2024

Analysis of bankruptcy prediction of shipping industry - Machine Learning Approach

Kwon, Minsu

http://dx.doi.org/10.24382/5216
University of Plymouth

[1] *E.g.* in the example of third party copyright materials reused in the thesis.
[2] In accordance with best practice principles such as, *Marking/Creators/Marking third party content* (2013). Available from: https://wiki.creativecommons.org/wiki/Marking/Creators/Marking_third_party_content [accessed 28th February 2022]

UNIVERSITY OF PLYMOUTH

ANALYSIS OF BANKRUPTCY PREDICTION OF SHIPPING INDUSTRY

- Machine Learning Approach

by

MINSU KWON

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Plymouth Business School

MAY 2024

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Dr. Sae-Yeon Roh and Dr. Min-Chul Sohn, for their unwavering support, guidance, and encouragement throughout the course of my research. Their invaluable insights and constructive feedback have been instrumental in shaping this thesis.

I am also immensely grateful to distinguished scholars for their financial support, which made this research possible. Special thanks to Dr. Keun-Sik Park, Dr. Jung-Hwan Kim, Dr. Dong-Joon Kang, Dr. Jin-Ho Noh, Dr. Ja-Yeon Lee, Dr. Jin-Ho Oh, Dr. Dae-Hwan Jeong, and Professor Su-Han Woo for their collaboration, advice, and assistance in various aspects of my study. Their expertise and contributions have significantly enriched my work.

I would also like to extend my heartfelt thanks to my colleagues and friends, Dr. Jae-Woo Park, Dr. Sang-Ho Cho, Dr. Yu-Na Kim, Dr. Jun-Seung Kim, and Dr. In-Hyeok Yeo. Their camaraderie, support, and intellectual discussions have made this journey enjoyable and stimulating.

I would like to express my eternal gratitude to my family members, who have been the endless motivation behind all of my achievements. None of my accomplishments would have been possible without their unwavering support.

Finally, to my beloved Su-Jin Kim, thank you for your unwavering support, understanding, and sacrifices. Your belief in me has been my greatest source of strength and motivation. This achievement would not have been possible without you.

I am profoundly grateful to everyone who has supported me in this endeavour. Your contributions have been crucial to the completion of this thesis, and I am deeply thankful to each one of you.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee. Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

The following external institutions were visited for research and consultation purposes:

Publications (or public presentation of creative research outputs):

Not Applicable

Word count of main body of thesis: 62992

Signed

Date: 25.03.2024

Analysis of Bankruptcy Prediction of Shipping Industry

- Machine Learning Approach

Minsu Kwon

Abstract

The Korean shipping industry's vulnerability to economic crises, such as the 2016 bankruptcy of Hanjin Shipping, highlights the need for robust bankruptcy prediction methods, particularly for small and medium-sized enterprises (SMEs). This research aims to identify key risk factors for predicting bankruptcy in shipping companies from Korean industry by leveraging financial, non-financial, and economic data through advanced machine learning models, including Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) networks.

A comprehensive literature review and interviews with industry practitioners were conducted to refine the variables used in the models. These models predict bankruptcy across 1, 3, and 5-year horizons, with Explainable Artificial Intelligence (XAI) techniques employed to interpret the impact of each variable. The findings reveal that non-financial and macroeconomic variables, such as LIBOR interest rates and trade volume growth rates, are significant predictors across all periods. Additionally, the importance of financial ratios, especially those related to profitability, increases with the length of the forecasting period. The research also highlights distinct risk factors between large shipping companies and SMEs, underscoring the need for tailored risk management strategies.

This study contributes valuable insights for stakeholders in the shipping market, including policymakers and financial institutions. The identified risk factors enable shipping companies to improve strategic planning and anticipate market cycles more effectively. Policymakers can use these insights to develop regulations that address the unique needs of shipping SMEs. Overall, this research provides a comprehensive understanding of the market dynamics, offering practical implications for managing bankruptcy risk in the shipping industry. Further approach can be held with wider geographical areas to reflect specific regional aspects with much advanced prediction models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Equation

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BDI	Baltic Dry Index
DART	Data Analysis Retrieval and Transfer System
DWT	Deadweight Tonnage
EBITDA	Earnings before interest, taxes, depreciation, and amortization
FN	False Negative
FP	False Positive
G/T	Gross Tonnage
GDP	Gross Domestic Product
IRONSTEEL	Dow Jones Iron & Steel Index
KMI	Korean Maritime Institution
KOBC	Korean Ocean Business Corporation
LDA	Linear Discriminant Analysis
LIBOR	London Interbank Offer Rate
LIME	Local Interpretable Model-agnostic Explanations
LR	Logistic Regression
LSTM	Long Short-Term Memory
MDA	Multiple Discriminant Analysis
RNN	Recurrent Neural Network
ROA	Return on Asset
ROC	Receiver operating characteristic
ROE	Return on Equity
SCFI	Shanghai Container Freight Index
SHAP	Shapley Additive Explanations
SME	Small and Medium Sized Enterprises
SMOTE	Synthesis Minority Oversampling Technique
SVM	Support Vector Machine
TP	True Positive
TN	True Negative
XAI	Explainable Artificial Intelligent
XGB	Extreme Gradient Boosting

1. INTRODUCTION

1.1. Background

The shipping industry is one of the most capital-intensive industries (Christiansen et al., 2007). For instance, shipping company's proportion of capital expenditure in their total asset averaged approximately 8% between 1999 and 2019, which placed the 8th sector among all industrial sectors (Lozinskaia et al., 2017). In this industry, a substantial portion of these investments is financed through debt, which can exceed 40% of a company's capital (Drobetz et al., 2013). This high reliance on debt financing, coupled with the sector's inherent high asset linkage and equity risk conditions, necessitates a strategic approach to financial management to mitigate bankruptcy risk (Yuen and Ko, 2018).

More recently, this industrial sector has reached a critical level of lending, with the top 40 banks in the world with $345 billion of exposure to the shipping industry and $150 billion in loans provided only by European banks (Clintworth et al., 2021). A high level of reliance on financing by shipping companies indicates that the success of shipping companies depends on successfully managing debt policies to avoid high financial distress costs (Makrominas, 2018). In these regards, it seems that proper financial management strategy is necessary for shipping companies to keep optimal capital structure and avoid potential bankruptcy risk.

Additionally, the shipping industry exhibits significant volatility, influenced by various global economic factors (Haider et al., 2019). Under the several crises such as financial crisis or COVID-19 pandemic, many shipping companies had gone bankrupt due to plunge in demand of cargoes and high level of liability of the company (Kamal et al., 2021). Because of fluctuation of freight rates caused by the global economy depression, shipping industry have struggled with their revenue. As growing demand for capital financed through loans in shipping industry, the banks raised credit standards to navigate through this volatile industry and respond predictive bankruptcy risk (Notteboom et al., 2021). These dynamics underscore the importance of effective financial management to maintain optimal capital structure and avoid potential bankruptcy.

Small and medium-sized enterprises (SMEs) in the shipping industry face a distinct set of challenges that complicate their financial sustainability and operational efficiency. Unlike large companies, which benefit from certified audited financial statements that support their credit scores and facilitate access to high-standard credit, SMEs struggle with information opacity and a lack of reliable data. This makes the assessment of credit risk by banks more complex and elevates the perceived risk of bankruptcy, significantly hindering their financing options (Mayr et al., 2021).

The financial vulnerability of shipping SMEs is further exacerbated by macroeconomic challenges such as shifts in domestic economic structures and international trade frictions. These macroeconomic dynamics hinder their ability to secure necessary financing and impair their capacity to manage bankruptcy risk effectively (Creazza et al., 2023). Moreover, mergers and alliances among large carriers have reshaped the competitive landscape to the disadvantage of SMEs. These strategic moves, aimed at enhancing operational efficiency and reducing costs, often result in market dominance by a few large carriers. This market dominance restricts the operational capabilities and financial health of SMEs by limiting their market access and skewing revenue distribution in favor of larger players (Chen et al., 2022; Tang & Sun, 2018).

Environmental regulations add another layer of complexity. The European Union’s stringent environmental standards require significant capital investments for compliance, such as adopting low-sulphur fuels and investing in green technologies (Zhou et al., 2023). These investments, although crucial for sustainability, pose financial challenges for SMEs with limited financing options, threatening their financial stability without innovative financial solutions or government support (Creazza et al., 2023).

The Korean shipping industry provides a compelling case for studying bankruptcy risk due to its significant global presence and recent financial turmoil. The 2016 bankruptcy of Hanjin Shipping, a major player in the industry, underscored the vulnerabilities and potential for catastrophic failures within the sector (Korea Maritime Institute, 2017). This event had far-reaching repercussions, increasing bankruptcy risks for smaller shipping companies and highlighting the urgent need for improved risk management strategies (Park et al., 2021a).

Korean SMEs in the shipping industry face similar pressures as their counterparts globally. For instance, Greek SMEs in the Mediterranean also contend with intense competition and restricted access to critical maritime routes due to the expansion strategies of larger carriers (Papana & Spyridou, 2020). Additionally, the economic crisis in Greece and ongoing adjustments in the European Union's trade policies post-Brexit have added layers of complexity. These SMEs must navigate variable economic policies and trade agreements, complicating stable operations and long-term planning (Yu et al., 2020). The unpredictability of trade flows and the increased administrative burden of new customs and regulatory frameworks impose significant operational and financial stress.

Given the critical role of the shipping industry in global trade and the significant challenges faced by SMEs, there is a pressing need for research focused on identifying and managing bankruptcy risk in this sector. Corporate bankruptcy and industrial recessions can have profound impacts not only on the companies themselves but also on the global market. Research on bankruptcy prediction has intensified, especially after the 2008 financial crisis (Shi & Li, 2019). Various stakeholders, including industry professionals, investors, and researchers, have sought to develop models to predict bankruptcy at the company, industry, and market levels (Lee, 2016). Advanced machine learning models have shown higher predictive performance compared to traditional linear statistical models, such as logistic regression or linear discriminant analysis (Mai et al., 2019). The current literature confirms that machine learning models are more efficient at predicting financial risk and bankruptcy (Murphy, 2022; Alam et al., 2021b; Wang & Liu, 2021). The complex nature of bankruptcy prediction makes machine learning a robust method for assessing corporate bankruptcy risk (Vochozka et al., 2020). Financial data combined with machine learning techniques allow for the replication of these models by researchers and practitioners, offering insights for improved bankruptcy risk management (Jones et al., 2017).

Therefore, this study aims to explore key risk factors in shipping industry, especially focusing on small and medium-sized companies, proactively evaluating and managing potential bankruptcy risk by considering the industry's unique characteristics. Unique aspects of shipping SMEs would be highlights through comparison with large firms. The development of advanced predictive models tailored to the shipping industry is essential to ensure the financial stability and sustainability of SMEs, which are vital to the global economy.

1.2. Key concept

Corporate bankruptcy is a legal process wherein a company unable to meet its debt obligations seeks relief through reorganization or liquidation under bankruptcy laws (Grunert et al., 2005). This process can have severe repercussions for creditors, employees, shareholders, and the broader economy (Altman, 1968). Understanding the triggers and indicators of bankruptcy is crucial for developing effective predictive models that can anticipate financial distress and enable timely interventions.

Risk factors for corporate bankruptcy are diverse and can be broadly categorized into financial, non-financial, and macroeconomic variables. Financial variables include liquidity ratios, such as the current ratio and quick ratio, profitability ratios like return on assets and return on equity, and leverage ratios, including the debt-to-equity ratio (Altman, 1968). These financial indicators provide critical insights into a company's fiscal health and operational performance (Beaver, 1966). Non-financial variables encompass qualitative attributes such as management quality, corporate governance structures, operational efficiency, and strategic decision-making processes. These factors play a pivotal role in influencing a company's resilience and stability (Ohlson, 1980). Macroeconomic variables pertain to broader economic conditions that affect the business environment, including interest rates, gross domestic product (GDP) growth, inflation rates, and trade volumes. Macroeconomic factors exert significant pressure on companies, especially in industries like shipping that are closely tied to global trade dynamics (Moyer, 1977; Tinoco & Wilson, 2013).

Predictive models are indispensable tools for forecasting the likelihood of corporate bankruptcy by analysing historical data and identifying precursory patterns of financial distress. The development of bankruptcy prediction models has evolved significantly, starting with discriminant analysis as the prevalent method in early models (Altman, 1968). Over time, more sophisticated methodologies have emerged, including logit analysis, neural networks, and decision trees, each offering distinct advantages (Luo et al., 2020). Advanced machine learning algorithms, including Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) networks, offer enhanced predictive accuracy by managing complex, non-linear relationships and large datasets. These models are particularly useful for uncovering nuanced insights into bankruptcy risk (Chen & Guestrin, 2016; Hochreiter & Schmidhuber, 1997).

The concept of forecasting horizons is also crucial in bankruptcy prediction. Forecasting horizons refer to the timeframes over which predictive models assess the risk of bankruptcy, typically categorized into short-term (1 year), medium-term (3 years), and long-term (5 years) periods prior to the bankruptcy event. Different forecasting horizons help capture the temporal patterns and trends in financial distress, as symptoms of bankruptcy may manifest differently over various time periods (Jones & Wang, 2019; Voda et al., 2021). This approach allows for a more comprehensive evaluation of bankruptcy risk, providing critical insights for both immediate and strategic decision-making.

The high predictive accuracy of machine learning models often comes at the cost of interpretability, rendering them "black boxes" (Molnar, 2020). Explainable AI (XAI) techniques address this challenge by elucidating the impact of individual variables on prediction outcomes. In the context of bankruptcy prediction, XAI helps identify the most influential risk factors and provides transparency into the decision-making process of predictive models. This transparency is crucial for gaining the trust of stakeholders and ensuring the practical applicability of the models (Ribeiro et al., 2016).

By addressing these key concepts, this section provides a foundation for understanding the theoretical and practical aspects of bankruptcy prediction in the shipping industry. It sets the stage for the subsequent sections that derive deeper into the research problem, aim, and objectives.

1.3. Research problem

The bankruptcy of Hanjin Shipping, a major player in the industry, underscored the vulnerabilities and high bankruptcy risk within the shipping sector (Song et al., 2019). This event highlighted the limitations of existing risk management practices and the urgent need for more robust predictive models that can anticipate financial distress before it escalates to bankruptcy. The fragility of the shipping industry is evident, but it is especially pronounced among small and medium-sized enterprises (SMEs). SMEs often operate with limited resources, lack access to high-quality financial data, and are more susceptible to economic fluctuations compared to larger firms (Kim & Park, 2018). These factors exacerbate their financial instability and increase the risk of bankruptcy.

Despite significant advancements in bankruptcy prediction models, substantial gaps remain in addressing the specific needs of the shipping industry as a whole. Traditional models, such as discriminant analysis (Altman, 1968), and more recent methodologies like logit analysis, neural networks, and decision trees, have been applied to various sectors like manufacturing and banking but do not sufficiently address the unique characteristics of the shipping industry (Edmister, 1972; Martin, 1977). Factors contributing to bankruptcy can vary significantly by industry and country (Luo et al., 2020). It is essential to consider financial, non-financial, and macroeconomic factors such as management quality, corporate governance, and broader economic conditions (Tinoco & Wilson, 2013). This comprehensive approach is crucial for accurately assessing the bankruptcy risk in the shipping industry, which is sensitive to global economic fluctuations and market dynamics.

Furthermore, the applicability of existing risk factors to the shipping industry, especially SMEs, remains debated (Alexandridis et al., 2020). Most studies focus on global shipping companies, often neglecting SMEs, which make up over 90% of the Korean shipping industry and face unique financial challenges exacerbated by repeated financial crises (Park et al., 2022). This highlights the need for tailored predictive models that address the distinct needs of both SMEs and larger shipping companies. Therefore, this research intends to address the unique challenges faced by the shipping industry by not only focusing on the entire industry but also conducting a comparative analysis between SMEs and larger firms. By developing models that cater to the diverse needs of all shipping companies, this research aims to provide comprehensive tools that can improve financial resilience and strategic planning across the entire industry.

Advanced machine learning models, such as Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) networks, offer enhanced predictive accuracy by handling complex, non-linear relationships and large datasets (Chen & Guestrin, 2016;

Hochreiter & Schmidhuber, 1997). However, their application in the shipping industry, particularly for SMEs, remains limited, necessitating further research. By assessing risk factors across different forecasting horizons (1 year, 3 years, 5 years), this research intends to enhance the practical utility of these models for stakeholders, including financial institutions and shipping companies of all sizes (Jones & Wang, 2019; Voda et al., 2021).

Finally, while machine learning models are known for their high predictive accuracy, they often lack interpretability, rendering them "black boxes" (Molnar, 2020). This limitation hinders their practical application, as stakeholders require transparent models to make informed decisions. Explainable AI (XAI) techniques can address this issue by elucidating the impact of individual variables on prediction outcomes, enhancing model transparency and stakeholder trust (Arrieta et al., 2020). Evaluating the influence of risk factors through XAI can provide managerial and policy implications to proactively respond to potential bankruptcy and minimize losses in the shipping industry. This research aims to deliver actionable insights that can benefit both SMEs and larger firms, ensuring a more stable and resilient shipping sector.

1.4. Research objective

In the light of research problem, this research aims to identify key risk factors for predicting bankruptcy risk in the shipping industry, particularly focusing on SMEs, by incorporating financial, non-financial, and macroeconomic data, and applying advanced machine learning models. The study will develop predictive models that address the unique challenges faced by different sizes of companies within the industry. This study specifically analyses the Korean shipping industry to identify these risk factors and validate the predictive models.

Based on the aim, the objectives of the research would be as follows:

1. To analyse the historical development of bankruptcy prediction models through a literature review and explore how these models can be applied to enhance bankruptcy prediction within the unique context of the shipping industry
2. To identify specific bankruptcy risk factors for small and medium-sized shipping companies to enhance predictive accuracy by combining a literature review and practitioner interviews, with a focus on the Korean shipping industry
3. To investigate the variation in bankruptcy risk factors across forecasting horizons of 1, 3, and 5 years prior to bankruptcy within the shipping industry using machine learning models, with the aim of enhancing the accuracy of bankruptcy predictions for strategic decision-making purposes
4. To assess the impact of identified risk factors by utilizing explainable AI over different forecasting horizons on bankruptcy risk in the shipping industry and explore their applicability in the development of practical policies and managerial strategies

The first objective is to explore the development trend of bankruptcy prediction models. This involves conducting a comprehensive literature review to explore the development of these models, from initial methods like discriminant analysis to more advanced techniques such as logit analysis, neural networks, and decision trees. Understanding these developments will help identify the most effective methodologies for the shipping industry.

The second objective is to identify specific risk factors that contribute to the bankruptcy prediction of shipping industry, specifically for SMEs. This will be achieved by combining theoretical background from existing literature and practical insights from semi-structured interviews with industry practitioners. The focus will be on uncovering relevant financial, non-financial, and macroeconomic variables that enhance the predictive accuracy of bankruptcy models for SMEs in this sector.

The third objective is to investigate how bankruptcy risk factors vary across different forecasting horizons. By applying historical datasets to machine learning models, the study will analyse risk factors over short-term (1 year), medium-term (3 years), and long-term (5 years) periods prior to bankruptcy events. This analysis aims to improve the accuracy of bankruptcy predictions and provide critical insights for strategic decision-making within the shipping industry.

The fourth objective is to assess the impact of identified risk factors using explainable AI techniques. This involves evaluating the influence of various risk factors over different forecasting horizons to enhance the transparency and interpretability of machine learning models. The findings will offer valuable managerial and policy implications, enabling stakeholders in the shipping industry to proactively respond to potential bankruptcy risks and minimize associated losses.

In summary, this research aims to develop a comprehensive bankruptcy prediction model for the shipping industry, especially focusing on SMEs. By incorporating a wide range of financial, non-financial, and macroeconomic variables, and leveraging advanced machine learning techniques, the study seeks to improve predictive accuracy and provide practical insights for stakeholders. This will contribute to more effective bankruptcy risk management for not only for SMEs but also entire shipping industry including large firms.

1.5. Research Design

The aim of this research is to explore the determinants for predicting bankruptcy risk in the shipping industry, focusing on SMEs by incorporating financial, non-financial, and economic data, using machine learning models, through an analysis of the Korean shipping industry. This involves identifying significant risk factors for bankruptcy prediction through exploratory research and evaluating the predictive ability of these variables via empirical analysis using machine learning models. The research design comprises two main stages as shown in Figure 1.1.

Exploratory studies aim to identify potential bankruptcy risk factors and justify those with strong academic backgrounds. The research model development mainly relies on existing literature reviews, because the existing methodologies should be justified by previous research and then it can be applied to new research models (Liang et al., 2016). It is argued that existing methodologies should be justified by previous research and then applied to new research development (Adams et al., 2014). However, in order to discover the variables focused on specific area, several approaches such as in-depth interviews or focus groups can be conducted (Alaka et al., 2018). Therefore, this research attempts to discover risk factors of bankruptcy in shipping industry through both existing literature reviews and conducting semi-structured interviews.

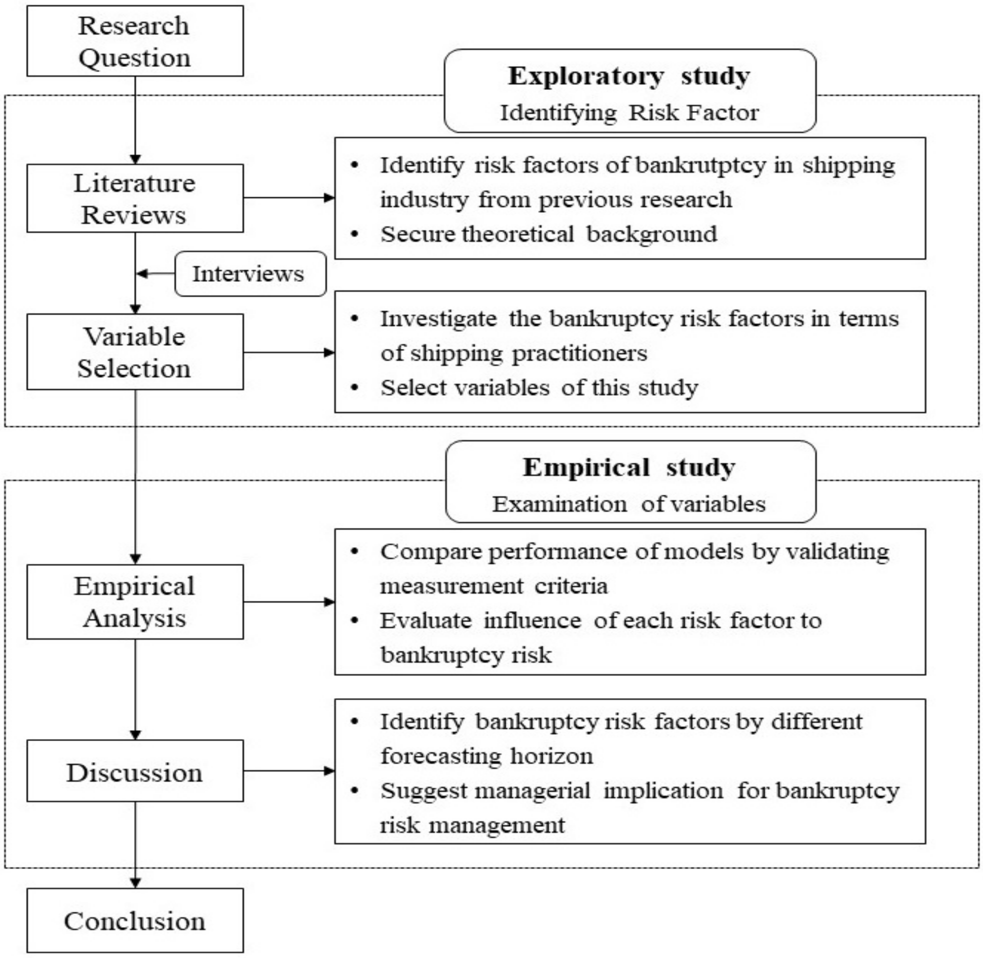
First of all, research variables would be identified to discover in not only financial factors but also other areas including economic or shipping index which can influence bankruptcy risk of shipping industry by previous literature reviews. Methodological approaches, including research modelling and data sampling techniques, are discussed in detail. To suggest specific risk factors, it would discuss characteristics and issues related with bankruptcy risk of shipping SMEs by analyzing status of Korean shipping industry. The research model can achieve a theoretical justification by the previous literature reviews and supplement through the interview approach (Adams et al., 2014). The explanatory variables constructed through the literature review would be verified and modified by interviewees who is part of shipping industry. Through the interview approach, new concept or new variables can also be identified in the perspective of practitioner in shipping industry.

The empirical study examines the predictive ability of variables proposed in the exploratory study by applying machine learning models. Initially, descriptive statistics such as mean values and trends of the collected data are provided to explore trends and outline the research process. Data preprocessing steps, including handling missing data, outliers, and data normalization, are conducted to prepare the input data for modelling. To avoid losing essential information, a feature selection process would be excluded.

The next step is to compare among machine learning models which proved its predictive ability from literature review. Two different types of models would be applied with taking into account each of strength in predicting bankruptcy: Extreme gradient boost and long short-term memory. These models offer high predictive accuracy and can handle statistical issues such as missing data, dimensionality, and skewness (Jones et al., 2017). The assessment of prediction ability uses criteria related to Type I and Type II errors, such as receiver operating characteristics (ROC). Different forecasting horizons (1 year, 3 years, and 5 years prior to bankruptcy) are applied to the models to identify varying sets of bankruptcy risk factors over time. This approach evaluates the short and long-term predictive ability of the models.

Finally, the influence of each risk factor on bankruptcy risk is assessed using explainable artificial intelligence (XAI) techniques. XAI plays a crucial role in revealing both the positive and negative impacts and the significance of relevant explanatory variables in predicting bankruptcy risk (Arrieta et al., 2020). This helps to uncover patterns and interactions between risk factors and bankruptcy risk. The sets of risk factors are categorized for large and small-sized shipping companies across different forecasting horizons. The findings provide policy implications for efficient bankruptcy risk management from the perspectives of various stakeholders, including managers, policymakers, and investors.

Figure 1.1. Research Design



Source: **Author**

1.6. Structure of this thesis

This thesis is organized into eight chapters, with the main body encompassing seven chapters (Chapters 2-7), which are divided into two key sections: the exploratory study and the empirical study, as illustrated in Figure 1.2.

Chapter 2 would discuss the overview of Korean shipping industry by suggesting specific industrial characteristics and emerged problem in the industry, which should be considered as bankruptcy risk factors. This chapter aims at identifying the main bankruptcy risk issues prevalent in the Korean shipping industry and justifying the relationship among them. The review of Korean shipping industry reveals some policy limitation and industrial problems which this thesis intends to overcome as the research objective.

Chapter 3 suggests literature reviews relevant to the bankruptcy prediction research. This chapter aims at suggesting theoretical background for targeting data, explanatory variables and research modelling. The discussion on the evolution of bankruptcy prediction models, detailed in Chapter 4, illuminates the methodological approaches used. This exploration also highlights the necessity for alternative bankruptcy predictors within the shipping industry, thereby motivating the selection of variables for this study. The identification of theoretical research gaps further justifies the necessity of this study.

Chapter 4 delves into various methodological approaches, covering aspects such as data sampling, research modelling, and analysis methods. The approach to bankruptcy prediction is critically examined in conjunction with the literature review. This comprehensive analysis serves to provide a methodological and modelling justification for this research.

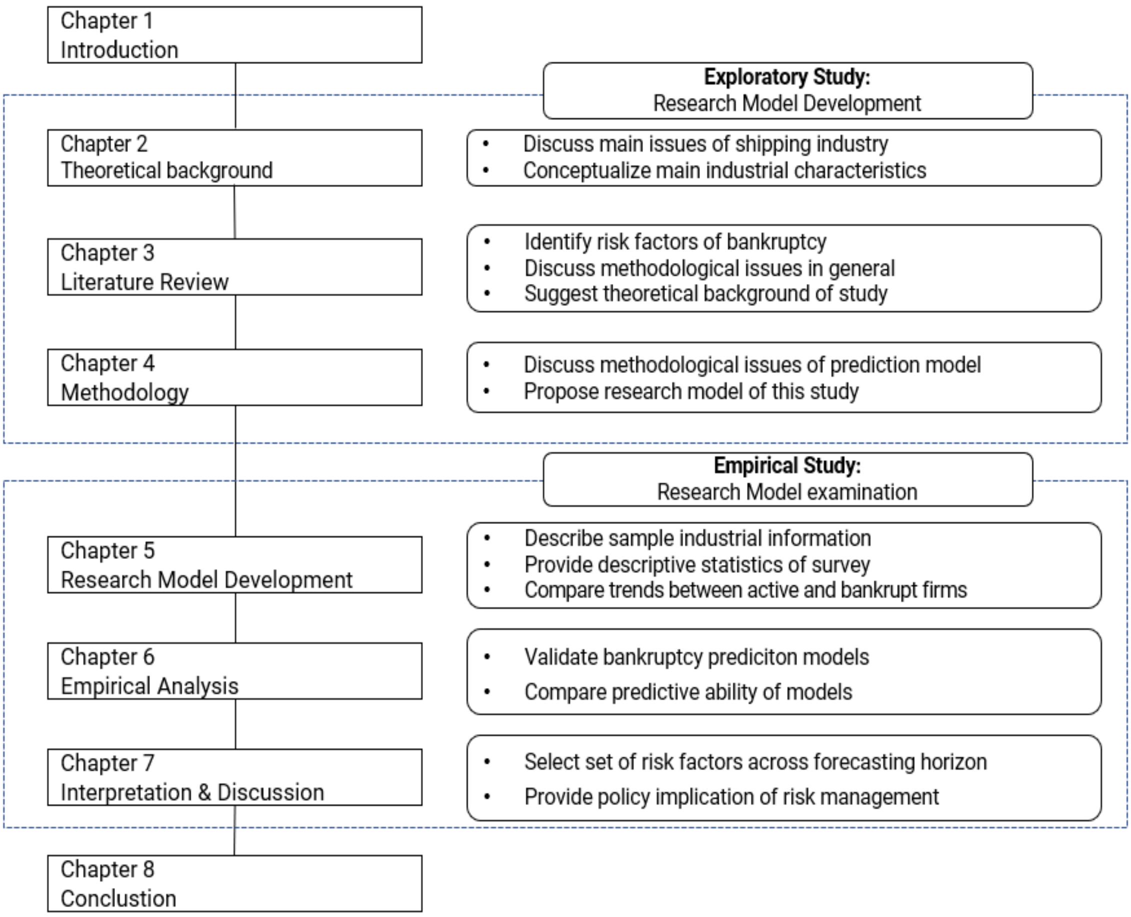
Chapter 5 presents developing process of research model in this study. To justify explanatory ability of variables for Korean shipping industry, semi-structured interview would be investigated how practitioners and academics perceive bankruptcy risk of shipping company. It follows that the all the explanatory variables are proposed. Then, descriptive statistics of the data would be suggested which collected from financial statement and interview study. It compares patterns by presenting descriptive statistics to differentiate between active and bankrupt firm years. Following this, the study discusses symptoms of bankruptcy observed over a five-year period preceding the event.

Chapter 6 presents the empirical results of two binary research model; extreme gradient boosting and long short-term memory. By comparing the predictive performances of two models with a confusion matrix, better models would be suggested in terms of different forecasting horizons.

Chapter 7 provides interpretation results with real data set of Korean shipping industry. By applying explainable artificial intelligent technique, the results of research model can be interpreted to present meaningful variables that have a high contribution to bankruptcy prediction. Afterwards, sets of risk factors would be presented divided into large and small-sized shipping companies in terms of different time periods (1, 3, 5 years prior to bankruptcy). This analysis would contribute to provide practical implication to various stakeholders for efficient bankruptcy risk management of Korean shipping industry.

Chapter 8 provides a comprehensive summary of the findings, discussing their implications for both theoretical frameworks and managerial practices, and details the contribution to different participants in shipping industry and limitations of this thesis.

Figure 1.2. Structure of thesis



Source: **Author**

2. CONTEXT BACKGROUND – KOREAN SHIPPING INDUSTRY

2.1. Introduction

The bankruptcy of Hanjin Shipping has been identified as a significant disruption to the logistics chain in the shipping industry, marking one of the most impactful events in recent years (Aydın and Kamal, 2022). Hanjin, once among the globe's largest shipping entities, accounted for an estimated 8% of the cargo turnover on the Pacific coast of the United States (Shin et al., 2019). In August 2016, the firm faced a judicial asset freeze request and was subsequently placed under external supervision, causing widespread disruption in global freight traffic (Song et al., 2019). This crisis was not limited to Hanjin Shipping alone. Following the 2008 financial crisis, other major companies like STX Pan Ocean and Korea Line, ranked among the top 10 based on ship gross tonnage, encountered severe financial distress, including court receivership due to bankruptcy risks (Lee, 2016). These events underscore the vulnerabilities within the Korean shipping industry and the critical need for improved risk management strategies.

Shipping industry's struggle with restructuring for financial viability was evident. SK Shipping, for instance, despite undertaking restructuring efforts such as divesting its bunkering business, could not avoid financial challenges, leading to its acquisition by Hahn & Company, a specialized investment entity (Kwon et al., 2023). Hyundai Merchant Marine (HMM) also endeavoured to fend off court receivership by liquidating its dedicated ship business and other group entities. Nevertheless, HMM was detached from the Hyundai Group and came under the control of its creditors, led by the Korea Development Bank, demonstrating the pervasive debt repayment challenges within Korea's top five shipping firms (Kim et al., 2022).

Bankruptcy not only precipitates severe macroeconomic implications for social welfare but also yields adverse microeconomic outcomes for the various stakeholders of the affected corporation (Ooghe and De Prijcker, 2008). The collapse of significant enterprises, particularly, exerts a pronounced impact on industries, more so on small and medium-sized enterprises (SMEs), compelling governments to alleviate the social and economic fallout of such bankruptcies (Eklund et al., 2020). The bankruptcy of Hanjin Shipping highlighted significant vulnerabilities within the Korean shipping industry, particularly affecting small and medium-sized enterprises (SMEs). This event underscored the prolonged downturn and oversupply of fleet capacity that has led to diminished freight rates and increased financial distress within the industry (Shin et al., 2019). Such conditions have been particularly challenging for smaller shipping companies that rely heavily on sea freight for their operational cash flows (Park et al., 2021c). The persistent decrease in freight rates, a direct consequence of the industry's extended slump, has emerged as a key indicator of increasing bankruptcy risk within the shipping sector, underscoring issues related to companies' external debt repayment abilities (Park et al., 2021b).

The Korean shipping industry has seen numerous companies go bankrupt due to their inability to effectively navigate market fluctuations, especially during economic downturns following growth periods (Park et al., 2022). The prolonged recessions have led to a continuous increase in ship supply, attributed to counter-cyclical investments, causing an enduring surplus of fleet capacity even as ship orders surged (Stopford, 2013). Furthermore, concerted efforts by China and Korea to stimulate their shipbuilding sectors resulted in an oversupply that disrupted the natural equilibrium of supply and demand dictated by freight rate economics (Choi et al., 2018b). From the shipping companies' standpoint, there's a pressing need to devise a framework capable of systematically managing structures that influence profit and loss volatility, a feature that is notably absent in all but a few of the major Korean shipping enterprises (Nam and An, 2017).

Throughout periods of downturn in the shipping industry, numerous small-sized shipping companies have struggled to effectively manage financial risks, often entering a detrimental cycle of restructuring and liquidating high-quality assets to offset liabilities (Choi et al., 2018b). Despite the government's efforts to implement policies aimed at stabilizing the shipping market, a targeted early warning system for the highly volatile shipping sector has yet to be developed, with existing measures primarily benefiting larger corporations. This oversight highlights the critical need to pinpoint specific characteristics of the Korean shipping industry that serve as indicators for bankruptcy risk among small and medium-sized enterprises (SMEs). Consequently, This study identifies risk factors unique to shipping SMEs through a detailed analysis of the Korean shipping industry, using it as a case study to illustrate broader trends and challenges.

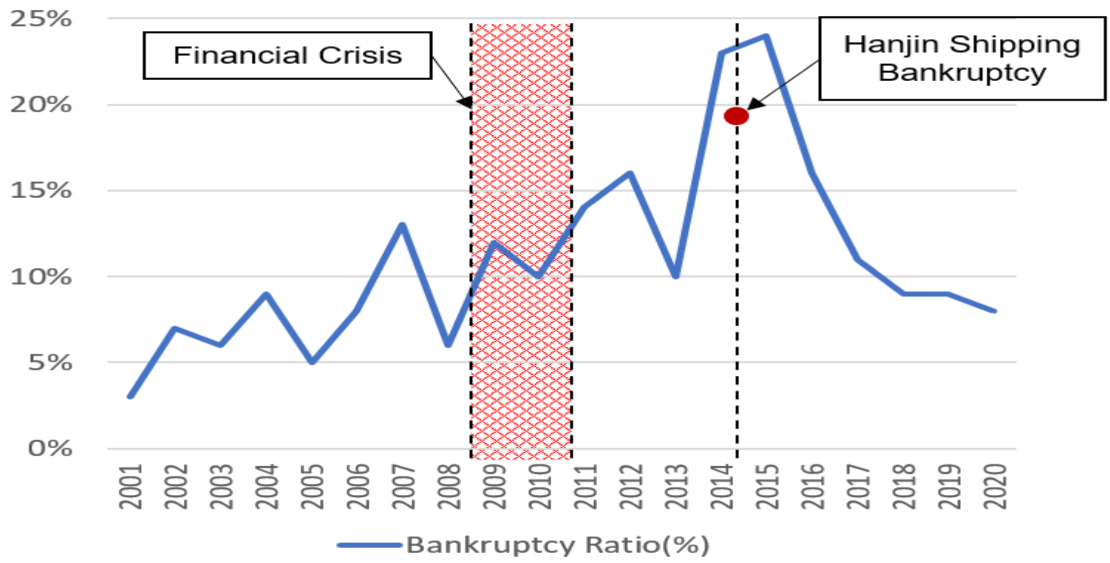
2.2. Current status of Korean shipping industry

After the 2008 financial crisis, the Korean shipping industry faced a severe economic downturn, which caused the default rate of shipping companies to increase significantly from 6% in 2008 to 24% in 2014 as shown in Figure 2.1. This data is based on official statistics provided by the Korean Ministry of Maritime Affairs & Fisheries[3] (2023), which oversees maritime policies and fisheries development. During this period, Hanjin Shipping, once Korea's largest and the world's seventh-largest shipping company, filed for bankruptcy. The liquidation of Hanjin's core assets to global shipping giants significantly diminished Korea’s fleet capacity and its presence in crucial shipping routes, leading to a decline in the overall revenue of the Korean shipping industry. The Samjung KPMG Research Institute[4] provides financial analysis into the Korean industry, highlighting the economic health and risks faced by companies. According to annual

[3] Available at: https:// https://www.mof.go.kr/statPortal/cate/partStat.do
[4] Available at: https://kpmg.com/kr/ko/home/industries/

report of KPMG (2021), Maersk, the world's No. 1 shipping company, took over 6 large vessels of 10,000 TEU capacity, and Mediterranean shipping co, the 2nd shipping company, acquired 3 large vessels and Long Beach terminal which is the core of US shipping route. As a result, Korea's ocean container capacity decreased by about 350,000 TEU from 1.05 million TEU in August 2016 to 700,000 TEU as of August 20, and its share in the Asia-Americas market also decreased from 12.2% to 7% during the same period.

Figure 2.1. Bankruptcy ratio of Korean shipping industry

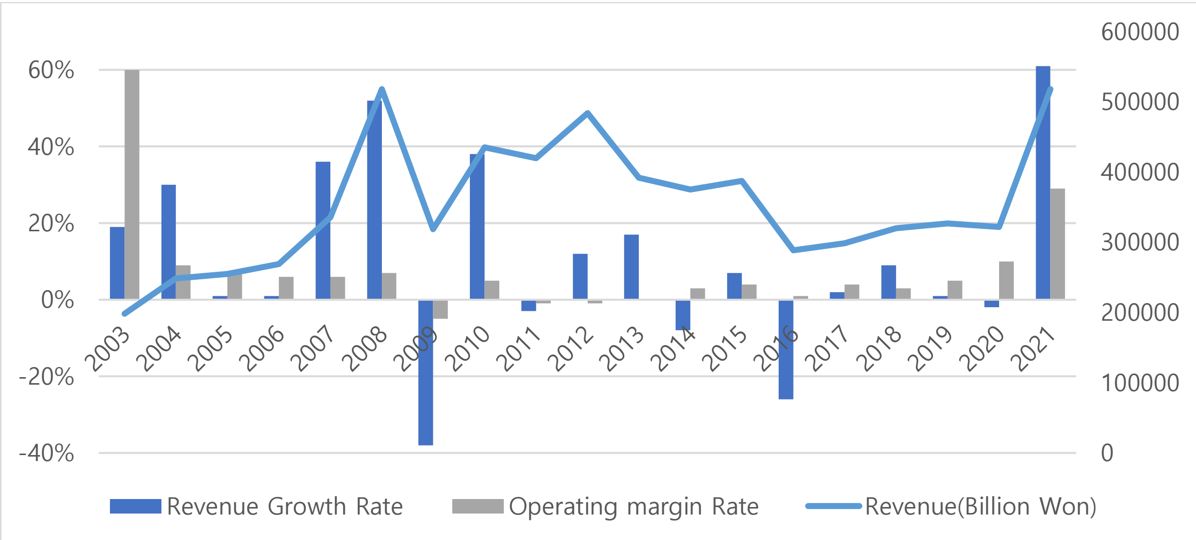


Source: **Korean Ministry of maritime affairs & fisheries (2023)**

Moreover, the global shipping industry has experienced a continuous decrease in overall freight rates, driven by a growth in shipping capacity that has outpaced the increase in freight volume since 2008 (Kalgora and Christian, 2016). This imbalance was deepened by aggressive investment during the mid-2000s boom, leading to an excess of shipping capacity without a corresponding rise in cargo volume, thereby causing a sharp fall in freight rates and financial burden for shipping companies (Lee, 2020). The Baltic Dry Index (BDI), a benchmark for bulk freight rates, and the China Container Freight Index (CCFI), indicative of container freight rates from China, both experienced significant declines as a result of the oversupply and fierce price competition (Notteboom et al., 2021). The BDI plummeted from 7,071 in 2007 to 920 in 2012, though it has since recovered slightly to 1,383 in 2021. Similarly, the CCFI fluctuated from 1,071 in 2007 to 993 in 2011 and 1,085 in 2014.

The Korean Ocean Business Corporation[5] (KOBC) is a key entity that oversees the operational aspects of Korea's maritime activities, providing comprehensive reviews and performance metrics. Since 2008, Korean shipping companies have consistently reported low operating margins, as depicted in Figure 2.2, with an average operating margin ratio of 3% from 2003 to 2019 (KOBC, 2023). Currently, the debt-to-equity ratio, a measure of corporate leverage, has been on the rise within the Korean shipping industry, peaking at 432% in 2012 and 2013, as shown in Figure 2.3. However, the industry has shown signs of recovery recently, with domestic shipping companies' total sales reaching KRW 51,797.1 billion in 2021, marking a 61% increase from 2020 and indicating a significant improvement from the average operating profit margin observed between 2003 and 2019 (KOBC, 2021). This recovery has also been reflected in the debt-to-equity ratio, which fell to 123%, the lowest since 2003, signalling a recent improvement in financial health.

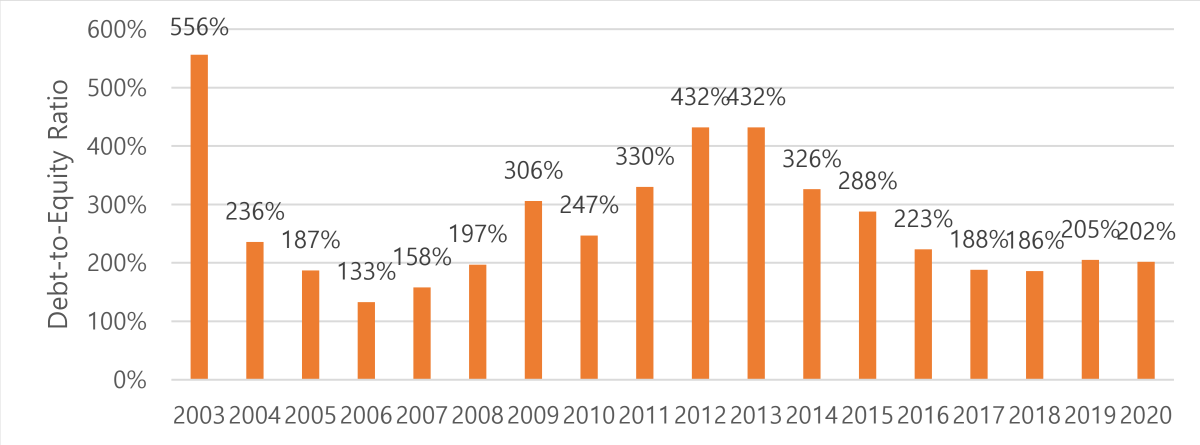
Figure 2.2. Revenue of Korean shipping industry over the last 20 years



Source: **Korean ocean business operation (2023)**

[5] Available at: https://www.kobc.or.kr/ebz/shippinginfo/main.do

Figure 2.3. Average Debt-to-Equity ratio of Korean shipping industry



Source: **Korean ocean business operation (2021)**

Despite these positive developments, Korean shipping companies face challenges in maintaining their global competitiveness, especially as global shipping giants continue to expand through mergers and acquisitions (M&A) and alliances. Major players like Maersk, COSCO, and CMA CGM have not only grown through M&A but have also sought to enhance cost efficiency by commissioning extra-large vessels. In contrast, Hyundai Merchant Marine (HMM), a leading Korean shipping firm, has seen a relative decrease in size due to internal restructuring. Moreover, with the construction of over 100 large vessels, including CMA-CGM's order for nine 22,000 TEU class ships and COSCO's six 20,143 TEU class vessels, global shipping companies are significantly boosting their capacities (Ghorbani et al., 2022). Compared to these expansions, HMM's fleet, including its super-large ships, remains modest, possessing only 16% of the capacity of Maersk, the world's largest shipping company, and 51% of Evergreen, ranked seventh (Song et al., 2019). Despite recent government-supported orders for 20 super-large vessels, HMM's capacity in the 12,000TEU class segment still lags behind global standards. The global shipping market has increasingly become an oligopoly, dominated by alliances that control 81.5% of the supply among the top nine shipping companies (KPMG, 2021). Although HMM joined The Alliance in April 2020, its fleet size and market position are relatively weak, necessitating the pursuit of new shippers and additional cargo to bolster its standing. Furthermore, SM Merchant Marine, which acquired Hanjin Shipping's US route, faces growth limitations due to its limited capacity and a focused shipping route, holding only a 0.2% supply share (Park et al., 2022).

The financial structure of small and medium-sized enterprises (SMEs) in the Korean shipping industry, defined as companies with total assets less than 50 billion won and average sales less than 8 billion won, presents unique challenges. Over 90% of shipping companies in Korea are SMEs, operating alongside larger firms but facing greater difficulties due to limited financial resources and less diversified business portfolios. SMEs in the shipping industry often rely heavily on short-term credits, making them particularly vulnerable to market fluctuations and economic downturns. This financial fragility is exacerbated by high debt ratios, with many SMEs reporting debt ratios exceeding 200% and some even above 1000%. This instability underscores the need for targeted support and robust risk management frameworks tailored to the unique conditions of SMEs in the shipping sector.

2.3. Inappropriate government response

In response to the global financial crisis of 2009, the Korean government unveiled a plan aimed at restructuring and boosting the competitiveness of the shipping industry. Recognizing the industry's crisis state, policies introduced in 2009 were designed to mitigate the immediate challenges through industry-specific restructuring, investment revitalization, and halting speculative charter ship operations (Jeon et al., 2017). Additionally, to lay the groundwork for sustained growth, the government promoted policies such as i) developing investment institutions specialized in shipping, ii) stabilizing the shipping tax framework with measures like the tonnage tax and international ship registration system, iii) establishing a reliable national cargo transport base, iv) supporting overseas market development, v) enhancing registration standards for outbound cargo transport businesses, and vi) improving shipping market analysis capabilities (Lee, 2020). Despite these efforts, the lack of immediate and effective policy execution meant that the anticipated revitalization of the shipping industry was not realized. As the financial health of major domestic shipping companies continued to decline, the government, starting in earnest in 2013, introduced measures like marine guarantee insurance and subsidies, which served as temporary fixes rather than foundational solutions (Park et al., 2022).

The bankruptcy of Hanjin Shipping led to a reduction in global shipping routes for domestic companies, constraining shippers' options and compelling them to opt for foreign shipping companies due to their lower freight rates (KOBC, 2019). Additionally, the long-term absence of a robust Korean shipping contender raises concerns over potential freight rate increases and diminished freight rate negotiation power for shippers. A reduction in ship supply by global shipping firms to enhance their bargaining position could jeopardize the availability of stable services (KOBC, 2019). As of 2019, the cargo load rate on national vessels was 47%, significantly lower than Japan's 64%. With countries like Japan and the United States prioritizing their vessels for strategic cargo, there's a pressing need for Korea to adopt strategies to boost container cargo loads on national ships, ensuring steady business for Korean shipping companies. In response, The government's “Five-Year Shipping Reconstruction Plan 2018-2022” aimed to increase the acquisition rate of national ships, support new shipbuilding, and provide ship management stabilization (Park et al., 2021b). This plan facilitated long-term contracts between shippers and Korean shipping companies and introduced the “Excellent Shippers Certification System” to incentivize the use of national carriers. Shippers who paid higher freight rates than in the previous year to Korean companies were rewarded with a 1% basic discount on shipping fees, plus an additional 3% off for any increase in freight rates compared to the prior year (Park et al., 2022). These measures aimed to reduce the financial burden and debt ratios linked to ship ownership, allowing companies to focus on enhancing their services. However, due to the continued low national cargo load rate and the lack of global operational routes, it is estimated that Korean shippers incur an additional KRW 1.4 trillion annually in costs compared to their Japanese counterparts (Park et al., 2021c).

Nevertheless, the Korean government's policies have been criticized for insufficiently reflecting the unique attributes of the shipping industry and lacking a proactive risk management system. Particularly during periods of high freight rates and ship prices, the industry saw significant orders for new ships and long-term charter contracts, which, in times of economic downturn, exacerbated the impact on the entire shipping sector (Jeong, 2021a). The government’s approach, overly focused on financial liquidity from a short-term perspective, failed to consider the industry's specific needs, thus not effectively enhancing competitiveness, especially for small and medium-sized enterprises (SMEs) in shipping (Jeong, 2021b). This failure is believed to have contributed to the current challenges, stemming from the absence of medium- to long-term support strategies and an inadequate assessment of the declining management health within shipping companies. In contrast, during crises such as the 2008 global financial downturn and the Covid-19 pandemic, countries like the United States and Japan have supported their shipping industries through policies tailored to their unique characteristics and strategic needs (Narvekar and Guha, 2021). In particular Major global shipping firms that suffered poor business performance and financial instability post-financial crisis benefited from both direct and indirect government assistance, as detailed in Table 2.1.

Table 2.1. Support policies for shipping industry by major countries

Country	Major Company	Support policy
China	**COSCO CSCL**	· Export-import bank of China provided $9.5 billion each for 5 years (2012~) · Enforcement of the ship dismantling subsidy support program estimated $1 billion in 2014 · Provide $18 billion in financial support to promote overseas M&A (2016)
Germany	**Hapag-Lloyd**	· Government payment guarantee of €1.2 billion (2009) · Cash grant of €750 million of Hamburg (2013)
Denmark	**Maersk**	· Supported $ 500 million in Export Credit Fund (2009) · Support for financial borrowing of $ 6.2 billion by policy financial institutions
Japan	**MOL, NYK**	· Issuance of 10-year corporate bonds with an interest rate of 1%
France	**CMA-CGM**	· $1.5 billion in loan guarantees from creditor banks · Additional $1.5 billion in support through the sovereign wealth fund · Additional support of €280 million for 3 years through the financial sector (2013~)
Singapore	**NOL**	· Government investment firm, Temasek acquires $1 billion of preferred stock
Korea	**HMM**	· Purchase ship by Korean Asset Management Corporation (KAMCO) · Introduction of the rapid underwriting system for corporate bonds · Establishment of ship fund ($ 1.2 billion)

Source: **Korean ocean business operation (2023)**

Maersk received a $500 million aid package from EKF, Denmark's export credit agency, and a $6.2 billion financial loan from policy banks in 2009. Similarly, CMA-CGM was supported with $150 million from the French sovereign wealth fund, alongside additional financial backing including loan guarantees. Notably, China's state-owned enterprises, COSCO and CSCL, were bolstered by a $9.5 billion injection from the Export-Import

Bank of China, along with governmental financial support for international mergers and acquisitions (M&A) to boost global competitiveness. When compared to these international examples, the support extended by the Korean government appears limited in both scale and approach. Shipping companies abroad that received substantial government support navigated through the crises more effectively and leveraged the opportunity to invest in mega-ships financed at relatively lower interest rates, thereby securing a competitive edge.

Contrastingly, the Korean government's approach, during a period of low freight rates and shipping costs, focused on encouraging the acquisition of second-hand ships to expand fleet capacity and offering tax incentives to support this strategy (Jeon et al., 2017). This policy anticipated a boost in shipping companies' sales and an improvement in global rankings through aggressive fleet expansion. However, this strategy diverged from international practices, particularly in its failure to adequately assess market conditions. This led to substantial orders for new shipbuilding at times of high freight rates and shipping prices, without appropriately accounting for investment risks, including those related to chartering and fleet management decisions (KPMG, 2021). The resultant increase in debt ratios from the mass acquisition of used ships further escalated interest expenses and compounded the tax burden, contrasting sharply with the direct and substantial support seen in other countries.

2.4. Characteristics of Korean shipping industry

2.4.1. Unstable Financial Structure

In terms of corporate finance, achieving an optimal capital structure is crucial for enhancing a company's financial health and resilience (Berk et al., 2013). The Korean shipping industry's primary challenge lies in its excessive reliance on debt financing, leading to a capital structure that severely impacts its financial stability. This deep dependency on debt not only increases the financial burden on shipping companies but also weakens their ability to endure economic downturns due to the heavy costs associated with servicing this debt. As presented in Table 2.2, recent statistical data from the Korean Maritime Institute[6] (KMI), a Korean government institution focused on maritime policy and economics, highlights that the average debt ratio of major Korean shipping companies over the last eight years stood at a staggering 469%, markedly higher than the 195.9% average observed among leading global shipping firms. This significant disparity, as presented in Table 2.2, underscores the Korean shipping industry's predominant reliance on debt for financing (KMI, 2021).

Hyundai Merchant Marine, a notable entity in Korea's ocean container shipping sector, exhibits an average debt-to-equity ratio exceeding 763%, nearly 2.7 times that of the highest among international shipping companies. This skewed capital structure towards debt financing starkly contrasts with healthier financial practices and places immense pressure on companies, especially during economic recessions. During such times, the obligation to meet debt repayments heightens the risk of bankruptcy.

This financial pressure affects not only large corporations but also impacts small and medium-sized enterprises (SMEs) within the Korean shipping industry. As illustrated in Figure 2.4, More than 52% of these companies report a debt ratio exceeding 200%, with 7% grappling with extremely high debt ratios above 1000% (KPMG, 2021). These figures highlight the pervasive financial risk due to elevated leverage levels, significantly increasing the likelihood of bankruptcy and the potential for credit rating downgrades across the sector.

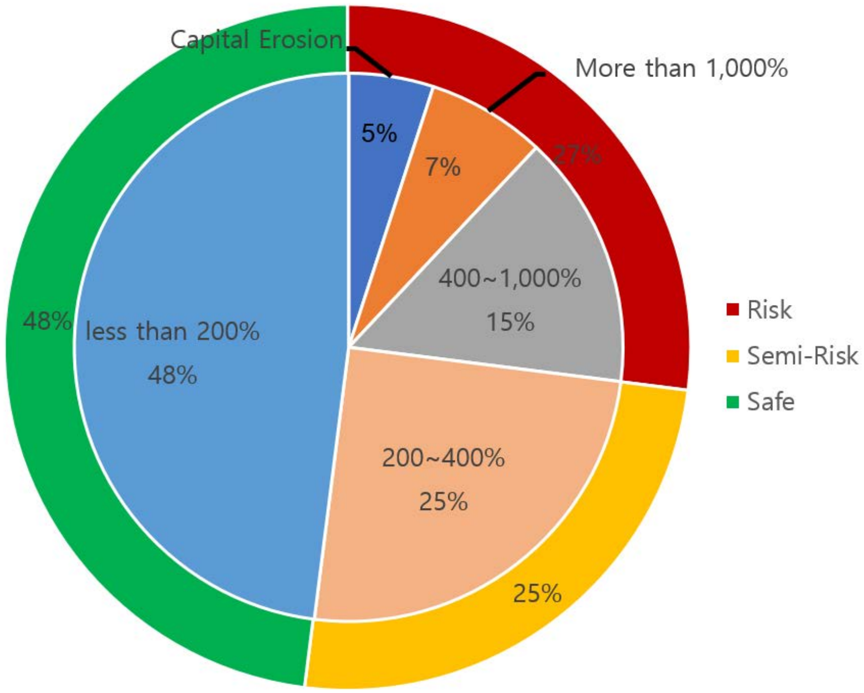
[6] Available at: https://www.kmi.re.kr/

Table 2.2. Debt-to-equity ratio of major shipping companies

Company	2013	2014	2015	2016	2017	2018	2019	2020	Average
Korea average	747.3	565.8	872.0	300.0	286.0	264.2	378.6	344.6	469.8
HMM	1185	959	2007	350	302	294	556	455	763.5
Korea Line	202	161	267	252	296	258	282	257	246.9
Hanjin Shipping	1445	968	816	-	-	-	-	-	1076.3
KSS Line	157	175	398	297.9	260.01	240.65	297.81	321.86	268.5
World average	193	170	170	218.7	204.6	231.7	212.4	166.4	195.9
Maersk	75	63	75	90	101	70	92	82	81.0
Hapag-Lloyd	138	142	120	124	145	145	145	126	135.6
CMA-CGM	214	188	164	279	158	201	171	138	189.1
COSCO	284	246	230	219	205	305	279	246	251.8
MOL	202	194	243	248	258	228	228	200	225.1
NYK	230	192	166	245	236	248	224	165	213.3
K-Line	206	162	194	326	329	425	348	208	274.8

Source: Korean maritime institute (2021)

Figure 2.4. Status of Korean shipping industry based on debt-ratio



Source: Samjung KPMG research institute (2021)

2.4.2. Simple Business Portfolio

The Korean shipping industry predominantly operates within the bulk carrier domain as summarized in Table 2.3. Of the total 157 shipping companies in Korea, 143 are engaged in bulk carrier operations, including tanker services, alongside 14 container carrier operators (KMI, 2022) Container shipping companies consist of 12 intra-Asian shipping companies, including 2 ocean shipping companies, and some of these container shipping companies also transport bulk cargo. As of 2022, about 49% of outbound shipping companies are bulk carriers, 37% are tankers, and only 14% of container ships. Furthermore, proportion of container vessels accounts for just 9% of the total deadweight tonnage capacity, illustrating the industry's lean towards small-scale operations and short-distance services. These narrow operations expose companies to significant risk from fluctuations in specific freight rates, especially given their reliance on a single cargo type.

Table 2.3. Types of vessels in Korean shipping industry in 2022

	No. of Vessel	Gross Tonnage	DWT
Container Vessel	143 (14%)	6,655,054 (16%)	7,108,836 (9%)
Bulk Carrier	502 (49%)	27,384,706 (66%)	45,672,829 (61%)
Tanker	374 (37%)	7,767,195 (18%)	22,061,819 (30%)
Total	1,019	41,806,955	74,843,484

Source: Korean Maritime Institute (2022)

Hanjin Shipping's bankruptcy in 2017 serves as a stark illustration of the risks associated with over-reliance on a narrow operational focus. In 2015, Hanjin's revenue was heavily skewed towards its container operations, contributing 92.4% (7.14 trillion won) of its total revenue of 7.73 trillion won, with the bulk carrier segment making up just 6.7% (Shin et al., 2019). The company's dependence on key routes, such as the Asia-Europe and Asia-North America lanes, which accounted for 60-70% of its operations, significantly contributed to its vulnerability. The downturn in the container shipping market during 2015 and 2016 led to substantial operational losses, demonstrating the inherent risks of a simplified business model (KPMG, 2021). While such a model can generate considerable profits in favourable market conditions, it equally poses a significant risk of losses during downturns.

Global container operators, particularly Japanese shipping companies, Chinese firms, and Maersk, demonstrate a strategic advantage through their diversified business portfolios. NS united Kaiun Kaish, Japanese shipping company, for instance, participate in a wide array of maritime activities, including dry cargo, tankers, car-only vessels, and cruise ships, beyond their container transport operations (Lee, 2020). The container segment constitutes only 42-57% of their business, with these companies also leading globally in the exclusive car carrier segment, securing stable profits. Their bulk cargo and tanker operations show robust performance, anchored by long-term contracts with key shippers like steel and power generation companies, as well as collaborations with Japanese oil refiners and petrochemical firms (KPMG, 2021). This diversification allows Japanese shipping companies to offset the reduced profitability in the container sector with revenue from other business areas (Park et al., 2021b).

Maersk, despite its significant reliance on container shipping for revenue, has effectively navigated economic fluctuations in the industry through its engagement in non-shipping activities. Notably, its oil development business, accounting for about 20% of total revenue between 2010 and 2014, boasted profit margins of approximately 60% (Lee, 2016). This emphasizes the importance for shipping companies to diversify their operations across different ship types and regions, leveraging profits from various sectors to remain competitive in the global market.

Conversely, Korean shipping companies primarily focus on either container or bulk shipping, a reflection of their smaller size. This narrow concentration heightens their vulnerability to market and freight rate fluctuations, increasing the risk of financial instability. The challenges faced by Hanjin Shipping and other major bulk cargo operators, which have undergone court receivership, restructuring, or mergers and acquisitions, illustrate the potential risks of a limited business model (Yoon et al., 2023). The reliance on a single sector significantly contributes to the risk of bankruptcy, highlighting the need for Korean shipping companies to consider broader diversification strategies.

2.4.3. Failed to specialize in business

Korean shipping companies engage in both shipowner and operation businesses, blending the roles of owning and operating vessels without specializing in either. This dual approach introduces complexities, as each segment carries its distinct risks. Shipowner business involves ship sales and purchase (S&P) and chartering, while the operation business focuses on generating revenue through shipping activities (Sirimanne et al., 2019). The concurrent management of these businesses amplifies the risk exposure, encompassing funding security, interest rate and ship price fluctuations, chartering challenges, and vessel accidents for shipowners, alongside freight rate variability, freight volume changes, and operational incidents for operators. This structure inherently doubles the risk for Korean shipping companies, complicating their operational landscape (Park et al., 2022). A significant hurdle for these companies is the specialization in ship-owning, attributed to the limited number of vessels owned. Among 165 Korean shipping companies, merely 23 own ten or more ships, with the majority relying on chartered vessels (KMI, 2022). As shown in Table 2.4, Over 50% of vessels and 83% of the deadweight tonnage are chartered, with owned vessels accounting for only 14% of deadweight tonnage despite comprising 37% of the fleet, which indicates a heavy reliance on chartering for cargo operations.

Table 2.4. Number of vessels according to ownership type

	No. of Vessel	Gross Tonnage	DWT
Chartering	537 (53%)	37,864,679 (78%)	61,982,294 (83%)
Owning	375 (37%)	9,035,331 (19%)	10,815,408 (14%)
Used	107 (11%)	1,403,085 (3%)	1,937,932 (3%)

Source: Korean ocean business operation (2021)

After the Asian financial crisis in 1997, the Korean government uniformly limited the debt-to-equity ratio at 200% across all industries (Kwak et al., 2012). However, the shipping industry is a capital-intensive industry in which most of the 80-90% of the capital are raised from banks as a loan to build a single vessel (Mok; and Ryoo, 2022). This regulation forced shipping companies to sell owned vessels and pivot to chartering, exacerbating financial vulnerabilities. For instance, Hanjin Shipping and Hyundai Merchant Marine, Korea's leading shipping firms, allocated approximately 13% and 33% of their total revenues to chartering costs, respectively, significantly impacting liquidity as a substantial portion of revenue (15~35%) was paid to foreign shipowners (Hwang et al., 2020). The situation worsened when companies, amid the global financial crisis, locked into long-term charter contracts at rates four to five times higher than pre-crisis levels, leading to soaring debts (Park et al., 2022). This strategy, coupled with the sale of high-quality assets to evade court receivership, precipitated a cycle of declining profitability. Ultimately, Hanjin Shipping declared bankruptcy, and HMM underwent restructuring, with these consequences rippling through the smaller and medium-sized segments of the Korean shipping industry.

While large shipping companies may initially withstand high debt ratios due to favourable exchange rates and strong performance, smaller and medium-sized firms face greater challenges under these financial burdens. The prospect of an extended period of high interest rates poses additional concerns across the shipping industry (KOBC, 2023). The Korean Ocean Business Corporation conducted an analysis to assess the impact of rising interest rates on 127 shipping companies, using the end of 2023 data and a baseline interest rate of 1.0%. As presented in Table 2.5, this analysis indicated that a 1.00% increase in the base rate in 2023 would lead to a 39% rise in total interest expenses for these companies, amounting to an additional KRW 552.5 billion. A further increase in interest expenses by 58%, or KRW 828.7 billion, is anticipated under this scenario.

Therefore, Korean shipping SMEs face amplified risks due to their dual role in shipowning and operations, compounded by heavy reliance on chartering and vulnerability to fluctuating financial regulations and market conditions.

Table 2.5. Scenario based on level of interest rates of Korean shipping industry

	Interest Rate 1% (2023)	100bp	125bp	150bp
Interest Expense	1,423,754	1,976,262	2,114,389	2,252,516
Increase	-	552,509	690,636	828,763
Growth Rate	-	39%	49%	58%

Source: Korean ocean business operation (2023)

2.4.4. Poor risk management system for small and medium sized shipping company

A crucial competency for shipping companies lies in their ability to navigate and manage market fluctuations. Effective management of these fluctuations is vital for stabilizing the inherently volatile shipping industry through proactive market condition assessments and predictions (Berk et al., 2013). The development of systems to systematically handle the volatility of profits and losses, by accurately measuring freight rate fluctuations and market volatility, is essential for the shipping sector (Yuen and Ko, 2018).

Despite widespread recognition of the importance of market forecasting, many companies struggle to establish systems that ensure the capability to make reliable predictions (Ashraf et al., 2019). Thus, only a few companies obtain this system to continuously maintain their capabilities, even in case of large sized shipping companies. As a result, many small sized shipping companies failed to manage market risk efficiently, resulting in bankruptcy during the recession after the economic boom. In other words, this means that when the market condition goes into a downturn due to the influence of supply increase during the economic boom, because of the failure to systematically manage the market risk caused by freight rate fluctuations, the loss of exceeds the tolerant range (Notteboom et al., 2021). Meanwhile, considering the high debt ratio of the shipping industry, financial institutions also need to secure the capability and system to respond expected financial risk by closely monitoring the market risk of shipping companies received financing (Park et al., 2022). In terms of shipping industry, it is important to secure stable financing source in that small and medium sized shipping companies suffered from liquidity deterioration during each of economic recession, and have been continued the vicious cycle of repaying liabilities by selling high-quality, profitable assets through restructuring. Therefore, it means that it is necessary to take a proactive approach by developing risk management system to prepare for liquidity deterioration due to a decrease in freight revenue during a shipping recession, especially focusing on the case of small and medium sized shipping companies.

2.5. Chapter Summary

This chapter provides a detailed examination of the Korean shipping industry, highlighting its systemic vulnerability to market fluctuations and financial crises. The discussion extends from the Asian financial crisis in the 1990s through the global financial disruption in 2008, illustrating the Korean government's policy measures' inadequacy in reducing industry volatility. The bankruptcy of Hanjin Shipping, along with wider industry challenges, emphasizes the profound impact of financial instability on the shipping sector and the significant vulnerability of SMEs to changes in the economy. The government's policy efforts, predominantly based on financial indicators, proved insufficient in protecting the industry from market instability and economic challenges, such as poor ship investment conditions. Efforts to implement an early warning system as part of the 'long-term development plan for the shipping industry' were limited by a lack of thorough industry analysis and failure to address the unique challenges faced by SMEs, despite their critical role in the industry (Park et al., 2022). This situation highlights the urgent need for a bankruptcy risk management framework that enables shipping SMEs to foresee market uncertainties and respond to bankruptcy risks effectively. The study suggests moving beyond traditional bankruptcy risk management methods by combining specific industry indicators with financial data, thus addressing the full spectrum of market instability and the complexities of business operations. Identifying potential risks and indicators of insolvency is essential, requiring a specialized risk management framework tailored to the shipping industry's unique challenges. The evolving nature of global economic crises, with anticipated future downturns likely to present new challenges influenced by a variety of risk factors such as trade rate disputes, trade conflicts, and the COVID-19 pandemic (Kamal et al., 2021).

This investigation into the Korean shipping industry's responses to past economic downturns and current risk management efforts not only highlight the specific difficulties encountered by shipping SMEs but also sets the Korean shipping sector as a crucial example for the global shipping industry. By focusing on risk factors unique to shipping SMEs, this research deepens understanding of the industrial vulnerabilities and establishes a basis for establishing more effective risk management strategies. Insights from the Korean shipping industry's experiences can be guidance for addressing the challenges of the global shipping sector, emphasizing the need for thorough risk assessment, strategic planning, and government support designed to meet the unique requirements of shipping SMEs.

In conclusion, this chapter highlights the critical risks faced by shipping SMEs and their effects on the bankruptcy rates of Korean shipping companies, presenting the Korean shipping industry as an important model for the maritime sector globally. It calls for the enhancement of risk management practices and the incorporation of strategic foresight, urging a holistic and proactive approach to risk management in the maritime industry.

This analysis underlines the importance of bankruptcy prediction for shipping SMEs through the study of the Korean shipping industry, offering valuable insights into crisis and risk management in the global shipping industry.

3. LITERATURE REVIEW

3.1. Chapter Introduction

There is a growing body of research to proactively manage and improve efficiency in response to potential bankruptcy risks, especially financial crises, by predicting corporate bankruptcies (Barboza et al., 2017). There have been various attempts to establish bankruptcy prediction models by using historical data to determine the extent of bankruptcy risk faced by a company and whether it will trigger a financial crisis. Once the bankruptcy risk analysis is clarified, appropriate preventive and corrective actions could be established immediately to minimize risk-related losses (Visvanathan, 2021). A corporate bankruptcy prediction is closely related to the uncertainty of the industry's internal and external environment. Because of capital is the essential part for corporate development, most companies need to raise capital despite of limited funds. Corporate bankruptcy forecasting can help management control the outbreak of financial crises and autonomously regulate business operations by detecting and resolving problems in production, operations or control as soon as possible. For predicting corporate bankruptcy effectively, the accurate, timely and diverse prediction models must set up which meet the needs of corporate, operators and department heads with timely and complete business data (Altman et al., 2020).

Traditional research on bankruptcy prediction initially utilized advanced statistical methodologies, including logistic regression, probit analysis, linear discriminant analysis, survival analysis, as well as linear and quadratic programming (Chen, 2011). However, empirical evidence has frequently pointed out the limitations of these approaches, particularly the violation of their foundational assumptions like multivariate normality and the independence of explanatory variables (Dangeti, 2017). In contrast, machine learning and computational intelligence methods offer a different paradigm. Those methods do not require specific prior assumptions but rather learn from historical data, making use of a wide array of explanatory variables that include financial ratios, macroeconomic indicators, and sociodemographic characteristic (Chen et al., 2019). These variables are either treated as continuous quantities or categorized into qualitative groups. Machine learning models such as artificial neural network, decision tree and support vector machine have gained prominence for their application in finance, especially in assessing credit risk, and corporate bankruptcy (Shetty et al., 2022).

This research aims to chart the course of study in this field by (1) reviewing how bankruptcy prediction has evolved, (2) situating this study within the existing knowledge base, (3) identifying research gaps in prior studies, and (4) justifying the need for further research. The primary focus is on the first objective, which seeks to position this research within the current trends of bankruptcy prediction studies. A systematic review of the literature, focusing on topics, theories, and methodologies within the field of bankruptcy prediction, offers a structured framework for situating this study within the larger context of bankruptcy research (Tranfield et al., 2003). Additionally, an examination of the evolution of research trends over time is crucial for understanding the current direction and identifying potential areas for future exploration.

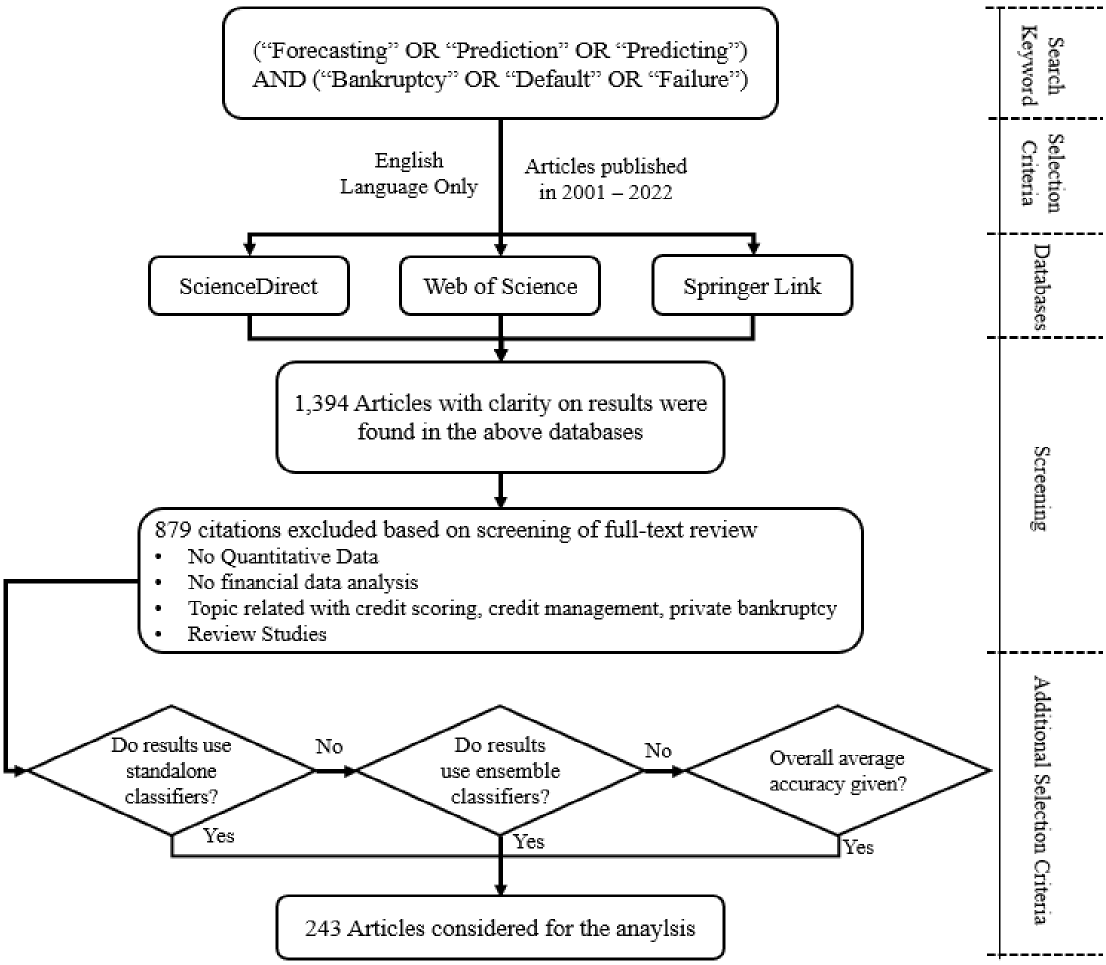
Alaka et al. (2018) analysed 49 journal papers in bankruptcy prediction during the period of 2010 to 2015 using systematic reviews. They constructed framework for selecting appropriate prediction model with identified 13 criteria within the research of bankruptcy prediction. In addition, Veganzones and Severin (2020) analysed 106 published papers in corporate bankruptcy prediction during the period between 2000 and 2017 using systematic reviews. They illustrated how to design corporate bankruptcy prediction model with the elements including sample data, prediction models, variables and evaluation metrics. However, they do not clearly discuss how the focus of their research has evolved in response to trends in the development of sophisticated tools such as hybrid and ensemble models. Moreover, even though there is not overall bankruptcy prediction model, little has been discussed the characteristics or differences in the industries used as the target data for model development. Consequently, this section employs a systematic literature review approach to analyse how bankruptcy prediction has been explored from the 2000s to the 2020s. This review will discuss the research on bankruptcy prediction in terms of thematic focus, disciplinary perspectives, explanatory variables, and methodologies, including research models, analysis methods, and interpretive techniques, highlighting technological advancements in the field. Afterwards, the bankruptcy prediction research focusing on the shipping industry would be analysed to illustrate specific industrial characteristics for improving predictive ability.

For the literature review presented in this section, a systematic procedure was employed, as illustrated in Figure 3.1. This process exclusively included papers published in academic journals due to their rigorous peer-review process, ensuring their suitability for investigating bankruptcy prediction research from theoretical and methodological viewpoints. Conference papers, contributions to edited books, dissertations, and these were excluded to maintain the quality and reliability of the reviewed literature. The target period for the review spanned the last two decades, specifically from 2001 to 2022, capturing the most relevant and contemporary insights into bankruptcy prediction.

To compile the literature, searches were conducted using bibliographic databases such as ISI Web of Science and ScienceDirect, focusing on publications from leading publishers like Springer and Elsevier. The search strategy was designed around a set of keywords to ensure comprehensiveness and relevance: (“Forecasting” OR “Prediction” OR “Predicting”) AND (“Bankruptcy” OR “Default” OR “Failure”). This approach aimed to capture a broad spectrum of studies pertinent to bankruptcy prediction while filtering out unrelated research topics, such as credit scoring, credit management, or personal bankruptcy cases.

Further refinement involved excluding papers that did not present experimental research or adopt a quantitative approach to corporate bankruptcy prediction. Additionally, the tables of contents of journals frequently publishing bankruptcy prediction studies were meticulously reviewed. As presented in Table 3.1, this process resulted in the identification of 243 papers across 140 journals, encompassing a wide range of disciplines including finance, accounting, economics, computer science, business, and mathematics. These selected papers provide a rich foundation for exploring the evolution, trends, and gaps in bankruptcy prediction research.

Figure 3.1 Process flowchart of the systematic literature review



Source: Author

Table 3.1 Number of papers by journal and decade

Journal Category	No. of Publications				
	2001-2005	2006-2010	2011-2015	2016-2022	Total
Accounting	0	1	0	2	3
Business	3	1	2	9	15
Computer Science	8	18	26	58	110
Economics	0	1	1	23	25
Finance	20	8	6	37	71
Mathematics	0	0	0	2	2
Shipping	1	4	2	6	13
Others	0	1	0	3	4
Total	32	34	37	140	243

***Source:* Author**

3.2. Descriptive analysis

The landscape of bankruptcy prediction research has evolved considerably over the past two decades, marked by a significant uptick in publication volume and diversification in the disciplines involved. This progression is highlighted by the substantial increase in the number of publications, peaking at 38 in 2022 from a mere single publication in 2000, as detailed in Figure 3.2. This trend indicates not just a growing interest in the field but also its critical relevance during economic disturbances, evidenced by spikes in publication rates coinciding with the 2008 financial crisis, the 2014 Russian financial crisis, and the 2020 Covid-19 pandemic. Particularly noteworthy is the research dedicated to the shipping industry, a sector that has faced unique financial challenges over the years. The orange bars in Figure 3.2 represent publications focused on shipping-specific bankruptcy prediction. While the overall number of shipping-related studies remains relatively small compared to the total number of publications, there has been a noticeable increase in recent years. This indicates a growing recognition of the distinct financial dynamics and risk factors inherent in the shipping industry.

The diversification of the field is further underscored by the increase in the number of journals publishing bankruptcy prediction research, expanding from nine in the early

2000s to 89 by 2016-2022, as shown in Table 3.2. This expansion signifies the field's broadening interdisciplinary approach and the incorporation of a wider array of methodologies and perspectives. The proportion of multi-authored papers grew from 78% in the 2000s to 88% in the 2010s, while the average number of papers per author increased, indicating a collective movement towards addressing the complexities of bankruptcy prediction.

Figure 3.2 The number of papers in bankruptcy prediction by year

Source: **Author**

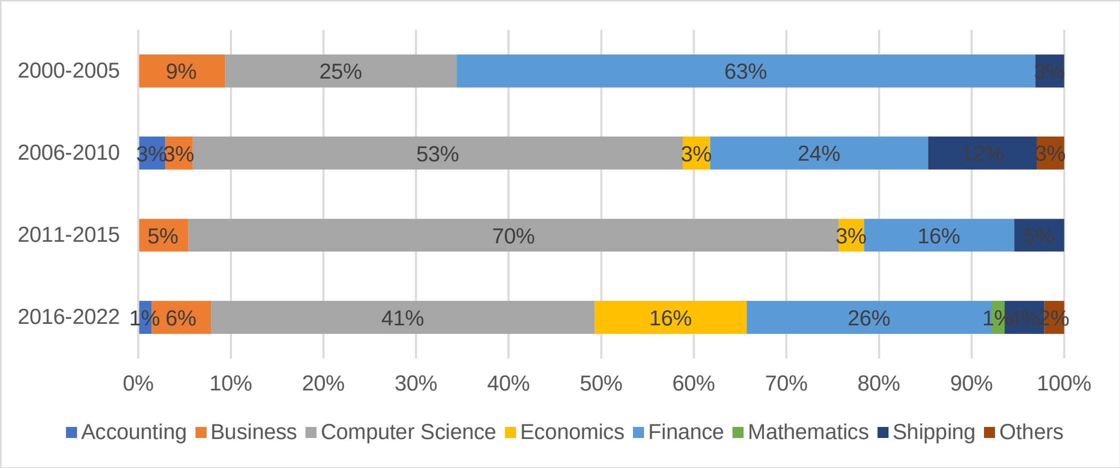
Table 3.2 Descriptive statistics of the data base by period

	2001-2005	2006-2010	2011-2015	2016-2022	Total
No. of journals	9	19	23	89	140
No. of authors	30	32	37	130	229
No. of papers	32	34	37	140	243
(Single-authored)	7 (22%)	5 (15%)	7 (19%)	17 (12%)	36 (15%)
(Multi-authored)	25 (78%)	29 (85%)	30 (81%)	123 (88%)	207 (85%)
Average No. of paper per author	1.07	1.06	1.00	1.08	1.06

Source: **Author**

Furthermore, the disciplinary focus within bankruptcy prediction research has broadened significantly. As presented in Figure 3.3, initially dominated by finance, which accounted for over 63% of studies, the field has seen growing contributions from computer science, particularly evident between 2011 and 2015, when it constituted 70% of research themes. Economic and business disciplines have also made substantial contributions, further illustrating the field's move towards integrating a wider range of theoretical and methodological approaches.

Figure 3.3 Change of proportion of bankruptcy prediction research themes



Source: **Author**

This interdisciplinary expansion is mirrored in the shift in research objectives as shown in Table 3.3. While approximately 70% of studies have concentrated on model development, closely related to information systems and operational research, there has been a notable increase in studies focusing on factor exploration. These studies, which now constitute around 30% of the research, focus on analysing elements of bankruptcy prediction models, such as risk factors and variable selection. This trend towards factor analysis highlights a growing recognition of the importance of dataset characteristics and variable selection in improving model performance, moving beyond the complexity of prediction methods.

Table 3.3. Research objective of the analysed publications

	No. of publications				
	2001-2005	2006-2010	2011-2015	2016-2022	Total
Model Development	27	25	23	84	159
Factor Exploration	5	9	12	53	79
Other	0	0	2	3	5
Total	32	34	37	140	243

Source: **Author**

In summary, the last two decades have witnessed a significant evolution in bankruptcy prediction research, characterized by an increase in publication volume, a diversification of publishing platforms and disciplinary focus, and a shift towards collaborative and qualitative research efforts. This evolution reflects the field's dynamic response to economic conditions and its progression towards a more interdisciplinary and comprehensive understanding of bankruptcy prediction.

3.1. Corporate Bankruptcy

Bankruptcy refers a situation in which a business cannot meet its debt obligations, prompting a petition for debt reorganization or asset liquidation in a federal district court (Schwarcz, 1999). Signs indicating a company's impending bankruptcy can often be observed well before bankruptcy occurs, thus the financial health of a company is evaluated through various indicators. Financial ratios commonly used for this assessment, which are unaffected by company size, facilitate comparisons across businesses of different scales (Altman, 1968). Bankruptcy not only results in severe macroeconomic repercussions, affecting societal well-being, but also causes significant microeconomic distress for the stakeholders of the firms involved (Tinoco and Wilson, 2013). Consequently, numerous efforts have been made to predict corporate bankruptcy. The ability to accurately forecast a decline in business activity leading to bankruptcy provides an opportunity for managers and creditors to implement remedial measures.

Following the empirical study by Altman (1968), bankruptcy prediction has been framed as a binary classification task distinguishing between bankrupt and solvent firms. Establishing a criterion for selecting the most suitable experimental samples for model development is crucial by defining corporate bankruptcy. Despite the lack of a universally accepted definition, corporate bankruptcy has been examined from multiple angles, including legal, economic, financial, and econometric perspectives, each offering a unique perspective on bankruptcy (Platt et al., 1994). These different perspectives contribute to a variety of definitions that can help distinguish between bankruptcy and non-bankruptcy companies. Moreover, corporate bankruptcy can be dissected from several dimensions, where various factors may influence a company's bankruptcy risk, particularly when focusing on specific data sets or industries (Cielen et al., 2004). By exploring the reasons behind bankruptcy, it becomes possible to identify explanatory variables that could enhance the predictive model's accuracy.

This section aims to review the literature on bankruptcy definitions, providing a foundation for selecting target companies within the research framework. Additionally, the discussion will cover the causes of bankruptcy, aiming to pinpoint risk factors that could serve as explanatory variables.

3.1.1. Definition of corporate bankruptcy

The definition of corporate bankruptcy has elicited considerable debate within the academic community, with research across diverse fields presenting varied interpretations. These definitions vary from broad conceptions, such as deviations from expected outcomes, cessation of operations to mitigate further losses, and managerial failure, to more specific terms like formal bankruptcy proceedings. As shown in Table 3.4, the definition of corporate failure across studies encompasses a broad spectrum, including but not limited to: loss to borrowers and guarantee recipients (Edmister, 1972), winding up by court order (McNamara et al., 1988, Sfakianakis, 2012); liquidation or cessation of trading (Keasey and Watson, 2019); prolonged non-compliance with banking obligations (Pindado et al., 2008, Dakovic et al., 2010); liquidation (Peel and Peel, 1987, Slotemaker, 2008); loan default (Grunert et al., 2005, Li et al., 2022), financial distress (Altman and Hotchkiss, 1993, Pindado et al., 2008), bankruptcy or default (Mitchell and Van Roy, 2007, Falkenstein et al., 2000), cash shortages (Mramor and Valentincic, 2003, Cultrera and Brédart, 2016) ; and a combination of bankruptcy, receivership, liquidation, inactivity, and special treatment (Jones and Wang, 2019, Altman et al., 2020).

Critiques have been levelled at the arbitrary nature of these definitions, pointing out the lack of a clear standard for determining corporate bankruptcy (Balcaen and Ooghe, 2006, Ooghe and De Prijcker, 2008). For instance, Keasey and Watson (2019) do not differentiate between business closure and bankruptcy, noting that small enterprises may cease trading for various reasons beyond insolvency, such as to avoid further losses or to capitalize on profits. Similarly, Falkenstein et al. (2000) argued that cash shortages may be temporary and not necessarily indicative of corporate bankruptcy. Jones and Wang (2019) point out the interchangeable use of terms such as failure, insolvency, and default, which may have different implications in the context of various national bankruptcy law. This diversity in definitions reflects the significant influence of legal frameworks on the conceptualization of corporate bankruptcy. Grunert et al. (2005) emphasized the role of a country’s legal system in defining bankruptcy, highlighting that these definitions can vary substantially across countries. For example, in the United States, the Securities Exchange Commission (SEC) identifies bankruptcy under two scenarios in the bankruptcy code: Chapter 11, which involves reorganization and requires court approval, and Chapter 7, which entails ceasing operations and liquidating the business (Hillegeist et al., 2004). In contrast, the UK insolvency law prescribes four different procedures in case of insolvency: Administration, Voluntary Arrangements, Receivership, and Liquidation (Tyagi, 2018).

The first three aim at rescuing the firm through restructuring or financial support, while the fourth involves selling the firm's assets (Mramor and Valentincic, 2003). This legal perspective suggests that "insolvency" is the technically correct term for companies unable to repay their debts.

Table 3.4. The definition of corporate bankruptcy

Division	Definition of Corporate Bankruptcy	Process of Firm Bankruptcy	Source
Economic Bankruptcy	- Total expenses exceeding total revenue. - Return on investment falling below the cost of capital. - Actual returns not meeting expected returns.	Deterioration of profitability	Peel and Peel (1987), Altman and Hotchkiss (1993)
Financial Bankruptcy	- Technical insolvency, characterized by the inability to repay debts at maturity because of insufficient corporate liquidity. - Substantive insolvency, i.e., capital impairment	Declined ability to pay	Grunert et al. (2005), Li et al. (2022), Mramor and Valentincic (2003)
Legal Bankruptcy[7]	- Chapter 7: Suspension of business activities as a result of substantive insolvency - Chapter 11: The court approved reorganization plan	liquidation	McNamara et al. (1988), Sfakianakis (2012)

Source: **Author**

From a financial standpoint, bankruptcy is recognized when a corporation is unable to meet its debt obligations upon maturity, due to either insufficient liquidity or significant capital impairment. The Basel Committee on Banking Supervision (2006) embraces an expansive interpretation of bankruptcy, considering a bankruptcy to have occurred if the bank believes the obligor will not fulfil its credit obligations in full without resorting to measures like asset liquidation, or if the obligor is over 90 days overdue on any credit obligation. In addition, Bureau van Dyke's ORBIS database defines bankruptcy with six

[7] Title 11 of the United States Code, also known as the United States Bankruptcy Code

classifications; (i) Default of payment; (ii) Undergoing insolvency proceedings, (iii) Bankruptcy proceeding, (iv) Dissolution via bankruptcy, (v) Liquidation, (vi) Inactive.

The variation in bankruptcy definitions across different studies poses challenges for interpreting and generalizing empirical results, highlighting the importance of adopting a uniform definition of corporate bankruptcy. Historical analyses often rely on the legal framework, viewing failed firms as those engaged in formal court proceedings, such as under Chapters 7 or 11 of the U.S. bankruptcy code. This legal perspective is aimed at aiding financially distressed firms through reorganization or credit support, thus identifying companies at significant risk of bankruptcy, including those that might successfully reorganize and survive (Slotemaker, 2008). Early stages of bankruptcy can lead to operational failures, triggering liquidity and solvency crises that could culminate in bankruptcy (Stef and Jabeur, 2018).

Consequently, this study adopts a definition of bankruptcy that encompasses firms which have filed for bankruptcy protection or have been reorganized under the Corporate Reorganization Act of Korea. This approach aligns with previous Korean studies that have adhered to the national bankruptcy law framework, ensuring consistency and relevance in the context of Korean corporate law and practices (Kwak et al., 2012, Choi et al., 2018b, Park et al., 2021b).

3.1.2. Causes of corporate bankruptcy

Bankruptcy occurs when a company is unable to service its debt obligations, with liabilities exceeding the current value of its assets and resulting in negative equity (Ohlson, 1980). Legally, a business is considered to have failed when it ceases operations or declares bankruptcy, allowing creditors to initiate legal action against the entity (Bracke et al., 2019). Bankruptcy declaration is not contingent solely on assets being less than liabilities but becomes apparent when a company stops meeting its debt and other financial obligations. Thus, identifying early signs of financial distress is crucial for pre-empting and managing potential bankruptcy risks. Research has identified a variety of factors that contribute to corporate bankruptcy risk, as summarized in Table 3.5.

Table 3.5. List of Causes of bankruptcy

Causes	Description	Source
Internal factors		
Financial distress (Lack of Equity)	· Insufficient equity for ongoing operations. · Lack of alternative financing due to an inadequate equity base. · Financing challenges leading to liquidity issues.	Bradley and Cowdery (2004); Novak and Sajter (2007); Aguiar-Díaz and Ruiz-Mallorquí (2015)
High-cost pressure	· Elevated operational costs, including staff or other expenses. · Increased variable costs from costly materials or subcontractors. · Inconsistent workloads, complicating fixed cost coverage.	Bradley and Cowdery (2004); Novak and Sajter (2007); Lukason and Hoffman (2014)
Poor Management issues	· Deficiencies in accounting or financial planning. · Inadequate management of accounts receivable. · Operational challenges in management and administration.	Korol (2017); Kücher et al. (2020); Mayr et al. (2021)
Operative problems	· Outdated technology or production processes. · High costs in production or operation. · Issues with suppliers, including supplier loss.	Korol (2017); Kücher et al. (2020); Mayr et al. (2021)
External Factors		
Competition	· Intensified industry competition from new entrants. · Price competition to secure additional market share. · Shifts in the competitive landscape within sectors.	Ooghe and De Prijcker (2008); Song et al. (2019); Liu et al. (2020)
General economic situation	· Deterioration of overall economic conditions. · Decline in consumer spending power. · Lack of investment incentives.	Kin et al. (2021); Lukáč et al. (2022)
Other issues	· Nature disasters · Fraud · Termination of contraction with customers	Ooghe and De Prijcker (2008); Grunert et al. (2005)

Source: **Author**

Studies have predominantly examined bankruptcy from an internal perspective, identifying various factors that contribute to corporate bankruptcy. Aguiar-Díaz and Ruiz-Mallorquí (2015) highlights internal factors like insufficient equity, demand volatility, cash shortages, and restricted financing access. Bradley and Cowdery (2004) point to liquidity issues from inadequate equity, poor financial planning, high operating costs, and management deficiencies as key internal risk factors. Conversely, Novak and Sajter (2007) emphasize external macroeconomic influences, including economic downturns, inflation rates, exchange rates, and legal regulations, as significant contributors to bankruptcy.

In addition to financial metrics, qualitative indicators can signal a corporate bankruptcy. Ooghe and De Prijcker (2008) suggested several qualitative factors, such as inexperienced management, risk-seeking behaviour, reluctance to adapt business models, and lack of commitment, as precursors to bankruptcy. Korol (2017) observed that management often overlooks early warning signs of bankruptcy. A. Kücher et al. (2020) noted the significance of a company's age and the application of financial controls in mitigating bankruptcy risk.

The presence of certain risk factors can lead to the development and escalation of crises in corporate management. It is broadly understood that the actualization of qualitative causes behind a company's distress is evident in worsening financial indicators over subsequent periods (Alaka et al., 2018). Research highlights the critical need to evaluate not just internal risk factors, such as a firm's financial health, but also external influences like macroeconomic conditions when analysing corporate bankruptcy. Accordingly, this study will identify quantifiable causes of bankruptcy to include them as explanatory variables. This method aims to provide a comprehensive assessment of bankruptcy risk by considering a wide array of factors, both within and outside the company, that impact its financial condition.

3.2. Variables of corporate bankruptcy prediction

3.2.1. Financial variables

The selection of financial ratios for analysis is grounded in three key considerations (Altman, 2018). Firstly, financial ratios are based on the solid theoretical principles of financial analysis and offer uniformity, making them easily comparable across different groups. This comparability stems from their standardized calculation methods within established regulatory frameworks. Secondly, compared to many qualitative factors that describe a company's organizational structure, strategy, product offerings, or market presence, financial ratios offer greater reliability. These ratios provide a more objective measure of a company's financial health than variables found in qualitative descriptions. Thirdly, financial ratios can be applied universally to all companies, unlike certain metrics, such as those related to financial markets, which are only applicable to publicly traded companies.

Historically, the use of financial ratios in forecasting corporate failure is well-established, with seminal works by Beaver (1968), Altman (1968), Zmijewski (1984). Beaver et al. (2005) highlighted the enduring effectiveness of accounting ratios over four decades, from 1962 to 2002, in signalling corporate bankruptcy. Key financial indicators, including those related to working capital, cash flow, profitability, and debt leverage, have consistently been identified as significant predictors of bankruptcy risk across numerous studies.

Beyond traditional financial metrics, the structure of a company's debt, including the volume and quality of its obligations, as well as its capacity to fulfil these liabilities on time, has been recognized as a valuable predictor of bankruptcy (Philosophov et al., 2008). These indicators have undergone extensive testing in previous research, confirming that changes in these financial ratios and metrics are closely linked to a company's financial health. The financial dimensions reflected by these ratios were specifically chosen for their direct relevance to bankruptcy prediction, illustrating the fundamental aspects that influence a company's likelihood of facing corporate bankruptcy.

Bankruptcy prediction models, fundamentally grounded in liquidity and solvency metrics, reveal that insolvent firms typically exhibit a lower equity book value relative to total assets, Return on Assets, and Cash to Current Liabilities ratio, alongside elevated Current Liabilities to Total Assets ratios (Dakovic et al., 2010). Beaver (1968) identified thirty ratios as critical in forecasting corporate bankruptcy, with an empirical analysis showing ratios including Cash Flow to Total Debt, Net Income to Total Assets, and Total Debt to Total Assets as exceptionally predictive, yielding over an 80% accuracy rate for one-year forecasts. Altman (1968) developed a model using five selected ratios and a multivariate discriminant analysis, achieving a 90% accuracy rate in classifying firms one year before bankruptcy. These ratios including Working Capital to Total Assets, Retained Earnings to Total Assets, EBIT to Total Assets, Market Value of Equity to Book Value of Total Debt, and Sales to Total Assets categorized into leverage, liquidity, profitability, efficiency, and market valuation groups, have demonstrated their predictive validity.

Incorporating these five financial ratios enhances the model's informational quality beyond individual analysis (Altman and Hotchkiss, 1993). The predictive function developed from these findings enables identifying bankruptcy risk up to three years in advance, offering company directors or boards ample time to implement preventative or corrective strategies. Following this approach, the categorization of financial ratios employed in bankruptcy prediction models is derived from their ability to assess various dimensions of a company's financial health: leverage, liquidity, profitability, and efficiency.

First of al. The leverage of a company is a critical indicator of its financial structure and its capability to fulfil long-term debt commitments while maintaining solvency (Liang et al., 2016). Leverage, as a financial metric, is instrumental in assessing the risk associated with a firm's capital structure, particularly its reliance on debt financing compared to equity (Altman and Sabato, 2007). High leverage levels are a concern because they indicate a greater dependency on borrowed funds, which, in turn, increases the firm's risk of insolvency and bankruptcy. Legal frameworks provide that a firm unable to repay its debts may be compelled to file for bankruptcy, highlighting the importance of managing leverage to avoid financial distress (Jaffee and Stiglitz, 1990). Adequate equity capital is essential not just for the borrowing entity but also for the lending institutions. It serves as a financial buffer, absorbing fluctuations in earnings and asset values, thus reducing the risk of bankruptcy (Alexandridis et al., 2020). The relationship between the amount borrowed and the risk of bankruptcy is direct: the higher the level of debt, the higher the likelihood of financial failure (Berk et al., 2013). The cash flow to debt ratio is another pivotal leverage metric, offering a perspective on a firm's capacity to cover its debt obligations with its annual cash flows. While it's unlikely for a company to meet all its debt obligations within a single year solely from cash flow, this ratio's essence lies in illustrating a firm's financial health and operational efficiency (Crutzen and Van Caillie, 2008). A poor cash flow to debt ratio is a red flag, suggesting potential challenges in sustaining operations without risking bankruptcy. These leverage ratios play pivotal tools for analysing financial leverage of company and its implications for bankruptcy and long-term viability.

Second, liquidity measures a company's financial robustness, focusing on its capacity to fulfil short-term liabilities using readily available cash and liquid assets. This financial metric serves as a bulwark against liquidity crises, allowing firms to address immediate payment demands efficiently (Tirole, 2010). However, maintaining a balance is crucial as excessive liquidity can potentially dilute investment returns, highlighting the need for strategic financial management (Mramor and Valentincic, 2003). A company's liquidity is often measured using the current ratio, which is a basic indicator that compares current assets and current liabilities (Altman, 2018). This ratio assesses a company's capacity to meet its short-term liabilities with its short-term assets, primarily consisting of cash and short-term investments (Beaver et al., 2005). A robust current ratio is indicative of a firm's financial agility, enabling it to liquidate assets swiftly without significant loss, thereby ensuring operational stability (Mossman et al., 1998). Liquidity not only provides a snapshot of a firm's financial health in the short term but also influences its long-term solvency (Behr and Weinblat, 2017). Firms with adequate liquidity are seen as less likely to face bankruptcy, as they possess the necessary resources to fulfil their obligations. The interplay between liquidity and leverage is significant for comprehensive financial assessment, allowing stakeholders to identify potential risks and take corrective actions timely (Liang et al., 2016). In essence, maintaining an optimal level of liquidity is crucial for a company's ongoing viability, enabling it to navigate through financial challenges and sustain its operations effectively.

Thirdly, the category of profitability is essential in evaluating a firm's capacity to convert its operations into profits effectively (Desai et al., 1996). This aspect is crucial for sustaining liquidity and leverage over the long term, underlining a company's economic viability. Altman and Hotchkiss (1993) emphasize that a firm's continued existence depends on the earning power of its assets, suggesting a negative correlation between profitability and the likelihood of bankruptcy. Cheng et al. (2018) explored the impact of investor trading activity on the efficacy of bankruptcy prediction models in U.S. firms, identifying economic profitability, indebtedness levels, and operational efficiency as critical indicators. Profitability ratios, by assessing the efficiency of asset use in generating earnings before interest and taxes, are indicative of a firm's success. High profitability firms are more likely to be deemed financially stable, reducing their bankruptcy risk. Among the ratios, Return on Assets (ROA) is favoured for gauging management's effectiveness in profit generation from assets. A superior ROA signifies the capability to produce greater earnings from smaller investments. Return on Equity (ROE) reflects the profitability achieved with shareholders' equity, and the ratio of Earnings Before Interest and Taxes (EBIT) to total assets measures the overall asset profitability. Altman (2018) emphasized that this profitability indicator is paramount in evaluating a firm's bankruptcy risk, consistently outperforming other metrics. Thus, profitability not only signifies the firm's current financial health but also its potential to maintain stability and growth, serving as a key determinant in the assessment of bankruptcy risk.

The last category, efficiency focuses on a firm's capability to efficiently convert its assets, liabilities, and equity into cash or sales (Berk et al., 2013). This aspect, referred to as turnover efficiency, illustrates the proficiency with which a company employs its assets to generate revenue. Crucial efficiency indicators like asset turnover and equity turnover evaluate the ability of a company's assets and equity to produce sales, respectively (Ooghe and De Prijcker, 2008). Typically, higher turnover ratios are indicative of efficient resource use, contributing to a firm's financial health. However, excessively high turnover ratios might signal a potential risk, suggesting that the company may lack sufficient assets or equity to sustain its sales levels, and may eventually go bankruptcy (Alam et al., 2021b).

Incorporating these categories into bankruptcy prediction models facilitates a thorough examination of a firm's financial health. By assessing leverage, liquidity, profitability, and efficiency, in addition to wider market and growth factors, stakeholders can gain a more comprehensive and nuanced understanding of a company's risk profile. This multifaceted approach not only aids in identifying potential financial distress but also provides a framework for strategic interventions to enhance financial stability and mitigate the risk of bankruptcy.

3.2.2. Market variables

Market variables related to a company's stock price, including market capitalization, market-to-book ratios, excess stock returns, and price volatility, are pivotal in evaluating bankruptcy risk. These indicators offer insights into a company's financial health that extend beyond the scope of traditional accounting data, showcasing their critical role in bankruptcy prediction (Jones et al., 2017). Market prices react more quickly to shifts in a firm's financial status than do financial statements, capturing subtle changes that accounting metrics may overlook (Reisz and Perlich, 2007).

Unlike financial statements, which are inherently historical, market prices provide a contemporary view of a company's financial condition, reflecting investor perceptions and future expectations (Agarwal and Taffler, 2008). The lag in financial reporting and the potential for earnings management through various techniques highlight the limitations of relying solely on financial statement data for predicting bankruptcy (Grammenos and Papapostolou, 2012, Kamal et al., 2021, Lombardo et al., 2022). Traditional accounting models, such as Altman's Z-score, fail to consider asset volatility, a significant factor in assessing bankruptcy risk (Hillegeist et al., 2004).

Empirical research comparing the effectiveness of accounting-based ratios and market-price variables in predicting bankruptcy shows mixed results. Some studies find market-price variables to be more predictive, while others advocate for models that combine both types of data for more accurate forecasts (Hillegeist et al., 2004, Shumway, 2001). Beaver et al. (2005) argue that market-price indicators, influenced by financial statement data, should complement rather than replace accounting information, enhancing the predictive power of bankruptcy models.

However, market-price indicators have limitations, particularly in efficiently reflecting all publicly available information. This issue may be more pronounced for smaller firms, which tend to be less monitored by market analysts and institutional investors, potentially affecting the reliability of market prices as predictors of corporate bankruptcy (Tinoco and Wilson, 2013, Lee et al., 2020). Given these considerations, this study will exclude market variables from the analysis. Despite their theoretical relevance, the mixed empirical evidence and specific challenges related to smaller firms suggest that market-price indicators may not consistently offer additional predictive value beyond that provided by accounting and other financial data (Agarwal and Taffler, 2008, Das et al., 2009)

3.2.3. Non-financial variables

The aim of this research is to explore the corporate bankruptcy prediction research over time and identify the key factors contributing to business closures due to bankruptcy. The analysis of bankruptcy risk has conventionally depended on the examination of financial data, observing its temporal changes and making comparative analyses with industry benchmarks. Within this analytical framework, the significance of ratios related to liquidity, leverage, and profitability is particularly emphasized.

Recent research, however, has begun to underscore the importance of considering the size and age of firms as critical determinants of bankruptcy risk. Empirical evidence suggests that smaller or younger firms are more prone to bankruptcy filings compared to their larger or older counterparts (Matin et al., 2018). This is corroborated by findings from Pompe and Bilderbeek (2005), who observed that predicting bankruptcy among young firms is notably more challenging than among well-established firms. Makrominas (2018) also highlighted the role of a company's age as a significant determinant of failure for SMEs within the UK, with a similar correlation between inadequate equity and company age being noted as a central factor in company bankruptcies in Austria.

Exploring the concept of a company's life cycle, Gordini (2014) suggests that the causes of firm bankruptcy are not singular but are influenced by a blend of factors related to size and age as companies navigate through their life cycles. Ooghe and De Prijcker (2008) highlighted the unique challenges startups encounter, including insufficient management or industry-specific expertise, inadequate control systems, and operational inefficiencies. Numerous new ventures find it challenging to provide high-quality products or services and to secure the initial capital required to navigate early-stage obstacles.

Furthermore, the size of a corporation has emerged as an important factor in determining company bankruptcy. Large firms, with their diversified operations, tend to exhibit a lower failure rate than smaller entities, suggesting a negative correlation between size and corporate bankruptcy (Kristóf and Virág, 2020). Large enterprises typically benefit from more effective management and organizational structures, including robust accounting information systems, making them less vulnerable to typical business risks or economic downturns. Conversely, smaller companies frequently face hurdles in obtaining extra equity or external financing during difficult periods. Reynolds and Francis (2000) propose using the logarithm of total assets as a proxy for a firm's size, highlighting that larger firms, with their ample resources, are better positioned to navigate financial difficulties. The financial ratios of SMEs can be particularly unstable over time, underscoring the relevance of incorporating non-financial variables into bankruptcy prediction models (Cultrera and Brédart, 2016, Tobback et al., 2017, Luo et al., 2020, Lee et al., 2020). However, the cross-country applicability of non-financial variables is limited due to issues of comparability and data availability, often restricting such analyses to single-country studies.

Additionally, shifts in the macroeconomic environment between the model estimation period and its application can significantly affect the model's predictive capabilities. Integrating macroeconomic indicators with traditional bankruptcy prediction models has been shown to enhance the reliability of predictions, confirming the direct or indirect impact of market conditions on individual firms' bankruptcy risk (Giriūniene et al., 2019, Olson et al., 2012). Valencia et al. (2019) employed the volatility of the foreign exchange rate to estimate bankruptcy probabilities directly using macroeconomic variables, given the financial crises' potential to induce significant foreign exchange deficiencies. Altman and Sabato (2007) argued for the improved predictive accuracy of models by incorporating non-financial variables such as GDP, firm size, or employee numbers, based on the premise that leading macroeconomic indicators including inflation or interest rates and company characteristics are foundational to bankruptcy prediction.

This comprehensive approach to understanding and predicting bankruptcy seeks to incorporate a broad spectrum of both financial and non-financial variables, including the size and age of firms as well as macroeconomic factors, which enable to develop a prediction model with high predictive accuracy tailored to the specific industries. Through this multifaceted analysis, the study aims to provide a more holistic understanding of bankruptcy risk, highlighting the importance of considering a wide range of factors in accurately predicting corporate bankruptcy.

3.2.4. Trend of variable in Bankruptcy Prediction Models

Entrepreneurs are increasingly focused on preventing business failures, highlighting the critical need for early detection of bankruptcy risks within entities. Within this scope, the most important goal is to increase the prediction accuracy of bankruptcy models, which is a key goal of numerous applications and research efforts (Mi et al., 2020). Researchers have dedicated efforts to refine these models, focusing significantly on identifying the most effective explanatory variables from financial and accounting perspectives.

Table 3.6 offers an insightful summary of financial variables recurrently chosen in bankruptcy prediction studies, underscoring their theoretical relevance and empirical validation across various research works. This aggregation ensures a solid theoretical foundation for their inclusion in predictive models by spotlighting variables with demonstrated explanatory power (Tsai et al., 2014). In Table 3.6, the "count" column indicates the number of times a particular variable has been selected in previous studies, while the "rank" column shows the relative importance or frequency of each variable compared to others. Financial variables are categorized into four main groups: profitability, liquidity, leverage, and efficiency, following suggestion of Altman’s Z-score (Altman, 1968). Each category plays a significant role in assessing different aspects of a firm's financial health and potential risk of bankruptcy.

According to Table 3.6, Return on Assets (ROA) is the most favoured variable, serving as a key indicator of profitability. Similarly, Return on Equity (ROE) and Return on Sales (ROS) are commonly used to assess a company's efficiency in generating profits. In terms of liquidity ratios, the Current Ratio and Working Capital Ratio are prominently featured, reflecting their significance in assessing a firm's short-term financial viability.

For leverage ratios, EBITDA to Assets and Retained Earnings to Total Assets are among the selected metrics, pointing to their utility in measuring a firm's financial structure and debt burden (Cheng et al., 2018). Efficiency ratios such as the Asset Turnover Ratio and Working Capital Turnover Ratio are also highlighted, emphasizing their role in assessing operational effectiveness.

The preference for these variables is largely attributed to their inclusion in or derivation from renowned models like Altman's Z-score model or Ohlson's O-score model. The consistent selection of these variables across numerous studies underscores their proven discriminant ability and their enduring significance in bankruptcy prediction models.

Given this context, the selection of explanatory variables for this study will be informed by the significant predictive ability proved in previous research. This approach serves as the initial phase in the feature selection process, aiming to integrate variables with established theoretical and empirical support for their predictive validity in assessing bankruptcy risk.

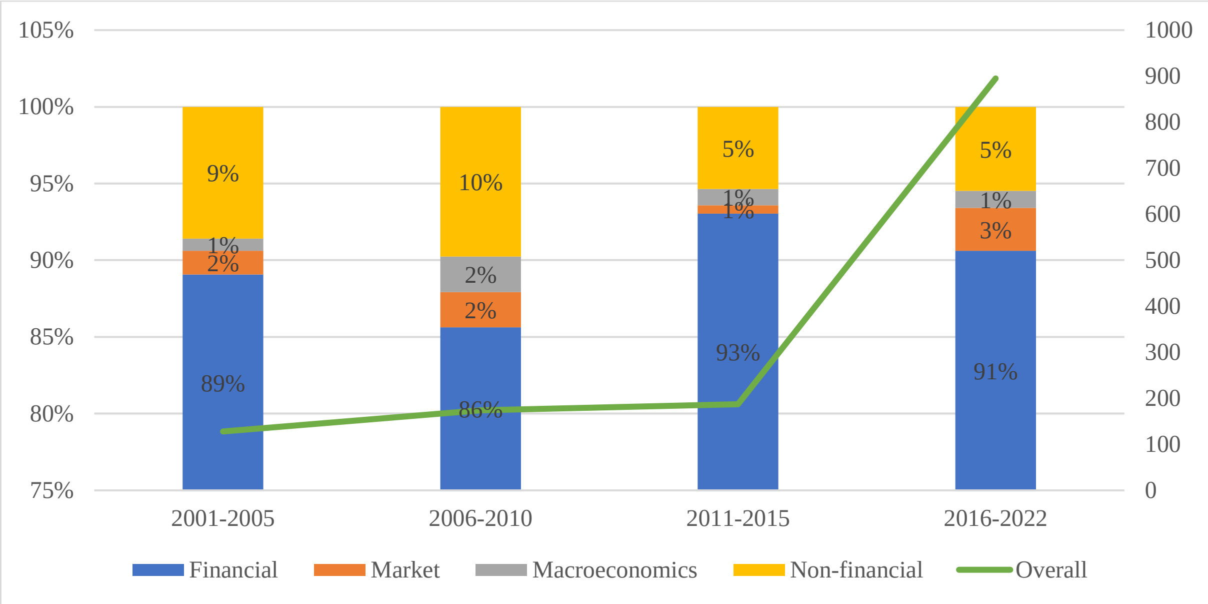
Table 3.6 Summary of Variables Chosen at Least Five Times in Previous Research

Category	Variable	Count	Rank	Category	Variable	Count	Rank
Financial (Leverage)	EBITDA/Total Assets	53	7	Financial (Liquidity)	Current Assets/Current Liabilities	67	2
	Retained Earnings/Total Assets	48	9		Working Capital/Total Assets	64	3
	Total debt/Total assets	41	10		Logarithm of Total Asset (SIZE)	54	6
	Total Liabilities/Total Assets	39	11		Cash/Total Assets	52	8
	Equity/Total Assets	21	18		Cash/Current Liabilities	37	12
	Total debt/Total equity	16	21		Current Assets/Total Assets	32	13
	Total Liabilities/Total Equity	15	22		Current Liabilities/Total Assets	24	15
	Equity/Total Debt	13	25		Quick Ratio	24	15
	Equity/Total Liabilities	13	25		Cash/Total Debt	13	25
	Total liabilities	8	39		Quick Assets/Current Liabilities	11	29
	Current Liabilities/Equity	7	42		Equity	9	32
	EBITDA/Total Liabilities	7	42		Current Liabilities/Total Liabilities	8	39
	EBITDA/Equity	7	42		Current Assets	6	47
	EBITDA	6	47		Cash/Sales	6	47
	Long-term Liabilities/Total Assets	5	57		Current Liabilities/Current Assets	6	47
	Net Income/Total liabilities	5	57		Working Capital/Current Assets	6	47
	Shareholder funds/Total Assets	5	57		Quick Assets/Total Assets	5	57
	Working capital	5	57		Working Capital/Equity	5	57
Financial (Profitability)	Net Income/Total Assets	89	1		Working Capital/Current Liabilities	5	57
	Net Income/Equity	55	5	Financial (Efficiency)	Sales/Total Assets	56	4
	Net Income/Sales	26	14		Working Capital/Sales	20	19
	EBITDA/Sales	23	17		Current Assets/Sales	15	22
	Sales/Current Assets	10	31		Inventory/Total Asset	14	24
	Cash flow/Sales	9	32		Current Liabilities/Sales	9	32
	Sales	9	32		Inventories/Sales	9	32
	EBITDA/Interest expenses	6	47		Sales/Inventories	9	32
	Gross Profit/Total Assets	6	47		Turnover/Total Assets	6	47
	Net interest margin	6	47		Operating Income/Sales	5	57
	Sales/Equity	6	47	Market	Price per share	11	29
	Net Income	5	57		Earnings per Share	7	42
	Operating Income/Total Assets	5	57	Macro-economics (Shipping Index)	GDP	9	32
	Return on sales	5	57		Market Share	8	39
	Return on investment	5	57	Non-financial	Age	17	20
	Sales/Total Liabilities	5	57		N of Employees	13	25

EBITDA: Earnings Before Interest, Taxes, Depreciation and Amortization
Source: **Author**

Figure 3.4 demonstrates the distribution and utilization of variables in bankruptcy prediction studies, categorizing them into four groups: Financial, Market, Macroeconomics, and Non-financial. The analysis reveals a dominant reliance on financial ratios, which constitute about 90% of the variables used, underscoring the continued relevance of accounting-based financial distress indicators. Macroeconomic variables, on the other hand, account for a mere 1% on average, reflecting their limited use in models. Market variables, including stock prices, slightly increased in representation from 2% on average to 3% between 2016 and 2022, indicating a growing but still modest interest.

Figure 3.4 The proportion of explanatory variable used in previous studies



Source: **Author**

Notably, the proportion of non-financial variables, which encompass company characteristics and market indices, was around 10% in the 2000s but has seen a decrease to 5% in recent years. This shift suggests that the relative decrease in non-financial variables is not due to a diminished importance but rather an expansion in the number of financial variables incorporated into models. Despite the reduction in their proportion, the attention to industrial factors and macroeconomic elements remains evident, highlighting their significance in bankruptcy prediction research.

Additionally, the overall number of variables incorporated into bankruptcy prediction models has more than tripled from 2001 to 2022. This expansion reflects a shift in the objectives of model development, from solely focusing on accuracy improvement to a broader aim of identifying bankruptcy risk factors through the application of a larger array of variables. The adoption of advanced prediction techniques, including machine learning and deep learning algorithms, has facilitated this broader analytical approach. These methods enable the integration of a vast spectrum of variables, encompassing financial ratios, market dynamics, macroeconomic indicators, and non-financial factors, thereby enriching the predictive capability and comprehensiveness of bankruptcy prediction models.

3.3. Bankruptcy prediction model

In the dynamic and highly competitive business environment, companies often engage in credit transactions with a range of trading partners, including procurement and credit sales activities. This makes the prediction of trading partners' bankruptcy crucial for managerial decision-making, highlighting the need for robust bankruptcy prediction models to guide investors and managers alike (Sun, 2007). Bankruptcy prediction is pivotal in safeguarding against significant financial losses by assessing the bankruptcy risk of firms prior to investment. (Chen et al., 2019). This area has generated considerable interest, leading to the exploration of numerous methodologies for developing predictive models that leverage financial statement data and market information to forecast a company's financial future as either bankrupt or non-bankrupt.

Drawing from the foundational work of Altman and Hotchkiss (1993), the development of bankruptcy prediction models typically follows a four-step process. This process begins with the analysis of failed and non-failed firms to identify distinctive financial characteristics prior to bankruptcy. The subsequent steps involve reclassifying firms based on these characteristics, validating the model's predictive accuracy on a separate dataset, and finally, employing the model for future bankruptcy forecasts.

Contemporary research predominantly treats bankruptcy prediction as a binary classification challenge, distinguishing between bankrupt and non-bankrupt firms (Li and

Miu, 2010, Pérez-Pons et al., 2022). To address this, a variety of classification methods have been proposed, all aimed at enhancing the predictive accuracy of bankruptcy risk. These methods range from structural models, which apply theoretical functions based on corporate insolvency theories and are often coupled with traditional statistical techniques, to data-driven empirical models, particularly machine learning models evaluated primarily on their predictive performance (Veganzones and Severin, 2020).

Machine learning, defined as a computer program's ability to improve its performance from past experiences, has been especially notable for its application in this domain (Shrivastav and Ramudu, 2020). Furthermore, there has been a rise in hybrid studies that integrate multiple modelling approaches, thereby categorizing these methods into three primary groups: (1) statistical method, (2) machine learning method, and (3) ensemble method. This diversification of methods reflects the field's evolving nature and its quest for more effective bankruptcy (Alaka et al., 2018).

This section explores the development of bankruptcy prediction models and compares various approaches utilized in prior research. Table 3.7 provides a comprehensive summary of these models including descriptions of the models, and recent studies among identified papers from the literature review. This table highlights the diversity of approaches used in the field and underscores the importance of selecting appropriate methods based on specific research objectives and data characteristics. This examination aids in establishing a robust theoretical basis for the prediction model employed in this study. Additionally, it offers guidance for selecting appropriate research models, taking into account the complexity, as well as the specific characteristics of the data and variables involved.

Table 3.7 Summary of Prediction models Identified in Literature Review

Category	Model	Description	Previous Studies
Linear Statistic	Multiple Discriminant Analysis (MDA)	A statistical technique used to classify a set of observations into predefined classes.	Altman (1968); Kliestik et al. (2018); Volkov et al. (2017)
	Logistic Regression (LR)	A statistical method for analysing datasets in which there are one or more independent variables that determine an outcome.	Iturriaga & Sanz (2015); Valencia et al. (2019); Stefko et al. (2020)
Machine Learning	Artificial Neural Networks (ANN)	Computing systems inspired by biological neural networks that learn to perform tasks by considering examples.	Kim et al. (2016); Bragoli et al. (2021); Wang & Liu (2021)
	Support Vector Machines (SVM)	Supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.	Zhou et al. (2014) Barboza et al. (2017); Rustam et al. (2018);
	Decision Tree	A decision support tool that uses a tree-like model of decisions and their possible consequences.	Tsai et al. (2014); Zieba et al. (2016); Pérez-Pons, et al. (2022);
Ensemble	Random Forest	An ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees.	Aliaj et al. (2018); Tang et al. (2019); Gregova et al. (2020)
	Gradient Boosting	A machine learning technique for regression and classification problems, which builds a model in a stage-wise fashion.	Cheng et al. (2018); Tabbakh et al. (2021); Zou et al. (2022)
	Extreme Gradient Boosting (XGBoost)	An optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable.	Carmona et al. (2019); Cao et al. (2020); Shetty et al. (2022)
Deep Learning	Recurrent Neural Networks (RNN)	A class of artificial neural networks where connections between nodes can create a cycle, allowing information to persist.	T.Korol (2020); Jang et al.(2020); Becerra-Vicario et al.(2020)
	Long Short-Term Memory (LSTM)	A type of recurrent neural network capable of learning long-term dependencies, particularly in sequence prediction problems.	Ciaccio & Cialone (2019); Valaskova et al. (2020); Kim et al. (2020)

3.3.1. Linear Statistic Model

Statistical methods have been extensively utilized to develop models for predicting bankruptcy. The field of bankruptcy prediction was pioneered by Beaver (1968) , who developed a binary classification model using a univariate method with financial ratios focusing on profitability, liquidity, and solvency to distinguish between bankruptcy and non-bankruptcy firms. While Beaver's method achieved significant results, it was later critiqued for the correlation among ratios and the conflicting signals that could be emitted by different ratios for the same firm (Dimitras et al., 1996).

Based on Beaver’s work, Altman (1968) introduced an approach that utilized multiple discriminant analysis (MDA) to advance the analysis beyond the univariate method shown in Equation 3.1. He created a corporate scoring model capable of estimating the probability of bankruptcy using a large dataset of both bankrupt and non-bankrupt groups, along with their financial statements. Altman selected specific financial ratios to serve as discriminants of bankruptcy, categorizing them into five key groups: (i) working capital to total assets, (ii) retained earnings to total assets, (iii) earnings before interest and taxes to total assets, (iv) market value of equity to total debt, and (v) sales to total assets. Even today, Altman’s Z-score model continues to be considered as the robust indicator of a corporate bankruptcy (Ayesha et al., 2020)

$$\boldsymbol{Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5} \tag{3.1}$$

X_1 = Working Capital / Total Assets

X_2 = Retained Earnings / Total Assets

X_3 = Earnings Before Interest & Tax / Total Assets

X_4 = Market Capitalization / Total Liabilities

X_5 = Sales / Total Assets

While Multiple Discriminant Analysis (MDA) marked a significant advancement in bankruptcy prediction methodologies, its application is constrained by specific assumptions that may not always mirror real-world scenarios, particularly in the context of financial ratios data (Shi and Li, 2019). Additionally, this model faces challenges such as multicollinearity, the presence of outliers, and non-normality of ratio values, which can violate its statistical assumptions and impact its predictive accuracy (Kücher et al., 2020). In contrast, the logit regression offers practical advantages by circumventing the restrictive assumptions required by MDA, such as multivariate normality and equal covariance matrices, and is adept at handling imbalanced dataset (Chi and Tang, 2006). The logit model, with its ability to incorporate non-linear parameters, is particularly well-suited for bankruptcy prediction as a binary classification problem where the categories are distinct and non-overlapping (Mihalovic, 2016). This model's flexibility allows for the inclusion of only significant variables through stepwise variable selection, improving efficiency.

Ohlson (1980) employed logistic regression to tackle the issue of bankruptcy prediction, incorporating a broad array of variables such as total assets, total liabilities, working capital, current liabilities, current assets, net income, and funds from operations as Equation 3.2. Additionally, he used two dummy variables to evaluate the probability of bankruptcy. The logistic regression method has gained widespread acceptance among researchers and practitioners alike, largely owing to the interpretability of the logistic score. This score can be straightforwardly converted into a statistical probability of bankruptcy, providing a clear and quantifiable measure of risk.

$$\begin{aligned} \boldsymbol{T} = -\mathbf{1.32} - \mathbf{0.407}\log\left(\frac{\boldsymbol{TA_t}}{\boldsymbol{GNP}}\right) + \mathbf{6.03}\frac{\boldsymbol{TL_t}}{\boldsymbol{TA_t}} - \mathbf{1.43}\frac{\boldsymbol{WC_t}}{\boldsymbol{TA_t}} + \mathbf{0.0757}\frac{\boldsymbol{CL_t}}{\boldsymbol{CA_t}} \\ - \mathbf{1.72}\boldsymbol{X} - \mathbf{2.37}\frac{\boldsymbol{NI_t}}{\boldsymbol{TA_t}} - \mathbf{1.83}\frac{\boldsymbol{FFO_t}}{\boldsymbol{TL_t}} + \mathbf{0.285}\boldsymbol{Y} \\ - \mathbf{0.521}\frac{\boldsymbol{NI_t} - \boldsymbol{NI_{t-1}}}{|\boldsymbol{NI_t}| + |\boldsymbol{NI_{t-1}}|} \end{aligned} \tag{3.2}$$

Where,

$TA = Total\ Asset, \qquad GNP = Gross\ National\ Product\ price\ index\ level,$

$TL = Total\ Liabilities\ , \qquad WC = Working\ Capital$

$CL = Current\ Liabilities, \qquad CA = Current\ Assets$

$X = 1\ if\ TL\ exceeds\ TA, \qquad 0\ otherwise,$

$NI = Net\ Income, \qquad FFO = Funds\ from\ operation,$

$Y = 1\ if\ a\ net\ loss\ for\ the\ last\ two\ years, 0\ otherwise$

Traditional statistical techniques like discriminant analysis and logistic regression have been foundational in predicting corporate bankruptcy, relying on the principle of distinguishing between two groups of companies based on data analysis (Zebari et al., 2020). Despite their simplicity and user-friendliness, these methods are limited by their relatively lower accuracy and assumptions of linear separability, variable independence, and multivariate normality. These assumptions frequently do not match the intricate characteristics of financial data. (Ohlson, 1980, Karels and Prakash, 1987). The reliance on fixed functions and preset assumptions further complicates the development of sophisticated financial models, underscoring the need for more adaptable and advanced methodologies in bankruptcy prediction (Gregova et al., 2020).

3.3.2. Machine Learning Model

Since the 1990s, the adoption of AI and data mining techniques has revolutionized the approach to management problems, including the binary classification challenge of bankruptcy prediction. Machine learning techniques have demonstrated a remarkable capacity to explore non-linear relationships between variables, offering an enhancement in prediction accuracy over traditional statistical methods (Park et al., 2021a)

Among these advanced methodologies, the neural network technique has stood out as a powerful tool in bankruptcy prediction, surpassing logistic regression and other computational classifiers like decision trees and genetic algorithms in performance.

Neural networks have gained recognition for their precision in predicting bankruptcy, achieving higher average accuracy than both statistical techniques and other computational classifiers (Kuhn and Johnson, 2013). Despite the significant explanatory power of neural networks, Ahn and Kim (2009) have identified challenges associated with their use, including the need for heuristic parameter setting and the risk of overfitting. Moreover, neural networks may struggle with multimodal data since all econometric metrics typically require normalization or standardization for effective training and error backpropagation. This standardization process, however, does not address the issue of data multimodality, which can significantly impact the predictive performance of neural networks (Zebari et al., 2019). Consequently, there's a growing interest in leveraging ensemble classifiers to mitigate these limitations.

Support Vector Machine (SVM) represents an innovative approach by using a linear model to establish nonlinear class boundaries, achieved by mapping input vectors into a higher-dimensional feature space (Murphy, 2022). Within this space, SVM constructs an optimal separating hyperplane, maximizing the margin between decision classes. As suggested by Min and Lee (2005), SVM became the alternative by merging the theoretical advantages of conventional statistical methods with the empirical strengths of machine learning. The simplicity of SVM has shown promise in various financial applications, including credit ratings, fraud detection in insurance claims, and corporate failure prediction (Rustam et al., 2018, Sanz et al., 2018, Pławiak et al., 2019).

The decision tree model have been widely recognized machine-learning technique due to its structured approach, comprising numerous nodes and branches (Mitchell and Mitchell, 1997). This model is constructed with each internal node representing a test on a specific attribute, and each branch corresponding to the outcome of that test (Bishop and Nasrabadi, 2006). The process begins with the root node, which is selected based on the highest information gain among attributes. Subsequent nodes are determined in a similar manner, with the attribute exhibiting the next highest information gain being chosen for the following node. This sequential selection continues until either all attributes have been evaluated or there are no further attributes that can subdivide the samples (Cho et al., 2010). The decision tree's information-theoretic foundation aims to reduce the number of tests needed for classification, ensuring the construction of a straightforward tree. This model is lauded for its computational efficiency and minimal memory requirements during both the training and prediction phases (Archer and Kimes, 2008). Furthermore, they provide clear interpretability, aiding in the understanding and analysis of how various features contribute to the classification process (Matin et al., 2018).

3.3.3. Ensemble models

In corporate bankruptcy predictions, managing large datasets poses considerable challenges for traditional classification models (Behr and Weinblat, 2017). To address these challenges, recent research has turned to ensemble algorithms, which combine multiple single classifiers to achieve a higher level of predictive performance. Ensemble learning, in particular, has been shown to offer greater accuracy and stability compared to individual base classifiers when applied to large datasets. Broadly, two main types of ensemble techniques have been suggested such as bagging (Breiman, 1996) and boosting (Freund and Schapire, 1997).

Bagging, short for Bootstrap Aggregating, entails creating multiple classifiers from randomly generated training sets by employing the same classification algorithm. Each classifier is trained independently, and their results are aggregated through a majority voting system to arrive at a final prediction (Breiman, 1996). Conversely, boosting uses the same classification algorithm for each classifier but trains them sequentially, not independently, distinguishing it from other methods. With boosting, each successive classifier focuses on correcting errors made by the previous ones, making each new classifier dependent on its predecessors and sensitive to their performance (Freund and Schapire, 1997)

Random forest algorithm, developed by Breiman (2001), extends the bagging technique by creating a model based on a multitude of decision trees. Each tree contributes a vote towards the final classification, with the most common outcome selected as the prediction The random selection of features at each node diminishes tree correlation within the forest, consequently reducing the overall error rate. This method has shown to be particularly effective in the presence of redundant features that contribute to class discrimination (Kim and Kang, 2010, Tsai et al., 2014). Random Forest models have demonstrated superior performance compared other machine learning algorithms and conventional statistical models, efficiently processing large data sets and managing thousands of input variables (Breiman, 2001, Archer and Kimes, 2008). Given the complexity and non-linear relationships between variables in bankruptcy prediction, Random Forest models are likely to surpass classical regression methods (Archer and Kimes, 2008). This model's ability to automatically exclude irrelevant features and its interpretability make it a popular choice in bankruptcy prediction studies.

Introduced by Friedman (2001), the gradient boosting model represents a significant advancement in the field of machine learning, particularly in bankruptcy prediction. This technique enhances predictive accuracy by aggregating the outputs of numerous weak classifiers to form a single strong classifier. Unlike approaches that rely on a limited number of strong predictors, gradient boosting focuses on incrementally training a multitude of weak predictors. Each new predictor corrects errors made by its predecessors, utilizing the full spectrum of predictive variables. This process continues until the model achieves a desired low error rate, with each predictor's coefficient optimized sequentially in a stage-wise manner, thereby enhancing accuracy without compromising the model's resistance to overfitting (Friedman, 2001, Schapire, 2013).

A key feature of gradient boosting involves assessing predictor contributions via the Relative Variable Importance (RVI) metric. This measure ranks predictors based on their weighted classification accuracy, summed up across all predictors in the model. The sum of these weighted improvements is then normalized, assigning the top performer a score of 100 and descending scores to others, indicating their relative contribution to the model's predictive power (Friedman, 2001). Additionally, gradient boosting offers partial dependence plots, aiding in the assessment of how predictors influence the outcome, clarifying the direction and magnitude of their effect (Behr and Weinblat, 2017).

One of the notable strengths of the gradient boosting model lies in its implementation, requiring minimal researcher intervention and data preparation, making it particularly beneficial for corporate failure prediction. Jones and Wang (2019) highlighted its suitability for analysing complex, high-dimensional, and nonlinear data sets, which closely mirror the conditions of businesses at risk of failure. Moreover, gradient boosting exhibits robustness against 'dirty data,' showing resilience to outliers, monotonic transformations, and missing values (Jones and Wang, 2019). Gradient boosting stands out not only for its methodological strengths but also for its practical applications in handling the complexities inherent in bankruptcy prediction, offering a sophisticated tool for navigating the challenges of predictive modelling in this critical domain (Dangeti, 2017).

Extreme Gradient Boosting (XGBoost) represents an advanced version of gradient boosting algorithms, delivering superior performance via a more regularized model formulation designed to prevent overfitting (Chen and Guestrin, 2016). As an efficient and scalable variant of the gradient boosting framework initially proposed by Friedman (2001), XGBoost stands out for its formal control over variable weights, making it a refined approach to regularization in predictive modelling. This method was applied by Carmona et al. (2019) to predict bankruptcy of bank within the Eurozone, demonstrating its utility in bankruptcy risk assessment. While both XGBoost and traditional gradient boosting operate on the foundational principles of gradient boosting, XGBoost differentiates itself with a regularization technique that significantly enhances model performance by effectively managing overfitting (Chen et al., 2015). Its efficiency and capacity to deliver state-of-the-art results have made XGBoost a popular choice among data scientists, evidenced by its prevalence in winning solutions for numerous machine learning competitions on platforms like Kaggle (Chen and Guestrin, 2016).

XGBoost not only provides an efficient framework for integrating diverse predictors, but also improves the interpretability of prediction models by assigning relative importance to individual features. This attribute is instrumental in dissecting the influence of various bankruptcy predictors, including both accounting-based variables and market-price indicators, within a unified statistical model (Zięba et al., 2016). It enables a comprehensive understanding of how different variables interact and contribute to the prediction of bankruptcy, sidelining less significant features (Molnar, 2020).

Using ensemble algorithm variations in bankruptcy prediction offers several benefits. Firstly, these models provide a robust approach for integrating various predictors and evaluating their performance within a unified statistical framework, eliminating the need for the normality assumption. Moreover, these algorithms enhance interpretability by ranking predictors according to their relative importance, aiding in the assessment of different bankruptcy indicators, such as accounting-based variables or market-price signals. In scenarios of high dimensionality, these variables may interact, serving both as competing and complementary information sources for bankruptcy prediction. This approach allows for the exclusion of non-essential features, thereby streamlining the prediction process (Valencia et al., 2019).

3.3.4. Deep Learning algorithms

With the improvements in computational capabilities, Bankruptcy prediction models have progressively embraced data-mining and deep-learning techniques, demonstrating that deep learning models can surpass traditional models in terms of performance (Jang et al., 2021). Deep learning, a field that has gained significant attention in the last decade, is widely applied across various domains of image recognition (Mohammed et al., 2018), voice recognition (Mai et al., 2019) and medical fields (Cha and Kang, 2018). Yet, its application in finance and management science remains relatively underexplored. A significant application of deep learning in finance involves utilizing Recurrent Neural Networks (RNN) for predicting stock price fluctuations, leveraging RNN's suitability for time series analysis (Mai et al., 2019). Deep learning has also been utilized to forecast bankruptcy by analysing textual data alongside traditional numerical data like financial ratios, significantly enhancing prediction accuracy (Cha and Kang, 2018).

Traditional classifiers such as logistic regression, Support Vector Machine (SVM), Artificial Neural Network (ANN), and decision tree tend to perform best with balanced datasets (Zhang et al., 2021). However, they often yield less optimal results with imbalanced data and can struggle with non-linear and complex data patterns (Rustam et al., 2018). Moreover, SVM and NN are noted for their sensitivity to missing values and the challenges they pose in training on large datasets (Alfaro et al., 2008). Deep learning techniques, however, enable effective model design for analysing large volumes of data, addressing classification and prediction in imbalanced datasets (Zhou and Lai, 2017).

However, RNNs face the vanishing gradient problem, which impairs learning ability over long sequences (Kim and Kang, 2019). Long Short-Term Memory (LSTM) networks are developed to overcome this limitation, memorizing information for extended periods and effectively capturing time dependencies in data (Vochozka et al., 2020). LSTM has demonstrated superior predictive accuracy and generalization capability in various research, compared to traditional machine learning models, achieving lower error rates and variability in predictions (Rundo, 2019). Cha and Kang (2018) found that LSTM models exhibit higher predictive accuracy, quicker response times, and enhanced generalization capabilities compared to other machine learning models. Bouktif et al. (2018) showed that LSTM models achieve lower RMSE scores and relative errors, highlighting their proficiency in time series analysis. Similarly, Pai and Ilango (2020) combined random forest for feature selection with LSTM for bankruptcy prediction, demonstrating LSTM's high performance and its ability to overcome the vanishing gradient problem associated with conventional deep learning models.

In summary, the advancement of deep learning, particularly LSTM networks, represents a significant leap forward in the predictive modelling of bankruptcy, offering sophisticated solutions to previously insurmountable challenges associated with traditional machine learning models.

3.3.5. Comparison of Models

Statistical models often face limitations in bankruptcy prediction due to the unique characteristics of bankruptcy data. Logistic regression, for example, utilizes maximum likelihood estimation to identify a concise set of predictors that enhance model fit, neglecting additional variables (Chen, 2011). This optimization of model parameters collectively leads to challenges in handling numerous variables without encountering issues of overfitting and convergence difficulties (Hastie et al., 2009). Given the infrequent occurrence of bankruptcy, research typically utilize multi-year samples to construct their models, yet the attributes of firms are subject to annual variations. However, conventional models are restricted to using a single set of explanatory variables for each firm, compelling researchers to selectively determine the timing for evaluating each firm's characteristics (Mitchell and Van Roy, 2007). Commonly, the focus is on data from the year preceding bankruptcy, inadvertently overlooking information on solvent firms that might eventually face bankruptcy (Zoričák et al. 2020). This arbitrary selection introduces a bias into the analysis, undermining the validity of findings derived from statistical models (Jones et al., 2017).

Moreover, traditional models do not adequately reflect the panel structure of financial statements or account for macroeconomic influences affecting all companies. Shumway (2001) highlighted issues associated with using single-period classification models with multi-period financial statement data, resulting in biased and inconsistent parameter estimates. Another limitation of conventional statistic methods is their assumption of failure process stability, which is incompatible with the reality of fluctuating economic and company conditions (Gregova et al., 2020). The dichotomous classification does not account for the timing of bankruptcy, treating all firms within each category uniformly without recognizing temporal variations in bankruptcy occurrence (Dangeti, 2017). This requirement for a stable failure process over an extended timeframe is impractical, further diminishing the applicability of traditional statistical models for accurate bankruptcy prediction.

Machine learning models, including statistical technique, are engineered to optimize prediction accuracy (Ribeiro et al., 2016). Those algorithms like decision tree or support vector machine are known for their predictive ability. However, their limited interpretability remains a challenge, as insights into how individual financial ratios influence bankruptcy risk are often obscured, reduced to mere measures of relative importance or marginal effects (Molnar, 2020). This lack of interpretability hinders their broader adoption, especially among managers and regulatory bodies who require clear explanations for a model's selection criteria, particularly for firms identified as at risk of bankruptcy.

The enhancement in bankruptcy prediction accuracy through these advanced techniques suggests that financial ratios' impact on a firm's viability exhibits nonlinear characteristics. Nonetheless, the adoption of such "black-box" models complicates the understanding of the prediction rationales (Schapire, 2013). Consequently, despite their lower predictive accuracy, traditional linear models like discriminant analysis and logistic regression remain preferred in scholarly research for their interpretability. These conventional models, while easier to interpret, often do not fully capture the complex functional relationships between significant financial ratios (Altman et al., 2020). Their effectiveness is heavily contingent upon the correct specification of independent variables by the researchers. A failure to identify and include all pertinent variables significantly diminishes logistic regression's predictive utility (Park et al., 2021a). This limitation indicates the balance between the need for predictive accuracy and the demand for model interpretability in the domain of bankruptcy prediction research (Altman et al., 2020).

Incorporating non-linear relationships into predictive models significantly enhances their generalization capabilities and predictive accuracy, facilitating the discernment of intricate patterns within the data (Sanz et al., 2018). Non-linear models structured as additive functions offer interpretability, as the function associated with each variable elucidates its marginal effect (Ribeiro et al., 2016). However, this clarity diminishes in non-additive models due to conditional effects based on specific predictor values. This presents a model selection challenge, balancing interpretability and predictive accuracy.

Despite of their predictive accuracy, Support Vector Machine (SVM) and Artificial Neural Network (ANN) face challenges with high-dimensional data, limiting their suitability for large bankruptcy datasets (Tseng and Hu, 2010, Chen, 2011). Despite their potential for greater predictive power, decision trees and SVMs suffer from limited interpretability, as they provide only relative importance or marginal effects of financial ratios on bankruptcy risk. Recent advances in ensemble techniques have shown promise in bankruptcy prediction. Zięba et al. (2016) suggested a new approach called composite functions that leverages eXtreme Gradient Boosting (XGB) to integrate econometric measures for bankruptcy prediction. This method, tested on a dataset of Polish companies, confirmed the superior performance of boosting techniques in predicting firm bankruptcy over other methods. The XGB algorithm efficiently addresses financial distress in balanced datasets, demonstrating the evolving landscape of predictive modelling in bankruptcy research.

Identifying the optimal artificial intelligence method for predicting corporate bankruptcy remains a challenge, with no single approach universally outperforming others. As shown in Table 3.8, the effectiveness of prediction models is contingent on various factors, including the nature of the dataset, industrial context, and the specific objectives of classification (Siswoyo et al., 2020). Recent advancements have highlighted the efficacy of ensemble techniques, which combine multiple predictors to enhance predictive performance (Choi et al., 2018a, Yang et al., 2021). This approach is particularly beneficial when the component models generate independent errors, offering a robust solution to corporate bankruptcy prediction (Lin et al., 2019, Sun et al., 2020).

Table 3.8 Advantages and disadvantages of prediction model by type of learning

Type	**Model**	**Advantage**	**Disadvantage**	Source
Conventional statistic	LR, MDA	• Effective with small datasets • Outputs interpreted as probabilities	• Specific data assumptions required • Limited to linear solutions	Altman (1968); Hosmer and Lemeshow (2000)
Machine learning	ANN, SVM, DT	• Offers non-linear solutions • Strong generation capability from small train data	• To achieve good performance, need to optimize its parameters • Risk to overfitting • Sensitive to noisy data and outliners	Haykin (1998); Vapnik and Cortes (1995)
Ensemble	RF, GB, XGB	• Handles categorical features. • Minimal parameters to tune. • Performs well with feature-rich datasets	• Ensemble interpretability can be challenging	Breiman (2001); Friedman (2001)
Deep learning	RNN, LSTM	• Learning from variable length inputs • Learning from long-term dependencies • Robust to data sparsity	• Complex configuration • Limited interpretability	Hochreiter and Schmidhuber (1997)

MDA: Multiple discriminant analysis, LR: Logistic regression, ANN: Artificial neural network, SVM: Support vector machine, DT: Decision tree, RF: Random Forest, GB: Gradient boosting, XGB: Extreme gradient boosting, RNN: Recurrent neural network, LSTM: Long short-term memory

Source: **Summarized by Author**

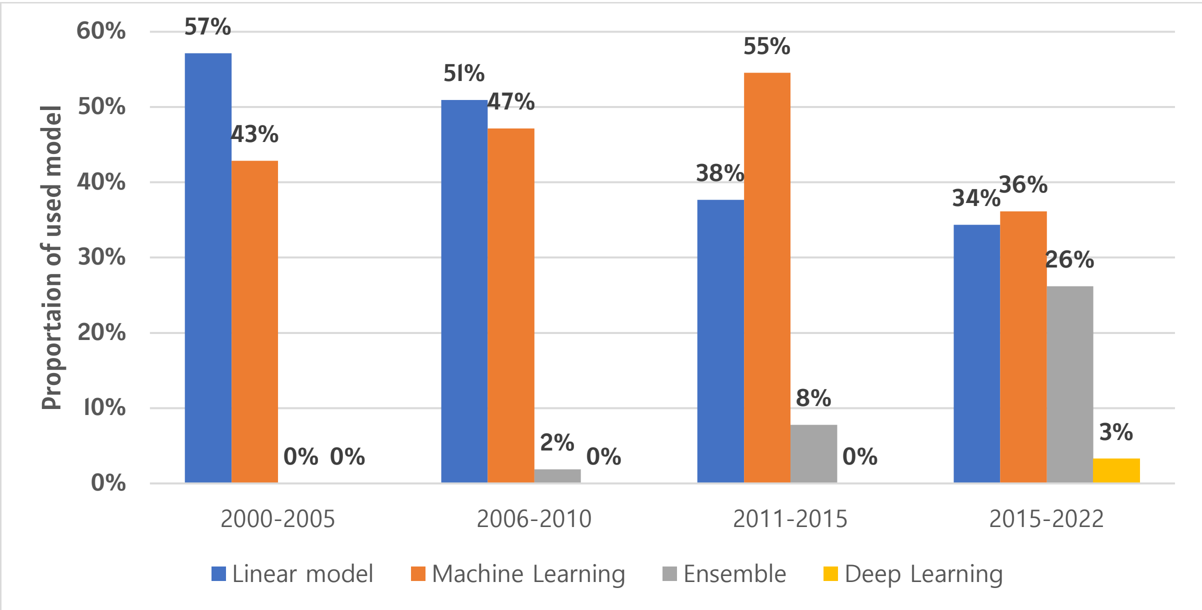
3.3.6. Research trend of bankruptcy prediction model

Historically, the task of predicting corporate bankruptcy was predominantly addressed through statistical methods like logistic regression (LR), linear discriminant analysis (LDA), and multiple discriminant analysis (MDA) (Veganzones and Severin, 2020). These traditional approaches offered a structured way to evaluate financial data against bankruptcy indicators (Tirole, 2010). However, with advancements in technology and the rise of machine learning (ML), the landscape of bankruptcy prediction research has shifted towards more sophisticated analytical methods. Machine learning, particularly its application in pattern recognition, has become an essential role in identifying financial bankruptcy patterns within corporate data, classifying companies into risk groups based on their financial condition (Son et al., 2019).

Machine learning techniques employed in bankruptcy prediction is vast, encompassing algorithms such as Naive Bayes, k-nearest neighbours, neural networks, support vector machine (SVM), random forest and extreme gradient boosting machines (XGB) (Barboza et al., 2017, Lombardo et al., 2022, Shetty et al., 2022). These methods have been complemented by the advent of deep learning approaches like recurrent neural networks (RNN) and long-short term memory (LSTM), which offer comprehensive insights into temporal financial data patterns (Alam et al., 2021a, Jang et al., 2021). Previous research suggests that machine learning models, particularly those based on non-linear algorithms, tend to outperform traditional statistical methods in predicting bankruptcy, thanks to their ability to capture complex patterns in the data (Kumar and Ravi, 2007, Son et al., 2019). Despite their superior predictive performance, machine learning models often face criticism for their lack of interpretability. The so-called 'black box' nature of these models hampers their interpretability, making it challenging to understand the reasoning behind their predictions (Behr and Weinblat, 2017, Giriūniene et al., 2019). This can be a significant drawback when stakeholders require clear explanations for decision-making processes. In contrast, traditional statistical models like MDA or logistic regression, despite being potentially less accurate, offer better transparency. They allow for the evaluation of each variable's weight in the model, providing insights into the determinants of bankruptcy risk (Jones, 2017).

As depicted in Figure 3.5, the evolution of bankruptcy prediction methodologies over the past two decades has been marked by significant shifts in the preference for certain types of models. During the 2000s, conventional statistical models predominated in bankruptcy prediction research, constituting about 60% of the methodologies employed. These models, known for their interpretability, allowed researchers to identify and assess the main risk factors contributing to bankruptcy (Linardatos et al., 2020). In contrast, machine learning techniques accounted for 40% of the approaches used, underscoring a growing interest in applying artificial intelligence to enhance the predictive accuracy of corporate bankruptcy models. This trend intensified from 2011 to 2015, with machine learning models being chosen in 55% of studies, marking a shift towards more sophisticated analytical techniques.

Figure 3.5 The development trend of bankruptcy prediction model



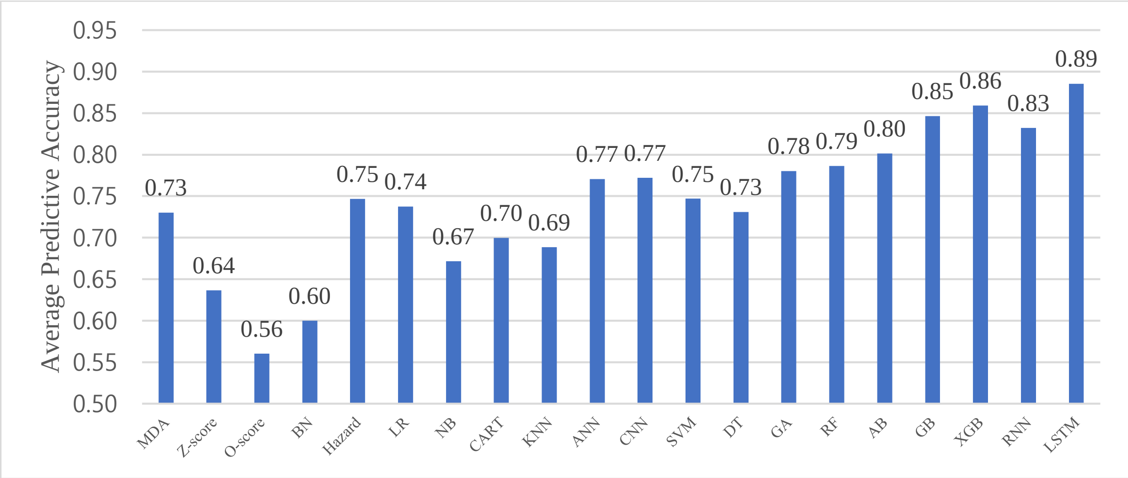
Source: **Author**

Recent years have seen a notable rise in the adoption of ensemble models, which combine multiple learning algorithms, such as random forests or boosting, to achieve better predictive performance (Jones and Wang, 2019). These models have steadily gained traction in bankruptcy prediction research, increasing to 30% usage between 2016 and 2022. Despite the relatively lower predictive ability of conventional linear statistical models, they continue to be employed in over 30% of studies due to their capacity to analyse bankruptcy's risk factors through clear model interpretability (Jones et al., 2017). While the proportion of machine learning techniques slightly declined to 40% from 2016 to 2022, this does not suggest a decrease in the significance of machine learning models. Instead, it reflects a strategic shift towards integrating these models within ensemble frameworks to leverage the strengths of various predictive approaches collectively. Additionally, deep learning algorithms, including recurrent neural networks (RNNs) and long-short term memory (LSTM) models, have started to make their mark, being applied in 3% of bankruptcy prediction studies from 2015 to 2022. This emerging trend highlights the field's continuous evolution towards employing more complex and capable models to predict corporate bankruptcy accurately.

Recent advances in bankruptcy prediction research have led to ongoing research on optimal prediction models, with the goal of increasing prediction accuracy by exploring and comparing different models. This pursuit acknowledges that the most effective model will depend on the characteristics of the target data set, the length of the forecast period, and the set of explanatory variables chosen (Shi and Li, 2019).

Figure 3.6 illustrates the average predictive accuracy of various bankruptcy prediction models based on an analysis of previous research gathered through a Systematic Literature Review (SLR) process. To construct this figure, results from previous studies were aggregated, summing the predictive accuracies reported for each model type. The mean accuracy for each model was then calculated to provide a comparative overview.

Figure 3.6 The average predictive accuracy of models applied in previous research



MDA: Multiple discriminant analysis, BN: Bayesian network, LR: Logistic regression, NB: Naïve bayes, CART: Classification and regression tree, KNN: K-nearest neighbour, ANN: Artificial neural network, CNN: Convolutional neural network, SVM: Support vector machine, DT: Decision tree, GA: Genetic algorithm, RF: Random Forest, AB: Adaboost, GB: Gradient boosting, XGB: Extreme gradient boosting, RNN: Recurrent neural network, LSTM: Long short-term memory

***Source:* Author**

The analysis reveals notable differences in the average accuracy of various prediction models. Simple scoring models, such as Altman's Z-Score or Ohlson's O-Score models, have shown the lowest performance compared to more advanced methodologies. Conventional linear models, including Multi Discriminant Analysis (MDA) and Logistic Regression (LR), also demonstrated relatively lower predictive performance when compared to machine learning models.

However, it is important to highlight that the difference in performance between conventional classifiers and certain sophisticated machine learning techniques, such as Artificial Neural Network (ANN) and Support Vector Machine (SVM), was not markedly significant. This finding suggests that simpler model structures could still be practical, especially when the research goals include statistical inference and model interpretability. In contrast, ensemble techniques and deep learning algorithms, notably Gradient Boosting (GB) and Long Short-Term Memory (LSTM), have shown superior performance, outperforming other models by approximately 10% in terms of predictive ability. This superior performance underscores the potential of these advanced methodologies in accurately predicting bankruptcy, highlighting a shift towards more complex models that can better manage the intricate relationships within financial data, ultimately justifying their advantages in bankruptcy prediction modelling.

3.4. Key Issues in Bankruptcy Prediction Models

3.4.1. Feature Selection Methods

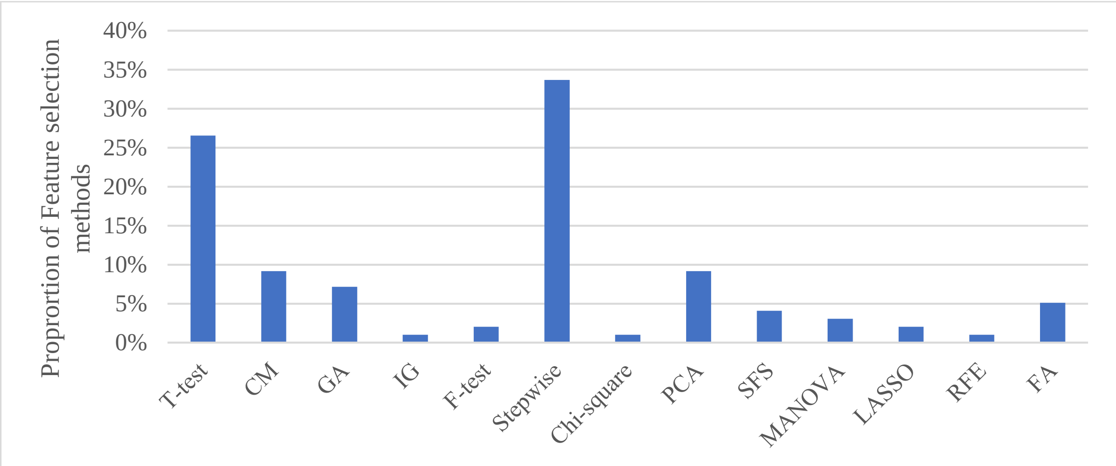
In the domain of bankruptcy prediction, the evolution from traditional statistical methodologies to advanced machine learning techniques signifies a strategic shift aimed at enhancing predictive accuracy. This transition is characterized by the adoption of artificial neural network, decision tree, support vector machine, ensemble techniques, and deep learning models, reflecting a broadened analytical scope that surpasses the confines of simpler statistical methods (Liang et al., 2016, Yuen and Ko, 2018). Despite certain studies showing the superior performance of contemporary methods over traditional ones, a consensus on the optimal approach remains elusive, highlighting the complexities inherent in bankruptcy prediction research.

The incorporation of a wide array of predictive variables has been a focal point of recent research endeavours. Traditionally centered around financial ratios following Altman's Z-score model, the range of variables has been significantly expanded to include governance indicators, macroeconomic variables (Tinoco and Wilson, 2013, Zelenkov and Volodarskiy, 2021), and industry-specific indices (Shome and Verma, 2020, Lukáč et al., 2022). This diversification of variables not only aims to refine the predictive models but also introduces challenges related to model complexity and interpretability (Tsai, 2009).

The process of feature selection has emerged as a pivotal research area within machine learning and data mining, focusing on enhancing model efficacy by identifying the most informative variables (Solorio-Fernández et al., 2020). This effort involves an examination of the dataset to retain variables with substantial predictive power while excluding those deemed irrelevant or redundant. Effective feature selection is instrumental in reducing computational costs, improving model accuracy, and facilitating a deeper understanding of the underlying predictive dynamics (Hira and Gillies, 2015). Among the various techniques employed for feature selection, statistical methods like correlation analysis and optimization approaches, including genetic algorithms, have been extensively utilized (Zelenkov et al., 2017).

To construct Figure 3.7, a comprehensive review of previous research was conducted, gathering data on the usage frequency of various feature selection methods in bankruptcy prediction studies. Each method's usage was counted across the studies, and the proportions were calculated based on the total number of occurrences.

Figure 3.7. Proportion of feature selection methods in previous studies



CM: Component matrix, GA: Genetic algorithm, IG: Information gain, PCA: Principal component analysis, SFS: Sequential forward selection, MANOVA: Multivariate Analysis Variance, RFE: Recursive feature elimination, FA: Feature analysis

***Source:* Author**

As Figure 3.7 illustrates, stepwise logistic regression as wrapper method and t-test as filter method have been most frequently used 33% and 27% respectively from the previous research, because of its efficiency. An insightful comparison of feature selection methodologies reveals the prominence of stepwise logistic regression and t-tests in previous studies, attributed to their efficiency in refining bankruptcy prediction models (Wang et al., 2014). The application of these methods has been shown to significantly improve model performance, emphasizing the critical role of integrating robust feature selection strategies with apt classification techniques to develop optimally predictive models (Tsai, 2009).

In summary, enhancing bankruptcy prediction models requires a strategic combination of advanced machine learning techniques, a comprehensive selection of predictive variables, and the careful application of feature selection methods. While this multifaceted approach aims to improve predictive accuracy, it must also balance model complexity with interpretability. However, it is important to recognize the potential risk of losing valuable predictive variables in the process, despite the efficiency of these methods. This approach contributes to the ongoing evolution of bankruptcy prediction research.

3.4.2. Forecasting Horizon

Altman et al. (2020) emphasized the importance of including companies undergoing insolvency proceedings, in addition to those declared bankrupt, in the analysis of bankruptcy. Their rationale is rooted in the observation that financial distress manifests in stages, with insolvency typically marking the onset and bankruptcy the culmination, assuming the company's financial imbalances remain unaddressed. They argue that financial data tend to exhibit more stability over shorter intervals, thereby enhancing the accuracy of predictions within a condensed timeframe. Consequently, forecasts become inherently more reliable as the prediction horizon shortens, given the reduced uncertainty. However, Volkov et al. (2017) suggest that leveraging data spanning multiple periods can significantly amplify the efficacy of ensemble models, such as random forests or extreme gradient boosting, in classifying corporate bankruptcy. This approach contrasts with reliance on single-point-in-time data, advocating for a more dynamic, longitudinal analysis to capture the evolving financial health of firms.

Table 3.9 presents the variation in bankruptcy prediction accuracy across different forecasting horizons in previous research. It reveals an approximate 85% prediction accuracy one year prior to bankruptcy, which declines to 70% three years before the event. The discrepancy in predictive accuracy from one year to two years pre-bankruptcy stands at 8%, expanding to 14.4% when extending the forecast to three years. This trend suggest that the predictive precision of bankruptcy models diminishes as the forecasting horizon extends, highlighting the strategic importance of choosing the appropriate time frame for model development to optimize accuracy.

Table 3.9. Bankruptcy prediction models according to different forecasting horizons

Studies	**Correct Classification rates**			**Differences between Correct classification rates**		
	1 year prior to failure (Y1)	2 years prior to failure (Y2)	3 years prior to failure (Y3)	Y1-Y2	Y2-Y3	Y1-Y3
Altman (1968)	95.0	72.0	48.0	23.0	24.0	47.0
Brabazon & Keenan (2004)	74.6	66.7	63.2	7.9	3.5	11.4
Brabazon & O'neil (2004)	78.4	76.0	73.2	2.4	2.8	5.2
Charalambous et al. (2000)	80.7	72.0	66.0	8.7	6.0	14.7
Charitou et al. (2004)	76.7	73.3	56.7	3.4	16.6	20.0
Dakovic et al. (2010)	82.6	73.3	70.9	9.3	2.4	11.7
Dewaelheyns & Hulle (2006)	83.3	76.2	75.0	7.1	1.2	8.3
Gepp & Kumar (2008)	90.1	89.5	89.3	0.6	0.2	0.8
Hu & Ansell (2007)	90.1	87.2	74.6	2.9	12.6	15.5
Hu & Chen (2011)	95.4	93.0	90.5	2.4	2.5	4.9
Korol (2013)	92.7	89.4	88.2	3.3	1.2	4.5
Lin et al. (2014)	92.7	76.9	70.5	15.8	6.4	22.2
Pompe & Bilderbeek (2005)	86.2	76.9	66.8	9.3	10.1	19.4
Quek et al. (2009)	96.2	88.7	76.1	7.5	12.6	20.1
Sun et al. (2011)	74.7	65.3	63.2	9.4	2.1	11.5
Xiao et al. (2012)	81.4	75.1	72.5	6.3	2.6	8.9
Zhu et al. (2007)	86.4	72.2	69.0	14.2	3.2	17.4
Huang et al. (2012)	91.2	89.8	85.4	1.4	4.4	5.8
Geng et al. (2015)	85.0	78.0	72.0	7.0	6.0	13.0
Berg (2007)	78.1	76.8	73.4	1.3	3.4	4.7
Tian et al. (2015)	64.8	39.4	29.7	25.4	9.7	35.1
Average	**84.6**	**76.6**	**70.2**	**8.0**	**6.4**	**14.4**

Source: **Author**

Traditional bankruptcy prediction models primarily focus on a one-year horizon, operating under the assumption that the bankruptcy process and warning indicators are uniform across all firms (Chen et al., 2021). They overlook the interconnections among financial variables across various stages of corporate bankruptcy (Ashraf et al., 2019). Some companies may rapidly transition to bankruptcy despite seemingly robust health, while others may linger on the brink of failure for extended periods despite presenting worse financial indicators.

Ooghe and De Prijcker (2008) demonstrates the dynamic nature of financial distress indicators preceding bankruptcy. Their analysis reveals that the relevance of specific financial ratios as predictors of bankruptcy shifts over time. For instance, in the year preceding bankruptcy (T–1), key indicators include liquidity measures like the Current ratio, efficiency metrics such as Fixed assets turnover, and solvency ratios like the Debt/Equity ratio. These variables, alongside Firm size, Firm's age, and Depreciation ratio, are pivotal in the immediate run-up to bankruptcy. Extending the temporal scope to two (T–2) and three (T–3) years before bankruptcy, the predictive value of certain ratios, such as Fixed assets turnover and Long-term capital/Fixed assets ratio, remains significant, albeit with variations in other indicators like Inventory turnover. This evolution in significant variables over time suggests a gradual degradation in financial health, marked by diminishing operational revenues and profitability. Such financial strain often necessitates increased reliance on credit, potential asset liquidation, and, ultimately, an inability to service debt obligations.

Tsai and Wu (2008) further elucidate this phenomenon, illustrating how a decline in operational efficiency precipitates a broader financial crisis. This spiral, characterized by a compounding inability to generate sufficient operational capital, leads to exacerbated financial distress and, eventually, insolvency. The pattern of significant financial ratios changing over the T–3 to T–1 period exemplifies how the distress process is a culmination of deteriorating operational capabilities.

This comprehensive understanding of corporate bankruptcy highlights the importance of considering a diverse array of bankruptcy risk factors across varying time horizons to enhance the predictive accuracy of bankruptcy models. Recognizing the differential impact of financial ratios over time allows for a more refined and temporally sensitive approach to bankruptcy prediction, facilitating earlier and more accurate identification of firms at risk of bankruptcy.

3.4.3. Imbalanced Dataset

In the perspective of machine learning, the challenge of predicting corporate bankruptcy is framed as a binary classification problem, where the objective is to distinguish between bankrupt and non-bankrupt firms. Given the real-world distribution of companies, the dataset for such predictions is inherently imbalanced which bankrupt firms constitute a minor fraction of the total population. Research of Beaver et al. (2011) indicates bankruptcies among firms listed on major stock exchanges like NYSEAMEX and NASDAQ are rare occurrences, typically less than 1% annually. This imbalance poses significant hurdles for machine learning models, thus often assume a relatively even distribution of classes within the training dataset (He and Garcia, 2009). The rarity of bankruptcy events, as highlighted by Branco et al. (2016) makes it more difficult to develop accurate prediction models. The scarcity of bankruptcy cases compared to the abundance of non-bankrupt company data skews model training, disproportionately focusing on the majority class and ignoring an important minority class, which refer bankrupt companies.

Son et al. (2019) contended that imbalanced datasets hinder the identification of causal relationships, complicating the analysis of variable effects on the dependent variable. Consequently, the scarcity of bankruptcy datasets poses significant challenges to developing highly accurate bankruptcy prediction models. This indicates that one class (bankrupt) is underrepresented relative to the other (non-bankrupt), leading to an imbalanced distribution of data. Such imbalance significantly impacts the performance of conventional machine-learning models, often deteriorating their effectiveness.

The degradation in model performance due to data imbalance stems from two core issues (Kim et al., 2015). Firstly, the inherent assumption of class distribution symmetry in many classification algorithms' objective functions does not hold true in the context of bankruptcy prediction. Secondly, the skewed class distribution distorts the decision boundaries, leading to a bias towards the majority (non-bankrupt) class. This results in models that prioritize reducing overall misclassification errors by predominantly classifying cases as non-bankrupt, thereby failing to adequately capture the minority (bankrupt) class (Sun et al., 2020).

The consequence of this imbalance is not merely a statistical challenge but a practical concern with tangible financial implications. Misclassifying a potentially bankrupt firm as financially healthy poses a greater risk than erroneously identifying a active firm as at risk of bankruptcy. As such, addressing the data imbalance in bankruptcy prediction models is crucial not only for improving algorithmic accuracy but also for minimizing the financial risks associated with incorrect classifications. Therefore, it is crucial to develop models which enhance the sensitivity to the underrepresented bankrupt class, thereby ensuring that models achieve early and accurate bankruptcy prediction (Zoričák et al., 2020, Le, 2022).

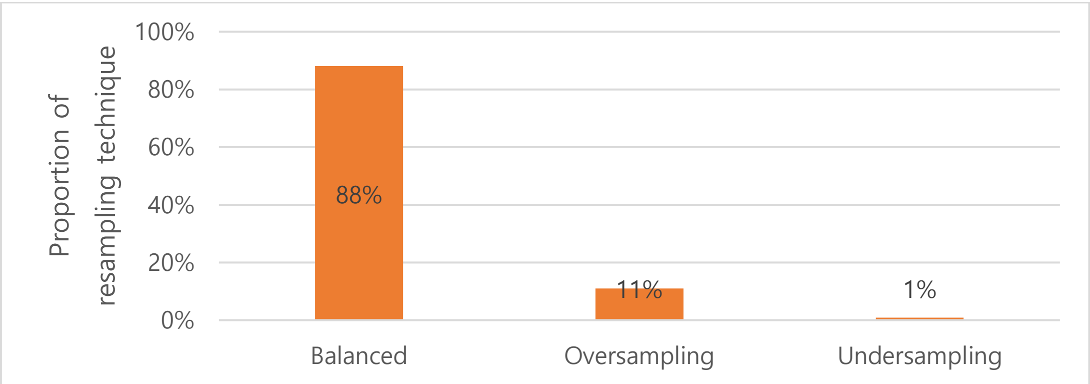
The challenge of imbalanced datasets in bankruptcy prediction significantly hampers the efficacy of classification models both during training and testing phases. During the training phase, models tend to favor accurate classification of the majority class, sometimes to the detriment of the minority class. This tendency arises because models are engineered to optimize total prediction accuracy, which results in a bias towards the majority class and overlooks instances of the minority class. Consequently, in the testing phase, this bias manifests as a predisposition towards the majority class, leading to a high rate of misclassification of minority class instances (Fernández et al., 2018). Therefore, this issue originates during the learning phase, wherein the classifiers' prediction performance is compromised, particularly concerning the minority class. In terms of bankruptcy prediction, the rarity of bankruptcy instances epitomizes an imbalance scenario, leading to models that fail to adequately reflect the characteristics of imbalanced datasets. This results in the creation of suboptimal classification models that are predisposed to unfavourable predictions across data classes (Sun et al., 2020). The discrepancy in dataset sizes, ranging from a few hundred to over 100,000 examples, further exacerbates the challenge posed by high imbalance rates.

To alleviate problems arising from unbalanced data sets, the main approach in previous research is employed as preprocessing techniques for achieving balanced class distribution. Resampling techniques, such as under-sampling and over-sampling, modify the class distribution in training data by decreasing instances of the majority class or increasing instances of the minority class (Ayesha et al., 2020). According to Haixiang et al. (2017), resampling technique is a prevalent strategy for addressing imbalanced data, with oversampling being notably more common than under-sampling. Zhou (2013) explored the application of both oversampling and under-sampling algorithms on datasets of U.S. and Japanese bankruptcies, demonstrating how oversampling replicates minority class instances to balance class distribution, whereas under-sampling selectively reduces majority class instances to achieve a similar balance. Chawla et al. (2002) explored several sampling methods, detailing Random Oversampling with Replication (ROWR) and Synthetic Minority Oversampling Technique (SMOTE) for oversampling, as well as Random Under-sampling (RU), Under-sampling Based on Clustering from the Nearest Neighbour (UBOCFNN), and Under-sampling Based on Clustering from a Gaussian Mixture Distribution (UBOCFMGD) for under-sampling.

Among these methods, SMOTE stands out as a particularly popular resampling technique due to its efficacy in balancing class distribution by generating synthetic samples (Le, 2022). This method ingeniously generates a balanced dataset with an equitable representation of bankrupt and non-bankrupt companies, facilitating the evaluation of diverse bankruptcy prediction models. The primary objective in employing such techniques is to identify the model that offers the highest classification accuracy, taking into account the unique characteristics of the dataset pertinent to bankruptcy prediction.

Figure 3.8 reveals that a significant majority of bankruptcy prediction research (88%) opts for datasets that maintain a balance between bankrupt and non-bankrupt companies. This approach primarily aims to ensure robust explanatory power and mitigate statistical challenges associated with imbalanced datasets by selecting sample sizes large enough to evenly distribute instances across both classes. Conversely, only a small portion of studies (12%) resort to resampling techniques to address dataset imbalances, with oversampling techniques being utilized in 11% of these instances and under-sampling in a mere 1%. The scarcity of datasets, particularly when focusing on specific industries or countries, constrains the ability to achieve adequate sample sizes, thereby amplifying the necessity for resampling techniques. In studies targeting specific industries or geographic regions, oversampling techniques, especially SMOTE, are generally preferred when data sets are significantly smaller and have significant imbalances. Through the strategic application of oversampling, researchers can effectively counteract the limitations imposed by imbalanced datasets, thereby enhancing the predictive accuracy and reliability of bankruptcy prediction models in specialized contexts (Wang and Liu, 2021).

Figure 3.8. Proportion of resampling techniques used in previous studies



Source: **Author**

3.5. Interpretability of bankruptcy prediction

Until recently, the focus in bankruptcy prediction was primarily on enhancing the accuracy of models, leading to increasingly complex designs like those seen in ensemble learning. While machine learning models have demonstrated superior predictive abilities due to their intricate algorithms, they have encountered a significant challenge known as the "black box" problem, which obscures the interpretability of their analysis (Jones, 2017). In contrast, conventional statistical methods are favoured for their clarity and ease of interpretation, as they rely on a select number of relevant variables for bankruptcy prediction, assessing their significance through feature selection algorithms. Despite the comparative lower predictive power of traditional statistical models against machine learning counterparts, their simplicity and interpretability have sustained their use in the field (Altman et al., 2020).

Accounting-based indicators of bankruptcy risk continue to be actively employed by researchers and practitioners, serving as a key criterion for assessing financial health (Das et al., 2009). The ability to identify significant financial indicators enables the modelling of failure probability and the prediction of financial distress, despite criticisms that these indicators are calculated after the fact (Gavurova et al., 2017). Variables with a substantial impact on bankruptcy prediction can serve as crucial tools for risk management, allowing companies to monitor their financial status closely and identify risks early to prevent bankruptcy (Ashraf et al., 2019). Financial institutions can leverage these indicators to avoid engaging with corporates at high risk of bankruptcy and make more informed decisions regarding financial support (Wang et al., 2021). However, the reliance on simple statistical assumptions and multivariate functions in these conventional methods often results in lower predictive accuracy when compared to machine learning model (Chen et al., 2021). This highlights the challenge in balancing the interpretability of statistical methods with the superior predictive capabilities of machine learning techniques.

Recent efforts in bankruptcy prediction have sought to balance the trade-off between model accuracy and interpretability by extracting comprehensible rules from complex "black box" machine learning models (Virág and Nyitrai, 2014, Obermann and Waack, 2015, Ribeiro et al., 2016). However, these approaches tend to approximate the original model's functionality, potentially compromising accuracy in favour of interpretability (Mi et al., 2020). Recently, Son et al. (2019) suggested a way to enhance the interpretability of machine-learning models, specifically boosting tree models, by employing feature importance techniques. This method involves assessing the Relative Variable Importance (RVI) of predictors within the model, ranking them according to their weighted classification accuracy, which is then averaged across all predictors (Chen and Guestrin, 2016). The calculation of RVIs considers how frequently a variable is used for splitting within the model, its contribution to improving the model through each split, and then averaging this contribution across all trees (Jones et al., 2017). Predictors are then scored based on their summed improvements, with the most impactful predictor scored at 100 and diminishing scores assigned to others, indicating their relative contribution to the model's predictive power. A low or zero RVI suggests minimal contribution by a variable to the model's overall effectiveness (Jones and Wang, 2019). However, interpreting RVIs poses challenges, as they provide limited insight into how specific financial ratios influence bankruptcy predictions beyond indicating their relative importance or marginal effects (Mi et al., 2020)

Finally, the emergence of explainable artificial intelligence (XAI) techniques marks a significant advancement in making complex machine learning models more understandable and transparent. Ribeiro et al. (2016) introduced the Local Interpretable Model-agnostic Explanation (LIME) technique, providing local linear approximations to explain the predictions of any machine learning model by perturbing input data samples. Similarly, Lundberg and Lee (2017) developed the SHAP (SHapley Additive exPlanations) method, integrating concepts from LIME and Shapley values to offer detailed interpretations of model prediction. While these methods aim to demystify the decisions made by machine learning models, they also seek to align these explanations with the needs of stakeholders, ensuring that the rationale behind predictions is accessible and meaningful. Applications of these XAI methods in bankruptcy prediction have demonstrated their potential to maintain high predictive accuracy while enhancing model transparency. For instance, Crosato et al. (2021) and Park et al. (2021a) utilized SHAP and LIME, respectively, to ascertain the importance of features within their bankruptcy prediction models. Ariza-Garzón et al. (2020) applied SHAP in the context of peer-to-peer lending to uncover complex relationships between features and the target variable, thereby improving interpretability.

Despite the advancements in developing sophisticated models for bankruptcy prediction, the challenge of balancing accuracy with interpretability remains. Traditional statistical models like logistic regression or linear discriminant analysis continue to be favoured for their simplicity and ease of understanding, despite potentially lower predictive performance. This efficiency and clarity is crucial in financial risk assessment, where firms rely on predictive insights to inform strategic decisions and risk management practices (Kang et al., 2022). Therefore, by incorporating XAI techniques into bankruptcy prediction models, researchers and practitioners can bridge the gap between complex model accuracy and the need for transparency. XAI enables the detailed examination of how specific variables influence model predictions, facilitating the identification of key bankruptcy risk factors. This approach not only enhances the utility of predictive models in practical applications but also supports stakeholders in making informed decisions based on comprehensive and interpretable insights.

3.6. Bankruptcy prediction in specific area

3.6.1. Bankruptcy prediction in Small and Medium-sized Enterprises

When developing bankruptcy prediction models for a specific country or region, it's crucial to focus on small and medium-sized enterprises (SMEs) due to their significant presence in the industrial sector (Zoričák et al., 2020). In Korea, SMEs are defined as companies with total assets less than 50 billion won and average sales less than 8 billion won. This definition encompasses SMEs across various industries, including the shipping sector, where SMEs operate alongside larger shipping companies. Large shipping companies, in contrast, are defined as those exceeding these asset and sales thresholds.

Predicting bankruptcy for SMEs is challenging because of their unique behaviours and fundamental characteristics that set them apart from larger companies (Cultrera and Brédart, 2016). In other words, recognizing these distinct characteristics is essential to effectively tailor bankruptcy prediction models to small and medium-sized businesses.

Previous research has explored the bankruptcy risk among SMEs by examining their characteristics from various angles, as summarized in Table 3.10. A common finding across these studies is the vulnerability of SMEs stemming from their financial structure. Often, these enterprises rely heavily on short-term credits and face substantial barriers to securing medium- and long-term financing (Ciampi and Gordini, 2008). This financial vulnerability has drawn criticism from governments and SME associations alike, raising concerns that excessive capital charges might result in the credit rationing of small firms. Such financial constraints on SMEs pose a risk to economic growth, considering the critical role these firms play in the economy (Altman and Sabato, 2007).

Additionally, other studies have delved into the complexities and opportunities within small business lending, analysing factors that influence SME profitability and the potential risks for banks, particularly in the U.S (Luo et al., 2020). Research has also investigated lending structures and strategies, providing insights into how financial institutions interact with SMEs (Park et al., 2021c). Therefore, developing a bankruptcy prediction model for SMEs requires a comprehensive understanding of their unique financial structures, market behaviours, and the challenges they face in accessing credit.

Table 3.10. Previous studies of bankruptcy prediction of SMEs

Model	No. of Firms	Variable		Author(year)
		Financial	Non-financial	
ANN	14966	Liquidity, Leverage, Profitability, Efficiency	Number of Employee	Di Ciaccio and Cialone (2019)
SVM	5840	Liquidity, Leverage, Profitability, Efficiency	Age, Size	Zoričák et al. (2020)
Logit	15605	Liquidity, Leverage, Profitability	Credit information, Age, Business Category	Luo et al. (2020)
Logit	102	Liquidity, Leverage, Profitability	Age, Size, Number of employees	Mayr et al. (2021)
Logit	30	Leverage	Age, Size, Number of employees	Padilla-Ospina et al. (2021)
Logit	2342	Liquidity, Leverage	Age, Credit status	Lee et al. (2020)
MDA	2120	Liquidity, Leverage, Profitability	Size, Business Category	Park et al. (2021c)
Logit	20000	Liquidity, Leverage, Profitability	Size, Number of managers	Tobback et al. (2017)
Logit	7152	Cash flow/total debt, Fiscal charges/added value, Current ratio, Earnings before interest and taxes/total assets, Equity/total assets	Size, Age, Business Category	Cultrera and Brédart (2016)
Logit	7424	Liquidity, Leverage, Profitability	Size, Age, Business Category	Ciampi and Gordini (2013)

Source: **Author**

The practice of differentiating between large corporations and small and medium-sized enterprises (SMEs) in credit risk modelling is already common among banks and consulting firms. This distinction is crucial, given the higher bankruptcy risk associated with SMEs, making a specialized credit model for these enterprises preferable for credit assessment purposes (Wang et al., 2022). In assessing credit risk for SMEs, researchers and financial institutions now incorporate a mix of financial, accounting-based, and non-financial variables.

Non-financial data play a significant role in this assessment and can be categorized into two main groups: firm-based non-financial data such as business sector, age, company patents, and corporate governance indicators, and external information including industrial and market information (Li et al., 2022). Variables indicative of firm size have shown a strong correlation with corporate bankruptcy probability, suggesting that smaller firms are less likely to face repayment difficulties compared to larger firms. Norden and Weber (2009) observed that the complexity of cash flows in checking accounts and the mechanisms of bankruptcy differ markedly across various company types. Larger firms often have the advantage of diverse funding sources and numerous bank relationships, unlike SMEs. Additionally, while financial data for large corporations and publicly listed companies are readily accessible, obtaining financial information for SMEs poses challenges, limiting its utility in bankruptcy prediction (Mayr et al., 2021).

Although financial information provides a standardized data format for assessing company status, it is subject to reporting delays, being disclosed only after quarterly or annual financial statements are finalized (Oliveira et al., 2017). This can be compounded in the case of SMEs, where obtaining accurate and comprehensive financial data is further complicated by their vulnerability and managerial challenges (Di Ciaccio and Cialone, 2019). Consequently, this has led to a scarcity of research focused on insolvency predictions specifically for SMEs.

Therefore, there's a pressing need to explore and integrate additional variables into bankruptcy prediction models for SMEs, taking into account their unique industrial characteristics. This approach will not only improve the accuracy of predictions but also support the development of tailored financial solutions to mitigate the heightened risk of bankruptcy faced by SMEs.

3.6.2. Bankruptcy prediction in Shipping industry

The shipping industry is known for its high cyclicality, volatility, capital intensity, and often significant leverage, making financial management within the sector particularly challenging (Christiansen et al., 2007). These characteristics can pose significant risks for shipping companies, especially when it comes to fulfilling obligations related to interest and capital repayments during downturns in the shipping market (Wang et al., 2017). The risk is further exacerbated for companies that predominantly operate their fleets in the spot market, as opposed to the more stable time-charter market. The late 1990s provide a case in point, where several shipping companies, heavily reliant on the spot market, ventured into the US high-yield bond market in 1997 and 1998. The subsequent recession in the shipping market in 1999 left these companies struggling to meet their liabilities, illustrating the vulnerabilities associated with such financial strategies (Lozinskaia et al., 2017). Additionally, the needs to expand company size and fleet renewal efforts during periods of high vessel prices significantly eroded the equity bases of many shipping companies during the late 1990s and early 2000s.

The financial crisis of 2008-2010, along with the banking crisis of the early 1980s, prompted a shift in strategic financial management within the shipping industry. To navigate temporary financial difficulties and fund newbuilding and second-hand purchases, companies have progressively utilized a combination of bank loans, private placements, and public issuances of debt and equity (Kavussanos and Tsouknidis, 2016). These developments have encouraged many shipping firms to regard capital markets not merely as a means of financing but as an integral part of their comprehensive strategy for enhancing financial management (Makrominas, 2018). Given the unique financial characteristics of the shipping industry, distinct from other sectors, it's critical to consider specific aspects when analysing financial data and interpreting bankruptcy prediction indicators for companies within this sector (Haider et al., 2019). One key aspect is leverage, such as the current ratio, which measures a company's capability to settle current liabilities with its current assets. A significant imbalance where current liabilities growth outpaces current assets growth signals potential liquidity issues, possibly leading to bankruptcy (Grammenos and Papapostolou, 2012).

Studies have shown that a high level of financial leverage, indicative of an equity shortage, signals an elevated bankruptcy risk (Barboza et al., 2017, Altman, 2018). This relationship between financial leverage and bankruptcy probability has been confirmed in various studies specific to the shipping industry as shown in Table 3.11. The gearing ratio, which compares long-term debt to the sum of long-term debt and shareholder's equity, provides insight into a company's debt burden and its resilience during income recession periods (Berk et al., 2013). An increasing gearing ratio suggests a growing dependency on debt for vessel acquisitions, which can lead to challenges in meeting interest and repayment obligations under adverse market conditions (Grammenos et al., 2007). Shipping companies with high leverage ratios often struggle with income generation, facing significant financial risks and challenges in servicing debt obligations to bondholders. The positive correlation between the leverage, as represented by gearing ratio, and bankruptcy risk highlights the financial instability of highly leveraged companies within the shipping sector. This finding aligns with other research in the field, such as the studies by Kavussanos and Tsouknidis (2016) or Lozinskaia et al. (2017), further emphasizing the critical need for shipping companies to manage their financial leverage carefully to mitigate bankruptcy risks.

The comprehensive survey of literature reveals that the probability of bankruptcy in shipping companies is influenced by combining financial, non-financial, industry-specific characteristics, and macroeconomic variables. Macroeconomic factors emerge as crucial predictors of bankruptcy, underscoring the shipping industry's vulnerability to global economic shift (Wang et al., 2017) Given the industry's significant role in the global economy, variations in economic conditions can significantly impact financial health and financing capabilities of shipping companies (Clintworth et al., 2021).

Table 3.11. Previous research of bankruptcy prediction in shipping industry

Model	Period	No. of Firms	Variable: Financial	Variable: Non-financial	Variable: Shipping index	Author (Year)
Logit	1998-2002	32	Market value	Inflation, Interest rates	ClackSea Index, Laid-up Tonnage	Grammenos et al. (2007)
Logit	1992-2004	50	Current ratio, EBITDA, Gearing ratio, Return on assets, Return on equity, Working capital to total assets,		Freight rate, Time charter rate	Grammenos et al. (2008)
MDA	1999-2007	48	Total assets, Cash ratio, Leverage, Profitability, Market-to-book, Stock return	Inflation, Interest rates, Currency	MSCI stock index, oil price,	Drobetz et al. (2010)
MDA	1997-2011	63	Coverage, Leverage, Liquidity, Profit margin	Age	ClarkSea index, MSCI stock index, Inactive tonnage to total fleet ratio, Time charter rate	Kavussanos and Tsouknidis (2016)
Logit	2005-2009	30	Leverage	Age	Baltic dry index (BDI), Fleet size, freight rate	Mitroussi et al. (2016)
Logit	2001-2016	192	Current ratio, EBITDA, Leverage, Return on assets, Return on equity, Tobin's Q,	Age, GDP	Baltic dry index (BDI), IRONSTEEL, MSCI stock index, Oil price	Lozinskaia et al. (2017)
MDA	2004-2016	12		Exchange rate	Baltic dry index (BDI), Container freight index rate, Oil price	Choi et al. (2018b)
Logit	1992-2014	40	Assets turnover, Cash ratio, Current ratio, EBITDA, Gearing ratio, Return on assets, Return on equity, Market value		Chartering cost, Freight income, Oil Price, Shipping cost, Ship rental income, Ship value, Voyage cost,	Haider et al. (2019)
MDA			Profitability, Market to book ratio, Stock return, Liquidity, Cash	Size, Cumulative abnormal returns, Daily stock price		Alexaandridis et al. (2020)
XGB	2000-2018	82	Current ratio, EBITDA, Liquidity ratio, Gearing ratio, Return on assets, Return on equity, Profit margin	Interest rate, Inflation, GDP	New build price index, Baltic dry index (BDI)	Clintworth et al. (2021)
Logit	2001-2019	74	Assets turnover, Current ratio, Return on assets, Quick ratio, Profitability		Chartering cost, Freight rate, Oil price, shipping cost, ship value, voyage cost	Park et al. (2021b)

Source: **Author**

The vulnerability of these companies to macroeconomic fluctuations emphasizes the importance of integrating a wide range of economic indicators into bankruptcy risk evaluations (Nguyen et al., 2022). Factors such as government regulations, policy shifts, and changes in business risk levels also play a significant role in affecting the financial health and stability of shipping companies (Choi et al., 2018b). Additionally, economic downturns are increasing the exposure of the shipping industry to bankruptcy risk, and the cyclical pattern of international trade, characterized by economic downturns and subsequent recoveries, has a direct impact on the demand for shipping services (Foreman, 2003). These results highlight the importance of responding to these macroeconomic fluctuations to effectively manage bankruptcy risk.

There have been also several attempts to apply non-financial variables as the risk factors of bankruptcy. Kavussanos and Tsouknidis (2016) highlighted that younger shipping companies exhibit a higher likelihood of bankruptcy compared to their more established counterparts, suggesting that industry experience plays a critical role in a firm's resilience. This is likely due to younger firms having less established customer bases, smaller networks, and potentially less access to capital. Similarly, Grammenos et al. (2008) found a negative correlation between a shipping company's size, as approximated by the logarithm of total assets, and its bankruptcy risk. This indicates that smaller or medium-sized shipping firms are at a greater risk of bankruptcy, possibly due to their limited financial resources and lower capacity to withstand economic shocks.

Finally, bankruptcy risk analysis in the shipping industry requires a specialized approach that is differentiated from general financial research due to the unique financial structure and operating characteristics of shipping companies. Previous studies have pointed out that the probability of bankruptcy is significantly higher for companies with certain financial configurations, emphasizing the need for industry-specific forecasting models (Martin et al., 2011). These models, which tailor independent variables to the specific features of each industry, tend to offer superior prediction accuracy compared to more generalized models (Park et al., 2021b). In contrast to the broader finance field where financial ratios and company characteristics predominate in assessing bankruptcy probabilities for bank loans, these factors play a different role in the shipping industry. The critical determinants of bankruptcy in shipping are not solely financial ratios but include market risk indicators, chartering policies, and other industry-specific risk factors (Grammenos et al., 2007). Consequently, financial statements, which carry outdated information, are often insufficient for making informed decisions in this rapidly changing environment (Haider et al., 2019).

Recent studies have emphasized the importance of shipping market indicators in predicting bankruptcy, particularly under turbulent market conditions and when funding options are constrained. For instance, the probability of bankruptcy has been shown to inversely correlate with the Baltic Dry Index, highlighting the impact of market conditions on bankruptcy risks (Mitroussi et al., 2016). Similarly, Kavussanos and Tsouknidis (2016) highlighted the necessity of considering both the current state and future outlook of the ship transport market when assessing the bankruptcy risk of shipping loans. Their analysis over a 14-year period revealed that profitability, leverage, and liquidity are closely tied to shipping indices, reflecting the cyclical nature of the shipping business and its influence on bank decision-making. Freight rates also play a essential role in determining bankruptcy risk, further illustrating the sector's dependency on market conditions (Grammenos et al., 2008). Haider et al. (2019) identified eight industry-specific indicators that capture the financial characteristics unique to the shipping industry, including ratios related to freight income, ship rental income, fuel cost, and ship value. These indicators reflect the capital-intensive nature of the industry, its exposure to volatile freight rates and ship prices, and the cyclicality and seasonality that define its operations. Therefore, evaluating bankruptcy risk in the shipping industry requires consideration of both static factors and evolving elements that affect financing choices over time. Macroeconomic conditions and industry-specific market indicators are deemed critical, while traditional financial ratios, though relevant, assume a secondary role in the analysis. This analysis highlights the necessity of a comprehensive understanding of the shipping industry's unique financial landscape to accurately evaluate and manage bankruptcy risk

3.7. Gaps of Research

Despite various attempts to predict corporate bankruptcy crisis, there are still necessary to explain the influence of determinants to bankruptcy risk by machine learning models (Kücher et al., 2020). From previous research about bankruptcy prediction, prediction models tend to be characterized with its trade-off features between accuracy and interpretability of result (Mayr et al., 2021). Despite of its relatively lower predictive power, conventional statistic models such as logistic regression or can explain the results and influences of input variables (Altman et al., 2020). In contrast, non-linear machine learning models such as support vector machine or artificial neural network show high level of predictive ability regardless of its lack of interpretability. This is because these models can capture non-linear relationship easily with their internal mathematical process hidden in the model so called "black boxes", which limit its practical interpretation of result. In these regards, as the classifier, a simpler and more explainable model would be preferred to complex model even though relative lower predictive ability (Voda et al., 2021). Hence, much of research of bankruptcy prediction has been still applied conventional statistic classifiers in terms of its practical values.

Up-to-date models such as ensemble machine learning or deep learning model have proved its practical value with high level of predictive ability in previous research as well as its simplicity in implementation from model structure or data pre-processing process (Kücher et al., 2020). To become a practical prediction model, it should overcome this trade-off feature which have high predictive ability as well as interpretability of the influence and role of variables on the overall model performance (Crosato et al., 2021). To overcome this limitation, explainable AI (XAI) would be applied as the evaluation method about the influence of risk factors to result. Through XAI, those “black box” models can be explainable which can help decision making process by detecting and correcting bias in the dataset (Lee et al., 2022). In addition, XAI can describe expected impact and potential biases by inferred only meaningful variables. Therefore, a set of risk factors which determine bankruptcy on Korean shipping industry can be identified by applying XAI in bankruptcy prediction models.

Concerning the wider impacts of corporate bankruptcy crisis to global economic components, there have been extensive attempts to discover the determinants of corporate bankruptcy risk by applying different databases from wide range of methodologies, industries or countries (Veganzones and Severin, 2020). From the extensive research about bankruptcy prediction of shipping industry, because most have focused on corporate's internal factors such as financial ratios, the external variables such as macroeconomic factors have required more attention (Kin et al., 2021). The internal factors of bankruptcy should be concerned alongside with corporate features such as age or size to assess the probability of bankruptcy (Kücher et al., 2020). Although several attempts have been made to explore the shipping industry's bankruptcy prediction using various methodologies, most studies on the shipping industry have focused on the bankruptcy risk of global shipping companies such as Maersk or MSC, despite the high proportion of SMEs in the entire shipping industry. This is because most of research focused on stock price or market value as the indicator of corporate value, which can reflect their financial condition. In contrast, most shipping SMEs are unlisted which faced limitation to evaluate corporate value reflecting market fluctuation. Therefore, it is evident that the approach to explore the specific perspective for shipping SMEs is required to identify risk factors for bankruptcy prediction. For instance, macroeconomic risks such as GDP would be considered because most shipping SMEs are much susceptible to these risks due to their lack of access to financing compared to major shipping companies (Lee et al., 2020). This study tries to fill the gap by applying industry variables extracted from the analysis of the Korean shipping industry, where SMEs account for more than 90% of the total shipping industry. In addition, semi-structured interviews with shipping industry practitioners will be conducted not only to justify predictive ability of those identified risk factors but also to discover risk factors which have not been used in previous research for bankruptcy of shipping SMEs. Through this approach, it is expected that specific bankruptcy risk factors in the shipping SMEs can be identified from the perspectives of academics and industry practitioners.

These identified risk factors would be analysed their predictive ability in not only short-term (1~2 years) but also mid-term (3~5 years) forecasting horizon. In previous research, forecasting horizon have been set to one or two years, which were defined as the event year and a year prior to bankruptcy (Shi and Li, 2019). Nonetheless, corporate bankruptcy is a long-term process, which its symptoms can be detected from corporate financial situation in few years prior to bankruptcy event (Rebetak and Bartosova, 2021). Furthermore, mid-term bankruptcy prediction is essential for banks or financial institutions in their decision process whether a loan should be granted to a corporate or not due to the financial situation over the 3 or 4 years (Mayr et al., 2021). Thus, forecasting horizon would be set on not only a year prior to bankruptcy but also three years as well as five years prior to the bankruptcy for evaluating its mid-term predictive ability. Identified different set of risk factors according to different forecasting horizon would be evaluate influence to bankruptcy risk of Korean shipping SMEs. It is expected to propose managerial implication to improve bankruptcy risk management by different sets of risk factors according to different forecasting horizon for small and medium sized shipping companies.

3.8. Chapter Summary

This chapter provides an in-depth review of studies in corporate bankruptcy prediction, charting the progression from classic statistical approaches such as Multiple Discriminant Analysis (MDA) and logistic regression, to contemporary machine learning techniques including Artificial Neural Networks (ANN) and Support Vector Machines (SVM), and advancing to state-of-the-art methodologies like boosting and deep learning models. This evolution marks a progressive relaxation of previous statistical assumptions, including skewness, outliers, and multicollinearity, and highlights the enhanced predictive ability of more sophisticated models.

The chapter also delves into the array of bankruptcy risk factors highlighted in past research, setting a foundation for the variable selection in this study. Predictors include accounting-based variables, market-price indicators, macroeconomic factors, corporate-specific variables, and shipping index variables, underlining the complex nature of assessing bankruptcy risk.

However, the literature review also identified several gaps and limitations within existing studies. One such limitation is the lack of consensus on the application of bankruptcy predictors, especially when targeting specific industries. To address this, the current study proposes conducting interviews with academics and practitioners within the Korean shipping industry to pinpoint industry-specific risk factors, thereby enhancing the theoretical foundation for variable selection.

Moreover, most research to date has been constrained to short-term prediction windows (one to two years prior to bankruptcy), underestimating the gradual nature of corporate decline. This study intends to extend the predictive horizon to one, three, and five years, aiming to capture medium-term predictive capabilities and address the decline in predictive accuracy over longer periods through the adoption of advanced predictive models.

Lastly, while advanced models such as boosting and deep learning offer heightened predictive accuracy, they often suffer from a lack of result interpretability. This study addresses this limitation by employing explainable AI techniques, aiming to elucidate the impact of various variables on bankruptcy risk. The objective is to delineate a comprehensive set of bankruptcy risk factors tailored to small and medium-sized shipping companies, thereby contributing valuable insights to the field of corporate bankruptcy prediction.

4. METHODOLOGY

4.1. Chapter Introduction

Bankruptcy prediction is fundamentally approached as a binary classification problem. Although numerous studies have demonstrated the effectiveness of various binary classification models in dealing with bankruptcy prediction, a common assumption across these studies is the complete availability of all necessary input features (Korol, 2017). Risk factors as input variables for bankruptcy prediction can be obtained from historical data. Access to financial data spanning several years empowers analysts to calculate average financial indicators tailored to specific sectors and to visually chart these indicators' fluctuations over time. This process allows experts to discern periods marked by significant industry downturns, leading to the financial distress and subsequent bankruptcy of several companies (Philosophov et al., 2008).

Moreover, when experts detect industry deterioration in a given year, they can infer that such decline is mainly attributable to particular changes in industrial conditions (Choi et al., 2018a). The single experience of one company is insufficient to conclusively predict the fates of others within the same sector, necessitating a reliance on aggregate industry indicators (Kim et al., 2020). Additionally, corporate bankruptcy during financial distress is often linked to external factors, such as force majeure events, shifts in international relations, legislative modifications, and other global occurrences (Kin et al., 2021).

Previous research typically developed prediction models based on data specific to a country or a group of countries, regardless of research method. Some models specifically draw on data pertinent to certain industries, reflecting the unique conditions of the country, its economic progress, and the evolution of its capital markets. Consequently, the adaptability of these models to different geopolitical contexts may be constrained, advising cautious application of country- or industry-specific models. Notably, models tailored to specific economic sectors can surpass the predictive capabilities of models devised elsewhere or for broader applications (Kovacova et al., 2019). It is, therefore, essential to identify distinct risk factors for refining bankruptcy prediction models tailored to particular industrial and economic context.

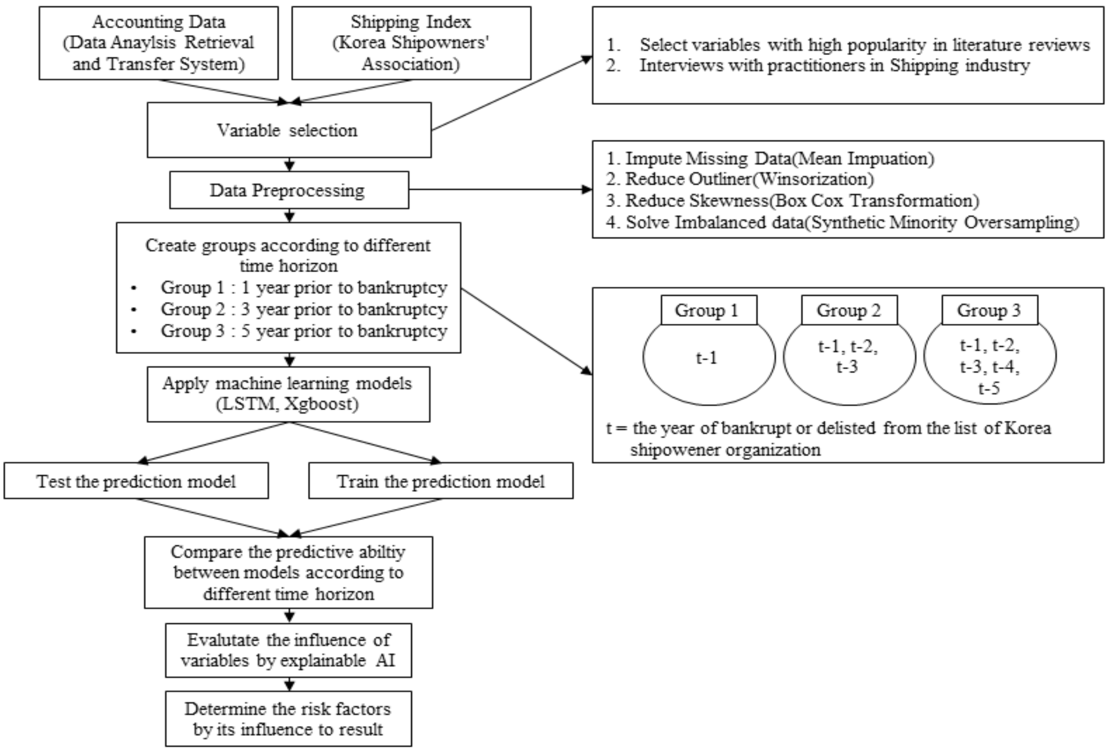
The complexity in bankruptcy prediction arises from the simultaneous occurrence of multiple risk factors within a single year, each impacting the industry's state to varying degrees (Altman et al., 2020). This complexity, compounded by data from diverse sources, poses challenges in model development and variable impact explanation (Li and Miu, 2010). In empirical research, handling observations with missing data typically involves either their exclusion or the imputation of missing values based on the most recent data in the series (García-Laencina et al., 2010). While listed companies typically uphold stringent finance and accounting management systems that reduce the likelihood of missing data, smaller and medium-sized non-listed companies may not have such extensive systems in place, resulting in incomplete financial statements (Pederzoli and Torricelli, 2010). Thus, bankruptcy prediction models must be robust enough to address both the presence and absence of missing data. Machine learning methods, known for their potent predictive capabilities, offer a solution for analysing these challenges across the industry spectrum. Through machine learning, a training set categorizing companies as either bankrupt or solvent can be developed, allowing for the extraction of insights and enabling precise bankruptcy predictions.

This section provides an overview of the research framework, detailing the statistical and mathematical principles utilized in modelling bankruptcy prediction. Figure 4.1 depicts how bankruptcy prediction modelling would proceed through the sections. Data would be extracted from the official statistics provided by Korean Shipowner Association[8]

[8] Available at: https://oneksa.kr/boards/statistics

(2023) an organization representing shipping companies in Korea, and DART[9] provided by Korean Financial Supervisory Service (2023), a government system for corporate financial statements. The core of this chapter is segmented into five primary sections, each representing a crucial stage of statistical modelling. The first section discusses the criteria for sampling datasets, focusing on the definition of bankruptcy and the forecasting horizon. Subsequent sections, 4.4 and 4.5, are dedicated to the selection of explanatory variables. The foundational theories of the classification algorithms employed in this research are outlined in section 4.6. Section 4.7 delves into data pre-processing issues, including theories on handling missing values, statistical data transformation, outliers, and class imbalance. The final section, 4.8, addresses the evaluation criteria for classification models. Last section would discuss how the interpretability of prediction model would be secured through up-to-date technique, explainable artificial intelligence.

Figure 4.1. Research procedure of bankruptcy prediction model in this study



Source: **Author**

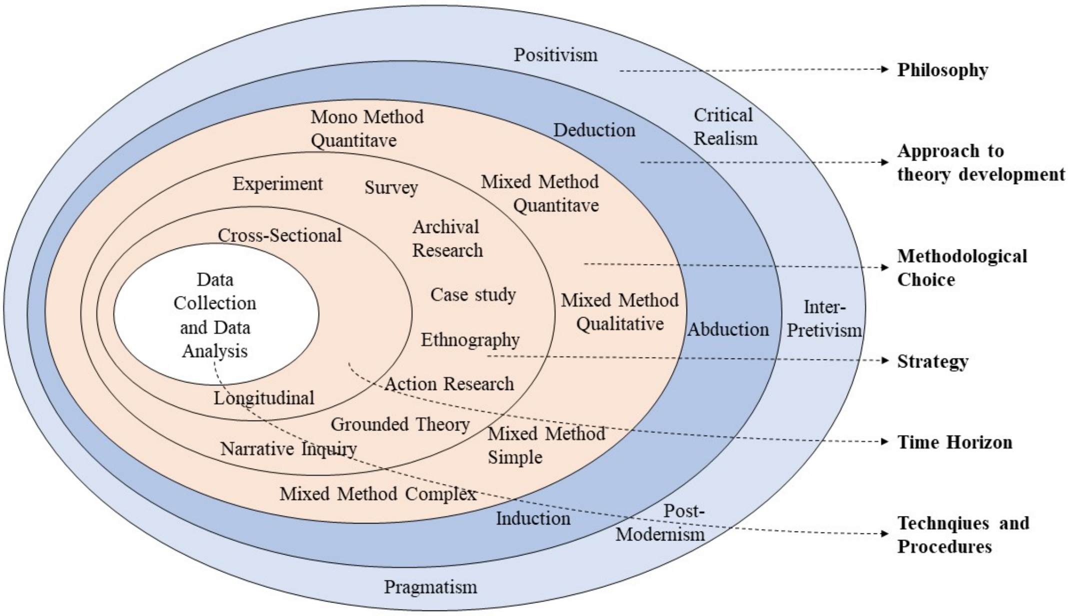
[9] DART: Data Analysis Retrieval and Transfer System provided by Korean financial supervisory service

4.2. Research Philosophy

Research philosophy is recognized as a fundamental parameter for addressing why a researcher decides to undertake a specific study. It requires developing a philosophical stance based on fundamental assumptions related to ontology, epistemology, and methodology. Saunders et al. (2009) introduce a spectrum of philosophical perspectives—positivism, critical realism, interpretivism, post-modernism, and pragmatism—through an analogy referred to as the "research onion," which integrates ontological, epistemological, and axiological stances, as illustrated in Figure 4.2.

Ontology concerns the foundational beliefs about the nature of reality and the essence of the phenomena under study. It essentially shapes a researcher's viewpoint regarding their subject matter, influencing the selection and approach to the research topic. Epistemology delves into the nature of knowledge itself, questioning what is considered acceptable, valid, and legitimate knowledge (Burrell and Morgan, 2017). Given the multidisciplinary nature of business management research, a spectrum of epistemological perspectives is adopted, each suggesting different methodological approaches while underscoring the importance of recognizing the implications and limitations that these foundational beliefs impose on the choice of research methods. Axiology, on the other hand, examines the role of values and ethics within the research process. It involves an introspective look at how researchers navigate their personal values in conjunction with those of their research participants (Adams et al., 2014). Through their research choices and methodologies, researchers manifest their axiological stance, reflecting on the ethical dimensions and value-driven aspects of their work (Heron, 1996). This philosophical grounding not only influences the direction and integrity of the research but also ensures a thoughtful engagement with the subject matter, participants, and broader academic community.

Figure 4.2. Research Onion



Source: Saunders et al. (2009)

This research adopts pragmatism as its philosophical stance. Pragmatism emphasizes practical outcomes and problem-solving, integrating different research methods to effectively address research objectives. Unlike a single ontological or epistemological position, pragmatism focuses on using the most suitable methods to answer research questions (Creswell & Poth, 2016). The pragmatic orientation of this study drives the use of a mixed methods approach, combining quantitative and qualitative methodologies. This integration is essential to comprehensively explore the determinants for predicting bankruptcy risk in shipping SMEs using financial, non-financial, and economic data through machine learning models.

The quantitative component adheres to positivist principles within a pragmatic framework, focusing on observable and measurable phenomena. The objective is to identify causal relationships between various risk factors and corporate bankruptcy, relying on highly structured statistical methodologies and quantifiable observed data. Advanced machine learning models such as XGBoost and LSTM are employed to ensure high predictive performance.

Complementing the quantitative analysis, the qualitative component aligns with interpretivist principles within a pragmatic framework. This approach acknowledges the subjective nature of human experiences and the importance of context in understanding complex phenomena (Saunders et al., 2009). Semi-structured interviews with industry practitioners were conducted to capture the nuanced perspectives of various stakeholders in the shipping industry. This method allows for the exploration of complex dynamics and the identification of new variables that may not be evident through quantitative data alone.

In essence, this study's philosophical approach embodies pragmatism by integrating positivism in its ontological stance and blending both positivism and interpretivism in its epistemological and axiological perspectives. This pragmatic approach signifies the acceptance and application of multiple methods, tailored to effectively meet the research objectives. It acknowledges that employing a variety of methodologies is not only possible but also highly suitable within a single study, catering to the complex and multifaceted nature of business research.

4.3. Research Design

This research adopts a pragmatist philosophy, which drives the selection of a mixed methods approach, integrating both quantitative and qualitative methodologies to effectively address the research objectives by identifying and analysing significant risk factors for predicting bankruptcy in the shipping industry, particularly focusing on SMEs in Korea. The quantitative component is grounded in positivist principles within a pragmatic framework, aiming to identify causal relationships between various risk factors and corporate bankruptcy by utilizing advanced machine learning models, such as Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) networks, to handle complex, non-linear relationships and large datasets due to their high predictive accuracy (Chen & Guestrin, 2016; Hochreiter & Schmidhuber, 1997). Data collection involves gathering financial data (e.g., liquidity ratios, profitability ratios, leverage ratios), non-financial data (e.g., management quality, corporate governance structures), and macroeconomic data (e.g., interest rates, GDP growth, inflation rates) from sources such as financial statements, industry reports, and economic databases (Beaver, 1966; Altman, 1968; Moyer, 1977; Tinoco & Wilson, 2013). Data preprocessing involves handling missing data, outliers, and ensuring data normalization to prepare the datasets for analysis. Feature selection focuses on identifying the most relevant features for the prediction models, ensuring that essential information is not lost. The models are trained on a training dataset and validated using a separate validation dataset, with cross-validation techniques employed to ensure robustness. The models' performance is evaluated using metrics such as AUC-ROC, precision, recall, and F1 score, providing a comprehensive assessment of their predictive accuracy (Alam et al., 2021a). The study examines bankruptcy risk across three forecasting horizons: 1 year, 3 years, and 5 years prior to bankruptcy, allowing for a detailed analysis of temporal patterns in financial distress (Jones & Wang, 2019; Voda et al., 2021). Explainable AI (XAI) techniques are used to interpret the results of the machine learning models, providing insights into the influence of individual variables on bankruptcy predictions and ensuring transparency and enhancing stakeholder trust (Ribeiro et al., 2016).

The qualitative component aligns with interpretivist principles within a pragmatic framework, focusing on capturing the comprehensive perspectives of stakeholders in the shipping industry through semi-structured interviews with industry practitioners to validate and supplement the findings from the quantitative analysis (Saunders et al., 2009). Participant selection involves identifying and selecting industry practitioners based on their expertise and experience, ensuring a diverse representation of insights. The interview process includes developing a semi-structured interview guide, conducting interviews in person, and recording and transcribing the conversations for accuracy. The data analysis method employed in this study was designed to quantify the practical insights for risk factors gathered from the interviews, ensuring a structured and objective assessment of the variables. The process began with compiling all the Likert scale responses into a dataset. Each variable evaluated by the participants was assigned a score based on the five-point Likert scale responses. This mixed methods design allows for the integration of quantitative and qualitative findings, validating quantitative results through qualitative insights and identifying additional risk factors or contextual nuances that may not be evident in quantitative data alone. This approach ensures a robust and holistic analysis, aligning with the pragmatic philosophy underpinning this research. The integration of quantitative and qualitative methodologies provides a comprehensive understanding of bankruptcy risk factors in the shipping industry, particularly for SMEs, supporting the development of robust predictive models and practical strategies for managing bankruptcy risk.

4.4. Data Sampling

4.4.1. Target data

The conceptualization of corporate bankruptcy in the literature encompasses a range of definitions. Following Altman (1968) approach, many studies synonymously use terms such as liquidation, restructuring, and failure. According to these criteria, a company is classified within the bankruptcy sample if it: (i) files for bankruptcy, (ii) ceases operations or undergoes liquidation, or (iii) initiates a restructuring process. This research adopts the legal definition of bankruptcy for its objective criteria, which facilitate straightforward company classification (Charitou et al., 2013). Considering country-specific legal framework, bankruptcy definitions have conformed to the guidelines set by the Korea

Financial Services Commission, which characterized a bankrupt firm based on conditions such as (i) legal management, (ii) payment deferrals, (iii) transaction suspensions, (iv)liquidation procedures, (v) reorganization applications, (vi) composition confirmations, or (vii) bankruptcy filings.

Therefore, this study chooses bankrupt firms within Korean shipping industry based on criteria such as; (i) Registrations cancelled by the Korea Shipowners' Association, (ii) court receivership, (iii) restructuring process including merger or acquisition, or reorganizations.

In this binary framework, bankruptcy status in Korea is coded as a binary dependent variable, with '1' representing companies that meet the aforementioned bankruptcy criteria and '0' for all others deemed active or safe. The dataset includes a list of Korean shipping companies from 2001 to 2022, and Table 4.1 shows the annual bankruptcy rates of the Korean shipping industry as provided by official statistics (Korea Shipowners' Association, 2023). Firms consistently reporting losses over the last three years are excluded to avoid misclassification of still-operating firms as failing, addressing concerns over sustained unprofitability (Kim et al., 2020). Additionally, firms lacking comprehensive financial data for at least two years preceding registration cancellation are omitted to ensure adequate data for bankruptcy risk assessment (Visvanathan, 2021). This study also excludes any firm-year observation post-initial delisting, thus maintaining the integrity of the dataset.

Table 4.1. The annual bankruptcy rates of Korean shipping industry

Year	No. of shipping companies	No. of bankrupt companies	Bankruptcy Rate (%)
2001	29	1	3%
2002	30	2	7%
2003	33	2	6%
2004	35	3	9%
2005	41	2	5%
2006	49	4	8%
2007	55	7	13%
2008	50	3	6%
2009	52	6	12%
2010	89	9	10%
2011	97	14	14%
2012	90	14	16%
2013	93	9	10%
2014	104	24	23%
2015	120	29	24%
2016	111	18	16%
2017	107	12	11%
2018	115	10	9%
2019	115	10	9%
2020	117	9	8%
2021	131	10	8%
2022	133	12	9%
Total	1796	210	12%
Mean	81.6	9.5	11%

Source: Korean shipowners Association (2023)

4.4.2. Forecasting Horizon

Generally, Bankruptcy unfolds as a protracted process, with initial indicators often manifesting several months or years prior to the formal filing for bankruptcy (Berk et al., 2013). Over time, various definitions of insolvency, bankruptcy, failure, financial distress, financial difficulties, financial soundness, or financial health have been established, each according to the legal process of bankruptcy. This diversity in definitions introduces challenges in accurately identifying the onset of financial distress, as highlighted by Platt et al. (1994) and further emphasized by Philosophov et al. (2008), who noted that the legal bankruptcy date does not necessarily coincide with the actual start of financial difficulties.

The gap between the emergence of financial distress and its legal acknowledgment can extend up to three years, complicating the prediction of corporate bankruptcy (Bauer and Agarwal, 2014). Early indicators of bankruptcy include deteriorating financial conditions, underscoring the necessity for risk managers to utilize timely data for analysis (Mramor and Valentincic, 2003). However, due to the unpredictability of the actual bankruptcy date and the static nature of data collection, short-term forecasts often fall short in providing stakeholders with adequate response time (Yuen and Ko, 2018).

Consequently, the capacity for medium- to long-term forecasting is crucial. Financial institutions, for instance, depend on these predictions when assessing loan applications, considering the long-term viability of the borrowing entity (Chauhan et al., 2009). Given that a significant portion of business loans are structured with medium to long-term maturities, predictive models must be equipped to evaluate a firm's financial stability over the entirety of the loan period (Bărbuță-Mișu and Madaleno, 2020).

Signs of economic and financial strain typically emerge three years before bankruptcy, indicating the impending crisis (Virág and Nyitrai, 2014). Firms nearing bankruptcy exhibit reduced productivity and profitability and increasingly rely on external financing, signifying a heavy debt burden (Liu et al., 2020). Previous research has established that the ideal forecasting horizon for bankruptcy prediction models is typically one year, during which the predictive accuracy of financial variables remains high but significantly diminishes beyond this period (Jones, 2017). Nonetheless, the precision of these predictions over medium-term horizons (2-5 years) also holds substantial importance, as financial institutions bear the risk of credit until the culmination of a customer's debt period (Philosophov et al., 2008). Traditional models, primarily designed for short-term predictions, often experience a decline in accuracy when extended to medium-term forecasts due to the inherent instability of financial ratios and the impact of economic fluctuations (Cheng et al., 2018). Consequently, considering a forecasting horizon that encompasses long-term predictive capabilities is crucial for capturing bankruptcy risk factors across varied forecast ranges.

For this research, firms classified as bankrupt from 2001 to 2022, as outlined in Chapter 4.1, were selected with the prerequisite that complete information on all relevant variables be available at the time of bankruptcy and for preceding years. Bankruptcy prediction research typically involves selecting bankrupt entities from a range of years to amass an extensive dataset. Therefore, the analysis of each variable is contextualized relative to the bankruptcy event (t), with preceding years labelled as t−1, t−3, t−5. Literature suggests that a five-year period prior to the bankruptcy event is optimally useful for such analyses (Mai et al., 2019). This study utilizes the most recent financial statement available before the bankruptcy (t−1) for assessment, recognizing that financial information is often not accessible in the year a company's registration is terminated. Accordingly, the forecasting horizon for potential bankruptcy in this research is determined as the likelihood of a firm entering bankruptcy after periods of 1, 3, or 5 years. Such an approach is essential for distinguishing between bankrupt and active firms well in advance of bankruptcy, thereby affirming the model's overall statistical integrity and the robust discriminative capability of the chosen variables. Consequently, the analysis will scrutinize each variable's capacity to distinguish between bankrupt and active firms over 1, 3, and 5-year intervals before the bankruptcy event, aiming to identify variables that provide meaningful predictive insights beyond merely immediate-term indicators.

4.4.3. Explanatory variables

As initial variable selection process, a wide array of financial variables was gathered from the literature to assess their predictive relevance to bankruptcy. These variables were categorized into four principal groups: liquidity, leverage, profitability, and efficiency, in alignment with the Bank of Korea’s Financial Statement Analysis guidelines. Out of over

250 financial ratios scrutinized, 68 variables were selected based on their demonstrated predictive significance. Additionally, non-financial variables detailing firm-specific characteristics and macroeconomic factors were incorporated, resulting in a comprehensive set of 80 variables as presented in Table 4.2.

Table 4.2. The list of explanatory variables selected from literature review

Category	Variable	Description	Category	Variable	Description
Financial (Leverage)	Return on total assets	$\frac{\text{EBITDA}}{\text{Total Assets}}$	Financial (Profitability)	Return on Assets	$\frac{\text{Net Income}}{\text{Total Assets}}$
	Retained Earnings to Total Assets	$\frac{\text{Retained Earnings}}{\text{Total Assets}}$		Return on Equity	$\frac{\text{Net Income}}{\text{Shareholder's Equity}}$
	Leverage	$\frac{\text{Total Debt}}{\text{Total Assets}}$		Return on Sales	$\frac{\text{Net Income}}{\text{Sales}}$
	Liabilities to Total Assets	$\frac{\text{Total Liabilities}}{\text{Total Assets}}$		EBITD to Sales	$\frac{\text{EBITDA}}{\text{Sales}}$
	Equity to Assets	$\frac{\text{Shareholder's Equity}}{\text{Total Assets}}$		Sales to Current Assets	$\frac{\text{Sales}}{\text{Current Assets}}$
	Debt to Equity	$\frac{\text{Total Debt}}{\text{Shareholder's Equity}}$		Cash flow to Sales	$\frac{\text{Cash Flow}}{\text{Sales}}$
	Liabilities to Equity	$\frac{\text{Total Liabilities}}{\text{Shareholder's Equity}}$		Sales	Total Sales
	Gearing ratio	$\frac{\text{Shareholder's Equity}}{\text{Total Debt}}$		EBITDA to Interest Coverage	$\frac{\text{EBITDA}}{\text{Interest expenses}}$
	Equity to liabilities	$\frac{\text{Shareholder's Equity}}{\text{Total Liabilities}}$		Gross Profit to Assets	$\frac{\text{Gross Profit}}{\text{Total Assets}}$
	Liabilities	Total Liabilities		Net interest margin	$\frac{\text{Interest returns} - \text{Interest paid}}{\text{Average Assets}}$
	Current Liabilities to Equity	$\frac{\text{Current Liabilities}}{\text{Equity}}$		Sales to equity	$\frac{\text{Total Sales}}{\text{Shareholder's Equity}}$
	EBITDA to Liabilities	$\frac{\text{EBITDA}}{\text{Total Liabilities}}$		Net Income	$\text{Total Revenue} - \text{Total Expenses}$
	EBITDA/EV	$\frac{\text{EBITDA}}{\text{Equity}}$		Operating Return on Assets	$\frac{\text{Operating Income}}{\text{Total Assets}}$
	EBITDA	$\text{Net Income} + \text{Interest} + \text{Taxes} + \text{Depreciation} + \text{Amortization}$		Return on investment	$\frac{\text{Net Income}}{\text{Investment}}$
	Long-term Liabilities to Assets	$\frac{\text{liabilities} - \text{current liabilities}}{\text{Total Assets}}$		Sales to Liabilities	$\frac{\text{Total Sales}}{\text{Total Liabilties}}$
	Net Income to Liabilities	$\frac{\text{Net Income}}{\text{Total Liabilities}}$		Gross Profit to Current Liabilities	$\frac{\text{Gross Profit}}{\text{Current Liabilities}}$
	Working Capital	$\text{Current Assets} - \text{Current Liabilities}$		Gross Profit to Liabilities	$\frac{\text{Gross Profit}}{\text{Total Liabilities}}$
	Current Assets to Liabilities	$\frac{\text{Current Assets}}{\text{Total Liabilities}}$		Gross Profit to Sales	$\frac{\text{Gross Profit}}{\text{Total Sales}}$
	EBITDA to Current Liabilities	$\frac{\text{EBITDA}}{\text{Current Liabilities}}$		ΔSales	$\frac{\text{Total sales}_t - \text{Total Sales}_{t-1}}{\text{Total Sales}_{t-1}}$
	Long-term Liabilities to Equity	$\frac{\text{liabilities} - \text{current liabilities}}{\text{Shareholder's Equity}}$		Sales to Current Liabilities	$\frac{\text{Total Sales}}{\text{Current Liabilities}}$
Financial (Liquidity)	Current Ratio	$\frac{\text{Current Assets}}{\text{Current Liabilities}}$	Financial (Efficiency)	Asset turnover	$\frac{\text{Total Sales}}{\text{Total Assets}}$
	Working Capital to Assets	$\frac{\text{Current Assets} - \text{Current Liabilitiesl}}{\text{Total Assets}}$		Working Capital to Sales	$\frac{\text{Working Captial}}{\text{Total Sales}}$
	Cash Assets Ratio	$\frac{\text{Cash}}{\text{Total Assets}}$		Current Assets to Sales	$\frac{\text{Current Assets}}{\text{Total Sales}}$
	Cash Ratio	$\frac{\text{Cash}}{\text{Current Liabilities}}$		Inventory to Assets	$\frac{\text{Inventory}}{\text{Total Assets}}$
	Current Assets to Assets	$\frac{\text{Current Assets}}{\text{Total Assets}}$		Current Liabilities to Sales	$\frac{\text{Current Liabilities}}{\text{Total Sales}}$
	Current Liabilities to Assets	$\frac{\text{Current Liabilities}}{\text{Total Assets}}$		Average Inventory Turnover	$\frac{\text{Average Inventories}}{\text{Total Sales}}$
	Quick Ratio	$\frac{\text{Cash} + \text{accounts receivable}}{\text{Current liabilities}}$		Inventory turnover	$\frac{\text{Average Inventories}}{\text{Total Sales}}$
	Cash to Debt Ratio	$\frac{\text{Cash}}{\text{Total Debt}}$		Operating Margin	$\frac{\text{Operating Income}}{\text{Sales}}$
	Shareholder's Equity	$\text{Total Assets} - \text{Total Liabilities}$	Non-financial	Size	*Proxied by Log* (Total Assets)
	Current Liabilities Ratio	$\frac{\text{Current Liabilities}}{\text{Total Liabilities}}$		Age	*Age of firm*
	Current Assets	$\text{Cash} + \text{Accounts Receivable} + \text{Inventory}$		Number of employees	*Number of full-time employees*
	Cash Sales Ratio	$\frac{\text{Cash}}{\text{Total Sales}}$		GDP	$\frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}$
	Current Liabilities to Current Assets	$\frac{\text{Current Liabilities}}{\text{Current Assets}}$	Macro-economics (Shipping Index)	Oil Price	*Current US$ price of Brent oil*
	Working Capital to Current Assets	$\frac{\text{Working Captial}}{\text{Current Assets}}$		Type of Operator	*Dummy value of Type: 1 (Container), 2 (Tanker), 3 (Bulk)*
	Quick Assets to Total Assets	$\frac{\text{Cash} + \text{accounts receivable}}{\text{Total Assets}}$		Chartering Cost	*Annual Voyage Chartering Cost*
	Working Capital to Equity	$\frac{\text{Current Assets} - \text{Current Liabilities}}{\text{Shareholder's Equity}}$		Freight Rate	*Average Transportation Revenue*
	Working Capital to Current Liabilities	$\frac{\text{Current Assets} - \text{Current Liabilities}}{\text{Current Liabilities}}$		Time Charter Rate	*Time charter rate index*
	Cash to Current Assets	$\frac{\text{Cash}}{\text{Current Assets}}$		Container Freight Index Rate	*Shanghai Containerized Freight Index*
	Retained Earnings to Current Liabilities	$\frac{\text{Retained Earnings}}{\text{Current Liabilities}}$		Baltic Dry Index (BDI),	*Annual Baltic dry index*
	ΔTotal Asset	$\frac{\text{Total Assets}_t - \text{Total Assets}_{t-1}}{\text{Total Assets}_{t-1}}$		IRONSTEEL	*Dow Jones U.S. Iron & Steel Index*

The framework established by Altman (1968), categorizing financial variables into leverage, liquidity, profitability, efficiency, and market segments, has been a cornerstone in bankruptcy research, providing a robust method to differentiate firms based on their bankruptcy risk. This study adopted Altman's approach to recognize its effectiveness in depicting the financial health of a company. Those explanatory variables in Table 4.2, have selected which already been extensively covered in the literature review as the initial variable section process.

In summary, leverage ratios such as the current ratio and gearing ratio are crucial in assessing a firm's financial risk and its ability to secure external financing (Mramor and Valentincic, 2003). Higher equity capital typically reduces financial risk, enhancing the probability of securing external funds (Shi and Li, 2019). Liquidity measures, including working capital and the current assets to liabilities ratio, provide insight into a company's short-term financial health (Lyandres and Zhdanov, 2013). Firms with stronger liquidity, indicated by higher working capital, are better positioned to avoid bankruptcy (Lukason and Andresson, 2019). This highlight significance of these factors in assessing the bankruptcy risk within the Korean shipping sector, predominantly comprised of small to medium-sized firms.

Profitability metrics, such as Return on Assets (ROA), Return on Equity (ROE), and Earnings Before Interest, Taxes, Depreciation, and Amortization (EBITDA), reflect the efficiency with which a company utilizes its equity and assets to generate returns (Lukáč et al., 2022). These variables are crucial for benchmarking the profitability of firms within similar sectors. Higher ROA levels correlate with a diminished likelihood of bankruptcy, while increased ROE and EBITDA levels indicate stronger financial health, particularly relevant in the asset-intensive shipping industry (Alexandridis et al., 2020).

Operational efficiency is assessed using the total asset turnover ratio, which measures a company's ability to generate revenue from its assets (El-Masry et al., 2010). Higher ratios indicate better operational efficiency. Previous findings highlight that non-bankrupt companies often present lower leverage, enhanced profitability, and superior liquidity metrics, alongside larger retained earnings and elevated past abnormal returns.

Given the predominance of small and medium-sized enterprises (SMEs) within the Korean shipping sector, company size is included as a key variable. Smaller companies face greater challenges in accessing external funding, suggesting a correlation between company size and bankruptcy risk (Kou et al., 2021). Larger firms are generally better equipped to withstand financial distress, implying a negative correlation between firm size and bankruptcy likelihood (Tobback et al., 2017). Additionally, the vulnerability of younger companies to bankruptcy, due to financing limitations or equity shortages, underscores the significance of corporate age as a risk factor.

In the context of bankruptcy prediction for the shipping industry, it is essential to consider non-financial factors such as industry-specific characteristics and macroeconomic indicators. Freight rates, shipping market indices, and oil prices significantly impact the bankruptcy risk of shipping companies (Clintworth et al., 2021). Higher oil prices and volatile freight rates can increase the financial instability of shipping firms.

By integrating both financial and non-financial variables, this study aims to provide a comprehensive analysis of bankruptcy risk in the Korean shipping industry, with a particular focus on SMEs. This approach enhances the robustness and accuracy of the bankruptcy prediction models, offering valuable insights for stakeholders. To refine this selection and gain deeper insights, semi-structured interviews were conducted with stakeholders from various sectors within the Korean shipping industry, including shipping companies, research institutions, universities, and government organizations. These interviews aimed to capture the industry's perspective and unearth previously unconsidered aspects. The selection of variables was guided by their explanatory power proven through past research and the practical experience of industry experts.

4.5. Interview

4.5.1. Conducting the Interviews

The literature review has identified several research gaps, particularly the absence of bankruptcy risk factors specific to the shipping industry. To bridge this gap, this study incorporates insights from industry practitioners through interviews, aiming to collect nuanced, relevant perspectives that augment the literature findings. Interviews are a potent tool in business research, providing a direct avenue to acquire data that may not be readily available in published formats such as journals, books, or online resources (Saunders et al., 2019). Moreover, interviews can capture specific, additional viewpoints, including the experiential knowledge of industry insiders, providing a depth of understanding that surpasses what can be obtained through surveys (Adams et al., 2014). This study opts for semi-structured interviews among the various interview types—structured, semi-structured, and unstructured. Structured interviews, with their standardized questions, tend to limit the depth of insight that can be gained, offering little flexibility to explore the interviewees' unique perspectives (Bell et al., 2022). While unstructured interviews allow for considerable freedom, they risk veering off-topic due to their open-ended nature (Kovalainen and Eriksson, 2015). Semi-structured interviews, however, strike a balance by providing a predefined set of topics while allowing adaptability in question sequencing and phrasing, accommodating the flow of conversation (Saunders et al., 2019). This flexibility makes semi-structured interviews particularly well-suited for exploring the "what" and "how" questions central to understanding the broaden range of risk factors associated with bankruptcy in the shipping industry.

Semi-structured interviews were chosen for their flexibility and focus, allowing for a predefined set of topics while accommodating the flow of conversation (Saunders et al., 2019). The interviews, planned to last approximately 60 minutes, were recorded for accuracy and conducted face-to-face to foster a conducive environment for in-depth conversation. The interview questionnaire was designed to align with the research objectives, focusing on bankruptcy risk factors specific to the shipping industry. The questions were distributed to the selected participants in advance via email. This advance distribution allowed participants ample time to prepare thoughtful and well-informed responses. By sending the questionnaire beforehand, participants could reflect on their experiences and insights, leading to more comprehensive and detailed discussions during the interviews.

The topics discussed during the interviews are provided in Table 4.3. During the interview, the interviewees would be asked to evaluate the importance of each of variables by a five-point Likert scale ranging from 1 (not at all important) to 5(very important) as shown in Appendix, which provide reliable result without confusion. From this evaluation process, variables above the mean score of 3.5 would be chosen, to ensure over moderate importance of model incorporating practical relevance and significant impact. This systematic approach ensures that the selected variables are not only grounded in theoretical and empirical research but also validated through the practical and experiential insights of industry experts (Xue and Hauskrecht, 2017).

Table 4.3. Interview questions

Topic	Question
Personnel Background	· Position in the organization · Years being worked · Type of business where belong
Bankruptcy risk factors	· Which risk factor do you think is particularly important in the list above? And why? · Which risk factor are important that are not included in the list above? And why? · How can risk factor can differ depending on size of shipping company?
Bankruptcy Evaluation	· Why do you evaluate bankruptcy risk of shipping company? · How do you think bankruptcy prediction can contribute to shipping company as managerial implications?

***Source*: Author**

The data analysis method employed in this study was designed to quantify the qualitative insights gathered from the interviews, ensuring a structured and objective assessment of the variables. The process began with compiling all the Likert scale responses into a dataset. Each variable evaluated by the participants was assigned a score based on the five-point Likert scale responses.

Initially, all responses from the Likert scale were systematically compiled into a comprehensive dataset. This step involved organizing the data to facilitate easy access and analysis. For each variable, the mean score was calculated to determine its overall importance. This involved averaging the scores given by all participants for each variable. Then, the variables were then ranked based on their mean scores. This ranking helped identify the most and least important factors as perceived by the industry experts. Variables with mean scores above 3.5 were identified as significant. These variables were considered relevant and impactful for further analysis in the study. This quantification process allowed the research to systematically evaluate the importance of each variable, ensuring that the selected variables were not only grounded in theoretical and empirical research but also validated through practical and experiential insights from industry experts. By converting qualitative feedback into quantifiable data, the study ensured a rigorous and objective analysis, aligning the qualitative insights with the quantitative needs of the research.

4.5.2. Selection of Interview Participants

The selection of interview participants meticulously targets a comprehensive spectrum of expertise within the South Korean shipping industry all with a minimum of five years of experience in their current roles, ensuring a broad-based and insightful exploration of the sector. The methodology for selecting interviewees in this study leverages their distinctive insights, expertise, or positions within their respective organizations or events, aiming to enrich the research with diverse perspectives (Adams et al., 2014). Participants were selected based on their distinctive insights, expertise, and positions within their respective organizations or events. The selection strategy aimed to enrich the research with diverse perspectives, targeting individuals directly engaged in operational facets such as port operations, ship management, or cargo logistics from not only large shipping companies but also SMEs. This approach ensured the collection of primary data encapsulating practitioners' viewpoints on industry-specific challenges.

In addition, the study included management-level professionals from governmental agencies or organizations. These individuals provided strategic insights into the broader implications of bankruptcy within the business landscape and the shipping industry, contributing to an understanding of policy enforcement, governance, and network-level dynamics that influence bankruptcy trends. Academic researchers and experts from universities or research institutions were also integral to the interviewee pool. Their participation ensured the incorporation of an academic lens, facilitating the inclusion of cutting-edge research findings and theoretical underpinnings in the variable selection process. This balanced approach amalgamated practical insights with scholarly research, enhancing the comprehensiveness and relevance of the study's findings.

Table 4.4 presents the diverse backgrounds of the respondents involved in this study, highlighting variations in the type of business within the shipping industry, work experience, and their respective positions. The distribution of respondents includes representatives from large shipping companies and SMEs, universities, and government authorities, ensuring a comprehensive perspective on the industry.

Table 4.4. Demographic data of respondents

	Category		Num	%
Type of Business	Shipping company	Large	2	15.4
		SMEs	2	15.4
	University		4	30.8
	Government authority		5	38.4
Working Period in current position (Years)	1-5 years		3	23.1
	6-10 years		5	38.4
	11-20 years		4	30.8
	More than 21 years		1	7.7
Position	Staff		3	23.1
	Middle Manager		7	53.8
	Senior Manager		3	23.1
	Director/CEO		0	0

Specifically, shipping companies are represented by 31% of the sample, with two respondents from large firms and two from SMEs, each comprising 15.4% of the total respondents. Universities also represent 31% of the sample, with four individuals. Additionally, government authorities and research institutions affiliated with the Korean government constitute 38.4% of the respondents, totalling five individuals. This diverse array of respondents ensures a well-rounded perspective on the shipping industry, leveraging their varied insights and experiences to enrich the research findings.

A significant majority of the respondents, 76.9%, have held their current positions for more than six years, indicating a deep-rooted understanding and extensive experience within the Korean shipping industry. Furthermore, 38.4% of the interviewees have accumulated over ten years of career experience, demonstrating their comprehensive knowledge and expertise in the field. Regarding their roles within their organizations, most respondents (53.8%) occupy middle management positions. Staff and senior managers are equally represented in the study, with three individuals from each category.

In conducting this case study, high ethical standards are required because there are many factors related to human affairs and real life. This included informed consent, ensuring privacy and confidentiality of data, and preventing any potential harm or deception. These ethical considerations form the backbone of the interview process, guiding it towards integrity and respect for all participants. Preparatory work for the interviews involves the careful design of questions aligned with the research objectives, tailored to the unique backgrounds of each participant to facilitate focused and substantive discussions. Participants are contacted in advance to schedule the interviews at mutually convenient times and venues, highlighting the research's participatory and respectful approach. The purpose and relevance of the interview to the research questions are clearly communicated to participants, ensuring transparency and informed consent. Interview recordings are proposed with explicit participant consent, safeguarding their autonomy and privacy. Acknowledging varying personal schedules and preferences, some interviews might be conducted via phone or email, accommodating a wide array of insights while respecting each interviewee's circumstances. Through this ethical and methodical approach to data collection, the study aims to draw upon the rich experiences and insights of interviewees, contributing valuable perspectives to the research. The information shared by participants is expected to significantly enhance the study's findings, providing a comprehensive understanding of the subject matter.

4.6. The research models

By applying two advanced models proved their superior predictive ability from literature review, XGBoost and LSTM, this research aims to harness the strengths of both models to achieve high predictive accuracy and robustness in bankruptcy prediction for the shipping industry. This combined approach addresses the limitations of traditional statistical methods and enhances the interpretability of machine learning models through the application of explainable AI techniques. The complementary nature of these models allows for comprehensive analysis, effectively capturing the complex dynamics of bankruptcy prediction.

4.6.1. Extreme gradient boosting

The ensemble method, specifically through bagging and boosting techniques, is crucial for credit scoring and bankruptcy prediction models. Extreme Gradient Boosting (XGBoost) is notable for its efficiency, scalability, and effectiveness in implementing gradient-boosted decision trees. XGBoost would be employed to analyze high-dimensional and nonlinear datasets, incorporating numerous predictors to accurately reflect the multifaceted nature of corporate bankruptcy (Carmona et al., 2019). This method enhances prediction accuracy through sequential parameter optimization, significantly reducing the risk of overfitting (Chen and Guestrin, 2016). XGBoost's ability to iteratively refine predictions based on prior errors makes it particularly effective in managing outliers, nonlinearities, and missing data (Bentéjac et al., 2021). This robustness ensures stable model performance, even in the presence of multicollinearity and heteroscedasticity. The application of XGBoost aligns with research objectives by providing a powerful tool for evaluating the explanatory ability of various financial and non-financial variables in predicting bankruptcy.

Proposed by Chen et al. (2015), XGBoost is acclaimed for its performance in various machine learning tasks. Its ability to manage complex data structures and uncover intricate patterns within datasets renders it indispensable for tasks requiring high accuracy in predictive modelling, including the critical domain of bankruptcy prediction. XGBoost uses K additive function $fk\ (x)$ to approximate the function of $Fk\ (x)$, it is written as follow:

$$\boldsymbol{F(x)} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{f_k(x), f_k \in F} \quad \textbf{(4.1)}$$

Source: Chen et al. (2015)

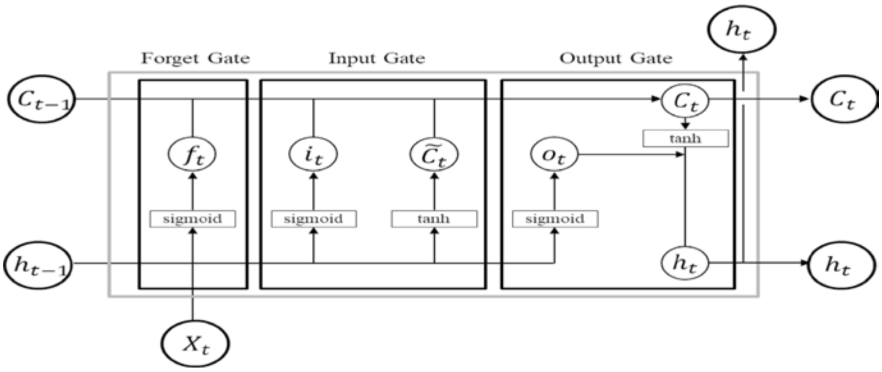
where K is the number of trees, $f_k(x)$is a function family F, and F is the set of all possible classification and regression trees (CART). Markedly, XGBoost utilizes a specific form of a base learner: $f_k(x)$ is a CART and can be denoted the as $\omega_{q\,(x)}, q\epsilon$, Where T represents the number of leaves in the tree, q denotes the decision rules of the tree, while ω represents a vector indicating the sample weight of leaf nodes.

In XGBoost, to find the minimum $f_k(x)$, the objective function is optimized using gradient descent, primarily concentrating on first-order gradient statistics. Despite its effectiveness, one limitation is the considerable time it takes to optimize parameters. A notable application of XGBoost in classification models is by Xia et al. (2017), who developed a sequential ensemble credit scoring model utilized XGBoost. This approach demonstrated superior performance on benchmark credit scoring datasets compared to other methods, showcasing XGBoost's strength in handling complex predictive tasks.

4.6.2. Long Short-Term Memory

Long Short-Term Memory (LSTM), introduced by Hochreiter and Schmidhuber (1997) addresses the long-term dependency problem inherent to traditional RNNs. Traditional RNNs struggle with learning from inputs separated by long time steps, often forgetting earlier inputs as the gap between input and output widens. LSTM's core innovation is the cell state, which carries information across long sequences, effectively retaining data from earlier inputs. Illustrated in Figure 4.3, the LSTM architecture includes a repeating module composed of multiple neural network layers that interact with each other. It employs memory cells C_t to modulate the hidden state h_t at each time step t, incorporating forget, input, and output gates. These gates enable to LSTM to maintain and access information over extended period, overcoming the limitation of simple RNNs by preventing long-term dependencies (Vochozka et al., 2020).

Figure 4.3. Logical structure of Long short-term memory



***Source*: Hochreiter and Schmidhuber (1997)**

The forget gate in an LSTM network plays a critical role in managing the cell state's information across long sequences. It decides how much of the existing information in the cell state should be retained or discarded as new input comes in. The operation of the forget gate is governed by the equation 4.2:

$$f_t = sigmoid\left(W_{hf}h_{t-1} + W_{xf}x_t + b_f\right) \quad \textbf{(4.2)}$$

where W_{hf} and W_{xf} are weight matrices, h_{t-1} represents the output of the recurrent neuron from the previous time step, x_t is a vector of explanatory variables from the current time step, and b_f denotes a bias term.

Next step, the input gate layer determines the amount of newly entered information that is used to update the cell state, which decide whether to add new information to the memory. First, the input gate decides which information to keep, guided by the following equations 4.3 and 4.4:

$$i_t = sigmoid\ (W_{hi}h_{t-1} + W_{xi}x_t + b_i) \quad \textbf{(4.3)}$$

$$C_t = sigmoid\ (W_{hC}h_{t-1} + W_{xC}x_t + b_C) \quad \textbf{(4.4)}$$

Where W_{hi}, W_{xi}, W_{hC} and W_{xC} are matrices of weights. b_i and b_c are biases that need to be learned during training process. Then current state of cell can be determined by input gate as following equation 4.5:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \qquad \textbf{(4.5)}$$

Finally, the output gate layer calculates the output h_t, using the following equations 4.6 and 4.7.:

$$o_t = sigmoid\left(W_{ho} h_{t-1} + W_{xo} x_t + b_f\right) \qquad \textbf{(4.6)}$$

$$h_t = o_t \times \tanh C_t \qquad \textbf{(4.7)}$$

Where W_{ho} and W_{xo} are matrices of weights and $b_o$is a bias. The activation function for all hidden layers was chosen as tanh, as it is a recommended sigmoid activation function for hidden layers by Bottou (2012).

LSTM models are particularly suited for long-term forecasting due to their design, which preserves relevant information over extended periods while maintaining short-term data integrity (Alam et al., 2021a). This capability is crucial for financial time series analysis. LSTMs autonomously discern optimal patterns from the dataset, eliminating the need for manual feature selection (Cha and Kang, 2018). Their directional nature, tailored to the sequential character of time series data, enhances the model's ability to capture variability within financial and non-financial variables (Becerra-Vicario et al., 2020. Utilizing LSTM supports our research objectives by providing a robust method for analysing extensive time-series datasets and capturing long-term dependencies crucial for bankruptcy prediction.

4.7. Preprocessing Procedure

4.7.1. Oversampling Process

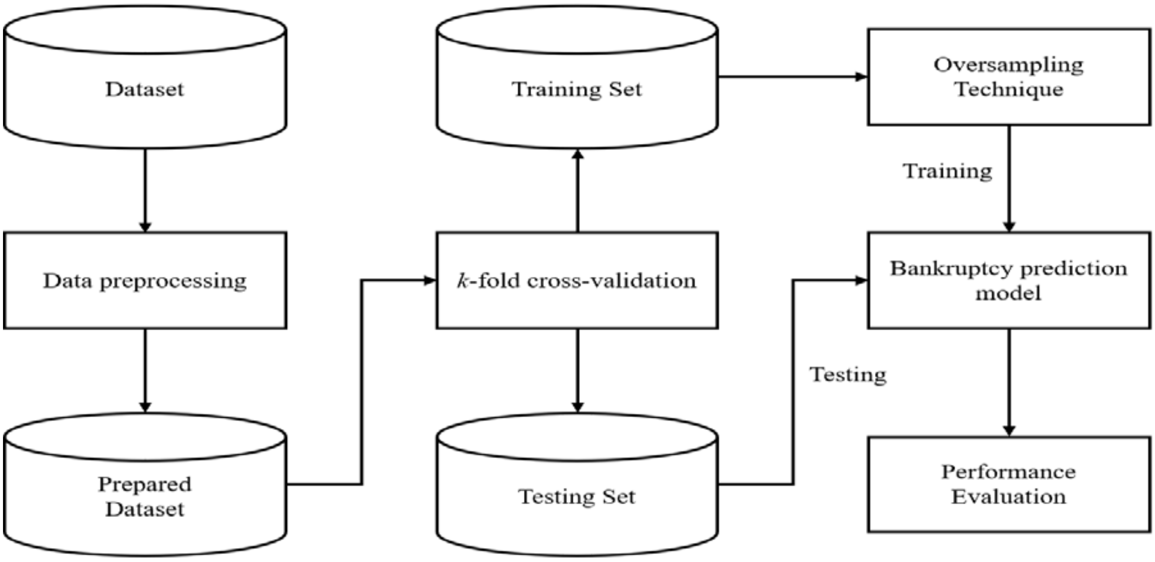
In bankruptcy prediction research, balancing the dataset between bankrupt and non-bankrupt firms is a critical methodological decision. Classical studies often employ equal samples of bankrupt and non-bankrupt firms to mitigate the imbalanced nature of real-world data, where non-bankrupt firms vastly outnumber bankrupt ones (He and Garcia, 2009). This approach aims to avoid the suboptimal classification bias towards the majority class (i.e., non-bankrupt firms) and improve the model's ability to predict bankruptcy events accurately. However, equalizing the samples might introduce bias, as the non-bankrupt sample may not accurately represent the broader population (Luo et al., 2020).

To address these concerns, this study incorporates all available data on bankrupt and non-bankrupt Korean small and medium-sized shipping companies, utilizing their annual financial reports. By considering the entire population, this approach reflects a more comprehensive and realistic scenario for financial institutions conducting credit analysis, which typically does not involve pre-selecting corporates based on their bankruptcy status (Kou et al., 2021). To manage the inherent data imbalance, this research will apply a weighting scheme to equalize the influence of both bankrupt and non-bankrupt samples in the analysis. This methodological choice aims to harness the predictive power of the dataset while minimizing the potential biases associated with imbalanced data.

To address the challenge of highly imbalanced datasets in bankruptcy prediction, this study employs the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE improves the representation of the minority class (bankrupt firms) by creating synthetic samples until the minority class size equals that of the majority class (non-bankrupt firms). This method proves especially beneficial in situations where the minority class is underrepresented, guaranteeing that the prediction model accurately addresses significant bankruptcy instances without bias from an imbalanced dataset (Sun et al., 2020). Unlike under-sampling methods, which reduce the size of the majority class potentially at the cost of losing informative data, oversampling retains all original data and enhances the dataset's balance without eliminating valuable information (García et al., 2019). However, a potential downside of oversampling, particularly with techniques like SMOTE, is the risk of introducing noise through the creation of closely similar synthetic samples. This can sometimes degrade the model's classification performance by focusing excessively on these artificially generated instances (Fernández et al., 2018). To mitigate this risk, advanced machine learning algorithms such as Extreme Gradient Boosting (XGBoost), which are adept at handling noise and overfitting, are particularly suitable for datasets augmented with SMOTE.

As illustrated in Figure 4.4, In the training phase of the prediction process, SMOTE will be applied to ensure a balanced distribution of classes, allowing the classification model to learn from a dataset where bankrupt and non-bankrupt cases are nearly equal in number. This balanced approach is crucial for improving binary classification performance, as models trained on imbalanced data tend to bias towards the majority class, neglecting the critical minority class instances. By implementing SMOTE in conjunction with robust machine learning algorithms, this study aims to develop a more effective and accurate bankruptcy prediction model that accounts for the challenges of imbalanced datasets.

Figure 4.4. Flowchart for the oversampling framework



***Source* : Smiti and Soui (2020)**

4.7.2. Missing data

Addressing missing values in datasets is a critical challenge in bankruptcy prediction research, where financial statement data might be incomplete. The presence of missing data is essentially an information loss, leading to decreased analytical efficiency. Broadly, there are two approaches to manage missing data: preprocessing methods and parallel methods (Murphy, 2022). Preprocessing methods involve strategies such as substituting missing values with the most frequent value of the feature, the mean for numerical features, or values from similar cases. These techniques are straightforward and allow for the application of any binary classification model post-imputation (Kim et al., 2020).

An example of an advanced preprocessing method is the Singular Value Decomposition (SVD) technique for imputing missing values, as suggested by Jones (2017). Despite its sophistication, Alam et al. (2021a) observed that SVD imputation has limited impact on enhancing the overall performance of classifiers.

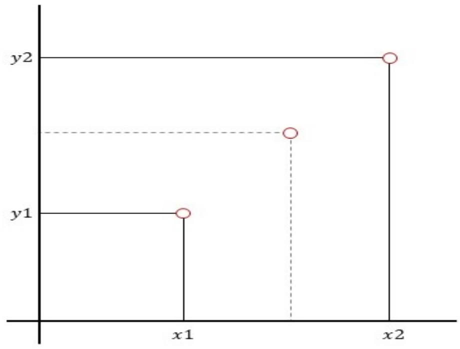
Parallel methods, on the other hand, integrate the treatment of missing values within the learning algorithm itself. This approach modifies the learning algorithms to directly handle missing data without prior imputation. It is notably effective with rule induction-based classification models, such as decision trees and learning from examples module, which inherently address missing data within their classification process (Tobback et al., 2017). Given the diverse nature of data sets and the absence of a one-size-fits-all solution, the choice between preprocessing and parallel methods for handling missing data should be tailored to the specifics of each dataset (Luo et al., 2020). The decision hinges on factors such as the nature of the missing data, the distribution of the dataset, the expected impact on predictive performance, and the computational efficiency of the chosen method.

Handling missing values is a critical aspect of preparing bankruptcy datasets, particularly for SMEs where financial statements may often be incomplete (Roy and Shaw, 2021). The challenge lies in addressing missing data without losing valuable information that could influence the accuracy of bankruptcy prediction. Direct deletion of missing values, while straightforward, can reduce the sample size and potentially bias results (García-Laencina et al., 2010). In machine learning applications, preserving the integrity of data without compromising its predictive capability is paramount.

To this end, this study adopts a two-pronged approach based on the extent of missing data within a dataset. For instances where a company's data exhibits more than 1% missing values, interpolation methods are employed for imputation. Linear interpolation, despite its simplicity, proves to be an effective method for this purpose (Qu et al., 2019). As shown in Figure 4.5, It involves replacing missing values with the mean of adjacent values, ensuring a smooth transition and minimizing information loss. For datasets with minimal missingness (less than 1%), a more conservative approach is taken by substituting missing values with those from the previous year. This method assumes that small gaps in data do not significantly impair the model's predictive power and allows for a practical solution to maintaining data continuity (Zhou and Lai, 2017). This approach is particularly justifiable in financial analysis, where year-on-year data often exhibits stability or gradual change, making previous year values a reasonable proxy for missing entries.

Therefore, those methods ensure that the dataset remains robust and comprehensive, reducing the risk of sample bias while maintaining the quality of predictive analysis. By carefully selecting imputation techniques tailored to the degree of missingness, this study aims to optimize the dataset's utility for accurate bankruptcy prediction without compromising the reliability of its findings.

Figure 4.5. The linear interpolant method between two given known points (x1, y1) and (x2, y2)



***Source* : Papana and Spyridou (2020)**

4.7.3. Outliner & Skewness

Data preprocessing is a critical phase in data mining and knowledge discovery, essential for enhancing the quality of data analysis. This process addresses the challenges posed by large datasets, which can lead to high memory usage, decreased processing speed, and increased sensitivity to noise, potentially compromising the performance of mining algorithm (Nyitrai and Virág, 2019). The primary goal of preprocessing is to remove or correct data that may misrepresent the true patterns in the dataset, such as noisy data or unrepresentative features.

To standardize the data, this study will utilize Box-Cox transformations by Guerrero and Johnson (1982), which enable to significantly reduce skewness in data distributions, thereby improving the symmetry and, consequently, the analytical results. The Box-Cox transformation, however, requires all input values to be positive. In cases where the dataset includes negative values, this study will adjust each feature by adding a constant to shift all values to the positive range before applying the Box-Cox transformation (Atkinson et al., 2021).

$$\boldsymbol{w_t} = \begin{cases} \log (\boldsymbol{y_t}) & (\boldsymbol{\lambda} = \boldsymbol{0}) \\ \dfrac{(\boldsymbol{y_t^{\lambda}} - \boldsymbol{1})}{\boldsymbol{\lambda}} & (\boldsymbol{\lambda} \neq \boldsymbol{0}) \end{cases} \tag{4.8}$$

***Source*: Guerrero and Johnson (1982)**

In this study, handling outliers is crucial due to the diverse business stages and periods over which the sampled companies have declared bankruptcy. To mitigate the potential underestimation of classifier performance caused by outliers, this research adopts a winsorization strategy. Specifically, independent variables will be winsorized at the 1st and 99th percentiles. This method effectively limits extreme values from skewing the data analysis, balancing the need for outlier management without overly distorting the dataset's natural variation (Nyitrai and Virág, 2019). This approach ensures that the analysis remains robust and reflective of underlying trends, despite the varying lengths and business cycle stages of the bankruptcy sample.

In bankruptcy prediction, there is a significant imbalance between the number of bankruptcy and non-bankruptcy datasets due to the rarity of corporate bankruptcy events. Therefore, it may overestimate the true error rate optimistically, as the best model is specifically tailored to fit one sub-sample (Dzik-Walczak and Odziemczyk, 2021). The problem of data imbalance, characterized by a significant difference in the number of instances between classes, is addressed through K-fold cross-validation. This technique ensures the model's ability to generalize across different data samples by dividing the dataset into ten equal parts, training on K-1 and testing on one, rotating until each part has been used for testing (Yu et al., 2014). In general, 5-fold cross-validation would be selected, which is quite reliable and sufficient for estimating true error rates (Smiti and Soui, 2020). This method not only assesses the model's predictive accuracy comprehensively but also reduces the potential bias from uneven class distributions common in bankruptcy prediction datasets. By implementing these strategies, the study aims to develop a more reliable and effective bankruptcy prediction model, capable of accurate financial data analysis.

4.8. Evaluation Metrics

In bankruptcy prediction models trained on imbalanced datasets, where instances of bankruptcy are relatively uncommon, reliance on accuracy as the sole metric can lead to misleading conclusions. This is particularly true in scenarios where incorrectly predicting a firm's bankruptcy (false negative) carries more severe implications than wrongly identifying a firm as solvent (false positive). As such, an overemphasis on accuracy can result in an overly optimistic evaluation of a model's true predictive capability, especially in datasets predominantly composed of non-bankrupt entities (Branco et al., 2016).

To address this issue, this study proposes employing a comprehensive set of evaluation metrics: sensitivity, specificity, Type I error, Type II error, and the Area Under the Receiver Operating Characteristic Curve (AUC). Each metric provides valuable insights into different facets of the model's performance, enabling a more thorough assessment of its predictive accuracy within the domain of bankruptcy prediction (Kumar and Ravi, 2007).

The foundation of this evaluation framework is based on the confusion matrix, outlined in Table 4.5, which classifies the model's predictions into four distinct outcomes: true positives, true negatives, false positives, and false negatives. This categorization facilitates the calculation of sensitivity (the true positive rate) and specificity (the true negative rate). Sensitivity assesses the model's capability to accurately identify actual bankruptcies in Equation 4.9, essential for preventing missed alerts of financial distress (Ooghe and De Prijcker, 2008). Specificity, on the other hand, quantifies the model's accuracy in recognizing solvent firms in Equation 4.10, thus reducing the likelihood of mistakenly classifying financially healthy firms as at risk (Kamal et al., 2021).

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{4.9}$$

$$Specificity = \frac{True\ Negative}{False\ Positive + True\ Negative} \tag{4.10}$$

Table 4.5. Confusion matrix for bankruptcy prediction

		Predicted Value		
		0 (non-bankruptcy)	**1 (bankruptcy)**	
Actual Value	**0 (non-bankruptcy)**	True Positive (TP)	False Negative (FN)	TP+FN
	1 (Bankruptcy)	False Negative (FP)	True Negative (TN)	FP+TN
		TP+FP	FN+TN	Total

***Source*: Hastie et al. (2009)**

Adopting a multi-metric approach to evaluation addresses the challenges posed by imbalanced datasets, enabling a more accurate depiction of the model's effectiveness in bankruptcy prediction. This methodology ensures the reliability of the model's outputs, making them valuable for decision-making processes within financial institutions by providing a comprehensive overview of the model's ability to distinguish between bankrupt and non-bankrupt firms accurately (Lee and Choi, 2013).

In terms of bankruptcy prediction models that are often confronted with imbalanced datasets, it is crucial to employ metrics that accurately reflect the data distribution. Traditional metrics, which primarily focus on the overall accuracy, may not adequately address the imbalance between the classes, particularly when the instances of bankruptcy are far less frequent than those of non-bankruptcy. To navigate this challenge, this study emphasizes metrics based on the rates of true positives (TPR) and true negatives (TNR), alongside their complements including the false positive rate (FPR = 1 - TNR) and the false negative rate (FNR = 1 – TPR). The FPR and FNR are critical as they measure the likelihood of Type I and Type II errors respectively, with Type I errors misclassifying bankrupt firms as non-bankrupt in Equation 4.11, and Type II errors doing the opposite as shown in Equation 4.12. (Ooghe and De Prijcker, 2008).

$$Type\ \ I\ Error = \frac{False\ Positive}{False\ Positive + True\ Negative} \tag{4.11}$$

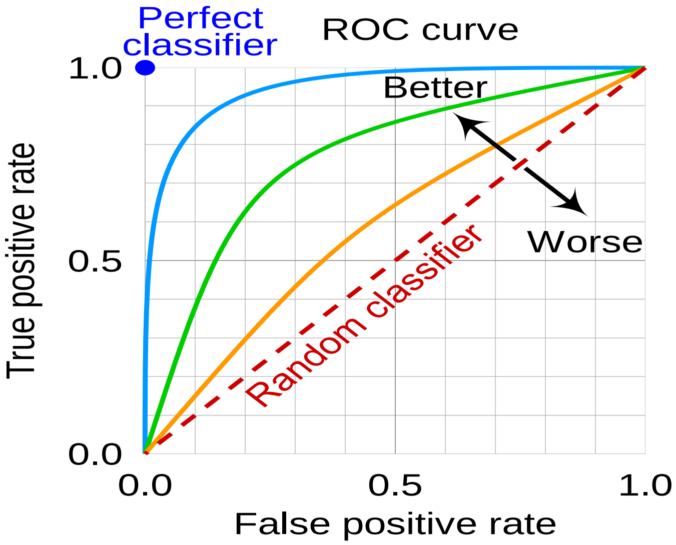
$$Type\ \ II\ Error = \frac{False\ Negative}{True\ Positive + False\ Negative} \tag{4.12}$$

Given the differential impacts of these errors, this research operates under the premise that not all misclassifications bear equal consequences. Specifically, the cost associated with Type II errors, where financially unstable firms are erroneously deemed stable, is posited to be higher than that of Type I errors due to the potential financial losses and procedural costs banks or financial institutions may incur in the event of a bankruptcy (Yeo, 2016). Conversely, Type I errors, which incorrectly classify safe firms as bankrupt, result in lost profit opportunities for credit institutions. This study, therefore, prioritizes the minimization of Type II errors in evaluating the performance of the proposed bankruptcy prediction model.

Finally, the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) are utilized as overarching measures of model performance in Equation as illustrated in Figure 4.6 (Hastie et al., 2019). The ROC curve plots the trade-off between true positive rates and false positive rates across different threshold settings as shown in Equation 4.13. This provides a holistic evaluation of a model's proficiency in distinguishing between bankrupt and non-bankrupt firms, regardless of class distribution or misclassification costs (Psillaki et al., 2010). An ideal ROC curve would closely approach the upper left corner of the plot, indicating a high true positive rate alongside a low false positive rate, culminating in an AUC value approaching 1. This metric serves as a robust indicator of a model's predictive accuracy across an imbalanced dataset, facilitating a direct comparison of performance across different models by encapsulating the balance between Type I and Type II errors in a single, comprehensive statistic (Son et al., 2019).

$$Receiver\ Operating\ Characteristics = \int_0^1 TPR\left(FPR^{-1}(x)\right)dx \tag{4.13}$$

Figure 4.6. Receiver Operating Characteristic (ROC) curve



Source **: Cook (2007)**

4.9. Interpretability

In the evolving landscape of machine learning research, the explication of model outcomes has emerged as a paramount concern, particularly in the context of complex models used for bankruptcy prediction. There have been used five primary methodologies for elucidating machine learning models, each catering to distinct facets of interpretability (Bracke et al., 2019). The initial category, surrogate models, encompasses inherently simple models like logistic regression and decision trees, which intrinsically facilitate interpretation without necessitating auxiliary explanatory frameworks. A second notable method involves feature importance metrics, notably employed within the random forest model framework (Schapire, 2013). This technique quantifies the impact on prediction variance when a specific feature is excluded, offering insight into the relative significance of model attributes.

Local surrogate models, the third approach, aim to approximate the predictions of complex models within specific data subsets. LIME (Local Interpretable Model-agnostic Explanations) is a prominent example of this technique, creating interpretable models that explain predictions on an instance-by-instance basis (Carmona et al., 2019). This local perspective allows for interpretations of how a model behaves under various conditions.

The fourth strategy, instance-based interpretation, eschews the creation of additional models, focusing instead on direct prediction analysis. This methodology's hallmark is its capacity to identify which variables markedly influence outcomes, with the Shapley value method standing out for its ability to account for interactions between features (Linardatos et al., 2020).

Lastly, SHapley Additive exPlanations (SHAP) integrate elements from both LIME and Shapley values to provide a unified framework for model interpretation (Antunes et al., 2017). SHAP values articulate the contribution of each feature to a particular prediction, grounded in the cooperative game theory principle of fair contribution. This method not only highlights the importance of individual features but also their interdependencies, thereby offering a detailed and equitable view of feature contributions across different instance (Linardatos et al., 2020).

$$\boldsymbol{g(z') = \varphi_0 + \sum_{i=0}^{M} \varphi_i z'_i} \tag{4.14}$$

***Source*: Molnar (2020)**

Where g represents an explanatory model, M denotes the maximum number of features, and z'_i is a binary variable that indicates whether *i*-th feature contributes $(z'_i = 1)$ or not $(z'_i = 0)$. φ_i is the SHAP value of the i-th feature, reflecting its contribution. SHAP values offer an effective solution to the computational demands of explanation models by considering all potential sequences of variables when assessing feature importance. Furthermore, a global interpretation can be achieved by summing the SHAP values across all samples (Ariza-Garzón et al., 2020). The feature importance is quantified by averaging the absolute values for each feature according to following equation:

$$\varphi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!\,(M - |z'| - 1)!}{M!} \left[f_x(z') - f_x\left(z' \backslash i\right)\right] \quad \textbf{(4.15)}$$

***Source*: Molnar (2020)**

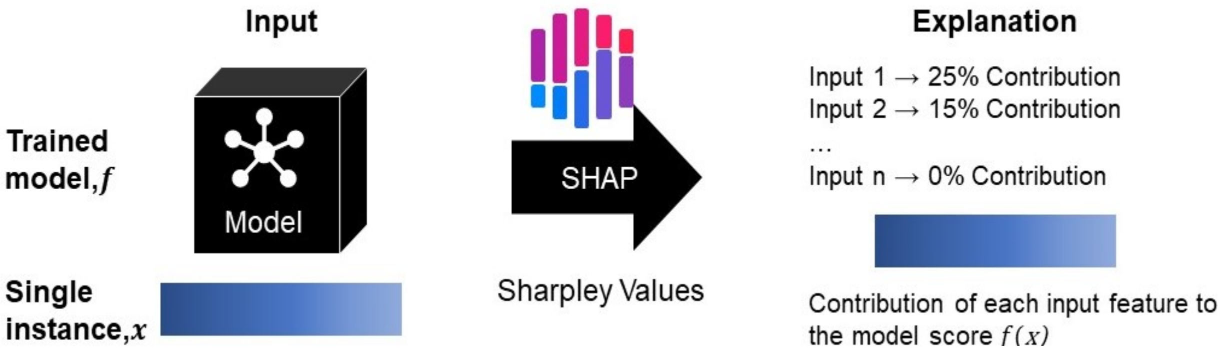
Where $\varphi_i(f,x)$ represents the Shapley value for feature i, M refer the number of features, z' represents the set of all features, $|z'|$ indicates the number of features in the feature subset z' excluding the i-th feature. $f_x(z')$ denotes the output of machine learning model f trained on the feature subset z'.

In this study, the SHapley Additive exPlanations (SHAP) explainer is employed to discern the key factors influencing the bankruptcy probability predictions for individual firms. Utilizing SHAP values, the study identifies and ranks the importance and impact of specific variables within the machine learning model, as depicted in Figure 4.7.

SHAP values clarify which features are most influential in the model's predictions, indicating their positive or negative effect on bankruptcy risk. This method allows for a detailed analysis of each firm's risk profile, especially those at higher risk, by pinpointing the critical factors contributing to their financial instability (Lee et al., 2022).

By incorporating explainable AI through SHAP, the study achieves a balance between the model's predictive accuracy and its interpretability. This approach enhances the model's transparency, enabling stakeholders to understand the rationale behind predictions. As a result, the study identifies a comprehensive set of risk factors that significantly affect the financial health of small and medium-sized enterprises (SMEs) in the shipping industry, providing valuable insights for risk management strategies.

Figure 4.7 Pipeline of SHarply addictive exPlanation(SHAP)



Source: Knapič et al. (2021)

4.10. Chapter Summary

This chapter has delineated the methodological framework employed in this research. It began with an overview of the theoretical underpinnings of the chosen research methodology. The study will sample corporate data based on various statuses including bankruptcy, restructuring, or mergers and acquisitions. Selection of explanatory variables will be guided by a comprehensive literature review, focusing on accounting-based, macroeconomic, non-financial, and shipping index variables. Market-based variables will be omitted due to the predominance of unlisted companies within the Korean shipping industry. Data spanning up to five years prior to the initial bankruptcy year will be gathered for bankrupt firms, employing a similar data collection approach for non-bankrupt firms up to the most recent year. The sample encompasses a 22-year period from 2001 to 2022. The selection of variables will be further refined through interviews with academics and industry practitioners to ensure theoretical validity.

For the empirical analysis, Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) models have been selected due to their proven predictive capabilities and their ability to address statistical challenges such as skewness, missing values, and outliers identified in the literature review. Moreover, their capacity to analyse time-series data sets acknowledges the long-term dependencies inherent in bankruptcy predictors.

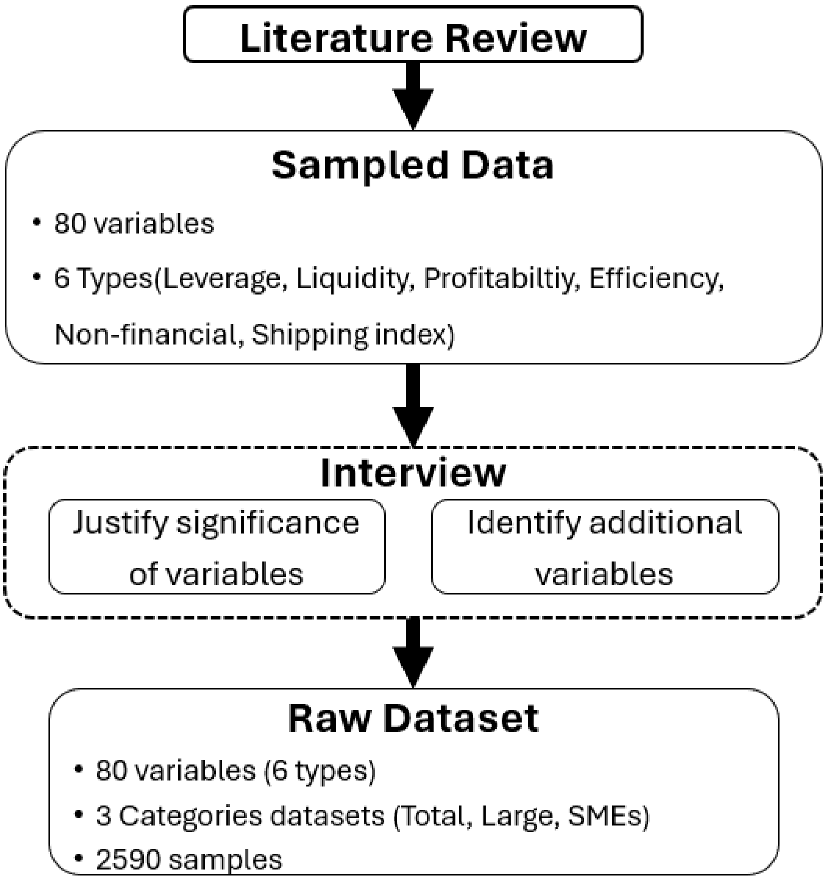
The study also addresses the imbalance issue between bankrupt and non-bankrupt firms using oversampling techniques. This approach mitigates the risk of the classification model being biased towards the more numerous non-bankrupt cases at the expense of accurately identifying bankrupt cases. To reconcile the trade-off between accuracy and interpretability, the SHAP explainer which is one of the explainable AI techniques will be deployed. This will facilitate a detailed evaluation of the impact of relevant bankruptcy risk factors for small and medium-sized shipping companies, enhancing the interpretability of the predictive models used in this study.

5. RESEARCH MODEL DEVELOPMENT

5.1. Introduction

This chapter discusses the methodology employed to develop the research model for this study, which is informed by both literature reviews and interviews, as depicted in Figure 5.1. The dataset analysed comprises 2,590 firm-year observations, including 62 bankrupt firms and 134 active shipping companies, spanning from 2001 to 2022. As detailed in Chapter 5, bankruptcy status is categorized into two states for each firm: ‘0’ representing active companies and ‘1’ indicating bankruptcy.

Figure 5.1. Research model development procedure



Source: **Author**

Consequently, the chapter is structured into two primary sections. The initial section presents the outcomes of interviews conducted with shipping industry practitioners.

These interviews served to validate the explanatory variables, with each participant asked to assess the importance of each variable on a five-point Likert scale. Additionally, the interviewees contributed insights on further variables that could enhance the model’s explanatory power.

Following this, the chapter delves into descriptive statistics for the sets of variables identified. This analysis sheds light on the distinctive characteristics of the variables across different groups and forecasting periods. By examining the relationship between bankruptcy risk and the explanatory variables, this study aims to provide effective bankruptcy prediction, serving as early warning signals for a potential bankruptcy crisis within the shipping industry.

5.2. Findings from Interview

During the interview process, respondents were requested to evaluate the importance of each variable listed in the Appendix. As delineated in Chapter 4.5, the resulting scores served as a basis for variable selection, with a threshold set at 3.5 or above from the interviewees for a variable to be considered significant. Table 5.1 compiles the average scores assigned to each variable, reflecting the respondents' perception of their importance. The outcomes from the interviews reveal that all variables identified through literature review achieved scores of 3.5 or higher, thereby affirming their substantial explanatory power in predicting bankruptcy within the shipping industry.

Interestingly, the respondents' opinions diverged based on their professional backgrounds. Academics and government authorities tended to emphasize the relevance of leverage and profitability metrics for shipping companies, whereas industry practitioners placed greater importance on shipping indices over the companies' internal financial metrics. Practitioners highlighted the need to recognize the significance of shipping indices, not only market indices like the Baltic Dry Index and China Container Freight Rate but also the current status of vessels owned by shipping companies. This latter aspect is considered reflective of a company's operational capacity.

Furthermore, interviewees in government authority also acknowledged the pivotal role of macroeconomic factors, such as global GDP growth rates and seaborne trade volume, in assessing bankruptcy risk within the shipping sector as determinants of maritime transport demand. Furthermore, multiple interviewees highlighted the influence of exchange rates and interest rates, particularly the London Inter-Bank Offered Rate (LIBOR), on financial risk. These factors are closely associated with financial leverage, investment financing plans, and cash flow management in the shipping industry (Kavussanos et al., 2021). Considering the capital-intensive nature of the shipping industry, interest rates play a crucial role in affecting cash flow and liquidity concerns for individual firms. Additionally, the foreign exchange rate directly impacts shipping companies' revenues, as most are denominated in USD (Alexandridis et al., 2020).

Consequently, these factors identified during the interviews will also be incorporated as explanatory variables in this research, enriching the analysis with a comprehensive set of determinants reflective of both internal company metrics and external market and economic conditions.

Table 5.1. The average importance score of variables evaluated by respondents

	Shipping company	**University**	**Government Authority**
Leverage	4.02	4.46	4.27
Liquidity	3.90	3.88	3.85
Profitability	3.99	4.18	4.21
Efficiency	3.94	3.81	3.85
Macroeconomic	4.06	4.09	4.08
Shipping Index	4.15	4.22	4.14

5.3. Description of Incorporate Variables

As discussed in the previous chapter, additional variables suggested by interviewees have been incorporated to evaluate their explanatory ability in predicting bankruptcy within the shipping industry. The global financial crisis significantly impacted the shipping sector, and this impact has been reflected in macroeconomic factors such as GDP and oil prices (Sousa et al., 2022).

As illustrated in Figure 5.2, currency exchange rates and LIBOR interest rates fluctuated significantly over the sample period, especially during the global financial crisis. GDP also showed an overall increase over the period but experienced notable fluctuations. Additionally, Figure 5.3 shows that shipping index variables also declined during the financial crisis. In particular, the Baltic Dry Index (BDI) plummeted by 94%, IRONSTEEL decreased by 47%, and oil prices fell by 71%. During this period, the growth rate of trade volume in Korea exhibited a pattern similar to global trends, but the extent of fluctuation was considerably large.

Figure 5.2. Annual macroeconomic factors across forecasting periods

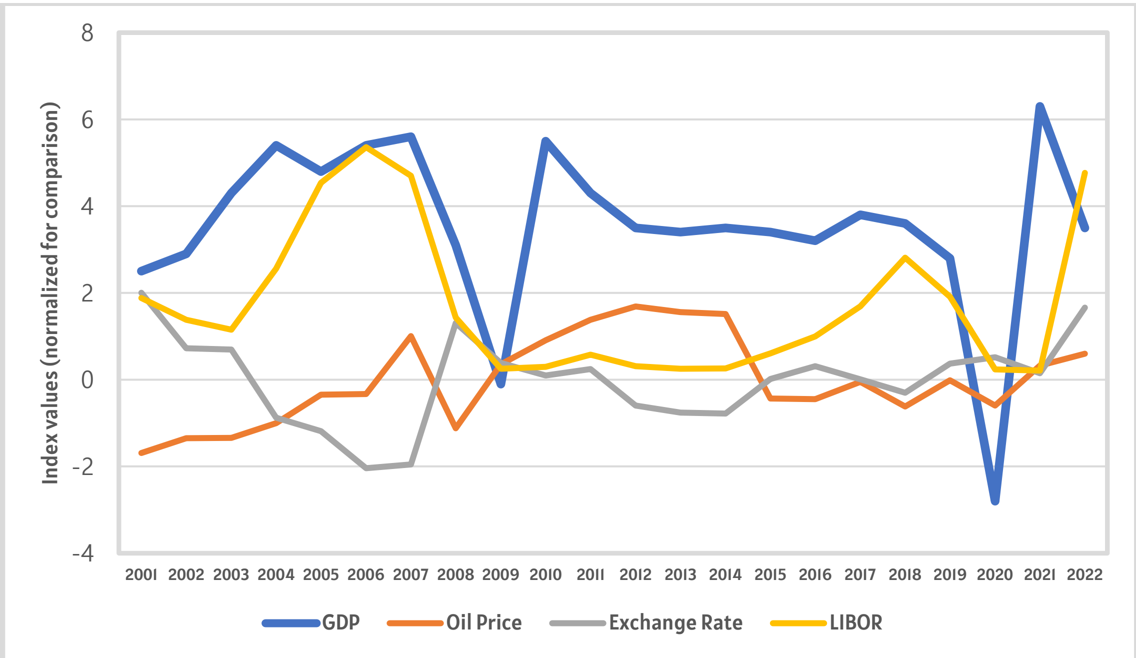
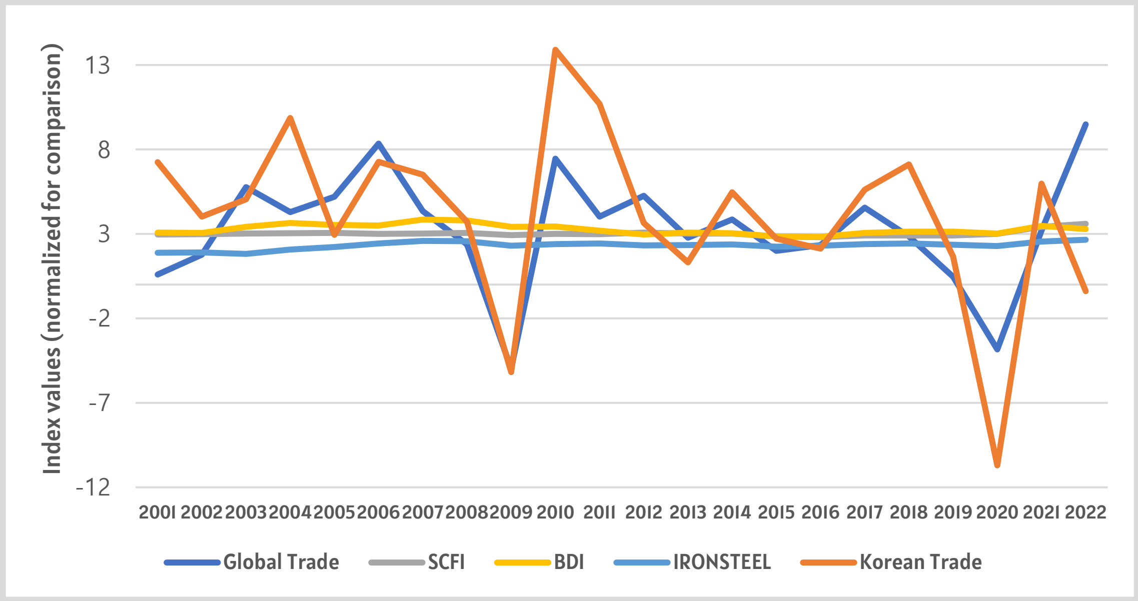


Figure 5.3. Annual shipping index across forecasting periods



In conclusion, the final set of explanatory variables were selected through a thorough literature review and validated by interviews with industry practitioners in Table 5.2. These variables were applied in the bankruptcy prediction models for the shipping industry to provide a comprehensive analysis. The study utilized a diverse array of data sources categorized into three primary groups: financial variables, non-financial variables, and shipping indices as presented in

Table 5.3. Each category encompassed specific types of data that are critical for assessing the financial health and operational status of shipping companies. Financial variables included leverage, liquidity, profitability, and efficiency, sourced from the Korean Financial Supervisory Service (2023). Non-financial variables such as size, age, GDP, oil prices, and currency were obtained from the Korean Financial Supervisory Service and the Korean Ministry of Maritime Affairs & Fisheries (2023). Shipping indices, including LIBOR, growth rates of global and Korean seaborne trade volumes, container freight index rates, Baltic Dry Index (BDI), IRONSTEEL, and metrics like the number of vessels, gross tonnage, and deadweight tonnage, were sourced from the Korean Ocean Business Operation (2023) and the Korean Ministry of Maritime Affairs & Fisheries (2023).

Table 5.2. Description of variables selected from explanatory study

Category	Variable	Description
Financial Metrics - Leverage		
A1	Return on total assets	$\frac{EBITDA}{Total\ Assets}$
A2	Retained Earnings to Total Assets	$\frac{Retained\ Earnings}{Total\ Assets}$
A3	Leverage	$\frac{Total\ Debt}{Total\ Assets}$
A4	Liabilities to Total Assets	$\frac{Total\ Liabilities}{Total\ Assets}$
A5	Equity to Assets	$\frac{Shareholder's\ Equity}{Total\ Assets}$
A6	Debt to Equity	$\frac{Total\ Debt}{Shareholder's\ Equity}$
A7	Liabilities to Equity	$\frac{Total\ Liabilities}{Shareholder's\ Equity}$
A8	Gearing ratio	$\frac{Shareholder's\ Equity}{Total\ Debt}$
A9	Equity to liabilities	$\frac{Shareholder's\ Equity}{Total\ Liabilities}$
A10	Liabilities	$Total\ Liabilities$
A11	Current Liabilities to Equity	$\frac{Current\ Liabilities}{Equity}$
A12	EBITDA to Liabilities	$\frac{EBITDA}{Total\ Liabilities}$
A13	EBITDA/EV	$\frac{EBITDA}{Equity}$
A14	EBITDA	$Net\ Income + Interest + Taxes + Depreciation + Amortization$
A15	Long-term Liabilities to Assets	$\frac{liabilities - current\ liabilities}{Total\ Assets}$
A16	Net Income to Liabilities	$\frac{Net\ Income}{Total\ Liabilities}$
A17	Working Capital	$Current\ Assets - Current\ Liabilities$
A18	Current Assets to Liabilities	$\frac{Current\ Assets}{Total\ Liabilities}$
A19	EBITDA to Current Liabilities	$\frac{EBITDA}{Current\ Liabilities}$
A20	Long-term Liabilities to Equity	$\frac{liabilities - current\ liabilities}{Shareholder's\ Equity}$
Financial Metrics - Liquidity		
B1	Current Ratio	$\frac{Current\ Assets}{Current\ Liabilities}$
B2	Working Capital to Assets	$\frac{Current\ Assets - Current\ Liabilitiesl}{Total\ Assets}$
B3	Cash Assets Ratio	$\frac{Cash}{Total\ Assets}$
B4	Cash Ratio	$\frac{Cash}{Current\ Liabilities}$
B5	Current Assets to Assets	$\frac{Current\ Assets}{Total\ Assets}$
B6	Current Liabilities to Assets	$\frac{Current\ Liabilities}{Total\ Assets}$
B7	Quick Ratio	$\frac{Cash + accounts\ receivable}{Current\ liabilities}$
B8	Cash to Debt Ratio	$\frac{Cash}{Total\ Debt}$
B9	Shareholder's Equity	$Total\ Assets - Total\ Liabilities$
B10	Current Liabilities Ratio	$\frac{Current\ Liabilities}{Total\ Liabilities}$
B11	Current Assets	$Cash + Accounts\ Receivable + Inventory$
B12	Cash Sales Ratio	$\frac{Cash}{Total\ Sales}$
B13	Current Liabilities to Current Assets	$\frac{Current\ Liabilities}{Current\ Assets}$
B14	Working Capital to Current Assets	$\frac{Working\ Captial}{Current\ Assets}$
B15	Quick Assets to Total Assets	$\frac{Cash + accounts\ receivable}{Total\ Assets}$
B16	Working Capital to Equity	$\frac{Current\ Assets - Current\ Liabilities}{Shareholder's\ Equity}$
B17	Working Capital to Current Liabilities	$\frac{Current\ Assets - Current\ Liabilities}{Current\ Liabilities}$
B18	Cash to Current Assets	$\frac{Cash}{Current\ Assets}$
B19	Retained Earnings to Current Liabilities	$\frac{Retained\ Earnings}{Current\ Liabilities}$
B20	ΔTotal Asset	$\frac{Total\ Assets_t - Total\ Assets_{t-1}}{Total\ Assets_{t-1}}$
Financial Metrics - Profitability		
C1	Return on Assets	$\frac{Net\ Income}{Total\ Assets}$
C2	Return on Equity	$\frac{Net\ Income}{Shareholder'sEquity}$
C3	Return on Sales	$\frac{Net\ Income}{Sales}$
C4	EBITDA to Sales	$\frac{EBITDA}{Sales}$
C5	Sales to Current Assets	$\frac{Sales}{Current\ Assets}$
C6	Sales	$Total\ Sales$
C7	EBITDA to Interest Coverage	$\frac{EBITDA}{Interest\ expenses}$
C8	Gross Profit to Assets	$\frac{Gross\ Profit}{Total\ Assets}$
C9	Net interest margin	$\frac{Interest\ returns - Interest\ paid}{Average\ Assets}$
C10	Sales to equity	$\frac{Total\ Sales}{Shareholder's\ Equity}$
C11	Net Income	$Total\ Revenue - Total\ Expenses$
C12	Operating Return on Assets	$\frac{Operating\ Income}{Total\ Assets}$
C13	Sales to Liabilities	$\frac{Total\ Sales}{Total\ Liabilties}$
C14	Gross Profit to Current Liabilities	$\frac{Gross\ Profit}{Current\ Liabilities}$
C15	Gross Profit to Liabilities	$\frac{Gross\ Profit}{Total\ Liabilities}$
C16	Gross Profit to Sales	$\frac{Gross\ Profit}{Total\ Sales}$
C17	ΔSales	$\frac{Total\ sales_t - Total\ Sales_{t-1}}{Total\ Sales_{t-1}}$
C18	Sales to Current Liabilities	$\frac{Total\ Sales}{Current\ Liabilities}$
Financial Metrics - Efficiency		
D1	Asset turnover	$\frac{Total\ Sales}{Total\ Assets}$
D2	Working Capital to Sales	$\frac{Working\ Captial}{Total\ Sales}$
D3	Current Assets to Sales	$\frac{Current\ Assets}{Total\ Sales}$
D4	Inventory to Assets	$\frac{Inventory}{Total\ Assets}$
D5	Current Liabilities to Sales	$\frac{Current\ Liabilities}{Total\ Sales}$
D6	Inventory turnover	$\frac{Average\ Inventories}{Total\ Sales}$
D7	Operating Margin	$\frac{Operating\ Income}{Sales}$
Non-financial		
E1	Size	*Proxied by Log (Total Assets)*
E2	Age	*Age of firm*
E3	Type of Operator	Dummy value of Type: 1 (Container), 2 (Tanker), 3 (Bulk)
E4	GDP	$\frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}$
E5	Oil Price	*Current US$ price of Brent oil*
E6	Currency	Annual exchange rate of KRW to USD
Shipping Index		
F1	LIBOR	*Annual average of Libor interest rates*
F2	Growth rate of Global Seaborne Trade Volume	$\frac{Trade\ volume_t - Trade\ volume_{t-1}}{Trade\ volume_{t-1}}$
F3	Growth rate of Korean seaborne Trade volume	$\frac{Trade\ volume_t - Trade\ volume_{t-1}}{Trade\ volume_{t-1}}$
F4	Container Freight Index Rate	*Annual average of Shanghai Containerized Freight Index*
F5	Baltic Dry Index (BDI),	*Annual average of Baltic dry index*
F6	IRONSTEEL	*Annual average of Dow Jones U.S. Iron & Steel Index*
F7	Number of vessels	*Number of owned vessels*
F8	Gross Tonnage	*Sum of gross tonnage of owned vessels*
F9	Deadweight Tonnage	*Sum of deadweight tonnage of owned vessels*

By systematically categorizing and sourcing data, this study ensures a comprehensive and reliable foundation for analysing bankruptcy risk factors in the Korean shipping industry. The references provided facilitate verification and further exploration of the data, contributing to the robustness and credibility of the research findings. This methodical approach ensures that the selected variables are not only theoretically sound but also practically relevant, enhancing the predictive bankruptcy models used in this study.

Table 5.3. Summary of Data Sources Used in the Research

Data		Sources
List of Shipping Companies		Korean Shipowners Association. (2023) - Available at: https://oneksa.kr/boards/statistics (accessed on 27th September, 2023)
Financial variables	Leverage	Korean Financial Supervisory Service (2023) - Available at http://dart.fss.or.kr (accessed on 9th September, 2023)
	Liquidity	Korean Financial Supervisory Service (2023) - Available at http://dart.fss.or.kr (accessed on 9th September, 2023)
	Profitability	Korean Financial Supervisory Service (2023) - Available at http://dart.fss.or.kr (accessed on 9th September, 2023)
	Efficiency	Korean Financial Supervisory Service (2023) - Available at http://dart.fss.or.kr (accessed on 9th September, 2023)
Non-financial	Size, Age	Korean Financial Supervisory Service (2023) - Available at http://dart.fss.or.kr (accessed on 9th September, 2023)
	GDP	Korean Ministry of Maritime Affairs & Fisheries (2023) - Available at: http://www.momaf.go.kr/statistics2023 (accessed on 23th October, 2023)
	Oil Price	Korean Ministry of Maritime Affairs & Fisheries (2023) - Available at: http://www.momaf.go.kr/statistics2023 (accessed on 23th October, 2023)
	Currency	Korean Ministry of Maritime Affairs & Fisheries (2023) - Available at: http://www.momaf.go.kr/statistics2023 (accessed on 23th October, 2023)
Shipping Index	LIBOR	Korean Ocean Business Operation (2023) - Available at: https://www.kobc.or.kr/ebz/shippinginfo/main (accessed on 12th October, 2023)
	Growth rate of Global Seaborne Trade Volume	Korean Ministry of Maritime Affairs & Fisheries (2023) - Available at: http://www.momaf.go.kr/statistics2023 (accessed on 23th October, 2023)
	Growth rate of Korean seaborne Trade volume	Korean Ministry of Maritime Affairs & Fisheries (2023) - Available at: http://www.momaf.go.kr/statistics2023 (accessed on 23th October, 2023)
	Container Freight Index Rate	Korean Ocean Business Operation (2023) - Available at: https://www.kobc.or.kr/ebz/shippinginfo/main (accessed on 12th October, 2023)
	Baltic Dry Index (BDI),	Korean Ocean Business Operation (2023) - Available at: https://www.kobc.or.kr/ebz/shippinginfo/main (accessed on 12th October, 2023)
	IRONSTEEL	Korean Ministry of Maritime Affairs & Fisheries (2023) - Available at: http://www.momaf.go.kr/statistics2023 (accessed on 23th October, 2023)
	Number of vessels, Gross Tonnage, Deadweight Tonnage	Korean Shipowners Association. (2023) - Available at: https://oneksa.kr/boards/statistics (accessed on 27th September, 2023)

5.4. Descriptive Statistics

5.4.1. Profile of target data

The final sample in this study comprises 2,590 firm-year observations, which include 62 bankrupt firms and 134 active shipping companies, spanning the sample period from 2001 to 2022. As outlined in Chapter 4, this study categorizes corporate bankruptcy into two states: 0 for active and 1 for bankruptcy. Table 5.4 presents the number of firm-month observations and the proportion of bankruptcy cases over the five years preceding bankruptcy for each forecasting horizon. In the five years leading up to bankruptcy, 2,273 active firm-month observations with 317 of which reflect bankrupt firms, which is accounted for 12.2%. In three years prior to bankruptcy, 2,391 active firm observations with 199 bankrupt firm-month observations, which accounted for 7.7%. As shown in Table 5.4, bankruptcy observations are highly rare cases which is accounted for 2.7% of the dataset in a year prior to bankruptcy, which results in high imbalanced between active and bankruptcy datasets.

Table 5.4. Distribution of active and bankrupt firm-year observations

Years prior to bankruptcy	**Bankrupt**		**Active**	
	No.	Percentage	No.	Percentage
1 year	70	2.7%	2520	97.3%
3 years	199	7.7%	2391	92.3%
5 years	317	12.2%	2273	87.8%

Table 5.5 shows distribution of active and bankrupt firms divided into 274 firm observation years of large firms and 2316 observation years of small and medium-sized companies over five-year observations. In case of large firms, there is a 7% bankrupt rate among sample periods, which is considered as the rare event. This means that there were only a few large firms which faced economic depressed can be observed. On the contrary,

bankrupt rates in SMEs was relatively high, which is accounted for approximately 8% average. In particular, when the forecast period is expanded to 5 years before bankruptcy, there are firm-year observations for about 12.4% of bankrupt rates during the sample period. This can be interpreted as SMEs being more vulnerable to bankruptcy risk compared to large firms.

Table 5.5. Comparison between large firms and bankrupt firm-year observations

Years prior to bankruptcy	**Large Firms**			**SMEs**		
	Active	Bankrupt	Bankrupt Rate	Active	Bankrupt	Bankrupt Rate
1 year	264	10	3.6%	1820	75	2.6%
3 years	254	20	7.3%	1824	192	7.7%
5 years	245	29	10.6%	1832	292	12.4%

5.4.2. Descriptive statistics of bankrupt and active firms

Table 5.6 provides a summary of the statistics, encompassing mean, standard deviation values for the variables chosen in this research, contrasting active and bankrupt firms in the shipping industry. For bankrupt firms, financial statements up to one year prior to bankruptcy were analysed. Generally, firms on the verge of bankruptcy exhibited higher leverage, lower profitability, and reduced efficiency in comparison to their active counterparts. Liquidity factors varied, with some ratios being higher and others lower, reflecting their inherent volatility.

Table 5.6 highlights distinct financial characteristics between active and bankrupt firms. Active firms, on average, demonstrated higher returns on total assets, indicating a superior ability to generate profit from assets. Ratios related to liabilities, such as the debt ratio, liabilities to total assets ratio, and equity to debt ratio, showed that bankrupt firms had a lower leverage level, implying diminished solvency. Regarding liquidity, active firms displayed higher average ratios concerning current assets, including the working capital to assets ratio, current liabilities to current assets ratio, and working capital to current assets ratios. This suggests that active firms are better equipped to fulfil their short-term liabilities. The t-statistic values underline a significant difference in shareholder's equity, pointing to a higher liquidity level in active firms, enabling them to cover their liabilities more effectively. Profitability ratios, including gross profit to liabilities, gross profit to sales ratios, sales, and net income ratio, exhibited a marked difference between bankrupt and active firms, as revealed by t-statistic values. This indicates that bankrupt firms have a lower capacity to generate sales and profit and are less efficient in managing labour and the costs of goods or services. The high standard deviation values for the EBITDA to interest coverage ratio among both active and bankrupt firms highlight its variability, suggesting diverse risks within the industry in generating sufficient profit to cover interest expenses. Efficiency ratios, such as current assets to sales and current liabilities to sales, were notably higher among active firms, indicating their stronger sales performance and greater efficiency in revenue generation from assets. The variability in these ratios, as suggested by higher standard deviation values, reflects differences in profit-generating capabilities even among active firms.

Additionally, active firms were, on average, older and larger than their bankrupt counterparts, a trend also mirrored in the status of vessel ownership, with larger firms operating more and larger vessels. In contrast, bankrupt firms, on average, operated fewer and smaller vessels. Regarding shipping market indices like the container freight index rate and the Baltic dry index, bankrupt firms recorded lower values than active firms, correlating with reduced revenue and potentially diminishing profitability.

In summary, Table 5.6 portrays active firms as having a more robust financial profile compared to bankrupt firms, highlighting significant differences in financial health and operational characteristics between the two groups within the shipping industry.

Table 5.6. Descriptive statistic of active and bankrupt firms

Variable	Total		Active		Bankruptcy		t-statistic
	Mean	Std	Mean	Std	Mean	Std	
A1	-0.54	29.31	0.04	0.20	-2.99	66.73	-1.014*
A2	-1.14	48.15	0.14	0.38	-6.47	109.53	-1.347
A3	1.83	45.74	0.72	0.31	6.50	104.08	1.240
A4	1.83	45.74	0.72	0.31	6.50	104.08	1.240
A5	-0.83	45.74	0.28	0.31	-5.50	104.08	-1.241
A6	6.73	205.18	7.94	227.20	1.38	9.63	-0.644
A7	6.29	69.93	6.49	32.70	5.34	144.42	-0.176
A8	6.29	69.93	6.49	32.70	5.34	144.42	-0.176
A9	6.73	205.18	7.94	227.20	1.38	9.63	-0.644
A10	10.48	0.94	10.60	0.91	9.97	0.86	-14.601***
A11	2.54	18.52	2.48	16.89	2.76	23.97	0.254
A12	0.10	1.68	0.10	1.82	0.13	0.80	0.369
A13	-0.14	6.73	-0.13	6.83	-0.17	6.14	-0.129
A14	0.01	0.27	0.02	0.28	-0.02	0.18	-3.029***
A15	1.03	31.83	0.37	0.26	3.78	72.45	1.049
A16	0.08	1.67	0.08	1.81	0.10	0.73	0.266
A17	0.00	0.04	0.00	0.04	0.00	0.04	-2.347*
A18	1.67	29.30	1.74	32.23	1.32	7.94	-0.290
A19	0.22	3.82	0.18	2.23	0.41	7.41	0.672
A20	3.75	61.82	4.01	23.08	2.58	132.60	-0.241
B1	3.05	37.55	2.41	32.28	5.72	53.84	1.318
B2	-0.50	15.13	-0.04	0.30	-2.41	34.40	-1.537
B3	0.08	0.11	0.08	0.10	0.09	0.15	1.561
B4	1.27	23.47	1.19	24.94	1.59	15.04	0.344
B5	0.30	0.25	0.30	0.24	0.32	0.30	0.817
B6	0.81	15.14	0.35	0.26	2.73	34.41	1.544
B7	2.97	37.54	2.34	32.28	5.61	53.82	1.303
B8	0.86	22.51	0.98	24.92	0.30	1.16	-0.615
B9	0.14	0.66	0.16	0.71	0.03	0.30	-6.265***
B10	0.52	0.28	0.51	0.26	0.56	0.32	3.122
B11	0.01	0.05	0.01	0.06	0.00	0.03	-4.464
B12	1.12	31.36	1.33	34.74	0.16	0.57	-0.753
B13	5.46	91.59	2.46	8.51	18.00	207.46	1.672
B14	-4.46	91.59	-1.46	8.51	-17.00	207.46	-1.672
B15	0.29	0.25	0.28	0.23	0.30	0.30	0.724
B16	-0.18	21.52	-0.53	5.56	1.30	47.64	0.855
B17	2.05	37.55	1.41	32.28	4.72	53.84	1.318
B18	0.27	0.23	0.27	0.22	0.29	0.27	1.904
B19	0.86	17.15	0.80	7.63	1.15	35.78	0.217
B20	0.92	16.81	1.09	18.64	0.41	3.40	-0.818
C1	-0.55	29.32	0.04	0.19	-2.99	66.75	-1.013
C2	-0.16	6.72	-0.15	6.82	-0.20	6.10	-0.142
C3	-0.54	17.23	-0.59	19.06	-0.26	1.86	0.395
C4	-0.22	9.44	-0.21	10.41	-0.25	1.86	-0.098
C5	7.41	34.50	6.86	29.85	9.55	49.01	1.177

Variable	Total		Active		Bankruptcy		t-statistic
	Mean	Std	Mean	Std	Mean	Std	
C6	0.04	0.15	0.04	0.16	0.02	0.11	-3.722***
C7	3874.71	141654.85	4602.72	156826.61	640.15	9598.83	-0.564
C8	0.17	3.18	0.25	0.74	-0.15	7.08	-1.240
C9	0.07	1.68	0.03	0.18	0.23	3.80	1.163
C10	7.81	69.07	7.71	74.78	8.01	33.32	0.087
C11	0.01	0.04	0.01	0.04	0.01	0.04	-2.631**
C12	-0.02	3.37	0.07	0.19	-0.37	7.67	-1.268
C13	3.54	23.10	3.30	24.01	4.51	18.13	1.250
C14	1.65	18.03	0.99	3.50	4.41	40.32	1.891
C15	0.80	6.61	0.56	2.43	1.79	14.16	1.926*
C16	0.15	0.47	0.16	0.49	0.10	0.38	-3.045*
C17	5.79	136.89	7.03	151.60	0.54	3.15	-0.956
C18	6.23	29.66	5.50	25.46	9.24	42.59	1.880
D1	1.46	3.85	1.36	2.21	1.86	7.49	1.470
D2	0.70	98.64	1.01	109.21	-0.59	6.48	-0.326
D3	5.97	216.61	7.10	239.85	1.03	11.46	-0.564
D4	0.02	0.04	0.02	0.04	0.02	0.06	0.579
D5	5.27	134.04	6.09	148.43	1.62	6.23	-0.672
D6	0.17	4.68	0.20	5.19	0.03	0.13	-0.743
D7	-0.18	8.39	-0.21	9.29	-0.05	0.35	0.368
E1	10.65	0.90	10.79	0.87	10.07	0.73	-19.750***
E2	17.44	14.33	19.08	15.00	11.34	8.11	-19.040***
E3	2.16	0.58	2.15	0.61	2.12	0.41	0.990
E4	3.43	1.95	1355.68	338.09	1453.23	181.59	-2.497*
E5	73.91	27.12	72.07	26.51	84.82	26.28	6.973*
E6	1128.07	85.38	1132.39	87.38	1111.51	48.08	-5.573*
F1	1.49	1.56	1.57	1.59	0.80	0.82	-6.283*
F2	3.18	3.26	1117.84	242.36	1195.56	128.63	-2.249*
F3	3.76	5.18	3.61	5.25	4.00	4.82	3.038*
F4	1188.09	757.27	1234.32	824.00	983.55	344.82	-10.977*
F5	2019.45	1563.91	2049.85	1561.06	1308.12	781.16	-1.806
F6	240.30	86.09	242.60	90.77	230.26	38.59	-3.379*
F7	6.34	10.69	6.94	10.24	4.12	13.69	-5.783*
F8	228.08	753.88	248.79	754.47	154.38	826.29	-3.344*
F9	354.70	1224.24	387.77	1244.60	241.62	1249.29	-3.413*

***P<.05, **P<.01, ***P<.001**

5.4.3. Descriptive statistics over five years prior to bankruptcy

Table 5.7 presents summary statistics for variables of bankruptcy firms over five years forecasting horizons, which would be divided into 1, 3, 5 years. According to Table 5.7, when bankrupt firms approached to bankrupt, the mean values of almost every value of ratios, especially among the ratios related with liabilities are deteriorated continuously over five years prior to bankruptcy. Thus, this can provide initial indications of explanatory power of variables encompassing leverage, liquidity, profitability, efficiency and firm-specific factors over one-to-five-year forecasting horizon.

According to Table 5.7, There are significant trends in leverage ratios over five years prior to bankruptcy. As firms approached to bankrupt, the mean values of all leverage ratios deteriorated continuously except equity to debt, total liabilities, and equity to liabilities. In particular, in case of retained earnings to total assets(A2) and debt ratio(A3), this trend can be observed specifically according to Figure 5.4 and Figure 5.5. In contrast, equity to debt, total liabilities, and equity to liabilities also showed similar trends, but they also showed slightly recovery a year prior to bankruptcy.

Figure 5.4. Retained Earnings to Total Assets ratio over five years prior to bankruptcy

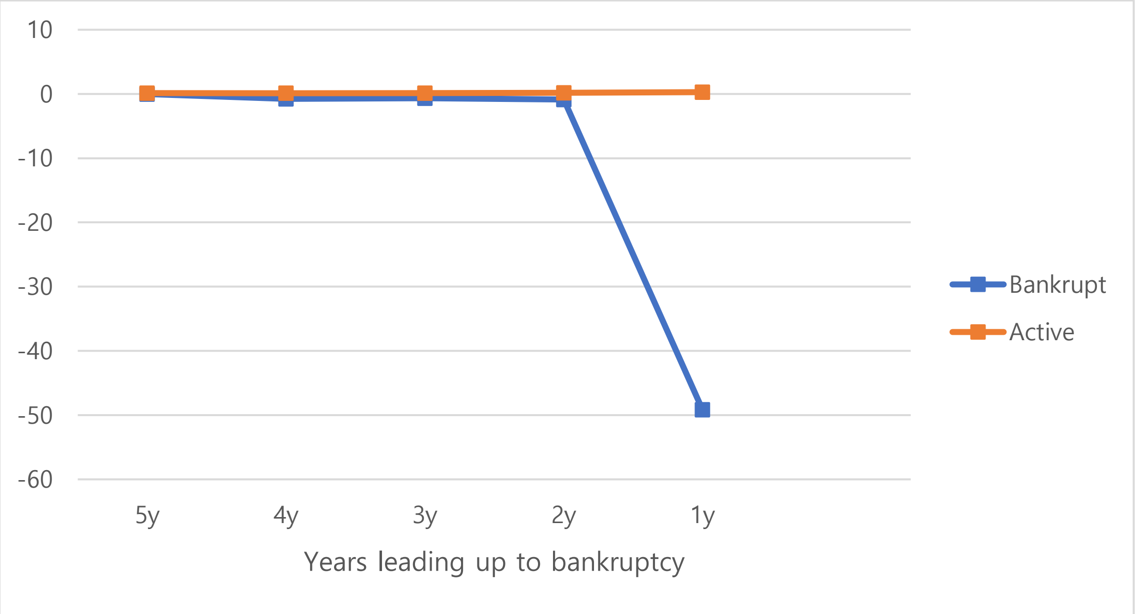
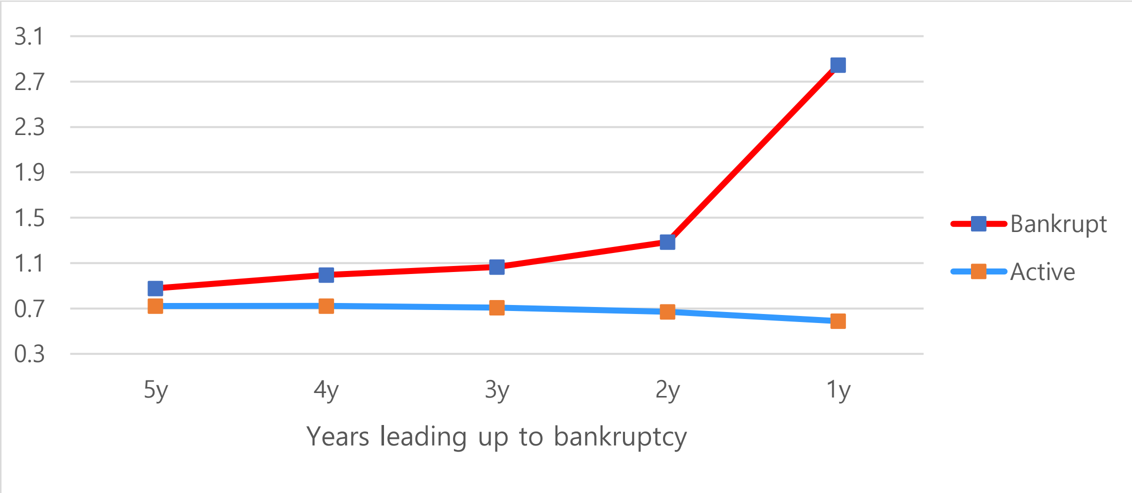


Figure 5.5. Debt ratio over five years prior to bankruptcy



When it comes to liquidity ratio, this trend also can be observed among all variables except cash sales ratio and working capital to equity ratio which show slightly stable during the sample period. Figure 5.6 indicates that the ratio of working capital to total assets is statistically higher in active firms compared to bankrupt firms. The ratios among bankrupt firms show continuously decreased over the sample period and nosedive in one year prior to bankruptcy. In case of current ratio, which is indicate the ability of firm to pay their short-term liabilities, show similar trends with working capital to assets ratios as displayed in Figure 5.7. The current ratio of active firms gradually increased during the sample period, while the ratio of bankrupt firms experienced to decline over the five years.

Figure 5.6. Working capital-to-assets over five years prior to bankruptcy

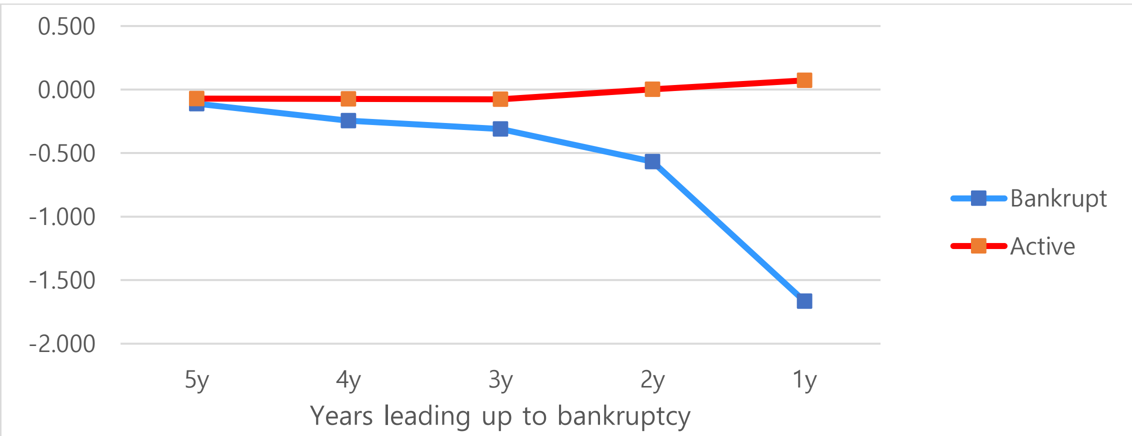
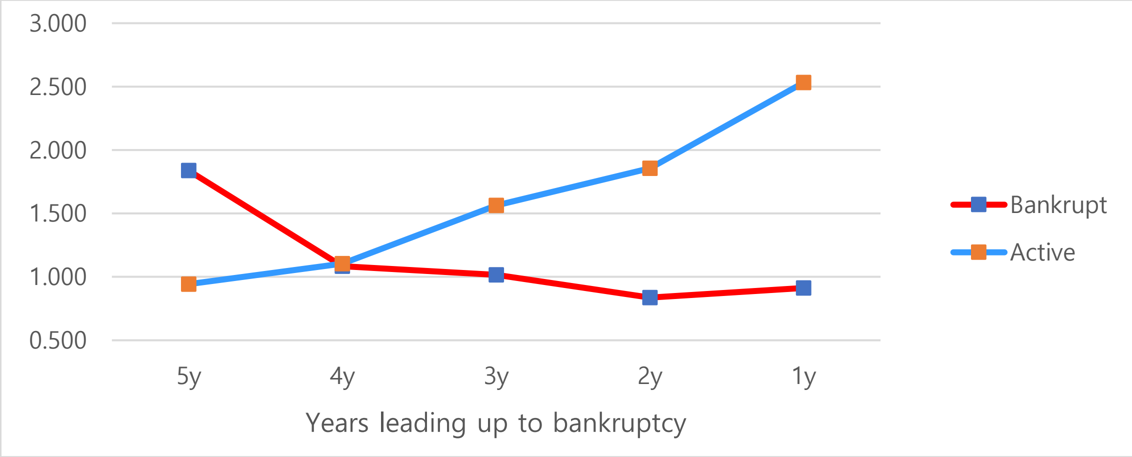


Figure 5.7. Current ratio over five years prior to bankruptcy



Compared to other variables, profitability ratios show greater fluctuations over the sample periods. Among variables, ratios related with profit with bankrupt firms illustrated continuously deterioration trends over the periods. In contrast, mean value of profitability ratios among active firms steadily rise during the sample periods. Figure 5.8 illustrates that the ratio of return on assets, indicating a firm's profitability, was statistically higher in active firms than in bankrupt firms during the observed periods. The return on assets of bankrupt firms has steadily declined over the past four years, and has plummeted in the year before the bankruptcy event. Furthermore, the levels of sales among active firms were statistically higher than among bankrupt firms over the periods. In Figure 5.9, the average value of sales continued to decrease for five years in bankrupt companies, while it steadily increased over time in active companies.

Figure 5.8. Return on assets over five years prior to bankruptcy

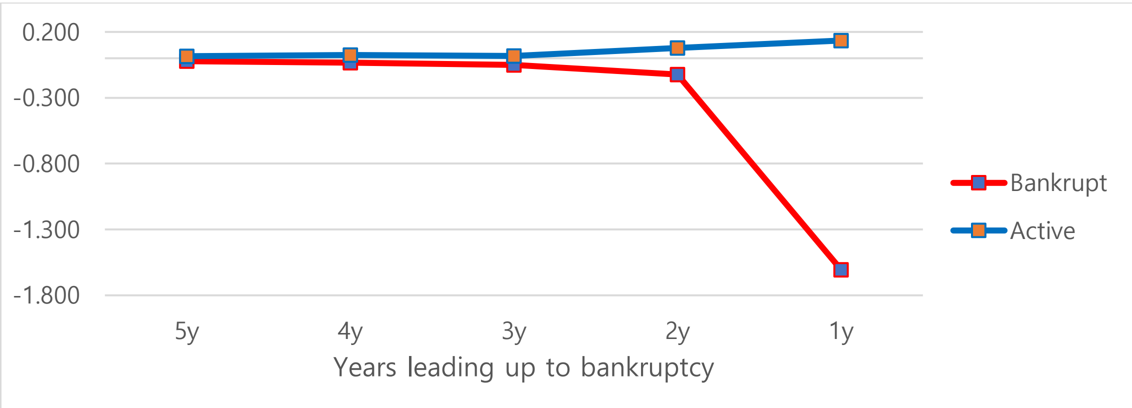
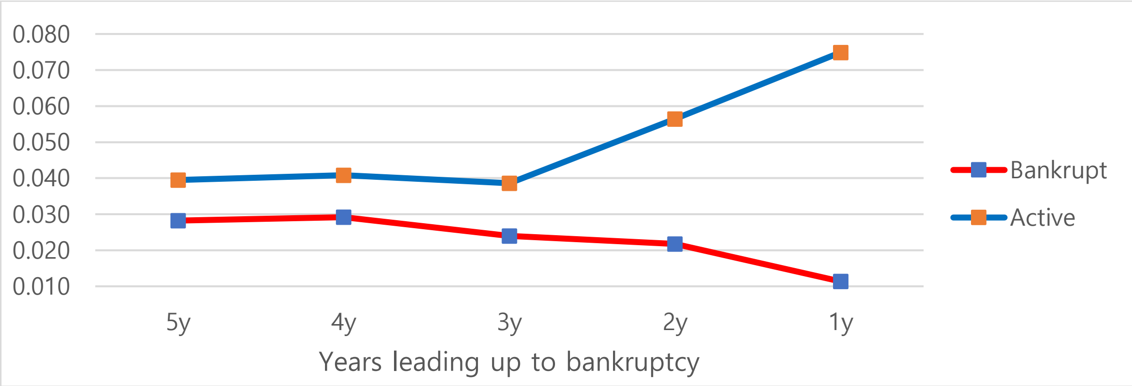


Figure 5.9. Sales over five years prior to bankruptcy



Mean values of ratios in efficiency category either fluctuated or worsen continuously during the periods among bankrupt firms. In Figure 5.10, the mean value of asset turnover ratio among bankrupt slightly fluctuated and decreased over time but active firms managed to stabilized. The gap in the asset turnover ratio between bankrupt and active firms has gradually increased over time. Figure 5.11 illustrates that the ratio of operating margin, which measure how efficiently a firm can generate profit, was statistically greater among active firms than ratios of bankrupt firms over the sample periods. Active firms experienced to slightly decline in two-year periods, but it improved again.

Figure 5.10. Asset turnover ratio over five years prior to bankruptcy

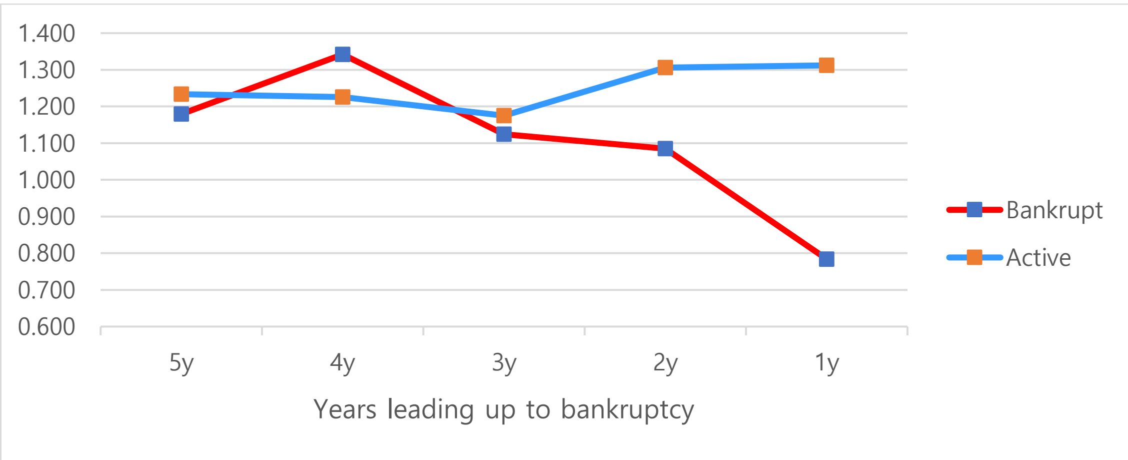
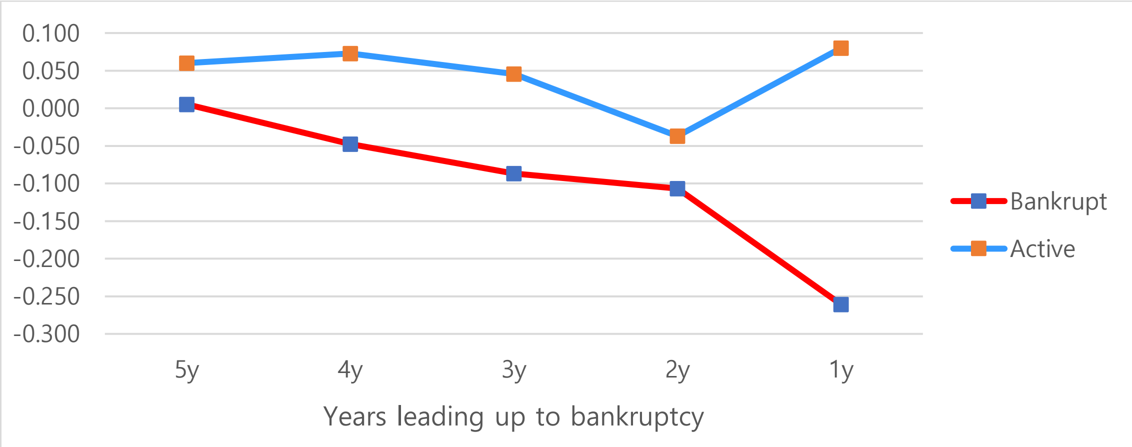
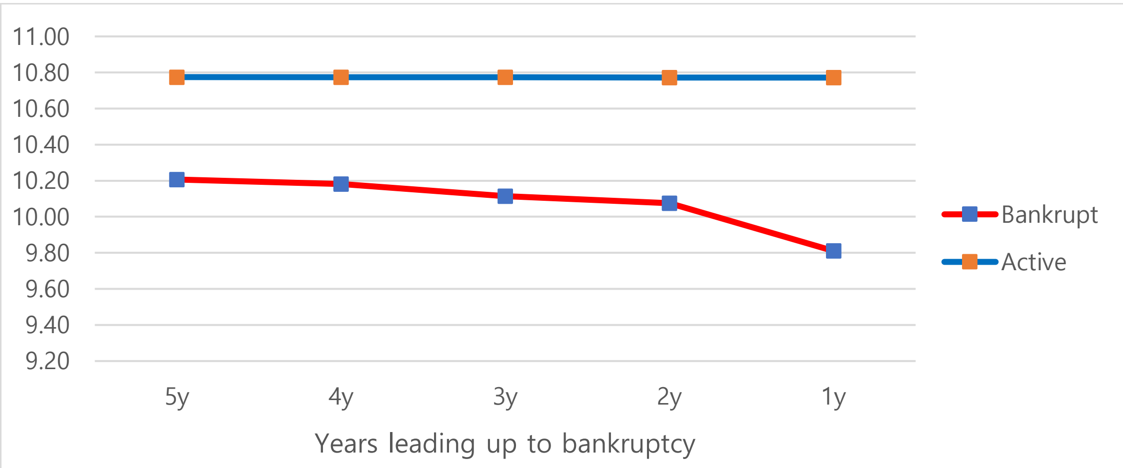


Figure 5.11. Operating margin over five years prior to bankruptcy



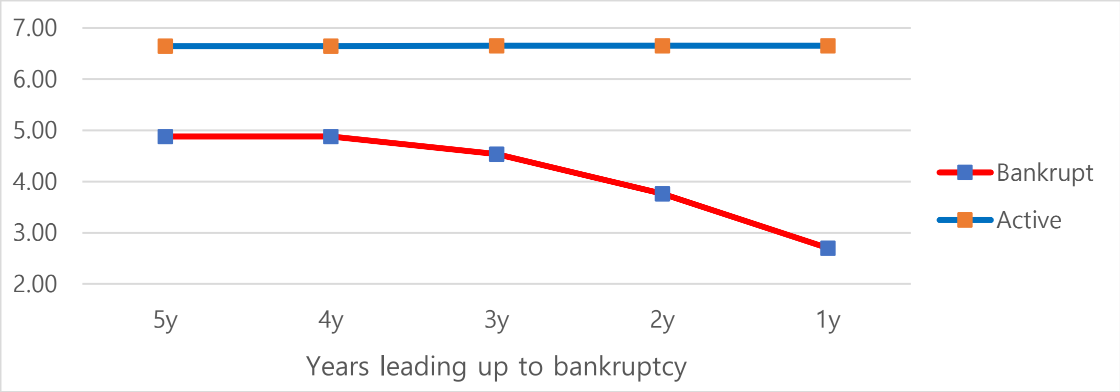
Finally, firm-specific variable showed significant difference between bankrupt and active firms. As the early symptoms of bankruptcy, constant value of natural logarithm of total assets as the proxy variable of size can be suggested over a five-year period leading up to bankruptcy. In Figure 5.12, the mean of logarithm of total assets of bankrupt firms showed steadily declined over the period of five to two years and plummeted a year prior to bankrupt. In contrast, the size of active firms showed greater than value of bankrupt firms consistently.

Figure 5.12. Size of firms over five years prior to bankruptcy



This pattern also can be observed from the number of vessels that company owned during the sample period. Figure 5.13 shows that while the number of vessels owned by active companies has stabilized, the number of vessels owned by bankrupt companies has decreased by approximately 60% over the period. Consequently, the disparity between these two groups of sampled firms expanded over the five years.

Figure 5.13. Number of owned vessels over five years prior to bankruptcy



In conclusion, as bankrupt companies approached bankruptcy, the severity of the financial depression of each firm became more severe, and the gaps in most variables between the two sample groups of companies widened over time. Additionally, each selected variable covering leverage, liquidity, profitability, efficiency, and macroeconomic factors show significant individual characteristics over the sample period. Therefore, it may provide significant insight into early warning signals of bankruptcy in the shipping industry by analysing explanatory power of each variable in different forecasting horizons.

Table 5.7. Descriptive statistics over a five-year prior to bankruptcy

Variable	Year prior to bankruptcy					
	5 years		3 years		1 year	
	Mean	Std	Mean	Std	Mean	Std
A1	-4.95	85.57	-7.96	108.53	-23.98	187.97
A2	-10.68	140.41	-16.97	177.93	-49.95	307.17
A3	10.17	133.44	15.81	169.15	45.56	292.30
A4	10.17	133.44	15.81	169.15	45.56	292.30
A5	-9.17	133.44	-14.81	169.15	-44.56	292.30
A6	0.72	4.01	0.60	3.09	0.78	4.34
A7	15.56	112.65	9.84	87.36	2.01	10.67
A8	15.56	112.65	9.84	87.36	2.01	10.67
A9	0.72	4.01	0.60	3.09	0.78	4.34
A10	10.06	0.79	10.04	0.77	9.99	0.76
A11	4.14	28.67	1.79	8.35	0.33	6.70
A12	0.04	0.73	0.02	0.77	-0.01	1.07
A13	-0.54	7.56	-0.27	4.46	-0.12	2.02
A14	-0.03	0.23	-0.04	0.28	-0.09	0.47
A15	5.96	92.90	9.31	117.82	27.27	203.98
A16	0.04	0.70	0.02	0.74	-0.01	1.07
A17	-0.01	0.05	-0.01	0.05	-0.01	0.06
A18	0.73	2.89	0.64	2.27	0.67	2.75
A19	-0.28	3.47	-0.41	4.31	-1.06	7.36
A20	11.43	92.40	8.05	83.46	1.67	6.57
B1	1.37	5.12	1.31	5.80	2.09	9.65
B2	-3.92	44.08	-6.21	55.84	-17.95	96.16
B3	0.08	0.15	0.08	0.15	0.10	0.18
B4	0.50	2.61	0.54	3.21	0.92	5.21
B5	0.29	0.29	0.29	0.29	0.33	0.31
B6	4.21	44.10	6.50	55.85	18.29	96.18
B7	1.31	5.12	1.26	5.80	2.07	9.65
B8	0.24	1.05	0.24	1.17	0.18	0.61
B9	0.03	0.31	0.02	0.29	-0.04	0.30
B10	0.56	0.32	0.60	0.31	0.67	0.32
B11	0.01	0.03	0.00	0.02	0.00	0.02
B12	0.17	0.64	0.13	0.28	0.15	0.28
B13	27.53	265.75	40.99	336.52	99.66	579.45
B14	-26.53	265.75	-39.99	336.52	-98.66	579.45
B15	0.27	0.28	0.27	0.28	0.32	0.31
B16	2.22	60.96	-0.39	5.42	0.44	4.56
B17	0.37	5.12	0.31	5.80	1.09	9.65
B18	0.28	0.28	0.29	0.28	0.31	0.32
B19	-2.01	30.59	-3.05	38.47	-8.57	66.20
B20	0.15	2.09	0.16	2.58	0.08	2.65

Variable	Year prior to bankruptcy					
	5 years		3 years		1 year	
	Mean	Std	Mean	Std	Mean	Std
C1	-4.95	85.60	-7.97	108.56	-23.99	188.02
C2	-0.57	7.51	-0.28	4.46	-0.13	2.05
C3	-0.43	2.34	-0.64	2.95	-1.11	4.02
C4	-0.43	2.34	-0.63	2.95	-1.11	4.02
C5	11.71	62.53	14.45	78.96	24.37	131.33
C6	0.02	0.12	0.02	0.11	0.01	0.06
C7	911.04	12158.21	726.60	11589.76	2550.92	19837.35
C8	-0.44	9.05	-0.83	11.45	-2.68	19.81
C9	0.36	4.87	0.55	6.17	1.59	10.68
C10	8.03	33.49	6.66	32.86	3.63	13.69
C11	0.01	0.04	0.01	0.04	0.01	0.04
C12	-0.64	9.83	-1.03	12.46	-3.08	21.55
C13	2.32	8.86	1.74	3.70	1.90	4.53
C14	0.73	9.00	0.09	4.42	-0.48	7.47
C15	0.78	8.33	0.27	1.05	0.23	0.90
C16	0.03	0.40	0.00	0.45	-0.10	0.60
C17	0.35	2.45	0.16	2.06	0.37	3.20
C18	4.07	9.58	3.17	5.64	3.80	8.28
D1	1.82	9.08	2.16	11.45	4.31	19.66
D2	-0.76	8.27	-1.48	3.03	-2.31	4.27
D3	1.40	14.68	0.51	1.09	0.59	1.41
D4	0.02	0.04	0.02	0.04	0.01	0.03
D5	2.16	7.88	1.99	3.18	2.91	4.42
D6	0.02	0.10	0.03	0.12	0.03	0.16
D7	-0.10	0.40	-0.15	0.47	-0.27	0.66
E1	10.07	0.73	10	0.75	9.81	0.86
E2	11.34	8.11	12.09	8.21	13.15	8.31
E3	2.12	0.41	2.12	0.41	2.13	0.42
E4	1453.23	181.59	1512.35	149.41	1586.06	133.67
E5	84.82	26.28	79.59	26.54	73.11	23.09
E6	1111.51	48.08	1114.93	47.35	1120.8	42.64
F1	0.8	0.82	0.81	0.76	0.85	0.81
F2	1195.56	128.63	1236.25	105.38	1269.43	80.26
F3	4	4.82	2.73	4.52	3.27	4.52
F4	983.55	344.82	990.78	422.13	1132.61	666.68
F5	1308.12	781.16	1175.5	629.2	1313.59	742.54
F6	230.26	38.59	225.64	40.51	240.54	54.24
F7	4.12	13.69	3.66	12.83	2.68	10.13
F8	154.38	826.29	138.06	751.79	95.44	572.51
F9	241.62	1249.29	218.35	1158.99	172.63	1055.16

5.4.4. Descriptive statistics of large companies and SMEs

Table 5.8 and Table 5.9 provides a summary of the statistics including mean and standard deviation value for variables chosen in this research, contrasting large firms and SMEs in shipping industry. Following same approach in section 5.4.2. financial statement one year prior to bankruptcy were analysed for bankrupt companies in large and SMEs. The analysis reveals stark financial disparities between bankrupt and active firms in both large companies and SMEs. In the case of large companies, bankrupt firms have higher debt ratios, lower retained earnings, and negative equity, highlighting their reliance on debt and accumulated losses. However, bankrupt SMEs demonstrate extreme financial distress with highly negative profitability and solvency ratios, reflecting their higher vulnerability and instability. Bankrupt firms consistently exhibit poorer financial metrics across all indicators, with SMEs showing more severe negative values compared to large companies. Return on assets (ROA) is a crucial indicator of how efficiently a company utilizes its assets to generate earnings. The data reveals a moderate disparity in ROA between bankrupt and non-bankrupt large companies. However, the disparity is markedly pronounced in SMEs, where bankrupt firms exhibit a severe decline in asset utilization compared to their non-bankrupt counterparts. This suggests that SMEs experience a more acute inefficiency in asset utilization when facing financial distress, potentially due to their limited resources and less diversified asset base. Retained earnings provide insight into a company's profitability and its ability to reinvest in its operations. For large companies, bankrupt firms have significantly lower retained earnings compared to non-bankrupt ones. This trend is even more pronounced in SMEs, where bankrupt companies report a dramatically negative ratio, compared to non-bankrupt SMEs. This stark contrast highlights the severe impact of past losses on the financial health of SMEs, indicating that they are less capable of absorbing financial shocks than larger firms.

The debt ratio, which measures the proportion of a company's assets financed by debt, is higher in bankrupt large companies compared to non-bankrupt ones. In SMEs, this ratio is not only higher for bankrupt firms but also exhibits greater variability, indicating extreme financial instability. This substantial difference underscores the heightened vulnerability of SMEs to financial leverage, as excessive debt can quickly lead to insolvency in the absence of sufficient earnings. Equity to Assets ratio is a measure of financial solvency. Bankrupt large companies show negative equity compared to their non-bankrupt counterparts. In SMEs, bankrupt companies display an even more alarming negative ratio, whereas non-bankrupt SMEs maintain a positive ratio. This suggests that SMEs face a more severe depletion of equity during financial distress, which severely undermines their solvency.

Liquidity ratios such as the Current Ratio and Quick Ratio are vital for assessing a company's ability to meet short-term obligations. The current ratio of large bankrupt companies is slightly higher than that of non-bankrupt companies, indicating better short-term liquidity. In contrast, non-bankrupt SMEs show a significantly higher current ratio compared to bankrupt SMEs, suggesting better liquidity management among solvent SMEs. However, the extreme variability in these ratios among SMEs indicates a greater overall liquidity risk.

This comparative analysis highlights significant financial disparities between large companies and SMEs during periods of bankruptcy and non-bankruptcy. SMEs exhibit more extreme financial distress indicators, such as drastically negative ROA, retained earnings, and higher debt ratios, compared to large companies. These findings suggest that SMEs are more vulnerable to financial instability and less resilient to economic shocks. The pronounced financial volatility in SMEs underscores the need for tailored risk management strategies and financial support mechanisms to enhance their stability and sustainability.

Table 5.8. Descriptive statistic of large shipping companies

Variable	Large Companies					
	Total		Bankrupt		Active	
	Mean	Std	Mean	Std	Mean	Std
A1	0.03	0.15	-0.29	0.37	0.04	0.12
A2	0.08	0.43	-0.85	0.78	0.11	0.37
A3	0.71	0.25	1.05	0.26	0.69	0.24
A4	0.71	0.25	1.05	0.26	0.69	0.24
A5	0.30	0.26	-0.05	0.26	0.31	0.25
A6	1.33	6.16	0.02	0.26	1.38	6.27
A7	6.14	23.58	1.74	10.41	6.31	23.93
A8	6.14	23.58	1.74	10.41	6.31	23.93
A9	1.33	6.16	0.02	0.26	1.38	6.27
A10	11.71	0.85	12.07	0.46	11.70	0.86
A11	2.47	10.63	0.37	3.33	2.55	10.81
A12	0.17	0.97	-0.22	0.37	0.18	0.98
A13	0.04	2.14	-0.80	4.74	0.07	1.96
A14	0.09	0.80	-0.59	1.15	0.11	0.77
A15	0.38	0.24	0.68	0.36	0.36	0.23
A16	0.15	0.96	-0.22	0.37	0.17	0.98
A17	0.01	0.11	-0.05	0.15	0.01	0.11
A18	0.69	2.21	0.22	0.04	0.71	2.25
A19	0.17	1.85	-3.21	6.38	0.30	1.25
A20	3.67	14.29	1.36	7.72	3.75	14.47
B1	1.38	3.07	1.95	2.50	1.35	3.08
B2	-0.03	0.21	-0.14	0.38	-0.03	0.20
B3	0.07	0.08	0.05	0.02	0.08	0.09
B4	0.57	2.52	0.29	0.25	0.58	2.57
B5	0.30	0.19	0.22	0.06	0.30	0.19
B6	0.33	0.21	0.36	0.41	0.33	0.20
B7	1.26	2.98	1.82	2.43	1.24	3.00
B8	0.30	1.68	0.05	0.03	0.30	1.71
B9	0.81	1.79	-0.05	0.88	0.84	1.80
B10	0.49	0.25	0.33	0.27	0.50	0.25
B11	0.07	0.15	0.04	0.05	0.08	0.15
B12	0.15	0.54	0.13	0.05	0.15	0.55
B13	1.56	1.74	1.53	1.23	1.56	1.76
B14	-0.56	1.74	-0.53	1.23	-0.56	1.76
B15	0.27	0.17	0.20	0.06	0.27	0.18
B16	-0.42	3.95	0.01	1.84	-0.43	4.01
B17	0.38	3.07	0.95	2.50	0.35	3.08
B18	0.26	0.19	0.22	0.11	0.27	0.19
B19	0.11	7.76	-12.80	23.72	0.60	5.88
B20	0.26	0.84	-0.32	0.14	0.28	0.85
C1	0.03	0.15	-0.29	0.37	0.04	0.12
C2	0.08	0.43	-0.85	0.78	0.11	0.37
C3	0.71	0.25	1.05	0.26	0.69	0.24

Variable	Large Companies					
	Total		**Bankrupt**		**Active**	
	Mean	**Std**	**Mean**	**Mean**	**Std**	**Mean**
C4	0.02	0.15	-0.29	0.37	0.03	0.12
C5	0.00	2.20	-0.86	4.81	0.04	2.02
C6	-0.70	8.22	-1.00	1.68	-0.68	8.36
C7	-0.68	8.12	-1.00	1.68	-0.67	8.26
C8	3.73	2.44	1.91	0.89	3.80	2.46
C9	0.25	0.40	0.11	0.13	0.25	0.41
C10	138.29	1180.80	-5.73	7.37	143.75	1202.61
C11	0.10	0.10	-0.01	0.07	0.11	0.10
C12	0.03	0.02	0.06	0.03	0.02	0.02
C13	7.93	50.86	1.49	4.55	8.17	51.79
C14	0.06	0.09	0.10	0.07	0.06	0.09
C15	0.05	0.08	-0.05	0.07	0.06	0.08
C16	1.98	2.57	0.40	0.19	2.04	2.60
C17	0.48	1.25	-0.11	0.59	0.51	1.26
C18	0.28	1.01	0.00	0.07	0.29	1.03
D1	0.09	0.43	-0.02	0.19	0.10	0.44
D2	0.49	2.68	-0.33	0.20	0.52	2.73
D3	3.77	3.36	2.14	1.35	3.83	3.40
D4	1.08	0.89	0.40	0.21	1.10	0.89
D5	-3.37	36.70	-0.08	0.90	-3.50	37.39
D6	0.94	5.02	0.74	0.51	0.95	5.11
D7	0.03	0.04	0.02	0.02	0.03	0.04
E1	4.32	41.56	0.82	0.65	4.45	42.33
E2	0.05	0.16	0.08	0.08	0.05	0.16
E3	-0.04	1.30	-0.18	0.28	-0.03	1.32
E4	11.91	0.77	12.06	0.48	11.90	0.77
E5	21.65	15.42	34.60	17.16	21.16	15.13
E6	1.80	0.75	1.50	0.50	1.81	0.76
F1	3.47	1.95	2.83	1.91	3.49	1.95
F2	70.92	27.67	85.38	26.57	70.38	27.57
F3	1129.68	91.44	1118.58	50.40	1130.10	92.62
F4	1.61	1.61	0.75	0.60	1.64	1.62
F5	3.21	3.30	2.45	2.46	3.24	3.33
F6	3.86	5.13	2.33	5.15	3.91	5.12
F7	1190.57	751.22	918.54	175.90	1200.87	762.64
F8	2142.62	1627.32	1090.10	261.79	2182.48	1643.88
F9	234.95	92.17	221.40	23.30	235.46	93.75

Table 5.9. Descriptive statistic of shipping SMEs

Variable	Shipping SMEs					
	Total		Bankrupt		Active	
	Mean	Std	Mean	Std	Mean	Std
A1	-0.61	31.00	-5.21	87.76	0.05	0.20
A2	-1.28	50.92	-11.26	143.99	0.14	0.38
A3	1.97	48.37	10.68	136.85	0.73	0.33
A4	1.97	48.37	10.68	136.85	0.73	0.33
A5	-0.97	48.37	-9.68	136.85	0.27	0.33
A6	7.37	216.96	0.54	2.62	8.33	231.83
A7	6.30	73.50	16.20	115.49	4.90	65.26
A8	6.30	73.50	16.20	115.49	4.90	65.26
A9	7.37	216.96	0.54	2.62	8.33	231.83
A10	10.34	0.84	10.00	0.57	10.39	0.86
A11	2.54	19.24	4.29	29.38	2.30	17.32
A12	0.09	1.74	0.02	0.63	0.10	1.85
A13	-0.16	7.08	-0.56	7.74	-0.11	6.98
A14	0.01	0.07	0.00	0.00	0.01	0.07
A15	1.11	33.66	6.28	95.28	0.37	0.27
A16	0.08	1.73	0.02	0.61	0.09	1.84
A17	0.00	0.01	0.00	0.00	0.00	0.01
A18	1.79	30.98	0.60	1.96	1.95	33.09
A19	0.23	3.99	-0.31	3.53	0.30	4.05
A20	3.76	65.19	11.91	94.75	2.60	59.73
B1	3.24	39.69	1.26	4.78	3.53	42.36
B2	-0.56	16.00	-4.12	45.21	-0.05	0.33
B3	0.08	0.12	0.08	0.15	0.08	0.11
B4	1.35	24.80	0.50	2.66	1.48	26.48
B5	0.31	0.26	0.29	0.29	0.31	0.25
B6	0.86	16.00	4.41	45.22	0.36	0.29
B7	3.17	39.68	1.20	4.77	3.45	42.36
B8	0.92	23.79	0.23	1.04	1.02	25.42
B9	0.06	0.21	0.00	0.01	0.06	0.23
B10	0.52	0.28	0.55	0.32	0.52	0.27
B11	0.00	0.01	0.00	0.00	0.00	0.02
B12	1.23	33.17	0.17	0.65	1.38	35.44
B13	5.93	96.84	28.58	272.52	2.71	8.97
B14	-4.93	96.84	-27.58	272.52	-1.71	8.97
B15	0.29	0.25	0.27	0.28	0.29	0.25
B16	-0.15	22.71	2.31	62.52	-0.50	5.76
B17	2.24	39.69	0.26	4.78	2.53	42.36
B18	0.27	0.24	0.29	0.28	0.27	0.23
B19	0.95	17.94	-2.22	31.34	1.40	15.04
B20	0.99	17.78	0.04	1.35	1.13	18.99
C1	-0.61	31.00	-5.21	87.76	0.05	0.20
C2	-1.28	50.92	-11.26	143.99	0.14	0.38
C3	1.97	48.37	10.68	136.85	0.73	0.33

Variable	Shipping SMEs					
	Total		Bankrupt		Active	
	Mean	Std	Mean	Mean	Std	Mean
C4	-0.61	31.01	-5.22	87.79	0.04	0.19
C5	-0.18	7.06	-0.59	7.68	-0.13	6.97
C6	-0.52	18.00	-0.45	2.40	-0.53	19.21
C7	-0.16	9.58	-0.45	2.40	-0.12	10.19
C8	7.84	36.45	12.17	64.11	7.23	30.50
C9	0.01	0.04	0.00	0.00	0.01	0.04
C10	4316.76	149793.37	958.54	12468.98	4793.67	160001.92
C11	0.18	3.36	-0.51	9.26	0.28	0.81
C12	0.08	1.77	0.37	4.99	0.03	0.18
C13	7.79	70.92	8.07	34.34	7.75	74.67
C14	0.01	0.02	0.00	0.00	0.01	0.02
C15	-0.03	3.57	-0.67	10.08	0.06	0.19
C16	3.73	24.40	1.84	3.46	4.00	26.04
C17	1.78	19.06	0.20	3.60	2.01	20.31
C18	0.86	6.98	0.24	0.89	0.95	7.44
D1	0.15	0.48	0.02	0.40	0.17	0.48
D2	6.42	144.75	0.31	2.40	7.29	154.66
D3	6.52	31.33	3.69	5.30	6.93	33.40
D4	1.51	4.06	1.92	9.33	1.45	2.53
D5	1.19	103.53	-0.78	8.48	1.47	110.59
D6	6.57	229.05	1.45	15.06	7.30	244.70
D7	0.02	0.04	0.02	0.04	0.02	0.04
E1	5.38	141.03	2.23	8.08	5.83	150.67
E2	0.18	4.95	0.02	0.10	0.21	5.29
E3	-0.20	8.86	-0.11	0.41	-0.21	9.47
E4	10.50	0.79	10.00	0.52	10.57	0.79
E5	16.94	14.12	11.05	6.86	17.78	14.67
E6	2.21	0.54	2.14	0.35	2.21	0.56
F1	3.43	1.95	3.34	1.69	3.44	1.98
F2	74.26	27.03	84.54	26.23	72.80	26.83
F3	1127.88	84.63	1111.59	48.79	1130.20	88.31
F4	1.47	1.55	0.82	0.87	1.56	1.61
F5	3.17	3.25	3.05	2.36	3.19	3.36
F6	3.75	5.19	3.92	4.86	3.73	5.24
F7	1187.80	757.98	1001.99	402.91	1214.18	792.13
F8	2004.88	1555.59	1291.10	727.12	2106.25	1614.25
F9	240.93	85.32	231.40	40.85	242.28	89.78

5.5. Chapter Summary

This chapter has outlined the development of bankruptcy prediction models for the Korean shipping industry, leveraging both interviews and descriptive analysis. Initially, engagements with industry experts were undertaken to not only validate the reliability of explanatory variables identified from literature reviews but also to gather deeper insights into the bankruptcy risks facing the shipping sector. The significance of each variable was assessed based on the interviewees' evaluations of their importance, leading to the identification of key variables such as the LIBOR interest rate, the status of vessel ownership, and Korea's annual trade volume as critical macroeconomic factors. These identified variables have thus been incorporated into the explanatory variable sets for model development.

Descriptive statistical analysis of these variable sets revealed distinct financial characteristics of firms facing bankruptcy, including higher leverage, reduced profitability and liquidity, and diminished efficiency in profit generation compared to their non-bankrupt firms. In particular, as companies neared bankruptcy, their financial distress intensified, and the disparity in most financial indicators between active and bankrupt firms became more pronounced.

The descriptive statistics highlight the significant individual characteristics of selected variables across various forecasting periods, covering aspects such as leverage, liquidity, profitability, efficiency, and macroeconomic factors. Therefore, this analysis offers valuable insights, serving as a preliminary indication of potential early warning signs of bankruptcy within the shipping industry by examining the explanatory power of each variable across different forecasting horizons.

6. EMPIRICAL ANALYSIS

6.1. Introduction

This chapter advances the descriptive analysis presented in Chapter 5, which preliminarily explored the predictive capacity of risk factors identified through literature reviews and interviews, aiming to forecast bankruptcy within the shipping industry using these variables. Continuing from the discussions in the previous chapter, this investigation deploys two sophisticated machine learning algorithms—extreme gradient boosting and long short-term memory (LSTM) models—as classification mechanisms over diverse datasets. The objective is to anticipate bankruptcy occurrences in the target year "t" based on variables from the preceding year "t-s," where "s" equals 1, 3, or 5 years, thereby covering a spectrum of forecasting intervals.

The study examines four specialized datasets that incorporate both financial ratios and macroeconomic elements, like shipping indices, designed to address the nuances of both large corporations and small to medium-sized enterprises (SMEs). The operational status of the companies under scrutiny is categorized through a binary system: '0' signifies an active or healthy company, and '1' indicates a company that has encountered bankruptcy, delisting, or has undergone significant economic depression such as mergers, acquisitions, restructuring, or has been excluded from the Korean Shipping Owner Organization.

The predictive efficacy of each model is evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC). The inherent imbalance found in bankruptcy datasets renders traditional accuracy metrics insufficient for thorough evaluation. Conversely, the ROC curve provides a holistic assessment of a model's performance throughout its operational range and enables an intuitive comparison across different models. It is essential to recognize that while the ROC curve serves as a testament to the predictive prowess of classification algorithms, it does not encapsulate the entirety of a classifier's predictive capacity. Consequently, the occurrence of Type-II errors, where a financially distressed company is misclassified as solvent, becomes a critical measure for gauging model performance. Type-II errors carry significant implications in bankruptcy prediction, indicating potential financial losses in terms of principal and interest for investors or financial institutions, in addition to the costs incurred from bankruptcy proceedings (Muller et al., 2009). Hence, this type of error is accorded paramount importance over other evaluative criteria for its profound impact. In conclusion, a prediction model that achieves the highest ROC value while minimizing Type-II errors is considered optimal for this study.

6.2. Data Preprocessing

Data preprocessing is a vital process that converts raw data into a format ready for analysis, tackling prevalent issues such as incompleteness, inconsistencies, and errors commonly found in real-world financial datasets (Agarwal and Taffler, 2008). These datasets often contain null values, making them unsuitable for direct application in data mining processes (Lukáč et al., 2022). The rapid increase in the generation and size of datasets across business, industrial, and academic fields necessitate sophisticated research models capable of analysing larger and more complex datasets while effectively handling errors.

The application of data preprocessing techniques enables the adjustment of datasets for compatibility with various machine learning algorithms, thereby facilitating the analysis of previously infeasible data. Furthermore, certain machine learning models demonstrate enhanced performance with specific data distributions, underscoring the importance of preprocessing to ensure data suitability (Kim et al., 2022). A meticulous preparation and screening of original data for missing values, normality of data distribution, and outliers are critical steps prior to the generation of an input matrix and the subsequent analysis within any machine learning framework (García et al., 2016). As a result, this study employs several preprocessing techniques to not only enhance the prediction performance of the model, but also protect the robust prediction capabilities of the machine learning model through careful examination the collected data.

The initial phase of preprocessing in this predictive modelling study involves addressing missing data through imputation techniques. Missing data in datasets can lead to several statistical issues, such as non-convergence, biased parameter estimates, and inflated fit indices, necessitating diverse imputation methods based on the proportion of missing data within the dataset (Wang et al., 2022). Subsequently, the process involves adjusting for outliers and skewness in the dataset. Outliers are mitigated through winsorization, a statistical transformation that limits extreme data values to reduce the influence of potential outliers, with 0.1% of data from both tails being winsorized.

Finally, to address imbalances in datasets, particularly in the context of bankruptcy datasets characterized by a high imbalance between bankrupt and active firms due to the rarity of bankruptcy cases, the Synthetic Minority Oversampling Technique (SMOTE) is applied. Specifically, the dataset comprises, on average, 13 bankrupt firms and 109 active firms annually, underscoring the significant imbalance challenge. Through the application of SMOTE, classification models can not only enhance their performance but also discern complex patterns among explanatory variables, thus ensuring a high level of predictive accuracy and evaluating the impact of each variable (Le, 2022).

6.2.1. Missing data imputation

In the real-world financial datasets of sampled firms, the occurrence of missing data in various variables is a common issue. This primarily stems from the substantial costs associated with reporting complete financial statements to national registries, which can be particularly burdensome for Small and Medium-sized Enterprises (SMEs) due to the complexity of accounting procedures required annually (Calabrese et al., 2016). Consequently, SMEs that do not meet certain size criteria often opt not to report complete information, resulting in datasets with incomplete financial factors, such as the number of employees, level of inventories, or cash flow (Mayr et al., 2021). These gaps necessitate specific handling strategies to ensure the integrity of the dataset for analysis.

To address these missing data, our methodology involves two distinct imputation techniques, contingent upon the proportion of missing data within the total dataset, as detailed in Chapter 4.7. For variables where more than 1% of data is missing, this study employs interpolation techniques, replacing missing values with the average of the neighbouring values before and after the gap. Conversely, for variables with missing data constituting less than 1% of the total, these values would be substituted with those from the preceding year. This approach is underpinned by the observation that financial ratios in year t can typically be derived from data in years $t-1$ and $t-2$, mitigating potential econometric issues (Kim et al., 2022).

An analysis of our datasets reveals that the proportion of missing values for most variables is relatively minor, typically less than 1%, with exceptions noted in certain key variables such as net interest margin (C9), inventory to assets ratio (D4), and inventory turnover (D6), which exhibit a significant frequency of missing values. Specifically, Table 6.1 delineates the proportion of missing values among explanatory variables, distinguishing between groups of bankrupt and active firms. It was found that financial ratios pertaining to interest expense and inventories had noticeable gaps, especially among bankrupt firms, with missing values accounted for 3.3% and 7.2%, respectively. Although such gaps were also present in active firms, they were considerably less prevalent, remaining under 1% and thus comparatively low.

Given the marked disparity in the distribution of missing values between bankrupt and active firms, a differentiated approach to imputation is warranted. For bankrupt firms with missing values exceeding 1%, mean values derived via interpolation will be utilized for substitution. Meanwhile, for the remainder of the missing data, values from the previous year will be adopted. This tailored strategy acknowledges the distinct patterns of missing information across firm types, ensuring a more nuanced and effective data preparation process (Zhou and Lai, 2017).

Table 6.1. Proportion of missing values by group

Variable	Bankrupt	Active
Net interest margin(C9)	3.3%	0.7%
Inventory to Assets(D4)	7.2%	1.1%
Inventory turnover(D6)	7.2%	1.1%

6.2.2. Reduce skewness & outliner

Outliers are observations that significantly deviate from the bulk of data in a dataset, introducing potential biases in statistical analyses (Lohmann et al., 2023). Hair (2009) suggests that researchers must critically assess whether to retain or exclude outliers from their dataset. In the context of predicting bankruptcy within the shipping industry, this study recognizes that outliers may often represent early warning signs. Consequently, a conservative approach towards outlier criteria is adopted, minimizing the exclusion of potentially predictive outliers.

To mitigate the impact of outliers on the statistical integrity of our results, this research follows the method proposed by Shumway (2001), where values beyond the 99th percentile and below the 1st percentile for each variable are capped at these thresholds. This technique, known as winsorization, is applied to both active and bankrupt firms, ensuring that outliers at both the 1st and 99th percentiles are adjusted to reduce their influence on the analysis.

Furthermore, bankruptcy risk factors typically exhibit a highly skewed distribution, as depicted in Figure 6.1. Such skewness can impair the predictive accuracy of machine learning algorithms, which often assume data normality. To address this, the study employs the Box-Cox transformation technique as a preprocessing step to normalize the distribution of financial variables. This method effectively reduces skewness, as evidenced by the variables' distribution post-transformation, illustrated in Figure 6.2.

By applying the Box-Cox transformation, this study not only normalizes the distribution of variables but also enhances the suitability of the transformed data for machine learning models, thereby improving the predictive performance in bankruptcy detection. The careful handling of outliers and skewness demonstrates the comprehensive approach taken to maintain the predictive relevance of the dataset while optimizing it for analysis.

Figure 6.1. Histograms of variables before preprocessing techniques

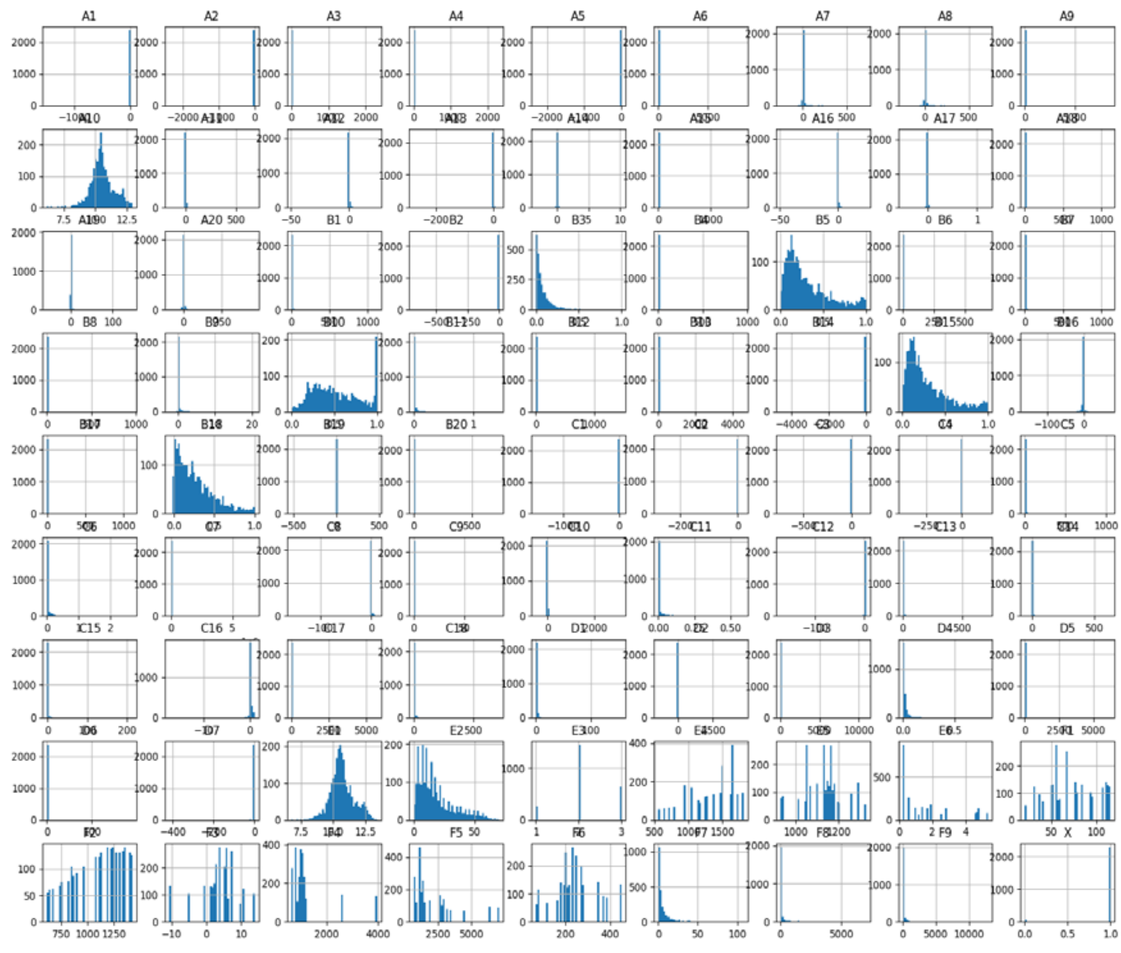
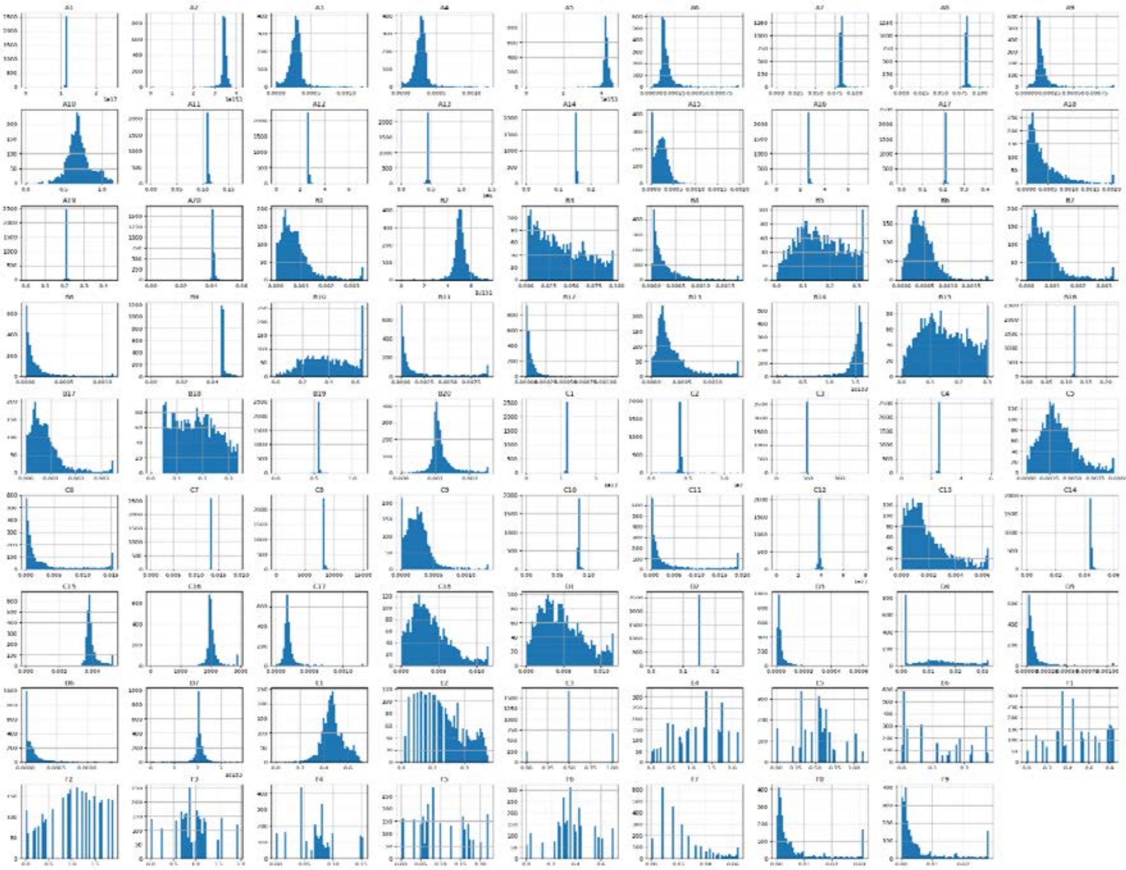


Figure 6.2. Histograms of variables after preprocessing techniques



6.2.3. Handling imbalanced dataset

This research faces a pronounced imbalance in the dataset used for bankruptcy prediction, with a clear gap observed between the number of bankrupt and active firms. Specifically, within a one-year forecasting horizon, the dataset comprises 46 records for the minority (bankrupt) class and 1,727 records for the majority (active) class, representing 2.5% and 97.5% of the dataset, respectively. This imbalance extends to longer forecasting horizons as well, with the minority class constituting 8% and 12% of the datasets for three-year and five-year forecasts, respectively. As outlined in the Figure 6.3, the dataset is initially partitioned into training and testing subsets, maintaining an 80% and 20% split. Subsequently, within the training set, the minority class is oversampled to match the number of records in the majority class, effectively correcting the initial imbalance. Table

6.2 presents the number of datasets before and after oversampling techniques over the forecasting horizons.

Finally, Figure 6.4 illustrates a discriminant plot clustering into bankrupt and active firms using variables of bankruptcy prediction. The challenges caused by the dataset's imbalance are significant overlap between the two classes, complicating the classification task and potentially diminishing the predictive performance of machine learning algorithms (Zoričák et al., 2020). Based on these observations, the use of the Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset is recommended. This strategy is anticipated to enhance the model's predictive accuracy by allowing for more intricate modelling techniques and effectively addressing the inherent complexities associated with imbalanced datasets in bankruptcy prediction (Veganzones and Séverin, 2018).

Figure 6.3. Flow chart of SMOTE Method

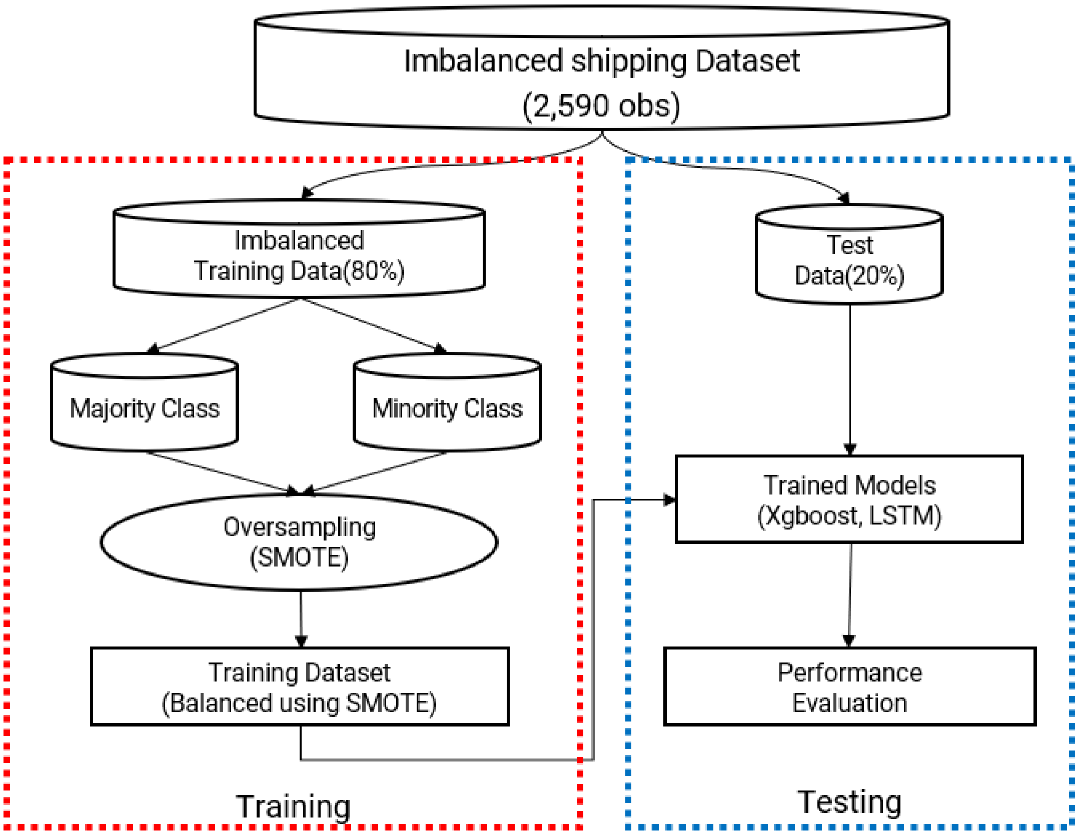
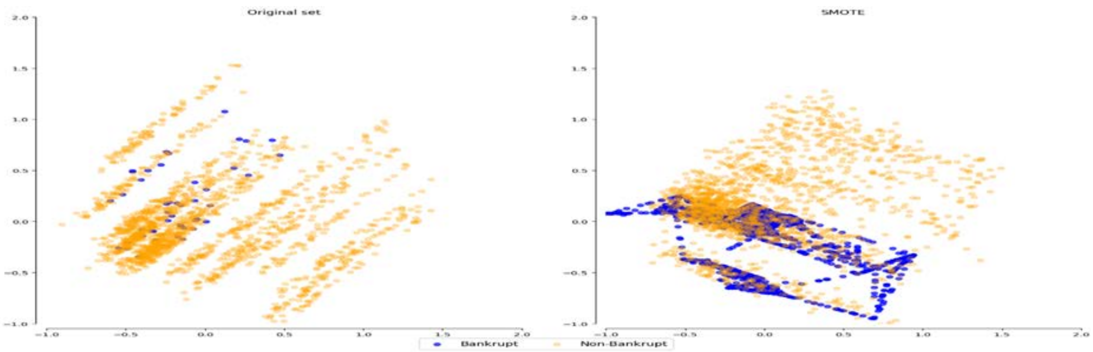


Table 6.2. Number of datasets before and after Oversampling technique

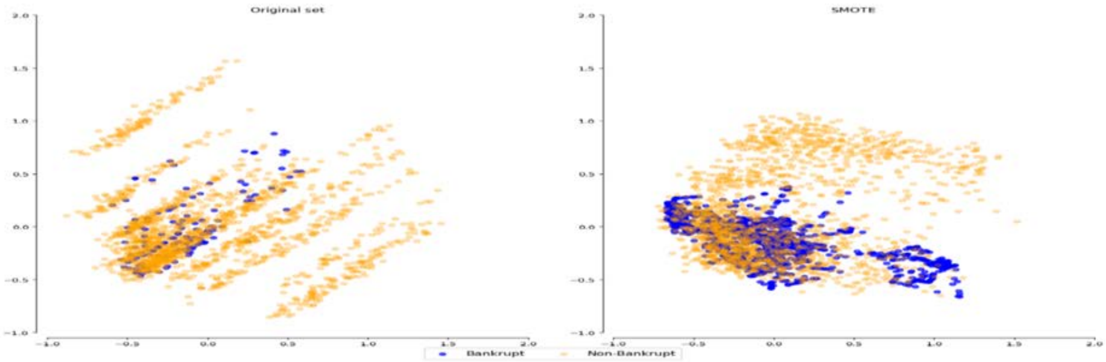
		Before	After
1 year	Active	1727	1727
	Bankrupt	46	1727
3 years	Active	1833	1833
	Bankrupt	159	1833
5 years	Active	1831	1831
	Bankrupt	253	1831

Figure 6.4. Scatter charts of each dataset by different forecasting horizon

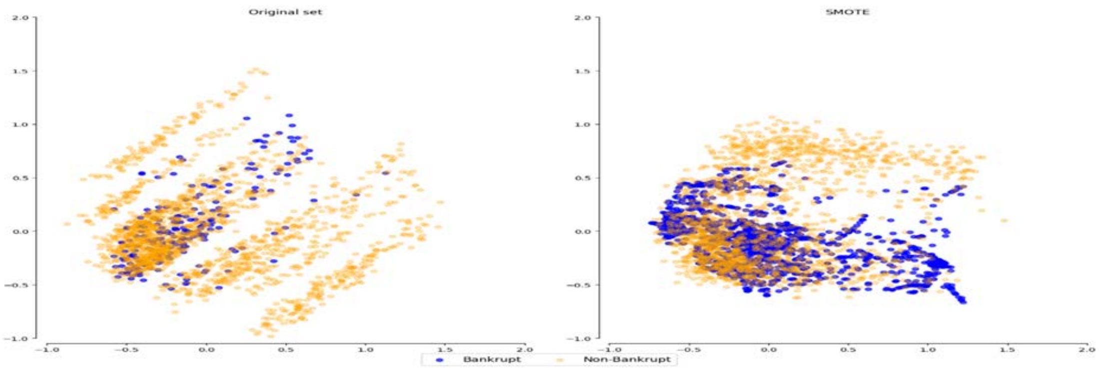
Dataset(1 year)



Dataset(3 years)



Dataset(5 years)



6.3. Empirical result

This research trained two different models of extreme gradient boosting (Xgboost) and Long short-term memory (LSTM) selected from literature review on three different datasets (all shipping firms, large shipping firms and SMEs), segmented over three forecasting periods. Each of the models consisted of different compositions as suggested in Table 6.3. While, Model 1 includes only financial metrics such as leverage, profitability, liquidity, and efficiency ratios. Model 2 incorporates not only financial metrics but also macroeconomic variables, including shipping indices and non-financial variables. This approach highlights the significance of industry-specific variables for efficient bankruptcy prediction tailored to the specific industry. Furthermore, Model 3 and Model 4 focus on large firms and SMEs, respectively, to emphasize their distinct characteristics and provide practical insights for different approaches based on the size of the company.

Table 6.3. Summary of Model Compositions and Variables

Model	Target Data	Variables
Model 1: Financial Metrics Bankruptcy Prediction Model	All shipping firms	Financial metrics (Leverage, Profitability, Liquidity, Efficiency)
Model 2: Integrated Financial and Industry-Specific Model	All shipping firms	Financial metrics (Leverage, Profitability, Liquidity, Efficiency), Macroeconomic, Non-financial
Model 3: Large Firm Bankruptcy Prediction Model	Large shipping firms	Financial metrics (Leverage, Profitability, Liquidity, Efficiency), Macroeconomic, Non-financial
Model 4: SME-Specific Bankruptcy Prediction Model	Shipping SMEs	Financial metrics (Leverage, Profitability, Liquidity, Efficiency), Macroeconomic, Non-financial

Typically, machine learning algorithms necessitate setting learning parameters and hyperparameters. While learning parameters are determined during the model training process, hyperparameters must be established before training begins, with specific values chosen in advance. Setting hyperparameters fitting each model has a large influence on the performance and behaviour of machine learning algorithms (Shetty et al., 2022).

To tune the hyperparameters efficiently, Bayesian optimization was applied as the method to find parameters by maximizing an arbitrary objective function. Bayesian optimization iteratively uses the estimated value of the objective function from the previous parameter to determine the value of the next parameter, leveraging prior knowledge to inform the search process (Xia et al., 2017). This approach was chosen for its ability to efficiently explore the hyperparameter space and its effectiveness in handling complex, multimodal functions, which allows it to focus on promising regions of the hyperparameter space more effectively than random or grid search methods.

Bayesian optimization was implemented using the hyperopt package in Python. During the tuning process, multiple iterations were performed to refine the hyperparameters, ensuring that the model's performance was maximized. The selected hyperparameters were validated through cross-validation to confirm their efficacy. All hyperparameters identified through this rigorous tuning and validation process are presented in Table 6.4., showcasing the values that led to the best model performance.

Table 6.4. The hyperparameters for machine-learning models

Model	Hyper parameters
XGBoost	colsample_bytree=0.812 learning_rate=0.197 max_depth=7 min_child_weight=1
LSTM	Learning rate=0.001 Epochs=1000 Dropout=0.2 Batch_size=64 Activation Function=Relu Recurrent Activation Function=Sigmond Optimization Function=Adam Loss= binary_crossentropy

To build a prediction model, training and test data were randomly divided in an 8:2 ratio. The training data underwent evaluation and verification using K-fold cross-validation to prevent overfitting, a situation where a machine learning model predicts accurately on training data but fails with new data (Alam et al., 2021b). In K-fold cross-validation, samples are randomly divided into k equally sized folds. While there's no universally optimal number for k, this study employs a 5-fold cross-validation process, commonly preferred in prior research.

6.3.1. Model 1 – Financial Metrics Bankruptcy Prediction Model

This section presents the results for each of both classification models with analysing only financial metrics across three forecasting horizons. The results of each model show in Table 6.5 with 6 different criteria.

Table 6.5. Results of model 1 with three different forecasting horizons

Year	Model	Performance					
		ROC	Accuracy	Sensitivity	Specificity	Type-I Error	Type-II Error
1y	Xgboost	**0.8861**	0.8977	**0.7895**	0.9018	0.0982	**0.2105**
	LSTM	0.7515	**0.9286**	0.2632	**0.9539**	**0.0461**	0.7368
3y	Xgboost	**0.9027**	**0.8861**	**0.5769**	**0.9206**	**0.0794**	**0.4231**
	LSTM	0.7439	0.8436	0.4038	0.8927	0.1073	0.5962
5y	Xgboost	**0.8698**	**0.8571**	**0.7027**	**0.8829**	**0.1171**	**0.2973**
	LSTM	0.7013	0.7625	0.4730	0.8108	0.1892	0.5270

According to Table 6.5, Xgboost models performed better than LSTM models for one year prior to bankruptcy. The ROC curve value of Xgboost and LSTM models are 0.8861 and 0.7515, respectively. In case of type-II errors, which is referred classify the bankrupt firms as the active firms, Xgboost model shows better performance compared to LSTM model, which produced 73.6%. Although LSTM produced better type-I errors (4.6%) than

Xgboost (9.8%) in one year forecasting period, this result suggests strong ability of Xgboost as the bankruptcy prediction model considering imbalanced dataset.

Table 6.6 shows the classification result of using Xgboost for one year prior to bankruptcy. The model correctly predicts 450 active firms and 15 bankrupt firms. Out of 518 firms from the dataset, 90.2% of active firms are correctly classified, and 78.9% of firms which failed after one-year are correctly classified. Therefore, this model can correctly classify 89.8% of firms.

Table 6.6. Confusion matrix of Model 1 - One year prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	450	49	499
Failed	4	15	19
Total	454	64	518

These results also can be observed from the dataset for three years prior to bankruptcy. Performance of Xgboost showed better than LSTM models according to ROC value, 0.9027 and 0.7439, respectively. Xgboost model shows relatively lower level of type-I errors, 7.9% and type-II errors, 42.3%, compared to LSTM model, which produced 10.7% of type-I errors and 59.6% of type-II errors. Table 6.7 shows classification performance of Xgboost model for three years prior to bankruptcy. This model can predict 429 active firms and 30 bankrupt firms correctly. Among 518 firms in this model, 92.1% of active firms are correctly classified, and 57.1% of firms which failed to bankrupt after three years are correctly classified. Therefore, this model can classify 88.6% of the overall firms correctly.

Table 6.7. Confusion matrix of Model 1 - Three years prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	429	37	466
Failed	22	30	52
Total	451	67	518

In case of models for five years prior to bankruptcy, Xgboost also performed better prediction ability compared to LSTM model according to ROC curve, 0.8698 and 0.7013 respectively. Also, Xgboost model produced 11.7% of type-I errors and 29.7% of type-II errors, which is lower lever than 18.9% and 52.7% of LSTM models. This can be referred as the higher prediction ability to avoid risk from classifying financially distressed firms which can be bankrupt five years later as the active firms.

Table 6.8 shows classification performance of Xgboost model for five years prior to bankruptcy. Among 518 total firms, this model can correctly predict 392 active firms and 52 bankrupt firms. From the dataset, 88.2% of active firms are correctly classified, and 70.3% of firms which failed to bankrupt after three years are correctly classified. Overall, 85.7% of firms are correctly classified by this model.

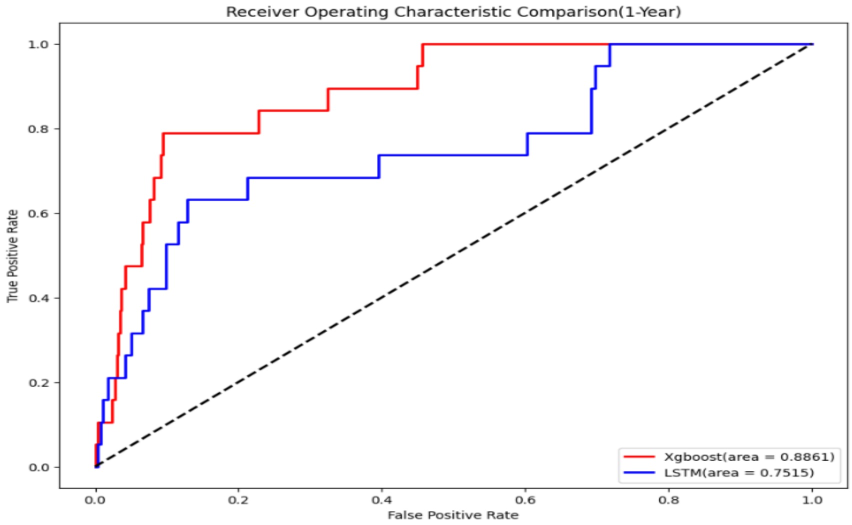
Table 6.8. Confusion matrix of Model 1 - Five years prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	392	52	444
Failed	22	52	74
Total	414	104	518

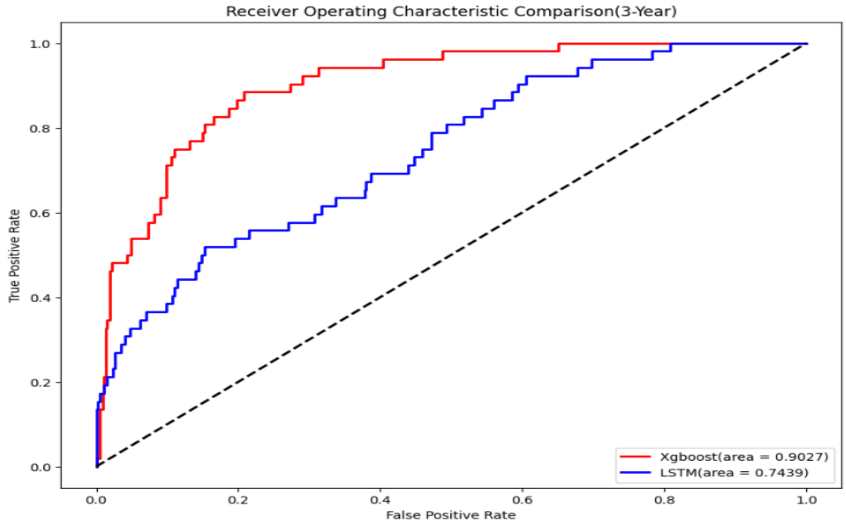
The ROC curve offers a visual representation of the classification model's performance across all thresholds (Noh, 2023). As shown in Figure 6.5, AUC of all three datasets shows higher performance of Xgboost compared to LSTM models, which refers that Xgboost can provide excellent ability in bankruptcy prediction.

Figure 6.5. Comparison of ROC graphs for Model 1 by forecasting horizon

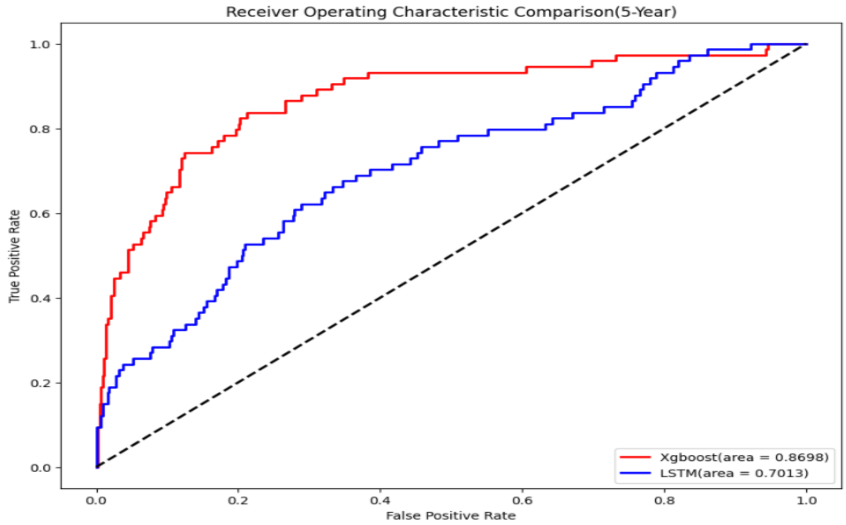
1 year



3 years



5 years



6.3.2. Model 2 – Integrated Financial and Industry-Specific Model

incorporate macroeconomic variables, including the shipping index, across three forecasting periods. The performance of each model is evaluated based on six distinct criteria. As delineated in Table 6.9, models employing XGBoost demonstrate superior predictive capability, as evidenced by higher ROC values, in comparison to LSTM models across all forecasting horizons. Despite LSTM models exhibiting enhanced accuracy and reduced Type I error at the 1-year and 5-year marks, the XGBoost models are deemed more effective over the entire range of forecasting horizons. This assessment takes into account the cost of misclassification errors and the pronounced imbalance between bankrupt and non-bankrupt groups.

Table 6.9. Result of model 2 with three different forecasting horizons

Year	Model	Performance					
		ROC	Accuracy	Sensitivity	Specificity	Type-I Error	Type-II Error
1y	Xgboost	**0.8674**	0.9363	**0.4211**	0.9559	0.0441	**0.5789**
	LSTM	0.7390	**0.9382**	0.1579	**0.9679**	**0.0321**	0.8421
3y	Xgboost	**0.9482**	**0.9382**	**0.6731**	**0.9678**	**0.0322**	**0.3269**
	LSTM	0.8407	0.8764	0.4808	0.9206	0.0794	0.5192
5y	Xgboost	**0.9118**	0.8784	**0.7027**	0.9077	0.0923	**0.2973**
	LSTM	0.7356	**0.9208**	0.3158	**0.9439**	**0.0561**	0.6842

According to Table 6.9, Xgboost models performed better than LSTM models for one year prior to bankruptcy. The ROC curve value of Xgboost and LSTM models are 0.8674 and 0.7390, respectively. This model also produced only 4.4% of type-I errors and 57.9% of type-II errors, while LSTM model produced 3.2% of type-I errors and 84.2% of type-II errors.

Table 6.10 presents classification performance of Xgboost model for one years prior to bankruptcy. This model can predict 477 active firms and 8 bankrupt firms correctly.

Among 518 firms in this model, 95.6% of active firms are correctly classified, and 42% of firms which failed to bankrupt after a year are correctly predicted. Therefore, this model can be considered as a model that has secured prediction performance of 93.6% accuracy.

Table 6.10. Confusion matrix of Model - One year prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	477	22	499
Failed	11	8	19
Total	488	30	518

Table 6.11 displays classification performance of Xgboost model for three years prior to bankruptcy. Among 518 firms in this model, 96.8% of active firms are correctly classified, and 67.3% of firms which failed to bankrupt after three years are correctly classified. Therefore, this model can classify 93.8% of the overall firms correctly. Furthermore, for three years forecasting horizon, Xgboost model showed better performance than LSTM models according to ROC values, which were 0.9482 of Xgboost and 0.8407 of LSTM. Xgboost model produces relatively lower type-I errors of 3.2% and type-II errors of 32.7%, compared to LSTM model, which produced type-I errors of 7.9% and type-II errors of 51.9%.

Table 6.11. Confusion matrix of Model 2 - Three years prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	451	15	466
Failed	17	35	52
Total	468	50	518

In case of models for five years prior to bankruptcy, Xgboost performed better prediction ability compared to LSTM model according to ROC curve, 0.9118 and 0.7356 respectively. Also, Xgboost model produces relatively lower type-I errors of 9.23% and type-II errors of 29.7%, compared to LSTM model, which produced type-I errors of 5.61% and type-II errors of 68.4%. Despite of better accuracy of LSTM model with lower type-I error, Xgboost model still produced higher predictive performance with better ROC score and type-II errors considering highly imbalanced number of firms.

Table 6.12 presents classification performance of Xgboost model for five years prior to bankruptcy. Out of a total of 518 firms, this model accurately predicts 403 active firms and 52 bankrupt firms. According to the dataset, 90.8% of active firms are correctly classified, and 70.3% of firms that did not go bankrupt within five years are correctly identified. Overall, 87.8% of firms are correctly classified by this model.

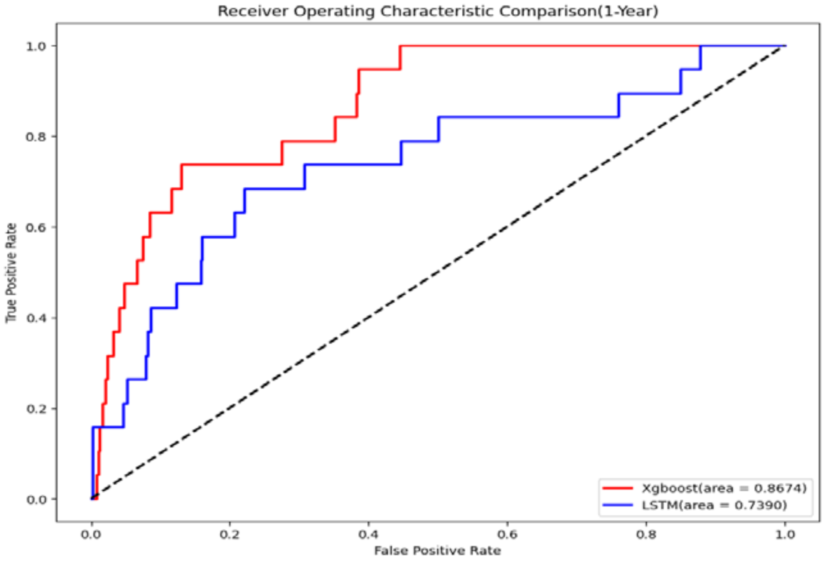
Table 6.12. Confusion matrix of Model 2 - Five years prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	403	41	444
Failed	22	52	74
Total	425	93	518

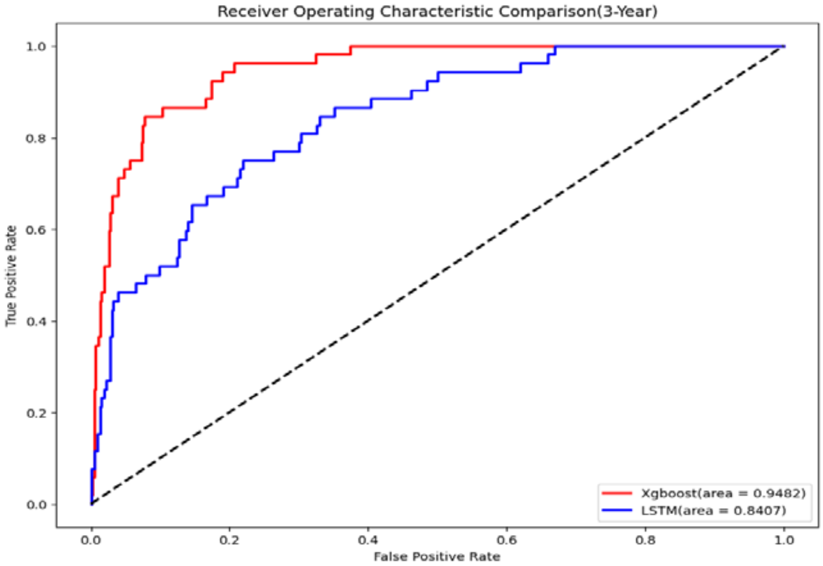
As shown in Figure 6.6, AUC of all three datasets shows higher performance of Xgboost compared to LSTM models, which refers that Xgboost model can attribute to secure significant predictive ability for all forecasting horizons including macroeconomic factors with shipping market index.

Figure 6.6. Comparison of ROC graphs for Model 2 by forecasting period

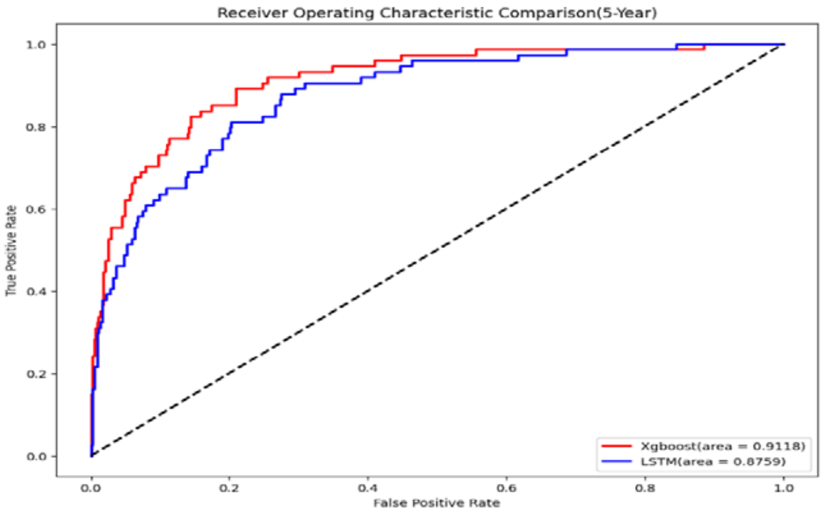
1 year



3 years



5 years



6.3.3. Model 3 – Large Firm Bankruptcy Prediction Model

Table 6.13 presents the results of bankruptcy prediction model for large shipping companies with six different criteria. In this tableXgboost models show higher ROC values with lower type-II error which are indicate better prediction performance than LSTM models for entire forecasting horizons. Although, LSTM models show better accuracy with lower type-I error, which can refer better ability to classify active firms, considering potential risk caused by misclassifying bankrupt firm, Xgboost model can be interpreted with better bankruptcy prediction ability.

Table 6.13 Result of bankruptcy prediction of large firms with three forecasting horizons

Year	**Model**	**Performance**					
		ROC	Accuracy	Sensitivity	Specificity	Type-I Error	Type-II Error
1y	Xgboost	**0.9701**	**0.9710**	**1.0000**	**0.9701**	**0.0299**	**0.0000**
	LSTM	0.5149	0.9420	0.0000	0.9701	0.0299	1.0000
3y	Xgboost	**0.9545**	0.8841	**1.0000**	0.8788	0.1212	**0.0000**
	LSTM	0.5455	**0.9275**	0.0000	**0.9697**	**0.0303**	1.0000
5y	Xgboost	**0.8074**	0.8261	**0.7500**	0.8361	0.1639	**0.2500**
	LSTM	0.7561	**0.8551**	0.1250	**0.9508**	**0.0492**	0.8750

Table 6.14. presents classification performance of Xgboost model for a year prior to bankruptcy of large firms. Among 69 large firms, this model can correctly predict 65 active firms and 2 bankrupt firms. Even though LSTM model shows better type-I error which refer lower error of misclassification active firms as bankrupt, Xgboost model obtains better predictive ability which correctly classify bankrupt firms, which refer 0 type-II error. Therefore, this model can correctly classify 97.1% of firms.

Table 6.14. Confusion matrix of Model 3 – a year prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	65	2	67
Failed	0	2	2
Total	65	4	69

Table 6.15 presents classification performance of Xgboost model for three years prior to bankruptcy. Among 69 total firms, this model can correctly predict 58 active firms and 2 bankrupt firms. From the dataset, 87.8% of active firms are correctly classified, and all of firms which failed to bankrupt after three years are correctly classified. Overall, 88.4% of firms are correctly classified by this model.

Table 6.15. Confusion matrix of Model 3 – three years prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	58	8	66
Failed	0	3	3
Total	58	11	69

Table 6.16 presents classification performance of Xgboost model for five years prior to bankruptcy. Among 69 total firms, this model can correctly predict 51 active firms and 6 bankrupt firms. From the dataset, 83.6% of active firms are correctly classified, and 75% of firms which failed to bankrupt after five years are correctly classified. Overall, 82.6% of firms are correctly classified by this model.

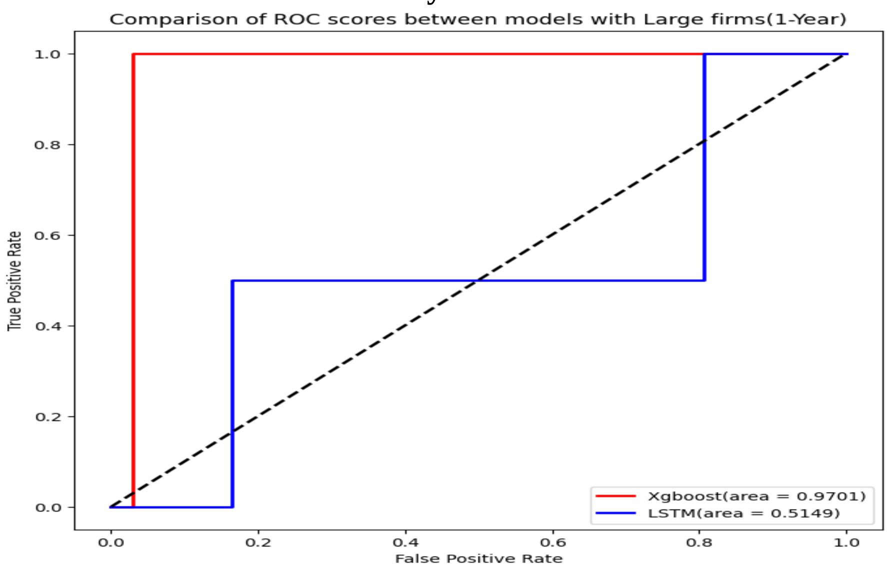
Table 6.16. Confusion matrix of Model 3 – five years prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	51	10	61
Failed	2	6	8
Total	53	16	69

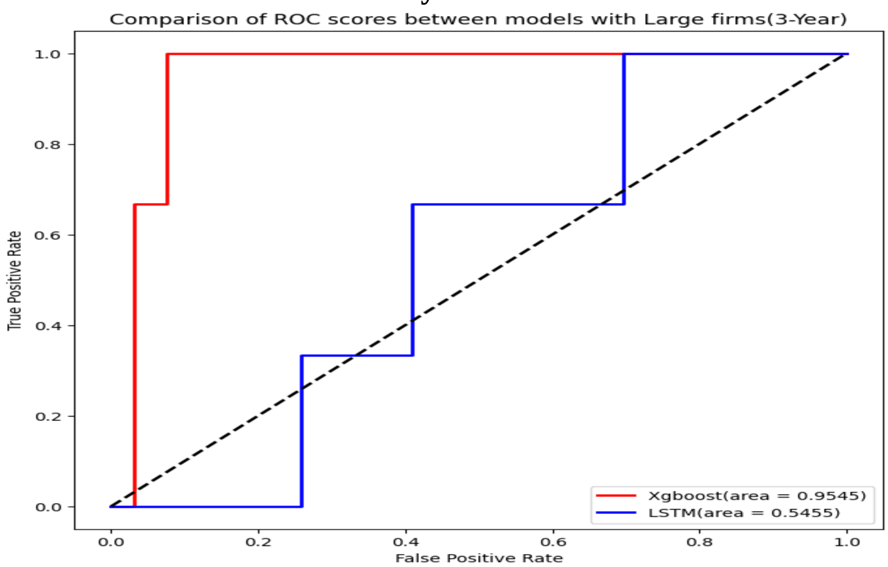
In conclusion, because these models have only applied large companies, which were highly rare events among all firm-year observations, prediction performance of models are relatively lower than results from other model despite of complex machine learning techniques. This is because not only bankruptcy cases of large companies rarely occur compared to other datasets, but also number of large companies is relatively small sized cases. Nevertheless, advanced machine learning model, Xgboost model proved its significant predictive ability by capturing complex relationship among highly imbalanced small sized dataset as shown in Figure 6.7.

Figure 6.7. Comparison of ROC graphs by forecasting horizons of large firms

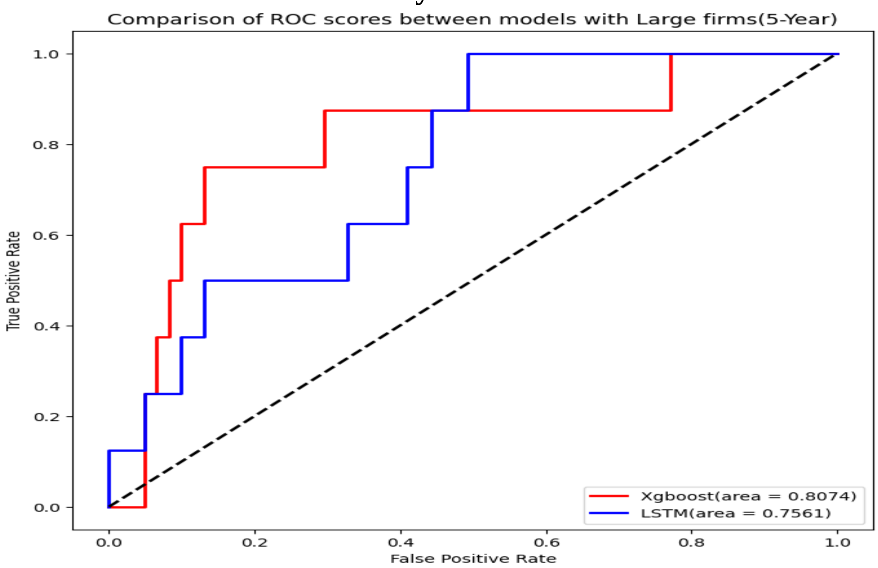
1 year



3 years



5 years



6.3.4. Model 4 – SME-Specific Bankruptcy Prediction Model

This section outlines the outcomes of classification models for small and medium-sized companies across three forecasting periods. Table 6.17 displays the results of bankruptcy prediction model for small-sized shipping companies with six different criteria. According to Table 6.17, Xgboost models present better classification performance with higher ROC values compared to LSTM models for all forecasting horizons. Moreover, Xgboost model produces relatively lower type-I errors of 6.2% and type-II errors of 24.7%, compared to LSTM model, which produced type-I errors of 5.77% and type-II errors of 52.4% on average.

Table 6.17. Result of model 4 with three forecasting horizons

Year	Model	Performance					
		ROC	Accuracy	Sensitivity	Specificity	Type-I Error	Type-II Error
1y	Xgboost	**0.9344**	0.9418	**0.7143**	0.9489	0.0511	**0.2857**
	LSTM	0.7317	**0.9612**	0.2857	**0.9822**	**0.0178**	0.7143
3y	Xgboost	**0.9340**	**0.9181**	**0.7568**	0.9321	0.0679	**0.2432**
	LSTM	0.8632	0.9073	0.4865	**0.9438**	**0.0562**	0.5135
5y	Xgboost	**0.9498**	**0.9138**	**0.7869**	**0.9330**	**0.0670**	**0.2131**
	LSTM	0.8789	0.8685	0.6557	0.9007	0.0993	0.3443

Table 6.18 presents classification performance of Xgboost model for a one year prior to bankruptcy of small sized firms. Among 464 total SMEs, this model can correctly predict 427 active firms and 10 bankrupt firms. From the dataset, 94.9% of the active companies in our dataset have been correctly classified, and 71.4% of the companies that failed to go bankrupt after one year were classified correctly. Therefore, this model can correctly classify 94.2% of firms.

Table 6.18. Confusion matrix of Model 4 - a year prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	427	23	450
Failed	4	10	14
Total	431	33	464

Table 6.19 shows classification performance of Xgboost model for three years prior to bankruptcy. Among 464 total firms, this model can correctly predict 398 active firms and 28 bankrupt firms. From the dataset, 93.2% of active firms are correctly classified, and 75.7% of firms which failed to bankrupt after three years are correctly classified. Overall, 91.8% of firms are correctly classified by this model.

Table 6.19. Confusion matrix of Model 4 - three years prior to bankruptcy

	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	398	29	427
Failed	9	28	37
Total	407	57	464

Table 6.20 presents classification performance of Xgboost model for five years prior to bankruptcy. Among 464 total firms, this model can correctly predict 376 active firms and 48 bankrupt firms. According to the dataset, 93.3% of active firms are accurately classified, and 78.7% of firms that did not go bankrupt within five years are correctly identified. Overall, this model correctly classifies 91.4% of the firms.

Table 6.20. Confusion matrix of Model 3 - five years prior to bankruptcy

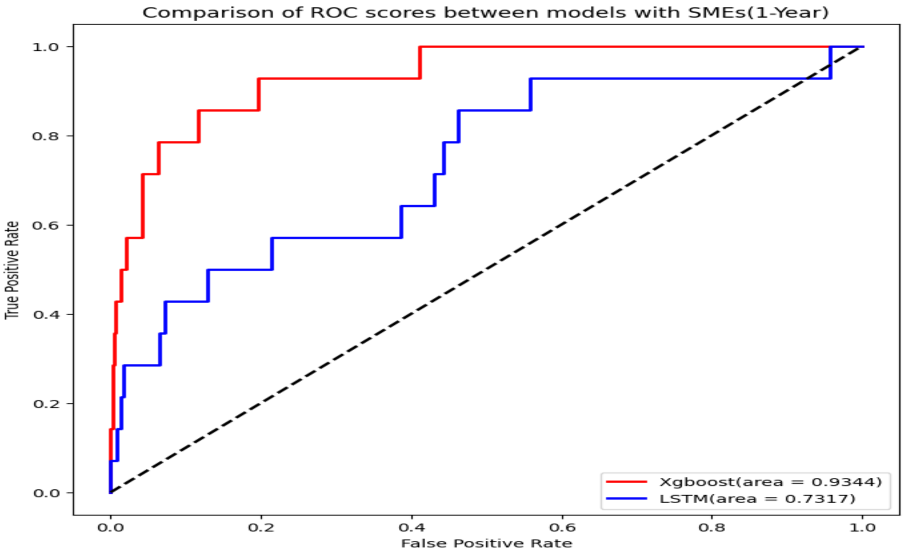
	Predicted		
Observed	**Active**	**Failed**	**Total**
Active	376	27	403
Failed	13	48	61
Total	389	75	464

This can be also seen from the results of ROC curve which suggest visual display of the performance of the classification model at all threshold as shown in

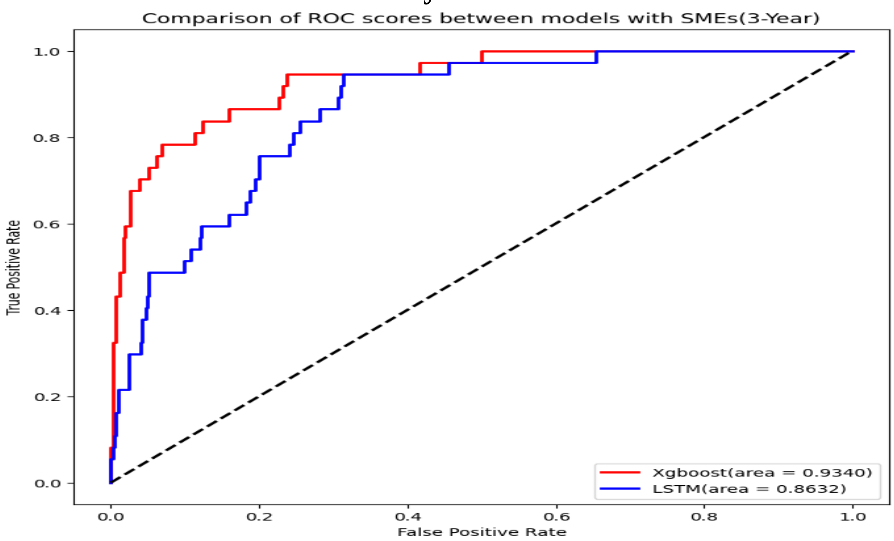
., which indicate better predictive performance for all forecasting horizons for SMEs. In conclusion, Xgboost models show significant predictive performance of bankruptcy of SMEs for all forecasting horizons, which is accounted for average ROC values of 0.9394. Considering unique challenges of SMEs with their highly imbalanced small sized dataset including noisy dataset, this model proved their significant efficiency to predict much complex relationship.

Figure 6.8. Comparison of ROC graphs by forecasting period of SMEs

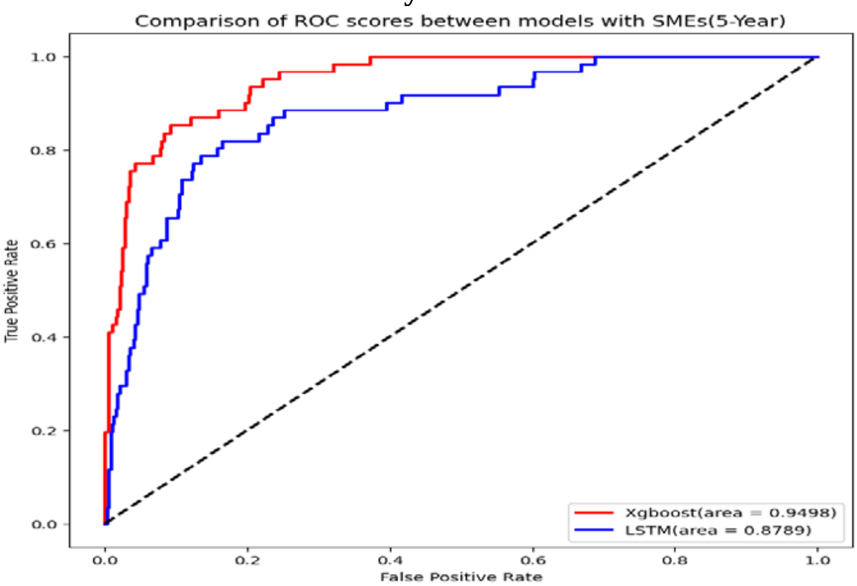
1 year



3 years



5 years



6.4. Chapter Summary

This chapter elucidates the outcomes of bankruptcy prediction analyses conducted using two machine learning models: Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) models, across varying forecasting horizons. The evaluation of these binary classification models was predicated on two key metrics: the Receiver Operating Characteristic (ROC) curve and the Type-II error rate, facilitating a comprehensive comparison of their predictive abilities.

The empirical findings show that XGBoost models consistently achieve higher predictive accuracy than LSTM models across all forecasting periods. This superiority is particularly notable given the infrequent occurrence of firm bankruptcies, underscoring the XGBoost model's adeptness at discerning pivotal relationships within markedly imbalanced, small-sized datasets.

This study assumed different weights on each misclassification error produced by prediction model. Particularly, a Type-II error, which involves incorrectly identifying a bankrupt firm as non-bankrupt, carries substantial consequences for credit institutions. These include potential losses in interest and principal, along with additional expenses related to bankruptcy proceedings. Conversely, a Type-I error, which occurs when a non-bankrupt firm is incorrectly classified as bankrupt, mainly results in missed profit opportunities. Hence, within the banking and financial sectors, the consequences of Type-II errors are more critical than those of Type-I errors. This highlights the need for a model evaluation approach that emphasizes reducing Type-II errors to safeguard against significant financial repercussions.

In conclusion, the models presented in this study provide crucial insights for the proactive identification of bankruptcy risks within the shipping industry, offering stakeholders the ability to make informed decisions well in advance of potential bankruptcy risk. Despite the variability in model performance and forecasting accuracy, the results which characterized by consistently high predictive accuracy suggest that these models proficiently encapsulate the historical data trends and behaviour of firms. This analytical paradigm enhances the identification of vulnerabilities that may critically impact the sustainability and operational viability of firms within the short to medium term in the shipping industry. This enables banks, financial institutions, and investors to execute informed decisions, thereby permitting investors to delineate their investment objectives with greater precision.

7. INTERPRETATION AND DISCUSSION

7.1. Introduction

This chapter investigates the determinants of bankruptcy within the shipping industry, utilizing the analytical frameworks outlined in Chapter 7. The analysis is stratified across three forecasting horizons: one, three, and five years preceding a firm's bankruptcy. This approach facilitates a comprehensive assessment of both short- and long-term risk factors of corporate bankruptcy. Additionally, the influence of various financial and economic indicators on the bankruptcy risk is evaluated across these divergent forecasting horizons. The SHAP (SHapley Additive exPlanations), which is explainable artificial intelligence technique is employed to assess the impact of each determinant on the likelihood of bankruptcy. SHAP analysis offers a visual representation of the effect each variable has on bankruptcy predictions. In the SHAP summary plot, individual points represent the impact (Shapley value) of variables, with the colour coding—red for high impact and blue for low impact—indicating the magnitude of influence. The plot positions these impacts on the y-axis according to the variable, and on the x-axis according to the Shapley value, with a jittering effect applied to points to illustrate the distribution of impacts. Variables that predominantly display red points towards the right of the plot are interpreted as having a positive correlation with the risk of bankruptcy; that is, an increase in these variables' values is illustrated with an elevated risk of bankruptcy.

An initial application of the SHAP method identifies the most significant variables influencing bankruptcy risk. The analysis prioritizes the top 10 variables by their SHAP values, enabling a comprehensive comparison of their relative importance across different models. Such comparative analysis highlights the distinct characteristics of various subsets within the shipping industry, including large firms and SMEs, thereby elucidating the critical factors unique to each category.

Ultimately, this chapter aims to delineate an optimal set of predictors for accurately forecasting bankruptcy at varying intervals before its occurrence, thus enhancing the predictive capabilities of models tailored to the shipping industry's specific risks.

7.2. Interpretation of result

7.2.1. Model I – Financial Metrics Bankruptcy Prediction Model

The SHAP (SHapley Additive exPlanations) method is employed to assess the impact of financial variables within bankruptcy prediction models, quantifying their influence through Shapley values and establishing a hierarchy of the top 10 variables. According to Table 7.1, the variable with the paramount importance for predicting bankruptcy within a one-year forecast horizon is current assets, with a Shapley value of 0.3607. Subsequent rankings highlight net income and working capital as significant predictors. In particular, the classification of the top ten variables includes three ratios each from the categories of leverage, liquidity and profitability, along with a single efficiency ratio. The Shapley value attributed to current assets, occupying the leading position, is markedly higher at approximately 109 times than that of sales, ranked 10th.

In the analysis of a three-year forecasting horizon, current assets persist as the variable with the highest Shapley value, recorded at 0.1579, followed by the growth rate of sales and total assets. Notably, net income experiences a demotion to fifth place, decreasing in its Shapley value by approximately 73%. The composition of the top ten variables remains consistent in terms of the leverage and efficiency categories, despite of the increased profitability ratios and a decreased presence of liquidity ratios. The disparity in Shapley values within this horizon is evident, with current assets being approximately 10.4 times more influential than the debt ratio, which ranks tenth.

Conversely, within a five-year forecasting horizon, the prominence of shareholder’s equity surpasses that of current assets. The leading ten variables are distributed among three leverage, four liquidity, two profitability, and one efficiency ratios. The Shapley value of equity, ranking first, is approximately 4.25 times greater than that of the retained earnings to current liabilities ratio, which ranks tenth.

Throughout the varying forecasting horizons, the Shapley value of Current assets ratio demonstrates a gradual decline from 0.3607 to 0.0535, albeit maintaining a high rank. Conversely, the value attributed to total liabilities exhibits an eightfold increase from 0.0058 to 0.0474, a pattern observed in the values of EBITDA, Sales, and the Growth rates of total assets. As the forecast horizon extends, the variance between the highest and lowest Shapley values narrows, indicating a convergence in the predictive influence of variables.

Table 7.1. SHAP value ranking of explanatory variables in Model I

Rank	1 year			3 years			5 years		
	Variable		SHAP value	Variable		SHAP value	Variable		SHAP value
1	B11	Current Assets	0.3607	B11	Current Assets	0.1579	B9	Shareholder’s Equity	0.0773
2	C11	Net Income	0.1079	C17	∆Sales	0.0306	B11	Current Assets	0.0535
3	A17	Working Capital	0.0110	B20	∆Total Asset	0.0297	A10	Total Liabilities	0.0474
4	A14	EBITDA	0.0074	C6	Sales	0.0294	A14	EBITDA	0.0377
5	B20	∆Total Asset	0.0068	C11	Net Income	0.0290	B20	∆Total Asset	0.0352
6	B6	Current Liabilities to Assets	0.0060	A10	Total Liabilities	0.0289	C6	Sales	0.0341
7	A10	Total Liabilities	0.0058	A14	EBITDA	0.0173	C17	∆Sales	0.0300
8	D6	Inventory turnover	0.0041	C9	Interest expense to total asset	0.0171	D6	Inventory turnover	0.0227
9	C9	Interest expense to total asset	0.0038	D3	Current Assets to Sales	0.0165	A3	Debt ratio	0.0184
10	C6	Sales	0.0033	A3	Debt ratio	0.0151	B19	Retained Earnings to Current Liabilities	0.0182

Figure 7.1 presents summary plots of SHAP values for variables critically influencing bankruptcy predictions within the shipping industry. These plots rank the top 10 variables across all datasets by their Shapley values, with the variables' feature values denoted by a colour gradient from blue (low) to red (high).

In the 1-year prediction model, all variables generally exhibit a positive correlation with the prediction of bankruptcy, implying that higher values of these variables tend to decrease the bankruptcy risk for shipping companies. However, the current liabilities-to-asset ratio presents an inverse relationship, which an increase in this ratio suggests a diminished capacity for debt repayment, potentially exacerbating bankruptcy risk. Similarly, the ratio of interest expenses to total assets is inversely related to bankruptcy risk, indicating a potential exacerbation of financial distress.

This pattern of relationships persists in the 3-year prediction model, where variables with high Shapley values typically correlate positively with bankruptcy prediction, signalling a reduction in bankruptcy risk. Conversely, as noted in the one-year model, the ratio of interest expenses to total assets has a negative impact on bankruptcy risk. Furthermore, it is observed that an increase in the current asset ratio and debt ratio leads to a higher probability of default, indicating a negative correlation.

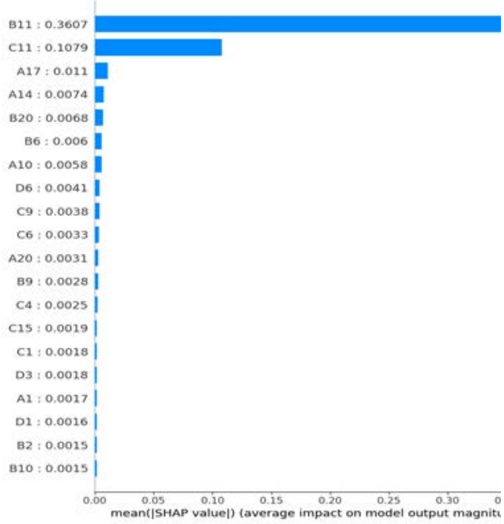
In the five-year forecast model, the debt ratio and current liabilities-to-retained earnings ratio, which indicate a company's debt level, are negatively correlated with bankruptcy risk, with higher ratios indicating a higher risk of insolvency. Conversely, other variables maintain a positive relationship with bankruptcy risk.

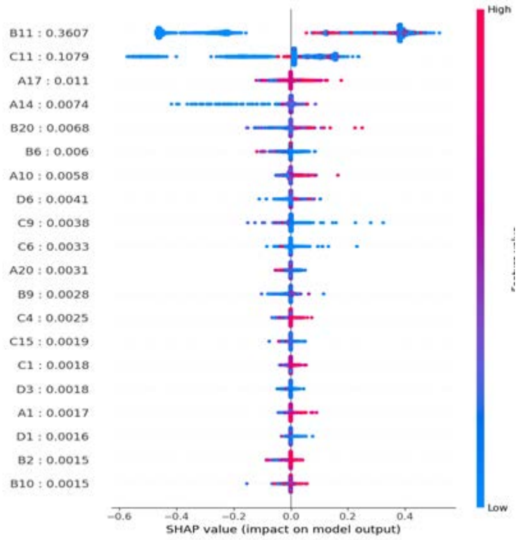
These results shed light on the intricate relationship between different financial ratios and bankruptcy risk in the shipping industry, underscoring how certain variables can act as markers of financial health or bankruptcy risk over various forecasting periods.

Figure 7.1. Summary plots of SHAP value in Model I

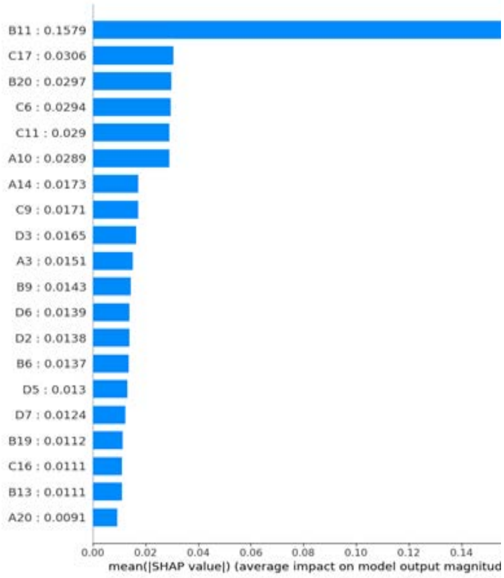
Year | Bar Plot | Summary Plot

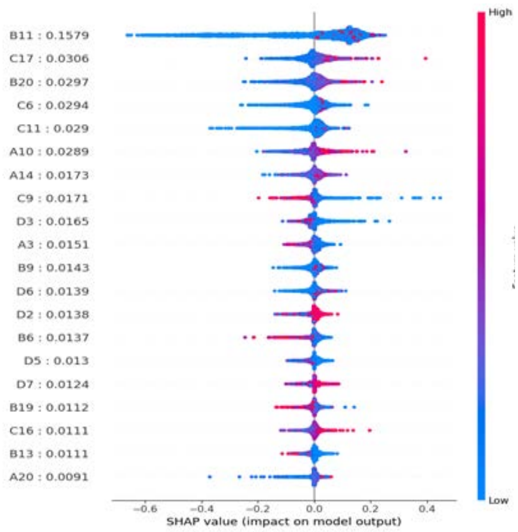
1 Year



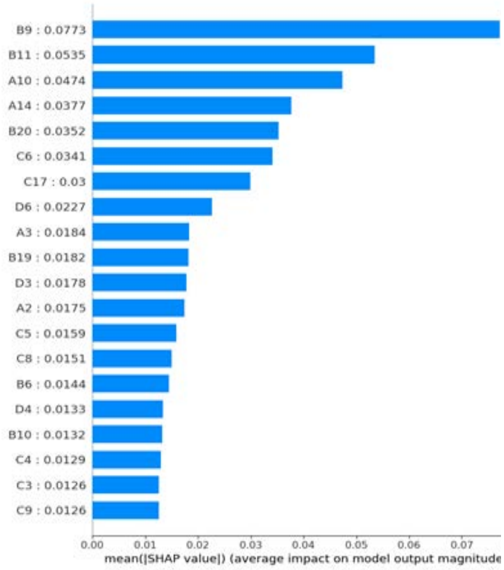


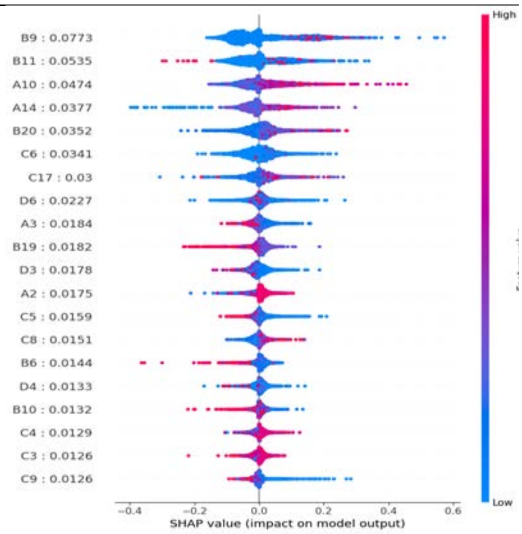
3 Year





5 Year





7.2.2. Model II – Integrated Financial and Industry-Specific Model

To assess the contribution of macroeconomic factors, including shipping indices, to bankruptcy predictions within the shipping industry, these elements were incorporated into Model 1. The SHAP (SHapley Additive exPlanations) methodology was then utilized to assess their influence through Shapley values.

Table 7.2 demonstrated the paramount variables influencing bankruptcy prediction over a one-year forecasting period. Notably, the company size emerged as the most critical variable, marked by the highest Shapley value of 0.0906. This was followed by the growth rate of global seaborne trade volume and the interest expense-to-assets ratio. Interestingly, the top ten variables consisted of only two financial ratios, alongside four shipping indices and four non-financial variables. The Shapley value by company size, which ranked first, was about 4.5 times much higher than the LIBOR interest rate, which ranked 10th.

Similar to the findings from preceding models, the size of firms retained its dominance in the three-year forecasting model with the highest Shapley value of 0.0705, succeeded by the growth rates of global and Korean seaborne trade volume. The interest expense-to-assets ratio, however, did not make it to the top ten list. This model's top ten encompassed three financial ratios with all falling within the profitability category, while the presence of non-financial variables was reduced by half. The quantity of shipping indices remained constant, though their average Shapley value experienced a marginal reduction of 10%. The disparity between the top ten rankings was pronounced, with the Shapley value of company size being approximately 2.87 times that of the gross profit-to-sales ratio, which secured the tenth rank.

Contrastingly, in the five-year forecasting model, the influence of the LIBOR interest rate surpassed that of company size. This ranking saw an even distribution of financial ratios, including four pertaining to profitability and one to leverage. Conversely, the count of shipping index variables witnessed a decline, while their average Shapley value saw a notable increment of 40%. The LIBOR interest rate's Shapley value, leading the rank, was significantly higher about 4.16 times than that of the growth rate of global seaborne trade volume, ranked tenth.

Throughout the forecasting intervals, the size of the firm consistently demonstrated significant explanatory power, albeit with a slight decrement in its Shapley value over time. Moreover, shipping indices revealed a robust capability to elucidate bankruptcy within the shipping sector, particularly highlighted by the ascending influence of the LIBOR interest rate, signifying its enduring predictive utility for bankruptcy within the industry.

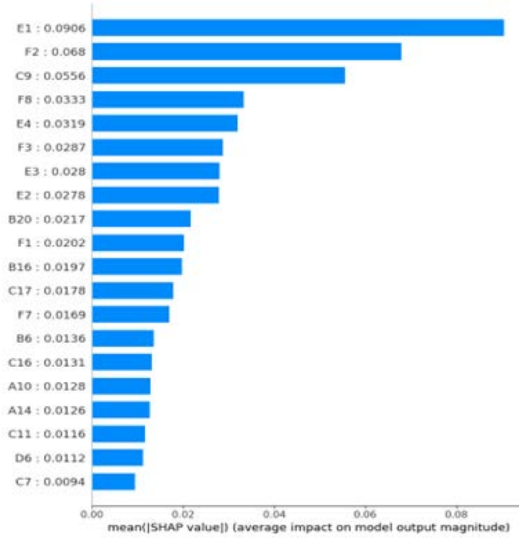
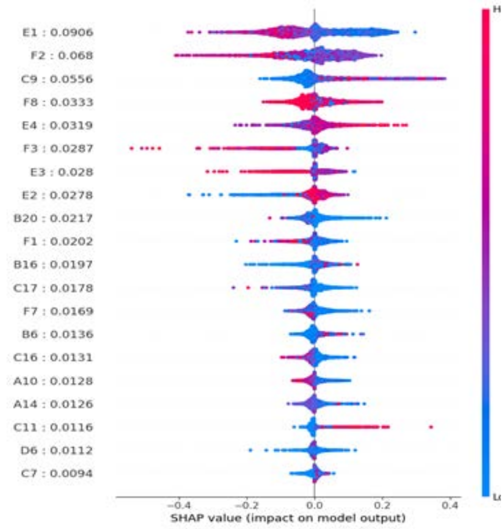
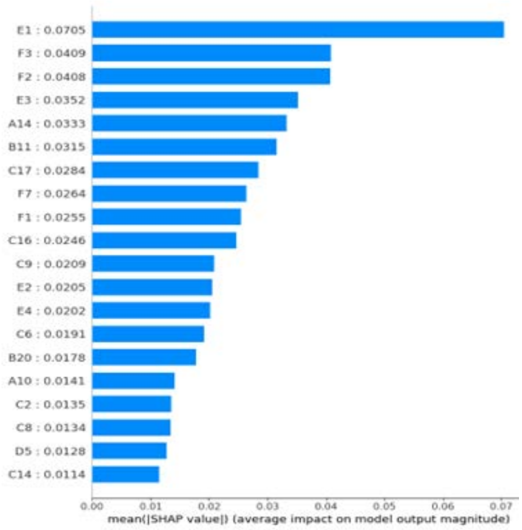
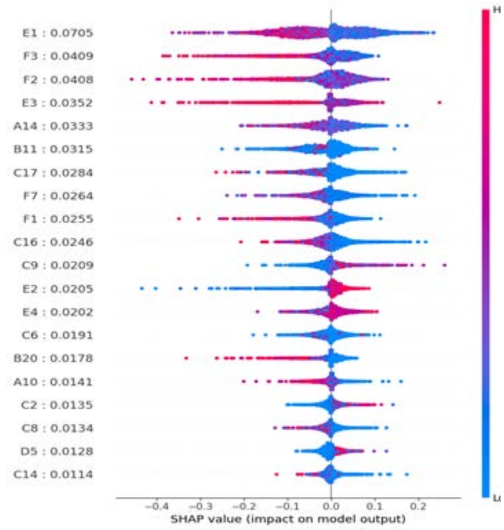
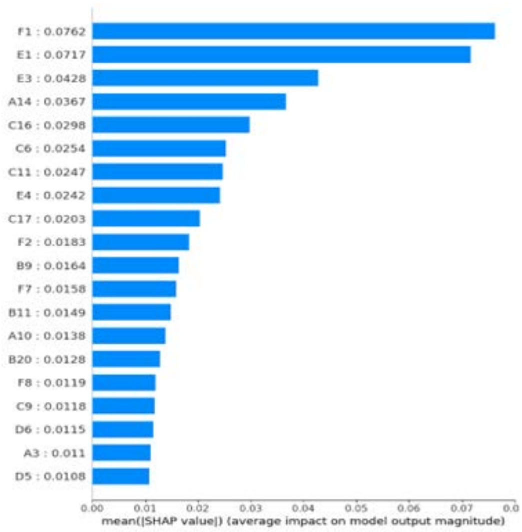
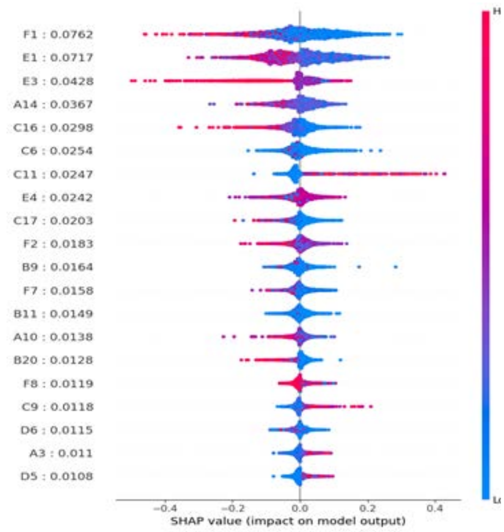
Table 7.2. SHAP value Ranking of explanatory variables in Model II

Rank	1 year			3 years			5 years		
	Variable		SHAP value	Variable		SHAP value	Variable		SHAP value
1	E1	Size	0.0906	E1	Size	0.0705	F1	LIBOR	0.0762
2	F2	Growth rate of global seaborne trade Volume	0.0680	F3	Growth rate of Korean seaborne trade Volume	0.0409	E1	Size	0.0717
3	C9	Interest expense to assets	0.0556	F2	Growth rate of global seaborne trade Volume	0.0408	E3	Type of Operator	0.0428
4	F8	G/T	0.0333	E3	Type of Operator	0.0352	A14	EBITDA	0.0367
5	E4	GDP Growth rate	0.0319	A14	EBITDA	0.0333	C16	Gross Profit to Sales	0.0298
6	F3	Growth rate of Korean seaborne trade Volume	0.0287	B11	Current Assets	0.0315	C6	Sales	0.0254
7	E3	Type of Operator	0.0280	C17	ΔSales	0.0284	C11	Net Income	0.0247
8	E2	Age	0.0278	F7	No. of vessel	0.0264	E4	GDP Growth rate	0.0242
9	B20	ΔTotal Asset	0.0217	F1	LIBOR	0.0255	C17	ΔSales	0.0203
10	F1	LIBOR	0.0202	C16	Gross Profit to Sales	0.0246	F2	Growth rate of global seaborne trade Volume	0.0183

Figure 7.2 shows that most financial variables among the top ten rankings inversely correlate with bankruptcy risk, implying that an increase in these variables corresponds with a diminished risk of bankruptcy for the shipping company. In contrast, the ratio of interest expenses to assets exhibits a direct correlation with bankruptcy risk, reflecting decreased profitability due to heightened interest expenses. The LIBOR interest rate, a pivotal shipping index, demonstrated substantial negative correlation with bankruptcy risk, underscoring the critical role of the Eurodollar base rate in financing shipping operations. Additionally, the global GDP growth rates variable indicated significant volatility across forecasting horizons, suggesting its profound impact on the shipping industry's susceptibility to bankruptcy. In addition, it was analysed that as the classification variables according to the type of shipping company decreased, the risk of bankruptcy of shipping companies increased, indicating a higher bankruptcy risk among container shipping companies compared to bulk and tanker firms.

The contribution of the top 10 variables, as quantified by the sum of their SHAP values, accounted for 40.58% in the one-year prediction model, 35.71% in the three-year model, and 37.1% in the five-year model. These percentages reflect the extent to which the model's explanatory power for bankruptcy occurrences in the shipping industry is attributable to these top variables. The data indicates that, on average, the top 10 variables explain approximately 37.7% of the variations in bankruptcy predictions, with their relative importance increasing as the forecast period to potential bankruptcy shortens.

Figure 7.2. Summary plots of SHAP value in Model II

<table>
<tr><th>Year</th><th>Bar Plot</th><th>Summary Plot</th></tr>
<tr><td>1 Year</td><td>
</td><td>
</td></tr>
<tr><td>3 Year</td><td>
</td><td>
</td></tr>
<tr><td>5 Year</td><td>
</td><td>
</td></tr>
</table>

7.2.3. Model III – Large Firm Bankruptcy Prediction Model

In this analysis, the focus shifts towards identifying the critical factors that predict bankruptcy among large shipping companies, differentiating them from small and medium-sized enterprises (SMEs). Table 7.3 presents the top 10 variables by Shapley value across three distinct forecasting horizons.

For the 1-year forecasting period, the Sales-to-liabilities ratio, which reflects profitability, emerges as the most significant predictor with the highest Shapley value of 0.1313. Subsequent variables, also linked to profitability, include the Retained earnings-to-total assets and the Interest expense-to-assets ratio. The composition of the top ten is predominantly financial ratios (nine in total) with a single non-financial variable. Notably, profitability ratios dominate among those with the highest Shapley values, emphasizing their substantial predictive capability for the bankruptcy of large shipping firms. The Shapley value of the gross profit to current liabilities ratio, which ranks first, is approximately 17.7 times that of the sales to current liabilities ratio, positioned tenth.

In the 3-year forecasting model, the gross profit-to-current liabilities ratio attains the highest Shapley value at 0.1522, followed by net income and the growth rate of total assets. On the other hand, Shapley's value to debt-to-sales ratio decreased by 88%, ranking 9th. Profitability ratios again constitute the majority within the top ten, accounting for 50% of the variables, with nine financial ratios and one shipping index present. The Gross profit-to-current liabilities ratio's Shapley value stands approximately 10.6 times above that of the return on sales ratio, which ranks 10th.

The trend continues in the five-year forecasting model, where Net income, representing profitability ratios, exhibits the highest Shapley value, followed by the gross profit-to-total liabilities ratio and the gross profit-to-asset ratio. This model's top ten is exclusively financial ratios, comprising seven profitability, two leverage, and one liquidity ratio. The Shapley value of net income, leading the rank, is about 6.26 times that of the Quick assets-to-total assets ratio, ranked tenth.

Throughout the forecasting periods, profitability ratios demonstrate significant predictive power, particularly those related to Gross profit and Net income. Among the top 10 variables, profitability ratios constitute approximately 53% with an average Shapley value of 0.0516. Compared to the broader analysis of the shipping industry, the Net income ratio maintains its substantial predictive value with an average Shapley value of 0.0729 across the periods. Meanwhile, shipping indices and non-financial variables exhibit limited explanatory power in predicting bankruptcy for large shipping firms. As bankrupt firm approaches to its bankruptcy year, importance of efficiency ratios increased such as asset turnover or inventory to assets, which reflects firms' ability to generate revenue from their assets. Furthermore, the disparity between the highest and lowest Shapley values diminishes with extended forecasting horizons.

Table 7.3. SHAP value Ranking of explanatory variables in Model III

Rank	1 year			3 years			5 years		
	Variable		SHAP value	Variable		SHAP value	Variable		SHAP value
1	C13	Sales to Liabilities	0.1313	C14	Gross Profit to Current Liabilities	0.1522	C11	Net Income	0.1014
2	A2	Retained Earnings to Total Assets	0.1012	C11	Net Income	0.0758	C15	Gross Profit to Liabilities	0.0627
3	C9	Interest expense to assets	0.0785	B20	ΔTotal Asset	0.0289	C8	Gross Profit to Assets	0.0495
4	B20	ΔTotal Asset	0.0555	F7	No. of vessel	0.0289	C12	Operating Return on Assets	0.045
5	C11	Net Income	0.0417	C8	Gross Profit to Assets	0.028	A18	Current Assets to Liabilities	0.0396
6	D3	Current Assets to Sales	0.0246	B3	Cash Assets Ratio	0.0255	A16	Net Income to Liabilities	0.026
7	B19	Retained Earnings to Current Liabilities	0.0237	A18	Current Assets to Liabilities	0.0216	C17	ΔSales	0.0212
8	E3	Type of Operator	0.0128	D4	Inventory to Assets	0.0185	C16	Gross Profit to Sales	0.0186
9	C5	Sales to Current Assets	0.0108	C13	Sales to Liabilities	0.0147	C3	Return on Sales	0.017
10	D1	Asset turnover	0.0074	C3	Return on Sales	0.0143	B15	Quick Assets to Total Assets	0.0162

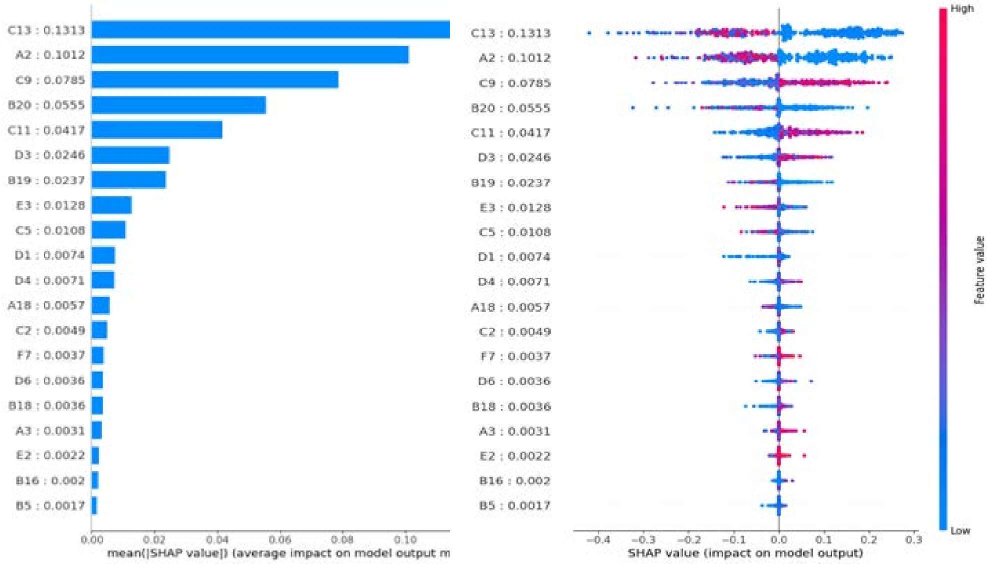
The predictive contribution of the models, highlighted by the Shapley values of the top 10 variables, illuminates their ability to enhance bankruptcy predictions in the shipping industry in Figure 7.3. For the one-year forecast model, the Shapley values of the top 10 variables elucidate 48.75% of the variance in bankruptcy predictions. This explanatory power slightly decreases to 40.84% in the three-year forecast model and to 39.7% in the five-year model. These statistics validate the models' robust ability to account for approximately 43.1% of the variability in bankruptcy predictions solely through these pivotal variables.

This analysis reveals a slight improvement in the models' predictive contribution as the time horizon shortens from three to one year, yet the contribution levels off between the three and five-year forecasts. Such consistency emphasizes the models' proficiency in mid-term bankruptcy forecasting, particularly considering that accurately discerning a company's financial distress through bankruptcy processes might require up to three years (Perboli and Arabnezhad, 2021). This insight highlights the importance of strategic financial management in large shipping companies and suggests a refined understanding of financial health and risk across various forecasting periods.

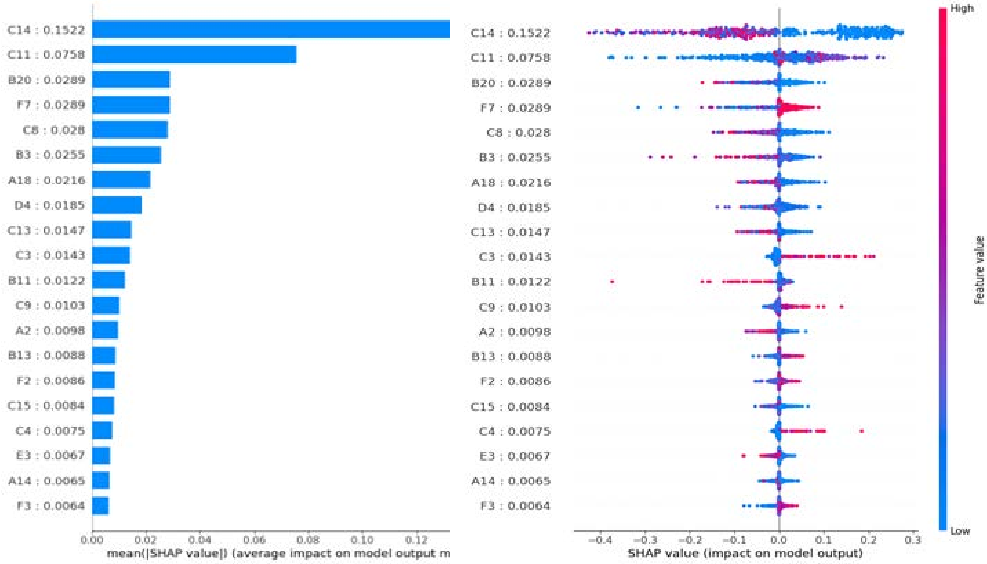
Figure 7.3. Summary plots of SHAP value in Model III

	Bar Plot	Summary Plot

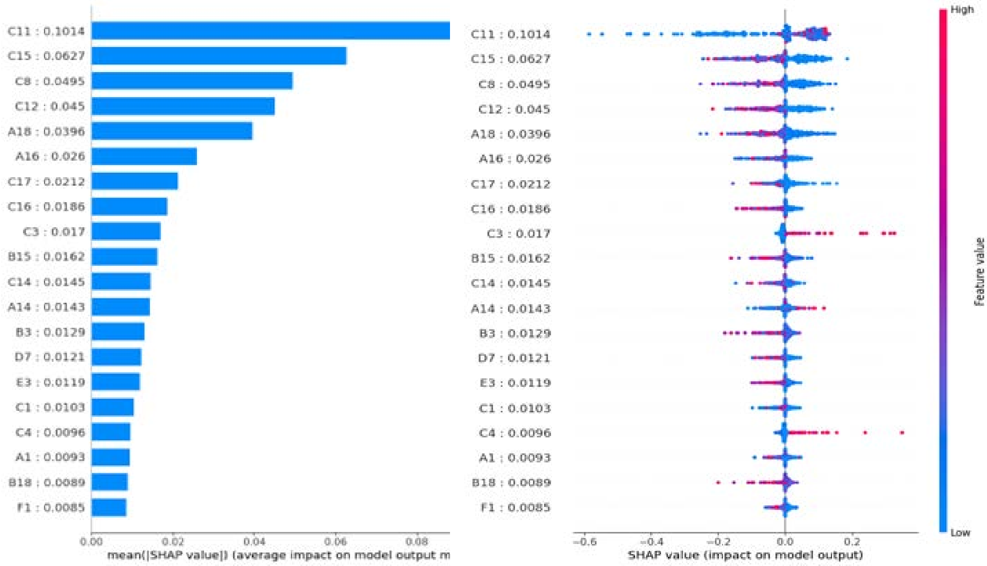
1 Year



3 Year



5 Year



7.2.4. Model IV – SME-Specific Bankruptcy Prediction Model

In this analysis, three models across different forecasting horizons assess key factors influencing bankruptcy predictions for SMEs, with a focus on Shapley values as presented in Table 7.4.

For the one-year forecast period, the size of firms emerges as the most critical variable, exhibiting the highest Shapley value of 0.0909, followed by the number of vessels owned and current assets. In the top 10 variables, shipping indices related to the shipping industry predominate with four variables, complemented by three each in non-financial and financial ratios. The Shapley value for firm size, positioned at the top, is about five times greater than the growth rate of sales, which is ranked tenth.

Similar to the findings from the previous analysis, the model for the three-year forecast period also highlights firm size as having the highest Shapley value at 0.0789, succeeded by sales and current assets. However, the Shapley value for the number of vessels owned experienced a 62% decrease. Among the top 10 variables, four are shipping indices, making up the largest category, followed by four financial ratios and two non-financial ratios. The Shapley value of firm size is approximately 3.74 times that of the growth rate of sales, ranked tenth.

In the five-year forecast model, firm size continues to display the highest Shapley value, followed by LIBOR interest rates and shareholder's equity. The top 10 variables include three each from shipping indices and non-financial factors, alongside four financial ratios. The Shapley value of firm size, leading the list, is roughly 3.80 times that of EBITDA, positioned tenth.

Throughout the forecast periods, the size of firms, represented by the logarithms of total assets, consistently shows the highest Shapley value, aligning with findings from the broader shipping industry analysis. Shipping indices, such as the growth rates of global seaborne trade volume, occupy a significant portion of the explanatory variables and exhibit high Shapley values. This suggests that SMEs in the shipping sector are particularly susceptible to fluctuations in global seaborne trade or LIBOR interest rates. Specifically, Korea's seaborne trade volume growth rate presents notable Shapley values in the one- and three-year forecasts but is absent in the five-year model. Contrary to prior results, current assets ratio demonstrates their significant explanatory power for firms' ability to meet short-term liabilities, though their average Shapley value is relatively modest at 0.0431. Lastly, the disparity between the highest and lowest Shapley values expands as bankrupt firm approaches to its bankruptcy year.

Table 7.4. SHAP value Ranking of explanatory variables in Model IV

Rank	1 year			3 years			5 years		
	Variable		SHAP value	Variable		SHAP value	Variable		SHAP value
1	E1	Size	0.0909	E1	Size	0.0789	E1	Size	0.0643
2	F7	No. of vessel	0.0702	C6	Sales	0.0407	F1	LIBOR	0.0612
3	B11	Current Assets	0.0519	B11	Current Assets	0.0406	B9	Shareholder’s Equity	0.0467
4	F2	Growth rate of global seaborne trade Volume	0.0451	F2	Growth rate of global seaborne trade Volume	0.0350	B11	Current Assets	0.037
5	F8	G/T of vessel	0.0257	F3	Growth rate of Korean seaborne trade Volume	0.0333	C6	Sales	0.0348
6	E4	GDP Growth rate	0.0257	F7	No. of vessel	0.0308	E3	Type of Operator	0.0328
7	F3	Growth rate of Korean seaborne trade Volume	0.0236	E3	Type of Operator	0.0271	F7	No. of vessel	0.0323
8	B16	Working Capital to Equity	0.0193	F1	LIBOR	0.0234	E4	GDP Growth rate	0.0238
9	E2	Age	0.0183	B9	Shareholder’s Equity	0.0218	F2	Growth rate of global seaborne trade Volume	0.0202
10	C17	ΔSales	0.0183	C17	ΔSales	0.0211	A14	EBITDA	0.0169

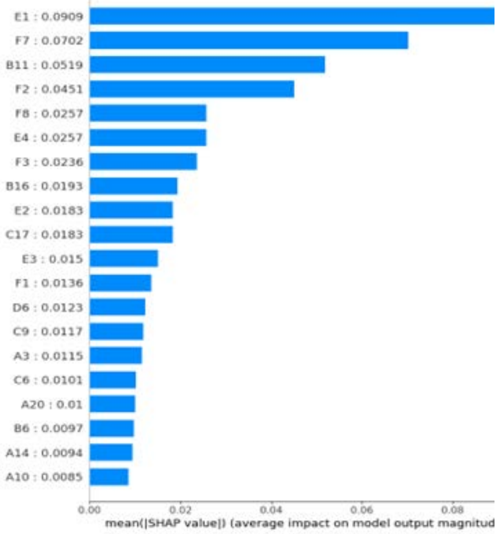
Figure 7.4 displays the connection between the bankruptcy risk of shipping SMEs and explanatory variables with high Shapley values. It is observed that all variables exhibit a positive correlation with bankruptcy risk, suggesting that the likelihood of bankruptcy for small and medium-sized shipping enterprises diminishes as these variables exhibit improvement. Especially, the shipping index, which includes the growth rate of global seaborne trade volume, displays considerable volatility throughout the forecasting periods, despite its negative correlation with bankruptcy risk. This suggests that shipping SMEs are significantly impacted by macroeconomic factors, including fluctuations in the shipping market and demand for sea transport, more so than their larger counterparts.

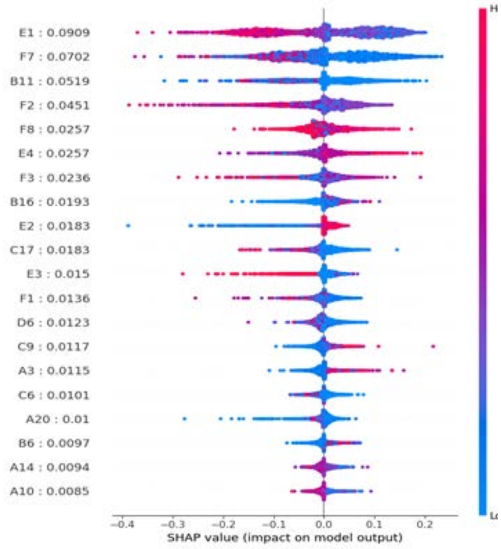
The model's predictive capability, as highlighted by the Shapley values of the top 10 variables, indicates that these variables contribute to 38.9% of the prediction capability in the one-year prediction model. In the three-year prediction model, the contribution is 35.27%, and for the five-year model, it increases slightly to 37.1%. These percentages underscore the model's substantial ability to account for approximately 37.1% of the bankruptcy events among shipping SMEs through the top 10 variables alone. Moreover, the stability of this predictive contribution across the various forecasting horizons demonstrates the model's consistent effectiveness in both short- and long-term bankruptcy prediction for shipping SMEs.

Figure 7.4. Summary plots of SHAP value in Model III

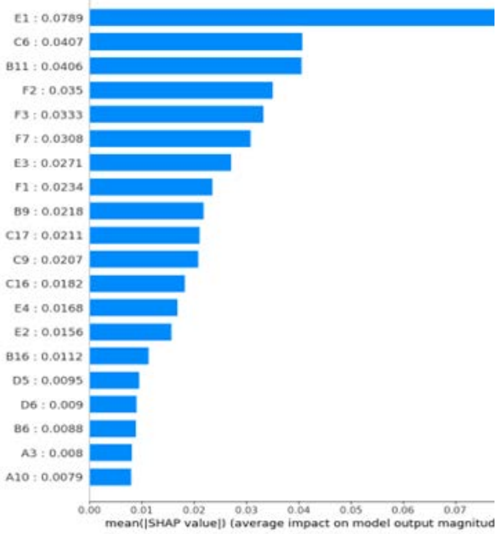
Year	Bar Plot	Summary Plot

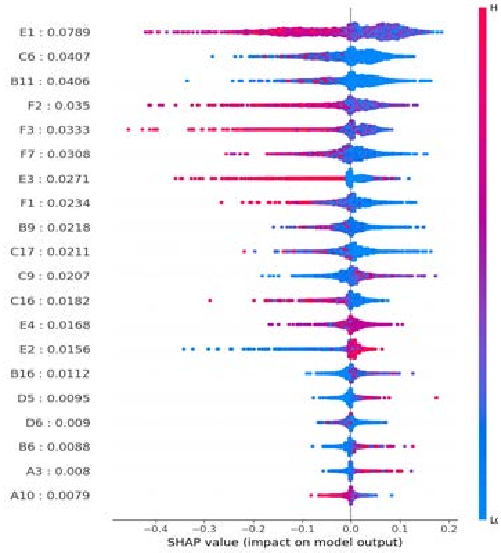
1 Year



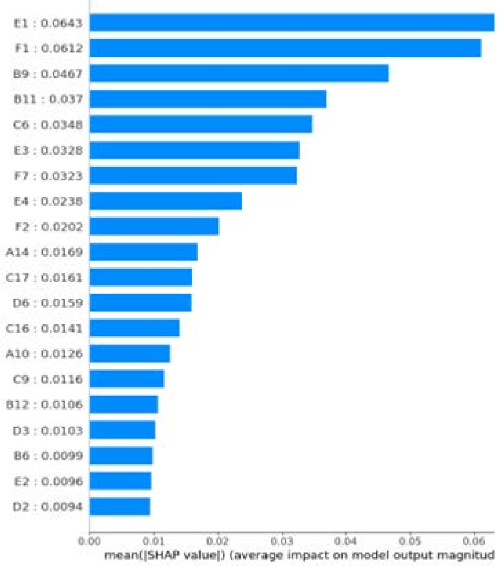


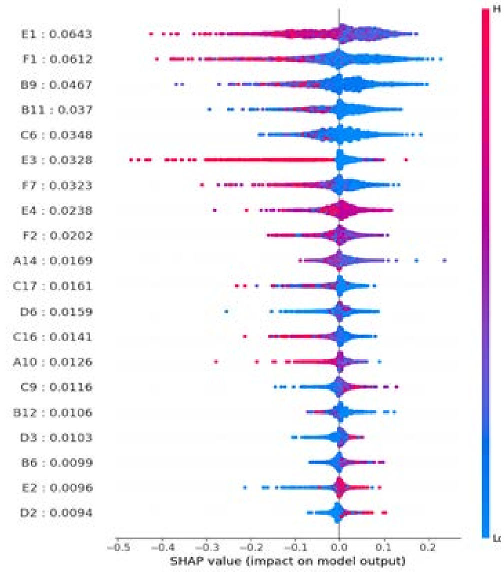
3 Year





5 Year





7.3. Discussion

7.3.1. Comparison between financial ratios and macroeconomic variables

As evidenced by Table 7.5, this study highlights a significant distinction between models that exclusively incorporate financial ratios and those that integrate macroeconomic variables, such as shipping indices. The column "Count" refers to the number of times each variable was selected as important across different runs or folds, while the "SHAP" column represents the sum of SHAP scores for each selected variable, indicating its overall impact on the model's predictions.

In the financial ratio only model, liquidity ratios were observed to have a substantial impact on bankruptcy risk. However, when macroeconomic variables and shipping indices are combined, their importance decreases, and the explanatory power of the latter increases compared to financial ratios. Especially, as firm approaches to its bankruptcy, the influence of non-financial variables and shipping indices, particularly the growth rates of global and Korean seaborne trade volumes and the LIBOR interest rate, becomes more pronounced. Firm-specific variables, including the type of operator, the number of vessels owned, and the size of the company, also exhibit significant explanatory power across the forecasting periods. Despite the relatively modest role of financial ratios, the relevance of profitability ratios grows with the extension of the forecasting horizon. The Shapley values of profitability metrics, such as gross profit-to-sales, sales growth rate, and net income ratios, show enhancement in the 3-to-5-year forecasting period, highlighting their mid-term predictive capacity.

The variables crucial in predicting bankruptcy within the shipping industry maintain consistent significance across the forecast periods. However, their impact varies with the length of the forecasting horizon. The relative importance of shipping indices, like the LIBOR interest rate, demonstrates fluctuations due to the dual-edged nature of falling and rising interest rates affecting borrowers' financial obligations. Conversely, the influence of global seaborn trade volume growth rate, indicative of the shipping industry's demand and subsequent profit impacts, intensifies as firms near bankruptcy.

In conclusion, the findings emphasize the critical need to monitor the fluctuations of shipping indices to manage bankruptcy risks effectively within the shipping industry over different forecasting horizons. Furthermore, the results from Table 6.2 and Table 6.7 suggest that enhancing the predictive performance of the shipping industry bankruptcy prediction model necessitates the inclusion of macroeconomic variables and non-financial factors that encapsulate the unique attributes of the industry.

Table 7.5. Comparison of financial and macroeconomic models by forecast periods

	1 year				**3 years**				**5 years**			
	Model I		**Model II**		**Model I**		**Model II**		**Model I**		**Model II**	
	Count	SHAP	Count	SHAP	Count	SHAP	Count	SHAP	Count	SHAP	Count	SHAP
Leverage	3	0.0242	0	0	3	0.0613	1	0.0333	3	0.0998	1	0.0367
Liquidity	3	0.3735	1	0.0217	2	0.1876	1	0.0315	4	0.1941	0	0
Profitability	3	0.1150	1	0.0556	4	0.1061	2	0.053	2	0.0586	4	0.1002
Efficiency	1	0.0041	0	0	1	0.0165	0	0	1	0.0265	0	0
Non-financial	0	0	4	0.1783	0	0	2	0.1057	0	0	3	0.1387
Shipping	0	0	4	0.1502	0	0	4	0.1336	0	0	2	0.0945

7.3.2. Comparison between large shipping firms vs SMEs

This study aims to pinpoint distinct explanatory variables critical for forecasting bankruptcy among small-sized shipping companies. Table 7.6 contrasts the distribution of variables with high Shapley values between large firms and small to medium-sized enterprises (SMEs). The column "Count" refers to the number of times each variable was selected as important across different runs or folds. The "Total SHAP" column represents the sum of SHAP scores for each selected variable, indicating its overall impact on the model's predictions. Generally, as approached to bankrupt year, those relationship has been stronger with SMEs, but those of large firms shows opposite. This means that the influence of shipping index and non-financial ratios such as gross domestic product (GDP) and global trade volume, as measured by Shapley value, have increased as the forecast period is shortened. In addition, among financial ratios, current assets had continuously great explanatory power for bankruptcy prediction of small sized shipping companies, which measures the capacity of firms to repay their short-term liabilities. This indicated that a liquidity depression could be a serious problem to make them difficult to meet their short-term liabilities, and may eventually lead to bankruptcy crisis.

Conversely, in the case of large companies, financial ratios related with gross profit and net income of firms had great effect on bankruptcy prediction of large shipping firms. In other words, this can indicate that large shipping company needs to manage their bankruptcy risk by improving their efficiency in generating profit and reducing the cost. However, the influence of the profitability ratio decreased and appeared to spread to other variables as bankrupt firm approaches to its bankruptcy. Leverage ratios such as retained earnings-to-total assets ratio and current assets-to-liabilities ratio had the great impact on bankruptcy risk of large firms depending on forecasting horizon. The lower leverage ratios increase the bankruptcy risk of companies, because lower values of these variables imply that the company much relies on financing capital expenditures through borrowing rather than retained earnings (Altman, 2018).

Table 7.6. Most influential variables in Large firm vs SMEs by forecasting horizon

	1 year				3 years				5 years			
	Large		SMEs		Large		SMEs		Large		SMEs	
	Count	Total SHAP	Count	Total SHAP	Count	Total SHAP	Count	Total SHAP	Count	Total SHAP	Count	Total SHAP
Leverage	1	0.1012	-	-	1	0.0216	-	-	2	0.0656	1	0.0169
Liquidity	2	0.0792	2	0.0712	2	0.0272	2	0.0624	1	0.0162	2	0.0837
Profitability	4	0.2623	1	0.0183	5	0.2850	2	0.0618	7	0.3154	1	0.0348
Efficiency	2	0.0320	-	-	1	0.0185	-	-	-	-	-	-
Non-financial	1	0.0128	3	0.1349	-	-	2	0.1060	-	-	3	0.1209
Shipping	-	-	4	0.1646	1	0.0289	4	0.1225	-	-	3	0.1137

Table 7.7 provides a detailed comparison of variables with the highest SHAP values, identifying key predictors of bankruptcy across different forecasting periods for both large firms and SMEs within the shipping industry. Each variable is selected based on the sum of SHAP values, as suggested in the previous section, and ranked in order of SHAP values across different forecasting horizons.

This comparison emphasizes distinct variances in the predictive variables for large firms versus SMEs, though the differences across forecasting periods remain relatively stable. Aligning with the overarching analysis of the shipping industry, it is evident that macroeconomic factors, such as GDP growth rates and global trade volume, exert a substantial influence on the bankruptcy risk of small-sized shipping companies. Additionally, firm-specific attributes, including the number of vessels owned and the overall size of the company, significantly affect SMEs, potentially indicating liquidity management challenges during periods of market volatility. These indicators could imply a propensity among SMEs to manage their liabilities through the sale of high-value assets, like vessels, or by reducing their scale to ease financial burdens (Michail, 2020).

Conversely, large shipping companies exhibit a heightened sensitivity to profitability variables rather than shipping indices or non-financial factors, drawing on their broad experience to safeguard against financial uncertainties and fluctuations in the market. This leads them to focus on diversifying their business operations as a strategy to enhance profit margins (Sousa et al., 2022). This comparative analysis advocates for the customization of bankruptcy prediction models to reflect the distinct characteristics inherent to shipping companies of varying sizes. By integrating variables that accurately represent the specific challenges and operational realities faced by large firms and SMEs, the efficacy of bankruptcy prediction models can be significantly improved.

Table 7.7. Comprehensive variable set of large firm and SMEs

Rank	Large firms		SMEs	
	Category	Variable	Category	Variable
1	Profitability	Net Income	Non-financial	Size
2	Profitability	Gross Profit to Current Liabilities	Shipping index	No. of vessel
3	Profitability	Sales to Liabilities	Liquidity	Current Assets
4	Leverage	Retained Earnings to Total Assets	Shipping index	Growth rate of global seaborne trade Volume
5	Profitability	Interest expense to assets	Shipping index	LIBOR
6	Liquidity	ΔTotal Asset	Profitability	Sales
7	Profitability	Gross Profit to Assets	Non-financial	Type of Operator
8	Profitability	Gross Profit to Liabilities	Liquidity	Shareholder's Equity
9	Leverage	Current Assets to Liabilities	Non-financial	GDP Growth rate
10	Profitability	Operating Return on Assets	Shipping index	Growth rate of Korean seaborne trade Volume

7.3.3. Comprehensively proposed variables for bankruptcy prediction

From an analysis of 80 input variables, 35 have been identified as having a significant impact on bankruptcy prediction within the entire shipping industry. However, the significance and influence of these variables vary across different forecasting periods. To address research question regarding the construction of a set of risk factors tailored to different forecasting horizons, Table 7.8 illustrates the frequency with which each variable was selected as important in previous analyses across different forecasting periods.

This table provides a comprehensive summary of variables that significantly influence bankruptcy predictions. It highlights that the effect of each variable on a shipping company's bankruptcy risk changes over time. Identifying an optimal set of explanatory variables that can act as early warning signals for bankruptcy, tailored to specific forecasting periods, is crucial for effective risk management and strategic planning in the shipping industry.

Table 7.8. Frequency of Variables with High Importance in Bankruptcy Prediction Across Different Forecasting Horizons

	Variable	1 year	3 years	5 years
Leverage		**1**	**2**	**4**
A2	Retained Earnings to Total Assets	1	0	0
A14	EBITDA	0	1	2
A16	Net Income to Liabilities	0	0	1
A18	Current Assets to Liabilities	0	1	1
Liquidity		**5**	**5**	**3**
B3	Cash Assets Ratio	0	1	0
B9	Shareholder's Equity	0	1	1
B11	Current Assets	1	2	1
B15	Quick Assets to Total Assets	0	0	1
B16	Working Capital to Equity	1	0	0
B19	Retained Earnings to Current Liabilities	1	0	0
B20	ΔTotal Asset	2	1	0
Profitability		**6**	**9**	**12**
C3	Return on Sales	0	1	1
C5	Sales to Current Assets	1	0	0
C6	Sales	0	1	2
C8	Gross Profit to Assets	0	1	1
C9	Interest expense to Assets	2	0	0
C11	Net Income	1	1	2
C12	Operating Return on Assets	0	0	1
C13	Sales to Liabilities	1	1	0
C14	Gross Profit to Current Liabilities	0	1	0
C15	Gross Profit to Liabilities	0	0	1
C16	Gross Profit to Sales	0	1	2
C17	ΔSales	1	2	2
Efficiency		**2**	**1**	**0**
D1	Asset turnover	1	0	0
D3	Current Assets to Sales	1	0	0
D4	Inventory to Assets	0	1	0
Non-financial		**8**	**4**	**6**
E1	Size	2	2	2
E2	Age	2	0	0
E3	Type of Operator	2	2	2
E4	GDP Growth rate	2	0	2
Shipping index		**8**	**9**	**5**
F1	LIBOR	1	2	2
F2	Growth rate of global seaborne trade Volume	2	2	2
F3	Growth rate of Korean seaborne trade Volume	2	2	0
F7	No. of vessel	1	3	1
F8	G/T	2	0	0

For instance, leverage ratios such as the current asset-to-liabilities ratio and profitability ratios like net income frequently exhibit high Shapley values in the 5-year forecasting period. This suggests their utility in long-term bankruptcy predictions. Conversely, shipping indices and liquidity ratios, including the growth rate of total assets and the gross tonnage of ships owned, exert a significant influence in the one-year forecast period, indicating their value for short-term forecasting. Additionally, certain variables, like the size of firms or the growth rate of global seaborne trade volume, maintain a significant impact on bankruptcy predictions across all forecasting horizons.

This analysis highlights the importance of tailoring the selection of predictive variables to the specific time frame of the bankruptcy forecast, facilitating more accurate and timely predictions within the shipping industry.

Given the varied predictive abilities and influences measured by the SHAP value of each variable, Table 7.9 proposes comprehensive sets of variables for bankruptcy prediction in the shipping industry, segmented by forecasting horizons.

Table 7.9. Proposed sets of variables for bankruptcy prediction by forecasting horizon

Rank	1 year		3 years		5 years	
	Type	**Variable**	**Type**	**Variable**	**Type**	**Variable**
1	Non-financial	Size	Shipping index	No. of vessel	Shipping index	LIBOR
2	Non-financial	Type of Operator	Non-financial	Size	Non-financial	Size
3	Profitability	Interest expense to assets	Shipping index	Growth rate of global seaborne trade Volume	Profitability	Net Income
4	Shipping index	Growth rate of global seaborne trade Volume	Shipping index	Growth rate of Korean seaborne trade Volume	Non-financial	Type of Operator
5	Liquidity	ΔTotal Asset	Liquidity	Current Assets	Shipping index	No. of vessel
6	Shipping index	G/T	Non-financial	Type of Operator	Leverage	EBITDA
7	Non-financial	GDP Growth rate	Profitability	ΔSales	Profitability	Gross Profit to liabilities
8	Shipping index	Growth rate of Korean seaborne trade Volume	Shipping index	LIBOR	Non-financial	GDP Growth rate
9	Non-financial	Age	Profitability	Gross Profit to Current Liabilities	Profitability	ΔSales
10	Shipping index	No. of vessel	Profitability	Net Income	Shipping index	Growth rate of global seaborne trade Volume

Short-term predictive power is particularly highlighted through variables such as the ratio of interest expenses to assets and the company's age. Newly established firms, defined as those in operation for less than five years, often face a higher bankruptcy risk due to internal deficiencies, including inadequate management, a lack of industry knowledge, inaccurate financial calculations, and an underestimation of necessary capital for operations (Jang et al., 2021). The shipping industry's reliance on debt financing further exacerbates this risk, thus an increase in interest expenses can place companies with floating interest rates in immediate bankruptcy risk (Lin et al., 2005). Specifically, small-sized shipping companies, with their limited diversification in business portfolios, are especially vulnerable to bankruptcy due to elevated debt financing (Park et al., 2022). The significance of interest expenses also ties into the predictive ability of the LIBOR interest rate for the long term. Rising LIBOR rates can reduce shipping costs, affecting vessel break-even points and lease payments for shipping finance negatively (Grammenos, 2013).

Moreover, unexpected interest rate fluctuations account for a significant portion of shipping risk management, potentially leading to cash flow and liquidity issues (Sousa et al., 2022). Such problems may arise as companies become incapable of meeting their debt obligations, often because shipping financing demands substantial asset capital, typically sourced from international commercial bank loans (Mok; and Ryoo, 2022). Bankruptcy issues in shipping companies thus stem from decreased liquidity, evident in the growth rate of current or total assets. To mitigate financial risk, shipping companies might need to undergo significant operational changes and drastic restructuring, including asset reduction (Yoon et al., 2023).

The size of firms, represented by the logarithm of total assets and the number of vessels owned, also plays a critical role in predictive accuracy across forecasting periods. Additionally, the shipping market's inherent volatility introduces unpredictable shifts in demand and freight rates, posing severe financial risks (Kamal et al., 2021). Such volatility, often influenced by economic shifts like global GDP growth or changes in maritime trade volumes, underscores the necessity for shipping companies to focus on profitability measures, such as the gross profit-to-liabilities ratio, to enhance profit generation efficiency.

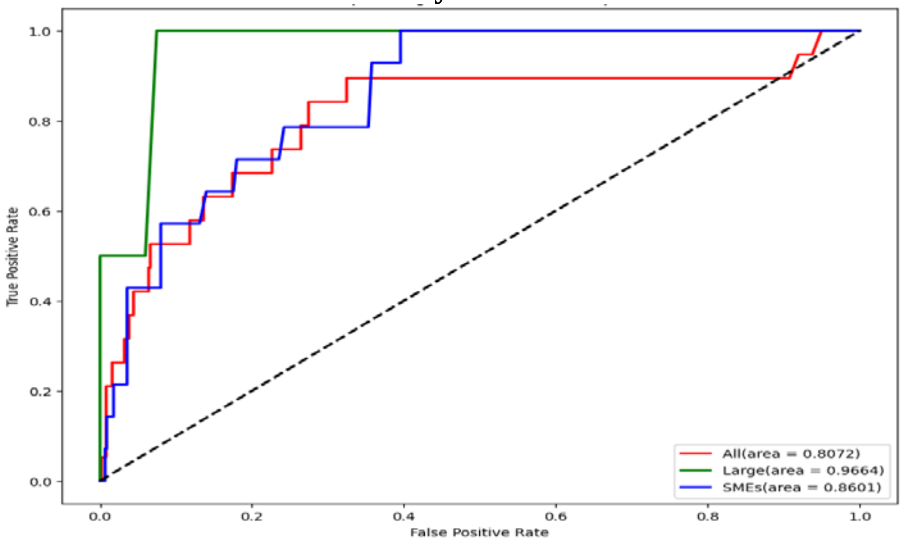
The assessment of the predictive ability of risk factors identified in Table 7.9 was conducted through the development and implementation of prediction models for different segments of the shipping industry which covering the entire sector, large corporations, and small to medium-sized enterprises (SMEs) across various forecasting horizons (1, 3, and 5 years). The findings, presented in Table 7.10, demonstrate that the predictive performance of each model segment remains impressively strong, as evidenced by Receiver Operating Characteristic (ROC) values and accuracy rates exceeding 90% throughout all forecasting periods. These findings point out the effectiveness of the chosen variable sets in enhancing the decision-making processes of stakeholders in the shipping industry, particularly with regard to bankruptcy prediction. Furthermore, Figure 7.5 displays the model's consistently high performance across the forecasting periods, indicating that the unique contribution of each variable, which varies with the forecasting timeframe, significantly enhances the predictive precision of the formulated variable sets in bankruptcy prediction.

Table 7.10. Result of suggested prediction model with sets of variables

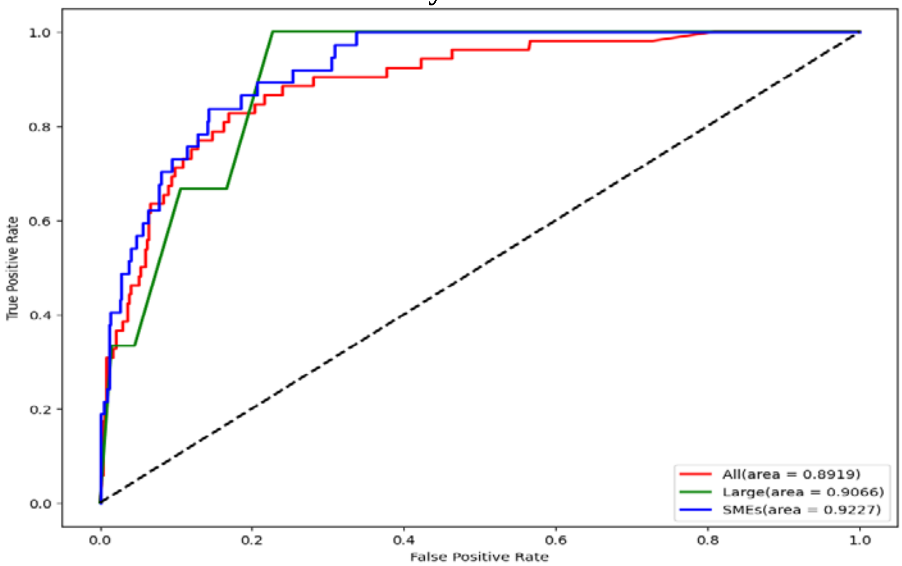
Period	**Shipping**		**Large firms**		**SMEs**	
	ROC score	Accuracy	ROC score	Accuracy	ROC score	Accuracy
1 year	0.8072	95.6%	0.9664	97.1%	0.8601	95.9%
3 years	0.8919	89.6%	0.9066	95.7%	0.9227	89.4%
5 years	0.8974	87.1%	0.9516	92.8%	0.9288	90.9%
Average	0.8655	90.8%	0.9415	95.2%	0.8977	92.1%

Figure 7.5. ROC curve of suggested prediction models by forecasting horizons

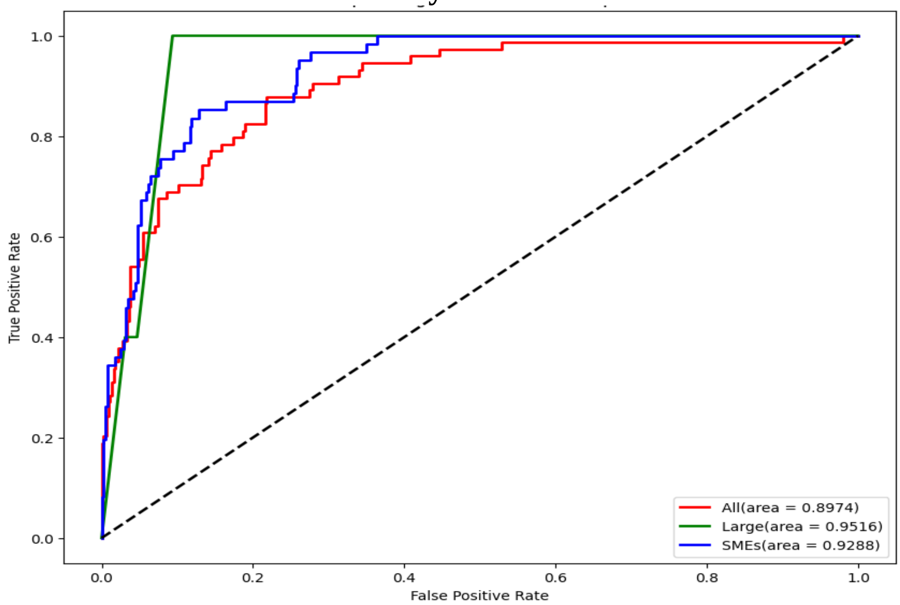
1 year



3 years



5 years



7.4. Chapter Summary

The aim of this chapter is to evaluate the influence of different variables on bankruptcy risk within the shipping industry, as per the models proposed in Chapter 6. This evaluation utilizes the SHAP method, an explainable AI technique, to clarify the decision-making processes of machine learning models. Variables were ranked by their SHAP values across diverse forecasting horizons and categorized into three segments: the overall shipping industry, large shipping corporations, and small to medium-sized shipping enterprises (SMEs).

Overall, there were significant differences in influence of variables across different prediction horizons. With the extension of the forecasting horizon, the disparity in influence among variables narrowed, with financial ratios, particularly profitability metrics such as net income and sales growth rate, seeing an enhancement in their impact. Conversely, macroeconomic factors, including the LIBOR interest rate and the growth rates of global and Korean trade volumes, maintained significant explanatory capacity throughout all forecasting periods. Despite the high volatility associated with these variables in the models, their elevated SHAP values across the forecasting horizons underscore their substantial contribution to bankruptcy prediction in the shipping sector. Additionally, empirical findings highlighted the importance of non-financial variables, like the size of firms represented by the logarithms of total assets or the number of vessels owned, in forecasting bankruptcy. For large shipping entities, profitability ratios emerged as crucial predictors of bankruptcy, whereas small shipping companies were predominantly influenced by macroeconomic factors, such as shipping indices and non-financial variables.

Drawing on the insights from the interpretation of results, this research outlines an optimal set of variables for each forecasting period. The application of the research model, as validated through empirical analysis, demonstrated that the selected variable sets possess considerable explanatory capacity across forecast horizons, not only for the shipping industry at large but also for both large firms and SMEs within the sector. Consequently, the curated set of variables effectively encapsulates the variance in variable impact across forecast periods and distinct characteristics of both large companies and SMEs in the shipping industry.

8. CONCLUSION

8.1. Introduction

The major objective of this research aims to explore the key risk factor for predicting bankruptcy risk of the small and medium-sized shipping companies. To fulfil this research goal, identifying bankruptcy risk factors specific to the shipping industry was essential, achieved through literature review and interviews. Subsequently, the predictive capability of the selected variables was assessed by deploying predictive models utilizing machine learning techniques. These two primary tasks were executed via exploratory research and empirical analysis, as depicted in Figure 1.1. The main objective of the exploratory research aimed to identify risk factors from conventional research models which can improve bankruptcy prediction in terms of small and medium-sized shipping companies. Theoretical justification of research was proved by prevalent bankruptcy risk issues identified in the Korean shipping industry (Chapter 2). Literature review conducted to provide theoretical background for target data, explanatory variables and research modelling (Chapter 3). The research model for this study was formulated based on insights from the literature review and methodological modelling (Chapter 4).

The empirical study then assessed the predictive power of various variables using the research model proposed in the exploratory study. To secure the validity of the identified explanatory variables, interview was conducted in Korean shipping industry. The general patterns of variables were provided by descriptive analysis to identify bankruptcy symptoms over a five-years periods (Chapter 5). Subsequently the research model evaluated predictive ability of variables with machine learning models; extreme gradient boosting and long short-term memory (Chapter 6). These models were assessed impacts of each variable over the different forecasting horizons by applying explainable artificial intelligent techniques (Chapter 7). Afterwards, the risk factors for large shipping firms and small and medium-sized enterprises are presented with each predictive horizon as short-term (1 year), medium-term (3 years), and long-term (5 years).

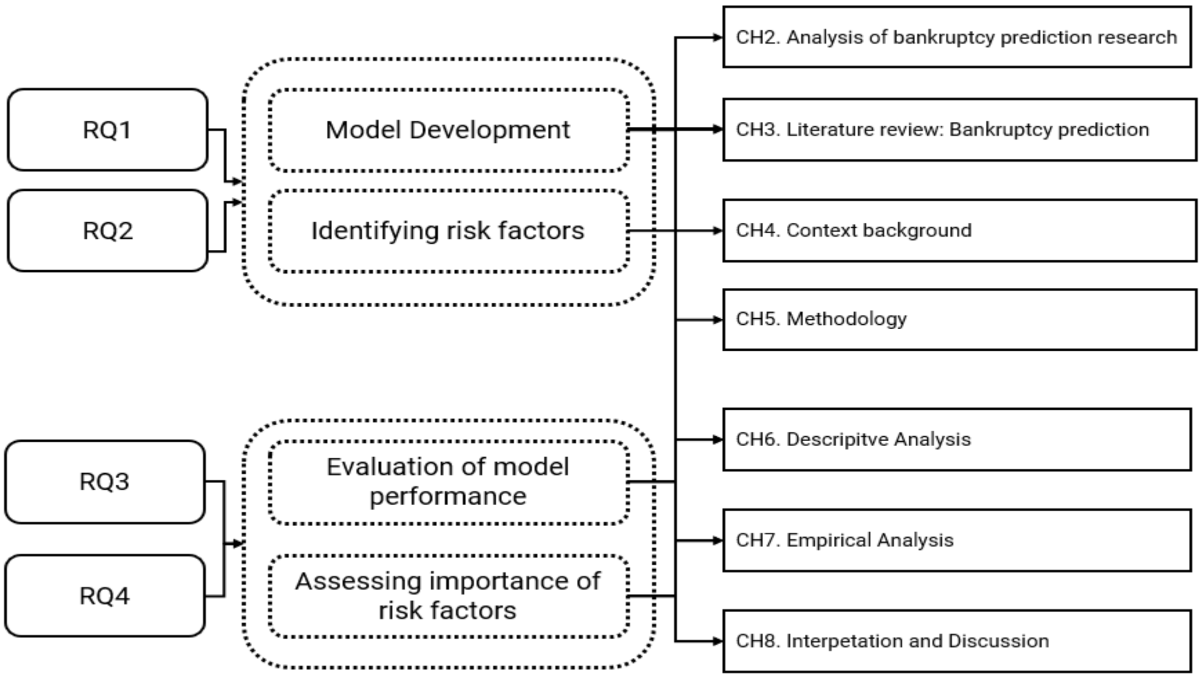
8.2. Research finding

The research questions of this study presented in Chapter 1 are as follows:

1. How has the development of bankruptcy prediction models evolved over time, and what factors and research methodologies can be employed in this evolution within the context of the shipping industry?
2. Which risk factors can be identified for bankruptcy prediction of small and medium-sized shipping companies in the Korean shipping industry through a combination of literature review and practitioner interviews?
3. How do bankruptcy risk factors vary across different forecasting horizons (1 year, 3 years, 5 years prior to bankruptcy) when utilizing machine learning models for improving bankruptcy prediction in the shipping industry?
4. How can be identified key risk factors by assessing the contribution through explainable AI over the different forecasting horizons, and how can these findings contribute to the enhancement of practical policies and managerial strategies in shipping industry?

Figure 8.1 demonstrates the approach taken to address the research questions in this study. RQ1 and RQ2, focusing on identifying potential bankruptcy risk factors, were tackled during the variable identification process through literature review and interviews, as conducted in Chapters 2 and 3. The justification for those variables, supported by robust academic foundations, was provided in Chapter 4. The research questions, RQ3 and 4 were addressed through the empirical analysis. RQ3 was dealt with the research model evaluation process with applying machine learning models from previous researches with different forecasting horizons in Chapter 7. The predictive ability of those identified variables were evaluated by the research model and explainable AI techniques in Chapter 8. The research findings and implications are described in detail in this order.

Figure 8.1. Conceptual flow of research questions



8.2.1. Model development and identifying risk factors: RQ1 and RQ2

The research questions 1 and 2 aim to construct a robust model for predicting bankruptcy within the shipping industry, with a specific focus on identifying variables that encapsulate the distinct characteristics of small and medium-sized shipping companies. The predictive capacity of these variables was meticulously established through extensive literature reviews and further corroborated by interviews with seasoned practitioners in the shipping sector.

Post the 2008 global financial crisis, bankruptcy prediction research garnered heightened attention due to its critical role in mitigating the adverse economic impacts on various sectors, including the pivotal survival of numerous companies (Shi and Li, 2019). A key strategy employed by companies to cushion against such crises involves seeking financial support from institutions, which necessitates these institutions to adeptly pinpoint firms with a high bankruptcy risk. Consequently, developing diverse, effective predictive models to identify early warning signs has become paramount for stakeholders like businesses, investors, government bodies, and banks (Wang et al., 2021). Historically, this research has concentrated on assessing a company's financial health and performance prospects, primarily using financial ratios as the cornerstone for bankruptcy prediction models, attributed to their proven high predictive ability (Jones and Wang, 2019).

However, the evolution of these models since the 2008 crisis reflects a paradigm shift from a sole reliance on financial ratios to an integration of macroeconomic factors and other non-financial indicators. This shift acknowledges the complex economic environments of the modern era (Alam et al., 2021b). Subsequent research underscores the significance of incorporating macroeconomic variables into bankruptcy prediction models, particularly in the wake of unpredictable crises like the COVID-19 pandemic (Cheng et al., 2018). The shipping industry, specifically, experienced a seismic disruption with the bankruptcy of Hanjin Shipping, which accounted for a considerable 5% of global shipping volume (Aydın and Kamal, 2022). This recession precipitated widespread vessel immobilization, inflicting significant financial strain on Korean SMEs in the shipping industry, which still remained to this day. Despite the heightened financial risks associated with SMEs, attributed to their reliance on short-term credit and challenges in accessing longer-term financing, bankruptcy prediction for these enterprises has been relatively underexplored (Ciampi and Gordini, 2013). This lack of focus is partly due to the limited attention these firms receive from market participants, contrasting with larger corporations (Cultrera and Brédart, 2016).

In addressing this gap, it becomes imperative to pinpoint variables that accurately reflect the unique attributes of shipping SMEs. Through a series of interviews with Korean shipping industry experts, this study validates variables identified from prior research and introduces alternative indicators such as the LIBOR interest rate, vessel ownership status, and Korea's annual trade volume as pertinent macroeconomic factors. These variables were carefully integrated into the development of bankruptcy prediction models.

The evolution of these models, especially within the shipping industry, marks a significant shift from traditional statistical methods to more complex machine learning and ensemble techniques. Although conventional models, such as discriminant analysis and logistic regression, are valued for their interpretability, they frequently fall short in accurately predicting bankruptcies in the complex shipping industry context, primarily due to their restricted handling of financial ratios. In contrast, machine learning algorithms exhibit enhanced predictive accuracy but are often criticized for their opacity, which poses challenges for managerial decision-making as the vital consideration in the shipping industry. The applicability and efficacy of these models in predicting bankruptcy for Korean shipping SMEs hinge on a multitude of factors, including the specific data structure, the industry's unique features, and the targeted objectives of the prediction model. Notably, ensemble methods like eXtreme Gradient Boosting (XGB), which amalgamate multiple predictors, demonstrate potential in augmenting performance, though no single technique uniformly excels in all scenarios. Therefore, selecting the appropriate model for predicting bankruptcy in Korean shipping SMEs entails a nuanced balance between predictive accuracy and interpretability, tailored to the distinct needs and peculiarities of the industry.

In conclusion, this study significantly enhances the development of more effective bankruptcy prediction models for the shipping industry. These models are specifically designed to meet the unique challenges encountered by shipping SMEs in Korea, offering a thorough understanding of the factors that influence bankruptcy risk within this sector.

8.2.2. Evaluation of model and assessing importance of risk factors: RQ3 and RQ4

Research question 3 and 4 aimed to evaluate the predictive ability of variables by applying bankruptcy prediction models with machine learning techniques over the different forecasting horizons. As big data and machine learning algorithms evolve, corporate bankruptcy prediction methods are transitioning from traditional statistical approaches to data-driven machine learning techniques. Compared to conventional statistic models, machine learning models focus more on improving prediction accuracy through multi-dimensional dataset rather than pay much attention to structure of research modelling (Jones, 2017). Even though, those complex machine learning models can achieve high level of predictive ability, those models can not suggest the interpretation of results, which makes to reduce the practical application value of these techniques. To overcome this limitation, the explainable ai technique, SHAP methods had been applied to evaluate the contribution of its variables.

Bankruptcy prediction models generally necessitate a binary target variable, necessitating the association of each financial statement with an indicator variable that signifies whether the company went bankrupt. In practice, there can be a delay of up to three years between a company's economic default and its formal declaration of bankruptcy (Tinoco and Wilson, 2013). Notably, in our dataset, the majority of bankrupt companies filed for bankruptcy within one or two years, or underwent mergers and acquisitions (M&A) or restructuring following their last financial statement. Therefore, it is necessary to identify risk factors from financial datasets by different forecasting horizons for bankruptcy prediction. In particular, most of the datasets are consisted of unlisted small-sized companies which that are not obligated to report financial statements annually, making it very important to find meaningful indicators among the available datasets. Furthermore, compared to large companies, SMEs are more vulnerable to bankruptcy risk due to their weaker financial structures, lower capitalisation, and difficult to raise external financing by their lower credit, thus they are affected immediately. Therefore, those identified variables can contribute to proactive response and management for shipping companies as the risk signal factors in different forecasting horizons. As suggested from Chapter 8, macroeconomic variables including shipping index significantly contribute to bankruptcy prediction for shipping SMEs, while profitability such as net income or gross profit-liabilities ratios affected more to large shipping companies. Although there was no apparent difference in the composition of the set of variables, the impact of each variable varied depending on the prediction horizon. For mid-term predictive periods (5-year), leverage ratios such as current asset-to-liabilities ratio and profitability ratios such as net income showed great contribution to bankruptcy prediction for shipping industry. On the other hand, shipping indices such as the growth rate of total assets or the number of ships owned showed high influence over the one-year forecast period, which can be considered to have short-term forecasting power. Finally, some variables contributed to have a significant impact on bankruptcy predictions over the entire forecasting horizons such as size of firms or growth rate of global seaborne trade volume.

The indicators derived from the research model can offer valuable insights for policymakers and market participants in the shipping industry. By understanding the current market state and predicting its cyclical fluctuations, shipping companies can refine their business strategies. Specifically, if the composite indicators signal an impending recession, companies could mitigate potential losses by securing vessel space in advance or opting for time charters. Policymakers, on the other hand, could use these diagnostic insights to enact regulations aimed at stimulating market conditions. The findings of this study are particularly relevant to shipowners, shipping companies, and banks involved in risk management within the shipping sector, as they can leverage the composite indicators for a deeper understanding of market. Additionally, this information can guide policymakers in revising existing policies to better support the industry.

8.3. Contribution

8.3.1. Research Contribution

This study marks a significant contribution in the field of bankruptcy prediction, particularly within the context of the shipping industry. First of all, it addresses a notable gap in the literature, as identified by Shi and Li (2019), regarding the no common agreement on the risk factors that accurately predict bankruptcy within this sector. By embracing a holistic approach that includes a diverse array of variables, the research emphasizes the critical importance of tailoring analysis to the specific operational and economic realities of the shipping industry. This methodological innovation extends beyond traditional financial metrics to incorporate macroeconomic indicators, non-financial markers, and shipping indices, thereby offering a more accurate and industry-specific framework for bankruptcy prediction.

Second, this study broadens the analytical scope beyond the commonly studied immediate pre-bankruptcy period by examining risk factors over extended forecasting horizons, specifically, 1, 3, and 5 years prior to bankruptcy. This longitudinal approach reveals significant variations in the relevance of different risk factors depending on the timeframe considered, shedding light on the dynamic nature of financial distress within the shipping industry. This can provide insight into incorporating longer-term perspectives in bankruptcy prediction models, thus providing a much critical understanding of corporate bankruptcy risk.

The employment of advanced machine learning techniques, including extreme gradient boosting and long-short term memory models, represents a methodological improvement in this sector. These sophisticated analytical tools have proven exceptionally adept at navigating the complexities of the shipping industry's typically small and imbalanced datasets. By successfully addressing inherent data challenges such as missing values, outliers, and skewness, the study not only demonstrates the practical applicability of these techniques but also significantly enhances the predictive accuracy of bankruptcy models. This contribution is particularly noteworthy, as it illustrates the potential of cutting-edge computational methods to refine and improve traditional models of financial analysis within specific industry contexts.

Finally, the application of the SHAP method which is an explainable AI technique addresses a crucial challenge in the use of machine learning models: the opacity of their predictive processes. By providing a transparent and detailed interpretation of how each variable influences bankruptcy risk predictions, the study demystifies the analytical process and offers a prioritized list of risk factors. This contribution is invaluable for stakeholders within the shipping industry, including investors, managers, and policy-makers, as it offers clear, actionable insights into which variables warrant close monitoring over different forecasting periods. Such clarity not only enhances the utility of bankruptcy prediction models but also facilitates more informed decision-making, ultimately contributing to more robust risk management and strategic planning within the shipping sector.

Overall, these contributions represent a significant step forward in the academic field of bankruptcy prediction, offering both theoretical insights and practical tools that are specifically tailored to the unique challenges and characteristics of the shipping industry. Through its comprehensive approach and methodological advancements, this study not only enriches the academic discourse but also provides a valuable framework for industry practitioners seeking to navigate the complex area of financial risk assessment.

8.3.2. Managerial contribution

Considering the capital-intensive nature, shipping industry requires substantial capital investment and maintains a high asset proportion. Therefore, efficient financial risk management is paramount for companies to manage highly volatile relationships with banks and financial institutions, and it is necessary to establish a bankruptcy risk management tool which is capable of monitoring the current state of the industry and predicting future trends. Notably, in terms of the shipping industry, only a small number of large companies operate their own risk management system, which makes them susceptible to bankruptcy risk (Park et al., 2022).

This study contributes to the enhancement of corporate risk management by identifying the influence of changes in internal and external environments on corporate financing activities. This analysis extends beyond traditional financial ratios to include macroeconomic factors and shipping indices. The research reveals that most small and medium-sized shipping companies, which are typically more vulnerable to market fluctuations such as changes in global seaborne trade volume or LIBOR interest rates, primarily engage in a single shipping business. As indicated by results from this research, unlike large shipping companies where profitability is paramount, small and medium-sized shipping companies that are greatly affected by market volatility must seek to mitigate economic fluctuations by diversifying their business portfolio. Beyond traditional shipping transportation service, these companies should expand their investment into new industrial sectors such as forwarding operations or alternative energy in pursuit of sustainable revenue generations.

In the context of the Korean shipping industry, there has been a tendency for new shipbuilding investments to concentrate on large and container shipping companies. From the past three years, 80% of shipbuilding orders received by Korean shipping companies have come from large container shipping companies (Kwon et al., 2023). These large shipping companies, possessing a diverse fleet composition including containers and bulk carriers, have currently engaged in shipbuilding investments driven by improving profitability factors such as net income or gross profit ratios. In contrast, the shipbuilding orders for small and medium-sized shipping companies, except for container shipping companies, appear to be at an insufficient level, due to their lack of financial capacity. In particular, most ships owned by these smaller companies are aging bulk carriers that have not been replaced, indicating a lag in investment for fleet renewal. As the value of ships is increased due to these technological advancements, it is not desirable to retain older ships like in the past and focus solely on maritime transportation services. Considering economic feasibility, there is a pressing need for these companies to acquire the latest ships, utilize them efficiently in their business operations, sell them opportunistically, and invest in competitive vessels as needed. This requires a strategic approach, grounded in experience and data analysis, with a long-term view towards maximizing profits and considering potential losses from decreased competitiveness associated with maintaining older vessels. Moreover, the recent strengthening of environmental regulations has created an urgent need for the replacement investment in new ships, as the normal operation and navigation of aging vessels have become increasingly challenging. Considering the high proportion of small and medium-sized ships in the fleet of these companies, the focus should be on phasing out uncompetitive vessels and increasing investment in new shipbuilding. In preparation for environmental regulations, proactive measures are needed to introduce and invest in eco-friendly ships for improving their profitability.

However, the majority of small and medium-sized shipping companies appear to have an insufficient financial capacity to fully engage in shipbuilding investments, and financing issues also emerged as evidenced by current assets and equity, which have been identified as key risk factors in this research. Since the bankruptcy of Hanjin Shipping in 2017, the Korean shipping industry has experienced a significant recovery period, with a challenging investment environment. Furthermore, there have been concerns that competitiveness of Korean shipping companies would be reduced due to the absence of a powerful major shipping company which can operate deep sea routes. If global shipping companies reduce supply of their fleets to Korean market to strengthen their negotiating power in freight rates, small sized shipping companies which are unable to operate the intercontinental routes, would struggle to maintain stable shipping operations. In fact, Korean shipping companies are estimated to have incurred an additional annual cost of 1.4 trillion won compared to Japanese companies following the receivership of Hanjin Shipping (KPMG, 2021).

The risk factors presented in this research, particularly those reflecting the size of the company such as the number of owned vessels, total assets, and current assets, have been shown to substantially affect the bankruptcy risk management strategies of shipping firms. Consequently, to address these challenges, it is imperative for shipping Small and Medium-sized Enterprises (SMEs) to engage in collaborative endeavours and pursue mergers and acquisitions (M&A). Such strategic alliances are essential to expand and consolidate the size of the shipping industry, achieving cost savings and operational efficiencies, and thereby enhancing network competitiveness across both established and emerging markets. Aggressive M&A activities and strategic alliances can also help secure freight rate competitiveness and achieve economies of scale. By expanding fleet size and company scale, particularly for smaller shipping companies, economies of scale can be realized, ensuring stable profitability in preparation for market fluctuations.

8.3.3. Policy Contribution

Various Korean shipping companies have faced bankruptcy during downturns following periods of prosperity due to a failure in effectively managing market risks associated with freight rate fluctuations. This pattern is evident from significant risk factors such as the growth rate of shipping indices or GDP, highlighted in this research. In particular, concerns arose within the global shipping industry during the COVID-19 pandemic crisis regarding global economic downturns (Kwon et al., 2023). However, these concerns were mitigated somewhat by expansive fiscal policies in major countries, spurring consumer spending and freight rate increases. Despite the unprecedented prosperity experienced by the shipping industry, policy responses for the Korean shipping market have primarily been constrained response by deficiencies in monitoring systems (KPMG, 2021). As suggested in this study, it is imperative for shipping companies to establish systems that systematically assess market volatility, and manage the profitability. Nevertheless, only a few Korean shipping companies have implemented such systems. Despite the Korean government's formulation of the '5-Year Shipping Reconstruction Plan' since the bankruptcy declaration of Hanjin Shipping in 2017, support has predominantly focused on financial assistance to large shipping companies, leaving inadequate support for small and medium-sized shipping companies (Park et al., 2021b).

Considering the risk factors identified in this study, such as the Current asset or Total asset ratio that reflects liquidity of company, it is essential to emphasize the potential occurrence of investment shortages and insolvency issues in small and medium-sized shipping companies due to chronic liquidity deterioration. Small and medium sized shipping companies, despite having inferior financial structures, low assets, and low credit ratings compared to large companies, face limited options for financing and policy support in financial markets (Luo et al., 2020). The economic environment surrounding the shipping industry continues to heavily rely on financial institutions as external sources of financing (Alexandridis et al., 2020). Shipbuilding investments over the past three years have predominantly been undertaken by large shipping companies, while investments by small and medium sized companies have been insufficient, indicating a lack of funding capacity (Park et al., 2021b). Therefore, maintaining the stability of financial structures through expanding financing sources can be a crucial condition for small and medium sized shipping companies to sustain competitiveness during economic downturns. This implies the need for financial instruments capable of covering funds in preparation for liquidity deterioration in response to reduced freight income during shipping downturns. In Korean shipping industry, numerous shipping companies have faced liquidity deterioration during shipping downturns, and a cycle of selling profitable assets such as ships to repay debt has persisted through restructuring (Park et al., 2022). In this regard, the government should implement measures to facilitate financial access for small shipping companies by securing stable financial instruments. For vulnerable small and medium-sized shipping companies in the highly volatile shipping market, there is a need for continuous safety mechanisms, serving as financial instruments to cover funds in anticipation of liquidity deterioration during shipping downturns. During sudden economic downturns leading to liquidity crises for shipping companies, the provision of emergency management funds with special interest rates and no credit rating restrictions can help manage such crises. For instance, banks and credit rating agencies should reflect the characteristics of shipping companies with relatively high debt ratios, assess insolvency risks, and enable efficient financing for shipping companies.

Furthermore, acknowledging the pivotal significance of size of company and number of owned vessels as elucidated in this study, it is imperative that the government facilitates competitiveness enhancement by endorsing the expansion of fleet sizes, particularly through the subsidization or support of new vessel construction. After the bankruptcy of Hanjin Shipping, the global operating routes of Korean shipping companies have been reduced, resulting in restrictions on the use of Korean-flagged vessels. Korean-flagged fleets, especially those of smaller shipping companies, have not fully recovered and remain insufficient, failing to reach the levels seen in 2016. When Hanjin Shipping declared bankruptcy in 2016, 35 out of its 61 owned vessels were sold and leased to overseas shipping companies. Currently, Hyundai Merchant Marine (HMM), the largest container shipping company in Korea, exhibits a smaller scale and inadequate ownership of mega vessels compared to major shipping companies. Its fleet size accounts for only

16% of Maersk, the world's top shipping company, and 51% of Evergreen, the seventh-ranked shipping company (KSA, 2021). While major shipping companies have expanded their scale through mergers and acquisitions and cost-cutting efforts via mega-vessel orders, HMM has been marginalized due to internal restructuring. Even though SM Line acquired Hanjin Shipping's North American routes, they still faced limitations in fleet size growth with a global ranking of 26th, which is accounted for 0.2% of supply market share (Hwang et al., 2017). Particularly, after the outbreak of COVID-19, a phenomenon occurred where global liners bypassed Korea and directly sailed from Shanghai to North America, causing significant disruptions in exports due to the inability of domestic shippers to secure vessel capacity. Therefore, by expanding shipping financing support from the government, initiative would allow shipping companies to introduce new vessels, improve profitability by cost competitiveness, and strengthen their position within the industry through route diversification.

Additionally, special guarantees and expanded support for ship investments are necessary for small and medium-sized shipping companies with poor financial conditions and limited access to ship financing. Applying special interest rates when raising funds from financial institutions can promote the acquisition of new and used vessels. For shipping companies which sold their assets to repay their liability, the government can also assist them in securing sound finances through the acquisition of second-hand vessels and restructuring through Sale and Leaseback (S&LB) policy. This approach allows for efficient fleet expansion and financial stabilization, enabling shipping companies to reinvest a portion of their earnings into the shipping industry. Especially, the necessity of improving government financial support has been increased for proactive responses to tightened global environmental regulations recently. Many small and medium sized shipping companies operate ships that are over 20 years old and operate only one or two vessels, facing cost pressures due to the need for eco-friendly capabilities such as LNG engines or exhaust scrubbers (Yeo et al., 2022). Particularly financial support policies for the transition to eco-friendly vessels should be needed for small and medium-sized companies with limited liquidity and financing capabilities. Given the anticipated surge in demand for eco-friendly ships due to environmental regulations from 2023, this approach would allow even small sized shipping company with relatively lower liquidity to effectively respond to high-level environmental regulations.

Finally, it is necessary to differentiate shipping companies by vessel type, route, and size, and to establish bankruptcy prediction systems tailored to the economic conditions and market fluctuations affecting each group. As demonstrated in this study, the stark differences in risk factors by different predictive periods underscore the potential to enhance the predictive capabilities of crisis response systems by incorporating such distinctions. Moreover, early warnings can be issued to vulnerable groups of shipping companies based on identified risk causes and types during different prediction periods. It allows to customize that takes into account the financial and operational situation of each shipping company, including customized financial support and policy support. Companies detecting bankruptcy risk early can utilize management strategies like fleet redeployment, contributing to fair and stable logistics support for international trade through transportation contracts based on actual freight rates. Furthermore, the establishment of a sophisticated crisis response framework within the shipping industry is imperative. This system should comprehensively integrate market surveillance, predictive analytics, early warning mechanisms, and methodologies for both industry-wide and company-specific insolvency forecasting. Such a comprehensive system would allow for effective management and minimize national economic losses by enabling timely government interventions to detect shipping market risks and corporate insolvency. Continuous market oversight could also contribute to Korean shipping companies making informed decisions regarding investment, business expansion, or contraction. Through this information as presenting risk factors as leading indicators of an impending economic downturn, financial investors, banks, or lending institutions can make much accurate prediction based on advanced bankruptcy prediction models to make well-informed decisions. Through this information, banks and financial institutions can avoid the costs associated with inaccurately assessing bankruptcy risks for shipping companies, especially to small sized companies which may pose higher risks compared to large corporations (Luo et al., 2020).

8.4. Limitation and future implication

Despite the significant contributions of this study, there are limitations that necessitate further discussion and exploration in future research. First, because this study was geographically limited to one country, interpretation of the findings should be generalized with caution. Even though the subject of the study is Korea, which shows clear industrial characteristics, it can change depending on regional characteristics. Although the study was conducted on the Korean shipping industry, which clearly showed the characteristics of the industry with a high proportion of small and medium-sized shipping industry, the results presented in future studies may change depending on the research subject and regional characteristics. Furthermore, due to the fact that small and medium-sized enterprises (SMEs) are not obligated to regularly disclose their financial status to the government, the absence of data precludes the real-time monitoring of bankruptcy risk. Especially, since shipping market index are announced daily with great volatility, the prediction model also required to reflect daily data patterns for improving assessment. Finally, as machine learning models continue to improve to capture more complex relationships between factors, it is noteworthy that these models should also be applied to predictive model development.

From these limitations, this study may suggest the significance for future research. Firstly, this model can be applied wider geographical areas to reflect specific regional characteristics such as country-specific features or shipping operation routes. By applying this study to different countries, it may provide insights to identify for identifying risk factors of bankruptcy prediction for the overall shipping industry. In addition, from this limitation, it can suggest indications to develop further improvement of predictive models. By applying much complex machine learning techniques such as LightGBM or CatBoost, it allows to improve predictive ability to capture explanatory relationships between variables. These techniques can also provide more efficient analysis using large amounts of data, such as daily financial data sets.

REFERENCE

ADAMS, J., KHAN, H. T. & RAESIDE, R. 2014. *Research methods for business and social science students,* SAGE Publications India.

AGARWAL, V. & TAFFLER, R. 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance,* 32, 1541-1551.

AGUIAR-DïAZ, I. & RUIZ-MALLORQUï, M. V. 2015. Causes and resolution of bankruptcy: The efficiency of the law. *The Spanish Review of Financial Economics,* 13, 71-80.

AHN, H. & KIM, K.-J. 2009. Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach. *Applied soft computing,* 9, 599-607.

ALAKA, H. A., OYEDELE, L. O., OWOLABI, H. A., KUMAR, V., AJAYI, S. O., AKINADE, O. O. & BILAL, M. 2018. Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications,* 94, 164-184.

ALAM, N., GAO, J. & JONES, S. 2021a. Corporate failure prediction: An evaluation of deep learning vs discrete hazard models. *Journal of International Financial Markets, Institutions and Money,* 75, 101455.

ALAM, T. M., SHAUKAT, K., MUSHTAQ, M., ALI, Y., KHUSHI, M., LUO, S. & WAHAB, A. 2021b. Corporate bankruptcy prediction: An approach towards better corporate world. *The Computer Journal,* 64, 1731-1746.

ALEXANDRIDIS, G., ANTYPAS, N., GULNUR, A. & VISVIKIS, I. 2020. Corporate financial leverage and M&As choices: Evidence from the shipping industry. *Transportation Research Part E: Logistics and Transportation Review,* 133, 101828.

ALFARO, E., GARCïA, N., GãMEZ, M. & ELIZONDO, D. 2008. Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems,* 45, 110-122.

ALIAJ, T., ANAGNOSTOPOULOS, A. & PIERSANTI, S. Firms default prediction with machine learning. Workshop on Mining Data for Financial Applications, 2019. Springer, 47-59.

ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance,* 23, 589-609.

ALTMAN, E. I. 2018. Applications of distress prediction models: what have we learned after 50 years from the Z-score models? *International Journal of Financial Studies,* 6, 70.

ALTMAN, E. I. & HOTCHKISS, E. 1993. *Corporate financial distress and bankruptcy,* New York: John Wiley & Sons.

ALTMAN, E. I., IWANICZ-DROZDOWSKA, M., LAITINEN, E. K. & SUVAS, A. 2020. A race for long horizon bankruptcy prediction. *Applied Economics,* 52, 4092-4111.

ALTMAN, E. I. & SABATO, G. 2007. Modelling credit risk for SMEs: Evidence from the US market. *Abacus,* 43, 332-357.

ANTUNES, F., RIBEIRO, B. & PEREIRA, F. 2017. Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing,* 60, 831-843.

ARCHER, K. J. & KIMES, R. V. 2008. Empirical characterization of random forest variable importance measures. *Computational statistics & data analysis,* 52, 2249-2260.

ARIZA-GARZõN, M. J., ARROYO, J., CAPARRINI, A. & SEGOVIA-VARGAS, M.-J. 2020. Explainability of a machine learning granting scoring model in peer-to-peer lending. *Ieee Access,* 8, 64873-64890.

ARRIETA, A. B., DïAZ-RODRïGUEZ, N., DEL SER, J., BENNETOT, A., TABIK, S., BARBADO, A., GARCïA, S., GIL-LõPEZ, S., MOLINA, D. & BENJAMINS, R. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion,* 58, 82-115.

ASHRAF, S., GS FëLIX, E. & SERRASQUEIRO, Z. 2019. Do traditional financial distress prediction models predict the early warning signs of financial distress? *Journal of Risk and Financial Management,* 12, 55.

ATKINSON, A. C., RIANI, M. & CORBELLINI, A. 2021. The Box–Cox Transformation: Review and Extensions. *Statistical Science,* 36, 239-255.

AYDıN, M. & KAMAL, B. 2022. A Fuzzy-Bayesian Approach on the Bankruptcy of Hanjin Shipping. *Journal of ETA Maritime Science,* 10, 2-15.

AYESHA, S., HANIF, M. K. & TALIB, R. 2020. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion,* 59, 44-58.

BALCAEN, S. & OOGHE, H. 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review,* 38, 63-93.

BARBOZA, F., KIMURA, H. & ALTMAN, E. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications,* 83, 405-417.

BĂRBUȚĂ-MIȘU, N. & MADALENO, M. 2020. Assessment of bankruptcy risk of large companies: European countries evolution analysis. *Journal of Risk and Financial Management,* 13, 58.

BAUER, J. & AGARWAL, V. 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance,* 40, 432-442.

BEAVER, W. H. 1968. Market prices, financial ratios, and the prediction of failure. *Journal of accounting research*, 179-192.

BEAVER, W. H., CORREIA, M. & MCNICHOLS, M. F. 2011. Financial statement analysis and the prediction of financial distress. *Foundations and Trends® in Accounting,* 5, 99-173.

BEAVER, W. H., MCNICHOLS, M. F. & RHIE, J.-W. 2005. Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting studies,* 10, 93-122.

BECERRA-VICARIO, R., ALAMINOS, D., ARANDA, E. & FERNãNDEZ-GãMEZ, M. A. 2020. Deep recurrent convolutional neural network for bankruptcy prediction: A case of the restaurant industry. *Sustainability,* 12, 5180.

BEHR, A. & WEINBLAT, J. 2017. Default prediction using balance-sheet data: a comparison of models. *The Journal of Risk Finance.*

BELL, E., BRYMAN, A. & HARLEY, B. 2022. *Business research methods,* Oxford university press.

Bergstra, J., Yamins, D., & Cox, D. D. (2013). Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms. In Proceedings of the 12th Python in Science Conference (pp. 13-20).

BENTëJAC, C., CSøRGŐ, A. & MARTïNEZ-MUóOZ, G. 2021. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review,* 54, 1937-1967.

BERK, J., DEMARZO, P., HARFORD, J., FORD, G., MOLLICA, V. & FINCH, N. 2013. *Fundamentals of corporate finance,* Pearson Higher Education AU.

BISHOP, C. M. & NASRABADI, N. M. 2006. *Pattern recognition and machine learning,* Springer.

BOTTOU, L. 2012. Stochastic gradient descent tricks. *Neural networks: Tricks of the trade.* Springer.

BOUKTIF, S., FIAZ, A., OUNI, A. & SERHANI, M. A. 2018. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies,* 11, 1636.

BRACKE, P., DATTA, A., JUNG, C. & SEN, S. 2019. Machine learning explainability in finance: an application to default risk analysis.

BRADLEY, D. B. & COWDERY, C. 2004. Small business: Causes of bankruptcy. *SBANC: Small Business Advancement National Center. University of Central Arkansas*.

BRANCO, P., TORGO, L. & RIBEIRO, R. P. 2016. A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys (CSUR),* 49**,** 1-50.

BREIMAN, L. 1996. Bagging predictors. *Machine learning,* 24**,** 123-140.

BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.

BURRELL, G. & MORGAN, G. 2017. *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*, Routledge.

CALABRESE, R., MARRA, G. & ANGELA OSMETTI, S. 2016. Bankruptcy prediction of small and medium enterprises using a flexible binary generalized extreme value model. *Journal of the Operational Research Society,* 67**,** 604-615.

CARMONA, P., CLIMENT, F. & MOMPARLER, A. 2019. Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance,* 61**,** 304-323.

CHA, S. & KANG, J. 2018. Corporate Default Prediction Model Using Deep Learning Time Series Algorithm, RNN and LSTM. *Journal of Intelligence and Information Systems,* 24**,** 1-32.

CHARITOU, A., DIONYSIOU, D., LAMBERTIDES, N. & TRIGEORGIS, L. 2013. Alternative bankruptcy prediction models using option-pricing theory. *Journal of Banking & Finance,* 37**,** 2329-2341.

CHAUHAN, N., RAVI, V. & CHANDRA, D. K. 2009. Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks. *Expert Systems with Applications,* 36**,** 7659-7665.

CHAWLA, N. V., BOWYER, K. W., HALL, L. O. & KEGELMEYER, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research,* 16**,** 321-357.

CHEN, J., YE, J., ZHUANG, C., QIN, Q. & SHU, Y. 2022. Liner shipping alliance management: overview and future research directions. *Ocean & Coastal Management,* 219**,** 106039.

CHEN, M.-Y. 2011. Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers & Mathematics with Applications,* 62**,** 4514-4524.

CHEN, T.-H., HO, R.-J. & LIU, Y.-W. 2019. Investor personality predicts investment performance? A statistics and machine learning model investigation. *Computers in Human Behavior,* 101**,** 409-416.

CHEN, T. & GUESTRIN, C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016. 785-794.

CHEN, T., HE, T., BENESTY, M., KHOTILOVICH, V., TANG, Y., CHO, H. & CHEN, K. 2015. Xgboost: extreme gradient boosting. *R package version 0.4-2,* 1**,** 1-4.

CHEN, Y.-S., LIN, C.-K., LO, C.-M., CHEN, S.-F. & LIAO, Q.-J. 2021. Comparable Studies of Financial Bankruptcy Prediction Using Advanced Hybrid Intelligent Classification Models to Provide Early Warning in the Electronics Industry. *Mathematics,* 9**,** 2622.

CHENG, C., JONES, S. & MOSER, W. J. 2018. Abnormal trading behavior of specific types of shareholders before US firm bankruptcy and its implications for firm bankruptcy prediction. *Journal of Business Finance & Accounting,* 45**,** 1100-1138.

CHI, L.-C. & TANG, T.-C. 2006. Bankruptcy prediction: Application of logit analysis in export credit risks. *Australian Journal of Management,* 31**,** 17-27.

CHO, S., HONG, H. & HA, B.-C. 2010. A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction. *Expert Systems with Applications,* 37**,** 3482-3488.

CHOI, H., SON, H. & KIM, C. 2018a. Predicting financial distress of contractors in the construction industry using ensemble learning. *Expert Systems with Applications,* 110**,** 1-10.

CHOI, J.-S., KIM, K.-H. & HAN, H.-J. 2018b. Development of early warning index in Korean shipping industry by using signal approach. *Maritime Policy & Management,* 45**,** 1007-1020.

CHRISTIANSEN, M., FAGERHOLT, K., NYGREEN, B. & RONEN, D. 2007. Maritime transportation. *Handbooks in operations research and management science,* 14**,** 189-284.

CIAMPI, F. & GORDINI, N. Using economic-financial ratios for small enterprise default prediction modeling: An empirical analysis. 2008 Oxford Business & Economics Conference Proceedings, Association for Business and Economics Research (ABER), 2008. 1-21.

CIAMPI, F. & GORDINI, N. 2013. Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of I talian Small Enterprises. *Journal of Small Business Management,* 51**,** 23-45.

CIELEN, A., PEETERS, L. & VANHOOF, K. 2004. Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research,* 154**,** 526-532.

CLINTWORTH, M., LYRIDIS, D. & BOULOUGOURIS, E. 2021. Financial risk assessment in shipping: a holistic machine learning based methodology. *Maritime Economics & Logistics***,** 1-32.

COOK, N. R. 2007. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation,* 115**,** 928-935.

CREAZZA, A., COLICCHIA, C. & EVANGELISTA, P. 2023. Leveraging shippers-logistics providers relationships for better sustainability in logistics: the perspective of SMEs. *The International Journal of Logistics Management*.

CRESWELL, J. W., & POTH, C. N. (2016). Qualitative inquiry and research design: Choosing among five approaches. *Sage publications.*

CROSATO, L., LIBERATI, C. & REPETTO, M. 2021. Look Who's Talking: Interpretable Machine Learning for Assessing Italian SMEs Credit Default. *arXiv preprint arXiv:2108.13914*.

CRUTZEN, N. & VAN CAILLIE, D. 2008. The business failure process: an integrative model of the literature. *Review of Business and Economics,* 53**,** 287-316.

CULTRERA, L. & BRëDART, X. 2016. Bankruptcy prediction: the case of Belgian SMEs. *Review of Accounting and Finance*.

DAKOVIC, R., CZADO, C. & BERG, D. 2010. Bankruptcy prediction in Norway: a comparison study. *Applied Economics Letters,* 17**,** 1739-1746.

DANGETI, P. 2017. *Statistics for machine learning*, Packt Publishing Ltd.

DAS, S. R., HANOUNA, P. & SARIN, A. 2009. Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking & Finance,* 33**,** 719-730.

DESAI, V. S., CROOK, J. N. & OVERSTREET JR, G. A. 1996. A comparison of neural networks and linear scoring models in the credit union environment. *European journal of operational research,* 95**,** 24-37.

DI CIACCIO, A. & CIALONE, G. 2019. INSOLVENCY PREDICTION ANALYSIS OF ITALIAN SMALL FIRMS BY DEEP LEARNING.

DIMITRAS, A. I., ZANAKIS, S. H. & ZOPOUNIDIS, C. 1996. A survey of business failures with an emphasis on prediction methods and industrial applications. *European journal of operational research,* 90**,** 487-513.
DROBETZ, W., GOUNOPOULOS, D., MERIKAS, A. & SCHRøDER, H. 2013. Capital structure decisions of globally-listed shipping companies. *Transportation Research Part E: Logistics and Transportation Review,* 52**,** 49-76.
DROBETZ, W., SCHILLING, D. & TEGTMEIER, L. 2010. Common risk factors in the returns of shipping stocks. *Maritime Policy & Management,* 37**,** 93-120.
DZIK-WALCZAK, A. & ODZIEMCZYK, M. 2021. Modelling cross-sectional tabular data using convolutional neural networks: Prediction of corporate bankruptcy in Poland. *Central European Economic Journal,* 8**,** 352-377.
EDMISTER, R. O. 1972. An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative analysis,* 7**,** 1477-1493.
EKLUND, J., LEVRATTO, N. & RAMELLO, G. B. 2020. Entrepreneurship and failure: two sides of the same coin? *Small Business Economics,* 54**,** 373-382.
EL-MASRY, A. A., OLUGBODE, M. & POINTON, J. 2010. The exposure of shipping firms' stock returns to financial risks and oil prices: a global perspective. *Maritime Policy & Management,* 37**,** 453-473.
FALKENSTEIN, E. G., BORAL, A. & CARTY, L. V. 2000. RiskCalc for private companies: Moody's default model. *Available at SSRN 236011*.
FERNãNDEZ, A., GARCIA, S., HERRERA, F. & CHAWLA, N. V. 2018. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research,* 61**,** 863-905.
FOREMAN, R. D. 2003. A logistic analysis of bankruptcy within the US local telecommunications industry. *Journal of Economics and Business,* 55**,** 135-166.
FREUND, Y. & SCHAPIRE, R. E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences,* 55**,** 119-139.
FRIEDMAN, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics***,** 1189-1232.
GARCïA-LAENCINA, P. J., SANCHO-GõMEZ, J.-L. & FIGUEIRAS-VIDAL, A. R. 2010. Pattern classification with missing data: a review. *Neural Computing and Applications,* 19**,** 263-282.
GARCïA, S., RAMïREZ-GALLEGO, S., LUENGO, J., BENïTEZ, J. M. & HERRERA, F. 2016. Big data preprocessing: methods and prospects. *Big Data Analytics,* 1**,** 1-22.
GARCïA, V., MARQUëS, A. I. & SãNCHEZ, J. S. 2019. Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction. *Information Fusion,* 47**,** 88-101.
GAVUROVA, B., PACKOVA, M., MISANKOVA, M. & SMRCKA, L. 2017. Predictive potential and risks of selected bankruptcy prediction models in the Slovak business environment. *Journal of Business Economics and Management,* 18**,** 1156-1173.
GHORBANI, M., ACCIARO, M., TRANSCHEL, S. & CARIOU, P. 2022. Strategic alliances in container shipping: A review of the literature and future research agenda. *Maritime Economics & Logistics,* 24**,** 439-465.
GIRIŪNIENE, G., GIRIŪNAS, L., MORKUNAS, M. & BRUCAITE, L. 2019. A comparison on leading methodologies for bankruptcy prediction: The case of the construction sector in Lithuania. *Economies,* 7**,** 82.
GORDINI, N. 2014. A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Expert systems with applications,* 41**,** 6433-6445.
GRAMMENOS, C. T. 2013. Revisiting credit risk, analysis and policy in bank shipping finance. *The handbook of maritime economics and business.* Informa Law from Routledge.
GRAMMENOS, C. T., ALIZADEH, A. H. & PAPAPOSTOLOU, N. C. 2007. Factors affecting the dynamics of yield premia on shipping seasoned high yield bonds. *Transportation Research Part E: Logistics and Transportation Review,* 43**,** 549-564.
GRAMMENOS, C. T., NOMIKOS, N. K. & PAPAPOSTOLOU, N. C. 2008. Estimating the probability of default for shipping high yield bond issues. *Transportation Research Part E: Logistics and Transportation Review,* 44**,** 1123-1138.
GRAMMENOS, C. T. & PAPAPOSTOLOU, N. C. 2012. US shipping initial public offerings: Do prospectus and market information matter? *Transportation Research Part E: Logistics and Transportation Review,* 48**,** 276-295.
GREGOVA, E., VALASKOVA, K., ADAMKO, P., TUMPACH, M. & JAROS, J. 2020. Predicting financial distress of slovak enterprises: Comparison of selected traditional and learning algorithms methods. *Sustainability,* 12**,** 3954.
GRUNERT, J., NORDEN, L. & WEBER, M. 2005. The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance,* 29**,** 509-531.
GUERRERO, V. M. & JOHNSON, R. A. 1982. Use of the Box-Cox transformation with binary response models. *Biometrika,* 69**,** 309-314.
HAIDER, J., OU, Z. & PETTIT, S. 2019. Predicting corporate failure for listed shipping companies. *Maritime Economics & Logistics,* 21**,** 415-438.
HAIR, J. F. 2009. Multivariate data analysis.
HAIXIANG, G., YIJING, L., SHANG, J., MINGYUN, G., YUANYUE, H. & BING, G. 2017. Learning from class-imbalanced data: Review of methods and applications. *Expert systems with applications,* 73**,** 220-239.
HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. & FRIEDMAN, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction,* Springer.
HE, H. & GARCIA, E. A. 2009. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering,* 21**,** 1263-1284.
HERON, J. 1996. *Co-operative inquiry: Research into the human condition,* Sage.
HILLEGEIST, S. A., KEATING, E. K., CRAM, D. P. & LUNDSTEDT, K. G. 2004. Assessing the probability of bankruptcy. *Review of accounting studies,* 9**,** 5-34.
HIRA, Z. M. & GILLIES, D. F. 2015. A review of feature selection and feature extraction methods applied on microarray data. *Advances in bioinformatics,* 2015.
HOCHREITER, S. & SCHMIDHUBER, J. 1997. Long short-term memory. *Neural computation,* 9**,** 1735-1780.
HWANG, J., YOON, J., AHN, Y. & KIM, J. 2017. Policy cases of Restructuring in Korea and its implication Korea Maritime Institute.
HWANG, S., KO, B., CHOI, G., AHN, Y.-G. & KIM, B.-J. 2020. A Study on the Prediction of Dry Bulk Freight Rate based on Long Term Short Term Memory. Korea Maritime Institute.
JAFFEE, D. & STIGLITZ, J. 1990. Credit rationing. *Handbook of monetary economics,* 2**,** 837-888.
JANG, Y., JEONG, I. & CHO, Y. K. 2021. Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors. *Engineering, Construction and Architectural Management.*
JEON, H., YOON, H. & CHOI, Y. 2017. A study on the limitation and development direction of shipping finance in Korea. Korea Maritime Institute.
JEONG, K. 2021a. Current status and comparison of container ship charter market. Korea Ocean Business Corporation.
JEONG, K. 2021b. Current status and comparison of container shipping rate index. Korea Ocean Business Corporation.

JONES, S. 2017. Corporate bankruptcy prediction: a high dimensional analysis. *Review of Accounting Studies,* 22**,** 1366-1422.

JONES, S., JOHNSTONE, D. & WILSON, R. 2017. Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting,* 44**,** 3-34.

JONES, S. & WANG, T. 2019. Predicting private company failure: A multi-class analysis. *Journal of International Financial Markets, Institutions and Money,* 61**,** 161-188.

KALGORA, B. & CHRISTIAN, T. M. 2016. The financial and economic crisis, its impacts on the shipping industry, lessons to learn: the container-ships market analysis. *Open Journal of Social Sciences,* 4**,** 38-44.

KAMAL, M. R., CHOWDHURY, M. A. F. & HOSAIN, M. M. 2021. Stock market reactions of maritime shipping industry in the time of COVID-19 pandemic crisis: an empirical investigation. *Maritime Policy & Management***,** 1-16.

KANG, C.-M., WANG, M.-C. & LIN, L. 2022. Financial Distress Prediction of Cooperative Financial Institutions—Evidence for Taiwan Credit Unions. *International Journal of Financial Studies,* 10**,** 30.

KARELS, G. V. & PRAKASH, A. J. 1987. Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance & Accounting,* 14**,** 573-593.

KAVUSSANOS, M. G. & TSOUKNIDIS, D. A. 2016. Default risk drivers in shipping bank loans. *Transportation Research Part E: Logistics and Transportation Review,* 94**,** 71-94.

KAVUSSANOS, M. G., TSOUKNIDIS, D. A. & VISVIKIS, I. D. 2021. *Freight derivatives and risk management in shipping*, Routledge.

KEASEY, K. & WATSON, R. 2019. Financial distress prediction models: a review of their usefulness 1. *Risk Management***,** 35-48.

KIM, H., CHO, H. & RYU, D. 2020. Corporate default predictions using machine learning: Literature review. *Sustainability,* 12**,** 6325.

KIM, H., CHO, H. & RYU, D. 2022. Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data. *Computational Economics,* 59**,** 1231-1249.

KIM, M.-J. & KANG, D.-K. 2010. Ensemble with neural networks for bankruptcy prediction. *Expert systems with applications,* 37**,** 3373-3379.

KIM, M.-J., KANG, D.-K. & KIM, H. B. 2015. Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications,* 42**,** 1074-1082.

KIM, S. & KANG, M. 2019. Financial series prediction using Attention LSTM. *arXiv preprint arXiv:1902.10877*.

KIN, T. Y., AHMAD AIZAM, A. H., ABU HASAN, S., ARIFFIN, A. F. & MAHAT, N. 2021. Bankruptcy prediction model with risk factors using fuzzy logic approach. *Journal of Computing Research and Innovation (JCRINN),* 6**,** 102-110.

KMI 2021. KMI Annual report 2021. Korea Maritime Institute.

KMI 2022. 2022 Shipping Statistics Handbook. Korea Maritime Institute: Korea Maritime Institute.

KNAPIČ, S., MALHI, A., SALUJA, R. & FRæMLING, K. 2021. Explainable artificial intelligence for human decision support system in the medical domain. *Machine Learning and Knowledge Extraction,* 3**,** 740-770.

KOBC 2019. KOBC annual report. Korea Ocean Business Corporation.

KOBC 2021. KOBC Annual report 2021. Korea Ocean Business Corporation.

KOBC 2023. KOBC Annual report 2023. Korea Ocean Business Corporation: Korea Ocean Business Corporation.

KOREAN MINISTRY OF MARITIME AFFAIRS & FISHERIES. (2023). Shipping Statistics 2023. Korean Ministry of Maritime Affairs & Fisheries. Available at: http://www.momaf.go.kr/statistics2020 (accessed on 23th October, 2023).

KOREAN MARITIME INSTITUE. (2023). Maritime Economic Review. Korean Maritime Institute. Available at: https://www.kmi.re.kr/ (accessed on 21th November, 2023)

KOREAN OCEAN BUSINESS OPERATION. (2023). Operational Review 2021. Korean Ocean Business Operation. Available at: https://www.kobc.or.kr/ebz/shippinginfo/main.do (accessed on 12th October, 2023)

KOREAN FINANCIAL SUPERVISORY SERVICE. (2023). Data Analysis, Retrieval, and Transfer System (DART). Available at http://dart.fss.or.kr (accessed on 9th September, 2023)

KOREAN SHIPOWNERS ASSOCIATION. (2023). Shipping Statistics Report. Korean Shipowners Association. Available at: https://oneksa.kr/boards/statistics (accessed on 27th September, 2023).

KOROL, T. 2017. Evaluation of the factors influencing business bankruptcy risk in Polad. *e-Finanse: Financial Internet Quarterly,* 13**,** 22-35.

KOTCHARIN, S. & MANEENOP, S. 2018. The Giant's Pull: How Macroeconomic Conditions in China Explain Leverage Decisions in Thailand's Shipping SMEs. *The Asian Journal of Shipping and Logistics,* 34**,** 337-344.

KOU, G., XU, Y., PENG, Y., SHEN, F., CHEN, Y., CHANG, K. & KOU, S. 2021. Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems,* 140**,** 113429.

KOVACOVA, M., KLIESTIK, T., VALASKOVA, K., DURANA, P. & JUHASZOVA, Z. 2019. Systematic review of variables applied in bankruptcy prediction models of Visegrad group countries. *Oeconomia Copernicana,* 10**,** 743-772.

KOVALAINEN, A. & ERIKSSON, P. 2015. Qualitative methods in business research: A practical guide to social research. *Qualitative Methods in Business Research***,** 1-376.

KPMG 2021. Current status and future prospects of the Korean shipping market. Samjung KPMG research institute.

KPMG. (2023). Annual Report on the Shipping Industry. KPMG. available at: http://www.kpmg.com/shippingreport (accessed on 10th October, 2023).

KRISTõF, T. & VIRãG, M. 2020. A comprehensive review of corporate bankruptcy prediction in Hungary. *Journal of Risk and Financial Management,* 13**,** 35.

KSA 2021. 2021 Maritime Statistics. Korea Shipowner's Association.

KþCHER, A., MAYR, S., MITTER, C., DULLER, C. & FELDBAUER-DURSTMþLLER, B. 2020. Firm age dynamics and causes of corporate bankruptcy: age dependent explanations for business failure. *Review of Managerial Science,* 14**,** 633-661.

KUHN, M. & JOHNSON, K. 2013. A short tour of the predictive modeling process. *Applied Predictive Modeling.* Springer.

KUMAR, P. R. & RAVI, V. 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European journal of operational research,* 180**,** 1-28.

KWAK, W., SHI, Y. & KOU, G. 2012. Bankruptcy prediction for Korean firms after the 1997 financial crisis: using a multiple criteria linear programming data mining approach. *Review of Quantitative Finance and Accounting,* 38**,** 441-453.

KWON, J., PARK, S., KIM, T. & KIM, H. 2023. A novel business cycle indicator of the Korean shipping industry. *The Asian Journal of Shipping and Logistics*.

LAHMIRI, S. & BEKIROS, S. 2019. Can machine learning approaches predict corporate bankruptcy? Evidence from a qualitative experimental design. *Quantitative Finance,* 19**,** 1569-1577.

LE, T. 2022. A comprehensive survey of imbalanced learning methods for bankruptcy prediction. *IET Communications,* 16**,** 433-441.

LEE, H. 2020. Current status and implications of sluggish shipping service exports. Institute for international trade.

LEE, K., AYYASAMY, M. V., JI, Y. & BALACHANDRAN, P. V. 2022. A Comparison of Explainable Artificial Intelligence Methods in the Phase Classification of Multi-Principal Element Alloys.

LEE, S.-Y. 2016. The Effects of Economic Conditions on Capital Structure: Evidence from Korean Shipping Firms. *Journal of Navigation and Port Research,* 40**,** 451-458.
LEE, S., CHOI, K. & YOO, D. 2020. Predicting the Insolvency of SMEs Using Technological Feasibility Assessment Information and Data Mining Techniques. *Sustainability,* 12**,** 9790.
LEE, S. & CHOI, W. S. 2013. A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications,* 40**,** 2941-2946.
LI, M.-Y. L. & MIU, P. 2010. A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. *Journal of Empirical Finance,* 17**,** 818-833.
LI, Y., STASINAKIS, C. & YEO, W. M. 2022. A hybrid XGBoost-MLP model for credit risk assessment on digital supply chain finance. *Forecasting,* 4**,** 184-207.
LIANG, D., LU, C.-C., TSAI, C.-F. & SHIH, G.-A. 2016. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research,* 252**,** 561-572.
LIN, W.-C., LIU, C.-F. & CHU, C.-W. Performance efficiency evaluation of the Taiwan's shipping industry: an application of data envelopment analysis. Proceedings of the Eastern Asia Society for Transportation Studies, 2005. Citeseer, 467-476.
LIN, W. C., LU, Y. H. & TSAI, C. F. 2019. Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems,* 36**,** e12335.
LINARDATOS, P., PAPASTEFANOPOULOS, V. & KOTSIANTIS, S. 2020. Explainable ai: A review of machine learning interpretability methods. *Entropy,* 23**,** 18.
LIU, J., SONG, Q., QI, Y., RAHMAN, S. & SRIBOONCHITTA, S. 2020. Measurement of systemic risk in global financial markets and its application in forecasting trading decisions. *Sustainability,* 12**,** 4000.
LOHMANN, C., MøLLENHOFF, S. & OHLIGER, T. 2023. Nonlinear relationships in bankruptcy prediction and their effect on the profitability of bankruptcy prediction models. *Journal of Business Economics,* 93**,** 1661-1690.
LOMBARDO, G., PELLEGRINO, M., ADOSOGLOU, G., CAGNONI, S., PARDALOS, P. M. & POGGI, A. 2022. Machine Learning for Bankruptcy Prediction in the American Stock Market: Dataset and Benchmarks. *Future Internet,* 14**,** 244.
LOZINSKAIA, A., MERIKAS, A., MERIKA, A. & PENIKAS, H. 2017. Determinants of the probability of default: the case of the internationally listed shipping corporations. *Maritime Policy & Management,* 44**,** 837-858.
LUKãČ, J., OLEXOVã, C. & KUDLOVã, Z. 2022. Factors predicting companies' crisis in the engineering industry from the point of view of financial analysis. *PLoS One,* 17**,** e0264016.
LUKASON, O. & ANDRESSON, A. 2019. Tax arrears versus financial ratios in bankruptcy prediction. *Journal of Risk and Financial Management,* 12**,** 187.
LUKASON, O. & HOFFMAN, R. C. 2014. Firm bankruptcy probability and causes: An integrated study. *International Journal of Business and Management,* 9**,** 80-91.
LUNDBERG, S. M. & LEE, S.-I. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems,* 30.
LUO, Z., HSU, P. & XU, N. 2020. SME Default Prediction Framework with the Effective Use of External Public Credit Data. *Sustainability,* 12**,** 7575.
LYANDRES, E. & ZHDANOV, A. 2013. Investment opportunities and bankruptcy prediction. *Journal of Financial Markets,* 16**,** 439-476.
MAI, F., TIAN, S., LEE, C. & MA, L. 2019. Deep learning models for bankruptcy prediction using textual disclosures. *European journal of operational research,* 274**,** 743-758.
MAKROMINAS, M. 2018. Estimating the implied risk premium of US-listed shipping firms. *Maritime Policy & Management,* 45**,** 1021-1041.
MARTIN, A., GAYATHRI, V., SARANYA, G., GAYATHRI, P. & VENKATESAN, P. 2011. Hybrid Model for Bankruptcy Reduction Using Genetic Algorithm. *Fuzzy C-Means and Mars, Research Scholar, Department of Banking Technology, Pondicherry University, Pondicherry*.
MARTIN, D. 1977. Early warning of bank failure: A logit regression approach. *Journal of banking & finance,* 1**,** 249-276.
MATIN, S., FARAHZADI, L., MAKAREMI, S., CHELGANI, S. C. & SATTARI, G. 2018. Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest. *Applied Soft Computing,* 70**,** 980-987.
MAYR, S., MITTER, C., KþCHER, A. & DULLER, C. 2021. Entrepreneur characteristics and differences in reasons for business failure: evidence from bankrupt Austrian SMEs. *Journal of Small Business & Entrepreneurship,* 33**,** 539-558.
MCNAMARA, R. P., COCKS, N. J. & HAMILTON, D. F. 1988. Predicting private company failure. *Accounting & Finance,* 28**,** 53-64.
MI, J.-X., LI, A.-D. & ZHOU, L.-F. 2020. Review study of interpretation methods for future interpretable machine learning. *IEEE Access,* 8**,** 191969-191985.
MICHAIL, N. A. 2020. World economic growth and seaborne trade volume: quantifying the relationship. *Transportation Research Interdisciplinary Perspectives,* 4**,** 100108.
MIHALOVIC, M. 2016. Performance comparison of multiple discriminant analysis and logit models in bankruptcy prediction. *Economics & Sociology,* 9**,** 101.
MIN, J. H. & LEE, Y.-C. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications,* 28**,** 603-614.
MITCHELL, J. & VAN ROY, P. 2007. Failure prediction models: performance, disagreements, and internal rating systems. *National Bank of Belgium Working Paper*.
MITCHELL, T. M. & MITCHELL, T. M. 1997. *Machine learning,* McGraw-hill New York.
MITROUSSI, K., ABOUARGHOUB, W., HAIDER, J. J., PETTIT, S. J. & TIGKA, N. 2016. Performance drivers of shipping loans: An empirical investigation. *International Journal of Production Economics,* 171**,** 438-452.
MOHAMMED, M. A., AL-KHATEEB, B., RASHID, A. N., IBRAHIM, D. A., ABD GHANI, M. K. & MOSTAFA, S. A. 2018. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers & Electrical Engineering,* 70**,** 871-882.
MOK;, I. & RYOO, D. 2022. The Economic Cycle and Contributing Factors to the Operating Profit Ratio of KoreanLiner Shipping. *Journal of Navigation and Port Research,* 46**,** 375-384.
MOLNAR, C. 2020. *Interpretable machine learning,* Lulu. com.
MOSSMANC, C. E., BELL, G. G., SWARTZ, L. M. & TURTLE, H. 1998. An empirical comparison of bankruptcy models. *Financial Review,* 33**,** 35-54.
MOYER, R. C. (1977). Forecasting financial failure: a re-examination. *Financial management*, 6(1), 11.
MRAMOR, D. & VALENTINCIC, A. 2003. Forecasting the liquidity of very small private companies. *Journal of Business Venturing,* 18**,** 745-771.

MULLER, G., STEYN-BRUWER, B. & HAMMAN, W. 2009. Predicting financial distress of companies listed on the JSE-A comparison of techniques. *South African Journal of Business Management,* 40**,** 21-32.

MURPHY, K. P. 2022. *Probabilistic Machine Learning: An introduction*, MIT Press.

NAM, H. J. & AN, Y. H. 2017. Default risk and firm value of shipping & logistics firms in Korea. *The Asian Journal of Shipping and Logistics,* 33**,** 61-65.

NARVEKAR, A. & GUHA, D. 2021. Bankruptcy prediction using machine learning and an application to the case of the COVID-19 recession. *Data Science in Finance and Economics,* 1**,** 180-195.

NGUYEN, T.-T., TRAN, D. T. M., DUC, T. T. H. & THAI, V. V. 2022. Managing disruptions in the maritime industry–a systematic literature review. *Maritime Business Review*.

NOH, S.-H. 2023. Comparing the Performance of Corporate Bankruptcy Prediction Models Based on Imbalanced Financial Data. *Sustainability,* 15**,** 4794.

NORDEN, L. & WEBER, M. 2009. The co-movement of credit default swap, bond and stock markets: An empirical analysis. *European financial management,* 15**,** 529-562.

NOTTEBOOM, T., PALLIS, T. & RODRIGUE, J.-P. 2021. Disruptions and resilience in global container shipping and ports: the COVID-19 pandemic versus the 2008–2009 financial crisis. *Maritime Economics & Logistics,* 23**,** 179-210.

NOVAK, B. & SAJTER, D. 2007. Causes of bankruptcy in Europe and Croatia.

NYITRAI, T. & VIRãG, M. 2019. The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences,* 67**,** 34-42.

OBERMANN, L. & WAACK, S. 2015. Demonstrating non-inferiority of easy interpretable methods for insolvency prediction. *Expert Systems with Applications,* 42**,** 9117-9128.

OHLSON, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research***,** 109-131.

OLIVEIRA, M. D., FERREIRA, F. A., PëREZ-BUSTAMANTE ILANDER, G. O. & JALALI, M. S. 2017. Integrating cognitive mapping and MCDA for bankruptcy prediction in small-and medium-sized enterprises. *Journal of the Operational Research Society,* 68**,** 985-997.

OLSON, D. L., DELEN, D. & MENG, Y. 2012. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems,* 52**,** 464-473.

OOGHE, H. & DE PRIJCKER, S. 2008. Failure processes and causes of company bankruptcy: a typology. *Management Decision,* 46**,** 223-242.

PADILLA-OSPINA, A. M., MEDINA-VãSQUEZ, J. E. & OSPINA-HOLGUïN, J. H. 2021. Financial determinants of innovation in SMEs: A machine learning approach. *Journal of Small Business Strategy,* 31**,** 117-131.

PAI, N. & ILANGO, V. Lstm neural network model with feature selection for financial time series prediction. 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2020. IEEE, 672-677.

PAPANA, A. & SPYRIDOU, A. 2020. Bankruptcy prediction: The case of the Greek market. *Forecasting,* 2**,** 505-525.

PARK, M. S., SON, H., HYUN, C. & HWANG, H. J. 2021a. Explainability of Machine Learning Models for Bankruptcy Prediction. *IEEE Access,* 9**,** 124887-124899.

PARK, S., KIM, H. & CHO, S. 2022. Crisis is repeated. Shipping companies need to prepare for recession. Korea Maritime Institute.

PARK, S., KIM, H., KWON, J. & KIM, T. 2021b. Empirics of Korean Shipping Companies' Default Predictions. *Risks,* 9**,** 159.

PARK, S. B., KIM, S.-K. & LEE, S. 2021c. Earnings management of insolvent firms and the prediction of corporate defaults via discretionary accruals. *International Journal of Financial Studies,* 9**,** 17.

PEDERZOLI, C. & TORRICELLI, C. 2010. A parsimonious default prediction model for Italian SMEs.

PEEL, M. J. & PEEL, D. A. 1987. Some further empirical evidence on predicting private company failure. *Accounting and Business Research,* 18**,** 57-66.

PERBOLI, G. & ARABNEZHAD, E. 2021. A Machine Learning-based DSS for mid and long-term company crisis prediction. *Expert Systems with Applications,* 174**,** 114758.

PëREZ-PONS, M. E., PARRA-DOMINGUEZ, J., HERNãNDEZ, G., HERRERA-VIEDMA, E. & CORCHADO, J. M. 2022. Evaluation metrics and dimensional reduction for binary classification algorithms: a case study on bankruptcy prediction. *The Knowledge Engineering Review,* 37.

PHILOSOPHOV, L. V., BATTEN, J. A. & PHILOSOPHOV, V. L. 2008. Predicting the event and time horizon of bankruptcy using financial ratios and the maturity schedule of long-term debt. *Mathematics and Financial Economics,* 1**,** 181-212.

PINDADO, J., RODRIGUES, L. & DE LA TORRE, C. 2008. Estimating financial distress likelihood. *Journal of Business Research,* 61**,** 995-1003.

PLATT, H. D., PLATT, M. B. & PEDERSEN, J. G. 1994. Bankruptcy discrimination with real variables. *Journal of Business Finance & Accounting,* 21**,** 491-510.

PŁAWIAK, P., ABDAR, M. & ACHARYA, U. R. 2019. Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Applied Soft Computing,* 84**,** 105740.

POMPE, P. P. & BILDERBEEK, J. 2005. The prediction of bankruptcy of small-and medium-sized industrial firms. *Journal of Business venturing,* 20**,** 847-868.

PRUSAK, B. 2018. Review of research into enterprise bankruptcy prediction in selected central and eastern European countries. *International Journal of Financial Studies,* 6**,** 60.

PSILLAKI, M., TSOLAS, I. E. & MARGARITIS, D. 2010. Evaluation of credit risk based on firm performance. *European journal of operational research,* 201**,** 873-881.

QU, Y., QUAN, P., LEI, M. & SHI, Y. 2019. Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science,* 162**,** 895-899.

REBETAK, F. & BARTOSOVA, V. Financial Distress and Bankruptcy Prediction in Conditions of Slovak Republic. SHS Web of Conferences, 2021. EDP Sciences.

REISZ, A. S. & PERLICH, C. 2007. A market-based framework for bankruptcy prediction. *Journal of financial stability,* 3**,** 85-131.

REYNOLDS, J. K. & FRANCIS, J. R. 2000. Does size matter? The influence of large clients on office-level auditor reporting decisions. *Journal of accounting and economics,* 30**,** 375-400.

RIBEIRO, M. T., SINGH, S. & GUESTRIN, C. 2016. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.

ROY, P. K. & SHAW, K. 2021. A multicriteria credit scoring model for SMEs using hybrid BWM and TOPSIS. *Financial Innovation,* 7**,** 1-27.

RUNDO, F. 2019. Deep LSTM with reinforcement learning layer for financial trend prediction in FX high frequency trading systems. *Applied Sciences,* 9**,** 4460.

RUSTAM, Z., NADHIFA, F. & ACAR, M. Comparison of SVM and FSVM for predicting bank failures using chi-square feature selection. Journal of Physics: Conference Series, 2018. IOP Publishing, 012115.

SANZ, H., VALIM, C., VEGAS, E., OLLER, J. M. & REVERTER, F. 2018. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC bioinformatics,* 19**,** 1-18.
SAUNDERS, M., LEWIS, P. & THORNHILL, A. 2009. *Research methods for business students*, Pearson education.
SAUNDERS, M., LEWIS, P. & THORNHILL, A. 2019. Research Methods for Business Students.
SCHAPIRE, R. E. 2013. Explaining adaboost. *Empirical inference.* Springer.
SCHWARCZ, S. L. 1999. Sovereign debt restructuring: a bankruptcy reorganization approach. *Cornell L. Rev.,* 85**,** 956.
SFAKIANAKIS, E. 2012. Bankruptcy prediction model for listed companies in Greece. *Investment Management and Financial Innovations,* 18**,** 166-80.
SHETTY, S., MUSA, M. & BRëDART, X. 2022. Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management,* 15**,** 35.
SHI, Y. & LI, X. 2019. An overview of bankruptcy prediction models for corporate firms: A systematic literature review. *Intangible Capital,* 15**,** 114-127.
SHIN, S.-H., LEE, P. T.-W. & LEE, S.-W. 2019. Lessons from bankruptcy of Hanjin Shipping Company in chartering. *Maritime Policy & Management,* 46**,** 136-155.
SHOME, S. & VERMA, S. 2020. Financial distress in Indian aviation industry: Investigation using bankruptcy prediction models. *Eurasian Journal of Business and Economics,* 13**,** 91-109.
SHRIVASTAV, S. K. & RAMUDU, P. J. 2020. Bankruptcy Prediction and Stress Quantification Using Support Vector Machine: Evidence from Indian Banks. *Risks,* 8**,** 52.
SHUMWAY, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The journal of business,* 74**,** 101-124.
SIRIMANNE, S. N., HOFFMAN, J., JUAN, W., ASARIOTIS, R., ASSAF, M., AYALA, G., BENAMARA, H., CHANTREL, D., HOFFMANN, J. & PREMTI, A. Review of maritime transport 2019. United Nations Conference on Trade and Development, Geneva, Switzerland, 2019.
SISWOYO, B., HERMAN, N. & DEWI, D. Machine learning approach as an alternative tool to build a bankruptcy prediction model in banking industry. IOP Conference Series: Materials Science and Engineering, 2020. IOP Publishing, 022083.
SLOTEMAKER, R. 2008. Prediction of Corporate Bankruptcy.
SMITI, S. & SOUI, M. 2020. Bankruptcy prediction using deep learning approach based on borderline SMOTE. *Information Systems Frontiers,* 22**,** 1067-1083.
SOLORIO-FERNãNDEZ, S., CARRASCO-OCHOA, J. A. & MARTïNEZ-TRINIDAD, J. F. 2020. A review of unsupervised feature selection methods. *Artificial Intelligence Review,* 53**,** 907-948.
SON, H., HYUN, C., PHAN, D. & HWANG, H. J. 2019. Data analytic approach for bankruptcy prediction. *Expert Systems with Applications,* 138**,** 112816.
SONG, D.-W., SEO, Y.-J. & KWAK, D.-W. 2019. Learning from Hanjin Shipping's failure: A holistic interpretation on its causes and reasons. *Transport Policy,* 82**,** 77-87.
SOUSA, A., BRAGA, A. & CUNHA, J. 2022. Impact of macroeconomic indicators on bankruptcy prediction models: Case of the Portuguese construction sector. *Quantitative Finance and Economics,* 6**,** 405-432.
STEF, N. & JABEUR, S. B. 2018. The bankruptcy prediction power of new entrants. *International Journal of the Economics of Business,* 25**,** 421-440.
STOPFORD, M. 2013. Shipping market cycles. *The handbook of maritime economics and business.* Informa Law from Routledge.
SUN, J., LI, H., FUJITA, H., FU, B. & AI, W. 2020. Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. *Information Fusion,* 54**,** 128-144.
SUN, L. 2007. A re-evaluation of auditors' opinions versus statistical models in bankruptcy prediction. *Review of Quantitative Finance and Accounting,* 28**,** 55-78.
TANG, O. & SUN, P.-W. 2018. Anti-competition of ocean shipping alliances: a legal perspective. *Maritime Business Review,* 3**,** 4-19.
TINOCO, M. H. & WILSON, N. 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International review of financial analysis,* 30**,** 394-419.
TIROLE, J. 2010. *The theory of corporate finance*, Princeton university press.
TOBBACK, E., BELLOTTI, T., MOEYERSOMS, J., STANKOVA, M. & MARTENS, D. 2017. Bankruptcy prediction for SMEs using relational data. *Decision Support Systems,* 102**,** 69-81.
TRANFIELD, D., DENYER, D. & SMART, P. 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management,* 14**,** 207-222.
TSAI, C.-F. 2009. Feature selection in bankruptcy prediction. *Knowledge-Based Systems,* 22**,** 120-127.
TSAI, C.-F., HSU, Y.-F. & YEN, D. C. 2014. A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing,* 24**,** 977-984.
TSAI, C.-F. & WU, J.-W. 2008. Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications,* 34**,** 2639-2649.
TSENG, F.-M. & HU, Y.-C. 2010. Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks. *Expert systems with applications,* 37**,** 1846-1853.
TYAGI, V. 2018. Corporate Insolvency Resolution Procedure under Indian Insolvency and Bankruptcy Code, 2016: A Comparative Perspective'. *The Chartered Accountant,* 66**,** 37-44.
VALENCIA, C., CABRALES, S., GARCIA, L., RAMIREZ, J. & CALDERONA, D. 2019. Generalized additive model with embedded variable selection for bankruptcy prediction: Prediction versus interpretation. *Cogent Economics & Finance,* 7**,** 1597956.
VEGANZONES, D. & SEVERIN, E. 2020. Corporate failure prediction models in the twenty-first century: a review. *European Business Review,* 33**,** 204-226.
VEGANZONES, D. & SëVERIN, E. 2018. An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems,* 112**,** 111-124.
VIRãG, M. & NYITRAI, T. 2014. Is there a trade-off between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model. *Acta Oeconomica,* 64**,** 419-440.
VISVANATHAN, G. 2021. The Information Content of Management Disclosures on Going Concern: An Industry Level Analysis of Bankruptcy Prediction and Investor Reaction. *Journal of Managerial Issues,* 33.
VOCHOZKA, M., VRBKA, J. & SULER, P. 2020. Bankruptcy or success? the effective prediction of a company's financial development using LSTM. *Sustainability,* 12**,** 7529.
VODA, A. D., DOBROTĂ, G., ŢǒRCĂ, D. M., DUMITRAȘCU, D. D. & DOBROTĂ, D. 2021. Corporate bankruptcy and insolvency prediction model. *Technological and Economic Development of Economy,* 27**,** 1039-1056.
VOLKOV, A., BENOIT, D. F. & VAN DEN POEL, D. 2017. Incorporating sequential information in bankruptcy prediction with predictors based on Markov for discrimination. *Decision Support Systems,* 98**,** 59-68.

WANG, G., MA, J. & YANG, S. 2014. An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications,* 41**,** 2353-2361.

WANG, G. W., YANG, Z., ZHANG, D., HUANG, A. & YANG, Z. 2017. Application of Bayesian networks in analysing tanker shipping bankruptcy risks. *Maritime Business Review*.

WANG, H. & LIU, X. 2021. Undersampling bankruptcy prediction: Taiwan bankruptcy data. *Plos one,* 16**,** e0254030.

WANG, H., TANG, J., WU, M., WANG, X. & ZHANG, T. 2022. Application of machine learning missing data imputation techniques in clinical decision making: taking the discharge assessment of patients with spontaneous supratentorial intracerebral hemorrhage as an example. *BMC Medical Informatics and Decision Making,* 22**,** 1-14.

WANG, T., ZHAO, S., ZHU, G. & ZHENG, H. 2021. A machine learning-based early warning system for systemic banking crises. *Applied Economics,* 53**,** 2974-2992.

XIA, Y., LIU, C., LI, Y. & LIU, N. 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert systems with applications,* 78**,** 225-241.

XUE, Y. & HAUSKRECHT, M. Active learning of classification models with likert-scale feedback. Proceedings of the 2017 SIAM International Conference on Data Mining, 2017. SIAM, 28-35.

YANG, D., ZHANG, W., WU, X., ABLANEDO-ROSAS, J. H., YANG, L. & YU, W. 2021. A novel multi-stage ensemble model with fuzzy clustering and optimized classifier composition for corporate bankruptcy prediction. *Journal of Intelligent & Fuzzy Systems,* 40**,** 4169-4185.

YEO, H. 2016. Solvency and liquidity in shipping companies. *The Asian Journal of Shipping and Logistics,* 32**,** 235-241.

YEO, S.-J., KIM, J. & LEE, W.-J. 2022. Potential economic and environmental advantages of liquid petroleum gas as a marine fuel through analysis of registered ships in South Korea. *Journal of Cleaner Production,* 330**,** 129955.

YOON, S., KIM, C. Y. & SEO, Y.-J. 2023. The motives for shipping asset securitisation: sale-leaseback transactions in the shipping industry. *Maritime Policy & Management,* 50**,** 287-300.

YU, B., LI, Y. & SHI, Y. 2020. Cooperation Mechanism of the Shipping SMEs Based on Field Theory. *Journal of Coastal Research,* 112**,** 472-476.

YU, Q., MICHE, Y., SëVERIN, E. & LENDASSE, A. 2014. Bankruptcy prediction using extreme learning machine and financial expertise. *Neurocomputing,* 128**,** 296-302.

YUEN, S. & KO, A. 2018. Corporate Governance in Shipping Logistics Industry: A Practitioner's Perspectives. *Journal of Business & Economic Policy,* 5.

ZEBARI, D. A., HARON, H., ZEEBAREE, D. Q. & ZAIN, A. M. A simultaneous approach for compression and encryption techniques using deoxyribonucleic acid. 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 2019. IEEE, 1-6.

ZEBARI, R., ABDULAZEEZ, A., ZEEBAREE, D., ZEBARI, D. & SAEED, J. 2020. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends,* 1**,** 56-70.

ZELENKOV, Y., FEDOROVA, E. & CHEKRIZOV, D. 2017. Two-step classification method based on genetic algorithm for bankruptcy forecasting. *Expert Systems with Applications,* 88**,** 393-401.

ZELENKOV, Y. & VOLODARSKIY, N. 2021. Bankruptcy prediction on the base of the unbalanced data using multi-objective selection of classifiers. *Expert Systems with Applications,* 185**,** 115559.

ZHANG, Y., LIU, R., HEIDARI, A. A., WANG, X., CHEN, Y., WANG, M. & CHEN, H. 2021. Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis. *Neurocomputing,* 430**,** 185-212.

ZHOU, L. 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems,* 41**,** 16-25.

ZHOU, L. & LAI, K. K. 2017. AdaBoost models for corporate bankruptcy prediction with missing data. *Computational Economics,* 50**,** 69-94.

ZHOU, Y., LI, X. & YUEN, K. F. 2023. Sustainable shipping: a critical review for a unified framework and future research agenda. *Marine Policy,* 148**,** 105478.

ZIĘBA, M., TOMCZAK, S. K. & TOMCZAK, J. M. 2016. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert systems with applications,* 58**,** 93-101.

ZMIJEWSKI, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research***,** 59-82.

ZORIČăK, M., GNIP, P., DROTăR, P. & GAZDA, V. 2020. Bankruptcy prediction for small-and medium-sized companies using severely imbalanced datasets. *Economic Modelling,* 84**,** 165-176.

APPENDIX 1. Interview Survey in English

☑ The results of this survey are protected as confidential, and all responses to the questionnaire, as well as personal matters, are handled with strict confidentiality and anonymity. They will never be used for purposes other than statistical analysis.

Questionnaire for Expert Survey on Predicting Bankruptcy of Shipping Companies

Hello,

Thank you very much for taking your precious time out of your busy schedule. My name is Minsu Kwon, and I am currently a Ph.D. student at the University of Plymouth in the UK.

The objective of my research topic is to explore the determinants for predicting and responding to the bankruptcy risk of shipping companies in advance. The purpose of this study is to contribute to enabling business managers or policy makers to take proactive measures by recognizing the magnitude and intensity of economic downturns in the shipping industry.

In relation to this, the survey is conducted solely for research purposes, and thus your responses will be treated confidentially and not disclosed to third parties. The survey is expected to take about 15 minutes. We would appreciate it if you could make your best estimate in your answers, even if you are not certain about the answers to the questions. If you would like to know the results of the survey, please send me an email, and I will send you a summary after the research is completed. Thank you for your valuable time. If you have any questions or comments about this study, please feel free to contact me at any time.

University of Plymouth
Ph.D. in International Logistics, Supply Chain, and Shipping Management
Minsu Kwon

Research Director: Minsu Kwon Tel: 7444-238084 e-mail: minsu.kwon@plymouth.ac.uk

Reference Material for Expert Survey

□ Research Background and Purpose

- ◆ Numerous shipping companies have gone bankrupt in crises such as the financial crisis and COVID-19, due to a sharp decrease in cargo demand and high debt burden.
  - The shipping industry is highly volatile and affected by global economic aspects.
  - Shipping companies, especially SMEs, face a high risk of bankruptcy due to their lack of ability to predict market volatility and manage financial risks.
- ◆ The purpose of this study is to explore the risk factors of shipping companies to proactively assess and manage potential bankruptcy risks.
  - This allows managers and policy makers to pre-emptively understand and react to the scale and intensity of economic downturns in the shipping industry, enabling them to respond to bankruptcy risks in advance.
  - The identified risk factors are expected to contribute to policy aspects in building a bankruptcy risk system for shipping companies.

Section 1. Corporate bankruptcy risk factors

Please indicate how each of variables are important to evaluate corporate bankruptcy

(1=not at all important, 2=Slightly important, 3=important, 4= fairly important, 5= very important)

How important are the following factors to evaluate corporate bankruptcy?			Importance of Factors				
			1	2	3	4	5
Leverage							
A1	Return on total assets	$\frac{EBITDA}{Total\ Assets}$					
A2	Retained Earnings to Total Assets	$\frac{Retained\ Earnings}{Total\ Assets}$					
A3	Debt ratio	$\frac{Total\ Debt}{Total\ Assets}$					
A4	Liabilities to Total Assets	$\frac{Total\ Liabilities}{Total\ Assets}$					
A5	Equity to Assets	$\frac{Shareholder's\ Equity}{Total\ Assets}$					
A6	Debt to Equity	$\frac{Total\ Debt}{Shareholder's\ Equity}$					
A7	Liabilities to Equity	$\frac{Total\ Liabilities}{Shareholder's\ Equity}$					
A8	Gearing ratio	$\frac{Shareholder's\ Equity}{Total\ Debt}$					
A9	Equity to liabilities	$\frac{Shareholder's\ Equity}{Total\ Liabilities}$					
A10	Liabilities	$Total\ Liabilities$					
A11	Current Liabilities to Equity	$\frac{Current\ Liabilities}{Equity}$					
A12	EBITDA to Liabilities	$\frac{EBITDA}{Total\ Liabilities}$					
A13	EBITDA/EV	$\frac{EBITDA}{Equity}$					
A14	EBITDA	$Net\ Income + Interest + Taxes + Depreciation + Amortization$					
A15	Long-term Liabilities to Assets	$\frac{liabilities - current\ liabilities}{Total\ Assets}$					
A16	Net Income to Liabilities	$\frac{Net\ Income}{Total\ Liabilities}$					
A17	Working Capital	$Current\ Assets - Current\ Liabilities$					
A18	Current Assets to Liabilities	$\frac{Current\ Assets}{Total\ Liabilities}$					
A19	EBITDA to Current Liabilities	$\frac{EBITDA}{Current\ Liabilities}$					
A20	Long-term Liabilities to Equity	$\frac{liabilities - current\ liabilities}{Shareholder's\ Equity}$					
Liquidity							
B1	Current Ratio	$\frac{Current\ Assets}{Current\ Liabilities}$					
B2	Working Capital to Assets	$\frac{Current\ Assets - Current\ Liabilitiesl}{Total\ Assets}$					
B3	Cash Assets Ratio	$\frac{Cash}{Total\ Assets}$					
B4	Cash Ratio	$\frac{Cash}{Current\ Liabilities}$					
B5	Current Assets to Assets	$\frac{Current\ Assets}{Total\ Assets}$					
B6	Current Liabilities to Assets	$\frac{Current\ Liabilities}{Total\ Assets}$					

How important are the following factors to evaluate corporate bankruptcy?			Importance of Factors				
			1	**2**	**3**	**4**	**5**
Liquidity							
B7	Quick Ratio	$\frac{Cash + \ accounts\ receivable}{Current\ liabilities}$					
B8	Cash to Debt Ratio	$\frac{Cash}{Total\ Debt}$					
B9	Shareholder's Equity	$Total\ Assets - Total\ Liabilities$					
B10	Current Liabilities Ratio	$\frac{Current\ Liabilities}{Total\ Liabilities}$					
B11	Current Assets	$Cash\ +\ Accounts\ Receivable + Inventory$					
B12	Cash Sales Ratio	$\frac{Cash}{Total\ Sales}$					
B13	Current Liabilities to Current Assets	$\frac{Current\ Liabilities}{Current\ Assets}$					
B14	Working Capital to Current Assets	$\frac{Working\ Captial}{Current\ Assets}$					
B15	Quick Assets to Total Assets	$\frac{Cash + \ accounts\ receivable}{Total\ Assets}$					
B16	Working Capital to Equity	$\frac{Current\ Assets - Current\ Liabilities}{Shareholder's\ Equity}$					
B17	Working Capital to Current Liabilities	$\frac{Current\ Assets - Current\ Liabilities}{Current\ Liabilities}$					
B18	Cash to Current Assets	$\frac{Cash}{Current\ Assets}$					
B19	Retained Earnings to Current Liabilities	$\frac{Retained\ Earnings}{Current\ Liabilities}$					
B20	ΔTotal Asset	$\frac{Total\ Assets_t - Total\ Assets_{t-1}}{Total\ Assets_{t-1}}$					
Profitability							
C1	Return on Assets	$\frac{Net\ Income}{Total\ Assets}$					
C2	Return on Equity	$\frac{Net\ Income}{Shareholder's Equity}$					
C3	Return on Sales	$\frac{Net\ Income}{Sales}$					
C4	EBITD to Sales	$\frac{EBITDA}{Sales}$					
C5	Sales to Current Assets	$\frac{Sales}{Current\ Assets}$					
C7	Sales	$Total\ Sales$					
C8	EBITDA to Interest Coverage	$\frac{EBITDA}{Interest\ expenses}$					
C9	Gross Profit to Assets	$\frac{Gross\ Profit}{Total\ Assets}$					
C10	Net interest margin	$\frac{Interest\ returns - Interest\ paid}{Average\ Assets}$					
C11	Sales to equity	$\frac{Total\ Sales}{Shareholder's\ Equity}$					
C12	Net Income	$Total\ Revenue\ -\ Total\ Expenses$					
C13	Operating Return on Assets	$\frac{Operating\ Income}{Total\ Assets}$					
C14	Return on investment	$\frac{Net\ Income}{Investment}$					

How important are the following factors to evaluate corporate bankruptcy?			Importance of Factors				
			1	**2**	**3**	**4**	**5**
Profitability							
C15	Sales to Liabilities	$\frac{Total\ Sales}{Total\ Liabilties}$					
C16	Gross Profit to Current Liabilities	$\frac{Gross\ Profit}{Current\ Liabilities}$					
C17	Gross Profit to Liabilities	$\frac{Gross\ Profit}{Total\ Liabilities}$					
C18	Gross Profit to Sales	$\frac{Gross\ Profit}{Total\ Sales}$					
C19	ΔSales	$\frac{Total\ sales_t - Total\ Sales_{t-1}}{Total\ Sales_{t-1}}$					
C20	Sales to Current Liabilities	$\frac{Total\ Sales}{Current\ Liabilities}$					
Efficiency							
D1	Asset turnover	$\frac{Total\ Sales}{Total\ Assets}$					
D2	Working Capital to Sales	$\frac{Working\ Captial}{Total\ Sales}$					
D3	Current Assets to Sales	$\frac{Current\ Assets}{Total\ Sales}$					
D4	Inventory to Assets	$\frac{Inventory}{Total\ Assets}$					
D5	Current Liabilities to Sales	$\frac{Current\ Liabilities}{Total\ Sales}$					
D6	Average Inventory Turnover	$\frac{Average\ Inventories}{Total\ Sales}$					
D7	Inventory turnover	$\frac{Average\ Inventories}{Total\ Sales}$					
D8	Operating Margin	$\frac{Operating\ Income}{Sales}$					
Non-financial variables							
E1	Size	*Proxied by Log (Total Assets)*					
E2	Age	*Age of firm*					
E3	Number of employees	*Number of full-time employees*					
E4	GDP	$\frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}$					
Shipping Index							
F1	Oil Price	*Current US$ price of Brent oil*					
F2	Type of Operator	*Type of operator (Container, Tanker, Bulk)*					
F3	Chartering Cost						
F4	Freight Rate						
F5	Time Charter Rate	*Time charter rate index*					
F6	Container Freight Index Rate	*Shanghai Containerized Freight Index*					
F7	Baltic Dry Index	*Baltic dry index per month*					
F8	IRONSTEEL	*Dow Jones U.S. Iron & Steel Index*					

Section 2. Interview Questions

1. What do you consider to be especially important factors for assessing the possibility of bankruptcy or the level of financial risk of domestic shipping companies, and why? Please specify.

☞ Risk Factors:

※ ex) Debt ratio, status of liquid assets, net profit margin, etc.

☞ Reason:

2. What do you consider to be important risk factors not listed above, and why? Please specify.

☞ Risk Factors:

※ ex) Charter hire proportion, interest rate fluctuations, etc.

☞ Reason:

3. Please provide your opinion on how risk factors vary according to the size of shipping companies.

☞ Risk Factors:

※ ex) The total debt, the status of capital reserves, etc.

4. Based on your experience, please describe your experience or the tasks you have performed in assessing the bankruptcy risk or financial status of shipping companies.

※ ex) Selecting partner companies, analysing trends in the shipping industry, etc.

5. Please describe your opinion on what role bankruptcy prediction analysis will play in the management activities of shipping companies.

※ex) early warning signal, assisting in corporate strategy planning, and improving financial stability, etc.

6. Please provide your opinion regarding the prediction and analysis of bankruptcy for shipping companies.

Section 3. Respondent Profile

Please check the option that applies to you.

1. Please select the industry in which you work.

1) Corporate	( )	2) University	( )	3) Research Association	( )

2. Please write your current position.

3. Please select the duration of your employment at your current affiliation.

1) Less than 3 years	( )	2) 3 ~ 5 years	( )	3) 5 ~ 10 years	( )	4) More than 10 years	( )

Thank you for responding to the consultation

APPENDIX 2. Interview Survey in Korean

☑ 본 조사의 결과는 통계법 제33조에 의거하여 비밀이 보장되며, 설문에 대한 모든 응답과 개인적인 사항은 철저히 비밀과 무기명으로 처리되고 통계분석의 목적 이외에는 절대 사용되지 않습니다.

해운기업 파산예측 분석을 위한 전문가 조사 질문지

안녕하십니까?

여러 가지 업무로 인해 바쁘신 와중에 귀중한 시간을 내주셔서 대단히 감사드립니다. 저는 영국 플리머스 대학교 박사과정에 재학 중인 권민수입니다.

제 연구주제는 해운기업의 파산 예측에 관해 파산 위험을 사전에 예측 및 대응하기 위한 결정 요인을 탐색하는 것을 목표로 합니다. 이 연구는 비즈니스 관리자나 정책 결정자들이 해운업체에서의 경기 하강의 규모와 강도를 인식하고 대응하기 위해 사전 조치를 취할 수 있도록 기여하는데에 그 목적이 있습니다.

이와 관련하여 본 설문은 연구 목적으로만 진행되며, 따라서 귀하의 응답은 기밀로 처리되고 제 3자에게 공개되지 않습니다. 본 설문은 약 15분 정도 소요될 예정입니다. 질문에 대한 답에 대해 확신이 없으시다 하더라도 최대한 추정하여 답을 해주시면 감사하겠습니다. 만약 설문결과에 대해 알고 싶으신 분은 제게 이메일을 보내주시면 연구가 종료된 후에 요약을 보내드릴 수 있도록 하겠습니다. 소중한 시간 할애해 주심에 감사드립니다. 본 연구에 대한 질문이나 의견이 있으시다면 언제든지 제게 연락을 주시길 바랍니다.

영국 플리머스 대학
국제 물류, 공급망 및 해운 관리 박사과정
권민수

2023년 7월

연구책임자: 권민수 전화: 010-5444-9485 e-mail: minsu.kwon@plymouth.ac.uk

전문가 조사를 위한 참고 자료

□ 연구배경 및 목적

◆ 금융 위기나 COVID-19와 같은 위기 속 다수 해운업체가 화물 수요의 급감과 높은 부채 부담으로 인해 파산
- 해운업은 글로벌 경제적 측면에 영향을 받는 매우 변동성이 높은 산업
- 해운업체, 특히 중소기업의 시장 변동성 예측 및 재무 위험 관리 능력 부족으로 높은 파산 위험 상존

◆ 본 연구의 목적은 해운업체의 위험 요인을 탐색하여 잠재적인 파산 위험을 사전에 평가하고 관리하기 위한 것임
- 관리자나 정책 결정자들이 해운업계의 경기 하강의 규모와 강도를 사전에 파악하고 반응하여 파산 위험에 사전 대응 가능
- 식별된 위험 요인은 해운기업의 파산 위험 시스템 구축을 위해 정책적 측면에서 기여 기대

제 1부

다음은 기존 연구에서 지목된 해운기업의 파산위험요인 목록입니다. 귀 사의 경험을 바탕으로 해운기업의 파산위험을 평가하는데 각 요인의 중요성의 정도를 표시하여 주시기 바랍니다.

(1 = 중요하지 않음, 2 = 중요하지 않음, 3 = 중립, 4 = 중요함, 5 = 매우 중요함)

해운기업 파산위험요인			요인 중요도				
			1	2	3	4	5
레버리지							
A1	법인세 이자 감가상각비 차감 전 영업이익(EBITDA)	기업이 영업활동을 통해 벌어들이는 현금창출 능력					
A2	총자산수익률	EBITDA/총자산					
A3	총자산 대비 이익잉여금 비율	이익잉여금/총자산					
A4	채무비율	총채무/총자산					
A5	총자산 대비 총부채 비율	총부채/총자산					
A6	총자산 대비 자기자본 비율	자기자본/총자산					
A7	자기자본 대비 채무 비율	총채무/자기자본					
A8	자기자본 대비 부채 비율	총부채/자기자본					
A9	레버리지 비율(Gearing 비율)	자기자본/총채무					
A10	부채 대비 자기자본 비율	자기자본/총부채					
A11	부채	총부채					
A12	자기자본 대비 유동부채 비율	유동부채/자기자본					
A13	자기자본 대비 EBITDA 비율	EBITDA/자기자본					
A14	EBITDA 대비 기업가치 비율	기업가치/EBITDA					
A15	총자산 대비 장기부채 비율	(총부채-유동부채)/총자산					
A16	부채 대비 순이익 비율	순이익/총부채					
A17	운전자본	유동자산-유동부채					
A18	부채 대비 유동자산 비율	유동자산/총부채					
A19	유동부채 대비 EBITDA	EBITDA/유동부채					
A20	자기자본 대비 장기부채 비율	(총부채-유동부채)/자기자본					
유동성							
B1	유동비율	유동자산-유동부채					
B2	총자산 대비 운전자본 비율	(유동자산-유동부채)/총자산					
B3	현금자산 비율	현금/총자산					

해운기업 파산위험요인			요인 중요도				
			1	2	3	4	5
B4	현금 비율	현금/유동부채					
B5	총자산 대비 유동자산 비율	유동자산/총자산					
B6	총자산 대비 유동부채 비율	유동부채/총자산					
B7	당좌비율	현금+미수금/유동부채					
B8	채무 대비 현금 비율	현금/총채무					
B9	자기자본	총자산 - 총부채					
B10	유동부채비율	유동부채/총부채					
B11	유동자산	1년 이내에 환금할 수 있는 자산					
B12	현금매출비율	현금/총매출					
B13	유동자산 대비 유동부채 비율	유동부채/유동자산					
B14	유동자산 대비 운전자본 비율	(유동자산-유동부채)/유동자산					
B15	총자산 대비 당좌자산 비율	(현금+미수금)/총자산					
B16	자기자본 대비 운전자본 비율	(유동자산-유동부채)/자기자본					
B17	유동부채 대비 운전자본 비율	(유동자산-유동부채)/유동부채					
B18	유동자산 대비 현금 비율	현금/유동자산					
B19	유동부채 대비 이익잉여금 비율	이익잉여금/유동부채					
B20	총자산 변화율	연간 총자산 변화율					
수익률							
C1	총자산 이익률	순이익/총자산					
C2	자기자본 이익률	순이익/자기자본					
C3	매출 수익률	총이익/매출액					
C4	매출 대비 EBITDA 비율	EBITDA/매출액					
C5	유동자산 대비 매출 비율	매출액/유동자산					
C6	매출액	총매출액					
C7	EBITDA 이자보상 비율	EBITDA/이자비용					
C8	자산 대비 총이익 비율	총이익/총자산					
C9	순이자마진	(이자운용수익-이자조달비용) /운용자산총액					
C10	자기자본 대비 매출 비율	총매출/자기자본					

해운기업 파산위험요인			요인 중요도				
			1	2	3	4	5
C11	순이익	총수익-총비용					
C12	총자산 영업이익률	영업이익/총자산					
C13	투자수익률	순이익/투자액					
C14	부채 대비 매출 비율	총매출/총부채					
C15	유동부채 대비 매출총이익 비율	(매출액-매출원가)/유동부채					
C16	부채 대비 매출총이익 비율	(매출액-매출원가)/총부채					
C17	매출 대비 매출총이익 비율	(매출액-매출원가)/매출액					
C18	매출변화율	연간 매출액 변화율					
C19	유동부채 대비 매출 비율	총매출액/유동부채					
효율성							
D1	자산회전율	매출액/총자산					
D2	매출 대비 운전자본 비율	(유동자산-유동부채)/매출액					
D3	매출 대비 유동자산 비율	유동자산/매출액					
D4	자산 대비 재고자산 비율	재고자산/총자산					
D5	매출 대비 유동부채 비율	유동부채/매출액					
D6	재고회전률	재고자산/매출액					
D7	영업이익률	영업이익/매출액					
비재무요인							
E1	기업규모	Log(총 자산) 기준					
E2	기업연혁	기업 설립 연차					
E3	직원 수	정규 직원 수					
E4	GDP	연간 GDP 변화율					
해운지표							
F1	유가	브렌트 원유 가격($)					
F2	선사업체 유형	제공서비스(컨테이너, 탱커, 벌크)					
F3	용선비용	기업의 연간 용선비용					
F4	운임요율						
F5	정기용선요율						
F6	컨테이너 운임 지수 요율	상하이 컨테이너 운임지수(SCFI) 기준					
F7	발틱운임지수(BDI)	연간 BDI 기준					
F8	철강운임지수	다우존스 미국 철강 지수					

제 2부

1. 국내 해운기업의 **파산 가능성 또는 재무 위험 수준을 판단하기 위해 특히 중요하다고 생각하시는 요인**에는 무엇이 있으며 그 이유는 무엇인지 기재해 주시기 바랍니다.

☞ 위험요인:

※예) 부채비율, 유동자산 보유현황, 순이익률 등

☞ 이　유:

2. 위 목록에 **기재되어 있지 않은 위험 요인 중 중요하다고 생각하시는 요인**은 무엇이 있으며 그 이유는 무엇인지 기재해 주시기 바랍니다.

☞ 위험요인:

※예) 용선비중, 금리변동 등

☞ 이　유:

3. **해운기업의 규모에 따라 위험 요인이 어떻게 달라지는지**에 대한 의견을 기재해 주시기 바랍니다.

☞ 위험요인:

※예) 부채규모, 자본금 보유 현황 등

4. 귀하의 경험을 바탕으로 해운기업의 파산 위험 또는 재무 수준 등을 파악했던 경험이나 수행하셨던 업무에 대해 의견을 기술해 주시기 바랍니다.

※예) 파트너사 선정시, 해운산업 동향분석 등

5. 파산 예측 분석이 해운기업의 경영활동에 있어 어떠한 역할을 할 것인지에 대한 의견을 기술해 주시기 바랍니다.

※예) 조기경보신호 역할, 기업전략 기획, 재무안정성 향상 등

6. 해운기업의 파산예측 및 분석과 관련하여 자유의견을 기재 부탁드립니다.

제 3부

귀하가 해당하시는 곳을 체크해주시기 바랍니다.

1. 귀하가 근무하시는 업종을 선택해주시기 바랍니다.

1) 기업 ( ) 2) 대학 ( ) 3) 연구소, 기관 ( )

2..귀하의 현재 직위에 대해 작성해주시기 바랍니다.

3. 현재 소속한 곳에서 근무하신 기간을 선택하여 주시기 바랍니다.

1) 3년 미만 ( ) 2) 3 ~ 5년 ( ) 3) 5 ~ 10년 ( ) 4) 10년 이상 ( )

자문에 응답해 주셔서 감사합니다.