

COMPARISON BETWEEN ROUGH SET THEORY AND LOGISTIC REGRESSION FOR CLASSIFYING FIRM'S PERFORMANCE

BAHTIAR JAMILI ZAINI, SITI MARIYAM SHAMSUDDIN & SAIFUL HAFIZAH JAAMAN

ABSTRACT

Superior prediction and classification in determining company's performance are major concerns for practitioners and academic researchers due to the importance of providing useful information to shareholders and potential investors. The general practice is to analyze firm's performance based on financial indicators measured on the financial statements published in the annual report, i.e. the balance sheet, income and cash flow statements. To do so, numerous financial indicators are used in order to classify the performance of firm accordingly involving many financial variables in order to determine firm's performance. Some of these financial ratios may be irrelevant and may correlate with each other giving redundant information for classification. Hence, this study investigates and determines the financial ratios that notably affect firm's performance so as to predict future performance. In this study, firm is considered as high performance firm if its share returns perform above returns provided by Kuala Lumpur Composite Index. Using various financial ratios as the independent variables, this study investigates and determines the financial indicators that significantly affect firm's share performance so as to predict its future performance. Financial performance predictions normally involve large information to explore, using traditional statistical methods such as multiple discriminant analysis or binary logistic regression analysis. Parametric statistical models are based on normality assumption requirements for the interpretation of the tests of significance, and if the data does not satisfy this assumption, the results obtained may be biased. However, it is noted that most financial ratios are skewed and non-normally distributed suggesting that non-parametric test is a superior alternative. Non-parametric model does not assume the data to have any specific characteristics. In this study comparison between rough set and logistic regression methodology is employed to identify the most significant indicators in classifying firm's performance. It was found that rough set model could accurately predict 86% of the testing dataset whereas logistic regression model shows 61.6% accuracy rate.

Keywords: feature selection; financial ratios; rough set theory; logistic regression

ABSTRAK

Kepentingan melaporkan maklumat yang tepat kepada pemegang saham dan bakal pelabur telah menjadikan keunggulan meramal dan mengelas prestasi firma satu keutamaan kepada penyelidik dan pengamal industri. Amalan yang sering digunakan ialah dengan menganalisis prestasi firma secara mengukur penentu kewangan berasaskan penyata kewangan yang diterbitkan dalam Penyata Tahunan firma, iaitu Kunci Kira-Kira, Penyata Pendapatan dan Penyata Aliran Tunai. Pelbagai penyata kewangan digunakan bagi tujuan mengelas prestasi firma melibatkan berbagai pemboleh ubah kewangan bagi menentukan kedudukan prestasi firma. Namun begitu, kemungkinan terdapat nisbah kewangan yang tidak relevan dan berkolerasi antara satu sama lain, justeru memberi maklumat yang kurang tepat. Oleh itu, dalam kajian ini kami menyiasat dan menentukan nisbah kewangan yang ketara penting mempengaruhi prestasi firma bagi tujuan meramal prestasi firma di masa hadapan. Dalam kajian ini, sebuah firma dianggap sebagai firma berprestasi tinggi jika pulangan sahamnya melebihi yang ditawarkan oleh Indeks Komposit Kuala Lumpur. Dengan menggunakan pelbagai nisbah kewangan sebagai pemboleh ubah tak bersandar, kajian ini menentukan nisbah kewangan yang secara signifikan mempengaruhi prestasi saham firma bagi meramal prestasi masa hadapan. Peramalan prestasi kewangan biasanya memerlukan pengkajian maklumat yang banyak menggunakan kaedah statistik konvensional seperti analisis diskriminan berganda atau regresi logit. Ujian signifikan

model statistik parametrik adalah berasas kepada keperluan andaian normal, jika data tidak memenuhi andaian ini, keputusan terhasil berkemungkinan adalah bias. Adalah diketahui umum, kebanyakan nisbah kewangan bersifat pencong dan tidak bertabur secara normal mencadangkan ujian bukan-parametrik merupakan alternatif terbaik. Model bukan-parametrik tidak mengandaikan data perlu memenuhi sebarang kriteria. Kajian ini membandingkan metodologi set kabur dengan regresi logistik bagi mengenal pasti penentu yang paling signifikan dalam mengklasifikasi prestasi firma mendapati model set kabur dapat meramal dengan tepat 86% set data ujian manakala model regresi logistik menunjukkan kadar ketepatan 61.6%.

Kata kunci: pemilihan ciri; nisbah kewangan; teori set kabur; regresi logistik

References

- Ahn B.S., Cho S.S. & Kim C.Y. 2000. The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction. *Expert Systems with Application* **18**: 65 – 74.
- Atiya A.F. 2001. Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. *IEEE Transactions on Neural Networks* **12**(4): 929 - 935.
- Beynon M., Curry B. & Morgan P. 2000. Classification and Rule Induction using Rough Set Theory. *Expert Systems* **17**(3): 136 – 148.
- Bose I. 2006. Deciding the Financial Health of Dot-Coms Using Rough Sets. *Information and Management* **43**: 835 – 846.
- Dimitras A., Slowinski R., Susmaga R. & Zopounidis C. 1999. Business Failure Prediction Using Rough Sets. *European Journal of Operational Research* **114**: 263-280.
- Golan R. H. & Ziarko W. 1995. A Methodology for Stock Market Analysis Utilizing Rough Set Theory. *IEEE/IAFE Conference on Computational Intelligence in Financial Engineering*. New York: 32 – 40.
- Hua Z., Wang Y., Xu X. & Liang L. 2007. Predicting Corporate Financial Distress Based on Integration of Support Vector Machine and Logistic Regression. *Expert Systems with Applications* **33**: 434 – 440.
- Kim K. & Han I. 2001. The Extraction of Trading Rules from Stock Market data Using Rough Sets. *Expert Systems* **18**(4): 194 – 202.
- Karbhari Y. & Muhamad S. Z. 2004. Prediction of Corporate Financial Distress: Evidence from Malaysian Listed Firms during the Asian Financial Crisis: <http://ssrn.com/abstract=596607> (30 July 2007).
- Ko P. C. & Lin P. C. 2006. An Evolution-Based Approach with Modularized Evaluations to Forecast Financial Distress. *Knowledge-Based Systems* **19**: 84 – 91.
- Kumar P. R. & Ravi V. 2007. Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques – A Review. *European Journal of Operational Research* **180**: 1 – 28.
- Kusiak A. 2001. Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing. *IEEE Transactions on Electronic Packaging Manufacturing*. January **24** (1): 44 – 50.
- Lam M. 2004. Neural Network Techniques for Financial Performance Prediction: Integrating Fundamental and Technical Analysis. *Decision Support Systems* **37**: 567 – 581.
- Lendasse A., Lee E., de Bondt E., Wertz V. & Verleysen M. 2001. Input Data Reduction for the Prediction of Financial Time Series. *Proceeding European Symposium on Artificial Neural Networks Bruges (Belgium)*. 25 – 27 April 2001, 237-244.
- Lin F. Y. & McClean S. 2001. A Data Mining Approach to the Prediction of Corporate Failure. *Knowledge-Based Systems* **14**: 189 – 195.
- Liu H. & Yu L. 2005. Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering* **17**(4): 491 – 502.
- Nur Adiana Hiau Abdullah & Abd Halim @ Hamilton Ahmad. 2005. Detecting Financial Distress. *International Journal of Management Studies*. **12** (1): 77 – 95.
- Olson D. & Mossman C. 2003. Neural Network Forecasts of Canadian Stock Returns Using Accounting Ratios. *International Journal of Forecasting* **19**: 453 – 465.
- Paik J. & Suh Y. 2005. Performance Comparison of Various Data Mining Models to Predict Delisting of Firms. *Proceeding of the 2005 KMIS International Comperence*. 24-26 November 2005, Jeju South Korea.
- Pawlak Z., Grzymala-Bussr J., Slowinski R. & Ziarko W. 1995. Rough Set. *Communication of the ACM* **38** (11): 88-95.

Comparison between rough set theory and logistic regression for classifying firm's performance

- Pechenizkiy M., Puuronen S. & Tsymbal A. 2006. The Impact of Sample Reduction on PCA-based Feature Extraction for Supervised Learning. *Proceeding of 21st ACM Symposium on Applied Computing*, ACM Press.
- Slowinski R., Zopounidis C. & Dimitras A. 1997. Prediction of Company Acquisition in Greece by Means of the Rough Set Approach. *European Journal of Operational Research* **100**(1): 1 – 15.
- Tay F. E. H. & Shen L. 2002. Economic and Financial Prediction Using Rough Sets Model. *European Journal of Operational Research* **141**: 641 – 659.
- Walczak B. & Massart D. L. 1999. Tutorial Rough Sets Theory. *Chemometrics and Intelligent Laboratory Systems* **47**: 1 – 16.
- Wang T. C. & Chen Y. H. 2006. Applying Rough Sets Theory to Corporate Credit Ratings. *IEEE* : 132 – 136.
- Wu D., Liang L. & Yang Z. 2008. An Empirical Study of Financial Distress of Chinese Public Companies Using Probabilistic Neural Networks and Multivariate Discriminate Analysis. *Socio-Economic Planning Sciences* **42**(3): 206 – 220.
- Zhou J., Wu Z., Yang C. & Zhao Q. 2006. The Integrated Methodology of Rough Set Theory and Support Vector Machine for Credit Risk Assessment. *IEEE the Computer Society. Proceeding of the 6th International Conference on Intelligent Systems design and Application* **1**: 1173 – 1179.

*Pusat Pengajian Sains Matematik
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor D.E.
MALAYSIA
E-mail: shj@ukm.my or shj.2006@yahoo.com*