Information Resources Management Journal
Volume 39 • Issue 1 • January-December 2026

Credit Scoring Prediction for Small and Medium-Sized Enterprises Based on Alternative Data and Gradient Boosting Algorithms

Shu Chen
https://orcid.org/0009-0002-6189-6288
Zhengzhou SIAS University, China
Received: November 9th, 2025 | **Accepted:** January 4th, 2026

ABSTRACT

In today's complex economic landscape, small and medium-sized enterprises (SMEs) are crucial drivers of growth, yet traditional credit scoring models often fail to capture their true creditworthiness because they are limited by narrow data sources and poor data adaptability. With the rise of big data and fintech, alternative data opens a richer avenue for SME credit assessment. This study leverages real-world, publicly available data, which includes operational behavior, supply chain interactions, and online transactions, to help build a more inclusive and forward-looking credit scoring framework for SMEs. The authors enhance the model's nonlinear fitting and feature representation capabilities by employing gradient boosting algorithms to significantly improve credit risk prediction accuracy. They compare the performance of various machine learning models and discuss trade-offs between predictive power, generalizability, and interpretability. The results offer financial institutions a dynamic, multidimensional risk assessment tool able to provide actionable insights for policy and practice.

KEYWORDS

Alternative Data, Gradient Boosting, Small and Medium-Sized Enterprises (SMEs), Credit Scoring, Prediction

INTRODUCTION

As the most dynamic micro-subject in China's economic system, small and medium-sized enterprises (SMEs) play an irreplaceable strategic role in promoting market innovation, absorbing employment, optimizing industrial structure and promoting regional coordinated development (Cunningham, 2011; Ma et al., 2023). However, the long-term dilemma of "difficult and costly financing" is still the key bottleneck restricting its survival and development (Bakhtiari et al., 2020; Naradda Gamage et al., 2020). Due to its inherent limitations, such as small scale, weak financial system, and insufficient information disclosure, it is difficult for financial institutions to accurately evaluate the credit status of SMEs through traditional methods (Hossain, 2023). From the perspective of financial institutions, providing credit to information-deficient SMEs will bring significant risk exposure (DeYoung et al., 2011). Without reliable and timely information support, lending institutions will face higher uncertainty in terms of enterprise default probability prediction, risk pricing, and

DOI: 10.4018/IRMJ.400759

portfolio stability control, which may lead to an increase in non-performing loan ratio, thus aggravating the pressure on capital adequacy ratio. Therefore, some SMEs are excluded from formal financial services, which not only reflects the inefficiency of the supply side, but also reflects the prudent risk avoidance behavior adopted by financial institutions under the constraints of supervision and market. As a result, many SMEs with high growth potential cannot obtain formal financial services because of "credit barriers," resulting in a structural mismatch between the supply of financial resources and the real demand of the real economy.

The traditional credit scoring system mainly relies on structured data, such as financial statements and central bank credit records, are constructed by classical algorithms such as logistic regression and linear discriminant analysis (Abdou & Pointon, 2011). These methods have certain effects on large enterprises, but they face obvious limitations in the context of SMEs. First, SME financial data are often lagging, incomplete, or even completely missing, especially for start-ups, which leads to a serious information gap in model input (Ciampi et al., 2021). Second, the linear model is difficult to capture the inherent nonlinear dynamic characteristics of SMEs and the complex risk association, so it cannot fully reflect their real operating conditions and potential credit risks. At present, with the intensification of macroeconomic fluctuations, the rapid rise of SMEs in the new economy, and the subversive impact of financial technology on the traditional risk management model, the relevance and effectiveness of the traditional credit scoring system are further weakened, and it is urgent to achieve innovation and upgrading through technological innovation and data dimension expansion. This transformation is essentially driven by the research and practice in the field of information systems (IS).

In this context, rise of alternative data provides a revolutionary path to alleviate the information asymmetry in SME credit evaluation (Djeundje et al., 2021; Hlongwane et al., 2024; Óskarsdóttir et al., 2019), and IS is the basic infrastructure to realize such data collection, integration, and value extraction. Different from traditional static data, alternative data comes from all kinds of real-time scenes in the whole life cycle of enterprises, covering online transaction records, supply chain interaction information, social media dynamics and digital operation indicators, etc. These dynamic signals can reflect the business vitality, market reputation and risk resilience of SMEs in time, effectively making up for the shortcomings of traditional data coverage. At the same time, the advancement of machine learning technology makes it possible to deeply explore the value of such data, among which the gradient boosting algorithm (Bentéjac et al., 2021) stands out with its excellent nonlinear modeling ability, automatic feature selection mechanism and good adaptability to heterogeneous data. The algorithm can minimize residual error by iteratively combining weak decision trees, accurately capture the hidden risk patterns in alternative data, and significantly improve the accuracy and stability of credit risk prediction (Chang et al., 2018).

The latest research at home and abroad is gradually exploring the integrated application of alternative data and machine learning. Some studies optimize the credit scoring model by introducing transaction features or social features, while others improve the model performance in small sample scenarios by integrating methods (Putra et al., 2020). However, key challenges remain, including the lack of standardized information system framework for cross-organizational data governance for SME credit evaluation, insufficient connection between algorithm model and existing information system workflow of financial institutions, and limited research on embedding ethical protection mechanism into information system-driven credit scoring system.

To make up for the above gap, this study is based on the actual risk management needs of financial institutions and the theoretical lens of IS-enabled financial innovation. Using real and publicly available data, we systematically evaluate the model from the dimensions of risk differentiation ability, generalization performance and practical feasibility, aiming at providing more inclusive and forward-looking credit evaluation solutions for SMEs. At the same time, the findings of this study also provide solid theoretical insights and empirical support for policy makers who are committed to promoting the development of inclusive finance and optimizing the financial service ecology of SMEs.