ISSN 2616-5902 Vol. 6, Issue 11: 214-218, DOI: 10.25236/AJBM.2024.061131

Research on Financial Credit Risk of Manufacturing Enterprises under Heterogeneous Data Based on Machine Learning

Yunpeng Zhao

Treasury Department, Bank of China, New York, NY 10018, USA
JackZ20241@outlook.com

Abstract: *This research focuses on a new method to evaluate the financial credit risk of manufacturing enterprises in a heterogeneous data environment, taking Chinese manufacturing enterprises as an example. This paper uses principal component analysis and K-means clustering technology to quantitatively score and grades the credit status of enterprises and systematically evaluates the relative importance of each credit index. To solve the challenge caused by the imbalance of data categories, the study introduced the SMOTE oversampling method to optimize the prediction accuracy of the machine learning model. After comparing the performance of various machine learning models, the best-performing model is selected as the core tool of stress testing, which is used to evaluate the coping ability of enterprises in various industries under different stress scenarios. The results show that: (1) there are significant differences in the effects of various credit indicators on the financial credit risk of manufacturing enterprises. The solvency of the industry is the most critical risk factor, while the impact of enterprise operation ability is relatively small; (2) Under the test condition of increasing pressure step by step, the MLP model shows the optimal stability and prediction accuracy, which is better than other models; (3) Enterprises in different subsectors show obvious differences in anti-pressure ability when facing external pressure. General equipment manufacturing enterprises have strong anti-pressure ability, while special equipment manufacturing enterprises show greater vulnerability. The results of this study have improved the understanding of risk assessment mechanisms, expanded the scope of application of methods, and made a significant contribution to the theoretical basis of financial credit risk assessment, which has greatly contributed to the development of this area. These discoveries have led to the creation of several innovative strategies designed to effectively manage credit risks within model companies, which provide stakeholders with valuable insights into the complexity of navigation solutions in a multifaceted environment. Using these ideas, stakeholders can make more accurate and wise decisions, effectively reducing the risk of exposure and contributing to the sustainable growth and stability of enterprises. This doubling of theoretical understanding and practical application enhances the overall effectiveness of credit risk management practices, strengthening a strong and adaptive strategy to ensure that the subtle and ever-changing challenges of today's financial landscape are specifically designed.*

Keywords: *Manufacturing Enterprises, Financial Credit Risk, Machine Learning, Disequilibrium Data, Principal Component Analysis, K-Means Clustering*

1. Introduction

With the continuous development of the global economy, manufacturing enterprises play an increasingly important role in the economic system. The "Made in China 2025" strategy clearly defines the goal of promoting the modernization of the manufacturing industry. Still, these enterprises face multiple financial credit risks such as capital shortage, high financing costs, and supply chain disruptions, which can cause serious shocks to the enterprises and the financial system. Traditional credit risk assessment methods make it difficult to deal with these complex and dynamic risks effectively, and more advanced techniques are needed. Stress testing, as a tool to assess the performance of companies under extreme economic scenarios, can effectively simulate potential crises. The research on integrating it with modern machine-learning techniques is still limited. Machine learning techniques, due to their ability to handle complex non-linear relationships and use big data to provide accurate risk assessment, are gradually showing their advantages in financial credit risk assessment. In this study, principal component analysis and K-means clustering are used to assess the

ISSN 2616-5902 Vol. 6, Issue 11: 214-218, DOI: 10.25236/AJBM.2024.061131

credit risk of manufacturing enterprises, and the SMOTE oversampling technique is used to solve the imbalance of data categories, to improve the prediction accuracy of the machine learning model. In addition, the study also conducted a stress test on financial credit risk and selected the MLP model with excellent performance for stress conduction analysis to evaluate its anti-stress ability under different stress scenarios. This research not only enriched the methods of financial credit risk warning but also provided scientific risk management tools for enterprises and financial institutions.

2. Related Research

As the core component of the national economy, the credit management of manufacturing enterprises is critical to achieving long-term sustainable development. To further study the financial credit risk of these enterprises, 1,987 manufacturing enterprises in the Shenzhen A-share market during 2013-2022 were selected as data samples. In terms of data processing, first of all, the enterprise samples whose index missing rate is more than one-third are eliminated, and some missing values are filled with the group mean method. After processing, the final form of 8,254 valid samples, involving 1,749 companies, all data from the CSMAR database. H Chen, Y Yang, and C Shao proposed a multi-task learning method to realize efficient spatio-temporal modeling of data. A hierarchical Bayesian inference structure based on the Gaussian process is constructed for knowledge transfer between multiple similar but not identical measurement tasks [1]. Varatharajah's team discusses the potential use of reinforcement learning (RL) --based human-in-loop recommendation systems to support the clinical management of COVID-19 [2]. TR Fan team proposed an interoperation framework for distributed heterogeneous Web databases based on mobile agents [3]. This paper analyse and benefit from heterogeneous data in applications, the Kamm team needed to integrate, store, and manage data from different sources to apply machine learning algorithms [4].

3. Application Design of Stress Testing In Machine Learning-Driven Financial Credit Risk Warning Model for Manufacturing Enterprises

3.1 Application of Comprehensive Score and Risk Classification in Machine Learning Financial Credit Risk Assessment

In the area of credit risk assessment, traditional classification technologies (e.g., "st" and "st" dichotomies) are often insufficient in the face of complex and multifaceted problems associated with corporate loans; while these techniques are easily applied, they are often insufficiently precise and detailed to ensure that credit risk is fully assessed. Given these limitations, the study attempts to address the shortcomings by developing a multi-stage forecasting model integrated into a more rigorous classification method aimed at improving the accuracy and practicality of credit risk assessment. Using a complex and detailed classification method, the study aims to provide a comprehensive analysis, a more accurate representation of the inherent complexity and variability of credit risks, and stimulate decision-making to achieve a more efficient and reliable risk assessment structure. This advanced approach would like to propose a stronger, better credit risk assessment and management system that supports a more complex, more sensible way of assessing risks and managing them [5].

In this study, the core component analysis (CSA) is used professionally to calculate a complex enterprise risk assessment, which results in a complex and multidimensional set of data becoming a single and consistent indicator that sums up a common list of credit risks for each company; This approach achieves its objective by identifying the key risk factors in a single, integrated assessment, which is an indicator of the level of corporate credit risk at which the lower the level of risk, the lower the risk it will receive. With the help of the k-mean polygon algorithm, these complex assessments are systematically divided into separate risk categories, dividing enterprises into five different categories: from a minimum to a high credit risk level. This dual approach not only promotes accurate and detailed credit risk between regional subenterprises but significantly improves the accuracy and depth of risk assessment by providing a more detailed and timely understanding of the credit situation. Through the integration of CSA and k-means classes, this approach provides a comprehensive framework for careful assessment and separation of credit risks at different levels, ensuring a reliable and reliable assessment process [6].

This paper investigates a large number of manufacturing enterprises, identifies and analyzes the correlation of credit risks of different enterprises, manages risks in an orderly manner, and formulates strategies. This strategy addresses various potential risks of the enterprise. The Russian academy of

ISSN 2616-5902 Vol. 6, Issue 11: 214-218, DOI: 10.25236/AJBM.2024.061131

natural resources warns against the transformation of the multi-stage assessment model into a tool needed by financial institutions and investors. This paper provides detailed risk descriptions to help improve risk management practices, as well as support reasonable, scientific process care, risk management strategies accurate and enterprise risk real conditions, using transfer framework analysis to optimize financial norms and strategists' overall investment effectiveness.

3.2 Application of Comprehensive Scoring and Multi-Level Risk Classification in Financial Credit Risk Assessment of Manufacturing Enterprises

In the financial complex, credit estimates of risk, where the impact of different credit on fashion predictions shows significant differences, indicating that each risk outcome is more affected. This survey examines the relative importance of different risks to business composition plans in the processing industry. In doing so, the machine trains algorithms designed to capture user information, providing mutually beneficial, scientifically acceptable support. The algorithm is convenient to predict and contribute to the risk of various social indicators through its complex analysis ability. This survey adjusts the assessment to accurately assess risk, provides information and integrity, and is functional. This survey develops smarter and more effective decision-making processes for financial risk management.

This study used several complex algorithms of machine learning, including random forests (rf), xgboost, and boost, each of which has unique advantages and a specific contribution to improving predicted accuracy and common models. Random forests significantly increase the predicted stability and accuracy through the polymerization of several trees of decision-making, competently handling the complex non-linear relationships inherent in high dimensional data collectors and capturing them. Xgboost is known for its effective gradient efficiency enhancement methods, with sophisticated weighting techniques for large data sets, improved model predictions, and enhanced awareness and performance optimization. Catbust has been successful in managing classification characteristics, effectively reducing retraining and improving learning, through advanced gradient enhancement techniques designed specifically for data classification. The use of these algorithms has created a strong, multifaceted framework for addressing various aspects of credit risk assessment, providing more accurate, rapid, and insightful stakeholder analysis.

Through the extensive and careful application of these advanced machine learning algorithms, this study systematically assesses the importance of each credit index, revealing its relative contribution to the model's projected results, and promoting a comprehensive understanding of its role. This in-depth Uber analysis is a key indicator of effective credit risk estimation in itself, and it shows the meticulousness of the corporate risk profile poured into various risks. Risk protectors have shown that they can show significant predictability in a recognized risk tray, but its impact can be significantly diminished when pushed into low-risk entities. By providing a detailed and detailed perspective on the functionality and importance of each credit indicator, such a rigorous review will allow for a more accurate and accurate assessment of their actual effectiveness in risk forecasts, improve the overall accuracy of credit risk assessment, and facilitate a wiser strategic financial risk management decision-making process.

The salient importance analysis presented in this article provides valuable empirical experience that greatly improves the accuracy of investment strategies and risk management protocols that carefully recognize and prioritize the most important credit indicators in risk forecasting. Through this approach, financial institutions and investors can develop a more scientific, rigorous, and effective risk management strategy, optimize risk early warning systems, and improve credit risk assessment. This approach has contributed to more strategic and wise financial decisions by providing important ideas, improving the accuracy of credit risk assessments, and providing important recommendations for future research and practical applications. The findings of this analysis have contributed significantly to the gradual improvement of risk management and financial decision support systems so that stakeholders can better align their practices with empirical evidence and enhance the stability and efficiency of financial business.

3.3 Application of Machine Learning Model Selection and Optimization in Financial Credit Risk Assessment of Manufacturing Enterprises

In the area of corporate credit risk warning, the main models usually fall into two main categories: traditional statistical methods and advanced machine learning methods. Traditional statistical models,

ISSN 2616-5902 Vol. 6, Issue 11: 214-218, DOI: 10.25236/AJBM.2024.061131

including linear discrimination analysis (LDA), probability regression and logic analysis, have historically provided valuable insights and analysis under certain conditions; However, when applied to complex and non-linear data structures that are common in today's business environment, they often face significant limitations. These problems arise from assumptions inherent in traditional statistical technologies that may not fully solve the variety and complexity of scenes in the real world. Such models are often unable to accurately capture the subtle model represented by the concentration of complex data, undermining the effectiveness of credit risk projections in an evolving and multidimensional business environment. Traditional statistical methods make a valuable contribution to history, but their usefulness in the context of modern risk assessment needs to be critically assessed by the increasingly complex and dynamic characteristics of the modern data environment.

Traditional statistical methods are commonly used, relying on strict assumptions about the rights allocation of information providers. Machines trained technologist as a neighbour to foreign countries (KNN), supporting vector machines (SVM) and trees solved, in the wide and complex set of Dan, in particular, distributed in front of persons uneven or non-linear Dan structures, gave the ability to recognize and use the prices of fashion transfers." For example, KNN calculates the distance between data points to enhance classification and reintegration, making them particularly suitable for data sets that do not comply with strict distribution rules. The support of the vector machine in solving classification problems by accurately defining the optimal hyperplane in the high-dimensional characteristics space has demonstrated extraordinary flexibility in the use of various nuclear functions to solve non-linear relationships. Integrated learning methods, such as random forests (rf), xgboost, and light gum, improve model accuracy and reliability by effectively managing data complexity and variability by combining forecasts of different tree decisions. Training technologies for these machines not only provide strong alternatives to traditional approaches but also contribute to a more thorough and flexible analysis of complex data structures, contributing to a better understanding of potential models and relationships.

In this study, comprehensive machine learning algorithms (KNN), linear support vectors (SVM), radial SVM function, random forest (rf), xgboost, light gum, and MLP are used to develop and evaluate credit risk early warning models, each algorithm has a unique advantage: KNN USES the proximity of a neighboring sample to inform forecasts; The linear support vector improves sorting accuracy by optimizing the decision boundaries; The radial fundamental function supports a vector that strengthens model adaptation to non-linear data; Rf improves predictability by selecting trees for decision-making; Xgboost USES gradient boost to maximize the productivity of large data sets; Lightgbm provides efficiency and scalability through the gradient frame; MLP USES its multi-layer neural structure to recognize complex patterns in the data. This diverse approach is aimed at strictly assessing the performance of algorithms in managing isometric data, determining their actual effectiveness, and creating a more reliable and accurate structure for assessing corporate credit risks. With a thorough analysis of the advantages and limits inherent in each algorithm, this study would like to improve the accuracy and reliability of credit risk assessment practices, providing stronger and more discerning tools for wise financial decisions.

4. Conclusion

Production companies play an important role in the national economy, which requires a complex financial credit risk assessment system that could be addressed through several advanced analytical techniques, including core components (CSAs), k-medium term test sites, and tropical laws that provide a strong basis for risk assessment. To address data imbalances, synthetic minority over-sampling technology (smote) has been introduced to provide a more representative and balanced set of data for risk assessment. Subsequent analysis showed that the industry's solvency had the most significant impact on credit risk, and corporate performance was variable depending on the specific level of risk it faced. With the help of various stress tests, the MLP model has been recognized as the most effective, demonstrating considerable stability and elasticity, especially in the production of general equipment. Based on these findings, this article proposes the development of a differentiated risk management strategy aimed at a specific risk profile that would support the use of machine learning techniques in financial risk management and recommend a thorough analysis of stress test scenarios to improve and improve the effectiveness of risk assessment methods. Such a comprehensive approach seeks to better understand credit risks and strengthen the strategic management of financial uncertainty.

ISSN 2616-5902 Vol. 6, Issue 11: 214-218, DOI: 10.25236/AJBM.2024.061131

References

[1] Wang F, Hou Z, Juan R, et al. Research on Influencing Factors and Countermeasures of Risk Control of State-Owned Enterprises under the Background of Big Data [J]. Proceedings of Business and Economic Studies, 2023, 6(6): 128-133.

[2] Peng C, Da F, Wu J, et al. China's local government debt risk assessment and countermeasures under the influence of COVID-19[J]. Procedia computer science, 2022, 199: 354-360. DOI:10.1016/j. procs. 2022.01.043.

[3] Fan T R, Liu M S, Tong K Z, et al. Research of interoperation model for heterogeneous database based on Internet[C]. International Conference on Machine Learning & Cybernetics, 2003. DOI:10. 1109/ ICMLC.2003.1259838.

[4] Kamm S, Veekati S S, Mueller T,et al. A survey on machine learning based analysis of heterogeneous data in industrial automation[J]. Computers in Industry, 2023.

[5] Chen H, Ma K, & Shen J. Interpretable Machine Learning Facilitates Disease Prognosis: Applications on COVID-19 and Onward[J]. International Journal of Computer Science and Information Technology, 2024, 3(3), 428-436.

[6] Chen H, Varatharajah Y, De Ramirez S S, Arnold P, Frankenberger C, Hota B, & Iyer R. A retrospective longitudinal study of COVID-19 as seen by a large urban hospital in Chicago[J]. MedRxiv, 2020, 11.