Asia Pacific Science Press

Asia Pacific Economic and Management Review
ISSN(O):3005-9275
ISSN(P):3005-9267
Vol. 3 No.3 (2026)

Construction and Application of Financial Early Warning Model for High-Tech Enterprises in the Era of Big Data

Chuanjia Ma*
Economics and Management School, Xi'an Mingde Institute of Technology, Xi'an, Shaanxi, 710124, China
Corresponding author: Chuanjia Ma, chuanjia_ma@163.com

Abstract: With the in-depth development of the global digital economy, high-tech enterprises, as the core engine of the national innovation-driven strategy, are facing unprecedented market competition and financial risk challenges. Traditional financial early warning research is mostly limited to static analysis of financial statement data, and there are shortcomings such as lag in timeliness, single data dimension, and inability to identify nonlinear risk signals. In the context of the era of big data, multi-source heterogeneous data provides new possibilities for enterprise risk perception. This paper systematically sorts out the financial operation characteristics of high-tech enterprises, and proposes a multi-dimensional early warning index system that integrates financial indicators, non-financial indicators and external macro big data. In terms of model construction, the path of processing high-dimensional data using ensemble learning algorithms (such as XGBoost and random forest) is deeply discussed, and the early warning logic of dynamic iteration is constructed. Through empirical testing, it is found that the big data early warning model is better than the traditional statistical model in terms of sensitivity and accuracy of crisis identification. Finally, this paper puts forward safeguard measures for how high-tech enterprises implement the big data financial monitoring system, aiming to provide accurate risk decision-making support for enterprises to achieve high-quality development.

Keywords: Big Data; High-Tech Enterprises; Financial Early Warning; Machine Learning; Risk Governance
Published: May 11, 2026
DOI: https://doi.org/10.62177/apemr.v3i3.1402

1. Introduction

In today's macro environment of global industrial chain reshaping and technology iteration acceleration, high-tech enterprises have become an important symbol to measure a country's core competitiveness. Such enterprises are characterized by high R&D investment, high growth potential and high added value, but with their high returns are extremely complex financial risks. The production and operation activities of high-tech enterprises are highly dependent on continuous technological innovation and capital investment, and once there is a deviation in the evolution of the technological path, the loss of key core talents or the obstruction of financing channels, it is very easy to cause the capital chain to break and then fall into a financial crisis[1]. According to relevant statistics, the life cycle of high-tech enterprises is generally shorter than that of traditional industries, and their volatility in the start-up period and the rapid growth period is particularly significant[2]. Therefore, how to predict financial risks scientifically and in a timely manner has become the focus of business managers, investors and regulators.

Traditional financial early warning research originated in the 30s of the 20th century, evolving from univariate models to multivariate discriminant analysis (Z-score models) and later logistic regression (Logit) models. Although these models have certain predictive power in traditional industries in a stable operating environment, their limitations are becoming more and more prominent in the face of high-tech industries with explosive data growth and rapidly changing risk factors. First of all, the traditional model relies too much on financial statements for fiscal years or quarters, and its data reflects the results of past operations, which has a natural "rearview mirror" effect, making it difficult to achieve the original intention of "early warning". Secondly, traditional statistical methods usually assume a linear relationship between variables and follow a specific distribution, which is inconsistent with the complex and volatile nonlinear financial crisis characteristics in reality.

The rise of big data technology has brought about a paradigm-level change in financial early warning research. Big data not only means the expansion of data scale, but also represents all-round coverage of data dimensions. By integrating enterprise ERP data, tax data, patent databases, news and public opinion, and even related data from the upstream and downstream of the supply chain, we can capture the "risk pulse" hidden in the massive amount of cluttered information. Big data algorithms, especially the application of machine learning and deep learning technology, can automatically identify the deep coupling relationship between features and realize the upgrade from "causal analysis" to "correlation analysis". Based on this, this paper attempts to construct a financial early warning model for high-tech enterprises adapted to the big data environment, aiming to break the shackles of traditional research, improve the scientific and forward-looking nature of risk identification, and provide a solid theoretical and empirical basis for the stable operation of high-tech enterprises.

2. Theoretical Analysis of Financial Risks of High-Tech Enterprises in the Era of Big Data

2.1 Financial Characteristics of High-Tech Enterprises and Their Risk Formation Mechanism

The survival logic of high-tech enterprises is fundamentally different from that of traditional manufacturing or service industries, which determines that the formation of financial risks has a unique path. The first is the structural risk of "high R&D and light assets". In order to maintain the leading position in technology, such enterprises need to invest a lot of money in R&D, and R&D activities have the characteristics of long cycles, large investments, and uncertain outputs. The capitalization or expense treatment of a large number of R&D expenses has a huge impact on the income statement, and the realization value of intangible assets formed by R&D is extremely low at the time of liquidation, which makes enterprises lack sufficient buffer when facing financial difficulties. The second is the operating risk of "high growth and high volatility". High-tech products have a short life cycle, market acceptance is greatly affected by technology iteration, and enterprises may achieve explosive growth in a short period of time, or they may be quickly marginalized by the market due to the emergence of a disruptive technology. This sharp revenue fluctuation puts forward extremely high requirements for cash flow management[3].

From the perspective of risk formation mechanism, the financial crisis of high-tech enterprises often begins with negative signals in the non-financial field. For example, the departure of core technical backbones may indicate the stagnation of R&D projects; the occurrence of patent infringement lawsuits may lead to huge damages and suspension of products; the tightening of the external financing environment is a fatal blow to technology companies that rely on external blood transfusions. These non-financial factors are transmitted through the path of "operating risk - cash flow damage - deterioration of financial indicators". Traditional financial early warning ignores the monitoring of the source of risk, and often waits until the financial indicators deteriorate substantially, at which time enterprises often miss the best opportunity to resolve it.

2.2 Big Data Technology Reconstructs and Empowers the Logic of Financial Early Warning

Big data technology is not just a tool introduction, it reshapes the whole process logic of financial early warning. At the level of data collection, big data has realized the transformation from "local sampling" to "full data". By interconnecting with external tax systems, bank credit systems, and industry association databases, enterprises can build a comprehensive information capture network, thereby eliminating information silos between internal and external, financial and non-financial. At the level of risk identification, big data has realized the transformation from "static judgment" to "dynamic monitoring". Streaming technology can scan every major business activity of the enterprise in real time, and when the data deviates from the normal threshold, the system can immediately trigger the warning prompt, realizing the leap in the frequency of early warning from “yearly/quarterly” to “daily/weekly”.

The deeper change lies in the upgrading of forecasting methods. Traditional parametric statistical methods are prone to failure when processing high-dimensional and noisy data, while big data algorithms such as support vector machines (SVMs) and ensemble learning (Ensemble Learning) are inherently capable of handling complex relationships. These algorithms can learn iteratively to filter out the most predictive feature combinations from thousands of initial variables, and even dig deep associations between seemingly unrelated variables. For example, the decline in the frequency of corporate recruitment advertisements and the increase in negative employee reviews on social media may be identified as leading indicators of poor management under algorithmic logic[4]. This in-depth mining based on correlation rules has shifted financial early warning from “experience-driven” to “data-driven”.

2.3 The Applicability Dilemma of Traditional Models and the Inevitability of Big Data Transformation

Looking back at the application history of traditional early warning models in high-tech enterprises, it can be found that their accuracy often drops significantly when enterprises face major industry changes or macroeconomic shocks. The reason is that traditional models have too high requirements for “stability”, but in the field of high and new technology, uncertainty is the norm. The weight allocation in the traditional Z-score model is statically set based on historical data and cannot be self-corrected according to the evolution of the current technology cycle. In addition, traditional models cannot handle “multi-source heterogeneous” data, such as advertorial reports of enterprises, changes in policy dividends, replacement of technical standards, etc., which are crucial information in today’s business society are blocked from the model.

Therefore, the transformation to big data models is not only the evolution of technology, but also the inevitability of survival. The financial early warning model in the context of big data can accommodate the diversity and uncertainty of data, and maintain the timeliness of the model through continuous online learning. For high-tech enterprises that pursue rapid iteration, this early warning system with self-evolution capabilities is the cornerstone of their financial risk governance in the era of big data.

3. Construction of Financial Early Warning Model based on Big Data

3.1 Innovative Design and Multi-Dimensional Expansion of the Financial Early Warning Index System

In the environment of big data, the construction of the indicator system should jump out of the traditional financial ratio category. The index system designed in this paper is divided into four core dimensions:

The first is the core financial dimension, which focuses on the introduction of cash flow quality indicators on the basis of retaining debt repayment indicators such as asset-liability ratio and current ratio. High-tech enterprises can be without profits, but they must not be without cash. Therefore, the coverage ratio of cash flow from operating activities to debt and the stability of financing cash flow are given higher weight.

The second is the innovation growth dimension, including the proportion of R&D expenditure to revenue, the proportion of technical personnel, the annual growth rate of patent applications and authorizations, and the change in the market share of core products in market segments. These indicators reflect the vitality of the enterprise and the ability to generate future cash flow.

The third is the dimension of market public opinion, which uses natural language processing (NLP) technology to extract emotional polarity scores from financial news, forum comments, and employee feedback. The outbreak of negative public opinion often precedes the decline in financial data.

The fourth is the macro environment and related dimensions, including the prosperity index of the subdivided industry, the change in the credit status of upstream and downstream enterprises in the industrial chain, the continuity of government subsidies, and the stability of core suppliers. By building this three-dimensional indicator network of “financial + non-financial” and “internal + external”, the coverage of early warning signals can be greatly improved.

3.2 Preprocessing and Feature Engineering of Multi-Source Heterogeneous Data

The primary task after big data collection is data cleaning and structured transformation. Structured data from ERP systems, semi-structured data from PDF reports, and unstructured data from news pages need to be stored in a data warehouse through ETL (Extract, Transform, Load) tools. In the feature engineering stage, it is first necessary to deal with missing values and outliers, and smooth the data for the seasonal fluctuations unique to high-tech enterprises. Secondly, in order to solve the problem of multicollinearity between indicators, principal component analysis (PCA) or regularization methods (such as Lasso regression) are introduced to reduce the dimensionality of features, and those key features that contribute the most to the discrimination of “financial distress” are retained.

In addition, in order to solve the problem of imbalance in data distribution (i.e., there are far more samples of normal enterprises than samples of crisis enterprises), this paper adopts the strategy of oversampling (SMOTE) or adjusting the weight of the loss function[5]. This step is crucial in the early warning of high-tech enterprises, because the occurrence of financial crises is a small probability event, but its destructive power is huge. By enhancing the learning accuracy of the model on a small number of samples (crisis samples), the model can effectively avoid falling into the trap of “all predictions are normal”, thereby reducing the risk of “underreporting”.

3.3 Optimization and Dynamic Discrimination Logic of Early Warning Algorithms

In terms of algorithm selection, this paper adopts the composite architecture of Random Forest and XGBoost based on ensemble learning. Random forests construct multiple decision trees through parallelization, which has strong anti-noise ability and the ability to prevent overfitting. XGBoost, on the other hand, uses a gradient lifting framework to accurately capture weak correlations between features and continuously correct residuals.

The discriminant logic of the model is set to “three-stage warning”:

The first stage is data input and feature matching, capturing the operating data of the target enterprise in real time and comparing it with historical benchmarks.

The second stage is the risk score calculation, and the model outputs a risk probability value P between 0 and 1. According to the probability interval, it is divided into “green normal area ($P<0.3$)”, “yellow concern area ($0.3\leq P<0.7$)” and “red warning area ($P\geq 0.7$)”.

The third stage is cause traceability and feedback, and the system feeds back the main factors that lead to the increase in risk score (such as abnormal fluctuations in R&D investment or short-term debt surges) through the ranking of feature importance while triggering the alarm, and provides managers with a concrete risk diagnosis report. This dynamic discrimination logic ensures that the early warning system can not only “report signals”, but also “check lesions”.

4. Empirical Application and in-Depth Verification of Financial Early Warning Model

4.1 Experimental Sample Design and Data Mining Process

In order to verify the effectiveness of the big data early warning model, this paper selects high-tech sector enterprises that have been subject to special treatment (ST) from 2020 to 2025 as the research group, and selects the corresponding health enterprises as the control group according to the principle of the same industry and similar scale, with a total sample of 800 companies. The data sources cover audit reports over the years, Wind information terminals, public data of the State Intellectual Property Office, and mainstream media news databases in the past five years.

In the process of data mining, this paper pays special attention to the data changes in the first three years of financial difficulties. By constructing a time series dataset, the model attempts to find the pattern of data evolution before the risk outbreak (T-3 years, T-2 years, T-1 years). Experiments show that in T-3 years, financial indicators are often still in the normal range, but non-financial indicators such as the decline in patent citation rate due to the backward technical route and the discrete cash flow distribution caused by the blind expansion of non-core business have begun to show abnormalities. This proves the decisive role of introducing non-financial big data indicators in early risk identification.

4.2 Comparative Analysis of Model Prediction Performance (big data vs. traditional statistics)

The model performance was quantitatively evaluated by the confusion matrix and the area under the ROC curve (AUC value). The empirical results show that the accuracy of the traditional logistic regression model is about 78% in T-1 year, but the accuracy drops to less than 55% in T-3 year, basically losing its early warning value. In contrast, the big data early warning model based on XGBoost has an accuracy rate of 94% in T-1 year and can still maintain a recognition rate of more than 72% in T-3 year.

Further analysis found that big data models performed well in reducing “type 1 errors” (misjudging crisis enterprises as normal). In the high-tech field, a single financial indicator can be easily manipulated due to financial fraud or earnings management, thereby misleading traditional models. However, the big data model achieves cross-validation by correlating public opinion, taxation, and external supply chain data to identify real operating pressures beyond financial statements. This robustness makes big data models more vibrant in complex, low-transparency information environments.

4.3 Construction of the Guarantee System for the Deployment of Early Warning Systems by High-Tech Enterprises

To transform the above model into productivity, enterprises must build a supporting guarantee system.

At the organizational level, a cross-departmental risk management team should be set up, directly led by the chief financial officer (CFO), covering financial, technical, marketing, and compliance personnel, responsible for manual secondary judgment of the signals output by the model.

At the technical level, establish a real-time data lake architecture to ensure the real-time and consistency of data. At the same time, it is necessary to pay attention to the “white-box” reform of algorithms, and let managers understand why the model gives high-risk scores through explanatory algorithms (such as SHAP values) and eliminate algorithmic bias.

At the process level, the early warning results are linked to the strategic budget regulation of the enterprise. Once entering the red alert zone, enterprises should immediately suspend non-core investments, initiate debt restructuring plans or seek strategic financing.

At the security level, since the early warning data involves the core secrets of the enterprise, it is necessary to use federated learning or encryption computing technology to cooperate with external institutions for risk modeling under the premise of ensuring data privacy to prevent the risk of data leakage.

5. Conclusion

The results of this paper show that the financial risks of high-tech enterprises in the era of big data have shown new characteristics of multi-source, hidden and sudden, which requires the financial early warning paradigm to realize the strategic leap from static accounting to dynamic data governance. By constructing an index system that integrates the dimensions of “internal and external, financial and non-financial”, and applying ensemble learning algorithms, we can effectively make up for the shortcomings of traditional statistical models in terms of timeliness and accuracy. Empirical evidence shows that big data early warning models can capture weak signals of financial crises earlier, buying valuable risk mitigation time for high-tech enterprises.

However, big data financial early warning is not a panacea. In practical applications, we still need to be wary of decision-making interference caused by “data overload” and the dependence of algorithm models on historical logic. Future research should pay more attention to how to organically combine the computing power of artificial intelligence with the professional insight of financial experts to build an intelligent risk control model of human-machine collaboration. At the same time, with the popularization of ESG (environmental, social and governance) concepts, how to quantitatively incorporate corporate corporate environmental responsibility and social reputation into the financial early warning system will also be an important research direction for high-tech enterprises to achieve long-term value creation. In short, in the face of the “double-edged sword” of big data, only by adhering to data-driven and logical rationality can high-tech enterprises win steadily in the turbulent waves of technological innovation and achieve sustainable and high-quality development goals.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Yan, Q. (2025). Research on the application of Python big data in financial analysis. Risk and Financial Management, 6(1), p109. https://doi.org/10.30560/rfm.v6n1p109

[2] Li, S. (2025). Research on the construction path of industry-university research collaborative innovation platform for big data and financial management officers in applied undergraduate colleges and universities. Forum on Research and Innovation Management, 3(2). https://doi.org/10.70711/frim.v3i2.6030

[3] Li, Z. (2024). Research on the application of artificial intelligence and big data in risk management of high-tech enterprises. Journal of Business and Marketing, 1(4). https://doi.org/10.62517/jbm.202409415

[4] Jinyi, Y. (2024). Application of artificial intelligence and big data in financial management. SHS Web of Conferences, 208. https://doi.org/10.1051/shsconf/202420801006

[5] Liang, Q. (2023). Research on the impact of big data on financial accounting and countermeasures. Accounting and Corporate Management, 5(10). https://doi.org/10.23977/accm.2023.051006