ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
https://www.jiisonline.org
J Intell Inform Syst 2023 June: 29(2): 241~265 http://dx.doi.org/10.13088/jiis.2023.29.2.241

Corporate Bankruptcy Prediction Model using Explainable AI-based Feature Selection

Gundoo Moon
Department of MIS, Graduate School, Dongguk University_Seoul, Korea
(sidhid@dongguk.edu)

Kyoung-jae Kim
Department of MIS, Business School, Dongguk University_Seoul, Korea
(kjkim@dongguk.edu)

A corporate insolvency prediction model serves as a vital tool for objectively monitoring the financial condition of companies. It enables timely warnings, facilitates responsive actions, and supports the formulation of effective management strategies to mitigate bankruptcy risks and enhance performance. Investors and financial institutions utilize default prediction models to minimize financial losses. As the interest in utilizing artificial intelligence (AI) technology for corporate insolvency prediction grows, extensive research has been conducted in this domain. However, there is an increasing demand for explainable AI models in corporate insolvency prediction, emphasizing interpretability and reliability. The SHAP (SHapley Additive exPlanations) technique has gained significant popularity and has demonstrated strong performance in various applications. Nonetheless, it has limitations such as computational cost, processing time, and scalability concerns based on the number of variables. This study introduces a novel approach to variable selection that reduces the number of variables by averaging SHAP values from bootstrapped data subsets instead of using the entire dataset. This technique aims to improve computational efficiency while maintaining excellent predictive performance. To obtain classification results, we aim to train random forest, XGBoost, and C5.0 models using carefully selected variables with high interpretability. The classification accuracy of the ensemble model, generated through soft voting as the goal of high-performance model design, is compared with the individual models. The study leverages data from 1,698 Korean light industrial companies and employs bootstrapping to create distinct data groups. Logistic Regression is employed to calculate SHAP values for each data group, and their averages are computed to derive the final SHAP values. The proposed model enhances interpretability and aims to achieve superior predictive performance.

Keywords : Explainable AI, Corporate Bankruptcy Prediction, SHAP, Random Forest, XGBoost

Received : May 24, 2023 Revised : June 9, 2023 Accepted : June 12, 2023
Publication Type : Regular Track Corresponding Author : Kyoung-jae Kim

1. Introduction

Credit assessment is a critical process in the financial industry, traditionally relying on subjective judgments from experts. However, with the emergence of big data, there has been a growing trend towards utilizing quantitative data analysis for credit assessment. Machine learning techniques, specifically those capable of handling large-scale datasets, have gained significant traction and recognition within this field. Despite the success of machine learning in credit assessment, a key limitation remains: the lack of interpretability. Machine learning algorithms learn from data to create predictive models but do not offer explicit explanations of the criteria used for predictions. This lack of interpretability presents significant challenges in practical applications, such as explaining credit decisions to customers or regulators.

Enterprises play a vital role in national economies, as they generate profits, create jobs, and drive innovation. However, when enterprises experience failures, the consequences can have a significant impact on the economy. Stakeholders, including employees, creditors, investors, and financial institutions, may suffer economic losses. Moreover, the overall productivity declines, and creditworthiness deteriorates, negatively affecting the national economy. Predicting corporate insolvency not only enables the development of improved management strategies but also enhances overall performance and mitigates insolvency risks. Additionally, it allows stakeholders to minimize losses through timely interventions, empowers financial institutions to make well-informed decisions regarding fund allocation to fortify the stability of their assets, and enables governments to leverage insolvency prediction models for financial regulation and institutional design. Against this backdrop, the academic community has actively pursued research on predicting corporate insolvency, employing various methodologies over an extended period. Traditional approaches have predominantly focused on machine learning and deep learning techniques to design models for forecasting corporate insolvency. While traditional machine learning techniques may encounter computational constraints and high costs associated with feature extraction, they have demonstrated exceptional problem-solving capabilities and have found widespread applications across diverse domains, including numerous studies on predicting corporate insolvency.

However, despite their remarkable performance, the intricate information processing involved in machine learning poses challenges in terms of interpretability. The inherent “black box” nature of machine learning models impedes the comprehension and acceptance of output results without providing rational explanations regarding the influence of input variables on the outcomes (Guidotti et al., 2018; Hwang, 2021). The limited ability to offer clear explanations of the internal workings and predicted results diminishes the trustworthiness of these outcomes for users (Lysaght et al., 2019; Bastani et al., 2017). Given these challenges of machine learning and the increasing potential for uncontrollable ramifications and social risks associated with the high impact of AI, the need for explainable artificial intelligence (XAI) has been proposed as a means to overcome the limitations of machine learning by providing statistically driven predictions with interpretability (Chun et al., 2021).

The primary objective of this study is to develop a predictive model for corporate bankruptcy using XAI techniques that focus on selecting variables with high explanatory power. The research utilizes Korean light industrial data and employs the bootstrap technique to create data groups. Subsequently, Logistic Regression is utilized to estimate SHAP values. The final classification and prediction outcomes are evaluated by applying the Random Forest, XGBoost, and C5.0 algorithms. To construct the optimal prediction model, an ensemble model is created through Soft Voting. This methodological approach draws inspiration from the research conducted by Utkin and Konstantinov (2022) and incorporates a modified version of the SHAP technique based on the Random Forest framework.

2. Research Background

2.1. Explainable AI

The concept of "Explainable AI" (XAI) originated in 1975 to describe systems capable of providing explanations for their decision-making processes. However, it was formally established as the term "XAI" in 2004 through the work of Michel Van Lent, William Fisher, and Michael Mancuso (Shortliffe and Buchanan, 2004; Van Lent et al., 2004). The importance of explainability gained prominence in the mid-1970s as researchers explored explanations within expert systems (Ahn and Cho, 2021). The emergence of deep machine learning models employing complex neural networks has brought about significant difficulties when compared to earlier machine learning approaches grounded in human-designed logic. The inherent complexity of deep machine learning, with its autonomous learning and construction of intricate neural networks, resulted in reduced comprehensibility of the underlying processes and outputs. Additionally, identifying specific causal factors in cases of model failure became challenging. To address these limitations, the field of XAI emerged, aiming to deepen users' understanding of the model's results through analysis of the learning process and factors influencing decision-making. This approach aims to overcome the black-box problem and bridge the understanding gap between users and developers (Linardatos et al., 2020).

The origins of XAI can be traced back to the Defense Advanced Research Projects Agency (DARPA) in the United States. DARPA initiated XAI development by modifying non-player character (NPC) artificial intelligence within a military simulation program to provide explanations for NPC behavior. The XAI system consisted of separate command and control AI components. The control AI organized situational data into vector format, while the command AI perceived the situation and issued orders based on the resulting values. By analyzing and acting upon these instructions, military officers engaged in simulated combat were able to make informed decisions supported by insights derived from XAI (Oh et al., 2022). Since then, DARPA has actively pursued XAI research to integrate explainable AI into existing machine learning models, incorporate human-computer interaction (HCI) capabilities into machine learning systems, and leverage XAI to enhance situational understanding (Oh et al., 2022).

XAI, also referred to as Interpretable AI or Transparent AI, has emerged as a prominent research field since 2016, attracting significant attention. The primary objective of XAI is to enhance the transparency of AI models by providing explanations to users, facilitating a deeper understanding of the underlying decision-making processes (Adadi and

Berrada, 2018). In the present era of widespread AI application across diverse domains, comprehending the inner workings of advanced AI systems has become increasingly important. As a result, the significance of XAI continues to grow steadily. XAI plays a pivotal role in improving the comprehensibility of complex AI systems for both developers and users, facilitating effective communication and bridging the gap between domain experts and non-experts (Lee et al., 2021; Oh et al., 2021). By promoting transparency and interpretability, XAI empowers users to embrace and leverage sophisticated AI technologies with greater confidence. It enhances the trustworthiness and readability of AI-generated outputs, promoting rational interaction between humans and AI. Furthermore, XAI assists users in making well-informed decisions by providing explanations for the decision-making processes of AI models. Additionally, XAI aids in identifying and rectifying flaws in decision-making systems, thereby optimizing performance and mitigating issues related to bias and discrimination (Tjoa and Guan, 2020).

XAI techniques can be broadly categorized into feature-based models and example-based models, as discussed by Ahn and Cho (2021) and Chu et al. (2022). Feature-based models provide insight into the decision-making process by explicitly highlighting the importance of individual features contributing to the model's output. Conversely, example-based models offer explanations by quantitatively measuring the impact of specific cases on the model's output. Incorporating these interpretative features into AI models enhances their reliability and fosters a comprehensive understanding. Consequently, XAI exhibits the potential to enhance the predictive performance of AI models in real-world applications, leading to increased accuracy and effectiveness.

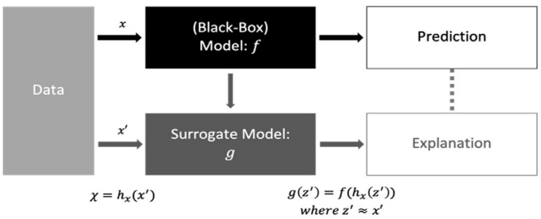
Several techniques are available for XAI, with LIME and SHAP serving as representative surrogate analysis methods. In this study, SHAP was employed as a method for quantifying the contribution of variables and representing them as values. The SHAP method, introduced by Lundberg and Lee (2017), leverages surrogate analysis to enhance causality and provide interpretability for predictions in black box models. This is accomplished through the utilization of intuitive and comprehensible Shapley values. SHAP possesses the capability to offer both local and global explanations, thereby promoting transparency and revealing the underlying mechanisms of the model (Son, 2022). By determining feature importance based on the derived SHAP values, the method effectively showcases the positive or negative contributions of each feature towards the final prediction.

<Figure 1> outlines the holistic process of the SHAP method. Initially, a black box model denoted as f is utilized to generate predictions using input values x. Subsequently, the process involves the identification of a surrogate model g by introducing transformed input values x', which differ from the original data. This strategy facilitates the interpretation of the entire dataset. SHAP values are defined by calculating the conditional mean of f, rather than relying on the actual prediction, utilizing simplified input values. Lundberg and Lee (2017) proposed a formulation for SHAP values that satisfies the

desired properties of local accuracy, missingness, and consistency as indicated below:

$$\varnothing_i(f,x) = \sum_{z \subseteq x'} \frac{|z'|!\,(M - |z'| - 1)!}{M!}[f_x(z') - f_x(z' \setminus i)]$$

In the provided expression, $\varnothing_i$ represents the SHAP value, f denotes the original model, x represents the input for f, and $f_x(z') - f_x(z' \setminus i)$ signifies the contribution of variable i(Bussmann et al., 2020).



〈Figure 1〉 SHAP Process

The utilization of SHAP values for variable selection confers the advantage of quantifying the influence exerted by each variable on the prediction outcome, thereby facilitating the determination of variable importance. This approach proves valuable in improving the predictive performance of the model through the selection of variables based on their statistical significance. Variable selection assumes a pivotal role in the development and optimization of prediction models, enabling the reduction of model complexity and enhancement of prediction accuracy. Consequently, the adoption of SHAP-based variable selection yields significant advantages over subjective methods reliant on experience and intuition, as it effectively enhances the predictive performance of the model.

2.2. Prior Research on Corporate Failure Prediction

2.2.1. Feature Selection

The selection of variables for the corporate bankruptcy prediction model relies on objective and statistically tractable characteristics found in financial statements. Considering the vast number of potential variables introduces complexity to the model, necessitates extensive computation time and cost, and can lead to overfitting issues. Hence, feature selection is employed during the modeling process to enhance the overall model.

In <Table 1>, various variable selection methods employed in corporate bankruptcy prediction studies are summarized. Lee and Choi (2013) utilized t-tests and correlation analysis to identify important variables associated with bankruptcy, recognizing that the significance of variable groups varies across industries. Kasgari et al. (2013) developed a corporate failure prediction model for Iranian companies, comparing artificial neural networks (ANN) with other techniques and using Garson’s Algorithm for variable selection. Zhou et al. (2012) employed the accurate Genetic Algorithm (GA) to select variables for constructing an empirical model for corporate financial distress prediction. Gordini (2014) also utilized GA to design a bankruptcy prediction model for SMEs. Liang et al. (2015) extensively examined the effectiveness of filter and wrapper-based feature selection methods for financial crisis prediction research, highlighting the improved prediction performance achieved through feature selection using GA and Logistic Regression. Jeong et al. (2012) constructed a bankruptcy prediction model using a neural network and employed GAM and GA for variable extraction. Yeh et al. (2014) incorporated Intellectual Capital as a predictive variable and utilized Random Forest (RF) and Random Subspace Trees (RST) to enhance the accuracy of corporate survival prediction. Khademolqorani et al. (2015) adopted the Multiple Attribute decision-making method to develop a bankruptcy prediction model and employed Factor Analysis for variable selection. Bems et al. (2015) introduced the Gini Index as a variable in credit scoring used for bankruptcy prediction. Iturriaga and Sanz (2015) presented a neural network-based bankruptcy prediction model with variable selection using the Gini Index and Mann-Whitney test. Various feature selection methods have been employed in corporate failure prediction studies, emphasizing the substantial impact of feature selection on prediction performance. Brenes et al. (2022) employed Univariate Analysis, parametric tests, and Hybrid variable selection to create a bankruptcy detection model using a multilayer perceptron. Kwon et al. (2017) applied Recurrent Neural Networks (RNN) to investigate the dynamic changes in accounting information for corporate bankruptcy prediction models and conducted variable selection based on Altman's Z-score. Cho et al. (2022) compared the performance of a variable selection method utilizing GA for generating counterfactual explanatory examples in bankruptcy prediction models, which were subsequently trained using Artificial Neural Networks, along with the LIME technique.

Feature selection constitutes a crucial phase in the modeling process as it aims to identify the most impactful features that significantly influence the model's performance. Traditional feature selection methods possess limitations in terms of explaining individual predictions made by the model. Consequently, discerning the specific features that contribute most to a prediction becomes challenging, thereby impeding interpretability. To address this issue and enhance the interpretability of the model's predictions, extracting features with substantial contributions, known as SHAP values, proves advantageous.

In this study, the significance of variable selection was duly acknowledged, with the objective of constructing an efficient and accurate bankruptcy prediction model through the identification of variables exhibiting high SHAP values, indicative of their explanatory power. We adopted a feature selection method based on SHAP values, which has received limited attention in previous research. SHAP values belong to a class of additive feature contribution values that possess desirable properties in terms of accuracy, missingness, and consistency. They allocate feature importance in a manner more consistent with human intuition (Gebreyesus et al., 2023). Furthermore, employing a feature selection method based on SHAP values not only facilitates effective modeling and provides insights into the data but also reduces computational costs while enhancing interpretability. A summary of previous studies utilizing SHAP feature selection is presented in <Table 2>. The data collection for these studies spanned from 2021 onwards, since this methos is relatively recent.

〈Table 1〉 Prior Research on Feature Selection

Year	Author	Variable Selection Method
2012	Jeong, Min, Kim	GAM, GA
2013	Kasgari, Divsalar, Javid, Ebrahimian	Garson's algorithm
2013	Lee, Choi	T-test, Correlation Analysis
2014	Gordini	GA
2014	Yeh, Chi, Lin	RF, RST
2014	Zhou, Lai, Yen	GA
2015	Bems, Macas, Zegklitz, Posik	Gini Index
2015	Khademolqorani, Hamadani, Refiei	Factor Analysis
2015	Iturrjaga, Sanz	Mann-Whitney Test, Gini Index
2015	Liang, Tsai, Wu	GA
2017	Kwon, Lee, Shin	Altman's Z-score
2022	Brenes, Johannssen, Chukhrova	Univariate Analysis, Parametric Test, Hybrid Variable Selection
2022	Cho, Shin	GA

GA: Genetic Algorithm GAM: Generalized Additive Model RF: Random Forest RST: Rough Set Theory

〈Table 2〉 Prior Research on SHAP Feature Selection

Year	Researcher	Research Topic
2021	Rouhi, Clausel, Oster, Lauer	Atrial fibrillation detection using interpretable feature selection using SHAP
2022	Le, Ho, Nguyen, Chang	Predicting an improved sequence-based DNA promoter using BERT pre-trained model and SHAP feature selection
2022	Liu, Liu, Luo, Zhao	Dianosis of Parkinson's disease, using SHAP value feature selection of Parkinson's disease medical dataset.
2022	Liu, Shen, Wang, Zhang, Zhu	Predicting RNA m5C sites based on feature selection of XGBoost and SHAP
2022	Chinnici, Gebreyesus, Dalton, Nixon, Chiara	Introduction of SHAP feature selection using Data Center dataset to optimize machine learning for Data Center
2023	Ni, Chen, Zhu, Pang, Wang, Yang	Prediction and interpretation of gamma pass rate based on SHAP feature selection

2.2.2. Corporate Failure Prediction Methodology

Corporate failure refers to a state wherein a company encounters a deteriorating financial condition, rendering it unable to recover due to reasons such as bankruptcy or insolvency. It signifies a situation where a company's total liabilities surpass the total value of its assets, resulting in significant financial challenges. The causes of corporate failure can vary, including factors such as mismanagement decisions, economic downturns, and industry transformations. Research in the field of corporate bankruptcy prediction primarily focuses on two key areas: exploring input variables that serve as indicators of corporate insolvency causes and developing models utilizing novel techniques. However, with concerns surrounding the reliability of variables and data collection, there has been a growing emphasis on the development of AI-based models, rather than solely relying on variable exploration.

Hwangbo and Moon (2016) conducted a research study utilizing financial data from delisted small manufacturing companies in the KOSDAQ market. Their study involved the development of Multivariate Discriminant Analysis models, Logistic Regression analysis models, and Artificial Neural Network analysis models. The sample consisted of a total of 166 companies, comprising 83 insolvent companies and 83 solvent companies that were delisted between 2009 and 2012. Comparative analysis of the models' predictive power revealed that the Logistic Regression analysis model exhibited the highest classification accuracy for the training sample, while the Artificial Neural Network model demonstrated the highest classification accuracy for the validation sample.

Kim (2016) proposed a study with the aim of developing a credit rating evaluation model for 1,295 domestic listed companies. The chosen methodology for this study was Random Forest. To compare and evaluate the performance of the proposed technique, traditional classification methods such as Multivariate discriminant analysis, artificial Neural Networks, and multiclass SVM models were employed. The empirical analysis confirmed the superiority of the proposed approach, as it yielded more accurate prediction outcomes compared to the traditional methods.

In a study by Mattsson and Steinert (2017), three machine learning algorithms, namely Random Forest, gradient boosting, and artificial Neural Network, were utilized to predict corporate bankruptcies. The study focused on Polish companies from 2000 to 2013, using 64 different financial ratios as the basis for prediction. The results indicated that tree-based algorithms demonstrated superior performance.

Park et al. (2017) conducted a research study using a Logistic Regression model to predict the insolvency of Korean shipping companies. The focus of the study was to validate the predictive accuracy by examining cases of shipping companies that had filed for court receivership. The findings revealed that the Logistic Regression model, tailored to the unique characteristics of the shipping industry, exhibited exceptional predictive power.

In the research conducted by Wagenmans (2017), the study aimed to analyze the predictive strength of payment behavior data and evaluate the performance of machine learning techniques in a realistic context. The available data from a pension fund was appropriately structured, and predictive models were developed using Logistic Regression, Random Forest, and Decision Tree algorithms. The findings underscored the remarkable performance of Random Forest, ultimately concluding its superiority over the other models.

Joshi et al. (2018) developed a predictive model for corporate bankruptcy by employing a genetic algorithm to select the most influential ratios. They utilized the Random Forest algorithm to classify companies into bankrupt and non-bankrupt categories.

The research conducted by Rustam and Satagih (2018) aimed to predict bank failures in Turkey by employing Random Forest as a classifier. The results revealed that Random Forest demonstrated an impressive training performance, achieving a perfect accuracy rate of 100%. During the testing phase, Random Forest exhibited an accuracy of 94% for the overall ratio and an accuracy of 96% when considering a subset of six ratios. These outcomes highlight the ability of Random Forest to effectively measure and identify significant variables for predicting bank failures.

Qu et al. (2019) proposed a bankruptcy prediction model that incorporates various techniques such as MDA, Logistic Regression, Ensemble methods, Neural Networks, Support Vector Machines, DBN, and CNN. Son (2019) designed bankruptcy prediction models based on financial statements using Logistic Regression, Random Forest, gradient boosting, and Neural Networks. Both studies commonly employed Logistic Regression and Ensemble techniques.

Kim (2019) developed a bankruptcy prediction model for domestic restaurant businesses utilizing Logistic Regression analysis. Similarly, Choi (2019) constructed an AI-based bankruptcy prediction model that incorporated news information as a foundational input, employing Logistic Regression analysis as well. Zhao (2020) collected financial datasets and proposed predictive models using deep learning techniques specifically designed for financial institutions. Kim and Upneja. (2021) aimed to construct prediction models for American restaurant data using Ensemble methods. Furthermore, Song et al. (2021) proposed research on bankruptcy prediction for both external and non-external firms, employing various techniques such as discriminant analysis, Logistic Regression, Decision Trees, Random Forest, adaptive boosting, gradient boosting, and Neural Networks. The research findings revealed that discriminant analysis and Random Forest demonstrated superior predictive power for external firm prediction, while Neural Networks exhibited strong predictive power for non-external firms.

A comprehensive analysis indicates that machine learning techniques outperform traditional model estimation methods such as discriminant analysis and Logistic Regression in the context of predicting corporate failure. Cho et al. (2021) utilized deep learning techniques to predict bankruptcy for selected distressed companies and compared the predictive performance against other machine learning Ensemble models. The experimental results demonstrated that the Random Forest model outperformed in terms of accuracy and precision, surpassing 90% across all evaluation metrics, while the deep learning model exhibited exceptional recall performance.

Brenes et al. (2022) conducted a literature review on bankruptcy prediction models utilizing Taiwanese corporate datasets and investigated the identification capability of multilayer perceptron. Carmona (2022) developed a bankruptcy prediction model using XGBoost applied to a sample of French companies and validated its performance. Kang (2023) proposed a study comparing Altman's K-Score with artificial Neural Networks for corporate bankruptcy prediction. The experimental results demonstrated that artificial

〈Table 3〉 Prior Research on Corporate Failure Prediction

Year	Researcher	LR	DT	NN	RF	SVM	RG	LS	BT
2016	Kim, Ahn			●	●	●			
2016	Hwangbo, Moon	●		●					
2017	Mattsson, Steinert			●	●				●
2017	Park, Kim, Kwon	●							
2017	Wagenmans	●	●		●				
2018	Joshi, Ramesh, Tahsildar				●				
2018	Rustam, Saragih				●				
2019	Qu, Quan, Lei, Shi	●		●		●			
2019	Son, Hyun, Phan, Hwang	●			●				●
2019	Kim	●							
2019	Choi	●							
2020	Zhao			●					
2021	Kim, Upneja	●	●						
2021	Song, Park, Lee		●	●	●				●
2021	Cho, Ahn, Kim	●			●	●			
2022	Brenes, Johannssen, Chukhrova			●		●			
2022	Carmona								●
2022	Kwon, Park	●							
2022	Kwon, Park	●			●	●	●	●	
2022	Kim, Jo, Yoo								●
2022	Heo, Baek			●	●	●			●
2023	Kang			●					

LR: Logistic Regression
DT: Decision Tree
NN: Neural Network
RF: Random Forest
SVM: Support Vector Machine
RG: Ridge
LS: Lasso
BT: Boosting

Neural Networks achieved a prediction accuracy 17.88% higher than that of K-Score. Kwon and Park (2022) conducted a study to assess the risk of bankruptcy in the hotel industry, taking into consideration the financial stability of hotels affected by the COVID-19 pandemic. They developed a bankruptcy prediction model specifically tailored to the domestic hotel industry using financial variables, ratios, and control variables. By setting a threshold probability, companies were identified as failed when the bankruptcy probability exceeded the threshold probability.

Kwon and Park (2022) also conducted a study focusing on detecting and proactively responding to repetitive bankruptcy cases in the shipping industry, specifically South Korean shipping companies. They utilized financial indicators and shipping market conditions as variables and employed five machine learning techniques for bankruptcy prediction in the shipping industry. The performance comparison revealed that Lasso and Logistic Regression exhibited superior predictive performance compared to other techniques. The integration of machine learning techniques enhanced the predictive power of the bankruptcy prediction model.

Kim et al. (2022) conducted a comparative study on boosting techniques with the aim of enhancing the performance of corporate bankruptcy prediction models through geometric mean optimization. Heo et al. (2022) focused on developing an explainable artificial intelligence-based approach for predicting corporate bankruptcy, taking into consideration the challenges posed by imbalanced data. They addressed the data imbalance issue using the Synthetic Minority Over-sampling Technique (SMOTE) and selected various machine learning techniques as candidate classification models, optimizing the hyperparameters for each model. The comparative analysis revealed that Random Forest emerged as the best-performing model, and to address model interpretability, it was combined with Local Interpretable Model-agnostic Explanations (LIME). The proposed methodology was validated through a case study.

To provide an overview of the techniques employed in recent studies on corporate bankruptcy prediction, a comprehensive compilation was made in <Table 3>. The primary analysis techniques utilized were Logistic Regression and Random Forest, alongside Neural Network, Support Vector Machine (SVM), and boosting techniques, which were also widely adopted.

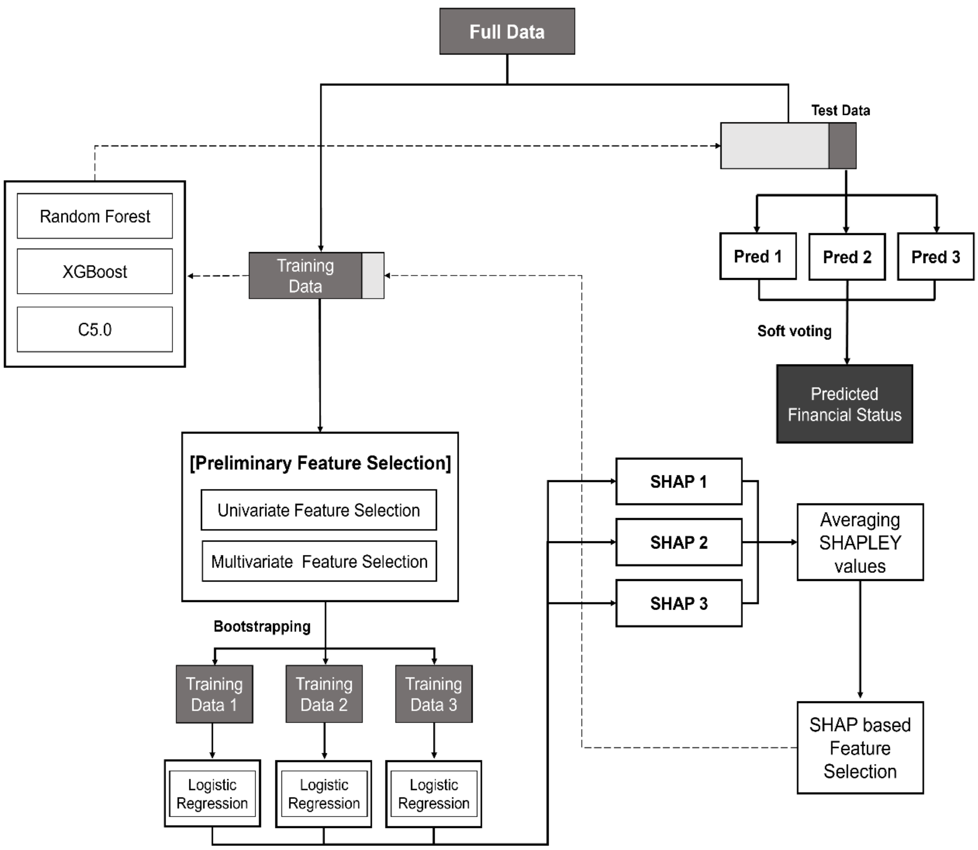
3. Experiment

In this study, the proposed research model was conducted in three sequential steps, following the framework depicted in <Figure 2>. The research dataset consisted of data from 1,698 Korean light industrial companies.

The first step, referred to as "Preliminary Feature Selection," involved performing Univariate Feature Selection on the training dataset to identify the linear relationship between each independent variable and the dependent variable. Variables with p-values below 0.05 (at a significance level of 5%) were retained. Subsequently, Multivariate Feature Selection was conducted using the Stepwise Feature Selection process based on Logistic Regression. This process assessed the classification contribution of the remaining variables to the dependent variable.

In the second step, three bootstrap datasets were created using the bootstrapping technique applied to the training dataset. Logistic Regression models were constructed for each bootstrap dataset to develop classification models. The Shapley Additive Explanations (SHAP) method was then applied to these models to derive three sets of Shapley values, representing the contributions of the independent variables used in each model. By calculating the average Shapley value for each independent variable, integrated variable-specific Shapley values were obtained. Based on these values, a “SHAP-based Feature Selection” was conducted to select the final variables. The final Shapley values for each variable were sorted in descending order, and the top 10 variables with the highest Shapley values were selected as the main variables. This selection corresponded to choosing the variables with the top 30% of Shapley values among all variables.

In the third step, the results of the SHAP-based feature selection were applied to Random Forest, XGBoost, and C5.0 models. These models were then applied to the Test Data to generate three classification values. Finally, the final classification result value was generated using Soft Voting. Soft voting involved calculating the average of the predicted values per observation obtained from the three aforementioned models. By performing these calculations on all observations, the default prediction value for the entire dataset of the Ensemble model could be generated.



〈Figure 2〉 Research Framework

〈Table 4〉 Initial Independent Variables

Variable Name	Variable Name	Variable Name
Net Sales Growth Rate	Interest Expenses to Total Borrowings and Bonds Payable	Total Cash Flow to Debt Ratio
Operating Profit Growth Rate	Return on Asset	Total Cash Flow to Current Liabilities
Net Profit Growth Rate	Total Asset Operating Income	Total Cash Flow to Borrowing Ratio
Total Asset Growth Rate	EBITDA(in thousands won)	Total Cash Flow to Total Capital Ratio
Capital adequacy Growth Rate	Non-operating Income Ratio	Non-current Asset Ratio
Net Income Growth Rate before Corporate Tax	Operating Income to Business Capital	Short-term Debt Repayment Ability
Growth Rate of Current Assets	Balance Ratio	Debt Service Coverage Ratio
Growth Rate of Tangible Assets	Cash Operating Profit/Sales	Total Assets to Debt Ratio
Growth Rate of Inventories	Cash Flow/Sales	Reserves to Total Assets Ratio
Cost of Sales Ratio	Free Cash Flow/Sales	Reserves 2
Cost Income Ratio	Interest Coverage Ratio	Cash Ratio
Operating Profit Margin	EBITDA/Interest Expense	Cash Operating Profit to Total Borrowings
Net Profit Margin Ratio before Corporate Tax	EBITDA/Total Borrowings	Receivable Turnover (in times)
Net Profit Margin	Total Borrowings to Total Assets	Inventory Turnover (in times)
ROIC	Short-Term Borrowings/Total Borrowings	Net Working Capital Turnover (in times)
ROE	Fixed Asset to Long-Term Capital	Total Capital Turnover (in times)
EBITA Margin	Financial Expenses to Sales	Turnover Period
Material Cost/Sales	Debt to Equity Ratio	Non-Current Asset Turnover (in times)
Labor Cost/Sales	Capital Adequacy Ratio	Payables Turnover Period
Expenses/Revenue	Borrowings/Sales	Payables Turnover (in times)
Total Asset Net Profit Margin before Corporate Tax	Current Ratio	Receivables Turnover Period
Net Financial Expenses to Sales	Borrowings/Capital Adequacy Ratio	Net Operating Capital Turnover (in times)
Capital Net Profit Margin before Corporate Tax	Short and Long Term Borrowings/Total Assets	Operating Capital Turnover (in times)
Depreciation Rate	Reserve Ratio	Net Working Capital Turnover Period
Depreciation Expense to Total Costs Ratio	Non-current Liabilities to Net Operating Capital	Current Asset Turnover (in times)
Net Income Interest Coverage Ratio before Corporate Tax	Non-current Liabilities Ratio	Non-Current Asset Turnover (in times)
Financial Expense to Total Debt Ratio	Non-current Ratio	Equity Turnover (in times)
Financial Expense to Total Cost Ratio	Quick Ratio	Capital Stock Turnover (in times)
Gross Profit Margin	Payable to Inventory	Inventory Turnover Period
Dividend Payout Ratio	Receivable to Payable	Invested Asset Turnover (in times)
Dividend Yield	Corporate Tax Burden Ratio	Labor Income Distribution Ratio
Retained Earnings to Equity Ratio	Net Operating Capital to Total Capital	Value Added (in thousands won)
Retained Earnings Ratio	Current Liabilities Ratio	Value Added Ratio
Net Gain on Foreign Currency Transactions and Translation to Sales	Current Asset Composition Ratio	Equipment Investment Efficiency
Research and Development Costs to Sales	Reserves to Paid in Capital Ratio	Profit Distribution Ratio
Labor Cost to Total Expenses	Reserves to Total Assets Ratio	Total Capital Investment Efficiency
Income before Corporate Taxes to Equity	Inventory to Net Operating Capital Ratio	Capital Distribution Ratio
Dividends to Equity	Inventory to Current Assets Ratio	Business Experience
Net Income to Capital Stock	Financial Leverage	Operating Capital
Taxes and Dues to Total Expenses	Borrowings to Monthly Sales	Value Added
Total Tax to Net Income before Taxes	Total Cash Flow to Sales Ratio	

3.1. Experimental Data

This study conducted experiments involving the collection of data from 1,698 Korean light industrial companies. The dataset comprised 122 independent variables related to financial ratios, including factors such as growth, profitability, stability, activity, productivity, as well as variables related to experience, managerial capital, and value added. Among these companies, 849 were classified as financially "non-bankrupt," while the remaining 849 were categorized as "bankrupt" companies. The classification of companies was based on the Standard Industrial Code, and both bankrupt and non-bankrupt companies from the same industry were selected for the experiment. The composed 122 variables are as shown in the <Table 4>.

3.1.1. Data

The dataset is partitioned into separate subsets. For this purpose, 80% of the total collected data, consisting of 1,358 instances, is designated as the training set, while the remaining 20%, comprising 340 instances, is allocated for testing. The distribution and count of the data are summarized in <Table 5>.

〈Table 5〉 Data Distribution

Data Split	Data Count	Data Distribution
Training Data	1,358	80%
Testing Data	340	20%
Total	1,698	100%

3.1.2. Univariate Analysis

In the variable selection process, variables with a presence in more than 95% of the overall dataset and statistical significance (p-value < 0.05) in the one-sample t-test are given priority. The main objective of this initial selection is to explore the relationship between the independent and dependent variables. Furthermore, supplementary variables that may not exhibit significant associations but have been consistently utilized in previous studies or are commonly used in credit rating assessments are also considered. By incorporating these considerations, the aim is to extract relevant variables during the initial selection stage. The selected 54 variables are as follows: cost of goods sold ratio, operating profit margin, pre-tax net profit margin, net profit margin, roe, ebita margin, material cost to sales ratio, labor cost to sales ratio, overhead cost to sales ratio, total asset pre-tax net profit margin, capital stock pre-tax net profit margin, depreciation rate, depreciation expense to total costs ratio, financial expense to total debt ratio, financial expense to total cost ratio, gross profit margin, dividend payout ratio, dividend yield, retained earnings ratio, research and development expense to sales ratio, labor cost to total expenses ratio, equity pre-tax net profit margin, equity dividend payout ratio, capital stock net profit margin, average borrowing interest rate, return on asset, total asset operating profit margin, EBITDA(in thousands won), operating profit to capital employed ratio, balance ratio, cash operating profit to sales ratio, cash flow to sales ratio, debt ratio, short-term borrowings/total borrowings, equity ratio, current ratio, quick ratio, net operating capital to total capital ratio, current asset composition ratio, retained earnings to contributed capital ratio, retained earnings to total assets ratio, inventory to current assets ratio, total cash flow to sales ratio, total cash flow to total capital ratio, total assets to debt ratio, retained earnings to total assets ratio, cash ratio, cash flow from operations to total borrowings, accounts payable turnover period, current asset turnover (in times), capital stock turnover (in times), value added (in thousands won), business experience, value added.

The t-test classification performance is a crucial metric for evaluating the effectiveness of variable selection using this method. It provides an assessment of the model's classification capability after identifying significant variables through the t-test. The classification performance results obtained from the t-test variable selection are presented in <Table 6>.

Out of the initial 122 variables, 54 were selected based on univariate analysis. The models using the selected variables demonstrated the following performance: Random Forest and XGBoost achieved perfect classification accuracy on the training dataset, while C5.0 achieved an accuracy of 95.66%. On the test dataset consisting of 340 data points, Random Forest achieved 86.47% accuracy, XGBoost achieved 84.71% accuracy, and C5.0 achieved 83.53% accuracy. These results indicate that Random Forest exhibits the highest classification accuracy, followed by XGBoost and C5.0.

〈Table 6〉 Classification Performance after Variable Selection using T-test

Method	Training		Testing	
	Accuracy (%)	Correctly Classified Instance	Accuracy (%)	Correctly Classified Instance
RF	100	1358	86.47	294
XGBoost	100	1358	84.71	288
C5.0	95.66	1299	83.53	284

3.1.3. Multivariate Analysis

The selected variables from the initial variable selection process undergo Multivariate Analysis to address multicollinearity among the identified independent variables. Logistic Regression analysis with stepwise variable selection is employed for this analysis, as it is suitable for models with binary dependent variables, such as constructing a corporate insolvency prediction model.

Binary Logistic Regression using the Conditional Forward Stepwise method is utilized in this study to select the final set of variables. This method sequentially adds variables based on their significant contributions, allowing for efficient computation. A total of 29 variables are selected through Multivariate Analysis as follows: operating profit margin, EBITDA margin, material cost/sales, labor cost/sales, expenses/revenue, total asset net profit margin before corporate tax, depreciation rate, financial expense to total debt ratio, financial expense to total cost ratio, dividend payout ratio, dividend yield, research and development costs to sales, labor cost to total expenses, income before corporate taxes to equity, net income to capital stock, return on asset, total asset operating income, operating income to business capital, cash operating profit/sales, total borrowings to total assets, short-term borrowings/total borrowings, current asset composition ratio, reserves to paid in capital ratio, reserves to total assets ratio, inventory to current assets ratio, total cash flow to sales ratio, total cash flow to total capital ratio, payables turnover period, business experience.

The Stepwise method combines forward selection, backward elimination, and stepwise selection techniques to perform variable selection. Starting from an empty model, variables are added or removed one by one to enhance the model. The evaluation metric after variable selection using Stepwise for predictive performance is shown in <Table 7>.

The classification performance of the variable selection using Stepwise is as follows. In the training dataset, both the Random Forest and XGBoost models achieved excellent performance, perfectly classifying all 1,358 data points. The C5.0 model exhibited an accuracy of 94.55%. In the test dataset, the Random Forest, XGBoost, and C5.0 models achieved classification accuracies of 85.88%, 84.71%, and 82.06% respectively. The variable selection method using Stepwise also demonstrated that the Random Forest model had the highest accuracy, while the C5.0 model showed the lowest performance.

〈Table 7〉 Classification Performance after Variable Selection using Stepwise

Method	Training		Testing	
	Accuracy (%)	Correctly Classified Instance	Accuracy (%)	Correctly Classified Instance
RF	100	1358	85.88	292
XGBoost	100	1358	84.71	288
C5.0	94.55	1284	82.06	279

3.1.4. Data Sampling

Data sampling is conducted using the bootstrapping technique, which is a method employed to overcome the limitations imposed by sample size. The bootstrapping procedure involves creating multiple datasets by randomly selecting observations from the original sample with replacement. This approach allows for the estimation of various statistics and constructs with high reliability and flexibility, irrespective of the sample's size, without relying on assumptions about the distribution. Bootstrapping is known for its low bias and variance, resulting in accurate results. Additionally, it effectively addresses the challenge of imbalanced data by generating new datasets through repeated sampling. In this research, bootstrapping is utilized to generate three distinct data groups by randomly sampling from the training data.

3.2. Experimental Result

3.2.1. SHAP

The results of the XAI analysis utilizing SHAP and the selected variables are presented in the following <Table 8>. It showcases the SHAP values for three distinct groups that were chosen based on variables and bootstrapping. These derived SHAP values are subsequently averaged to establish the ultimate SHAP value.

The variables were selected based on the averaged Shapley values. The SHAP summary plot in <Figure 3> illustrates the contributions of the bankruptcy prediction model, highlighting variables with high SHAP values. In the Shapley summary plot, the feature values are represented by the relative magnitude of each variable's value using colors. The color spectrum ranges from yellow, indicating higher values, to purple, indicating lower values. On the X-axis, SHAP Value represents the extent to which a positive value contributes to increasing

〈Table 8〉 Result of XAI Analysis

Variable Name	SHAP value 1	SHAP value 2	SHAP value 3
Operating Profit Margin	0.13595	0.134901	0.014308
Depreciation Rate	0.011452	0.011867	0.016308
Financial Expense to Total Debt Ratio	0.020602	0.017909	0.031818
Financial Expense to Total Cost Ratio	0.060684	0.028143	0.02969
Income before Corporate Taxes to Equity	0.064807	0.005785	0.014518
Net Income to Capital Stock	0.012745	0.024373	0.073762
Reserves to Total Assets Ratio	0.027228	0.020905	0.057502
Total Cash Flow to Total Capital Ratio	0.010009	0.017191	0.003696
Dividend Yield	-0.00431	0.035118	0.093716
Business Experience	0.024809	0.028155	0.01812

〈Figure 3〉 SHAP Summary Plot

the likelihood of the outcome, which is corporate bankruptcy in this case, while a negative value decreases the likelihood. On the Y-axis, variables higher up indicate a higher contribution to predicting the outcome, while variables lower down have a lower contribution.

A higher reserves to total assets ratio tends to be associated with a lower probability of bankruptcy. An increase in the depreciation rate tends to increase the probability of bankruptcy, whereas a decrease in labor cost to sales is associated with a lower probability of bankruptcy. A lower total asset net profit margin before corporate tax is related to a higher probability of bankruptcy, while a higher margin is associated with a lower probability of bankruptcy.

3.2.2. Comparison of the Classification Performance

The model was constructed by selecting the top 10 variables with the highest SHAP values, including reserves to total assets ratio, total borrowings to total assets, depreciation rate, labor cost to sales, total asset net profit margin before corporate tax, financial expense to total debt ratio, net income to capital stock, expenses/revenue, current asset composition ratio, and return on asset. These variables were then applied to Random Forest, XGBoost, and C5.0 using the test data, resulting in three classification values. Finally, Soft Voting was employed to generate the final bankruptcy prediction value. The classification performance of the SHAP-based feature selection is presented in <Table 9>.

In the training data, Random Forest, XGBoost, and the Ensemble model achieved a prediction accuracy of 100%, while C5.0 exhibited an accuracy of 88.14%. In the test data, Random Forest, XGBoost, and C5.0 demonstrated similar performance with an accuracy of 82%, while the Ensemble model displayed the highest accuracy of 84%.

〈Table 9〉 Classification Performance after Variable Selection using SHAP

Method	Training		Testing	
	Accuracy (%)	Correctly Classified Instance	Accuracy (%)	Correctly Classified Instance
RF	100	1358	82.94	282
XGBoost	100	1358	82.06	279
C5.0	88.14	1197	82.35	280
Ensemble	100	1358	84.12	286

4. Conclusion

Corporate bankruptcy prediction is a significant research area that focuses on enhancing prediction accuracy through data mining techniques and evaluating predictive performance. However, the predominant approach in most studies is to rely on variable selection methods driven by statistical analysis. These methods lack interpretability, are sensitive to data distribution, and overlook variable interactions. Moreover, the large number of variables selected using these methods can lead to overfitting or computational complexity issues.

This study proposes a novel approach that integrates explainable artificial intelligence (XAI) to identify interpretable variables. It evaluates the performance of three AI models, namely Random Forest, XGBoost, and C5.0, using a dataset collected from 1,698 Korean light industrial companies. The performance of these models is compared with an Ensemble model formed through Soft Voting to design an efficient corporate bankruptcy prediction model. Logistic Regression is employed to calculate SHAP values for each group of data samples using bootstrapping, following Univariate and Multivariate analyses. The 10 top-ranked variables based on averaged SHAP values are selected to enhance the model's performance. These selected variables are then trained on the three machine learning models to evaluate their classification performance. The performance of the Ensemble model is compared with the individual models to design a model with superior predictive ability. The experimental results demonstrate that the Ensemble model exhibits the best predictive performance.

This study has significant implications as it integrates XAI into a corporate bankruptcy prediction model, resulting in the selection of interpretable variables and providing interpretability for the prediction results. Furthermore, the proposed model is computationally efficient, achieving robust classification performance with a smaller number of variables. The findings reveal that SHAP-based variable selection can yield similar prediction accuracy with significantly fewer variables.

However, there are limitations to this study. Generalization may be challenging as it relies solely on data from Korean light industrial companies and does not encompass diverse data sources or multiple countries. Additionally, the model is developed based on quantitative factors derived solely from financial statements. Another limitation lies in the use of Logistic Regression for SHAP calculation, which may restrict the model's flexibility. The proposed variable selection method, based on SHAP values generated from a randomly created dataset using bootstrap, demonstrates the effectiveness of the approach across different algorithms. Future studies could address these limitations by incorporating data from diverse industries and countries and employing techniques to calculate SHAP values which are suitable for each algorithm.

Reference

Adadi, A. and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI).", *IEEE access,* 6 (2018), 52138~52160.

Ahn, Y. A. and H. J. Cho, "A Study on XAI-based Clinical Decision Support System", *Journal of the Korea Contents Association*, Vol.21, No.12(2021), 13~22.

Bastani, O., C. Kim, and H. Bastani., "Interpreting blackbox models via model extraction." *arXiv preprint arXiv:1705.08504* (2017).

Bemš, J., O. Starý, M. Macaš, J. Žegklitz and P. Pošík, P, "Innovative default prediction approach.", *Expert Systems with Applications,* 42.17-18 (2015): 6277~6285.

Brenes, R. F., A. Johannssen, and N. Chukhrova, "An intelligent bankruptcy prediction model using a multilayer perceptron.", *Intelligent Systems with Applications,* (2022), 200136.

Brenes, R., A. Johannssen and N. Chukhrova, "An intelligent bankruptcy prediction model using a multilayer perceptron.", *Intelligent Systems with Applications,* (2022), 200136.

Bussmann, N., P. Giudici, D. Marinelli and J. Papenbrock, "Explainable AI in fintech risk management.", *Frontiers in Artificial Intelligence* 3, (2020), 26.

Carmona, P., A. Dwekat and Z. Mardawi, "No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure.", *Research in International Business and Finance,* 61 (2022), 101649.

Chinnici, M., Y. G. Y. Gebreyesus, D. Dalton and S. Nixon, "Machine Learning for Data Center Optimizations: Feature Selection Using SHapley Additive exPlanation (SHAP)", *Future Internet* 15.3 (2023): 88.

Cho, J. H., E. J. Ahn and S. S. Kim, "A Study on the Prediction Model for Insolvent Companies Based on Deep Learning", Journal of Business Research, Vol.36, No1(2021), 99~113.

Cho, S. and K. Shin. "Domain Knowledge Incorporated Counterfactual Example-Based Explanation for Bankruptcy Prediction Model", *Journal of Intelligence and Information Systems*, 28(2), 2022, 307-332.

Choi, J. W., "Forecasting corporate default using artificial intelligence based on news information", *Konkuk University*, 2019.

Chu, H. G., H. S. Shin, S. K. Cho, Y. S. Yoo and C. S. Park, "Sensitivity Analysis Using Explainable AI for Building Energy Use", Journal of the Architectural Institute of Korea, Vol.36, No11(2022), 279~287.

Chun Y. E., S. B. Kim, J. Y. Lee and J. H. Woo, "Study on credit rating model using explainable AI", Journal of the Korean Data And Information Science Society, Vol.32, No2(2021), 283~295.

Gebreyesus, Y., D. Dalton, S. Nixon, D. Chiara and M. Chinnici, "Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP)." *Future Internet,* Vol. 15, No. 3, (2023), 88.

Gordini, N., "A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy." *Expert systems with applications,* Vol.41, No.14 (2014), 6433~6445.

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, F. Giannotti and D. Pedreschi, "A survey of methods for explaining black box models.", *ACM computing surveys (CSUR),* 51.5 (2018), 1~42.

Heo, S. W. and D. H. Baek, "A Methodology for Bankruptcy Prediction in Imbalanced Datasets using eXplainable AI", *Journal of the Society of Korea Industrial and Systems Engineering*, Vol.45, No.2(2022), 65~76.

Hwang, S. H., "Knowledge Graph-based UI for Explainable Recommendation", *Yonsei University*, 2019

Hwangbo, Y., and J. G. Moon. "A Comparative Study on Failure Pprediction Models for Small and Medium Manufacturing Company." *Asia-Pacific Journal of Business Venturing and Entrepreneurship,* Vol11, No.3, 2016, 1~15.

Iturriaga, F. L. and I. P. Sanz. "Bankruptcy visualization and prediction using neural networks: A study of US commercial banks." *Expert Systems with applications,* Vol. 42, No. 6, (2015), 2857~2869.

Jeong, C. W., J. H. Min and M. S. Kim, "A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction.", *Expert Systems with Applications,* Vol.39, No.3 (2012), 3650~3658.

Joshi, S., R. Ramesh and S. Tahsildar, "A Bankruptcy Prediction Model Using Random Forest," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 1~6, doi: 10.1109/ICCONS.2018.8663128.

Kang, B. "Comparison of Altman's K-Score and Neural Network's Corporate Bankruptcy Prediction", *Chonnam National University*, 2023

Kasgari A. A., M. Divsalar, M. R. Javid and S. J. Ebrahimian, "Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses.", *Neural Computing and Applications,* 23 (2013), 927~936.

Khademolqorani, S., A. Z. Hamadani, and F. M. Rafiei. "A hybrid analysis approach to improve financial distress forecasting: Empirical evidence from Iran.", *Mathematical Problems in Engineering,* (2015),1~9

Kim, M. J., S. Y. Jo and K. M. Yoo, "Geometric Mean Optimization-based Boosting for Bankruptcy Prediction", *Journal of The Korean Institute of Intelligent Systems*, Vol.32, No.4(2022), 346~358.

Kim, S. J. and H. C. Ahn, "Application of Random Forests to Corporate Credit Rating Prediction", *The Journal of Business and Economics*,32(1), 2016, 187~211.

Kim, S. J., "Evaluation of Distress Prediction Model for Food Service Industry in Korea: Using the Logit Analysis", *Journal of the Korea Academia-Industrial cooperation Society*, Vol.20, No.11 (2019), 151~156.

Kim, S. Y. and A. Upneja, "Majority voting ensemble with a Decision Tree for business failure prediction during economic downturns.", *Journal of Innovation & Knowledge* 6.2, (2021): 112~123.

Kwon, H., D. Lee and M. Shin, "Dynamic forecasts of bankruptcy with Recurrent Neural Network model", *Journal of Intelligence and Information Systems*,23(3),2017, 139~153.

Kwon, J. H. and S. H. Park, "Machine Learning Models and Failure Prediction on Shipping Companies", *Journal of shipping and logistics*, Vol.38, No.2(2022), 399~418.

Kwon, U. J. and S. Y. Park, "Survival prediction of hospitality businesses during the pandemic", *Journal of The Korean Data Analysis Society,* Vol.24, No.5(2022), 1791~1809

Le, N. Q. K., Q. T. Ho, V. N. Nguyen and J. S. Chang, "BERT-Promoter: An improved sequence-based predictor of DNA promoter using BERT pre-trained model and SHAP feature selection", *Computational Biology and Chemistry,* 99, (2022), 107732.

Lee, S. G., S. K. Son., S. R. Shin and K. J. Yee., "Airfoil Inverse Design using XAI (eXplainable Artificial Intelligence)", *Journal of the Korean Society for Aeronautical and Space Sciences*, 2021, 50~51.

Lee, S. J. and W. S. Choi, "A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis.", *Expert Systems with Applications*, Vol 40, No.8(2013), 2941~2946.

Liang, D., C. Tsai and H. Wu, "The effect of feature selection on financial distress prediction.", *Knowledge-Based Systems.* 73 (2015), 289~297.

Linardatos, P., V. Papastefanopoulos and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods.", *Entropy,* 23.1 (2020), 18.

Liu, Y., Y. Shen, H. Wang, Y. Zhang and X. Zhu, "m5Cpred-XS: a new method for predicting RNA m5C sites based on XGBoost and SHAP", *Frontiers in Genetics*, (2022), 13.

Liu, Y., Z. Liu, X. Luo and H. Zhao, "Diagnosis of Parkinson's disease based on SHAP value feature selection.", *Biocybernetics and Biomedical Engineering,* 42(3), (2022). 856-869.

Lundberg, S. M. and S. I. Lee, "A unified approach to interpreting model predictions.", *Advances in neural information processing systems,* 30, (2017).

Lysaght, T., H. Y. Lim, V. Xafis and K.Y. Ngiam, "AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research.", *Asian Bioethics Review,* 11, (2019), 299~314.

Mattsson, B., and O. Steinert, "Corporate bankruptcy prediction using Machine Learning techniques.", Gothenburg University, 2017.

Ni, Q., L. Chen, J. Zhu, J. Pang, Z. Wang and X. Yang, "Prediction and interpretation of gamma pass rate based on SHAP value feature selection", *Radiation Oncology,* 2023

Oh, D. H., "Utilization of Artificial Intelligence Technology in the Military and Suggestion of XAI Technology Application Direction", *Journal of Digital Contents Society*, Vol.23, No.5(2022), 943~951.

Oh, H. R., A. L. Son and Z. K. Lee, "Occupational accident prediction modeling and analysis using SHAP", *The Journal of Contents Computing*, Vol.22, No.7(2021), 1115~1123.

Park, S. H., T. I. Kim and C. H. Kwon, "A Study of Default Prediction for Korean Shipping Companies", *Journal of Shipping and Logistics*, Vol.33, No.4,2017,823~852.

Qu, Y., P. Quan, M. Lei and Y. Shi, "Review of bankruptcy prediction using machine learning and deep learning techniques.", *Procedia Computer Science,* 162, (2019), 895~899.

Rouhi, R., M., Clausel, J., Oster and F. Lauer, "An interpretable hand-crafted feature-based model for atrial fibrillation detection", *Frontiers in Physiology,* 12, (2021). 657304.

Rustam Z. and G. S. Saragih, "Predicting Bank Financial Failures using Random Forest," *2018 International Workshop on Big Data and Information Security (IWBIS)*, Jakarta, Indonesia, 2018, pp. 81-86, doi: 10.1109/IWBIS.2018.8471718.

Shortliffe, E. H. and B. G. Buchanan, "A model of inexact reasoning in medicine.", *Readings in uncertain reasoning*, 1990, 259~275.

Son, H., C. Hyun, D. Phan and H. J. Hwang, "Data analytic approach for bankruptcy prediction.", *Expert Systems with Applications,* 138 (2019), 112816.

Son, J. H., S. J. Woo, H. G. Baek, B. W. Hwang and S. G. Choi, "False positive reduction in anomaly detection using XAI", *Journal of Computing Science and Engineering*, 2022, 609~611.

Song, H. J., D. J. Park and Z. K. Lee, "An Empirical Comparison of Bankruptcy Prediction of External Auditing and Non-External Auditing Companies Using Machine Learning Methods", *Journal of The Korea Society of Information Technology Policy & Management*, Vol.13, No.3(2021), 2521~2527.

Tjoa, E. and C. Guan. "A survey on explainable artificial intelligence (XAI): Toward medical XAI.", *IEEE transactions on Neural Networks and learning systems,* 32.11, (2020), 4793~4813.

Utkin, L. and A. Konstantinov, "Ensembles of random SHAPs.", *Algorithms,* 15.11, (2022), 431.

Van Lent, M., W. Fisher and M. Mancuso, "An explainable artificial intelligence system for small-unit tactical behavior.", *Proceedings of the national conference on artificial intelligence*, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.

Wagenmans, F., "Machine learning in bankruptcy prediction", Utrecht University, 2017.

Yeh, C., D. Chi and Y. Lin, "Going-concern prediction using hybrid random forests and rough set approach.", *Information Sciences,* 254 (2014), 98~110.

Zhao, J., "Corporate financial risk prediction based on embedded system and deep learning.", *Microprocessors and Microsystems,* (2020), 103405.

Zhou L., K. K. Lai, and J. Yen, "Empirical models based on features ranking techniques for corporate financial distress prediction.", *Computers & mathematics with applications,* Vol.64, No.8 (2012), 2484~2496.

국문요약

설명가능 AI 기반의 변수선정을 이용한 기업부실예측모형

문건두* · 김경재**

기업의 부실 예측 모델은 기업의 재무 상태를 객관적으로 모니터링하는 데 필수적인 도구 역할을 한다. 적시에 경고하고 대응 조치를 용이하게 하며 파산 위험을 완화하고 성과를 개선하기 위한 효과적인 관리 전략을 수립할 수 있도록 지원한다. 투자자와 금융 기관은 금융 손실을 최소화하기 위해 부실 예측 모델을 이용한다. 기업 부실 예측을 위한 인공지능(AI) 기술 활용에 대한 관심이 높아지면서 이 분야에 대한 광범위한 연구가 진행되고 있다. 해석 가능성과 신뢰성이 강조되며 기업 부실 예측에서 설명 가능한 AI 모델에 대한 수요가 증가하고 있다. 널리 채택된 SHAP(SHapley Additive exPlanations) 기법은 유망한 성능을 보여주었으나 변수 수에 따른 계산 비용, 처리 시간, 확장성 문제 등의 한계가 있다. 이 연구는 전체 데이터 세트를 사용하는 대신 부트스트랩 된 데이터 하위 집합에서 SHAP 값을 평균화하여 변수 수를 줄이는 새로운 변수 선택 접근법을 소개한다. 이 기술은 뛰어난 예측 성능을 유지하면서 계산 효율을 향상시키는 것을 목표로 한다. 해석 가능성이 높은 선택된 변수를 사용하여 랜덤 포레스트, XGBoost 및 C5.0 모델을 훈련하여 분류 결과를 얻고자 한다. 분류 결과는 고성능 모델 설계를 목표로 soft voting을 통해 생성된 앙상블 모델의 분류 정확성과 비교한다. 이 연구는 1,698개 한국 경공업 기업의 데이터를 활용하고 부트스트래핑을 사용하여 고유한 데이터 그룹을 생성한다. 로지스틱 회귀 분석은 각 데이터 그룹의 SHAP 값을 계산하는 데 사용되며, SHAP 값 평균은 최종 SHAP 값을 도출하기 위해 계산된다. 제안된 모델은 해석 가능성을 향상시키고 우수한 예측 성능을 달성하는 것을 목표로 한다.

주제어 : 설명가능 인공지능, 기업부실예측, SHAP, Random Forest, XGBoost

논문접수일 : 2023년 5월 24일　논문수정일 : 2023년 6월 9일　게재확정일 : 2023년 6월 12일
원고유형 : Regular Track　교신저자 : 김경재

* 동국대학교_서울 일반대학원 경영정보학과
** 교신저자 : 김경재
동국대학교_서울 경영대학 경영정보학과
(04620) 서울특별시 중구 필동로1-30
Tel: +82-2-2260-3324, E-mail: kjkim@dongguk.edu

저 자 소 개

문건두
현재 KCA에서 이사로 재직 중이다. University of Illinois Urbana-Champaign에서 경영학 석사를 동국대학교에서 경영학박사 과정을 수료하였다. 관심분야는 XAI, 기업신용평가 시스템이다.

김경재
현재 동국대학교 경영대학 경영정보학과 교수로 재직 중이다. KAIST에서 경영정보시스템을 전공으로 박사학위를 취득하였으며, 연구 관심분야는 비즈니스 애널리틱스, CRM, 추천기술, 빅데이터 등이다.