Early Detection on Company Bankruptcy: A Comparison of Neural Networks and Logistic Regression

Muhammad Fairus Ahmad Shukri[1] , Nor Hafizah Abdul Razak[2], MA MAT DIN[3]

[1,2,3]*College of Computing, Informatics and Mathematics, UiTM Kampus Sungai Petani, Malaysia*

Article Information

Received: 21-11-2024
Revised: 28-11-2024
Published: 05-12-2024

Keywords
Bankruptcy prediction; Neural Networks; Logistic Regression

****Correspondence Email:***
2023382969@student.uitm.edu.my

Abstract

Detecting firm insolvency at an early stage is crucial for financial analysis and risk management. This study compares the efficacy of two widely used bankruptcy prediction techniques: Neural Networks (NN) and Logistic Regression (LR). We evaluate each approach based on its accuracy, computing efficiency, and interpretability, aiming to identify a suitable predictive model that aligns with specific objectives, data characteristics, and the need for interpretability in financial decision-making. This research indicates that NN provides superior prediction accuracy but is accompanied by increased computing demands and reduced interpretability. In contrast, LR offers more speed, requires less processing resources, and provides explicit understanding of variable correlations; however, it may not perform well with intricate and nonlinear data. This study confirms the significance of choosing a suitable predictive model that balances competing demands of accuracy, efficiency, and interpretability in financial decision-making.

1. Introduction

The prediction of bankruptcy is a crucial aspect of financial decision-making that has been extensively studied over the years. The development of this topic began with researchers focusing on univariate analysis until the 1960s, where the first multivariate study emerged (Gissel et al., 2007) Since then, more than 500 studies have been published in this field using nearly 20 different statistical and machine learning tech including Neural Networks and logistic regression model (Shi & Li, 2019) as cited by (Brenes et al., 2022). It highlights that these methods have been successfully employed by researchers, such as by (Gissel et al., 2007), (Vochozka et al., 2020), and (Horak et al., 2020), to predict financial distress and bankruptcy.

The paper also notes the potential benefit of developing a dashboard to compare model performance, providing stakeholders with an easy-to-interpret overview of a model's effectiveness. The proposed financial analysis application aims to assist financial analysts, traders, investors, and those starting in finance by offering explanatory financial analyses combining fundamental and technical analysis. The study intends to evaluate the efficacy of neural networks and logistic regression models in predicting bankruptcy, considering criteria such as predicted accuracy, model interpretability, computational efficiency, and scalability.

This evaluation will provide insights into the relative strengths and weaknesses of each method for financial risk assessment in the context of bankruptcy prediction.

1.1 Literature Review

Financial stability is a crucial factor in the success of any company. The prediction of financially troubled enterprises has been extensively investigated in the financial arena, with the first studies using univariate analysis (Brenes et al., 2022). However, multivariate approaches have shown better results, and machine learning algorithms have emerged as powerful tools for predicting corporate failure.

The ability to predict corporate failure is essential for stakeholders, including investors, creditors, and regulators. Accurate predictions can help prevent financial crises, enable informed decision-making, and promote economic stability. Various studies have employed different methods, including logistic regression, neural networks, and genetic algorithms, to develop models that can accurately predict company failures. Chinedu et al. (2022) evaluated the efficacy of genetic algorithms vs neural networks in forecasting insolvency among Nigerian industrial firms. The genetic algorithm demonstrated superior performance compared to the neural network models; nevertheless, its efficacy may be constrained by inconsistent outcomes across different sectors.

Extensive research using cuckoo search algorithm which was powered by feed forward neural networks (FNN) by Marso & El Merouani, 2020. They found that FNN outperformed logistic regression models in predicting company failures based on Polish firms' data. However, they noted that the models developed would not be able to predict a company's failure in the long term (upwards of 3 years) as it is still difficult to predict. Gavurova et al., 2022 conducted research on predicting bankruptcy among non-financial corporations using a dataset from the Slovak Republic. They successfully identified key indicators, such as return on sales (ROS), quick ratio (QR), Net Working Capital to Assets ratio (NWC/A), and Personnel costs/sales (NWC/A), which can reduce the likelihood of bankruptcy. They utilized both logistic regression and neural networks to develop the model and found success with the highest success rate using neural networks and scaling techniques.

A comparative study on the performance of LR and ANN in predicting bankruptcy among small and medium sized enterprises in Morocco has been done by Zizi et al., 2021. They found that both methods were effective, but noted that ANN offers advantages in terms of its adaptive learning properties and high predictive capability and accuracy rates. Finally, Mishraz et al., 2021 conducted a similar study on the banking sector in India. They concluded that Artificial Neural Networks (ANN) is the best prediction model due to its effectiveness in dealing with complex data sets and ability to handle non-linear relationships.

2. Research Methods

The target population for this study is companies from the Taiwanese Bankruptcy Prediction dataset, obtained from the UC Irvine Machine Learning Repository. This dataset provides a rich source of information on factors contributing to corporate bankruptcy. The unit of analysis is individual companies or entities. In order to collect data for this study, the Taiwanese Bankruptcy Prediction dataset from the publicly available repository at UC Irvine Machine Learning Repository been retrieved. The dataset contains 95 features and 6819 instances with no missing values. This dataset will be the foundation for this analysis, providing valuable insights into factors that contribute to corporate bankruptcy.

Two primary methods to predict corporate failure namely; Logistic Regression (LR) and Artificial Neural Networks (ANN) been utilized in this work. Reasons wise in choosing these methods is LR is a widely used technique for binary classification problems, while ANN has been shown to be effective in handling complex data sets. Specifically, these two methods have been proven to be effective in similar studies. This phase involves gathering information related to the parameter selection that would allow for hyperparameter tuning methods and evaluation criteria that will be used to assess and forecast accuracy. This step includes the various

approaches for tuning the parameters in logistic regression and neural networks, such as solver, regularization parameters, number of layers and nodes per layer, activation function, optimizer. to ensure accurate model performance.

The pre-processing steps on the dataset been conducted to prepare the data for both modelling. First, identification and removal of outliers is executed. Feature selection was then performed to identify the most relevant factors that contribute to corporate bankruptcy. By analysing the correlation between each feature and the target variable (bankruptcy prediction), a selection of features subset that provided the most value for the modelling process is determined. Finally, division of the dataset into training and testing in 80:20, 70:30, 60:40 ratio been made to evaluate the models' performance effectively. Figure 1 depicts the outlines of the research process.

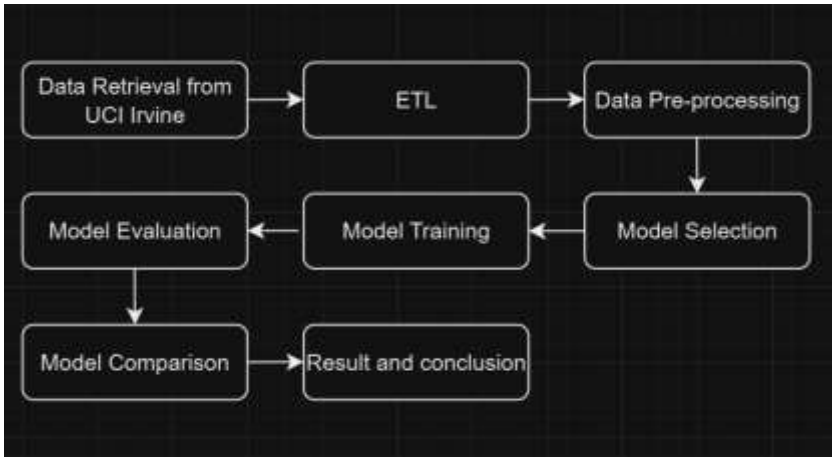

Fig. 1 Research Process

3. Results and Discussions

Discussion of the findings in this research was based on parameter settings, which are decided upon executing the built-up configurations of weka. Those parameters that were set was tabulated in Table 1, These are done in order to produce results based on each configuration by using the default settings for each model.

Table 1. Overview of Parameters

Parameters	NN	LR
Dataset Size	6.6k+	6.6k+
Train, Test Ratio	60:40, 70:30, 80:20	60:40, 70:30, 80:20
Attributes selection	Best First	Best First

The integration of the models will include rigorous testing to attain optimal results, while also assessing the model's responsiveness to various splits and its applicability to real-world data. Table 2 presents a general summary of the accuracy attained through the parameters.

Table 2. Accuracy Summary

Split	NN	LR
60:40	87.57%	92.08%
70:30	94.23%	92.03%
80:20	90.24%	92.45%

Table 2 presents initial performance measures, emphasising the efficacy of two machine learning models (NN and LR) across various data partitioning situations. Although these results exhibit potential, they are not conclusive, since more adjustments will be implemented to enhance model performance. We want to enhance these first findings by executing a series of experiments focused on optimising model parameters and mitigating class imbalance using strategic sampling methods. By implementing these upgrades, we expect more precise and dependable findings that more accurately represent the underlying connections in the data. This is to ensure that we take into account of the imbalance amount of data that is present in this data that is where the data of bankrupt companies is less compared to the companies that are not bankrupt.

4. Conclusions

This study provided valuable insights into the factors contributing to corporate bankruptcy in the Taiwanese context. By utilizing logistic regression and artificial neural network models, I would be able to identify key predictors of bankruptcy and develop effective prediction models. The findings might highlight the importance of financial ratios, market conditions, and operational metrics in assessing a company's risk of bankruptcy as the findings are still being worked on. The data pre-processing techniques employed, including outlier removal and feature selection, would demonstrate the significance of careful data preparation for accurate model performance. This study serves as a foundation for further research in this area, particularly in exploring additional factors that may influence corporate failure and improving prediction models. In future research, it would be worthwhile to examine more advanced machine learning techniques and compare their performance with the methods used in this study. Additionally, incorporating dynamic data and real-time analysis could provide more timely predictions and allow for early intervention strategies to prevent bankruptcy. Overall, this research contributes to the growing body of knowledge on corporate failure prediction and provides a basis for practitioners and researchers to develop proactive strategies for mitigating financial distress.

5. References

Brenes, R. F., Johannssen, A., & Chukhrova, N. (2022). An intelligent bankruptcy prediction model using a multilayer perceptron. *Intelligent Systems with Applications*, *16*, 200136. https://doi.org/10.1016/j.iswa.2022.200136

Gavurova, B., Jencova, S., Bacik, R., Miskufova, M., & Letkovsky, S. (2022). Artificial intelligence in predicting the bankruptcy of non-financial corporations. *Oeconomia Copernicana*, *13*(4), 1215–1251. https://doi.org/10.24136/oc.2022.035

Gissel, J. L., Giacomino, D., & Akers, M. D. (2007). A Review of Bankruptcy Prediction Studies: 1930-Present. *2007*, *5*.

Horak, J., Krulicky, T., Rowland, Z., & Machova, V. (2020). Creating a Comprehensive Method for the Evaluation of a Company. *Sustainability*, *12*(21), 9114. https://doi.org/10.3390/su12219114

Marso, S., & El Merouani, M. (2020). Predicting financial distress using hybrid feedforward neural network with cuckoo search algorithm. *Procedia Computer Science*, *170*, 1134–1140. https://doi.org/10.1016/j.procs.2020.03.054

Mishraz, N., Ashok, S., & Tandon, D. (2021). Predicting Financial Distress in the Indian Banking Sector: A Comparative Study Between the Logistic Regression, LDA and ANN Models. *Global Business Review*, 09721509211026785. https://doi.org/10.1177/09721509211026785

Shi, Y., & Li, X. (2019). An overview of bankruptcy prediction models for corporate firms: A Systematic literature review. *Intangible Capital*, *15*(2), 114. https://doi.org/10.3926/ic.1354

Vochozka, M., Vrbka, J., & Suler, P. (2020). Bankruptcy or Success? The Effective Prediction of a Company's Financial Development Using LSTM. *Sustainability*, *12*(18), 7529. https://doi.org/10.3390/su12187529

Zizi, Y., Jamali-Alaoui, A., El Goumi, B., Oudgou, M., & El Moudden, A. (2021). An Optimal Model of Financial Distress Prediction: A Comparative Study between Neural Networks and Logistic Regression. *Risks*, *9*(11), 200. https://doi.org/10.3390/risks9110200