A Proposed Classification of Data Mining Techniques in Credit Scoring

Abbas Keramati
Department of Industrial Engineering
University of Tehran, Tehran 1473984954, Iran

Niloofar Yousefi
Department of Industrial Engineering
University of Tehran, Tehran 1439955961, Iran

Abstract

Credit scoring has become very important issue due to the recent growth of the credit industry, so the credit department of the bank faces a large amount of credit data. Clearly it is impossible analyzing this huge amount of data both in economic and manpower terms, so data mining techniques were employed for this purpose. So far many data mining methods are proposed to handle credit scoring problems that each of them, has some prominences and limitations than the others, but there is no a comprehensive reference introducing most used data mining method in credit scoring problem. The aim of this study is providing a comprehensive literature survey related to applied data mining techniques in credit scoring context. Such reference can help the researchers to be aware of most common methods in credit scoring evaluation, find their limitations, improve them and suggest new method with better capabilities. At the end we notice the limitation of the most proposed methods and suggest the more applicable method than other proposed.

Keywords
Data mining, credit risk management, credit scoring, literature review.

1. Introduction
Increasing the demand for consumer credit has led to the competition in credit industry. So credit managers have to develop and apply machine learning methods to handle analyzing credit data in order to saving time and reduction errors. Credit scoring can be defined as a technique that helps lenders decide whether to grant credit to the applicants with respect to the applicants' characteristics such as age, income and marital status[1]. In recent years, several quantitative methods have been proposed for credit risk evaluation. Among all existent approaches, data mining methods have found more popularity than the others because of their ability in discovering practical knowledge from the database and transforming them into useful information. The first researches into credit scoring were done by Fisher and Durand, who applied linear and quadratic discriminant analysis respectively to categorize credit applications as good or bad ones. This study aims to prepare a literature survey in data mining technique applied in credit risk evaluation problem from 2000 to 2010. The main purpose of this study is helping to researchers to be aware of the present methods, find their limitations and suggest more efficient methods.

2. Credit Scoring
Credit scoring models are known as statistical models which have been widely used to predict the default risk of individuals or companies. These are multivariate models which use the main economic and financial indicators of a company or individuals' characteristic such as age, income and marital status as input, assign them a weight which reflect its relative importance in predicting default. The result is an index of creditworthiness that is expressed as a numerical score, which measures the borrower's probability of default. The initial credit scoring models are devised in the 1930s by authors such as [2] and [3], and developed with studies by [4], [5] and others from 1960. [6] Has done a comprehensive survey on credit and behavioral scoring.

2.1 Type of scoring
Based on [7] there are different kinds of scoring:

- Application Scoring: This king of scoring consists of estimation creditworthiness of a new applicant who applies for credit. It estimates credit risk with respect to social, demographic financial conditions of a new applicant to decide whether to grand credit to them or not.
- Behavioral Scoring: It is similar to application scoring with difference that it involves existing customers, so the lender has some evidences about borrower's behavior to dynamic management of portfolio process.
- Collection Scoring: Collection scoring classifies customers into different groups according to their levels of insolvency. In the other words, it separates the customers who need more decisive actions from those who do not require to be attended to immediately. These models are used in order to management of delinquent customers from the first signs of delinquency.
- Fraud Detection: It categorizes the applicants according to the probability that an applicant be guilty.

2.2 Application of Credit Scoring Models

Reference [8] Introduce three major credit scoring application:

- Credit scoring for credit cards
- Credit scoring for mortgages
- Credit scoring for small business lending

3. Research Methodology

For doing this research, we searched online journal databases which some of them are referenced here. These databases include Science Direct; Scopus, Emerald, Springer, IEEE Explore, Jstore, John Wiley, Academic Search Primier. The searched Journals include Expert Systems with Applications, European Journal of Operational Research, Computers & Operations Research, Computer Engineering and Applications, Journal of Empirical Finance, Journal of Banking & Finance, Journal of Business Economics and Management, Mathematics and Economics, IEEE international conference papers. It is also used some books in credit risk and data mining context.

Credit Scoring, Credit Risk, Credit Rating, Data Mining, Classification, Neural Network, Bayesian Classifier, Discriminant Analysis, Logistic Regression, Decision Tree, K-Nearest Neighbor, Support Vector Machine, Fuzzy Rule-Based System, Survival Analysis and Hybrid Model were the most useful keywords to find articles.

This survey is done in two stages: in first stage we selected the articles related to the credit scoring, by reviewing abstracts and titles. In this stage, it is identified approximately two hundred articles based on the descriptors "credit scoring" and "data mining". Then we extract most used data mining methods in credit scoring context from the articles and some books. At next stage, the articles associated with each data mining method applied for credit scoring task were collected from 2000 to 2010.

4. Literature Survey

4.1 Data Mining

Nowadays each individual and organization – business, family or institution – produces and collects huge volume of data about itself and its environment. Based on [9] " Data mining is the process of selection, exploration, and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database." Data mining is a useful tool for taking strategic decision and play important role in market segmentation, customer services, fraud detection, credit and behavior scoring and benchmarking [10], [6]. In this study we consider ten data mining method employed for credit scoring.

4.2 Data Mining Techniques Employed For Credit Scoring

Several single and hybrid data mining methods are applied for credit scoring problem [11], [12], [13], [14], [15]. The most used applied methods for doing credit scoring task are derived from classification technique. Classification can involve any context in which some decision or forecast is made on the basis of available information. It can be defined as a method which classifies the members of a given set of instances into some groups in terms of their characteristics. Classification task is very suited to data mining methods and techniques.

Reference [16] Prepare a study that investigates several classification techniques in term of their advantages and limitations in credit risk assessment. They consider Probit, Neural Network, decision tree and k-NN models and some combination of these methods for this purpose. The k-NN method obtains the best result among the mentioned methods. [17]Employ Multi-group Hierarchical DIScrimination (M.H.DIS) method for classifying purpose in credit risk context and measure the efficiency of this model against some traditional methods such as discriminant analysis, logit analysis, and probit analysis. They conclude that the proffered model has more classification ability versus benchmarked methods. [18] Assess the ability of some classification methods in handling credit scoring task. These methods include linear discriminant analysis, logistic regression, neural networks, k-nearest neighbor, support vector machines, classification and regression tree and multivariate adaptive regression splines. Based on this study, SVM, MARS, logistic regression and neural network prepare very good classification, though LDA and CART are very user-friendly tools in building a credit scoring model. [19] Probe the suitability of self organization maps (SOMs) for credit scoring. They also note the advantages of application integrated SOM with some supervised classifier instead of solely SOMs method. [20]Peruse eleven classifiers of five different types included Bayesian theory, Neural networks, Statistical tools, Machine learning methods, Kernel based model. They do this study to finding the best method in term of prediction accuracy of default probability. It is shown in this study that the kernel based RBF neural network is the best method in identifying the true positives. [14] Compare the profitability of seven data mining classification techniques: naive Bayes, logistic regression, neural network, decision table, decision tree, k-nearest neighbor, and support vector machine with and without unifying domain knowledge. For this purpose, misclassification cost and area under the curve (AUC) are used for analyzing the results. Their study findings show that incorporating domain knowledge improves the effectiveness of some data mining methods. [21]Use a mathematical programming (MP) discriminant analysis method and compare its performance with logistic regression, discriminant analysis, k-NN classifier and support vector machine technique. Evaluate the ability of discriminant analysis, logistic regression, neural network and classification and regression tree in prediction and classification task. CART and neural network outperform the others method based on this study. [22] Review the applicability of some binary classification methods in term of predictive and classification accuracy using fourteen datasets. [23] Explore the classification accuracy of K-Nearest Neighbor, Support Vector Machine and Neural Network, without features selection. The results indicate that integrating these methods with effective feature selection approaches lead to more accurate classification. [24] Investigate the capability of five classifier and pairs of classifier ensembles in handling of credit risk prediction. The results show that combining the individual classifier can improve the accuracy of predictive models. [7] Design an ensemble credit scoring model that able to manage missing information, unbalanced data and non ii-d data points. It is brought several studies which used data mining techniques in credit scoring in Table 1.

- Neural network: Artificial neural networks (ANNs) are non-linear statistical modeling based on the function of the human brain. They are powerful tools for unknown data relationship modeling. ANNs able to recognize the complex pattern between input and output variables then predict the outcome of new independent input data.
- Bayesian classifier: A Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions and is particularly suited when the dimensionality of the inputs is high. A naive Bayes classifier assumes that the existence (or nonexistence) of a specific feature of a class is unrelated to the existence (or nonexistence) of any other feature. The major disadvantage of this model is that the predictive accuracy is highly correlated with this assumption. An advantage of this method is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.
- Discriminant analysis: Discriminant analysis which was studied by Fisher as early as 1936 is an alternative to logistic regression that assumes the explanatory variables follow a multivariate normal distribution and have a common variance-covariance matrix. This method is used to classifying observations in two classes. The objective of this method is to minimize the distance within each group and maximize the distance between different groups using discriminant function. The main drawbacks of this model are its unrealistic assumptions.
- Logistic regression: Logistic regression is a form of linear regression. This model can predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Generally, the dependent or response variable is dichotomous. .The relationship between the predictor and response variables is not a linear function in logistic regression. The advantages of this method are that the logistic regression does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables and the weakness of the model is that the independent variables be linearly related to the logit of the dependent variable.

- K-nearest neighbor: K-nearest neighbor is a nonparametric classifier based on learning by similarity. A training data set is collected, for this training data set, a distance function is introduced between the explanatory variable of observations. For each new observation this method explores the pattern space for the K nearest neighbors that are closest to the new observation in term of distance between the explanatory variables. The new observation is assigned to the class which its most KNN belong to that class.
- Decision tree: A classification tree is a tree-like graph of decisions and their possible consequences. Top-most node in this tree is the root node which a decision is supposed to take on it. In each inner node, it is done a test on an attribute or input variable. Each branch which follows the node lead to the result of the test, and the classes are represented by leaf nodes. Classification trees are used when the response variable is quantitative discrete or qualitative. CT is based on maximizing purity measure of the response variables of the observations. The advantage of this method is that it is a white box model and so it is simple to understand and explanation, but the limitation of this model is that, it can not be generalized a designed structure for one context to the other contexts.
- Survival analysis: This method is a new credit scoring model. The conventional method can distinguish good borrowers from bed ones at the time of loan application, but this model can compute the profitability of the customers over the customers' lifetime and perform profit scoring [25]. SA can predict the time until the event will occur instead of predicting the probability of occurrence an event.
- Fuzzy rule-based system: It helps to creditors in designing rules that accurately derive the credit score with explanation, while most of credits scoring models focus on estimating a score without explanation how the results obtained. The advantages of this model are that the fuzzy rules are capable of handling both quantitative and qualitative factors, so if there are a large set of inputs, scoring results will be less sensitive to small measurement errors.
- Support vector machine: Support vector machine is a classifier technique, first proposed by [26]. This method involves three elements. A score formula which is a linear combination of features selected for the classification problem, an objective function which considers both training and test samples to optimize the classification of new data, an optimizing algorithm for determining the optimal parameters of training sample objective function. The advantages of the method are that, in the nonparametric case, SVM requires no data structure assumptions such as normal distribution and continuity. SVM can perform a nonlinear mapping from an original input space into a high dimensional feature space and this method is capable of handling both continuous and categorical predictions. The weaknesses of this method are that, it is difficult to interpret unless the features interpretable and standard formulations do not contain specification of business constraints [8].
- Hybrid models: These models are credit scoring models which are developed by integrating two or more existing models. The advantage of these models is that the creditor can benefit form the advantages of two or more models and also they can remove the weakness of a model by combining them with the other models, but this type of methods are difficult to formulate and implement than simple methods[8].

TABLE 1: Data mining approaches in credit scoring

		References
Data mining techniques	Neural network	[27], [28], [29], [30], [31], [32], [33] and[34]
	Bayesian classifier	[35], [36], [37], [38], [39], [40] and [41]
	Discriminant analysis	[5], [42], [43], [44], [45] and [46]
	Logistic regression	[47], [48], [49], [50], [51], [52], [53] and [54]
	K-nearest neighbor	[55], [56], [57], [58] and [59]
	Decision tree	[60], [61], [18], [62], [63], [64], [65] and [66]
	Survival analysis	[67], [68], [25], [69], [70], [71], [72], [73], [74], [75] and [76]
	Fuzzy rule-based system	[77], [78], [50], [79], [80], [81], [82], [83] and [84]
	Support vector machine	[85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99] and [100]
	Hybrid models	[101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111] and [112]

5. Conclusion

Credit scoring has become very important issue due to the recent growth of the credit industry, so the credit department of the bank faces the huge numbers of consumers' credit data to process, but it is impossible analyzing this huge amount of data both in economic and manpower terms. In this study we reviewed the papers which have applied data mining methods in credit risk evaluation problem. Ten data mining technique which were most used method in the credit risk evaluation context were extracted, and then we searched almost all papers which had focused on these ten methods form 2000 to 2010. It is concluded that the support vector machine has been widely applied in recent years. Since to improve the performance of this model, it is necessary a method for reduction the feature subset, many hybrid SVM based model are proposed. Moreover the hybrid models have been attended in the last decade because of its enjoying from advantages of two or more models. Many of these proposed models can only classify customers into two classes "good" or "bad" ones. Form the perspective of risk management, prediction the probability of default for each applicant, who apply for a credit, will be more meaningful than classifying them into the binary classes. For this reason we propose the models which have the ability of estimation the probability of default, and also are simple to interpret and understand. It is noticeable that if we can reach this goal, in fact we have could classify customers into "good" or "bad" classes according to their probability of default.

References

1. Chen, M. C., and Huang, S. H., 2003, "Credit scoring and rejected instances reassigning through evolutionary computation techniques." Expert Systems with Applications, 24(4), 433–441.
2. Fisher, R., 1936, "The use of multiple measurements in taxonomic problems." Annals of Eugenics, 7, 179–188.
3. Durand, D., 1941, Risk elements in consumer instalments financing.
4. Beaver, W., 1967, "Financial ratios as predictors of failures." Empirical Research in Accounting: selected, 38(1), 63-93.
5. Altman, E. I., 1968, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." Journal of Finance, 23(4), 589–609.
6. Lyn C, T., 2000, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers." International Journal of Forecasting 16 (2), 149–172.
7. Paleologo, G., Elisseeff, A., and Antonini, G., 2010, "Subagging for credit scoring models." Journal of Operational Research 201(2), 490–499.
8. Ravi, V., 2008, advances in banking technology and management: impacts of ICT and CRM: information science reference.
9. Giudici, P., 2003, "Applied Data Mining: Statistical Methods for Business and Industry". John Wiley & Sons.
10. Giudici, P., 2001, "Bayesian data mining, with application to benchmarking and credit scoring." Applied Stochastic Models in Business and Society, 17, 69–81.
11. Zurada, J., and Lonial, S., 2005, "Comparison of the Performance of Several Data Mining Methods for Bad Debt Recovery in the Healthcare Industry." the Journal of Applied Business Research, 21(2), 37-53.
12. Chye Koh, H., Chin Tan, W., and Peng Goh, C., 2006, "A Two-step Method to Construct Credit Scoring Models with Data Mining Techniques." Journal of Business and Information, 1, 96-118.
13. Kirkos, E., Spathis, C., and Manolopoulos., Y., 2007, "Data Mining techniques for the detection of fraudulent financial statements." Expert Systems with Applications 32(4), 995-1003.
14. Atish P, S., and Huimin, Z., 2008, "Incorporating domain knowledge into data mining classifiers: An application in indirect lending." Decision Support Systems 46(1), 287–299.
15. Yeh, I. C., and Lien, C. h., 2009, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients." Expert Systems with Applications 36(2), 2473–2480.
16. Galindo, J., and Tamayo, P., 2000, "Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications." Basic Methodology and Risk Modeling Applications 15(1-2), 107–143.
17. Doumpos, M., Kosmidou, K., Baourakis, G., and Zopounidis, C., 2002, "Credit risk assessment using a multicriteria hierarchical discrimination approach: A comparative analysis." European Journal of Operational Research, 138(2), 392–412.
18. Xiao, W., Zhao, Q., and Fei, Q., 2006, "a comparative study of data mining methods in consumer loans credit scoring management." journal of systems science and systems engineering, 15(4), 419-435.

19. Huysmans, J., Baesens, B., Vanthienen, J., and Gestel, T. v., 2006, "Failure prediction with self organizing maps." Expert Systems with Applications 30(3), 479-487.
20. Satchidanand, S. S., and Jay, B. S., 2007, "Empirical evaluation of sampling and algorithm selection for predictive modeling for default risk." ISIS.
21. Ince, H., and Aktan, B., 2009, "A Comparison of Data mining Techniques for Credit Scoring in Banking: A managerial Perspective." Journal of Business Economics and Management, 10(3), 233–240.
22. Zhou, L., and Lai, K. K., 2009, "Benchmarking binary classification models on data sets with different degrees of imbalance." Frontiers of Computer Science in China, 3(2), 205-216.
23. Li, F. C., 2009a, "Comparison of the Primitive Classifiers without Features Selection in Credit Scoring"International Conference on Management and Service Science. City: IEEE.
24. Twala, B., 2010, "Multiple classifier application to credit risk assessment." Expert Systems with Applications 37(4), 3326–3336.
25. Baesens, B., Gestel, T. V., Stepanova, M., Poel, D. V. d., and Vanthienen, J., 2005, "Neural Network Survival Analysis for Personal Loan Data." The Journal of the Operational Research Society, 56(9), 1089-1098.
26. Vapnik, V., 1995, the Nature of Statistical Learning Theory, New York: Springer
27. West, D., 2000, "Neural network credit scoring models." Computers & Operations Research 27(11-12), 1131-1152.
28. Malhotra, R., and Malhotra, D. K., 2003, "Evaluating consumer loans using neural networks." Omega, 31(2), 83-96.
29. Hsieh, N. C., 2004, "An integrated data mining and behavioral scoring model for analyzing bank customers." Expert Systems with Applications 27(4), 623-633.
30. Angelini, E., Tollo, G. d., and Roli, A., 2008, "A neural network approach for credit risk evaluation." The Quarterly Review of Economics and Finance 48(4), 733–755.
31. Yu, L., Wang, S., and Lai, K. K., 2008a, "Credit risk assessment with a multistage neural network ensemble learning approach." Expert Systems with Applications 34(2), 1434-1444.
32. Abdou, H., Pointon, J., and El-Masry, A., 2008, "Neural nets versus conventional techniques in credit scoring in Egyptian banking." Expert Systems with Applications 35(3), 1275-1292.
33. Tsai, M. C., Lin, S. P., Cheng, C. C., and Lin, Y. P., 2009, "The consumer loan default predicting model – An application of DEA–DA and neural network." Expert Systems with Applications, 36(9), 11682–11690.
34. Khashman, A., 2010, "Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes." Expert Systems with Applications, 37(9), 6233-6239.
35. Baesens, B., Egmont-Petersen, M., Castelo, R., and Vanthienen, J., 2002, "Learning Bayesian network classifiers for credit scoring using Markov Chain Monte Carlo search"International Congress on Pattern Recognition. City: IEEE Computer Society pp. 49-52.
36. Li, X. S., and Guo, Y. H., 2006, "Personal credit scoring models on naive Bayesian classifier." Computer Engineering and Applications, 42(1), 197-201.
37. McNeil, A. J., and Wendin, J. P., 2007, "Bayesian inference for generalized linear mixed models of portfolio credit risk." Journal of Empirical Finance 14(2), 131–149.
38. Kadam, A., and Lenk, P., 2008, "Bayesian inference for issuer heterogeneity in credit ratings migration." Journal of Banking & Finance 32 (10), 2267–2274.
39. Panigrahi, S., Kundu, A., Sural, S., and Majumdar, A. K., 2009, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning." Information Fusion 10(4), 354-363.
40. Stefanescu, C., Tunaru, R., and Turnbull, S., 2009, "The credit rating process and estimation of transition probabilities: A Bayesian approach." Journal of Empirical Finance 16(2), 216–234.
41. Antonakis, A. C., and Sfakianakis, M. E., 2009, "Assessing naive Bayes as a method for screening credit applicants.". Journal of Applied Statistics, 36(5), 537-545.
42. Eisenbeis, R. A., 1977, "Pitfalls in the application of discriminant analysis in business, finance, and economics." The Journal of Finance, 32(3), 875-900.
43. Taffler, R. J., and Abassi, B., 1984, "A Model for Predicting Debt Servicing Problems in Developing Countries." Journal of the Royal Statistical Society, 147(4), 541-568.
44. Yobas, M. B., Crook, J. N., and Ross, P., 2000, "Credit scoring using neural and evolutionary techniques." IMA Journal of Management Mathematics 11(2), 111-125.
45. Kumar, K., and Bhattacharya, S., 2006, "Artificial neural network vs linear discriminant analysis in credit ratings forecast: A comparative study of prediction performances." Review of Accounting and Finance, 5(3), 216-227.

46. Abdou, H. A., 2009, "An evaluation of alternative scoring models in private banking." The Journal of Risk Finance, 10(1), 38-53.
47. Steenackers, A., and Goovaerts, M. J., 1989, "A credit scoring model for personal loans." Insurance: Mathematics and Economics, 8(1), 31-34.
48. Laitinen, E. K., 1999, "Predicting a corporate credit analyst's risk estimate by logistic and linear models." International Review of Financial Analysis 8(2), 97-121.
49. Alfo, M., Caiazza, S., and Trovato, G., 2005, "Extending a Logistic Approach to Risk Modeling through Semiparametric Mixing." Journal of Financial Services Research, 28(1), 163-176.
50. Tang, T. C., and Chi, L. C., 2005, "Predicting multilateral trade credit risks: comparisons of Logit and Fuzzy Logic models using ROC curve analysis." Expert Systems with Applications, 28(3), 547–556.
51. Ma, R. w., and Tang, C. y., 2007, "Building up Default Predicting Model based on Logistic Model and Misclassification Loss." Systems Engineering - Theory & Practice, 27(8).
52. Sohn, S. Y., and Kim, H. S., 2007, "Random effects logistic regression model for default prediction of technology credit guarantee fund." European Journal of Operational Research, 183(1), 427-478.
53. Luo, J. h., and Lei, H. y., 2008, "Empirical study of corporation credit default probability based on Logit model"Wireless Communications, Networking and Mobile Computing., IEEE.
54. Liang, Y., and Xin, H., 2009, "Application of Discretization in the Use of Logistic Financial Rating"International Conference on Business Intelligence and Financial Engineering, IEEE.
55. Paredes, R., and Vidal, E., 2000, "A class-dependent weighted dissimilarity measure for nearest neighbor classification problems." Pattern Recognition Letters 21(12), 1027-1036.
56. Hand, D. J., and Vinciotti, V., 2003, "Choosing k for two-class nearest neighbor classifiers with unbalanced classes." Pattern Recognition Letters 24(9-10), 1555-1562.
57. Islam, M. J., Wu, Q. M. J., Ahmadi, M., and Sid-Ahmed, M. A., 2007, "Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers"International Conference on Convergence Information Technology. IEEE Computer Society.
58. Marinakis, Y., Marinaki, M., Doumpos, M., and Matsatsinis, N., 2008, "Constantin Zopounidis, Optimization of nearest neighbor classifiers via metaheuristic algorithms for credit risk assessment." Journal of Global Optimization 42(2), 279-293.
59. Li, F. C., 2009b, "The Hybrid Credit Scoring Model based on KNN Classifier"Sixth International Conference on Fuzzy Systems and Knowledge Discovery. IEEE Computer Society.
60. Mues, C., Baesens, B., Files, C. M., and Vanthienen, J., 2004, "Decision diagrams in machine learning: an empirical study on real-life credit-risk data." Expert Systems with Applications 27(2), 257-264.
61. Lee, T. S., Chiu, C. C., Chou, Y. C., and Lu, C. J., 2006, "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines." Computational Statistics & Data Analysis 50(4), 1113-1130.
62. Zhao, H., 2007, "A multi-objective genetic programming approach to developing Pareto optimal decision trees." Decision Support Systems 43(3), 809-826.
63. Lopez, R. F., 2007, "Modeling of insurers' rating determinants. An application of machine learning techniques and statistical models." European Journal of Operational Research 183(2), 1488-1512.
64. Yeh, H. C., Yang, M. L., and Lee, L. C., 2007, "An Empirical Study of Credit Scoring Model for Credit Card"International Conference on Innovative Computing, Information and Control City, pp. 216.
65. Bastos, J. A., 2008, "Credit scoring with boosted decision trees". City: Munich Personal RePEc Archive, pp. 262-273.
66. Li, H., Sun, J., and Wu, J., 2010, "Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods." Expert Systems with Applications, 37(8), 5895-5904.
67. Thomas, L. C., Ho, J., and Scherer, W. T., 2001, "Time will tell: Behavioural scoring and the dynamics of consumer credit assessment." IMA Journal of Management Mathematics 12(1), 89–103.
68. Stepanova, M., and Thomas, L. C., 2002, "survival Analysis Methods for Personal Loan Data." Operations Research 50(2), 277–289.
69. Noh, H. J., Roh, T. H., and Han, I., 2005,. "Prognostic personal credit risk model considering censored information." Expert Systems with Applications 28(4), 753–762.
70. Sohn, S. Y., and Shin, H. W., 2006, "Reject inference in credit operations based on survival analysis." Expert Systems with Applications 31(1), 26–29.
71. Carling, K., Jacobson, T., Linde, J., and Roszbach, K., 2007, "Corporate credit risk modeling and the macroeconomy." Journal of Banking & Finance, 31(3), 845–868.

72. Beran, J., and Djaidja, A. Y. K., 2007, "Credit risk modeling based on survival analysis with immunes." Statistical Methodology, 4(3), 251–276.
73. Andreeva, G., Ansell, J., and Crook, J., 2007, "Modelling profitability using survival combination scores." European Journal of Operational Research 183(3), 1537–1549.
74. Bellotti, T., and Crook, J., 2009a, "Credit scoring with Macroeconomic Variables using Survival Analysis." Journal of the Operational Research Society, 60(3), 1699-1707.
75. Cao, R., Vilar, J. M., and Devia, A., 2009, "Modelling consumer credit risk via survival analysis." Statistics & Operations Research Transactions, 33(1), 3-30.
76. Sarlija, N., Bensic, M., and Susac, M. Z., 2009, "Comparison procedure of predicting the time to default in behavioural scoring." Expert Systems with Applications 36(5), 8778–8788.
77. Baetge, J., and Heitmann, C., 2000, "creating a fuzzy rule-based indicator for the review of credit standing." Schmalenbach Business Review, 52(3), 318 – 343.
78. Tung, W. L., Quek, C., Cheng, P., and EWS, G., 2004, "a novel neural-fuzzy based early warning system for predicting bank failures." Neural Networks 17(4), 567–587.
79. Laha, A., 2007, "Building contextual classifiers by integrating fuzzy rule based classification technique and k-nn method for credit scoring." Advanced Engineering Informatics 21(3), 281–291.
80. Hoffmann, F., Baesens, B., Mues, C., Gestel, T. V., and Vanthienen, J., 2007, "Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms." European Journal of Operational Research 177(1), 540–555.
81. Jiao, Y., Syau, Y. R., and Lee, E. S., 2007, "Modelling credit rating by fuzzy adaptive network." Mathematical and Computer Modelling 45(5-6), 717–731.
82. Lahsasna, A., Ainon, R. N., and Wah, T. Y., 2008, "Credit Risk Evaluation Decision Modeling Through Optimized Fuzzy Classifier"International Symposium on information technology, ITSim. IEEE Computer Society.
83. Liu, K., Lai, K. K., and Guu, S. M., 2009, "Dynamic Credit Scoring on Consumer Behavior using Fuzzy Markov Model"Fourth International Multi-Conference on Computing in the Global Information Technology. IEEE Computer Society.
84. Xinhui, C., and Zhong, Q., 2009, "Consumer Credit Scoring Based on Multi-criteria Fuzzy Logic"International Conference on Business Intelligence and Financial Engineering. IEEE Computer Society: Beijing, China
85. Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., and Wu, S., 2004, "Credit rating analysis with support vector machines and neural networks: a market comparative study." Decision Support Systems 37(4), 543– 558.
86. Gestel, T. V., Bart Baesens, Dijcke, P. V., Garcia, J., Suykens, J. A. K., and Vanthienen, J., 2006b, "A process model to develop an internal rating system: Sovereign credit ratings." Decision Support Systems 42(2), 1131-1151.
87. Chen, W. H., and Shih, J. Y., 2006, "A study of Taiwan's issuer credit rating systems using support vector machines." Expert Systems with Applications 30(3), 427-435.
88. Gestel, T. V., Baesens, B., Suykens, J. A. K., Poel, D. V. d., Baestaens, D. E., and Wil, M., 2006a, "Bayesian kernel based classification for financial distress detection,." European Journal of Operational Research 172(3), 979-1003.
89. Yang, Y., 2007, "Adaptive credit scoring with kernel learning methods." European Journal of Operational Research 183(3), 1521-1536.
90. Martens, D., Baesens, B., Gestel, T. V., and Vanthienen, J., 2007, "Comprehensible credit scoring models using rule extraction from support vector machines." European Journal of Operational Research 183(3), 1466-1476.
91. Huang, C. L., Chen, M. C., and Wang, C. J., 2007, "Credit scoring with a data mining approach based on support vector machines." Expert Systems with Applications 33(4), 847–856.
92. Xu, X., Zhou, C., and Wang, Z., 2009, "Credit scoring algorithm based on link analysis ranking with support vector machine." Expert Systems with Applications 36(2), 2625-2632.
93. Zhou, L., Lai, K. K., and Yu, L., 2009, "Credit scoring using support vector machines with direct search for parameters selection." Soft Computing 13(2), 149-155.
94. Chen, W., Ma, C., and Ma, L., 2009, "Mining the customer credit using hybrid support vector machine technique." Expert Systems with Applications 36(4), 7611-7616.
95. Luo, S. T., Cheng, B. W., and Hsieh, C. H., 2009, "Prediction model building with clustering-launched classification and support vector machines in credit scoring." Expert Systems with Applications 36(4), 7562–7566.

96. Bellotti, T., and Crook, J., 2009b, "Support vector machines for credit scoring and discovery of significant features." Expert Systems with Applications 36(2), 3302–3308.
97. HÄRDLE, W., Lee, Y. J., Schafer, D., and Yeh, Y. R., 2009, "Variable Selection and Oversampling in the Use of Smooth Support Vector Machines for Predicting the Default Risk of Companies." Journal of Forecasting 28(2), 512–534
98. Zhou, L., Lai, K. K., and Yu, L., 2010, "Least squares support vector machines ensemble models for credit scoring." Expert Systems with Applications 37(1), 127–133.
99. Yu, L., Yue, W., Wang, S., and Lai, K. K., 2010, "Support vector machine based multiagent ensemble learning for credit risk evaluation." Expert Systems with Applications 37(2), 1351–1360.
100. Kim, H. S., and Sohn, S. Y., 2010, "support vector machines for default prediction of SMEs based on technology credit." European Journal of Operational Research 201(3), 838–846.
101. Lee, T. S., Chiu, C. C., Lu, C. J., and Chen, I. F., 2002, "Credit scoring using the hybrid neural discriminant technique." Expert Systems with Applications 23(3), 245–254.
102. Malhotra, R., and Malhotra, D.K. ., 2002, "differentiating between good credits and bad credits using neuro-fuzzy systems." European Journal of Operational Research, 136(1), 190-211.
103. Wang, Y., Wang, S., and Lai, K. K., 2005, "A New Fuzzy Support Vector Machine to Evaluate Credit Risk"TRANSACTIONS ON FUZZY SYSTEMS. IEEE Computer Society.
104. Lee, T. S., and Chen, I. F., 2005, "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines." Expert Systems with Applications 28(4), 743–752.
105. Hsieh, N. C., 2005, "Hybrid mining approach in the design of credit scoring models." Expert Systems with Applications 28(4), 655–665.
106. Huang, J. J., Tzeng, G. H., and Ong, C. S., 2006, "Two-stage genetic programming (2SGP) for the credit scoring model." Applied Mathematics and Computation 174(2), 1039–1053.
107. Zhou, J., and Bai, T., 2008, "Credit Risk Assessment using Rough Set Theory and GA-based SVM"The 3rd International Conference on Grid and Pervasive Computing – Workshops. IEEE Computer Society: Kunming pp. 320-325.
108. Zhang, D., Hifi, M., Chen, Q., and Ye, W., 2008, "A Hybrid Credit Scoring Model Based on Genetic Programming and Support Vector Machines"Fourth International Conference on Natural Computation. IEEE Computer Society, pp. 8-12.
109. Yu, L., Wang, S., Wen, F., Lai, K. K., and He, S., 2008b, "Designing A Hybrid Intelligent Mining System for Credit Risk Evaluation." Journal of Systems Science and Complexity 21(4), 527-539.
110. Lin, S. L., 2009, "A new two-stage hybrid approach of credit risk in banking industry." Expert Systems with Applications 36(4), 8333–8341.
111. Chen, F. L., and Li, F. C., 2010, "Combination of feature selection approaches with SVM in credit scoring"Expert Systems with Applications.
112. Hsieh, N. C., and Hung, L. P., 2010, "A data driven ensemble classifier for credit scoring analysis." Expert Systems with Applications, 37(1), 534–545.