy rate for good credit (Type I accuracy rate), and accuracy rate for bad credit (Type II accuracy rate). These criteria are also used to select the best combined model. The results indicate that the best combined model using the first combination method is logistic regression with the backward selection method.

For the second combination method, VOTE1, when any individual model included in the combined model predicts an observation to be a bad risk, VOTE1 will indicate that the particular record is a bad risk. This is a conservative form of voting. For the third combination method, VOTE3, when all the three individual models included in the combined model predict an observation to be a bad risk, VOTE3 will indicate that the particular record is a bad risk. This is a most liberal form of voting. Finally, for the fourth combination, VOTE2, when any two individual models predict an observation to be a bad risk, VOTE2 will indicate that the particular record is a bad risk. This is a compromise between VOTE1 and VOTE3.

The combination results are summarized in Table 1, Panel B. Comparing the combined models with the individual models they are based on (see Panel A), the logistic regression combined model outperforms all the individual models in terms of the overall accuracy rate and Type I accuracy rate. It is also the most stable in terms of the performance indicators. VOTE1 outperforms the individual models in terms of proportion of responses captured and Type II accuracy rate. VOTE2 did not

outperform any individual models in terms of the criteria. VOTE3 outperforms all the individual models in terms of the overall accuracy rate and Type I accuracy rate.

After the combined models are being built and the results tabulated, it is important to ascertain which criterion is the most important in order to choose a best combined model. If the most important criterion is the overall accuracy rate, then the logistic regression combined model should be chosen as the model to be deployed. However, if the most important criterion is the proportion of responses captured or Type II accuracy rate, then VOTE1 should be chosen as the model to be deployed. Finally, if the most important criterion is the Type I accuracy rate, then VOTE3 should be chosen as the model to be deployed. Thus, the final combined model to be deployed depends on the criterion that the user deems the most important.

To further assess the results, it is noted that Zhu et al. [2001] reported an improvement of between 0.5% and 3% in accuracy rates for their logistic regression combined model. For the current study, the improvement in overall accuracy rate for the logistic regression combined model vis-à-vis the individual models is 2.1% (i.e., 78.8% – 76.7%). (See Table 1.) Hence, the current results are consistent with those of Zhu et al. [2001].

Finally, there have been a few prior studies in which the German credit database was used, as in this study. The comparable results are as follows. Kaburlasos and Kazarlis [2002] reported an overall accuracy rate of 71.6% for a decision tree credit scoring model, and Cho and Withrich [2002] reported an overall accuracy rate of 61.6% for a model based on multiple probabilistic rules. Also, Merkevicius and Garsva [2004] reported

an overall accuracy rate of 83.7% on the 20% validation sample for a neural network (clustering) model. However, this model has very unbalanced accuracy rates that make application rather doubtful. In particular, whereas the Type I accuracy rate (for good credit) is very high at 95.3%, the Type II accuracy (for bad credit) is only 18.1%. Hence, the results of the current study compare favorably with these prior results.

The above comparative results on the German credit database are summarized in Table 1, Panel C. (Saar-Tsechansky and Provost [2004] also used the German credit database in their study. However, they focused on sampling issues and no accuracy rates were reported.)

5.4 An Illustrative Algorithm

Finally, to capture the process of the construction and combination of credit scoring models, an illustrative algorithm is presented in Figure 2. As shown, the process starts with the construction of individual models. This stage comprises the following five steps: (1) define objective, (2) select variables, (3) select sample and collect data, (4) select modeling tools and construct models, and (5) validate and assess models.

With reference to the earlier illustration, its objective is to develop a credit scoring model to predict the credit risk of loan applicants as bad or good risk. To construct the model, 20 attributes comprising demographic characteristics and credit details are used. The sample and data are contained in the German data set taken from the UCI Machine Learning Repository [Blake and Merz, 1998]. Logistic regression, neural network, and decision tree are used to construct the credit scoring model. The resulting models are validated and

118

International Journal of Business and Information
 A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

assessed, with the main findings summarized in Panel A of Table 1. At this stage, if no combination of models is intended, the best individual model will be selected for deployment.

If the individual models are to be combined, the second stage is effected (see middle part of Figure 2). Five steps are used to combine the models: (1) select models for combination, (2) select modeling methods to combine models, (3) combine models, (4) validate and assess combined models, and (5) select best combined model for deployment.

Construction of Individual Models:

Combination of Models:

Post-Implementation:

```
                          ┌─────────────────────┐
                          │  Define Objective   │
                          └─────────────────────┘
                                    │
                          ┌─────────────────────┐
                          │   Select Variables  │
                          └─────────────────────┘
                                    │
                     ┌──────────────────────────────┐
                     │ Select Sample and Collect Data│
                     └──────────────────────────────┘
                                    │
            ┌────────────────────────────────────────────┐
            │ Select Modeling Tools and Construct Models  │
            └────────────────────────────────────────────┘
                                    │
                  ┌───────────────────────────────┐
                  │   Validate and Assess Models  │
                  └───────────────────────────────┘
                        │                   │
                        │        ┌─────────────────────────────────────┐
                        │        │ Select Best Model for Deployment if No │
                        │        │     Combination is Intended          │
                        │        └─────────────────────────────────────┘
                        │                   │
            ┌───────────────────────────────┐
            │  Select Models for Combination │
            └───────────────────────────────┘
                        │
            ┌───────────────────────────────────┐
            │ Select Modeling Methods to Combine Models │
            └───────────────────────────────────┘
                        │
                  ┌──────────────────┐
                  │  Combine Models  │
                  └──────────────────┘
                        │
            ┌────────────────────────────────┐
            │ Validate and Assess Combined Models │
            └────────────────────────────────┘
                        │
            ┌────────────────────────────────────┐
            │ Select Best Combined Model for Deployment │
            └────────────────────────────────────┘
                        │                   │
            ┌──────────────────────────────────────┐
            │   Monitor Performance of Deployed Model │
            └──────────────────────────────────────┘
                        │
            ┌──────────────────────────────────────┐
            │  Refresh/Reconstruct Model if Necessary │
            └──────────────────────────────────────┘
```
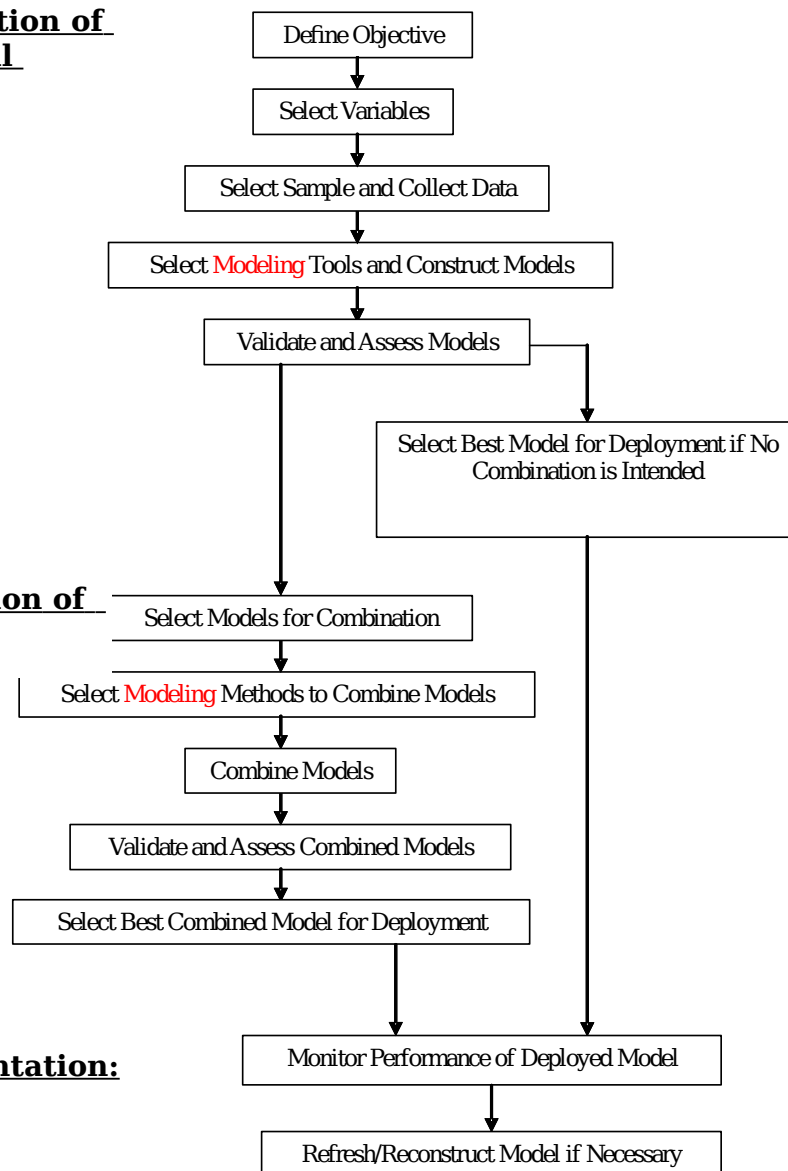
Figure 2. Illustrative Algorithm for the Construction and Combination of Credit Scoring Models

Again, with reference to the earlier illustration, the best

120

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining Techniques

logistic regression, decision tree, and neural network models are selected for combination. Four combination methods are used to combine the models; namely, logistic regression, VOTE1, VOTE2 and VOTE3. After combining the methods, the combined models are validated and assessed. The results are summarized in Panel B of Table 1. Finally, the best combined model is selected for deployment.

It is noted from the bottom part of Figure 2 that, after the deployment of a model (whether it is an individual or combined model), it is necessary to monitor the performance of the deployed model. If necessary (for example, when the performance becomes unacceptable), the credit scoring model needs to be refreshed (e.g., by incorporating recent data into the modeling process) or reconstructed (i.e., to build up a final model from the beginning of the model construction process).

6. Conclusion

This paper discusses and illustrates the use of data mining techniques in the construction and combination of credit scoring models. In recent years, data mining has gained widespread attention and increasing popularity in the commercial world. Besides credit scoring, there are other potential data mining applications for businesses. For example, data mining can be used to: (1) perform churn modeling to identify customers who are likely to churn, (2) construct fraud detection models to give early warning signals of possible fraudulent transactions, (3) understand consumers and customers better (e.g., via market basket analysis), (4) segment customers (e.g., via clustering), or

(5) construct models to predict the probability of purchasing certain products or services in order to facilitate cross-selling or up-selling. The findings can then be used, say, to prepare mail catalogs, advertisements, promotion campaigns, etc.

6.1 Limitations of Credit Scoring

Although credit scoring has significant benefits, its limitations should also be noted. One of the major problems that can arise when constructing a credit scoring model is that the model may be built using a biased sample of consumers and customers who have been granted credit [Hand, 2001]. This may occur because applicants (i.e., potential customers) who are rejected will not be included in the data for constructing the model. Hence, the sample will be biased (i.e., different from the general population) as good customers are too heavily represented. The credit scoring model built using this sample may not perform well on the entire population since the data used to build the model is different from the data that the model will be applied to.

The second problem that can arise when constructing credit scoring models is the change of patterns over time. The key assumption for any predictive modeling is that the past can predict the future [Berry and Linoff, 2000]. In credit scoring, this means that the characteristics of past applicants who are subsequently classified as "good" or "bad" creditors can be used to predict the credit status of new applicants. Sometimes, the tendency for the distribution of the characteristics to change over time is so fast that it requires constant refreshing of the credit scoring model to stay relevant.

One of the consequences of credit scoring is the possibility

that end-users become so reliant on the technology that they reduce the need for prudent judgment and the need to exercise their knowledge on special cases. In other instances, end-users may unintentionally apply more resources than necessary to work the entire portfolio. This could run into the risk of a self-fulfilling prophecy [Lucas, 2002b]. In the U.S., a new industry has emerged that is dedicated to help borrowers improve their credit scores by rearranging finances [Timmons, 2002], rather than obeying the simple rule: pay your bills on time and keep your debt low. Such score-polishing actions could potentially distort the patterns of credit default.

Despite the limitations highlighted above, there is no doubt that credit scoring will continue to be a major tool in predicting credit risk in consumer lending. It is envisaged that organizations using credit scoring appropriately will gain important strategic advantage and competitive edge over their rivals.

6.2 Limitations of Combining Credit Scoring Models

Although combining individual credit scoring models may lead to combined models that perform better than the individual models they are based on in terms of accuracy rate and/or other criteria, the process does has its pitfalls. One of the limitations is that there may be difficulties in interpreting the rules generated by the combined model. Furthermore, the combined model may not make sense since it is built on individual models that may generate different rules.

Another limitation is that more time may be needed to construct the final model. Also, time taken to score the

combined model online may also be longer as more models are being run. The extra length of time taken may not be justifiable if the improvement of the combined model in terms of accuracy rate and/or other criteria is minimal. The third limitation is that it may be difficult, if not infeasible, to build a combined model such that it significantly outperforms the individual models that it is based on. Time and effort spent on trying to build a superior combined model may be futile. It is always important to understand the limitations of building a credit scoring model and a combined model so that the right expectations can be set.

Despite the above, it is noted that the combination of credit scoring models can lead to improved credit scoring results. More generally, from a management perspective, it is worthwhile noting that credit scoring has the following benefits: (1) reduced subjectivity and increased objectivity in risk assessment, (2) increased speed and consistency of risk assessment, (3) possible automation of the risk assessment process, (4) reduced cost of risk assessment, (5) better allocation of resources, (6) better determination of interest rate, loan duration, loan amount, etc., (7) better risk management, (8) better collection/recovery strategies, and (9) possible development of non-traditional markets. It is hoped that this paper can make a contribution to the credit scoring literature.

Acknowledgments

124

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

References

Akhavein, J., W.S. Frame, and L.J. White. 2005. The diffusion of financial innovations: An examination of the adoption of small business credit scoring by large banking organisations, *The Journal of Business* 78(2) 577-596.

Avery, R.B., R.W. Rostic, P.S. Calem, and G.B. Canner. 2000. Credit scoring: Statistical issues and evidence from credit-bureau files, *Real Estate Economics* 28(3) 523-547.

Berry, M.J.A., and G.S. Linoff. 2000. *Mastering Data Mining: The Art and Science of Customer Relationship Management*, New York: John Wiley & Sons, Inc.

Blake, C.L., and C.J. Merz. 1998. UCI repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html], University of California, Department of Information and Computer Science.

Chung, H.M., and P. Gray. 1999. Data mining, *Journal of Management Information Systems* 16(1) 11-13.

Cho, V., and B. Withrich. 2002. Distributed mining of classification rules, *Knowledge and Information Systems* 4(1) 1-30.

Consumer Federation of America. 2002. *Credit Scores Accuracy and Implications for Consumers*. New York: Consumer Federation of America and National Credit Reporting Association.

Cundiff, K. 2004. Closing the loop: How credit scoring drives performance improvements along the financial value chain, *Business Credit* 106(3) 38-42.

Diana, T. 2005. Credit risk analysis and credit scoring – now

and in the future, *Business Credit* 107(3) 12-16.

Fensterstock, A. 2005. Credit scoring and the next step, *Business Credit* 107(3) 46-49.

Hand, D.J. 2001. Modelling consumer credit risk, *IMA Journal of Management Mathematics* 12(1) 139-155.

Jacobson, T., and K. Roszback. 2003. Bank lending policy, credit scoring and value-at-risk, *Journal of Banking and Finance* 27(4) 615-633.

Johnson-Speck, C. 2005. Abstracts of significant cases bearing on the regulation of insurance, *Journal of Insurance Regulation* 23(4) 81-84.

Kaburlasos, V.G., and S. Kazarlis. 2002. σ-FLNMAP with voting (σ-FLNMAPwV): A genetically optimised ensemble of classifiers with the capacity to deal with partially-ordered, disparate types of data, 4th International Conference on Technology & Automation, Greece, 5-6 October.

Kellison, B., and P. Brockett. 2003. Check the score: credit scoring and insurance losses: Is there a connection? *Texas Business Review* Special Issue, 1-6.

Lee, T.H., and S.H. Jung. 1999/2000. Forecasting creditworthiness: logistic vs artificial neural net, *The Journal of Business Forecasting Methods & Systems* 18(4) 28-30.

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

Lucas, P. 2000a. Shedding light on credit scores, *Credit Card Management* (August) 78-80.

Lucas, P. 2000b. Why recoveries are on the rise, *Credit Card Management* (October) 71-72.

Lucas, P. 2002. Score updates, *Collections & Credit Risk* 7(10)

126

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

22-25.

Merkeviscius, E., and G. Garsva. 2004. Forecasting of credit classes with the self-organizing maps, *Informacines Technologoijos Jr Valdymas* 4(33) 61-66.

Mester, L.J. 1997. What's the point of credit scoring? *Business Review* (September) 3-16.

Miller, M. 2003. Research confirms value of credit scoring, *National Underwriter* 107(42) 30.

Morrison, J. 2004. Introduction to survival analysis in business, *The Journal of Business Forecasting Methods & Systems* 23(1) 18-22.

Park, S. 2004. Solving the mystery of credit scoring models, *Business Credit* 106(3) 43-47.

PBI Media. 2002. Banks must remain focussed on risk management, Greenspan says, *EFT Report* 25(23) 6-7.

Perin, M. 1998. Risky business: sub-prime market growth attracts host of new players, *Houston Business Journal* (August). Downloadable from website http://www.bizjournals.com/houston/stories/1998/08/31/newscolum n4.html.

Prakash, S. 1995. Mortgage lenders see credit scoring as key to hacking through red tape, *American Banker* (August) 1.

Quittner, J. 2003. Credit cards: sub-prime's tech dilemma: with delinquencies and charge-offs on the rise, the industry examines the role of automated decisioning, *Bank Technology* 16(1) 19-23.

Rimmer, J. 2005. Contemporary changes in credit scoring, *Credit Control* 26(4) 56-60.

Rowland, J. B. 2003. Confidently evaluate small businesses with

credit scoring, *Business Credit* (March) 26-31.

Saar-Tsechansky, M., and F. Provost. 2004. Active sampling for class probability estimation and ranking, *Machine Learning* 54(2) 153-178.

Sandler, A. L., S.E. McGinn, and J.L. Barloon. 2000. Fair lending scrutiny of credit score-based underwriting systems, *ABA Bank Compliance* (March/April) 38.

Schiff, S. 2003. Two views on 2003: credit scoring likely to be year's "hot topic," *Rough Notes* 146(1) 105.

Stanton, J. 2001. Opportunities for reducing delinquencies and defaults in federal housing credit programs: A review of new technologies and promising practices, *Journal of Public Budgeting, Accounting & Financial Management* 13(2) 157-192.

Thomas, L.C. 2000. A survey of credit and behavioral scoring – forecasting financial risk of lending to consumers, *International Journal of Forecasting* 16(2) 163-167.

Timmons, H. 2002. The cracks in credit scoring: As loan defaults rise, worries grow about how creditors pick borrowers, *Business Week* (November) 25.

Wagner, H. 2005. The use of credit scoring in the mortgage industry, *Journal of Financial Services Marketing* 9(2) 179-183.

Wendel, C., and M. Harvey. 2003. Credit scoring: Best practices and approaches, *Commercial Lending Review* 18(3) 4-7.

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

Zuckerman, S. 1996. Taking small business competition nationwide, *U.S. Banker* (August) 24-28, 72.

Zhu, H., P.A. Beling, and G.A. Overstreet. 2001. A study in the

128

International Journal of Business and Information
A Two-step Method to Construct Credit Scoring Models with Data Mining
Techniques

combination of two consumer credit scores, *Journal of the Operational Research Society* 52(9) 974-980.

Hian Chye Koh is an associate professor and dean of the School of Business, SIM University (Singapore). He received his doctoral degree from Virginia Tech and is a Fellow Certified Public Accountant. He has published widely in the areas of data mining and statistical modeling. He has also served as a consultant to several small and medium enterprises, large organizations, and government bodies. His major research and teaching interest is in data mining.

Wei Chin Tan is a statistical reporting manager in Singapore Telecommunications Limited. She obtained her bachelor of science and her master of science degrees in statistics from the National University of Singapore. She has been involved as a consultant and developer in major data mining and data warehousing projects for large organizations and government agencies. With vast experience in data mining, she has current research interests in the use of data mining in marketing applications.

Chwee Peng Goh is a data miner at Great Eastern Life Assurance Co Ltd (Singapore). She has previously held appointments as a data mining consultant and statistician at National Computer Systems Pte Ltd (Singapore) and the Building and Construction Authority (Singapore), respectively. She received both her bachelor of social science degree (with honors) and her master of science degree from the National University of Singapore. She majored in statistics in both the degrees.