NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF ECONOMICS AND LAW

THESIS

APPLICATION OF DATA ANALYSIS IN FINANCIAL RISK MANAGEMENT AT LISTED COMPANIES ON VIETNAM STOCK EXCHANGE

Course: Machine Learning

Instructor: **MSc. PHAN HUY TÂM**
Student's name: **NGUYỄN HOÀNG VI**
Student ID: **K214142099**

JUNE 10, 2024

Ho Chi Minh City

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Master of Science Phan Huy Tam, for his invaluable guidance, insightful feedback, and constant encouragement throughout the course of this research. His expertise and dedication were instrumental in shaping the direction and outcome of this study.

I am also profoundly thankful to the University of Economics and Law for providing an excellent academic environment and resources that facilitated this research. Special thanks go to my colleagues and friends for their support and constructive discussions that enriched my understanding of data analysis and financial risk management.

Lastly, I am indebted to my family for their unwavering support and patience, which has been a cornerstone in the completion of this thesis. Their belief in me has always been a source of strength and motivation.

This study would not have been possible without the collective efforts and contributions of all these individuals and institutions.

Thank you.

LIST OF TABLES

TABLE OF FIGURES

ABSTRACT

By LASSO variable selection method and machine learning to predict financial risk, the paper aims to discover factors that have a major impact on financial risk in manufacturing enterprises listed on the Vietnamese stock exchange. Research results indicates that the LASSO model identifies 5 financial indicators affecting financial risks which risky businesses faced during 5 years (2018-2022), they are Short-term solvency ratio (Short-term assets/Short-term liabilities); Total asset turnover (Net revenue/Total assets); ROA (Profit after tax/Total assets); Long-term debt to total asset ratio; Short-term debt to total debt ratio. Artificial Neural Network (ANN) shows the highest efficiency in predicting financial risk. From those finding, the paper proposes some solutions and policies for businesses.

Keywords: *financial risk management, machine learning, data analysis, predictive analysis*

TABLE OF CONTENT

CHAPTER 1. INTRODUCTION

1.1. Reason for choosing the topic

In the rapidly globalizing world, enhancing the operational efficiency of Vietnamese enterprises is crucial for maintaining a competitive edge against international firms. This improvement is not just a pathway to growth but also a shield against unforeseen risks. The COVID-19 pandemic has underscored the urgency of this need, as the government's social distancing measures and closure of non-essential businesses have severely disrupted supply chains and everyday life, causing widespread economic turmoil.

The wave of corporate bankruptcies that began in mid-2020 paints a stark picture. Globally, Euler Hermes predicted a 35% increase in company bankruptcies between 2019 and 2021. The United States, a global economic leader, faced a 57% increase in businesses at risk of bankruptcy in 2021 compared to pre-pandemic levels in 2019. Brazil, the UK, and Spain saw increases of 45%, 43%, and 41% respectively. Even China, where the pandemic first emerged, anticipated a 20% rise in bankruptcies. By the end of 2021, every region experienced a double-digit increase in default rates, with North America leading at +56%, followed by Central and Eastern Europe (+34%), Latin America (+33%), Western Europe (+32%), and Asia (+31%).

Vietnam, though resilient, was not immune. According to the General Statistics Office of Vietnam, the country's GDP saw remarkable growth, reaching over USD 2,700 in 2019. Despite the pandemic, Vietnam's GDP grew by 2.9% in 2020. However, the first half of 2021 witnessed a significant economic strain, with 70,209 businesses exiting the market—a 24.9% increase from the same period in 2020. The number of temporarily suspended operations rose by 22.1%, and business dissolutions surged by 33.8% (Ministry of Industry and Trade, 2021).

Despite these challenges, many businesses have not only survived but thrived. This resilience is often attributed to substantial capital reserves and essential industry status, but this view is overly simplistic. Even well-capitalized companies can collapse quickly without effective capital management and accurate bankruptcy risk forecasting.

Accurate prediction of bankruptcy risk is vital in both crisis and normal conditions. It enables companies to make informed decisions about capital utilization and operational strategies, ensuring their continued strength and preventing bankruptcy. The current complex post-pandemic environment amplifies the importance of this capability.

This research focuses on applying advanced data analysis techniques to manage financial risks for companies listed on the Vietnam Stock Exchange. By leveraging sophisticated data analysis methods, this study aims to provide actionable insights for predicting and mitigating financial risks. These insights are essential for enhancing the resilience and success of Vietnamese companies in an increasingly volatile economic landscape.

1.2. Research objectives

The primary objective of this research is to identify the key factors that significantly impact financial risk in companies listed on the Vietnam Stock Exchange. By analyzing these factors, the study aims to propose effective solutions and policies that can assist managers and investors in making informed and strategic decisions to mitigate risks and enhance financial stability.

1.3. Research subjects and scope

The research subjects are non-financial companies listed on the Vietnamese stock market. The research scope covers ten years, from 2014 to 2023.

1.4. Research methods

This study adopts a systematic approach to evaluate financial risk:

- **S-score Model**: Calculates the bankruptcy risk of businesses, forming the basis for classifying companies into risky and non-risky categories.
- **LASSO Model**: Identifies the top five most influential independent variables from a pool of 29 variables derived from the collected data.

- **Artificial Neural Network (ANN) Model:** Predicts and classifies businesses into risky or non-risky categories based on the identified variables.
- **Confusion Matrix:** Assesses the accuracy and performance of the ANN model in classifying the risk levels of businesses.

By combining these methodologies, the study aims to provide a comprehensive framework for financial risk assessment and management in companies listed on the Vietnam Stock Exchange.

CHAPTER 2. LITERATURE REVIEW

2.1. Studies on bankruptcy risk prediction using traditional models

Traditional research methods, starting with Alman (1968) and Beaver (1966), used financial indicators to predict business bankruptcy risk. Lin (2009) examined the financial distress prediction capabilities of discriminant analysis, logit, and probit models for Taiwanese companies after the 2009 financial crisis, showing favorable results for traditional methods. Serrano-Cinca & Gutiérrez-Nieto (2013) used discriminant analysis with partial least squares to predict the financial crisis of U.S. banks in 2008, demonstrating prediction performance equivalent to that of machine learning models. Liang et al. (2015) used discriminant analysis and logistic regression models to select variables for financial distress prediction, using them as inputs for machine learning models. The main advantage of traditional methods is their interpretability regarding predictive variables and business bankruptcy risk, although they require stringent data conditions.

2.2. Studies on bankruptcy risk prediction using intelligent models

Intelligent models were developed relatively early, with neural network models first emerging in the 1990s (Serrano-Cinca, 1996). Advances in technology have enabled the processing of complex algorithms in a short time, allowing the development of machine learning models with self-improving performance, capable of handling highly complex problems efficiently without demanding much data. Random Forest, proposed by Breiman (2001), involves creating an ensemble of decision trees during the bootstrap process and providing results based on majority voting. In finance, Random Forest has been successfully applied to detect credit fraud (Whitrow et al., 2009) and predict customer churn for banks (Xie et al., 2009).

Research by Zhao et al. (2009) demonstrated that machine learning models are more effective than traditional methods. Similarly, Barboza et al. (2017) showed that Random Forest, bagging, and boosting outperform SVM, logit, and discriminant analysis in effectiveness.

2.3. Studies on bankruptcy risk prediction in Vietnam

In Vietnam, predicting business bankruptcy risk has also attracted considerable interest. Bùi Phúc Trung (2012) used the traditional Z-score method to assess the bankruptcy risk of listed companies. Nguyễn Thị Cành & Phạm Chí Khoa (2014) evaluated Vietcombank's corporate customers to predict bankruptcy probability using the KVM-Merton method. Huỳnh Thị Cẩm Hà et al. (2017) applied a classification tree model in machine learning to predict financial distress of Vietnamese companies, achieving over 90% accuracy. The use of Altman's Z-score for 60 Vietnamese enterprises was demonstrated in Hoàng Thị Hồng Vân's (2020) study, showing a prediction accuracy of 76.67% using indicators such as average assets, ROA, and ROE.

However, studies on the bankruptcy risk of Vietnamese enterprises mainly use traditional models, with limited use of machine learning models. Therefore, this research compares the bankruptcy prediction performance of Vietnamese enterprises using both traditional methods and modern machine learning models.

CHAPTER 3. DATA AND RESEARCH METHODOLOGY

3.1. Data and variables

Data: The enterprises included in the study are manufacturing enterprises listed on the Vietnamese stock exchange. The financial indicators used to predict the financial risk of these enterprises are calculated from the metrics and indicators in publicly audited financial statements (balance sheets, cash flow statements, income statements) at the end of the year for manufacturing enterprises listed on the Vietnamese stock exchange from 2015 to the end of 2023, totaling 2,152 observations.

Dependent variable: The study uses the Z-score (Altman, 1968) to classify enterprises as either at risk or not at risk. The dependent variable (y) is assigned a value of 1 if the Z-score < 0.862 (enterprise faces financial risk), and y is assigned a value of 0 if the Z-score ≥ 0.862 (enterprise is financially safe). This classification is consistent with the results of previous research. The classification of 2,152 observations results in 564 observations in the financial risk group and 1,588 observations in the no financial risk group.

Independent variables: Based on the available data sources and suggestions from studies by Zeytinoglu et al. (2013), Valaskova et al. (2018), etc., as well as the actual operations of enterprises, the study constructed and selected 29 variables (Table 1).

Table 1. Independent Variables in the Model

Variable	Variable Description	Variable	Variable Description
X1	Short-term assets/Short-term liabilities	X16	Revenue/Equity capital
X2	(Short-term assets - Inventory)/Short-term liabilities	X17	Short-term liabilities/Total assets
X3	Revenue/Inventory	X18	Long-term debt/Total assets
X4	Receivables/(Revenue/365)	X19	Profit before tax/Equity capital

X5	Revenue/Fixed assets	X20	log(Market capitalization)
X6	Revenue/Total assets	X21	Inventory/Sales
X7	Total debt/Total assets	X22	Inventory turnover ratio
X8	Equity/Total assets	X23	Net profit after tax/Net revenue
X9	Total debt/Equity capital	X24	Cost of goods sold/Receivables
X10	ROS (Return on Sales)	X25	Fixed assets/Total assets
X11	ROA (Return on Assets)	X26	Short-term liabilities/Payables
X12	ROE (Return on Equity)	X27	Payables/Total assets
X13	log(Sales)	X28	log(Total assets)
X14	(Short-term assets - Short-term liabilities)/Total assets	X29	Working capital/Total assets
X15	Short-term receivables/Short-term assets		

To predict bankruptcy risk for Vietnamese enterprises, I used the financial indicators of 648 Vietnamese enterprises from 2014 - 2023, collected from Institute for Research and Development of Banking Technology, National University - Ho Chi Minh City. Barboza et al. (2017) demonstrated the effectiveness of classifying enterprises with and without risk using financial indicators characteristic of leverage, liquidity, profitability, company size, and growth groups. This assertion is also reflected in the empirical research by Zoričák et al. (2020) in predicting bankruptcy risk of small and medium-sized enterprises. Collectively, Tian & Yu (2017) used 26 financial indicators to predict corporate bankruptcy, with results indicating that liquidity and financial leverage indicators have the greatest impact on predicting financial risk. This study selects financial variables to predict bankruptcy risk of Vietnamese enterprises, researched and referenced from the comprehensive studies of Tian & Yu (2017), and presented in Table 1.

The data consists of 6228 observations with 29 characteristics representing explanatory variables. Enterprises are classified into two groups: 130 bankrupt enterprises and 170 non-bankrupt enterprises, corresponding to the encoded values of 0 and 1. In practice, the number of bankrupt enterprises is smaller than the number of non-bankrupt enterprises, thus data is often unbalanced. However, the disparity ratio in the research dataset is not significant, so the data is used to build the model without additional data balancing techniques. The cleaned data is randomly divided into training and testing sets in a 75% and 25% ratio. The training dataset is used to build the model, and the testing dataset is used to evaluate the model's performance. In this paper, models are used to predict the bankruptcy probability of enterprises one year in advance, with explanatory variables being the indicators of the previous year and the dependent variable being the status of the enterprise in the next year.

Table 2 provides descriptive statistics of the explanatory variables used in the model. Most of the explanatory variables in the model have relatively high standard deviation values compared to the mean values. This indicates the diversity in the dataset of enterprises used to predict bankruptcy risk. This is reflected in indicators of enterprise profitability such as earnings before interest and taxes, net income to total revenue, or indicators reflecting the efficiency of inventory growth management, and operating income to sales revenue. This diversity allows for a more accurate and representative evaluation of the bankruptcy risk prediction model for Vietnamese enterprises.

Table 2. Descriptive Statistics of the Variables

	X1	**X6**	**X11**	**X18**	**X26**
count	7306.000000	7306.000000	7306.000000	7306.000000	7306.000000
mean	2.729195	1.152448	0.028076	0.107426	0.519560
std	8.064433	2.547350	0.381798	0.264845	1.791078
min	0.001221	-0.050494	-21.143065	-0.066669	0.000622
25%	1.060950	0.351314	0.006911	0.000862	0.194984
50%	1.411576	0.813709	0.033866	0.026167	0.367520
75%	2.401049	1.413081	0.076003	0.138554	0.577949
max	408.731402	185.731414	2.873531	9.415347	79.214671

3.2. Research methods

3.2.1. S-Score (1978) Model for Constructing the Dependent Variable

The model was proposed by Springrate [14], to predict distress of financial distress, Companies with a Springate score lower than 0.862 are classified as "high" probability of bankruptcy. The S-Score formula was as follows:

$$\textbf{S-Score} = 1.03 \cdot X1 + 3.07 \cdot X2 + 0.66 \cdot X3 + 0.4 \cdot X4$$

Where:

- X1: Working Capital / Total Assets
- X2: Retained Earnings / Total Assets
- X3: Earnings Before Interest and Taxes (EBIT) / Total Assets
- X4: Earnings Before Taxes / Short - Term Liabilities
- X5: Revenue / Total Assets

Table 3. S-score Interpretation Table

Z	Interpretation
S > 0.862	The firm is safe, based solely on the financial indicators used in the calculation.
S = 0.862	The firm is in a danger zone, requiring attention regarding bankruptcy risk.
S < 0.862	The firm has serious financial problems and is highly likely to go bankrupt.

3.2.2. LASSO model for selecting independent variables

Tibshirani (1996) introduced and developed the LASSO penalty function model for selecting explanatory variables highly correlated with the explained variable in predictive modeling.

The LASSO model is formulated as:

$$\sum_{i=1}^{n}\left(\left(-Y_{i,t}\left(\beta_0+\sum_{k=1}^{p}x_{i,t-1,k}\beta_k\right)\right)+\log(1+\ e^{\beta_0+\sum_{k=1}^{p}x_{i,t-1,k}\beta_k}\right)$$

Here, $y_{i,t}$ represents the binary variable indicating the status of firm i at time t, $x_{i,t-1,k}$ are the kk-th financial indicators of firm i at time t−1, and n is the number of firms in the sample.

The LASSO penalty function is represented by:

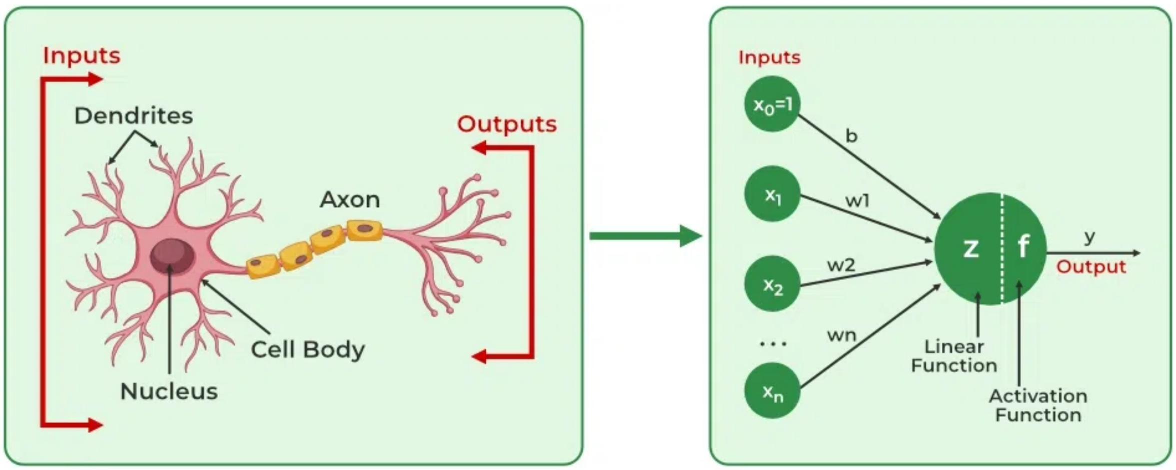
$$\sum_{k=1}^{p}\omega_k\ |\beta_k| <\ \lambda$$

which imposes restrictions on the parameter estimates constrained by the tuning parameter λ. A smaller λ value retains fewer explanatory variables in the predictive model (Le Hai Trung et al., 2023). This penalty function, also known as the "L1" penalty, automatically sets parameter estimates of insignificant explanatory variables to 0 and reduces the estimates of less meaningful explanatory variables. Another advantage of the LASSO method is its ability to handle the multicollinearity issue among explanatory variables. This is particularly beneficial for financial risk forecasting models, as they often utilize multiple highly correlated financial variables (Tian et al., 2015).

3.2.3. Artificial neural network model for predicting financial risk

Artificial Neural Network (ANN) is a computational model that simulates the neural network functions of the human brain. Each neuron is an information processing unit and a basic component of a neural network. The structure of a neuron is described as in Figure 1.

Figure 1. Artificial Neural Network



The basic components of an artificial neuron include:

- Input set: These are the input signals of the neuron, usually represented as an m - dimensional vector.

- Link set: Each link is represented by a weight (called synaptic weight). The synaptic weight between the j input signal and neuron k is typically denoted as w_{jk}. Typically, these weights are randomly initialized at the network initialization and continuously updated during the learning process.

- Summing function: Usually used to calculate the sum of the products of inputs with their synaptic weights.

- Threshold (also known as bias): This threshold is often introduced as a component of the transfer function.

- Transfer function (Activation function): This function is used to limit the output range of each neuron. It takes the input as the result of the summing function and the given threshold.

Typically, the output range of each neuron is limited to [0,1] or [-1, 1]. Various transfer functions are listed in Table 1.1, which can be linear or nonlinear functions. The selection of the transfer function depends on the specific problem and the experience of the network designer.

- Output: This is the output signal of a neuron, with each neuron having at most one output.

Artificial neurons, like biological neurons, receive input signals, process them by multiplying these signals with synaptic weights, calculate the sum of the obtained products, and then send the result to the transfer function, where the output signal is produced.

$$Y = (X_1w_1 + \ b_1) + (X_2w_2 + \ b_2) + \ldots + (X_nw_n + \ b_n)$$

While individual neurons have specific information processing functions, the computational ability of neurons is mainly achieved through the structured connections of neurons, allowing them to handle complex computations and produce accurate results.

3.2.4. Confusion matrix for evaluating prediction performance

In this study, we employ the method of comparing financial risk forecasting models based on the confusion matrix. This method evaluates the classification performance of observations into two risk classes: risk and non-risk, by accuracy and the comprehensiveness levels of the classification. The risk class is assigned a value of 1, while the non-risk class is assigned a value of 0. The confusion matrix consists of the following indices (Figure 2):

- TP (true positive) is the number of positive predictions, meaning the number of enterprises at financial risk correctly predicted to be at financial risk;
- TN (true negative) is the number of enterprises not at financial risk correctly predicted not to be at financial risk;
- FP (false positive) is the number of enterprises not at financial risk but predicted to be at financial risk;

- FN (false negative) is the number of enterprises at financial risk but predicted not to be at financial risk.

Figure 2. Confusion Matrix

	True Value		
		1	**0**
Predicted Result	**1**	**TP**	**FP**
	0	**FN**	**TN**

The accuracy of the model is the proportion of correct predictions, calculated by the following formula:

Accuracy = (TP + TN) / (TP + FP + TN + FN)

Therefore, to enhance the predictive efficiency of the model, in addition to the model's accuracy, the following two criteria are used to evaluate predictive effectiveness: Precision and Recall.

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

Precision indicates the proportion of actual financial risk enterprises predicted correctly to be at financial risk, while Recall indicates the proportion of correctly predicted financial risk enterprises among the total financial risk enterprises.

CHAPTER 4. RESEARCH RESULTS

In the financial risk forecasting for listed manufacturing companies on the stock exchange in Vietnam, I divided the dataset into two sets: the training set consisting of 4 years from 2018 to 2021 and the test set, which is the data from the year 2022. The training set is used for machine learning, and the test set is used to evaluate the learning capability of the machine. In some cases, to avoid overfitting, I used the method of selection based on the LASSO model to select variables that affect a company's finances and have the potential to pose high financial risks to the company. The research process is divided into three parts:

4.1. Applying the LASSO Model

Using LASSO regression to find important explanatory variables. After applying LASSO regression, the selected variables are X1, X6, X11, X18, X26. These variables represent: Short-term Assets/Short-term Liabilities (X1), Revenue/Total Assets (X6), ROA (Net Profit/Total Assets) (X11), Long-term Debt/Total Assets (X18), Short-term Debt/Total Assets (X26). The results show that these variables, representing liquidity, financial efficiency, profitability, leverage, and financial risk of the company, help in forecasting and distinguishing companies with high or low financial risk.

4.2. Selecting Machine Learning Models

Re-forecasted financial risk for companies using machine learning models with the financial variables selected through LASSO regression. The performance of these models is shown in Table 2.

Table 4. Forecast results with LASSO model variables of machine learning models

	Logistic Regression	Decision Tree	Random Forest	K-Nearest Neighbors	Support Vector Machine	Artificial Neural Network
Accuracy	0.886457	0.936389	0.949384	0.925445	0.924761	0.971272
Precision	0.805797	0.865979	0.924370	0.864130	0.886957	0.956403
Recall	0.737401	0.891247	0.875332	0.843501	0.811671	0.931034
F1 Score	0.770083	0.878431	0.899183	0.853691	0.847645	0.943548

The results (Table 2) show that all models forecast the classification of companies in the test file with an accuracy of over 92%, with the artificial neural network (ANN) model having the highest accuracy at approximately 96.3%, higher than other intelligent machine learning models such as Random Forest or SVC. Surprisingly, the traditional Logistic Regression model, with an accuracy of over 94%, ranks fourth after other intelligent models.

The high Recall_score indicates a lower rate of missed companies at risk, and the high Precision_score reflects high predictive accuracy. The best model not only has high accuracy but also has both Recall_score and Precision_score above 90%. In these two criteria, the artificial neural network (ANN) model continues to have the highest predictive performance. The Decision Tree model shows a higher ability to miss companies at risk and lower predictive accuracy (both Recall_score and Precision_score are <85%), indicating the robustness of the Decision Tree model is not significant.

Based on the comparative evaluation results, I decided to select the artificial neural network (ANN) model to continue the evaluation and forecasting, identifying financial risks in listed manufacturing companies on the stock exchange.

4.3. Training and Testing Data with Artificial Neural Network Model

The artificial neural network (ANN) model is a popular type of neural network in machine learning, with at least one hidden layer between the input and output layers.

Training on the dataset

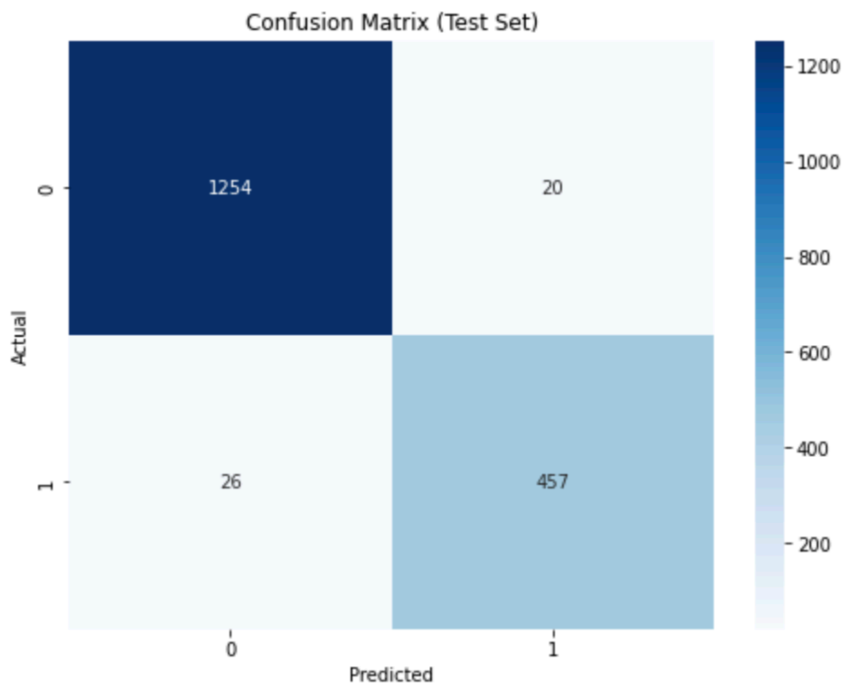
After selecting the neural network model as the main model, I retrained the neural network model based on the dataset of manufacturing companies listed on the stock market from 2017 to 2020 with the selected variables X1, X6, X11, X18, X26 (Table 1) chosen through LASSO regression testing results. I further randomly split the training dataset from 2017 to 2020, with 70% for 'train' data and 30% for 'test' data, and then ran the neural network model. The accuracy results of the model on the 'train' set are 97.83%, and the accuracy on the 'test' set is 97.38%.

Model accuracy is used to evaluate the performance of the neural network model, and for classification capability, accuracy is usually an important index to assess the accuracy level after training the model on the test set. However, to clearly see the confidence level of the results achieved by the model and detect errors that the model encounters, thereby improving and adjusting the model to enhance prediction performance and minimize errors during the training and testing process of the model, I used the confusion matrix to evaluate and understand the performance of a classification model, as well as to support decision-making in improving the model.

The obtained confusion matrix results with the y-axis as the actual results and the x-axis as the predicted results are as follows:

- Actual: 0 represents companies without financial risk.
- Actual: 1 represents companies with financial risk.
- Predicted: 0 represents the predicted result of companies without financial risk.
- Predicted: 1 represents the predicted result of companies with financial risk. From the results of the confusion matrix (Figure 3), we can see:

Figure 3. Confusion Matrix results on the dataset from 2018 to 2021



As mentioned in the previous section, a high Recall_score implies a lower rate of missed companies at risk, while a high Precision_score reflects high predictive accuracy. After retraining on the dataset, the Recall_score and Precision_score achieved are 94.62% and 95.81%, respectively.

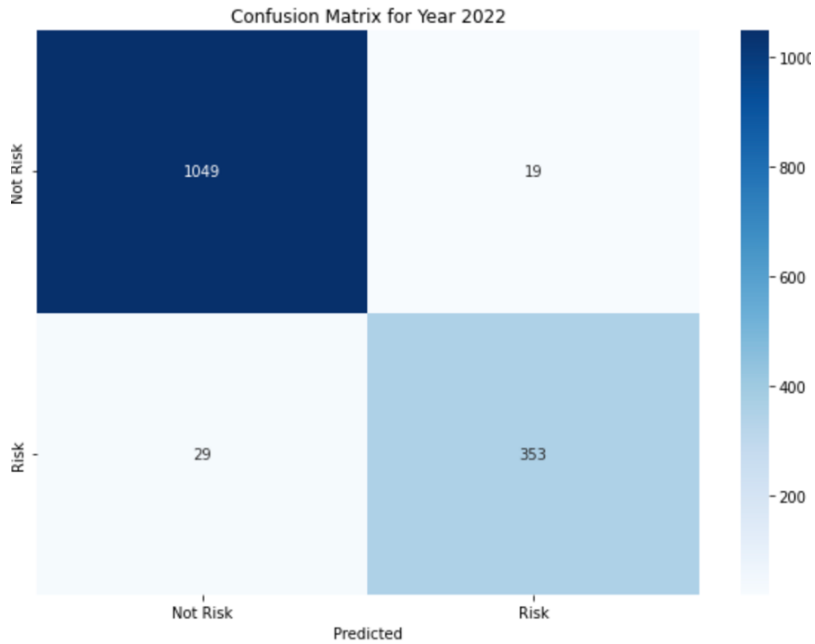
Re-testing on real data from 2022

To verify that the model still performs well and does not suffer from overfitting, we conduct a test on the dataset from the year 2022. The accuracy of the model on the dataset of manufacturing companies listed on the stock market from 2018 to 2021 is 97,38%. Therefore, for the data that the model has not been trained on and tested, it will yield an accuracy result of how many percentages?

To validate the performance of the model, we continue to run the model on the real dataset of manufacturing companies listed on the stock market in 2021 with the variables X1, X6, X11, X18, X26. The result shows an accuracy of up to 96,69%.

And the result when using the confusion matrix (Figure 4) indicates:

Figure 4. Confusion Matrix results on the dataset from the year 2022



4.4. Discussion

The application of data analysis in financial risk management has proven to be a crucial component for listed companies on the Vietnam Stock Exchange. The implementation of predictive models, particularly those utilizing machine learning techniques like LASSO regression, has allowed for a more nuanced and accurate assessment of financial risk. This section discusses the results obtained from training and re-testing the predictive model, highlighting its efficacy and areas for improvement.

Training Phase Results

The predictive model was initially trained on a comprehensive dataset, and the results were encouraging. The confusion matrix from the training phase is as follows:

- Coordinate (Actual: 0, Predicted: 0): The model correctly identified 1254 companies without financial risk.

- Coordinate (Actual: 0, Predicted: 1): The model incorrectly predicted financial risk for 20 companies that were not at risk.

- Coordinate (Actual: 1, Predicted: 0): The model incorrectly identified 26 companies with financial risk as not being at risk.

- Coordinate (Actual: 1, Predicted: 1): The model correctly identified 457 companies with financial risk.

These results demonstrate the model's high accuracy, particularly in correctly identifying companies with financial risk (95.9% accuracy for companies with financial risk). However, the false positive and false negative rates indicate areas where the model can be refined. Specifically, the model misclassified 20 companies as having financial risk when they did not, and failed to identify financial risk in 26 companies that were at risk.

Re-Testing on 2022 Data

To validate the model's robustness, it was re-tested on real data from 2022. The confusion matrix for this phase is as follows:

- Coordinate (Actual: 0, Predicted: 0): The model correctly identified 1049 companies without financial risk.

- Coordinate (Actual: 0, Predicted: 1): The model incorrectly predicted financial risk for 19 companies that were not at risk.

- Coordinate (Actual: 1, Predicted: 0): The model incorrectly identified 29 companies with financial risk as not being at risk.

- Coordinate (Actual: 1, Predicted: 1): The model correctly identified 353 companies with financial risk.

The re-testing results are consistent with the training phase, reinforcing the model's reliability. The slight increase in the number of false negatives (29 compared to 26 in the training phase) and the slight decrease in false positives (19 compared to 20) suggest that while the model maintains its overall accuracy, there is a need for continuous improvement to minimize misclassifications further.

The implementation of predictive analytics in financial risk management offers several significant advantages:

> Enhanced Predictive Accuracy: The high accuracy rates of the model in both training and re-testing phases indicate that predictive analytics can effectively identify financial risks, allowing for timely intervention and mitigation strategies.
> Proactive Risk Management: Companies can use these insights to adopt a proactive approach to risk management, identifying potential issues before they escalate into significant financial distress.
> Strategic Decision-Making: Data-driven decision-making allows for more informed and strategic planning, improving overall financial health and resilience.
> Regulatory Compliance: Accurate risk assessments ensure that companies remain compliant with financial regulations, thereby avoiding legal and financial penalties.

Despite the positive results, there are several challenges and areas for future improvement:

- Model Adaptation: The financial environment is dynamic, and the model must continuously adapt to changing conditions and new data to maintain its accuracy and relevance.

- Data Quality: The accuracy of predictive models is heavily dependent on the quality and completeness of the data. Ensuring high-quality data collection and preprocessing is essential.

- Integration with Business Processes: For maximum effectiveness, predictive analytics should be seamlessly integrated into existing business processes and decision-making frameworks. This requires investment in technology and training for personnel.

Future directions for enhancing financial risk management through data analysis include incorporating more sophisticated techniques such as deep learning and exploring additional financial and macroeconomic variables to provide a more comprehensive risk assessment.

The application of data analysis in financial risk management for listed companies on the Vietnam Stock Exchange demonstrates the significant potential of predictive analytics. By accurately identifying financial risks, companies can take proactive measures to safeguard their financial health and ensure sustainable growth. Continuous improvement and adaptation of predictive models, along with high-quality data and seamless integration into business processes, will be key to leveraging the full potential of data analysis in financial risk management.

CHAPTER 5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Based on the dataset of 750 manufacturing companies listed on the Vietnam stock exchange from 2018 to 2022, with 29 financial indicators influencing financial risk, the research results show that out of these 29 indicators, the LASSO model selects 5 key indicators: Liquidity, Operational efficiency, Profitability, Debt structure, and Long-term debt to asset ratio, which have the most significant impact on the financial risk of companies over the 5-year period (2018-2022). This is consistent with reality as these indicators are commonly used by companies to assess their financial health.

In this study, the authors also compared various forecasting models and found that the artificial neural network (ANN) is the best model for predicting whether a company will encounter financial risk. This is in line with the research findings of Dao Trong Thinh and Doan Van Toan (2016) and also aligns with the research results of Qin (2022). The artificial neural network model demonstrates strong adaptability to different types of data and robust operational capabilities. The model I have developed achieves an accuracy of 97.83% on the training dataset and 96.69% accurate prediction of whether companies will face financial risk based on real-world data. The model also performs well on machine learning evaluation metrics, such as recall at 94.62% and precision at 95.81%.

The research results can support departments within companies in detecting financial risks, aiding investors as well as business owners in making informed decisions. Additionally, these findings serve as a reference for further research groups.

5.2 Recommendations

Financial risk management is a crucial factor in enterprise risk management to ensure successful operations, including activities such as identifying potential risks, assessing their impact on business operations, and preparing plans to cope with adverse events. However, managing financial risks for companies requires significant resources and time. Companies need to establish financial risk management departments, adhere to financial risk management processes, utilize various risk identification techniques, and seek funding sources for financial risk management activities.

Applying data analysis in financial risk management can help companies assess themselves more thoroughly and comprehensively. This maximizes resource efficiency, saves time while still yielding accurate results, greatly contributing to the development and survival of businesses. Companies need to focus more on human resources and resources in developing data analysis, combining experience and data understanding to make the right decisions. Because data is the most authentic reflection of all business issues.

REFERENCES

[1]. Altman, E. (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, Journal of Finance, 23, 589-609.

[2]. Đào Trọng Thịnh, Doãn Văn Toàn (2016), Ứng dụng mô hình mạng thần kinh nhân tạo trong quản trị rủi ro tín dụng tại một số ngân hàng thương mại trên địa bàn TP. Biên Hòa, Tạp chí Khoa học Lạc Hồng, số 5, 31-35

[3]. Đoàn Khánh Hưng và Trần Thị Hiền (2019), Tác động của cấu trúc tài chính đến rủi ro tài chính của các doanh nghiệp dịch vụ du lịch niêm yết tại Việt Nam, Tạp chí Khoa học Đại học Huế: Kinh tế và Phát triển, 128(5A), 93-104.

[4]. Huang, B., Wei, J., Tang, Y., and Liu, C. (2021), Enterprise risk assessment based on machine learning, Computational Intelligence and Neuroscience.

[5]. Murugan, M. S. (2023), Large-scale data-driven financial risk management & analysis using machine learning strategies, Measurement: Sensors, 27.

[6]. Lê Hải Trung, Trương Thị Thùy Dương (2023), Ứng dụng phương pháp LASSO trong dự báo rủi ro phá sản của các doanh nghiệp Việt Nam, Tạp chí Khoa học & Đào tạo Ngân hàng, số 250.

[7]. Qin, W. (2022), Research on financial risk forecast model of listed companies based on convolutional neural network, Scientific Programming, 1-10, DOI 10.1155/2022/3652931.

[8]. Shrivastava, S., Jeyanthi, P. M., and Singh, S. (2020), Failure prediction of Indian Banks using SMOTE, LASSO regression, bagging and boosting, Cogent Economics & Finance, 8(1).

[9]. Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. Journal of Royal Statistical Society, B, 58, 267-288.

[10]. Valaskova, K., Kliestik, T., and Kovacova, M. (2018), Management of financial risks in Slovak enterprises using regression analysis, Oeconomia copernicana, 9(1), 105-121.

[11]. Valaskova, K., Kliestik, T., Svabova, L., and Adamko, P. (2018), Financial risk measurement and prediction modelling for sustainable development of business entities using regression analysis, Sustainability, 10(7).

[12]. Vuong, Q. H. (2023). Mindsponge Theory. Walter de Gruyter GmbH.

[13]. Zeytinoglu, E., and Akarim, Y. D. (2013), Financial failure prediction using financial ratios: An empirical application on Istanbul Stock Exchange, Journal of Applied Finance and Banking, 3(3).

[14] Springate, G. L. V. (1978). Predicting the possibility of failure in a Canadian firm (Unpublished master's thesis). Simon Fraser University, Canada.