Quest for Research Excellence On Computing, Mathematics and Statistics

Editors

Kor Liew Kee

Kamarul Ariffin Mansor

Asmahani Nayan

Shahida Farhan Zakaria

Zanariah Idrus

Faculty of Computer and Mathematical Sciences

Quest for Research Excellence on Computing, Mathematics and Statistics

Chapters in Book

The 2nd International Conference on Computing, Mathematics and Statistics (iCMS2015)

Editors:

Kor Liew Lee
Kamarul Ariffin Mansor
Asmahani Nayan
Shahida Farhan Zakaria
Zanariah Idrus

Quest for Research Excellence on Computing, Mathematics and Statistics

Chapters in Book

The 2nd International Conference on Computing, Mathematics and Statistics

(iCMS2015)

4-5 November 2015
Langkawi Lagoon Resort
Langkawi Island, Kedah
Malaysia

Publication by
Faculty of Computer & Mathematical Sciences
UiTM Kedah

ISBN 978-967-0314-26-6

Content

CHAPTER 25
Data Mining in Predicting Firms Failure: A Comparative Study Using Artificial Neural Networks and Classification and Regression Tree

Norashikin Nasaruddin, Wan-Siti-Esah Che-Hussain,
Asmahani Nayan, and Abd-Razak Ahmad

Abstract. Financial Institutions and investors alike are very much interested in the accuracy of predicting the potential failures of firms. These financial institutions believe accurate prediction will lead to a low default rate in servicing their financial loans. The aim of this study is to find a better model to classify firms that is more likely to fail. Bad prediction model will lead to a high default rate. Using financial and non-financial information, this paper illustrates the construction and comparison of two models – artificial neural networks (NN) and classification and regression tree (CART) models to classify the failed from the non-failed firms. This study found that based on the training sample's result (NN = 94.03% & CART = 94.69%) the overall accuracy result of CART is higher than the NN model. Similar result can be drawn for the validation sample with CART leading at 92.93% overall accuracy rate compared to NN's 91.92%.

Keywords: data mining; artificial neural networks; regression tree; firms failure

Norashikin Nasaruddin (✉) • Wan-Siti-Esah Che-Hussain • Asmahani Nayan • Abd-Razak Ahmad
Universiti Teknologi MARA (UiTM) Kedah
e-mail: norashikin116@kedah.uitm.edu.my, wsech569@kedah.uitm.edu.my, asmahanin@kedah.uitm.edu.my, ara@kedah.uitm.edu.my

1 Introduction

The financial health of a company is one important factor that determines whether a company is eligible for securing a loan. Thus predicting the failure of a business entity is essential and has become the focus of many researches. Financial institutions use various statistical and non-statistical models to predict the chances firms fail. The core objective is to obtain the best model that can predict firm's failure or bankruptcy to the highest degree. However, in the real world, there are different types of businesses with different sets of information that can be used as indicators to identify financially distressed firms. Due to this, it creates a limitation on the approach to develop a holistic prediction model of bankruptcy that can be used for all. For that reason each prediction model developed by the researchers is unique in their scope and specialization of business [1]. They have their own strengths and weaknesses. Over the years, researchers had been developing new models using new methods with significant financial distress indicators in order to produce the best prediction model [2] and [3].

In this paper, two methods to assess the possibility of a firm fails were used. The methods are the CART and the neural networks models. These models were chosen as each one of the models has a high rate of acceptable prediction in previous studies. Thirty-three financial data including financial ratios that are commonly used in bankruptcy studies were used in this study to measure the financial situation of companies grouped under the consumer product industry category by the Companies Commission of Malaysia (Suruhanjaya Syarikat Malaysia – SSM).

The paper is divided into five sections. Section 2 reviews previous research on the models of bankruptcy prediction. Section 3 describes the methodologies. In Section 4, we discuss the analysis of the result. Finally, Section 5 concludes.

2 Literature Review

There are a number of key models that have been developed by various researchers and used in the bankruptcy prediction studies in the past century. The first significant type of bankruptcy prediction model is univariate analysis by Fitzpatrick [4]. Over the years, Multiple Discriminant Analysis (MDA) becomes more popular and replaces univariate analysis as a tool in the study of firms' failure. In 1968, Altman published the Z score formula using multiple discriminant analysis for predicting bankruptcy accurately [3]. This statistical model that combines five financial ratios to produce a product called a Z-score has become the best known predictor of bankruptcy. Altman's Z score model has the ability to predict distress and bankruptcy up to two to three years in advance. The next type of prediction models is the logic/probit analysis. There were a few researches done using the logic/probit model [5] and [6]. The Decision Tree and the Artificial Neural Network (NN) are two of the later day bankruptcy prediction models. One of the first studies to apply NNs to the bankruptcy prediction problem was the work by Odom and Sharda [7]. Their work showed that the degree of prediction was very encouraging with an 85% overall accuracy rate. Many banks are using default prediction products that are based on neural network such as Moody's Public Firm Risk Model to determine the financial situation of a firm before giving out loan.

Fundamentally, there are two main types of methods in developing firms' failure prediction models. The first one is the statistical techniques among which are the regression analysis, correlation analysis, discriminant analysis, the logit model, and the probit model. The second techniques involved computational intelligence such as the artificial neural networks (ANNs), support vector machines, genetic programming techniques and data mining. As such, a lot of studies have been done on comparing the accuracy of predicting corporate failure using both machine learning techniques and the statistical approach. Which model becomes the most accurate model to predict bankruptcies depends mostly on the types of business or firm nature being investigated.

A study by Jae and Chulwoo, compared the multi-discriminant analysis, logistic regression, decision tree and artificial neural network model with binary classification model to predict corporate failure based on genetic algorithm [8]. This study used financial ratios of 2542 externally audited small and medium-sized manufacturing representing bankrupt and non-bankrupt companies from 2001 to 2004. The prediction accuracy model showed that for the training sample, the classification accuracy of the proposed model is slightly lower than the neural network model while the classification accuracy for the validation sample shows that the classification accuracy of the proposed model is the best. Another study used three credit scoring models, logistic regression, CART and artificial neural network to classify applicants or borrowers into different level of loan eligibility [9]. The eligibility of applicants to borrow money will depend on their financial situation. The results indicate that artificial neural network predicts slightly better in the classification accuracy compared to logistic regression or CART. And recently, David, Dunsun & Yanyan did a comparative study that also used data mining models to predict bankruptcy [10].

A study by Yap, Ong & Nor compared three credit scoring models which were scorecard model, logistic regression and decision tree model to the members of a recreational club, it was found that the classification error rates for the three models are about the same for all three models [11]. With respect to that factor, they concluded that both scorecards and decision trees are superior predictor over logistic regression model as the two models were more applicable and easier to understand. Another study has done by Guangli et al using the logit model and decision tree to predict credit card churn in China's banking industry [12]. This study used misclassification cost measurement by taking two types of error and economic sense to evaluate the two classification models. The result shows that the logit model is slightly better than decision tree in predicting credit card churn.

3 The Methods

The data used in this study are failed and non-failed firms categorized under the consumer product industry category by the Companies Commission of Malaysia (Suruhanjaya Syarikat Malaysia – SSM). The sample consists of five years data of 130 companies whereby there are 118 (90.8%) non-failed and 12 (9.2%) failed firms. The sample data was divided into two parts called training and validation sample. The training sample (70%) data is used to build the models, while the validation sample (30%) data is for the validation of the models. Fig. 1 depicts the data modeling process using SPSS Clementine.

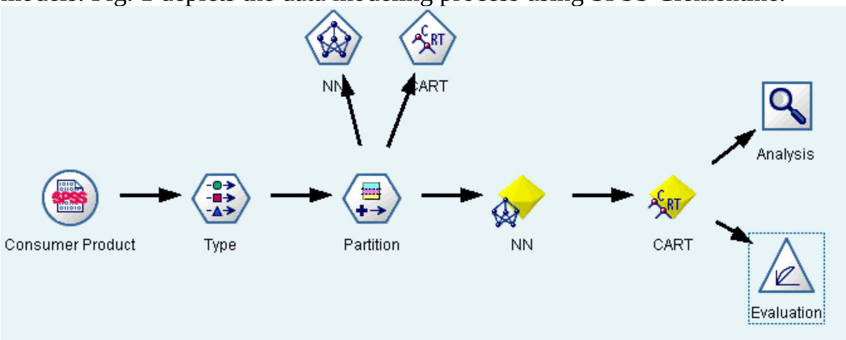

Fig. 1. Data mining process flow diagram.

The process starts with a ‘SPSS’ node represent the Consumer Product data. The two pentagon-shaped nodes represent the construction of the models using decision trees (CART) and neural network. Finally, the two predictive models were then connected to the ‘Analysis’ node which generate the computation for accuracy rates while the ‘Evaluation’ nodes produces the lift charts.

4 Results

The results of the analysis for both methods are discussed below.

4.1 Regression Tree (CART)

A decision tree is easy to understand and has the ability to be converted to a set of rules. Furthermore, decision tree model can classified categorical and numerical data without any prior assumptions to be met. The CART model finds three variables to be influential on the company status and the decision tree rules are listed in Table 2 while Fig. 2 shows the CART model.

Table 8. CART rules.

Healthy	Delisted
• Operating margin is greater than -44.947 • Book value per share is less than or equal 8.350	• Operating margin is less than or equal -44.947
• Operating margin is greater than -44.947 • Book value per share is greater than 8.350 • Prefix last is greater than 7.032	• Operating margin is greater than -44.947 • Book value per share is greater than 8.350 • Prefix last is less than or equal 7.032

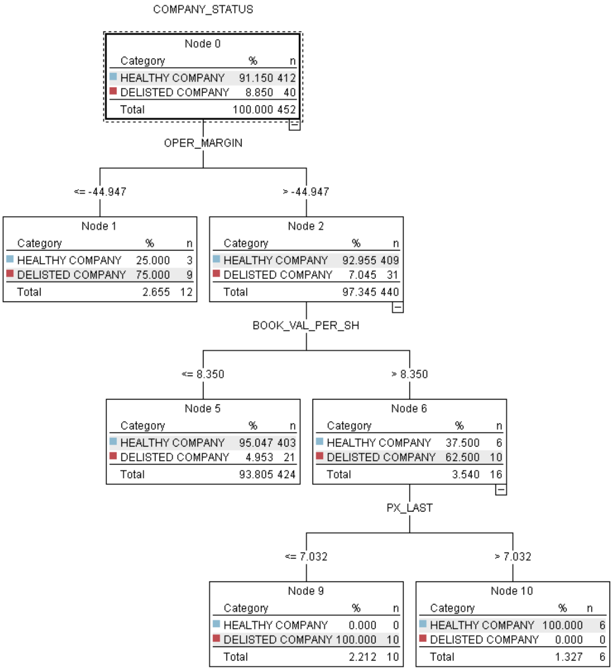

Fig. 2. CART Model

4.2 Neural Network

The neural network, on the other hand, is an interconnected group of nodes. To classify a company using a Neural Network model, a back-propagation network with the following features was developed:

1. *Thirty three input variables:* Operating Income per Share, Book Value per Share, Net Debt to Shareholder Equity, Total Debt to Total Equity, Inventory to Sales, Sales Revenue Turnover, Operating Margin, Balance Sheet Current Asset Report, Total Debt to Total Capital, Debt to Market Capital, Profit Margin, Balance Sheet Total Asset, Total Debt to Total Asset, Operating Income, Inventory to Total Asset, Net Income, Last Price, Net Fix Asset Turnover, Capital Expenditure to Sales, Working Capital, Balance Sheet Total Liability, Asset Turnover, Sales Growth, Inventory to Current Assets, Return on Asset, Net Debt, Price/Book Value per Share, Net Change

Total Equity, Return on Company Equity, Current Ratio, Price/T12M Cash Flow per Share, Operating Income to Total Debt, Cash Flow per Share.

2. *One output variable:* Company status that takes two values. 1 for delisted and 0 for healthy company.

The simple illustration of neural network is presented in Fig. 3. The neural networks model has 32 neurons in the input layer, 3 neurons for hidden layer and 1 neuron for the output layer.

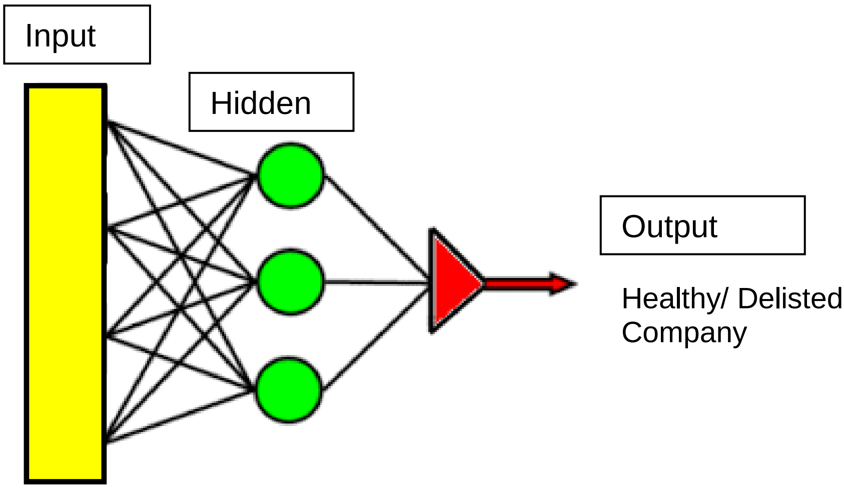

Fig. 3. Simple Representation of back-propagation neural network.

In neural networks, no mathematical model is produced and only the importance of the input variables in descending order is given as shown in Table 3. The top three important variables are Book Value per Share, Net Debt to Shareholder Equity and Total Debt to Total Equity.

Table 2. Relative Importance of Input Variables.

Book Value per Share	0.1665
Net Debt to Shareholder Equity	0.1510
Total Debt to Total Equity	0.1425
Sales Revenue Turnover	0.0805
Inventory to Sales	0.0740

Cash Flow per Share	0.0636
Operating Margin	0.0627
Profit Margin	0.0350
Balance Sheet Current Asset Report	0.0232
Total Debt to Total Asset	0.0201
Net Income	0.0182
Operating Income	0.0170
Total Debt to Total Asset	0.0141
Last Price	0.0133
Inventory to Current Assets	0.0125
Balance Sheet Total Liability	0.0121
Net Fix Asset Turnover	0.0106
Net Change Total Equity	0.0102
Return on Company Equity	0.0081
Inventory to Total Asset	0.0075
Price/Book Value per Share	0.0075
Debt to Market Capital	0.0073
Working Capital	0.0065
Capital Expenditure to Sales	0.0064
Asset Turnover	0.0055
Sales Growth	0.0054
Net Debt	0.0041
Price/T12M Cash Flow per Share	0.0038
Current Ratio	0.0038
Balance Sheet Total Asset	0.0029
Operating Income to Total Debt	0.0022
Return on Asset	0.0019

4.3 Model Comparison

A comparison between these two models was made to determine a more predictive model. The overall accuracy rates for training and validation samples are given in Table 3. The result shows that CART performs better than the Neural Network model in both the training and validation samples.

Table 3. Accuracy rate.

Model	Training	Validation
CART	94.69%	92.93%
Neural Network	94.03%	91.92%

5 Conclusions

As a conclusion, the CART model was found to predict better than the neural network model both during the training and also validation process. However, both models achieve very high accuracy rates that signify their high predictive ability. This study also revealed that whichever model to use in predicting bankruptcy would eventually relies more towards the input variables. Book Value per Share, Net Debt to Shareholder Equity and Total Debt to Total Equity were the outstanding input variables for the neural network model while Book Value per Share, Operating Margin and Prefix Last in the CART model. These six relatively important variables were refined from a set of thirty three input variables initially used in the testing.

Acknowledgements.

The authors appreciate the financial support funded by the Ministry of Higher Education of Malaysia through RAGS grant.

References

[1] Adnan, A., Humayon, A.: Predicting corporate bankruptcy: where we stand?. [online], Available from EMERALD, 18(2006)

[2] Rahimpoor, A., Ansari, A., Alinezhad, S.M.: A Comparative Study of Bankruptcy Prediction Models of Fulmer and Springate in Accepted Companies in Tehran Stock Exchange. Caspian Journal of Applied Sciences Research, 1(13), 76-81(2012)

[3] Altman, E.: Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance, 23(4), 589-609 (1968)

[4] Fitzpatrick, P.J.: A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Companies. Certified Public Accountant, 598-605, 656-662 & 721-731(1932)

[5] Ohlson, J.: Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, 18(1), 109-131(1980)

[6] Zmijewski, M.E.: Methodological Issues Related to the Estimation of Financial Distress Prediction Models. Journal of Accounting Research, 22, 59-82(1984)

[7] Odom, M., Sharda, R.: A Neural Network Model for Bankruptcy Prediction. In: Proceedings of International Joint Conference on Neural Networks, pp. 163-168. (1990)

[8] Jae, H.M., Chulwoo, J.: A Binary Classification Method for Bankruptcy Prediction. Expert Systems with Applications, 36(3), 5256-5263(2009)

[9] Yap, B.W., Irma, R. I.: Using Data Mining Predictive Models to Classify Credit Card Applicants. In: IEEE the IMS International Conference Proceedings, pp. 394-398. (2010)
[10] David, L. O., Dursun, D., Yanyan, M.: Comparative Analysis of Data Mining Methods for Bankruptcy Prediction. Decision Support Systems, 52(2), 464-473(2012)
[11] Yap, B. W., Ong, S. H., Nor, H. M.: Using Data Mining to Improve Assessment of Credit Worthiness via Credit Scoring Models. Expert Systems with Applications, 38, 13274-13283(2011)
[12] Guangli, N., Wei, R., Lingling, Z., Yingjie, T., Yong, S.: Credit Card Churn Forecasting by Logistic Regression and Decision Tree. Expert Systems with Applications, 38(12), 15273-15285(2011)

CHAPTER 26
Risks of Divorce: Comparison between Cox and Parametric Models

Sanizah Ahmad, Norin Rahayu Shamsuddin, Nur Niswah Naslina Azid @ Maarof, and Hasfariza Farizad

Abstract. Divorce or also known as the dissolution of marriage occurs when the bond of matrimony between married couples is dissolved. Since the rate of divorce is on the rise all around the world, this study aims to identify potential risk factors contributing to divorce by making comparisons between Cox Proportional Hazards (PH) model (a semi-parametric method) with Weibull and Lognormal models (parametric methods) using survival data. We retrospectively studied 531 secondary data of the Muslim couples who filed for divorce in Selangor, Malaysia. The age at marriage of husband and wife, the presence of children, duration of marriage, couples' educational level and employment status, household income and counseling session were identified as potential risk factors. The AIC (Akaike Information Criterion) were used to compare the efficiency of models between the three methods. The Cox PH model gives the best fit with respect to the lower AIC value. The survival result from the Cox model showed that age at marriage of husband and attending counseling session significantly affect the decision to divorce.

Keywords: akaike information criterion (AIC); Cox proportional hazards; divorce; weibull; lognormal

Sanizah Ahmad (✉) • Norin Rahayu Shamsuddin • Nur Niswah Naslina Azid @ Maarof • Hasfariza Farizad
Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA.
e-mail: sanizah@tmsk.uitm.edu.my, norinrahayu@kedah.uitm.edu.my, niswah@kelantan.uitm.edu.my

iCMS 2015

ISBN 978-967-0314-25-6