Asian Journal of Economics, Business and Accounting

16(2): 1-12, 2020; Article no.AJEBA.59058
ISSN: 2456-639X

Financial Distress Prediction Using Hybrid Machine Learning Techniques

Suduan Chen[1] and Zong-De Shen[2*]

[1]Department of Accounting Information, National Taipei University of Business, No.321, Sec. 1, Jinan Road, Zhongzheng District, Taipei, 100, Taiwan.
[2]Department of Accounting, Chinese Culture University, No.55, Hwa-Kang Road, Yang-Ming-Shan, Taipei, 11114, Taiwan.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJEBA/2020/v16i230231
Editor(s):
(1) Chun-Chien Kuo, National Taipei University of Business, Taiwan.
Reviewers:
(1) Ika Prayanthi, Universitas Klabat, Indonesia.
(2) Poonam Rani, International Institute of Research and Studies, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/59058

Original Research Article

Received 04 May 2020
Accepted 10 July 2020
Published 13 July 2020

ABSTRACT

The purpose of this study is to establish an effective financial distress prediction model by applying hybrid machine learning techniques. The sample set is 262 financially distressed companies and 786 non-financially distressed companies, listed on the Taiwan Stock Exchange between 2012 and 2018. This study deploys multiple machine learning techniques. The first step is to screen out important variables with stepwise regression (SR) and the least absolute shrinkage and selection operator (LASSO), followed by the construction of prediction models, as based on classification and regression trees (CART) and random forests (RF). Both financial variables and non-financial variables are incorporated. This study finds that the financial distress prediction model built with CART and variables screened by LASSO has the highest accuracy of 89.74%.

Keywords: Machine learning approach; financial distress prediction; least absolute shrinkage and selection operator; classification and regression tree; random forests.

**Corresponding author: E-mail: szd2@ulive.pccu.edu.tw;*

1. INTRODUCTION

Both the U.S. and Taiwan have reported major corporate financial distress over the past 20-plus years after the Asian financial crisis in 1997. The examples are Enron in 2001, Xeron and K-Mart in 2002, WorldCom in 2003, AIG in 2005, and IBM in 2008 in the U.S., as well as Chou Chin Industrial and Pacific Electric Wire & Cable in 2003, Procomp Informatics, Infodisc Technology, Summit Technology, and ABIT Computers in 2004, Rebar in 2007, and Tah Chung Steel in 2008 in Taiwan. The global financial tsunami in 2007-2008 was caused by the subprime mortgage crisis in the U.S. In 2008, Lehman Brothers sank into bankruptcy, Merrill Lynch was acquired by the Bank of America, and the American International Group (AIG) asked for a federal bail-out. Taiwan was not insulated from this global financial crisis, and many companies and factories shut down, which resulted in heavy losses for investors and the loss of jobs. The trade war between China and the U.S. in 2018-2019 has increased the probability of financial distress for numerous companies in Taiwan.

Business operations are closely concerned with social dynamics, and corporate crises cost dearly to the economy. Such events not only damage the rights of stakeholders, but also lead to massive losses for society. If managers and CPAs can identify problems and issue warnings early, necessary measures should be taken to prevent financial distress or stop the further deterioration of financial troubles.

In the era of globalization, companies are expanding and increasingly internationalized, thus, business risks and challenges are also increasing. In fact, there are signs before the outbreak of any financial crisis. In order to determine the probability of financial distresses before they occur, it is necessary to construct a robust pre-warning model, which can help to identify problems early, improve management, and prevent fraud, while minimizing the damages caused by bankruptcies and protecting shareholders, employees, creditors, and other stakeholders.

The earliest research on financial distress predictions was Fitzpatrick [1], who noticed that net profit/equity and equity/debt are the two indicators with the highest predictability. After that, similar research was based on heuristic analysis and the comparison of financial ratios between failed companies and normal companies; however, it was not until the early 1960s that financial distress predictions became more systematic.

Beaver [2] predicted the probability of corporate financial crises by measuring profitability, liquidity, and solvency, and the results showed that the best variables to spot any financial distress is operating cash flows/debts (successfully identified 90% of bankrupt companies one year before) and net profits/assets (88% success rate during the same period). Altman [3] used the multivariate discriminant approach (MDA) to examine bankruptcy problems by selecting 33 bankrupt companies and 33 non-bankrupt in the same industries and of comparable sizes, and constructed the discriminant function: $Z = .012X1 + .014X2 + .033X3 + .006X4 + .999X5$, where X1: working capital/total assets, X2: retained earnings/total assets, X3: earnings before interest and taxes/total assets, X4: market value of equity/book value of total debt, X5: sales/total assets.

However, Moyer [4] indicated that Altman’s prediction model was not applicable to all periods. Empirical evidence also indicates that most financial ratios do not satisfy the robust presumptions required for the MDA approach. The famous study by Ohlson [5] used the Logit model to predict financial crises and incorporated asset scales into the model. The research subjects were 105 bankrupt companies and 2,058 normal companies in 1970-1976, a total of nine financial variables were used to estimate the model, and the empirical results indicated a few financial metrics that are statistically significant for financial distress predictions, with an accuracy rate of 92% or above: total assets/logarithm of GNP Price Index; total Liabilities/total assets; (return on total assets, ROA) or working capital/total Liabilities; working capital/total assets, current liabilities/current assets. His study incorporates two dummy variables, OPNEG and INTWO. The former is 1 if total assets exceed total liabilities and 0 if not. The latter is 1 in the case of net losses during the two years before bankruptcy and 0 if not. Zmijewski [6] deployed the Probit model for the prediction of financial distresses and selected ROA, D/E ratio, and current ratio as the research variables. While both Probit and Logit are regression techniques, the Logit model assumes the probabilities are in the logistic distribution, while the Probit model assumes a normal distribution, which is a much stricter criterion.

Shumway [7] constructed the discrete time hazard model as a major progress for the use of logistic regression analysis in financial distress prediction.

In addition to the traditional approaches of financial distress predictions, many techniques have emerged due to the rapid development of computers and software, including data mining, machine learning, deep learning, and artificial intelligence [8-18].

Machine learning is the use of most data (typically 50% to 80%) for training before making predictions, which enhances prediction accuracy and reduces judgment errors [19-22]. Machine learning techniques are also very suitable for financial distress prediction [8,11,13,17,23-25]. This study adopts multiple machine learning techniques, first, by selecting key variables with the least absolute shrinkage and selection operator (LASSO) and stepwise regression (SR). This is followed by the construction of financial distress prediction models using classification and regression trees (CART) and random forests (RF). The purpose is to compare the accuracy rates of different models to choose the optimal model.

2. MATERIALS AND METHODS

This study sources data from the Taiwan Economic Journal (TEJ), and multiple machine learning techniques are employed. The first step is to select the important variables with stepwise regression (SR) and the least absolute shrinkage and selection operator (LASSO). This is followed by the construction of financial distress prediction models by applying classification and regression trees (CART) and random forests (RF). The purpose is to compare accuracy rates of different models to choose the optimal model.

SR is a frequently used statistic technique. The LASSO algorithms seek to minimize the squared sum of residuals under the condition that the sum of the absolute values of coefficients is smaller than a specific constant. Its advantage lies in significant variable compression in the case of large parameter estimates, and the variable is compressed to zero if parameter estimates are small. Meanwhile, the parameter estimates are continuous for LASSO analysis. This is a suitable model for high-dimensional data.

CART and RF, which are the two most representative decision tree algorithms, with the following advantages [20,26-27]: (1) The decision tree models can be expressed in graphs or rules, and the rules are easy to explain and comprehend; (2) It is possible to handle continuous or categorical data and indicate the relative importance of variables; (3) It processes large data sets well. Decision trees can be constructed for many variables.

2.1 Stepwise Regression

Stepwise regression (SR) is the comparison of variables entered one by one and the subsequent development of analytical models [28]. The procedures are, as follows: confirmation of the initial model, comparison of parameters to include or exclude variables from the analytical model, and completion of the process until iterations can no longer improve the model.

2.2 LASSO

The least absolute shrinkage and selection operator (LASSO) was developed by Tibshirani [29]. It is the selection of appropriate explanatory variables by restricting the regression parameters with a penalized sum of squares, which is similar to ridge regression, and expressed as Eq. (1).

$$\min \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \ldots \beta_k x_{k,t})^2 \tag{1}$$

The parameter restriction is expressed as Eq. (2)

$$S.T \sum_{j=1}^{k} \left|\beta_j\right|, \leq \lambda \tag{2}$$

where if λ->∞, parameter estimate β̂ is not subject to the limitation of Eq. 2, and the estimated value will be equal to the estimate with the least square method. If λ is zero, all the parameter estimates are zero. If λ is increased, the coefficient of the explanatory variables with stronger correlations with dependent variables will change, and will not equal zero. In contrast, the explanatory variables with smaller correlations and dependent variables will maintain the corresponding coefficients at zero. Therefore, it is possible to determine whether coefficients are zero as the criterion for variables screening. Given the above features, LASSO is a tool for the processing of collinearity and screening of variables [30].

2.3 CART

Classifications and regression trees (CART) were developed by Breiman et al. [31]. As a binary decision tree, CART decomposes each variable into different binary combinations based on Gini impurity.

Gini impurity measures the percentage of erroneous labels for the elements randomly selected in the same set, and is defined, as follows:

$$Gini(p) = 1 - \sum_{i=1}^{J} p_i^2 \tag{3}$$

where p denotes the percentage of a certain category as the number of total categories.

In the development of the tree structures, entropy is subject to the inherent risks of the main function models; therefore, the flexibility of Gini impurity is a big advantage. This is the reason that CART algorithms, as proposed by Breiman et al. [31], have become the criteria for node selection.

2.4 Random Forests

A single decision tree may not have sufficient predictive power or suffer from overfitting, thus, the ensemble of multiple decision trees can enhance the objectivity and comprehensiveness of the prediction model. Random forests, as proposed by Breiman [32], is an approach to resolve single classification and prediction issues by ensemble learning with a combination of multiple models. Model establishment repeats the extraction of certain data and variables for the development of the tree structures, thus, the multiple tree structures at the bottom are random and final predictions are based on the majority. Therefore, the feature selection and prediction accuracy of random forests are superior to those produced with prediction models comprised of a single tree structure [33].

Random forecasts use the boot strapping method to extract data for modeling. In addition to the n number of bagging data points for each branch, a tree structure requires the m number of features for each branch and the number of branches. The development criteria are, as follows:

$$n \le N, m \le M, 0 < ntree < \infty \tag{4}$$

where N denotes the number of training data points, M is the number of variables for the training data, and ntree is the total number of decision trees.

2.5 Sampling and Variable Selection

2.5.1 Data sources

Samples of this study are the companies listed on the Taiwan Stock Exchange. Data is sourced from the Taiwan Economic Journal (TEJ) by classifying the financial distresses into the following types: (1) closure and bankruptcy; (2) restructuring; (3) failure to honor checks; (4) asking for a bail-out; (5) takeovers; (6) de-listings; (7) suspension of operations due to liquidity crunch; (8) doubts from CPAs regarding the company as a going concern; (9) negative net worth. The signals of financial distress include (1) asset embezzlement; (2) trading suspension; (3) withdrawn of credit facilities by banks; (4) significant losses; (5) dishonoring of checks issued by the chairperson; (6) suspension of operations due to poor business; (7) asset impairments; (8) financial distress of affiliated companies; (9) poor internal control; (10) insider trading. The research period is 2012-2018, and the sample pool consists of 262 listed companies in Taiwan with financial distress and 786 companies without financial distress. The matching data is sourced according to the techniques suggested by several authors and from companies in the same industries and of comparable sizes. The financial distress companies (FD) to normal companies (non-financial distress companies, NFD) ratio is 1 to 3 [19-20,22]. The data distribution across the different years is shown in Table 1.

2.5.2 Variable source and variable definition

The research variables used in this study are generally used by many researchers in financial distress prediction related research [3,5,8,11,13-15,17,20,22,24,34-36] and practice. Variables are defined as follows:

(1) Dependent variable: The dependent variable is a dummy variable, with 1 indicating financial distress and 0 indicating non-financial distress.

(2) Independent variables: This study selects a total of 20 independent variables commonly used to measure financial distress. This list consists of 15 financial variables and 5 non-financial variables (also known as corporate governance variables in literature). The definitions of the variables are summarized in Table 2.

Table 1. Research sample distribution

Year	Number of Financial Distress Companies (FD)	Number of Normal Companies (NFD)
2012	22	66
2013	13	39
2014	23	69
2015	17	51
2016	37	111
2017	73	219
2018	77	231
Total	262	786

Table 2. Research variables

No.	Variable Description	Definition or Formula
X01	Quick ratio	Quick assets ÷ Current liabilities
X02	Current ratio	Current assets ÷ Current liabilities
X03	Debt ratio	Total liabilities ÷ Total assets
X04	Debt-to-equity ratio	Total liabilities ÷ Total equity
X05	Accounts receivable turnover	Net sales ÷ Average accounts receivable
X06	Inventory turnover	Cost of goods sold ÷ Average inventory
X07	Operating cash flow ratio	Operating cash flow ÷ Current liabilities
X08	Gross profit ratio	Gross profit ÷ Net sales
X09	Pretax profit ratio	Pretax profit ÷ Net sales
X10	Net income ratio	Net income÷ Net sales
X11	Sales revenue growth rate	(Current Year's sales revenue - Last Year's sales revenue) ÷ Last Year's sales revenue
X12	ROE	Return on total equity: Net income ÷ average total equity
X13	ROA	[After-tax net profit + interest expense × (1-20%)] ÷ average total assets
X14	Times interest earned ratio	EBIT÷ Interest
X15	Long-term funds appropriate rate	(Total stockholders' equity + long term liabilities) ÷ Total fixed assets
X16	Audited by Big4 (the big four CPA firms)	1 for companies audited by BIG4, otherwise, it is 0
X17	The ratio of stocks held by directors and supervisors	Number of stocks held by directors and supervisors ÷Total number of common stock outstanding
X18	The ratio of stocks held by the major stockholders	Number of stocks held by the major stockholders ÷Total number of common stock outstanding
X19	The ratio of pledged stocks held by directors and supervisors	The number of pledged stocks held by directors and supervisors ÷ Number of stocks held by directors and supervisors
X20	The ratio of pledged stocks held by the major stockholders	The number of pledged stocks held by the major stockholders ÷ Number of stocks held by the major stockholders

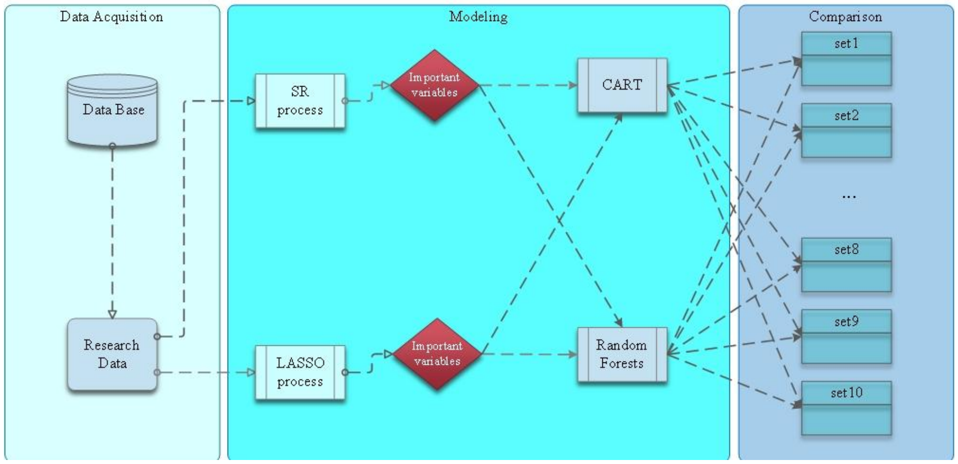

Fig. 1. Research flows

2.5.3 Research process

Samples of this study are companies listed on the Taiwan Stock Exchange, the data is sourced from the Taiwan Economic Journal (TEJ), and the research period is 2012-2018. The sample pool is comprised of 262 financial distress companies and 786 normal companies. This study selects a total of 20 variables typically used to measure financial distress. The list consists of 15 financial variables and 5 non-financial variables. This is followed by the screening of key variables with LASSO and stepwise regression for the classification models. The next step is the establishment of four prediction models: LASSO-CART, LASSO-RF, SR-CART, and SR-RF. This is followed with the comparison and analysis of prediction results, in order to identify the model with the highest accuracy. The research flows are depicted in Fig. 1.

3. RESULTS

The establishment of the prediction model for financial distress occurs in two stages, i.e. variable selections with SR and LASSO, and the construction of the four prediction models (LASSO-CART; LASSO-RF; SR-CART; SR-RF). The selection results are described in detail, as follows.

3.1 Stepwise Regression Selection

This study sources a total of 22,008 data for 1,048 listed companies (262 with financial distress and 786 without financial distress), as based on the 20 research variables over the research period of seven years (2002-2018). SR is used to select important variables. The coefficients and AIC value changes are shown in Fig. 2.

Four important variables are selected with SR. These four variables in the order of selection are X12: ROE; X3: debt ratio; X7: operating cash flow ratio and X15: long-term funds appropriate rate. Table 3 shows the variable estimates and AICC values.

3.2 LASSO Selection

As in the SR process, this study screens the important variables for the 22,008 data using the Lasso method. The coefficients and AIC value changes are shown in Fig. 3.

Table 3. Selection result of the SR

Variable	Estimate	AICC Value
X12 ROE	0.004019	-860.6563
X3 Debt ratio	-0.000617	-907.9149
X7 Inventory turnover	-0.003838	-916.7870
X15 Long-term funds appropriate rate	0.00000136	-924.0550*

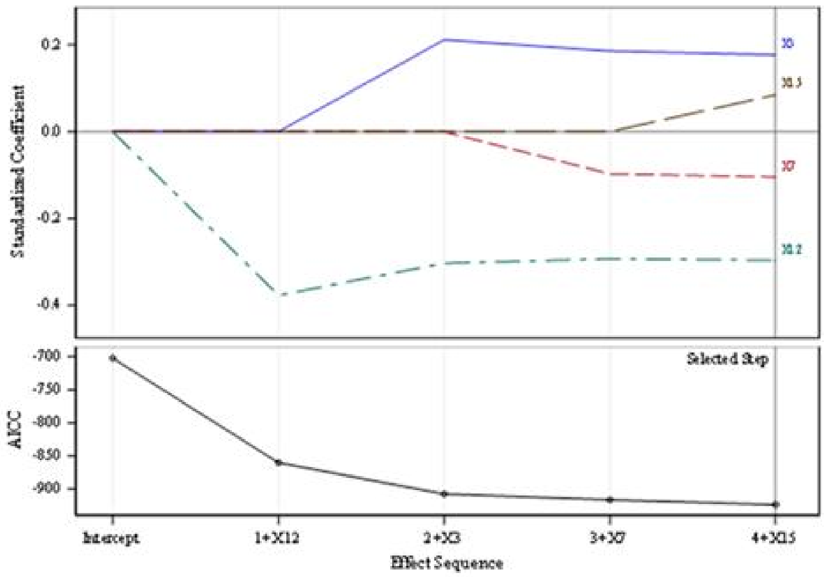

Fig. 2. SR selection coefficient and AICC value

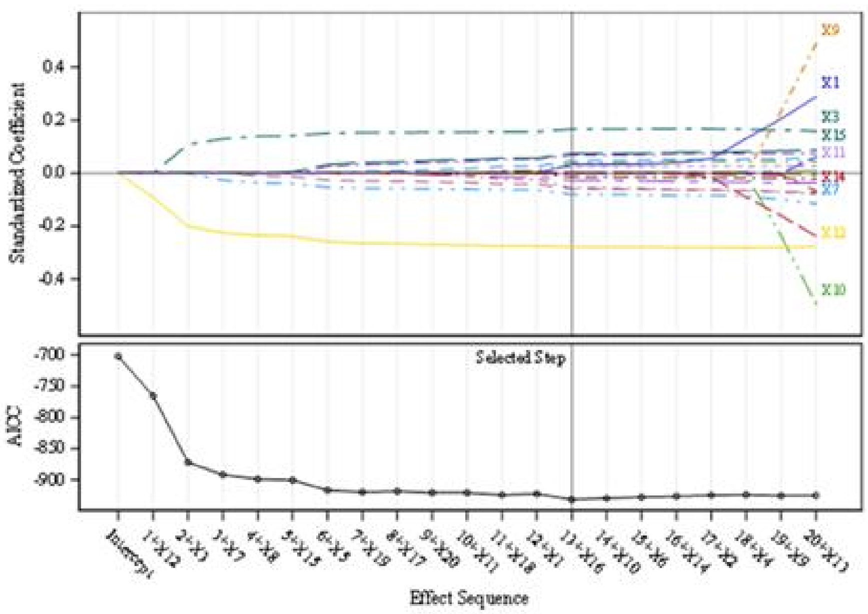

Fig. 3. LASSO selection coefficient and AICC value

The LASSO method selects a total of 13 important variables in the sequence of selection, as follows: X12:ROE; X3: Debt ratio; X7: Operating cash flow ratio; X8: Gross profit ratio; X15: Long-term funds appropriate rate; X5: Accounts receivable turnover; X19: The ratio of pledged stocks held by directors and supervisors; X17: The ratio of stocks held by directors and supervisors; X20: The ratio of pledged stocks held by the major stockholders; X11: Sales revenue growth rate; X18: The ratio of stocks held by the major stockholders; X1: Quick ratio; X16: Audited by Big4. The coefficients and AIC value changes are shown in Table 4.

3.3 Predication Models and Cross-Validation

The two sets of important variables, as selected with the SR method and the LASSO method, are then combined with CART and RF, respectively, for the construction of the prediction models. This study normalizes these variables and conducts random sampling without replacement. The data is divided into the training group and testing group for modeling and prediction in machine learning. This study adopts ten-fold cross-validation, which is considered by academics and researchers to be more robust to derive accurate predictions [19, 20-22]. This means the processes of modeling and verification are conducted 10 times, and the means are used to measure accuracy, as based on the results of the 10 iterations. The dataset is first randomly segmented into 10 parts. Nine out of the 10 parts are used for modeling, and one is used to verify prediction accuracy. This process is repeated 10 times to derive the means of the accuracy rates.

3.3.1 CART models

This study uses the important variables screened by SR and LASSO, and constructs prediction models with CART for financial distress. Table 5 shows the average accuracy, as based on the ten-fold cross-validation results. The LASSO-CART model has the higher accuracy of 89.74%, followed by the SR-CART model at 85.44%.

3.3.2 RF models

This study uses the important variables screened by SR and LASSO and constructs prediction models with RF for financial distress. Table 6 shows the average accuracy, as based on the ten-fold cross-validation results. The LASSO-RF model reports the higher accuracy of 86.30%, followed by the SR-RF model at 84.13%.

4. DISCUSSION

This study presents a new approach for screening variables with stepwise regression and LASSO, and then, combines the results with CART and RF as decision tree algorithms for model construction. This study integrates multiple machine learning techniques and samples a total of 1,048 listed companies (262 financial distress companies and 786 normal companies). A total of 20 research variables (15 financial variables and 5 non-financial variables) frequently used for the measurement of financial distress are selected. This study sources 22,008 data over the research period of seven years (2002-2018) and selects four important variables with the SR method and 13 with the LASSO method. These important variables are then used for the construction of prediction models for financial distresses with CART and RF. A robust ten-fold cross-validation is performed in the modelling process, in order to ensure prediction accuracy.

All four prediction models built by this study yielded an accuracy of over 80%. Among them, three models report an accuracy of higher than 85%. The mean accuracy of all four models is 86.40%, thus, the prediction accuracy is fairly high. The research results indicate that the CART and RF prediction model for financial distress, as built with the important variables selected by the LASSO method, demonstrate better accuracy than their counterparts, where the key variables were selected with the SR method.

Table 4. Selection result of LASSO

Variable	Estimate	AICC Value
X12 ROE	-0.003615	-765.564
X3 Debt ratio	0.003775	-871.799
X7 Operating cash flow ratio	-0.000479	-891.636
X8 Gross profit ratio	-0.001395	-898.695
X15 Long-term funds appropriate rate	0.00000119	-900.361
X5 Accounts receivable turnover	0.000046527	-916.416
X19 The ratio of pledged stocks held by directors and supervisors	0.001208	-919.255
X17 The ratio of stocks held by directors and supervisors	-0.000799	-917.971
X20 The ratio of pledged stocks held by the major stockholders	-0.001559	-920.354
X11 Sales revenue growth rate	0.000003356	-920.351
X18 The ratio of stocks held by the major stockholders	0.001793	-924.035
X1 Quick ratio	0.000014421	-922.187
X16 Audited by Big4	-0.015855	-931.2705*

Table 5. CART model accuracy using ten-fold cross-validation

Model	Accuracy
SR-CART	85.44%
LASSO-CART	89.74%

Table 6. RF model accuracy using ten-fold cross-validation

Model	Accuracy
SR-RF	84.13%
LASSO-RF	86.30%

The accuracy of the four models, as constructed by this paper, from high to low are LASSO-CART (89.74%), LASSO-RF (86.30%), SR-CART (85.44%), and SR-RF (84.13%).

Based on the above discussion, this study provides a way to construct rigorous and effective models for the prediction of financial distress.

5. CONCLUSIONS

Financial distress and bankruptcies have become a frequent occurrence since the Asian financial crisis in 1997 and the global financial tsunami in 2008. This has caused heavy losses to the investing public, the economies in different countries, and even around the world. It is, however, difficult for investors to spot the signs before the information of any financial distress becomes public, and by that time, it is impossible to make amends. This is the reason why the construction of effective models to predict financial distress is an important issue for both academics and practitioners.

According to the summary of relevant laws and regulations, as issued by the Taiwan Stock Exchange, the following situations are the most frequently seen warning signals of financial distress for companies listed in Taiwan: (1) the issuance of qualified opinions from auditors; (2) net value lower than half of the paid-in capital; (3) no shareholders' meetings within six months after the end of a fiscal year unless there are legitimate reasons; (4) defaults on straight corporate bonds or convertible bonds, or investors asking to redeem corporate bonds or convertible bonds; (5) checks dishonored due to insufficient funds in the bank; (6) more than half of the directors and supervisors under suspension of power or under provisional injunction due to financial or non-financial factors.

This study conducts research in a different approach from those mentioned in literature, meaning it integrates a few machine learning techniques by selecting important variables with the SR method and LASSO method. These variables are then used for the construction of prediction models for financial distress with CART and RF algorithms, respectively. Robust ten-fold cross-validation is performed to boost prediction accuracy. Among the four prediction models built by this study, the LASSO-CART (based on variables selected with the LASSO method and CART algorithms) reports the highest accuracy of 89.74%. The accuracy of the other three models from high to low are: LASSO-RF (86.30%), SR-CART (85.44%), and SR-RF (84.13%).

In addition to the 15 financial variables typically used for the measurement of financial distress, this study selects 5 non-financial variables (also known as corporate governance variables in literature). As major financial distress reflect the importance of corporate governance, it is necessary to establish a healthy monitoring system to prevent frauds or illegal transactions by directors or supervisors. A comprehensive corporate governance mechanism also boosts the investing public's confidence and willingness to invest, encourages the development of stock markets, and attracts the in-flow of international capital. It will also enhance the internal control and board effectiveness of a company, the functioning of supervisors (Audit Committee), the protection of rights for shareholders and stakeholders, the transparency of information disclosure, as well as key policies and action plans, such as the selection of external auditors and lawyers.

In the era of Big Data, this study combines a few machine learning techniques to establish robust prediction models for financial distress. The research findings can serve as a reference to competent authorities, CPAs, institutional investors, securities analysts, company executives, and academics.

Finally, it must be mentioned that every model cannot be perfect. The weaknesses or the limitations of our optimal model-- LASSO-CART are: (1) the biggest shortcoming of LASSO in practice is that it is difficult to "minimize". In addition, LASSO seems to be not good enough when it comes to the selection of high-dimensional variables; (2) in CART classification, the Gini index (Gini) is used to select the best data segmentation characteristics. A small change in the dataset can make the tree structure unstable and each iteration process in CART will reduce the GINI.

ACKNOWLEDGEMENTS

The authors would like to thank the the editors and the anonymous reviewers of this journal.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Fitzpatrick FA. Comparison of ratios of successful industrial enterprises with those of failed firm. Certified Public Accountant, 6. 1932;6:727-731.
2. Beaver WH. Financial Ratios as Predictors of Failure. J. Account. Res. 1966;4:71-111.
   Available:https://doi.org/10.2307/2490171
3. Altman EI. Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. J Finance. 1968;23:589-609.
   Available:https://doi.org/10.2307/2978933
4. Moyer RC. Forecasting financial failure: A re-examination. Financ. Manage. 1977;23 (4):11-17.
   Available:https://doi.org/10.2307/3665489
5. Ohlson JA. Financial ratios and the probability prediction of bankruptcy. J. Account. Res. 1980;18(1):109-131.
   Available:https://doi.org/10.2307/2490395
6. Zmijewski ME. Methodological issues related to the estimation of financial distress prediction models. J. Account. Res. 1984;22:59-82.
   Available:http://dx.doi.org/10.2307/2490859
7. Shumway T. Forecasting bankruptcy more accurately: A simple hazard model, J. Bus. 2001;74(1):101-124.
   Available:https://doi.org/10.1086/209665
8. Odom MD, Sharda RA. neural network model for bankruptcy prediction., Proceedings of the IEEE International Joint Conference on Neural Network,, San Diego, CA, USA, 17-21 June 1990; IEEE: City, Country; 1990.
   Available:https://doi.org/10.1109/IJCNN.1990.137710
9. Tam K, Kiang M. Managerial applications of neural networks: The case of bank failure predictions. Manage Sci. 1992;38(7):926-947.
   Available:https://doi.org/10.1287/mnsc.38.7.926
10. Shin KS, Lee TS, Kim HJ. An application of support vector machines in bankruptcy prediction model. Expert Syst. Appl. 2005; 28(1):127-135.
    Available:https://doi.org/10.1016/j.eswa.2004.08.009
11. Sun J, Li H. Data mining method for listed companies' financial distress prediction, Knowl Based Syst. 2008;21:1-5.
    Available:https://doi.org/10.1016/j.knosys.2006.11.003
12. Sung TK, Chang N, Lee G. Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction. J Manage Inform Syst. 1999;16(1):63-85.
    Available:https://doi.org/10.1080/07421222.1999.11518234
13. Yeh CC, Chi DJ, Hsu MF. A hybrid approach of DEA, rough set and support vector machines for business failure prediction. Expert Syst. Appl. 2010;37(2): 1535-1541.
    Available:https://doi.org/10.1016/j.eswa.2009.06.088
14. Chen MY. Predicting corporate financial distress based on integration of decision tree classification and logistic regression. Expert Syst. Appl. 2011;38:11261-11272.
    Available:https://doi.org/10.1016/j.eswa.2011.02.173
15. Sun J, He KY, Li H. SFFS-PC-NN optimized by genetic algorithm for dynamic prediction of financial distress with longitudinal data streams. Knowl Based Syst. 2011;24:1013–1023.

Available:https://doi.org/10.1016/j.knosys.2011.04.013

16. Hsieh TJ, Haiso HF, Yeh WC. Mining financial distress trend data using penalty guided support vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithm. Neurocomputing. 2012;82:196-206.
Available:https://doi.org/10.1016/j.neucom.2011.11.020

17. Chen SD, Jhuang S. Financial distress prediction using data mining techniques, ICIC Express Letters, Part B: Applications. 2018;9(2):131-136.
Available:https://doi.org/10.24507/icicelb.09.02.131

18. Jan CL, Haiso D. Detection of fraudulent financial statements using decision tree and artificial neural network. icic express letters, Part B: Applications. 2018;9(4): 347-352
Available:https://doi.org/10.24507/icicelb.09.04.347

19. Yeh CC, Chi DJ, Lin YR. Going-concern prediction using hybrid random forests and rough set approach. Inf. Sci. 2014;254:98–110.
Available:https://doi.org/10.1016/j.ins.2013.07.011

20. Goo YJ, Chi DJ, Shen ZD. Improving the prediction of going concern of Taiwanese listed companies using a hybrid of LASSO with data mining techniques. Springer Plus. 2016;5.
Available:https://doi.org/10.1186/s40064-016-2186-5

21. Jan CL. An Effective Financial Statements Fraud Detection Model for the Sustainable Development of Financial Markets: Evidence from Taiwan. Sustainability. 2018;10(2):1-14.
Available:https://doi.org/10.3390/su10020513

22. Chen SD. An effective going concern prediction model for the sustainability of enterprises and capital market development. Appl Econ. 2019;51(31): 3376-3388.
Available:https://doi.org/10.1080/00036846.2019.1578855

23. Sun J, Li H, Fujita H, Fu B, Ai W. Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. Inf Fusion. 2020;54:128–144.
Available:https://doi.org/10.1016/j.inffus.2019.07.006

24. Shen F, Liu Y, Wang R, Zhou W. A dynamic financial distress forecast model with multiple forecast results under unbalanced data environment. Knowl Based Syst. 2020;192:105365-105381.
Available:https://doi.org/10.1016/j.knosys.2019.105365

25. Huang YP, Yen MF. A new perspective of performance comparison among machine learning algorithms for financial distress prediction. Appl. Soft Comput. 2019;83:105663-105677.
Available:https://doi.org/10.1016/j.asoc.2019.105663

26. Quinlan JR. Introduction of decision trees. Mach Learn. 1986;1(1):81–106.
Available:https://doi.org/10.1007/BF00116251

27. Barros RC, Basgalupp M, Carvalho A, Freitas A. A survey of evolutionary algorithms for decision tree induction. IEEE T SYST MAN CY C. 2012;42(3):291-312.
Available:https://doi.org/10.1109/TSMCC.2011.2157494

28. Hocking RR. The analysis and selection of variables in linear regression. Biometrics. 1976;32:1–49.
Available:https://doi.org/10.2307/2529336

29. Tibshirani R. Regression Shrinkage and Selection via the LASSO. J R Stat Soc Series B Stat Methodol. 1996;58:267–288.
Available:http://doi.org/10.1111/j.2517-6161.1996.tb02080.x

30. Oyeyemi GM, Ogunjobi EO, Folorunsho AI. On performance of shrinkage methods — A monte carlo study. Int J Stat Appl. 2015;5(2):72–76.
Available:http://doi.org/10.5923/j.statistics.20150502.04

31. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Wadsworth International Group, Belmont, California; 1984

32. Breiman L. Random forests. Mach Learn. 2001;45(1):5－32.
Available:http://doi.org/10.1023/A:1010933404324

33. Hapfelmeier A, Ulm KA. New variable selection approach using Random Forests. Comput Stat Data Anal. 2013;60(C):50-69. Available:https://doi.org/10.1016/j.csda.2012.09.020
34. Choi H, Son H, Kim C. Predicting financial distress of contractors in the construction industry using ensemble learning. Expert Syst. Appl. 2018;110:1–10. Available:https://doi.org/10.1016/j.eswa.2018.05.026
35. Balasubramanian SA, Radhakrishna GS, Sridevi P. Natarajan T. Modeling corporate financial distress using financial and non-financial variables: The case of Indian listed companies. Int J Law Manag. 2019;61(3-4):457-484. DOI: 10.1108/IJLMA-04-2018-0078
36. Li Y, Li X, Xiang E, Djajadikerta HG. Financial distress, internal control and earnings management: Evidence from China. J Contemp Account Econ. 2020;16: 100210-100227. Available:https://doi.org/10.1016/j.jcae.2020.100210

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/59058