

Research Article

Adaboost Ensemble Classifiers for Corporate Default Prediction

¹Suresh Ramakrishnan, ¹Maryam Mirzaei and ²Mahmoud Bekri

¹Department of Management, Universiti Teknologi Malaysia, Malaysia

²Economic and Statistic Institute, Karlsruhe Institute of Technology, Germany

Abstract: This study aims to show a substitute technique to corporate default prediction. Data mining techniques have been extensively applied for this task, due to its ability to notice non-linear relationships and show a good performance in presence of noisy information, as it usually happens in corporate default prediction problems. In spite of several progressive methods that have widely been proposed, this area of research is not out dated and still needs further examination. In this study, the performance of multiple classifier systems is assessed in terms of their capability to appropriately classify default and non-default Malaysian firms listed in Bursa Malaysia. Multi-stage combination classifiers provided significant improvements over the single classifiers. In addition, Adaboost shows improvement in performance over the single classifiers.

Keywords: Data Mining, default prediction, ensemble classifier

INTRODUCTION

Due to the significant consequences which default imposes on different groups of society as well the noteworthy troubles qualified by firms during the Global Financial Crisis, the crucial importance of measuring and providing for credit risk have highlighted. Since the mid-1990s, there has been growing concern in emerging and developing economies among researchers. Regarding the growth in financial services, there have been swelling sufferers from off ending loans. Therefore, default risk forecasting is a critical part of a financial institution's loan approval decision processes. Default risk prediction is a procedure that determines how likely applicants are to default with their repayments. Review of literature on the subject confirmed hand full of studies conducted in the last four decades. Despite of these studies, the recent credit crisis indicated that yet there are areas of the study that needs researchers' attention. Moreover, emerging of the regulatory changes such as Basel III accord and the need for more precise and comprehensive risk management procedures justifies need of research in area of credit risk modeling and banking supervision. This requirement like these pushes companies especially banks and insurance companies to have a very robust and transparent risk management system.

As a valuable implement for scientific decision making, corporate default prediction takes an imperative role in the prevention of corporate default. From this point of view, the accuracy of default

prediction model is an essential issue and many researchers have focused on how to build efficient models. In supervised classification tasks, the mixture or ensemble of classifiers represent a remarkable method of merging information that can present a superior accuracy than each individual method. To improve model accuracy, classifier ensemble is a capable technique for default prediction. In fact, the high classification accuracy performance of these combined techniques makes them appropriate in terms of real world applications, such as default prediction. However, research on ensemble methods for default prediction just begins recently and warrants to be considered comprehensively.

Former researches on ensemble classifier for default prediction used DT or NN as base learner and were both compared to single NN classifier. This study further explores Ada Boost and bagging ensemble for default prediction to compare with various baseline classifiers including learning Logistic Regression (LR), Decision Tree (DT), artificial Neural Networks (NN) and support vector machine (SVM) as base learner.

LITERATURE REVIEW

Significant advances have been made in the past few decades regarding methodologies for default prediction. Beaver (1966) introduced the Naïve Bayes approach using a single variable and Altman (1968, 1973) suggested the use of Linear Discriminant Analysis (LDA). Since then several contributions have been made to improve the Altman's results, using different techniques. The use of data mining techniques

such as Artificial Neural Networks (ANN), decision trees and Support Vector Machine (SVM) for bankruptcy prediction started in the late 1980s (Pompe and Feelders, 1997; Shin *et al.*, 2005).

Frydman *et al.* (1985) used Decision Trees first time for default prediction. Using this model, they classified firms to failed and non-failed based on firm-level and country-level factors. According to their results, this technique allows for an easy identification of the most significant characteristics in default prediction. In another study, Quinlan (1986) noted that decision trees method can deal with noise or non-systematic errors in the values of features. There are some other studies which predicted default using this method such as, (Messier Jr. and Hansen, 1988; Pompe and Feelders, 1997). Detailed examination of corporate default prediction by Lin and McClean (2001) showed a better performance of the hybrid model. They used four different techniques to predict corporate default, which two of the methods were statistical and the outstanding two models were machine learning techniques. In different but related work, Shin and Lee (2002) suggested a model using genetic algorithms technique. Some other related studies have employed Artificial Neural Networks to predict default.

Artificial Neural Networks was first demonstrated experimentally by Hertz *et al.* (1991) to analyze bankrupt companies. Since then the method became a common accuracy amongst. Recently, some of the main commercial loan default prediction products applied ANN technique. For example, Moody's public firm risk model ANN and many banks and financial institutions have developed this method for default prediction (Atiya, 2001). More recently, the support vector machine was commenced for default risk investigation. This technique which is based on statistical learning theory compared with the traditional methods is more accurate in predicting default likelihood (Hårdle *et al.*, 2005). In a major study on default prediction Gestel *et al.* (2005) employed SVM and logistic regression. The results based on combination of both techniques showed more stability in prediction power which is necessary for rating banks.

The limited research undertaken into the application of classifier combination to default prediction problems has arguably generated better results. In this regard, Myers and Forgy (1963) implemented a multi-stage methodology in which they employed a two stage discriminate analysis model. The second stage model was constructed using the lowest scoring of the development sample used in the first stage. They reported that the second stage model identified 70% more bad cases than the first stage model. In another study, Lin (2002) conveyed up to 3% improvement when employing a logistic model, followed by a neural network. There has been relatively little research effort to compare different classification methodologies within the credit risk area. Only in the study by West *et al.* (2005) was more than a single combination strategy given consideration and in this

case only one type of classifier which is neural network has been employed. In another study, Abellán and Masegosa (2012) showed that using bagging ensembles on a special type of decision trees, called Credal Decision Trees (CDTs), provides an appealing tool for the classification task.

METHODOLOGY

Framework of Adaboost ensemble method: The key idea of multiple classifier systems is to employ ensemble of classifiers and combine them in various approaches. Theoretically, in an ensemble of N independent classifiers with uncorrelated error areas, the error of an overall classifier obtained by simply averaging/voting their output can be reduced by a factor of N . Boosting is a meta-learning algorithm and the most broadly used ensemble method and one of the most powerful learning ideas introduced in the last twenty years. The original boosting algorithm has been proposed by Robert Schapire (a recursive majority gate formulation and Yoav Freund (boost by majority) in 1990. In this type, each new classifier is trained on a data set in which samples misclassified by the previous model are given more weight while samples that are classified correctly are given less weight. Classifiers are weighted according to their accuracy and outputs are combined using a voting representation. The most popular boosting algorithm is AdaBoost (Freund and Schapire, 1997). Adaboost applies the classification system repeatedly to the training data, but at each application, the learning attention is focused on different examples of this set using adaptive weights ($\omega_b(i)$). Once the training procedure has completed, the single classifiers are combined to a final, highly accurate classifier based on the training set. A training set is given by:

$$T_n = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$$

where, y takes values of $\{-1, 1\}$. The weight $\omega_b(i)$ is allocated to each observation X_i and is initially set to $1/n$. this value will be updated after each step. A basic classifier denoted $C_b(X_i)$ is built on this new training set, T_b and is applied to each training sample. The error of this classifier is represented by ξ_b and is calculated as:

$$\xi_b = \sum \omega_b(i) \xi_b(i)$$

where,

$$\xi_b(i) = \begin{cases} 0 & C_b(X_i) = y_i \\ 1 & C_b(X_i) \neq y_i \end{cases}$$

The new weight for the $(b+1)$ -th iteration will be:

$$\omega_{b+1}(i) = \omega_b(i) \cdot \exp(\alpha_b \xi_b(i))$$

where, α_i is a constant calculated from the error of the classifier in the b -th iteration. This process is repeated

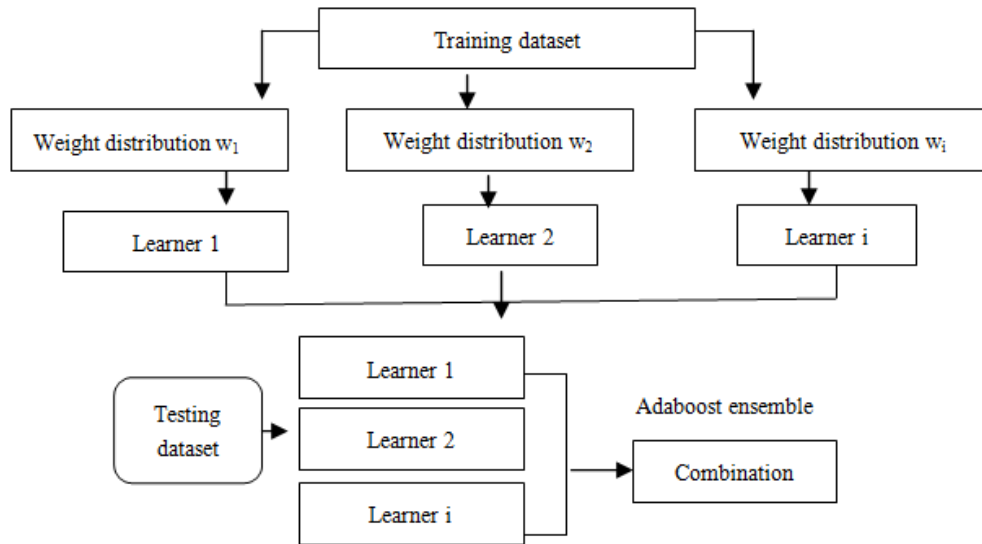


Fig. 1: The framework of Adaboost algorithm

in every step for $b = 1, 2, 3, \dots, B$. Finally, the ensemble classifier is built as a linear combination of the single classifiers weighted by the corresponding constant α_b . The framework of AdaBoost algorithm, weak learning algorithm and combination mechanism for default prediction is shown in Fig. 1.

Supervised learners:

Logistic regression: Logistic regression is a type of regression methods (Allison, 2001; Hosmer and Lemeshow, 2000) where the dependent variable is discrete or categorical, for instance, default (1) and non-default (0). Logistic regression examines the effect of multiple independent variables to forecast the association between them and dependent variable categories. According to Morris (1997) and Martin (1977) was the first researcher who used logistic technique in corporate default perspective. He employed this technique to examine failures in the U.S. banking sector. Subsequently, Ohlson (1980) applied logistic regression more generally to a sample of 105 bankrupt firm and 2,000 non-bankrupt companies. His model did not discriminate between failed and non-failed companies as well as the Multiple Discriminate Analysis (MDA) models reported in previous studies. According to Dimitras *et al.* (1996), logistic regression is in the second place, after MDA, in default prediction models.

Decision tree: Decision trees are the most popular and powerful techniques for classification and prediction. The foremost cause behind their recognition is their simplicity and transparency and consequently relative improvement in terms of interpretability. Decision tree is a non-parametric and introductory technique, which is capable to learn from examples by a procedure of

simplification. Frydman *et al.* (1985) first time employed decision trees to forecast default. Soon after, some researchers applied this technique to predict default and bankruptcy including (Carter and Catlett, 1987; Gepp *et al.*, 2010; Messier Jr. and Hansen, 1988; Pompe and Feelders, 1997).

Neural networks: Neural Networks (NNs), usually non-parametric techniques have been used for a variety of classification and regression problems. They are characterized by associates among a very large number of simple computing processors or elements (neurons). Corporate default have predicted using neural networks in early 1990s and since then more researchers have used this model to predict default. As a result, there are some main profitable loan default prediction products which are based on neural network models. Also, there are different evidence from many banks which have already expanded or in the procedure of developing default prediction models using neural network (Atiya, 2001). This technique is flexible to the data characteristics and can deal with different non-linear functions and parameters also compound prototypes. Therefore, neural networks have the ability to deal with missing or incomplete data (Smith and Stulz, 1985; Smith and Winakor, 1935).

Support vector machines: Among different classification techniques, Support Vector Machines are considered as the best classification tools accessible nowadays. There are a number of empirical results attained on a diversity of classification (and regression) tasks complement the highly appreciated theoretical properties of SVMs. A Support Vector Machine (SVM) produces a binary classifier, the so-called optimal

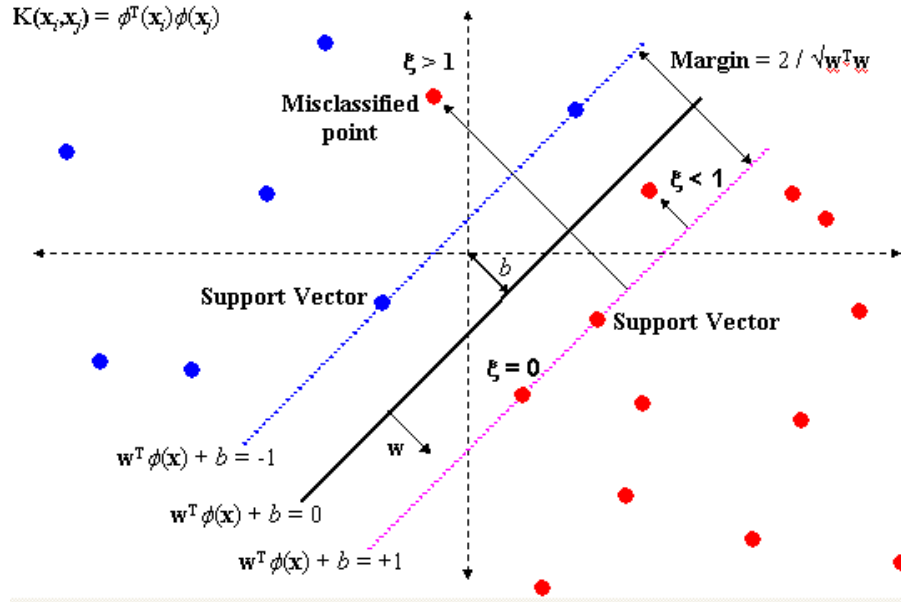


Fig. 2: The SVM learns a hyperplane which best separates the two classes

separating hyper planes, through extremely nonlinear mapping the input vectors into the high-dimensional feature space. SVM constructs linear model to estimate the decision function using non-linear class boundaries based on support vectors. Support vector machine is based on a linear model with a kernel function to implement non-linear class boundaries by mapping input vectors non-linearly into a high-dimensional feature space.

Based on conceptual elements of statistical learning and the potential of SVMs for firm rating, for the linear classification problem a SVM is defined and this method is simplified for nonlinear cases. In the linear case (Fig. 2) the following inequalities hold for all n points of the training set:

$$\begin{aligned} \text{Min } & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ x_i^T w + b & \geq 1 - \xi_i \quad \text{for } y_i = 1 \\ x_i^T w + b & \leq -1 + \xi_i \quad \text{for } y_i = -1 \\ \xi_i & \geq 0 \end{aligned}$$

This can be combined into two constraints:

$$\begin{aligned} Y_i (x_i^T w + b) & \geq 1 - \xi_i \\ \xi_i & \geq 0 \end{aligned}$$

The basic idea of the SVM classification is to find such a separating hyperplane that corresponds to the largest possible margin between the points of different classes.

EMPIRICAL RESULTS

Data description: The dataset was used to classify a set of firms into those that would default and those that would not default on loan payments. It consists of 285

observations of Malaysian companies. Of the 285 cases for training, 121 belong to the default case under the requirements of PN4, PN17 and Amended PN17 respectively and the other 164 to non-default case. Consulting an extensive review of existing literature on corporate default models, the most common financial ratios that are examined by various studies were identified. The variable selection procedure should be largely based on the existing theory. The field of default prediction, however, suffers from a lack of agreement as for which variables should be used. The first step in this empirical search for the best model is therefore the correlation analysis. If high correlation is detected, the most commonly used and best performing ratios in the literature are prioritized. Therefore, the choice of variables entering the models is made by looking at the significance of ratios.

The components of the financial ratios which are estimated from data are explained below and Table 1 shows the summary statistics for selected variables for default and non-default firms. To select the variables, two approaches including linear regression and decision tree analysis were used. The most significant variables based on two methods were identified. These variables selected from the significant indicators for the model which could best discriminate the default firms from the non-default firms. These selected financial ratios include: Profitability ratios, liquidity and growth opportunity (Fig. 3).

RESULTS AND DISCUSSION

In this experiment study, the main goal is to compare ensemble classifiers. To obtain comparable experimental results, the same default prediction problem is solved by four different classification

Table 1: The summary statistics for selected variables for default and non-default firms

	Definition of variable	Means of non-default companies	Means of default companies	Test of equality of group means
1	EBIT /total assets	0.155647	-0.026080	0.000
2	Cash/total assets	0.046677	-0.191281	0.137
3	Current asset/current liabilities	1.854502	1.178482	0.000
4	Size	6.181860	5.844030	0.271
5	Total current liabilities to total assets	0.151896	-0.05514	0.000
6	Net profit/Net sale	0.098689	-0.87107	0.174
7	Growth opportunity	0.095607	-0.06198	0.000
8	Net profit/total assets	0.095607	-0.06197	0.000

Table 2: Performance of classifier systems

Baseline			Adaboost		
Classifier system	% Accuracy	% ROC area	Classifier system	% Accuracy	% ROC area
LR	74.39	70.2	LR (adaboost)	86.17	71.2
NN	58.20	54.8	NN (adaboost)	70.50	71.9
DT	69.12	61.2	DT A(adaboost)	82.94	83.4
SVM	83.41	71.9	SVM (adaboost)	86.83	82.9

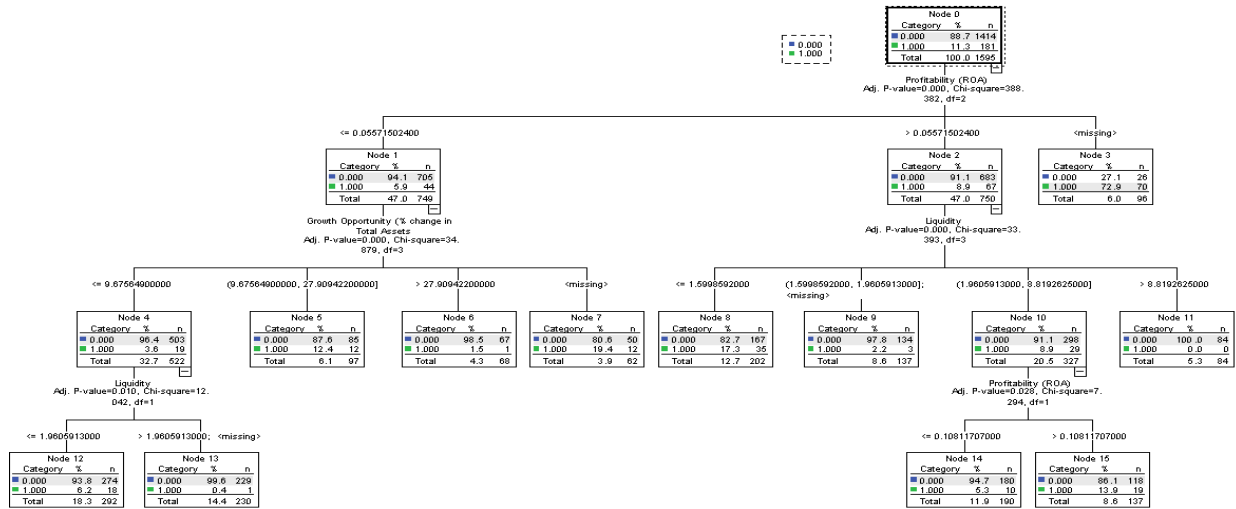
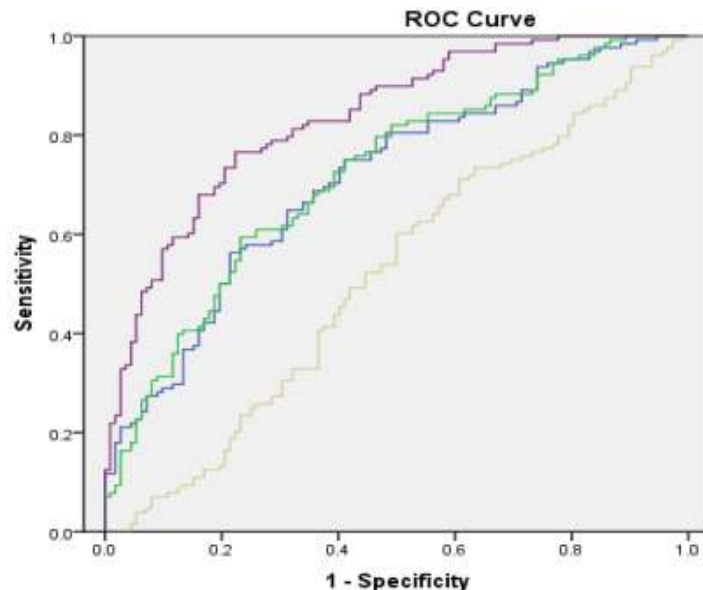
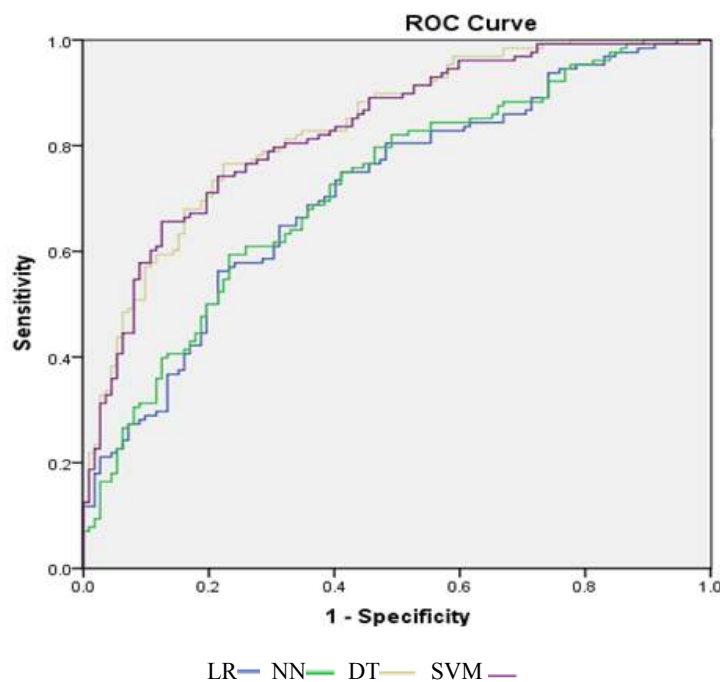


Fig. 3: Decision tree



(a): Baseline



(b): Adaboost

Fig. 4: ROC curve

methods, i.e., AdaBoost ensemble with logistic regression (represented as LA), AdaBoost ensemble with decision tree (represented as DTA), single classifier on DT and single classifier on SVM.

The results are presented in two parts. First part of this section displays the percent of accuracy rate for each classifier system. Then, the enhancement over the baselines has been shown for ensemble classifiers. Table 2 shows the percent of model accuracy and the area under ROC curve for each classifier system. Comparison of forecasting accuracy reveals that the SVM has a lower model risk than other models. According to the results, SVM is the best. The performance of Neural Network is significantly worse than other approaches. Generally, the findings for the baseline classifiers are not predominantly unexpected and are well-matched with previous empirical researches of classifier performance for default risk data sets especially in case of SVM classifier. SVM with a high generalization capacity seems to be a capable technique for default prediction in Malaysia as an emerging economy. Also, Table 2 shows the performance accuracy of multi-stage classifiers in compare with baselines.

The Adaboost classifiers considerably outperform the baseline. By the results, all multi-stage systems outperform the baseline including adaboost with Logistic regression, naïve bayes and neural network, J48 and support vector machines. Roc curve plots the type II error against one minus the type I error. In the case of default prediction in this study, it describes the

percentage of non-defaulting firms that must be inadvertently denied credit (Type II) in order to avoid lending to a specific percentage of defaulting firms (1-Type I) when using a specific model. Figure 4, shows the ROC curve for baseline and Adaboost classifiers.

CONCLUSION

Default prediction takes an important role in the prevention of corporate default, which makes the accuracy of default prediction model be widely concerned by researchers. Appropriate identification of firms 'approaching default is undeniably required. By this time, various methods have been used for predicting default. The use of ensemble classifiers has become common in many fields in the last few years. According to various studies, diverse individual classifiers make errors on different instances (Polikar, 2006; Rokach, 2010). The variety is supposed to improve classification accuracy. According to (Brown *et al.*, 2005; Rokach, 2010), diversity creation can be obtained in several ways and the approaches to classify them vary. The selection of a particular technique can have important consequences on the data analysis and subsequent interpretation of findings in models of credit risk prediction, especially if the quality of data is not good. This study further explores Adaboost ensemble method and makes an empirical comparison. In addition, Adaboost outperforms the other three methods with statistical significance and especially suits for Malaysian listed companies. Therefore, this study

contributes to provide incremental evidence for default prediction research based on AdaBoost and guide the real world practice of DP to some extent. However, this study also has the limitation that the experimental data sets are only collected from Malaysian listed companies and further investigation can be done based on other countries' real world data sets in future study.

REFERENCES

- Abellán, J. and A. Masegosa, 2012. Bagging schemes on the presence of noise in classification. *Exp. Syst. Appl.*, 39(8): 6827-6837.
- Allison, P.D., 2001. *Logistic Regression Using the SAS System: Theory and Application*. SAS Publishing, BBU Press, Cary, NC.
- Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Financ.*, 23(4): 589-609.
- Altman, E.I., 1973. Predicting railroad bankruptcies in America. *Bell J. Econ. Manage. Sci.*, 4(1): 184-211.
- Atiya, A., 2001. Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE T. Neural Networ.*, 12: 929-935.
- Beaver, W.H., 1966. Financial ratios as predictors of failure. *J. Accounting Res.*, 4(3): 71-111.
- Brown, G., J.L. Wyatt, R. Harris and X. Yao, 2005. Diversity creation methods: A survey and categorization. *Inform. Fusion*, 6(1): 5-20.
- Carter, C. and J. Catlett, 1987. Assessing credit card applications using machine learning. *IEEE Expert*, 2: 71-79.
- Dimitras, A.I., S.H. Zanakis and C. Zopounidis, 1996. A survey of business failure with an emphasis on prediction methods and industrial application. *Eur. J. Oper. Res.*, 90: 487-513.
- Freund, Y. and R.E. Schapire, 1997. A Decision-theoretic generalisation of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1): 119-139.
- Frydman, H., E. Altman and D. Kao, 1985. Introducing recursive partitioning for financial classification: The case of financial distress. *J. Financ.*, 40(1): 269-291.
- Gepp, A., K. Kumar and S. Bhattacharya, 2010. Business failure prediction using decision trees. *J. Forecasting*, 29(6): 536-555.
- Gestel, T.V., B. Baesens, P.V. Dijke, J. Suykens, J. Garcia and T. Alderweireld, 2005. Linear and nonlinear credit scoring by combining logistic regression and support vector machines. *J. Credit Risk*, 1(4): 31-60.
- Härdle, W., M.A. Rouslan and D. Schäfer, 2005. Predicting Bankruptcy with Support Vector Machines. SFB 649 Discussion Paper 2005-009, pp: 225-248.
- Hertz, J., A. Krogh and R.G. Palmer, 1991. *The Theory of Neural Network Computation*. Addison Welsey, Redwood, CA.
- Hosmer, D.W. and S. Lemeshow, 2000. *Applied Logistic Regression*. Wiley, New York.
- Lin, F. and S. McClean, 2001. A data mining approach to the prediction of corporate failure. *Knowl-Based Sys.*, 14(3-4): 189-195.
- Lin, Y., 2002. Improvement on behavioural scores by dual-model scoring system. *Int. J. Inf. Tech. Decis.*, 1: 153-165.
- Martin, D., 1977. Early warnings of bank failure: A logit regression approach. *J. Bank. Financ.*, 1: 249-276.
- Messier Jr., W. and J. Hansen, 1988. Inducing rules for expert system development: An example using default and bankruptcy data. *Manage. Sci.*, 34: 1403-1415.
- Morris, R., 1997. *Early Warning Indicators of Corporate Failure: A Critical Review of Previous Research and Further Empirical Evidence*. Ashgate Publishing, Brookfield.
- Myers, J.H. and E.W. Forgy, 1963. The development of numerical credit evaluation systems. *J. Am. Stat. Assoc.*, 58(303): 799-806.
- Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *J. Account. Res.*, 18(1).
- Polikar, R., 2006. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.*, 6(3): 21-45.
- Pompe, P. and A. Feelders, 1997. Using machine learning, neural networks and statistics to predict corporate bankruptcy. *Microcomput. Civil Eng.*, 12: 267-276.
- Quinlan, J.R., 1986. Induction of decision trees. *Mach. Learn.*, 1: 81-106.
- Rokach, L., 2010. Ensemble methods in supervised learning. In: Maimon, O. and L. Rokach (Eds.), *Data Mining Knowledge Discovery Handbook*, 2nd Edn., Springer Science+Business Media, pp: 959-979.
- Shin, K.S. and Y.J. Lee, 2002. A genetic algorithm application in bankruptcy prediction modeling. *Expert Syst. Appl.*, 9: 503-512.
- Shin, K.S., T.S. Lee and H.J. Kim, 2005. An application of support vector machines in bankruptcy prediction model. *Expert Syst. Appl.*, 28(1), 127-135.
- Smith, C.W. and R.M. Stulz, 1985. The determinants of firms' hedging policies. *J. Financ. Quant. Anal.*, 20(4): 391-405.
- Smith, R. and A. Winakor, 1935. *Changes in Financial Structure of Unsuccessful Industrial Corporations*. Bureau of business research. Bulletin Urbna University of Illinois Press, pp: 51.
- West, D., S. Dellana and J. Qian, 2005. Neural network ensemble strategies for financial decision applications. *Comput. Oper. Res.*, 32: 2543-2559.