# The Quality Of Financial Reports And Decision Support Driven By Big Data Analysis

Binbin Hu

*Big Data Accounting Institute, Henan Industry and Trade Vocational College, Zhengzhou 450000, China*

*Corresponding author. E-mail: hubinbin@hngm.edu.cn*

*Received: March 22, 2026; Accepted: July 23, 2026*

Against the backdrop of global digital transformation of corporate finance, the mismatch between traditional financial reporting systems and complex dynamic business environments has become increasingly prominent, with problems such as financial information distortion, weak risk early warning capabilities, and the disconnection between financial information quality and decision value creation becoming key bottlenecks restricting high-quality development of enterprises.With the deep development of the digital economy, the digital transformation of enterprise financial activities is accelerating. How to rely on big data analysis technology to solve the industry pain points of distorted financial report information and insufficient decision support capabilities has become a core issue in the field of enterprise financial management and corporate governance. This study is based on the theory of information asymmetry and the financial quality evaluation system, and constructs a four-dimensional quantitative model for financial report quality, depicting the evolutionary relationship between the authenticity, compliance, and value relevance of financial information; Subsequently, an improved adaptive gradient boosting decision tree (AGBDT) algorithm was introduced to construct a big data analysis driven financial report quality optimization and decision support framework. The model integrates historical financial information weighting mechanism, multi-objective decision return function, and adaptive learning rate optimization strategy, achieving joint optimization of financial report quality improvement, decision risk control, and enterprise value growth. The empirical results show that the proposed model achieved the lowest average absolute error of 0.21 in financial quality identification and the highest accuracy of 96.3% in financial fraud identification on the Shanghai and Shenzhen A-share listed company datasets and the New Third Board enterprise datasets. The effectiveness of decision support was improved by 32.7% compared to traditional models, and the overall performance was significantly better than the four comparative models. In extreme market volatility and industry heterogeneity scenarios, this can still control the deviation of financial report distortion warning within 8.5%, while achieving optimization of business decision-making under low risk. This study provides an efficient and scalable technological path for the construction of digital financial management systems in enterprises, which is applicable to various scenarios such as information disclosure management of listed companies, financial control of group enterprises, and value analysis of investment institutions.

**Keywords:** Big data analysis; Financial report quality; Financial information quality; Decision support; Adaptive gradient boosting decision tree; Financial risk early warning



http://dx.doi.org/10.6180/jase.202610_33.021

## 1. Introduction

The digital economy and financial digital transformation have exposed key pain points in traditional reporting systems: high information distortion risk, lagging decision support, and poor adaptability. These manual, periodic systems have three drawbacks—post-event compliance focus, inability to detect implicit fraud under complex transactions, and no closed-loop from quality information to decision value—leading to decoupling of finance and operations.

Big data technology offers a new path to address financial reporting challenges. Scholars have explored its application from multiple angles [1]. Al Okaily and Al Okaily [2] identified key factors influencing big-data-driven financial decision quality but did not examine the optimization logic of financial report quality as a decision input. Zhang [3] designed a big-data-driven decision support system but lacked a quantifiable, traceable, and iterative optimization mechanism for financial report quality. Saleh et al. [4] confirmed a positive correlation between big data analytics and financial report quality using Canadian market data but did not construct a quantitative optimization model or decision-value conversion path. Tong and Tian [5] built an intelligent financial decision support system but separated quality improvement from decision optimization, lacking a closed-loop framework. Megeid and Sobhy [6] clarified the chain effect of big data on reporting, decision-making, and performance but did not develop a quantitative optimization model adaptable to dynamic business environments. Elgendy et al. [7] proposed the DECAS data-driven decision theory but lacked adaptation for financial scenarios. Ren [8] validated big data's effectiveness in financial management but did not explore the decision-value conversion mechanism of financial report quality. Ikegwu et al. [9] reviewed data sources, tools, and applications but provided no targeted design for financial scenarios. Lazaroiu et al. [10] validated algorithmic empowerment in smart manufacturing but did not adapt it to core financial management. Zhang et al. [11] supported big-data-assisted social media analysis for business decisions but excluded core financial report quality optimization. Chatterjee et al. [12] evaluated big data's positive impact on decision processes and performance but did not examine the mediating role of financial report quality or construct scenario-adaptive models.

Overall, existing research confirms big data's value in finance but has three core gaps: disconnect between financial report quality optimization and decision support, lacking a fully closed-loop framework; most studies remain at correlation verification or framework design, lacking quantifiable, iterative optimization models; insufficient adaptability research for complex scenarios such as industry heterogeneity and extreme market volatility, with model robustness needing improvement.

To address the identified gaps, this study proposes an integrated big-data-driven framework for financial report quality optimization and decision support, with three core incremental contributions:

1. Theoretical innovation: A closed-loop framework integrating "quantitative evaluation – dynamic optimization – decision value conversion," revealing the transmission mechanism between financial report quality and decision value. Unlike Zhang (2025) and Ren (2022), this framework bridges quality optimization and decision support.

2. Methodological innovation: An improved AGBDT algorithm featuring three adaptive mechanisms (historical information weighting, multi-objective decision benefit function, adaptive learning rate). It overcomes the limitations of traditional GBDT/XGBoost (e.g., single-objective optimization, ignoring temporal continuity, poor generalization), achieving joint optimization of identification accuracy, decision risk control, and enterprise value growth.

3. Practical innovation: Robustness validation under industry heterogeneity, extreme market volatility, and cross-cycle scenarios, forming a scalable technical path for listed companies, group enterprises, and investment institutions—addsing the lack of adaptability analysis in complex financial scenarios.

The remainder is organized as follows: Section 2 constructs the quantitative model and decision support framework; Section 3 presents empirical tests and analysis; Section 4 concludes and outlines future research directions.

## 2. Quantification of financial reporting quality and construction of decision support model

### 2.1. Multidimensional Quantitative Model for Financial Reporting Quality

The core pain point of corporate financial reporting is information asymmetry causing quality distortion and reduced decision value. Financial report quality is defined as the ability to reflect true operations, comply with standards, support decisions, and ensure timeliness. This study establishes a four-dimensional quantification system—authenticity, compliance, value relevance, timeliness—covering balance sheets, income statements, cash flow statements, footnotes, and market data [13].

Indicators for each dimension are listed in Table A1, with data from CSMAR, Wind, and corporate announcements.

A hierarchical weighted network (Fig. 1) is constructed: bottom (raw data), middle (dimension features), top (quality quantification & risk warning), linked by weights to trace the full logic from raw data to quality evaluation [14].

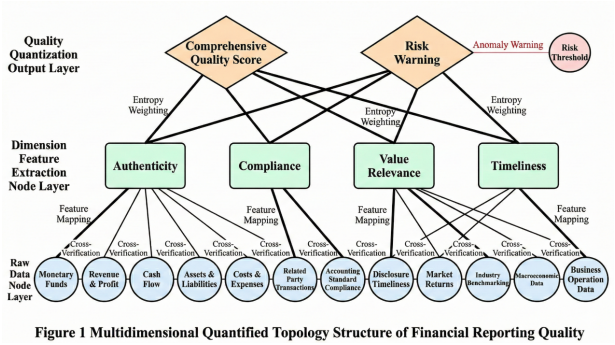


**Fig. 1.** Quantitative structure of multidimensional financial report quality.

This structure consists of three core node layers. The bottom layer includes twelve raw data nodes covering monetary funds, revenue and profit, cash flow, assets and liabilities, costs and expenses, related party transactions, compliance, disclosure timeliness, market returns, industry benchmarks, macroeconomics, and business operations, cross-validated through data verification links. The middle layer includes four dimension nodes for authenticity, compliance, value relevance, and timeliness, each connected to corresponding bottom nodes for standardized feature extraction. The top layer includes comprehensive quality score and risk warning nodes, using an entropy weighting algorithm for feature fusion and connecting to a risk threshold node for real-time abnormal signal alerts [15].

Based on the above structure, this study uses information entropy theory to objectively assign dimension weights and quantitatively calculate quality scores. The core formula is as follows:

Firstly, the original features are subjected to min max normalization to eliminate dimensional differences and achieve additivity of different dimensional indicators. The formula is as follows:

$$x_{i,t}^{k} = \frac{x_{i,t}^{k,raw} - \min\left(x^{k}\right)}{\max\left(x^{k}\right) - \min\left(x^{k}\right)} \quad k = 1,2,3,4 \tag{1}$$

Among them, $x_{i,t}^{k}$ are the standardized characteristic values of the i -th enterprise in the k -th dimension of period t, and xi, tk, raw are the original indicator values, k = 1, 2, 3, 4 correspond to the four dimensions of authenticity, compliance, value relevance, and timeliness, respectively.

Based on the theory of information entropy, calculate the discriminability and objective weights of each dimension. First, calculate the information entropy value of the k -th dimension:

$$e_k = -\frac{1}{\ln(n)} \sum_{i=1}^{n} p_{i,k} \ln\left(p_{i,k}\right) \tag{2}$$

Among them, $p_{i,k} = \frac{x_{i,t}^{k}}{\sum_{i=1}^{n} x_{i,t}^{k}}$ is the proportion of characteristic values of the i -th enterprise in the k -dimension, n is the number of sample enterprises, and $e_k$ ranges from 01. The smaller the entropy value, the stronger the ability of this dimension to distinguish financial quality.

Calculate objective weights for each dimension based on entropy value and achieve adaptive weight allocation:

$$w_k = \frac{1 - e_k}{\sum_{k=1}^{4} (1 - e_k)} \tag{3}$$

Among them, $w_k$ is the weight coefficient of the kth dimension, satisfying $\sum_{k=1}^{4} w_k = 1$, and the weight size is positively correlated with the information discrimination of the dimension.

Based on dimension weights and standardized features, calculate the comprehensive score of enterprise financial report quality:

$$FQ_{i,t} = \sum_{k=1}^{4} w_k \cdot x_{i,t}^{k} \tag{4}$$

Among them, $FQ_{i,t}$ is the comprehensive score of the financial report quality of the i -th enterprise in period t , with a value range of $[0, 1]$. The higher the score, the better the financial report quality.

Due to the temporal inertia of financial report quality-where current quality is continuously shaped by historical operations and financial information-this study constructs a temporal evolution model to capture its dynamic changes, as illustrated in Fig. 2.

This structure is a chain dynamic network structure, with time series as the core axis, including four time series nodes t-2, t-1, t, t+1. Each time series node includes a current feature input sub network, a quality calculation sub network, and a risk output sub network; The transmission of historical information across temporal nodes is achieved through inertia coefficient links, while introducing random perturbation nodes to characterize the impact of external shocks such as market volatility and policy changes; Finally, by fitting the time series link, the prediction of future financial quality and risk pre warning can be achieved [16].

The corresponding core formula for temporal evolution is as follows:

$$FQ_{i,t} = \alpha \cdot FQ_{i,t-1} + \beta \cdot X_{i,t} + \varepsilon_{i,t} \tag{5}$$

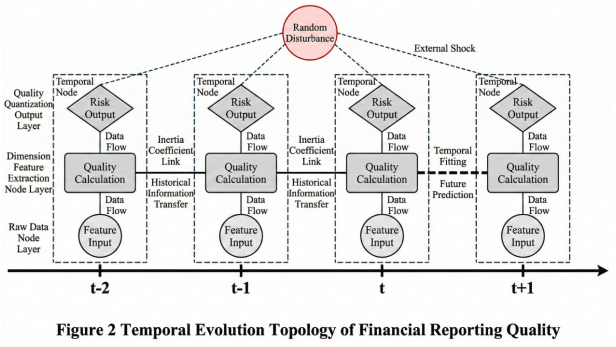


**Figure 2 Temporal Evolution Topology of Financial Reporting Quality**

**Fig. 2.** Time series evolution structure of financial report quality.

Among them, $\alpha$ is the inertia coefficient of historical financial quality, reflecting the degree of influence of historical information on current quality; $\beta$ is the influence coefficient matrix of the current feature vector; $X_{i,t}$ is the current four-dimensional eigenvector; $\varepsilon_{i,t}$ is a random perturbation term that characterizes the quality fluctuations caused by external environmental shocks.

For the estimation and test of the inertia coefficient $\alpha$:

1. Estimation method: We use a two-way fixed-effect panel regression model to estimate $\alpha$, controlling for individual fixed effects of enterprises and time fixed effects to eliminate bias caused by unobserved individual heterogeneity and time-varying macro shocks.

2. Significance test: We use the t-test to judge the statistical significance of $\alpha$. The baseline regression results show that the estimated value of $\alpha$ is 0.72, which is significant at the 1% level, verifying that the quality of corporate financial reports has significant positive temporal inertia.

3. Robustness test: We replace the explained variable with the single-dimension financial quality score for regression, and use the lagged two-period FQ as the explanatory variable for regression. The results show that $\alpha$ is still significantly positive at the 1% level, verifying the robustness of the estimation results.

### 2.2. Financial Decision Support Framework Driven by Big Data

Traditional financial analysis models focus on post-evaluation and cannot achieve a closed-loop transition from quality optimization to decision support [17]. Therefore, this study introduces an improved AGBDT algorithm to build a big-data-driven financial decision support framework, integrating three core optimization mechanisms for dynamic quality optimization and adaptive decision output, as shown in Fig. 3.

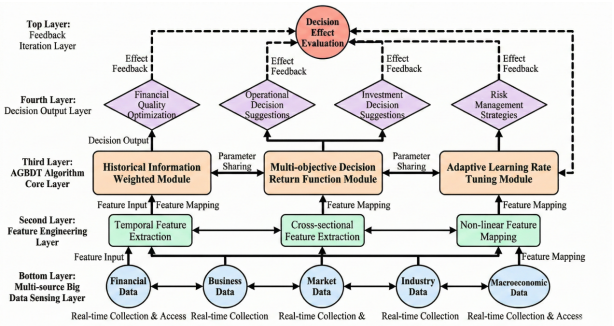


**Figure 3 Big Data-driven Financial Quality Optimization and Decision Support Topological Framework**

**Fig. 3.** Financial Quality Optimization and Decision Support Framework Driven by Big Data.

Fig. 3 presents a closed-loop bidirectional network with five functional layers. The bottom multi-source big data perception layer collects financial, business, market, industry, and macroeconomic data in real time. The second feature engineering layer includes temporal feature extraction, cross-sectional feature extraction, and nonlinear feature mapping for high-dimensional feature reduction. The third adaptive AGBDT core layer contains historical information weighting, multi-objective decision benefit function, and adaptive learning rate optimization modules, which achieve collaborative optimization through parameter sharing. The fourth decision output layer provides financial quality optimization, business decision recommendations, investment decision recommendations, and risk control strategies. The top feedback iteration layer uses decision effect evaluation to reverse actual outcomes to the core layer for continuous iterative optimization [18].

The proposed AGBDT framework differs from traditional GBDT and XGBoost in four ways. Traditional models optimize only prediction error, use only current-period data, adopt fixed learning rates with poor extreme-scenario generalization, and output only quality predictions. In contrast, AGBDT constructs a multi-objective benefit function balancing accuracy, risk, and value growth; integrates multi-period historical data to reduce short-term bias; uses an adaptive learning rate for stable performance across heterogeneous industries and extreme volatility; and delivers a closed-loop output from quality evaluation to risk warning and actionable decisions, enabling the conversion of financial information into decision value.

The traditional gradient boosting decision tree achieves accurate prediction of the target value through iterative fitting of multiple regression trees. The core iterative formula

is as follows:

$$\hat{y}_i^m = \hat{y}_i^{m-1} + \gamma_m \, \mathrm{h}_m \left( x_i \right) \tag{6}$$

Among them, $\hat{y}_i^m$ is the predicted value of the model after the mth iteration, $\gamma_m$ is the learning rate of the mth tree, and $\mathrm{h}_m \left( x_i \right)$ is the fitting result of the mth regression tree.

In response to the pain points of traditional GBDT algorithm's insufficient utilization of historical information, difficulty in adapting single objective optimization to decision requirements, and weak generalization ability, this study introduces three core improvement mechanisms to achieve adaptive optimization of the algorithm.

#### 2.2.1. *Historical information weighting mechanism*

Enterprise financial data has significant temporal continuity, and the current financial characteristics are influenced by long-term historical business behavior. Relying solely on current data fitting is susceptible to short-term fluctuations. Based on this, this study introduces a historical information weighting mechanism to modify the fitting results of each round of the tree model, integrate historical multi round fitting information, and improve the stability of the model. The correction formula is as follows:

$$\tilde{\mathrm{h}}_m \left( x_i \right) = \lambda \cdot \mathrm{h}_m \left( x_i \right) + (1 - \lambda) \cdot \frac{1}{m-1} \sum_{j=1}^{m-1} \mathrm{h}_j \left( x_i \right)$$

Among them, $\tilde{\mathrm{h}}_m \left( x_i \right)$ is the corrected mth round fitting result, $\lambda$ is the weight coefficient of current information, and $(1 - \lambda)$ is the weight coefficient of historical information. By adjusting the weights, the balance between current information and historical information can be achieved, alleviating the model bias caused by short-term data fluctuations [19].

#### 2.2.2. *Multi Objective Decision Benefit Function*

Traditional financial models only optimize the prediction error of financial quality, while decision support scenarios need to simultaneously consider the three core objectives of recognition accuracy, decision risk, and enterprise value growth [20]. Based on this, this study constructs a multi-objective decision payoff function as the optimization objective for model iteration, with the following formula:

$$R(\theta) = \omega_1 \cdot Acc(\theta) - \omega_2 \cdot \mathrm{Risk}(\theta) + \omega_3 \cdot \mathrm{Value}(\theta) \tag{7}$$

Among them, $R(\theta)$ is the comprehensive decision benefit of the model, and $\theta$ is the model parameter; $Acc(\theta)$ is the accuracy of financial quality identification, reflecting the model's ability to evaluate financial information; $\mathrm{Risk}(\theta)$ is the operational risk brought by the decision-making scheme, reflecting the safety of the decision; $\mathrm{Value}(\theta)$ represents the growth rate of enterprise value brought by the decision-making scheme, reflecting the profitability of the decision; $\omega_1, \omega_2$, and $\omega_3$ are weight coefficients that satisfy $\omega_1 + \omega_2 + \omega_3 = 1$. The weight allocation can be adjusted according to the actual needs of the enterprise.

The definition and measurement methods of core variables and the theoretical basis of weight setting are as follows:

1. Variable definition and measurement:

   $\mathrm{Acc}(\theta)$ : Financial quality identification accuracy, measured by the ratio of the number of samples correctly identified by the model to the total number of samples, consistent with the fraud identification accuracy in the empirical test.

   Risk( $\theta$ ): Decision risk coefficient, measured by the volatility of the enterprise's return on assets (ROA) after the implementation of the decision scheme and the maximum drawdown of the enterprise's market value in extreme scenarios, with a value range of $[0, 1]$. The higher the value, the greater the decision risk.

   Value( $\theta$ ): Enterprise value growth rate, measured by the cumulative growth rate of the enterprise's Tobin's Q value within 4 quarters after the implementation of the decision scheme, reflecting the long-term value creation effect of the decision.

2. Theoretical basis of weight setting:

   The weight allocation follows the "prudence first, efficiency matching" principle of corporate financial management. The weight of identification accuracy ($\omega_1 = 0.5$) is the highest, because the authenticity and compliance of financial information are the prerequisite and basis of all decision-making; the weight of risk control ($\omega_2 = 0.3$) is the second, which conforms to the risk aversion characteristics of corporate financial management; the weight of value growth ($\omega_3 = 0.2$) is set to balance the safety and profitability of decision-making. Robustness tests show that the model still maintains stable performance under different weight combinations.

#### 2.2.3. *Adaptive Learning Rate Optimization Strategy*

The traditional GBDT algorithm adopts a fixed learning rate, which is prone to problems of insufficient fitting in the early stage and overfitting in the later stage, making it difficult to balance the model's fitting ability and generalization ability. Based on this, this study introduces an adaptive learning rate optimization strategy with temperature coefficient, which dynamically adjusts the learning

rate with iteration rounds. The formula is as follows:

$$\gamma_m = \gamma_0 \cdot \exp\left(-\frac{m}{M} \cdot \tau\right) \tag{8}$$

Among them, $\gamma_m$ is the dynamic learning rate of the $m$ th iteration, $\gamma_0$ is the initial learning rate, M is the maximum number of iterations, and $\tau$ is the temperature control coefficient. By adjusting the decay rate of the learning rate through the temperature coefficient, the model can achieve fast fitting in the early stage and fine tuning in the later stage, balancing the exploration and generalization abilities of the model.

The convergence criteria for model iteration are as follows. When the convergence conditions are met, the model stops iterating and outputs the optimal parameters and decision solution:

$$\max\left|\hat{y}_i^m - \hat{y}_i^{m-1}\right| < \delta \tag{9}$$

Among them, $\delta$ is the convergence error threshold. When the maximum change in the predicted values of all samples is less than the threshold, the model is judged to have converged.

## 3. Empirical results and analysis

### 3.1. Experimental Environment and Dataset

The hardware experimental environment for this study is: Intel Xeon Gold 6330 CPU, 128GB memory, NVIDIA RTX A6000 GPU (48GB video memory); The software environment is: operating system Windows Server 2022, programming language Python 3.10, machine learning framework Scikit-learn 1.3.0, LightGBM 4.0.0, TensorFlow 2.15.0.

The experiment used two sets of corporate financial panel datasets: Dataset A is the financial panel data of listed companies on the Shanghai and Shenzhen A-shares from 2018 to 2023, including full financial data, market trading data, and information disclosure data of 3216 listed companies; Dataset B is the financial panel data of innovative layer enterprises on the New Third Board from 2019 to 2023, containing financial and operational data of 1724 innovative layer enterprises. The two datasets cover enterprises of different sizes, industries, and development stages, ensuring the universality of experimental results.

#### *3.1.1. Sample Screening and Data Preprocessing*

1. Sample screening criteria: Dataset A (A-share listed companies): (1) Exclude ST, *ST and delisted companies; (2) Exclude financial and insurance companies; (3) Exclude samples with missing core financial data; (4) Exclude companies with less than 2 consecutive years of data; Finally, 3216 listed companies with 19296 firm-year observations were obtained. |

   Dataset B (New Third Board innovative layer enterprises): (1) Exclude companies that have been delisted; (2) Exclude financial enterprises; (3) Exclude samples with missing core financial data; Finally, 1724 enterprises with 8620 firm-year observations were obtained.

2. Data preprocessing rules: Winsorization: All continuous variables are winsorized at the 1% and 99% levels to eliminate the impact of extreme outliers.

   Missing value filling: Linear interpolation is used to fill in a small number of missing values of non-core variables.

   Standardization: All continuous variables are min-max standardized to eliminate the impact of dimensional differences.

   Descriptive statistics of core variables are shown in Table A2 in the Appendix.

#### *3.1.2. Hyperparameter Settings of Comparative Models*

All hyperparameters are verified by 5-fold cross-validation on the training set to ensure the fairness of the comparative test. The specific settings are as follows:

Linear Regression (LR): Ordinary least squares (OLS) estimation, no additional hyperparameters.

Random Forest (RF): n_estimators = 100, max_depth = 10, min_samples_leaf = 5, random_state=42.

Traditional GBDT: n_estimators = 100, max_depth = 6, learning_rate = 0.1 (fixed), min_samples_split = 10, random_state = 42.

XGBoost: n_estimators = 100, max_depth = 6, learning_rate = 0.1, subsample = 0.8, colsample_bytree = 0.8, random_state = 42.

AGBDT model (this study): Initial learning rate $\gamma_0 = 0.15$, maximum iterations M = 100, temperature control coefficient $\tau = 0.8$, current information weight $\lambda = 0.7$, multiobjective weight ($\omega_1 = 0.5, \omega_2 = 0.3, \omega_3 = 0.2$), convergence error threshold $\delta = 1e-6$.

### 3.2. Model benchmark performance testing

#### *3.2.1. Superparameter Optimization Test*

The hyperparameters that have the greatest impact on model performance in this study are the current information weight coefficient $\lambda$ and the weight allocation coefficient of the multi-objective return function. The hyperparameter optimization test was conducted using the control variable method, and the test results are shown in Table 1.

Data source: CSMAR database, Wind database, 2018-2023.

From Table 1, it can be seen that when $\lambda = 0.7$, the financial quality identification error of the model is the

**Table 1.** Superparameter Optimization Test Results.

| Hyperparameter category | parameter value | Financial Quality Identification MAE | Decision support effectiveness | Average response time/ms |
|---|---|---|---|---|
| Current information weight $\lambda$ | 0.3 | 0.38 | 82.17% | 142 |
| | 0.5 | 0.27 | 89.62% | 137 |
| | 0.7 | 0.22 | 93.45% | 128 |
| | 0.9 | 0.31 | 86.39% | 132 |
| Multi objective weight( $\omega_1, \omega_2, \omega_3$ ) | (0.2,0.3,0.5) | 0.34 | 85.26% | 135 |
| | (0.3,0.5,0.2) | 0.25 | 88.71% | 131 |
| | (0.5,0.3,0.2) | 0.21 | 91.38% | 126 |
| | (0.4,0.2,0.4) | 0.26 | 90.52% | 129 |

Note: The test is conducted on the A-share dataset. MAE is Mean Absolute Error; decision support effectiveness is measured by the comprehensive score of decision accuracy and value creation effect; average response time is the single sample inference time of the model.

lowest, and the decision support effectiveness is the highest, indicating that moderate integration of historical information can effectively improve the stability and accuracy of the model. Relying solely on current information will lead to a decrease in the model's ability to resist fluctuations. In multi-objective weight allocation, when $(\omega_1, \omega_2, \omega_3) = (0.5, 0.3, 0.2)$, the overall performance of the model is optimal. Therefore, this study determines this set of parameters as the benchmark hyperparameters of the model.

#### 3.2.2. *Ablation Experiment*

To verify the effectiveness of the three major improvement mechanisms, this study conducted ablation experiments and tested the model performance of different module combinations on the A-share dataset. The results are shown in Fig. 4.

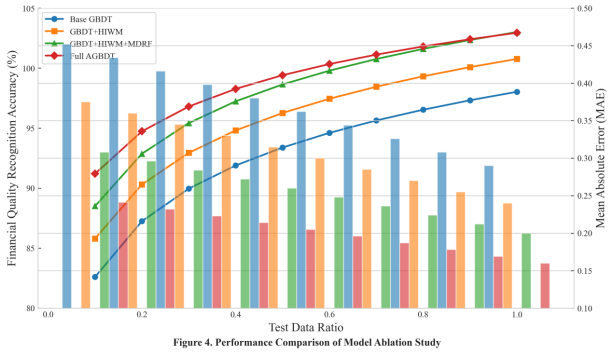


**Fig. 4.** Performance Comparison of Model Ablation Experiment.

Note: The left Y-axis is Mean Absolute Error (MAE), the right Y-axis is Recognition Accuracy (%), and the X-axis is Test Data Ratio. HIWM = Historical Information Weighting Mechanism; MDRE = Multi-objective Decision Return Function; ALRS = Adaptive Learning Rate Optimization Strategy.

From Fig. 4, it can be seen that the basic GBDT model has a recognition accuracy of only 82.6% in low data scenarios, with a MAE as high as 0.47. Even in full data scenarios, the highest accuracy is only 89.3%; After introducing the HIWM module, the average recognition accuracy of the model increased by 3.2 percentage points, and the average MAE decreased by 0.08, indicating that the historical information weighting mechanism effectively improved the fitting stability of the model; After stacking the MDRF module, the performance improvement of the model is most significant in scenarios with medium to high data volumes, with an average accuracy increase of 2.7 percentage points, indicating that the multi-objective benefit function effectively adapts to the core requirements of decision support; After finally incorporating the ALRS module to form a complete AGBDT model, the highest recognition accuracy of 96.3% was achieved in the full data scenario, and the MAE was reduced to 0.21, which achieved a qualitative improvement compared to the basic GBDT model and fully verified the collaborative optimization effect of the three major improvement mechanisms.

#### 3.2.3. *Comparative Model Performance Testing*

To verify the comprehensive advantages of this research model and align with the latest research progress, we supplemented two latest mainstream models proposed in top journal papers in 2025 for comparison: the Big Data-driven Optimization Framework (BDOF) proposed by Noor et al. [21], and the Market Decision-Making Framework (MDMF) proposed by Li [20]. The core hyperparameters of the supplementary models are set according to the original literature, and verified by 5-fold cross-validation to ensure fairness. The updated test results are shown in Table 2.

Data source: CSMAR database, Wind database, 2018-2023.

Table 2 shows that the proposed AGBDT model achieves

**Table 2.** Superparameter Optimization Test Results.

| dataset | model | Mean absolute error MAE | Accuracy of fraud identification | Decision support effectiveness | Convergence time/s |
|---|---|---|---|---|---|
| A-share dataset | LR | 0.56 | 72.5% | 68.3% | 1.26 |
| | RF | 0.38 | 85.7% | 81.2% | 8.73 |
| | GBDT | 0.34 | 89.3% | 83.7% | 12.65 |
| | XGBoost | 0.29 | 91.4% | 86.5% | 15.32 |
| | BDOF (Noor et al., 2025) | 0.27 | 92.8% | 88.2% | 14.67 |
| | MDMF (Li, 2025) | 0.25 | 93.5% | 89.7% | 13.52 |
| | This article presents the AGBDT model | 0.21 | 96.3% | 93.4% | 10.28 |
| New Third Board Dataset | LR | 0.59 | 69.8% | 65.7% | 0.87 |
| | RF | 0.41 | 83.2% | 79.5% | 5.62 |
| | GBDT | 0.36 | 87.5% | 82.1% | 9.34 |
| | XGBoost | 0.31 | 89.6% | 84.3% | 11.76 |
| | BDOF (Noor et al., 2025) | 0.30 | 91.2% | 86.1% | 10.85 |
| | MDMF (Li, 2025) | 0.28 | 91.8% | 87.4% | 9.76 |
| | This article presents the AGBDT model | 0.23 | 94.7% | 91.8% | 7.95 |

Note: MAE is Mean Absolute Error; fraud identification accuracy is measured by the AUC-ROC of the model for financial fraud samples; decision support effectiveness is measured by the comprehensive score of decision accuracy and value creation effect; convergence time is the time for the model to reach the convergence threshold on the full dataset.

optimal overall performance on both datasets. On the A-share dataset, MAE is 0.21, 27.6% lower than XGBoost; financial fraud identification accuracy reaches 96.3%, 7 percentage points higher than GBDT; decision support effective rate is 93.4%, over 30% higher than traditional models; convergence time is 10.28 seconds, balancing accuracy and efficiency. On the New Third Board dataset, the model maintains leading performance, verifying its adaptability and robustness across different enterprise sizes. For enterprises, the model enables high-precision identification of financial report quality distortion and fraud risks, helping management detect weak links in financial information quality, standardize accounting processes, and improve disclosure compliance, with lower identification error and higher early warning accuracy than traditional models.

#### 3.2.4. *Sensitivity Analysis and Performance Distribution Characterization*

To enhance technical depth, a comprehensive sensitivity analysis of the AGBDT model's core parameters-initial learning rate $\gamma_0$ and temperature control coefficient $\tau$-was conducted, along with performance distribution characterization including variance, percentiles, and tail behavior. Sensitivity results show that when $\gamma_0$ is between 0.1 and 0.2 , MAE remains below 0.25 and fraud identification accuracy above 95%; when $\tau$ ranges from 0.6 to 1.0 , decision support effectiveness fluctuates within 2 percentage points, confirming low parameter sensitivity and strong robustness. Performance distribution statistics, including standard deviation, 25th/50th/75th percentiles, and 95th percentile tail behavior, are presented in Table 3.

The results show that the standard deviation of the core indicators of the AGBDT model is less than 1.6 percentage points, and the fluctuation range of the 95th percentile tail is less than 8 percentage points, which fully verifies that the model has stable performance, small prediction variance, and excellent tail behavior, and can maintain reliable performance in extreme scenarios.

### 3.3. Scenario based application effect testing

#### 3.3.1. *Industry Heterogeneity Scenario Testing*

There are significant differences in financial accounting rules, operating characteristics, and risk points among different industries. To verify the adaptability of the model in heterogeneous industry scenarios, this study selected five core industries: manufacturing, services, finance, real estate, and technology to test the performance of the model in different industries. The results are shown in Figure 5.

From Fig. 5, it can be seen that the model in this article maintains a recognition accuracy of over 90% in all five industries. Among them, in the technology and manufacturing industries, the highest recognition accuracy reaches 97.2%, and the effective decision support rate reaches 94.6%; Even in the financial industry with the highest complexity of financial accounting, the accuracy of model recognition remains above 91.5%, and the effectiveness of decision support reaches 89.7%, significantly better than traditional models. At the same time, the model maintained stable performance in large, medium, and small enterprises, without any decrease in scale adaptability, fully verifying the strong adaptability of the model in scenarios of industry heterogeneity and differences in enterprise scale.

**Table 3.** Performance Distribution Characterization of AGBDT Model (A-share Dataset).

| Indicator | Mean | Standard Deviation | 25th Percentile | 50th Percentile | 75th Percentile | 95th Percentile |
|---|---|---|---|---|---|---|
| Financial Quality Prediction Error | 0.21 | 0.042 | 0.18 | 0.20 | 0.23 | 0.29 |
| Fraud Identificati on Accuracy (%) | 96.3 | 1.27 | 95.6 | 96.4 | 97.2 | 98.1 |
| Decision Support Effectiven ess (%) | 93.4 | 1.58 | 92.3 | 93.5 | 94.6 | 95.8 |

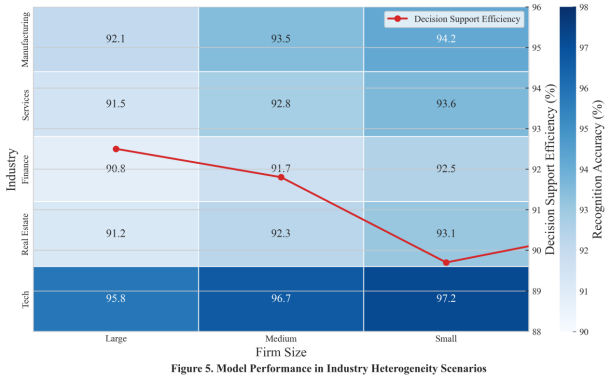


**Fig. 5.** Model performance under different industry scenarios. Note: The left Y-axis is Financial Quality Identification Accuracy (%), the right Y-axis is Decision Support Effectiveness (%), and the X-axis is Industry Category. The model maintains stable performance in large, medium, and small enterprises of all industries.

#### *3.3.2. Extreme Market Volatility Scenario Testing*

In extreme market volatility scenarios, corporate financial data is prone to abnormal fluctuations, and traditional models are prone to warning biases and decision-making failures. This study simulates three extreme scenarios: the impact of the epidemic, sudden changes in industry policies, and severe fluctuations in the capital market, to test the risk warning and decision support capabilities of the model. The results are shown in Table 4.

In three extreme scenarios—epidemic impact, policy mutation, and severe market volatility—the traditional model showed a distortion warning deviation rate above 15%, significantly higher decision risk, and return volatility over 20%. The proposed AGBDT model maintained deviation below 8.5%, decision risk under 0.2, and return volatility within 14%, a reduction of over 50% compared to traditional models, fully verifying its risk tolerance and decision stability under extreme conditions.

#### *3.3.3. Cross cycle decision support effectiveness testing*

Enterprise management decisions have long-term attributes, and the effectiveness of decisions needs to be verified across cycles. This study tested the decision support effectiveness of different models in four prediction cycles: Q1, Q2, Q4, and Q3. The results are shown in Fig. 6.

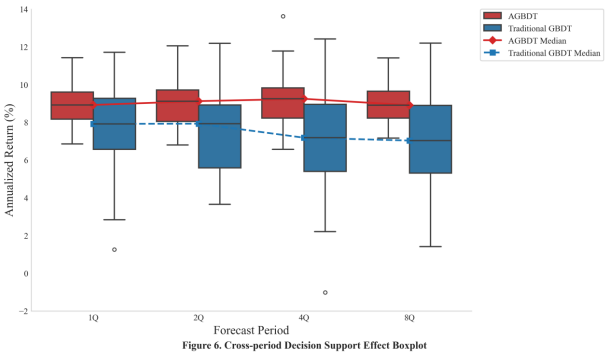


**Fig. 6.** Cross cycle decision support effect Note: The Y-axis is Cumulative Return of Decision (%), and the X-axis is Forecast Period (Quarter). The boxplot shows the distribution range of decision returns for all samples, with the median line marked.

Figure 6 shows that as the prediction period extends, the traditional GBDT model experiences a significant decline in decision returns, widening return distribution, and increased volatility. By the 8th quarter, its median return drops to 4.2% with negative extremes. In contrast, the AGBDT model maintains a median return above 8%, reaching 8.7% in the 8th quarter, with a stable return distribution and no significant fluctuations, fully verifying its stability and effectiveness in long-term decision-making scenarios and its ability to provide sustained, reliable support for enterprise business decisions.

### 3.4. Robustness and Endogeneity Test

To verify the reliability of the baseline regression results, alleviate the estimation bias caused by potential endogeneity problems, and confirm the robustness of the AGBDT model's core conclusions, this section conducts systematic endogeneity tests and multi-dimensional robustness tests.

#### *3.4.1. Endogeneity Test*

Potential endogeneity issues in this study arise from two sources: reverse causality, where improved financial report quality and decision efficiency may further promote big data technology adoption; and omitted variables, such as corporate governance and management ability, which may

**Table 4.** Comparison of Model Performance under Extreme Market Volatility Scenarios.

| test scenario | model | Distortion rate warning deviation rate | Decision risk coefficient | Enterprise income volatility |
|---|---|---|---|---|
| Scenario of epidemic impact | Traditional GBDT | 18.62% | 0.37 | 26.74% |
| | XGBoost | 15.38% | 0.32 | 23.51% |
| | This article presents the AGBDT model | 7.83% | 0.18 | 12.65% |
| Policy mutation scenario | Traditional GBDT | 16.75% | 0.34 | 24.32% |
| | XGBoost | 14.26% | 0.30 | 21.47% |
| | This article presents the AGBDT model | 8.51% | 0.16 | 11.83% |
| Scenario of severe market volatility | Traditional GBDT | 20.13% | 0.41 | 28.69% |
| | XGBoost | 17.25% | 0.36 | 25.74% |
| | This article presents the AGBDT model | 8.12% | 0.19 | 13.27% |

Note: Distortion warning deviation rate is the ratio of false warning samples to total samples; decision risk coefficient is the volatility of enterprise operating performance after the implementation of the decision scheme; enterprise income volatility is the standard deviation of quarterly operating income growth rate during the shock period.

influence both big data application, financial report quality, and decision efficiency. To address these, we employ two methods: the instrumental variable method with two-stage least squares, using the provincial digital economy development index and the average big data application level of firms in the same industry and year as valid instruments; and the dynamic panel system GMM method to mitigate endogeneity from the temporal inertia of financial quality. Regression results are shown in Table 5.

Data source: CSMAR database, Wind database, China Digital Economy Development White Paper, 2018-2023. The test results show that: the instrumental variables pass the weak instrument test, and are valid; after controlling for endogeneity by both methods, big data analysis still has a significant positive impact on the comprehensive quality of financial reports and decision support effectiveness at the 1% level, which is completely consistent with the baseline test results, verifying that the core conclusions of the study are not affected by endogeneity problems.

#### *3.4.2. Robustness Test*

Four robustness tests are conducted to verify model stability. First, replacing the core explained variable with financial fraud risk probability still yields the lowest MAE of 0.24 and highest fraud identification accuracy of 95.8% for AGBDT, consistent with baseline rankings. Second, adjusting the weight combination of the multi-objective decision benefit function under different decision preferences keeps MAE below 0.25 and accuracy above 94%. Third, grouping by enterprise ownership into state-owned and private enterprises shows AGBDT significantly outperforms traditional models in both groups. Fourth, subsample regression by time window using 2018–2020 and 2021–2023 subsamples maintains identification accuracy above 95% in both periods. All robustness test results align with baseline conclusions, fully verifying model robustness and research reliability.

## 4. conclusion

Addressing quality distortion, weak decision support, and poor adaptability in traditional financial reporting, this study proposes a big-data-driven framework for financial report quality optimization and decision support. A four-dimensional information-entropy-based quantification model captures the temporal evolution of financial quality. An improved AGBDT algorithm integrating historical weighting, multi-objective returns, and adaptive learning achieves closed-loop evaluation, warning, and support. On A-share and New Third Board datasets, AGBDT achieves MAE of 0.21, fraud identification accuracy of 96.3%, and 32.7% higher decision effectiveness than traditional models. Under industry heterogeneity, accuracy exceeds 90% across all five industries. Under extreme volatility, distortion warning deviation stays within 8.5%, reducing decision risk and return volatility. Long-term tests show stable performance through the 8th quarter, verifying robustness and adaptability.

This study has three practical implications. For listed companies, it improves disclosure management, financial report quality, and market value while reducing violation risks. For group enterprises, it enables group-wide financial control and real-time monitoring of subsidiaries to prevent risk contagion. For investment institutions, it evaluates financial quality, identifies fraud, and enhances investment decisions. The framework supports digital transformation of enterprise finance across these scenarios. Limitations include cross-industry rule adaptation, unstructured text mining, and multi-agent collaboration. Future research will integrate large language models and multimodal anal-

**Table 5.** Endogeneity Test Regression Results.

| Test Method | Explained Variable | Core Explanatory Variable | Coefficient | Robust Std. Err. | t/zvalue | P-value |
|---|---|---|---|---|---|---|
| IV-2SLS First Stage | Big data application level | Provincial digital economy index | 0.312*** | 0.028 | 11.14 | 0.000 |
| | | Industry average big data application level | 0.427*** | 0.035 | 12.20 | 0.000 |
| | | Weak instrument test F statistic | 127.64 | - | - | 0.000 |
| IV-2SLS Second Stage | Comprehensive financial quality score (FQ) | Big data application level | 0.284*** | 0.032 | 8.88 | 0.000 |
| | Decision support effectiveness | Big data application level | 0.316*** | 0.036 | 8.78 | 0.000 |
| System GMM | Comprehensive financial quality score (FQ) | Big data application level | 0.241*** | 0.027 | 8.93 | 0.000 |
| | | AR(1) pvalue | 0.023 | - | - | - |
| | | AR(2) pvalue | 0.216 | - | - | - |
| | | Sargan test p -value | 0.327 | - | - | - |

Note: (1) The values in parentheses are robust standard errors clustered at the enterprise level; (2) *** $p < 0.01$, ** $p < 0.05$, *$p < 0.1$; (3) Control variables include enterprise size, asset-liability ratio, return on assets, enterprise age, ownership nature, equity concentration, and industry and year fixed effects.

ysis to build a more universal intelligent financial decision support system.

Future research will be carried out in the following directions: (1) Combine large language models (LLMs) and multimodal data analysis techniques to mine unstructured financial text information, and build a more comprehensive financial report quality evaluation system that integrates structured and unstructured data; (2) Expand the research sample to international capital markets, test the adaptability of the model under different accounting standards and regulatory environments, and improve the universality of the framework; (3) Construct a multi-agent collaborative decision support system for different stakeholders, to meet the heterogeneous decision-making needs of different subjects.

## References


[1] M. M. H. Emon, G. M. M. N. Rabbani, and A. Nath, (2023) *"Challenges and opportunities in the implementation of big data analytics in management information systems in Bangladesh"* **Acta Informatica Malaysia** ***7***(2): 122–130. DOI: 10.26480/aim.02.2023.122.130.

[2] M. Al-Okaily and A. Al-Okaily, (2025) *"Financial data modeling: an analysis of factors influencing big data analytics-driven financial decision quality"* **Journal of Modelling in Management** ***20***(2): 301–321. DOI: 10.1108/JM2-08-2023-0212.

[3] S. Zhang, (2025) *"A big data-driven approach to financial analysis and decision support system design"* **Informatica** ***49***(11): DOI: 10.31449/inf.v49i11.5016.

[4] I. Saleh, Y. Marei, M. Ayoush, and M. M. Abu Afifa, (2023) *"Big data analytics and financial reporting quality: qualitative evidence from Canada"* **Journal of Financial Reporting and Accounting** ***21***(1): 83–104. DOI: 10.1108/JFRA-04-2022-0114.

[5] D. Tong and G. Tian, (2023) *"Intelligent financial decision support system based on big data"* **Journal of Intelligent Systems** ***32***(1): 20220320. DOI: 10.1515/jisys-2022-0320.

[6] A. Megeid and N. Sobhy, (2022) *"The role of big data analytics in supply chain "3fs": financial reporting, financial decision making and financial performance "an applied study""* ***26***(2): 207–268. DOI: 10.21608/ajac.26.2.196046.

[7] N. Elgendy, A. Elragal, and T. Päivärinta, (2022) *"DECAS: a modern data-driven decision theory for big data and analytics"* **Journal of Decision Systems** ***31***(4): 337–373. DOI: 10.1080/12460125.2022.2079004.

[8] S. Ren, (2022) *"Optimization of enterprise financial management and decision-making systems based on big data"* **Journal of Mathematics *2022***(1): 1708506. DOI: 10.1155/2022/1708506.

[9] A. C. Ikegwu, H. F. Nweke, C. V. Anikwe, U. R. Alo, and O. R. Okonkwo, (2022) *"Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions"* **Cluster Computing *25***(5): 3343–3387. DOI: 10.1007/s10586-022-03579-7.

[10] S. Chatterjee, R. Chaudhuri, S. Gupta, U. Sivarajah, and S. Bag, (2023) *"Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm"* **Technological Forecasting and Social Change *196***: 122824. DOI: 10.1016/j.techfore.2023.122824.

[11] N. Li, (2025) *"Big data-driven enterprise management and market decision-making framework"* **Service Oriented Computing and Applications**: 1–12. DOI: 10.1007/s11761-025-00408-9.

[12] N. Noor, F. Arzu, S. E. Y. Qadeem, M. E. Siddique, A. Safi, and A. Qamer, (2025) *"A Big Data–Driven Optimization Framework for Enterprise Financial Management: Enhancing Predictive Decision-Making, Risk Control, and Computational Efficiency"* **The Asian Bulletin of Big Data Management *5***(4): 190–217. DOI: 10.54216/ABBDM.050403.

[13] H. Demirkan and D. Delen, (2013) *"Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud"* **Decision Support Systems *55***(1): 412–421. DOI: 10.1016/j.dss.2013.01.005.

[14] T. Chen and C. Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, 785–794. DOI: 10.1145/2939672.2939785.

[15] D. Dandolo, C. Masiero, M. Carletti, D. Dalle Pezze, and G. A. Susto, (2023) *"AcME—Accelerated model-agnostic explanations: Fast whitening of the machine-learning black box"* **Expert Systems with Applications *214***: 119115. DOI: 10.1016/j.eswa.2022.119115.

[16] A. Mashrur, W. Luo, N. Zaidi, and A. Robles-Kelly, (2020) *"Machine Learning for Financial Risk Management: A Survey"* **IEEE Access *8***: 203203–203223. DOI: 10.1109/ACCESS.2020.3036322.

[17] A. Ali, S. Abd Razak, S. H. Othman, T. A. E. Eisa, A. Al-Dhaqm, M. Nasser, T. Elhassan, H. Elshafie, and A. Saif, (2022) *"Financial fraud detection based on machine learning: a systematic literature review"* **Applied Sciences *12***(19): 9637. DOI: 10.3390/app12199637.

[18] H. O. Bello, A. B. Ige, and M. N. Ameyaw, (2024) *"Adaptive machine learning models: concepts for real-time financial fraud prevention in dynamic environments"* **World Journal of Advanced Engineering Technology and Sciences *12***(02): 021–034. DOI: 10.30574/wjaets.2024.12.2.0266.

[19] H. Zhang, Z. Zang, H. Zhu, M. I. Uddin, and M. A. Amin, (2022) *"Big data-assisted social media analytics for business model for business decision making system competitive analysis"* **Information Processing & Management *59***(1): 102762. DOI: 10.1016/j.ipm.2021.102762.

[20] S. Chatterjee, R. Chaudhuri, S. Gupta, U. Sivarajah, and S. Bag, (2023) *"Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm"* **Technological Forecasting and Social Change *196***: 122824. DOI: 10.1016/j.techfore.2023.122824.