Extracting Rules from Imbalanced Data: The Case of Credit Scoring

Seyed Mahdi Sadatrasoul*
Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
Sadatrasoul@iust.ac.ir

Mohammad Reza Gholamian
Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
Gholamian@iust.ac.ir

Kamran Shahanaghi
Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
Shahanaghi@iust.ac.ir

Received: 25/Mar/2014 Revised: 08/Aug/2014 Accepted: 15/Sep/2014

Abstract

Credit scoring is an important topic, and banks collect different data from their loan applicant to make an appropriate and correct decision. Rule bases are of more attention in credit decision making because of their ability to explicitly distinguish between good and bad applicants. The credit scoring datasets are usually imbalanced. This is mainly because the number of good applicants in a portfolio of loan is usually much higher than the number of loans that default. This paper use previous applied rule bases in credit scoring, including RIPPER, OneR, Decision table, PART and C4.5 to study the reliability and results of sampling on its own dataset.

A real database of one of an Iranian export development bank is used and, imbalanced data issues are investigated by randomly Oversampling the minority class of defaulters, and three times under sampling of majority of non-defaulters class. The performance criterion chosen to measure the reliability of rule extractors is the area under the receiver operating characteristic curve (AUC), accuracy and number of rules. Friedman's statistic is used to test for significance differences between techniques and datasets. The results from study show that PART is better and good and bad samples of data affect its results less.

Keywords: Credit Scoring; Banking Industry; Rule Extraction; Imbalanced Data; Sampling

1. Introduction

In today's competitive economy, credit scoring is widely used in banking industry. Every day, individual's and company's records of past borrowing and repaying actions are gathered and analyzed by information systems. Banks use this information to determine the individual's and company's profit. Application (credit) scoring is one of the main issues in the process of lending[1]. Credit scoring is used to answer one key question – what is the probability of default within a fixed period, usually one year. Credit scoring use banks historical loans data to classify customer as good or bad.

There are many techniques suggested to perform classification in the credit scoring problems including statistical and intelligent techniques. Logistic regression is the most favorite statistical and traditional method used to assess the credit scores[2]. Harrell applied Linear discriminant analysis and he shown that it is as efficient as logistic regression[3]. There are also many intelligent techniques applied to the problem including neural networks, Bayesian networks, support vector machines, case based reasoning, decision trees, and etc. Some studies have shown that neural networks, SVM, decision trees and other intelligent techniques, are superior to statistical techniques [4-6].

In recent years hybrid techniques are also proposed and they are the main focus of many researchers. Hybrid techniques usually use different algorithms strengths to improve the other algorithms weaknesses. In some hybrid techniques both statistical and intelligent techniques are used together. There are so many miscellaneous hybridization algorithms used in the literature. Lee et al used a hybrid neural discriminant technique with BP neural network and discriminant analysis, the hybrid model showed better accuracy than the BP neural network and discriminant analysis individually[7]. In another study Lee and Chen introduced a two-staged hybrid procedure with artificial neural networks and multivariate adaptive regression[8]. Tsai and Chen divide hybrid approaches into four main categories, they also organized 4 experiments with different combinations of clustering algorithms and classifiers; among their experiments logistic regression and neural network hybrid shown the best accuracy[9]. Huang, Chen and Wang studied using

* Corresponding Author

Meta heuristic techniques in order to tune intelligent techniques parameters, An application of support vector machines, genetic algorithms and F-score is studied and showed better results than using the pure SVM model[10]. In the last decade, using Ensemble techniques increased in the area and in some cases they give better accuracy rate[11, 12].West, Dellena and Qian used Neural network ensemble strategies including cross validation, bagging and boosting for financial decision applications, it shown better accuracy rate and generalization ability[11]. Ensemble learning is an open issue in recent year's studies[13, 14].

Because of robustness and transparency needs and also the auditing process done by regulators on the credit scoring in some countries, Banks cannot use many of mentioned techniques [15].By using rule bases, banks can easily interpret the results and explore the rejecting reasons to the applicant and regulatory auditors. There is actually a little literature in the field of rule based credit scoring. Ben-Davide provides a new method for rule pruning and examined his method on the credit scoring data set[16]. Hoffmann et.al introduced a new learning method for fuzzy rule induction based on the evolutionary algorithms[17].Martens et al used the support vector machine for rule induction in the credit scoring problems[18]. Malhotra et. al. used the adaptive neuro fuzzy inference systems(ANFIS) for rule induction and showed that this method works betters from discriminant analysis on their own credit scoring dataset which is gathered from credit unions[19], they used the back propagation method to learn their Rules membership function to fit on the data. Baesens et.al. use and evaluate three neural network rule extraction techniques including Neurorule, Trepan, and Nefclass, for rule extraction in three real life data bases including German credit database, Bene1 and Bene2 credit database[20]. They showed Nerorule and Trepan yield better classification accuracy compared to the C4.5 algorithm and the logistic regression. Finally they visualize the extracted rule sets using decision table[21].

In the credit scoring context, imbalanced data sets frequently occur as the number of good loans in a portfolio of loans are usually much higher than the number of loans that default[22]. Its reported that defaults ratio are ten percent of the whole bank’s loan portfolio on average[23]. As mentioned practical studies show that the real credit scoring datasets are imbalanced. There are some but few studies which investigate imbalanced credit scoring data sets. Huang, Hung and Jiau proposed a strategy of data cleaning for handling imbalanced distribution of credit data is in order to avoiding problems of over fitting and relevance of classifiers[24]. Brown and Mues run several experiments based on different classifiers on five UCI and non UCI credit datasets, they balanced their samples on 70(good)/30(bad) [22]. Their experiments show that random forest and gradient boosting classifiers perform very well in the credit scoring context.

The aim of this paper is to use previous applied rule bases in credit scoring, including RIPPER, OneR, Decision table, PART and C4.5 to study the reliability and results of sampling on its own dataset. In order to extract invaluable rules bases the results are compared in terms of area under the receiver operating characteristic curve (AUC), Accuracy and number of rules.

This study is divided into four other major parts: section 2 describes the classification techniques used. Section 3 introduces the data, experiments settings, Section 4 discussed their results and finally study concluded in section 5.

2. Overview of Classification Techniques

This paper aims to extract the best rules from imbalanced data in the credit scoring context. For this purpose 5 rule based and tree induction (with the aim of rule induction) classifiers are selected. A brief description of these techniques is presented below.

2-1- C4.5

Decision trees split the data into smaller subsets using their nodes and at the end of each node there is a series of leaf nodes assigning a class to each of the observations. C4.5 build trees based on the concept of information theory[25].the entropy of a sample of K, can be computed by[25]:

$$\text{Entropy (k)} = -\, p_1 log_2(p_1) - \, p_2 log_2(p_2) \qquad (1)$$

Where $p_1(p_0)$ are the proportions of the class values 1(0) in the sample K, respectively. The attribute with the highest normalized information is used for division. The algorithm then occurs on the smaller subsets iteratively.

2-2- RIPPER

Repeated Incremental Pruning toProduce Error Reduction (RIPPER), is a rule based learning that builds a setof rules by considering minimizing the amount of error[26]. In the optimization step if the modified rule is better according to an MDL heuristic, rules are replaced with a modified one in order to reach a small rule set.

2-3- OneR

OneR is a one-level decision tree algorithm, which selects attributes one-by-one from a dataset and generates a different set of rules based on error rate. At last the attribute and its appropriate rule set with minimum error is selected[27].

2-4- Decision table

Decision Table algorithm build tables using a simple decision table majority classifier[28].It uses a ‘decision table’ to summarize the dataset. After finding the line in the decision table that fits the non-class values, a new data item is assigned a category. Then the wrapper method is employed to find a good subset of attributes for inclusion in the table. The likelihood of over-fitting is reduced by eliminating attributes that contribute little or nothing to a model of the dataset and at last a smaller, well-defined decision table is reached.

2-5- PART

Partial decision tree algorithm (PART) is a developed version of RIPPER and C4.5[29]. Its main improvement is that it does not need to perform global optimization like C4.5and RIPPER to produce rules. It uses the standard covering algorithm to generate a decision list, and avoids over pruning by inducing rules from partial decision trees.

3. Empirical Evaluation

In this section first the data set characteristics is described. Secondly dataset samples are explained and finally the performance analyses are done.

3-1- Data sets characteristics

An Iranian commercial bank real export loan dataset is used to evaluate the proposed algorithm. Table (1) shows the characteristics of the dataset. The initial dataset include 1109 corporate applicants' and 46 financial and non financial data in the period from 2007 to 2012. First, the data cleaning is done; it includes removing redundant, outlier's data and missing values. There were a few missing Values for some corporate, some of them lack financial data and others lack the result of their loans, in fact they were in the process of debt repay, some of them haven't applied for loan yet. So 387corporate are excluded. From 722 remainedcorporate,652 are credit worthy (90.3%) and other 70 was unworthy (9.7%). Dummy variables were created for the categorical variables (ex. Type of industry).Using dummy variables number of variables increased to 55.Table (1) summarizes the dataset characteristics before and after cleaning step.

Table 1: dataset description

status	Data size	Inputs variables		
		Total	Continuous	Categorical
Before cleaning	1109	46	38	8
After cleaning	722	55	34	21

Delinquency status was defined by Basel committee definition of "default" and used to generate a 1/0 target variable for modeling purposes (good = 1, bad = 0). Accounts with no more than three months or more in arrears were classified as good. Those that were currently three or more months in arrears, or had been three months in arrears, were classified as bad. The results and descriptions of the variables used are shown in table (5) in appendix (1).

3-2- Re-sampling setup

Table (2) shows the main imbalanced dataset and samples built in order to consider imbalanced issue. The main dataset has a 90/10 class distribution, a 75/25 ratio in percent class distribution is selected for balancing the data and the main database is altered in different scenarios to meet this distribution. The two most common preprocessing techniques are random minority oversampling (ROS) and random majority under sampling (RUS). In ROS, instances of the minority class (bad applicants) are randomly duplicated in the dataset. In RUS, instances of the majority class (good applicants) are randomly discarded from the dataset.

In this study four different balanced datasets are created using two mentioned techniques. First using ROS bad instances are duplicated and the "Oversampled dataset" is created. This duplication is done until the distribution of good/bad meets to 75/25 so the number of bad instances increased from 70 to 217 samples. In another re sampling scenario, using RUS, three different "Under sampled datasets" are created. In order to use all of the datasets, simple random sample without replacement is done. The Under sampled dataset are designed in a manner that each good applicant in the main dataset is included in one and only one of three different under sampled datasets. This reduction is done until the distribution of good/bad meets nearly to 75/25 so the number of good instances decreased for these three under sampled datasets sequentially to 218,226 and 208 samples.

Table 2: Different samples of dataset used

Dataset name	Data size	Good	Bad	Good/All percent
Main imbalanced dataset	722	652	70	90.3
Oversampled dataset	869	652	217	75.02
Under sampled dataset No.1	288	218	70	75.74
Under sampled dataset No.2	297	226	70	76.9
Under sampled dataset No.3	278	208	70	74.82

3-3- Performance analysis

Five different measures are used to analysis the performance of the constructed rule bases. The performance criterion chosen to measure the effect of significant difference in number of observations is the area under the receiver operator characteristic curve(AUC) statistic[22]. Confusion matrix is another favorable instrument used in performance evaluations as shown in table (3).Overall accuracy, Good precision and bad precision are important measures after the ROC measure, as they shown the classifications quality as another dimension.

The overall accuracy of successfully identifying loans is computed using equation (2)

$$\text{Overall accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \tag{2}$$

Table 3: The confusion matrix

		PREDICTED CLASS	
ACTUAL CLASS		Class= Worthy	Class= Unworthy
	Class=Worthy	a(TP)	b(FN)
	Class= Unworthy	c(FP)	d(TN)

The precision of successfully identifying non-default loans is computed using equation (3)

$$\text{Good precision}=\frac{TP}{TP+FP} \qquad (3)$$

The precision of successfully identifying default loans is computed using equation (4)

$$\text{Bad precision}=\frac{TN}{TN+FN} \qquad (4)$$

Compactness of rules is another issue in rule base systems. At a defined level of ROC and accuracy measures for two rule bases, the rule base which has lower number of rules is preferred.

4. Results and Discussions

All the experiments in this paper are done using 10 fold cross validation. Table (4) shows classification accuracy, number of rules and area under curve for five datasets. The best classification accuracy, the lowest number of rules and area under curve for each data set are bolded. The best results for all of experiments are also underlined. Three groups of experiments are done and their results are presented below:

Table 4: Performance measures on different datasets and classifiers

dataset	Method	AUC	Accuracy(ALL)%	Precision(Bad)%	Precision (good)%	Number of rules
Main imbalanced dataset	RIPPER	0.531	89.47	**31.3**	90.8	2
	Decision table	0.499	**90.3**	0	90.3	**1**
	OneR	0.494	89.20	0	90.2	3
	PART	**0.612**	87.40	27.7	**91.6**	28
	C4.5	0.574	87.11	20.5	90.9	19
Over sampled dataset	RIPPER	0.881	87.45	72.3	93.3	**15**
	Decision table	0.887	80.21	57.5	92.3	575
	OneR	0.643	76.87	55.2	81.5	45
	PART	<u>**0.941**</u>	<u>**90.22**</u>	75.8	<u>**96.2**</u>	22
	C4.5	0.93	90.1	<u>**76.1**</u>	95.8	48
Under sampled dataset No.1	RIPPER	0.594	72.92	37.5	77.3	3
	Decision table	0.492	**73.95**	0	75.3	**1**
	OneR	0.544	73.61	40	**77.5**	7
	PART	**0.667**	72.22	**42.6**	71.4	22
	C4.5	0.595	69.79	36.1	78.9	24
Under sampled dataset No.2	RIPPER	0.517	73.99	34.8	77.3	**1**
	Decision table	0.511	**75.67**	25	76.4	**1**
	OneR	0.518	71.62	29.4	77.1	6
	PART	**0.656**	71.28	**38.8**	**80.8**	17
	C4.5	0.535	69.93	32.7	78.4	25
Under sampled dataset No.3	RIPPER	0.538	71.94	38.9	76.9	2
	Decision table	0.525	**73.02**	22.2	74.7	**1**
	OneR	0.504	71.22	27.3	75	7
	PART	0.581	71.58	**42.4**	**79.5**	20
	C4.5	**0.596**	68.70	38	79.2	20

4-1-First group experiments (Data sets performance comparisons)

First a test set at the 5% level of importance from the best performer using Friedman's testis done against different datasets for all of performance measures. Its findings are as follows:

- It shows that the results of oversampling data set have a significant difference compared to other four datasets; it can be seen that oversampling and increasing the number of observations increase the results performance compared with other reduction techniques at a defined level of good/bad ratio (75/25).
- The three under sampled datasets haven't any significant difference in their results; it can be concluded that different good observations in three different datasets don't have an importance issue in the results.

- The main dataset and three under sampled datasets haven't any significant difference; another separated Friedman test for AUC confirmed this hypothesis.
- Number of rules doesn't have significant change in all of the datasets and techniques, exclude decision table. It shows significant difference and increase in number of rules in oversampled dataset.

The results can be also used to evaluate sampling for different scenarios. Fig. (1) Shows that from one hand oversampling enhance the performance measures totally except number of rules and on the other hand under sampled datasets have no important difference in performance measures. It can be concluded that although the under sampled datasets have lower records but this do not affected their results comparing to the main imbalanced data set.

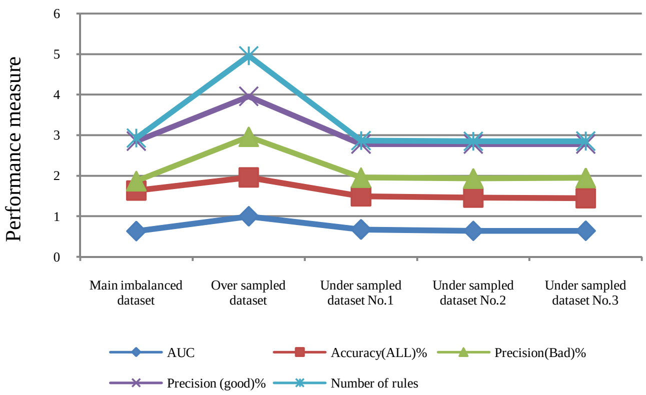

Fig. 1. Status of different performance measures for different samples (the results are standardized).

4-2-Second group experiments (Classifiers performance comparisons)

Second test set at the 5% level of importance using Friedman's test is done for different classifiers against all of performance measures. Its findings are as follows:

- The OneR, decision table and RIPPER haven't any significant difference between each other but have significant difference with other classifiers, they are the worst performers.
- The PART have significant difference with other classifiers, it is the best performer.

4-3-Third group experiments (Classifiers performance comparisons)

Third test set at the 5% level of importance using Friedman's testis done for different classifiers against three main performance measures. Its findings are sorted by their importance and presented below:

- **AUC measure:** The OneR, decision table and RIPPER haven't any significant difference between each other and they are the worst players, but PART and C4.5 have major difference with worst players and with each other. PART is the best performer throughout this measure.
- **Accuracy measure:** All of the classifiers haven't any significant difference between each other under the accuracy measure.
- **Number of rules measure:** The OneR, decision table and RIPPER haven't any significant difference between each other, they are the best players, also PART and C4.5haven't any significant difference between each other and they are the best players.

In brief, when considering different measures based on their importance it can be concluded that PART and after it C4.5 have a very good performance in different levels of class imbalance. However decision table, OneR and RIPPER are the worst performers. The mentioned results were attractive in the oversampled dataset and the results of two best classifiers on this dataset can be used for credit scoring classification. Fig (2) shows the results in brief.

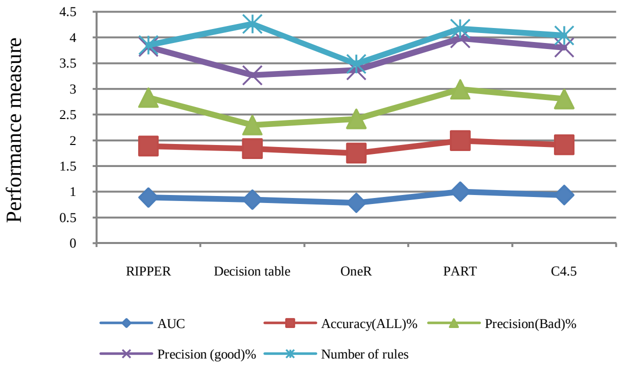

Fig. 2. Status of different performance measures for different classifiers (the results are standardized).

Table (5) shows the ranks against each performance measure for different classifiers. Note that statistical significant test do not checked for this table.

Table 5: Classifiers rank against different measures

Performance measure name	Classifiers rank
AUC	PART> C4.5>RIPPER> Decision table> OneR
Accuracy(ALL)%	RIPPER> Decision table>PART> C4.5> OneR
Precision(Bad)%	PART>RIPPER> C4.5> OneR> Decision table
Precision (good)%	C4.5> PART>RIPPER> Decision table> OneR
Number of rules	RIPPER> OneR>PART> C4.5> Decision table

5. Conclusion

In this paper, a number of different classifiers are used and compared on various balanced and imbalanced datasets. These techniques include RIPPER,C4.5, PART, OneR and Decision table. An imbalanced dataset from a major Iranian bank is applied and balanced using oversampling and several random under sampling techniques. Classifiers and datasets are compared using five different performance measures and Friedman's test. The results of the study shows that random oversampling of bad loans yield to better performance measures for all of the classifiers. It is also found that

PART classifier is perform better on imbalanced data than other classifiers and that it's the best performer at all of the experiments and performance measures except number of rules. On the other hand OneR and decision table techniques are the worst classifiers at all.

Next researches can focus on using other oversampling methods and their effect on the classifiers training. Studying the effect of different sampling methods on feature selection is also another open area of future researches.

Acknowledgement

The authors' kindly acknowledge MR. zekavat for his kind cooperation.

Appendix (1) Variables included in Iran credit dataset and their types are shown in table (6).

Table 6: list of variables in Iran commercial bank credit dataset

Variable	Type	Variable	Type
Net profit	Continuous	Type of industry: industry and mine (=1, other =0)	Categorical
Active in internal market	Categorical	Type of industry: agricultural (=1, other =0)	Categorical
number of countries that the company export to	Categorical	Type of industry: oil and petrochemical (=1, other =0)	Categorical
Sales growth	Categorical	Type of industry: infrastructure and service(=1, other =0)	Categorical
Target market risk (from 1 to 5)	Categorical	Type of industry: chemical (=1, other =0)	Categorical
Seasonal factors	Categorical	Year of financial ratio	Continuous
Company history(number of years)	Categorical	Type of book: Tax declaration(=1,other=0)	Categorical
Top Mangers history	Categorical	Type of book: Audit Organization (=1,other=0)	Categorical
Type of company: Cooperative (=1, other =0)	Categorical	Type of book: Accredited auditor (=1,other=0)	Categorical
Type of company: Stock Exchange(LLP) (=1, other =0)	Categorical	Inventory cash	Continuous
Type of company: Generic join stock(PJS) (=1, other =0)	Categorical	Accounts receivable	Continuous
Type of company: Limited and others (=1, other =0)	Categorical	Other Accounts receivable	Continuous
Type of company: Stock Exchange (=1, other =0)	Categorical	Stock	Continuous
Experience with Bank (number of years in 5 categories)	Categorical	Current assets	Continuous
Audit report Reliability	Categorical (binary)	Non-current assets	Continuous
Current period sales	Continuous	Total assets	Continuous
Prior period sales	Continuous	Short-term financial liabilities	Continuous
Two-Prior period sales	Continuous	Current liabilities	Continuous
Current period assets	Continuous	Long-term financial liabilities	Continuous
Prior period assets	Continuous	Non-current liabilities	Continuous
Two-Prior period assets	Continuous	Total liabilities	Continuous
Current period shareholder Equity	Continuous	Capital	Continuous
Prior period shareholder Equity	Continuous	Accumulated gains or losses	Continuous
Two-Prior period share holder Equity	Continuous	shareholder Equity	Continuous
checking accounts creditor turn over	Continuous	Sale	Continuous
checking Account Weighted Average	Continuous	Gross profit	Continuous
Average exports over the past three years	Continuous	Financial costs	Continuous
Last three years average imports	Continuous	y (nonworthy/worthy)	Categorical (binary)

References

[1] Van Gestel, T. and B. Baesens, Credit risk management: basic concepts: financial risk components, rating analysis, models, economic and regulatory capital. 2009: Oxford University Press, USA.

[2] Wiginton, J.C., A note on the comparison of logit and discriminant models of consumer credit behavior. Journal of Financial and Quantitative Analysis, 1980. 15(03): p. 757-770.

[3] Harrell, F.E. and K.L. Lee, A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality. Biostatistics: Statistics in Biomedical; Public Health; and Environmental Sciences. The Bernard G. Greenberg Volume. New York: North-Holland, 1985: p. 333–343.

[4] Crook, J.N., D.B. Edelman, and L.C. Thomas, Recent developments in consumer credit risk assessment. European Journal of Operational Research, 2007. 183(3): p. 1447-1465.

[5] Huang, Z., et al., Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision support systems, 2004. 37(4): p. 543-558.

[6] Ong, C.S., J.J. Huang, and G.H. Tzeng, Building credit scoring models using genetic programming. Expert Systems with Applications, 2005. 29(1): p. 41-47.

[7] Lee, T.S., et al., Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications, 2002. 23(3): p. 245-254.

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

[9] Tsai, C.F. and M.L. Chen, Credit rating by hybrid machine learning techniques. Applied soft computing, 2010. 10(2): p. 374-380.

[10] Huang, C.L., M.C. Chen, and C.J. Wang, Credit scoring with a data mining approach based on support vector machines. Expert Systems with Applications, 2007. 33(4): p. 847-856.

[11] West, D., S. Dellana, and J. Qian, Neural network ensemble strategies for financial decision applications. Computers & operations research, 2005. 32(10): p. 2543-2559.

[12] Tsai, C.F. and J.W. Wu, Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Systems with Applications, 2008. 34 (4): p. 2639-2649.

[13] Louzada, F., et al., Poly-bagging predictors for classification modelling for credit scoring. Expert Systems with Applications: An International Journal, 2011. 38(10): p. 12717-12720.

[14] Finlay, S., Multiple classifier architectures and their application to credit risk assessment. European Journal of Operational Research, 2010.

[15] Thomas, L.C., Consumer credit models: pricing, profit, and portfolios. 2009: Oxford University Press, USA.

[16] Ben-David, A., Rule effectiveness in rule-based systems: A credit scoring case study. Expert Systems with Applications, 2008. 34(4): p. 2783-2788.

[17] Hoffmann, F., et al., Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. European Journal of Operational Research, 2007. 177(1): p. 540-555.

[18] Martens, D., et al., Comprehensible credit scoring models using rule extraction from support vector machines. European Journal of Operational Research, 2007. 183(3): p. 1466-1476.

[19] Malhotra, R. and D.K. Malhotra, Differentiating between good credits and bad credits using neuro-fuzzy systems. European Journal of Operational Research, 2002. 136(1): p. 190-211.

[20] Baesens, B., et al., Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, 2003: p. 312-329.

[21] Baesens, B., et al., Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, 2003. 49(3): p. 312-329.

[22] Brown, I. and C. Mues, An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 2011.

[23] Dinh, T.H.T. and S. Kleimeier, A credit scoring model for Vietnam's retail banking market. International Review of Financial Analysis, 2007. 16(5): p. 471-495.

[24] Huang, Y.M., C.M. Hung, and H.C. Jiau, Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. Nonlinear Analysis: Real World Applications, 2006. 7(4): p. 720-747.

[25] Quinlan, J.R., C4. 5: programs for machine learning. 1993: Morgan kaufmann.

[26] Cohen, W.W., Learning Trees an ules with Set-val ed Features. 1996.

[27] Holte, R.C., Very simple classification rules perform well on most commonly used datasets. Machine learning, 1993. 11(1): p. 63-90.

[28] Kohavi, R., The power of decision tables. Machine Learning: ECML-95, 1995: p. 174-189.

[29] Frank, E. and I.H. Witten, Generating accurate rule sets without global optimization. 1998.

Seyed mahdi sadatrasoul is a Ph.D student in industrial engineering and systems management at Iran University of Science and Technology (IUST), Tehran. He received his Bs degree in Industrial Engineering from IUST, in 2006 and obtained M.S. degree in information technology management from Tarbiat modares university (TMU), Tehran, in 2009. Presently he is the assistant of faculty member at IT Group in School of Industrial Engineering and is actively engaged in conducting academic, research and development programs in the field of data and process mining. He has contributed more than 20 research papers to many national and international journals and conferences. He has also published two books by reputed publishers. His research interests are including data mining and its synergies with operation research, credit allocation and scoring, e- commerce and financial information systems (FIS).

Mohammad Reza Gholamian is an Assistant Professor in School of Industrial Engineering at the Iran University of Science and Technology (IUST), Tehran. He received his M.S. degree in Industrial Engineering from Isfahan University of Technology (IUT), Isfahan in 1998 and obtained Ph.D. degree in Industrial Engineering from Amirkabir University of Technology (AUT), Tehran in 2005 for the work in the field of Hybrid Intelligent Decision Making Systems. Presently he is a faculty member of IT Group in School of Industrial Engineering and is actively engaged in conducting academic, research and development programs in the field of Industrial Engineering and Information Technology. He has contributed more than 105 research papers to many national and international journals and conferences. Besides this, he has published four books by reputed publishers. His research interests include data mining, soft computing, and decision theory and e-business models.

Kamran Shahanaghi is an Assistant Professor in School of Industrial Engineering at the Iran University of Science and Technology (IUST), Tehran. He received his M.S. degree in Industrial Engineering from IUST in 1986 and obtained Ph.D. degree in 2000. Presently he is a faculty member of optimization Group in School of Industrial Engineering and is actively engaged in conducting academic, research and development programs in the field of Industrial Engineering and optimization. He has contributed more than 140 research papers to many national and international journals and conferences. His research interests include operation research and uncertainty.