

Model Selection for Financial Distress Prediction by Aggregating TOPSIS and PROMETHEE Rankings

V. García¹, A.I. Marqués², L. Cleofas-Sánchez³, and J.S. Sánchez³

¹ División Multidisciplinaria de Ciudad Universitaria, Universidad Autónoma de Ciudad Juárez, 32310 Ciudad Juárez, Chihuahua (Mexico)

² Dept. Business Administration and Marketing

³ Institute of New Imaging Technologies, Dept. Computer Languages and Systems Universitat Jaume I, 12071 Castelló de la Plana, Spain

Abstract. Many models have been explored for financial distress prediction, but no consistent conclusions have been drawn on which method shows the best behavior when different performance evaluation measures are employed. Accordingly, this paper proposes the integration of the ranking scores given by two popular multiple-criteria decision-making tools as an important step to help decision makers in selecting the model(s) properly. Selection of the most appropriate prediction method is here shaped as a multiple-criteria decision-making problem that involves a number of performance measures (criteria) and a set of techniques (alternatives). An empirical study is carried out to assess the performance of ten algorithms over six real-life bankruptcy and credit risk databases. The results reveal that the use of a unique performance measure often leads to contradictory conclusions, while the multiple-criteria decision-making techniques may yield a more reliable analysis. Besides, these allow the decision makers to weight the relevance of the individual performance metrics as a function of each particular problem.

Keywords: Model selection, multi-criteria decision-making, financial distress, TOPSIS, PROMETHEE

1 Introduction

The lingering international financial crisis and the recommendations issued by the Basel Committee on Banking Supervision have led to increasing attention of financial institutions on credit and operational risk assessment, converting this into a critical task because of the heavy losses associated with wrong decisions. One major risk for financial institutions comes from the difficulty to distinguish the creditworthy applicants from those who will probably default on repayments.

The most traditional approaches to bankruptcy and credit risk management have been based upon subjective judgments made by human financial experts, using past experiences and some standard guiding principles, but the considerable increase in business demands and database sizes have brought about the

development of more formal and accurate methods to efficiently assess the financial risk. The adoption of statistical and operations research models (e.g., logistic regression, discriminant analysis, or linear and quadratic programming) represented a first step in that direction [18,23,30]. However, some assumptions of the statistical procedures, such as the multivariate normality for explanatory variables, are frequently violated in practice, thus making them theoretically invalid for finite samples [10]. During the last decade, efforts have focused on the deployment of data mining techniques such as artificial neural networks [3,5,14], support vector machines [4,20,28] and classifier ensembles [19,29,31,32], to design and implement solutions for financial risk prediction. In contrast with statistical models, data mining methods do not assume any specific prior knowledge, but automatically extract information from the examples available.

From a practical point of view, financial distress prediction can be defined as a binary classification problem where a new observation must be categorized into one of the two predefined classes based on a number of input variables. These collect a variety of information that describes socio-demographic characteristics and economic conditions of the applicants, and then the model has to produce the output in terms of their creditworthiness. In its most usual form, financial distress prediction aims at assigning customers to either good (those who are liable to reimburse the financial obligation) or bad (those debtors with high probability of defaulting on repayments). Assuming a set of m past observations $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$, where each sample x_i is characterized by D explanatory variables, $x_{i1}, x_{i2}, \dots, x_{iD}$, and y_i denotes the class (good/bad, defaulter/non-defaulter, positive/negative), then a model F has to predict the value y for a new case \mathbf{x} , that is, $F(\mathbf{x}) = y$.

Many comparative studies of a variety of financial risk prediction models have been conducted, but their conclusions may vary depending on the performance measure evaluated. For example, Desai et al. [7] concluded that customized neural networks perform better than linear models when measuring the percentage of bad applicants correctly classified, whereas logistic regression yields better results in terms of percentage of good and bad applicants correctly classified. Yobas et al. [33] found that linear discriminant analysis outperforms neural networks, genetic algorithms and decision trees in the proportion of samples correctly classified. Baesens et al. [4] showed that the support vector machines achieve the highest accuracy rate, while the neural networks perform the best in terms of the area under the ROC curve. Bensic et al. [5] suggested that the predictive accuracy of probabilistic neural networks is superior to that of logistic regression, CART decision trees, radial basis function, multi-layer perceptron and learning vector quantization. Antonakis and Sfakianakis [2] evaluated the performance of k -nearest neighbors decision rule, multi-layer perceptron, decision trees, logistic regression, linear discriminant analysis and naïve Bayes, showing that the k -nearest neighbors rule achieved the highest accuracy and the neural network was the best method in terms of the Gini coefficient. Wang [32] found that stacking and bagging using a decision tree as base classifier achieve the best performance in terms of accuracy, type-I error and type-II error.

Disagreement of these and many other related studies suggests that there exists no prediction algorithm that could be superior for any performance measure. However, model selection constitutes an issue of particular interest in financial risk management, which indicates the need of more powerful methods to evaluate the prediction performance of a set of classifiers. Bearing the limitations of individual performance measures in mind, the purpose of this paper is to introduce an aggregated multiple-criteria decision-making (MCDM) method for a more comprehensive assessment of financial distress prediction models. To this end, the TOPSIS and PROMETHEE algorithms are employed to rank a collection of classifiers based on a set of metrics and then their ranking scores are combined to provide a single scalar, demonstrating that the use of this technique allows for stronger conclusions regarding the performance of prediction models.

2 Multiple-criteria decision-making

Assessing the performance of prediction models means to take more than one criterion of interest into account, usually weighting this against the gains of other complementary criteria. Under this consideration, model selection can be viewed as an example of MCDM problems. Generally speaking, MCDM constitutes a branch of operations research that comprises several analytical tools to judge the pros and cons of a finite set of alternatives (or choices) based on a finite set of criteria (or attributes), with the aim of making a reliable decision [15,27].

Within the context of financial risk prediction, the MCDM methods should allow decision makers to choose the model that achieves a closely optimal trade-off of the assessment criteria of interest. A general MCDM problem can be expressed in the form of a $(M \times N)$ decision matrix as that given in Table 1, where M is the number of alternatives (prediction models) and N denotes the number of decision criteria (performance measures).

Table 1. Decision matrix for an MCDM problem (z_{ij} indicates the value of alternative A_i when evaluated in terms of criterion C_j)

	C_1	C_1	\dots	C_N
A_1	z_{11}	z_{12}	\dots	z_{1N}
A_2	z_{21}	z_{22}	\dots	z_{2N}
\vdots	\vdots	\vdots	\ddots	\vdots
A_M	z_{M1}	z_{M2}	\dots	z_{MN}

Popular examples of the many MCDM methods proposed in the literature are TOPSIS and PROMETHEE, which provide interesting advantages over other techniques [24]: (i) a solution that represents the rationale of human choice; (ii) a scalar value that accounts for both the best and worst alternatives; and (iii) a simple computation procedure that can be easily implemented into a spreadsheet.

2.1 TOPSIS

The basic principle behind TOPSIS is to find the best alternative by simultaneously minimizing the distance to the positive ideal solution and maximizing the distance to the negative ideal solution [11]. The positive ideal solution is formed as a composite of the best performance values exhibited by any alternative for each criterion, whereas the negative ideal solution is the composite of the worst performance values.

Assuming a problem with M alternatives and N criteria defined as in Table 1, the stepwise procedure of TOPSIS can be implemented as follows:

Algorithm 1 TOPSIS

- 1: Calculate the normalized decision matrix, where the normalized value n_{ij} of the original score z_{ij} is calculated as

$$n_{ij} = \frac{z_{ij}}{\sqrt{\sum_{i=1}^M z_{ij}^2}} \quad i = 1, \dots, M \quad j = 1, \dots, N$$

- 2: Calculate the weighted normalized values $v_{ij} = w_j z_{ij}$, where w_j is the weight of the criterion C_j and $\sum_{j=1}^N w_j = 1$

- 3: Determine the positive ideal A^+ and negative ideal A^- solutions

$$A^+ = \{v_1^+, \dots, v_N^+\} = \{(\max_j v_{ij} | i \in I), (\min_j v_{ij} | i \in J)\}$$

$$A^- = \{v_1^-, \dots, v_N^-\} = \{(\min_j v_{ij} | i \in I), (\max_j v_{ij} | i \in J)\}$$

where I and J are associated with benefit and cost criteria, respectively

- 4: Calculate the separation of each alternative from the positive ideal solution and that from the negative ideal solution using the N -dimensional Euclidean distance

$$d_j^+ = \sqrt{\sum_{j=1}^N (v_{ij} - v_j^+)^2} \quad \text{and} \quad d_j^- = \sqrt{\sum_{j=1}^N (v_{ij} - v_j^-)^2} \quad i = 1, \dots, M$$

- 5: Calculate the relative closeness to the ideal solution. The relative closeness of the alternative A_i with respect to A^+ is defined as $R_i^+ = \frac{d_i^-}{d_i^+ + d_i^-} \quad i = 1, \dots, M$

- 6: Rank alternatives according to the descending order of the index R_i^+
-

Note that the TOPSIS model provides a complete ranking of alternatives according to their global utilities and it does not require criterion preferences to be independent [11].

2.2 PROMETHEE

The objective of the PROMETHEE method [6] is to rank alternatives based on their values over different criteria. As an outranking technique, it quantifies a ranking through the pairwise comparisons (differences) between the criterion values describing the alternatives. It makes use of the concept of preference flow: the positive preference flow indicates how an alternative is outranking all the other alternatives, whereas the negative preference flow indicates how an alternative is outranked by the remaining alternatives. The procedure of PROMETHEE can be expressed in a series of steps as follows:

Algorithm 2 PROMETHEE

- 1: For each pair (A_i, A_j) of a finite set of alternatives $A = \{A_1, A_2, \dots, A_M\}$, calculate aggregated preference indices as follows:

$$\pi(A_i, A_j) = \sum_{k=1}^N P_k(A_i, A_j)w_k \quad \text{and} \quad \pi(A_j, A_i) = \sum_{k=1}^N P_k(A_j, A_i)w_k$$

where w_k is the normalized weight of the criterion C_k . $\pi(A_i, A_j)$ indicates how A_i is preferred to A_j and $\pi(A_j, A_i)$ indicates how A_j is preferred to A_i . $P_k(A_i, A_j)$ and $P_k(A_j, A_i)$ are the preference functions for alternatives A_i and A_j , respectively

- 2: Define the positive and the negative preference flows as

$$\phi^+(A_i) = \frac{1}{M-1} \sum_{a \in A} \pi(A_i, a) \quad \text{and} \quad \phi^-(A_i) = \frac{1}{M-1} \sum_{a \in A} \pi(a, A_i)$$

- 3: Compute the net preference flow for each alternative as $\phi(A_i) = \phi^+(A_i) - \phi^-(A_i)$
-

When $\phi(A_i) > 0$, A_i is more outranking all the alternatives on all evaluation criteria. Conversely, when $\phi(A_i) < 0$, A_i is more outranked. The alternative A_i with the maximum net preference flow $\phi(A_i)$ is deemed as the best.

3 Experimental methodology

This study evaluates the performance of a set of financial distress prediction models by means of TOPSIS and PROMETHEE, with the aim of demonstrating that the MCDM tools are superior to a unique metric for making better decisions about which prediction technique is the most appropriate for a particular financial problem. The TOPSIS and PROMETHEE models have been conducted through the Sanna software [12], while the classifiers here tested have been implemented using the WEKA toolkit [8] with the default parameter values: Bayesian network (BNet), multi-layer perceptron (MLP), radial basis function (RBF), support vector machine (SVM), naïve Bayes classifier (NBC), logistic regression (logR), nearest neighbor rule (1NN), RIPPER propositional rule learner, and C4.5 and CART decision trees.

3.1 Data sets

Experiments have been carried out on six real-life financial data sets. The Australian and German databases have been taken from the UCI Machine Learning Database Repository (<http://archive.ics.uci.edu/ml/>). The Australian data set consists of 690 instances of MasterCard applicants, 307 of which have been identified as creditworthy and 383 as non-creditworthy applicants; each example has 14 explanatory variables (6 numerical and 8 categorical). The German credit data set refers to a credit screening application in a German bank, containing observations on 24 numerical attributes for 1000 past applicants: 700 examples of creditworthy applicants and 300 cases whose credit should not be granted. The Iranian data set comes from a modification to a corporate client database of a small private bank in Iran [22]. It consists of 950 examples labeled

as good customers and 50 as bad customers, each one described by 27 input attributes. The Polish data set contains bankruptcy information of 120 companies recorded over a two-year period [21], with 112 bankruptcy accounts and 128 good status accounts. The Thomas database, which comes with the book by Thomas et al. [26], collects 12 input variables to describe the data of 1125 applicants for a credit product. Finally, the UCSD data set corresponds to a subset with samples randomly chosen from the database used in the 2007 Data Mining Contest organized by the University of California San Diego and Fair Isaac Corporation. This is a completely balanced data set with 2500 examples from each class and 38 variables.

3.2 Experimental set-up

As data are rather limited, the performance of the prediction models has been assessed with 5-fold cross-validation because this is quite reliable. Each original database has been divided into five blocks, using four subset for training and the remaining one for testing purposes. Ten repetitions have been run for each trial. Thus the results from classifying the test samples have been averaged across the 50 runs and then evaluated with some standard performance measures and analyzed with TOPSIS and PROMETHEE.

Typical performance evaluation scores in the field of financial distress include accuracy, Gini coefficient, Kolmogorov-Smirnov statistic, root mean squared error, area under the ROC curve, geometric mean of accuracies, or type-I and type-II errors [26,9,1], among others. For a two-class problem, as is the case of the databases here experimented, most of these metrics can be easily derived from a (2×2) confusion matrix where each entry contains the number of correct (true positive, true negative) or incorrect (false positive, false negative) predictions.

Many financial systems often use the accuracy to evaluate the performance of the prediction models, thus representing the proportion of the correctly predicted cases on a particular data set. However, empirical and theoretical evidences show that this measure is strongly biased with respect to data imbalance and proportions of correct and incorrect predictions. Because financial data are commonly imbalanced, the area under the ROC curve (AUC) has been suggested as an appropriate evaluator without regard to class distribution or misclassification costs [4,16]. The AUC for a binary problem can be roughly defined as the arithmetic average of the mean predictions for each class [25]:

$$AUC = \frac{\textit{sensitivity} + \textit{specificity}}{2} \quad (1)$$

where the sensitivity (or true positive rate, TPrate) measures the percentage of good applicants that have been predicted correctly, and the specificity (or true negative rate, TNrate) is the percentage of bad applicants predicted as bad.

Another measure often used in skewed domains is the geometric mean of accuracies, which intends to maximize the predictive accuracy on each class while keeping them balanced (the difference between TPrate and TNrate is small). It

punishes those models that produce large disparities in accuracy between classes. It is worth noting that the geometric mean of accuracies closely relates with the distance to perfect classification in the ROC space.

$$Gmean = \sqrt{sensitivity \cdot specificity} \quad (2)$$

The root mean squared error (RMSE) is another common performance measure used in general prediction problems. Let p_1, p_2, \dots, p_m and a_1, a_2, \dots, a_m be the predicted and actual outputs on the test samples, respectively. The RMSE allows to quantify the difference between the predictions and the true labels, measuring the deviation of the classification model from the target value [13].

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (p_i - a_i)^2} \quad (3)$$

4 Results

Figure 1 plots the average results of five performance measures achieved with each prediction model. The values of RMSE has not been included because the meaning of this metric is opposite to that of the others, that is, lower values of RMSE indicate a better performing model.

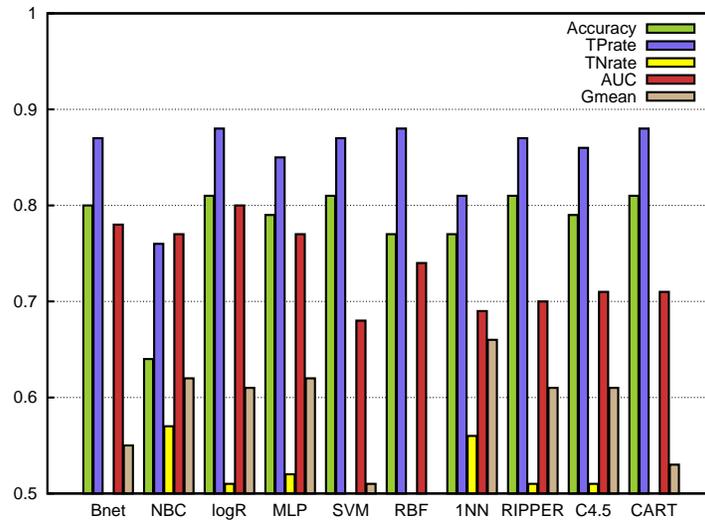


Fig. 1. Average values of five performance measures over different prediction models

Table 2 shows the results of six performance evaluation measures (accuracy, RMSE, TPrate, TNrate, AUC, and Gmean) averaged over the six experimental

Table 2. Performance results averaged over the six experimental databases

	Accuracy	RMSE	TPrate	TNrate	AUC	Gmean
BNet	0.80	0.39	0.87	0.48	0.78	0.55
NBC	0.64	0.54	0.76	0.57	0.77	0.62
logR	0.81	0.37	0.88	0.51	0.80	0.61
MLP	0.79	0.40	0.85	0.52	0.77	0.62
SVM	0.81	0.43	0.87	0.50	0.68	0.51
RBF	0.77	0.39	0.88	0.42	0.74	0.49
1NN	0.77	0.47	0.81	0.56	0.69	0.66
RIPPER	0.81	0.38	0.87	0.51	0.70	0.61
C4.5	0.79	0.40	0.86	0.51	0.71	0.61
CART	0.81	0.37	0.88	0.48	0.71	0.53
Weight	0.04762	0.23810	0.09524	0.14286	0.19048	0.28571

data sets. For each measure, the best performing model has been highlighted in bold. When analyzing the behavior by means of the predictive accuracy, logR, SVM, RIPPER and CART were the classifiers with the highest average rates. With the RMSE measure, logistic regression and CART appeared to be the best classifiers. Assessment of performance by means of the true positive rate suggested that logR, the RBF neural network and the CART decision tree were the best performing algorithm. The naïve Bayes classifier yielded the highest true negative rate. By using AUC, the best model corresponded to logistic regression, whereas the 1NN decision rule outperformed the remaining classifiers in terms of the geometric mean of accuracies.

From Figure 1 and Table 2, it seems obvious that there is no classification algorithm that achieves the best results across all measures and therefore, one might draw different conclusions about the best performing model depending on the performance evaluation measure used. For example, the TNrate indicates that the naïve Bayes classifier is the most suitable method, the AUC proposes the logistic regression model as the best algorithm, and the geometric mean of accuracies suggests that the 1NN rule is the most accurate technique. These contradictory outcomes describe a realistic situation in which two analysts might make radically different decisions depending on how they have evaluated the performance of a given financial risk prediction system. These are good examples of practical situations where we believe that MCDM methods should be taken into consideration in order to make more reliable, consistent decisions.

The problems associated with the use of individual performance evaluation measures led to experiment with the two MCDM techniques chosen for the present study. Although assigning weights to alternatives is nontrivial, here the weight of each performance measure used in TOPSIS and PROMETHEE methods were set according to its relative relevance for the financial risk prediction task. For example, AUC and Gmean are significant metrics in this practical domain because they select optimal models independently of the class distribution and the cost associated to each class [13]. On the other hand, the misclassi-

fication cost associated with false positives is generally much higher than the misclassification cost associated with false negatives [17]. The weights were then normalized in the interval $[0, 1]$ such that the sum of all weights was equal to 1 (see the last row in Table 2).

Table 3 summarizes the ranking of prediction models given by TOPSIS and PROMETHEE, respectively. The results are straightforward: the higher the ranking, the better the classifier. From the analysis with these two MCDM tools, the logistic regression model seemed to be the best performing algorithm, whereas the RIPPER rule learner and the MLP neural network were among the top-three ranked classifiers. These results indicate that TOPSIS and PROMETHEE, which gave similar rankings, can be useful to make accurate decisions in financial risk prediction problems.

Table 3. Preference ranking with TOPSIS and PROMETHEE

TOPSIS		PROMETHEE	
Rank Alternative	Score	Rank Alternative	Score
(4) BNet	0.78659	(4) BNet	0.06349
(10) NBC	0.20299	(5) NBC	0.00529
(1) logR	0.91444	(1) logR	0.61905
(3) MLP	0.80115	(2) MLP	0.29101
(8) SVM	0.59485	(10) SVM	-0.57672
(7) RBF	0.72369	(9) RBF	-0.31746
(9) 1NN	0.44631	(7) 1NN	-0.04762
(2) RIPPER	0.87712	(3) RIPPER	0.19577
(6) C4.5	0.76144	(8) C4.5	-0.18519
(5) CART	0.78476	(6) CART	-0.04762

Paradoxically, despite the claims of many authors, the SVM appeared to be one of the worst alternatives for financial distress prediction when TOPSIS and PROMETHEE were used for model selection. Such disagreement may be due to the use of inappropriate performance evaluation measures, while the MCDM tools are able to correct their misleading results.

Even though the ranks given by TOPSIS and PROMETHEE were quite similar, we further defined an aggregated ranking score as the mean of the values of the two methods for each prediction model (or alternative) i . This aggregated score allows to combine the rankings of TOPSIS and PROMETHEE in an easy, fair manner as follows:

$$Aggregated_Score_i = \frac{score_i(t) + score_i(p)}{2} \quad (4)$$

where $score_i(t)$ and $score_i(p)$ are the scores given by TOPSIS and PROMETHEE, respectively.

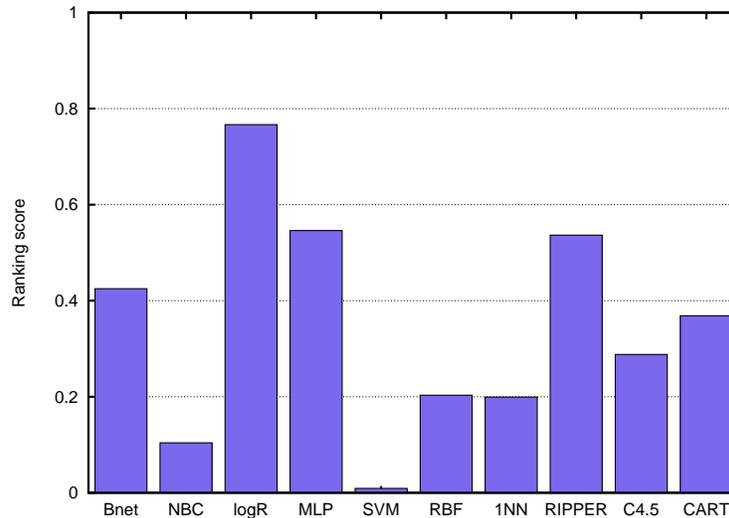


Fig. 2. Aggregated ranking scores

By plotting the aggregated score, Figure 2 provides a better representation of the decisions made, showing the superiority of the logistic regression model over the other methods and the poor performance of SVM.

5 Concluding remarks

This paper advocates the application of MCDM methods to evaluate the performance of financial risk prediction models. It has empirically been demonstrated that the use of single performance evaluation measures may lead to unreliable conclusions regarding the best performing algorithm, thus making difficult the selection of the most accurate model for a particular financial problem.

Two popular MCDM techniques, TOPSIS and PROMETHEE, have been tested in the experiments over six real-life financial data sets, using ten prediction models (alternatives) and six performance evaluation measures (criteria). The assessment of classifiers through single performance measures has given contradictory results, in the sense that different metrics have proposed different algorithms as the best alternative. This suggests that financial risk prediction is a real-world problem where MCDM tools should be applied to consistently evaluate a set of models and select the most appropriate one for each particular problem. Both TOPSIS and PROMETHEE have indicated that logistic regression, RIPPER and MLP are the best prediction models when the performance is evaluated with a composite of measures.

Besides, an aggregated score defined as a linear combination of the ranking values given by TOPSIS and PROMETHEE has also been introduced. The most interesting advantages of this new score are: (i) it transforms the individual

rankings of two MCDM methods into a single scalar which allows to make more consistent decisions; (ii) its graphical representation allows to clearly understand the decisions made by an analyst.

Acknowledgement

This work has partially been supported by the Spanish Ministry of Economy [TIN2013-46522-P], the Generalitat Valenciana [PROMETEOII/2014/062], the Mexican PRODEP [DSA/103.5/15/7004] and the Mexican Science and Technology Council through the Postdoctoral Fellowship Program [232167].

References

1. Abdou, H., Pointon, J.: Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance & Management* 18(2–3), 59–88 (2011)
2. Antonakis, A., Sfakianakis, M.E.: Assessing naïve Bayes as a method for screening credit applicants. *Journal of Applied Statistics* 36(5), 537–545 (2009)
3. Atiya, A.: Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Trans. on Neural Networks* 12(4), 929–935 (2001)
4. Baesens, B., Gestel, T.V., Viaene, S., Stepanova, M., Suykens, J., Vanthienen, J.: Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society* 54(6), 627–635 (2003)
5. Bensic, M., Sarlija, N., Zekic-Susac, M.: Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management* 13(3), 133–150 (2005)
6. Brans, J.P., Vincke, P.: A preference ranking organisation method: The PROMETHEE method for multiple criteria decision-making. *Management Science* 31(6), 647–656 (1985)
7. Desai, V., Crook, J., Overstreet, G.: A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research* 95(1), 24–37 (1996)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The WEKA data mining software: an update. *SIGKDD Explorations Newsletter* 11(1), 10–18 (2009)
9. Hand, D.: Good practice in retail credit scorecard assessment. *Journal of the Operational Research Society* 56(9), 1109–1117 (2005)
10. Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., Wu, S.: Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems* 37(4), 543–558 (2004)
11. Hwang, C.L., Yoon, K.: *Multiple Attribute Decision Making – Methods and Applications*. Springer-Verlag, New York (1981)
12. Jablonsky, J.: Software support for multiple criteria decision making problems. *Management Information Systems* 4(2), 29–34 (2009)
13. Japkowicz, N., Shah, M.: *Evaluating Learning Algorithms: A Classifier Perspective*. Cambridge University Press, New York, NY (2011)

14. du Jardin, P.: Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing* 73(10-12), 2047–2060 (2010)
15. Köksalan, M., Wallenius, J., Zionts, S.: *Multiple Criteria Decision Making: From Early History to the 21st Century*. World Scientific, Singapore (2011)
16. Lee, J.S., Zhu, D.: When costs are unequal and unknown: A subtree grafting approach for unbalanced data classification. *Decision Sciences* 42(4), 803–829 (2011)
17. Lee, T.S., Chen, I.F.: A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications* 28(4), 743–752 (2005)
18. Lee, T.S., Chiu, C.C., Chou, Y.C., Lu, C.J.: Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis* 50(4), 1113–1130 (2006)
19. Marqués, A., García, V., Sánchez, J.: Two-level classifier ensembles for credit risk assessment. *Expert Systems with Applications* 39(12), 10916–10922 (2012)
20. Min, J., Lee, Y.C.: Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications* 28(4), 603–614 (2005)
21. Pietruszkiewicz, W.: Dynamical systems and nonlinear Kalman filtering applied in classification. In: *Proceedings of the 7th IEEE International Conference on Cybernetic Intelligent Systems*. pp. 263–268. London, UK (2008)
22. Sabzevari, H., Soleymani, M., Noorbakhsh, E.: A comparison between statistical and data mining methods for credit scoring in case of limited available data. In: *Proceedings of the 3rd CRC Credit Scoring Conference*. Edinburgh, UK (2007)
23. Shi, Y., Peng, Y., Kou, G., Chen, Z.: Classifying credit card accounts for business intelligence and decision making: A multiple-criteria quadratic programming approach. *International Journal of Information Technology & Decision Making* 4(4), 581–599 (2005)
24. Shih, H.S., Shyur, H.J., Lee, E.: An extension of TOPSIS for group decision making. *Mathematical and Computer Modelling* 45(7–8), 801–813 (2007)
25. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45(4), 427–437 (2009)
26. Thomas, L., Edelman, D., Crook, J.: *Credit Scoring and Its Applications*. SIAM, Philadelphia, PA (2002)
27. Triantaphyllou, E.: Multi-criteria decision making methods. In: *Multi-Criteria Decision Making Methods: A Comparative Study*, Applied Optimization, vol. 44, pp. 5–21. Springer (2000)
28. Trustorff, J.H., Konrad, P., Leker, J.: Credit risk prediction using support vector machines. *Review of Quantitative Finance and Accounting* 36(4), 565–581 (2011)
29. Tsai, C.F., Wu, J.W.: Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications* 34(4), 2639–2649 (2008)
30. Tseng, F., Lin, L.: A quadratic interval logit model for forecasting bankruptcy. *Omega* 13(1), 85–91 (2005)
31. Twala, B.: Combining classifiers for credit risk prediction. *Journal of Systems Science and Systems Engineering* 18(3), 292–311 (2009)
32. Wang, G., Hao, J., Ma, J., Jiang, H.: A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications* 38(1), 223–230 (2011)
33. Yobas, M., Crook, J., Ross, P.: Credit scoring using neural and evolutionary techniques. *IMA Journal of Mathematics Applied in Business and Industry* 11(4), 111–125 (2000)