INDO-ASIAN JOURNAL OF INFORMATION RESEARCH AND TECHNOLOGY VOL 2 (01), JANUARY – JUNE 2025, PP. 1-13

AN EVALUATION OF MACHINE LEARNING TECHNIQUES FOR FORECASTING BANKRUPTCY IN INDIAN MANUFACTURING FIRMS POST-IBC 2016

Simrat Kaur
Research Scholar, Amity University Noida, Uttar Pradesh, India
Email Id : hks6999@gmail.com

ABSTRACT:

This paper's primary goal is to compare different machine-learning prediction techniques to discover how financial variables affect the likelihood of bankruptcy for Indian manufacturing enterprises. To anticipate bankruptcy following the implementation of the Insolvency and Bankruptcy Code (IBC) 2016, 15 financial variables are used. This article employs Logistic Regression, Artificial Neural Networks, Decision Trees, and Random Forest as machine learning predictive techniques for comparative analysis.

Keywords: *Bankruptcy, Machine learning, Logistic Regression, ANN, Decision Tree, Random Forest.*

INTRODUCTION

Given the impact of the global financial crisis on the global economy during the last decade, timely prediction of a business firm's failure is an essential issue in today's economic system (Di donato & Nieddu, 2016). Wruck in his paper contends that a firm may experience numerous stages before being declared dead, including financial distress, insolvency, filing for bankruptcy, and administrative receivership to avoid filing (Wruck, 1990).

In India, an industrial firm (one that has been in operation for at least five years) that has incurred losses equal to or greater than its whole net worth at the end of any fiscal year is sent to the Board for Industrial and Financial Reconstruction (BIFR) (Arup Roychoudhury, 2016). A series of Rules and Regulations created under this Act build up the legal framework. The legal system is still in development. On October 1, 2016, the Insolvency and Bankruptcy Board of India (IBBI) was formed as the regulator (BCAS, 2022).

The IBC (Insolvency and Bankruptcy Code) is a bankruptcy law that aims to reform and unify the following legal frameworks linked to the reorganization and insolvency resolution of businesses, partnership firms, and people, among others. Few research gaps have been identified while doing research on this research topic. One of the research gaps is that due to the implementation of IBC 2016, the research on bankruptcy data post-IBC 2016 is very limited. This paper focuses on the prediction of bankruptcy on the data post-IBC 2016. Another research gap is that research on bankruptcy prediction using machine learning techniques in India is very limited, especially in the field of finance. There is a need for more research using Machine learning techniques in other fields too. This paper demonstrates the significance and application of machine learning in today's world. The purpose of this paper is to demonstrate how machine learning has evolved for the benefit of society by demonstrating how it can be used in bankruptcy prediction. The models developed in this study could be utilized by investors, creditors, auditors, and others involved with a company to anticipate business failure.

Based on identified research gaps following research objectives are made:

R1: To identify the financial variables for the prediction of bankruptcy.

R2: To study the impact of financial variables on the bankruptcy prediction of manufacturing companies in India.

R3: To conduct a comparative analysis of different machine learning methods among Indian manufacturing companies.

LITERATURE REVIEW

Background

There has been significant research into the ability to foresee financial trouble for financial companies since FitzPatrick's initial work in the early 1930s (FitzPatrick, 1932). Beaver began by arguing that financial ratios can be useful in models for anticipating bankruptcy, financial distress, and individual firm failure (Beaver, 1966). Altman created the first model to predict bankruptcy in 1968. Altman created the Altman Z-score by combining five variables (Altman, 1968). According to his paper, the model's short-term accuracy was 95%, but when applied to two or more years prior to the bankruptcy, that figure drops to 72%. Ohlson and Zmijewski investigated the risk of bankruptcy using logit and probit models respectively (Ohlson, 1980; Zmijewski, 1984).

Machine Learning Techniques

Many methods, including artificial intelligence and statistical methods, have been used to anticipate company insolvency and financial crises; many researchers have demonstrated that artificial intelligence outperforms traditional statistical methods (Jerez et al., 2010). Logistic regression has been used in several corporate Financial Distress Prediction (FDP) research. Neural Networks are the new method nowadays to detect bankruptcy. According to Hertz's 1991 research, algorithm-based computer networks known as ANNs (Artificial Neural Networks) can be designed to imitate the human brain's internal functioning (Hertz et al., 1991). Wilson and Sharda created a revised neural network model to forecast corporate bankruptcy and show that it outperforms discriminant analysis (Wilson & Sharda, 1994). Because of its great mapping ability based on network design, neural network technology has several benefits over traditional statistical methods (Jo et al., 1997). Another method i.e., Decision Trees (DT) for FDP has been utilized in a number of researches, including Chen (Chen, 2011). By constructing a sequence of tree-based classification rules, DT models eventually divide a set of data into smaller sections (Halteh et al., 2018). It has been discovered that decision trees do not require much training, therefore the models they generate are simple to understand and can handle missing data and prevent data overfitting via tree pruning (Chi & Shen, 2022).

RESEARCH METHODOLOGY

Data Collection

Information about bankrupt companies is gathered from the IBBI website, where 115 listed companies were discovered to be bankrupt. 68 of the 115 companies were in the manufacturing sector. During the data collection process, it was discovered that 32 companies' data was missing. 36 remaining companies had data three years prior to bankruptcy.

TABLE 1: VARIABLES

Variables		
Dependent Variable: *Bankrupt or* *Non-Bankrupt*	Bankrupt Companies declared by NCLT from 2017 to 2023. Three-year prior data is taken. From 2017 to 2023, non-bankrupt companies based on the same industry and the total amount of assets selected, and three-year prior data is used.	
Independent Variables		
Working Capital to Total Assets	"Working Capital/Total Assets"	(Altman, 1968), (Ohlson, 1980)
Retained Earnings to Total Assets	"Retained Earnings/Total Assets"	(Altman, 1968)
EBIT to total assets	"EBIT/Total Assets"	(Altman, 1968)
Sales to Total Assets	"Sales/Total Assets"	(Altman, 1968), (Springate, 1978)
EBT to Current Liabilities	"EBT/Current Liabilities"	(Springate, 1978)
Debt to Asset ratio	"Total Debt/ Total Assets"	(Ohlson, 1980), (Zmijewski, 1984)
Current ratio	"Current Assets/ Current Liabilities"	(Zmijewski, 1984)
ROA (Return on Assets)	"Net Income/ Total Assets"	(Ohlson, 1980), (Zmijewski, 1984)
Quick ratio	"(Current Assets-inventory-prepaid expense)/ Current Liabilities"	(Mselmi et al., 2017)
Cash Flow ratio	"Cash Flow from Operations/ Total Debt"	(Ong et al., 2011)
Interest Coverage ratio	"EBITDA/ Interest Expense"	(Hu & Ansell, 2007)

Variables		
ROE (Return on Equity)	"Net Income/ Shareholders Equity"	(Chen, 2011), (Fallahpour et al., 2017)
EPS (Earning per share)	"(Net Income- Preferred Dividend)/ Average outstanding shares of the company"	(Chen, 2011)
Debt to Equity ratio	"Total Debt/ Shareholders Equity"	(Mselmi et al., 2017), (Hu & Ansell, 2007)
Cash ratio	"(Cash and cash equivalents+ short-term investments)/ Current Liabilities"	(Hu & Ansell, 2007), (Fallahpour et al., 2017)

Methodology

To predict bankruptcy, 15 financial variables are selected, represented in Table 1. Based on earlier studies, variables are selected. Based on the sector of the bankrupt company and the total value of the bankrupt company's assets, non-bankrupt companies are selected (Lakshan & Wijekoon, 2012).

Data from a total of 72 companies are being gathered for this study. Data from the previous three years is gathered from the annual reports of both non-bankrupt and bankrupt companies. 36 companies consist of 17 sectors of the manufacturing industry. Three companies are in the cable sector and paper sector. Four companies are in the textile industry and gas and petroleum industry. Four companies are in the steel industry, while two are in the mining industry. Three companies are in the auto ancillaries industries. 1 company from the automobile, chemical, and consumer durables. Three companies are in the agro-processing industry. 1 company each in the Pharma, electronics, Alcoholic Beverages, Plastics, and FMCG sectors. The final two companies are in the non-ferrous metal industry.

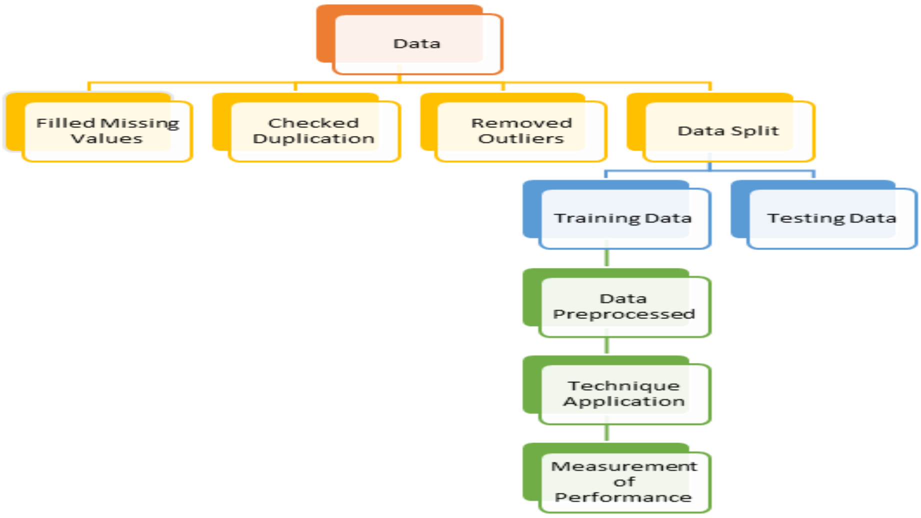

Figure 1: Methodology Applied

According to earlier research, Decision Tree-based models (DTs) are more accurate at predicting bankruptcy than more conventional models like Logistic Regression. In Python, supervised learning is used to predict bankruptcy. As shown in Fig. 1, the methodology applied for prediction is mentioned. Missing values were checked after importing the data into Python. There was a total of 8 missing values discovered in the dataset. The sum average was used to fill in the missing numbers. Duplication in the data was examined after missing data was sorted. It was discovered that there is no data duplication. Outliers were then removed. Boxplot was developed to assess whether or not there are outliers. For each of the four models, the data were split into training and testing periods. 30% of the data was tested for prediction, and 70% of the data was trained. The random state number was set to 42 (Luo et al., 2016). Standard Scaler was used to scale the data. This paper compares the accuracy of the selected predictive techniques i.e., Logistic Regression, Artificial Neural Networks, Decision Tree, and Random Forest.

DATA ANALYSIS

Binary logistic regression with a p-value of 0.05 is used to assess the impact of all identified financial variables on company bankruptcy. Table 2 shows the results of all the variables.

TABLE 2: BINARY LOGISTIC REGRESSION RESULTS

Variables	***B***	***Sig.***	***Decision***
Working Capital to Total Assets	-2.694	0.005	***Accepted***
Retained Earnings to Total Assets	0.000	0.534	Rejected

Variables	***B***	***Sig.***	***Decision***
EBIT to total assets	-28.964	0.000	***Accepted***
Sales to Total Assets	-1.28	0.593	Rejected
EBT to Current Liabilities	-20.887	0.000	***Accepted***
Debt to Asset ratio	2.818	0.001	***Accepted***
Current ratio	-1.265	0.002	***Accepted***
ROA	-46.641	0.001	***Accepted***
Quick ratio	-1.783	0.003	***Accepted***
Cash Flow ratio	-4.528	0.010	***Accepted***
Interest Coverage ratio	0.000	0.686	Rejected
ROE	0.129	0.378	Rejected
EPS	-0.476	0.002	***Accepted***
Debt to Equity ratio	-0.058	0.235	Rejected
Cash ratio	-19.976	0.014	***Accepted***

B values are the coefficients for the logistic regression equation that forecasts the dependent variable based on the independent variable. If it is negative, then there is an inverse relation between the dependent and independent variables. It is found that 5 financial variables have no impact on bankruptcy. Those variables are Sales to total assets, Retained earnings to total assets, ROE, Interest Coverage ratio, and Debt to equity. Other variables were found to be significant to the bankruptcy.

- There is a negative impact of Working capital to total assets on bankruptcy.
- There is a negative impact of EBIT to total assets on bankruptcy.
- There is a negative impact of EBT to Current liabilities on Bankruptcy.
- There is a negative impact of Current ratio on bankruptcy.
- There is a negative impact of ROA on bankruptcy.
- There is a negative impact of Quick ratio on bankruptcy.
- There is a negative impact of Cash Flow ratio on bankruptcy.
- There is a negative impact of EPS on bankruptcy.
- There is a negative impact of cash ratio on bankruptcy.
- The debt to asset ratio showed a positive impact on bankruptcy.

Ten financial variables are included for further comparative study in the prediction of bankruptcy.

EMPIRICAL FINDINGS

The outcomes of using machine learning methods in Python on testing data are shown in Table 3. If a company is bankrupt it is donated as yes and if not bankrupt, then it is donated as no. The performance was evaluated using the confusion matrix. As per the confusion matrix, the predicted data can be divided into True Positive (TP), False Positive (FP) which is also known as type-1 error, False Negative (FN) which is also known as type-2 error, and True Negative values. The degree to which the model accurately predicts whether the company is bankrupt (positive) is measured by the True Positive. False Positive is when a model incorrectly predicts a company is bankrupt (positive) when it actually does not (negative). When a model predicts a company as non-bankrupt (negative) when it is actually bankrupt (positive), it is called a false negative. The instances that the model correctly identifies a company as non-bankrupt (negative) are known as true negatives. The less there will be type-1 and type-2 error, the more accurate the model will be. Accuracy value has been the most used performance measurement but due to its non-identification of difference between different classes, in this paper, other performance measurements like Precision, Recall, and AUC (Area under the curve) value have also been used.

$$\textbf{Accuracy}=(TP+TN)/(TP+FN+TN+FP) \quad (1)$$

The proportion of true positives and true negatives to all the positive and negative data is referred to as the Accuracy of the model.

$$\textbf{Precision Value}=TP/(FP+TP) \quad (2)$$

The percentage of positively predicted instances that are actually correct is represented by the precision value of a model.

$$\textbf{Recall Value}=TP/(FN+TP) \quad (3)$$

The recall value of a model is a measurement of the model's accuracy in predicting positive instances from actual positive results.

AUC value is the area under the curve which represents the accuracy of the model's class predictions. Precision value, Recall value, Accuracy, and AUC values of all four models are mentioned. The results from the various predictive techniques are:

TABLE 3: PREDICTIVE TECHNIQUES

Techniques	***Precision***	***Recall***	***Accuracy (%)***	***AUC***
Logistic Regression	0.900	0.818	85.71	0.945
Artificial Neural Networks	0.833	0.909	85.45	0.945
Decision Tree	0.9090	0.909	90.47	0.945
Random Forest	1.000	0.818	90.47	0.950

As we can see from Table 3, the random forest and decision tree technique, which had the best accuracy of 90.47 percent, surpassed the other two predicting techniques, as shown by the findings. Logistic Regression outperformed the ANN techniques as Logistic regression has an accuracy of 85.71 whereas ANN has an accuracy of 85.45. In terms of Precision value, Random Forest has the highest value followed by Decision tree, Logistic Regression, and ANN. In terms of Recall, the decision tree having a value of 0.9 outperformed random forest and logistic regression, with ANN having the same value. Random forest have the highest AUC value i.e., 0.95 whereas all the other three models have the same AUC value of 0.945.

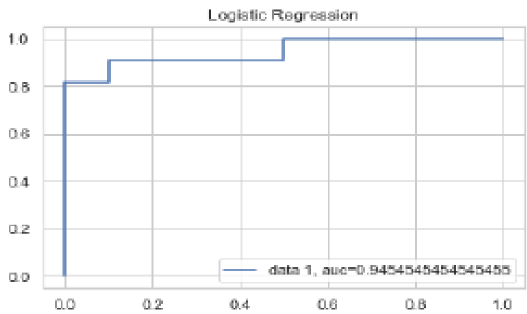

Figure 2: AUC value of Logistic Regression

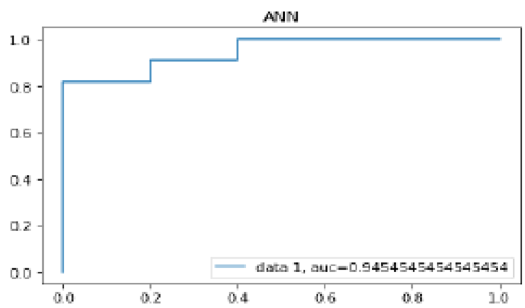

Figure 3: AUC value of ANN

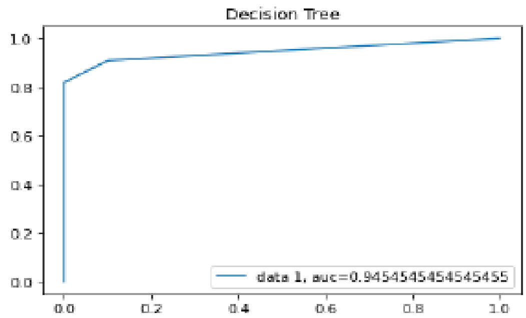

Figure 4: AUC value of Decision Tree

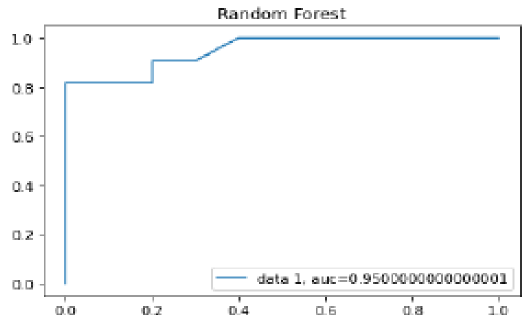

Figure 5: AUC value of Random Forest

Fig. 2, 3, 4, and 5 represent the AUC value of all four models under testing data. The X-axis represents the False-Positive rates (FPR) whereas the y-axis denotes True-positive rates (TPR).

$$\mathbf{FPR=FP/(TN+FP)} \quad \mathbf{(4)}$$

$$\mathbf{TPR=TP/(FN+TP)} \quad \mathbf{(5)}$$

False positive rate is the proportion of negative cases that were incorrectly classified, whereas true positive rate is the proportion of positive cases that were correctly classified also known as Recall value.

If the AUC value is more than 0.7 it is considered to be good. When the AUC is higher, the model performs better at differentiating between bankrupt and non-bankrupt companies. All four models, as seen in the figures, have AUC values of more than 0.7, however, the random forest has the highest AUC value at 0.95, whereas all the other techniques have the same AUC value of 0.94. There is not a very high difference between the AUC values of all four models which shows that all four models can be used for bankruptcy prediction of manufacturing companies in India.

CONCLUSION

One of the most essential and increasing exponentially areas of finance is bankruptcy prediction. The prediction methods employed must be reliable. Companies will be able to respond to an advanced warning of bankruptcy due to the accuracy of predictive methodologies. If a mechanism for forecasting insolvency is developed, investors will be able to choose whether or not to invest in a company. As per this research, Working Capital to Total Assets, Current Ratio, EBIT to Total Assets, EBT to Current Liabilities, Debt to Asset Ratio, ROA, Quick Ratio, Cash Flow Ratio, EPS, and Cash Ratio can all be utilized to predict bankruptcy in India's manufacturing industry. Sales to Total Assets, a turnover ratio that has been extensively used in prior research and the well-established models of Altman and Springate, did not affect this study.

In terms of predicting bankruptcy, random forest outperformed ANN, decision tree, and logistic regression in terms of AUC but based on accuracy, Random Forest, and Decision Tree both showed higher accuracy than ANN and Logistic Regression. ANN and Logistic regression showed the lowest results but in terms of Precision value, Logistic regression outperformed ANN whereas in previous research it was seen that ANN outperformed Logistic regression, In this study, we found out that logistic regression in terms of accuracy as well as Precision, surpassed ANN. For recall value, ANN surpassed Logistic Regression. For overall performance, Random Forest is the better prediction technique than others as it showed the highest accuracy, high precision value as well and high AUC value. After the random forest, the decision tree is another good predictive technique as it showed a higher recall value than the random forest and the same accuracy as the Random Forest. Therefore, it is shown through this study that DT models are better in prediction than other models.

FUTURE SCOPE

It has been observed that there is a difficulty in data availability in India, as data is not readily available following the implementation of IBC 2016. Because the data in this work is so restricted, researchers can anticipate bankruptcy using larger data and a longer time period in future studies. This study attempts to propose future research on the prediction of bankruptcy in other Indian sectors. In this paper, only financial variables are considered. When predicting bankruptcy in India, macroeconomic factors, market variables, and corporate governance indicators can also be considered as independent variables.

REFERENCES

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23(4), 589–609.

Arup Roychoudhury. (2016). Rajya Sabha passes Bankruptcy Code. Business Standard. https://www.business-standard.com/article/economy-policy/rajya-sabha-passes-bankruptcy-code-116051200075_1.html

BCAS. (2022). Insolvency and Bankruptcy Code, 2016 (IBC). BCAS REFERENCER. https://www.bcasonline.org/Referencer2018-19/part5/insolvency-and-bankruptcy-code-2016-ibc.html

Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of Accounting Research, 4, 71–111.

Chen, M. Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. Expert Systems with Applications, 38(9), 11261–11272. https://doi.org/10.1016/j.eswa.2011.02.173

Chi, D. J., & Shen, Z. De. (2022). Using Hybrid Artificial Intelligence and Machine Learning Technologies for Sustainability in Going-Concern Prediction. Sustainability, 14(3), 1810. https://doi.org/10.3390/su14031810

Di donato, F., & Nieddu, L. (2016). A new proposal to predict corporate bankruptcy in Italy during the 2008 economic crisis. In Causal Inference in Econometrics, 213–223. https://doi.org/10.1007/978-3-319-27284-9_13

Fallahpour, S., Lakvan, E. N., & Zadeh, M. H. (2017). Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem. Journal of Retailing and Consumer Services, 34(March 2016), 159–167. https://doi.org/10.1016/j.jretconser.2016.10.002

FitzPatrick, P. J. (1932). A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms. The Certified Public Accountant, 598–605.

Halteh, K., Kumar, K., & Gepp, A. (2018). Financial distress prediction of Islamic banks using tree-based stochastic techniques. Managerial Finance, 44(6), 759–773. https://doi.org/10.1108/MF-12-2016-0372

Hertz, J., Krogh, A., Palmer, R. G., & Horner, H. (1991). Introduction to the Theory of Neural Computation. Physics Today, 44(12), 70. https://doi.org/10.1063/1.2810360

Hu, Y. C., & Ansell, J. (2007). Measuring retail company performance using credit scoring techniques. European Journal of Operational Research, 183(3), 1595–1606. https://doi.org/10.1016/j.ejor.2006.09.101

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Artificial Intelligence in Medicine, 50(2), 105–115. https://doi.org/10.1016/J.ARTMED.2010.05.002

Jo, H., Han, I., & Lee, H. (1997). Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. Expert Systems with Applications, 13(2), 97–108. https://doi.org/10.1016/S0957-4174(97)00011-0

Lakshan, A. M. I., & Wijekoon, W. M. H. N. (2012). Predicting corporate failure of listed companies in Sri Lanka. GSTF Business Review (GBR), 2(1), 180–185.

Luo, W., Phung, D., Tran, T., Gupta, S., Rana, S., Karmakar, C., Shilton, A., Yearwood, J., Dimitrova, N., Ho, T. B., Venkatesh, S., & Berk, M. (2016). Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. Journal of Medical Internet Research, 18(12). https://doi.org/10.2196/jmir.5870

Mselmi, N., Lahiani, A., & Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. International Review of Financial Analysis, 50, 67–80. https://doi.org/10.1016/j.irfa.2017.02.004

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, 18(1), 109–131. https://doi.org/10.2307/2490395

Ong, S. W., Choong Yap, V., & Khong, R. W. L. (2011). Corporate failure prediction: a study of public listed companies in Malaysia. Managerial Finance, 37(6), 553–564. https://doi.org/10.1108/03074351111134745

Springate, G. L. V. (1978). Predicting the possibility of failure in a Canadian firm: Unpublished MBA Research Project/Simon Fraser University.

Wilson, R. L., & Sharda, R. (1994). Bankruptcy prediction using neural networks. Decision Support Systems, 11(5), 545–557. https://doi.org/10.1016/0167-9236(94)90024-8

Wruck, K. H. (1990). Financial distress, reorganization, and organizational efficiency. Journal of Financial Economics, 27(2), 419–444. https://doi.org/10.1016/0304-405X(90)90063-6

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