Indonesian Banking Review
Volume 2, Number 2, 2023

Machine Learning Model for Small Bank Bankruptcy Prediction

Afdholul Ihsan Sundawa[1*]

[1] Universitas Diponegoro, Semarang, Indonesia

Abstract

Article history:
Received: July 9, 2023
Revised: August 22, 2023
Accepted: October 21, 2023
Published: December 30, 2023

Keywords:
Bank Bankruptcy,
Financial Prediction,
Machine Learning,
Small Banks.

Identifier:
Zera Open
Page: 126-143
https://zeraopen.com/journal/ibr

The bankruptcy of small banks can create serious disruptions to overall financial stability, especially in developing economies where the financial system is heavily supported by medium- and small-scale banking institutions. Unlike large commercial banks that often have diversified portfolios and greater access to capital buffers, small banks are typically more vulnerable to liquidity shocks, credit risks, and external market fluctuations. In this regard, advances in machine learning provide an opportunity to construct more robust prediction models that outperform traditional approaches, such as financial ratio analysis and logistic regression. This study explores various machine learning models, including Support Vector Machine, Random Forest, and Neural Network, while highlighting their strengths, weaknesses, and practical relevance in the context of small banks. By employing a Systematic Literature Review (SLR) method, the research systematically evaluates studies published and focus on bankruptcy prediction within the banking sector. The findings indicate that ensemble learning and deep learning models achieve higher predictive accuracy, though challenges in interpretability and transparency remain crucial for regulators and practitioners.

*Corresponding author:
(Afdholul Ihsan Sundawa)

1. Introduction

The bankruptcy of small banks is a crucial issue in maintaining the sustainability and stability of the global financial system. Banks not only function as custodians of public funds but also play an important role as financial intermediaries, liquidity providers, and financing channels, especially for small and medium-sized enterprises (SMEs) which are the backbone of the economy. If a small bank fails, the impact is not limited to that institution but can spread throughout the entire financial ecosystem. This condition has the potential to create broader instability and even disrupt public and investor confidence in the financial system. This situation is even more relevant, especially in developing countries, where small banks often dominate the formal financial sector structure. Therefore, the ability to predict small bank bankruptcy early is not just a preventive effort but also an urgent need for regulators, investors, and the bank managers themselves (Bucher et al., 2019).

Traditionally, the methods used to predict bank bankruptcy rely on financial ratio analysis and conventional statistical approaches, such as logistic regression. This model is indeed able to provide a clear enough picture of the risk factors and financial performance of a bank. However, prediction accuracy is often relatively low due to its limitations in capturing non-linear patterns and complex interactions in financial data. Along with the development of information technology, new opportunities have emerged through the application of Machine Learning (ML). ML algorithms have the ability to process large amounts of data, detect complex patterns, and produce predictions with a higher level of accuracy compared to classical approaches (Shetty et al., 2022).

Various recent studies show that the use of ML models such as Random Forest, Support Vector Machine, Gradient Boosting, and Neural Networks can significantly improve the performance of bank bankruptcy prediction (Petropoulos et al., 2020). Furthermore, ensemble learning approaches that combine the strengths of several algorithms have also been shown to produce superior predictive performance compared to single models (Siswoyo et al., 2022). However, the application of ML is not without challenges, especially concerning the aspect of interpretability. For regulators and policymakers, understanding the reasons or factors behind the prediction results is as important as the model's accuracy. Therefore, recent research has begun to integrate the concept of explainable AI (XAI) in bankruptcy prediction so that the results are more transparent and accountable (Park et al., 2021). In the context of small banks, the challenge of bankruptcy prediction is even more complex due to their limited operational scale, high dependence on the local economic sector, and high vulnerability to economic shocks.

Bragoli et al. (2022) emphasized that industry variables and local context must be seriously considered when building a prediction model. This indicates that the use of standard financial indicators alone is not enough. The prediction model also needs to incorporate external factors such as macroeconomic conditions, real sector dynamics, as well as monetary and fiscal policies. In addition, recent literature highlights the importance of a comparative evaluation between various ML models to understand the advantages and limitations of each method. For example, Shrivastav and Ramudu (2020) proved the effectiveness of Support Vector Machine

(SVM) in predicting bank bankruptcy in India. Meanwhile, Clement (2020) emphasized that traditional models, such as logistic regression, still have relevance when combined with modern techniques to strengthen prediction results. Thus, this study focuses on synthesizing recent literature through a Systematic Literature Review (SLR) approach to provide a comprehensive understanding of the trends, methods, and implications of using machine learning in predicting the bankruptcy of small banks.

2. Literature Review

2.1. Machine Learning in Bankruptcy Prediction

The development of recent literature shows a significant increase in the application of Machine Learning (ML) to predict bankruptcy, both at the corporate and banking institution levels. This shift occurred due to the ability of ML algorithms to overcome the various limitations that traditional methods have had. Conventional models such as logistic regression and financial ratio analysis are generally only able to capture linear relationships between variables, so they often fail to describe the real complexity in financial data (Balogun et al., 2022). In contrast, ML algorithms have the advantage of processing large and complex data, detecting non-linear patterns, and identifying interactions between variables that are difficult to see with classical approaches. Thus, the resulting prediction accuracy tends to be higher and more reliable (Boughaci & Alkhawaldeh, 2020).

Some popular algorithms that are often used in bankruptcy prediction research include Random Forest, Gradient Boosting, and Neural Networks (Du

Jardin, 2021). Random Forest works by combining many decision trees to produce more stable and consistent predictions. Gradient Boosting, on the other hand, improves model weaknesses gradually to improve overall performance. Meanwhile, Neural Networks are known to be effective at capturing complex patterns hidden in layered data. Various studies prove that these algorithms are not only effective in predicting the failure of large-scale banks but are also relevant for small banks that have limited historical data. With these advantages, machine learning is increasingly seen as a potential approach to improving early warning systems for bankruptcy in the banking sector.

2.2. Challenges and Issues of Interpretability

Although machine learning (ML) has been shown to have significant predictive advantages in detecting bankruptcy risk, a new and equally important problem has emerged: the issue of interpretability. High prediction accuracy is beneficial, but without a clear understanding of the reasons behind the prediction results, it is difficult for regulators and bank managers to make the right decisions. In the context of the financial sector, transparency and accountability are fundamental aspects, so the need to explain what factors contribute to the potential for bankruptcy is becoming more urgent. This makes interpretability a priority in modern research. In line with this need, recent literature emphasizes the importance of applying explainable AI (XAI) in the development of prediction models. The XAI concept aims to bridge the gap between the complexity of ML algorithms and the human need to understand the logic of the predictions produced (Machlev et al., 2022).

Thus, the prediction results are not only more accurate but also accountable to the supervisory authorities (Park et al., 2021). In addition, research by Lahmiri and Bekiros (2019) shows that the use of ML-based cognitive models can provide a more comprehensive explanation of bankruptcy risk. This model can map risk factors comprehensively, making the strategic decision-making process easier. However, major challenges remain, especially regarding data validation, which is the main foundation for ensuring the reliability of the model. Without valid data, the resulting interpretation risks being misleading, thus demanding a serious effort in financial data management and verification.

3. Methods

This study adopts the Systematic Literature Review (SLR) method as the main framework to identify, evaluate, and synthesize various literature related to small bank bankruptcy prediction with a Machine Learning (ML) approach. The choice of the SLR method was made because this approach offers a systematic, structured, and transparent process for assessing the development of recent research. In this way, the results obtained are not just a summary of the literature, but also a critical analysis that is able to provide academic and practical contributions. The implementation stages of the SLR begin with the process of identifying literature from various scientific databases, with the main focus on Google Scholar, Elsevier, and Researchgate as repositories of academic publications. The literature search was carried out using specific keywords such as “bankruptcy prediction”, “small banks”, and “machine learning”. The publication time frame was limited to the last five years, to ensure the relevance of the research to the current context, given the very rapid development of ML technology. The inclusion criteria were set strictly, namely only articles in the form of empirical research or literature reviews that explicitly discuss the application of machine learning to predict the bankruptcy of small-scale banks or financial institutions were considered.

The second stage involves the screening of articles. Screening is carried out based on the title, abstract, and keywords to ensure the suitability of the theme. Articles discussing non-bank companies or publications before 2020 were excluded from the analysis list. After this stage, a quality assessment was carried out by considering aspects of the research methodology, sample size, type of ML algorithm used, as well as the empirical and theoretical contributions presented by the author. The quality assessment is important to maintain the credibility of the synthesis results. In the final stage, data from the selected articles were extracted to be analyzed narratively. This analysis focused on three main dimensions, namely the trends in the research methodology used, the level of prediction accuracy achieved, and the issue of model interpretability which is often a challenge in the application of ML. With this approach, this study not only presents a map of literature development but also presents a critical synthesis of the various opportunities and obstacles that arise in the application of machine learning to support an early warning system for small bank bankruptcy.

4. Results

The results of a systematic review of the latest literature reveal that the application of Machine Learning (ML) in predicting the bankruptcy of small banks has experienced quite significant developments, especially in the last five years. From a series of analyses of articles that met the inclusion criteria, it can be concluded that the direction of current research is no longer limited to achieving higher prediction accuracy. Instead, the developing trend has moved to a broader domain, which includes other important issues such as the transparency of prediction results, model interpretability, and the relevance of the model to the industrial context and operational scale of the bank being studied. In other words, research on bankruptcy prediction through ML is now trying to answer needs not only from a technical perspective but also from a practical and regulatory perspective (Chen, 2018).

One of the main findings in the results of this study is that ensemble learning-based machine learning models have been shown to be able to provide more stable prediction results when compared to single models. A study conducted by Siswoyo et al. (2022) showed that a combination of several algorithms, for example, Random Forest, Gradient Boosting, and Neural Networks, can result in higher accuracy in detecting bankruptcy than if only one model is used. The main advantage of the ensemble learning method lies in its ability to reduce the level of variability or uncertainty in prediction results, which usually arises when a standalone model is applied. This becomes even more important in the context of small banks, because the available data is often limited in both quantity and quality. With the combination of models, the weaknesses in a certain algorithm can be compensated for by the strengths of other algorithms, so the results produced tend to be more balanced and reliable.

In addition to the use of ensemble learning, the review also showed that Support Vector Machine (SVM) remains one of the most widely used algorithms in bankruptcy prediction research. A study conducted by Shrivastav and Ramudu (2020) confirmed that SVM has the advantage of producing more precise classifications, especially for banks that are facing financial pressure in India. This finding is in line with the results of other studies that emphasize that SVM is very superior in handling datasets with a small size but a high dimension of variables. This characteristic is very relevant for small banks, considering that the number of their customers is usually limited, but the related financial data can be very diverse. Therefore, SVM is often seen as one of the best choices in the context of bankruptcy prediction in small-scale financial institutions.

However, although SVM is still widely used, there has also been a shift in research trends towards the use of deep learning models, especially Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM). These models are becoming more popular because of their ability to detect more complex non-linear patterns and capture temporal relationships in financial data. For example, Shetty et al. (2022) showed that neural networks can significantly increase the accuracy of bankruptcy prediction compared to classical methods such as logistic regression. This is especially true when the dataset includes dynamic macroeconomic variables, such as changes in inflation rates, capital market volatility, or interest rate policies. However, a major challenge of implementing deep learning is the limitation in terms of interpretability. For banking regulators, it's not just the prediction result that's important, but also the reason behind the prediction. The problem is that deep learning models are often seen as "black boxes" whose logic is difficult for non-technical users to understand.

The issue of interpretability has become one of the main concerns in recent literature. Park et al. (2021), for example, proposed the application of Explainable Artificial Intelligence (XAI) techniques to provide transparency to prediction results. By using XAI, the prediction model can show which variables contribute the most to the increase in bankruptcy risk. This allows regulators to understand more clearly, for example, whether the main factor causing the bankruptcy prediction comes from the non-performing loan ratio, a decrease in the level of liquidity, or excessive exposure to market volatility. With such an explanation, the prediction model is not only more accurate but also more accountable. This transparency also strengthens the accountability of the model and increases the level of user acceptance in the banking regulatory environment.

In addition to technical aspects, this systematic review also confirmed that local context and industry characteristics are very influential in building a bankruptcy prediction model. A study conducted by Bragoli et al. (2022) on small banks in Italy found that local industry variables have a significant influence on prediction accuracy. This finding shows that a universal prediction model cannot be directly applied to all countries or sectors. Instead, there needs to be an adjustment by considering local macroeconomic conditions, local regulations, and specific industry characteristics. In the context of Indonesia, for example, where Rural Credit Banks

(BPR) dominate the small banking sector, local factors such as dependence on the SME sector, domestic monetary policy, and the stability of the regional economy become very relevant additional variables to be included in the prediction model. Thus, generalizing the model without considering local aspects can actually reduce the accuracy and validity of the prediction results.

In addition to local relevance, recent research also reviewed real bankruptcy cases that have broad implications for the stability of the global financial system. For example, a study by Lahmiri and Bekiros (2019) which discusses the failure of Silicon Valley Bank (SVB) shows how machine learning models can be used to analyze the bankruptcy risk of even a large bank. Although SVB is not a small bank, the lesson that can be learned from the case is that external factors such as market volatility, sudden changes in depositor behavior, and exposure to certain sectors can increase a bank's vulnerability to bankruptcy. This lesson is important for the context of small banks because it confirms that prediction models should not only focus on internal financial indicators but must also be able to capture external dynamics that can affect operational stability.

Furthermore, the results of the literature review also highlight comparative studies between various ML algorithms. Boughaci and Alkhawaldeh (2020), for example, conducted a systematic comparison of several machine learning techniques in the banking context. Their research results show that no single algorithm is always superior in all conditions. For example, Random Forest was proven to be more effective when applied to large datasets with heterogeneous variables, while SVM showed superior performance on small datasets with more complex patterns. This finding confirms that the choice of algorithm should be adjusted to the characteristics of the available data, not based on the popularity of the algorithm alone. In addition, other research emphasizes the need to integrate financial stress variables into the prediction model. Petropoulos et al. (2020) stated that bankruptcy prediction models that only rely on internal bank financial data tend to have limitations.

Conversely, combining macroeconomic variables, market indicators, and internal financial data can produce more comprehensive predictions. This type of integrative approach is very relevant for regulators, because it allows them to anticipate potential financial crises with a broader and more comprehensive perspective. Although the accuracy of ML models shows a significant increase from time to time, this review also identifies that the challenge of data validation remains a very important issue. The financial data of small banks is often limited, poorly documented, or even has quality problems. This can lead to the risk of bias in the model training process. Clement (2020) emphasized the importance of applying cross-validation techniques and using independent datasets as an effort to reduce the risk of overfitting. Without adequate validation steps, the model may indeed show high accuracy on the test data, but the results could be misleading when applied in real conditions.

Another trend that emerges from the literature is the use of hybrid methods that combine classic statistical techniques with modern machine learning. The hybrid approach aims to maintain the transparency of traditional models while taking advantage of the predictive power of ML algorithms. For example, research by

Lahmiri and Bekiros (2019) shows that combining logistic regression with neural networks can result in a more balanced model between prediction accuracy and interpretability. This approach is becoming increasingly important because regulators not only demand accurate results but also an understandable explanation for the reasons behind the predictions. The results confirm that the application of machine learning provides a very significant contribution to improving the ability to predict the bankruptcy of small banks. Ensemble and deep learning models have been proven to have a higher level of accuracy, although challenges of interpretability and data limitations remain major obstacles. The use of explainable AI and the integration of local variables are proven to be potential solutions to answer these challenges.

5. Discussion

The results of this systematic review confirm that machine learning (ML) has brought about significant changes in the approach to predicting the bankruptcy of small banks. However, the findings from the literature also highlight a number of aspects that need to be discussed in more depth, both from the perspective of theoretical contributions, practical implications, and research limitations. Theoretically, the research analyzed enriches the literature by showing how ML algorithms are able to overcome the limitations of traditional methods. Models such as Random Forest, Gradient Boosting, SVM, and deep learning have been proven to be able to detect non-linear patterns that were previously difficult to capture with logistic regression or simple financial ratio analysis (Petropoulos et al., 2020; Shetty et al., 2022). Another theoretical contribution is the recognition that local context plays an important role in shaping the prediction model, as proven by Bragoli et al. (2022) in the case of Italian banking. Thus, the theory of bankruptcy prediction cannot be separated from macroeconomic factors and the industrial sector where small banks operate.

From a practical perspective, the application of ML offers tangible benefits for regulators and policymakers. With higher accuracy, the model can be used to provide an early warning of potential bankruptcy, so that mitigation steps can be taken before the risk spreads widely. The use of explainable AI (XAI) also provides an opportunity for regulators to understand the reasons behind the predictions, so that data-based decisions can be more accountable (Park et al., 2021). This is important because the banking sector has strict regulations, and every decision related to the stability of the financial system must be supported by clear arguments. However, this research also uncovers a number of limitations. One of the most prominent is the issue of data limitations. Small banks often do not have a data reporting system as good as large banks, so data quality and completeness are major challenges. This condition can lead to the risk of bias in the ML model, which has the potential to reduce the validity of the prediction results (Clement, 2020).

Another challenge is overfitting, which is a condition when the model shows very good performance on training data but fails when applied in the real world. To overcome this problem, several studies recommend cross-validation and the use of independent datasets in model testing. In addition, although deep learning models show high accuracy, interpretability remains a serious problem. Regulators need an understandable explanation for every prediction generated. Therefore, a combination of traditional methods with ML or the use of hybrid models is an interesting alternative that needs to be developed further (Lahmiri & Bekiros, 2019). This hybrid approach offers a balance between accuracy and transparency, both of which are very important in the banking context.

The discussion also emphasizes that future research needs to broaden its scope by including non-financial variables, such as customer behavior, regional socio-economic conditions, and the influence of government policies. This is especially relevant for small banks that often have strong exposure to the SME sector and the local economic conditions. By including these variables, the prediction model can become more comprehensive and relevant for practical decision-making. Thus, this discussion underscores that although ML has opened up great opportunities in small bank bankruptcy prediction, effective implementation still requires contextual adjustments, strict data validation strategies, and the integration of interpretability approaches. The future of research will likely move toward hybrid models and the integration of XAI, which can answer the needs of regulators while maintaining the high accuracy offered by machine learning.

6. Conclusion

This systematic review confirms that the application of machine learning has brought about significant advances in predicting the bankruptcy of small banks. Unlike traditional methods which tend to be limited to linear analysis, machine learning-based models are able to capture more complex patterns and produce more accurate predictions. A number of modern algorithms, such as Support Vector Machine, Random Forest, and various deep learning models, show strong performance in improving the accuracy of risk classification. In fact, the ensemble approach which combines more than one algorithm is proven to be able to provide more stable results because it can balance the weaknesses of each single model.

However, this review also found that there are a number of challenges that still need to be addressed. Data limitations at small banks are often a major obstacle, in addition to the risk of overfitting and the low level of interpretability of the prediction results. To answer this, recent research encourages the importance of model transparency, including through the integration of explainable AI which can help explain risk factors more clearly for regulators and decision-makers. Machine learning has great potential as an instrument for predicting the bankruptcy of small banks. However, in order to be implemented effectively, the model needs to be designed by considering a balance between accuracy, interpretability, and relevant local context.

References

Balogun, E. D., Ogunsola, K. O., & Ogunmokun, A. S. (2022). Developing an advanced predictive model for financial planning and analysis using machine learning. *IRE Journals*, *5*(11), 320-328.

Boughaci, D., & Alkhawaldeh, A. A. (2020). Appropriate machine learning techniques for credit scoring and bankruptcy prediction in banking and finance: A comparative study. *Risk and Decision Analysis*, *8*(1-2), 15-24.

Bragoli, D., Ferretti, C., Ganugi, P., Marseguerra, G., Mezzogori, D., & Zammori, F. (2022). Machine-learning models for bankruptcy prediction: do industrial variables matter?. *Spatial Economic Analysis*, *17*(2), 156-177.

Chen, J. M. (2018). Models for predicting business bankruptcies and their application to banking and financial regulation. *Penn St. L. Rev.*, *123*, 735.

Clement, C. (2020). Machine learning in bankruptcy prediction–a review. *Journal of Public Administration, Finance and Law*, (17), 178-196.

Du Jardin, P. (2021). Forecasting bankruptcy using biclustering and neural network-based ensembles. *Annals of Operations Research*, *299*(1), 531-566.

Lahmiri, S., & Bekiros, S. (2019). Can machine learning approaches predict corporate bankruptcy? Evidence from a qualitative experimental design. *Quantitative Finance*, *19*(9), 1569-1577.

Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI*, *9*, 100169.

Park, M. S., Son, H., Hyun, C., & Hwang, H. J. (2021). Explainability of machine learning models for bankruptcy prediction. *Ieee Access*, *9*, 124887-124899.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, *36*(3), 1092-1113.

Shetty, S., Musa, M., & Brédart, X. (2022). Bankruptcy prediction using machine learning techniques. *Journal of Risk and Financial Management*, *15*(1), 35-47.

Shrivastav, S. K., & Ramudu, P. J. (2020). Bankruptcy prediction and stress quantification using support vector machine: Evidence from Indian banks. *Risks*, *8*(2), 52-74.

Siswoyo, B., Abas, Z. A., Pee, A. N. C., Komalasari, R., & Suyatna, N. (2022). Ensemble machine learning algorithm optimization of bankruptcy prediction of bank. *IAES International Journal of Artificial Intelligence*, *11*(2), 679-686.