Journal of Engineering, Project, and Production Management
2026, 16(3), 2026-46

Precise Enterprise Budget Control by Integrating BP Neural Networks and Decision Tree Algorithms

Li Chen
Associate Professor, School of Economics and Management, Anyang University, Anyang, 455000, China, E-mail: chenli-lichen@outlook.com

Project Management
Received January 13, 2026; revised April 12, 2026; May 6, 2026; accepted May 7, 2026
Available online May 29, 2026

Abstract: Precise budget regulation ensures resilience against future crises. However, large-scale and disorganized enterprise datasets often reduce prediction accuracy and weaken budget control. To address this, a decision tree algorithm is integrated with an improved Back Propagation (BP) algorithm enhanced by an additional momentum term, enabling more effective processing of enterprise data. The proposed fusion approach achieved prediction errors below 2% on the Iris dataset. Applied to enterprise budgeting, it achieved over 95% accuracy in forecasting net profit, monthly savings, and risk tolerance. These results demonstrate that the model can accurately predict departmental budgets, enabling precise budget control. Consequently, management can allocate funds more rationally, reduce resource waste, and enhance overall operational efficiency. This study provides a robust method for improving budget prediction accuracy and control, offering significant benefits for enterprise financial planning and long-term stability.

Keywords: Decision tree algorithm, back propagation algorithm, enterprise budget, prediction, precision control.

DOI 10.32738/JEPPM-2026-46

---

1. Introduction

With the continuous expansion in scale of most enterprises, enterprise resource allocation is becoming increasingly complex, and market competitiveness is also constantly increasing (Fathima et al., 2025). The above issues require enterprises to finely control costs, predict risks, and forecast various internal budgets, to control various funds of the enterprise based on the budget forecast results, thereby ensuring that the enterprise can cope with various economic risks in the future and optimize internal resource allocation (Okeke, Bakare, and Achumie, 2024) Currently, many scholars have implemented budget control measures for enterprises, attempting to predict various project budgets to achieve precise budget management.

Edge computing and data-driven methods have been used in enterprise budget control, but both still suffer from high network latency and insufficient prediction accuracy (Zhang, 2024). In addition, multiple linear regression has been used in enterprise budget control, but it has poor predictive performance on large-scale datasets (Khun-anod et al. 2023). Multi-factor analysis has also been used in enterprise budget control, but it is computationally intensive and requires long model time (Abbasov et al. 2025). Subsequently, some scholars have introduced the Back Propagation (BP) algorithm into enterprise budget control to improve its budget accuracy. However, the BP algorithm is prone to inaccurate analysis of complex data, so it is not feasible to use it alone to analyze enterprise budgets (Oyedokun et al. 2024).

To address the aforementioned issues, this study proposes a precise budget control model for enterprises that integrates C5.0 decision trees and Back Propagation Neural Networks (BPNNs). The proposed model utilizes C5.0 decision trees to pre-classify multi-departmental budget data in enterprises, dividing large-scale datasets into homogeneous subsets based on departmental characteristics to reduce data heterogeneity interference. Afterward, the BPNN improved with an additional momentum term, which is used to independently predict each subset, accelerating convergence and avoiding local optima through a momentum mechanism to improve prediction stability. Finally, a prediction decision conversion mechanism is constructed, and budget deviation tolerance intervals and inter-departmental resource allocation rules are designed to achieve effective linkage from numerical prediction to management action. The main contributions and innovations of this study are as follows: (1) Using C5.0 decision trees for departmental pre-classification of budget data, establishing a cascade architecture of classification prediction, and solving the problem of heterogeneity in large-scale data. (2) Build an additional momentum BP prediction model, optimize the weight update path through the accumulation of historical gradient information, and improve the convergence speed and accuracy of prediction. (3) Establish a collaborative mechanism between front-end data classification and back-end decision support, establish rules for converting predicted results into budget adjustment actions, and ensure the operability of control.

2. Literature Review

To control the enterprise budget, many scholars have analyzed the enterprise budget control model. For example, Wang et al.(2023) designed a budget control model based on edge computing to solve problems such as excessive enterprise resource data and network latency in the enterprise budget control model. The test findings denoted that the network latency of the control model was reduced by 12.4% when budgeting and controlling the enterprise budget, but the model paid less attention to budget accuracy. Liu et al. (2023) designed a data-driven control model to manage the raw material procurement budget for non-ferrous metals. The analysis findings denoted that the prediction accuracy of the control model for the metal material cost budget could reach 89.3%, but its prediction accuracy did not meet the expected requirements. In addition, to evaluate and improve enterprise budgets, Lu et al. (2024) designed an enterprise budget optimization model based on multiple linear regression. The test results showed that the model accurately controlled various budgets of the enterprise, but in practical applications, the model had poor analysis performance on large-scale data. Wang et al. (2024) designed a financial crisis prediction model based on multi-factor analysis to predict crises in enterprise financial budgets and control them. The test findings denoted that the prediction error of the model was only 8.4%, but when using the model, the computational resource consumption was relatively high. Ye et al. (2023) designed a budget control method based on multi-objective optimization to regulate the emergency crisis budget of petrochemical enterprises. The findings denoted that the model improved the ability of petrochemical enterprises to respond to crises by 16.4% after regulation, but the budget time of the model was relatively long.

The C5.0 DT algorithm can classify data and is often used in various models (Rezki et al., 2024). For example, Xin et al. (2024) proposed a credit risk management method based on C5.0 DT to manage bank credit risk. The test findings denoted that the accuracy of this method in analyzing bank credit could reach 90.2%. To analyze vegetation characteristics in non-agricultural environments, Wang et al. (2024) used the C5.0 DT algorithm to analyze vegetation satellite data. The results showed that this method had an error of only 2.1% in vegetation feature analysis. The BP algorithm is one of the core algorithms in deep learning and is often used in various models(Simon et al., 2023). For example, Yin et al. (2024) designed a refined oil transportation path prediction model based on the BP algorithm to reduce the sediment pollution of gasoline generated during the transportation of refined oil. The findings denoted that the model could accurately predict the production of sediment contaminated gasoline in the refined oil transportation path. In addition, to accurately identify the maturity of chili peppers, Wu et al. (2024) designed a maturity recognition model based on the BP algorithm and extreme learning machine. The findings denoted that the recognition accuracy of the model could reach 92.1%.

In summary, although many enterprise budget control models are currently available, their analysis of large-scale internal datasets often results in low predictive accuracy and poor budget control effectiveness due to data confusion. To address the above problems, this study uses C5.0 DT to optimize the BP algorithm and constructs an enterprise budget control model based on the C5.0-BP algorithm, thereby improving the accuracy of budget prediction and regulating budgets accordingly.

3. Enterprise Budget Control Model Based on BP and C5.0 DT

3.1. BPNN Algorithm based on C5.0 DT Optimization

Enterprise budget management can help companies effectively allocate and schedule resources, ensuring that the funding needs of various departments and projects match business needs, thereby improving the efficiency of resource utilization in the enterprise. Through budgeting, enterprises can set cost limits, strictly monitor expenses, reduce expenditures, and thereby improve profitability. It is also possible to reserve risk reserves through budgeting to cope with various uncertainties. At present, there are many enterprise budget control models, but due to the excessive use of enterprise data, many models still have poor data feature extraction, inaccurate data classification, and large budget errors. So, it is necessary to optimize the current enterprise budget control model. Back Propagation Neural Network (BPNN) is a multi-layer feedforward network that achieves self-optimization through error BP algorithm. It can process nonlinear relationships in data, learn patterns in data, and generate predictive models with generalization ability based on data patterns (Hunjra et al., 2024; Mahmud et al., 2024). The basic process of BPNN prediction is denoted in Fig. 1.

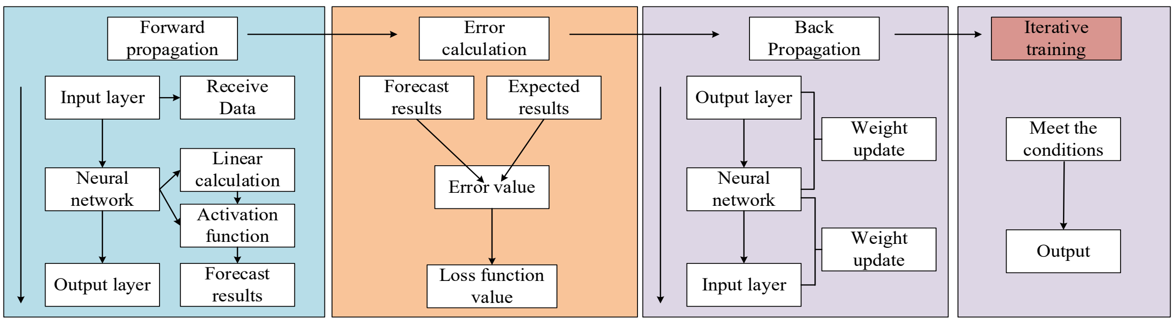


Fig. 1. BPNN prediction process

As shown in Fig. 1, BPNN prediction comprises four steps: forward propagation, error calculation, backpropagation, and iterative training. In forward propagation, input data is received and passed from the input layer to the output layer through a Neural Network (NN). The NN contains multiple hidden layers, and linear calculations are performed on the data in each hidden layer, followed by activation function processing to obtain the layer's output. Then it serves as input to the next hidden layer until the output prediction is obtained. The formula for linear calculation is shown in Eq. (1).

$$Z = Wx + b \tag{1}$$

In Eq. (1), Z is a linear function, W denotes the weight of the hidden layer, x denotes the input vector, and b represents the bias vector. The Sigmoid function is selected as the activation function for the hidden layer, and its expression is denoted in Eq. (2).

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

In Eq. (2), $\sigma(x)$ represents the activation function. In error calculation, the predicted results of forward propagation are compared with the expected output, and their mean square error is calculated. This error is used as the loss function value. Starting from the output layer, the partial derivative of the loss function value with respect to the weight parameters is calculated, and the weights are updated using the obtained weight gradient. However, during weight updates, the BP algorithm is prone to getting stuck in local minima, resulting in poor performance of the trained network. To address this issue, an additional momentum term is introduced in the weight update formula, which introduces the direction and magnitude of the previous weight update. This ensures that the weight update does not rely solely on the current gradient information, but can also be influenced by previous weights, avoiding BP from getting stuck in local minima. The update formula is shown in Eq. (3).

$$W' = \eta \frac{\partial E}{\partial W} + \alpha W \tag{3}$$

In Eq. (3), W' refers to the updated weight, η represents the learning rate, α means the momentum coefficient, W represents the pre-update weight, and E means the error. The above process is repeated iteratively to train the BP network until the calculated error meets the required threshold, or the maximum number of iterations is reached. BP networks can learn patterns in data and predict trends in data changes. However, when learning from large-scale datasets, the BP algorithm encounters a wide range of data types, which can lead to long computation time and high resource consumption for the BPNN. Moreover, the complexity of the data makes it difficult for it to learn patterns, thereby reducing its predictive performance. The C5.0 DT algorithm can use DTs to classify and extract key features of data (Chi et al., 2024). Before the BPNN predicts the data, the C5.0 algorithm is used to classify the data first and then input the required category data according to actual needs, thereby reducing the computation time and resource consumption of the BPNN. The basic process of C5.0 DT for data classification is shown in Fig. 2.

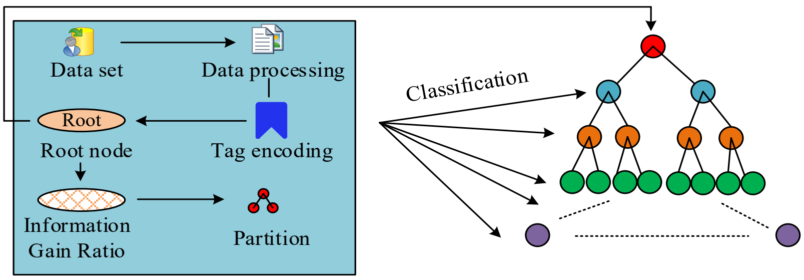


Fig. 2. Basic process of C5.0 DT classification

As shown in Fig. 2, after collecting a large amount of data, the algorithm first checks for missing, abnormal, and erroneous values in the data and processes them. Missing values are imputed using the mean, and outliers and erroneous values are deleted. The processed data is then tagged for subsequent classification by the DT. After the data is prepared, the entire training dataset is used as the root node of the DT. Starting from this node, the information gain ratio of each feature is calculated, and the feature with the highest information gain ratio is selected as the partition feature of the current node. Based on the selected partition features, the current node's dataset is divided into multiple subsets, each corresponding to a branch of the DT. Feature selection and data partitioning are repeated for each subset to construct the sub-nodes of the decision tree until all samples in the node belong to the same category. Through the above process, large-scale datasets can be divided into different types of small datasets, and suitable datasets on the DT can be selected according to actual needs and input into the BPNN for prediction. In this process, the formula for calculating the information gain ratio when the DT classifies data is shown in Eq. (4).

$$IG(A) = IG(D|A)S(A) \tag{4}$$

In Eq. (4), $IG(A)$ refers to the information gain ratio of feature A, $IG(D|A)$ represents the information gain, $S(A)$ is the classification information metric of feature A, and D is the raw data. The calculation formula for $IG(D|A)$ is denoted in Eq. (5).

$$IG(D|A) = H(D) - H(D|A) \tag{5}$$

In Eq. (5), $H(D)$ denotes the original data entropy, and $H(D|A)$ denotes the conditional entropy, which is the weighted entropy of each subset after the dataset is split. The calculation of $H(D)$ is shown in Eq. (6).

$$H(D) = -\sum_{i=1}^{k} pi \log_2(pi) \tag{6}$$

In Eq. (6), pi represents the probability of category i, and k represents the total number of categories. By calculating the information gain ratio of each data feature as described above, it can be classified accordingly. The basic process of the BPNN algorithm optimized by C5.0 DT algorithm is shown in Fig. 3.

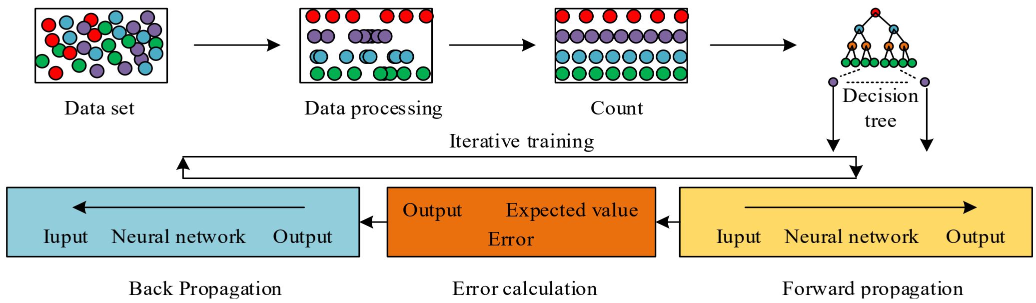


Fig. 3. Basic process of C5.0-BP algorithm

As shown in Fig. 3, when inputting a large-scale dataset, the C5.0 DT algorithm first classifies the large-scale dataset and assigns datasets of the same category to the same branch of the DT. When the BPNN predicts data, it selects an appropriate dataset to input into the BPNN according to actual needs, which simplifies the calculation process of the Neural Network (NN) and eliminates the interference of other data on the prediction results, ensuring the prediction effect.

3.2. Enterprise Budget Control Model Based on C5.0-BP Algorithm

The C5.0-BP algorithm can classify data in large-scale datasets and predict the patterns of data changes. In enterprise budget control, it is necessary to first predict the budget of the enterprise. The C5.0-BP algorithm can predict large-scale datasets in enterprises, thereby helping managers to control the budget of the enterprise. The basic structure of the prediction model based on C5.0-BPNN is shown in Fig. 4.

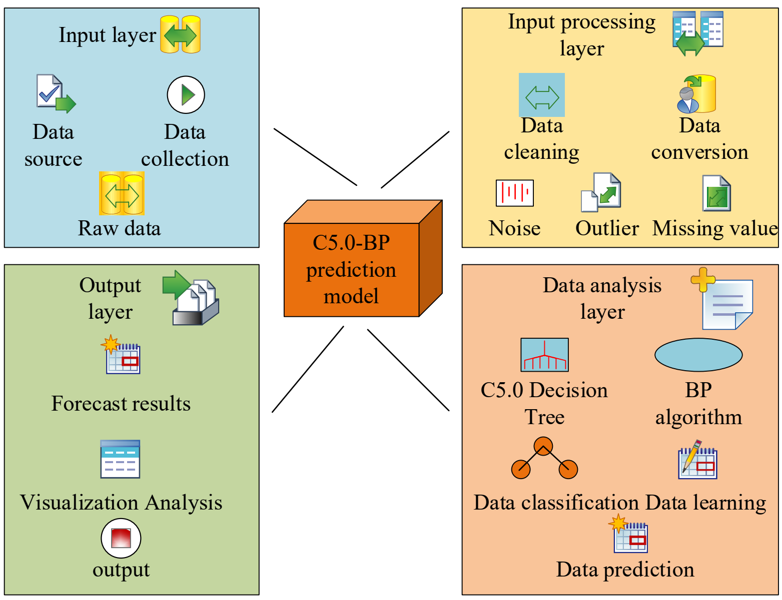


Fig. 4. Basic structure of C5.0-BP prediction model

As shown in Fig. 4, the model contains an input layer, a data processing layer, a data analysis layer, and an output layer. The input layer is responsible for collecting and predicting raw data related to the target from various data sources. The collected raw data is then cleaned and converted to remove noise, outliers, and missing values, and normalized to facilitate processing by the prediction model. The data processing layer is responsible for selecting features from preprocessed data. In this layer, the filtering method selects features by calculating the Pearson correlation coefficient of the data. The calculation formula is shown in Eq. (7).

$$r = \frac{\sum_{m=1}^{M}(x_m - \bar{x})(y_m - \bar{y})}{\sqrt{\sum_{m=1}^{M}(x_m - \bar{x})^2 \sum_{m=1}^{M}(y_m - \bar{y})^2}} \tag{7}$$

In Eq. (7), r represents the Pearson correlation coefficient, x_m represents the m th independent variable, $\bar{x}$ refers to the average value of the independent variable, y_m represents the m th dependent variable, $\bar{y}$ means the average value of the dependent variable, and M means the sample size, which is the total number of data points. The feature data with the highest absolute value of the correlation coefficient is selected. After feature selection, the selected features are input into the data analysis layer. In this layer, the data first enters the C5.0 DT, and the feature data is classified through the DT to divide the large-scale dataset into various sub-datasets. Then, each sub-dataset is utilized as input data for the BPNN. In the BPNN, each sub-dataset learns the change patterns of various sub-datasets through forward propagation, error calculation, backpropagation, and iterative training, obtaining the data change patterns and predicting the future trend of data changes. The final output layer is the prediction result of the output data analysis layer. Applying this model to enterprise budget forecasting, through the predicted results, enterprise managers can plan enterprise funds and ensure the safety of enterprise funds. The basic process of the enterprise budget control model based on the C5.0-BP prediction model is shown in Fig. 5.

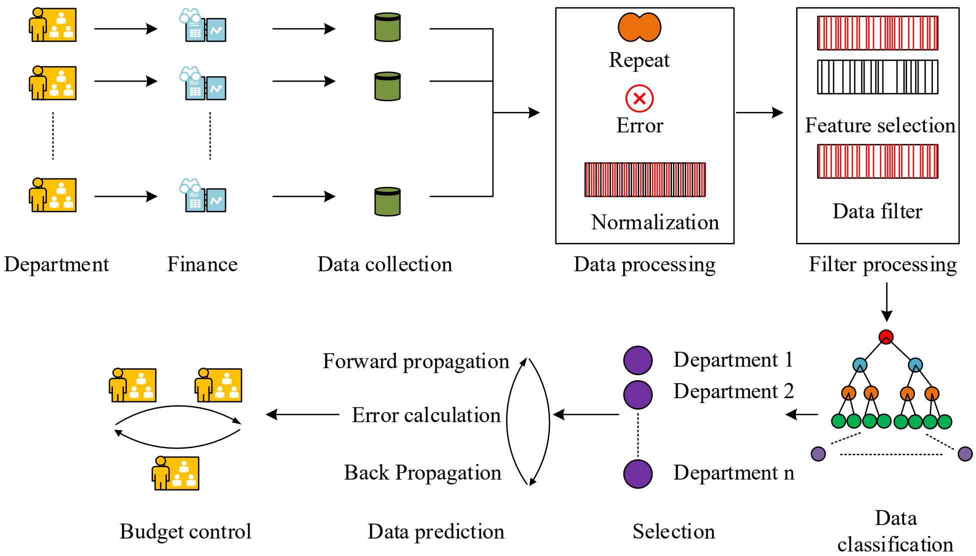


Fig. 5. Basic process of enterprise control model

In Fig. 5, when the control model controls the enterprise budget, it first collects data related to the enterprise budget from multiple data sources, such as financial systems and business systems of various departments within the enterprise, for example, actual income and expenditure data, market environment data and historical income and expenditure data. The collected data are cleaned. The duplicate, erroneous, and invalid data are removed. Finally, the data are normalized. The calculation formula is shown in Eq. (8).

$$X' = \frac{X - \bar{X}}{std} \tag{8}$$

In Eq. (8), X' refers to the normalized data, X represents the raw data, $\bar{X}$ represents the mean of the raw data, and std represents the standard deviation of the raw data. After processing the data related to the enterprise, the filtering method is used to select features from the processed data that are not related to the enterprise budget in the past. Afterwards, the C5.0 DT classifies the feature data, such as by department, budget purpose, or time, to divide the large-scale dataset into multiple small datasets. Based on the actual budget requirements of the enterprise, the sub-dataset is input into the BPNN for prediction. For example, to predict the budget of a certain department, the relevant sub-datasets of each department are input into a BPNN for calculation, thereby obtaining the budget prediction result of that department. The budget is adjusted based on the predicted results and the actual situation of the enterprise. For example, when the actual budget of a department in a company is less than the actual budget, funds will be allocated from the department whose actual budget is greater than the predicted budget, so that the actual budget of the entire company is the same as the predicted budget. Finally, it is output. This method reduces the scale of data input for the BPNN and alleviates its computational burden. It eliminates interference from other data to ensure prediction accuracy. In this process, the company budget needs to calculate the company's net profit, monthly savings, and risk tolerance to make more accurate predictions about the enterprise budget. The formula for calculating net profit is shown in Eq. (9).

$$NP = SV \times UP - VC + FC \tag{9}$$

In Eq. (9), NP represents the net profit of the enterprise, SV represents sales volume, UP represents unit price, VC represents variable cost, and FC represents fixed cost. The calculation formula for the monthly savings amount of the company is shown in Eq. (10).

$$MS = DI - E \tag{10}$$

In Eq. (10), MS represents monthly savings, DI represents disposable income, and E represents consumer spending. The calculation formula for the company's risk tolerance is shown in Eq. (11).

$$RT = \frac{EI - LI}{EI} \times 100\% \tag{11}$$

In Eq. (11), RT represents risk tolerance, EI represents target income, and LI represents the minimum income. Based on the above calculations, more accurate budget predictions can be made for the company.

3.3. Manager Decision Process Refactoring and Model Interaction Design

When managers plan the enterprise budget, their traditional budget decision-making process has obvious experience driven and post response characteristics. The budget gathering process consists of an annual preparation stage, a monthly regulation and calculation stage, and a risk reserve stage. Among them, the annual preparation adopts the "base method", and each department submits incremental declarations based on the budget limit of the previous year. After manual review by the finance department and management approval, a rigid annual budget is finalized, making it difficult to dynamically adjust throughout the year. Monthly regulation is a post-response mode, where departments submit adjustment requests after overspending. After verification by finance and special approval by management, the average time is over two weeks, missing the best opportunity for correction. Risk reserves are extracted and centrally controlled based on a fixed proportion of income. In times of crisis, fund allocation requires multiple levels of approval, making it difficult to cope with unexpected situations. The entire process relies on manual experience judgment and lacks data-driven predictive support and real-time monitoring mechanisms. The C5.0-BP model will reconstruct the three core budget decision-making processes of enterprises. After reconstruction, in the annual preparation stage, managers upload historical data and market parameters, and the model outputs departmental budget ranges and multi-scenario plans. After negotiation, the annual total budget is determined, and a monthly deviation tolerance threshold is set. During the monthly adjustment phase, the model automatically monitors the deviation between actual expenditures and forecasts, triggering a three-level warning mechanism. A yellow warning is pushed to departments for self-inspection, while a red warning triggers cross departmental resource allocation algorithms to generate fund transfer plans. Managers manually adjust within a ± 10% range before executing them. In the risk reserve stage, the model calculates each department's reserve ratio using differentiation and simulates crisis scenarios. When the crisis indicator exceeds the threshold, the reserve will be automatically released, and the manager will approve it afterward. Through the collaborative mechanism of "algorithm suggestion manual decision-making", managers approve major deployments at the strategic level, participate in departmental negotiations at the tactical level, and supervise data quality at the operational level, which not only improves decision-making efficiency but also retains management discretion.

4. Analysis of the Effect of Precise Budget Control in Enterprises Based on C5.0-BP Algorithm

4.1. Performance Analysis of C5.0-BP Algorithm

To analyze the effectiveness of the proposed algorithm, the classification and prediction performance of the C5.0-BP algorithm were studied through the following experimental environment configuration. The experimental environment configuration during the experiment is denoted in Table 1.

Table 1. Experimental environment configuration

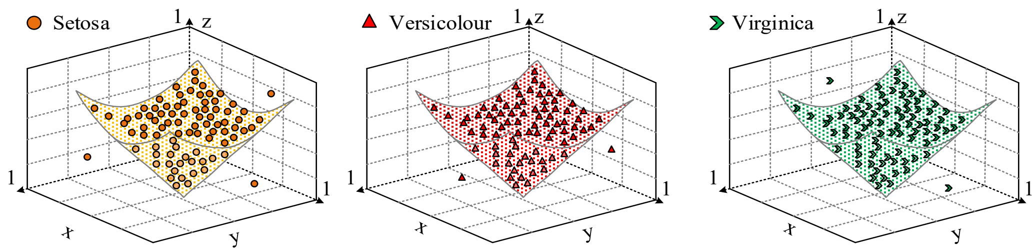
Project	Index	Allocation
Hardware environment	CPU	Intel Core i7-12700K
	GPU	NVIDIA RTX 309
	RAM	64GB DDR4
	OS	Windows 10
Software environment	Deep learning framework	PyTorch 1.12
	ML library	Scikit-learn 1.1
	Data analysis software	Python 3.9

The Iris dataset was chosen as the experimental dataset. This dataset included 150 samples, divided into three categories (Setosa, Versicolour, Virginica), each with 50 samples and four features. The study divided the dataset into a training dataset and a validation dataset in a 7:3 ratio to test the performance of the algorithm. The classification performance of three types of data and the prediction performance of the data were analyzed based on features. During the experiment, the learning rate of the BPNN was set to 0.001, the number of iterations was set to 500, the batch size was set to 100, and the confidence threshold of the C5.0 DT was set to 0.25. The Iris dataset contained 150 iris samples, divided into three categories: Setosa, Versicolour, and Virginica. Four features of sepal length, sepal width, petal length, and petal width were selected as input data. The Iris dataset is shown in Table 2.

Table 2. Sample iris dataset

Sample ID	Calyx length (cm)	Calyx width (cm)	Petal length (cm)	Petal width (cm)	Category label (output)
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
3	4.7	3.2	1.3	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa
5	5.4	3.9	1.7	0.4	Setosa
6	7.0	3.2	4.7	1.4	Versicolour
7	6.4	3.2	4.5	1.5	Versicolour
8	6.9	3.1	4.9	1.5	Versicolour
9	5.5	2.3	4.0	1.3	Versicolour
10	6.5	2.8	4.6	1.5	Versicolour
11	6.3	3.3	6.0	2.5	Virginica
12	5.8	2.7	5.1	1.9	Virginica
13	7.1	3.0	5.9	2.1	Virginica
14	6.3	2.9	5.6	1.8	Virginica
15	6.5	3.0	5.8	2.2	Virginica

According to Table 2, the input data of the Iris dataset consists of four numerical features: sepal length, sepal width, petal length, and petal width, all measured in centimeters. The output data are category labels, consisting of three categories: Setosa (mountain iris), Versicolour (color changing iris), and Virginica (Virginia iris), with 50 samples in each category. In the C5.0-BP algorithm, the first four columns of features are used as input vectors for the neural network. After pre-classification by the C5.0 decision tree, they are input into the additional momentum BPNN, and the output is the probability distribution of each category. The category corresponding to the highest probability is taken as the prediction result. Through the above settings, the algorithm was tested. First, the classification performance of the C5.0-BP algorithm on three types of samples in the Iris dataset was tested. The results are shown in Fig. 6.



(a) Setosa classification effect (b) Versicolour classification performance (c) Virginica classification effect

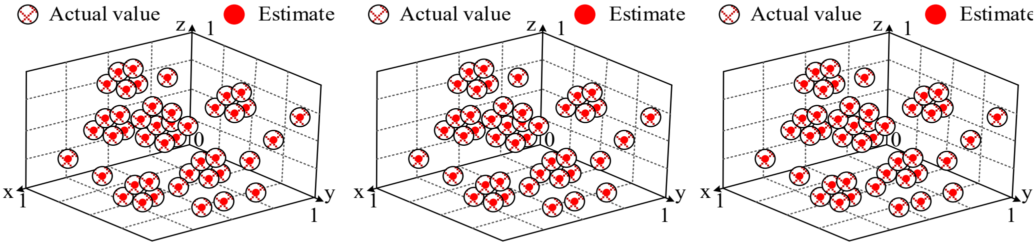
Fig. 6. Algorithm classification effect analysis

In Fig. 6(a), when using the C5.0-BP algorithm to classify various types of data in the Iris dataset, the algorithm performed well in classifying Setosa class data in the dataset. Most Setosa data points could be classified into this class, and only a very small number of data points were misclassified. From Fig. 6(b) and 6(c), the algorithm performed well in classifying Versicolour and Virginica data in the Iris dataset, with only a very small number of data points being misclassified. From this result, the algorithm has good classification performance and can accurately classify the data in the dataset. The predictive performance of the C5.0-BP algorithm was then analyzed. The C5.0-BP algorithm was employed to analyze the features in the dataset to predict the types of data in the dataset. The findings are denoted in Fig. 7.

As shown in Fig. 7(a), when the algorithm predicted Setosa data in the dataset, the predicted values were the same as the actual values, indicating that the algorithm had a high prediction accuracy. From Fig. 7(b) and 7(c), the algorithm performed well in predicting Versicolour and Virginica class data in the dataset. Finally, the prediction time and prediction error of C5.0-BP algorithm were analyzed. The findings are indicated in Fig. 8.

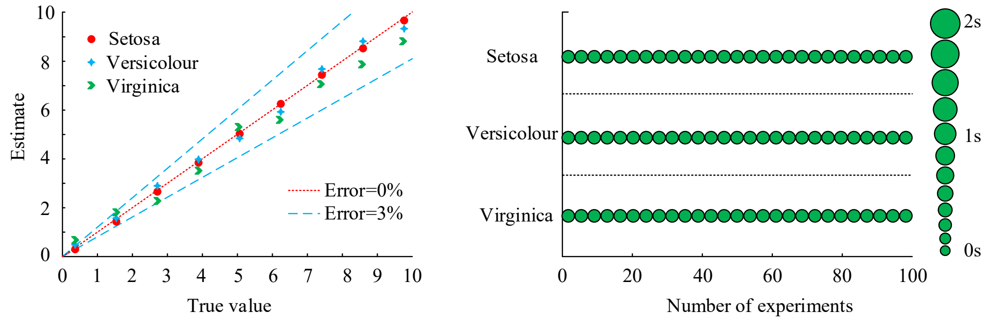
As shown in Fig. 8(a), when predicting the three types of data in the dataset, the algorithm had low prediction errors for Setosa, Versicolour, and Virginica, which were 1.2%, 0.9%, and 1.1%, respectively, with prediction errors all below 3%.

According to Fig. 8(b), when using this algorithm to predict three types of data, the prediction time was relatively short, with less than 1 second per data point. From the above analysis results, the proposed C5.0-BP algorithm can accurately classify and predict data, thereby providing support for the prediction and control model of enterprise budgets.



(a) Setosa prediction performance (b) Versicolour prediction performance (c) Virginica's prediction performance

Fig. 7. Algorithm prediction performance analysis



(a) Algorithm error rate (b) Algorithm time consumption

Fig. 8. Prediction error and prediction time

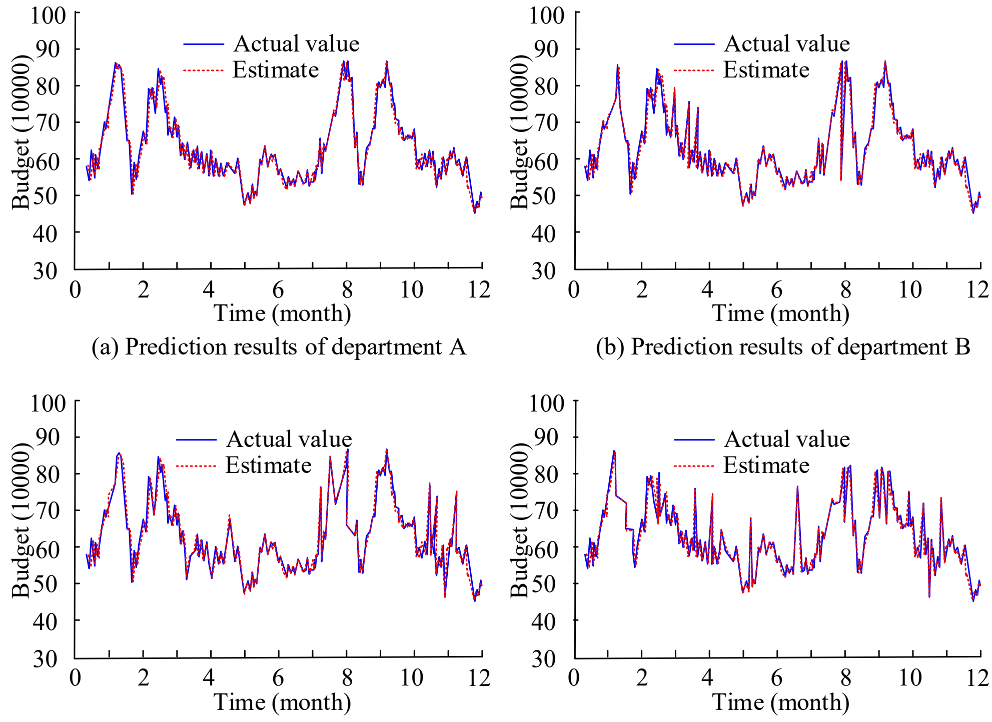
4.2. Analysis of Effectiveness for Enterprise Budget Control

After verifying the performance of the C5.0-BP algorithm, the actual impact of the enterprise budget control model constructed based on this algorithm was analyzed. For the analysis, the publicly available dataset of urban planning enterprise budgets on the Kaggle platform was selected to test methods of enterprise budget control. The Kaggle Enterprise Budget Dataset is shown in Table 3.

Table 3. Sample Kaggle enterprise budget dataset

Sample ID	Department ID	Month	Historical revenue and expenditure (10000 yuan)	Number of projects	Personnel scale (person)	Market Environment Index	Monthly total budget (RMB 10000)	Net profit (RMB 10000)	Monthly savings amount (10000 yuan)	Risk tolerance score
1	A	1	245.6	12	45	0.82	268.3	32.5	18.2	0.75
2	A	2	312.8	15	45	0.85	356.7	28.3	15.6	0.72
3	A	3	298.4	14	46	0.81	334.2	30.1	17.3	0.74
4	B	1	156.3	8	32	0.78	172.5	22.8	12.5	0.68
5	B	2	148.7	8	32	0.76	165.4	24.1	13.2	0.70
6	B	6	162.4	9	33	0.80	178.9	23.5	12.8	0.69
7	C	1	189.5	10	38	0.79	208.6	19.2	10.5	0.65
8	C	6	142.3	7	38	0.74	158.7	15.8	8.3	0.58
9	D	1	223.4	11	42	0.83	245.8	27.6	15.1	0.71
10	D	8	356.7	18	45	0.88	398.4	31.2	16.8	0.73
11	D	9	378.2	19	46	0.89	421.5	33.5	18.4	0.76
12	A	3	325.6	16	47	0.86	368.9	29.7	16.5	0.73

According to Table 3, this dataset contained various financial statements from various departments of the enterprise from 2019 to 2020. Four departments, namely the security department (A), health department (B), transportation department (C), and education department (D), were selected as the test dataset, and they were still divided into training and testing datasets in a 7:3 ratio. This dataset contained monthly operational data for four departments (Safety A, Health B, Transportation C, Education D) of a city planning enterprise from 2019 to 2020. Six dimensional features, including department ID, month, historical revenue and expenditure, project quantity, personnel size, and market environment index were selected as input data. After filling in missing values through mean interpolation, removing outliers through the 3 σ principle, and normalizing Z-score, they were pre-classified into four homogeneous subsets by department type using C5.0 decision tree. Subsequently, each subset is input into an additional momentum BPNN, which outputs four types of results: monthly total budget forecast, net profit forecast, monthly savings forecast, and risk tolerance score. By analyzing the actual effectiveness of the enterprise budget control model based on the above dataset, the prediction effect of the total budget of each department was first analyzed. The findings are denoted in Fig. 9.



(a) Prediction results of department A (b) Prediction results of department B

(c) Prediction results of department C (d) Prediction results of department D

Fig. 9. Analysis of the effectiveness of the total budget forecast for each department

According to Fig. 9(a), the C5.0-BP enterprise budget control model had a good predictive effect when analyzing the total budget of Department A. The actual total budget from January to December was the same as the predicted total budget. From Fig. 9(b), 9(c), and 9(d), when the model predicted the total budget of departments B, C, and D, the actual and predicted values were the same, indicating good prediction performance. And based on the actual situation in the dataset, from January to December 2019, Fig. 9(a) shows that Department A experiences a budget peak in the second to third months. Under the traditional budget process, this seasonal surge often leads to approval delays. C5.0-BP can predict the peak in advance, and managers can initiate emergency procurement fund approval in advance based on the predicted results, avoiding safety projects from being delayed due to budget approval. Fig. 9(c) shows that Department C falls into a budget trough in June, while Fig. 9(b) indicates that Department B is in a stable period during the same period. The model identifies the difference through real-time monitoring and triggers a cross departmental fund pool allocation mechanism, which can temporarily transfer the idle budget of department B to department C for the current month, filling the gap in transportation maintenance funds. This dynamic balance of "peak shaving and valley filling" cannot be achieved in traditional static budgeting, often resulting in the C department being forced to delay infrastructure maintenance. Fig. 9(d) shows that Department D experiences its highest peak of the year from August to September, forming resource competition with Department A during the same period. The model automatically generates multi-scenario deployment schemes through deviation tolerance interval design. On the premise of ensuring the security needs of Department A, some non-emergency expenses of Department D are smoothed out to the 10th month to avoid internal organizational friction caused by two key departments competing for budgets at the same time. The accuracy of predicting the net profit, monthly savings, and risk tolerance of each department is shown in Fig. 10.

As shown in Fig. 10 (a), when using the C5.0-BP budget control model to predict the net profit of the enterprise, the accuracy of the model in predicting the net profit of departments A, B, C, and D within the enterprise was 96.5%, 97.3%, 97.8%, and 96.7%, respectively, indicating a high prediction accuracy. As shown in Fig. 10(b), the model had a high prediction accuracy of over 95% when analyzing the monthly savings of various departments in the enterprise. From Fig. 10(c), when the model predicted the future risk tolerance of various departments in the enterprise, its prediction results were also good, exceeding 95%. From the actual situation, Fig. 10(a) shows that the net profit of Department B slightly decreases in the fourth month, indicating that the department has a one-time equipment procurement expenditure for that month. In traditional decision-making, such fluctuations may trigger economic contraction in the enterprise, but the model is inconsistent and marks this date as a short-term anomaly. Management may not adjust the annual strategy based on this decision. From the above results, the C5.0-BP enterprise budget control model can predict various data of various departments in the enterprise. Finally, to prove the superiority of the budget control model, the model was compared with the Genetic Algorithm-Back Propagation Algorithm (GA-BP) budget control model, the K-means-Convolutional Neural Network (K-means-CNN) budget control model, and the Least Absolute Shrinkage and Selection Operator-Support Vector Regression (Lasso-SVR) budget control model. The data from department A were selected as the experimental dataset for comparative analysis. The four models were used to forecast the various budgets for Department A for January to March, and their average values were taken as the final results, as denoted in Table 4.

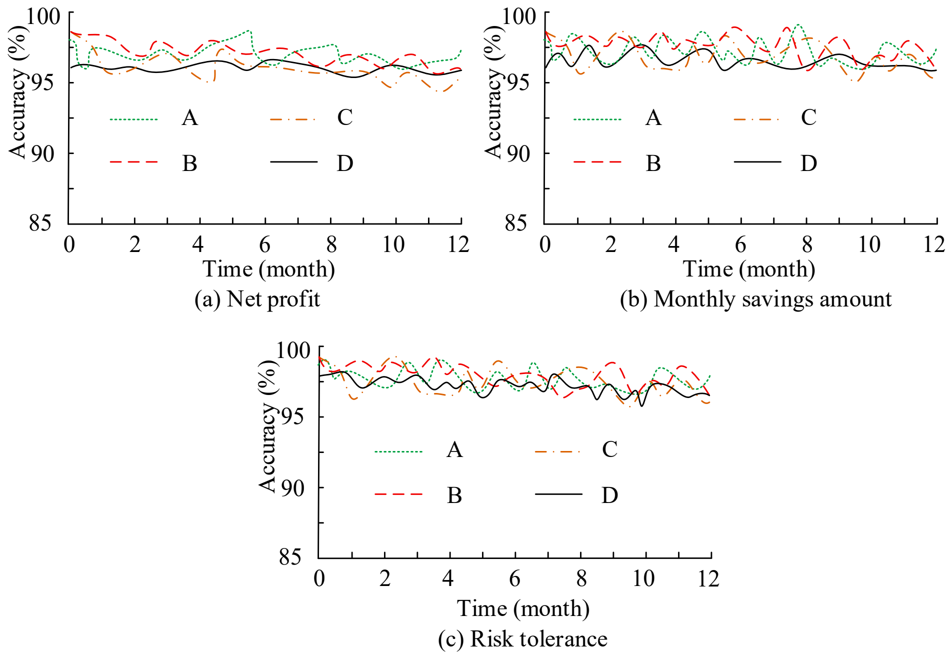


Fig. 10. Prediction accuracy analysis

Table 4. Comparison of model prediction accuracy

Model	Net profit	Monthly savings amount	Risk tolerance	General budget
C5.0-BP	98.4±0.9%	98.7±1.2%	97.4±1.3%	98.6±0.5%
GA-BP	90.5±1.2%	92.1±1.4%	89.7±2.4%	90.3±1.8%
K-means-CNN	83.9±1.6%	84.6±1.7%	83.8±1.9%	82.7±2.2%
Lasso-SVR	79.9±0.7%	80.2±2.9%	79.8±1.7%	79.3±2.1%

According to Table 4, when analyzing the performance of the four models using data from department A, the C5.0-BP model had the best prediction performance. The model could achieve an accuracy of 98.4% ± 0.9% in predicting the net profit of department A from January to March, which was higher than the GA-BP model's 90.5 ± 1.2%, the K-means-CNN model's 83.9 ± 1.6%, and the Lasso-SVR model's 79.9 ± 0.7%. The C5.0-BP model achieved over 97% accuracy in predicting monthly savings, risk tolerance, and total budget, whereas the other three models achieved below 95% accuracy. From this result, the C5.0-BP model can issue warnings and generate departmental budget plans based on its accurate predictions, and there will be sufficient time in management to complete various budget fund controls. The GA-BP model has an error rate of approximately 10%, leading to delayed and inaccurate warnings and forcing managers to urgently approve and pay fees during peak periods. However, Lasso SVR cannot accurately predict the budget peak due to a 20% error, which will lead to a shortage of funds in the department and force the suspension of the project, resulting in hidden security risks. Based on the above results, the C5.0-BP budget control model proposed in the study can accurately predict enterprise budgets, enabling precise budget control. To further verify the performance of the C5.0-BP enterprise budget control model in a real enterprise environment, this study selected different enterprises in a certain city for actual testing. Financial statement data and departmental income and expenditure details from manufacturing, retail trade, and technology service enterprises were selected for the dataset from 2019 to 2020. The C5.0-BP enterprise budget control model was tested for its accuracy in predicting net profit, monthly savings, and company risk tolerance in different enterprises. The results are shown in Table 5.

According to Table 5, the C5.0-BP model can achieve a prediction accuracy of over 96% in different types of enterprises, indicating good predictive performance.

Table 5. Prediction accuracy

Enterprise type	Net profit (%)	Monthly savings amount (%)	Wind direction tolerance (%)	Average accuracy (%)
Manufacturing	97.2±0.8	96.8±1.1	96.5±1.3	96.8±1.0
retail trade	96.4±1.2	97.5±0.9	95.9±1.6	96.6±1.2
Technology services	98.1±0.6	97.3±1.0	97.8±0.8	97.7±0.8

5. Conclusion

In response to the large dataset size and poor budget prediction performance in the current enterprise budget precision control model, which leads to unsatisfactory budget control effects, this study used C5.0 and BP algorithms to construct a prediction model, and based on this model, controlled the enterprise budget to improve its prediction accuracy and control effect. The study first analyzed the classification and prediction effects of the C5.0-BP algorithm on the data. The results showed that the algorithm had high classification accuracy for three types of data in the Iris dataset, and had low prediction errors for Setosa, Versicolour, and Virginica data, which were 1.2%, 0.9%, and 1.1%, respectively. The prediction time was also less than 1 second. Furthermore, comparing the C5.0-BP budget control model with the GA-BP budget control model, the K-means-CNN budget control model, and the Lasso-SVR budget control model, the C5.0-BP model had much higher prediction accuracy for various indicators in enterprises than the other budget control models. From the above analysis results, the proposed method for precise control of enterprise budgets can accurately predict the budgets of various departments of the enterprise, thereby enabling precise control of the budget based on the predicted results. This study established a closed-loop mechanism of “prediction control”. The model output is not only used as reference information but can trigger specific control actions. For example, when the model detects a gap between the actual budget of department C and the predicted budget, while the actual budget of department A is lower than the predicted budget and there is a surplus during the same period, the fund allocation mechanism is automatically triggered to transfer the budget from department A to department C, so that the actual budget and predicted budget of the entire company tend to be consistent. This conversion mechanism effectively transforms prediction accuracy into resource allocation efficiency, avoiding the disconnection dilemma of traditional budgeting between prediction and execution. According to the model, the manager will reconstruct three core decision-making processes. In the annual budget preparation stage, strategic choices are made based on the departmental budget range and multi-scenario plans output by the model, with a significantly shortened preparation cycle and dynamic adjustment capability. During the monthly budget adjustment phase, routine adjustments are automatically processed by algorithms, and managers focus on abnormal decisions, resulting in a significant improvement in response speed. In the risk reserve allocation stage, the model accurately allocates reserves according to the risk tolerance of each department, automatically releases funds in times of crisis, and balances fund efficiency and risk prevention. These three changes enable managers to achieve a balance between algorithmic efficiency and management discretion through strategic level approval of major deployments, tactical level participation in departmental negotiations, and operational level supervision of data quality. However, the C5.0 DT algorithm used in this study is sensitive to missing values in the dataset, which may affect the prediction performance. In the future, probability-based DT models can be adopted to consider the impact of the probability of missing values in classification results, thereby ensuring the predictive performance of data.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The author used Grammarly Premium solely for language editing and readability improvement. The author reviewed and verified all content and takes full responsibility for the accuracy and integrity of the manuscript.

References

Abbasov, R. (2025). Government Budgeting and Expenditure: A Multifaceted Analysis of Economic Growth, Fiscal Sustainability, and Social Impact. *iBusiness*, 17(1), 10.4236. https://doi.org/10.4236/ib.2025.171004

Chi, X., Bai, C., Yang, F., Qi, Q., Zhang, R., and Xu, H. (2024). Joint intra- and inter-channel nonlinear compensation scheme based on improved learned digital back propagation for WDM systems. *Optics Express*, 32(4), 22-34. https://doi.org/10.1364/OE.506995

Fathima, S. H., Susanto, B., Yuliani, N. L., Darmadi, R. A., and Utami, P. S. (2025). Determination analysis of village-owned enterprise sustainability. *Accounting and Finance Studies*, 5(1), 20-36. https://doi.org/10.47153/afs51.12682025

Hunjra, A. I., Bagh, T., Palma, A., and Goodell, J. W. (2024). Is enterprise risk-taking less sensitive to financial flexibility post COVID-19? Evidence from non-linear patterns. *International Review of Financial Analysis*, 95(5), 103432-103442. https://doi.org/10.1016/j.irfa.2024.103432

Khun-anod, K., Limsawasd, C., and Athigakunagorn, N. (2023). Predicting cost and schedule performance of green building projects based on preproject planning efforts using multiple linear regression analysis. *Journal of Architectural Engineering*, 29(3), 4023025-4023027. https://doi.org/10.1061/JAEIED.AEENG-142

Liu, Y., Liu, W., Lin, S., Yang, C., Huang, K., and Shen, Z. J. (2023). Data-driven raw material robust procurement for non-ferrous metal smelter under price and demand uncertainties. *IEEE Transactions on Automation Science and Engineering*, 21(4), 5852-5865. https://doi.org/10.1109/TASE.2023.3319390

Lu, X. (2024). Evaluation and improvement of enterprise budget performance based on multiple linear regression algorithm. *Informatica*, 48(6), 21-34. https://doi.org/10.31449/inf.v48i6.5510

Mahmud, M., Budiman, I., Indriani, F., Kartini, D., Faisal, M. R., Rozaq, H. A. A., and Caesarendra, W. (2024). Implementation of C5.0 algorithm using chi-square feature selection for early detection of hepatitis C disease. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 6(2), 116-124. https://doi.org/10.35882/jeeemi.v6i2.384

Okeke, N. I., Bakare, O. A., and Achumie, G. O. (2024). Forecasting financial stability in SMEs: A comprehensive analysis of strategic budgeting and revenue management. *Open Access Research Journal of Multidisciplinary Studies*, 8(1), 139-149. https://doi.org/10.53022/oarjms.2024.8.1.0055

Oyedokun, O., Ewim, S. E., and Oyeyemi, O. P. (2024). Leveraging advanced financial analytics for predictive risk management and strategic decision-making in global markets. *Global Journal of Research in Multidisciplinary Studies*, 2(2), 16-26. https://doi.org/10.58175/gjrms.2024.2.2.0051

Rezki, M. K., Mazdadi, M. I., Indriani, F., Muliadi, M., Saragih, T. H., and Athavale, V. A. (2024). Application of SMOTE to address class imbalance in diabetes disease classification utilizing C5.0, random forest, and SVM. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 6(4), 343-354. https://doi.org/10.35882/jeeemi.v6i4.434

Simon, K., Vicent, M., Addah, K., Bamutura, D., Atwiine, B., Nanjebe, D., and Mukama, A. O. (2023). Comparison of deep learning techniques in detection of sickle cell disease. *Artificial Intelligence Applications*, 1(4), 252-259. https://doi.org/10.47852/bonviewAIA3202853

Wang, P., Xu, J., Zhou, M., and Albeshri, A. (2023). Budget-constrained optimal deployment of redundant services in edge computing environment. *IEEE Internet of Things Journal*, 10(11), 9453-9464. https://doi.org/10.1109/JIOT.2023.3234966

Wang, R. (2024). Safeguarding enterprise prosperity: An in-depth analysis of financial management strategies. *Journal of the Knowledge Economy*, 15(4), 17676-17704. https://doi.org/10.1007/s13132-024-01752-z

Wang, T. (2024). Improved random forest classification model combined with C5.0 algorithm for vegetation feature analysis in non-agricultural environments. *Scientific Reports*, 14(11), 10367–10372. https://doi.org/10.1038/s41598-024-60066-x

Wu, X., Wu, X., and Huang, F. W. Y. (2024). Characterization of pepper ripeness in the field using hyperspectral imaging with back propagation neural network and kernel-based extreme learning machine models. *Analytical Letters*, 57(3), 409-424. https://doi.org/10.1080/00032719.2023.2210708

Xin, Q., Song, R., and Wang, Z. (2024). Enhancing bank credit risk management using the C5.0 decision tree algorithm. *Journal of Computer Technology and Applied Mathematics*, 1(4), 100-107. https://doi.org/10.5281/zenodo.14032041

Ye, J., Jiang, W., Yang, X., and Hong, B. (2023). Emergency materials response framework for petrochemical enterprises based on multi-objective optimization. *Energy*, 269(6), 126670–126682. https://doi.org/10.1016/j.energy.2023.126670

Yin, P., Xie, L., Zhang, H., Li, W., and Wang, W. (2024). Modelling wax deposition of diesel in sequential transportation of product oil pipeline using optimized back propagation neural network. *The Canadian Journal of Chemical Engineering*, 7(5), 102-114. https://doi.org/10.1002/cjce.25149

Zhang, Y. (2024). Financial optimization budget allocation model based on big data and data-driven technology. *Procedia Computer Science*, 247(6), 396-402. https://doi.org/10.1016/j.procs.2024.10.047

Li Chen graduated from Zhejiang Normal University in 2010, with a master's in Management and is a Chinese Certified Public Accountant (CPA). She is currently an associate professor at the School of Economics and Management of Anyang University. She has received numerous honors, including the Advanced Individual in "Three Educations" at Anyang University, the Advanced Individual in Private Education of Henan Province, the Advanced Individual in Teaching and Educating at Anyang University, and the Excellent Teacher at Anyang University. She has published articles in journals such as Communication of Finance and Accounting, presided over several projects, including the Educational Planning Project and the Employment and Entrepreneurship Project for Colleges and Universities of the Department of Education of Henan Province, and edited two textbooks. She has also guided students to win awards in various competitions. Her research interests include corporate finance, fiscal and taxation financial systems, industry-education integration, and innovation and entrepreneurship education.