Credit Risk Assessment and Credit Strategy Model Based on Gradient Descent Decision Tree Algorithm and Nonlinear Programming

Guoqing Chen

A Thesis Submitted in Partial Fulfillment of Requirements for
degree of Doctor of Philosophy in Statistical Management Science
March 2025

การประเมินความเสี่ยงและแบบจำลองกลยุทธ์ด้านเครดิตตามอัลกอริทึมแผนผังการตัดสินใจแบบไล่ระดับและโปรแกรมแบบไม่เชิงเส้นตรง

วิทยานิพนธ์

ของ

Guoqing Chen

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

มีนาคม 2568

Credit Risk Assessment and Credit Strategy Model Based on Gradient Descent Decision Tree Algorithm and Nonlinear Programming

Guoqing Chen

A Thesis Submitted in Partial Fulfillment of Requirements

for Doctor of Philosophy (Statistical Management Science)

March 2025

The examining committee has unanimously approved this Thesis, submitted by Mr. Guoqing Chen , as a partial fulfillment of the requirements for the Doctor of Philosophy Statistical Management Science at Mahasarakham University

Examining Committee

.... Chairman
(Assoc. Prof. Manad Khamkong , Ph.D.)

.... Advisor
(Assoc. Prof. Piyapatr Busababodhin , Ph.D.)

.... Co-advisor
(Assoc. Prof. Nipaporn Chutiman , Ph.D.)

.... Committee
(Asst. Prof. Monchaya Chiangpradit , Ph.D.)

.... External Committee
(Asst. Prof. Wuttichai Srisodaphol , Ph.D.)

Mahasarakham University has granted approval to accept this Thesis as a partial fulfillment of the requirements for the Doctor of Philosophy Statistical Management Science

....	....
(Prof. Pairot Pramual , Ph.D.)	(Prof. Anongrit Kangrang , Ph.D.)
Dean of The Faculty of Science	Acting Dean of Graduate School

TITLE	Credit Risk Assessment and Credit Strategy Model Based on Gradient Descent Decision Tree Algorithm and Nonlinear Programming
AUTHOR	Guoqing Chen
ADVISORS	Associate Professor Piyapatr Busababodhin , Ph.D. Associate Professor Nipaporn Chutiman , Ph.D.
DEGREE	Doctor of Philosophy **MAJOR** Statistical Management Science
UNIVERSITY	Mahasarakham University **YEAR** 2025

ABSTRACT

Small and medium-sized enterprises (SMEs) play a crucial role in economic development, driving innovation, employment, and economic growth. However, due to their relatively small scale, limited capital accumulation, and lack of sufficient collateral, SMEs often face significant challenges in accessing financing. Financial institutions encounter difficulties in accurately assessing credit risk and designing optimal credit strategies when extending loans to SMEs. Traditional credit decision-making approaches typically rely on financial indicators and rule-based judgments, failing to fully leverage transaction data and supply chain relationships, thereby reducing the efficiency of credit resource allocation. This research aims to develop a credit risk assessment and credit strategy optimization model based on the Gradient Descent Decision Tree (GDDT) algorithm and nonlinear programming. By incorporating machine learning techniques, the proposed model captures non-linear and complex relationships within credit data, enhancing the accuracy of credit risk evaluation. Simultaneously, optimizing credit strategies through nonlinear programming ensures a more rational allocation of credit resources, balancing risk control with profit maximization. This study provides technical support for financial institutions, advancing intelligent credit decision-making, enhancing SMEs' financing capabilities, and contributing to sustainable economic development.

For objective one, this study determines the risk evaluation index system for small and medium-sized enterprises with a history of credit risk, and select appropriate models to quantify the risks of these enterprises. The company's credit risk will be quantified by the entropy weight method and TOPSIS model, and specific calculation steps will be used. Through the mining and cleaning of original data, eleven risk assessment indicators are obtained, namely total input price and tax, total sales price and tax, total valid invoices, monthly input amount coefficient of variation, monthly sales amount coefficient of variation, number of upstream enterprises, number of downstream enterprises, total profit, operation time, percentage of valid invoices, and credit rating. This study adopts the entropy weight method based on

data-driven weight calculation method to assign weights. The weights of the above eleven indicators are 0.3078, 0.1819, 0.0011, 0.0018, 0.0058, 0.1017, 0.0023, 0.0939, 0.1679, 0.1110, and 0.0248 respectively. After obtaining the weights, TOPSIS method is used to quantify the credit risk of each enterprise. Then, the research compares the results of enterprise quantified risk with the credit rating provided by the original data to facilitate verification of the rationality and accuracy of the model. The changing trends of the two variables are basically the same, and the degree of fit is very high, which shows that the credit risk calculated in this paper is very accurate and the credit strategy can be analyzed based on this data.

For objective two, Based on the evaluation indicators determined in the previous research, this research improves the traditional decision tree algorithm by adding a regularization function to solve the overfitting problem of the algorithm. It establishes a credit risk assessment model with the default probability of the enterprise as the dependent variable, and then predicts the probability value of the enterprise's default. From the simulation results, the AUC value of the area under the ROC curve of Gradient Descent Decision Tree Algorithm reaches 0.99. At the same time, the AUC value of the AdaBoost algorithm is 0.97, and the AUC value of the traditional decision tree algorithm is 0.82, and The Gradient Descent Decision Tree Algorithm works well. Next, this study will use the Gradient Descent Decision Tree Algorithm to calculate the default probability of 302 companies without credit records. Finally, Using nonlinear programming model to formulate credit strategies for 302 enterprises without credit records.

For objective three, Considering that the production, operation and economic benefits of enterprises may be affected by unexpected factors, such as the COVID-19 epidemic, we have comprehensively considered the credit risks of different industries and enterprise categories and their response strategies to ensure that banks can flexibly adjust credit policies in different situations and reduce potential risks. In the end, the bank's total loan amount was 100 million yuan, and the expected return was 5.4004 million yuan. Compared with normal circumstances, considering that the total loan amount of the bank's loan strategy after the policy subsidy did not change, the expected return on the loan was reduced by 1.3502 million yuan. However, this adjustment strategy can help many small and medium-sized enterprises in dire straits to overcome difficulties. It reflects the bank's great sense of responsibility and social responsibility, has produced huge social benefits, and further improved the bank's reputation and customer retention rate. From this perspective, this strategic adjustment has a positive impact.

Keyword : Credit risk assessment, Small and medium-sized enterprises, Gradient descent decision tree algorithm, Nonlinear programming, Entropy weight method

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to my esteemed supervisor, Associate Professor Dr. Piyapatr Busababodhin and my co-supervisor, Associate Professor Dr. Nipaporn Chutiman, for their invaluable guidance, unwavering support, and genuine care throughout my doctoral journey. Your profound knowledge, insightful advice, and meticulous attention to detail have been instrumental in shaping my academic growth and research skills. Beyond scholarly pursuits, your encouragement and understanding have provided me with strength and resilience during challenging moments, both academically and personally. Your kindness and patience have made my academic experience truly enriching and memorable. Thank you for being a mentor, a role model, and a source of inspiration in my life. It has been a privilege and an honor to work under your guidance. I am sincerely grateful to Dr. Sujitta Suraphee, Dr. Monchaya Chiangpradit, Dr. Khemmanant Khamthong, and Dr. Supawadee Wichitchan, for their engaging lectures, thought-provoking discussions, and dedication to teaching have significantly enriched my knowledge and intellectual development. I would like to express my sincere gratitude to the members of the academic committee Associate Professor Dr. Manad Khamkong from Chiang Mai University and Assistant Professor Dr. Wuttichai Srisodaphol form Khon Kaen University,for their thoughtful questions, insightful suggestions, and valuable discussions during the defense process and beyond have significantly enriched my research.Last but not least, My thankful send to all academic staffs members of the Department of Mathematics, Mahasarakham University.

Guoqing Chen

TABLE OF CONTENTS

Page

List of Tables

List of Figures

Page

Chapter 1
INTRODUCTION

This introductory chapter presents a broad overview of the research, setting the stage with background and context. It defines the study's purpose and objectives, emphasizing its significance and relevance within today's financial environment. Additionally, the chapter explores the financing challenges encountered by small and medium-sized enterprises (SMEs) and underscores the importance of effective credit risk management strategies.

1.1 Overview

Small and medium-sized enterprises (SMEs), as the most numerous and dynamic actors in the market economy, play a pivotal role in fostering employment, enhancing public welfare, stabilizing society, driving economic growth, and spurring innovation. Recognizing their importance, the State Council of China issued the Opinions on Further Strengthening Financial Services for Small and Medium-sized Enterprises to boost credit support, improve the financial environment, and mitigate credit risks associated with SMEs. In tandem, the China Banking and Insurance Regulatory Commission implemented detailed regulations that compel banks to enhance credit risk management by refining loan approval processes and improving risk evaluation accuracy. These comprehensive policy initiatives highlight the relevance and urgency of studying credit risk assessment for SMEs in China and provide a significant policy context for such research.

According to the 2022 Credit Risk Report from the People's Bank of China, by the end of 2022, non-performing loans held by SMEs had surpassed 10 trillion yuan, with the average non-performing loan rate exceeding 5%. This substantial exposure to credit risk underscores the pressing need for thorough investigation and systematic evaluation to address the financial vulnerabilities faced by SMEs. Accurate credit risk assessment is a critical tool that allows financial institutions to gauge an enterprise's repayment capacity and loan default risk. Effective risk evaluation is fundamental for ensuring the sustainable development of SMEs and strengthening the risk management frameworks of financial entities.

This study aims to conduct a comprehensive analysis of the credit risk landscape for Chinese SMEs. It delves into core challenges influencing credit risk, including information asymmetry, incomplete or inaccurate financial data, imbalances in the supply and demand of capital, macroeconomic fluctuations, and policy-related factors. Additionally, the research evaluates the suitability of existing credit risk models when applied to SMEs within the Chinese market. By designing a refined, comprehensive, and scientifically sound risk assessment index system, the study seeks to offer precise and objective insights into the risk profile of SMEs. Such a system will provide a clearer understanding of current credit risks, facilitating the development of actionable strategies to reduce risk exposure, enhance SME sustainability, and strengthen financial institutions' risk management practices. Moreover, the findings will support policymakers, financial institutions, and SMEs in implementing effective credit risk control measures. The broader goal is to contribute valuable references for risk management improvement and to foster a healthier financial environment that promotes the stable, long-term growth of SMEs, thereby further advancing China's economic development and innovation-driven transformation.

1.2 Background

Small and medium-sized enterprises (SMEs) have a large scale and are full of innovative vitality. They are an important component and core driving force of China's economic and social development, playing a crucial role in promoting market prosperity, economic growth, supporting export earnings, expanding employment, and promoting technological innovation. However, due to factors such as small scale and low market share of small and medium-sized enterprises, it is difficult to meet the conditions for direct financing. Therefore, in the financing process, they mainly rely on indirect financing, such as commercial bank loans.

1.2.1 Analysis of the current development status of SMEs

In recent years, small and medium-sized enterprises (SMEs) in China have exhibited a clear trend of growth in both quantity and quality, reflecting their increasing importance within the national economy. From a quantitative perspective, the number of SMEs rose significantly, reaching approximately 433,000 in 2021, compared to around 363,000 in 2017—an impressive increase of about 70,000 enterprises. This expansion highlights the rapid development phase of China’s SME sector. Moreover, the total profits generated by SMEs have demonstrated a general upward trajectory over the past several years. The only notable exception occurred in 2018 when profits experienced a decline, primarily due to the adverse effects of the U.S. financial crisis.

Qualitatively, a growing number of high-caliber SMEs have emerged, contributing to industrial innovation and specialized development. Data from the Ministry of Industry and Information Technology (MIIT) reveals that nearly 9,000 so-called “little giant” enterprises—firms recognized for their specialization and technological expertise—have been nurtured in recent years. These enterprises, in turn, have driven the growth of over 80,000 provincial-level specialized, refined, and innovative SMEs across various regions. Notably, nearly 70 "little giant" enterprises have received the prestigious National Science and Technology Award, while more than 1,500 have actively participated in major national science and technology projects, underscoring their critical role in driving technological advancement and innovation.

SMEs have become the backbone of the Chinese economy, representing the largest and most dynamic group of enterprises. Their robust development and significant contributions to economic vitality are evident. A breakdown of industrial enterprises as of the end of 2022 shows that medium-sized enterprises accounted for 8.82% of the total, small enterprises made up 89.26%, and large enterprises constituted only 1.92%. Consequently, SMEs collectively represented a dominant 98.08% of all industrial enterprises. This overwhelming majority not only highlights their extensive market presence but also underscores their pivotal role in shaping China’s industrial structure, fostering employment, and driving innovation-led growth. The continued expansion and qualitative enhancement of SMEs are critical for sustaining the country’s economic momentum and achieving long-term development goals.

1.2.2 Analysis of the Current Situation of Loans for SMEs

Small and medium-sized enterprises (SMEs) in China continue to face significant challenges in accessing loans and financing, which hampers their growth and development. One of the primary obstacles is the severe shortage of credit allocation from commercial banks to SMEs. Banks, by nature, often prioritize lending to large enterprises due to their economies of scale, stable operations, higher creditworthiness, and lower associated risks. This preference leaves SMEs at a disadvantage. In addition, the loan application process for SMEs is notoriously cumbersome. It involves the submission of extensive financial documentation and detailed operational information, coupled with lengthy approval cycles. These bureaucratic procedures further complicate the ability of SMEs to secure necessary funding in a timely and efficient manner, exacerbating their financial constraints.

Another significant issue is the limited diversity of loan products available to SMEs. Chinese banks primarily offer fixed asset loans, working capital loans, and trade financing. While these forms of financing are essential, their narrow scope restricts SMEs' access to alternative financing options. The lack of varied loan types makes it difficult for enterprises at different growth stages or with unique financing needs to find suitable funding solutions. Without diversified financial products tailored to specific requirements, many SMEs struggle to secure adequate capital for innovation, expansion, and operational stability.

Furthermore, indirect financing channels for SMEs remain underdeveloped and highly concentrated. Although alternative financing methods, such as leasing, have proven effective in addressing medium- to long-term funding needs—especially for equipment upgrades and modernization—their adoption in China is still at an early stage. Only commercial banks possess sufficient financial resources to offer such financing, yet they suffer from outdated business models and a limited range of financial products. This lack of innovation and diversification in financial services prevents banks from effectively meeting the evolving and varied funding demands of SMEs in today's dynamic economic landscape.

In summary, SMEs in China face a multifaceted financing dilemma characterized by insufficient credit support, a narrow range of loan options, and underdeveloped alternative financing mechanisms. Addressing these challenges requires not only policy reforms to encourage greater credit allocation to SMEs but also the development of innovative, diversified financial products and services tailored to their diverse needs. Improving access to financing will empower SMEs to enhance their operational capacity, invest in innovation, and contribute more effectively to economic growth and industrial upgrading.

1.2.3 Analysis of loan disbursement for SMEs in commercial banks

(1) Analysis of Small and Medium sized Enterprise Loans from Commercial Banks in Various Provinces

By the end of 2022, the total outstanding loans extended by Chinese banking and financial institutions to small and medium-sized enterprises (SMEs)—including loans for SMEs, individual businesses, and small and micro enterprise owners—amounted to an impressive 59.70 trillion yuan. Within this figure, loans classified as *inclusive* for small and micro enterprises, where the total credit limit does not exceed 10 million yuan per borrower, reached 23.6 trillion yuan, reflecting a significant year-on-year growth rate of 23.6%. This rapid expansion demonstrates the increasing emphasis on supporting smaller businesses through targeted financial measures.

A geographical breakdown highlights five provinces leading in inclusive small and micro enterprise loans: Zhejiang, Guangdong, Jiangsu, Shandong, and Fujian, together accounting for over half of the nation's total loan balance for small and micro enterprises. Among these, Zhejiang Province stands out for its innovative approach in creating a comprehensive policy framework tailored to differentiated financial services. For the first time, Zhejiang has developed a specialized system featuring seven distinct loan models designed specifically for small and micro enterprises. Central to these efforts is the strategic application of technology-driven solutions to enhance financial inclusion. By leveraging local financial service platforms and big data analytics, Zhejiang has significantly increased the availability of "credit loans" that do not require traditional forms of collateral. This approach marks a departure from conventional risk management strategies heavily reliant on collateral-based lending and the pawn culture.

Through such innovations, Zhejiang has expanded the diversity and reach of credit loan products, thereby improving the financial ecosystem for SMEs. The use of big data not only optimizes risk assessment but also reduces the information asymmetry that historically hindered small businesses from accessing adequate financing. By embracing these advanced methodologies, financial institutions in Zhejiang have been able to broaden their business operations, promote product innovation, and provide more substantial support for SME growth and development.

Overall, these measures reflect a broader national trend toward fostering an inclusive financial environment for SMEs. The dynamic evolution of lending practices, particularly in Zhejiang and other leading provinces, underscores the importance of adopting technology-enhanced, data-driven models to ensure more efficient and accessible credit solutions. Such frameworks play a pivotal role in empowering SMEs to overcome financing barriers, stimulate economic growth, and drive innovation, thereby contributing to the long-term stability and sustainability of China's economy.

(2) Analysis of Small and Medium sized Enterprise Loans from Commercial Banks in Various Years

Between 2019 and 2022, the loan volume for small and medium-sized enterprises (SMEs) in China exhibited a consistent upward trajectory, characterized by steady and gradual growth without pronounced fluctuations. At the start of this period, the total SME loan balance stood at approximately 9.9 trillion yuan in the first quarter of 2019. By the fourth quarter of 2022, it had surged dramatically to around 23.6 trillion yuan, representing a substantial increase of roughly 13.7 trillion yuan over three years. This remarkable expansion underscores the persistent and robust demand for financing among SMEs in China. As these enterprises continue to develop and expand, their increasing need for credit will play a pivotal role in driving national economic growth and fostering industrial transformation and upgrading.

Despite the strong growth momentum and vibrant development of SMEs, a significant mismatch persists between the funding needs of these enterprises and the availability of bank loans, highlighting a persistent financing gap. Banks, while remaining the primary source of credit for SMEs, have struggled to fully meet their growing financial demands. Although alternative financing mechanisms, such as private equity, venture capital, and supply chain financing, offer additional funding avenues, they are often accompanied by high costs, limited accessibility, and various constraints, which prevent them from fully addressing the financing challenges faced by SMEs. Consequently, many SMEs remain underfunded, and the actual loans they secure are frequently insufficient to support their operational and expansion needs.

Moreover, access to capital markets through instruments such as stocks and bonds presents additional hurdles for SMEs. These challenges include stringent qualification requirements, complex issuance procedures, and difficulties with sales and distribution. As a result, many SMEs find it challenging to leverage these channels for financing. Given these limitations, bank loans remain the most practical, reliable, and efficient financial resource for SMEs. They offer comparatively lower costs, fewer risks, and higher accessibility, making them indispensable for securing working capital and investment funds.

In conclusion, while the demand for SME financing continues to rise in tandem with their growth and innovation, addressing the structural constraints of the current financial system is critical. Enhancing the availability of tailored bank loans, coupled with reforms to diversify and improve alternative financing channels, will be essential to fostering a more supportive financial environment. This will not only alleviate funding shortages but also strengthen the long-term sustainability of SMEs, ultimately contributing to economic resilience and industrial upgrading in China.

1.2.4 Current situation of credit risks for SMEs in commercial banks

(1) Overall situation of non-performing loans

Non-performing loans (NPLs) refer to loans issued by financial institutions, such as commercial banks, that are at risk of becoming irrecoverable due to the borrower's inability to fulfill repayment obligations or pay interest as originally agreed. These loans are typically categorized into five distinct stages based on factors like the borrower’s creditworthiness, the state of their business operations, and the loan's repayment status. These categories are: normal loans, special attention loans, substandard loans, doubtful loans, and loss loans. To gain a deeper understanding of the current credit risk landscape in Chinese commercial banks, this study compiled data on NPLs from 2018 to 2022, which is presented in Table 1. This analysis aims to explore the trends and implications of NPLs over the specified period, providing insights into the evolving credit risk environment within the sector.

Table 1 Statistical Table of Non-Performing Loans of Commercial Banks from 2018 to 2022 (Unit: 100 million yuan)

Project/Year	2018	2019	2020	2021	2022
Total loans	1306535	436405	563129	721118	888714
Balance of non-performing loans	77889	91731	108850	112595	118403
Subordinated loans	31949	37269	50790	51642	55316
Doubtful	34530	40058	42460	44396	45859
loss loan	11410	14404	15600	16557	17228

（**Data source:** State Administration of Financial Supervision and Administration of China）

According to Table 1, since 2018, the non-performing loans of China's commercial banks have been on the rise year by year. Especially in 2020, due to the impact of the COVID-19, the business environment of enterprises is bad, the operating income has declined significantly, and even some enterprises cannot operate normally, leading many enterprises to face the dilemma of tight cash flow, and it is difficult to repay loans on time. This is also an important reason for the increase of approximately 1.7 trillion yuan in non-performing loans in 2020 compared to 2019. The increase in loan growth in China indicates that commercial banks are facing increasingly high credit risks. Not only are there more and more defaulting enterprises, but it also indicates that there are obvious deficiencies in the risk management level of commercial banks.

(2) Analysis of credit risks for small and medium-sized enterprises

With the continuous expansion in both the scale and number of small and medium-sized enterprises (SMEs) in China, the growth in financing available to these businesses has not kept pace with their development, creating a significant and persistent mismatch between credit supply and demand. This imbalance has become a central issue in the lending relationship between commercial banks and SMEs, posing critical challenges for financial institutions aiming to balance profitability with risk management. Addressing this discrepancy requires a comprehensive analysis of the credit risk associated with lending to SMEs by commercial banks, as these businesses inherently carry unique financial vulnerabilities and operational uncertainties.

Among China's financial institutions, the six major state-owned banks—known for their vast financial resources, extensive networks, and skilled manpower—play a pivotal role in the nation's financial ecosystem. These banks represent essential channels for providing funding to SMEs. Enhancing their commitment to SME credit allocation would have far-reaching positive effects, including bolstering SMEs' access to financing, strengthening their capacity for innovation, boosting competitiveness, and ultimately accelerating the transformation and upgrading of the real economy.

As shown in Table 2, China Construction Bank leads the market in terms of inclusive small and micro enterprise loan volume, with a total loan balance exceeding 2 trillion yuan by the end of 2022. This remarkable performance places it ahead of other state-owned banks, including Agricultural Bank of China, Industrial and Commercial Bank of China, Bank of China, and Postal Savings Bank of China, which also maintain significant but comparatively smaller loan portfolios. In contrast, institutions such as Bank of Communications and China Merchants Bank exhibit lower volumes of loans to small and micro enterprises, reflecting variations in their strategic focus and lending priorities within the SME financing sector.

The disparities in loan distribution among these banks highlight the critical need for more balanced and proactive credit policies tailored to SMEs. If major financial institutions increase their lending to this sector, it could significantly enhance the financial resilience and operational growth of SMEs. Additionally, adopting innovative risk assessment models and leveraging financial technology to streamline credit processes would further reduce the barriers to financing. These measures would not only alleviate the long-standing financing difficulties of SMEs but also foster a more dynamic, sustainable, and competitive economic environment by ensuring that financial resources are more effectively aligned with the real economy's development needs.

Table 2 Statistical Table of Inclusive Small and Micro Enterprise Loans from Major Commercial Banks in China (Unit: 100 million yuan)

Project/Year	2018	2019	2020	2021	2022
China Construction Bank	6310.17	9631.55	14523.55	19022.99	23799.75
Agricultural Bank of China	4937.00	5923.00	9615.20	13219.62	17689.94
Industrial and Commercial Bank of China	3216.85	4715.21	7452.27	10990.12	15503.16
Bank of China	3042.00	4129.00	6117.00	8815.00	12283.00

Project/Year	2018	2019	2020	2021	2022
Postal Bank of China	5400.00	6531.85	8012.47	9606.02	11818.94
China Merchants Bank	3931.75	4533.29	5306.50	6011.00	6783.49
Bank of Communications of China	1081.33	1639.52	2607.53	3388.19	4562.39

（Data source: official websites of major commercial banks）

The credit risk faced by small and medium-sized enterprises (SMEs) encompasses a broad range of risk categories, including market risk, operational risk, credit risk, liquidity risk, and legal risk. These risks are not isolated but interconnected, creating a complex web of challenges that require financial institutions and commercial banks to adopt a holistic approach to risk management in order to mitigate the overall credit risk associated with SME lending effectively.

Firstly, market risk arises from fluctuations in the external business environment, which can lead to unstable operations and financial instability. Factors contributing to market risk include shifting consumer demand, intensified competition, and unpredictable price changes. Unlike large enterprises with diversified resources, SMEs typically operate with limited financial buffers. Consequently, any adverse market shift can result in substantial financial strain, directly increasing their credit risk exposure. This vulnerability makes it imperative for SMEs to develop agile strategies and for lenders to carefully evaluate market conditions when assessing creditworthiness.

Secondly, operational risk stems from internal inefficiencies, poor management practices, and errors in daily business operations. SMEs, due to their smaller management teams and limited organizational structures, often experience weaker control over critical functions, including workflow management, quality assurance, and risk monitoring. Failure to adapt to evolving market demands or identify key operational vulnerabilities exacerbates this risk. Financial institutions must consider these factors when designing tailored credit risk assessment models that account for the operational fragility of SMEs.

Thirdly, credit risk, perhaps the most direct and significant category, involves the borrower’s inability to repay the principal and interest on loans as agreed. SMEs, constrained by limited revenue streams and scarce financing options, often struggle to maintain stable cash flow. Factors such as funding shortages, declining profitability, or insolvency frequently lead to loan default. Credit institutions need to implement robust evaluation frameworks that account for the cyclical nature of SME revenues and proactively manage credit exposure.

Next, liquidity risk refers to a borrower's potential inability to meet short-term financial obligations due to an insufficient cash reserve. SMEs are particularly susceptible to liquidity crises due to their narrow capital bases and restricted access to diverse financing channels. A sudden disruption in the cash flow pipeline, such as delayed receivables or unexpected expenses, can trigger severe financial distress, potentially culminating in bankruptcy or forced liquidation. Financial institutions must prioritize liquidity risk analysis and offer flexible lending terms or revolving credit lines to mitigate this threat.

Finally, legal risk involves exposure to penalties, losses, or reputational damage due to non-compliance with laws, regulations, or contractual obligations. SMEs, with their limited legal expertise and smaller compliance infrastructures, often underestimate or misunderstand the complexities of regulatory frameworks and contract stipulations. When legal disputes arise, the repercussions can be disproportionately severe, affecting both the borrowing enterprise and the lending institution. Providing SMEs with legal guidance and incorporating legal risk assessments into credit evaluation practices can help reduce this exposure.

In summary, SMEs face a multifaceted credit risk landscape where various types of risks—market, operational, credit, liquidity, and legal—intertwine to shape their financial stability. Commercial banks and financial institutions must adopt comprehensive risk management strategies that not only address individual risk components but also consider their interdependencies. Enhancing risk prediction models with sector-specific insights, offering customized financial products, and integrating advanced financial technologies can improve the resilience of SMEs while safeguarding the interests of lenders. This approach will contribute to a more sustainable and robust financial environment, fostering growth and innovation among SMEs while minimizing systemic financial risks.

1.3 Problem Statement

(1) Lack of scientific in evaluation data

Small and medium-sized enterprises are important economic entities in China, but compared to large enterprises, their financial management capabilities and financial data openness are relatively low. Many small and medium-sized enterprises have almost no complete and accurate financial data records and reports, which directly affects the assessment of their credit risk by banks.

(a) Incomplete financial data for small and medium-sized enterprises

Small and medium-sized enterprises (SMEs) often face limitations in terms of resources and capabilities, which can result in deficiencies in their financial management practices. For instance, these businesses may not prepare financial statements in accordance with internationally recognized accounting standards, and their financial records may be incomplete or not regularly updated. Such inconsistencies make it challenging for banks to fully assess the financial health and operational performance of SMEs. As a result, banks are unable to acquire a clear and accurate picture of the businesses' financial status, which hinders their ability to make informed, data-driven decisions. The lack of reliable and comprehensive financial information further complicates the process of evaluating the creditworthiness of SMEs, making it difficult for financial institutions to assess risks with precision.

(b) Inaccurate financial data for small and medium-sized enterprises

Many small and medium-sized enterprises may encounter problems such as carelessness and lack of professional knowledge in financial management, such as inaccurate accounting processing and non-standard accounting records. This situation will affect commercial banks' understanding of accurate financial information of enterprises, leading to erroneous judgments in credit evaluation, and thus having a significant impact on whether banks can provide credit services such as loans to enterprises, thereby affecting the financing activities of enterprises. Due to the incompleteness and inaccuracy of financial data for small and medium-sized enterprises, it greatly increases the difficulty for banks to assess their credit risk and also weakens the acquisition rate of bank loans for small and medium-sized enterprises.

(c) Many small and medium-sized enterprises may even have false financial statement data

Firstly, if the financial statements of small and medium-sized enterprises contain situations such as concealing liabilities, exaggerating assets or income, it will cause commercial banks to lose accuracy in conducting credit evaluations, and cannot reflect the true financial status and risk level of the enterprise. Secondly, false financial statements can provide commercial banks with incorrect operational data, making it difficult for them to accurately assess key indicators such as a company's profitability, debt repayment ability, and cash flow status. As a result, they cannot predict the company's future repayment ability and operational stability, increasing the risk of credit defaults and losses. Finally, false financial data may lead to deviations in the risk management of small and medium-sized enterprises by commercial banks. If banks make incorrect assessments of enterprise risks based on false data, it may lead to excessive lending, exceeding credit limits, and increasing the risk of credit default and fund loss. This will bring important risks and adjustments to the financial stability and healthy operation of commercial banks.

(2) The singularity of evaluation methods

At present, commercial banks often use traditional qualitative and quantitative analysis for credit risk assessment of small and medium-sized enterprises, lacking more scientific and systematic methods and tools. For example, many commercial banks rely on financial data, industry conditions, and business experience to evaluate the credit risk of small and medium-sized enterprises. This method often lacks objectivity and scientificity, making it difficult to improve accurate and comprehensive risk prediction.

(a) Limitations of financial indicators

Financial indicators mainly reflect the past operating conditions of a company and cannot comprehensively predict and measure its future performance. Financial indicators such as net profit and debt to asset ratio can only provide relevant local information about the enterprise, and cannot fully understand key indicators such as profitability, solvency, and cash flow status of the enterprise. In addition, financial indicators are also easily affected by corporate financial fraud, false financial statements, etc., which will further reduce the accuracy of credit risk assessment to a certain extent.

(b) Neglecting the impact of non-financial factors

The operational risks of small and medium-sized enterprises are influenced by various factors, such as market competitiveness, industry development trends, business models, and management capabilities. However, traditional financial indicator evaluation methods only focus on the financial condition of the enterprise, ignoring the impact of these non-financial factors. For example, a company may have a good financial condition, but due to fierce industry competition or improper business strategies, there is still a high level of credit risk.

(c) Blindly applying credit risk assessment methods of large enterprises to small and medium-sized enterprises

The credit risk assessment model of commercial banks in China still focuses on large enterprises as the main research object, and lacks targeted applications for small and medium-sized enterprises. Small and medium-sized enterprises (SMEs) are the group with the highest number, employment, and activity in the market, playing a crucial role in economic development, employment, and entrepreneurship. However, due to the differences in business management, market size, and credit records between small and medium-sized enterprises and large enterprises, credit risk assessment models based on large enterprises cannot be fully applicable to small and medium-sized enterprises.

(3) Limitations of the evaluation process

The credit evaluation process for small and medium-sized enterprises (SMEs) conducted by commercial banks is intricate and time-consuming, typically encompassing several key stages: information collection and preliminary screening, credit assessment and loan demand analysis, risk evaluation and investigation, condition review and loan approval, as well as continuous monitoring and risk management. Given the complexity and length of this process, multiple challenges and inefficiencies can arise, affecting the accuracy and effectiveness of credit risk assessment for SMEs.

(a) Insufficient Technological Support for Evaluation Tools

One major limitation is the lack of advanced technological infrastructure to support automated credit evaluation. Currently, many commercial banks still rely heavily on manual operations for processing SME credit applications. In the absence of intelligent, data-driven risk assessment platforms, credit officers must invest substantial time and effort into evaluating risk. This manual approach makes the process cumbersome, inefficient, and prone to human error. The potential for inaccuracies, omissions, and inconsistencies not only depletes valuable human resources but also undermines the reliability of credit decisions, increasing the likelihood of misjudging an enterprise’s actual creditworthiness. Adopting technology-driven tools and artificial intelligence-based systems could significantly improve the speed, accuracy, and scalability of credit evaluations.

(b) Incomplete evaluation indicators

A prevalent issue in SME credit risk assessment is the overemphasis on financial indicators while other critical factors are insufficiently considered. Commercial banks often focus primarily on metrics such as revenue, profitability, and debt ratios, while neglecting qualitative aspects of business operations. Important dimensions like the enterprise's business model, market positioning, competitive advantage, innovation capabilities, and the quality of management receive inadequate attention. This narrow focus results in incomplete evaluations that fail to capture the holistic risk profile of SMEs. A more comprehensive evaluation framework that integrates both financial and non-financial factors would provide a clearer and more accurate assessment of credit risk.

(c) High Subjectivity and Individual Discrepancies in Evaluation Results

Another significant drawback is the heavy reliance on subjective judgment by credit officers, which introduces variability and inconsistency in assessment outcomes. Since credit evaluations are influenced by the experience, professional expertise, and personal judgment of individual credit staff, discrepancies often emerge in risk assessments of the same enterprise by different personnel. Furthermore, biases, attitudes, and emotional factors can inadvertently affect credit decisions, leading to inconsistency and unpredictability. This subjectivity compromises the objectivity of credit risk analysis and increases the likelihood of unfair or inaccurate evaluations. Standardizing assessment criteria, enhancing staff training, and implementing data-driven decision-making models can help mitigate these issues, promoting greater fairness and consistency in credit evaluations.

1.4 Research Questions

From the previous discussion, this study has the following research questions:

Question **1:** Initially, various business and financial indicators for each enterprise are extracted from the transaction bill data in the provided annexes. These indicators are then carefully filtered for relevance. Subsequently, this study enhances the traditional decision tree model by incorporating a regularization term, which helps to control the complexity of the decision tree. To further improve prediction accuracy, an ensemble learning approach is employed, wherein multiple weak decision tree classifiers are stacked together. This model is then optimized iteratively using a gradient descent algorithm, ultimately resulting in a strong classifier.

Based on a supervised learning framework, a credit risk assessment model is developed, with the objective function focused on calculating the default probability for each enterprise. This model is applied to determine the default probability of the enterprises listed in Annex 1, offering insights into their respective credit risks. The next step involves training the model using the data from Annex 1 (enterprises with existing credit records) and then applying the trained model to rate the enterprises in Annex 2 (enterprises without prior credit records).

Finally, the developed credit risk assessment and credit strategy models are used to evaluate the credit risk of each enterprise. The resulting analysis informs the bank's optimal credit strategy, maximizing the expected return for the bank. The model ultimately provides a detailed strategy for managing credit risk and determining the appropriate credit allocation.

Question **2:** The financial performance and operational stability of enterprises may be influenced by unforeseen external factors, which can vary in their impact across different industries and business categories. These unexpected events, such as the COVID-19 pandemic, may have significantly different effects on enterprises, depending on their industry and sector. In this context, the credit risk of the enterprises in Annex 2 (those without prior credit records) is reassessed, taking into account the potential disruptions caused by such sudden events. Given a total annual credit allocation of 100 million yuan, this research proposes an adjusted credit strategy for the bank. The strategy considers both the default risk of these enterprises and the anticipated effects of external factors, ensuring that the bank's credit decisions are dynamically aligned with the broader economic environment.

1.5 Research Objectives

(1) The first step involves establishing a comprehensive risk evaluation index system tailored specifically for small and medium-sized enterprises (SMEs). This system aims to capture the unique financial and operational risks faced by SMEs. Based on this system, appropriate models are selected to effectively quantify and assess the various risk factors that may influence the stability and creditworthiness of these enterprises.

(2) Utilizing statistical data, a robust mathematical model is then developed to inform the bank's credit strategy decisions. This model assists in determining key parameters such as loan limits, interest rates, and loan terms. The goal is to strike a balanced approach that maximizes the bank's profitability while minimizing associated risks. Through this process, the model ensures a careful alignment between risk management and return optimization, ultimately fostering sound credit practices.

(3) Recognizing that external and unexpected events, such as the COVID-19 pandemic, can significantly affect the production, operation, and overall financial health of businesses, it becomes crucial to account for these unpredictable disruptions in credit risk assessments. In this regard, the credit risks across different industries and business categories are thoroughly analyzed, alongside the development of corresponding strategies for mitigating these risks. This approach ensures that banks remain agile, able to adjust credit policies according to specific circumstances, and safeguard themselves against potential adverse outcomes.

1.6 Significance of the Study

Through the introduction of the research content in this article, the main significance and contribution of the research work are as follows:

(1) Theoretical significance

This research introduces a novel approach to credit risk assessment for small and micro enterprises (SMEs) by incorporating non-financial indicators. These indicators, specifically innovative information and business models, are crucial for evaluating the creditworthiness of SMEs, which have traditionally been underrepresented in financial assessments. The study proposes advanced models for processing unstructured data, including text feature extraction models, keyword extraction models, and semantic matching techniques, enabling the analysis of textual business indicators and other non-financial variables. Through an extensive review of literature, case studies of SMEs, and the analysis of enterprise information disclosures, four distinct risk characteristics of Chinese small and micro enterprises are identified.

Building upon this, the research introduces a more comprehensive set of non-financial indicators for credit risk evaluation, which includes key elements of innovation and business models. These enterprises play a pivotal role in economic development, but their small scale, lack of collateral, and often irregular financial data present significant challenges for traditional credit risk assessments. Conventional methods typically rely on limited financial data and subjective judgment, leading to inaccurate assessments, prolonged approval times, and inefficiencies. By integrating upstream and downstream invoice data, along with data mining and statistical models, this research enables the incorporation of diverse factors such as transaction behavior, operational performance, and supply chain stability into the credit risk evaluation framework. This leads to a more holistic, objective, and precise credit risk assessment model. Ultimately, the study aims to enhance the credit risk assessment system for small and micro enterprises, improving both the accuracy and efficiency of these evaluations.

(2) Practical significance

In its practical application, this research provides a structured framework and process for credit risk evaluation tailored specifically to small and micro enterprises. This framework can be directly applied to real-world scenarios, offering enhanced accuracy in the credit risk assessments of such enterprises. Furthermore, it provides innovative tools for government bodies and financial institutions, equipping them with more effective methods for evaluating credit risk. The research also includes empirical validation using actual data, ensuring the results are directly applicable to data mining practices within industrial sectors and financial institutions. This makes the research highly relevant for improving credit risk management practices within these industries.

In conclusion, the findings from this research significantly contribute to both the theoretical understanding and practical application of credit risk assessment for small and micro enterprises. The theoretical insights into the risk characteristics of these enterprises provide a deeper understanding of their unique financial and operational challenges. On the practical side, the proposed models and frameworks offer actionable solutions for financial institutions, enabling more accurate assessments of credit risk. The outcomes of this study hold substantial practical value for the financial sector, providing an important reference for enhancing enterprise credit risk management strategies.

1.7 Scope of the Study

A bank has allocated a loan fund of 100 million yuan to provide financing to enterprises, with interest rates ranging between 4% and 15% annually. The loan term is set at one year. To develop an effective credit strategy, the bank analyzes data from 123 enterprises that have established credit histories, alongside 302 enterprises lacking such records. Additionally, the bank incorporates 2019 statistical data that illustrates the relationship between loan interest rates and customer churn rates, as well as the actual data provided in the annexes.

To study the optimal credit strategy for small and medium-sized enterprises (SMEs), a mathematical model is developed. The primary objective of this model is to quantify the credit risk associated with the 123 enterprises with credit histories. This allows the bank to determine a tailored credit strategy for these businesses, considering the fixed annual total credit allocation. Through this analysis, the bank can more accurately define the optimal loan terms, including the appropriate loan limits and interest rates for each enterprise.

Building on this foundation, the model is then extended to analyze the 302 enterprises without prior credit records. By assessing the credit risk of these companies, the bank can establish a corresponding credit strategy under the same total annual credit cap of 100 million yuan. This process ensures that both groups of enterprises—those with established credit histories and those without—are considered within the bank's overall credit framework.

In addition, the study incorporates the potential impact of unforeseen external factors, such as the COVID-19 pandemic, which could significantly disrupt the production, operations, and financial stability of enterprises. The effect of such disruptions can vary across different industries and types of businesses. Therefore, the credit risk of the 302 enterprises is reassessed with these potential external factors in mind. By factoring in these risks, the bank is able to adjust its credit strategy to better account for such uncertainties.

Ultimately, the study provides a comprehensive revised credit adjustment strategy that ensures the bank optimally allocates the 100-million-yuan annual credit limit. This strategy accounts for both internal credit risks and external disruptions, allowing the bank to mitigate potential financial risks while supporting the growth of SMEs.

1.8 Structure of the Proposal

Chapter1: This opening chapter provides an overview of the research, introducing the background and context of the study. It outlines the research purpose and objectives, explaining the significance and relevance of the study in the current financial landscape. The chapter also highlights the challenges faced by small and medium-sized enterprises (SMEs) in securing financing and the need for effective credit risk management strategies.

Chapter 2: This chapter presents a comprehensive review of existing literature related to the credit risk of SMEs. It explores various definitions and frameworks of credit risk, with a focus on the specific challenges SMEs face in obtaining credit. Additionally, it discusses relevant statistical methods and theories used to assess credit risk, highlighting the key findings from previous research. The chapter concludes by identifying the research gaps in current literature, establishing the foundation for the methodology employed in this study.

Chapter 3: In this chapter, the research methodology is discussed in detail. It includes an overview of the data sources used in the study, the variables considered, and the steps involved in applying the methodology. The chapter also presents the workflow chart, outlining the process followed to construct the credit risk model and derive actionable insights. This section provides clarity on how the research was conducted and the rationale behind the chosen methods.

Chapter 4: This chapter presents the results of the research, including the development and quantification of the credit risk indicator system. It provides an in-depth analysis of the credit strategies based on the risk profiles of SMEs. Furthermore, the chapter examines how unexpected factors, such as economic disruptions or global crises, can influence credit strategies and the associated risk assessments. The discussion draws on the findings to propose effective strategies that balance risk and return for both SMEs and financial institutions.

Chapter 5: The final chapter discusses the conclusions drawn from the research. It synthesizes the key findings and presents strategic recommendations from multiple perspectives to help SMEs navigate the challenges of obtaining bank loans. The chapter emphasizes how the proposed strategies can reduce credit risk for banks while facilitating smoother access to financing for small and medium-sized enterprises. It also suggests areas for future research and potential improvements in credit risk management practices.

Chapter 2
LITERATURE REVIEW

This chapter provides an in-depth review of existing literature on SME credit risk. It examines different definitions and conceptual frameworks, emphasizing the unique challenges SMEs encounter when seeking credit. Additionally, it explores key statistical methods and theoretical approaches used for credit risk assessment, summarizing significant findings from prior research. The chapter concludes by identifying gaps in the current literature, laying the groundwork for the study's research methodology.

2.1 Definition of SMEs Credit Risk

The study of credit risk among small and medium-sized enterprises (SMEs) has garnered significant scholarly attention, driven by the need to address both economic and non-economic factors that influence their creditworthiness. Methodologically, while economic factors are easier to quantify and dominate much of the research, challenges persist in assessing the more complex and less tangible elements, such as interest rate risk causality and the impact of social barriers. These difficulties often hinder a thorough and nuanced understanding of SME credit risk. For example, Belás et al. (2018) conducted an empirical analysis of 352 firms operating in the Czech business environment, utilizing structural equation modeling (SEM) as the primary analytical tool. Their findings underscore that the management of credit risk in SMEs is influenced by a range of non-economic factors, with a particular emphasis on education and family background. These social dimensions, alongside more traditional economic factors like banking relationships, capital, and financial literacy, significantly shape the effectiveness of credit risk management strategies. Specifically, the study revealed that payment discipline and the ability to navigate financial systems are crucial in managing credit risk. In another significant study, Fernandes and Artes (2016) focused on Brazilian SMEs, investigating both their geographical location and the default risk associated with their operations. They applied a localized default risk metric, utilizing the ordinary kriging method in combination with a logistics credit scoring model. Their research demonstrated that the incorporation of a localized risk variable could greatly enhance the accuracy and performance of credit scoring models, offering valuable empirical evidence to support the refinement of SME credit risk assessment tools. Zhao et al. (2022) shifted the focus to SMEs involved in supply chain finance (SCF), proposing a new credit risk evaluation index system tailored to this context. By selecting a broad range of indicators that account for both qualitative and quantitative aspects of SMEs, the study established a comprehensive evaluation system. This model was designed to effectively manage and mitigate risks for banks involved in supply chain finance, offering a more holistic approach to credit risk assessment in this specialized domain. Gonçalves (2016) took an innovative approach by employing cognitive mapping techniques alongside the interactive multi-criteria decision-making (TODIM) method. This approach led to the development of a decision support system aimed at improving the credit evaluation process for SMEs. The model not only enhances the efficiency of credit decision-making but also introduces a more transparent and informed risk assessment, making it easier for financial institutions to evaluate SME credit applications with greater accuracy. Finally, considering the essential role of SMEs in national economies and their significant inclusion in the New Basel Capital Accord, Altman and Sabato (2007) devised a distress prediction model specifically designed for SMEs. This one-year default prediction model, built upon financial indicators unique to SMEs, was shown to outperform generic corporate models in terms of predictive accuracy. The study highlighted the importance of selecting the most relevant financial variables, underscoring the need for models that are specifically tailored to the distinctive characteristics of SMEs.

Together, these studies provide a multi-dimensional and multi-faceted approach to credit risk management for SMEs, thus laying a solid theoretical foundation for the introduction of dependency models for important social and economic factors.

2.2 Statistical Methods Theory

2.2.1 Decision Tree Algorithm

Decision tree is a machine learning algorithm based on a tree structure for classification and regression problems. The basic idea is to construct a tree structure by dividing the data layer by layer so that each leaf node corresponds to a category (classification problem) or a value (regression problem). One of the proposers of decision trees is Ross Quinlan, who proposed the ID3 (Iterative Dichotomise 3) algorithm in 1986. The theoretical basis of decision trees is derived from the concepts of Entropy and Information Gain in Information Theory. The goal of a decision tree is to reduce the purity (or uncertainty) of the target variable in each partitioned subset by selecting appropriate features to partition the data. Information Gain measures the reduction in entropy that results from the use of a particular feature for partitioning, and the feature with the greatest information gain is selected for partitioning. This usually involves a partitioning criterion, one common one being Entropy or Gini Impurity. The information gain is $Entropy(S) = -\sum_{i=1}^{c} p_i \, log_2(p_i)$, where S is the dataset of the current node, c is the number of categories, and p_i is the percentage of the ith category in the node. Gini impurity:$Gini(S) = 1 - \sum_{i=1}^{c} p_i^2$, where S is the dataset of the current node, c is the number of categories, and p_i is the percentage of the ith category in the node.

Decision trees are usually generated using a recursive approach. The optimal features and division points are selected at each node, and then the same operation is performed recursively on the child nodes until the stopping conditions are met (e.g., maximum depth is reached, the number of samples in the node is less than a certain threshold, etc.).

The advantage is that it is easy to understand and interpret, presenting an intuitive tree structure. It does not require much preprocessing of the data, and is able to handle numerical and sub-typical features. The disadvantage is that it is easy to overfitting, especially when dealing with data with complex relationships. Sensitive to outliers, which may lead to instability. May produce highly complex trees, leading to overfitting of training data. Decision trees are an intuitive and easy to understand machine learning algorithm for classification and regression problems. It generates a tree structure by recursively selecting the optimal features and division points, but requires attention to overfitting and sensitivity to outliers. Common decision tree algorithms include ID3, C4.5, and CART. Each algorithm has unique strengths, making them suitable for different types of datasets and tasks. For example, C4.5 is often preferred for handling mixed-type data, while CART is widely used for its simplicity and effectiveness in binary splits.

2.2.2 AdaBoost Algorithm

AdaBoost, short for Adaptive Boosting, is a meta-algorithm used in statistical classification. It was developed by Yoav Freund and Robert Shapira in 1995, a pioneering contribution that led to them receiving the prestigious Gödel Prize in 2003. The primary objective of AdaBoost is to improve the performance of weaker learning algorithms, often referred to as 'weak learners,' by combining the results of these learners in a weighted manner to produce a final, robust prediction. This final prediction is based on the weighted sum of the individual learners' outputs, resulting in a more accurate and reliable classifier.

Although AdaBoost is most commonly used in binary classification tasks, its versatility allows it to be extended to handle multiclass classification problems or even predict continuous values within a bounded range on the real number line. The key feature that makes AdaBoost adaptive is its ability to focus more on the instances that were misclassified by previous weak learners. This adaptation process allows AdaBoost to iteratively refine the model by emphasizing harder-to-classify data points, which helps to correct errors in the earlier stages of learning.

One of the advantages of AdaBoost is its potential to reduce the risk of overfitting compared to other machine learning algorithms, especially when used in appropriate scenarios. While the individual weak learners might not perform particularly well on their own, as long as each learner performs slightly better than random guessing, the overall AdaBoost model can converge into a strong classifier. This characteristic of combining weak learners to form a robust ensemble classifier makes AdaBoost a powerful and efficient tool in various machine learning applications.

A boosted classifier is a classifier of the form $F_T(x) = \sum_{t=1}^{T} f_t(x)$, where each f_t is a weak learner that takes an object x as input and returns a value indicating the class of the object.

AdaBoost achieves high accuracy by iteratively focusing on the misclassified samples and adjusting the weights of both samples and weak learners. This results in a final model that is a weighted combination of multiple weak learners, often outperforming individual strong learners. AdaBoost is highly flexible, allowing for the use of various types of weak learners, such as decision stumps (single-level decision trees), neural networks, or other classifiers. This flexibility enables AdaBoost to be adapted to different data types and problems. AdaBoost inherently performs feature selection by focusing on those features that contribute most to the correct classification of samples. This results in the algorithm ignoring irrelevant or noisy features, improving the robustness of the model. Despite its iterative and complex nature, AdaBoost often exhibits good resistance to overfitting. This is because it focuses on improving the performance on difficult samples, rather than simply memorizing the training data. Compared to other complex algorithms, AdaBoost is relatively easy to implement and can achieve good results with default parameters. This algorithm also has some disadvantages, such as: Training AdaBoost can be computationally expensive, especially when dealing with large datasets or complex weak learners. The sequential nature of AdaBoost's training process, where each weak learner is trained based on the performance of the previous one, contributes to this. The optimal number of weak learners to use in AdaBoost is not always clear and may require some experimentation or cross-validation to determine. Too few weak learners may result in underfitting, while too many can lead to overfitting or unnecessary computational overhead. AdaBoost's performance can be affected by imbalanced datasets, where one class is significantly more prevalent than the others. The algorithm may focus too much on the majority class, leading to poor performance on the minority class.

2.2.3 Gradient Descent Method

The Gradient Descent method (Gradient Descent) is an optimisation algorithm for minimising (or maximising) the objective function. The basic idea of this method is to adjust the model parameters step by step by iterating along the direction of the gradient of the objective function (or in the opposite direction), so that the objective function reaches the minimum (or maximum) value. The formulation of the gradient descent method dates back to a very early time, but one of the major contributors was Augustin-Louis Cauchy, a 19th century French mathematician who proposed a preliminary form of the gradient descent method in 1847. However, the development and application of the gradient descent method gradually became a central method in the field of optimisation in the following decades. The theoretical background of the gradient descent method is built on the foundations of calculus. The process is analogous to the process of choosing the direction of descent when descending a mountain based on information about the slope of the current position. Specifically, the model update formula of the gradient descent method is as follows as $\theta_{ij} = \theta_j - \alpha \cdot \nabla J(\theta_j)$, where θ_j denotes the model parameter of the iteration, α is the learning rate, and $\nabla J(\theta_j)$ is the gradient of the objective function J with respect to the model parameter θ_j. This formula indicates that for each iteration, the model parameters are updated by subtracting the current model parameters from the learning rate multiplied by the gradient, which makes the objective function decrease gradually.

The advantage is that it is simple and easy to implement and is a general optimisation method. It performs well on large-scale datasets and can handle high-dimensional data. The disadvantages are that it may fall into local minima and has limitations for non-convex functions. The choice of learning rate has a large impact on the performance of the algorithm, too large or too small may lead to performance degradation. Sensitive to the initial value, improper selection of the initial value may affect the convergence of the algorithm. Gradient descent is a commonly used optimisation algorithm to minimise the objective function by iteratively adjusting the model parameters. Despite some drawbacks, it is still widely used in practical applications. When using the gradient descent method, care needs to be taken to adjust parameters such as the learning rate and initial value to obtain better performance.

2.2.4 Nonlinear Programming

Nonlinear programming (NLP) is a mathematical approach for solving optimization problems where either the objective function or some of the constraints are nonlinear (Kuhn, 1991). In essence, NLP involves the process of finding the extrema—such as maximum, minimum, or stationary points—of an objective function. This is done by considering a set of unknown real variables, subject to a system of constraints, which can be both equalities and inequalities. These constraints form the boundaries within which the optimization process occurs.

Nonlinear programming (NLP), a subfield of mathematical optimization, focuses on solving problems in which the relationships between the decision variables, the objective function, and/or the constraints are nonlinear. Unlike linear programming, which deals with optimization problems where both the objective function and constraints are linear, NLP is designed to handle more complex scenarios. These scenarios involve situations where the interactions between variables are not confined to straight-line relationships. As a result, nonlinear programming opens the door to a wider range of optimization problems, many of which would be difficult or even impossible to solve using linear methods. By accommodating these more intricate relationships, NLP enables the exploration of optimization landscapes that are more diverse and challenging, making it a crucial tool for addressing real-world problems in fields like engineering, economics, and logistics.

Let $n, m,$ and p be positive integers. Let X be a subset of Rn (usually a box-constrained one), let f, gi, and h_j be real-valued functions on X for each i in $\{1, \dots, m\}$ and each j in $\{1, \dots, p\}$, with at least one of f, g_i, and h_j being nonlinear.

A nonlinear programming problem is an optimization problem of the form

Minimize $f(x)$

$$\text{Subject to } \begin{array}{c} g_i(x) \leq 0 \;\; for\; each\; i \in \{1, \cdots, m\} \\ h_j(x) = 0 \;\; for\; each\; j \in \{1, \cdots, p\} \\ x \in X \end{array}$$

Nonlinear programming encompasses diverse applications, notably in engineering design, control systems, data fitting, and economic planning. These applications frequently exhibit common characteristics in their problem structures, which render convex optimization algorithms highly efficient. Comprehending both the nature of these attributes and the manner in which algorithms interpret the problems can significantly facilitate optimization tasks, ranging from problem formulation to the selection of the most suitable solution method.

2.3 Review of Research

2.3.1 Various risk assessment statistical models

(1) Support vector machine (SVM) models and related improved models

Support Vector Machines (SVM) and their enhanced versions have garnered significant attention in the realm of credit risk assessment. In a notable study, Zhang et al. (2014) introduced an innovative multi-criteria optimization classifier, KFP-MCOC, which integrates kernel methods, defuzzification, and penalty factors. This classifier demonstrated superior performance over traditional SVMs and fuzzy SVMs, particularly in terms of separation, efficiency, and generalization based on empirical testing. Similarly, Harris (2013) developed a credit scoring model specifically tailored for the Barbados Credit Union. The model showed that using quantitative credit risk models substantially improved credit risk evaluations compared to traditional judgment-based methods.

Wang and Ma (2012) proposed a hybrid integration model known as RSB-SVM, which combines bagging and stochastic subspaces with SVM as the core learning mechanism. This model, validated using financial data from 239 companies in ICBC, proved to be an effective alternative for corporate credit risk assessment. Zhou et al. (2011) further advanced SVM-based approaches by introducing KASNP, a method designed to extend the nearest point problem of SVM to affine subspaces, providing effective solutions for non-linear classification tasks.

In their quest to enhance SVM performance, Yu et al. (2011) put forward the concept of a weighted least squares support vector machine (LSSVM) classifier, showcasing promising results in credit risk classification. Danenas and Garsva (2012) employed a hybrid approach that combined genetic algorithms and particle swarm optimization to evolve the parameters for a linear SVM classifier, applying this to credit risk assessment using a sliding window methodology. Additionally, addressing the financing challenges faced by micro, small, and medium-sized enterprises (MSMEs), Yao et al. (2023) enhanced the classical Slime Mold algorithm by integrating it with SVM for parameter optimization, resulting in the RF-LSMA-SVM model, which improved the classification performance in credit risk assessments.

Tang et al. (2021) introduced a cost-sensitive SVM (CSSVM) and a robust loss method, Blinex loss-sensitive kernel (CSKB), demonstrating superior performance across various evaluation metrics when tested on synthetic datasets. These innovations reflect the growing diversity of methodologies applied in credit risk assessment, contributing to more robust and reliable models.

(2) The KMV model in credit rating

The KMV model has significantly impacted credit rating practices. Kealhofer (2003) highlighted that the Expected Default Frequency (EDF) generated by the KMV model consistently outperformed agency debt ratings in predicting defaults, as demonstrated by power tests. The model's default probabilities, grounded in an option-theoretic framework, provided a clearer explanation of debt spreads when compared with other models, notably the generalized Merton methodology.

Katchova and Barry (2005) applied the KMV model to estimate capital requirements for agricultural lenders under the New Basel Capital Accord. Their findings, which compared Credit Metrics and Moody's KMV models, revealed notable differences in the capital requirements of agricultural lenders depending on portfolio risk and granularity. Guo and Zhou (2022) employed the KMV model to analyze the default distances of 36 Chinese banks from 2016 to 2019. Their empirical analysis demonstrated that while external factors like financial regulatory policies did impact credit risk, the internal factors, such as income structure and risk management practices, played a more substantial role. This underscores the need for banks to optimize their internal structures to mitigate credit risk.

(3) Logistic Model and neural network algorithm in Credit Rating

Logistic regression models have been extensively utilized in credit risk prediction. Zhan (2021) utilized enterprise index data to develop a Logistic regression model alongside the risk-adjusted return on capital (RAROC) model, calculating default probabilities to inform credit strategies. Similarly, Sun Yuchen (2021) employed a logistic regression model to evaluate MSME default rates, relying on accessible transaction information, credit ratings, and credit histories. Wang et al. (2022) constructed a personal credit risk assessment model by combining logistic regression with feature selection techniques and XGBoost, using real-world credit data from Bank X. This study aimed to optimize feature selection through three methods—Logistic regression, AIC-Logistic regression, and BIC-Logistic regression—and identified the model that provided the best accuracy and AUC to predict credit risk. Neural network algorithms have also been explored for credit risk assessment. Chandrappa and Gitesh (2009) compared the performance of fuzzy logic models with neural networks and classification tree models, concluding that the latter were more effective in predicting bankruptcy. Hu and Su (2022) designed an artificial neural network model to predict the bankruptcy risk of commercial banks. By selecting 14 financial indicators and integrating cluster and factor analysis, their model demonstrated solid performance in credit risk evaluation. In further developments, Liu, Zhang, and Fan (2022) proposed a hybrid model that first applies XGBoost to linearize the original features, followed by a graph-based neural network (forgeNet) for accurate credit risk predictions. Fan and Qin (2023) enhanced fuzzy neural networks to construct an effective corporate credit risk assessment model, and Huang et al. (2018) compared multiple neural network models in the context of Chinese SMEs, with probabilistic neural networks (PNN) emerging as the top performer in terms of both accuracy and robustness. Wang and Ma (2011) introduced the RS-Boosting method, which combines augmented subspaces and stochastic subspaces to optimize enterprise credit risk prediction. In a similar vein, Zhu et al. (2019) developed the RS-Multibooting method, leveraging multiple stochastic subspaces and Multibooting integrated machine learning to enhance SMEs' credit risk prediction accuracy. Corazza et al. (2021) employed the Elman network for bankruptcy prediction, demonstrating its robustness compared to traditional neural network models. Furthermore, Hayashi (2016) utilized the Re-RX algorithm family for multi-validation, establishing it as a powerful decision support tool for credit risk assessment. Oreski and Oreski (2014) also introduced a hybrid genetic algorithm for neural networks (HGA-NN) for feature selection, which enhanced classification accuracy in credit risk models. Khashman (2010) explored supervised neural networks with back-propagation for credit application decisions, showing significant improvements in decision-making accuracy. Similarly, Khashman (2011) demonstrated that the EmNN outperformed traditional neural networks in decision speed and accuracy. Chaojie Li (2022) integrated random forest and deep neural networks to construct an intelligent decision-making system, assessing credit default probability for MSMEs. Xiangzhou Chen and Lihong Tao (2021) developed an MLP neural network model for credit risk assessment in SME supply chain finance, proving its high accuracy and stability. Oskarsdottir and Bravo (2021) proposed a multilayer network model using personalized PageRank algorithms, quantifying default risk among borrowers. Lastly, Shudai Wang, Zihui Yang, and Pingmiao Zhang (2023) explored the evolution of credit risk contagion in municipal investment bonds using network analysis techniques, revealing the intricate dynamics of debt risk across regions.

(4) Decision tree models and Markov models in credit rating

Decision tree models have been widely explored for their application in credit rating systems due to their ability to handle complex, non-linear relationships and provide interpretable results. Addo et al. (2018) developed a binary classifier using both machine and deep learning models to predict loan default probabilities. Their study highlighted that the decision tree model, when compared to a multi-layer artificial neural network, demonstrated greater stability in prediction, especially when using the top 10 most important features. Arora et al. (2019) examined various classification algorithms enhanced by Bolasso (Bootstrap-based feature selection) in the context of credit ratings, finding that the BS-RF (Random Forest) model outperformed others in terms of accuracy and Area Under the Curve (AUC), significantly improving lender decision-making. Similarly, García-Céspedes and Moreno (2022) studied the application of machine learning techniques to calibrate the outputs of Vasicek's (1987) credit risk model, showing that tree-based models delivered rapid and accurate estimations of actual loss distributions. Cheng Yang et al. (2023) proposed a multi-level risk assessment framework that used bill transaction data to develop a credit scoring model for MSMEs. By employing a random forest algorithm to predict defaults and quantifying credit scores through a weighted average approach, they were able to achieve a more quantitative and effective assessment of MSMEs' credit risk. Alzamora et al. (2022) utilized the LightGBM decision tree model, which delivered high confidence levels, supporting decision-making processes in rural credit risk assessments and effectively reducing delinquency rates. In an effort to improve decision tree stability when dealing with imbalanced datasets, Chang et al. (2016) integrated bootstrap aggregation and synthetic minority oversampling techniques into a credit risk model, leading to more reliable predictions. Luo et al. (2016) enhanced the Cox model by combining it with multinomial logistic regression to improve prediction accuracy, especially for scoring predictions. Wu et al. (2012) developed an enhanced decision support model (EDSM) by merging decision trees with correlation vector machines, which improved the interpretability and reliability of predictions. García et al. (2012) applied filtering algorithms to enhance the accuracy of instance-based classifiers, showing that filtered datasets significantly outperformed unfiltered ones when using nearest-neighbor decision rules. Abdelmoula (2015) employed the K-nearest-neighbor classifier to predict defaults on short-term loans at Tunisian commercial banks, achieving an 88.63% classification rate by selecting optimal information sets related to accruals and cash flows. Liu et al. (2019) introduced an evolutionary multi-objective soft subspace clustering algorithm

(EMOSSC), demonstrating its superiority in credit risk classification. Zhou et al. (2010) explored the nearest subspace classification method, which extends the training set by using the same class of creditors across different subspaces, showing competitive performance in credit risk assessments.

In the domain of credit rating, Markov models have also been extensively studied for modeling credit risk, particularly in the context of rating transitions. D'amico et al. (2014) challenged the use of traditional Markov chain models in credit rating migration problems by introducing a multivariate semi-Markov chain model. This innovative approach accounts for the non-Markovian nature of credit ratings and successfully calculates transfer probabilities, reliability functions, and credit default swap prices. Siu (2005) applied a multivariate Markov chain model for credit transfer matrices, incorporating credibility theory for prior estimation of transfer matrices. This method offered a practical way to estimate dynamic credit rating changes and credit risk measures. Kubo and Sakai (2011) proposed a new model for assessing long-term credit risk, extending beyond stock prices to include factors such as economic cycles. Their model integrated long-term cash flow forecasts and credit risk propagation, providing valuable insights for calculating risk spreads and broadening the applicability of traditional models. Lu et al. (2012) combined data envelopment analysis (DEA) with Markov chain models to assess the credit risk of large Taiwanese corporations, offering banks and financial institutions a more efficient decision-support tool for lending decisions. Wozabal and Hochreiter (2012) enhanced the traditional Markov chain model by considering credit rating shifts and estimating parameters using maximum likelihood and heuristic global optimization techniques. Their model showed stronger risk sensitivity and greater dependency than the Generalized Linear Mixed Model (GLMM), offering a fresh perspective on modeling credit rating transitions.

(5) Loan default probability model

In the context of credit rating, Lang and Santomero (2003) investigated U.S. banks' practices concerning retail portfolio credit, suggesting that direct estimation of portfolio volatility could be a feasible approach to estimating economic capital. However, they highlighted that existing methodologies fail to meet the rigorous regulatory requirements. Consequently, alternative approaches that directly estimate structural risk parameters, such as default probabilities, were deemed more practical. Carey and Hrycay (2001) emphasized the critical importance of estimating the average default probability for borrowers within each internal credit rating class, as these probabilities are key components of portfolio credit risk models. They tested the main methods used to estimate these probabilities and uncovered potential issues, including bias, instability, and gaming, which can compromise the accuracy and reliability of these estimations. Ju and Sohn (2017) introduced a quantitative score-based credit risk rating model, which models loan default probabilities as a function of various loan attributes. Their model aids in distinguishing between different default patterns, offering a novel approach for more accurate credit scoring. Kanno (2020) focused on the credit risk of Japanese real estate investment trusts (J-REITs) across two related markets during 2008-2017. His study identified a significant correlation between the extent of J-REIT lending and a decrease in the credit risk associated with issued J-REITs, thus providing a fresh perspective on understanding REIT credit risk. Lu (2013) employed a parametric model under a risk-neutral measure to assess credit risk, using interpolation to estimate the risk-free yield curve. His empirical findings revealed that financially distressed firms have a higher probability of default, with the model closely linked to broader economic conditions. García-Céspedes and Moreno (2017) proposed a multi-period credit risk model that estimates multi-period loss distributions using a Taylor expansion approximation. They found that the approximation offered greater accuracy for portfolios with higher probabilities of recession or default. Spuchľakova and Cug (2015) put forward a methodology to estimate credit risk parameters based on market prices, including a simplified approach for calculating the loss-given-default (LGD) rate. This methodology is instrumental in pricing credit default swaps (CDS) and helps banks calculate regulatory capital, as well as aiding investors in pricing risky bonds. Finally, Chen Tianxin and Li Junshuai (2021) examined the impact of economic policy uncertainty on bank credit risk, using a default probability model. Their empirical analysis provides valuable insights into the mechanism through which economic policies influence credit risk.

The application of credit default swaps (CDS) in credit ratings has been the focus of several significant studies. Chamizo and Novales (2020) proposed a method for estimating the global credit risk factor (GCRF) by using information from the default-related components of CDS spreads. Their composite factor regression model was applied to stress-test credit portfolios and assess future credit risk scenarios. Zha et al. (2020) explored cross-country credit risk spillovers between sovereigns and firms using CDS data from Asian reference entities. Their findings revealed that financial institutions are more sensitive to shifts in the credit risk of foreign sovereign debtors, providing crucial insights into how these institutions respond to international credit risk. Dodd et al. (2021) analyzed the effect of leverage growth on corporate CDS by examining data on CDS from 350 firms in emerging markets. They discovered that leverage growth significantly increased CDS premiums, particularly for high-risk firms. Their study also highlighted that during the post-GFC period, CDS spreads were largely driven by global market risk factors, with emerging market CDS spreads primarily reflecting these factors during the crisis. Brownlees et al. (2021) developed a credit risk model for large financial institutions, where default intensity interdependence was influenced by both common risk factors and entity-specific idiosyncratic shocks. They recovered credit risk networks from CDS data using a lasso estimation process, which allowed them to study the interdependence of credit risks among European financial institutions. Palazzo and Yamarthy (2022) investigated the relationship between corporate credit risk and monetary policy shocks, using daily CDS data around Federal Reserve Open Market Committee (FOMC) announcements. Their research revealed that positive interest rate shocks led to increased expected losses and higher risk premiums in CDS spreads, shedding light on how monetary policy decisions affect the credit risk landscape. Dai et al. (2023) examined how CDS initiation influences the cost management strategies of reference firms. They concluded that CDS contracts enhance creditors' bargaining power, thereby reducing the likelihood of bankruptcy by mitigating cost stickiness after CDS initiation. Boudreault et al. (2015) proposed an estimation method for credit-sensitive instruments that minimizes noise and is highly parallelizable. Their simulation study demonstrated that this method outperforms other relevant estimators in terms of statistical properties. In their empirical research, they conducted a correlation analysis between CDS premiums and stock prices of firms listed in the CDX North America index, showing that ignoring noise could lead to an underestimation of correlations.

(6) Machine learning and game theory in credit rating

Research on the application of machine learning in credit rating covers a wide range of aspects, from credit risk assessment by BigTech to corporate credit risk modelling based on big data and credit risk assessment using methods such as clustering and fuzzy classification. Huang et al. (2023) compare the BigTech approach with the traditional banking approach by conducting a horse race analysis of large-scale loan transactions from online merchants at China's Medibank. traditional banking methods in loan default prediction. The results show that the BigTech approach performs better among borrowers with no bank account records, demonstrating its strengths in information and modelling. On the other hand, Huang and Wei (2023) constructed a quantitative model of corporate credit risk through big data analysis, investigated the relationship between credit rating and customer churn rate, determined the annual loan limit and annual interest rate of corporations, and provided banks with optimal credit strategies. Cheng and Qu (2020), through web crawler technology and word frequency analysis, investigated the bank fintech's credit risk impact and explored the application of fintech in the banking industry. Baser et al. (2023) proposed a clustering-based fuzzy classification (CBFC) credit risk assessment method, which uses a machine learning approach to compute the default risk and selects features that help to improve the prediction ability. Their results show that the CBFC model performs well in credit risk assessment and provides a new approach to credit rating. Machado and Karray (2022) used a combination of unsupervised and supervised machine learning methods to predict the credit scores of business customers using hybrid machine learning and found that the hybrid model outperformed a single model in prediction. Khandani et al. (2010) applied machine learning techniques to construct a nonlinear nonparametric predictive model to improve the classification rate of delinquency and default of credit card holders. In addition, Bao, Lianju, and Yue (2019) proposed a combinatorial strategy for credit risk assessment combining unsupervised and supervised learning, which achieved superior performance by applying unsupervised learning techniques through two stages. Doumpos and Zopounidis (2007) investigated the combinations of different classification methods, including machine learning and statistical techniques. found that the combined model outperformed a single model in credit risk analysis. Taken together, these studies suggest that machine learning plays a key role in credit ratings, improving the accuracy of predictions and the efficacy of credit risk analysis. The combination and innovative application of different methods provide financial institutions with more comprehensive and reliable credit assessment tools.

In the field of credit rating, the application of game theory provides a unique way to deeply understand and optimise credit decisions. Feng Bao (2022) used evolutionary game tools to construct a three-party game model of credit under the environment of government intervention, and analysed the impact of government intervention and internal factors on the credit game system through Matlab numerical simulation. The results show that the system is more likely to evolve to the optimal state with a higher initial proportion of positive credit strategies; government subsidies have a facilitating effect on the system's convergence to the optimal state; and government penalties can regulate the speed of evolution and have a tendency to maintain the positive state. Xue Chenhui and Gui Ping (2020) study the strategic choices of China's commercial banks under the implementation of green credit policy based on the game model, through the evolutionary game, complete information static game and repeated game models. The study points out that increasing the additional benefits from green credit, reducing the cost of issuance, and strengthening the regulation of commercial banks' environmental performance are the key factors to ensure the effective supply of green credit. Carling et al. (2007) propose a duration model of the pre-default survival time of a commercial loan borrower, which is used to analyse commercial defaults and can be used as an input to a portfolio credit risk model. The model not only allows for duration dependence but also increases inter-firm default risk correlation through macro explanatory variables. Feng Bao and Feng Chunfeng (2022) constructed a credit market model for green supply chain finance through an evolutionary game framework to examine the impact of government subsidies on the evolutionary path of the game system. It is found that the system is more likely to evolve towards the Pareto optimum when the proportion of initial positive strategies is high; government subsidies have a positive effect on the convergence of the system to the optimal state, but too large an initial subsidy may have a reverse effect. These studies not only provide a new perspective on game theory for credit rating, but also provide useful guidance for formulating effective policies and promoting sustainable financial development.

2.3.2 Research on Credit Risk of Small and Medium sized Enterprises

(1) P2P lending in credit rating

In the context of P2P lending, Guo et al. (2016) introduced a data-driven framework designed to overcome the limitations of traditional credit rating and assessment models, which often fail to meet the specific needs of individual investors. To address this gap, they proposed an instance-based credit risk assessment model that is capable of evaluating both the potential returns and the risks associated with each loan. By framing the investment decision-making process as a portfolio optimization problem with boundary constraints, their model provides a more tailored approach to managing investment risk. The authors conducted extensive testing using real datasets from two major P2P lending markets, and their findings revealed that the proposed model significantly improved investment performance when compared to traditional methods. In a similar vein, Zhou et al. (2021) developed four distinct gradient boosting techniques aimed at modeling the credit risk of borrowers on P2P lending platforms. Their research highlighted the critical importance of incorporating the timeliness of observations and features, leading them to segment the datasets using timestamps. Additionally, the study utilized Bayesian methods for hyperparameter optimization, which enhanced the precision and reliability of the resulting risk models. By refining the modeling techniques and integrating timely data, Zhou et al. (2021) were able to improve the accuracy of credit risk assessments, offering a more robust approach to risk management in P2P lending. The empirical results not only assess the performance of the model, but also reveal the importance of timestamps in assessing the performance of credit risk models in P2P lending. The role of psychological factors in borrowers' financial decisions has been increasingly emphasised. Yang et al. (2022) designed an online risk assessment framework called PsyCredit based on the theory of systemic functional language. It successfully improves the performance of credit risk assessment through deep learning and psychometric methods, which is innovative for solving the problem of mining psychological factors in P2P lending. Ma and Wang (2016) identified the factors affecting the credit risk of P2P lending from three aspects: the P2P lending platform, the borrower, and the environment, and revealed the intrinsic connection between these factors through an explanatory structural modelling approach, which provided a comprehensive perspective for managing and preventing the credit risk of P2P lending. Giudici et al. (2020) proposed a methodology that uses "alternative data" to enhance traditional credit scoring methods. Through methods such as centrality measures and financial ratio derivation, they extracted alternative data from borrower similarity networks to improve the prediction accuracy and model interpretability of credit risk. This provides a useful empirical study to improve the credit risk accuracy of P2P platforms, especially for SMEs.

(2) Credit portfolios in credit rating

In the research on credit rating applications of credit portfolios, researchers have used a variety of methods and models aiming to better understand and quantify financial risk as well as to improve the accuracy of credit loss forecasting. Çomakoğlu (2009) calibrated a multifactor Credit Metrics model with a regression approach and conducted a Carlo simulation for portfolios with different rating concentrations, revealing that the transformational risk and portfolio concentration on risk-return. Lopez and Saidenberg (2000) propose a cross-sectional simulation-based approach to credit risk model evaluation to assess the accuracy of model credit loss prediction by simulating credit portfolio forecasts at a given point in time. Fang et al. (2022) introduces a Fourier methodology by means of a Fourier domain in which the cumulative distribution function of total portfolio losses is calculated, improving the efficiency of credit risk quantification and allocation. Abdominal et al. (2015) enhance the methodology of scenario integration into risk models, particularly in stress-testing value-at-risk models, providing detailed application guidelines for market, credit and operational risk. Leitao and Ortiz-Gracia (2020) propose a wavelet-based nonparametric density estimation technique for measuring credit investment portfolio risk, which is capable of efficiently calculating risk measures such as marginal risk contribution, Var and ES. Wang et al. (2022) explored an optimisation algorithm for credit portfolios, which improves the optimisation efficiency of credit portfolios by minimising the default risk of the entire portfolio, introducing a bivariate intensity model and a multi-objective genetic algorithm. Together, these studies provide different perspectives and methods for risk management and optimisation of credit portfolios, providing powerful tools and guidance for financial institutions.

(3) Credit structure-related research

In the field of credit structure research, scholars have extensively investigated the influence of credit risk, financial technology, and cross-border capital flows on the credit systems of commercial banks from various perspectives. Jin, Yu, and Mi (2012) explored the impact of different industries on the credit risk faced by commercial banks, employing grey correlation analysis from both industry and macroeconomic perspectives. Their findings revealed that industry-specific and macroeconomic factors affect banks' impaired loan rates in varying degrees, offering fresh insights into grey wind control. The study's conclusions provide theoretical support for implementing a "differential treatment and regulation" approach, which could guide the optimization of credit structures. Bao Xing, Li Wei, and Li Quan (2022) developed a commercial bank fintech index using web crawlers and examined the relationship between fintech adoption and the non-performing loan (NPL) ratio of commercial banks. Their research concluded that fintech usage improves internal controls and reduces information asymmetry, ultimately leading to a lower NPL ratio. The benefits of fintech were particularly pronounced in banks with significant information asymmetry in their credit structures, highlighting the potential of technological advances in mitigating credit risk. Shi Jianhuai and Yang Yuqing (2022) focused on the effects of cross-border capital flows on the credit scale, risk, and structure of commercial banks in China. Using micro-level bank data, the study constructed a bank competition index to assess the moderating role of competition in the relationship between capital flows and bank credit. The results indicated that cross-border capital inflows not only expanded the size of bank credit but also led to an accumulation of credit risk, with bank competition exacerbating this effect. Additionally, capital inflows were found to influence credit structures by increasing credit allocation to the corporate sector while crowding out residential sector lending. The study further analyzed how bank-level heterogeneity influences credit responses to these capital flows.

These studies provide a multi-perspective and multi-level analysis to deepen the understanding of the factors influencing credit structure, and provide important references for formulating differentiated credit policies, guiding the development of fintech, and understanding the impact of cross-border capital flows on bank behaviour.

(4) Research related to credit risk contagion

Academic research on credit risk contagion has thoroughly examined its mechanisms, dimensions, and influencing factors. Fanelli and Maddalena (2020) explored the contagion of credit risk among default-prone financial agents by developing a nonlinear dynamic model that incorporates credit risk transfer markets and the interactions of various financial agents. The study utilized a time-delayed susceptibility-infection-recovery model to investigate credit risk contagion in the market, revealing steady states under various time delays and bank support policies. Chen Tingqiang et al. (2023) constructed a dynamic evolutionary model of credit risk contagion that integrates cost constraints and the influence of credit risk management expenses on banks' risk management decisions. Their findings suggest that both the level of credit risk management and banks' willingness to bear costs positively affect credit risk contagion, and they explored optimal strategies under varying conditions. Ericsson and Renault (2006) developed a structural bond valuation model to assess both liquidity and credit risk comprehensively. Their model revealed that renegotiation during financial distress is influenced by illiquidity in distressed debt markets, and they found that liquidity spreads exhibit a downward-sloping term structure with a positive correlation between illiquidity and default risk. Kiesel et al. (2001) expanded the ratings-based credit risk model by introducing the concept of spread risk. Their study demonstrated that volatility in spreads represents a major risk, especially for high-rated credits. Leippold and Vanini (2003) applied a financial economics approach to operational risk management in banks, presenting a framework to optimize profitability through capital allocation, network stability, and diversification, using both analytical and numerical methods. Kealhofer (2003) analyzed the influence of asset price bubbles on standard credit risk measures and proposed a new metric, the expected holding period credit loss (EHPCL). Through sensitivity analysis, the study demonstrated that EHPCL is effective in minimizing bias in credit loss estimates. Segoviano Basurto and Padilla (2007) combined the Conditional Probability of Default (CoPoD) method with Consistent Information

Multiple Density Optimization (CIMDO) to enhance credit risk measurement in investment portfolios. Their methods performed well in stress tests, providing strong support for better credit risk management. Belkin and Forest (1998) suggested a one-parameter representation for credit risk and transfer matrices, which were stress-tested to analyze how credit changes affect individual instruments or an entire portfolio. Kanno (2020) examined credit rating migration risk and interconnections in the Japanese corporate lending market by analyzing outstanding loan data and using portfolio credit risk techniques. His research, which included a detailed analysis of network structures and historical economic conditions, provides valuable insights into credit risk management in corporate lending.

Finally, Chih-Bing Lin and Yu-Wen Lee (2023) investigated the role of consumer credit in alleviating consumers' financial pressure and increasing the demand for green products by developing a Stackelberg model with immediate payment, bank-provided consumer credit, and retailer-provided consumer credit. The study explored the influencing factors of consumer credit strategy preference through numerical analysis with a mixed strategy model. The results show that under certain conditions, consumer credit has a positive effect on increasing both market demand for green products and channel members' profits.

(5) Empirical study of credit risk-return relationship and its investigation

In asset pricing theory, risk and return are usually positively correlated, however, the empirical results on the relationship between credit risk and return are controversial, creating a credit risk-return puzzle. Nedumparambil and Bhandari (2020) found that this puzzle does exist by using credit ratings as a measure of credit risk through a study of the Indian stock market. It is noteworthy that traditional asset pricing models fail to consider credit risk comprehensively, while some evidence supports behavioural and rational pricing explanations. Brei et al. (2020), on the other hand, examine the relationship between bank competition and stability in Sub-Saharan Africa. They find that credit risk is low when the level of bank competition is moderate, but when competition is excessive, the positive effects are cancelled out by the negative ones. Moreover, credit risk is not only related to macroeconomic and financial factors, but is also closely related to the business and regulatory environment, revealing a complex credit risk landscape. Djebali and Zaghdoudi

(2020) study conventional banks in the MENA region, and find a non-linear relationship between bank stability and both credit risk and liquidity risk. These two risks are detrimental to bank stability at high risk levels. They recommended that banks diversify their operations, strengthen their own funds, and adapt to new technologies to ensure the stability of the financial system. Focusing on banks in the Association of Southeast Asian Nations (ASEAN) region, Ovi (2020) examined the relationship between business cycles, income diversification and banks' capital buffers and credit risk. It is found that banks exhibit countercyclical responses in adjusting their capital buffers and credit risk, while income diversification helps to reduce credit risk. Zhou et al. (2021), on the other hand, focuses on online consumer credit services in the area of e-commerce and proposes an integrated scheme to infer online consumer credit risk. The scheme achieves finer-grained credit risk determination through data augmentation and model enhancement, providing a useful tool for the industry. In China's financial industry, Yuan et al. (2022) point out that the current credit rating market faces problems such as inflated ratings, insufficient grade differentiation, and poor prediction and warning performance. To this end, they propose a new methodology based on the hologram approach to deal with unstructured data, create a reasonable number of "bad samples", and establish the "CAFE Risk Assessment System". Pei et al. (2023), on the other hand, in terms of regulatory efficiency Pei et al. (2023) introduce corporate credit risk ratings and achieve rational allocation of regulatory resources through a hybrid decomposition strategy and a two-stage multi-objective feature selection multi-category integrated model. Their model is designed to effectively deal with the category imbalance problem, taking into account the balance of various demands and providing an efficient tool for regulation. Finally, Stefania et al. (2023) designed a risk scoring model through discriminant analysis techniques, which solved the problems encountered in models that read the credit histories of corporate customers, and provided useful references for the management of credit risk exposures. Ayayi (2012) emphasised the importance of both qualitative and quantitative risk management tools when discussing the assessment of credit risk, arguing that good governance practices and a sustainable financial performance have a direct impact on low credit risk. His study also suggests that the depth and breadth of outreach and underwriting play a key role in credit risk control in MFIs.

In summary, findings from different regions and industries suggest that credit risk is related to a variety of factors, including macroeconomics, competitive environment, and regulatory quality. These studies provide us with an in-depth insight into the relationship between credit risk and various aspects such as macroeconomics, competitive environment, and regulatory quality. This comprehensive understanding helps us to better understand and effectively manage the complexity and variability of credit risk. In the financial and corporate sectors, these studies provide strong support for the development of risk management strategies and the improvement of regulatory efficiency, which are important for maintaining the stability and sustainable development of the financial system.

2.4 Research Gaps

(1) Non-financial credit risk assessment indicators for technology-based small and micro enterprises have been proposed.

This article addresses the unique challenges of credit risk assessment in small and micro enterprises (SMEs) by identifying four key risk characteristics specific to these entities. Through a comprehensive literature review and research, the study highlights two primary factors that significantly influence the credit risk of small and micro enterprises: innovative information and business models. These insights provide novel solutions to the persistent problem of inaccurate or difficult credit risk assessments faced by SMEs (Zhu, Y., Chen, L., Hu, Y., et al., 2024). In particular, the article proposes non-financial credit risk assessment indicators, which can effectively mitigate the challenges posed by high information asymmetry in these businesses. These indicators serve as a crucial tool in overcoming the difficulties of evaluating credit risk in SMEs, where traditional financial data may often be incomplete or unreliable. To validate the efficacy of these non-financial factors, the study collected first-hand data from a variety of SMEs and compared the performance of traditional financial metrics with the newly proposed indicators (Zhu, Q., 2022). The experimental results demonstrate that incorporating non-financial factors enhances the accuracy and reliability of credit risk assessments for small and micro enterprises, offering a more robust and comprehensive approach to evaluating their creditworthiness. This research contributes to the field by providing actionable insights into improving credit risk assessment models for SMEs, thus aiding financial institutions in making more informed lending decisions.

(2) A new perspective on quantifying credit risk.

Credit risk in small and medium-sized enterprises (SMEs) has long been a challenging and significant issue in the financial sector. Traditional credit risk assessments typically prioritize large enterprises, leaving SMEs, which often struggle with limited scale, irregular financial data, and inadequate collateral, without sufficient credit support. This study focuses on addressing these challenges by specifically exploring the credit risk of SMEs, filling a crucial gap in the existing literature and offering an innovative approach to solving the financing difficulties faced by these businesses. A novel aspect of this research is the use of upstream and downstream invoice data as the primary source of information for credit risk assessment (Zhu, Z., & Du, S. 2024). Invoice data not only reflects the transactional behavior of SMEs but also carries valuable insights regarding their supply chain relationships, business stability, and market competitiveness. This information is instrumental in assessing credit risk, as it provides a more direct and timely representation of the enterprise's operational conditions compared to traditional financial statements. The dataset for this study primarily consists of authentic, real-time, and large-scale invoice data from SMEs, which offers a significant advantage over conventional financial data. Invoice data presents a more accurate and holistic picture of a company's actual business activities, making it a more reliable basis for assessing credit risk. Through the use of data mining techniques, this study aims to uncover the underlying patterns and trends within the invoice data, facilitating a deeper understanding of the factors influencing credit risk. By constructing a robust data mining model, this research enables a quantitative approach to credit risk assessment, providing valuable decision-making support for financial institutions. The study integrates statistical models, such as the Logit and Probit models, with machine learning algorithms, including neural networks, to predict SME credit risk more accurately. These advanced models are capable of handling complex nonlinear relationships within the data, thereby enhancing both the accuracy and efficiency of the credit risk evaluation process. This integrated approach not only improves the reliability of predictions but also offers a comprehensive tool for addressing the unique challenges of SME financing.

Chapter 3
RESEARCH METHODOLOGY

This chapter provides a detailed explanation of the research methodology. It covers the data sources utilized, the variables analyzed, and the step-by-step process of implementing the methodology. Additionally, it presents a workflow chart that illustrates the procedure for building the credit risk model and extracting meaningful insights. This section clarifies the research approach and the reasoning behind the selected methods.

3.1 Data and Variables

3.1.1 Data

This study focuses on small and medium-sized enterprises in China and collects data on related enterprises from 2016 to 2019. The data is sourced from the Wind database and is based on daily data. This study collected daily data from 425 small and medium-sized enterprises from 2016 to 2019. Among them, 123 enterprises had credit record data, while 302 enterprises did not have credit record data.

Table 3 Overview of input invoice data of companies with credit records

Record	Company Code	Invoice Number	Invoicing Date	Sales Unit Code	Amount /yuan	Tax /yuan	Price and tax total/yuan	Invoice Status
1	C1	3390939	2017/7/18	A00297	-943.4	-56.6	-1000	Valid
2	C1	3390940	2017/7/18	A00297	-4780.24	-286.81	-5067.05	Valid
3	C1	3390941	2017/7/18	A00297	943.4	56.6	1000	Valid
......								
3444	C1	3390942	2019/12/31	A00297	4780.24	286.81	5067.05	Valid
3445	C1	9902669	2019/12/31	A05061	326.21	9.79	336	Invalid
3446	C2	40826107	2017/8/8	A05991	170.94	29.06	200	Valid
3447	C2	4420531	2017/8/9	A03142	37735.85	2264.15	40000	Valid
3448	C2	4420532	2017/8/9	A03142	4716.98	283.02	5000	Valid
......								
35599	C2	11924153	2019/12/31	A11821	1946.02	252.98	2199	Valid
35600	C2	7757361	2019/12/31	A12601	601.94	18.06	620	Valid

Record	Company Code	Invoice Number	Invoicing Date	Sales Unit Code	Amount /yuan	Tax /yuan	Price and tax total/yuan	Invoice Status
35601	C2	8611117	2019/12/31	A13080	145.63	4.37	150	Valid
……								
210946	C123	95472001	2018/12/29	A03626	264.15	15.85	280	Valid
210947	C123	54469883	2019/12/18	A03626	264.15	15.85	280	Valid

Table 3 provides details regarding the enterprise's input and output invoices. The input invoice refers to the one issued by the seller when the enterprise purchases goods or products, while the corresponding output invoice is the one issued by the enterprise to the buyer when it sells products. Valid invoices are those issued for legitimate business transactions. Cancelled invoices, on the other hand, are those that were initially issued for trading activities but later rendered invalid due to the cancellation of the transaction for various reasons. Negative invoices are issued when a buyer returns goods, leading to a refund, after the tax has already been recorded by the enterprise. In such cases, negative invoices are required to adjust the original transaction.

The company issued a total of 210,947 input invoices and 162,484 output invoices. These invoices provide a comprehensive overview of the enterprise’s purchase and sales transactions, contributing to a better understanding of its trading activities and tax-related adjustments.

The credit record data of 123 enterprises are shown in Table 4.

Table 4 Credit rating and default situation of 123 companies

Company Code	Credit Rating	Default Situation
C1	A	NO
C2	A	NO
C3	C	NO
……		
C45	B	YES
C46	C	NO

Company Code	Credit Rating	Default Situation
	……	
C121	D	YES
C122	D	YES
C123	D	YES

From Table 4, the credit rating refers to the manual assessment made by the bank based on the actual situation of the enterprise. In principle, the bank will not lend to an enterprise with a credit rating of D.

Table 5 illustrates the variations in customer churn rates based on different loan interest rates and credit ratings categorized as A, B, and C. The customer churn rate refers to the proportion of potential customers that a bank loses, influenced by factors such as the interest rates on loans. The table provides a comparative analysis of how customer retention fluctuates across varying credit ratings and interest rate levels, offering insights into the impact of these factors on customer behavior and bank performance.

Table 5 The description of loan annual interest rate and customer churn rate

Loan Annual Interest Rate	**Credit Rating A**	**Credit Rating B**	**Credit Rating C**
0.04	0	0	0
0.0425	0.094574126	0.066799583	0.068725306
0.0465	0.135727183	0.13505206	0.122099029
0.0505	0.224603354	0.20658008	0.181252146
0.0545	0.302038102	0.276812293	0.263302863
0.0585	0.347315668	0.302883401	0.290189098
0.0625	0.41347177	0.370215852	0.34971559
0.0665	0.447890973	0.406296668	0.390771683
0.0705	0.497634453	0.458295295	0.45723807
0.0745	0.511096612	0.508718692	0.492660433
0.0785	0.573393087	0.544408837	0.513660239
0.0825	0.609492115	0.548493958	0.530248706
0.0865	0.652944774	0.588765696	0.587762408
0.0905	0.667541843	0.625764576	0.590097045
0.0945	0.694779921	0.635605146	0.642993656
0.0985	0.708302023	0.673527424	0.658839416
0.1025	0.731275401	0.696925431	0.696870573
0.1065	0.775091405	0.705315993	0.719103552
0.1105	0.798227368	0.742936326	0.711101237
0.1145	0.790527266	0.776400729	0.750627656
0.1185	0.815196986	0.762022595	0.776816043
0.1225	0.814421029	0.791503697	0.784480512
0.1265	0.854811097	0.814998933	0.795566274
0.1305	0.870317343	0.822297861	0.820051434
0.1345	0.871428085	0.835301602	0.832288422
0.1385	0.885925945	0.845747745	0.844089875
0.1425	0.874434682	0.842070844	0.836974326
0.1465	0.902725909	0.868159536	0.872558957
0.15	0.922060687	0.885864919	0.895164739

3.1.2 Variables

By consulting literature and principal component analysis models, this article selects eleven refined variables, including Total input price and tax *(x_1)*, Total sales price and tax *(x_2)*, Total valid invoices *(x_3)*, Monthly input amount coefficient of variation *(x_4)*, Monthly sales amount coefficient of variation *(x_5)*, Number of upstream enterprises *(x_6)*, Number of downstream enterprises*(x_7)*, Total profit *(x_8)*, Operation time *(x_9)*, Percentage of valid invoices *(x_{10})*, Credit rating *(x_{11})*, as measurement variables for risk assessment of small and medium-sized enterprises. The definitions and symbols of relevant variables are shown in Table 6.

Table 6 Definitions and symbols of relevant variables

Symbol	Definition and Description	Unit
j	Enterprise j	
k_j	The expected returns of J Bank for all customer enterprises with different credit ratings k	
Z	The expected return of the bank on the all customer enterprise j	yuan
t_j	The expected return of the bank on the current customer enterprise j	yuan
A_j	Possible loan amount for enterprise j	yuan
B_j	The actual loan amount of enterprise j	yuan
y_j	Indicator for determining whether the bank has lent to existing customer enterprises	
i_j	Enterprise J loan interest rate	%
P_j	The repayment probability of enterprise j	
q_j	The probability of customer churn in enterprise j	
M_j	The customer churn value of enterprise J	yuan
d	Annual total credit amount of the bank	
R_j	The default probability of enterprise j	
x_1	Total input price and tax	yuan
x_2	Total sales price and tax	yuan
x_3	Total valid invoices	
x_4	Monthly input amount coefficient of variation	
x_5	Monthly sales amount coefficient of variation	
x_6	Number of upstream enterprises	
x_7	Number of downstream enterprises	
x_8	Total profit	yuan
x_9	Operation time	days
x_{10}	Percentage of valid invoices	%
x_{11}	Credit rating	
h_{1j}	The coefficient of increase in loan limits for different enterprises by banks	
h_{2j}	Bank's interest rate discount coefficient for different enterprises	

3.2 Steps of the Methodology

3.2.1 Methodology steps for objective 1

Objective 1: To determine the risk evaluation index system for small and medium-sized enterprises, and select appropriate models to quantify the risks of these enterprises. The main framework diagram is shown in the Figure 1.

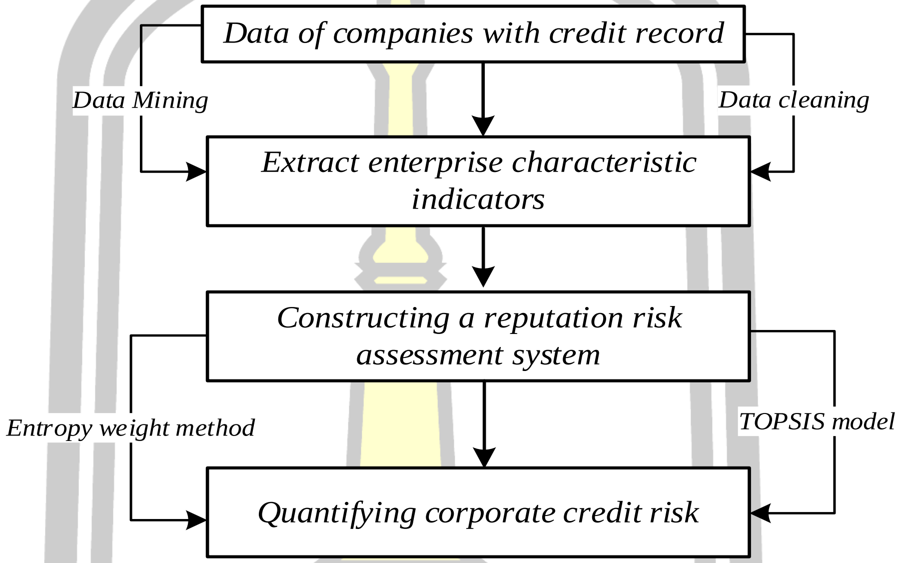

Figure 1 Framework diagram of objective 1

It can be seen from Figure 1 that the company's credit risk will be quantified by the Entropy weight method and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) model, and specific calculation steps will be used.

Methodology 1：***Entropy weight method to calculate weight***

The entropy weight method is an objective weighting method that can avoid the influence of subjective weighting on the credit risk quantification results to the greatest extent. The principle of the entropy weight method is the degree of variation of the indicator, that is, the higher the degree of variation, the higher the corresponding weight.

Step 1: First of all, this paper needs to process the enterprise indicator data positively and normalize it to ensure the non-negativity of the data:

$$z_{ij} = \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)}.$$

Among them, z_{ij} is the normalized variable, $\min(x_i)$and $\max(x_i)$are the minimum and maximum values of each indicator respectively, $x_i = (x_1, x_2, \cdots, x_{11})$.

Step 2: Calculate the weight of the j enterprise under the i credit risk indicator and regard it as the probability p_{ij} when calculating information:

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^{n} z_{ij}}.$$

Step 3: Calculate the information entropy j of the e_j credit risk indicator and calculate the corresponding information utility value d_j. The reason for the conversion here is that the larger the information entropy, the less information about the credit risk indicator. By introducing the information utility value d_j, the amount of information can be positively measured.

$$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^{n} p_{ij} \ln(p_{ij}),$$

$$d_j = 1 - e_j.$$

Step 4: Finally, the entropy weight w_j of each credit risk indicator is obtained by normalization.

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}.$$

Methodology 2 : *TOPSIS model*

The TOPSIS method is a method of sorting samples based on data. The basic idea is to construct an idealized target based on sample data. For example, in this case, it is to construct an enterprise with optimal credit risk indicators in all aspects, and then measure the degree of closeness between the actual enterprise and this idealized enterprise. The closer it is, the lower its credit risk.

Step 1: Find the maximum value of each column, that is, each credit risk indicator, recorded as $z_i^+ (i = 1,2. \dots, m)$, and form a vector:

$$Z^+ = \{z_1^+, z_2^+, \cdots, z_m^+\}.$$

This vector represents the ideal enterprise.

Step 2: Similarly, find the minimum value of each column, that is, each indicator, recorded as $z_i^- (i = 1,2, \dots m)$, to form a vector:

$$Z^- = \{z_1^-, z_2^-, \cdots, z_m^-\}.$$

This vector represents the least ideal enterprise, that is, each positively converted indicator reaches the minimum.

Step 3: Define the distance between the i sample and the ideal target as D_i^+, and the calculation formula is

$$D_i^+ = \sqrt{\sum_{j=1}^{m} (z_j^+ - z_{ij})^2}.$$

Define the distance between the i enterprise and the undesirable target as D_i^-, and the calculation formula is

$$D_i^- = \sqrt{\sum_{j=1}^{m} (z_j^- - z_{ij})^2}.$$

Define the score of the i enterprise as S_i, and the calculation formula is

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}.$$

Obviously, S_i is between $[0,1]$. When S_i is closer to 1, it means that enterprise i is closer to the ideal target, and the credit risk of the enterprise is lower. On the contrary, when S_i is closer to 0, it means that enterprise i is farther from the ideal target, and the credit risk of the enterprise is higher.

3.2.2 Methodology steps for objective 2

Objective 2: According to the credit risk measurement index system established above, the improved decision tree model is used to predict the credit rating and default record of 302 enterprises with no credit record, at the same time, the credit risk of these enterprises is quantified. Finally, the nonlinear programming model is used to calculate the bank's credit strategy for each enterprise. The main framework diagram is shown in the figure 2.

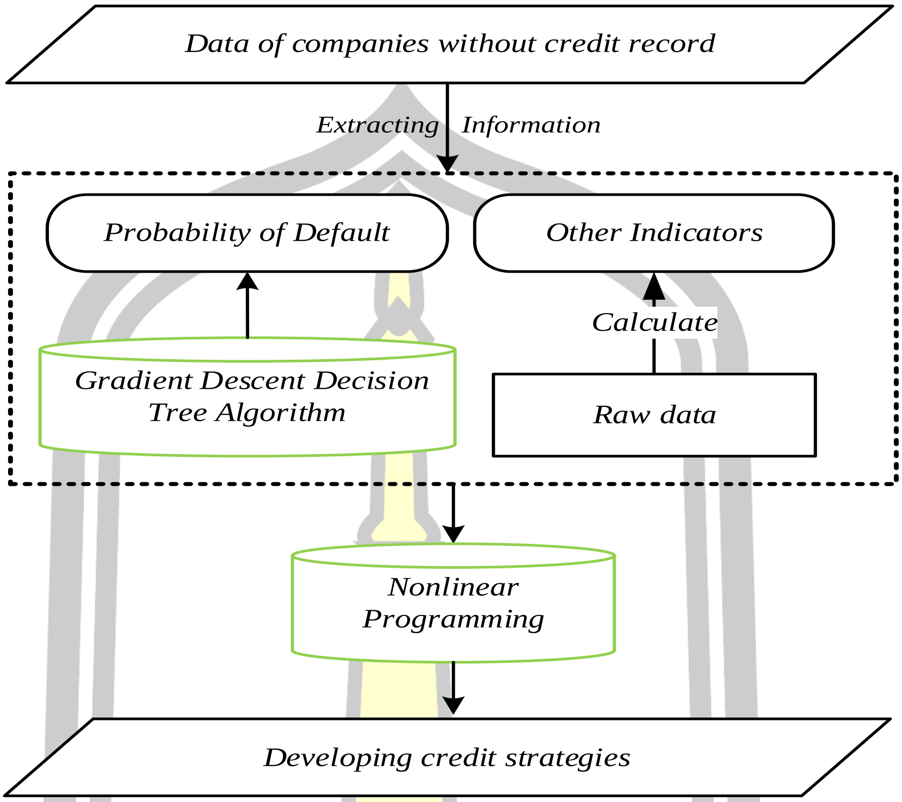

Figure 2 Framework diagram of objective 2

Methodology 1 : *Gradient Descent Decision Tree Algorithm*

Based on the evaluation indicators determined in the previous research, this research improves the traditional decision tree algorithm by adding a regularization function to solve the overfitting problem of the algorithm. It establishes a credit risk assessment model with the default probability of the enterprise as the dependent variable, and then predicts the probability value of the enterprise's default. The Gradient Boosted Decision Tree Algorithm is adopted here.

In order to further improve the prediction accuracy of the gradient boosting decision tree algorithm, this paper improves the original decision tree algorithm model and establishes a credit risk model. Assuming that for a company, its credit risk is characterized by the default probability of credit repayment, denoted as, based on the various evaluation indicators considered earlier, the default probability of credit repayment for each company can be expressed as equation:

$$R_j = \varphi(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}) \quad (1)$$

The optimization function can be expressed as follow:

$$R_j = \varphi(x_1, x_2, x_3, x_4, x_5, x_6, x_{7,}x_{8,}x_{9,}x_{10,}x_{11}) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right)$$

$$+\delta(f_t) + constant \qquad (2)$$

From equation (2), l is the loss function $\delta(ft)$ is the regularizerconstant is the constant term. The objective function construction and the specific iterative calculation process of the algorithm are as follows:

Step 1: Build a single decision tree

Let $f(x_i)$ be a decision tree function, and the number of leaf nodes is T. Then the decision tree feature selection is performed based on the information gain method: for the decision tree, assuming that the current node is denoted as C, the left child node after the split is denoted L, and the right child node is denoted R, then the benefit obtained by the split is defined as the objective function value of the current node minus Get the sum of the objective function values of the left and right child nodes:

$$Gain = f_C - f_L - f_R \qquad (3)$$

In the process of generating a decision tree, the features with the greatest benefit are selected as branches.

Step 2: Based on the additive model, build an ensemble learning decision tree

After establishing a single decision tree model, based on the idea of ensemble learning, multiple decision trees (weak classifiers) are integrated. This article uses an additive model to obtain a strong classifier. At this time, the objective function is an additive model composed of K trees:

$$\hat{y}_l = \sum_{k=1}^{K} f_k(x_i), f_k \epsilon F \qquad (4)$$

Where f is the *kth* decision tree, and then the Boosting algorithm is used to train and learn the model. Since the learning model is an additive model, the Boosting algorithm can learn the model from front to back, learning only one basis function and its coefficients (structure) at each step, gradually approaching the optimization objective function, and thus simplifying the complexity of the operation. The method starts with a constant prediction and learns a new function each time as follows:

$$\hat{y}_i^0 = 0 \qquad (5)$$

$$\hat{y}_i^1 = f_1(x_i) = \hat{y}_i^0 + f_1(x_i) \qquad (6)$$

$$\hat{y}_i^2 = f_1(x_i) + f_2(x_i) = \hat{y}_i^1 + f_2(x_i) \qquad (7)$$

$$\cdots$$

$$\hat{y}_i^t = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{t-1} + f_t(x_i) \qquad (8)$$

At step t, its objective function can be written as:

$$R_j^t = \sum_{i=1}^{n} l(y_i, \hat{y}_i^t) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) \qquad (9)$$

On the other hand, from the second-order expansion of Taylor's formula at point x, we can get:

$$f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)\Delta x^2 \qquad (10)$$

Then the above formula is transformed into:

$$R_j^t = \sum_{i=1}^{n} \left[l(y_i, \hat{y}_i^t) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] \qquad (11)$$

The loss function l selected in this article is the square loss function, and the objective function is:

$$R_j^t = \sum_{i=1}^{n} (y_i - (\hat{y}_i^{t-1} + f_t(x_i)))^2 = \sum_{i=1}^{n} [2(\hat{y}_i^{t-1} - y_i)f_t(x_i) + f_t(x_i)^2] \quad (12)$$

Among them, $(\hat{y}_i^{t-1} - y_i)$ is the residual. When using the square loss function, the integrated decision tree continuously fits the model by fitting the residual in the previous step model.

Step 3: Based on regular optimization, improved gradient descent decision tree model

Let $f(x_i)$be a decision tree function, and the number of leaf nodes is T. In order to avoid overfitting problems in the decision tree, this article adds regular terms to the original decision tree model. The complexity of the decision tree is expressed by the regular term as:

$$\delta(f_t) = \gamma T + \frac{1}{2}\lambda \textstyle\sum_{j=1}^{T} w_j^2 \quad (13)$$

Then at step t, its objective function can be written as:

$$R_j^t = \sum_{i=1}^{n} l(y_i, \hat{y}_i^t) + \sum_{i=i}^{t} \delta(f_i) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) + \delta(f_t) + constant \quad (14)$$

For a single decision tree, define the set $I_j = \{i|q(x_i) = j\}$ is the set of all training samples divided into leaf nodes. Then Equation (11) can be reorganized into the sum of T independent quadratic functions according to the leaf nodes of the tree:

$$R_j^t = \sum_{i=1}^{n} \left[g_i w_q(x_i) + \frac{1}{2} h_i w_{q(x_j)}^2\right] + \gamma T + \frac{1}{2}\gamma \sum_{j=1}^{T} w_j^2$$

$$= \sum_{j=1}^{T} \left[(\sum_{i\epsilon l} g_i)w_j + \frac{1}{2}(\sum_{i\epsilon I_j} h_i + \lambda)w_j^2\right] + \gamma T \quad (15)$$

Definition $G_j = \sum_{i\epsilon I_j} g_i, H_j = \sum_{i\epsilon I_j} h_t$, Then Equation (15) can be expressed as:

$$R_j^t = \sum_{j=1}^{T} \left[G_i w_j + \frac{1}{2}(H_i + \lambda)w_j^2\right] + \gamma T \quad (16)$$

Find the first derivative of Equation (16) and set the first derivative equal to 0, we have:

$$w_j^* = -\frac{G}{H_j + \lambda} \quad (17)$$

Then the value of the objective function at this time is:

$$R_j^t = -\frac{1}{2}\sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (18)$$

From this, the income for each decision tree split is:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \gamma \quad (19)$$

At each iteration, the profit and loss are calculated through Equation (19), a new decision tree is generated based on the principle of maximum benefit, and the predicted value corresponding to each leaf node is calculated through Equation (18), and the newly generated decision tree $f_t(x)$ is added to the model again, that is:

$$\hat{y}_i^t = \hat{y}_i^{t-1} + f_t(x_i) \tag{20}$$

Through the above process, iteration is continued, and each iteration generates a weak classifier. Each weak classifier is trained on the basis of the residuals of the previous classifier. In this way, iteration is continued until the target accuracy is reached. The model's The iteration flow chart is as Figure 3.

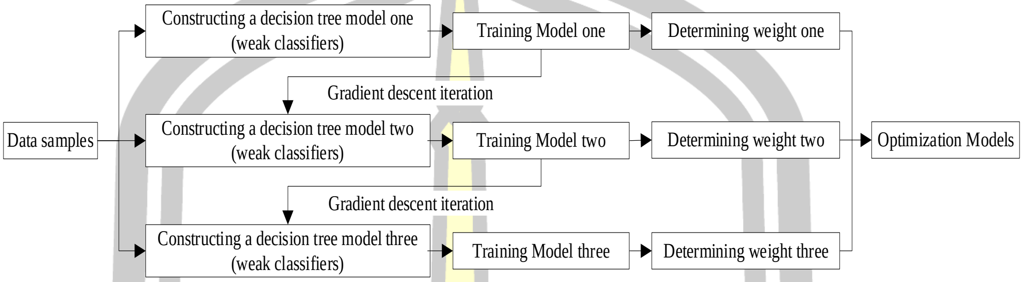

Figure 3 Model iteration flowchart

Methodology 2 : *Nonlinear Programming*

This study used data from 123 enterprises that had credit record and programmed an improved gradient decision tree model for supervised learning training. The final learning model was obtained through training set and validation set partitioning, ensemble learning, cross validation, and other methods. The prediction accuracy of the learning model can be obtained through computation. Compare the prediction accuracy with traditional decision tree models and AdaBoost algorithm. Finally, using the improved decision tree model, the default probabilities of 123 companies that had credit record were calculated. Next, I will calculate the default probability and credit rating of 302 companies, and start building the bank's credit strategy model. Banks are not allowed to lend to enterprises with a credit rating of D, and are not allowed to lend to enterprises with a high probability of default. Establishing a credit disbursement strategy model based on nonlinear programming algorithm. The bank's credit strategy is shown in Figure 4.

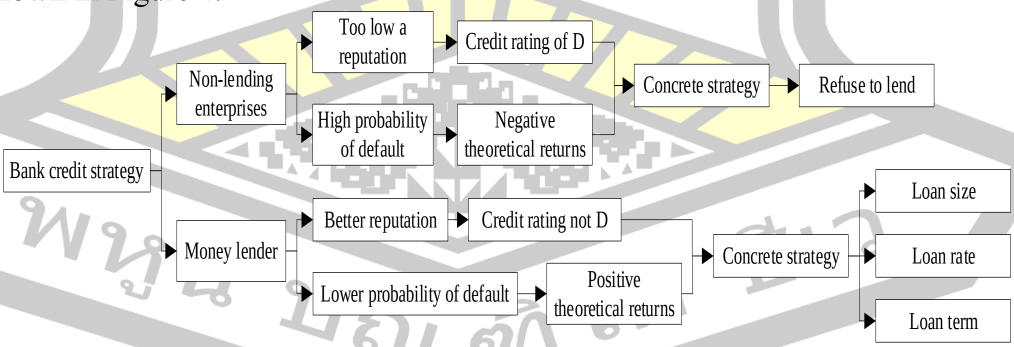

Figure 4 The bank's credit strategy

For the bank, the goal of lending to different enterprises is to maximize loan returns while minimizing customer churn and default risk. This way, the bank can obtain relatively stable and generous loan returns, as well as a stable customer base, which is conducive to the development and profitability of the bank's business.

Step 1: Do not grant loans to companies with a high probability of default

In the credit risk assessment model, the default probability of each enterprise has been obtained. When the default probability of the enterprise is too high, the enterprise is very likely to fail to repay, causing the bank to suffer losses. Therefore, we use the bank's expected return t_j for its current customer company j to measure the company's credit risk. Let A_j be the loan amount of the enterprise, i_j be the enterprise loan interest rate, and P_j be the repayment probability of the enterprise.

$$t_j = A_j \cdot i_j \cdot P_j - A_j \cdot 0.22 \cdot \left(1 - P_j\right).$$

Suppose $t_j \leq 0$, The company is considered to have a high risk of default and the bank will not lend, and $y_j = 0$. Suppose $t_j > 0$, It is believed that the risk of default of the company is low, and the bank lends, and $y_j = 1$. This formula represents the expected repayment amount of enterprise j after one year of borrowing, and the expected return that the bank can obtain. If the enterprise does not default and repays normally, the bank can obtain a profit on the loan amount, which is the return; if the enterprise defaults and cannot repay normally, the bank will suffer a large loss. Considering that the enterprise default does not mean that the principal and interest will not be repaid at all, and the default probability of small and micro enterprises is higher, if the calculation is based on the total loss of principal and interest, the expected return of almost all small and micro enterprises will be negative, and loans cannot be issued. Therefore, this article refers to relevant research and selects the one-year average loan default loss rate of 0.22 for calculation to avoid the default of the loan being completely lost, resulting in too many enterprises with high default risks being identified and unable to issue loans. Therefore, the final loan amount of the enterprise is expressed as B_j, then: $B_j = A_j \cdot y_j$. Only when $y_j = 1$, the bank will lend to the enterprise and set a specific loan amount and interest rate.

Step 2: Bank's credit strategy model

Therefore, this research establishes a multi-objective programming model with the goals of maximizing loan returns, minimizing customer churn rates, and minimizing customer default risks. Based on the known conditions and assumptions of the problem, the planning model needs to meet the following constraints:

(1) Banks have a certain limit on the loan amount for different enterprises, and the total loan amount for all enterprises does not exceed the bank's loan limit.

(2) Banks have a certain range of loan interest rates for different enterprises.

(3) Banks have the right to decide whether to lend to enterprises with different default risks.

(4) Customer churn in banks can cause certain losses to expected returns, and the value of customer churn varies among enterprises of different scales. The final objective function is:

$$maxZ = \sum_{j=1}^{302} y_j[A_j \cdot i_j \cdot (1 - R_j) - A_j \cdot 0.22 \cdot R_j](1 - q_j) - M_j \cdot q_j$$

The constraints that need to be met are:

$$\sum_{j=1}^{302} A_j \leq d$$
$$10 \leq A_j \leq E_j$$
$$0.04 \leq i_{kj} \leq 0.15$$
$$t_j = A_j \cdot i_j \cdot P_j - A_j \cdot 0.22 \cdot (1 - P_j)$$
$$y_j = \begin{cases} 0, t_j \leq 0 \\ 1, t_j > 0 \end{cases}$$
$$q = f(i)$$
$$M = \eta(x_1, x_2, x_3, x_4, x_5, x_6, x_{7,} x_{8,} x_{9,} x_{10,} x_{11})$$
$$R = \phi(x_1, x_2, x_3, x_4, x_5, x_6, x_{7,} x_{8,} x_{9,} x_{10,} x_{11})$$
$$P + R = 1$$

M, P, R, E are Representing customer churn value, enterprise repayment probability, enterprise default probability, and maximum loan limit respectively.
Finally, formulate credit strategies for 302 enterprises without credit records.

3.2.3 Methodology steps for objective 3

Objective 3: Considering that the production, operation and economic benefits of enterprises may be affected by unexpected factors, such as the COVID-19 epidemic, we have comprehensively considered the credit risks of different industries and enterprise categories and their response strategies to ensure that banks can flexibly adjust credit policies in different situations and reduce potential risks. The main framework diagram is shown in the Figure 5.

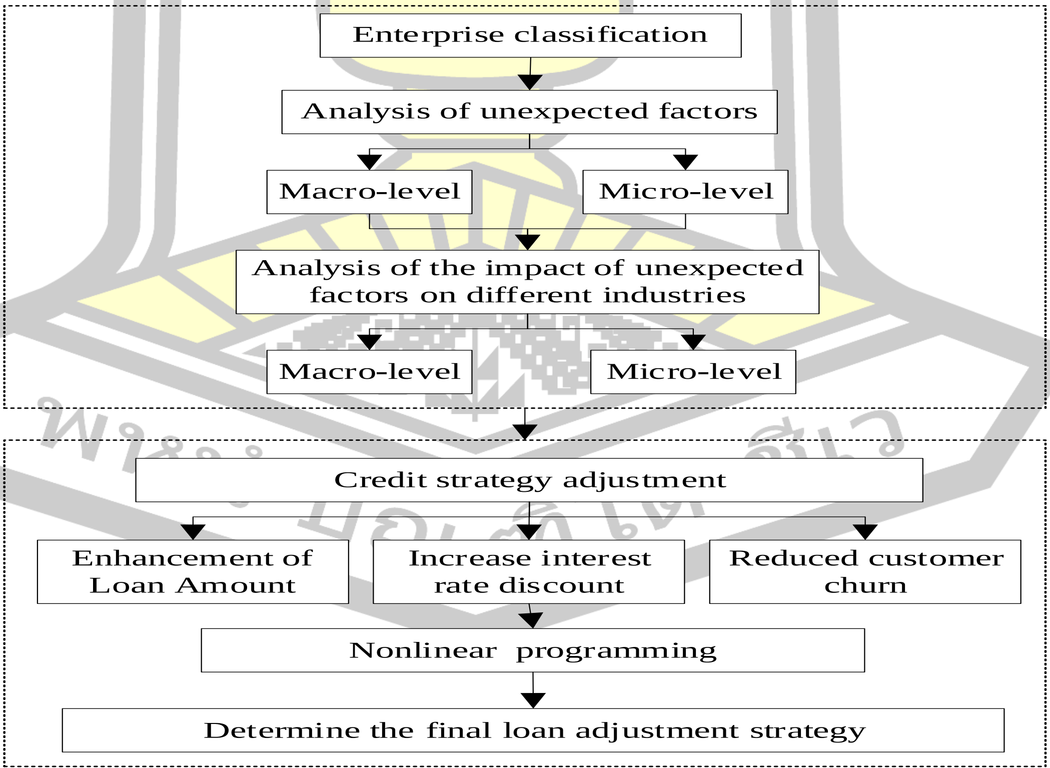

Figure 5 Framework diagram of objective 3

The production, operations, and profitability of enterprises are often influenced by a range of external factors, including the market environment, national policies, and unforeseen events. These factors can vary in their impact across different industries and types of businesses, and their effects are frequently transmitted to financial institutions through credit loans and other financial channels. As a result, such external influences can have a significant bearing on a bank's capital security and overall profitability.

Given these complexities, it is crucial for banks to thoroughly assess the credit risks associated with each enterprise before making decisions on lending. This evaluation process should take into account not only the financial health and stability of the business but also the potential effects of unexpected external factors on its performance. These considerations are essential for determining the appropriate loan amount and the most suitable interest rate, ensuring that the bank can effectively mitigate risks while safeguarding its capital. In short, banks must adopt a comprehensive approach to credit risk assessment, factoring in both the inherent financial characteristics of the enterprises and the broader external uncertainties that could impact their ability to repay loans.

Step 1: Based on the Statistical Classification Method of Large, Medium, Small and Micro Enterprises promulgated by the National Bureau of Statistics, the information of 302 enterprises is classified according to industry categories and enterprise scales, and is divided into 14 major industry categories and 4 major scale categories.

Step 2: Taking the COVID-19 epidemic as an example, the severity of the impact of the epidemic on enterprises of different industries and sizes is analyzed from both macro and micro levels. From a macro perspective, this study selects representative listed company stocks in various industries, obtains stock price information of various industries since the outbreak of the COVID-19 epidemic based on the WIND database, and uses stock price fluctuations as a basic indicator for analyzing the development and recovery of the industry. At the micro level, this study randomly selected 72 companies from 302 companies to analyze their year-on-year growth rates from January to February 2020 compared with the same period in 2018 and 2019 to determine the operating conditions of the companies;

Step 3: Based on a combination of macro and micro perspectives, this study divided various types of companies into four categories according to the degree of impact of the epidemic.

Step 4: Since the COVID-19 epidemic is a major emergency that seriously affects social and economic development, the state and governments at all levels have introduced a variety of support policies to focus on helping industries in the hardest-hit areas of the epidemic, especially to help small and medium-sized enterprises overcome difficulties. In terms of loans, banks are also required to increase credit support and the central government to arrange interest subsidy funds and other measures. When considering lending strategies, banks need to consider both national policy inclinations and the maximum profit of banks. Therefore, this study introduces market risk function, policy influencing factors and other indicators, and considers credit risk, market risk, national policy and social benefits at the same time, establishes a multi-objective nonlinear programming function, and solves it, and compares it with the results of credit risk analysis of 302 enterprises under the total credit amount strategy of 100 million yuan, and obtains the credit adjustment strategy of the bank when the annual credit limit is 100 million yuan. On the basis of Objective 2, this article considers factors such as national policy bias and the degree of influence on various industries, and quantifies them, adding them to the loan strategy adjustment model to obtain the final multi-objective nonlinear programming model. The constraints that the model needs to meet are the same as Objective 2. The objective function of this model is as follow:

$$maxZ = \sum_{j=1}^{302} y_j[A_j \cdot (1 + h_{1j})(1 - R_j)i_j \cdot (1 - h_{2j}) - A_j \cdot 0.22 \cdot (1 - h_{2j})R_j] - M_j \cdot q_j$$

h_{1j} is the coefficient of increase in loan quotas for different enterprises by banks, which is related to factors such as national policy subsidies, the industry of the enterprise, the scale of the enterprise, and customer churn rate, h_{2j} is the interest rate discount coefficient of the bank for different enterprises, and the influencing factors are the same as above; The constraints that need to be met are:

$$\sum_{j=1}^{302} y_i A_j \leq d$$
$$10 \leq A_j \leq E_j$$
$$0.04 \leq i_{kj} \leq 0.15$$
$$t_j = A_j \cdot i_j \cdot P_j - A_j \cdot 0.22 \cdot (1 - P_j)$$
$$y_j = \{ \begin{matrix} 0, t_j \leq 0 \\ 1, t_j > 0 \end{matrix}$$
$$q = f(i)$$
$$M = \eta(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11})$$
$$R = \phi(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11})$$
$$P + R = 1$$

M, P, R, E are Representing customer churn value, enterprise repayment probability, enterprise default probability, and maximum loan limit respectively. Finally, formulate credit strategies for 302 enterprises without credit records.

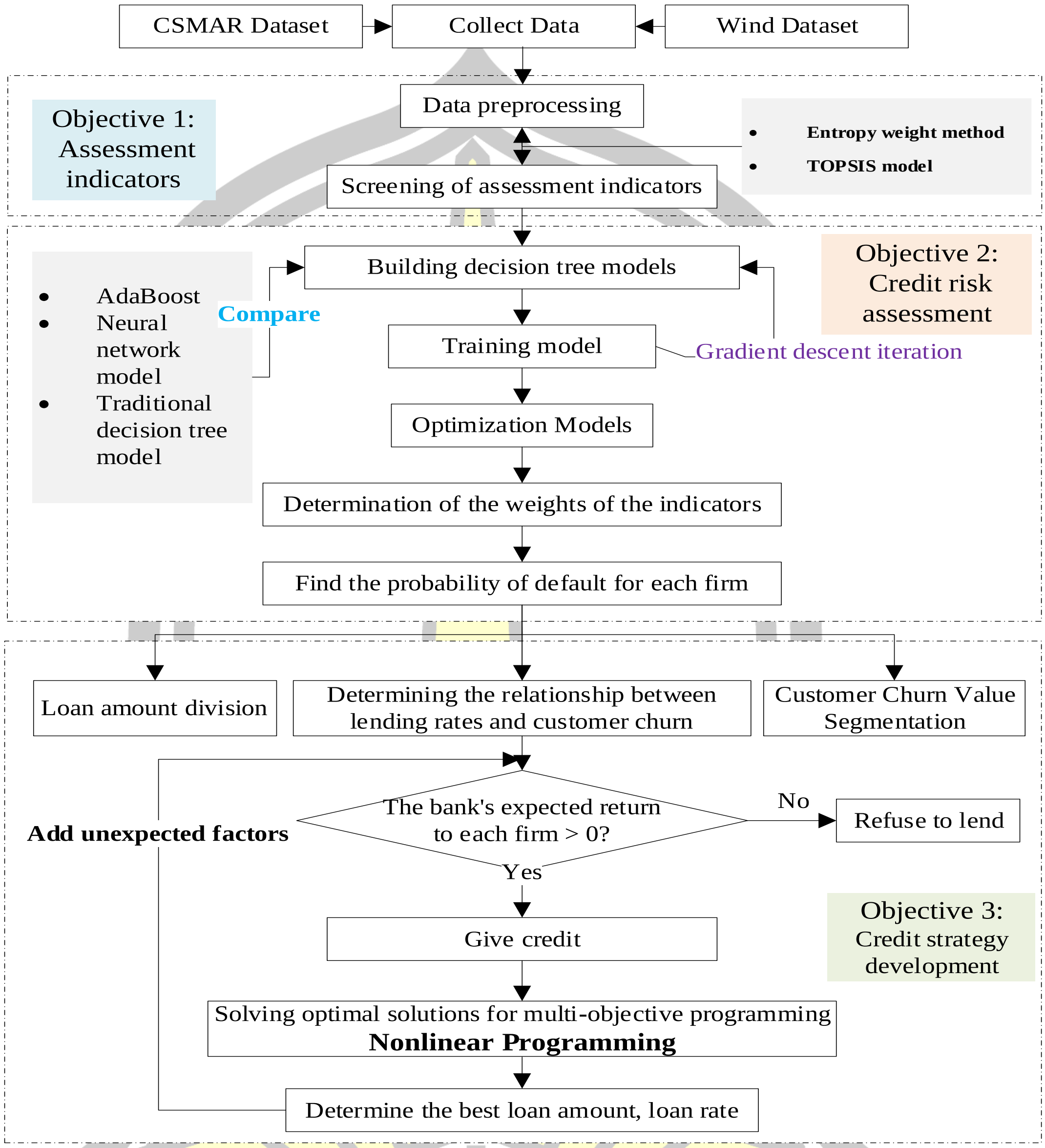

Figure 6 Map of the research methodology

Figure 6 provides a detailed overview of the operational procedures and analytical methodologies utilized in this study.

3.3 Work Flowchart

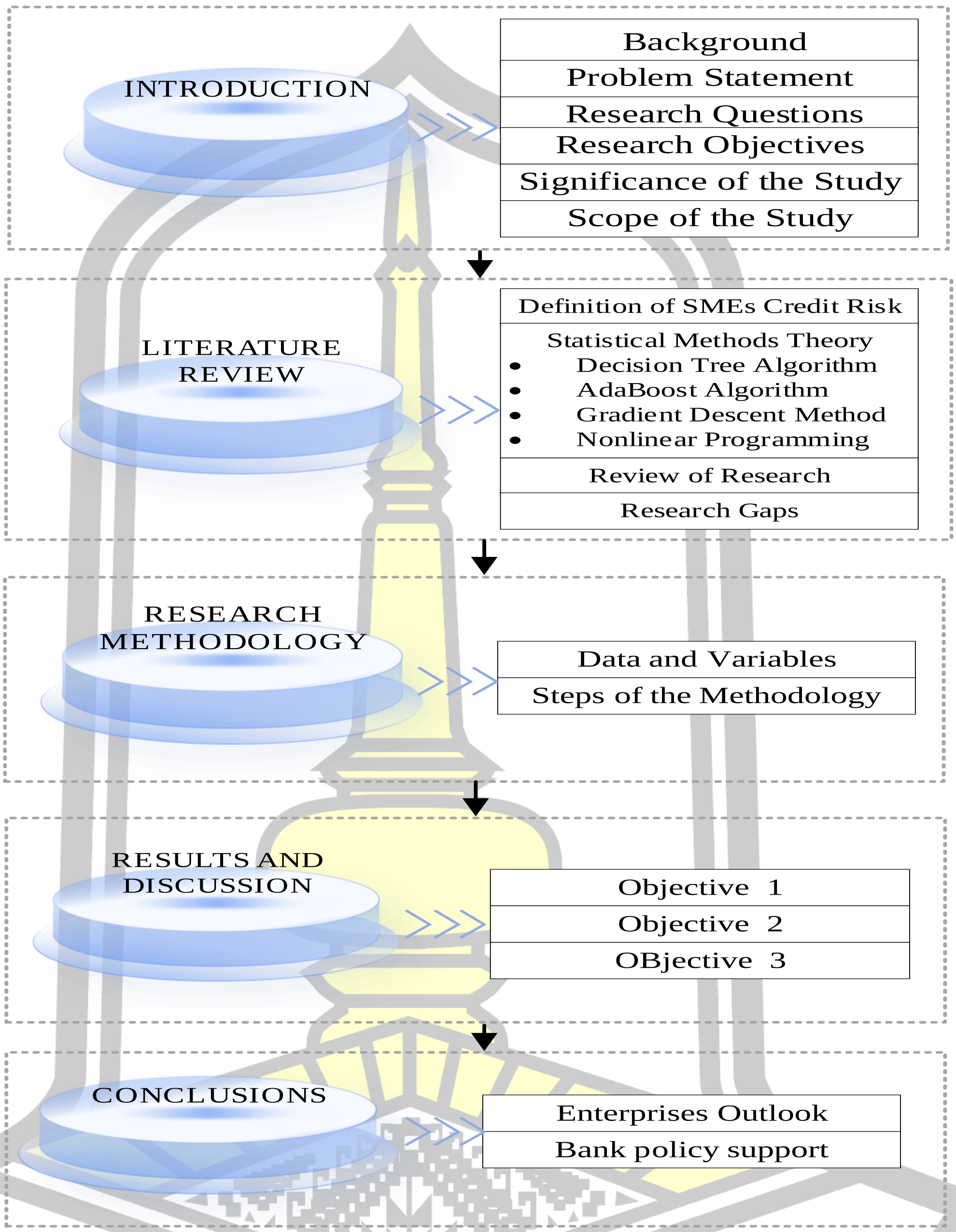

Figure 7 Work Flowchart

Figure 7 provides a detailed overview of the research work in this study. Mainly including research background, literature review, model construction, result analysis, conclusion and discussion.

Chapter 4
RESULTS AND DISCUSSION

This chapter presents the research findings, detailing the development and measurement of the credit risk indicator system. It offers a thorough analysis of credit strategies based on the risk profiles of SMEs. Additionally, the chapter explores the impact of unforeseen factors, such as economic disruptions or global crises, on credit strategies and risk assessments. Drawing from the results, the discussion proposes effective strategies that optimize the balance between risk and return for both SMEs and financial institutions.

4.1 Construction and quantification of credit risk indicator system

4.1.1 Data analysis and processing

(1) Data analysis

The data for enterprises with credit records is processed and analyzed, focusing on two key aspects: credit records and invoice information. The credit records are categorized into four levels: A, B, C, and D, with grade A representing the highest creditworthiness and grade D the lowest. The default record indicates whether an enterprise has ever defaulted on any obligation. Regarding invoice information, it is divided into input and output invoices. Input invoices refer to those issued by upstream enterprises when goods are purchased, while output invoices are issued by downstream enterprises when products are sold. Invoices are further classified into valid and invalid categories. Valid invoices represent successful transactions, whereas invalid invoices signify transactions that were canceled for various reasons, which may be tied to the enterprise's reputation. Additionally, some valid invoices are marked as negative, indicating that although a transaction initially occurred, it was later reversed, possibly due to issues related to the enterprise's reputation.

Specifically, the information contained in the invoice includes invoice number, invoice date, transaction unit code, amount, tax amount, price and tax total, and invoice status. The invoice date can be used to analyze the company's operating time and related time series data, the transaction unit code can be used to analyze the number and strength of the company's upstream and downstream companies, the amount and tax amount can be used to analyze the company's scale, strength and repayment ability, and the invoice status can be used to analyze the company's reputation. There are a total of 210947 input invoices and 162484 output invoices.

According to the credit level of the enterprise, the record of breach of contract, and the business time of the enterprise, the credibility status of an enterprise can be roughly judged. The article first analyzes the credit rating distribution of 123 companies with credit records.

①Enterprise's credibility distribution situation

According to the credit level of the enterprise, the record of breach of contract, and the business time of the enterprise, the credibility status of an enterprise can be roughly judged. This article first analyzes the credit rating distribution of 123 companies with credit records.

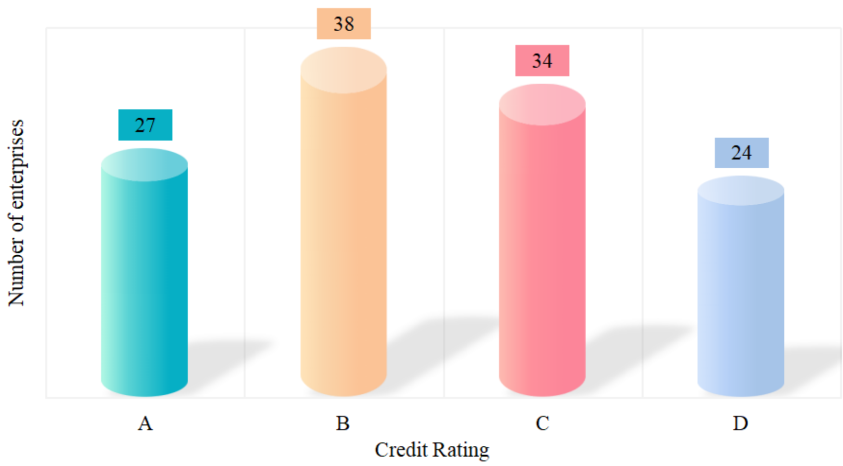

Figure 8 Credit level distribution

It can be seen through the distribution diagram of the credit grade that the distribution of four types of enterprises is relatively balanced. Among them, level B credit companies are the most, which shows that the reputation between enterprises is large. Large differences. Secondly, the article analyzes the operating duration of 123 credit records.

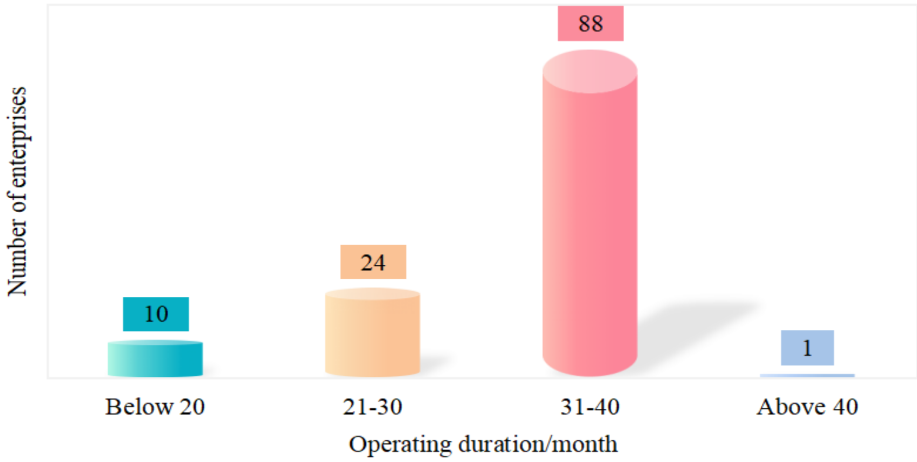

Figure 9 Distribution of operation duration

It can be seen from the distribution map of the operation duration that the operation time of most enterprises is within the scope of 31-40 months, that is, about three years or even less, and there is only one enterprise operating more than 40 months. It reflects the characteristics of weak qualifications for small and medium - sized enterprises and high operating risks.

②Enterprise profit distribution

Based on the total profit of the enterprise, it can be judged that the business status of an enterprise, and more importantly, the ability to repay the loan. Therefore, paying attention to the profit status of enterprises can effectively judge credit risk. Analysis of the profit of 123 companies.

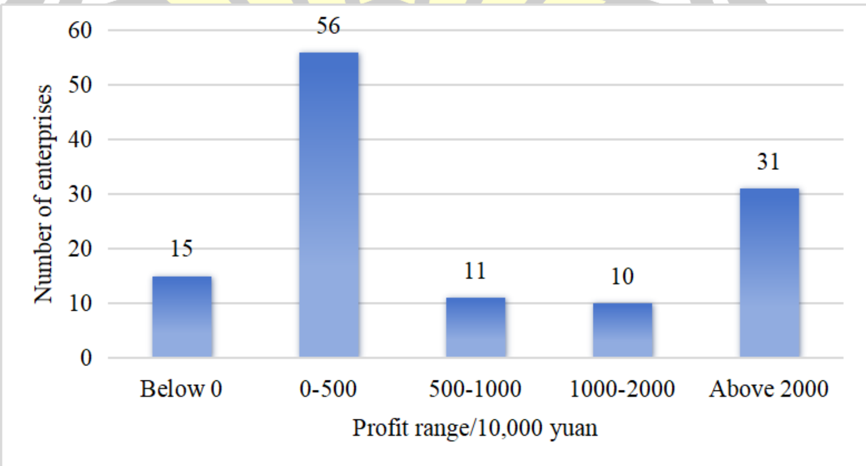

Figure 10 Profit range of the enterprises

It can be seen from Figure 10 that most of the companies' profits are greater than 0, and they are in a profitable state, and the profit of 31 companies has reached more than 20 million yuan, which shows that overall views of small and medium -sized enterprises have strong repayment capabilities, credit risk is relatively compared low.

③Enterprise scale distribution

The scale of an enterprise serves as an indicator of its financial strength. The greater the company's financial resources, the lower the likelihood of default. The total value-added tax (VAT) is often closely tied to the primary business income of the enterprise. As such, this study uses the total VAT paid by the enterprise as a proxy for measuring the company's financial strength.

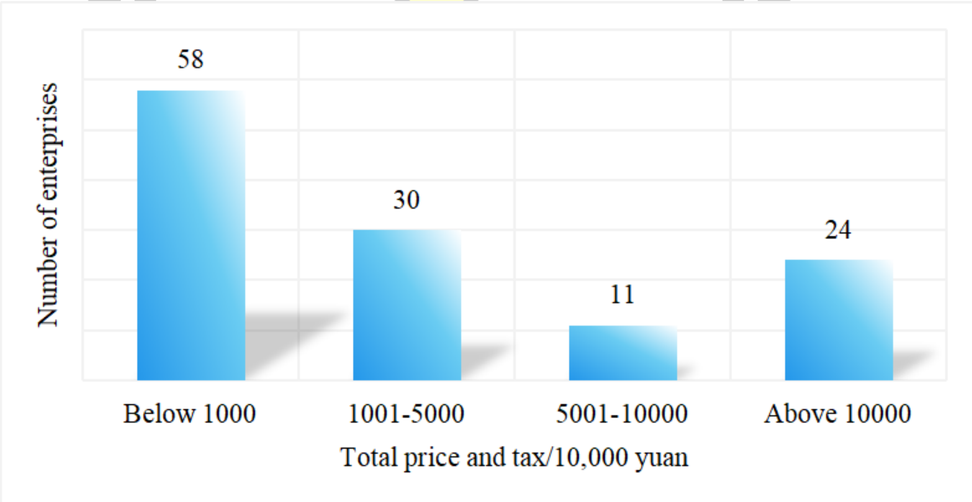

Figure 11 Total output price and tax

It can be seen from Figure 12 that the total output and price tax of most enterprises is less than 10 million, which reflects the small and medium -sized enterprises' small -scale characteristics. The scale indicates that the risk of defaults with these enterprises is low, which should be the main object of bank loan.

④Enterprise fluctuation level distribution situation

Whether the corporate fund fluctuations are stable is also an important aspect for banks to measure corporate credit risk. The smaller the degree of capital fluctuations, the more stable the business of the enterprise is, and it is not easy to go bankrupt to form bad debts. This article mainly observes the degree of volatility of the amount from the mutant coefficient of the monthly entry amount. The statistical results are shown in Figure 13.

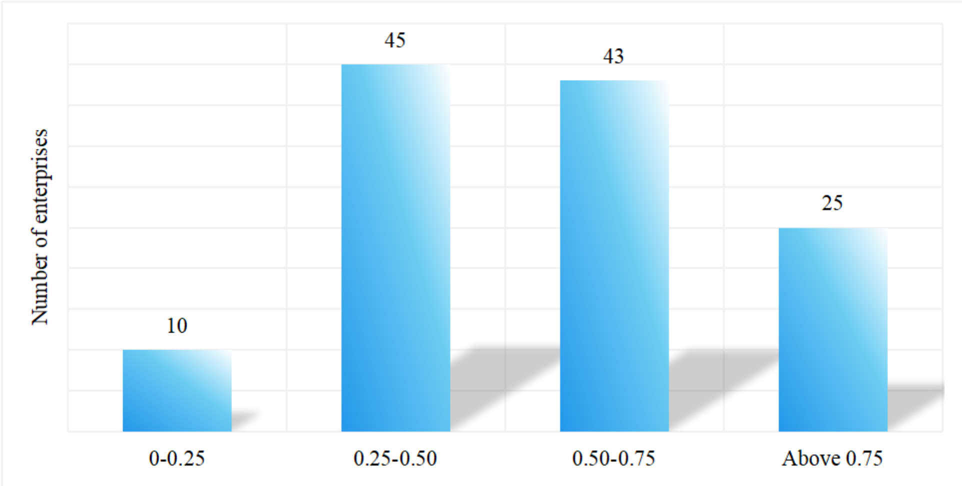

Figure 12 Mutant coefficient distribution of input amount

Generally speaking, the mutation coefficient is less than 0.25 that the data is relatively stable, but the amount of input amount of most enterprises that can be discovered is high, indicating that the flow of the enterprise is relatively unstable and has a large credit risk. There are only 10 companies from 10 companies. The amount of input amount is relatively stable.

⑤The distribution of the number of corporate enterprises

Through the upstream and downstream companies contacted by the enterprise, the influence of an enterprise in the industrial chain. The more companies they contacted, the greater the influence of this enterprise. Essence More upstream companies can improve the stability of the supply chain. When a supplier has a problem, the enterprise can obtain raw materials or services from other suppliers to reduce the risks caused by the interruption of supply. Having multiple suppliers can enhance the bargaining ability of the enterprise, as companies can use competition to reduce procurement costs. More downstream enterprises means that the company's market is more diversified, reducing dependence on single customers, thereby reducing financial risks. Diversified customer bases can provide enterprises with a more stable source of income. Even if some customers lose, the overall income can still remain stable.

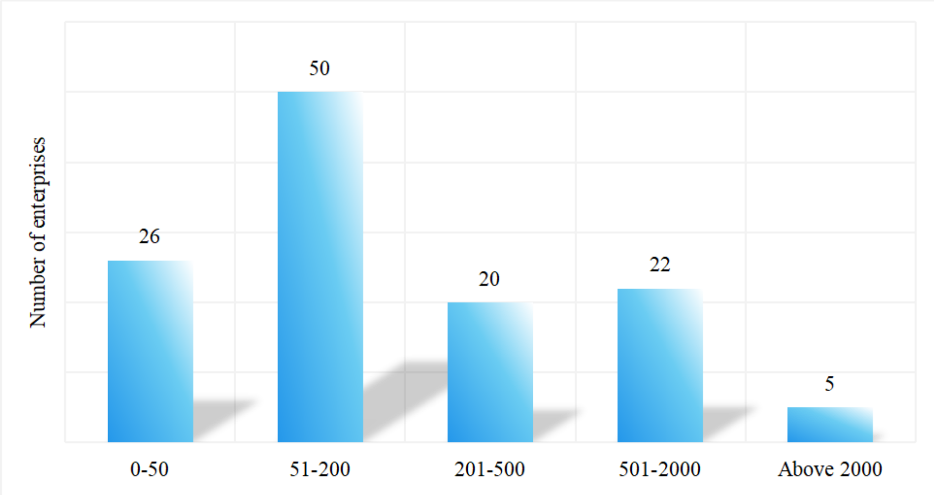

Figure 13 The number of upstream and downstream enterprises

(2) Data processing

According to the above analysis results, the amount of invoice data is huge, and the data must be simplified. Since invalid invoices are invoices for unfinished transactions, they can be eliminated and not included in the final data of each enterprise. In addition, since negative invoices are returns of transactions, they should also be eliminated during calculation. In theory, negative invoices should correspond to previous transactions and then be deleted accordingly. If there is no corresponding transaction invoice, it can be considered that the enterprise has evaded taxes, which will affect its credibility and need to be evaluated in subsequent calculations.

Since loans are not granted to D-level enterprises in principle, all D-level enterprises can be eliminated during calculations. Finally, after data processing, the simplified input invoices and output invoices greatly simplify the analysis difficulty.

Finally, the article processes the outliers that appear and replaces them with the sequence mean. For example, the profit of some credit A enterprises is negative 1.6 billion yuan, which is obviously unreasonable.

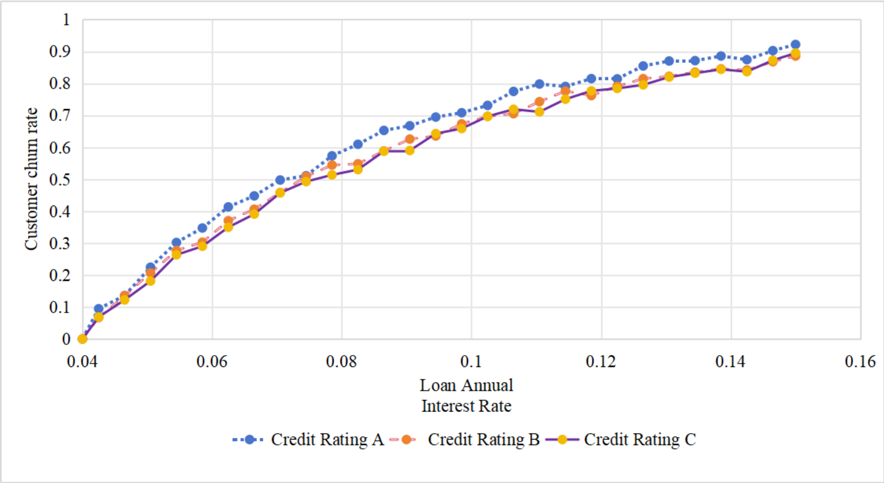

Figure 14 The relationship between loan interest rates and customer loss rates under different credit levels

It can be seen from Figure 15 that the interest rate and customer loss rate are positively correlated, and at the same interest rate level, the loss rate of different reputation rating companies is different, and the loss rate of Class A enterprises is higher.

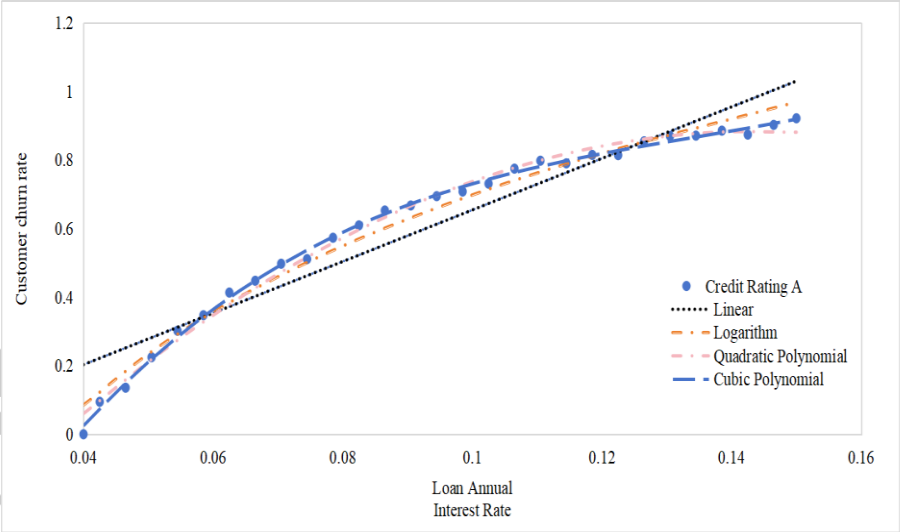

Figure 15 Customer loss rate (credibility rating A) and the annual interest rate relationship of loans

Observation can find that the customer loss rate L_r is a functional value depending on the credibility rating and loan interest rate. Considering that the minimum value is 0 due to variables (customer loss rates), the minimum value is 0. Unable to apply the number conversion. It is impossible to calculate the composite model, power model, S model, growth model, index distribution model, and logistic, so we mainly use the pair, one, two, and three models to fit the fitting test. Analysis of the impact of the annual interest rate of the loan on the customer's loss rate in the case of credibility rating, and obtained the regression function (Figure 16). Model Summary and Parameter Estimates is shown in Table 7.

Table 7 Model Summary and Parameter Estimates

Fitting Equation	**Model Summary**			**Parameter Estimates**			
	R^2	**F**	**Sig.**	**constant**	b_1	b_2	b_3
Linear	0.911	276.616	0.000	-0.098	7.524	/	/
Logarithm	0.982	1500.776	0.000	2.239	0.669	/	/
Quadratic polynomial	0.993	1847.861	0.000	-0.697	21.984	-76.410	/
Cubic Polynomial	**0.998**	3690.626	0.000	-1.121	37.970	-258.570	640.944

After observation, you can know that the decideral coefficient R^2 of the four fit models is 0.911, 0.982, 0.993, and 0.998, which meets the requirements of less than 0.05. (Tables are displayed as 0) The best. After analysis and comparison, the final cubic polynomial fitting equations are finally selected as its regression function:

$$\text{A}:L_{r1} = 640.944r^3 - 258.570r^2 + 37.970r - 1.121$$

In the same way, the impact of analyzing the annual interest rate of loans on credibility rating in B and C, respectively.

$$\text{B}:L_{r2} = 552.829r^3 - 225.051 + 33.995r - 1.017$$

$$\text{C}:L_{r3} = 504.717r^3 - 207.386r^2 + 32.157r - 0.973$$

In the above-mentioned forms, L_r represents the fitting estimation value of the loss rate; r represents the loan interest rate.

4.1.2 Selection of enterprise characteristic variables and construction of indicator evaluation system

The focus of the article is how to extract the operating characteristics that can measure the credit risk of the enterprise from the invoice information and credit records of the enterprise, and then quantify the credit risk of the enterprise based on these operating characteristics. According to the question, banks often provide loans to enterprises with strong strength and stable supply and demand relationships. In addition, characteristics such as repayment ability and high reputation are also important factors that affect whether banks lend to enterprises. The following article will start from these four aspects to select indicators.

(1) Enterprise strength

The strength of an enterprise can be judged based on its size and transaction volume. This article selects the total input price and tax, total output price and tax, and total number of valid invoices as indicators to measure the size and transaction volume of an enterprise.

① Total input price tax represents the overall value of products acquired by an enterprise within a given time frame. A higher total indicates a larger production and operational scale for the business. This metric serves as a useful indicator for assessing the size of an enterprise's production capacity. By analyzing the total input price tax, one can gain insights into the scale of an enterprise's operations and its level of economic activity. The specific formula for calculating this value is as follows:

$$TX_{mi} = \sum_{k=1}^{n_m} tx_{mik}.$$

Among them, TX_{mi} is the total input price and tax of the i enterprise, is the input price and tax of the k input invoice of the enterprise, and n_m is the number of input invoices.

The meaning of total sales price tax is similar to that of total input price tax. The specific calculation formula is:

$$TX_{ei} = \sum_{k=1}^{n_e} tx_{eik}.$$

Among them, TX_{ei} is the total sales price tax of the i enterprise, tx_{eik} is the sales price tax of the k sales invoice of the enterprise, and n_e is the number of sales invoices.

② The total amount of valid invoices refers to the combined value of valid sales and purchase invoices issued by a company over a specified period. A higher total indicates a greater volume of legitimate transactions conducted by the company, reflecting a larger overall transaction volume. This metric serves as a reliable indicator for assessing the company's transactional activity and business scale. By analyzing the total amount of valid invoices, one can gauge the extent of a company's operations. The specific calculation formula for determining this value is as follows:

$$n_i = n_{ei} + n_{mi}.$$

Among them, n_{mi} and n_{ei} are the number of valid invoices for purchase and sales respectively, and n_i is the total number of valid invoices for the i enterprise.

(2) Supply and demand stability

The stability of the supply and demand relationship refers to the consistency and reliability of the connections between an enterprise and its upstream and downstream partners. This stability can be assessed by examining fluctuations in transaction volumes between the enterprise and its suppliers and customers. Smaller fluctuations indicate a more stable supply-demand relationship. Additionally, the stability can be evaluated by considering the number of upstream and downstream enterprises the company interacts with. A larger number of contacts suggests more purchasing and sales channels, contributing to a more stable relationship. Thus, this paper uses two key indicators to measure supply and demand stability: the coefficient of variation for both monthly input and monthly sales amounts, and the number of upstream and downstream enterprises the company engages with.

① The coefficient of variation of the monthly input amount represents the variability in a company's monthly input expenditures, calculated on a month-to-month basis. This coefficient serves as a measure of the degree of fluctuation within the data, providing insight into the stability of the company's financial activities. By calculating the coefficient of variation, comparisons can be made between different companies, offering a standardized way to assess variability. A higher coefficient of variation indicates greater volatility in the company's monthly input amounts, suggesting more instability in the supply and demand dynamics. This measure helps in understanding the extent to which a company's financial activities fluctuate over time. The specific formula for calculating this coefficient is as follows:

$$c_{mi} = \frac{s_{mi}}{\bar{x}_{mi}}.$$

Where c_{mi} is the coefficient of variation of the monthly input amount of the i enterprise, s_{mi} is the standard deviation of the monthly input amount of the enterprise.

The calculation formula is $s_{mi} = \sqrt{\frac{\sum_{i=1}^{n}(x_{mi}-\bar{x}_{mi})^2}{n-1}}$, $\bar{x}_{mi}$ is the average monthly input amount, and the calculation formula is $\frac{1}{n}\sum_{i=1}^{n} x_{mi}$.

② The concept of the monthly sales amount variation coefficient is similar to that of the monthly input amount variation coefficient. It measures the extent of fluctuation in a company's monthly sales figures, providing an indicator of how stable or variable the company's sales performance is over time. Just like the input amount variation coefficient, this metric helps to assess the degree of volatility in sales, which can reflect underlying changes in market demand, pricing strategies, or other business factors. A higher variation coefficient indicates greater sales instability, suggesting that the company's revenue generation may be subject to larger fluctuations. The specific formula used to calculate this coefficient is as follows:

$$c_{ei} = \frac{s_{ei}}{\bar{x}_{ei}}.$$

Where c_{ei} is the coefficient of variation of the monthly sales amount of the i enterprise, s_{ei} is the standard deviation of the monthly sales amount of the enterprise, and $\bar{x}_{ei}$ is the mean monthly sales amount.

③ The number of upstream and downstream enterprises an enterprise engages with is determined by counting the total number of distinct businesses involved in the sales and purchase transactions reflected in the company's invoices. A higher number of connected upstream and downstream enterprises indicates a more extensive and diversified supply chain, which generally suggests a more stable supply and demand relationship. The greater the network of these interactions, the less dependent the enterprise is on any single supplier or customer, leading to reduced vulnerability to supply chain disruptions. Therefore, a robust network of upstream and downstream partners can contribute to greater operational stability for the enterprise. Define the number of upstream enterprises contacted by the ith enterprise as Q_{si}, and the number of downstream enterprises contacted as Q_{xi}.

(3) Ability to repay loan

The ability of an enterprise to repay a loan is directly related to its profitability. Only enterprises that can continue to make profits can repay their loans and avoid bad debt risks. This article selects the total profit of an enterprise as an indicator to measure the ability of an enterprise to repay a loan. The specific calculation formula is as follows:

$$\pi_i = TX_{ei} - TX_{mi} - T_i.$$

Among them, π_i is the total profit of the i enterprise, and T_i is the tax payable by the enterprise, which is equal to the output tax minus the input tax.

（4）Credibility

The creditworthiness of an enterprise is the most important factor in determining its credit risk. Creditworthiness refers to the possibility of an enterprise complying with a contract. The higher the creditworthiness of an enterprise, the lower its credit risk. Creditworthiness can be measured by indicators such as effective invoice rate, operating time, and credit rating.

① The effective invoice rate represents the percentage of valid invoices relative to the total number of invoices issued by an enterprise. A higher rate of effective invoices indicates greater reliability and credibility of the enterprise, as it reflects a higher volume of legitimate, completed transactions. This metric serves as an important indicator of the company's business integrity and operational transparency. The greater the proportion of valid invoices, the more trustworthy the enterprise is perceived to be, which can also positively impact its reputation and creditworthiness. The specific calculation formula for determining this rate is as follows:

$$R_i = \frac{n_i}{n_i'}.$$

Among them, R_i is the effective invoice rate of the enterprise, n_i is the number of effective invoices, and n_i' is the total number of invoices.

② The operating time of an enterprise is another key factor in assessing its reputation. Typically, a longer operational history is associated with a stronger reputation, as it suggests greater experience, stability, and trustworthiness in the market. In this study, the operating time of an enterprise is calculated by determining the difference between the date of its first invoice and the date of its most recent invoice. This method provides a clear indication of the enterprise's duration of activity and serves as a useful proxy for its operational stability and reputation in the industry. The longer the time span between the first and last invoices, the more established and reliable the enterprise is likely to be perceived, and defines the operating time of the i enterprise as t_i.

③ Credit rating is a reputation indicator that is comprehensively evaluated by bank professionals based on various indicators of the enterprise, which can effectively measure the reputation of the enterprise. In order to facilitate quantitative analysis, this paper converts the four ratings into numerical values, as shown in Table 8.

Table 8 Rating conversion

Credit Rating	A	B	C	D
Corresponding Value	80	60	40	20

In addition, enterprises with a record of default will be given a downgrade penalty and their credit scores will be reassessed. In summary, this article summarizes 11 indicators for assessing four aspects of corporate credit risk. The four aspects are strength, supply and demand relationship, repayment ability, and credibility. The 11 indicators are the total price and tax of import and export items, total profit, monthly import and export amount variation coefficient, total number of valid invoices, proportion of valid invoices, total number of upstream and downstream enterprises, and operating time. In this way, an enterprise credit risk assessment indicator system can be constructed, as shown in Figure 17.

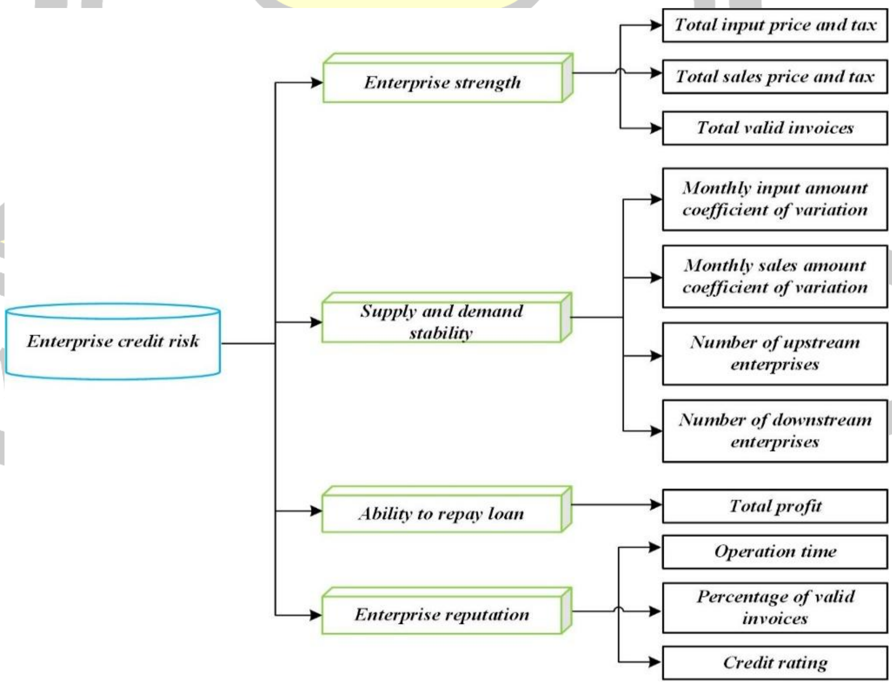

Figure 16 Enterprise Credit Risk Index System

Therefore, this paper obtains an enterprise credit risk indicator system, and then the credit risk of each enterprise can be quantified based on this system.

4.1.3 Quantifying corporate credit risk

(1) Entropy weight method result

Using the values of the various indicators and the evaluation system developed earlier, the credit risk of each enterprise can be quantified. The first step in this process is to assign appropriate weights to each indicator. To minimize the potential for subjectivity in the weighting process and ensure more objective results, this study utilizes the entropy weight method, which is a data-driven approach for calculating the weights. Once the weights are determined, the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method is employed to assess the credit risk of each enterprise. This approach helps quantify the relative credit risk of enterprises by comparing their performance against ideal and worst-case scenarios. The weighted results for the 11 indicators are presented in Table 9.

Table 9 Index weight results

Index	***Total input price and tax***	***Total sales price and tax***	***Total valid invoices***	***Monthly input amount coefficient of variation***	***Monthly sales amount coefficient of variation***	
Weights	0.3078	0.1819	0.0011	0.0018	0.0058	
Index	***Number of upstream enterprises***	***Number of downstream enterprises***	***Total profit***	***Operation time***	***Percentage of valid invoices***	***Credit rating***
Weights	0.1017	0.0023	0.0939	0.1679	0.1110	0.0248

（2）TOPSIS method to quantify corporate credit risk results

Finally, we obtained the quantitative credit risks of 123 enterprises. Due to space limitations, we only show the credit risks of the top 10 and bottom 10 enterprises here (Table 10). For specific results, see Appendix Table A.1.

Table 10 Quantified value of credit risk of some Companies

Company Code	**Credit Risk**	**TOPSIS Score**	**Company Code**	**Credit Risk**	**TOPSIS Score**
C1	0.3299	0.6701	C114	0.9979	0.0021
C2	0.7828	0.2172	C115	0.9987	0.0013
C3	0.8921	0.1079	C116	0.9980	0.0020
C4	0.8327	0.1673	C117	0.9987	0.0013
C5	0.9646	0.0354	C118	0.9938	0.0062
C6	0.9078	0.0922	C119	0.9976	0.0024
C7	0.8750	0.1250	C120	0.9982	0.0018
C8	0.8098	0.1902	C121	0.9934	0.0066
C9	0.9414	0.0586	C122	0.9961	0.0039
C10	0.9468	0.0532	C123	0.9984	0.0016

The article compares the results of enterprise quantified risk with the credit rating provided by the original data to facilitate verification of the rationality and accuracy of the model.

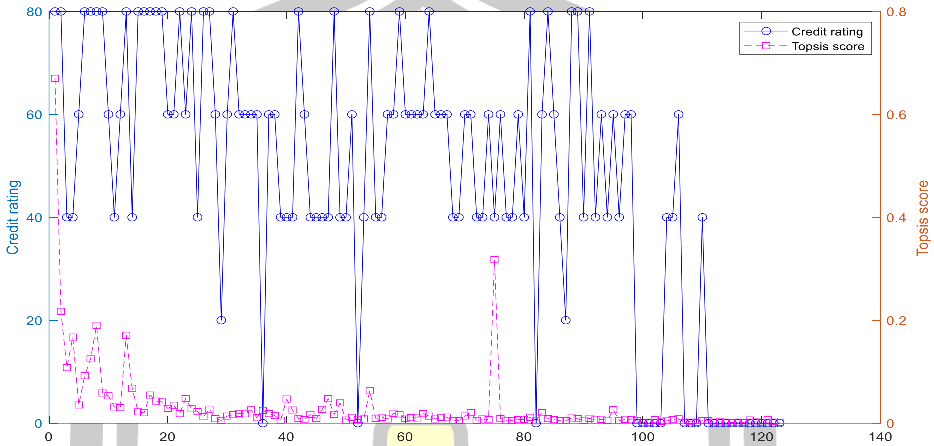

Figure 17 Credit rating vs. TOPSIS score

As shown in Figure 17, the changing trends of the two variables are basically the same, and the degree of fit is very high, which shows that the credit risk calculated in this paper is very accurate and the credit strategy can be analyzed based on this data.

4.2 Credit strategy analysis

4.1.1 Simulation

Accuracy is a direct indicator, but it may not be accurate enough when the data is unbalanced. Area Under Curve (AUC) provides a more comprehensive view of the model's discriminative ability and is suitable for scenarios with imbalanced datasets and multiple thresholds that need to be considered.

Divide the training set and test set into 70%, 80%, and 90% ratios respectively, and ensure that the test set is consistent each time. Use 70% of the training set to train the three models of Gradient Descent Decision Tree Algorithm an, AdaBoost algorithm, and a different decision tree algorithm, and record the training time and performance of each model. Repeat the above steps and use 80% and 90% of the training set to train models Decision Tree Algorithm, AdaBoost Algorithm, and Gradient Descent Decision Tree Algorithm respectively. In the test set, calculate the prediction results of each model, generate the Receiver Operating Characteristic (ROC) curve, and calculate the AUC value. Record the AUC values of each model under different training set ratios. Compare the AUC values of the three models under different training set ratios, and analyze their performance in unbalanced data. The simulation results are shown in Figure 18~20.

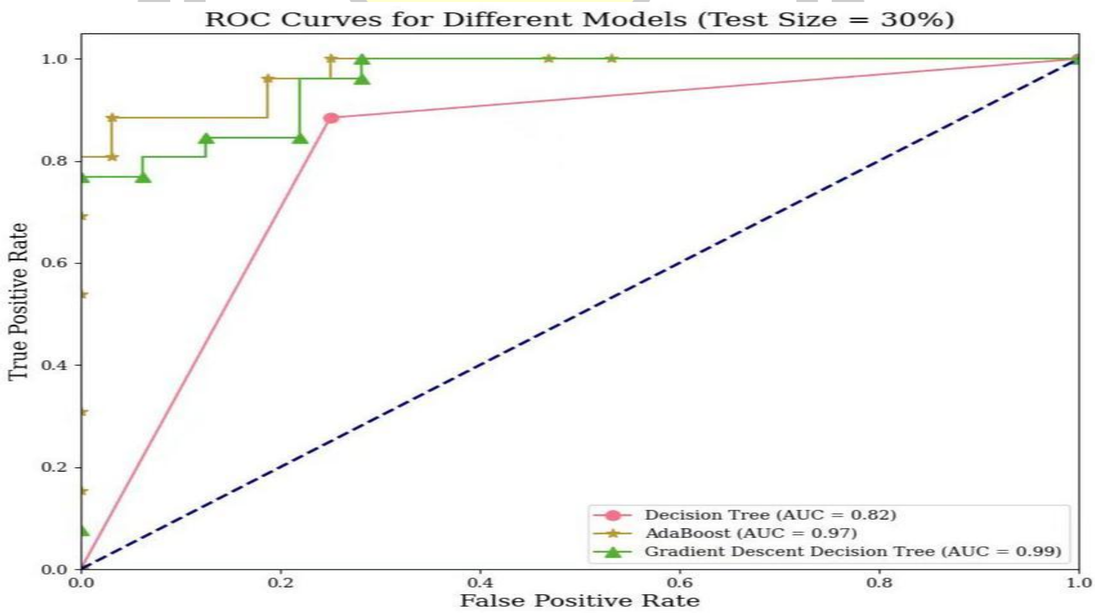

Figure 18 Comparison of results of different models with 70% training set, 30% test set

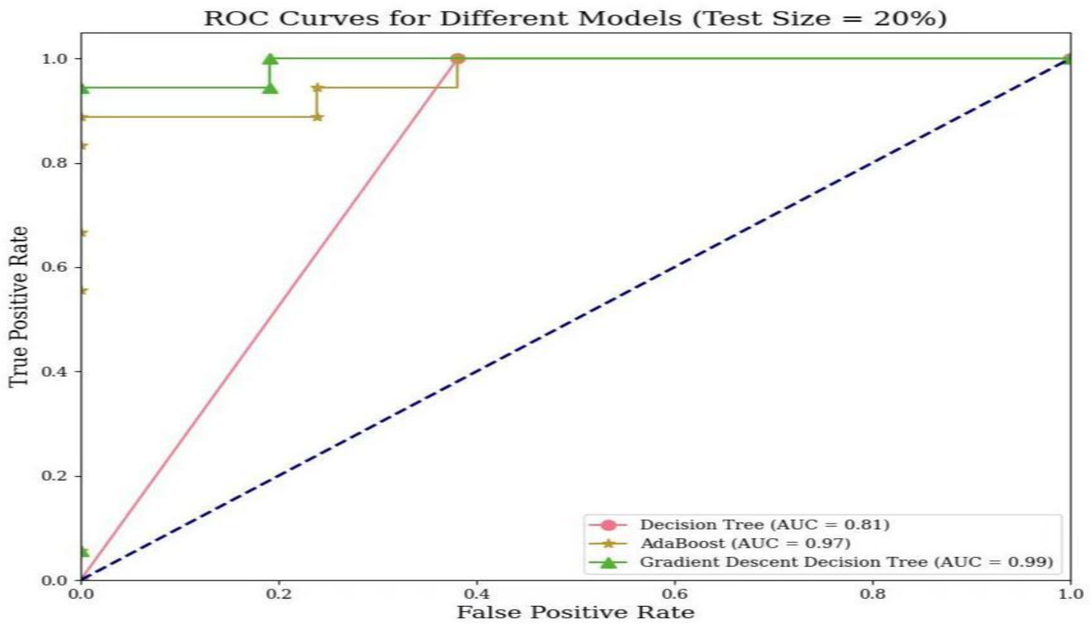

Figure 19 Comparison of results of different models with 80% training set, 20% test set

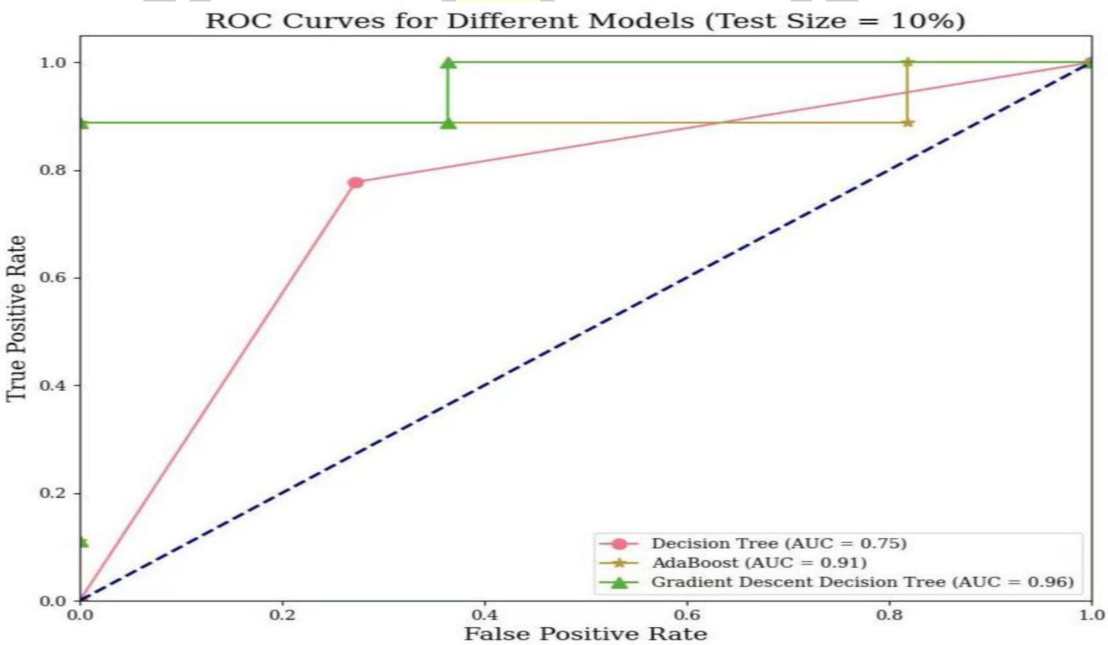

Figure 20 Comparison of results of different models with 90% training set, 10% test set

From the simulation results, we can see that the AUC value of the area under the ROC curve of Gradient Descent Decision Tree Algorithm reaches 0.99. At the same time, the AUC value of the AdaBoost algorithm is 0.97, and the AUC value of the traditional decision tree algorithm is 0.82, and The Gradient Descent Decision Tree Algorithm works well. Next, I will use the Gradient Descent Decision Tree Algorithm to calculate the default probability of 302 companies without credit records.

4.1.2 Results of default probability and credit rating of 302 enterprises

Through the Gradient Descent Decision Tree Algorithm, the results of default probability and credit rating of 302 enterprises are calculated and shown in Table 11.

Table 11 default probability and credit rating of 302 enterprises

Company Code	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*
D1	NO	A	D77	NO	B	D153	NO	B	D229	NO	D
D2	NO	A	D78	NO	B	D154	NO	B	D230	YES	C
D3	NO	B	D79	NO	B	D155	NO	B	D231	NO	A
D4	NO	A	D80	NO	B	D156	NO	C	D232	YES	B
D5	NO	C	D81	NO	A	D157	NO	B	D233	NO	B
D6	NO	A	D82	NO	B	D158	NO	B	D234	NO	D
D7	NO	B	D83	NO	C	D159	NO	A	D235	YES	B
D8	NO	C	D84	NO	B	D160	NO	A	D236	NO	B
D9	NO	A	D85	NO	B	D161	NO	B	D237	YES	C
D10	NO	B	D86	NO	A	D162	NO	C	D238	YES	C
D11	NO	A	D87	NO	B	D163	YES	B	D239	NO	C
D12	NO	B	D88	NO	A	D164	NO	B	D240	YES	D
D13	NO	B	D89	NO	A	D165	NO	B	D241	NO	A
D14	NO	B	D90	NO	C	D166	NO	A	D242	NO	C
D15	NO	A	D91	NO	B	D167	NO	B	D243	NO	C
D16	NO	A	D92	NO	B	D168	YES	B	D244	NO	C
D17	NO	B	D93	NO	B	D169	NO	A	D245	YES	C
D18	NO	A	D94	YES	B	D170	NO	C	D246	NO	A
D19	NO	B	D95	YES	B	D171	NO	B	D247	YES	B
D20	NO	C	D96	NO	C	D172	NO	B	D248	YES	D
D21	NO	C	D97	NO	C	D173	NO	C	D249	NO	B
D22	NO	C	D98	NO	C	D174	NO	B	D250	NO	D
D23	NO	A	D99	NO	C	D175	NO	A	D251	NO	C
D24	NO	A	D100	NO	C	D176	NO	C	D252	NO	B

Company Code	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*
D25	NO	B	D101	NO	B	D177	NO	B	D253	NO	D
D26	NO	B	D102	NO	A	D178	NO	B	D254	YES	D
D27	NO	A	D103	NO	B	D179	NO	B	D255	NO	B
D28	NO	B	D104	NO	B	D180	NO	B	D256	YES	B
D29	NO	A	D105	NO	C	D181	NO	B	D257	YES	C
D30	NO	C	D106	NO	B	D182	NO	B	D258	NO	D
D31	NO	A	D107	NO	B	D183	NO	B	D259	NO	D
D32	NO	B	D108	NO	B	D184	NO	C	D260	NO	C
D33	NO	C	D109	NO	B	D185	NO	B	D261	YES	D
D34	NO	B	D110	NO	B	D186	NO	B	D262	YES	C
D35	NO	B	D111	NO	B	D187	NO	B	D263	YES	C
D36	NO	B	D112	NO	B	D188	NO	B	D264	NO	C
D37	NO	A	D113	NO	B	D189	NO	C	D265	NO	B
D38	NO	B	D114	NO	B	D190	NO	C	D266	NO	D
D39	NO	A	D115	NO	A	D191	NO	B	D267	NO	C
D40	NO	B	D116	NO	B	D192	NO	A	D268	NO	B
D41	NO	A	D117	YES	B	D193	NO	B	D269	NO	B
D42	NO	A	D118	NO	B	D194	NO	B	D270	NO	B
D43	NO	A	D119	YES	B	D195	NO	B	D271	YES	D
D44	NO	C	D120	NO	B	D196	NO	A	D272	YES	D
D45	NO	B	D121	NO	B	D197	NO	C	D273	NO	D
D46	NO	A	D122	NO	C	D198	NO	C	D274	YES	D
D47	NO	B	D123	NO	C	D199	NO	B	D275	NO	D
D48	NO	C	D124	NO	B	D200	YES	B	D276	YES	D
D49	NO	B	D125	NO	C	D201	NO	C	D277	NO	D
D50	NO	A	D126	NO	B	D202	NO	C	D278	YES	D
D51	NO	B	D127	NO	B	D203	NO	A	D279	YES	D
D52	NO	C	D128	NO	C	D204	NO	B	D280	YES	D
D53	NO	A	D129	NO	B	D205	NO	C	D281	YES	C
D54	NO	C	D130	NO	B	D206	NO	C	D282	YES	D
D55	NO	B	D131	NO	B	D207	NO	C	D283	YES	D

Company Code	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*	*Company Code*	*Default Situation*	*Credit Rating*
D56	NO	B	D132	NO	B	D208	NO	C	D284	YES	C
D57	NO	A	D133	NO	B	D209	NO	B	D285	YES	D
D58	NO	A	D134	NO	B	D210	NO	B	D286	NO	D
D59	NO	B	D135	NO	A	D211	NO	B	D287	NO	D
D60	NO	C	D136	NO	B	D212	NO	B	D288	NO	D
D61	NO	B	D137	NO	A	D213	NO	C	D289	NO	D
D62	NO	A	D138	NO	A	D214	NO	B	D290	YES	D
D63	NO	B	D139	NO	B	D215	NO	C	D291	YES	D
D64	YES	B	D140	NO	B	D216	NO	A	D292	NO	D
D65	NO	A	D141	YES	C	D217	YES	C	D293	YES	D
D66	NO	B	D142	NO	C	D218	NO	C	D294	NO	D
D67	NO	C	D143	NO	C	D219	YES	B	D295	NO	D
D68	NO	A	D144	NO	B	D220	NO	B	D296	NO	D
D69	NO	C	D145	NO	B	D221	NO	D	D297	NO	D
D70	NO	A	D146	NO	C	D222	NO	D	D298	NO	D
D71	NO	A	D147	NO	C	D223	NO	B	D299	NO	D
D72	NO	A	D148	NO	C	D224	YES	A	D300	YES	D
D73	NO	A	D149	NO	C	D225	NO	C	D301	NO	D
D74	NO	A	D150	NO	C	D226	NO	C	D302	YES	D
D75	NO	B	D151	NO	B	D227	NO	D			
D76	NO	B	D152	NO	B	D228	NO	D			

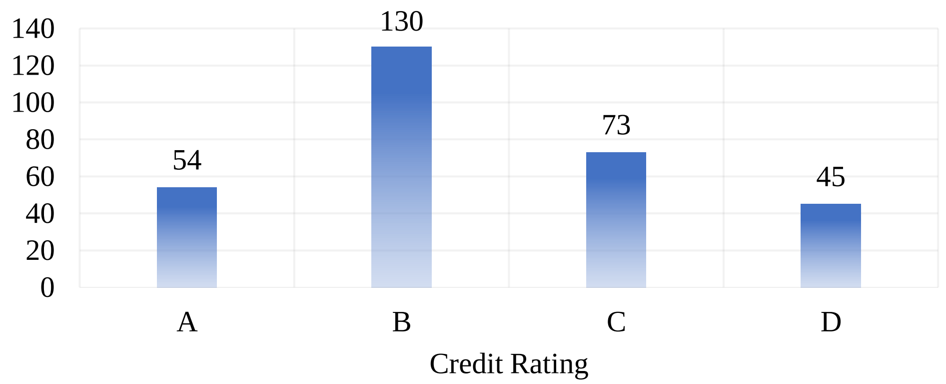

Figure 21 Credit level distribution

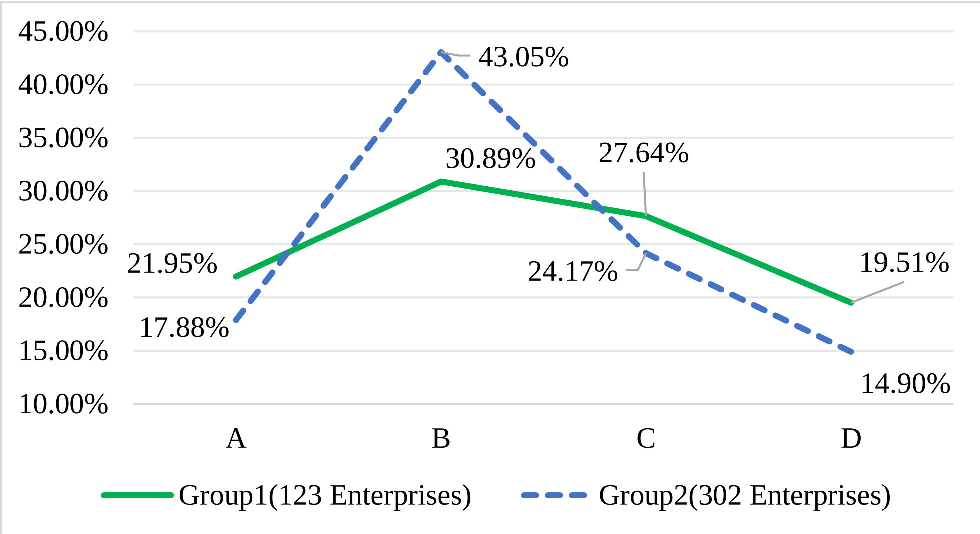

Figure 22 Proportion of different credit ratings in the two groups of samples

Table 11 presents a comprehensive overview of the default situations and credit ratings of a diverse set of companies, identified by unique company codes ranging from D1 to D302. Each company is assessed on two critical parameters: their current default situation, which indicates whether they have encountered a default event, and their credit rating, which reflects their creditworthiness.

The dataset reveals that the majority of enterprises have a default situation marked as "NO," signifying no occurrence of defaults. However, a smaller subset of companies is flagged with a "YES," indicating instances of default. This differentiation is vital for understanding the financial stability and risk profiles of the enterprises.

Credit ratings are categorized into several levels, including "A," "B," "C," and "D," which likely denote descending levels of credit quality or financial reliability. A noteworthy observation is that most enterprises fall under the "B" and "C" categories, suggesting moderate to lower credit quality. Companies rated "A" appear to exhibit higher creditworthiness, while those rated "D" may represent the highest risk of default.

Further, the interplay between default situations and credit ratings reveals intriguing insights. For instance, companies with a "YES" default situation predominantly fall within the lower credit rating tiers ("B," "C," and "D"), underscoring the correlation between poor credit ratings and higher default probabilities. Conversely, entities rated "A" generally exhibit "NO" default situations, reflecting stronger financial health.

4.1.3 Quantification results of credit risk of 302 enterprises

The research obtained the quantitative credit risks of 302 enterprises. Due to space limitations, the research only shows the credit risks of the top 10 and bottom 10 enterprises here (Table 12). For specific results, see Appendix Table A.2.

Table 12 Quantified value of credit risk of some Companies

Company Code	Credit Risk	Company Code	Credit Risk
D1	0.2859	D293	0.6876
D2	0.2859	D294	0.5727
D3	0.2773	D295	0.5572
D4	0.2727	D296	0.5243
D5	0.4307	D297	0.6930
D6	0.2602	D298	0.7101
D7	0.2818	D299	0.7146
D8	0.2754	D300	0.5616
D9	0.2870	D301	0.5724
D10	0.2857	D302	0.7061

The article compares the results of enterprise quantified risk with the credit rating provided by the original data to facilitate verification of the rationality and accuracy of the model.

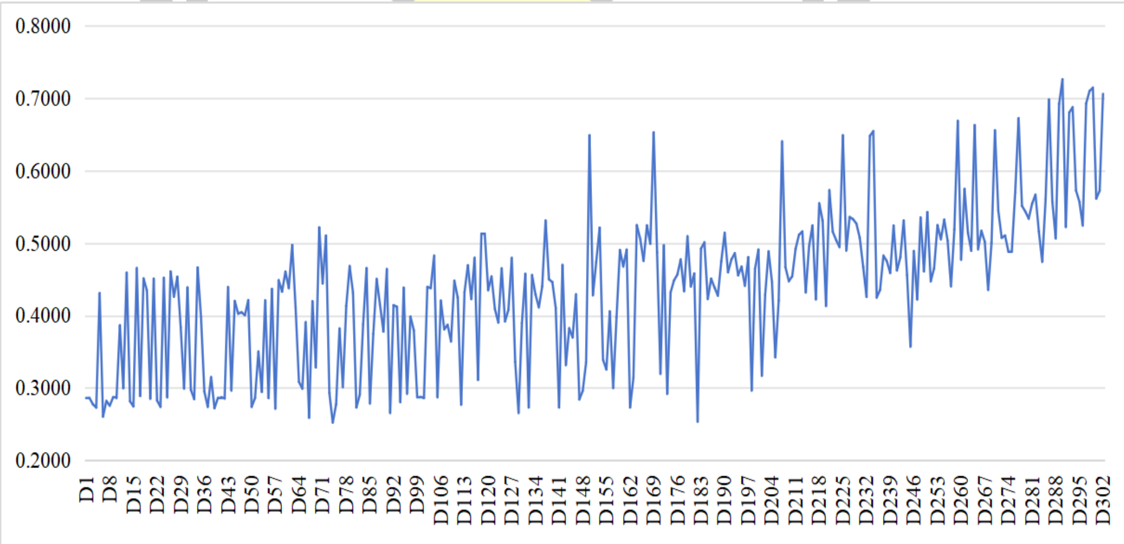

Figure 23 Credit risk of 302 enterprises

As shown in Figure 24, the variation in credit risk among a group of 302 enterprises, identified by their unique company codes (D1 to D302) on the x-axis. The y-axis represents the credit risk levels, with values ranging from approximately 0.2 to 0.8. The credit risk metric reflects the probability of default or the financial vulnerability of each enterprise.

The graph reveals significant fluctuations in credit risk across companies. In the initial section, the credit risk values exhibit relatively lower and consistent levels, indicating better financial stability among these enterprises. As the company codes progress, an upward trend becomes evident, with the amplitude of fluctuations also increasing. This suggests a higher degree of variability and a rise in credit risk for enterprises in the latter part of the dataset.

Notably, the graph includes sharp peaks and troughs at various intervals, highlighting significant disparities in risk levels even among closely positioned companies. The companies towards the far right demonstrate the highest credit risk values, surpassing 0.7, indicating a heightened likelihood of financial instability or default.

This visualization is a valuable tool for identifying patterns and outliers in credit risk distribution. Enterprises with extreme credit risk levels may warrant closer attention, either as potential default candidates or as examples of financial robustness. Such insights are critical for stakeholders in financial risk management, lending, and investment decision-making.

4.1.4 Credit strategies of 302 enterprises

The constructed model calculates the bank loan strategies of 302 enterprises, including whether to lend, loan amount and loan interest rate. The specific results are shown in Table 13.

Table 13 Credit strategies of 302 enterprises

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D1	Yes	11.67	6.53%	D152	Yes	63.83	15.00%
D2	Yes	21.07	5.47%	D153	Yes	40.76	9.50%
D3	Yes	21.06	5.78%	D154	Yes	33.04	4.96%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D4	Yes	21.05	6.27%	D155	Yes	40.76	9.50%
D5	Yes	21.06	5.78%	D156	Yes	34.01	14.99%
D6	Yes	21.05	6.27%	D157	Yes	13.06	5.77%
D7	Yes	21.05	6.44%	D158	Yes	40.76	9.50%
D8	Yes	21.05	6.27%	D159	Yes	73.02	4.33%
D9	Yes	21.05	6.27%	D160	Yes	33.04	4.79%
D10	Yes	21.06	5.78%	D161	Yes	33.05	4.65%
D11	Yes	41.04	4.79%	D162	Yes	13.05	6.27%
D12	Yes	61.03	4.13%	D163	Yes	33.05	4.61%
D13	Yes	61.03	4.07%	D164	Yes	40.76	9.50%
D14	Yes	61.04	4.01%	D165	Yes	13.07	5.65%
D15	Yes	21.06	5.78%	D166	Yes	33.04	4.96%
D16	Yes	21.05	6.27%	D167	Yes	33.04	4.96%
D17	Yes	61.02	4.33%	D168	Yes	40.76	9.50%
D18	Yes	41.04	4.96%	D169	Yes	40.76	9.50%
D19	Yes	41.04	4.79%	D170	Yes	40.76	9.50%
D20	Yes	61.03	4.13%	D171	Yes	33.03	5.86%
D21	Yes	41.04	4.79%	D172	Yes	20.77	9.50%
D22	Yes	41.04	4.79%	D173	Yes	33.09	4.38%
D23	Yes	61.04	4.01%	D174	Yes	40.76	9.50%
D24	Yes	71.03	4.12%	D175	Yes	40.76	9.50%
D25	Yes	41.04	4.96%	D176	Yes	33.04	4.95%
D26	Yes	71.03	4.13%	D177	Yes	40.76	9.50%
D27	Yes	71.04	4.01%	D178	Yes	40.76	9.50%
D28	Yes	71.03	4.13%	D179	Yes	40.76	9.50%
D29	Yes	41.04	4.79%	D180	Yes	40.76	9.50%
D30	Yes	21.05	6.27%	D181	Yes	40.76	9.50%
D31	Yes	71.03	4.12%	D182	Yes	33.04	4.96%
D32	Yes	21.07	5.65%	D183	Yes	40.76	9.50%
D33	Yes	21.86	14.99%	D184	Yes	20.77	9.50%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D34	Yes	41.84	14.99%	D185	Yes	40.76	9.50%
D35	Yes	71.04	4.01%	D186	Yes	40.76	9.50%
D36	Yes	21.06	5.77%	D187	Yes	20.77	9.50%
D37	Yes	71.03	4.12%	D188	Yes	33.05	4.69%
D38	Yes	71.02	4.33%	D189	Yes	40.76	9.50%
D39	Yes	71.97	15.00%	D190	Yes	40.76	9.50%
D40	Yes	71.02	4.33%	D191	Yes	40.76	9.50%
D41	Yes	21.05	6.27%	D192	Yes	40.76	9.50%
D42	Yes	41.04	4.79%	D193	Yes	40.76	9.50%
D43	Yes	41.8	14.99%	D194	Yes	20.77	9.50%
D44	Yes	41.04	4.95%	D195	Yes	40.76	9.50%
D45	Yes	71.03	4.13%	D196	Yes	40.76	9.50%
D46	Yes	71.03	4.13%	D197	Yes	20.77	9.50%
D47	Yes	41.04	4.79%	D198	Yes	40.76	9.50%
D48	Yes	41.04	4.96%	D199	Yes	33.88	14.99%
D49	Yes	71.03	4.13%	D200	Yes	40.76	9.50%
D50	Yes	41.04	4.95%	D201	Yes	40.76	9.50%
D51	Yes	41.04	4.79%	D202	Yes	40.76	9.50%
D52	Yes	41.04	4.95%	D203	Yes	40.76	9.50%
D53	Yes	71.02	4.33%	D204	Yes	33.03	5.86%
D54	Yes	71.02	4.33%	D205	Yes	20.77	9.50%
D55	Yes	41.04	4.78%	D206	Yes	40.76	9.50%
D56	Yes	71.03	4.13%	D207	Yes	40.76	9.50%
D57	Yes	41.05	4.65%	D208	Yes	40.76	9.50%
D58	Yes	41.04	4.78%	D209	Yes	33.04	4.96%
D59	Yes	41.05	4.65%	D210	Yes	40.76	9.50%
D60	Yes	41.04	4.79%	D211	Yes	40.76	9.50%
D61	Yes	41.05	4.65%	D212	Yes	40.76	9.50%
D62	Yes	71.03	4.13%	D213	Yes	40.76	9.50%
D63	No	41.04	4.79%	D214	Yes	40.76	9.50%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D64	No	0	9.50%	D215	Yes	20.77	9.50%
D65	No	0	4.76%	D216	Yes	40.76	9.50%
D66	No	0	4.79%	D217	Yes	33.1	4.48%
D67	Yes	41.04	4.96%	D218	Yes	40.76	9.50%
D68	Yes	71.02	4.40%	D219	Yes	40.76	9.50%
D69	Yes	41.04	4.79%	D220	Yes	40.76	9.50%
D70	Yes	21.05	6.27%	D221	No	0	14.99%
D71	Yes	41.04	4.96%	D222	No	0	9.50%
D72	Yes	71.03	4.12%	D223	Yes	20.77	9.50%
D73	Yes	41.04	4.96%	D224	Yes	40.76	9.50%
D74	Yes	71.03	4.13%	D225	Yes	40.76	9.50%
D75	Yes	41.99	14.99%	D226	Yes	40.76	9.50%
D76	Yes	41.84	14.99%	D227	No	0	9.50%
D77	Yes	21.06	5.75%	D228	No	0	9.50%
D78	Yes	41.05	4.65%	D229	No	0	9.50%
D79	Yes	21.05	6.27%	D230	Yes	40.76	9.50%
D80	Yes	41.04	4.79%	D231	Yes	0	9.50%
D81	Yes	71.02	4.33%	D232	Yes	40.76	9.50%
D82	Yes	21.05	6.27%	D233	Yes	40.76	9.50%
D83	Yes	71.04	4.01%	D234	No	0	9.50%
D84	Yes	21.06	5.77%	D235	Yes	33.06	4.49%
D85	Yes	21.06	6.23%	D236	No	0	9.50%
D86	Yes	41.05	4.65%	D237	Yes	40.76	9.50%
D87	Yes	41.04	4.79%	D238	Yes	40.76	9.50%
D88	Yes	21.06	5.78%	D239	Yes	40.76	9.50%
D89	Yes	41.04	4.96%	D240	No	0	9.50%
D90	Yes	71.03	4.12%	D241	Yes	40.76	9.50%
D91	Yes	41.04	4.96%	D242	Yes	40.76	9.50%
D92	Yes	41.03	5.02%	D243	Yes	40.76	9.50%
D93	Yes	41.04	4.79%	D244	No	0	9.50%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D94	Yes	22.77	9.50%	D245	Yes	33.06	4.49%
D95	Yes	41.03	5.00%	D246	No	0	9.50%
D96	Yes	41.14	14.99%	D247	Yes	40.76	9.50%
D97	Yes	71.03	4.12%	D248	No	0	9.50%
D98	Yes	41.04	4.96%	D249	No	0	9.50%
D99	Yes	71.04	4.01%	D250	No	0	9.50%
D100	Yes	71.03	4.12%	D251	No	0	9.50%
D101	Yes	71.02	4.31%	D252	Yes	40.76	9.50%
D102	Yes	41.04	4.79%	D253	No	0	9.50%
D103	Yes	41.04	4.78%	D254	No	0	9.50%
D104	Yes	41.05	4.75%	D255	Yes	33.03	5.18%
D105	Yes	81.83	15.00%	D256	Yes	40.76	9.50%
D106	Yes	41.04	4.95%	D257	Yes	33.05	4.65%
D107	Yes	41.04	4.96%	D258	No	0	4.45%
D108	Yes	41.03	5.00%	D259	No	0	9.50%
D109	Yes	41.73	14.99%	D260	Yes	33.06	4.49%
D110	Yes	41.03	5.02%	D261	No	0	9.50%
D111	Yes	41.04	4.79%	D262	Yes	40.76	9.50%
D112	Yes	21.06	5.78%	D263	Yes	13.1	4.95%
D113	Yes	21.05	6.27%	D264	Yes	40.76	9.50%
D114	Yes	21.07	5.47%	D265	No	0	9.50%
D115	Yes	21.86	14.99%	D266	No	0	9.50%
D116	Yes	21.07	5.41%	D267	Yes	40.76	9.50%
D117	Yes	28.77	9.50%	D268	No	0	9.50%
D118	Yes	21.05	6.27%	D269	Yes	40.76	9.50%
D119	Yes	28.77	9.50%	D270	No	0	9.50%
D120	Yes	81.02	4.33%	D271	No	0	9.50%
D121	Yes	21.05	6.27%	D272	No	0	9.50%
D122	Yes	41.84	14.99%	D273	No	0	9.50%
D123	Yes	41.04	4.96%	D274	No	0	9.50%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D124	Yes	41.04	4.79%	D275	No	0	9.50%
D125	Yes	71.04	4.01%	D276	No	0	9.50%
D126	Yes	41.04	4.96%	D277	No	0	9.50%
D127	Yes	41.04	4.96%	D278	No	0	9.50%
D128	Yes	41.08	4.37%	D279	No	0	4.53%
D129	Yes	41.04	4.79%	D280	No	0	9.50%
D130	Yes	41.05	4.65%	D281	No	0	9.50%
D131	Yes	41.04	4.96%	D282	No	0	9.50%
D132	Yes	21.05	6.27%	D283	No	0	9.50%
D133	Yes	41.06	14.22%	D284	Yes	33.06	4.49%
D134	Yes	41.04	4.79%	D285	No	0	9.50%
D135	Yes	71.02	4.33%	D286	No	0	9.50%
D136	Yes	71.02	4.33%	D287	No	0	9.50%
D137	Yes	71.04	4.01%	D288	No	0	4.61%
D138	Yes	41.04	4.79%	D289	No	0	9.50%
D139	Yes	21.06	5.77%	D290	No	0	9.50%
D140	Yes	41.04	4.79%	D291	No	0	9.50%
D141	No	0	9.50%	D292	No	0	9.50%
D142	Yes	41.05	4.75%	D293	No	0	9.50%
D143	Yes	41.04	4.96%	D294	No	0	9.50%
D144	Yes	71.02	4.33%	D295	No	0	9.50%
D145	Yes	41.04	4.96%	D296	No	0	9.50%
D146	Yes	41.04	4.96%	D297	No	0	9.50%
D147	Yes	21.06	5.78%	D298	No	0	9.50%
D148	Yes	82	15.00%	D299	No	0	9.50%
D149	Yes	21.14	4.71%	D300	No	0	9.50%
D150	Yes	21.05	6.27%	D301	No	0	9.50%
D151	Yes	41.04	4.79%	D302	No	0	9.50%

As shown in Table 13, 243 enterprises have obtained bank loans, accounting for 80.46% of the total number of enterprises, and the remaining 59 enterprises have not obtained loans, accounting for 19.54%. The loan decision tendency shows that banks are more strict in selecting enterprises, and may focus on certain specific industries or enterprise characteristics (such as profitability and asset quality). The average loan amount of enterprises that obtained loans was 331,100 yuan, the median was 407,600 yuan, and the distribution range of loan amounts was 0 yuan to 820,000 yuan. The proportion of enterprises with small loans (500,000 yuan) was the highest, reaching 85%, indicating that banks tend to provide enterprises with small-scale financial support with lower risks. The average level of loan interest rates was 7.37%, and the median was 6.27%. The highest loan interest rate reached 15%, and the lowest interest rate was 4.01%. Further analysis found that large loans are usually accompanied by lower interest rates, and small loan enterprises may bear higher interest costs.

4.3 Credit strategies under unexpected factors

4.3.1 The industry and scale classification of the enterprise

Because various types of emergencies may have different impacts on enterprises in different industries, different sizes, and different categories. It is necessary to classify various types of enterprises and study enterprises in different enterprises and industry categories. Based on the above classification, the statistics of the 302 enterprise classification results are shown in Table 14.

Table 14 The scale and industry statistics of 302 companies

Number	Industry category	Small enterprise	Micro - enterprise	Individual household	Total
1	Industry	9	21	3	33
2	Construction industry	5	33	0	38
3	Transportation industry	12	38	0	50
4	Retail industry	4	5	0	9
5	Agriculture, forestry and fisheries	2	12	0	14
6	Wholesale	3	3	0	6
7	Software and	12	10	0	22

Number	Industry category	Small enterprise	Micro - enterprise	Individual household	Total
	information technology service industry				
8	Property management	0	2	0	2
9	Information transmission industry	4	4	0	8
10	Accommodation	0	2	0	2
11	Leasing and business service industry	0	47	0	47
12	Catering industry	1	1	0	2
13	Real estate development and operation	0	3	0	3
14	Other unimpeded industries	0	12	54	66

From Table 14, we can see that most of the 302 enterprises are small or micro enterprises or individual business owners, with no medium-sized enterprises. Among them, micro enterprises in the Transportation industry, Construction industry, and Leasing and business service industry account for the highest proportion.

4.3.2 Analysis of the impact of unexpected factors on enterprises

Emergency is a general term for unexpected events. Such events usually cause huge economic losses or casualties, and even endanger the country's political security and economic security. Emergency events usually have the characteristics of incomplete information, suddenness, and high uncertainty. For small and medium-sized enterprises in the market economy environment, due to the lack of talent resources, cash flow shortages, and fragile capital chains, most small and medium-sized enterprises have low risk tolerance. If they do not get loans or external assistance when dealing with emergencies, it is difficult to overcome difficulties by themselves. The overall emergency factors can be divided into two categories: macro-level emergency factors and micro-level emergency factors. The former may often be caused by a chain reaction caused by major national or global political, economic, and public health events (such as Sino-US trade frictions, the outbreak of the new crown pneumonia epidemic, etc.), and the latter may be caused by a series of problems caused by the enterprise itself or the upstream and downstream supply chain (such as product quality defects, public relations crises, supply chain breaks, etc.). The various emergency factors are extremely complex. This study takes the outbreak of the new crown pneumonia epidemic as an example to analyze in detail the impact of this major public health emergency on different enterprises as shown in Figure 24.

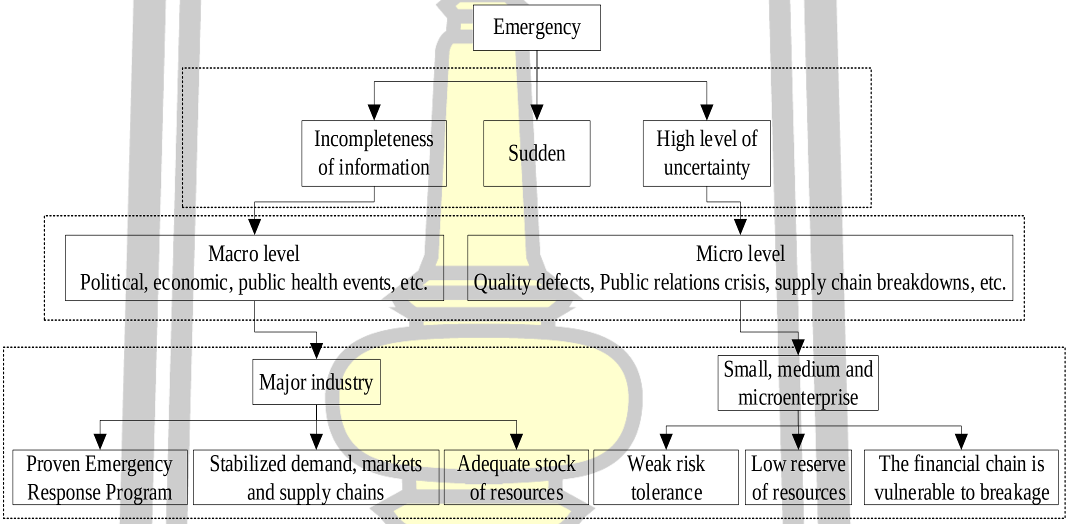

Figure 24 Emergency analysis process

In order to evaluate the impact of the COVID-19 epidemic on different industries and different types of enterprises, this study will analyze them from the macro and micro levels. At the macro level, the price fluctuations of the stock market can often reflect the development status of the entire industry very realistically and quickly, and can be regarded as the "barometer" of the development of each industry. Therefore, this study selects 1-3 typical listed company stocks in each industry as the analysis objects, and analyzes the stock price change trend, rise and fall of the target companies during the COVID-19 epidemic, so as to judge the impact of the COVID-19 epidemic on different industries. At the micro level, since the COVID-19 epidemic broke out in January 2020, and the data of 302 companies without credit records and credit ratings contain the invoice information of the companies from January to February 2020, the turnover of the 302 companies in January and February of each year is calculated, and the year-on-year growth analysis of the sales is carried out, and then the degree of impact of the COVID-19 epidemic on enterprises of different sizes and industries is analyzed. The specific idea is shown in Figure 25.

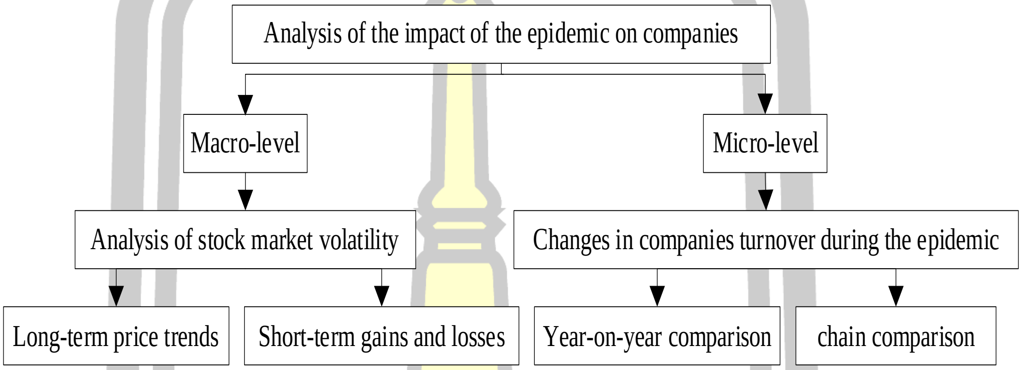

Figure 25 Analysis of the impact of the epidemic on enterprises

Through investigation and screening, this paper selects representative listed companies in various industries for analysis. This study uses well-known financial websites such as Sina Finance and East money to analyze the daily price change information of the above stocks. The outbreak of the epidemic was set to January 20, 2020. From January 20 to September 30, 2020, the maximum decline of the selected company's stocks, the time for the first recovery to rise, and the time required to restore the stock price before the epidemic were calculated. The maximum decline of the stock can reflect the risk tolerance of the company in the industry for public health events, the time for the first recovery to rise can reflect the market's confidence in the industry and the recovery of the industry itself, and the stock price recovery time can reflect the recovery of the company in the industry. The results of the calculation and analysis are shown in Table 15.

Table 15 Stock change information of the selected companies

Classification	Companies	Code	The beginning of the epidemic	First restoration date	Biggest decline	Remark
Industry	China Shenhua	sh.601088	January 20, 2020	July 6, 2020	18.87%	
Construction	China Communications Construction	sh.601800		March 2, 2020	17.92%	

Classification	Companies	Code	The beginning of the epidemic	First restoration date	Biggest decline	Remark
Transportation	Shanghai Port Group	sh.600018		September 11, 2020	28.16%	Not restored to original price
Retail	Heilan Home	sh.600398		September 11, 2020	26.18%	Not restored to original price
Agriculture, forestry, animal	Hainan Rubber	sh.601118		March 13, 2020	20.62%	
husbandry and fishery Wholesale	Suning.com	sz.002024		July 9, 2020	22.62%	
Software and information technology services	Hundsun Electronics	sh.600570		February 6, 2020	14.37%	
Property management	Nandu Property	sh.603506		February 19, 2020	18.30%	
Information transmission	Hengbao Shares	sz.002104		February 24, 2020	21.45%	
Accommodation	Huatian Hotel	sz.000428		March 2, 2020	17.94%	
Leasing and commercial services	CYTS	sh.601888		August 12, 2020	22.67%	
Catering	ST Yunwang	sz.002306		March 18, 2020	14.39%	
Real estate development and management	China Fortune Land Development	sh.600340		July 6, 2020	22.67%	

As shown in Table 15, the stock prices of companies in various industries fell to a large extent at the beginning of the outbreak, which was related to the huge panic caused by the unknown epidemic situation at that time. After a period of time, the stock prices of some industries rose rapidly and returned to the level before the epidemic, and even rose further, indicating that the industry was little affected by the epidemic, and even the epidemic had a positive effect on the development of the industry (such as software and information technology services, industry, construction, information transmission, etc.), while some industries were hit hard by the epidemic. This can be seen from two aspects of data. On the one hand, the maximum decline in stock prices of these industries during the epidemic was greater than that of other industries; on the other hand, the recovery speed of stock prices of companies in this industry (such as transportation, retail, leasing and business services, etc.) was also much slower than that of other industries, which shows that on the one hand, the industry has not fully recovered, and on the other hand, investors' investment information about the industry has not fully recovered.

Based on the above analysis, all companies are first divided into four levels according to the degree to which they are affected by the epidemic: severe, partially affected, no impact, and Positive. The specific classification results are shown in Table 16.

Table 16 Rating of the impact of the epidemic on various industries

Impact rating	Classification
Severe impact	Transportation, retail, leasing and commercial services
Partial impact	Catering, wholesale, agriculture, forestry, animal husbandry and fishery, accommodation, wholesale
No impact/Minimal impact	Industry, construction, real estate development, property management
Positive	Software and information technology services, information transmission

72 enterprises of various types were randomly selected from the 302 enterprises for analysis, and the year-on-year growth rates of profits in January and February 2020 compared with those in 2018 and 2019 were calculated. Some of the calculation results are shown in Table 17, and the complete data is shown in Appendix Table A.3.

Table 17 Comparison of the operating conditions of enterprises in January and February

Enterprise code	January		February		Enterprise Category	Enterprise Scale
	Sales	Profit	Sales	Profit		
	2018	2019	2018	2019		
D18	-80. 00%	-86.11%	38.81%	60.07%	Wholesale	micro
D20	-45.67%	-78.12%	-501.24%	10.51%	Software and information technology services	small
D39	-70. 14%	-79.29%	135.57%	-78.72%	Software and Information Technology Services	small
D40	-98. 53%	-99. 54%	101.52%	-99.48%	Transportation	small
D48	-99. 98%	-99. 98%	99.87%	99.06%	Other unspecified industries	micro
D52	-89. 04%	-94.47%	-89.04%	-94.47%	Wholesale	small
…	…	…	…	…	…	…
D287	-41.80%	-89.10%	106.47%	-89.10%	Software and Information Technology Services	micro
D294	-81.36%	-74.86%	-81.36%	-74.86%	Agriculture, forestry, animal husbandry and fishery	micro
D296	1603.33%	169.79%	1603.33%	169. 79%	Software and Information Technology Services	micro
D297	-37.20%	57.41%	-34.60%	67.07%	Retail	micro

It can be seen intuitively from Table 17 that the sales of most companies in 2020 have shrunk by 80%-90% compared with 2018 and 2019. Due to the impact of the epidemic, the input expenditure of small and medium-sized enterprises in 2020 has also decreased significantly. There are many companies that suffered losses in January and February 2018 or 2019. Due to the sharp decline in expenditure in 2020, they turned to profit or reduced losses in the same period of 2020. Therefore, there are a certain number of companies whose operating profits increased year-on-year in 2020. Therefore, it is more reasonable to analyze the year-on-year changes in turnover. In January 2020, the various companies were less affected by the epidemic, maintaining a general level of half of the companies with profit growth and half of the companies with profit reduction; while in February 2020, the impact of the epidemic on companies was clearly shown. It is estimated that about 80% of the companies have seen a year-on-year decline in sales. According to statistical results, the average sales shrinkage rate of enterprises whose sales decreased year-on-year compared with 2019 was as high as 86.94%. It can be seen that the business activities of most enterprises suffered serious damage during the epidemic. A considerable number of small and medium-sized enterprises will face the unfavorable situation of production stagnation and sales obstruction during the epidemic, and urgently need government policy support to overcome the difficulties.

4.3.3 Adjustment of credit strategy under unexpected factors

The production, operation and profit benefits of enterprises are often affected by the market environment, national policies and even some external sudden factors. These factors often have different impacts on enterprises in different industries and categories. The various impacts on enterprises will also be indirectly transmitted to banks through credit loans and other channels, thereby affecting the bank's capital security and profitability. According to existing domestic research, under the impact of the epidemic, the quality of bank loans has deteriorated. Due to the sharp decline in social and economic liquidity, and even when necessary, the "city closure" epidemic prevention measures have been carried out, which directly led to the obstruction of human resources and logistics transportation of small and medium-sized enterprises and production stagnation. At the same time, the epidemic has led to an increase in unemployment and a decrease in consumption capacity, which in turn has led to a sharp decrease in orders, a sharp decline in turnover, a decrease in repayment ability, and an increase in loan demand due to tight funds. Therefore, small and medium-sized enterprises have difficulty in repaying principal and interest, overdue loans have surged, subprime loans and non-performing loan rates have begun to rise, and banks' loan loss risks have increased.

During the epidemic, small and medium-sized enterprises were hit the hardest by the epidemic and suffered heavy losses. Many excellent companies in the past were even on the verge of bankruptcy during the epidemic. To this end, the state has introduced many support policies for small and medium-sized enterprises. According to the policy measures issued by various provinces, municipalities and autonomous regions, policy support is mainly provided to small and medium-sized enterprises from two aspects: reducing costs and providing financing. The main ways to reduce costs are tax reduction, lowering water and electricity prices, etc.; the main ways to provide financing are loan extension, lowering interest rates, lowering the threshold for small loans, etc. First of all, for these enterprises that are most severely affected by the epidemic, on the one hand, banks need to cooperate with national policies and actively provide more loan resources to related fields and related enterprises to help them overcome difficulties. On the other hand, they also need to consider the subsequent increase in loan delinquency and non-performing rates. Secondly, the credit focus of banks should also be tilted towards epidemic prevention and control related fields such as medical protection and people's livelihood security, meet the loan needs of enterprises, and support these enterprises to develop better. Finally, considering the long-term development and profitability of banks, banks are also looking for opportunities in the epidemic crisis and strategically deploying industries that may develop rapidly. For example, the epidemic may benefit and reconstruct certain industries. 5G industry, online education, cloud computing, online video, e-commerce, insurance industry, intelligent manufacturing and other industries will usher in new development opportunities, and banks should also seize new development opportunities for the future in a timely manner.

Based on the classification of the severity of the impact as described above, and taking into full consideration factors such as national policy inclination, bank profitability, and long-term development of banks, the specific adjustment strategies of banks are shown in Table 18.

Table 18 Specific adjustment strategies of banks

Impact rating	Loan limit adjustment policy	Loan interest rate adjustment policy
Severe impact	Loan limit increased by 30%	Loan interest rate discount 30%-50%
Partial impact	No change	Loan interest rate discount 10-30%
No impact/Minimal impact	No change	No change
Positive	Loan limit increased by 20%-50%	No change

Among them, for enterprises severely affected by the epidemic, according to national policies, appropriate preferential treatment and inclination should be given in terms of loan limit and loan interest rate. For enterprises that are less affected or not affected by the epidemic, the inclination should be appropriately reduced or even no preferential policies should be given. At the same time, based on long-term development and the interests of banks, for enterprises in certain favorable industries (such as cloud computing, e-commerce, intelligent manufacturing, etc.), the loan amount should be appropriately increased to improve customer retention rate and future profitability of banks. Based on the new bank loan adjustment policy and the solution of the multi-objective programming model, the loan strategies of 302 enterprises were calculated, including whether to lend, loan amount, loan interest rate and other information. The results are shown in Table 19.

Table 19 Adjustment of credit strategy under unexpected factors

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D1	Yes	10.01	6.75%	D152	Yes	26.04	4.90%
D2	Yes	11.13	5.50%	D153	Yes	92.69	7.22%
D3	Yes	11.13	5.50%	D154	Yes	61.81	9.62%
D4	Yes	11.13	5.45%	D155	Yes	25.97	6.44%
D5	Yes	11.13	5.44%	D156	Yes	43.81	9.62%
D6	Yes	11.14	5.43%	D157	Yes	45.11	7.22%
D7	Yes	11.13	5.44%	D158	Yes	6.13	5.45%
D8	Yes	11.13	5.43%	D159	Yes	26.05	4.84%
D9	Yes	11.13	5.43%	D160	Yes	95.99	2.02%
D10	Yes	28.81	9.62%	D161	Yes	43.97	5.23%
D11	Yes	49.02	2.31%	D162	Yes	25.99	5.74%
D12	Yes	70.98	4.62%	D163	Yes	6.13	5.43%
D13	Yes	81.3	9.62%	D164	Yes	43.81	9.62%
D14	Yes	70.99	4.50%	D165	Yes	43.81	9.62%
D15	Yes	11.13	5.45%	D166	Yes	23.81	9.62%
D16	Yes	11.13	5.53%	D167	Yes	26.01	5.32%
D17	Yes	70.99	4.48%	D168	Yes	25.97	6.47%
D18	Yes	31.05	3.91%	D169	Yes	43.81	9.62%
D19	Yes	31.04	4.92%	D170	Yes	26.04	4.94%
D20	Yes	90.99	8.93%	D171	Yes	61.81	4.81%
D21	Yes	49.02	2.31%	D172	Yes	43.94	3.14%
D22	Yes	49.02	2.32%	D173	Yes	23.81	9.62%
D23	Yes	71	4.37%	D174	Yes	26.03	5.01%
D24	Yes	71	4.37%	D175	Yes	25.95	7.68%
D25	Yes	49	2.44%	D176	Yes	44.02	2.25%
D26	Yes	71	4.46%	D177	Yes	26.04	4.91%
D27	Yes	71	4.37%	D178	Yes	25.98	5.88%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D28	Yes	90.91	2.85%	D179	Yes	26.05	4.84%
D29	Yes	49.02	2.34%	D180	Yes	43.99	2.44%
D30	Yes	11.13	5.46%	D181	Yes	26.05	4.86%
D31	Yes	90.99	2.52%	D182	Yes	26.04	4.88%
D32	Yes	11.12	5.56%	D183	Yes	26.04	4.91%
D33	Yes	12.22	14.42%	D184	Yes	43.98	2.61%
D34	Yes	32.13	14.43%	D185	Yes	23.81	9.62%
D35	Yes	71.01	4.36%	D186	Yes	44.02	4.45%
D36	Yes	11.11	5.72%	D187	Yes	43.81	7.70%
D37	Yes	71	4.41%	D188	Yes	6.13	5.37%
D38	Yes	70.98	4.67%	D189	Yes	44.02	2.31%
D39	Yes	102.1	14.45%	D190	Yes	44.02	2.25%
D40	Yes	100.99	2.17%	D191	Yes	43.94	3.26%
D41	Yes	11.13	5.44%	D192	Yes	44.01	2.36%
D42	Yes	49.02	2.32%	D193	Yes	26.01	5.31%
D43	Yes	32.15	14.43%	D194	Yes	43.98	2.56%
D44	Yes	31.04	4.91%	D195	Yes	6.13	5.43%
D45	Yes	70.99	4.53%	D196	Yes	43.95	5.68%
D46	Yes	71	4.36%	D197	Yes	44.01	2.34%
D47	Yes	49.02	2.32%	D198	Yes	23.81	9.62%
D48	Yes	31.04	4.90%	D199	Yes	26.04	4.89%
D49	Yes	90.99	3.06%	D200	Yes	45.12	14.44%
D50	Yes	49.02	2.34%	D201	Yes	43.81	9.62%
D51	Yes	31.04	4.91%	D202	Yes	26	5.41%
D52	Yes	31.03	4.00%	D203	Yes	44.01	2.33%
D53	Yes	90.99	4.01%	D204	Yes	61.81	9.62%
D54	Yes	71	4.36%	D205	Yes	43.81	9.62%
D55	Yes	49.01	2.35%	D206	Yes	23.81	9.62%
D56	Yes	71	3.49%	D207	Yes	44.02	4.47%
D57	Yes	31.04	4.90%	D208	Yes	26.04	3.91%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D58	Yes	31.02	4.09%	D209	Yes	43.81	7.70%
D59	Yes	49.02	4.65%	D210	Yes	43.98	2.60%
D60	Yes	31.04	4.89%	D211	Yes	40	2.38%
D61	Yes	31.04	4.90%	D212	Yes	43.81	9.62%
D62	Yes	71	4.38%	D213	Yes	26.02	5.17%
D63	No	0	3.93%	D214	Yes	44.01	2.33%
D64	No	0	4.81%	D215	Yes	26.04	3.96%
D65	No	0	4.95%	D216	Yes	23.81	9.62%
D66	No	0	4.90%	D217	Yes	61.81	4.81%
D67	Yes	31.04	4.91%	D218	Yes	26.03	5.00%
D68	Yes	90.99	2.08%	D219	Yes	61.81	4.81%
D69	Yes	66.81	4.81%	D220	Yes	43.81	9.62%
D70	Yes	28.81	9.62%	D221	No	0	4.84%
D71	Yes	31.03	5.05%	D222	No	0	4.81%
D72	Yes	100.99	2.03%	D223	Yes	43.81	9.62%
D73	Yes	49.01	2.40%	D224	Yes	18.03	5.67%
D74	Yes	71	4.36%	D225	Yes	43.81	9.62%
D75	Yes	32.07	14.43%	D226	Yes	26.04	4.93%
D76	Yes	50.12	7.22%	D227	No	0	7.70%
D77	Yes	23.06	2.60%	D228	No	0	3.34%
D78	Yes	31.04	4.90%	D229	No	0	4.81%
D79	Yes	23.06	2.56%	D230	Yes	43.96	2.73%
D80	Yes	31.04	4.90%	D231	Yes	26	5.43%
D81	Yes	71	4.36%	D232	Yes	0	4.81%
D82	Yes	23.06	5.21%	D233	Yes	44.01	2.30%
D83	Yes	71	4.36%	D234	No	0	2.34%
D84	Yes	11.13	5.45%	D235	Yes	25.99	5.67%
D85	Yes	11.13	5.48%	D236	No	0	2.32%
D86	Yes	31.05	4.89%	D237	Yes	61.81	4.81%
D87	Yes	31.04	4.89%	D238	Yes	43.92	3.91%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D88	Yes	11.13	5.44%	D239	Yes	43.99	2.51%
D89	Yes	31.05	4.89%	D240	No	0	5.93%
D90	Yes	100.99	2.26%	D241	Yes	43.81	9.62%
D91	Yes	49.01	4.68%	D242	Yes	24.98	6.11%
D92	Yes	31.05	4.89%	D243	Yes	43.95	5.72%
D93	Yes	31.04	4.89%	D244	No	0	9.62%
D94	Yes	11.66	14.33%	D245	Yes	43.81	9.62%
D95	Yes	30.99	5.61%	D246	No	0	4.81%
D96	Yes	31.79	14.40%	D247	Yes	0	9.62%
D97	Yes	51	4.36%	D248	No	0	4.45%
D98	Yes	49.02	4.63%	D249	No	0	9.62%
D99	Yes	71	4.37%	D250	No	0	9.62%
D100	Yes	90.99	4.55%	D251	No	0	9.62%
D101	Yes	100.99	4.01%	D252	Yes	0	9.62%
D102	Yes	31.04	4.91%	D253	No	0	2.31%
D103	Yes	31.04	4.90%	D254	No	0	4.81%
D104	Yes	31.04	4.93%	D255	Yes	43.81	9.62%
D105	Yes	72.1	14.45%	D256	Yes	61.81	4.81%
D106	Yes	49.01	4.69%	D257	Yes	61.81	4.81%
D107	Yes	49.02	2.32%	D258	No	0	2.32%
D108	Yes	49.01	2.36%	D259	No	0	9.62%
D109	Yes	48.81	9.62%	D260	Yes	61.81	9.62%
D110	Yes	49.01	2.34%	D261	No	0	4.81%
D111	Yes	49	2.43%	D262	Yes	35.81	4.81%
D112	Yes	23.06	2.57%	D263	Yes	43.81	9.62%
D113	Yes	11.14	5.42%	D264	Yes	6.12	5.62%
D114	Yes	11.11	5.81%	D265	No	0	9.62%
D115	Yes	12.22	14.42%	D266	No	0	7.70%
D116	Yes	11.09	6.29%	D267	Yes	61.81	4.81%
D117	Yes	28.81	9.62%	D268	No	61.81	4.81%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D118	Yes	11.11	5.71%	D269	Yes	23.81	9.62%
D119	Yes	28.81	9.62%	D270	No	0	2.48%
D120	Yes	71	4.39%	D271	No	0	9.62%
D121	Yes	23.06	5.12%	D272	No	0	4.81%
D122	Yes	32.07	14.43%	D273	No	0	9.62%
D123	Yes	49.02	2.31%	D274	No	0	4.81%
D124	Yes	31.03	5.03%	D275	No	0	7.70%
D125	Yes	71	4.42%	D276	No	0	4.81%
D126	Yes	49.02	4.63%	D277	No	0	9.62%
D127	Yes	31.04	4.97%	D278	No	0	4.81%
D128	Yes	48.81	9.62%	D279	No	0	9.62%
D129	Yes	31.04	4.94%	D280	No	0	4.81%
D130	Yes	31.04	4.97%	D281	No	0	9.62%
D131	Yes	31.04	4.90%	D282	No	0	9.62%
D132	Yes	11.14	5.43%	D283	No	0	4.81%
D133	Yes	31.68	14.46%	D284	Yes	43.93	3.13%
D134	Yes	31.02	5.11%	D285	No	0	5.76%
D135	Yes	71	4.41%	D286	No	0	4.81%
D136	Yes	100.98	2.14%	D287	No	0	9.62%
D137	Yes	100.99	4.08%	D288	No	0	9.62%
D138	Yes	48.99	4.94%	D289	No	0	9.62%
D139	Yes	11.11	5.79%	D290	No	0	4.81%
D140	Yes	31.04	4.89%	D291	No	0	9.62%
D141	No	0	9.62%	D292	No	0	4.81%
D142	Yes	31.04	4.91%	D293	No	0	4.81%
D143	Yes	31	5.43%	D294	No	0	9.62%
D144	Yes	71	4.36%	D295	No	0	7.70%
D145	Yes	31.04	3.95%	D296	No	0	4.81%
D146	Yes	48.92	3.90%	D297	No	0	9.62%
D147	Yes	11.1	5.99%	D298	No	0	4.81%

Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate	Company code	Lending or not	Amount (Ten thousand yuan)	Interest Rate
D148	Yes	72.09	14.45%	D299	No	0	7.70%
D149	Yes	23.03	5.67%	D300	No	0	4.81%
D150	Yes	11.13	5.46%	D301	No	0	9.62%
D151	Yes	21.13	5.46%	D302	No	0	9.62%

By comparing different loan strategies under normal conditions and under emergency situations, it can be found that the loan amount and interest rate in emergency situations show different trends depending on the type of enterprise, In Table 20, the results of the first 20 companies are displayed. For specific results, see Appendix Table A.4.

Based on the difference between the credit quota amount and the post-adjustment amount, this study examines the impact of unexpected situations on the loan acquisition amount of different enterprises. According to Table 20, not all enterprises experienced a reduction in loan acquisition amounts. This study defines a slight change as a fluctuation between 0 and 150,000 yuan, a moderate change as a fluctuation between 150,000 and 500,000 yuan, and a significant change as a fluctuation between 500,000 and 650,000 yuan. By calculating and categorizing the extent of changes among different enterprises, it was found that Slight Decrease accounted for the largest proportion, approximately 37%, followed by Slight Increase at around 24%. Moderate Decrease, Moderate Increase, and Significant Increase accounted for approximately 12%, 7%, and 1%, respectively.

Table 20 Comparison of loan amounts before and after emergencies (Unit: Ten thousand yuan)

Company code	**Normal(Q1)**	**After adjustment(Q2)**	**Q2-Q1**	**Loan category**
D1	11.67	10.01	-1.66	Slight Decrease
D2	21.07	11.13	-9.94	Slight Decrease
D3	21.06	11.13	-9.93	Slight Decrease
D4	21.05	11.13	-9.92	Slight Decrease
D5	21.06	11.13	-9.93	Slight Decrease
D6	21.05	11.14	-9.91	Slight Decrease
D7	21.05	11.13	-9.92	Slight Decrease
D8	21.05	11.13	-9.92	Slight Decrease
D9	21.05	11.13	-9.92	Slight Decrease
D10	21.06	28.81	7.75	Slight Increase
D11	41.04	49.02	7.98	Slight Increase
D12	61.03	70.98	9.95	Slight Increase
D13	61.03	81.3	20.27	Moderate Increase
D14	61.04	70.99	9.95	Slight Increase
D15	21.06	11.13	-9.93	Slight Decrease
D16	21.05	11.13	-9.92	Slight Decrease
D17	61.02	70.99	9.97	Slight Increase
D18	41.04	31.05	-9.99	Slight Decrease
D19	41.04	31.04	-10	Slight Decrease
D20	61.03	90.99	29.96	Moderate Increase

Based on the above changes in the degree of bank credit support for enterprises, we get the Credit rating cross-tab under normal and nonnormal circumstances in Table 21. Due to unexpected events, some of the companies with credit rating A were downgraded to credit rating B and C, with 2 and 4 companies respectively. Overall, companies with credit rating A have stronger risk resistance. Some of the companies that were originally rated B were downgraded to C, but at the same time a few were upgraded to A, with 17 and 8 companies respectively. Overall, companies with a credit rating of B have slightly inferior risk resistance to those with a credit rating of A. Seven companies that were originally rated C were downgraded to C, but at the same time, two companies were upgraded to A. As for enterprises with a credit rating of D, banks have always refused to provide loan support, so the number remains unchanged.

Table 21 Credit rating cross-tab under normal and nonnormal circumstances

Nonnormal / Normal	A	B	C	D	Total
A	48	2	4	0	54
B	8	105	17	0	130
C	0	2	64	7	73
D	0	0	0	45	45
Total	56	109	85	52	302

According to the results of the calculation, the credit lines of each enterprise are visually displayed in normal circumstances and emergencies, and the results are shown in Figure 26.

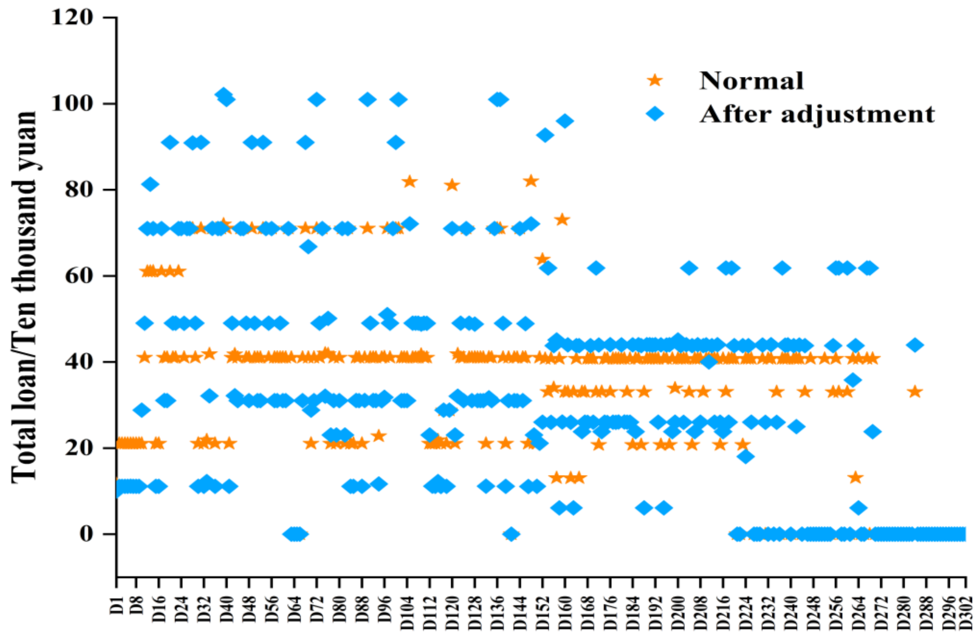

Figure 26 Comparison of loan amounts before and after emergencies

In the end, the bank's total loan amount was 100 million yuan, and the expected return was 5.4004 million yuan. Compared with normal circumstances, considering that the total loan amount of the bank's loan strategy after the policy subsidy did not change, the expected return on the loan was reduced by 1.3502 million yuan. However, this adjustment strategy can help many small and medium-sized enterprises in dire straits to overcome difficulties. It reflects the bank's great sense of responsibility and social responsibility, has produced huge social benefits, and further improved the bank's reputation and customer retention rate. From this perspective, this strategic adjustment has a positive impact.

Chapter 5
CONCLUSIONS

This dissertation provides a comprehensive exploration of how the Gradient Descent Decision Tree (GDDT) algorithm, when integrated with nonlinear programming techniques, can revolutionize the processes of credit risk assessment and credit strategy optimization in modern financial systems. The research is driven by the increasing complexities financial institutions face in assessing credit risk and formulating credit strategies, particularly for small and medium-sized enterprises (SMEs), which play a significant role in global economic development but often encounter significant barriers in obtaining financing. This work aims to bridge key gaps in traditional credit risk management methodologies, offering a robust, scalable, and adaptable framework to enhance credit assessment accuracy and optimize credit resource allocation.

5.1 Credit Risk Assessment in the Modern Financial Landscape

As the financial landscape becomes more intricate and dynamic, particularly with the advent of advanced technologies and data availability, the demand for more sophisticated tools for credit risk assessment has intensified. Traditional methods, which largely rely on financial indicators and expert judgment, often fail to fully capture the complexities inherent in credit evaluation, especially in cases involving SMEs. In response, this dissertation proposes the use of the Gradient Descent Decision Tree (GDDT) algorithm, a powerful machine learning model designed to improve the accuracy of credit risk predictions by modeling complex, nonlinear relationships within credit data.

The GDDT algorithm utilizes the power of gradient descent to iteratively optimize decision trees, allowing the model to learn and adapt to varying risk factors in the dataset. This iterative process improves the predictive accuracy of the model, making it more reliable in forecasting credit risks across a broad range of enterprises, especially those with limited credit history or non-standard financial profiles. One of the significant advantages of the GDDT model is its ability to handle large and diverse datasets, incorporating both financial and non-financial information, such as transaction data, supply chain relationships, and external economic factors. This flexibility allows the model to better account for the multifaceted nature of credit risk, offering a more nuanced understanding of the factors that contribute to an enterprise's creditworthiness.

Moreover, the interpretability of decision trees is a key strength of the GDDT algorithm. Unlike other machine learning techniques, which often operate as "black boxes," decision trees provide a clear, understandable representation of how decisions are made, which increases the transparency and trustworthiness of the model’s outputs. Stakeholders, including financial analysts, risk managers, and institutional decision-makers, can more easily interpret and act upon the insights provided by the model. This interpretability is particularly important in the context of credit risk assessment, where decision-makers must justify their lending choices to regulatory bodies, investors, and other stakeholders.

5.2 Integrating Nonlinear Programming for Credit Strategy Optimization

In parallel with the GDDT algorithm, this dissertation integrates nonlinear programming techniques to optimize credit strategies. Financial institutions must carefully balance competing objectives: minimizing risk, ensuring profitability, and complying with regulatory requirements. Nonlinear programming offers a powerful mathematical framework for optimizing these objectives, especially in situations where multiple constraints and goals need to be addressed simultaneously.

Nonlinear programming allows financial institutions to model complex relationships between various variables, such as the amount of credit to allocate to a given enterprise, the associated risk, and the expected return. By formulating these relationships as mathematical constraints and objectives, nonlinear programming can identify the optimal credit strategy that maximizes the expected return while minimizing risk exposure. This approach enables institutions to allocate credit resources more efficiently, directing funds to enterprises that present the best risk-return profiles.

The integration of GDDT with nonlinear programming offers a unique advantage in credit strategy formulation. The predictive insights derived from the GDDT algorithm can serve as inputs into the optimization model, guiding the allocation of credit resources in a way that aligns with the institution's risk tolerance and profitability goals. The nonlinear programming model allows for the incorporation of real-time data, such as changes in market conditions, the financial health of SMEs, and other economic factors, ensuring that credit strategies remain adaptable and responsive to shifting circumstances.

5.3 Empirical Testing and Validation of the Proposed Framework

To evaluate the effectiveness of the proposed framework, this research conducted extensive empirical testing using real-world credit datasets. The results of these tests demonstrate the significant improvements the GDDT algorithm offers over traditional methods of credit risk assessment. Specifically, the GDDT model outperforms conventional credit risk models in terms of prediction accuracy, risk categorization, and the ability to account for complex, nonlinear interactions between variables.

For instance, the research found that the GDDT model achieved an AUC (Area Under the Curve) value of 0.99 in predicting the default probability of SMEs, significantly outperforming the AdaBoost algorithm (AUC = 0.97) and the traditional decision tree algorithm (AUC = 0.82). These results highlight the superior performance of the GDDT algorithm in capturing the complexities of credit risk and demonstrate its potential as a reliable tool for financial institutions.

Moreover, the nonlinear programming model was tested in conjunction with the GDDT algorithm to optimize credit strategies for a sample of 302 SMEs. The optimization process showed that, by balancing risk and return, the model was able to enhance the overall credit allocation efficiency, ultimately leading to a more rational distribution of credit resources and reducing the likelihood of defaults. The empirical testing results underscore the practicality and scalability of the proposed framework, offering valuable insights for financial institutions seeking to optimize their credit strategies.

5.4 Broader Implications for Financial Decision-Making

This research has important implications for the future of credit risk management and financial decision-making. By demonstrating the power of combining machine learning algorithms with optimization techniques, the study highlights how advanced technologies can be leveraged to address some of the most pressing challenges in financial institutions today. The ability to accurately assess credit risk, particularly in dynamic and uncertain environments, is crucial for the stability and profitability of financial institutions. The GDDT algorithm, with its ability to model nonlinear relationships and handle large, complex datasets, represents a significant advancement in this regard.

Furthermore, the integration of nonlinear programming into credit strategy formulation provides financial institutions with a more flexible and efficient approach to resource allocation. As the financial landscape continues to evolve, with increasing reliance on data-driven decision-making and the rise of fintech innovations, the methodologies proposed in this dissertation offer a forward-looking framework for optimizing credit strategies. This approach not only helps financial institutions improve their operational efficiency but also contributes to their long-term sustainability and competitiveness in the market.

In addition, this research suggests that combining predictive analytics with optimization frameworks can be applied to other areas of financial decision-making, such as loan pricing, portfolio management, and risk assessment under uncertainty. By providing a more comprehensive understanding of how financial institutions can navigate complex risk environments, this work opens new avenues for research and development in the field of financial technologies.

5.5 Contributions to Academic and Practical Knowledge

The findings of this research make several important contributions to both academic theory and practical financial applications. Academically, the dissertation advances the field of credit risk management by introducing a novel methodological approach that combines machine learning, nonlinear programming, and credit strategy optimization. This approach bridges the gap between theoretical models and practical applications, offering a more accurate and efficient means of assessing credit risk and formulating credit strategies.

Practically, the research provides valuable insights into how financial institutions can leverage advanced technologies to improve their credit risk assessment processes and optimize credit strategies. By demonstrating the effectiveness of the GDDT algorithm and nonlinear programming, the study provides financial institutions with actionable tools to enhance their decision-making and improve their overall credit risk management frameworks.

Moreover, this research offers a foundational methodology for integrating macroeconomic factors, industry-specific risks, and behavioral insights into predictive models, which is an area of growing importance in financial decision-making. As financial institutions continue to adapt to changing market conditions, the ability to incorporate these broader factors into credit risk models will become increasingly important.

5.6 Future Research Directions

While this dissertation presents a comprehensive framework for credit risk assessment and strategy optimization, there are several areas for future research that could further enhance the model's capabilities and applicability. For instance, future research could focus on real-time credit assessment, where machine learning models are constantly updated with the latest financial and transactional data. This would enable financial institutions to make more agile and responsive credit decisions in a rapidly changing environment.

Another avenue for future research is the integration of behavioral data and sentiment analysis into credit risk models. The inclusion of customer behavior, market sentiment, and other non-traditional data sources could further improve the accuracy and robustness of credit risk assessments, especially in volatile economic conditions.

Finally, further exploration of adaptive credit strategy optimization models could be undertaken, considering the dynamic nature of the economy and the increasing influence of external shocks, such as pandemics, financial crises, or geopolitical events. Nonlinear programming models that incorporate scenario-based simulations and stress-testing techniques could help financial institutions better prepare for and mitigate risks arising from unforeseen circumstances.

In conclusion, this dissertation demonstrates the transformative potential of integrating the Gradient Descent Decision Tree algorithm with nonlinear programming in the realm of credit risk assessment and strategy optimization. By addressing both predictive and prescriptive challenges, the proposed approach offers a comprehensive solution that aligns with the evolving needs of financial institutions in a complex and dynamic financial landscape. The methodologies developed in this study not only provide an effective means of assessing credit risk but also enable financial institutions to formulate adaptive, optimized credit strategies. As such, the insights from this research hold the potential to significantly impact the financial industry, fostering innovation, enhancing resilience, and contributing to the sustainable growth of financial systems in an increasingly interconnected and unpredictable world.

REFERENCES

[1] Abdelmoula, A. K. (2015). Bank credit risk analysis with k-nearest-neighbor classifier: Case of Tunisian banks. Accounting and Management Information Systems, 14(1), 79-90.
[2] Abdymomunov, A., Blei, S., & Ergashev, B. (2015). Integrating stress scenarios into risk quantification models. Journal of Financial Services Research, 47(1), 57-79.
[3] Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. Risks, 6(2), 38-47.
[4] Alzamora, G. S., Aceituno-Rojo, M. R., & Condori-Alejo, H. I. (2022). An Assertive Machine Learning Model for Rural Micro Credit Assessment in Peru. Procedia Computer Science, 202, 301-306.
[5] Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the US market. Abacus, 43(3), 332-357.
[6] Arora, N., & Kaur, P. D. (2019). A bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. Applied Soft Computing, 35(2), 121-135.
[7] Ayayi, A. G. (2012). Credit risk assessment in the microfinance industry: An application to a selected group of Vietnamese microfinance institutions and an extension to East Asian and Pacific microfinance institutions 1. Economics of Transition, 20(1), 37-72.
[8] Boudreault, M., Gauthier, G., & Thomassin, T. (2015). Estimation of correlations in portfolio credit risk models based on noisy security prices. Journal of Economic Dynamics and Control, 23(7), 334-349.
[9] Bao, W., Lianju, N., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. Expert Systems with Applications, 12(8), 301-315.
[10] Baser, F., Koc, O., & Selcuk-Kestel, A. S. (2023). Credit risk evaluation using clustering based fuzzy classification method. Expert Systems with Applications, 22(9), 119-130.
[11] Belás, J., Smrcka, L., Gavurova, B., & Dvorsky, J. (2018). The impact of social and economic factors in the credit risk management of SME. Technological and Economic Development of Economy, 24(3), 1215-1230.
[12] Belkin, B., Suchower, S., & Forest, L. (1998). A one-parameter representation of credit risk and transition matrices. CreditMetrics monitor, 1(3), 46-56.
[13] Brei, M., Jacolin, L., & Noah, A. (2020). Credit risk and bank competition in Sub-Saharan Africa. Emerging Markets Review, 44(11), 100-112.
[14] Brownlees, C., Hans, C., & Nualart, E. (2021). Bank credit risk networks: Evidence from the Eurozone. Journal of Monetary Economics, 37(5), 585-599.
[15] Carling, K., Jacobson, T., Lindé, J., & Roszbach, K. (2007). Corporate credit risk modeling and the macroeconomy. Journal of banking & finance, 31(3), 845-868.
[16] Carey, M., & Hrycay, M. (2001). Parameterizing credit risk models with rating data. Journal of banking & finance, 25(1), 197-270.
[17] Caruso, G., Gattone, S. A., Fortuna, F., & Di Battista, T. (2021). Cluster Analysis for mixed data: An application to credit risk evaluation. Socio-Economic Planning Sciences, 36(12), 104-116.
[18] Chamizo, A., & Novales, A. (2020). Looking through systemic credit risk: Determinants, stress testing and market value. Journal of International Financial Markets, Institutions and Money, 46(2), 27-39.
[19] Chandrappa, S., & Gitesh, K. K. (2009). Credit risk quantification: application of neural networks and fuzzy logic. Procedia Economics and Finance, 26(7), 311-316.
[20] Chang, Y. C., Chang, K. H., Chu, H. H., & Tong, L. I. (2016). Establishing decision tree-based short-term default credit risk assessment models. Communications in Statistics-Theory and Methods, 45(23), 6803-6815.
[21] Cheng, M., & Qu, Y. (2020). Does bank FinTech reduce credit risk? Evidence from China. Pacific-Basin Finance Journal, 63(11), 131-143.
[22] Chi, B. W., & Hsu, C. C. (2012). A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. Expert systems with applications, 39(3), 2650-2661.
[23] Corazza, M., De March, D., & di Tollo, G. (2021). Design of adaptive Elman networks for credit risk assessment. Quantitative Finance, 21(2), 323-340.
[24] Çomakoğlu, A. (2009). Credit risk modelling and quantification. Istanbul Technical University, Turkey.

[25] D'amico, G., Manca, R., & Salvi, G. (2014). Bivariate semi-Markov process for counterparty credit risk. Communications in Statistics-Theory and Methods, 43(7), 1503-1522.
[26] Dai, J., Hu, N., Huang, R., & Yan, Y. (2023). How does credit risk affect cost management strategies? Evidence on the initiation of credit default swap and sticky cost behavior. Journal of Corporate Finance, 41(12), 102-115.
[27] Danenas, P., & Garsva, G. (2012). Credit risk evaluation modeling using evolutionary linear SVM classifiers and sliding window approach. Procedia Computer Science, 43(9), 1324-1333.
[28] Djebali, N., & Zaghdoudi, K. (2020). Threshold effects of liquidity risk and credit risk on bank stability in the MENA region. Journal of Policy Modeling, 42(5), 1049-1063.
[29] Dodd, O., Kalimipalli, M., & Chan, W. (2021). Evaluating corporate credit risks in emerging markets. International Review of Financial Analysis, 39(6), 32-46.
[30] Doumpos, M., & Zopounidis, C. (2007). Model combination for credit risk assessment: A stacked generalization approach. Annals of Operations Research, 19(1), 289-306.
[31] Ericsson, J., & Renault, O. (2006). Liquidity and credit risk. The Journal of Finance, 49(5), 2219-2250.
[32] Fan, B., & Qin, J. (2023). Design and Implementation of an Enterprise Credit Risk Assessment Model Based on Improved Fuzzy Neural Network. Applied Artificial Intelligence, 37(1), 218-226.
[33] Fang, F., Shen, X., & Qiu, C. (2022). A New and Efficient Fourier Method for Risk Quantification and Allocation of Credit Portfolios. Available at SSRN.
[34] Fernandes, G. B., & Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. European Journal of Operational Research, 49(2), 517-524.
[35] Figini, S., & Giudici, P. (2017). Credit risk assessment with Bayesian model averaging. Communications in Statistics-Theory and Methods, 46(19), 9507-9517.
[36] Fanelli, V., & Maddalena, L. (2020). A nonlinear dynamic model for credit risk contagion. Mathematics and Computers in Simulation, 17(4), 45-58.
[37] García-Céspedes, R., & Moreno, M. (2017). An approximate multi-period Vasicek credit risk model. Journal of Banking & Finance, 29(6), 105-113.
[38] García-Céspedes, R., & Moreno, M. (2022). The generalized Vasicek credit risk model: A Machine Learning approach. Finance Research Letters, 47(9), 126-139.
[39] García, V., Marqués, A. I., & Sánchez, J. S. (2012). On the use of data filtering techniques for credit risk prediction with instance-based models. Expert Systems with Applications, 39(18), 13267-13276.
[40] Gilchrist, S., Ortiz, A., & Zakrajsek, E. (2009). Credit risk and the macroeconomy: Evidence from an estimated dsge model. Unpublished manuscript, Boston University, 13.
[41] Giudici, P., Hadji-Misheva, B., & Spelta, A. (2020). Network based credit risk models. Quality Engineering, 32(2), 199-211.
[42] Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. European Journal of Operational Research, 49(2), 417-426.
[43] Gonçalves, T. S., Ferreira, F. A., Jalali, M. S., & Meidutė-Kavaliauskienė, I. (2016). An idiosyncratic decision support system for credit risk analysis of small and medium-sized enterprises. Technological and Economic Development of Economy, 22(4), 598-616.
[44] Hastie, Trevor; Rosset, Saharon; Zhu, Ji; Zou, Hui (2009). "Multi-class AdaBoost". Statistics and Its Interface. 2 (3): 349-360.
[45] Hamerle, A., & Rösch, D. (2006). Parameterizing credit risk models. Journal of Credit Risk, 2(4).
[46] Harris, T. (2013). Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions. Expert Systems with Applications, 40(11), 4404-4413.
[47] Hayashi, Y. (2016). Application of a rule extraction algorithm family based on the Re-RX algorithm to financial credit risk assessment from a Pareto optimal perspective. Operations Research Perspectives, 3, 32-42.
[48] Hu, Y., & Su, J. (2022). Research on credit risk evaluation of commercial banks based on artificial neural network model. Procedia Computer Science, 19(9), 1168-1176.
[49] Huang, K., & Wei, K. (2023). Research on Credit Decision of Small and Medium-Sized Enterprises by Risk Quantification Model. In Proceedings of the 4th International Conference on Economic Management and Model Engineering, ICEMME 2022, November 18-20, 2022, Nanjing, China.
[50] Huang, X., Liu, X., & Ren, Y. (2018). Enterprise credit risk evaluation based on neural network

algorithm. Cognitive Systems Research, 52, 317-324.
[51] Huang, Y., Li, Z., Qiu, H., Tao, S., Wang, X., & Zhang, L. (2023). BigTech credit risk assessment for SMEs. China Economic Review, 102016.
[52] Jaisingh, W., & Nanjundan, P. (2023). IoT Infrastructure to Energize Electromobility. In AI-Powered IoT in the Energy Industry: Digital Technology and Sustainable Energy Systems (pp. 75-97). Cham: Springer International Publishing.
[53] Jin, J., Yu, Z. and Mi, C. (2012), "Commercial bank credit risk management based on grey incidence analysis", Grey Systems: Theory and Application, 2(3) :385-394.
[54] Ju, Y., & Sohn, S. Y. (2017). Technology credit scoring based on a quantification method. Sustainability, 9(6), 1057-1069.
[55] Jesus, S., & Gabriel, J. (2006). Credit cycles, credit risk, and prudential regulation, International Journal of Central Banking, 12(3):65-98.
[56] Jarrow, R. A., & Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. The journal of finance, 50(1), 53-85.
[57] Kuhn, H. W. (1991). Nonlinear programming: a historical note. *History of Mathematical Programming (JK Lenstra et al., eds.) North Holland, Amsterdam*, *14*, 82-96.
[58] Kanno, M. (2020). Credit risk assessment in real estate investment trusts: A perspective on blockholding and lending networks. International Review of Financial Analysis, 35(7):556-568.
[59] Kanno, M. (2020). Credit rating migration risk and interconnectedness in a corporate lending network. Research in International Business and Finance, 45(12), 128-130.
[60] Katchova, A. L., & Barry, P. J. (2005). Credit risk models and agricultural lending. American Journal of Agricultural Economics,27(1), 194-205.
[61] Kealhofer, S. (2003). Quantifying credit risk I: default prediction. Financial Analysts Journal, 59(1), 30-44.
[62] Kealhofer, S. (2003). Quantifying credit risk II: Debt valuation. Financial Analysts Journal, 59(3), 78-92.
[63] Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767-2787.
[64] Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. Expert Systems with Applications, 37(9), 6233-6239.
[65] Khashman, A. (2011). Credit risk evaluation using neural networks: Emotional versus conventional models. Applied Soft Computing, 11(8), 5477-5484.
[66] Khalil, J., Martel, J. M., & Jutras, P. (1999). A multicriterion system for credit risk rating (No. 1999-14).
[67] Kiesel, R., Perraudin, W., & Taylor, A. (2001). The structure of credit risk: spread volatility and ratings transitions.
[68] Kubo, H., & Sakai, Y. (2011). On long-term credit risk assessment and rating: Towards a new set of models. Journal of Risk Research, 14(9), 1127-1141.
[69] Lang, W., & Santomero, A. M. (2003). Risk quantification of retail credit: current practices and future challenges. CEIS Tor Vergata Research Paper.
[70] Leitao, Á., & Ortiz-Gracia, L. (2020). Model-free computation of risk contributions in credit portfolios. Applied Mathematics and Computation, 38(2), 125-139.
[71] Leong, C. K. (2016). Credit risk scoring with bayesian network models. Computational Economics, 47(3), 423-446.
[72] Li, S., Tsang, I. W., & Chaudhari, N. S. (2012). Relevance vector machine based infinite decision agent ensemble learning for credit risk analysis. Expert Systems with Applications, 39(5), 4947-4953.
[73] Liu, C., Xie, J., Zhao, Q., Xie, Q., & Liu, C. (2019). Novel evolutionary multi-objective soft subspace clustering algorithm for credit risk assessment. Expert Systems with Applications, 13(8), 112-131.
[74] Luo, S., Kong, X., & Nie, T. (2016). Spline based survival model for credit risk modeling. European Journal of Operational Research, 253(3), 869-879.
[75] Lu, S. L. (2013). Measuring credit risk by using a parameterized model under risk-neutral measure. Applied Economics Letters, 20(8), 719-723.
[76] Lu, S. L., Lee, K. J., & Zou, M. L. (2012). How to gauge credit risk: an investigation based on data envelopment analysis and the Markov chain model. Applied Financial Economics, 22(11), 887-897.

[77] Leippold, M., & Vanini, P. (2003). The quantification of operational risk. Available at SSRN 481742.
[78] Lopez, J. A., & Saidenberg, M. R. (2000). Evaluating credit risk models. Journal of Banking & Finance, 24(1-2), 151-165.
[79] Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. Expert Systems with Applications, 20(6), 168-179.
[80] Ma, H. Z., & Wang, X. R. (2016). Influencing factor analysis of credit risk in P2P lending based on interpretative structural modeling. Journal of Discrete Mathematical Sciences and Cryptography, 19(3), 777-786.
[81] Mvula Chijoriga, M. (2011). Application of multiple discriminant analysis (MDA) as a credit scoring and risk assessment model. International Journal of Emerging Markets, 6(2), 132-147.
[82] Nedumparambil, E., & Bhandari, A. K. (2020). Credit risk–return puzzle: Evidence from India. Economic Modelling, 29(11), 195-206.
[83] Ovi, N., Bose, S., Gunasekarage, A., & Shams, S. (2020). Do the business cycle and revenue diversification matter for banks' capital buffer and credit risk: Evidence from ASEAN banks. Journal of Contemporary Accounting & Economics, 16(1), 1001-1024.
[84] Oreski, S., & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert systems with applications, 41(4), 2052-2064.
[85] Oskarsdottir, M., & Bravo, C. (2021). Multilayer network analysis for improved credit risk prediction. Omega, 32(5), 102-118.
[86] Palazzo, B., & Yamarthy, R. (2022). Credit risk and the transmission of interest rate shocks. Journal of Monetary Economics, 46(9), 120-136.
[87] Pei, X., Li, H., & Wu, A. (2023). A novel enterprise credit risk rating model with the perspective of multiple regulatory demands. Expert Systems with Applications, 22(9), 120-135.
[88] Siu, T. K., Ching, W. K., Fung, S. E., & Ng, M. K. (2005). On a multivariate Markov chain model for credit risk measurement. Quantitative Finance, 5(6), 543-556.
[89] Segoviano Basurto, M., & Padilla, P. (2007). Portfolio credit risk and macroeconomic shocks: Applications to stress testing under data-restricted environments.
[90] Spuchľakova, E., & Cug, J. (2015). Credit risk and LGD modelling. Procedia Economics and Finance, 23, 439-444.
[91] Stefania, G. C., Claudia, G. D., Guillermo, A. M. L., Gustavo, G., Dario, R. H. J., Alfonso, M. T. F., & Tatiana, A. N. S. (2023). Credit Risk Scoring Model Based on The Discriminant Analysis Technique. Procedia Computer Science, 22(10), 928-933.
[92] Tang, J., Li, J., Xu, W., Tian, Y., Ju, X., & Zhang, J. (2021). Robust cost-sensitive kernel method with blinex loss and its applications in credit risk evaluation. Neural Networks, 43(2), 327-344.
[93] Wang, G., & Ma, J. (2011). Study of corporate credit risk prediction based on integrating boosting and random subspace. Expert Systems with Applications, 38(11), 13871-13878.
[94] Wang, G., & Ma, J. (2012). A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine. Expert Systems with Applications, 39(5), 5325-5331.
[95] Wang, K., Li, M., Cheng, J., Zhou, X., & Li, G. (2022). Research on personal credit risk evaluation based on XGBoost. Procedia computer science, 19(9), 1128-1135.
[96] Wang, Z., Zhang, X., Zhang, Z., & Sheng, D. (2022). Credit portfolio optimization: a multi-objective genetic algorithm approach. Borsa Istanbul Review, 22(1), 69-76.
[97] Wu, T. C., & Hsu, M. F. (2012). Credit risk assessment and decision making by a fusion approach. Knowledge-Based Systems, 35(7), 102-110.
[98] Wozabal, D., & Hochreiter, R. (2012). A coupled Markov chain approach to credit risk modeling. Journal of Economic Dynamics and Control, 36(3), 403-415.
[99] Yu, B. L., & Chen, Y. J. (2017). Grey-fuzzy comprehensive evaluation of enterprise groups credit risk. Journal of Discrete Mathematical Sciences and Cryptography, 20(6), 1525-1529.
[100] Yu, L., Yao, X., Wang, S., & Lai, K. K. (2011). Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. Expert Systems with Applications, 38(12), 15392-15399.
[101] Yu, L., Yue, W., Wang, S., & Lai, K. K. (2010). Support vector machine based multiagent ensemble learning for credit risk evaluation. Expert Systems with Applications, 37(2), 1351-1360.
[102] Yang, K., Yuan, H., & Lau, R. Y. (2022). PsyCredit: An interpretable deep learning-based credit assessment approach facilitated by psychometric natural language processing. Expert Systems with Applications, 19(8), 153-169.

[103] Yuan, G. X., Zhou, Y., Liu, H., & Yan, C. (2022). The framework of CAFE credit risk assessment for financial markets in China. Procedia Computer Science, 20(2), 33-46.
[104] Zamore, S., Beisland, L. A., & Mersland, R. (2019). Geographic diversification and credit risk in microfinance. Journal of Banking & Finance, 10(9), 105-119.
[105] Zhao, Y., Yang, M., Peng, X., & Li, G. (2022). Research on SMEs' Credit Risk Evaluation Index System from the Perspective of Supply Chain Finance—R-cluster and coefficient of variation based on panel data. Procedia Computer Science, 21(4), 613-620.
[106] Zhang, Z., Gao, G., & Shi, Y. (2014). Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. European Journal of Operational Research, 237(1), 335-348.
[107] Zhou, X., Jiang, W., & Shi, Y. (2010). Credit risk evaluation by using nearest subspace method. Procedia computer science, 19(1), 2449-2455.
[108] Zhou, X., Jiang, W., Shi, Y., & Tian, Y. (2011). Credit risk evaluation with kernel-based affine subspace nearest points learning method. Expert Systems with Applications, 38(4), 4272-4279.
[109] Zhou, L., Fujita, H., Ding, H., & Ma, R. (2021). Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting. Applied Soft Computing, 11(2), 107-118.
[110] Zhan, X., Li, X., & Su, W. (2021). Comprehensive Study of Small Micro-Enterprise Credit Risk Quantification and Credit Decisions. Academic Journal of Business & Management, 3(5), 28-34.
[111] Zha, Y., Power, D., & Tantisantiwong, N. (2020). The cross-country transmission of credit risk between sovereigns and firms in Asia. The Quarterly Review of Economics and Finance, 38(5), 309-320.
[112] Zhou, J., Wang, C., Ren, F., & Chen, G. (2021). Inferring multi-stage risk for online consumer credit services: an integrated scheme using data augmentation and model enhancement. Decision Support Systems, 14(9), 113-128.
[113] Zhu, Y., Zhou, L., Xie, C., Wang, G. J., & Nguyen, T. V. (2019). Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. International Journal of Production Economics, 211, 22-33. Zhu, Y., Chen, L., Hu, Y., et al. (2024). Credit risk management for small and medium-sized enterprise loans in commercial banks: Research practice from Changde. China Foreign Capital, 46(5), 111-119.
[114] Zhu, Q. (2022). Enterprise credit risk assessment based on an improved TOPSIS algorithm: A case study of small, medium, and micro enterprises. Journal of Changchun University, 32(7), 24-33.
[115] Zhu, Z., & Du, S. (2024). Digital finance, bank competition, and credit constraints of small and medium-sized enterprises. Journal of Shandong Institute of Business and Technology, 38(2), 95 – 105.
[116] Zhu, M. (2020). Research on credit risk and prevention management of small and micro enterprises in commercial banks. Modern Marketing (Information Edition),42 (7), 76 – 77.

APPENDIXES

APPENDIXES A Results Tables

Table A. 1: Quantified value of credit risk of 123 Companies

Company Code	Credit Risk	Company Code	Credit Risk	Company Code	Credit Risk
C1	0.3299	C42	0.9909	C83	0.9795
C2	0.7828	C43	0.9931	C84	0.9910
C3	0.8921	C44	0.9831	C85	0.9930
C4	0.8327	C45	0.9909	C86	0.9952
C5	0.9646	C46	0.9733	C87	0.9950
C6	0.9078	C47	0.9523	C88	0.9900
C7	0.8750	C48	0.9831	C89	0.9913
C8	0.8098	C49	0.9605	C90	0.9945
C9	0.9414	C50	0.9939	C91	0.9905
C10	0.9468	C51	0.9887	C92	0.9937
C11	0.9685	C52	0.9927	C93	0.9926
C12	0.9698	C53	0.9917	C94	0.9945
C13	0.8292	C54	0.9371	C95	0.9738
C14	0.9321	C55	0.9905	C96	0.9954
C15	0.9779	C56	0.9886	C97	0.9926
C16	0.9798	C57	0.9918	C98	0.9933
C17	0.9454	C58	0.9808	C99	0.9948
C18	0.9571	C59	0.9837	C100	0.9981
C19	0.9586	C60	0.9905	C101	0.9982
C20	0.9713	C61	0.9892	C102	0.9931
C21	0.9651	C62	0.9899	C103	0.9964
C22	0.9814	C63	0.9815	C104	0.9956
C23	0.9521	C64	0.9863	C105	0.9930
C24	0.9720	C65	0.9919	C106	0.9915
C25	0.9775	C66	0.9888	C107	0.9990
C26	0.9875	C67	0.9882	C108	0.9961
C27	0.9728	C68	0.9953	C109	0.9986
C28	0.9913	C69	0.9949	C110	0.9952
C29	0.9947	C70	0.9861	C111	0.9950
C30	0.9863	C71	0.9794	C112	0.9979
C31	0.9834	C72	0.9947	C113	0.9975
C32	0.9810	C73	0.9918	C114	0.9979
C33	0.9819	C74	0.9931	C115	0.9987
C34	0.9740	C75	0.6822	C116	0.9980
C35	0.9894	C76	0.9911	C117	0.9987
C36	0.9753	C77	0.9954	C118	0.9938
C37	0.9814	C78	0.9949	C119	0.9976
C38	0.9852	C79	0.9931	C120	0.9982
C39	0.9942	C80	0.9928	C121	0.9934
C40	0.9530	C81	0.9887	C122	0.9961
C41	0.9745	C82	0.9956	C123	0.9984

Table A. 2: Quantified value of credit risk of 302 Companies

Company Code	Credit Risk	Company Code	Credit Risk	Company Code	Credit Risk	Company Code	Credit Risk	Company Code	Credit Risk
D1	0.4330	D62	0.4973	D123	0.3902	D184	0.5009	D245	0.3571
D2	0.4605	D63	0.4112	D124	0.4651	D185	0.4227	D246	0.4890
D3	0.4330	D64	0.3082	D125	0.3918	D186	0.4508	D247	0.4222
D4	0.4605	D65	0.2989	D126	0.4077	D187	0.4386	D248	0.5353
D5	0.4330	D66	0.3906	D127	0.4795	D188	0.4274	D249	0.4609
D6	0.4605	D67	0.2587	D128	0.3360	D189	0.4748	D250	0.5427
D7	0.4330	D68	0.4196	D129	0.2654	D190	0.5141	D251	0.4473
D8	0.4605	D69	0.3282	D130	0.3895	D191	0.4597	D252	0.4652
D9	0.4330	D70	0.5216	D131	0.4574	D192	0.4773	D253	0.5247
D10	0.4605	D71	0.4442	D132	0.2729	D193	0.4859	D254	0.5054
D11	0.4330	D72	0.5103	D133	0.4557	D194	0.4555	D255	0.5323
D12	0.4605	D73	0.2929	D134	0.4292	D195	0.4674	D256	0.5034
D13	0.4330	D74	0.2520	D135	0.4112	D196	0.4410	D257	0.4404
D14	0.4605	D75	0.2773	D136	0.4402	D197	0.4805	D258	0.5190
D15	0.4330	D76	0.3819	D137	0.5311	D198	0.2964	D259	0.6688
D16	0.4605	D77	0.3010	D138	0.4499	D199	0.4646	D260	0.4772
D17	0.4330	D78	0.4132	D139	0.4461	D200	0.4909	D261	0.5748
D18	0.4605	D79	0.4682	D140	0.4103	D201	0.3167	D262	0.5145
D19	0.4330	D80	0.4315	D141	0.2729	D202	0.4286	D263	0.4895
D20	0.4605	D81	0.2729	D142	0.4697	D203	0.4884	D264	0.6629
D21	0.4330	D82	0.2904	D143	0.3316	D204	0.4467	D265	0.4915
D22	0.4605	D83	0.3879	D144	0.3822	D205	0.3422	D266	0.5169
D23	0.4330	D84	0.4653	D145	0.3697	D206	0.4206	D267	0.5018
D24	0.4605	D85	0.2784	D146	0.4290	D207	0.6405	D268	0.4353
D25	0.4330	D86	0.3746	D147	0.2840	D208	0.4666	D269	0.5006
D26	0.4605	D87	0.4503	D148	0.2959	D209	0.4474	D270	0.6558
D27	0.4330	D88	0.4144	D149	0.3352	D210	0.4541	D271	0.5448
D28	0.4605	D89	0.3778	D150	0.6489	D211	0.4921	D272	0.5073
D29	0.4330	D90	0.4640	D151	0.4280	D212	0.5114	D273	0.5104
D30	0.4605	D91	0.2654	D152	0.4741	D213	0.5160	D274	0.4880

Company Code	Credit Risk	Company Code	Credit Risk	Company Code	Credit Risk	Company Code	Credit Risk	Company Code	Credit Risk
D31	0.4330	D92	0.4141	D153	0.5212	D214	0.4320	D275	0.4887
D32	0.4605	D93	0.4119	D154	0.3388	D215	0.4960	D276	0.5698
D33	0.4330	D94	0.2803	D155	0.3253	D216	0.5243	D277	0.6725
D34	0.4605	D95	0.4383	D156	0.4056	D217	0.4223	D278	0.5514
D35	0.4330	D96	0.2920	D157	0.2997	D218	0.5548	D279	0.5436
D36	0.4605	D97	0.3984	D158	0.3976	D219	0.5310	D280	0.5338
D37	0.4330	D98	0.3794	D159	0.4904	D220	0.4133	D281	0.5544
D38	0.4605	D99	0.2870	D160	0.4677	D221	0.5730	D282	0.5669
D39	0.4330	D100	0.2875	D161	0.4908	D222	0.5156	D283	0.5161
D40	0.4605	D101	0.2859	D162	0.2729	D223	0.5041	D284	0.4743
D41	0.4330	D102	0.4393	D163	0.3141	D224	0.4945	D285	0.5604
D42	0.4605	D103	0.4378	D164	0.5247	D225	0.6489	D286	0.6983
D43	0.4330	D104	0.4825	D165	0.5059	D226	0.4898	D287	0.5571
D44	0.4605	D105	0.2870	D166	0.4755	D227	0.5360	D288	0.5066
D45	0.4330	D106	0.4204	D167	0.5244	D228	0.5330	D289	0.6925
D46	0.4605	D107	0.3809	D168	0.4991	D229	0.5269	D290	0.7263
D47	0.4330	D108	0.3870	D169	0.6529	D230	0.5071	D291	0.5224
D48	0.4605	D109	0.3642	D170	0.4872	D231	0.4673	D292	0.6804
D49	0.4330	D110	0.4480	D171	0.3194	D232	0.4257	D293	0.6876
D50	0.4605	D111	0.4247	D172	0.4968	D233	0.6481	D294	0.5727
D51	0.4330	D112	0.2768	D173	0.2918	D234	0.6546	D295	0.5572
D52	0.4605	D113	0.4322	D174	0.4319	D235	0.4248	D296	0.5243
D53	0.4330	D114	0.4693	D175	0.4483	D236	0.4353	D297	0.6930
D54	0.4605	D115	0.4227	D176	0.4564	D237	0.4825	D298	0.7101
D55	0.4330	D116	0.4796	D177	0.4773	D238	0.4751	D299	0.7146
D56	0.4605	D117	0.3110	D178	0.4336	D239	0.4588	D300	0.5616
D57	0.4330	D118	0.5129	D179	0.5093	D240	0.5242	D301	0.5724
D58	0.4605	D119	0.5128	D180	0.4401	D241	0.4622	D302	0.7061
D59	0.4330	D120	0.4350	D181	0.4579	D242	0.4803		
D60	0.4605	D121	0.4539	D182	0.2532	D243	0.5312		
D61	0.4330	D122	0.4089	D183	0.4923	D244	0.4556		

Table A. 3: Comparison of the operating conditions of the selected enterprises in January and February

Enterprise code	January Sales	January Profit	February Sales	February Profit	Enterprise Category	Enterprise Scale
	2018	2019	2018	2019		
	-80. 00%	-86.11%	38.81%	60.07%		
D18	-45.67%	-78.12%	-501.24%	10.51%	Wholesale	micro
D20	-70. 14%	-79.29%	135.57%	-78.72%	Software and information technology services	small
D39	-98. 53%	-99. 54%	101.52%	-99.48%	Software and Information Technology Services	small
D40	-99. 98%	-99. 98%	99.87%	99.06%	Transportation	small
D48	-89. 04%	-94.47%	-89.04%	-94.47%	Other unspecified industries	micro
D52	-83. 59%	-88.22%	-57.63%	-236.83%	Wholesale	small
D53	-98. 93%	-99. 51%	-1101. 53%	-1647.50%	Software and Information Technology Services	small
D57	-91.98%	-86.20%	-74.97%	118.99%	Industrial	small
D71	-99. 87%	-99.94%	95.78%	87.60%	Leasing and business services	small
D73	224.65%	-96.77%	83.26%	17.92%	Other unspecified industries	micro
D89	-97. 25%	-94. 87%	-96. 36%	-94.87%	Leasing and business services	micro
D93	24.65%	-54.13%	91.54%	-193.93%	Construction	micro
D102	-96.27%	-95.64%	-96.27%	-95. 64%	Property Management	micro
D109	-100.00%	-100.00%	-483.54%	55.37%	Service Industry	micro
D120	-97. 96%	-98.42%	101.10%	llO.49%	Software and Information Technology Services	micro
D124	-27.68%	736.76%	51.74%	70.62%	Industrial	micro
D127	-97. 52%	-92.06%	64.68%	-131.87%	Industrial	small
D135	-95.81%	-94.12%	-203.46%	-1856.51%	Transportation	small
D142	0.00%	-99.97%	0.00%	-99.97%	Industry	small
D144	-100.00%	-100.00%	99.39%	80.51%	Agriculture, forestry, animal husbandry and fishery	small
D146	-88.90%	4.74%	-543.43%	-4283.92%	Construction Software and Information	small
D152	-97.97%	-96.64%	-97.97%	102.04%	Technology Services	small
D153	234.03%	-99.95%	100.71%	-99.86%	Construction Leasing and Business	small
D155	-99.80%	-99.81%	-0. 58%	-138.06%	Leasing and business services	small
D156	-9.91%	-12.91%	197.86%	282.56%	Industrials	small
D159	-98. 34%	-99.38%	102.63%	100.94%	Software and information technology services	small
D161	-96.27%	-98.10%	100.44%	100.41%	Industrial	small
D166	-99.29%	-99.66%	100.71%	103.26%	Other industries not listed	micro
D167	-99.91%	-99.58%	-99.91%	-99.58%	Industrials	micro
D168	0.00%	-97.9%	0.00%	-86.94%	Construction	small
D173	-100.00%	-100.00%	-3884.00%	-268.50%	Construction	micro
D175	-99. 69%	-99.79%	-7261.86%	-2134.98%	Industrials	micro
D176	-97.75%	-98.73%	102.53%	100.89%	Industrial	micro
D178	-93.96%	-95.93%	-93.96%	100.41%	Rental and business services	micro
D179	25.14%	-83.73%	12744.29%	77.81%	Industrial	micro
D180	-98.26%	-97.95%	-98.26%	-97.95%	Other industries not listed	micro
D182	-52.99%	-94.09%	-52.99%	100.09%	Leasing and business services	micro
D184	-98. 79%	-92.48%	-2.57%	35.17%	Information transmission	micro
D187	-79.82%	-84.94%	102.24%	114.56%	Leasing and business services	micro
D188	-98.28%	-94.74%	-207.23%	-524.90%	Leasing and business services	micro
D190	-67. 97%	-82.10%	100.68%	105.31%	Transportation	micro
D191	-99.31%	-95.69%	101.78%	100.83%	Real Estate Development	micro
D192	-57.29%	-43.8%	50.08%	46.74%	Retail Trade	micro
D195	-89.50%	-71.06%	-89.50%	100.48%	Leasing and Business Services	micro
D197	-100.00%	-100.00%	-102.68%	-101.82%	Construction	micro
D202	-81.58%	0.00%	148.84%	136.71%	Information Transmission	micro
D203	-82.15%	-84.74%	101.94%	104.09%	Industrial	micro
D207	-93.03%	-93.82%	101.30%	-93.82%	Accommodation	micro
D210	-73.40%	-38.74%	100.44%	-38. 74%	Real Estate Development	micro
D211	0.00%	824.54%	136.29%	824.54%	Industrial	micro
D212	394.08%	16.98%	1337.98%	16.98%	Leasing and Business Services	micro
D219	-75.35%	0.00%	10352.33%	124.78%	Industrial	micro
D221	-100.00%	-100.00%	79.91%	65.76%	Industrial	micro
D225	-67.70%	-88.94%	-67.70%	109. 73%	Accommodation	micro

Enterprise code	January		February		Enterprise Category	Enterprise Scale
	Sales	Profit	Sales	Profit		
D227	-90.80%	-85.94%	-90.80%	-85.94%	Rental and business services	micro
D229	-98.44%	-97.77%	-98.44%	-97.77%	Construction	micro
D230	-88.34%	-51.89%	-88. 34%	107.62%	Leasing and Business Services	micro
D241	0.00%	36.33%	0.00%	36.33%	Software and Information Technology Services	micro
D242	0.00%	0.00%	-866.50%	-904.24%	Software and Information Technology Services	micro
D247	-89.19%	-45.52%	101.22%	124.89%	Software and Information Technology Services	micro
D249	-94.17%	-77.27%	185.71%	-77.27%	Industrial	micro
D254	2444.18%	45.67%	752.46%	-8.64%	Leasing and business services	micro
D255	-87.62%	-69.73%	100.83%	-69.73%	Leasing and business services	micro
D256	0.00%	0.00%	-208.03%	-692.61%	Retail trade	micro
D268	223.64%	-60.34%	191.87%	122.39%	Retail	micro
D269	0.00%	-87.12%	-10.83%	-40.14%	Software and information technology services	micro
D280	-98.32%	-94.43%	-98.32%	100.75%	Industrial	micro
D284	-41.80%	-89.10%	106.47%	-89.10%	Construction	micro
D287	-81.36%	-74.86%	-81.36%	-74.86%	Software and Information Technology Services	micro
D294	1603.33%	169.79%	1603.33%	169. 79%	Agriculture, forestry, animal husbandry and fishery	micro
D296	-37.20%	57.41%	-34.60%	67.07%	Software and Information Technology Services	micro
D297					Retail	micro

Table A. 4: Comparison of loan amounts before and after emergencies (Unit: Ten thousand yuan)

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D1	11.67	10.01	-1.66	Slight Decrease
D2	21.07	11.13	-9.94	Slight Decrease
D3	21.06	11.13	-9.93	Slight Decrease
D4	21.05	11.13	-9.92	Slight Decrease
D5	21.06	11.13	-9.93	Slight Decrease
D6	21.05	11.14	-9.91	Slight Decrease
D7	21.05	11.13	-9.92	Slight Decrease
D8	21.05	11.13	-9.92	Slight Decrease
D9	21.05	11.13	-9.92	Slight Decrease
D10	21.06	28.81	7.75	Slight Increase
D11	41.04	49.02	7.98	Slight Increase
D12	61.03	70.98	9.95	Slight Increase
D13	61.03	81.3	20.27	Moderate Increase
D14	61.04	70.99	9.95	Slight Increase
D15	21.06	11.13	-9.93	Slight Decrease
D16	21.05	11.13	-9.92	Slight Decrease
D17	61.02	70.99	9.97	Slight Increase
D18	41.04	31.05	-9.99	Slight Decrease
D19	41.04	31.04	-10	Slight Decrease
D20	61.03	90.99	29.96	Moderate Increase
D21	41.04	49.02	7.98	Slight Increase
D22	41.04	49.02	7.98	Slight Increase
D23	61.04	71	9.96	Slight Increase
D24	71.03	71	-0.03	Slight Decrease
D25	41.04	49	7.96	Slight Increase
D26	71.03	71	-0.03	Slight Decrease
D27	71.04	71	-0.04	Slight Decrease
D28	71.03	90.91	19.88	Moderate Increase

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D29	41.04	49.02	7.98	Slight Increase
D30	21.05	11.13	-9.92	Slight Decrease
D31	71.03	90.99	19.96	Moderate Increase
D32	21.07	11.12	-9.95	Slight Decrease
D33	21.86	12.22	-9.64	Slight Decrease
D34	41.84	32.13	-9.71	Slight Decrease
D35	71.04	71.01	-0.03	Slight Decrease
D36	21.06	11.11	-9.95	Slight Decrease
D37	71.03	71	-0.03	Slight Decrease
D38	71.02	70.98	-0.04	Slight Decrease
D39	71.97	102.1	30.13	Moderate Increase
D40	71.02	100.99	29.97	Moderate Increase
D41	21.05	11.13	-9.92	Slight Decrease
D42	41.04	49.02	7.98	Slight Increase
D43	41.8	32.15	-9.65	Slight Decrease
D44	41.04	31.04	-10	Slight Decrease
D45	71.03	70.99	-0.04	Slight Decrease
D46	71.03	71	-0.03	Slight Decrease
D47	41.04	49.02	7.98	Slight Increase
D48	41.04	31.04	-10	Slight Decrease
D49	71.03	90.99	19.96	Moderate Increase
D50	41.04	49.02	7.98	Slight Increase
D51	41.04	31.04	-10	Slight Decrease
D52	41.04	31.03	-10.01	Slight Decrease
D53	71.02	90.99	19.97	Moderate Increase
D54	71.02	71	-0.02	Slight Decrease
D55	41.04	49.01	7.97	Slight Increase
D56	71.03	71	-0.03	Slight Decrease
D57	41.05	31.04	-10.01	Slight Decrease

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D58	41.04	31.02	-10.02	Slight Decrease
D59	41.05	49.02	7.97	Slight Increase
D60	41.04	31.04	-10	Slight Decrease
D61	41.05	31.04	-10.01	Slight Decrease
D62	71.03	71	-0.03	Slight Decrease
D63	41.04	0	-41.04	Moderate Decrease
D64	0	0	0	—
D65	0	0	0	—
D66	0	0	0	—
D67	41.04	31.04	-10	Slight Decrease
D68	71.02	90.99	19.97	Moderate Increase
D69	41.04	66.81	25.77	Moderate Increase
D70	21.05	28.81	7.76	Slight Increase
D71	41.04	31.03	-10.01	Slight Decrease
D72	71.03	100.99	29.96	Moderate Increase
D73	41.04	49.01	7.97	Slight Increase
D74	71.03	71	-0.03	Slight Decrease
D75	41.99	32.07	-9.92	Slight Decrease
D76	41.84	50.12	8.28	Slight Increase
D77	21.06	23.06	2	Slight Increase
D78	41.05	31.04	-10.01	Slight Decrease
D79	21.05	23.06	2.01	Slight Increase
D80	41.04	31.04	-10	Slight Decrease
D81	71.02	71	-0.02	Slight Decrease
D82	21.05	23.06	2.01	Slight Increase
D83	71.04	71	-0.04	Slight Decrease
D84	21.06	11.13	-9.93	Slight Decrease
D85	21.06	11.13	-9.93	Slight Decrease
D86	41.05	31.05	-10	Slight Decrease

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D87	41.04	31.04	-10	Slight Decrease
D88	21.06	11.13	-9.93	Slight Decrease
D89	41.04	31.05	-9.99	Slight Decrease
D90	71.03	100.99	29.96	Moderate Increase
D91	41.04	49.01	7.97	Slight Increase
D92	41.03	31.05	-9.98	Slight Decrease
D93	41.04	31.04	-10	Slight Decrease
D94	22.77	11.66	-11.11	Slight Decrease
D95	41.03	30.99	-10.04	Slight Decrease
D96	41.14	31.79	-9.35	Slight Decrease
D97	71.03	51	-20.03	Moderate Decrease
D98	41.04	49.02	7.98	Slight Increase
D99	71.04	71	-0.04	Slight Decrease
D100	71.03	90.99	19.96	Moderate Increase
D101	71.02	100.99	29.97	Moderate Increase
D102	41.04	31.04	-10	Slight Decrease
D103	41.04	31.04	-10	Slight Decrease
D104	41.05	31.04	-10.01	Slight Decrease
D105	81.83	72.1	-9.73	Slight Decrease
D106	41.04	49.01	7.97	Slight Increase
D107	41.04	49.02	7.98	Slight Increase
D108	41.03	49.01	7.98	Slight Increase
D109	41.73	48.81	7.08	Slight Increase
D110	41.03	49.01	7.98	Slight Increase
D111	41.04	49	7.96	Slight Increase
D112	21.06	23.06	2	Slight Increase
D113	21.05	11.14	-9.91	Slight Decrease
D114	21.07	11.11	-9.96	Slight Decrease
D115	21.86	12.22	-9.64	Slight Decrease

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D116	21.07	11.09	-9.98	Slight Decrease
D117	28.77	28.81	0.04	Slight Increase
D118	21.05	11.11	-9.94	Slight Decrease
D119	28.77	28.81	0.04	Slight Increase
D120	81.02	71	-10.02	Slight Decrease
D121	21.05	23.06	2.01	Slight Increase
D122	41.84	32.07	-9.77	Slight Decrease
D123	41.04	49.02	7.98	Slight Increase
D124	41.04	31.03	-10.01	Slight Decrease
D125	71.04	71	-0.04	Slight Decrease
D126	41.04	49.02	7.98	Slight Increase
D127	41.04	31.04	-10	Slight Decrease
D128	41.08	48.81	7.73	Slight Increase
D129	41.04	31.04	-10	Slight Decrease
D130	41.05	31.04	-10.01	Slight Decrease
D131	41.04	31.04	-10	Slight Decrease
D132	21.05	11.14	-9.91	Slight Decrease
D133	41.06	31.68	-9.38	Slight Decrease
D134	41.04	31.02	-10.02	Slight Decrease
D135	71.02	71	-0.02	Slight Decrease
D136	71.02	100.98	29.96	Moderate Increase
D137	71.04	100.99	29.95	Moderate Increase
D138	41.04	48.99	7.95	Slight Increase
D139	21.06	11.11	-9.95	Slight Decrease
D140	41.04	31.04	-10	Slight Decrease
D141	0	0	0	—
D142	41.05	31.04	-10.01	Slight Decrease
D143	41.04	31	-10.04	Slight Decrease
D144	71.02	71	-0.02	Slight Decrease

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D145	41.04	31.04	-10	Slight Decrease
D146	41.04	48.92	7.88	Slight Increase
D147	21.06	11.1	-9.96	Slight Decrease
D148	82	72.09	-9.91	Slight Decrease
D149	21.14	23.03	1.89	Slight Increase
D150	21.05	11.13	-9.92	Slight Decrease
D151	41.04	21.13	-19.91	Moderate Decrease
D152	63.83	26.04	-37.79	Moderate Decrease
D153	40.76	92.69	51.93	Significant Increase
D154	33.04	61.81	28.77	Moderate Increase
D155	40.76	25.97	-14.79	Slight Decrease
D156	34.01	43.81	9.8	Slight Increase
D157	13.06	45.11	32.05	Moderate Increase
D158	40.76	6.13	-34.63	Moderate Decrease
D159	73.02	26.05	-46.97	Moderate Decrease
D160	33.04	95.99	62.95	Significant Increase
D161	33.05	43.97	10.92	Slight Increase
D162	13.05	25.99	12.94	Slight Increase
D163	33.05	6.13	-26.92	Moderate Decrease
D164	40.76	43.81	3.05	Slight Increase
D165	13.07	43.81	30.74	Moderate Increase
D166	33.04	23.81	-9.23	Slight Decrease
D167	33.04	26.01	-7.03	Slight Decrease
D168	40.76	25.97	-14.79	Slight Decrease
D169	40.76	43.81	3.05	Slight Increase
D170	40.76	26.04	-14.72	Slight Decrease
D171	33.03	61.81	28.78	Moderate Increase
D172	20.77	43.94	23.17	Moderate Increase
D173	33.09	23.81	-9.28	Slight Decrease

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D174	40.76	26.03	-14.73	Slight Decrease
D175	40.76	25.95	-14.81	Slight Decrease
D176	33.04	44.02	10.98	Slight Increase
D177	40.76	26.04	-14.72	Slight Decrease
D178	40.76	25.98	-14.78	Slight Decrease
D179	40.76	26.05	-14.71	Slight Decrease
D180	40.76	43.99	3.23	Slight Increase
D181	40.76	26.05	-14.71	Slight Decrease
D182	33.04	26.04	-7	Slight Decrease
D183	40.76	26.04	-14.72	Slight Decrease
D184	20.77	43.98	23.21	Moderate Increase
D185	40.76	23.81	-16.95	Moderate Decrease
D186	40.76	44.02	3.26	Slight Increase
D187	20.77	43.81	23.04	Moderate Increase
D188	33.05	6.13	-26.92	Moderate Decrease
D189	40.76	44.02	3.26	Slight Increase
D190	40.76	44.02	3.26	Slight Increase
D191	40.76	43.94	3.18	Slight Increase
D192	40.76	44.01	3.25	Slight Increase
D193	40.76	26.01	-14.75	Slight Decrease
D194	20.77	43.98	23.21	Moderate Increase
D195	40.76	6.13	-34.63	Moderate Decrease
D196	40.76	43.95	3.19	Slight Increase
D197	20.77	44.01	23.24	Moderate Increase
D198	40.76	23.81	-16.95	Moderate Decrease
D199	33.88	26.04	-7.84	Slight Decrease
D200	40.76	45.12	4.36	Slight Increase
D201	40.76	43.81	3.05	Slight Increase
D202	40.76	26	-14.76	Slight Decrease

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D203	40.76	44.01	3.25	Slight Increase
D204	33.03	61.81	28.78	Moderate Increase
D205	20.77	43.81	23.04	Moderate Increase
D206	40.76	23.81	-16.95	Moderate Decrease
D207	40.76	44.02	3.26	Slight Increase
D208	40.76	26.04	-14.72	Slight Decrease
D209	33.04	43.81	10.77	Slight Increase
D210	40.76	43.98	3.22	Slight Increase
D211	40.76	40	-0.76	Slight Decrease
D212	40.76	43.81	3.05	Slight Increase
D213	40.76	26.02	-14.74	Slight Decrease
D214	40.76	44.01	3.25	Slight Increase
D215	20.77	26.04	5.27	Slight Increase
D216	40.76	23.81	-16.95	Moderate Decrease
D217	33.1	61.81	28.71	Moderate Increase
D218	40.76	26.03	-14.73	Slight Decrease
D219	40.76	61.81	21.05	Moderate Increase
D220	40.76	43.81	3.05	Slight Increase
D221	0	0	0	—
D222	0	0	0	—
D223	20.77	43.81	23.04	Moderate Increase
D224	40.76	18.03	-22.73	Moderate Decrease
D225	40.76	43.81	3.05	Slight Increase
D226	40.76	26.04	-14.72	Slight Decrease
D227	0	0	0	—
D228	0	0	0	—
D229	0	0	0	—
D230	40.76	43.96	3.2	Slight Increase
D231	0	26	26	Moderate Increase

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D232	40.76	0	-40.76	Moderate Decrease
D233	40.76	44.01	3.25	Slight Increase
D234	0	0	0	—
D235	33.06	25.99	-7.07	Slight Decrease
D236	0	0	0	—
D237	40.76	61.81	21.05	Moderate Increase
D238	40.76	43.92	3.16	Slight Increase
D239	40.76	43.99	3.23	Slight Increase
D240	0	0	0	—
D241	40.76	43.81	3.05	Slight Increase
D242	40.76	24.98	-15.78	Moderate Decrease
D243	40.76	43.95	3.19	Slight Increase
D244	0	0	0	—
D245	33.06	43.81	10.75	Slight Increase
D246	0	0	0	—
D247	40.76	0	-40.76	Moderate Decrease
D248	0	0	0	—
D249	0	0	0	—
D250	0	0	0	—
D251	0	0	0	—
D252	40.76	0	-40.76	Moderate Decrease
D253	0	0	0	—
D254	0	0	0	—
D255	33.03	43.81	10.78	Slight Increase
D256	40.76	61.81	21.05	Moderate Increase
D257	33.05	61.81	28.76	Moderate Increase
D258	0	0	0	—
D259	0	0	0	—
D260	33.06	61.81	28.75	Moderate Increase

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D261	0	0	0	—
D262	40.76	35.81	-4.95	Slight Decrease
D263	13.1	43.81	30.71	Moderate Increase
D264	40.76	6.12	-34.64	Moderate Decrease
D265	0	0	0	—
D266	0	0	0	—
D267	40.76	61.81	21.05	Moderate Increase
D268	0	61.81	61.81	Significant Increase
D269	40.76	23.81	-16.95	Moderate Decrease
D270	0	0	0	—
D271	0	0	0	—
D272	0	0	0	—
D273	0	0	0	—
D274	0	0	0	—
D275	0	0	0	—
D276	0	0	0	—
D277	0	0	0	—
D278	0	0	0	—
D279	0	0	0	—
D280	0	0	0	—
D281	0	0	0	—
D282	0	0	0	—
D283	0	0	0	—
D284	33.06	43.93	10.87	Slight Increase
D285	0	0	0	—
D286	0	0	0	—
D287	0	0	0	—
D288	0	0	0	—
D289	0	0	0	—

Company code	Normal(Q1)	After adjustment(Q2)	Q2-Q1	Loan category
D290	0	0	0	—
D291	0	0	0	—
D292	0	0	0	—
D293	0	0	0	—
D294	0	0	0	—
D295	0	0	0	—
D296	0	0	0	—
D297	0	0	0	—
D298	0	0	0	—
D299	0	0	0	—
D300	0	0	0	—
D301	0	0	0	—
D302	0	0	0	—

APPENDIXES B Main Program

```
####################################
##### Main Program ###############


###Data cleaning and indicator calculation
clc
close all
clc,clear
data=xlsread("C:\Users\cheng\Desktop\group 1);
result1=zeros(123,11);
for i=1:123
   m=find(data(:,1)==i);
   re=data(m(1,1):max(m),:);
   re=sortrows(re,[3,4,5]);
   void=find(re(:,10)==0);
   void=size(void,1);
   valid=size(m,1)-void;
   mm=find((re(:,10)==1)&(re(:,9)>0));
   ree=re(mm,:);
   ree=sortrows(ree,[3,4,5]);
   for j=7:9
      A=ree(:,j);
      B = filloutliers(A,'nearest','mean');
      ree(:,j)=B;
   end
   smonth(1,:)=ree(1,3:4);
   smonth(2,:)=ree(size(ree,1),3:4);
length=(((smonth(2,1)-2017)*12+smonth(2,2))-((smonth(1,1)-
2017)*12+smonth(1,2)))+1;%
   up=data(m(1,1):max(m),6);
   up=numel(unique(up));
   result1(i,1)=i;% company code
   result1(i,2)=sum(ree(:,7),1);
   result1(i,3)=sum(ree(:,8),1);
   result1(i,4)=sum(ree(:,9),1);
   result1(i,5)=numel(m);
   result1(i,6)=valid;
   result1(i,7)=void;
   result1(i,8)=valid/numel(m);
   result1(i,9)=up;
   result1(i,10)=length;
   stdj=std(ree(:,7))./mean(ree(:,7));
   result1(i,11)=stdj;
end
xlswrite("C:\Users\cheng\Desktop\group1-1);
########################################################################
####
```

```
### Entropy weight method and TOPSIS model
clc
close all
clc,clear
data=xlsread("C:\Users\cheng\Desktop\group1-1);
data1=data(:,2:12);
result=zeros(123,3);
for i=1:11
   maxx=max(data1(:,i));
   minn=min(data1(:,i));
   if i==4||i==5
      data1(:,i)=(maxx-data1(:,i))/(maxx-minn);
   else
      data1(:,i)=(data1(:,i)-minn)/(maxx-minn);
   end
end
X=data1;
data1(find(data1==0))=[0.0001];
data1(find(data1==1))=[0.9999];
summ=sum(data1);
p=data1./summ;
e=p.*log(p);
k=-1/log(123);
Ej=sum(e)*k;
Dj=1-Ej;
w=Dj/sum(Dj);
for i=1:11
   X(:,i)= X(:,i)./sqrt(sum(X(:,i).^2));
end
R=w.*X;
r_max=max(R);
r_min=min(R);
d_z = sqrt(sum([(R -repmat(r_max,size(R,1),1)).^2 ],2)) ;
d_f = sqrt(sum([(R -repmat(r_min,size(R,1),1)).^2 ],2));
s=d_f./(d_z+d_f );
f=1-s;
result(:,1)=data(:,1);
result(:,2)=f;
result(:,3)=s;
##########################################################################
#####
```

APPENDIXES C Paper Publications and International Conferences

Paper One Publication

https://www.scopus.com/record/display.uri?eid=2-s2.0-85200132389&doi=10.1145%2f3671151.3671357&origin=inward&txGid=d7d5f9a989a3cd1b84fa057af92eb49a

Credit Risk Analysis of Micro and Small Enterprises Based on Big Data-Driven Logistic Regression Models of Bank Credit Records

Guoqing Chen
Department of Mathematics, Mahasarakham University, Maha Sarakham 44150, Thailand
chenguoqing1139@163.com

Piyapatr Busababodhin*
Department of Mathematics, Mahasarakham University, Maha Sarakham 44150, Thailand
piyapatr.b@msu.ac.th

Nipaporn Chutiman
Department of Mathematics, Mahasarakham University, Maha Sarakham 44150, Thailand
nipaporn.c@msu.ac.th

ABSTRACT

This paper presents an analytical examination of a study that utilises a logistic regression model driven by large volumes of big data aggregated from bank credit records to assess the credit risk associated with small and medium-sized enterprises (SMEs). The model utilises a comprehensive dataset to substantially improve the predictive accuracy associated with credit defaults by SMEs. The model provides a careful analysis of key metrics such as credit history, trading behaviour and financial ratios to gain insight into the complex financial situation of SMEs. The study aims to enhance risk assessment methodologies and facilitate banking organisations to tailor credit solutions to the specific needs of SMEs, while effectively managing potential credit defaults. The study highlights the indispensable role played by big data and cutting-edge analytics to contribute to the field of financial risk management, particularly in the area of SME lending.

CCS CONCEPTS

• **Applied computing** → Electronic commerce; Electronic data interchange; Law, social and behavioral sciences; Economics; Operations research; Forecasting; Operations researchDecision analysis; Multi-criterion optimization and decision-making; Physical sciences and engineering; Mathematics and statistics.

KEYWORDS

Credit Risk, Big Data, Logistic Regression Models, Credit Records, Micro and Small Enterprises

ACM Reference Format:

Guoqing Chen, Piyapatr Busababodhin*, and Nipaporn Chutiman. 2024. Credit Risk Analysis of Micro and Small Enterprises Based on Big Data-Driven Logistic Regression Models of Bank Credit Records. In *5th International Conference on Computer Information and Big Data Applications (CIBDA 2024), April 26–28, 2024, Wuhan, China*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3671151.3671357

ACM ISBN 979-8-4007-1810-6/24/04
https://doi.org/10.1145/3671151.3671357

1 INTRODUCTION

In recent years, the role of micro and small enterprises (MSEs) in driving economic growth, innovation and employment has been increasingly recognised. However, one of the persistent challenges faced by these enterprises is access to finance. Traditional credit scoring systems, with their reliance on historical financial statements and collateral, often fall short in assessing the creditworthiness of MSEs, which many MSEs lack. This financial exclusion not only hinders the growth of microenterprises, but also limits their contribution to the wider economy [1]. The emergence of big data technologies has opened up new avenues for credit risk analysis. Unlike traditional methods, big data-driven approaches can provide a more comprehensive assessment of credit risk using a wide range of data sources, including bank credit records, transactional data, social media activity and mobile phone usage patterns. In this context, logistic regression modelling has emerged as a powerful tool for analysing large amounts of unstructured data and transforming it into meaningful predictors of credit default [2]. This statistical approach is capable of handling binary classification tasks and is particularly suitable for predicting the likelihood of default on microenterprise loans.

The focus of this study is on the use of big data-driven logistic regression modelling to improve the accuracy and efficiency of credit risk assessment for MSMEs. By utilising a large dataset of bank credit records, this study aims to develop a logistic regression model that predicts the probability of default based on various characteristics that affect credit risk. The objectives of the study include collecting and pre-processing a large data set of bank credit records of MSMEs, selecting and designing key features affecting credit risk, developing and evaluating a logistic regression model, and exploring the application of the model results in credit decision-making and risk management practices of financial institutions [3].

The significance of this study is that it has the potential to provide financial institutions with more effective tools for assessing and managing MSME credit risk. By improving the predictive accuracy of credit defaults and providing a more nuanced understanding of risk factors, this study aims to promote sustainable growth and financial accessibility for microenterprises [4]. Furthermore, this study highlights the indispensable role of big data and advanced analytics in revolutionising the field of financial risk management, particularly in SME lending.

In summary, this study aims to harness the power of big data-driven logistic regression modelling to address the daunting challenge of credit risk assessment for SMEs. By providing a more comprehensive and accurate credit risk assessment, this study aims to enhance the lending decisions and risk management strategies of financial institutions, ultimately contributing to the growth and financial inclusion of microenterprises in the global economy.

2 RELATED WORK

In the field of credit risk analysis of micro enterprises (MSEs), to adapt to the challenge of lack of traditional financial data, research efforts have turned to the use of big data-based logistic regression models. These models utilize bank credit records to provide a comprehensive perspective that traditional methods cannot capture. Previous research has shown that MSEs often face financial exclusion due to the stringent requirements of traditional credit scoring systems, which rely heavily on historical financial statements and collateral that many MSEs are unable to provide. In response, researchers have developed innovative methods using big data analytics, mining a wide range of data sets including transaction data, social media analysis and mobile phone usage patterns to predict credit risk with greater accuracy. In particular, logistic regression models have attracted attention for their ability to process large amounts of unstructured data and transform it into meaningful predictors of credit defaults. This statistical method has been the subject of numerous studies aimed at refining its algorithms and enhancing its predictive power specifically for credit risk assessment. Researchers focus on variable selection, model tuning, and integration of non-traditional data sources to improve model effectiveness.

For data collection and pre-processing, the first step is to collect a large dataset of bank credit records of Micro and Small Enterprises (MSEs). This dataset should include various characteristics such as loan amount, repayment history, interest rate, collateral and other relevant financial indicators. Data preprocessing is essential for cleaning and normalising the data, dealing with missing values and preparing the data for analysis. Feature selection and engineering to identify the most relevant features that affect microenterprise credit risk. Feature engineering may be required to create new features that better capture risk factors. Logistic regression modelling, Logistic regression is a commonly used statistical model for binary classification tasks such as predicting whether a loan will default (high risk) or not (low risk). In this case, the logistic regression model will be trained to predict the probability of default based on the selected features. Big data analytics, given the potentially large number of credit records, big data techniques may be employed to process and analyse the data efficiently. This may involve the use of distributed computing frameworks such as Apache Hadoop or Apache Spark to process and analyse data at scale. Model evaluation and validation requires the use of appropriate metrics to evaluate the performance of the logistic regression model such as accuracy, precision, recall, and area under the receiver operating characteristic curve (ROC). Cross-validation techniques can be used to validate model performance on unseen data. Risk assessment and decision making, logistic regression models can provide a quantitative assessment of the credit risk of each MSE. Financial institutions can use these risk assessments to make informed lending decisions, set interest rates, and manage their overall credit risk. Ongoing monitoring and model updating, credit risk models should be regularly updated with new data and reassessed to ensure their accuracy and relevance. This may involve regular retraining of the model or adjusting the model to changes in the economic environment or lending policies. In conclusion, this study aims to use big data and logistic regression models to improve the accuracy and efficiency of credit risk assessment for MSMEs, and ultimately help financial institutions make better lending decisions and risk management.

Table 1: Data relating to 100 credit-recorded enterprises.

Enterprise code	Credit rating	Default or not
GS-1	I	No
GS-2	II	No
GS-3	III	No
GS-4	III	No
...		
GS-97	II	No
GS-98	II	No
GS-99	IV	Yes
GS-100	IV	Yes

3 DATA ANALYSIS

The dataset provided provides a comprehensive overview of the creditworthiness of 100 companies, ranging from GS-1 to GS-100, and analysis of the data shows a clear correlation between a company's credit rating and its likelihood of default. Companies with credit ratings of I and II generally have low default rates, suggesting a favourable credit profile. On the contrary, companies with credit rating IV have a higher propensity to default, indicating that the lower the credit rating, the higher the risk [5]. This trend suggests that credit ratings are a reliable indicator of default risk, with lower ratings indicating a greater likelihood that a company will fail to honour its contractual obligations. The findings highlight the importance of credit assessment as a tool for predicting financial reliability and risk management in the corporate sector.

As shown in Table 1, the data provided encapsulates the credit history of 100 MSMEs, including the unique identifiers, credit ratings, and default status of these businesses. In order to quantify the credit risk of these businesses with credit history (a common challenge in big data analysis for credit decision making), the analysis first quantifies the risk into the probability of customer performance [6]. A logistic regression model is then constructed to quantify the credit risk associated with these 100 businesses. The model allows the probability of compliance (p_i) to be calculated for each business. Subsequently, a mapping function is applied to normalise the p_i value so that it is in the range of 0 to 1. This normalisation clearly explains the results: p_i values close to 1 indicate higher creditworthiness and lower risk, while values close to 0 imply poorer creditworthiness and higher risk. These findings illustrate the efficacy of logistic regression models in assessing and predicting credit risk, providing a valuable measure for financial decision makers [7].

4 MODELLING

In the quantitative analysis of credit risk, we reconstruct credit risk as a function of the firm's performance rate. Thus, a bank's credit risk loss rate is represented by the complement of the firm's

Table 2: Construction and description of indicators.

Indicators	Description
Proportion of input invoices cancelled by enterprise x_{1i}	The lower the ratio, the higher the credit and the lower the credit risk.
Percentage of cancelled sales invoices by enterprise x_{2i}	The lower the ratio, the higher the creditworthiness and the lower the credit risk
Creditworthiness of each enterprise x_{3i}	The better the creditworthiness, the higher the creditworthiness and the lower the credit risk.
Total net invoice amount x_{4i}	The larger the amount, the more capital the enterprise has and the stronger the enterprise is
Coefficient of variation of invoice amount x_{5i}	The higher the coefficient of variation, the higher the deviation of the amount and the higher the credit risk
Fund turnover rate x_{6i}	Fast capital turnover, the faster the turnover rate, the better the utilisation of funds, the higher the credit risk of the enterprise

compliance rate, denoted by $1 - p_i$. This formulation assumes that the higher the compliance rate, the better the creditworthiness of the firm, the lower the credit risk, and thus the lower the credit loss rate [8]. Conversely, the lower the compliance rate, the worse the credit, the higher the credit risk and the higher the credit loss rate.

The quantitative analysis of the credit risk of these 100 enterprises was transformed into a credit risk assessment model using Logistic regression. The model argues that the higher the credit of the enterprise, the lower the default rate and therefore the lower the credit risk. The construction of the enterprise credit risk assessment model mainly includes three stages: the construction and selection of indicators, the collection and processing of sample data, and the determination of default [9].

In terms of indicator construction and selection, the selected indicators must adhere to the principles of scientificity, comprehensiveness, fairness, relevance, legality and practicality. These principles are the basis for constructing an appropriate evaluation indicator system [10]. The evaluation system consists of five indicators, as detailed in Table 2 below. Each indicator has been carefully selected and described to ensure that the evaluation system accurately reflects the multifaceted nature of credit risk and provides a robust framework for corporate credit assessment.

Various indicators used to assess the creditworthiness of the company are listed according to table 2. The cancellation ratios of input invoices (x_{1i}) and output invoices (x_{2i}) are inversely related to creditworthiness; lower ratios indicate higher creditworthiness and lower credit risk. The company's creditworthiness (x_{3i}) follows a similar trend; the better the creditworthiness, the higher the credit and the lower the risk. The net invoice amount (x_{4i}) is directly proportional to the financial strength of the company; the larger the net invoice amount, the stronger the company's financial capability. The coefficient of variation of invoice amount (x_{5i}) is a measure of deviation; the smaller the coefficient of variation, the more stable the earnings and the lower the credit risk; the larger the coefficient of variation, the more unstable the earnings and the higher the risk [11]. Finally, the capital turnover ratio (x_{6i}) reflects the efficiency of capital use; the faster the capital turnover ratio, the more efficient the capital utilisation, the higher the credit rating and the lower the risk. By analysing these indicators, a comprehensive understanding of the company's credit risk can be obtained [12].

A Logistic regression model is built from the relevant data, noting that the firm's compliance rate is p_i and the credit risk loss rate is $1 - p_i$, where denotes the first type of firm ($i = 1, 2..100$). The correlation coefficient of the indicators in the regression equation is fitted through the indicators, and the model is.

$$\ln \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} \quad (1)$$

In the realm of statistical modeling, Logistic Regression emerges as an extension of the linear regression framework, positioned within the broader family of generalized linear models. The underlying philosophy of Logistic Regression is rooted in the principles of linear regression, adapted to accommodate binary or categorical outcome variables [13]. The quintessential equation governing Logistic Regression is formalized as:

$$h_\beta(x) = \frac{1}{1 + e^{-\ln \frac{p_i}{1-p_i}}} \quad (2)$$

The firm's compliance rate after converting Equation 1 is p_i.

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i})}} \quad (3)$$

In this expression, (p) represents the probability of the occurrence of the event of interest. The left-hand side, known as the logit function, effectively transforms the probability (p) into an unbounded continuous variable. The coefficients ($\beta_0, \beta_1, \ldots, \beta_{100}$) embody the estimated impacts of the predictor variables ($x_1, x_2, \ldots, x_{100}$) on the log odds of the probability in question [14]. This transformation from linear regression to Logistic Regression is pivotal for analyzing binary outcomes, as it ensures the model-generated probabilities are constrained between 0 and 1.

$$P(y = 1|x;\beta) = h_\beta(x) = p_i \quad (4)$$

$$P(y = 0|x;\beta) = 1 - h_\beta(x) = 1 - p_i \quad (5)$$

During the training phase of a logistic regression model, the objective is to maximize the likelihood of the observed outcomes given the predicted probabilities. The training dataset consists of 100 samples, each associated with a predicted probability $h_\beta(x)$for the positive class, and consequently, $1 - h_\beta(x)$ for the negative

Table 3: Values of regression coefficients β.

Regression coefficient	Numerical value
β_0	0.4328
β_1	-10.982
β_2	0.3824
β_3	-0.9132
...	
β_{97}	-0.8892
β_{98}	-0.8963
β_{99}	0.1764
β_{100}	0.1532

class. To align the model's predictions with the actual class labels of the samples, the goal is to maximize the likelihood function that represents the probability of the training dataset being classified correctly.

The likelihood function for the dataset under the logistic regression model is given as:

$$L(\beta) = \prod_{i=1}^{100} P\left(y^{(i)} | x^{(i)}; \beta\right) \tag{6}$$

$$L(\beta) = \prod_{i=1}^{100} h_\beta\left(x^{(i)}\right)^{y^{(i)}} \left(1 - h_\beta\left(x^{(i)}\right)\right)^{1-y^{(i)}} \tag{7}$$

The number of samples is 100, x_i represents the feature vector of the i^{th} sample, y_i is the binary class label for the i^{th} sample, and $h_\beta(x_i)$ is the predicted probability for the i^{th} sample being the positive class.

Maximizing the likelihood function is equivalent to maximizing its logarithm, which is a more feasible computational task due to the numerical stability and simplicity of summing log probabilities over taking the product of probabilities. This log-likelihood function is expressed as:

$$l(\beta) = \log(L(\beta)) = \sum_{i=1}^{m} \log\left(\left(h_\beta\left(x^{(i)}\right)\right)\right)^{y^{(i)}} + \log\left(\left(1 - h_\beta\left(x^{(i)}\right)\right)\right)^{1-y^{(i)}} \tag{8}$$

$$l(\beta) = \log(L(\beta)) = \sum_{i=1}^{m} y^{(i)} \log\left(h_\beta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\beta\left(x^{(i)}\right)\right) \tag{9}$$

The maximization of this log-likelihood function is typically achieved through optimization algorithms such as gradient descent [15]. By doing so, the logistic regression model parameters are tuned to provide the best fit for the observed data, thereby yielding the most accurate predictions possible given the model structure.

$$J(\beta) = -\frac{1}{100} l(\beta) \tag{10}$$

Bringing the calculated coefficients into equation yields the coefficients as shown in Table 3 below.

As can be seen in Table 3, the regression coefficients have both positive and negative values, reflecting the different effects of each predictor on the response variable. The positive intercept β_0 indicates that the baseline level of the dependent variable is above zero when all predictors are absent. The compliance rate of the 100 firms is found by the formula and then mapped to between 0 and 1 by the mapping. We can conclude that the closer the value of P_i is to 1, the better the creditworthiness of the enterprise applying for the loan, and the closer the value of P is to 0, the worse the creditworthiness of the enterprise applying for the loan. The better the creditworthiness of the lending firm, the lower the credit risk, and the worse the creditworthiness of the lending firm, the higher the credit risk. 100 firms' compliance rate P_i is shown in Table 4 below.

Table 4: Compliance rate of 100 enterprises P_i.

100 Enterprises	Compliance rate
GS1	0.8956
GS2	0.8931
GS3	0.7131
GS4	0.8012
...	
GS97	0.7621
GS98	0.8132
GS99	0.4123
GS100	0.3125

An assessment of the compliance rates (P_i) of the 100 companies listed in Table 4 reveals that these rates vary significantly. The majority of companies maintain compliance rates above 0.7, indicating a general tendency towards high levels of compliance with protocols or standards. However, companies such as GS99 and GS100 in particular have significantly lower compliance rates, suggesting that a subset of companies have significantly lower compliance rates compared to their peers. This variation can be attributed to differences in company policies, management efficiency or industry-specific challenges.

5 CONCLUSIONS

In this study, we utilise a logistic regression model driven by a large amount of big data aggregated from bank credit records to assess the credit risk associated with small and medium-sized enterprises (SMEs). The model utilises a comprehensive dataset that provides detailed analysis of key metrics such as credit history, transaction behaviour and financial ratios, which significantly improves the predictive accuracy associated with credit defaults by SMEs. The results of this study highlight the important role that big data and advanced analytics play in the field of financial risk management, particularly in the area of SME lending. By employing logistic regression modelling, we were able to take into account a wider range of risk factors and process large amounts of data, thereby improving the quality and efficiency of credit decisions. Analyses of a dataset containing the creditworthiness of 100 companies show a clear correlation between a company's credit rating and its likelihood of default. This correlation highlights the importance of credit assessment as a tool for predicting financial reliability and risk management in the corporate sector. In addition, the study demonstrated the effectiveness of logistic regression models in assessing and predicting credit risk, providing valuable metrics for financial decision makers.

In our quantitative analysis of credit risk, we reconstruct credit risk as a function of the firm's performance rate. The logistic regression model suggests that the higher the creditworthiness of the firm, the lower the default rate and hence the lower the credit risk. This study emphasises that the selection of indicators must adhere to the principles of science, comprehensiveness, fairness, relevance, legitimacy and practicality. These principles are the basis for constructing a suitable evaluation index system that accurately reflects the multifaceted nature of credit risk. The regression coefficients obtained from the logistic regression model have both positive and negative values, reflecting the different effects of each predictor on the response variable. The compliance rates of the 100 companies calculated using the logistic regression model varied considerably. Most of the companies maintained a compliance rate above 0.7, indicating a high level of compliance with protocols or standards. However, some companies had significantly lower compliance rates, indicating the need to tailor credit solutions to the specific needs of SMEs. In conclusion, this study demonstrates the effectiveness of using big data-driven logistic regression models for SME credit risk analysis. The insights gained from this study can help financial institutions to enhance their risk assessment methodologies and tailor credit solutions to the specific needs of SMEs while effectively managing potential credit defaults. This study contributes to the continuous improvement of SMEs' access to finance and highlights the indispensable role of big data and cutting-edge analytics in financial risk management.

6 DISCUSSIONS

In this study, we employ a logistic regression model to assess the credit risk associated with small and micro enterprises (SMEs) through a large dataset of bank credit records. An important aspect of our analysis is the emphasis on the compliance rate of firms as an important indicator of firm reputation. By cleverly using the inverse of the compliance rate ($1-P_i$) as a robust predictor of a bank's potential credit loss rate, we reveal the intricate relationship between corporate credit behaviour and risk. We skilfully employ a logistic regression framework to operationalise this dynamic interaction through a set of indicators that are scientifically accurate, fair and neutral. This ensures the comprehensiveness, relevance, legitimacy, and real-world practical applicability of the selected indicators. The rigorous development process of these sophisticated models not only strengthens risk assessment protocols, but also provides valuable insights into the formulation and implementation of prudent credit policies.

The application of these sophisticated models has significantly improved the financial resilience and risk mitigation capabilities of banking institutions that extend credit to SMEs. By providing a more nuanced understanding of credit risk and its determinants, this study represents a major advance in the field of financial risk management. It provides banks with the tools to make more informed lending decisions, thereby contributing to the sustainable growth of small and micro-enterprises and to the overall stability of the financial system. This study uses big data-driven logistic regression models to highlight the indispensable role that big data and cutting-edge analytics play in the field of financial risk management, especially in the area of SME lending. The ability of these models to process large amounts of data and consider a wider range of risk factors greatly improves the quality and efficiency of credit decision-making. One of the main contributions of this study is the development of a scientifically accurate, fair and neutral indicator system that can effectively capture the multifaceted nature of credit risk. This comprehensive risk assessment methodology has the potential to revolutionise the way financial institutions assess and manage SME credit risk.

In conclusion, this study demonstrates the effectiveness of using a big data-driven logistic regression model for credit risk analysis of MSMEs. The insights gained from this study can help financial institutions improve their risk assessment methods and tailor credit solutions to the specific needs of SMEs, while effectively managing potential credit defaults. The findings of this study are of great significance to the field of financial risk management and can contribute to ongoing efforts to improve access to finance for SMEs.

REFERENCES

[1] Sun, Y., Zeng, X., Cui, X., Zhang, G., & Bie, R. 2021. An active and dynamic credit reporting system for SMEs in China. *Personal and Ubiquitous Computing, 25*, 989-1000.

[2] Xiao, P., Salleh, M. I. B., & Cheng, J. 2022. Research on Factors Affecting SMEs' Credit Risk Based on Blockchain-Driven Supply Chain Finance. Information, 13(10), 455.

[3] Roeder, J., Palmer, M., & Muntermann, J. 2022. Data-driven decision-making in credit risk management: The information value of analyst reports. Decision Support Systems, 158, 113770.

[4] Yıldırım, M., Okay, F. Y., & Özdemir, S. 2021. Big data analytics for default prediction using graph theory. Expert Systems with Applications, 176, 114840.

[5] Cheraghali, H., & Molnár, P. 2023. SME default prediction: A systematic methodology-focused review. *Journal of Small Business Management*, 1-59.

[6] Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., & Shen, H. 2021. Combating emerging financial risks in the big data era: A perspective review. *Fundamental Research, 1*(5), 595-606.

[7] Barddal, J. P., Loezer, L., Enembreck, F., & Lanzuolo, R. 2020. Lessons learned from data stream classification applied to credit scoring. *Expert Systems with Applications, 162*, 113899.

[8] Zhang, L., & Song, Q. 2022. Credit evaluation of SMEs based on GBDT-CNN-LR hybrid integrated model. Wireless Communications and Mobile Computing, 2022, 1-8.

[9] Li, G. 2022. The prediction algorithm of credit risk of science and technology finance based on cloud computing. *Journal of Computational Methods in Sciences and Engineering, 22*(1), 235-251.

[10] Hasan, M. M., Popp, J., & Oláh, J. 2020. Current landscape and influence of big data on finance. *Journal of Big Data, 7*(1), 1-17.

[11] Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. 2020. Corporate default forecasting with machine learning. Expert Systems with Applications, 161, 113567.

[12] Gu, Z., Lv, J., Wu, B., Hu, Z., & Yu, X. 2024. Credit risk assessment of small and micro enterprise based on machine learning. *Heliyon, 10*(5).

[13] Yang, F., & Masron, T. A. 2024. Role of financial inclusion and digital transformation on bank credit risk. *Journal of International Financial Markets, Institutions and Money, 91*, 101934.

[14] Lin, Y. H., Zheng, R., Wu, F., Zeng, N., Li, J., & Tao, X. 2024. Blockchain-driven framework for financing credit in small and medium-sized real estate enterprises. *Journal of Enterprise Information Management, 37*(1), 201-229.

[15] Du, Y., Xu, H., & Chen, Y. 2024. Digital empowerment and innovation in risk control strategies for fishery supply chain finance—a case study of Puhui agriculture and animal husbandry financing guarantee company limited. *Marine Development, 2*(1), 1-20.

Paper Two Publication

https://www.scopus.com/record/display.uri?eid=2-s2.0-85212696405&origin=resultslist&sort=plf-f&src=s&sot=a&s=ORCID%280000-0002-1483-7720%29&sessionSearchId=9f6506592c129e7a1de6a6ce303263b4

Micro and Small Enterprises Credit Risk and Response Strategies to Emerging Factors Based on Big Data-Driven Multi-Objective Nonlinear Model

Guoqing Chen
Department of Mathematics, Faculty of Science, Mahasarakham University
Thailand
chenguoqing1139@163.com

Piyapatr Busababodhin*
Department of Mathematics, Faculty of Science, Mahasarakham University
Thailand
piyapatr.b@msu.ac.th

Nipaporn Chutiman
Department of Mathematics, Faculty of Science, Mahasarakham University
Thailand
nipaporn.c@msu.ac.th

Abstract

This study constructs a framework based on a multi-objective nonlinear programming model, aiming to systematically explore the credit risk and contingency factors coping strategies for SMEs. By taking into account a variety of factors such as national policy inclination, the degree of industry influence, enterprise size and customer turnover rate, the model provides banks with optimal credit decisions in a complex economic environment. The results show that in the context of emergencies such as epidemics, banks can support affected SMEs by adjusting loan amounts and interest rate preferences, while effectively controlling credit risks. Although the model shows high effectiveness in theoretical and empirical analyses, it still faces challenges such as data acquisition and model solving in practical applications. Future research can combine machine learning and big data technologies to further improve the prediction accuracy and applicability of the model. This study not only enriches the theoretical framework of financial risk management, but also provides scientific and systematic guidance for banks' lending decisions in special economic environments, which has important academic and practical value.

CCS Concepts

• **Mathematics of computing**; • **Mathematical analysis**; • **Nonlinear equations**; • **Mathematical software**; • **Statistical software**; • **Applied computing**; • **Enterprise computing**; • **Enterprise modeling**;

Keywords

Multi-objective nonlinear programming, small and medium-sized enterprises, credit risk, emerging factors, bank credit decisions

ACM Reference Format:
Guoqing Chen, Piyapatr Busababodhin, and Nipaporn Chutiman. 2024. Micro and Small Enterprises Credit Risk and Response Strategies to Emerging Factors Based on Big Data-Driven Multi-Objective Nonlinear Model. In *2024 International Conference on Big Data and Digital Management (ICBDDM 2024), August 16–18, 2024, Shanghai, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3696500.3696547

*Corresponding author

ICBDDM 2024, August 16–18, 2024, Shanghai, China

ACM ISBN 979-8-4007-1027-8/24/08
https://doi.org/10.1145/3696500.3696547

1 INTRODUCTION

Small and medium-sized enterprises (SMEs) are a key force in economic development, but because of their small size and weak risk-resistance, they are vulnerable to the market environment, national policies and emergencies (e.g., the New Crown Epidemic), which indirectly affects the safety of banks' capital and profitability. Studies have shown that the epidemic has led to the deterioration of bank loan quality and the increase of loan delinquency rate and non-performing loan rate, therefore, banks need to support SMEs and at the same time effectively control the credit risk [1]. This paper proposes a credit risk management strategy for SMEs based on a multi-objective nonlinear programming model, which takes into account factors such as national policy, industry impact, enterprise scale and customer churn rate, and constructs a multi-objective optimisation framework to maximise bank returns and take into account risk control. By solving the model, the bank can obtain specific suggestions for lending decisions, including whether to lend or not, loan amount and interest rate, which helps the bank to achieve sound operation in the complex economic environment [2]. The study enriches the financial risk management theory and provides support for banks' lending decisions. The results show that banks can effectively respond to SMEs' funding needs by flexibly adjusting their credit strategies, while controlling risks and achieving long-term profitability. Future research can further optimise the model and dynamically adjust it with empirical data to enhance its application value and provide support for the synergistic development of banks and SMEs [3].

2 RELATED WORK

Various types of emergencies have different impacts on enterprises of different industries, scales and categories, so it is necessary to classify these enterprises systematically and study their scale and industry characteristics in depth. Based on the classification statistics of 100 enterprises with no credit records, this paper reveals their scale and industry distribution characteristics to provide reference for policy formulation and enterprise management. In the sample, small and micro enterprises occupy the majority share, while medium-sized enterprises and self-employed households are relatively few [4]. This reflects the disadvantage of small and microenterprises in terms of access to finance, which makes them more likely to be in a no-credit-record status.

Table 1: Size of enterprises and industry statistics.

Serial number	Industry Classification	Medium Enterprise	Small Enterprises	Micro Enterprise	Self-employed	Total
1	Industrial	2	1	0	0	3
2	Construction	0	2	7	0	9
3	Transport	1	4	6	0	11
4	Retail	1	4	5	0	10
5	Agriculture, forestry, animal husbandry and fishery	0	1	6	0	7
6	Wholesale	3	1	2	0	6
7	Software and Information Technology Services	0	6	5	0	11
8	Property Management	0	1	3	0	4
9	Information Transmission	0	3	5	0	8
10	Accommodation	1	2	1	0	4
11	Rental and business services	0	1	6	0	7
12	Catering Industry	0	1	2	0	3
13	Property Development and Operation	0	0	3	0	3
14	Other industries not listed	0	0	6	8	14

Table 2: Year-on-Year Comparison of Operating Conditions of Enterprises in the 1st Quarter.

Enterprise code	Q1				Business Category	Enterprise size
	Sales		Profit			
	Year 2018	Year 2019	Year 2018	Year 2019		
Co1	-79.99%	-85.01%	37.71%	61.18%	Wholesale	Micro
Co2	-44.56%	-77.01%	-500.13%	11.62%	Software and Information Services	Small
Co3	-69.03%	-78.18%	136.68%	-77.61%	Software and Information Services	Small
Co4	-97.42%	-98.43%	102.63%	-98.37%	Transportation	Small
Co5	-98.87%	-98.87%	100.98%	100.17%	Other unspecified industries	Micro
...						
Co96	-88.03%	-93.36%	-88.03%	-93.36%	Wholesale	Micro
Co97	-82.48%	-87.11%	-56.52%	-235.72%	Software and Information Services	Small
Co98	-97.82%	-98.41%	-1100.42%	-1646.41%	Industrial	Micro
Co99	-90.87%	-85.18%	-73.86%	119.99%	Rental and business services	Micro
Co100	-98.76%	-98.83%	96.89%	88.71%	Other unspecified industries	Micro

An analysis of Table 1 reveals the distribution of various types of enterprises across different sizes and industry categories. First, as a whole, micro and small enterprises account for the majority, while medium-sized enterprises and self-employed persons are relatively few [5]. This reflects the fact that small and micro-enterprises are more common among enterprises with no credit history, probably because these enterprises are more restricted in their access to finance.

In the industrial sector, there were two medium-sized enterprises and one small enterprise, while there were no micro-enterprises or self-employed persons. This suggests that the industrial sector is characterised by larger firms and smaller firms are less likely to be involved. The construction sector, on the other hand, is dominated by small and micro-enterprises, with nine enterprises, two of which are small and seven are micro-enterprises, indicating that enterprises in the construction sector are generally smaller in size [6]. The transport sector is also dominated by small and micro enterprises, with a total of 11 enterprises, of which 4 are small and 6 are micro enterprises, indicating that small, medium and micro enterprises are more active in the transport sector.

The 100 businesses were analysed to cover all industry categories except catering. By calculating the year-on-year growth rate of profits in the first quarter of 2020 versus 2018 and 2019, it reveals how each enterprise operates in different years. Table 2 provides the operating data of selected enterprises in Q1 and their year-on-year growth rates. Analysing these data, it is evident that there is a significant decline in sales items and profits in general across the firms in Q1 2020.

Analysing data from 100 businesses with no credit history shows significant differences in resilience by sector and size. Wholesale

Table 3: Ratios of year-on-year reductions in sales and profits by firms, 1Q 2020.

Business Situation	Sales		Profit	
Year	Year 2018	Year 2019	Year 2018	Year 2019
Year-on-year reduction ratio	61.12%	69.45%	47.92%	47.22%

microenterprise Co1 saw sales fall by 79.99 per cent and 85.01 per cent in 2018 and 2019 respectively, but profits increased by 37.71 per cent and 61.18 per cent. Sales in Co2, a software and information services microenterprise, fell by 44.56 per cent and 77.01 per cent, while profits rose from -500.13 per cent to 11.62 per cent. On the contrary, Co3 and Co4, small enterprises in the transport sector, saw a sharp decline in profits in 2019. Microenterprises Co5 in other unspecified industries experienced a 98.87 per cent drop in sales but less volatile profits. The deteriorating economic environment in the first quarter of 2020 led to a significant drop in sales and profits for most enterprises, particularly Co96 in wholesale and Co97 in software and information services, demonstrating the vulnerability of the enterprises. Industrial microenterprises Co98 saw profits fall from -1100.42 per cent to -1646.41 per cent in the face of market shocks, while Co99 in leasing and business services and Co100 in other unspecified sectors showed some resilience with less volatility in profits despite falling sales.

Table 3 shows that firms' sell-through and profit decreased significantly in the first quarter of 2020 compared to 2018 and 2019. In terms of sales, the year-on-year reduction ratios are 61.12% and 69.45% in 2018 and 2019, respectively, reflecting a significant reduction in market demand in early 2020. On the profit side, the year-on-year reduction ratios for 2018 and 2019 are 47.92 per cent and 47.22 per cent, respectively, indicating that corporate profitability has been severely hit. Although the decline in profit is slightly lower than that in sales, which may indicate that some companies have made adjustments in cost control or high-margin businesses, the overall profitability is still not optimistic.

In the first quarter of 2020, companies experienced significant year-on-year declines in both sales and profits. This trend was particularly pronounced in 2019, with sales declining at a rate of nearly 70% and profits declining at a rate of nearly 50%, suggesting that the deteriorating market environment has had a broad and far-reaching impact on business operations. This situation can stem from a variety of complex factors including, but not limited to, sharp shifts in market demand, disruptions in the global supply chain, and changes in the policy environment. For enterprises, how to effectively adjust their business strategies, optimise resource allocation, and improve risk resistance in such an environment has become an important issue to be addressed [7].

Table 4 shows the year-on-year changes in turnover of different types of enterprises in the first quarter of 2020, specifically including the number of enterprises with year-on-year increases and decreases in turnover, as well as the frequency of enterprises with year-on-year increases in turnover [8]. By analysing this data, we can gain insights into the performance of each sector in the face of economic volatility.

In the first quarter of 2020, the turnover of companies in most sectors fell significantly year-on-year. The real estate development and business sector rose at a frequency of 20 per cent, industry at 13.04 per cent, construction at only 7.69 per cent and transport at 0 per cent. The retail trade sector performed better, with a rise frequency of 40 per cent. The agriculture, forestry, animal husbandry and fisheries industry was 25 per cent, the wholesale trade and information transmission industry were both 40 per cent, the software and information technology services industry was 28.57 per cent, the property management industry was 33.33 per cent, the accommodation industry and the leasing and business services industry were both 16.67 per cent, and the unspecified industry was 25 per cent. Overall, the retail, wholesale and information transmission industries performed relatively well, while the transport and construction industries were the most vulnerable [9].

The new crown epidemic has severely affected MSMEs, leading to stagnant production, rising unemployment, declining consumer power, declining turnover, declining repayment capacity, and increased demand for loans, but rising overdue and non-performing loan rates and increased risk of bank loan losses. The state has introduced policies to support MSMEs such as tax cuts, utility bill reductions, loan extensions and interest rate reductions. Banks need to cooperate with the policies to provide loan support and manage loan risks [10]. Credit strategies should be tilted towards areas such as pharmaceutical protection and livelihood protection, while seizing opportunities in emerging industries such as 5G and online education, and adjusting credit strategies to support industrial development and achieve long-term profitability.

Table 5 demonstrates the specific loan amount and interest rate adjustment strategies of banks under different impact level ratings. By analysing the data in the table, it can be seen that banks have made differentiated adjustments to their lending policies according to the degree of impact of the epidemic on enterprises, as a way to support the capital needs of the affected enterprises and at the same time seize potential market opportunities.

For enterprises seriously affected by the epidemic, banks adopted a proactive support policy by raising the ceiling of the loan amount by 30 per cent and offering interest rate concessions of 35 to 55 per cent to alleviate the pressure on capital and maintain operations, while easing the financial burden, improving repayment capacity and reducing credit risks. Some of the affected enterprises were offered interest rate concessions of 25-35 per cent, but the loan amount remained unchanged, providing the necessary financial support while avoiding excessive lending and ensuring financial security [11]. For enterprises with no or minimal impact, the bank's lending quota and interest rate policy remained unchanged to control overall credit risk and rationally allocate resources. For enterprises favoured by the epidemic, banks raised the upper limit of the loan amount by 25 to 55 per cent, but kept interest rates unchanged, seizing market opportunities, supporting rapid development and improving asset quality and profitability.

3 MODEL BUILDING

This paper constructs a multi-objective non-linear planning model to optimise the bank's lending decisions by considering a variety

Table 4: Frequency of year-on-year turnover decrease in Q1 2020 for different types of enterprises.

Business Type	Number of companies with year-on-year increase in turnover	Number of companies with year-on-year decrease in turnover	Frequency of enterprises with year-on-year turnover increase
Property Development	1	4	20.00%
Industrial	3	20	13.04%
Construction	1	12	7.69%
Transport	0	4	0.00%
Retail	4	6	40.00%
Agriculture, forestry, animal husbandry and fishery	1	3	25.00%
Wholesale	2	3	40.00%
Software and Information Technology Services	4	10	28.57%
Property Management	1	2	33.33%
Information Transmission	2	3	40.00%
Accommodation	1	4	16.67%
Leasing and business services	2	10	16.67%
Unspecified industry	2	6	25.00%

Table 5: Banks' specific adjustment strategies.

Impact Rating	Loan Limit Adjustment Policy	Loan Rate Adjustment Policy
Serious impact	30% increase in the maximum loan amount	Loan Rate Offer 35-55 per cent
Partial impact	No change	Loan Interest Rate Offer 25%-35
No Impact / Minimal Impact	No change	No change
Favourable	25%-55% increase in loan limit ceiling	No change

of factors such as national policy tilts and the degree to which each industry is affected, which are quantified and introduced into the lending strategy adjustment model. The model aims to maximise the bank's returns while taking into account risk control and policy alignment, with a specific objective function:

$$maxZ = \sum_{j=1}^{100} y_i \left[A_j \left(1+h_{1j}\right)\left(1-R_j\right) i_j \left(1-h_{2j}\right) - 0.22 \cdot A_j \left(1-h_{2j}\right) R_j\right] - M_j \cdot q_j \quad (1)$$

Where h_{1j} denotes the enhancement coefficient of the bank's loan amount to different enterprises, which is affected by factors such as national policy subsidies, the industry to which the enterprise belongs, the enterprise's size, and the customer turnover rate, etc.; h_{2j} denotes the coefficient of interest rate preference of the bank to different enterprises, and the influencing factors are similar to those of h_{1j}.

The model needs to satisfy the following constraints:

$$\sum_{j=1}^{100} y_i A_j \le d \quad (2)$$

$$10 \le A_j \le E_j \quad (3)$$

$$0.04 \le i_{kj} \le 0.15 \quad (4)$$

In the above constraints, y_i is a dichotomous variable and $y_i = 1$ when $t_j > 0$; otherwise, $y_i = 0$. t_j denotes the firm's net income over the loan period, which is calculated as:

$$t_j = A_j \cdot i_j \cdot P_j - A_j \cdot 0.22 \cdot \left(1 - P_j\right) \quad (5)$$

$$y_j = \begin{cases} 0, t_j \le 0 \\ 1, t_j > 0 \end{cases} \quad (6)$$

where P_j is the probability of repayment by the firm; M_j is the value of customer churn; and q_j is the loan quality function, denoted as.

$$q = f(i) \quad (7)$$

M, P, R, and E in the model denote the value of customer churn, the probability of repayment by the firm, the probability of default by the firm, and the maximum loan limit, respectively, and these parameters are described and computed through a series of variables $x_1, x_2, x_3, x_4, x_5, x_6, x_7$, which take the following functional form:

$$M = \eta\left(x_1, x_2, x_3, x_4, x_5, x_6, x_7\right) \quad (8)$$

$$R = \phi\left(x_1, x_2, x_3, x_4, x_5, x_6, x_7\right) \quad (9)$$

$$E = \tau\left(x_1, x_2, x_3, x_4, x_5, x_6, x_7\right) \quad (10)$$

This multi-objective non-linear planning model seeks to optimise the bank's loan portfolio while minimising risk and improving returns by taking multiple factors into account. The quantification of national policy tilts and the degree of industry impact makes the model more realistic and instructive, especially in the current complex economic environment, and provides banks with a scientific decision-making tool that helps them to maintain their own

Table 6: Optimal credit strategies for banks.

Enterprise	Whether to lend	Amount/Ten thousand	Interest Rate
Co1	1	11.2081	0.0581
Co2	1	25.8444	0.0478
Co3	1	25.8367	0.0508
Co4	1	25.8294	0.0441
Co5	1	25.8367	0.0407
...			
Co96	0	39.8794	0.0841
Co97	0	77.8741	0.0364
Co98	1	63.7892	0.0134
Co99	0	77.8741	0.0364
Co100	0	77.8741	0.0364

sound development while supporting enterprises affected by the epidemic.

4 MODEL SOLVING AND ANALYSIS

Based on the new bank loan adjustment policy identified in Table 5 and the solution of the multi-objective planning model, this paper calculates the optimal lending strategy for all firms, including information on whether or not to lend, the amount of the loan, and the interest rate of the loan. Table 6 shows the specific lending strategies for some enterprises, reflecting the decision-making process of banks in a complex economic environment.

The data in Table 6 shows that the bank has adopted differentiated credit strategies for different enterprises based on the multi-objective non-linear programming model. For example, the bank lent RMB 112,081 to Co1 at an interest rate of 0.0581; it lent RMB 258,444 and RMB 258,367 to Co2 and Co3 at an interest rate of 0.0478 and 0.0508, respectively, indicating that the bank comprehensively considered the enterprises' capital demand, repayment ability and market risk when lending. For Co96, Co97, Co99 and Co100, the banks chose not to lend despite the higher calculated loan amounts, reflecting a cautious attitude towards high default risk and repayment uncertainty. In contrast, Co98 loan interest rate was as low as 0.0134 and the bank still chose to lend RMB 637,892 due to its better repayment ability and market outlook, which reduces credit risk. Overall, the bank flexibly adjusted its lending strategy in the complex market environment through the multi-objective non-linear planning model to satisfy the enterprise's capital demand while controlling the credit risk to achieve the goals of sound operation and long-term profitability.

5 DISCUSSIONS

Based on a multi-objective non-linear programming model, this study explores the credit risk and unexpected factors coping strategies of SMEs, providing banks with new credit decision-making methods in a complex economic environment. The results show that banks can accurately adjust their credit strategies by taking into account factors such as national policies, industry impacts, enterprise size and customer turnover, especially in the context of emergencies such as epidemics, where flexible adjustments are particularly important. The lending strategies suggested by the model help banks to effectively control credit risks and avoid rising non-performing loan ratios while supporting enterprises to tide over difficult times. The study finds that there are significant differences in credit demand and risk characteristics across industries and firm sizes, and that banks can adjust their loan amounts and interest rate preferences according to the degree of impact in order to reduce financial pressure or seize market opportunities. Although the model faces challenges in practical application such as data acquisition and computational power, its prediction accuracy and applicability can be improved in the future by combining machine learning and big data technologies. Overall, this study provides scientific and systematic solutions for SME credit risk management, enriches the theory of financial risk management, provides powerful support for bank credit decision-making, and is of great significance for enhancing banks' ability to cope with emergencies and promoting the sustainable development of SMEs.

6 CONCLUSION

This study constructs a multi-objective non-linear planning model to explore the credit risk and unexpected factors coping strategies for SMEs. The results show that by integrating national policies, industry impacts, enterprise size and customer turnover, banks can formulate flexible and precise credit strategies to support SMEs in a complex environment. The application of the model shows that banks can adjust the loan amount and interest rate preference under the epidemic to help the impacted enterprises tide over the difficulties, and control the credit risk through differentiated strategies to achieve sound operation. Despite the validity of the model, the practical application faces data and computing power challenges. Future research should combine machine learning and big data technologies to improve model accuracy and applicability. This study enriches the theory of financial risk management and is of great practical significance in providing decision-making guidance for banks to achieve a balance between risk control and profitability in a special economic environment.

References

[1] Abbasi, D., Ashrafi, M., & Ghodsypour, S. H. **2020**. A multi objective-BSC model for new product development project portfolio selection. Expert Systems with Applications, 162, 113757. https://doi.org/10.1016/j.eswa.2020.113757.

[2] Chen, Z. S., Zhou, J., Zhu, C. Y., Wang, Z. J., Xiong, S. H., Rodríguez, R. M., ... & Skibniewski, M. J. **2023**. Prioritizing real estate enterprises based on credit risk assessment: an integrated multi-criteria group decision support framework. Financial Innovation, 9(1), 120. https://doi.org/10.1186/s40854-023-00412-x.

[3] Ni, Z., Bai, M., & Dong, H. **2022**. An improved multi-objective grey target decision model for corporate decision making. Highlights in Science, Engineering and Technology, 4, 192-205. https://doi.org/10.54097/hset.v4i.192.

[4] Gheitasi, M., Feylizadeh, M. R., & Ahari, R. M. **2022**. A multi-objective mathematical model of financial flows in Omni-Channel distribution systems. Journal of Intelligent & Fuzzy Systems, 42(6), 4851-4879. https://doi.org/10.3233/JIFS-211875.

[5] Ghaemi-Zadeh, N., & Eghbali-Zarch, M. **2024**. Evaluation of business strategies based on the financial performance of the corporation and investors' behavior using D-CRITIC and fuzzy MULTI-MOORA techniques: A real case study. Expert Systems with Applications, 247, 123183. https://doi.org/10.1016/j.eswa.2023.123183.

[6] Sukumar, A., Mahdiraji, H. A., & Jafari-Sadeghi, V. **2023**. Cyber risk assessment in small and medium-sized enterprises: A multilevel decision-making approach for small e-tailors. Risk Analysis, 43(10), 2082-2098. https://doi.org/10.1111/risa.14043.

[7] Emtehani, F., Nahavandi, N., & Rafiei, F. M. **2023**. Trade credit financing for supply chain coordination under financial challenges: a multi-leader–follower

game approach. Financial Innovation, 9(1), 6. https://doi.org/10.1186/s40854-023-00422-9.

[8] Song, X., Wu, X., Liu, C., & Yang, H. **2022**. The Order Financing Risk Evaluation under the Financial Supply Chain Model Based on the Logistic Model. Journal of Mathematics, 2022(1), 1854436. https://doi.org/10.1155/2022/1854436.

[9] Amaral, F. V. A., & Gartner, I. R. **2021**. Portfolio impact investment management using multi-objective optimization. Pesquisa Operacional, 41, e245832. https://doi.org/10.1590/0101-7438.2021.041.00245832.

[10] Wan, J., & Yu, B. **2023**. Early warning of enterprise financial risk based on improved BP neural network model in low-carbon economy. Frontiers in Energy Research, 10, 1087526. https://doi.org/10.3389/fenrg.2022.1087526.

[11] Sahoo, P. B. B., & Thakur, V. **2023**. Enhancing the performance of Indian micro, small and medium enterprises by implementing supply chain finance: challenges emerging from COVID-19 pandemic. Benchmarking: An International Journal, 30(6), 2110-2138. https://doi.org/10.1108/BIJ-09-2022-0531.

International Conference One

International Conference on Applied Statistics (ICAS2022)

International Conference on Applied Statistics 2022
3 - 4 November, 2022 a Virtual Conference

21 October 2022

Paper ID: 8314

Title: Combined prediction model of gasoline octane loss value based on XGboost-GBDT combined prediction

Authors: Guo-Qing Chen, Zi-Qiang Fang, Hai-Yang Peng, Busababodhin P and Huiyan Wang

Dear Authors,

Your manuscript entitled “Combined prediction model of gasoline octane loss value based on XGboost-GBDT combined prediction” has been carefully reviewed by three ICAS2022 academic peer reviewers. On behalf of ICAS2022 academic committee, it is my pleasure to inform you that your paper has been accepted for **oral presentation** in ICAS2022. Your full paper will also be published in the conference proceedings.

Please proceed the online registration on the conference webpage. The presentation schedule will be announced online. You are required to give a presentation at the designated time.

Thank you for submitting your paper to ICAS2022. We look forward to seeing you at ICAS2022 in November 3-4, 2022.

Best Regards,

Prof. Dr. Yupaporn Areepong

EDITOR : International Conference on Applied Statistics 2022 (ICAS2022)

International Conference Two

International Conference on Applied Statistics (ICAS2024)


ICAS 2024

The 7th International Conference on Applied Statistics 2024 (ICAS2024)

this certificate is proudly presented to:

Guoqing Chen, Piyapatr Busababodhin, Nipaporn Chutiman and Tossapol Phoophiwfa

for presenting the paper entitled:

Gradient Descent Decision Tree Algorithm and Nonlinear Programming for Credit Risk Assessment and Credit Strategy

October 24 - 25, 2024 , Chiang Mai, Thailand

Dr.Siwiga Dusadenoad
DEAN
GRADUATE SCHOOL OF APPLIED STATISTICS
NATIONAL INSTITUTE OF DEVELOPMENT ADMINISTRATION

Prof. Dr. Torranin Chairuangsri
DEAN
FACULTY OF SCIENCE
CHIANG MAI UNIVERSITY

NIDA GSAS Graduate School of Applied Statistics

HART

REFERENCES

BIOGRAPHY

NAME	Chen Guoqing
DATE OF BIRTH	1993/05/01
PLACE OF BIRTH	Anhui province
ADDRESS	Xipu Street, Pidu District, Chengdu City, Sichuan Province
POSITION	No. 33 Fuzi Road
PLACE OF WORK	No. 1 Xiyuan Avenue, High-tech West District, Chengdu City, Sichuan Province
EDUCATION	2006.09-2012.07 Middle School in Hexian Second Middle School, Anhui Province, China 2012.09-2016.07 Bachelor of Audit in School of Business of Anhui University of Finance and Economics,China 2016.09-2019.07 Master of Corporate Finance in Xihua University,China 2022.07-2025.04 Doctor of Philosophy in Statistical Management Science, Mahasarakham University,Thailand