View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Universiti Utara Malaysia: UUM eTheses

ANALYSIS OF BANKRUPTCY USING DATA MINING APPROACH

ONG AI PING

UNIVERSITI UTARA MALAYSIA

2009

ANALYSIS OF BANKRUPTCY USING DATA MINING APPROACH

A project submitted to the Faculty of Information Technology in partial

fulfillment of the requirement for the degree

Master of Science (Information Technology)

Universiti Utara Malaysia

By

ONG AI PING

© Ong Ai Ping, 2009.

All Rights Reserved.

PERMISSION TO USE

In presenting this project in partial fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor, in her absence, by the Dean of the Faculty of Information Technology. It is understood that any copying or publication or use of this project or parts thereof for financial gain should not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my project.

Request for permission to copy or to make use of material in this project, in whole or in part should be addressed to:

Dean of the Faculty of Information Technology

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman

ABSTRAK

Kajian ini berkaitan dengan pembangunan model ramalan rangkaian neural untuk syarikat yang ingin membuat ramalan tahap syarikat sama ada akan menghadapi kebankrapan. yang Dalam pada itu, Jumalah 367 data set adalah diperolehi daripada *Kuala Lumpur Stock Exchange* (KLSE) and Bank Negara Malaysia. Data ini seterusnya dianalisis dengan menggunakan asas statistic, *frequency* dan *cross tabulation* untuk mendapatkan lebih banyak maklumat berkaitan data. Pada peringkat awal, data adalah diklasifaikan dengan menggunakan *logistic regression.* Seterusnya ianya ditrain dengan rangkaian neural untuk mendapatkan model kebankrupan. Dimana, capaian menunjukkan adalah lebih sesuai dengan model yang mengandungi 12 nod *input*, 6 nod *hidden layer* dan 1 nod untuk *output*. Model yang dipilih menunjukkan generalisasi 100%. Metodologi ini sepatutnya memperolehi pendekatan baru kepada paten yang wujud dalam data ini. Oleh itu, rangkaian neural amat berpotensi untuk menyokong ramalan kebankrupan ini.

ABSTRACT

This study involves the development of neural network prediction model to predict the stage of bankruptcy of a company. A total of 367 data was attained from the Registrar of Business and Companies, Kuala Lumpur Stock Exchange (KLSE) and Bank Negara Malaysia (Central Bank of Malaysia). The data was then analyzed by considering the basic statistics, frequency and cross tabulation in order to get more information about the data. Initially, the data was classified using logistic regression. In addition, it was also trained using neural network in order to obtain the bankruptcy model. The findings show that the most suitable prediction model consist of 12 nodes of input , hidden layer 6 node and one output layer. The generalization performance of the selected model is100%. This methodology should be able to provide some new insight into the type of pattern that exists in the data. Thus, neural network has a great potential in supporting for predicting bankruptcy.

ACKNOWLEDGEMENT

Sincere thanks are due to my both supervisors, Associate Professor Fadzilah Siraj and Puan Nur Azzah Binti Abu Bakar for their patiently navigating and generously sharing their rich source and knowledge with me.

Evenly thanks to my course mate that zealously advice me and helping me all over this project.

Further gratitude must go to my beloved family members who have supporting and encourage me during completing this project.

Last and not least, I would like to express my deep appreciation to all who are helped me to completed this project whether in direct or indirect way.

Thanks

TABLE OF CONTENTS

PERMISSION TO USE **I**

ABSTRAK **ii**

ABSTRACT **iii**

ACKNOWLEDGEMENT **iv**

TABLE OF CONTENTS **v**

LIST OF TABLES **ix**

LIST OF FIGURES **x**

LIST OF ABBREVIATIONS **xii**

CHAPTER 1: INTRODUCTION **1**

1.1 Background 1

1.2 Problem Statement 3

1.3 Research Objectives 4

1.4 Scope of the Study 4

1.5 Research Question 5

1.6 Significance of the Study 5

1.7 Conclusion 5

CHAPTER 2: LITERATURE REVIEW **7**

2.1 Factor analysis of Finance and Banking 7

2.2 Bankruptcy Prediction Modeling and Technique 9

2.3 Data Mining 14

2.4 Business using Data Mining 17

2.5 Loan using Data Mining 20

2.6 Banking using Data Mining 21

2.7 Other Area using Data Mining 22

2.8 Conclusion 24

CHAPTER 3: METHODOLOGY **25**

3.1 Business understanding 27

3.2 Data understanding 27

3.3 Data preparation 27

3.3.1 Data Preparation 28

3.3.2 Data Collection 28

3.3.3 Data cleansing 29

3.3.4 Data selection 30

3.3.5 Data Preprocessing 31

3.4 Modeling 33

3.5 Evaluation 34

3.6 Deployment 34

3.7 Conclusion 35

CHAPTER 4: RESULTS OF THE STUDY 36

4.1 Attribute of Bankruptcy Model 36

4.1.1 Frequency of Attribute 40

4.2 Descriptive Data Mining Approach 47

4.2.1 Cross Tabulation 47

4.2.2 Scatter Plot 55

4.3 Correlation 57

4.4 Logistic Regression 59

4.4.1 Examining the Variables 59

4.4.2 Case Processing Summary 61

4.4.3 Omnibus Tests of Model Coefficients 62

4.4.4 Variables in the Equation 62

4.4.5 Model Summary 64

4.4.6 Classification 64

4.4.7 Final Accuracy 65

4.5 Neural Networks 66

4.5.1 Neural Network Tool 66

4.5.2 The Experiments 73

4.5.2.1 To determine the most suitable number of hidden units 74

4.5.2.2 To determine the most suitable learning rate 77

4.5.2.3 To determine the most suitable Momentum Rate 79

4.5.2.4 To determine the best Activation Function 80

4.6 Conclusion 82

CHAPTER 5: CONCLUSION **83**

5.1 Conclusion 83

5.2 Limitation 83

5.3 Recommendation 84

REFERENCES **85**

LIST OF TABLES

Tables	Pages
Table 3.1: Input Features	28
Table 3.2: Ratios with the data sign	30
Table 3.3: Relationship between internal and external variables with risk of bankruptcy	30
Table 3.4: Binary and Symbolic Representation	32
Table 4.1: Type of Attributes	37
Table 4.2: Ratio of Attributes	38
Table 4.3: The missing attribute	39
Table 4.4: Number of Attributes	39
Table 4.5: Target	47
Table 4.6 Correlations	57
Table 4.7: Case Processing Summary	61
Table 4.8: Omnibus Tests of Model Coefficients	62
Table 4.9: Variables in the Equation	62
Table 4.10:Model Summary of accuracy	64
Table 4.11: Classification Table(a,b)	65
Table 4.12: Classification Table (a)	65
Table 4.13: The training, validation and test results using various number of hidden units	74
Table 4.14: The Weight seed using various number of hidden units	75
Table 4.15: The number of hidden units using various number of epoch	75
Table 4.16: The weight seed using various number of hidden units	76
Table 4.17: The training and test results using various number of learning rate	78
Table 4.18: The weight seed using various number of learning rate	78
Table 4.19: The training and test results using various momentum rate	79
Table 4.20: The weight seed using various number of momentum rates	80
Table 4.21: Result to determine the best Activation Function	81

LIST OF FIGURES

Figures	Pages
Figure 3.1: Phases of CRISP-DM Reference Model	26
Figure 3.2: Node and Layers of Neural Network	34
Figure 4.1: Working Capital	40
Figure 4.2: Retained Earning	40
Figure 4.3: Earning before Income Tax	41
Figure 4.4: Total Sales	42
Figure 4.5: Total Debts	42
Figure 4.6: Type of Industries	43
Figure 4.7: Gross Domestic Product	43
Figure 4.8: Age	44
Figure 4.9: Size	45
Figure 4.10: Bank Rate	45
Figure 4.11: Inflation Rate	46
Figure 4.12: Target	46
Figure 4.13: Cross Tabulation of Working Capital	48
Figure 4.14 : Cross Tabulation of Retained Earning	49
Figure 4.15 : Cross Tabulation of Earning before Income Tax	49
Figure 4.16 : Cross Tabulation of Total Sales	50
Figure 4.17 : Cross Tabulation of Total Debts	51
Figure 4.18 : Cross Tabulation of Type of Industries	51
Figure 4.19 : Cross Tabulation of Gross Domestic Product	52
Figure 4.20 : Cross Tabulation of Age	53
Figure 4.21 : Cross Tabulation of Size	53
Figure 4.22 : Cross Tabulation of Bank Rate	54
Figure 4.23 : Cross Tabulation of Inflation Rate	55
Figure 4.24 : Scatterplot Diagram	56
Figure 4.25 : Correlation	58
Figure 4.26: Binary Logistic	60
Figure 4.27: Dialog Box of Logistic Regression	61

Figure 4.28: Neural Connection	67
Figure 4.29: Data Viewer	68
Figure 4.30: Data Allocation	68
Figure 4.31: Denoted of Data Allocation	69
Figure 4.32: Multilayer Perceptron Network (MLP)	70
Figure 4.33: MLP Training Stages	72
Figure 4.34: Confusion matrix for output	77

LIST OF ABBREVIATIONS

ANN	Artificial Neural Networks
BPN	Back-Propagation Neural Network
CRISP	Cross Industry Standard Process
GA-BP	Genetic Algorithm And Back Propagation
KDD	Knowledge Discovery In Databases
KLSE	Kuala Lumpur Stock Exchange
LRA	Logistic Regression Analysis
LSA	Latent Semantic Analysis
MDA	Multivariate Discriminant Analysis
MLP	Multi-Layer Perceptron
PLSA	Probabilistic Latent Semantic Analysis
RBAC	Role-Based Access Control
ROB	Registrar Of Business
ROC	Registrar Of Companies
RST	Rough Set Theory
SVM	Support Vector Machines

CHAPTER 1

INTRODUCTION

This study focuses on using data mining approach for analysis of bankruptcy. The aim of the study is to alert and give the warning signs in earlier stage to the company's that facing financial problem and almost to bankruptcy.

1.0 Background

Bankruptcy refers to the firms which are unable to pay debts and are either declared bankrupt in terms of Commercial Code (1857, Part 111, Title 1) or dissolved and wound under the Companies Act 1995 (Section 35(c) and Section 214 subsection 2 paragraph (a)(ii)). Bankruptcy is a proceeding in which a court administers the estate (i.e. the property and other assets) of the debtor for the benefit of the creditors. A debtor (i.e. a person or business who owes money to others) may choose to file a bankruptcy proceeding to resolve a hopeless financial situation, or to stave off the collection of debts for a period of time to allow for financial re-organization.

Recently, the number of companies declared as bankrupt due to the recession has increased. Therefore, the development of the bankruptcy prediction model has been considered as important, as bankruptcy prediction can have major impact on lending

The contents of the thesis is for internal user only

REFERENCE

Ahmed, K. M., El-Makky, N. M. and Taha, Y. (1998). Effective data mining: a data warehouse-backboned architecture

Abdelwahed, T. and Amir, M. (2005). New evolutionary bankruptcy forecasting model based on genetic algorithms and neural networks

Agarwal, S., Ambrose, B. W., & Chomsisengphet, S. (2005). Asymmetric information and the automobile loan market.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Preceding of The Journal of Finance, 23, 589-609.

Altman, E. (1983). Corporate financial distress - a complete guide to predicting avoiding and dealing with bankruptcy. New York: Willey.

Angelidis, D. and Lyroudi, K. (2006). Efficiency in the Italian banking industry: data envelopment analysis and neural networks

Apte, C., Liu, B., Pednault, E. P.D., Smyth, P. (2002). Business applications of data mining. Proceeding of Communications of The ACM Vol. 45, No. 8

Bloemer, J., Brijs, T., Swinnen, G., Vanhoof, K. (2002). Identifying latently dissatisfied customers and measures for dissatisfaction management. Proceeding of International Journal of Bank Marketing. pp. 27-37.

Cavaretta, M. (2006). Data Mining Challenges in the Automotive Domain

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Thomas, R., Shearer, C., & Wirth, R., (2000). CRISP-DM 1.0 Step-by-step data mining guide.

Charalambous, C. and Martzoukos, S. H. (2005). Hybrid Artificial Neural Networks for Efficient Valuation of Real Options and Financial Derivatives

Chen, R. S., Wu, R. C. & Chen, J. Y. (2005). Data Mining Application in Customer Relationship Management Of Credit Card Business. Proceedings of the 29th Annual International Computer Software and Applications Conference. Institute of Information Management, National Chiao Tung University, Taiwan.

Chen, Y., Tsai, F. S., and Chan, K. L. (2007). Blog Search and Mining in the Business Domain proceeding of ACM SIGKDD Workshop on Domain Driven Data Mining

Cipollini, A. and Missaglia, G. (2005). Business cycle effects on capital requirements: scenario generation through Dynamic Factor analysis

Cumby, C., Fano, A., Ghani, R. and Krema, M. (2004). Predicting Customer Shopping Lists from PointofSale Purchase Data

Dass, R. (2007).Data Mining In Banking And Finance: A Note For Banker. Indian Institute of Management Ahmedabad

Ettl, M., Zadrozny, B., Chowdhary, P., & Abe, N. (2005). Business Performance Management System for CRM and Sales Execution. International Workshop on Business Process Monitoring & Performance Management. IBM T.J. Watson Research Center, USA.

Foster, D. P. and Stine, R. A. (2001). Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy

Fang, R. and Tuladhar, S. (2004). Teaching Data Warehousing And Data Mining In Graduate Program Of Information Technology. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)

Ghani, R. and Soares, C. (2006). Data Mining for Business Applications. Proceeding of KDD-2006 Workshop

Goodarzi, A., Kohavi, R., Harmon, R., Senkut, A. (1998). Loan Prepayment Modeling. American Association for Artificial Intelligence.

He, D. (2002): Resolving Non-performing Assets of the Indian Banking System.

Hekanaho, J., Back, B., Sere, K. and Laitinen, T. (1998). Analysing Bankruptcy Data with Multiple Methods

Hueglin, C. and Vannotti, F. (2001). Data Mining Techniques to Improve Forecast Accuracy in Airline Business

Hunziker, P., Maier, A., Nippe, A., Tresch, M., Weers, D., & Zemp, P. (1999). Data mining at a major bank: Lessons from a large marketing application. Credit Suisse.

Ikizle, N., & Guvenir, H. A. (2002). Mining Interesting Rules in Bank Loans Data. Bilkent University,Department of Computer Engineering.

Kalos, A. and Rey, T. (2005) Data Mining in the Chemical Industry

Kuhlmann, M., Shohat, D. and Schimpf, G. (2003). Role Mining - Revealing Business Roles for Security Administration using Data Mining Technology

Lei, H., Chan, C.C.(2003). Rule-Based Classifier for Bankruptcy Prediction

Liao, S. H., & Chen, Y. J. (2004). Mining customer knowledge for electronic catalog marketing. Expert Systems with Applications. 27. pp.521–532. Department of Management Sciences, Decision Making, Tamkang University.

Lin, Z. and Wu, J. (2005) Research on Audit Informatization under the Environment of E-business.

Lo, V. S.Y. (2002) The True Lift Model - A Novel Data Mining Approach to Response Modeling in Database Marketing Proceeding of SIGKDD Explorations, 4(2), p.p. 78-86

Marchiori, E. (2002). Data Mining. Free University Amsterdam.

McLaren, I. (1999).Designing the Data Warehouse for Effective Data Mining. Searchspace Limited.

Mierzejewski, F. (2006): Economic capital allocation under liquidity constraints. Published in: Proceedings of the 4th Actuarial and Financial Mathematics Day (2006): pp. 107-116.

Mitra, S., Pal, S. K., Mitra, P. (2002). Data Mining in Soft Computing Framework:A Survey. Proceedings of IEEE Transactions on Neural Networks. v.13. pp3-14.

Mody, A., & Patro, D. (1996). Methods of Loan Guarantee Valuation and Accounting.

Molloy, I., Chen, H., Li, T., Wang, Q., Li, N. and Bertino, E. (2008) Mining Roles with Semantic Meanings

Nakaoka, I., Tani, K., Hoshino, Y.and Kamei, K. (2006). A Bankruptcy Prediction Method Based on Cash flow Using SOM

Nasir, M.L., John,R.I., Bennett, S.C..(2005). Predicting Corporate Bankruptcy Using Modular Neural Networks

Park, Y. (2008): Banking Market Concentration and Credit Availability to Small Businesses

Sai, Y., Zhong, C. J., Nie, P. Y.(2007). A Hybrid RST and GA-BP Model for Chinese Listed Company Bankruptcy Prediction

Schied, A. and Schoeneborn, T. (2008): Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets.

Schmidt, F. (2009): The Undervaluation of Distressed Company's Equity.

Scott, R. I., Svinterikou, S., Tjortjis, C., Keane, J. A. (1999). Experiences of using Data Mining in a Banking Application. Department of Computation, UMIST, Manchester, UK.

Shi, A., Long, A. & Newcomb, D. (2001). Enhancing e-Business Through Web Data Mining.

Shin K. S. , Lee, T. S., Kim, H. J. (2004). An application of support vector machines in bankruptcy prediction model

Topaloglou, N., Vladimirou, H., and Zenios, S. A. (2005) Controlling Currency Risk with Options or Forwards

Vieira, A. S., Ribeiro, B., Mukkamala, S.,. Neves, J. C and Sung, A. H. (2004). On the Performance of Learning Machines for Bankruptcy Detection

Yeung, D. S., Ng, W. W. Y., Chan, A. P. F., Chan, P. P. K., Firth, M., Tsang, E. C. C. (1998). Bankruptcy Prediction Using Multiple Intelligent Agent System via a LocalizedGeneralization Error Approach

Yoon, J. S., Kwon, Y. S., Roh, T. H. (2007). Performance Improvement of Bankruptcy Prediction using Credit Card Sales Information of Small & Micro Business

Zhao, Y., Zhang, H., Figueiredo, F., Cao, L., Zhang, C. (2007). Mining for Combined Association Rules on Multiple Datasets. Proceeding of ACM SIGKDD Workshop on Domain Driven Data Mining