# THE ANALYSIS OF CORPORATE SOCIAL RESPONSIBILITY AND FINANCIAL PERFORMANCE USING ROUGH SET THEORY AND DECISION TREE – EVIDENCE FROM CHINA

Kuang-Hua Hu[1], Ming-Fu Hsu[2,*], Fu-Hsiang Chen[3] and Kai-Ya Chang[3]

[1]Accounting School
Nanfang College of Sun Yat-sen University
Wenquan Town, Conghua, Guangzhou 510970, P. R. China
khhu0622@gmail.com

[2]English Program of Global Business
[3]Department of Accounting
Chinese Culture University
No. 55, Hwa-Kang Rd., Yang-Ming-Shan, Taipei 11114, Taiwan
{ chenfuhsiang1; onlykaikai }@gmail.com; *Corresponding author: hsumf0222@gmail.com

Abstract. *Corporate social responsibility (CSR) would increase corporate value. Especially, the positive comments on a company's performance in environment, society, morality and responsibility often help increase operation efficiency, improve corporate image and reputation, enhance brand value and raise revenue. Unlike most of previous academic papers which discussed the relationship between CSR performance and financial performance merely through the overall CSR performance, this study tried dividing CSR performance into such indicators as "Responsibility management", "Market responsibility", "Social responsibility" and "Environmental responsibility" to elaborate on the effects of financial performance. Moreover, the rough set theory and the decision tree, which are more effective than the traditional statistical analysis in terms of the accuracy of prediction, were adopted to estimate and predict the relationship between the CSR performance and the financial performance. According to the research results, the overall accuracy rate is about 80%; "Responsibility management" is relatively the most important variable among the CSR performance variables, while "Corporate growth rate" and "Corporate scale" are relatively important variables among the non-CSR performance variables. The findings can be useful for enterprises in the implementation of CSR plans and for the government in the making of CSR policies.*
**Keywords:** Corporate social responsibility (CSR), Data mining, Financial performance, Rough set theory, Decision tree

1. **Introduction.** Many food safety events have been reported worldwide in recent years, such as the poisonous milk powder of Sanlu Group in China and the illegal cooking oil of Ting Hisin International Group in Taiwan, and they caused great panic among the consumers. Apart from reducing consumers' trust in enterprises, food safety problems affect the financial performance of enterprises. Because of these serious CSR problems, the public expect enterprises to not only make profits but also shoulder the responsibility for society, economy and environment. Therefore, CSR performance has been gaining increasing public attention. Gradually, many countries have formulated a large number of international norms and regulations about CSR. Aside from helping enterprises practice their social responsibility, these norms and regulations serve as an essential tool for enterprises to implement CSR plans. Since 2006, the Chinese government has been active to promote CSR and required enterprises to shoulder social responsibility according to relevant administration decrees. As is speculated in Article 5 of Company Law of People's

Republic of China, "A company shall obey law and administrative regulations, follow social and commercial morality, keep honest and trustworthy, receive the governmental and public supervision, and shoulder social responsibility". On May 14, 2008, Shanghai Stock Exchange issued the Notice of Motivating Listed Companies to Shoulder Social Responsibility and the Guidelines of Disclosing the Environmental Information about the Listed Companies of Shanghai Stock Exchange. By doing so, it aimed to encourage listed companies to reveal their annual CSR reports on official websites and respect their decisions to make CSR strategic plans and work mechanisms.

Moore and Spence [1] and Scherer et al. [2] pointed out that the implementation of CSR was beneficial for national economy: on the one hand, it met the social expectation on the role of enterprise as a citizen; on the other hand, it increased the social demand for consumption, which not only helped enhance economic consumption but also enabled enterprises to promote national economy with their financial performance while making profits. From the perspective of business, the CSR implementation can increase corporate value. Especially, the positive comments on a company's performance in environment, society, morality and responsibility often help increase operation efficiency, improve corporate image and reputation, enhance brand value and raise revenue [3-5]. According to the previous academic papers about CSR and financial performance, there is a positive relationship between the two [5-7]. The relationship is a causal one, for great financial performance can offer a company capital and enable it to make an investment to improve its environmental and social performance; additionally, high environmental and social performance can generate remarkable financial performance and enable enterprises to make use of resources and keep their promises for stakeholders.

Most of the previous academic papers about the CSR and financial performance focused on the overall CSR performance alone [6,8]. However, CSR includes many different dimensions; the industrial features of enterprises are different, which leads to the difference in focus. Therefore, it is difficult to delve into the effects of the CSR dimensions on financial performance. Given the problem, this study adopted the Research Report on Corporate Social Responsibility of China in the Blue Book of Corporate Social Responsibility issued by Chinese Academy of Social Sciences and used such indicators as "Responsibility management", "Market responsibility", "Social responsibility" and "Environmental responsibility" to measure the CSR performance. Unlike the previous studies which merely explored the overall CSR performance, this study probed into the effects of the above four indicators on financial performance.

Most of the previous studies on CSR and financial performance adopted the traditional statistical analyses, including the logistic regression analysis and the multivariate discriminant analysis [5,6,8,9]. These analyses cannot be used unless they meet specific statistical hypotheses (such as meeting the normal distribution). Therefore, the data mining for any statistical hypothesis of data portfolios was not adopted. Gradually, a huge number of scholars have applied it to such areas as finance and accounting [10,11]. As data mining can meet the demand for massive operation and the result of the classification can be used as the basis for decision-making and prediction, an increasing number of experts and scholars have used it to solve problems, academically and practically. This study aims to modularize data mining through training and establish a system model, so as to estimate and predict the relevance between CSR performance and financial performance.

The contributions of this study are as follows. 1) Unlike the previous studies, this study divided CSR into such indicators as "Responsibility management", "Market responsibility", "Social responsibility" and "Environmental responsibility" and probed into their effects on financial performance. 2) The accuracy of the data mining-based prediction model of this study is normally higher than that of the prediction models based on the traditional logistic regression analysis and the discriminant analysis, which demonstrates that data mining can be applied to the topics about finance and accounting. 3) With

the data mining, this study proposed a new idea in the exploration into the relationship between CSR performance and financial performance, and the finding can be useful for enterprises' CSR plans and for the government's CSR policies.

The remainder of this paper is organized as follows. In Section 2, the relationship of CSR and financial performance and data mining are introduced. In Section 3, rough set theory and decision tree are proposed for detecting the analysis of CSR and financial performance. Section 4 discourses the empirical results. Finally, Section 5 offers conclusions with observations and remarks.

2. **Literature Review.**

2.1. **Relationship between CSR and financial performance.** From the perspective of stakeholder, Freeman [12] believed that there was a positive relationship between CSR and financial performance. Tsoutsoura [13] argued that stronger CSR might contribute to greater financial performance because CSR helped enterprises improve their reputation and build a positive public image. Barnett and Salomon [6] took "return on assets" and "net profit" as the indicators of financial performance and used the American KLD social responsibility data to measure CSR performance in the analysis of the relationship between CSR and financial performance. The findings showed that the CSR implementation would require cost and have negative effects on profit making in the early stage; but later, CSR would improve the image and reputation of enterprises and contribute to more profits. A summary of the above studies on the relationship between CSR and financial performance, the CSR implementation is beneficial for enterprises. Currently, many enterprises have listed CSR as an operation objective. However, CSR has diverse dimensions. More remarkable achievements will be obtained if CSR performance is divided into smaller items in the analysis of the relationship between CSR and financial performance. Therefore, this paper divided CSR into such indicators as "Responsibility management", "Market responsibility", "Social responsibility" and "Environmental responsibility" in accordance with Research Report on Corporate Social Responsibility of China (2015) [14] to delve into the relevance between each indicator and financial performance.

2.2. **Rough set theory.** Pawlak and Slowinski [15] proposed that rough set was used to solve the problems of vague and imprecise information; it could simplify data and obtain the minimum expression of knowledge while maintaining key information in the data; then, it could make decision-making norms and detect the knowledge hidden in the data set.

2.3. **Decision tree.** Decision tree [10] can be presented in the form of a tree diagram which is regular and easy to understand. Aside from generating the rules which are easy to understand, decision tree can process continual and category items, achieving classification without too much computation. As decision tree establishes rules according to different variables, the nodes are generated through the division of data, and the division is repeated until the rules of selecting important variables are found or the division becomes impossible.

3. **Study Methodology.**

3.1. **Variable measurement.** In this study, "Return on assets", "Return on equity" and "Earnings per share" were taken as the proxy variables of financial performance, and the data were collected from the Taiwan Economic Journal (TEJ). The details about the selection of CSR variables and non-CSR variables are as follows: CSR variables included "Responsibility management", "Market responsibility", "Social responsibility" and "Environmental responsibility", and the selection method has been mentioned above; the non-CSR variables, including "Corporate scale", "Debt-to-equity ratio", "Development

TABLE 1. Variable measurement

| Variable | Variable Measurement | References |
|---|---|---|
| Return on assets (ROA) | Net profit after tax + Interest * (1 – Tax rate)/Average total assets | [6] |
| Return on equity (ROE) | Net profit after tax/Average net shareholder's equity | [6] |
| Earnings per share (EPS) | (Net profit after tax – Dividend of special shares)/Weighted average of the number of issued common shares | [6] |
| Responsibility management ($X_1$) | Reveal current CSR management | [14] |
| Market responsibility ($X_2$) | Reveal the market responsibility performance of enterprises | |
| Social responsibility ($X_3$) | Reveal the social responsibility performance of enterprises | |
| Environmental responsibility ($X_4$) | Reveal the environmental responsibility performance of enterprises | |
| Corporate scale ($X_5$) | Measure corporate scale with the quantity of employees | [6] |
| Debt ratio ($X_6$) | Total debts/total assets | [6,16] |
| R&D intensity ($X_7$) | (Research and development expenditure/Net amount of sales income) | [16] |
| Growth rate of enterprise ($X_8$) | (Net amount of sales of the current year – Sales of the previous year)/Sales of the previous year | [16] |
| Management competence ($X_9$) | Net amount of sales/Average total assets | [16] |

intensity", "Growth rate of enterprise" and "Management competence", were selected according to the previous academic papers. The details are shown in Table 1.

3.2. **Using rough set theory to classify.** Pawlak [17] was the first to propose that the rough set theory was a new way to infer discrete data; the set theory was the mathematical foundation of the rough set and was effective in seeking uncertainty and the knowledge hidden in fuzzy data. The rough set theory can be analyzed in a large quantity of data to seek rules. Its advantage is that the data do not need to be processed in advance and that no statistical hypotheses are needed to present data. The rough set algorithm is based on the reduction of accuracy of data and makes data pattern more obvious, while the core of the rough set theory lies in the knowledge of classification. In other words, classification or a group of cases are used to infer the form of the obtained decision-making rules.

3.3. **Three decision tree model.** This study adopted REPTree, CART and C4.5, which are detailed as follows.

3.3.1. *REPTree.* Information profit-making and variance are used to build the decision tree; meanwhile, the reduced-error pruning is applied to quickening the search for the tree. As for missing value, the corresponding cases are used for slicing [18].

3.3.2. *CART.* The binary division rule is used to deduce and analyze massive data for different predictions and to determine the selected prediction variables. CART is the most effective method to process missing data. If a record in the prediction data is missing, it will be impossible to choose the best division in the tree building [10].

3.3.3. *C4.5.* It was based on ID3 and developed by Quinlan [19] in 1993. Due to the limitation on the use of ID3 (which could merely process discrete and category-styled data rather than continual numerical data), the C4.5 decision tree was developed. Moreover, C4.5 features higher classification accuracy.

## 4. Empirical Results.

4.1. **Study data.** The samples of this study were the Top 100 state-owned, private and foreign-funded listed enterprises according to the Research Report on Corporate Social Responsibility of China (2013-2015). The variables of financial performance and the non-CSR variables were obtained from TEJ; the CSR variables were collected from the Research Report on Corporate Social Responsibility of China of the Blue Book of Corporate Social Responsibility. In case of a missing numerical value, the sample concerned would be removed. According to Table 2, there were 248 samples, including 131 state-owned enterprises and 117 private ones. As the foreign-funded enterprises in China are all not listed company, the number of the samples of foreign-funded companies was 0.

TABLE 2. Rules of sample selection

| Process of sample selection | Number of samples |
|---|---|
| Number of Top 100 state-owned, private and foreign-funded enterprises (2013-2015) in the Blue Book of Corporate Social | 900 |
| Number of the listed companies of Top 100 state-owned, private and foreign-funded enterprises (2013-2015) in the Blue Book of Corporate Social Responsibility | 301 |
| Number of the companies whose financial data could not be obtained in TEJ | (53) |
| Final number of samples | 248 |

4.2. **Descriptive statistics of variables.** The narrative statistics of all variables are shown in Table 3. The minimum value of the EPS is $-1.2$ and the maximum value is 605; the minimum value of the ROA is $-6.71\%$ and the maximum is $19.05\%$; the ROE minimum value is $-39.67\%$, and the maximum value is $49.68\%$; and in the CSR variables, the minimum value of the responsibility management is 0, and the maximum is 100; the minimum value of the market responsibility is 0, and the maximum is 100; the minimum value of social responsibility is 0, and the maximum value is 95.5; the minimum value of environmental responsibility is 0, and the maximum is 100; in non-CSR variables, the minimum size of the corporate scale is 282 people, and the maximum is 544083 people; the minimum value of debt ratio is $9.14\%$, the maximum is $88.04\%$, the minimum value of R&D intensity is $0\%$ and the maximum value is $288.17\%$. Therefore, the difference of R&D intensity is a bit big. The minimum growth rate of enterprise is $-32.21\%$ and the maximum value is $29254.09\%$. The growth rate of the enterprise is large, which shows that the future opportunities of the enterprise are different, management competence minimum value is $-0.12\%$, and the maximum value is $3.04\%$.

4.3. **Empirical results analysis.** The research adopted WEKA to execute rough set theory and decision tree. In this study, "Return on assets", "Return on equity" and "Earnings per share" were taken as the proxy variables of financial performance, and these variables were ranked according to the way Pai et al. [20] measured independent variables. The first 50% were set as "1", while the other 50% as "0". The rough set theory and the decision tree (REPTree, CART and C4.5) were adopted to obtain the

TABLE 3. Descriptive statistics

| Variables | Average mean | Min | Max | Standard deviation | Median | Variation |
|---|---|---|---|---|---|---|
| EPS (TWD) | 9.52 | −1.2 | 605 | 61.9 | 0.37 | 3831.37 |
| ROA (%) | 4.25 | −6.71 | 19.05 | 4.13 | 3.23 | 17.041 |
| ROE (%) | 8.62 | −39.67 | 49.68 | 10.76 | 8 | 115.77 |
| $X_1$ | 36.36 | 0 | 100 | 33.84 | 23 | 1145.17 |
| $X_2$ | 41.31 | 0 | 100 | 28.42 | 40.050 | 807.81 |
| $X_3$ | 37.89 | 0 | 95.5 | 28.13 | 33.2 | 791.27 |
| $X_4$ | 31.57 | 0 | 100 | 28.13 | 25.1 | 791.37 |
| $X_5$ (person) | 57408 | 282 | 544083 | 90837 | 24432 | 8251525590 |
| $X_6$ (%) | 59.97 | 9.14 | 88.04 | 18.13 | 63.515 | 328.54 |
| $X_7$ (%) | 2.70 | 0 | 288.17 | 18.35 | 0.5012 | 336.69 |
| $X_8$ (%) | 123.24 | −32.21 | 29254 | 1857.6 | 3.2341 | 3450690.35 |
| $X_9$ (%) | 0.79 | 0.12 | 3.04 | 0.49 | 0.67 | 0.24 |

TABLE 4. General accuracy of the rough set theory and the decision tree

| | | | EPS | ROA | ROE |
|---|---|---|---|---|---|
| Overall accuracy rate | Decision tree | C4.5 | 80% | 76% | 79.2% |
| | | REP Tree | 82% | 80.8% | 79.6% |
| | | CART | 82% | 79.2% | 78.4% |
| | Rough Set Theory | | 80% | 74.8% | 75.2% |

TABLE 5. Ranking of relatively important variables of the models

| | EPS | ROA | ROE |
|---|---|---|---|
| C4.5 | $X_8, X_7, X_5, X_6, X_1$ | $X_6, X_1, X_5$ | $X_8, X_9, X_1, X_5$ |
| REP Tree | $X_8, X_5, X_1, X_4$ | $X_8, X_6, X_9, X_1, X_7$ | $X_8, X_5, X_6, X_2, X_1$ |
| CART | $X_8, X_5, X_6, X_1$ | $X_8, X_6, X_5, X_1$ | $X_8, X_5$ |
| Rough Set Theory | $X_1, X_8$ | $X_1, X_5$ | $X_1, X_8$ |

overall accuracy rate and the relatively important variables, which are shown in Tables 4 and 5.

According to the summary of the research results of the above proxy variables, "Earnings per share" had the highest accuracy in the evaluation based on the rough set theory and the decision tree, with all of its accuracy rate being above 80%. As for the variables which influence financial performance, "Responsibility management" ($X_1$) had the highest frequency of appearance among the CSR variables. This shows that it has greater effect on financial performance. According to similar studies, effective corporate governance (or responsibility management) promotes legal management and the disclosure of social responsibility information [21]. It has also been pointed out that management policies in CSR's responsibility management will increase financial resources and will lead to greater social benefits [22]. This demonstrates that responsibility management is more important among the four CSR dimensions. However, among the non-CSR variables, "Growth rate of enterprise" ($X_8$) and "Corporate scale" ($X_5$) had the highest frequency of appearance, which shows that these two indicators are more important for financial performance.

5. **Conclusion.** In this study, the data mining was employed to explore the relationship between CSR and financial performance according to the financial performance of Chinese listed enterprises as well as the CSR and non-CSR variables. According to Table 4, most of the general accuracies were above 80%, which indicates that CSR indeed has great effect

on financial performance, with "Responsibility management" being relatively the most important variable. Liu and Zhang [21] and Rodriguez-Fernandez [22] also pointed out that a complete corporate governance mechanism (responsibility management) helped enterprises implement CSR. Additionally, "Growth rate of enterprise" and "Corporate scale" are relatively important among the non-CSR variables. In this study, the state-owned and private enterprises in China were taken as the samples. It is suggested that future studies explore the relationship between state-own enterprises and financial performance and between private ones and financial performance respectively. As these two kinds of enterprises involve diverse industries, future studies can delve into one of the industries and analyze the relationship between the CSR of different industries and financial performance, so that enterprises be able to explore how to enhance financial performance through CSR, solve their problems and thus strengthen their competitiveness.

## REFERENCES

[1] G. Moore and L. Spence, Editorial: Responsibility and small business, *Journal of Business Ethics*, vol.67, no.3, pp.219-226, 2006.

[2] A. G. Scherer, A. Rasche, G. Palazzo and A. Spicer, Managing for political corporate social responsibility: New challenges and directions for PCSR 2.0, *Journal of Management Studies*, vol.53, no.3, pp.273-298, 2016.

[3] M. P. Lee, Does ownership form matter for corporate social responsibility? A longitudinal comparison of environmental performance between public, private, and joint-venture firms, *Business and Society Review*, vol.114, no.4, pp.435-456, 2009.

[4] M. Weber, The business case for corporate social responsibility: A company-level measurement approach for CSR, *European Management Journal*, vol.26, no.4, pp.247-261, 2008.

[5] Q. Zhu, J. Liu and K. H. Lai, Corporate social responsibility practices and performance improvement among Chinese national state-owned enterprises, *International Journal of Production Economics*, vol.171, pp.417-426, 2016.

[6] M. L. Barnett and R. M. Salomon, Does it pay to be really good? Addressing the shape of the relationship between social and financial performance, *Strategic Management Journal*, vol.33, no.11, pp.1304-1320, 2012.

[7] C. Wickert, A. G. Scherer and L. J. Spence, Walking and talking corporate social responsibility: Implications of firm size and organizational cost, *Journal of Management Studies*, vol.53, no.7, pp.1169-1196, 2016.

[8] J. J. Griffin and J. F. Mahon, The corporate social performance and corporate financial performance debate, *Business & Society*, vol.36, no.1, pp.5-31, 1997.

[9] T. Jain and D. Jamali, Looking inside the black box: The effect of corporate governance on corporate social responsibility, *Corporate Governance: An International Review*, vol.24, no.3, pp.253-273, 2016.

[10] F. H. Chen and H. Hu, An alternative model for the analysis of detecting electronic industries earnings management using stepwise regression, random forest, and decision tree, *Soft Computing*, vol.20, no.5, pp.1945-1960, 2016.

[11] F. H. Chen, D. J. Chi and Y. C. Wang, Detecting biotechnology industry's earnings management using Bayesian network, principal components analysis, back propagation neural network, and decision tree, *Economic Modeling*, vol.46, pp.1-10, 2014.

[12] R. E. Freeman, *Strategic Management: A Stakeholder Perspective*, Prentice-Hall, Boston, USA, 1984.

[13] M. Tsoutsoura, *Corporate Social Responsibility and Financial Performance*, Center for Responsible Business, Haas, University of California, Berkeley, California, 2004.

[14] J. Chen, Q. Huang, H. Peng and H. Zhong, *Blue Book of Corporate Social Responsibility: Research Report on Corporate Social Responsibility of China*, Beijing, China, 2015.

[15] Z. Pawlak and R. Slowinski, Rough set approach to multiattribute decision analysis, *European Journal of Operational Research*, vol.72, pp.443-459, 1994.

[16] A. McWilliams and D. Siegel, Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, vol.21, pp.603-609, 2000.

[17] Z. Pawlak, Rough sets, *International Journal of Parallel Programming*, vol.11, no.5, pp.341-356, 1982.

[18] S. Kalmegh, Analysis of WeKA data mining algorithm REPTree, simple cart and random tree for classification of Indian news, *International Journal of Innovative Science, Engineering & Technology*, vol.2, no.2, pp.438-446, 2015.

[19] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

[20] P. F. Pai, M. F. Hsu and L. Lin, Enhancing decisions with life cycle analysis for risk management, *Neural Computing and Applications*, vol.24, nos.7-8, pp.1717-1724, 2014.
[21] X. Liu and C. Zhang, Corporate governance, social responsibility information disclosure, and enterprise value in China, *Journal of Cleaner Production*, vol.142, pp.1075-1084, 2017.
[22] M. Rodriguez-Fernandez, Social responsibility and financial performance: The role of good corporate governance, *BRQ Business Research Quarterly*, vol.19, no.2, pp.137-151, 2016.