Information Sciences 254 (2014) 98–110

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Going-concern prediction using hybrid random forests and rough set approach

Ching-Chiang Yeh [a,*], Der-Jang Chi [b], Yi-Rong Lin [b]

[a] Department of Business Administration, National Taipei College of Business, No. 321, Sec. 1, Ji-Nan Rd., Zhongzheng District, Taipei 10051, Taiwan, ROC
[b] Department of Accounting, Chinese Culture University, No. 55, Hwa-Kang Road, Yang-Ming-Shan, Taipei City 11114, Taiwan, ROC

ARTICLE INFO

Article history:
Received 7 June 2012
Received in revised form 7 May 2013
Accepted 26 July 2013
Available online 6 August 2013

Keywords:
Going-concern prediction
Intellectual capital
Random forest
Rough set theory

ABSTRACT

Corporate going-concern opinions are not only useful in predicting bankruptcy but also provide some explanatory power in predicting bankruptcy resolution. The prediction of a firm's ability to remain a going concern is an important and challenging issue that has served as the impetus for many academic studies over the last few decades. Although intellectual capital (IC) is generally acknowledged as the key factor contributing to a corporation's ability to remain a going concern, it has not been considered in early prediction models. The objective of this study is to increase the accuracy of going-concern prediction by using a hybrid random forest (RF) and rough set theory (RST) approach, while adopting IC as a predictive variable. The results show that this proposed hybrid approach has the best classification rate and the lowest occurrence of Types I and II errors, and that IC is indeed valuable for going-concern prediction.

1. Introduction

The going-concern principle is one of the most important accounting assumptions in the preparation of financial statements. According to this principle, an entity or organization will continue its operations into the foreseeable future, at least, or in perpetuity. A going-concern opinion implies that the entity is not at risk of liquidation or even of reducing the scale of its operations substantially, whether voluntarily or involuntarily. Although auditors are not responsible for predicting bankruptcy or future events, according to Chen and Church [15], "going-concern opinions are useful in predicting bankruptcy and provide some explanatory power in predicting bankruptcy resolution." Thus, going-concern prediction has been the focus of rigorous research efforts for decades. In particular, several researchers have suggested prediction models to aid auditors in conducting going-concern assessments of firms.

Prior studies on going-concern prediction are based primarily on conventional statistical techniques [22,34] such as univariate statistical methods, multiple discriminant analysis (MDA), and logit and probit analyses. These conventional statistical methods, however, have some restrictive assumptions such as the linearity, normality, and independence of predictor or input variables. Considering that the violation of these assumptions occurs frequently within financial data [18], the methods have intrinsic limitations in terms of effectiveness and validity. Recently, Bellovary et al. [6] had undertaken an extensive review of going-concern prediction.

Artificial intelligence (AI) approaches such as inductive learning are less vulnerable to these violations. Moreover, AI aims to identify valid, novel, potentially useful, and understandable correlations and patterns in data [6]. AI can be an alternative solution to classification problems, given that data mining has shown to have better predictive capability than conventional

* Corresponding author. Tel.: +886 2 2322 6161; fax: +886 2 2322 6323.
E-mail addresses: ychinc@webmail.ntcb.edu.tw (C.-C. Yeh), qq1011108@gmail.com (D.-J. Chi), linyirong823@gmail.com (Y.-R. Lin).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.07.011

statistical methods of going-concern prediction [2,23,26,35,37,40]. Although the above-mentioned AI techniques have generally been shown to be effective in going-concern prediction, they are not without limitations.

First, since McKee [42] used financial ratios for going-concern prediction, most studies have primarily used financial ratios as independent variables. Financial ratios, originating in the financial statements of firms, can reflect some characteristics of a corporation. However, in the current knowledge era, the core competences of firms are derived from the knowledge and skills of their employees, and the value of intellectual capital (IC) now exceeds that of some tangible assets. Moreover, numerous researchers have recognized that the nature of a firm's IC plays an important role in its risk of bankruptcy [57,59]. Despite the growing importance of IC for firms in the knowledge era, it is usually excluded from early prediction models. Therefore, in this study, we believe that IC, which reflects the status of the corporation in going-concern predictions, will be a decisive factor that influences the predictive capability. Second, the studies mentioned above show that different researchers have used different independent variables as inputs for going-concern prediction. However, a few of the researchers used independent variables as a module of their going-concern prediction. These researchers did not pay much attention to finding and selecting important independent variables. Moreover, few studies employed these variables to generate the appropriate rules for going-concern decisions.

Recently, rough set theory (RST) [48–51] is a relatively new approach in AI that has been extensively used for knowledge reasoning and knowledge acquisition. Using the concepts of lower and upper approximations in rough sets, the knowledge hidden in the information systems can be unraveled and expressed in the form of "if ..., then ..." decision rules [1,47,50,51,60,62]. The extracted rules are easily interpretable, permitting complex relationships to be represented in an intuitive and comprehensible manner. The rules establish a relationship between descriptions of objects based on attributes and their assignment to specific classes. Moreover, the rules can be used for the classification of new objects [36]. Recently, it has found its application in a wide variety of fields including credit rating [17,64], business failure [7,61], knowledge acquisition [60], market decision-making [48], and early warning [7,54], etc. Therefore, we attempt to investigate the effectiveness of RST approach in conducting the going-concern prediction tasks and to predict the characteristics of going-concern so decision-makers can understand the rules of going-concern.

Moreover, to determine and select important independent variables in the development of a going-concern prediction model, the random forest (RF) method is used in this study. RF is a relatively newer ensemble method that combines trees grown on bootstrap samples of data and a random subset bagging of predictor variables [10]. During the randomization of features, RF can provide an importance index of independent variables by calculating accuracy and the Gini index. Furthermore, the importance index captures the interactions among predictors through the randomizations of predictors [40]. In terms of robustness to outliers and noise, and calculation time, RF is superior to other machine learning methods such as bagging or boosting.

In order to make great use of the advantages of RF in preprocessing the business data, and further improve the classification accuracy of the RST predictor model, RF + RST is proposed to predict the going-concern in this work. Moreover, we examine whether an assessment of a corporation's IC conveys any useful additional predictors in going-concern prediction. First, we use IC as a predictive (independent) variable. Second, RF is used to conduct variable selection because of its reliability in obtaining the significant independent variables. Third, the obtained significant independent variables from RF are used as inputs for the RST model. Fourth, we generate meaningful rules using RST for going-concern prediction. Fifth, to validate the effectiveness of our model, comparative experiments are conducted. Finally, to examine the effect of IC, we also compare the obtained results to see whether the model including IC gives better classification accuracy.

The remainder of this paper is organized as follows: In Section 2, we conduct a literature review about going-concern prediction and IC. In Section 3, we present the methodologies used in previous research, which are relevant to our paper for RF and RST. In Section 4, we describe the experimental design of this study. In Section 5, we summarize and discuss the empirical results. Finally, in Section 6, we present the conclusions of this study and discuss the future research directions.

2. Literature review

2.1. Going-concern

Going-concern prediction as a concept remains one of the most controversial areas of the auditing profession and has received much criticism since its origin in the eighteenth century. Lenard et al. [37] state that the auditor must provide the annual audit report on the financial condition of a company, which is consolidated with the company's financial statements. One of the important things that an audit report should address is the likelihood of the survival of the company (remain a going-concern). A modified audit report on going-concerns indicates that an auditor's evaluation can show whether the survival of the company is under threat or not.

There is substantial literature on going-concern prediction. We categorized the methods extensively used in prior research into statistical methods and AI methods.

2.2. Statistical methods

In light of the difficulty of going-concern assessment, numerous prior studies on going-concern prediction using publicly available information have employed statistical techniques such as univariate statistical methods, multivariate discriminant

analysis (MDA) [3,32,41,45], logit analysis [12,15,16,25,26,37,43,46] and probit analysis [5,22,34]. These studies concede that besides financial ratios, default status is also a significant variable that can explain auditors' choices [5,45]. The financial ratios typically selected included the ratio of current assets to current liabilities (current ratio) and the ratio of net income to total assets (return on assets). Moreover, after SAS 59 was released in 1988, there is a tendency for ascribing greater consideration to non-financial ratios. For example, the use of auditor reputation as an indicator of a company's ability to remain a going-concern, as used in Morris and Strawser [44] and Carcello and Neal [13], emphasizes the importance of audit committee independence besides the use of financial ratios in going-concern assessment.

The general conclusion from these efforts is that going-concern prediction using conventional statistical methods involves some restrictive assumptions such as linearity, normality, and independence among predictor or input variables. Considering that these assumptions for independent variables are frequently violated with financial data [18], the methods may be limited in terms of obtaining effectiveness and validity.

2.3. Artificial intelligence methods

Recently, AI techniques such as NN [2,24,26,37], support vector machines [41], DT [33,35,41], *k*-nearest neighbors [26], and AntMiner+ [41] have also been applied in this context. These techniques are considered (at least) supplementary to the traditional statistical going-concern prediction models.

In summary, the previous literature has made extensive efforts to apply NN to the going-concern prediction problem; many researchers have compared NN with other statistical and machine learning methods. The general conclusion in most prior studies is that NN outperform conventional statistical methods and inductive learning methods. However, NN act as "black boxes" and require expertise for defining the network's topology.

2.4. Intellectual capital

Stewart [53] defined IC as intellectual material (knowledge, information, intellectual property, and experience) that can be exploited to generate wealth. It offers a quantitative perspective and is linked to the identification and measurement of existing intangible assets created within the firm. IC may be conceptualized as comprising three distinct subcategories: human capital, organizational (structural) capital, and relational capital [9,31].

Human capital (HC) is defined as the qualities of employees, such as their knowledge and skills. Such capabilities become the source of difference within organizations [65]. Organizational capital is the institutionalized knowledge and codified experience residing within and utilized through databases, patents, manuals, structures, systems, and processes [65]. Relational capital is the value assigned by the organization to its relationships with public administrations, the mass media (corporate image), its corporate reputation, and social relationships.

According to De Saa-Perez and Garcıa-Falcon [19], HC is the knowledge and skills embodied in individuals. HC contributes to economic performance by increasing the productivity of physical capital; thus, it can be easily controlled through good strategic managerial practices. The availability of suitable HC that can provide guaranteed higher performance is very rare. It is important to note that the alignment between human beings with various types of skills and competencies and organizations is not uniform, which makes organizational HC very distinctive. Moreover, current studies are also examining the relationship between organizations' HC and their financial performance. It has been generally accepted that HC plays an important role in ensuring superior financial performance [65]. Nevertheless, thus far, no study has investigated the combined effects of HC on going-concern prediction. This study aims to conduct such an analysis.

2.5. Discussion

Regarding the above review, focus on designing more sophisticated classifiers although features of going-concern play an important role to affect the later prediction result. In particular, NN and DT are the most widely used model for going-concern prediction [6,41]. However, to the best of our knowledge, RST have been used extensively in different applications as classifiers, but have not been applied to the problem of going-concern so far.

With regard to Bellovary et al. [6], who provide a detailed review of AI techniques for going-concern-related problems, one important trend is to build a hybrid system. Moreover, the above-related works show that different researchers took different independent (predictive) variables as input for going-concern prediction. Based on these predictive variables, few of them took predictive variables as a module of their going-concern prediction. They also did not pay much attention to finding and selecting important predictive variables based on their importance, IC especially. Moreover, fewer studies applied these variables to extract useful information for going-concern decisions. Finally, most studies only examine average prediction performance of their models without considering the Type I and Type II errors.

Therefore, this paper proposes a novel hybrid model for going-concern prediction by integrating the RF and RST techniques, while adopting IC as a predictive variable. The RF method is used to conduct variable selection to obtain the significant independent variables, and RST can generate meaningful rules for going-concern. In order to evaluate the performance of the proposed framework, comparative experiments are conducted. Besides NN and DT, were also employed the state-of-the-art classification technique SVM [11,38,54,63] as the benchmarks and considering the Type I and Type II errors.

3. Methodology

3.1. Random forest

RF is another advanced method of machine learning. The classification is achieved by constructing an ensemble of randomized classification and regression trees (CART) [10]. The RF algorithm uses a combination of independent decision trees to model data and measure variable importance [10]. Each decision tree in a forest is constructed using a bootstrap sample from the data. Approximately one-third of the data instances are not used to grow the tree; these instances are termed the out-of bag (OOB) data for the tree. At each node of the tree, m variables out of all the n input variables are randomly selected, and each of the tree nodes is split using the selected m variables. The random selection of features at each node decreases the correlation between the trees in the forest. Thus, the RF algorithm can handle many redundant features and avoid model over-fitting. It has been shown that RF outperforms AdaBoost ensembles on noisy datasets, and can perform well on data with many weak input variables [10].

To evaluate the importance of variable x, its values in the OOB instances associated with each tree in the forest are permuted randomly. The permuted OOB instances and the original OOB instances are then classified using the tree. The number of correct classifications among the original OOB instances is subtracted from the number of predictions of the correct class among the permuted OOB instances to calculate a raw score based on the tree. The importance score of variable x is defined as the average of the raw scores over all the trees in the forest. For a fixed number of trees in the forest, the larger the importance score of a variable, the greater its importance for classification. In addition, a z-score can be obtained by dividing the importance score of the variable by its standard error, and a statistical significance level may be assigned to thez-score assuming normality [10].

In RF, cross validation or a test set is not required to obtain an unbiased estimate of the test error [66,67], because RF provides *de facto* test set predictions. Each tree in a forest is based on a bootstrap sample of the data that, on average, includes approximately two-thirds of the data points [23]. This means that approximately one-third of the data points will not be used in the training set for any given tree; conversely, any given point is omitted from the training set of approximately one-third of the trees. This subset of trees, none of which has been trained using the point, can predict a value for that point as though it were in a test set. Thus, RF automatically provides a form of cross-validation [58].

RF has several advantages over other statistical modeling methods [39]. Its variables can be both continuous and categorical. Because a large number of trees are induced and averaged during the run, RF produces low-bias and low-variation results, but highly accurate classification and good predictions. RF's OOB error estimates test the classifications by voting on a small number of samples, which further strengthens the model. This relative performance of this somewhat counterintuitive strategy is significantly better than many other classifiers, including discriminant analysis, support vector machines, and NN, and it is robust against over-fitting [39]. Moreover, RF has only two hyper-parameters (the number of variables in the random subset at each node and the number of trees in the forest), and it is usually not very sensitive to their values. The RF algorithm is becoming increasingly popular and is, apparently, very powerful in many different applications [20], although this has not been elucidated from a mathematical perspective [8].

3.2. Rough set theory

RST was developed by Pawlak [48] in the early 1980s. It was developed as a mathematical tool to deal with uncertain or vague knowledge and enables clear classificatory analysis of data tables. RST can be used to deal with quantitative and qualitative attributes simultaneously without requiring any *a priori* information about the probability distribution of the data. To use the RST process, one begins with a relational database, a table of objects with attributes, and the attribute values for each object. One attribute is selected as the decision attribute, making the rest of the attributes the conditional attributes [48]. RST addresses the problem of vagueness by applying the concept of equivalence classes to partition training instances according to specified criteria. Two partitions are formed in the mining process. The members of the partition can be formally described by unary set-theoretic operators or by successor functions for upper and lower approximation spaces, from which both possible and certain rules can be easily derived [50].

Suppose we are given an information system $S = (U, A)$, $X \subseteq U$, and $P \subseteq A$. The set X is approximated using information contained in P by constructing lower and upper approximation sets:

$$P_*(X) = \{x \in U : P(x) \subseteq X\} \text{ (Lower approximation)}$$

and

$$P^*(X) = \{x \in U : P(x) \cap X \neq \phi\} \text{ (Upper approximation)}.$$

The elements in $P_*(X)$ can be classified as members of X by the knowledge in P. However, the elements in $P^*(X)$ can only be classified as possible members of X by the knowledge in P. The set $PN_P(X) = P^*(X) - P_*(X)$ is termed the P-boundary region of X, and it comprises those objects that cannot be classified with certainty as members of X with the knowledge in P. The set X is termed "rough" with respect to the knowledge in P if the boundary region is non-empty. RST classifiers usually apply the

concept of RST to reduce the number of attributes in a decision table [49] and to extract valid data from inconsistent decision tables.

RST rule induction algorithms were implemented for the first time in a Learning from Examples based on Rough Sets (LERS) system [28]. A local covering is induced by exploring the search space of blocks of attribute-value pairs that are then converted into the rule set. The Learning from Examples Module, version 2 (LEM2) algorithm calculates the local covering for each approximation of the decision table concept [27] and subsequently converts these into decision rules. This algorithm has been successfully used in many problems.

4. Experimental design

4.1. Experimental process

In the study, we propose a hybrid model (RF + RST), which combines RF and RST to improve the accuracy rate of going-concern prediction. Moreover, we investigate whether IC is useful as an additional predictor in going-concern opinions. Initially, we incorporate IC as a potential predictive variable. Then, we use RF to perform variable selection because of its reliability in obtaining the relative importance of the predictive variables. Subsequently, we use the important predictive variables obtained from RF as input variables for RST. Moreover, the RST approach comprises the following two steps: (1) discretizing the attributes, and (2) extracting the decision rules using the LEM2 algorithm. Then, we generate meaningful rules using RST for going-concern prediction.

In order to evaluate the performance of proposed approach, we compared it with other three methods: Back-propagation neural network (BPN), C4.5 and SVM. BPN is a feedforward network and is probably the most commonly used class of NN in business applications [56]. C4.5 is the popular DT [52] builder is often treated as the benchmark for predicting classification problem, and has the advantages of dealing with missing data, continuous data, pruning, and generates rule from the DT which is built. SVMs are currently state-of-the-art for the classification task and generally speaking exhibit good predictive performance, due to its ability to capture non-linearities [11,38,54,63]. We use LIBSVM package [14] to construct the classifier. The data is scaled by scale subprogram. The kernel function radial basis function and cross-validation method are used; and the hyperparameters set by a grid-search subprogram.

Therefore, to evaluate the accuracy rate of these generated rules, we compare them with different classifiers-DT with the C4.5 algorithm [52] and NN with the BPN algorithm and use the pure RST, RF + DT, RF + NN, and RF + SVM models as the benchmarks. Finally, to test whether IC enables going-concern prediction, we analyze the IC before and after the going-concern prediction model is considered. Subsequently, the obtained results can be compared to determine whether the model including IC will provide better classification accuracy.

4.2. Data and samples

The initial dataset comprises companies listed in the Taiwan Economic Journal (TEJ) for a period of five years from 2004 to 2008. TEJ is an important source of data on the securities markets in Taiwan. After eliminating companies with multiple-year going-concern opinions, and financial institutions, insurance institutions, and companies with missing values, our final sample-selection procedures yielded a sample of 220 observations, including 55 companies with a first-time going-concern opinion and 165 clean audit opinions.

4.3. Potential predictive variables

To apply prediction methods to going-concern prediction, first, potential predictive variables should be selected. In this paper, we employ potential predictive variables that were used in prior research on going-concern prediction [2,4,15,29,30,33,35,41]. Initially, we prepare a set of 27 variables (comprising 18 financial ratios, 2 non-financial ratios, and 7 HC variables) and these variables are placed in decision classes using binary assignment (clean or non-clean audit opinions, coded as 1 or 0, respectively), as shown in Table 1. These variables were available in the TEJ database.

4.4. Performance evaluation

An auditor may make two kinds of errors [37]. To minimize Type I errors, a firm may over-audit, which is costly to the client. In addition, when the audit report is modified unnecessarily, it may be costly to the client firm if stockholders decide to sell their stock. As a result, the client may change its auditors. Type II errors may be a consequence of under-auditing, and in this case, the firm's financial statements are either prepared fraudulently or are not prepared in accordance with the generally accepted accounting principles (GAAP). One outcome of an unqualified opinion in this situation could be a lawsuit against the auditors.

Hence, prediction accuracy and Type I and II errors are considered when evaluating the performance of the developed going-concern prediction models. They can be measured using the confusion matrix shown in Table 2.

Then, prediction accuracy can be obtained as follows:

Table 1
Independent variables.

Categories	Variable	Definition
Financial ratios	F01	Current assets/current liabilities
	F02	Change in current assets/current liabilities
	F03	Current assets/total assets
	F04	Current assets/net income
	F05	Total assets (log or natural log)
	F06	Cash flow from operations/total liabilities
	F07	Long-term debt/total assets
	F08	Total liabilities/total assets
	F09	Change in total liabilities/total assets
	F 10	Net income before tax/net sales
	F 11	Net income before tax/total assets
	F 12	Change in net income/total assets
	F 13	Retained earnings/total assets
	F 14	Net worth divided/total liabilities
	F 15	Atlman's Z-score
	F 16	1 if negative net income, 0 otherwise
	F 17	1 if negative operating income, 0 otherwise
	F 18	1 if negative operating income in two years, 0 otherwise
Non-financial ratios	G01	1 if a big 4 auditor performs the audit, 0 otherwise
	G02	1 if auditor changes in this year, 0 otherwise
Intellectual capital	I01	Education/qualifications of workforce
	I02	Age of workforce
	I03	Net sale/workforce
	I04	Net income/workforce
	I05	Net income before tax/workforce
	I06	R&D expense/net sale
	I07	R&D expense/total assets

Table 2
Confusion matrix.

Actual	Predicted	
	GC modified opinion	Not GC modified opinion
GC modified opinion	(a)	II (b)
Not GC modified opinion GC, going concern	I (c)	(d)

$$\text{Prediction accuracy} = \frac{a+d}{a+b+c+d}$$

Type I error rate indicates the model's rate of prediction errors that incorrectly classify firms as having a "clean" unqualified opinion. Type II error rate, on the other hand, presents the model's rate of prediction errors that incorrectly classify firms belonging to the unqualified opinion group as a part of the qualified opinion group. Undoubtedly, an effective going-concern prediction model is one that has a lower incidence of Type II errors.

5. Results and discussion

5.1. Importance ranking of variables

To determine the relative importance of the potential predictive variables, we employed the RF available in the R package random forest [39] to calculate the value of variable importance. This implementation is based on the original Fortran code authored by Leo Breiman, the inventor of RF.

Following Liaw and Wiener [39], our preliminary tests indicated that the performance of RF barely depends on the actual value of its hyper-parameters within a large interval, which is consistent with the results of some previous studies [10]. To speed up the training, we considered that the forest comprises 13,000 trees. At each split, we randomly selected and considered five features, which is approximately the square root of the total 27 features employed, for splitting. Fig. 1 illustrates the convergence of the RF algorithm: it depicts the evolution of the OOB error as a function of the number of trees used. Fig. 1 shows the effect of the number of trees used when five features are selected at each split. When the number of trees is

approximately 10,300, the error rate is almost 0.102. Both plots indicate that the results are not sensitive to the selection of these two parameters. Therefore, we decided to use 10,300 trees. The black lines present the actual error rate, and the green and red lines are the upper and lower confidence bounds, respectively.

Moreover, to estimate the importance of variable m, the number of votes for the correct class is counted using the OOB cases in every tree. Subsequently, the number of correct votes is counted again after randomly permuting the values of variable m in the OOB cases. The average of the margin between these two numbers over all the trees in the forest is the raw importance score for variable m. The raw score is divided by its standard error to obtain a z-score, and the value of variable importance is the negative z-score for variable m. Fig. 2 plots the values of variable importance for all variables, sorted in descending order. The figure shows that the importance values of the last 12 variables (variables I06, F07, F17, F02, F06, F18, I03, I02, I01, F12, F04, and G02) are much lower than those of the other variables. Therefore, these variables were removed, and the rest of the 15 variables (including 12 financial ratios, 1 non-financial ratio, and 2 HC variables) were selected as the potential predictor variables were taken as input variables for the RST, DT, NN, and SVM classifier.

5.2. Comparisons of different methods

After the significant independent variables were determined, we implemented the RST classifiers. The 15 most important variables were selected as conditional attributes for classifying going-concern opinions, and they were discretized. Discretization can convert continuous attributes into discretized ones. There is no general way to define the optimal boundary values and expert opinions based on experience and knowledge are the best tools for identifying a set of decision problems [21]. Therefore, we asked expert auditors to discretize these conditional attributes, providing norms according to their professional knowledge and experience. The intervals were determined by experts based on the average and standard deviations from the initial data. The interval peer assessment was conducted in three rounds during the RST mining processes.

Moreover, we employed the RST-based application Rough Set Exploration System (RSES) system. This system (currently in version 2.2) was created by the so-called Group of Logic at Warsaw University (http://logic.mimuw.edu.pl/) under the supervision of Andrzej Skowron. For verification, comparison, and enhancement of the accuracy rate, we set the following pre-conditions: (1) Let coverage = 0.6 and 0.9, and then run the experiment (coverage refers to the parameter value from the LEM2 algorithm; the default value of coverage is 0.9). (2) The dataset was split into two sub-datasets with the 67% dataset used as a training set and the other 33% used as a testing set. (3) The experiment with the 67% and 33% random split was repeated 10 times, and subsequently, the average accuracy rate was computed. Using these 15 attributes in addition to a class enabled us to build a decision table to generate rules using RST (the LEM2 algorithm) to extract classification rules for going-concern opinions. The results of the experiment are shown in Table 3.

For further analysis of the above methods, t-test is used to analyze significant differences among the above methods (Table 3). The t-test method is used to determine whether there is a significant difference between two group's means. It helps to answer the underlying question: do the two groups come from the same population, and only appear differently because of chance errors, or is there some significant difference between these two groups [55]. Since data used for prediction in all models are same, we carried out t-test between coverage = 0.9 and coverage = 0.6 to verify whether the above results are statistically significant. The p-values are reported in Table 3. Since p-values < 0.05, the results indicate that coverage = 0.6 statistically significantly outperforms coverage = 0.9 on the average correct classification rate. Moreover, Tables 4 and 5 present the partial rule sets for going-concern prediction on coverage = 0.9 and coverage = 0.6, respectively.

In Table 4, "Rules" refer to the decision rules generated by the LEM2 algorithm for classifying going-concern opinions, and "Support" refers to real examples coinciding with the generated decision rules in the dataset. Furthermore, in Table 4, two class rules regarded as intelligent auditing systems were generated; an example of rule 1 is presented as follows:

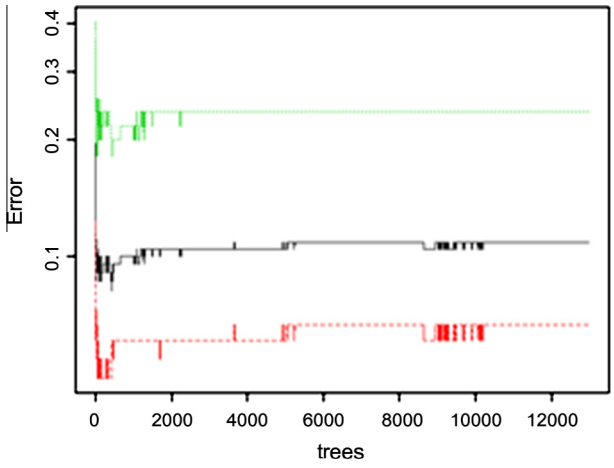

Fig. 1. Error rate (OOB) plot versus the number of trees.

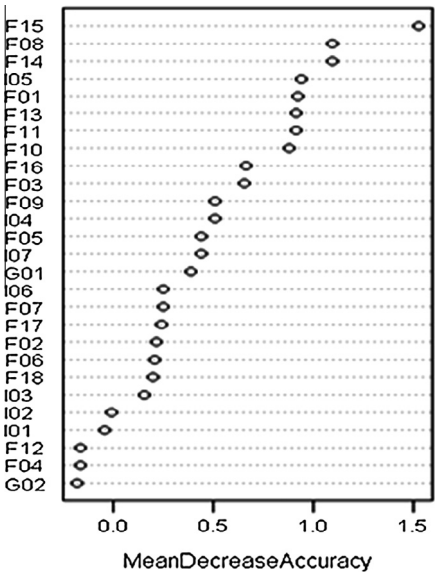

Fig. 2. Variables' importance ranking.

Table 3
The results of experiment for RF + RST.

Rounds	Coverage = 0.9[a]		Coverage = 0.6		*p*-Value[b]
	Accuracy	Rules	Accuracy	Rules	
1	0.963	24	0.974	5	
2	0.918	19	0.952	8	
3	0.939	20	0.976	7	
4	0.98	23	0.973	7	
5	0.979	19	0.951	6	
6	0.926	21	0.973	9	
7	0.925	23	1	9	
8	0.909	20	1	8	
9	0.951	21	0.95	3	
10	0.977	18	0.95	5	
Average	0.9467	20.8	0.9699	6.7	−0.0402*

[a] Coverage refers to the parameter value from LEM2 algorithm.
[b] * Indicates a significant difference of the average correct classification rate between two models at the 0.05 significance level.

Table 4
The partial rule sets by LEM2 algorithm using coverage = 0.9.

No.	Rules	Support
1	(G01=1)&(F16=0)=>(PYG={0[66]})	66
2	(F16=1)&(G01=1)&(F11=2)&(F10=2)&(F13=2)&(F03=3)=>(PYG={0[8]})	8
3	(F16=1)&(F11=2)&(F10=2)&(I04=3)&(F08=1)=>(PYG={0[7]})	7
4	(F16=1)&(F15=1)&(F13=1)&(F11=1)&(F10=1)&(I04=1)&(I05=1)&(F08=4)&(F14=1)&(F01=1)&(F09=4)&(G01=1)=>(PYG={1[6]})	6
5	(F03=4)&(F05=1)&(F13=1)&(F16=1)=>(PYG={1[6]})	6
6	(G01=0)&(F16=0)&(F10=3)=>(PYG={0[5]})	5
7	(F16=1)&(F15=1)&(F13=1)&(F11=1)&(F10=1)&(G01=1)&(I04=1)&(I05=1)&(F05=1)&(F01=1)=>(PYG={1[5]})	5
8	(F16=1)&(F08=4)&(F14=1)&(F15=1)&(F01=1)&(F10=2)=>(PYG={1[5]})	5

Table 5
The partial rule sets by LEM2 algorithm using coverage = 0.6.

No.	Rules	Support
1	(G01=1)&(F16=0)=>(PYG={0[75]})	75
2	(F16=1)&(F15=1)&(F13=1)&(F08=4)&(F14=1)&(F11=1)&(I04=1)&(F09=4)&(F10=1)&(I05=1)=>(PYG={1[11]})	11
3	(F16=1)&(F15=1)&(F13=1)&(F08=4)&(F14=1)&(F01=1)&(G01=1)&(I04=1)&(I05=1)=>(PYG={1[11]})	11
4	(F16=1)&(F08=4)&(F14=1)&(F15=1)&(F01=1)&(F10=2)=>(PYG={1[7]})	7

Table 6
Comparisons of experiment results of the constructed model.

Criteria	Model						
	RF + RST (coverage = 0.9)	RST	RF + RST (coverage = 0.6)	RST	RF + DT	RF + NN	RF + SVM
Accuracy	0.9467	0.9266	0.9699	0.9374	0.91549	0.9295	0.9458

Table 7
The *t*-test of results.

Methods		RF+ RST (coverage=0.9)	RST	RF+RST (coverage=0.6)	RST	RF+DT	RF+NN	RR+SVM
RF + RST	(coverage = 0.9)		0.0284*	−0.0317*	0.0485*	0.0173*	0.0313*	0.0507
RST				−0.0116*	−0.0377*	0.0374*	−0.0456*	−0.0293*
RF + RST	(coverage = 0.6)				0.0224*	0.0005*	0.0145*	0.0308*
RST						0.0266*	0.0406*	−0.0401*
RF + DT							−0.0345*	−0.0182*
RF + NN								−0.0322*

* Indicates a significant difference of the average correct classification rate between these models at the 0.05 significance level, and the values are the corresponding *p*-values.

Table 8
Type I and Type II errors of the constructed model.

Model	Performance assessment (%)	
	Type I error	Type II error
RF + RST (coverage = 0.9)	4.10	12.80
RST	4.30	13.25
RF + RST (coverage = 0.6)	3.50	10.60
RST	3.80	11.84
RF + DT	1.90	26.0
RF + NN	4.32	18.02
RF + SVM	3.74	11.02

Table 9
Comparisons of experimental results with different variables for RF + RST (coverage = 0.9).

Rounds	Predictive variables					
	IC + financial ratios		Non-financial ratios + financial ratios		Financial ratios	
	Accuracy	Rules	Accuracy	Rules	Accuracy	Rules
1	0.957	27	0.941	31	0.925	33
2	0.929	27	0.963	27	0.909	28
3	0.964	23	0.915	26	0.939	30
4	0.918	25	0.942	21	0.962	24
5	0.935	19	0.980	26	0.961	38
6	0.931	17	0.911	28	0.945	29
7	0.961	22	0.930	26	0.898	29
8	0.920	22	0.915	27	0.839	23
9	0.957	30	0.918	25	0.936	30
10	0.933	21	0.933	24	0.920	28
Average	0.9405	23.3	0.9348	26.1	0.9234	29.2

If (G01 = 1) and (F16 = 0), then PYG = 0 (support = 66). This indicates that when G01 = 1 and F16 = 0 occur simultaneously, the PYG (class) is "Clean." In other words, the going-concern opinion is clean when a Big 4 auditor performs the audit and, concurrently, the firm does not have negative net income. A total of 66 examples support this rule.

From Table 3, the average accuracy rate of the experiment results are 0.9467 and 0.9699 based on coverage = 0.9 and coverage = 0.6, respectively. For the performance evaluation of the proposed hybrid model, the experiment uses the pure RST, RF + DT, RF + DT, and RF + SVM models as benchmarks. Table 6 presents a comparison of the experiment results obtained using different methods. The empirical results indicated that the proposed procedure outperforms the other methods listed in this paper in terms of accuracy rate. Moreover, to verify whether the above models are statistically significant, we also conducted *t*-tests on the average correct classification rates for these models. The *p*-values are reported in Table 7. Although there is no significant difference between RF + RST with coverage = 0.9 and RF + SVM in terms of prediction accuracy. Since *p*-values < 0.05, the proposed hybrid model (RF + RST with coverage = 0.6) performs the best in terms of prediction accuracy.

Table 10
Comparisons of experimental results with different variables for RF + RST (coverage = 0.6).

Rounds	Predictive variables					
	IC + financial ratios		Non-financial ratios + financial ratios		Financial ratios	
	Accuracy	Rules	Accuracy	Rules	Accuracy	Rules
1	0.974	11	1	9	0.913	12
2	1	11	0.95	7	0.973	14
3	0.952	8	0.956	8	0.950	16
4	1	13	0.952	7	0.975	15
5	0.975	16	0.946	11	0.950	15
6	0.955	7	0.938	10	0.975	16
7	1	13	0.971	10	0.971	14
8	0.933	8	0.947	8	0.949	13
9	0.917	9	0.974	8	0.943	11
10	0.955	11	0.976	12	0.971	11
Average	0.9661	10.7	0.961	9	0.957	13.7

Furthermore, as shown in Table 7, RF improves the classification accuracy of the going-concern prediction, regardless of the coverage used.

Moreover, Table 8 summarizes the rate of Type I and Type II errors of the models discussed. As shown in Table 8, our hybrid model has fewer Type II errors as compared to the other models. Hence, we can conclude that the "RF + RST" model not only has the best classification rate but also has the lowest incidence of Type II errors.

Moreover, for testing whether IC is helpful for going-concern prediction, the experiments also repeatedly tested these models using three predictive variable methods: IC + financial and non-financial ratios, financial and non-financial ratios, and the baseline (which uses financial ratios). Tables 9 and 10 show the experimental results for average accuracy rate with three performance measurements of predictive variables for RF + RST method on coverage = 0.9 and coverage = 0.6, respectively. As shown in Tables 9 and 10, using IC and financial ratios as the predictors for the RF + RST method creates the best performance for prediction accuracy.

Finally, this paper employed *t*-tests to examine whether there are any significant differences between these predictors in terms of prediction accuracy. The results in Tables 11 show that coverage = 0.9 and coverage = 0.6 using IC as a predictor statistically significantly outperforms those without IC, respectively. Therefore, from the improved classification rate of the model, we can conclude that IC improves the classification accuracy of the prediction model, regardless of coverage and variables.

5.3. Discussion and findings

According to the experiments discussed above, the analysis results and implications for going-concern prediction are presented as follows:

(1) There were numerous predictive variables that could be considered. Therefore, finding the important predictive variables was crucial, as this would affect the accuracy and classification of the model developed. Instead of selecting variables based on domain knowledge, we selected the variables according to their importance as calculated by RF. Tables 6–8, prove that the RF method can effectively improve the accuracy rate of going-concern prediction, regardless of the methods and variables used. The analysis presented above suggests that variable selection can enable researchers in making going-concern predictions without possessing special domain knowledge.

(2) Based on the predictive variables discussed above and a decision attribute, we generated decision rules using the LEM2 algorithm for classifying going-concern opinions. This yielded a set of comprehensible and meaningful rules that can be applied readily in intelligent systems for predicting going-concern opinion. Through the proposed hybrid model, the study attempted to determine the hidden information in decision rules, which can offer an intelligent way to predict going-concern opinions.

(3) In summary, based on Tables 6–8, the ranking of the models in order of accuracy is as follows: RF + RST, RST + SVM, RF + NN, and RF + DT. The results of the experiments offer insight into why the proposed hybrid model was superior in this study. The results prove that the proposed hybrid model is stable in terms of accuracy because it is the best model, regardless of accuracy, Type II errors, and predictive variables.

(4) This study also shows that the models including IC and financial ratios provide better classification results than the predictors using non-financial and financial ratios. Moreover, IC improves the classification accuracy of the prediction model. Overall, this shows that IC is more important than the non-financial ratios in predicting the going-concern opinion.

(5) Finally, Tables 9–11, in which the proposed hybrid model is compared with the other models in terms of accuracy and predictive variables, show that proposed hybrid approach are superior to AI techniques in different hybrid models. Moreover, when the RF + RST method on coverage = 0.9 is compared with the RF + SVM method, the accuracy rate

Table 11
The *t*-test of results for RF + RST.

Predictive variables	Non-financial ratios + financial ratios	Financial ratios
Coverage = 0.9		
IC + non-financial ratios + financial ratios	0.0381*	0.0267*
IC + financial ratios	0.0443*	0.0329*
Non-financial ratios + financial ratios		0.0386*
Coverage = 0.6		
IC + non-financial ratios + financial ratios	0.0411*	0.0371*
IC + financial ratios	0.0449*	0.0409*
Non-financial ratios + financial ratios		0.0460*

* Indicates a significant difference of the average correct classification rate between these models at the 0.05 significance level, and the values are the corresponding *p*-values.

is 0.9467 vs. 0.9458 in the dataset, respectively. Clearly, the difference of the two methods is very small and of no significance even. Therefore, owing to the limitations of AI techniques in the study, it is worthwhile to propose a newer hybrid model to amplify the advantages of the hybrid models and minimize their limitations.

6. Conclusions

Corporate going-concern prediction plays a significant role in accounting and audit decisions. In the current knowledge era, the core competences of firms are derived from the knowledge and skills of their employees, and the value of IC now exceeds that of some tangible assets. In current prediction models, many researchers have focused on the financial ratios rather than IC. In this study, we use IC as the predictive variable and propose a hybrid model that combines RF and RST (the LEM2 algorithm) to not only enhance classification accuracy but also extract meaningful rules for going-concern prediction. To demonstrate the proposed approach, we employed the pure RST, RF + DT, RF + NN, and RF + SVM models as benchmarks. According to the experiments, the results of this study can be summarized as follows.

First, IC is a useful *ex ante* determinant of going-concern prediction. The new predictive variable, that is, IC, improves the classification accuracy of going-concern prediction. We obtained this result by emphasizing the linkage between IC and going-concern prediction. In particular, we have shown that HC variables are important *ex ante* indicators of the corporation's going-concern prediction. Second, the experimental results indicate that the proposed model surpasses the other listed methods in terms of both higher accuracy and fewer variables, and the proposed procedure yields a set of easily understandable decision rules that facilitate the interpretation of audit information and enable auditors to identify the variables that are most important to an intelligent audit system for predicting going-concern opinion. Nonetheless, currently there are no predictive models for this purpose that can be generalized for all settings. The proposed approach is an alternative for the development of such going-concern decision rules, but other AI techniques, such as NN, DT, and SVM, are available and more studies are needed to define which would have the best performance for predicting going-concern. In particular, these meaningful rules enable clients to understand going-concern opinions. This study may aid auditors in focusing on the main variable and making useful decisions for clients.

Based on the findings of this empirical case study, we can positively conclude that the proposed hybrid approach using RF and RST is more efficient than the listed methods for classifying the going-concern opinions of clients. For future research, other applications can be considered for evaluating the approach proposed in this study, such as business failure or credit rating, or other predictive variables can be used as attributes for classifying the going-concern opinion. Moreover, we should continue to compare our proposed approach with the state-of-the-art algorithms for classification in predicting going-concern.

Acknowledgments

The authors would like to thank the Editor-in-Chief and reviewers for their useful comments and suggestions, which were very helpful in improving this manuscript.

References

[1] E.H. Aboul, E.A. Mohamed, S.O. Hala, Rough sets data analysis in knowledge discovery: a case of Kuwaiti diabetic children patients, Advance in Fuzzy Systems 8 (2008) 1–13.
[2] M. Anandarajan, A. Anandarajan, A comparison of machine learning techniques with a qualitative response model for auditors' going concern reporting, Expert Systems with Applications 16 (4) (1999) 385–392.
[3] P. Barnes, H. Huan, The auditor's going concern decision: some UK evidence concerning independence and competence, Journal of Business Finance & Accounting 20 (2) (1993) 213–228.
[4] B.K. Behn, S.E. Kaplan, K.P. Krumwiede, Further evidence on the auditor's going-concern report: the influence of management plans, Auditing: A Journal of Practice & Theory 20 (1) (2001) 13–29.
[5] T. Bell, R. Tabor, Empirical analysis of audit uncertainty qualifications, Journal of Accounting Research 29 (2) (1991) 350–370.

[6] J.L. Bellovary, D.E. Giacomino, M.D. Akers, A review of going concern prediction studies: 1976 to present, Journal of Business & Economic Research 5 (2007) (1976) 9–28.
[7] M. Beynon, M. Peel, Variable precision rough set theory and data discrimination: an application to corporate failure prediction, Omega 29 (2001) 561–576.
[8] G. Biau, L. Devroye, G. Lugosi, Consistency of random forests and other averaging classifiers, Journal of Machine Learning Research 9 (2008) 2015–2033.
[9] N. Bontis, Intellectual capital: an exploratory study that develops measures and models, Management Decision 36 (2) (1998) 63–76.
[10] L. Breiman, Random forests, Machine Learning 45 (2001) 5–32.
[11] Y. Cao, G. Wan, F. Wang, Predicting financial distress of Chinese listed companies using rough set theory and support vector machine, Asia-Pacific Journal of Operational Research 28 (1) (2011) 95–109.
[12] J.V. Carcello, T.L. Neal, Audit committee composition and auditor reporting, Accounting Review 75 (4) (2000) 453–467.
[13] J.R. Casterella, B.L. Lewis, P.L. Walker, Modeling the audit opinions issued to bankrupt companies: a two-stage empirical, Journal of Business Finance & Accounting 32 (2000) 204–229.
[14] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, 2001. <http://www.csie.ntu.edu.tw/cjlin/libsvm> (accessed 29.04.13).
[15] K.C.W. Chen, B.K. Church, Default on debt obligations and the issuance of going-concern opinions, Auditing: A Journal of Practice & Theory 11 (2) (1992) 30–50.
[16] D. Cornier, M. Magnan, B. Morard, The auditor's consideration of the going concern assumption: a diagnostic model, Journal of Accounting, Auditing & Finance 10 (2) (1995) 201–221.
[17] M. Daubie, P. Levecq, N. Meskens, A comparison of rough sets and recursive partitioning induction approaches: an application to commercial loans, International Transactions in Operational Research 9 (2002) 681–694.
[18] E.B. Deakin, A discriminant analysis of predictors of business failure, Journal of Accounting Research 10 (1972) 167–179.
[19] P. De Saa-Perez, J.M. Garcia-Falcon, A resource-based view of human resource management and organizational capability development, International Journal of Human Resource Management 13 (1) (2002) 123–140.
[20] R. Diaz-Uriarte, S. Alvarez de Andres, Gene selection and classification of microarray data using random forest, BMC Bioinformatics 7 (2006) 3.
[21] A. Dimitras, R. Slowinski, R. Susmaga, C. Zopounidis, Business failure using rough sets, European Journal Operational Research 114 (1999) 263–280.
[22] N. Dopuch, R. Holthausen, R. Leftwich, Predicting audit qualifications with financial and market variables, Accounting Review 63 (3) (1987) 431–453.
[23] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC, New York, 1993.
[24] H.L. Etheridge, R.S. Sriram, H.Y.K. Hsu, A comparison of selected artificial neural networks that help auditors evaluate client financial viability, Decision Sciences 31 (2) (2000) 531–550.
[25] B. Foster, T. Ward, J. Woodroof, An analysis of the usefulness of debt defaults and going concern opinions in bankruptcy risk assessment, Journal of Accounting, Auditing & Finance 13 (3) (1998) 351–371.
[26] C. Gaganis, F. Pasiouras, M. Doumpos, Probabilistic neural networks for the identification of qualified audit opinions, Expert Systems with Applications 32 (2007) 114–124.
[27] J.W. Grzymala-Busse, A new version of the rule induction system LERS, Fundamental Informaticae 31 (1997) 27–39.
[28] J.W. Grzymala-Busse, R. Slowinski, LERS – a system for learning from examples based on rough sets, in: R. Slowinski (Ed.), Intelligent Decision Support: Handbook of Applications and Advances in Rough Set Theory, Kluwer Academic Publishers, London, 1992.
[29] J.V. Hansen, J.B. McDonald, J.D. Stice, Artificial intelligence and generalized qualitative-response models: an empirical test on two audit decision-making domains, Decision Science 23 (3) (1992) 708–723.
[30] W. Hopwood, J.C. Mckeown, J.F. Mutchler, A reexamination of auditor versus model accuracy within the context of the going-concern opinion decision, Contemporary Accounting Research 10 (1994) 409–431.
[31] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, New York, 1998.
[32] T. Kida, An investigation into auditors' continuity and related qualification judgments, Journal of Accounting Research 18 (2) (1980) 506–523.
[33] E. Kirkos, C. Spathis, A. Nanopoulos, Y. Manolopoulos, Identifying qualified auditors' opinions: a data mining approach, Journal of Emerging Technologies in Accounting 4 (1) (2007) 183–197.
[34] H.C. Koh, R. Brown, Probit prediction of going and non-going concerns, Managerial Auditing Journal 6 (3) (1991) 18–23.
[35] H.C. Koh, C.K. Low, Going concern prediction using data mining techniques, Managerial Auditing Journal 19 (3) (2004) 462–476.
[36] E. Krusinska, R. Slowinski, J. Stefanowski, Discriminant versus rough set approach to vague data analysis, Applied Stochastic Models and Data Analysis 8 (1) (1992) 43–56.
[37] M.J. Lenard, P. Alam, G.R. Madey, The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision, Decision Sciences 26 (2) (1995) 209–227.
[38] H. Li, J. Sun, Predicting business failure using multiple case-based reasoning combined with support vector machine, Expert Systems with Applications 36 (2009) 10085–10096.
[39] A. Liaw, M. Wiener, Classification and regression by random forest, R News 2–3 (2002) 18–22.
[40] K.L. Lunetta, L.B. Hayward, J. Segal, P.V. Eerdewegh, Screening large-scale association study data: exploiting interactions using random forests, BMC Genetics 5 (2004) 32.
[41] D. Martens, L. Bruynseels, B. Baesens, M. Willekens, J. Vanthienen, Predicting going concern opinion with data mining, Decision Support Systems 45 (4) (2008) 765–777.
[42] T. McKee, Discriminant prediction of going concern status: a model for auditors, in: Proc. 6th American Accounting Association Annual Meeting, 1996, pp. 294–321.
[43] J. McKeown, J. Mutchler, W. Hopwood, Towards an explanation of auditor failure to modify the audit opinions of bankrupt companies, Auditing: A Journal of Practice & Theory 10 (1991) 1–13.
[44] R.E. Morris, R.S. Jerry, An examination of the effect of CPA firm type on bank regulators' closure decisions, Auditing: A Journal of Practice & Theory 18 (2) (1999) 143–158.
[45] J.F. Mutchler, Auditor's perceptions of the going-concern opinion decision, Auditing: A Journal of Practice & Theory 3 (1984) 17–29.
[46] J.F. Mutchler, W. Hopwood, J. McKeown, The influence of contrary information and mitigating factors on audit opinion decisions on bankrupt companies, Journal of Accounting Research 35 (2) (1997) 295–310.
[47] S.K. Pal, P. Mitra, Case generation using rough sets with fuzzy representation, IEEE Transactions on Knowledge and Data Engineering 16 (2004) 293–300.
[48] Z. Pawlak, Rough sets, International Journal of Information Computation Science 11 (5) (1982) 341–356.
[49] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Springer-Verlag, New York, 1991.
[50] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.
[51] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (2007) 28–40.
[52] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1993.
[53] T.A. Stewart, Intellectual Capital: The New Wealth of Organizations, Bantam Doubleday Dell Publishing Group, New York, 1997.
[54] A. Sun, E.-P. Lim, Y. Liu, On strategies for imbalanced text classification using SVM: a comparative study, Decision Support Systems 48 (2009). pp. 191-20.
[55] C.-F. Tsai, Feature selection in bankruptcy prediction, Knowledge-Based Systems 22 (2009) 120–127.
[56] A. Vellido, P.J.G. Lisboa, J. Vaughan, Neural networks in business: a survey of applications (1992–1998), Expert Systems with Applications 17 (1999) 51–70.
[57] P. Verwijmeren, J.M.M. Derwall, Employee well-being, firm leverage, and bankruptcy risk, Journal of Banking & Finance 34 (5) (2010) 956–964.

[58] M.C. Wiener, L. Obando, J. O'Neill, Building Process Understanding for Vaccine Manufacturing Using Data Mining, Quality Engineering 22 (3) (2010) 157–168.
[59] Y. Wu, C. Gaunt, S. Gray, A comparison of alternative bankruptcy prediction models, Journal of Contemporary Accounting & Economics 6 (2010) 34–45.
[60] W.Z. Wu, W.X. Zhang, H.Z. Li, Knowledge acquisition in incomplete fuzzy information systems via the rough set approach, Expert Systems 20 (2003) 280–286.
[61] Z. Xiao, X. Yang, Y. Pang, X. Dang, The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster– Shafer evidence theory, Knowledge-Based Systems 26 (2012) 196–206.
[62] Y.Y. Yao, Semantics of fuzzy sets in rough set theory, Transactions on Rough Sets 2 (2004) 310–331.
[63] C.-C. Yeh, D.-J. Chi, M.-F. Hsu, A hybrid approach of DEA, rough set and support vector machines for business failure prediction, Expert Systems with Applications 37 (2) (2010) 1535–1541.
[64] C.-C. Yeh, F. Lin, C.-Y. Hsu, A hybrid KMV model, random forests and rough set theory approach for credit rating, Knowledge-Based Systems 33 (2012) 166–172.
[65] M.A. Youndt, M. Subramaniam, S.A. Snell, Intellectual capital profiles: an examination of investments and returns, Journal of Management Studies 41 (2) (2004) 335–361.
[66] J. Zhang, M. Zulkernine, Network intrusion detection using random forests, Proceedings of the Third Annual Conference on Privacy, Security and Trust, St. Andrews, New Brunswick, Canada, 2005. October.
[67] J. Zhang, M. Zulkernine, A hybrid network intrusion detection technique using random forests, in: The First International Conference on Availability, Reliability and Security, 2006 (ARES"06), IEEE, Austria, 2006.