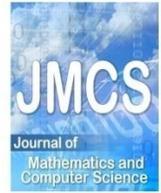


Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com



Utility based Credit Scoring for Banks and Financial Institutions: Case Study of a Major Iranian Bank

Seyed Mahdi Sadatrasoul¹, Mohammad Reza Gholamian², Zeynab Hajimohammadi³, Mahdi Hosseini⁴

^{1,2}*School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran*

³*Amirkabir University of Technology, Department of computer science, Tehran, Iran*

⁴*Advanced Information Systems Research Group, ICT Research Institute of ACECR, Tehran, Iran*

Email: ¹sadatrasoul@iust.ac.ir, ²gholamian@iust.ac.ir, ³zhajimohammadi@aut.ac.ir, ⁴mahdi.Hosseini@ictrc.ir,

Article history:

Received April 2014

Accepted September 2014

Available online October 2014

Abstract

Credit scoring mainly distinguishes good customers from the bad ones; therefore it is a classification problem. There are many techniques introduced to solve the problem such as support vector machines, neural networks and rule based classifiers. The main objective of this process is to maximize the profit of bank or financial institute. However these traditional methods of classification seem not to support this objective well. This paper investigates this issue and shows that the best classification model is not necessarily the most profitable model. The applications of the models are shown on an ironing real credit dataset since 2007 to 2012.

Keywords: Credit Scoring; Banking Industry; Classification; Utility based data mining.

1.0 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, 2]. In this paper, we will address the credit scoring problem. Credit scoring is used to answer one key question, i.e., what is the probability of default within a fixed period, usually one year. Credit scoring uses banks historical loans data to classify customers in terms of creditworthiness as good or bad.

There are many techniques suggested to perform classification in the credit scoring problems include statistical and intelligent techniques. Logistic regression is the most favorite statistical and traditional method used to assess the credit score[3]. Linear discriminant analysis is also applied and it is shown that it is as efficient as logistic regression[4]. There are also various intelligent techniques applied to the

problem including neural networks, Bayesian networks, support vector machines, case based reasoning, decision trees, etc. There are also studies which used clustering for scoring customers credit[5]. Some studies have shown that neural networks, SVM, decision trees and other intelligent techniques, are superior to statistical techniques [4, 6-8].

In recent years hybrid techniques are also proposed and they are the main focus of many researchers. Hybrid techniques usually use the strengths of different algorithms to improve the weaknesses of other algorithms. In some hybrid techniques both statistical and intelligent techniques are used together. There is various hybridization algorithms used in the literature. A hybrid neural discriminant technique with BP neural network and discriminant analysis proposed, and showed better accuracy than the BP neural network and discriminant analysis[9]. A two-stage hybrid procedure with artificial neural networks and multivariate adaptive regression is also proposed in[10]. In a study hybrid approaches are divided into four main areas and different combinations of clustering algorithms and classifiers are tested; logistic regression and neural network hybrid shown the best accuracy[11]. There are also studies which hybrid Meta heuristic techniques with intelligent techniques. An integration of support vector machines, genetic algorithms and F-score is studied[12]. In the last decade, using Ensemble techniques increased in this area and in some cases it gives better accuracy rate[13, 14]. Neural network ensemble strategies including cross validation, bagging and boosting for financial decision applications, are studied and shown better accuracy rate and generalization ability[13]. However, Ensemble learning is yet an open issue in recent year's studies[15, 16].

Although these methods help banks distinguish risk of customers, they do not effective for maximizing the banks profit. Since the probability of default is the main issue in building such models, it can be asserted that the higher the accuracy of a model, the better the quality is. The main objective of this paper is to estimate the utility value of classification models. It is obvious that evaluating the classifiers using utility value has much more value for banks and it is more useful in this problem. Utility based data mining was introduced as a cost sensitive learning. Utility based data mining is concerned with profit or cost and it opposes to the concept of accuracy which is measured using correctly classified instances.

Utility-based data mining is almost a new field of study in machine learning discipline. Knowledge extraction and application usually takes place under complex circumstances that were rarely under consideration through early work in predictive data mining and machine learning. It is often assumed that training data is freely available and focused on simple objectives, e.g. predictive accuracy. Over time, there has been a growing interest in developing methodologies for economical data acquisition and evaluating utility-based data mining techniques. Different problems in the field of utility-based data mining were studied, e.g. cost-sensitive learning, cost of data acquisition, and active query learning [17]. Several studies used the concept of utility to assess the financial value of classification models. Arnt and Zilberstein[18] developed a classification model using three different types of cost simultaneously, namely, cost—error cost, feature measurement cost and response time cost. Ciraco et al.[19] evaluated the relation between change in misclassification cost ratios and the improvement in the utility derived from a classification model by setting the ratio as 1:10, 2:10, 3:10, ..., 10:10, 10:9, 10:8, ..., 10:1. In another work, Zadrozny used a cost-sensitive learning approach in order to find the optimal classification model by maximizing the expected benefit[20]. Another approach was proposed by Chawla and Li [21]in which the profit of customers is considered based on their individual probability of default. Additionally, some researches considered the prediction of null attributes in the preprocessing phase based on attribute cost [22, 23]. Other studies, done on different fields, include model performance assessment based on utility-based measure[24], ROI maximization [25] data quantity assessment in the case of high learning costs[26], Sampling methods for heavily skewed target value distributions[27] and Selection of High utility item sets [28].

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.

2.0 Building Classification Models to Estimate Banks Profit

This paper aims to develop classification models for credit scoring problem and evaluate profit or utility of models for banks. For this purpose, four steps are designed and described as follows.

2.1. Data Cleaning

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 includes 1109 corporate applicants' and 46 financial and non-financial data in the period from 2007 to 2012. The first step, namely data cleaning, includes removing redundant, outlier's data and missing values. There were few missing values for some corporate, some of them lack financial data and others lack the result of their loan, most of which were in the process of debt repay, some of them haven't applied for loan yet. Therefore the missing value elimination technique is used and a total number of 387 corporates with mentioned characters are excluded[29]. From 722 remained corporates, 652 were credit worthy (90.3%) and other 70 were unworthy (9.9%). 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.

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

Table 1: Dataset description

Status	Data Size	Inputs Variables		
		Total	Continuous	Categorical
Before cleaning	1109	46	38	8
After cleaning	722	55	34	21

2.2. Sample balancing

The main dataset has a 90/10 class distribution. It can be seen that it is imbalanced and model building on imbalanced data has many problems including over fitting and having poor rate of learning. Using at least 1000 Goods and 500 Bads generally leads to an acceptable model. In general, the more the balance between Goods and Bad ones, the more accurate the resulting score is[15]. To avoid over fitting, the G/B odds ratio of 3:1 is used in this paper[30, 31]. Also since over-sampling generally gives better performance than under-sampling, random minority oversampling (ROS) is used to oversample the applicants labeled as Bad. Finally, the total number of data, number of Goods and number of Bads are reported in table (2).

Table 2: Tuned dataset

<i>Dataset Name</i>	<i>Data Size</i>	<i>Good</i>	<i>Bad</i>	<i>Good / All (%)</i>
Main imbalanced dataset	722	652	70	90.3
Tuned dataset	869	652	217	75.02

2.3 Selecting the classifiers

Different classifiers are built to assess the utility derived from the models versus accuracies. These classifiers include CHAID, Quest, C&R, C5, neural networks and linear regression. Feature selection methods are also used with all of the classifiers. It is expected that in some cases it increase the accuracy rate[2, 32]. Clementine 11 is used to build models. Pearson test at the significant level of 95% is used to include the attributes in the model. Other classifiers are built on the default setting in Clementine.

2.4. Analysis of Performance

Two different measures are used to analyze the performance of the models. Confusion matrix, shown in table (3), as a favorable instrument is used in performance evaluations. Overall accuracy, computed using (1), is used as the measure of validity.

$$\text{Overall accuracy} = (TP + TN) / (TP + TN + FN + FP) \quad (1)$$

For evaluating the profit, first the models were needed to be built. Profit models for customers are complex and differ for different types of loans and financial services. A typical profit model is explained in[33]. The proposed profit model, as shown in (2), is an enhanced version of the typical model presented in[33].

$$\text{Total Profit} = \text{Profit of Loan (Interest Rate)} + \text{Cross Selling Profit} - \text{Cost of Fund} - \text{Default Cost} - \text{Overhead Cost} \quad (2)$$

Turning (2) to the real world for each cell of confusion matrix yields to (3-6).

$$U(A) = L \times I_L + CA \times Y \times I_{CA} - L \times I_{FC} - k \times OC \quad (3)$$

$$U(B) = -L \times I_L + CA \times Y \times I_{CA} + L \times I_{FC} + k \times OC \quad (4)$$

$$U(C) = -L - L \times I_L + CA \times Y \times I_{CA} - L \times I_{FC} - k \times OC \quad (5)$$

$$U(D) = L + L \times I_L + CA \times Y \times I_{CA} - L \times I_{FC} - k \times OC \quad (6)$$

Where, L is the amount of loan, IL is the interest rate, IFC is the cost of fund interest rate, OC is the overhead cost, K is the coefficient for overhead cost, CA is checking account yearly weighted average, Y is the number of years cooperating with bank, and ICA is checking account interest rate.

Table 3: The confusion matrix

	Predicted Class	
	Label of the Class	
Actual Class	Worthy	A (TP) B (FN)
	Unworthy	C (FP) D (TN)

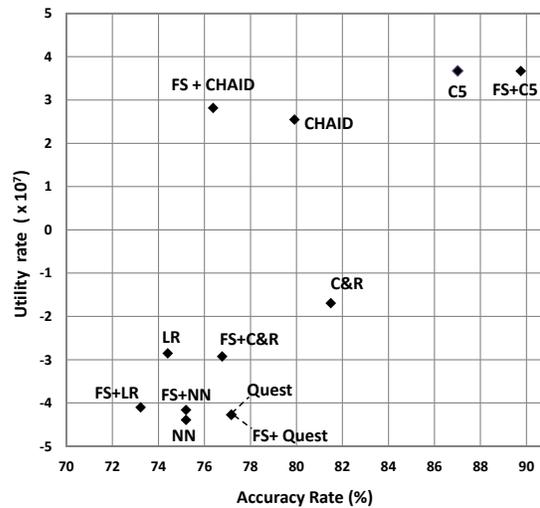
3. Results and Discussion

Table (4) shows the twelve classification techniques evaluated in this research. Accuracy rate and utility value of each model is also reported. Furthermore, results of the models which outperformed others are bolded. It can be seen that using C5 classifier, added with feature selection, had the best performance in terms of accuracy. But from the utility point of view, application of C5 without any feature selection resulted in the highest level of performance. As it can be seen accuracy of the C5 model is lower than the model with feature selection as preprocessing.

Table 4: Dataset description

Method	LR	NN	C5	C&R	Quest	CHAID
Accuracy	74.41	75.2	87.01	81.5	77.17	79.92
Utility	-28559469.63	-43942579.33	36703569	-16968923.58	-42717742.4	25477624
Method	FS+LR	FS+NN	FS+C5	FS+C&R	FS+ Quest	FS+CHAID
Accuracy	73.23	75.2	89.76	76.77	77.17	76.38
Utility	-41040561.96	-41622519.44	36677091	-29279887.31	-42717742.4	28157101

Fig. 1 shows a better view of classifiers. X axis shows the accuracy rate and Y axis shows the utility value. The classifiers which are placed at the top right side of the figure performed better and the classifiers with less satisfactory results are placed at the bottom left side. Next to C5, CHAID added with feature selection, and CHAID are the best performers among all. In this case it can be seen that although the combination of CHAID and feature selection is better in terms of utility but it shows the worst results in terms of accuracy rate. In the case of C&R, the simple model without feature selection performs better in terms of both indexes, namely, utility and accuracy. In other words, the model which has the better accuracy rate has better utility. Furthermore, other classifiers, including linear regression, show



results similar to C&R.

Fig 1: Accuracy rates versus utility for each classifier

4. Conclusion

In this paper, a number of different classifiers were used and compared based on accuracy rate and Utility of the financial institutions derived from models. These Classifiers include C5, Linear regression, neural network, Quest, CHAID and C & R. Experiments are setup based on an imbalanced dataset from a major Iranian bank. The dataset was balanced using random oversampling of bad applicants by 3:1 ratio

of goods/bads. The results reveal that the model with the best accuracy rate does not necessarily show the best utility rate. Therefore the bank's risk score is related to the profit score because in general there is a significant correlation but they are not permanently related to each other. The C5 algorithm with feature selection was the best in terms of accuracy rate but simple C5 without feature selection was the best performer in terms of model utility for banks. Therefore it is recommended that the banks use utility based model evaluation to maximize their profit.

Next researches can focus on using Multi attribute utility theory to construct the utility value of customers and evaluate the results with it.

Acknowledgement

The authors wish to acknowledge Mr. zekavat the manger of risk and Mr. soleimani the R&D manager of export development bank of Iran (EDBI) for their kind cooperation.

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. Somol, P., et al., *Filter - versus wrapper - based feature selection for credit scoring*. International journal of intelligent systems, 2005. **20**(10): p. 985-999.
3. 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.
4. 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', North-Holland, New York, United States, 1985: p. 333-343.
5. sadatrasoul, s.m., et al., *A New Method for Clustering in Credit Scoring Problems*. Journal of Mathematics and Computer Science, 2013. **6**(2): p. 97-106.
6. 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.
7. 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.
8. 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.
9. Lee, T.S., et al., *Credit scoring using the hybrid neural discriminant technique*. Expert Systems with Applications, 2002. **23**(3): p. 245-254.
10. Lee, T.S. and I.F. 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.
11. Tsai, C.F. and M.L. Chen, *Credit rating by hybrid machine learning techniques*. Applied soft computing, 2010. **10**(2): p. 374-380.
12. 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.
13. West, D., S. Dellana, and J. Qian, *Neural network ensemble strategies for financial decision applications*. Computers & operations research, 2005. **32**(10): p. 2543-2559.
14. 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.

15. Finlay, S., *Multiple classifier architectures and their application to credit risk assessment*. European Journal of Operational Research, 2011. **210**(2): p. 368-378.
16. Louzada-Neto, F., et al., *Poly-bagging predictors for classification modelling for credit scoring*. Expert Systems with Applications, 2011.
17. Abe, N. *Machine learning paradigms for utility-based data mining*. 2005: ACM.
18. Arnt, A. and S. Zilberstein. *Learning policies for sequential time and cost sensitive classification*. 2005: ACM.
19. Ciraco, M., M. Rogalewski, and G. Weiss. *Improving classifier utility by altering the misclassification cost ratio*. 2005: ACM.
20. Zadrozny, B. *One-Benefit learning: cost-sensitive learning with restricted cost information*. 2005: ACM.
21. Chawla, N. and X. Li. *Pricing based framework for benefit scoring*. 2006.
22. Melville, P., et al. *Economical active feature-value acquisition through expected utility estimation*. 2005: ACM.
23. Melville, P., et al. *An expected utility approach to active feature-value acquisition*. 2005: IEEE.
24. Yao, H., H.J. Hamilton, and L. Geng. *A unified framework for utility-based measures for mining itemsets*. 2006: Citeseer.
25. Yan, L. and P. Baldasare. *Beyond classification and ranking: constrained optimization of the ROI*. 2006: ACM.
26. Weiss, G.M. and Y. Tian, *Maximizing classifier utility when training data is costly*. ACM SIGKDD Explorations Newsletter, 2006. **8**(2): p. 31-38.
27. McCarthy, K., B. Zabar, and G. Weiss. *Does cost-sensitive learning beat sampling for classifying rare classes?* 2005: ACM.
28. Tseng, V.S., C.J. Chu, and T. Liang. *Efficient mining of temporal high utility itemsets from data streams*. 2006: Citeseer.
29. Tan, P.N., *Introduction to data mining*. 2007: Pearson Education India.
30. Chuang, R.J. and M.J. Chen, *The building of credit scoring system on the residential mortgage finance*. International Journal of Forecasting, 2006. **15**(2): p. 65–90.
31. Chi, B.W. and C.C. Hsu, *A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model*. Expert Systems with Applications, 2011.
32. Liu, Y. and M. Schumann, *Data mining feature selection for credit scoring models*. Journal of the Operational Research Society, 2005. **56**(9): p. 1099-1108.
33. 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.