2012 International Conference on Economics, Business Innovation
IPEDR vol.38 (2012) © (2012) IACSIT Press, Singapore

Validating a Bankruptcy Prediction by Using Naïve Bayesian Network Model: A case from Malaysian Firms

Suzaida Bakar [1 +] and Muhammad Zuhairi Abd Hamid[1]

[1]Department of Finance and Economics, Universiti Tenaga Nasional

Abstract. The purpose of this paper is to propose and validate the Naïve Bayesian model for bankruptcy prediction for the Malaysian firms. First, we suggest heuristic methods that guide the selection of bankruptcy potential variables. Based on the correlations and partial correlations among the variables, the method goal is to eliminate variables that provide little or no additional information beyond that subsumed by the remaining variables. A Naïve Bayesian model is developed using the proposed heuristic method and it is found to perform well based on logistic regression which is used to validate the model. The proposed Naïve Bayesian model consists of three first-order variables and seven second-order variables. Our results show that the proposed model performance is the best fits when the percentage corrects show 91.7%. Finally, the results of this study could also be applicable to business and investors decision making contexts other than bankruptcy prediction model.

Keywords: Bankruptcy prediction, Naive Bayesian model, Logistic regression, financial distress

1. Introduction

In today dynamic economic environment, the number and magnitude of bankruptcy filing are increasing significantly. Even auditors, who have good knowledge about firm's financial position, often fail to make an accurate judgment on firms' going concern conditions [6]. The prediction of financial distress is an important and challenging issue that has served as the impetus for many academic studies over the past three decades [2]. The financial distress and finally bankruptcy can cause some great damages to shareholders, virtual investors, creditors, managers, employers, suppliers of early materials and clients. The aim of this study is to validate bankruptcy prediction of listed companies in Bursa Malaysia Berhad under PN17 and GN3 companies by using Naïve Bayes models. Naïve Bayes model is gaining an increasing popularity as modeling tools for complex problems involving probabilistic reasoning under certainty. Naïve Bayes model are probabilistic graphical models that represent a set of random variables for given problem and the probabilistic relationships between them. The major advantage of naïve Bayes model is that the output is explicitly a probability, which can be easily interpreted. In order to evaluate bankruptcy prediction model, we will consider fifteen predictor variables including liquidity ratios, leverage ratios, profitability ratios and other factors like firm's size and then we use two methods for choosing variables that adopted from Sun and Shenoy[6]. The purpose of this paper is to measure it is possible to predict bankruptcy by using Naïve Bayes model. Some research shows that Altman Z-score model has a certain limitation such as limitation of the test samples is that the samples were small and not proportional to actual bankruptcy rates [12, 13]. Previous study has been shown that there a potential search bias in the variable selection technique used by Altman [10]. The lack of a theory of bankruptcy invites the researcher to consider a multitude of variables and then to reduce the original set to the most accurate subset. A very comprehensive study using a variety of financial ratios has been conducted by Beaver [2]. He concludes that the cash flow to debt ratio was the single best predictor. Later on, a development of Bayesian Network (BN) models for early warning of bank failure by Sarkar and Sriram [9]. They found that both Naïve BN model and a composite attribute BN model have comparable performance to the well known induced decision tree classification algorithm. Due to that,

[+] Corresponding author. Tel.: +6 012 5673641 ; fax: +609 4552006
E-mail address: Suzaida@uniten.edu.my

the main objective of this project paper is to model and validate a bankruptcy prediction model by using Naïve Bayes model and logistic regression and to provide a good method to guide the selection of variables in Naïve Bayes model. The significance of this study will provide a benchmark for most of the investors to prevent from investing in the company that having financial distress and facing with the possibility of bankruptcy. It is hope that this proposed Bayesian Network models can help Bursa Malaysia to produce new bankruptcy model in predicting bankruptcy of the firm. In term of practical contribution to the industry it is hope that it can contribute to the novelty in modeling a new intelligent bankruptcy forecasting system for the firms.

2. Previous studies

Salehi and Abedini[8] has proposed the ability of financial ratios for prediction of financial distress of the listed companies in Tehran Stock Exchange (TES) was investigated. The assessment of the model was done by utilizing the data of two groups. The presented model was according to five the ratios, namely; ratios indicate liquidity, profitability, managing of debt and managing of property. The results of the model indicate the validity of that model and the selected ratios. The results of the test of the ability of model prediction indicate the reality that the model designed four years before financial distress in companies; present a correct prediction about the financial distress. A study by Sun and Shenoy [6] provided operational guidance for building Naı¨ve Bayesian network (BN) models for bankruptcy prediction. A Naı¨ve Bayes model is developed using the proposed heuristic method and is found to perform well based on a 10-fold validation analysis. The results show that the model's performance is the best when the number of states for discretization is either two or three. They experiment whether modeling continuous variables with continuous distributions instead of discretizing them can improve the model's performance. The finding suggests that this is not true. Finally, the results of this study could also be applicable to business decision-making contexts other than bankruptcy prediction. Again, study by Sarkar and Sriram [9] has demonstrated how probabilistic models may be used to provide early warnings for bank failure. The automatic system examines the financial ratios as predictors of bank performance and assesses the posterior probability of bank financial health. Both models are able to make accurate predictions with the help of historical data to estimate the required probabilities. In particular, the more complex model is found to be very well calibrated in its probability estimates. Two different probability models were examined in this context, the Naïve Bayes classification model and a CA model. Although out of two models, the Naïve Bayes model appeared sharper while the CA model was better calibrated.

3. Methodology

This section present in details the data and methodology that are being adopted in this study. The methodology used in this study includes a correlation and partial correlation, heuristic methods; a Naïve Bayes model and logistic regression.The first method is correlation and partial correlation between variables and the second method based on heuristic methods. Then models for predicting financial distress are developed using Naïve Bayes model and logistic regression is uses to measure Naïve Bayes models performance. The Naïve Bayesian classifier has been used extensively for classification because of its simplicity and because it embodies the strong independence assumption that, given the value of the class, the attributes are independent of each other. Sample firm used in this study are companies that listed in Bursa Malaysia Berhad under PN17 and GN3 companies across various industries during the period 2001 until 2010.

3.1. Data

The sources of the data for this study are taken from Bursa Malaysia Berhad and company official website. The data consist of fifteen predictor variables of twenty eight companies under PN17 companies and nine companies under GN3 companies for ten years from 2001 until 2010. To avoid double counting information, we analyze whether one variable is dependent with bankruptcy, given the other variable in the pair by examining the partial correlations between that variable and bankruptcy, while controlling the other variable in the pair. The variables used in this study are X1(Current Assets-Current Liabilities)/Total Assets; X2 Operating Cash Flow/Total Liabilities; X3 Cash/Total Assets; X4 Total Liabilities/Total Assets; X5 Long

Term Debt/Total Assets; X6 Sales/Total Assets; X7 Earnings before Interest and Taxes/Total Assets; X8 Net Income/Total Assets; X9 Net income/Sales; X10 Current Assets/Sales; X11 Current Assets/Total Assets; X12 Current Assets/Current Liabilities; X13 Market Value/Total Liabilities; X14 Natural log of total assets ; X15 Natural log of (total Assets/CPI Index). Grounded on prior feature selection literature [5], the goal is to eliminate variables that provide little or no additional information beyond that subsumed by the remaining variables. To achieve the goal, the proposed heuristic relies on correlations and partial correlations among variables. This heuristic is based on the assumption that the dependence between every pair of variables is linear and measured by the correlation coefficient.

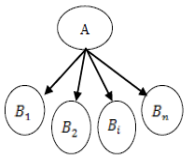

Fig.1: Naïve Bayes Model Nodes

Figure 1 presents a graphical representation of Naïve Bayes model that named by [11] because of its simplicity. In a naïve Bayes model, the node of interest has to be the root node, which means, it has no parent nodes. In a bankruptcy prediction context, A represents the bankruptcy variable. *B1, B2, Bi* and B_n represent *n* bankruptcy predictor variables. The above assumption says that predictors, $B_1, B_2, \ldots, B_n$ are conditionally mutually independent given the state of bankruptcy [6]. Logistic Regression is use to test Naïve Bayes models performance. Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Generally, the dependent or response variable is dichotomous, such as presence or absence or success or failure.

4. Results

4.1. Heuristic Method

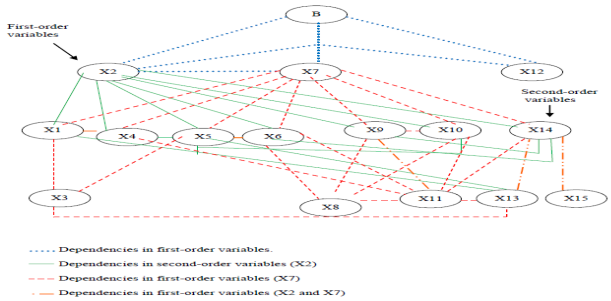

Fig. 2: Dependencies' among Variables

As shown in Fig. 2, we describe how the proposed heuristic works. First, we obtain the correlations among all variables, including fifteen potential predictors and the variable of interest, firms' bankruptcy status (Altman Z-score). Variables that have significant correlations (Pearson correlation coefficient ≤ 0.05) are assumed to be dependent and therefore connected. We use the cutoff of 0.05 to help identify a small subset of most important predictors while excluding the unimportant ones.

4.2. Naïve Bayes Model

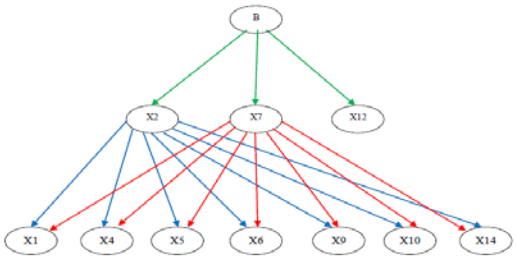

Fig. 3: Proposed Naïve Bayes Model

Figure.3 above shows only three out of fifteen variables are connected with B which is operating cash flow over total liabilities (X2), earnings before interest, tax and depreciation amortization over total assets (X7) and current ratio (X12) because of p-value is less than 0.05. Meanwhile, there are seven variables are selected as second order variables such include X1, X4, X5, X6, X9, X10 and X14. X12 is not connected with any variables in second order variables because the p-value is greater than 0.05 and measures they don't has bidirectional significant relationship even though X12 is connected with B. Based on our assumption, those variables are directly having significant correlation ($P \leq 0.05$) with bankruptcy status will be first order variable and the remaining non-first order variable that have positively significant relationship with first order variable consider as second order variable. Only variable that has significant bidirectional relationship will be remain in second order variable.

4.3. Logistic Regression

The goal of logistic regression is to find the best fitting and most parsimonious, yet biologically reasonable model to describe the relationship between dependent variable and set of independent variables. These independent variables are often called covariates. The difference between a logistic regression model from the linear regression model is that the outcome variable in logistic regression is binary or dichotomous [4]. Based on Table 2 below, we ran a logistic regression model to predict the validity of the Naïve Bayesian model based on variables X1, X2, X4, X5, X6, X7, X9, X10, X12 and X14. From the results given in the Table 2, it can be seen that the predictive capability of the Naïve Bayes model is good where the capacity of the model to predict correctly is 91.7% compared to the threshold value which is only 88.3%. The coefficient indicates that only X12 which is current ratio is significant which the Wald statistic is 11.557. From this result, the logistic regression equation is as follows:

$$\text{ProbBN} = \frac{e^{13.121(0.745*X1)-(0.638*X2)+(0.128*X4)+(2.470*X5)-(0.05*X6)-(0.206*X7)-(0.143*X9)+(1.748*X10)-(1.039*X12)-(2.138*X14)}}{1+e^{13.121(0.745*X1)-(0.638*X2)+(0.128*X4)+(2.470*X5)-(0.05*X6)-(0.206*X7)-(0.143*X9)+(1.748*X10)-(1.039*X12)-(2.138*X14)}}$$

Table 2 Summary of the model

Nagelkarke R^2 = 0.424 ; Cox & Snell R^2 = 0.218; Hosmer & Lemershow Chi Square = 0.561(p-value = 6.776) ; Wald statistic (X12 p-value)* = 11.557(0.001)*; Threshold value= 88.3 % ; Overall Percentage of prediction by the Model = **91.7%**

5. Conclusion

A lot of factors can contribute to bankruptcy problem either directly or indirectly. Ten potential variables were identified as the factors of bankruptcy based on naïve Bayes model. The variables are working capital divided by total assets (X1), operating cash flow over total liability (X2), Total liabilities over total assets (X4), long term debts over total assets (X5) sales over total assets (X6), Earnings before Interest, tax, depreciation and amortization over total assets (X7), net income over sales (X9), current assets over total assets (X10), current ratio (X12) and natural log of total assets (X14). More importantly, the above reported results shows that naïve Bayes models can be as a tool in predicting bankruptcy in Malaysia. Based upon results, we can conclude that our proposed Naïve Bayesian network model is simple to implement and performs well according to the overall percent of cases that are correctly predicted by the model which indicates 91.7% accurate in predicting bankruptcy for Malaysian firms. We would like to recommend Bursa

Malaysia Berhad to use these Naïve Bayes model for listing firm and evaluating them. It is important because to make sure that only qualified company can be listing in Bursa Malaysia and help investors to make a right choice in investing decision. Naïve Bayes model was not only important to investors and Bursa Malaysia Berhad but also financial institutions such as bank. Bank can uses Naïve Bayes model for loan making decisions. The purpose is to avoid any default repayment of loan disbursement.

6. References

[1] Altman, E. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance 23*, 1968 (September), pp. 589–609.

[2] Beaver, W.H. Financial ratios as predictors of failure. *Journal of Accounting Research 4 (Suppl.)*, 1966, pp.71–111.

[3] H. Etemadi, A.A. Rostamy and H.F Dehkordi. A Genetic Programming Model for Bankruptcy Prediction: Empirical Evidence from Iran. *Expert Systems Applications 36*, 2009, pp. 3199-3207.

[4] Hosmer DW, Lemeshow S. *Applied Logistic Regression. 2*. New York, USA: John Wiley and Sons; 2000.

[5] Koller, D., Sahami, M., 1996. Toward optimal feature selection. *In: Proceedings of the Thirteenth International Conference in Machine Learning*. Morgan Kaufmann Publishers, San Francisco, CA, pp. 284–292.

[6] L. Sun and P. Shenoy. Using Bayesian Networks for Bankruptcy Prediction. *European Journal of Operational Research,* Vol. 180, Issue 2, 2007, pp. 738-753.

[7] R.Estevam, E.R. Hruschka, and Ebecken, N.F.F. Towards Efficient Variables Ordering for Bayesian Networks Classifier. *Data & Knowledge Engineering*, Vol. 63, 2007, pp. 258–269.

[8] Salehi, M & Abedini, B. Financial Distress Prediction in Emerging Market: Empirical Evidences from Iran. *Journal of Contemporary Research in Business,* Vol. 1(1), 2009, pp. 6-26.

[9] Sarkar, S & Sriram, R. Bayesian Models for Early Warning of Bank Failures. M*anagement Science*, 2001, pp .1457-1475.

[10] Scott J. The probability of bankruptcy: a comparison of empirical predictions and theoretical models. *Journal Banking Finance; 5*, 1981(September)., pp. 317– 44

[11] Titterington, D.M., Murray, G.D., Murray, L.S., Spiegelhalter, A.M., Skene, A.M., Habbema, J.D.F., Gelpke, G.J., Comparison of discrimination techniques applied to a complex data-set of head-injured patients (with discussion). *Journal of the Royal Statistical Society*, Series A 144, 1981, pp.145–175.

[12] Zmijewski ME. Essays on corporate bankruptcy. *PhD Dissertation,* State University of New York at Buffalo, 1983.

[13] Zmijewski, M. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research 22 (Suppl.)*, 1984, pp. 59–82.