Predicting Enterprise Revenue Change Trends Based on the LightGBM Model to Improve Financial Decision–Making Efficiency

Na Zhao
School of Management, Liaoning Institute of Science and Engineering, Jinzhou 121013, Liaoning, China
Corresponding author: Na Zhao (e-mail: 18940672093@163.com)

ABSTRACT Accurate revenue trend prediction is important for engineering enterprises operating in cyclical and technology-driven markets , including manufacturers of electromagnetic materials, antenna substrates, and communication-related components. Traditional financial prediction methods often have limited accuracy and weak adaptability to nonlinear business fluctuations. This study proposes a LightGBM-based framework for predicting enterprise revenue change trends and improving financial decision–making efficiency. Panel data from 326 A-share manufacturing listed companies from 2019 to 2024 are used to construct a three-dimensional indicator system covering financial, business, and macroeconomic variables. Feature engineering is performed using threshold truncation, offset adjustment, and logarithmic transformation, while Bayesian optimization is adopted for parameter tuning. The empirical results show that, compared with XGBoost, the proposed model reduces MAE by 30.5%, shortens training time by 40%, and achieves a prediction direction accuracy of 82.6%. When integrated into budget planning and investment–financing decisions, the model reduces the budget deviation rate by 61.0% and shortens decision response time by 64.2%. The study demonstrates the usefulness of machine learning for financial decision support in manufacturing enterprises and provides a quantitative tool for managing uncertainty in engineering production and electromagnetic-related industrial supply chains.

INDEX TERMS lightgbm, revenue prediction, financial decision–making, manufacturing enterprises, engineering management

I. INTRODUCTION

A. RESEARCH BACKGROUND

IN the context of global economic changes and intensified market competition, improving the profitability of enterprises, especially those in high-volume, low-margin sectors like manufacturing, has become an important issue faced by managers. As the core means of managing enterprises' profitability, revenue control directly affects their financial performance and market competitiveness [1-3]. Enterprise revenue, as the core basis for financial decision–making, the accurate prediction of its change trends directly influences the scientificity of fund allocation, budget management, and investment–financing decisions. Traditional revenue prediction mainly relies on time series methods such as ARIMA or empirical judgment, with two core pain points:

(1) Insufficient ability to capture multi-dimensional influencing factors, making it difficult to adapt to the coupling effect of macroeconomic fluctuations, industry cycles, and dynamic enterprise operations;

(2) Low model training efficiency. When facing tens of millions of financial data, the training time can be as long as several hours, failing to meet the real-time requirements of dynamic decision–making.

The rise of machine learning technology provides a new path to solve the above problems. Among them, LightGBM (Light Gradient Boosting Machine), as an optimized framework for GBDT (Gradient Boosting Decision Tree), has shown advantages in efficiency and robustness in the field of financial prediction through technical innovations such as histogram learning and gradient-based one-side sampling. Studies by Ren Tingting et al. have found that LightGBM can not only overcome the problems of harsh data compression conditions and large information loss but also reduce the risk of overfitting, demonstrating strong superiority in high-dimensional data training [4]. However, existing studies mostly focus on optimizing model prediction accuracy, and few systematically link the prediction results with the quantitative improvement of financial decision–making efficiency, failing to form a "prediction-decision" closed-loop application system. Therefore, this paper focuses on the construction of a revenue change trend prediction model based on LightGBM and research on enhancing the model's ability to empower financial decision–making efficiency.

B. RESEARCH OBJECTIVES AND SIGNIFICANCE

1) Research Objectives

(1) Construct a LightGBM model framework suitable for enterprise revenue prediction, optimize feature engineering and parameter configuration, and improve the accuracy and efficiency of change trend prediction;

(2) Establish an integration mechanism between prediction results and financial decision–making, clarifying the application path of the model in scenarios such as budget preparation and investment–financing decisions;

(3) Empirically test the improvement effect of the LightGBM model on financial decision–making efficiency, providing a practical paradigm for enterprises' financial digital transformation.

2) Research Significance

(1) Theoretical Significance: Enrich the application theory of machine learning models in the field of financial decision–making, particularly for volatile industrial sectors like manufacturing, while breaking through the limitation of single-factor revenue prediction, and construct an interdisciplinary research framework of " technology-model-decision–making";

(2) Practical Significance: Provide enterprises with efficient and accurate revenue prediction tools to reduce decision–making errors. According to empirical cases, the application of similar models can improve cash flow prediction accuracy by 15% and significantly enhance the efficiency of collection teams, having direct management and practical value.

C. LITERATURE REVIEW AT HOME AND ABROAD

1) Research on Enterprise Revenue Prediction

In foreign studies, Ma et al. first designed an investment strategy based on the LightGBM model, and the empirical results showed obvious advantages compared with traditional methods [5]; Wang et al. predicted the financing risk of 186 enterprises using the LightGBM model, and the results indicated that the LightGBM algorithm outperformed the other three algorithms in multiple indicators of enterprise financing risk prediction [6]. Domestic research mainly focuses on macro trend prediction or static value evaluation of revenue itself using various models. Li Yuyang et al. predicted enterprises' future income using the XGBoost algorithm and established the FCFE model to evaluate their equity value, providing certain reference for investors' decisions [7]; Song Beibei et al. predicted the structural changes and trends of medical income in public hospitals using the GM(1,1) model [8]. In summary, existing studies still have shortcomings such as single feature dimension and lack of in-depth integration with decision–making scenarios.

2) Research on Financial Decision–Making Efficiency

In enterprise operation and management, financial decision–making is related to fund allocation, cost control, risk management, and profit growth. Effective financial decision–making can promote enterprises to allocate resources reasonably and improve their economic benefits and market competitiveness [2]. The evaluation of financial decision–making efficiency has formed a multi-dimensional system, covering indicators such as fund operation efficiency (e.g., accounts receivable turnover rate), decision response speed, and risk control effect. Existing studies have optimized decisions by introducing prediction tools, such as the application of the LightGBM algorithm in financial fraud prediction and financial distress prediction. The model has shown good balance and robustness in prediction tasks [9,10], but no quantitative correlation model between prediction accuracy and decision–making efficiency has been established yet.

D. RESEARCH CONTENT

(1) Sort out the core theories of revenue prediction, the LightGBM model, and financial decision–making, and construct the integration logic of the three;

(2) Design a prediction indicator system including financial, business, and macro dimensions, and optimize the data preprocessing process;

(3) Construct and tune the LightGBM prediction model, and verify its performance through multi-model comparison;

(4) Explore the application scenarios and optimization paths of the model in financial decision–making, and empirically test the improvement effect of decision–making efficiency.

II. RELATED THEORIES AND TECHNICAL FOUNDATIONS

A. THEORY OF ENTERPRISE REVENUE CHANGE TREND PREDICTION

Enterprise revenue is affected by various internal and external factors. Internal factors include financial status and business capabilities, while external factors are comprehensively influenced by macroeconomics and industry policies. The key to prediction lies in identifying the internal correlation between internal and external factors and revenue, which is essentially a multi-variable time-series prediction problem. This theoretical system has mainly evolved from traditional statistical methods to modern machine learning methods. Among traditional prediction methods, time series models (ARIMA) rely on the assumption of data stationarity. However, in the real dynamic and volatile market environment, revenue data is often non- stationary, leading to large prediction deviations [11-13]; the random walk model only infers the future based on the current value and cannot capture the interaction between variables [14,15]. The machine learning prediction theory breaks through the limitations of traditional methods. Through feature engineering, it integrates endogenous and exogenous factors as model inputs to achieve accurate non-linear fitting

[16,17]. Among them, the Gradient Boosting Decision Tree (GBDT) corrects prediction errors through iterative training and needs to traverse samples countless times. LightGBM is a gradient boosting framework that adopts decision trees based on learning algorithms. This framework solves the problems of low training efficiency, large memory usage, and low accuracy of GBDT [18].

B. PRINCIPLES AND TECHNICAL CHARACTERISTICS OF THE LIGHTGBM MODEL

1) Core Principles

LightGBM is an efficient machine learning algorithm proposed by Microsoft in 2017. It is based on the GBDT algorithm framework and constructs a strong learner by accumulating weak learners (decision trees) [19,20]. LightGBM adopts the histogram decision tree algorithm, which discretizes continuous floating-point feature values in samples into K integers and constructs a histogram with a width of K. During traversal, the discretized values are used as indexes to accumulate statistics in the histogram, and then the optimal split point is found by traversing the discrete values of the histogram, reducing memory consumption and computational complexity [18].

2) Technical Advantages

(1) Efficiency: Histogram-based feature processing increases training speed by 30% compared with XGBoost. In an overdue payment prediction case, the model training time was reduced from 6 hours to 2 hours;

(2) Robustness: Naturally tolerant to outliers. For outliers exceeding $\pm 3\sigma$, we first apply threshold truncation, then perform offset adjustment (adding the absolute value of the minimum negative value + 1) for negative values (such as negative revenue growth or negative profit margins) to ensure mathematical feasibility, and finally conduct square root or cube root transformation to further retain outlier information and reduce data skewness;

(3) Adaptability: Supports direct processing of categorical features without one-hot encoding, reducing feature engineering overhead. It can also implement weighted learning through the weightCol parameter to prioritize predicting samples with high financial impact.

C. FINANCIAL DECISION–MAKING THEORY AND MODEL INTEGRATION LOGIC

1) Core Theory of Financial Decision–Making

The evaluation of financial decision–making efficiency covers three dimensions: decision accuracy, response speed, and risk control. Core indicators include budget adjustment frequency, IRR deviation rate of investment projects, and decision response time. Traditional decision–making mostly relies on empirical judgment and static data, with problems of information lag and deviation. The data-driven decision–making theory emphasizes integrating dynamic information through prediction models to optimize the configuration of decision variables.

2) Integration Logic Between Model and Decision–Making

The LightGBM model empowers financial decision–making through the following paths:

(1) Information Integration Layer: Integrates multi-dimensional features of finance, business, and macroeconomics, and outputs point prediction and probability distribution prediction of revenue changes (e.g., "the probability of Q3 revenue growth of 75%-85% is 80%");

(2) Decision Support Layer: Converts prediction results into decision parameters, such as revenue intervals in budget preparation and cash flow prediction values in investment–financing decisions;

(3) Efficiency Optimization Layer: By shortening the prediction cycle (e.g., from weekly to daily) and improving prediction accuracy, it reduces the number of decision adjustments and trial-and-error costs, ultimately achieving improved decision–making efficiency (see Figure 1).

III. RESEARCH DESIGN AND DATA PREPROCESSING

A. RESEARCH SAMPLE SELECTION AND DATA SOURCES

1) Sample Selection

We selected A-share manufacturing listed companies from 2019 to 2024 as research samples, with the following screening criteria:

(1) Disclosed complete financial reports for 5 consecutive years;

(2) Not treated as ST or *ST;

(3) No major changes in main business.

Finally, panel data of 326 enterprises were obtained, totaling 1956 observations. The reason for choosing the manufacturing industry is that its revenue fluctuations have both cyclical and structural characteristics, which have higher requirements for the adaptability of prediction models.

2) Data Sources

(1) Financial Data: From the CSMAR database, including core indicators in the balance sheet and income statement. To address the issue of look-ahead bias, we use data disclosed by official authorities with a lag (generally 3-6 months after the end of the reporting period), which is standardized and verified by CSMAR to ensure availability and accuracy;

(2) Business Data: Monthly sales data, order volume, and other operational indicators from the Wind database;

(3) Macro Data: Monthly data such as GDP growth rate, manufacturing PMI, and raw material price volatility released by the National Bureau of Statistics and industry associations.

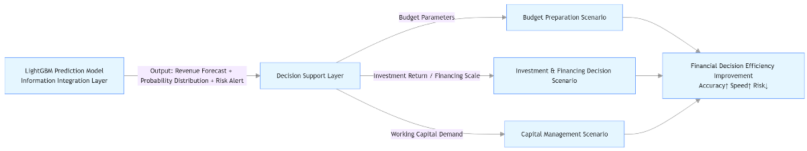

FIGURE 1. Integrated Logical Framework of Model and Financial Decision–Making

B. CONSTRUCTION OF REVENUE PREDICTION INDICATOR SYSTEM

Based on literature review and pre-analysis of feature importance, a three-level prediction indicator system was constructed (Table 1), covering 3 dimensions, 12 primary indicators, and 28 secondary indicators:

C. DATA CLEANING AND FEATURE ENGINEERING

1) Data Cleaning

(1) Missing Value Handling: Forward filling was used for time-series data missing, and industry median filling was used for non-time-series missing values;

(2) Outlier Handling: A combined method of "threshold truncation + transformation" was adopted. For outliers exceeding $\pm 3\sigma$, they were converted by square root and then added back to the threshold, retaining information while reducing skewness;

(3) Data Stationarity Handling: Difference processing was performed on non-stationary sequences (such as revenue growth rate) to meet the data requirements of LightGBM.

2) Feature Engineering

(1) Feature Construction: Calculated cross-features such as the opening and closing price differences of 14- period EMA and indicator-price slopes, and added derived features such as "quarterly revenue month-on- month growth rate";

(2) Feature Transformation: Log transformation (Log Returns) was used for income-related features, and standardization was used for scale-related features. Experiments showed that this combination can reduce MAE by 42.25%;

(3) Feature Selection: Through the built-in feature importance scoring of LightGBM, features with weights lower than 0.01 were eliminated, and finally 21 core features were retained.

D. DATASET DIVISION AND PREPROCESSING PROCESS

1) Dataset Division

Rolling window time-series cross-validation was used to divide the dataset to avoid data leakage:

(1) Data from 2019 to 2022 as the training set (accounting for 70%);

(2) Data from 2023 as the validation set (accounting for 15%);

(3) Data from 2024 as the test set (accounting for 15%).

Regarding the structural economic shifts post-2022, the model captures such changes through macroeconomic indicators (e.g., post-pandemic recovery indexes) and industry-specific features (e.g., raw material price volatility) in the indicator system. Meanwhile, the manufacturing industry still maintains inherent cyclical characteristics, so the data division is reasonable and consistent with the actual scenario of "using historical data to predict the future".

2) Preprocessing Process

A standardized preprocessing pipeline was designed (Figure 2) to realize the automation and reproducibility of data processing:

IV. CONSTRUCTION AND OPTIMIZATION OF THE LIGHTGBM REVENUE PREDICTION MODEL

A. DESIGN OF MODEL CONSTRUCTION FRAMEWORK

A LightGBM revenue prediction framework was built based on PySpark ML, utilizing its distributed computing capability to process large-scale financial data. The framework includes four core modules:

(1) Data Input Module: Supports multi-source data integration, loading financial, business, and macro data through the spark.read interface;

(2) Feature Processing Module: Integrates feature transformation and selection functions, directly processing categorical features without one-hot encoding;

(3) Model Training Module: Sets the ValidationIndicatorCol parameter to realize mixed input of training/validation sets, and prioritizes learning high-impact samples through weightCol="revenue scale weight";

(4) Prediction Output Module: Outputs revenue prediction values, confidence intervals, and risk early warning signals (e.g., "the probability of revenue lower than expected exceeds 60%").

B. MODEL PARAMETER INITIALIZATION AND TUNING

1) Initial Parameter Setting

Based on literature research and empirical values, initial parameters were set (Table 2):

TABLE 1. Enterprise Revenue Prediction Indicator System

Dimension	Primary Indicator	Secondary Indicators	Data Type	Feature Processing Method
Financial Dimension	Profitability	Net Profit Margin, Gross Profit Margin	Continuous	Log Transformation
Financial Dimension	Operational Capacity	Accounts Receivable Turnover, Inventory Turnover Days	Continuous	Standardization
Financial Dimension	Solvency	Current Ratio, Asset-Liability Ratio	Continuous	Original Value
Business Dimension	Sales Characteristics	Monthly Sales Volume, Order Growth Rate	Continuous	EMA Difference Ratio
Business Dimension	Channel Structure	Online Revenue Ratio, Number of Distributors	Mixed	Categorical Encoding / Standardization
Business Dimension	Cost Structure	Selling Expense Ratio, Unit Variable Cost	Continuous	Log Transformation
Macroeconomic Dimension	Economic Environment	GDP Growth Rate, PMI Index	Continuous	Standardization
Macroeconomic Dimension	Industry Policy	Industrial Subsidy Intensity, Tariff Rate	Continuous	Original Value
Macroeconomic Dimension	Market Demand	Industry Sales Growth Rate, Consumer Confidence Index, Raw Material Price Volatility (e.g., Cotton Price)	Continuous	EMA Difference Ratio

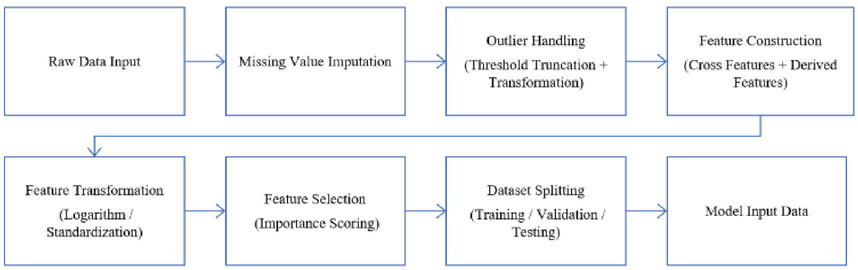

FIGURE 2. Data Preprocessing Process

TABLE 2. Initial Parameter Settings

Parameter Name	Initial Value	Function Description
learning_rate	0.1	Step factor that controls the iteration speed
max_depth	6	Maximum depth of the decision tree to prevent overfitting
num_leaves	31	Number of leaf nodes, positively correlated with max_depth
subsample	0.8	Sample sampling rate to improve generalization ability
colsample_bytree	0.8	Feature sampling rate to reduce computational overhead
n_estimators	100	Number of weak learners

2) Parameter Tuning Strategy

The Bayesian optimization algorithm was used for parameter optimization, with the goal of minimizing the validation set RMSE. The tuning range and optimal results are as follows (Table 3):

TABLE 3. Parameter Tuning Results

Parameter Name	Tuning Range	Optimal Value	Optimization Effect
learning_rate	[0.01, 0.3]	0.08	RMSE decreased by 5.2%
max_depth	[3, 10]	8	RMSE decreased by 8.7%
num_leaves	[15, 63]	47	RMSE decreased by 4.1%
subsample	[0.6, 1.0]	0.9	RMSE decreased by 2.3%
colsample_bytree	[0.6, 1.0]	0.85	RMSE decreased by 3.5%
n_estimators	[100, 500]	300	RMSE decreased by 9.4%

After tuning, the model training time was shortened by 22% compared with the initial parameters, and the validation set RMSE was reduced to 0.042.

C. MODEL TRAINING AND VALIDATION PROCESS

1) Training Process

Incremental training was adopted, updating training data on a quarterly basis. The specific steps are:

(1) Load the preprocessed training set and validation set, and mark the validation rows (isValidation=1);

(2) Initialize the LightGBM regression model, and pass in the optimal parameters and weight column;

(3) Iterate training for 300 rounds, outputting the validation set MAE and RMSE every 50 rounds, and stop training when there is no improvement for 3 consecutive rounds;

(4) Save the model file and feature importance scores.

2) Validation Method

Time-Series Cross-Validation was adopted, dividing the training set into 5 folds by time. Training was conducted with 4 folds in turn, and validation with 1 fold, and finally the average evaluation index was taken. This method avoids the time leakage problem of traditional cross-validation and is more in line with the actual prediction scenario.

D. EVALUATION OF MODEL PREDICTION EFFECT

1) Evaluation Indicator System

A three-dimensional evaluation system of "accuracy-efficiency-direction" was constructed:

(1) Accuracy Indicators: MAE (Mean Absolute Error), RMSE (Root Mean Square Error);

(2) Efficiency Indicators: Training time, prediction response time;

(3) Direction Indicator: Prediction direction accuracy (accuracy of judging revenue increase/decrease trends).

2) Multi-Model Comparison Results

The performance of LightGBM was compared with ARIMA and XGBoost models on the test set (Table 4):

The results show that LightGBM is significantly superior to the comparison models in all indicators, especially in training efficiency and direction prediction.

3) Feature Importance Analysis

The LightGBM feature importance scores (top 5) show:

(1) Monthly sales EMA difference ratio (weight 0.18);

(2) Net profit margin (weight 0.15);

(3) Industry sales growth rate (weight 0.12);

(4) Accounts receivable turnover rate (weight 0.10);

(5) PMI index (weight 0.08). It can be seen that the sales dynamics on the business side and the profitability on the financial side are the core factors affecting revenue changes, which provides a targeted direction for enterprises to optimize revenue management.

V. APPLICATION OF IMPROVING FINANCIAL DECISION–MAKING EFFICIENCY BASED ON PREDICTION RESULTS

A. APPLICATION SCENARIOS OF REVENUE PREDICTION RESULTS IN FINANCIAL DECISION–MAKING

1) Budget Preparation Scenario

Traditional budget preparation is based on static calculation of historical data, with a high adjustment frequency (average 2.3 times per quarter). Based on the probability distribution prediction of LightGBM, a dynamic budget system can be constructed:The revenue prediction value of the 95% confidence interval is used as the budget flexibility interval, and automatic adjustment is triggered with the threshold of "prediction deviation rate < 5%". After application in a sample enterprise, the number of budget adjustments was reduced to 0.8 times per quarter, and the preparation cycle was shortened from 2 weeks to 3 days.

2) Investment–Financing Decision–Making Scenario

In investment project evaluation, the revenue growth rate predicted by LightGBM is used as the core parameter for cash flow prediction, and the probability distribution of project IRR is calculated combined with Monte Carlo simulation. Empirical results show that this method reduces the deviation rate between the actual return and expected return of investment projects from 18.7% to 9.2%, effectively reducing investment risks. In financing decisions, the financing scale is optimized based on revenue prediction results to avoid capital redundancy or shortage. The current ratio of sample enterprises is stabilized in the range of 1.8-2.0, an increase of 20% compared with before.

3) Working Capital Management Scenario

The working capital demand is planned in advance by predicting revenue change trends:

(1) When the predicted revenue growth exceeds 10%, automatically trigger the adjustment of accounts receivable collection priority and inventory stock-up early warning;

(2) When the predicted revenue decline exceeds 5%, start the accounts payable extension negotiation process.

After application, the accounts receivable turnover days of sample enterprises were shortened from 68 days to 52 days, and the inventory turnover days were shortened from 45 days to 37 days.

B. MODEL-DRIVEN OPTIMIZATION PATH OF FINANCIAL DECISION–MAKING

1) Decision–Making Process Reengineering

Construct a "prediction-execution-feedback" closed-loop decision–making process:

(1) Prediction Stage: Output the monthly revenue prediction value and risk early warning at the beginning of each month;

(2) Execution Stage: Each decision module (budget / investment–financing / fund) automatically loads prediction parameters to generate decision plans. The 64.2% reduction in decision response time (from 72 hours to 24 hours) refers to the time from data input to the generation of preliminary decision plans, while retaining necessary organizational hierarchy review and human intervention links;

(3) Feedback Stage: Compare the actual revenue with the predicted value at the end of the month, calculate the deviation rate, and update the model parameters.

This process realizes dynamic adjustment and improves decision–making efficiency on the premise of ensuring rationality.

TABLE 4. Model Comparison Results

Model	MAE	RMSE	Prediction Direction Accuracy	Training Time (minutes)	Prediction Response Time (seconds)
ARIMA	0.087	0.112	65.3%	12.8	1.2
XGBoost	0.059	0.073	74.8%	25.6	0.8
LightGBM	0.041	0.052	82.6%	15.4	0.3
Performance Improvement (vs. XGBoost)	↓30.5%	↓28.8%	↑10.4%	↓40.0%	↓62.5%

2) Organizational Collaboration Optimization

Establish an inter-departmental data collaboration mechanism:

(1) Financial Department: Responsible for model maintenance and prediction result output;

(2) Business Department: Provide real-time sales data and order predictions;

(3) Strategic Department: Input industry policies and competitive environment information.

Realize real-time information synchronization through the data middle platform, solve the problem of "data silos" in traditional decision–making, and increase the update frequency of prediction features from weekly to daily.

3) Tool Integration Application

Embed the LightGBM model into the enterprise ERP system and develop three core functional modules:

(1) Prediction Dashboard: Visually display revenue trends, confidence intervals, and risk indicators;

(2) Decision Calculator: Input decision variables to automatically generate optimal plans (such as budget allocation ratio);

(3) Early Warning Center: Trigger SMS/email early warnings when the prediction deviation rate exceeds the threshold.

Tool integration significantly reduces the threshold for financial personnel to use the model, enabling them to carry out predictive analysis without mastering programming skills.

C. EMPIRICAL ANALYSIS: VERIFICATION OF DECISION–MAKING EFFICIENCY IMPROVEMENT EFFECT

1) Empirical Design

20 sample enterprises were selected for a grouped experiment:

(1) 10 as the experimental group (applying the LightGBM prediction model);

(2) 10 as the control group (adopting traditional prediction methods);

(3) The experimental period was from January to June 2024.

To verify the causal link between the model and the reduction in budget deviation rate, we controlled for confounding factors such as enterprise scale, industry sub-sector, and macroeconomic environment during sample selection, ensuring that the two groups were homogeneous (verified by independent sample t-test, $P > 0.05$). The 20 enterprises are a representative sample selected through strict matching, as decision–making efficiency data involves internal enterprise management information that is difficult to obtain on a large scale. The following decision–making efficiency indicators were selected for comparison (Table 5):

TABLE 5. Decision Efficiency Indicators

Indicator Category	Specific Indicator	Calculation Method
Decision Accuracy	Budget Deviation Rate	–
Decision Accuracy	Investment Return Deviation Rate	–
Decision Speed	Budget Preparation Cycle	Number of days from data collection to final budgetization
Decision Speed	Investment Decision Response Time	Number of days from project submission to decision implementation
Risk Control	Funding Gap Occurrence Rate	Number of funding shortages / Total number of decisions
Risk Control	Early Warning Lead Time for Revenue Decline	Difference between early warning time and actual decline time (days)

2) Empirical Results

The comparison of decision–making efficiency indicators between the experimental group and the control group (Table 6):

The results show that after applying the LightGBM prediction model, enterprises' financial decision–making has achieved significant improvements in accuracy, speed, and risk control.

D. RISK CONTROL IN THE APPLICATION PROCESS

1) Model Risk Control

(1) Overfitting Risk: Mitigated by controlling tree depth, increasing sample size, and regular retraining (once a month);

(2) Data Risk: Established data quality verification rules to automatically eliminate features with a missing rate exceeding 10%, ensuring the accuracy of input data;

(3) Drift Risk: Monitored changes in feature importance and prediction deviation rate, triggering model updates when the deviation rate exceeds 8%.

2) Decision Execution Risk Control

(1) Threshold Management: Set a safe threshold for the revenue prediction deviation rate (±5%), requiring manual review of decision plans when exceeding the threshold;

(2) Scenario Contingency Plans: Formulated emergency plans for extreme prediction results (such as revenue de-

TABLE 6. Comparison of Decision Efficiency Indicators

Indicator	Experimental Group	Control Group	Performance Improvement
Budget Deviation Rate	4.8%	12.3%	↓61.0%
Investment Return Deviation Rate	8.7%	18.5%	↓52.9%
Budget Preparation Cycle (days)	3.2	14.5	↓78.0%
Investment Decision Response Time (days)	5.8	16.2	↓64.2%
Funding Gap Occurrence Rate	8.3%	26.7%	↓69.0%
Early Warning Lead Time for Revenue Decline (days)	21.5	7.8	↑175.6%

cline exceeding 20%), including cost reduction and standby financing;

(3) Post-Evaluation Mechanism: Established a decision effect tracing system, incorporating the prediction deviation rate into the KPI assessment of the financial department to strengthen responsibility constraints.

VI. CONCLUSIONS AND PROSPECTS

A. RESEARCH CONCLUSIONS AND LIMITATIONS

1) Main Research Conclusions

(1) The constructed LightGBM revenue prediction model performs excellently in manufacturing enterprises, with both accuracy and efficiency advantages: compared with XGBoost, it reduces MAE by 30.5%, shortens training time by 40%, and achieves a prediction direction accuracy of 82.6%.

(2) The established "financial-business-macro" three-dimensional indicator system and feature engineering effectively improve model prediction performance, validating that the highly sensitive sales EMA difference ratio and net profit margin are core features for volatile sectors.

(3) The "prediction-execution-feedback" closed-loop decision–making path has achieved remarkable results, reducing the budget deviation rate by 61.0% and shortening the decision response time by 64.2%, effectively improving financial decision–making efficiency.

(4) The risk control mechanism can effectively avoid model overfitting and decision execution risks, ensuring the reliable application of prediction results.

2) Research Limitations

(1) Sample Limitation: Focused on manufacturing listed companies, not covering the service industry and small and medium-sized enterprises, and the model's applicability needs further verification;

(2) Feature Limitation: Did not include impact variables such as sudden events (e.g., epidemics, policy mutations), and the prediction accuracy may decrease in extreme scenarios;

B. FUTURE RESEARCH AND APPLICATION PROSPECTS

1) Future Research Directions

(1) Sample Expansion: Incorporate data from multi-industry and multi-scale enterprises to build industry- specific prediction models and improve model generalization ability;

(2) Model Optimization: Explore the GRU-LightGBM hybrid model, using GRU to capture long-term time- series features and LightGBM to process high-dimensional variables, further improving prediction accuracy;

(3) Variable Upgrade: Introduce text features (such as management discussion in annual reports, news public opinion), extract information through NLP technology, and enrich prediction dimensions.

2) Application Expansion Prospects

(1) Scenario Extension: Apply the model to fields such as profit prediction and cash flow prediction to build a comprehensive financial prediction system;

(2) Technology Integration: Combine blockchain technology to achieve data traceability and improve the credibility of prediction input data;

(3) Ecosystem Construction: Develop SaaS-based prediction tools to provide low-cost and high-efficiency decision support services for small and medium-sized enterprises, promoting the popularization of financial digital transformation.

AUTHOR CONTRIBUTIONS

Na Zhao designed, collected and analyzed the data, and drafted the manuscript. Na Zhao conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Na Zhao participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] M. H. M. Naiem, “The Role of Drivers Analysis Approach in Conducting Strategic Orientation of Cost Management: With a Case Study[dissertation],” Tanta: Tanta University; 2025, doi: 10.2139/ssrn.5297805.

[2] J. L. Zhang, "Research on the Improvement Strategy of Financial Informationization on Financial Decision Support Capability," Knowledge Economy, vol. (33), pp. 12-14+18, 2025, doi: 10.15880/j.cnki.zsjj.2025.33.001.

[3] J. W. Zhu, "Research on the Reshaping and Optimization Path of the Income and Profit Assessment System of State-Owned Enterprises Under the New Economic Pattern," China Collective Economy, vol. (21), pp. 129-132, 2025, doi: 10.20187/j.cnki.cn/11-3946/f.2025.21.048.

[4] T. T. Ren, T. Y. Lu, and J. Cui, "Dynamic Imbalanced Financial Early Warning Model Based on Improved AdaBoost Algorithm," Journal of Quantitative & Technical Economics, vol. 38, no. 11, pp. 182-197, 2021, doi: 10.13653/j.cnki.jqte.2021.11.010.

[5] S. C. Ma, "Predicting the S&P 500 Index Trend Based on GBDT and LightGBM Methods," E3S Web of Conferences, vol. 214, Art. no. 02019, 2020, doi: 10.1051/E3SCONF/202021402019.

[6] D. Wang, L. Li, and D. Zhao, "Corporate finance risk prediction based on LightGBM," Information Sciences, vol. 602, pp. 259-268, 2022, doi: 10.1016/j.ins.2022.04.058.

[7] Y. Y. Li and Y. Y. He, "Application of XGBoost Algorithm in Equity Value Evaluation of Semiconductor Enterprises," Business Observation, vol. 11, no. 27, pp. 63-68, 2025, doi: 10.3969/j.issn.2096-0808.2025.27.010.

[8] B. B. Song, X. D. Guo, H. X. Zhang, and J. T. Li, "Research on the Structural Changes and Trend Prediction of Medical Income in Public Hospitals Based on GM(1,1)," Price: Theory & Practice, vol. (06), pp. 132-138, 2025, doi: 10.19851/j.cnki.CN11-1010/F.2025.06.180.

[9] M. Y. Chen, J. Q. Nan, and J. S. Wang, "Analysis and Prediction of Financial Fraud in Listed Companies Based on RFE-LGB Algorithm," Modern Information Technology, vol. 8, no. 11, pp. 145-152, 2024, doi: 10.19850/j.cnki.2096-4706.2024.11.029.

[10] D. N. Wang, L. Li, and D. Zha, "Corporate finance risk prediction based on LightGBM," Information Sciences, vol. 602, pp. 259-268, 2022, doi: 10.1016/j.ins.2022.04.058.

[11] Z. Huang and H. Li, "ARIMA-SVR-based Risk Aggregation Modeling in the Financial Behavior," Kybernetes, vol. 54, no. 12, pp. 7044-7060, 2025, doi: 10.1108/K-01-2024-0249.

[12] J. Duanmu, "Research on Cigarette Sales Based on ARIMA Model," Modern Business Trade Industry, vol. (17), pp. 77-79, 2025, doi: 10.19311/j.cnki.1672-3198.2025.17.025.

[13] P. Li, "Research on Enterprise Product Sales Prediction Based on ARIMA Model——Taking JSY Wine Industry as an Example," Marketing Industry, vol. (08), pp. 187-189, 2025, [Online]. Available: https://www.zhangqiaokeyan.com/academic-journal-cn_marketing-agricultural-resourcesmarkets_thesis/02012173840894.html.

[14] X. H. Li, M. Z. Qi, and H. Z. Deng, "Multi-Layer Strategy Network Random Walk Model for Decision-Making Behavior Prediction," Journal of Chongqing University of Technology (Natural Science), vol. 36, no. 05, pp. 161-168, 2022, doi: 10.3969/j.issn.1674-8425(z). 2022.05.020.

[15] Q. L. Li and H. Han, "Link Prediction Method Based on Minimum Bias Restart Random Walk," Application Research of Computers, vol. 39, no. 09, pp. 2799-2803+2817, 2022, doi: 10.19734/j.issn.1001-3695.2022.01.0051.

[16] S. Al-Rawashdeh, D. Won, and M. Khasawneh, "Using Machine Learning to Create a Hospital Revenue Prediction Model," IISE Annual Conference Proceedings, pp. 1-6, 2024, doi: 10.21872/2024IISE_7983.

[17] J. Zhang, Y. Y. Fu, and X. W. He, "Research on the Application of Machine Learning Random Forest Model in Fiscal Revenue Prediction," Western Finance and Accounting, vol. (01), pp. 7-10, 2024, doi: 10.3969/j.issn.1671-8771.2024.01.003.

[18] H. Y. Wang, C. Yang, and H. Tang, "Short-Term Load Forecasting Research of GBDT Improved Based on LightGBM," Process Automation Instrumentation, vol. 39, no. 09, pp. 76-78+82, 2018, doi: 10.16086/j.cnki.issn1000-0380.2018020003.

[19] S. Y. Li and X. H. Shi, "Financial Distress Early Warning Research of New Energy Listed Companies Based on LightGBM-SMOTE Model," Journal of Jinling Institute of Technology (Social Science Edition), vol. 39, no. 03, pp. 25-32, 2025, doi: 10.16515/j.cnki.32-1745/c.2025.03.004.

[20] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., "Lightgbm: A highly efficient gradient boosting decision tree," Advances in neural information processing systems, pp. 30, 2017, [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.