Journal of Computers Vol. 37 No. 5, October 2026, pp. 373-389
https://doi.org/10.63367/19911599202610370502 3

# Optimization of Financial Budget and Design of Budget System for Vocational Colleges Based on Data Analysis

Kai-Hua Li[1,2*], Qiang Guo[1], Ya-Di Shi[1], and Xing-Jie Wang[3]

[1] Hebei Mechanical and Electrical Polytechnic,
Xingtai City 054000, Hebei Province, China
465116587@qq.com, guoqiang@hbjd.edu.cn, 502870507@qq.com

[2] Research Base for Digital-Intelligent Marketing Operation,
Xingtai City 054000, Hebei Province, China

[3] JinNiu Hotel of Jizhong Energy Co., Ltd,
Xingtai City 054000, Hebei Province, China
934385509@qq.com

*Received 10 July 2026; Revised 4 August 2026; Accepted 10 September 2026*

**Abstract.** The traditional financial budget management in vocational colleges is generally featured by inadequate data integration, weak analytical abilities, incorrect budget planning, and ineffective management of budget execution monitoring and dynamic adjustment. In order to solve the above problems, a data-driven financial budget management and optimization system in vocational colleges is designed in this study. The system takes into account budget planning, budget execution, monitoring, adjustment, and performance analysis into one management process. Multi-source financial and business data are extracted, cleaned, transformed and loaded into the financial data warehouse through the ETL process. The multidimensional and tabular analytical models are built for analyzing financial overview, financial income and expenditure, balance analysis, comparison analysis, financial risk warning and business management. Moreover, an ensemble decision tree regression model with randomized feature selection and split point generation is proposed to predict major financial indices. In particular, quarterly final-account data and annual budget data of a vocational college are used to evaluate the proposed model. The experimental results show that the prediction accuracies of the six major financial indices including operating revenue, operating cost, operating profit, net cash flow, accounts receivable and accounts payable vary from 85.8% to 97.5%. Meanwhile, the corresponding cross validation scores range from 0.90 to 0.96. As can be seen from the learning curve results, the model has relatively stable generalization ability with the increase of training sample sizes. Therefore, the proposed system can help realize integrated financial data management, multidimensional visualization analysis, financial budget forecast and dynamic adjustment of budgets.

**Keywords:** machine learning, budget optimization, system design

## 1 Introduction

In recent years, China's higher education industry has continued to transform and upgrade, entering a new stage of leapfrog development. In 2022, the total investment in national fiscal education funds exceeded 6 trillion yuan for the first time, a year-on-year increase of 5.97%. Although the national finance has gradually increased the allocation of education funds, the balance of resource allocation and the efficiency of education fund utilization in universities have not been unified, and the contradiction between supply and demand of education funds caused by large-scale expansion still exists widely. In December 2024, the Central Economic Work Conference once again proposed that party and government organs should habitually tighten their belts and optimize the use of fiscal funds [1]. The report of the National People's Congress and the Chinese People's Political Consultative Conference in March 2025 proposes that local governments should deepen the reform of zero based budgeting, strengthen budget constraints and performance management. With the transformation of the macroeconomic environment, the extensive and simple mode of using university funds has become a thing of the past, and budget

* Corresponding Author

management under careful calculation, strict control and supervision has become the trend. In response, domestic universities are continuously promoting budget management and internal control construction [2].

How to solve the contradiction between supply and demand of funds, allocate resources reasonably and effectively has become an issue that almost all universities cannot ignore. Therefore, each university needs to make corresponding budgets for its future development plan every year. The budget is the overall planning of the university's demand and use of funds, resource allocation, and tuition income [3].

Currently, the external market environment of universities has changed, and budget management and internal control are one of the six major components of internal control in universities. The new "Implementation Regulations of the Budget Law" establish the legal status of internal control in budget execution, and the "Financial System of Higher Education Institutions" also strengthens the significance of control and supervision in the budget process. The two are closely connected and interact with each other. Examining the budget from an internal control perspective can improve the efficiency of budget fund utilization, reduce fund allocation risks, and improve the budget performance evaluation system. Therefore, universities should start from reality, analyze external environmental trends, optimize budget management, implement effective budget internal control, safeguard sustained high-quality development, and contribute to the smooth reform of the financial system [4].

In the current budget management system of universities, various departments lack effective constraints during the budget declaration stage, often resulting in overestimation of budget requirements. In the budget preparation stage, the scientificity and rationality of budget preparation are difficult to fully guarantee. In the budget execution stage, the standardization of the budget execution process is insufficient, and the budget adjustment procedures are not rigorous enough. Some departments have the phenomenon of arbitrarily adjusting budgets and increasing budget expenditures to complete the execution progress. At the same time, the existing internal control of university budget management has not played an effective role in budget constraints, which cannot effectively constrain and supervise the operation of university funds. Specifically, the main issues are as follows:

1) The budget preparation method is not scientific enough. In vocational colleges, the budget preparation mainly adopts the zero based budget method and incremental budget method. When implementing, rigid expenditures such as personnel salaries, water and electricity expenses, etc. are prioritized to maintain the basic operating expenses of the school. The promotion of the zero based budget method is difficult, mainly because it will affect the budget interests of some units. Therefore, the incremental budget method still exists [5].

2) There are deficiencies in the budget preparation review, and there is a lack of professional technical demonstration and comprehensive benefit demonstration during the review.

3) The supervision of budget execution is not strict, and there is generally a budget execution control process, but there are weak points in some links, and the control effect is somewhat vague, resulting in lax supervision.

4) The budget performance evaluation system is not scientific. Firstly, the quantification of performance evaluation indicators is insufficient, resulting in relatively rough evaluation results. Secondly, budget performance evaluation is mainly completed by departments through self-evaluation, which may lead to inaccurate and formalistic evaluation results. Finally, there is a lack of reward and punishment system for budget performance evaluation.

Therefore, this article focuses on the current situation of financial budget management in vocational colleges. Firstly, a budget management system framework is designed, and then big data analysis and reinforcement learning models are added to the system to improve the ability of budget management and process optimization.

## 2 Related Work

This chapter summarizes the relevant research results on university financial systems, some of which have conducted innovative research in the intelligent reimbursement process, some have innovated in intelligent budgeting, and combined with artificial intelligence technology. In terms of technological innovation in reimbursement systems, Yaoke Yang proposed a project-based blockchain university reimbursement platform model. The platform is based on the dynamic role of participants in a multi project environment and introduces an attribute based access control model to achieve fine-grained permission management. At the same time, to solve the problem of complex logical relationships between projects and invoices, a dual chain storage structure of main chain and secondary chain and corresponding logical transaction algorithms were designed to solve the complex correspondence between different reimbursement states and projects. Considering the query efficiency of the reimbursement process, a Merkle tree index table query algorithm was designed to improve the overall reimbursement efficiency of the system [6].

Yingnan Wu's research field is budget optimization in the construction process, using deep learning engineering cost prediction models. The model combines data preprocessing, feature extraction, and model optimization techniques to construct a deep learning framework including convolutional neural networks (CNN), long short-term memory networks (LSTM), and their hybrid models. At the same time, an experimental platform was built to verify that the framework can effectively process the spatial and temporal characteristics of engineering cost data, improve prediction accuracy and efficiency, and enhance the applicability and effectiveness of the model in different engineering fields. This provides a new technical path and theoretical support for the intelligent development of engineering cost management, and its core framework structure has reference significance for the construction of this framework [7].

Yang Xu, starting from the development history, system architecture and system construction of smart finance, systematically elaborated how to use network technology as the carrier, third-party settlement platforms (banks, UnionPay, WeChat and Alipay, etc.) as the support, industry finance integration as the direction, data sharing as the basis, unify the information portal entrance, unify the information identification code, develop shared data interfaces, embed artificial intelligence modules, and integrate financial businesses such as financial management rules, accounting systems, and accounting processing specifications with artificial intelligence (AI) The deep integration of information technologies such as mobile payment and blockchain enables the networked filling of reimbursement documents, automatic matching of shared data systems, artificial intelligence conversion of accounting vouchers, electronic payment methods, and automation of reimbursement logistics information. It builds a new intelligent financial reimbursement management system for universities that is informationized, networked, and intelligent, with human-computer interaction. It innovates financial management methods and improves the level and efficiency of financial management in universities [8].

In terms of budget management, Chunhui Zhang aims to manage hospital financial budgets and proposes the design of a public hospital financial budget information management system based on principal component analysis (PCA). Design the overall framework of the financial budget information management system for public hospitals, and analyze the performance of various modules in the financial budget information management system; Analyze financial data through PCA, complete system database design, and design software program implementation process based on specific hardware module functions; Integrating the designed hardware and software to complete the design of a financial budget information management system for public hospitals, its handling of budget methods has reference significance for the budget processing methods in this article [9].

## 3 System Design and Model Construction

The original budget management system for higher vocational education based on numerical analysis designed in this article is shown in Fig. 1.

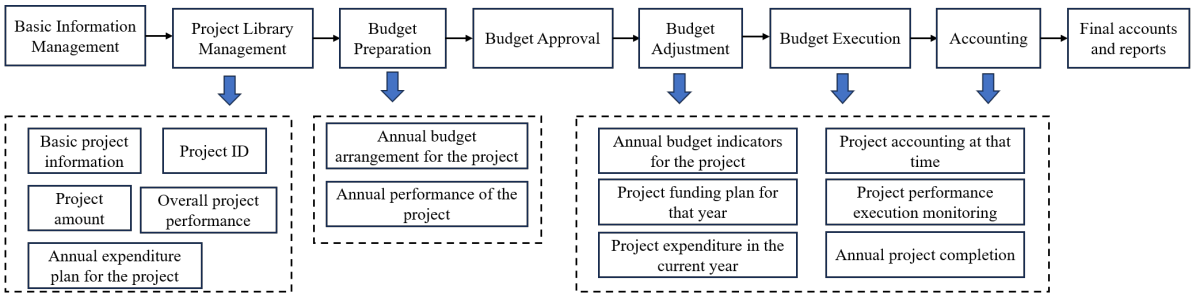



**Fig. 1.** Overall system architecture

The financial management system constructed in this article constructs a multidimensional budget classification model, implements refined classification management of income and expenditure projects in universities, and configures budget declaration permissions to each responsible unit as needed. The main research object of this article is vocational colleges. Based on the current internal management situation, when preparing the budget, the finance department first predicts the annual funding needs of the university based on the business

parameters of functional departments and the budget requirements reported by each functional department. A benchmark plan is generated for the budget unit to confirm [10]. Then, the budget unit synchronously fills in the income and expenditure forecast table in the system (the specific management measures of each university are different, and there are differences in the forms that need to be filled in). After the initial review by the business supervisory department, it is submitted to the finance department for summary, forming a school wide budget draft and submitting it to the school level decision-making. After the budget takes effect, an indicator ledger is generated, and the budget is automatically decomposed to each responsibility center for control linkage. Finally, a dynamic adjustment mechanism is established. The budget unit can submit an adjustment application form online. After the dynamic application form is, Through the above approval process, the budget is iteratively updated. The system has achieved the full online process of budget preparation, execution, monitoring, and adjustment through information technology, significantly improving the efficiency of budget management. At the same time, data analysis models and artificial intelligence models are integrated into the performance preparation function module. The data analysis model determines the rationality of budget preparation by analyzing budget data, while machine learning provides financial budget prediction results based on previous years' data [11].

### 3.1 Machine Learning based Business Financial Budget Prediction Model

Machine learning models have the advantage of processing multidimensional data and can simultaneously provide information on historical financial performance, project characteristics, and other aspects of vocational colleges, overcoming the limitations of traditional budgeting methods that are difficult to fully integrate multiple factors [12]. The budget preparation of vocational colleges involves complex relationships between numerous variables, and machine learning algorithms such as decision trees and random forests can effectively identify nonlinear patterns and hidden rules in data, improving prediction accuracy. Secondly, machine learning models can continuously learn and optimize with the addition of new data, enabling budget forecasting to be dynamically adjusted according to the development of vocational colleges and improving the adaptability of budget management [13]. Compared to traditional methods that rely on subjective judgments, machine learning models provide more objective data support and reduce the interference of human bias in budget preparation [14]. The network structure of the algorithm is shown in Fig. 2.

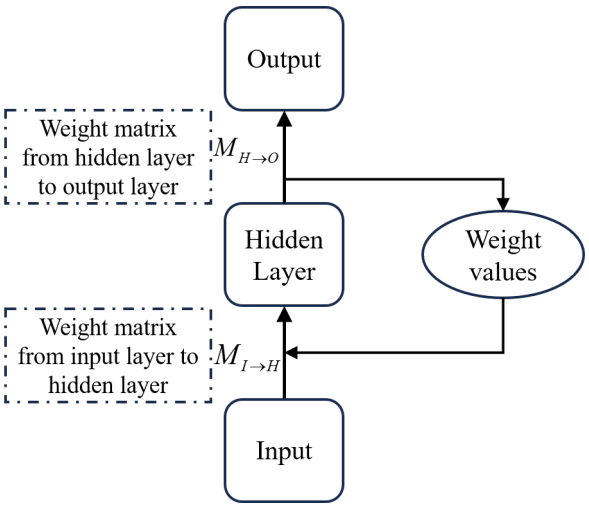



**Fig. 2.** Network structure diagram

The output of the algorithm [15] is publicly displayed as follows:

$$o_t = f\left(M_{I\to H}\cdot i_i + W\cdot o_{t-1} + \alpha\right) \tag{1}$$

$$o_{final} = soft\max\left(M_{H\to O}\cdot o_i + \beta\right) \tag{2}$$

Among them, $\alpha$ and $\beta$ are bias parameters, f( ) and *soft*max ( ) are activation functions, $o_t$ represents the output at the current time, and $o_{final}$ represents the final output. The core algorithm of this paper includes three gate control mechanisms, namely input gate, forget gate, and output gate, which are used to control the input, output, and forget of information, thus solving the gradient vanishing and exploding problems in traditional recurrent neural networks [16]. At the same time, the algorithm also includes a cell state variable for memory and transmission of information, which can effectively handle long-range dependencies in time series data. The calculation of the forget gate is as follows:

$$f_{gate,t} = \sigma\left(w_f\left[o_{t-1}, i_t\right] + \alpha_{f_{gate}}\right) \tag{3}$$

The input gate determines how much new information needs to be added to the unit state, and its implementation is divided into two steps. Firstly, the input gate determines the information that needs to be updated based on the current input and the previous output. Secondly, layer tanh generates a vector to update the temporary cell state $s_t$:

$$i_{gate,t} = \sigma\left(w_i\left[o_{t-1}, i_t\right] + \alpha_{i_{gate}}\right) \tag{4}$$

$$s_t = \tanh\left(w_s\left[o_{t-1}, i_t\right] + \alpha_s\right) \tag{5}$$

Then, it is necessary to update the old cell state and replace $s_{t-1}$ with $s_t$. Multiply the forget gate with the previous cell state by element to determine the preservation of the previous cell state. At the same time, multiply the input gate with the current unit state element by element to determine how much input of the network is saved at the current time [17]. Finally, combine the current temporary cell state with the previous unit state:

$$s_t = i_{gate,t} \circ s_t + f_{gate,t} \circ s_{t-1} \tag{6}$$

To determine the final output, an output gate is used in the algorithm, which determines which cell states should be outputted. The output gate calculation formula is expressed as:

$$o_{gate,t} = \sigma\left(w_o\left[o_{t-1}, i_t\right] + b_{o_{gate}}\right) \tag{7}$$

Finally, input the cell state into an tanh activation function for processing, obtaining a numerical value within the range of [−1,1]. Finally, multiply this value by the output of gate *sigmoid* to obtain the final output value, which is expressed as follows:

$$o_t = o_{gate,t} \circ \tanh\left(s_t\right) \tag{8}$$

In the process of constructing the financial budget prediction model in this article, the selection and processing of raw financial data is a crucial step. To ensure the scientificity of the model and the accuracy of the prediction results, this article uses the original financial data from the quarterly final accounts and 2023 annual budget reports of a vocational college in Xingtai City, Hebei Province from 2000 to 2022. These data provide detailed records of the company's financial operations in each quarter, covering multiple important projects and corresponding values involved in the company's financial budget. Not only does it provide the company's operational performance in different economic environments, but it also reflects the historical trends in the company's financial management. The framework for comprehensive budget preparation mainly consists of

business budget, special budget, and financial budget. The financial budget mainly includes the balance sheet, profit and loss budget, and cash flow budget. The items involved in each table are shown in Table 1.

**Table 1.** List of framework content

| Serial number | Project name | Serial number | Project name |
|---|---|---|---|
| 1 | Taxes and Surcharges | 12 | Cash inflows from financing activities |
| 2 | Sales expenses | 13 | Cash outflow from financing activities |
| 3 | Administrative expenses | 14 | The impact of exchange rates on cash and cash equivalents |
| 4 | R&D expenses | 15 | Net increase in cash and cash equivalents |
| 5 | Financial expenses | 16 | Closing balance of cash and cash equivalents |
| 6 | Other income | 17 | Net cash flows generated from financing activities |
| 7 | Gains from changes in fair value | 18 | Other receivables |
| 8 | Monetary funds | 19 | Inventory |
| 9 | Notes receivable | 20 | Contract asset |
| 10 | Accounts receivable | 21 | Other current assets |
| 11 | Prepayments | 22 | Long term equity investment |

Throughout the internal financial data of universities, it is generally found that the amount of raw data is insufficient to support the construction of subsequent predictive models [18]. Therefore, while ensuring data continuity and integrity, this article has conducted necessary data expansion on the raw financial dataset through interpolation. This article uses the mean imputation method to calculate the missing values of numerical features in the original financial dataset [19]. By calculating the mean of existing data and using this mean to replace the missing values in the features. Let $A_1$, $A_2$, $A_3$, …, $A_i$ be a known data point in a feature, and $n$ be the number of missing values. The formula for mean interpolation is:

$$\hat{A} = \sum_{i=1}^{m} A_i \Big/ m \tag{9}$$

In the formula, $m$ is the number of non missing values, $A_i$ is the observation value of the $i$-th sample, and $\hat{A}$ is the mean of the feature used to replace missing values [20]. The missing values of classification features were interpolated using mode interpolation method, which maintains the trend of category distribution in the dataset and makes the interpolated data more in line with the known distribution. Let the set of classification features be represented as:

$$S = \left\{F_{C1}, F_{C2}, F_{C3}, \cdots, F_{Ci}\right\} \tag{10}$$

If the frequency of each category is set to $f(F_{Ci})$, the definition of mode is:

$$\hat{F}_C = \arg\max_{F_{Ci}} f\left(F_{Ci}\right) \tag{11}$$

In the formula, $\hat{F}_C$ is the mode of the feature, which selects frequency $f(F_{Ci})$ and the largest category $F_{Ci}$ as a replacement for missing values. This process will be applied to all positions of missing values. Through the above interpolation method, the existing data volume can be expanded by 10 times. The data modeling process is shown in Fig. 3.

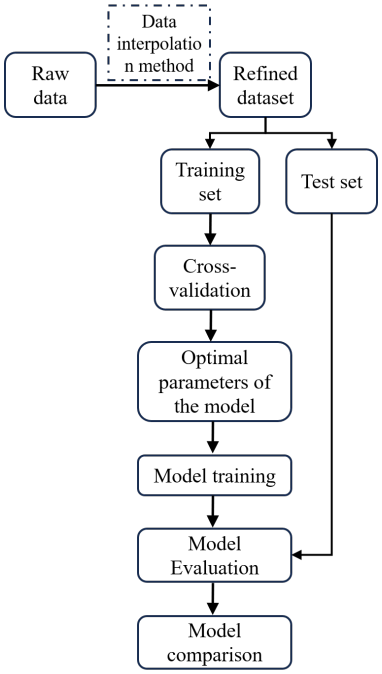



**Fig. 3.** Data modeling flowchart

In the process of constructing the model, this article uses the ensemble learning method for decision trees, which constructs a regression model by randomly selecting features and split points, similar to a random forest. However, more randomness is introduced in the process of constructing the tree, thereby reducing the risk of overfitting and improving the generalization ability of the model. It constructs a powerful regression model by combining multiple weak learners, namely decision trees. Its main advantage lies in its ability to calculate quickly and its suitability for processing datasets with a lot of noise [21].

1) Data preparation: When building each tree, use the entire dataset without performing resampling operations in the random forest, i.e. without Bootstrap [22]. The training dataset is:

$$S_{data} = \{(a_1, b_1), (a_2, b_2), \cdots, (a_m, b_m)\} \tag{12}$$

In the formula, $S_{data}$ is the dataset containing all training samples, $a_i$ is the feature vector of the $i$-th sample, and $b_i$ is the target value of the $i$-th sample.

2) This article adopts the method of randomly selecting feature subsets based on decision tree ensemble learning. More randomness is introduced in the splitting process of each network node. On each network node, a feature subset is randomly selected instead of finding the optimal splitting point in all feature subsets. The process of randomly selecting feature subsets is expressed by the formula:

$$F_{Random} = RandomSubset(S_{feature}, T) \tag{13}$$

$S_{feature}$ is the set of all features, with a total of $p$ features, and $F_{Random}$ is a random selection of $T$ features from $S_{feature}$.

3) Randomly select a splitting point, and the selection of splitting points still uses a random method. For each selected feature subset, randomly select one as the splitting point instead of finding the optimal splitting point on that feature. By randomly selecting features and splitting points, the calculation process is optimized. For each feature subset, randomly select a splitting point $p_{split,i}$. The selection process is represented as:

$$p_{split.i} = Random\left(a_i\right), \min\left(a_i\right) \le p_{split.i} \le \max\left(a_i\right) \tag{14}$$

In the formula, $p_{split.i}$ is the random splitting point of feature $i$, and $a_i$ is the value of feature $i$ for all samples in the current node.

4) Calculate variance reduction, which is mainly used to evaluate the degree of improvement in the distribution of target values within nodes for each split. It helps to select the best split in each split step and construct a decision tree model that can more accurately predict target values [23]. For each candidate feature $i$ and $p_{split.i}$ splitting point, the variance before and after splitting is reduced to:

$$\Delta Var\left(i, p_{split.i}\right) = Var\left(S\right) - \left(\frac{\left|S_L\right|}{\left|S\right|} Var\left(S_L\right) + \frac{\left|S_R\right|}{\left|S\right|} Var\left(S_R\right)\right) \tag{15}$$

$\Delta Var(i,split.i)$ is the variance reduction caused by feature $i$ and splitting point $p_{split.i}$, $Var(S)$ is the variance of the samples in the current node $S$, $S_L$ and $S_R$ are the sample sets of the left and right child nodes after splitting, $|S_L|$ and $|S_R|$ are the sample quantities of the left and right child nodes, and $Var(S_L)$ and $Var(S_R)$ are the variances of the target values of the samples in the left and right child nodes, respectively.

5) During the solving process, the model adopts a recursive splitting method, so it is necessary to set a stopping criterion for recursive splitting. The construction process of the decision tree recursively divides the data until the stopping condition is met. If the maximum depth of the tree is reached or the number of samples in each leaf node is lower than the minimum value, recursive splitting stops.

6) The final prediction is that the decision tree model constructed in this chapter can perform regression prediction on new input data. The formula used in the prediction process is expressed as:

$$V_{predicted}\left(a\right) = \frac{\sum_{i \in L(a)} V_{predicted.i}}{L\left(a\right)} \tag{16}$$

The algorithm prediction model includes multiple decision trees and is trained by fully utilizing the interpolated and supplemented dataset samples to fully utilize the overall characteristics and diversity of the data. Splitting nodes uses mean square error as a measure, which helps minimize prediction error in regression tasks and does not limit the depth of the decision tree. Therefore, the decision tree model can capture complex relationships between sample data.

In terms of splitting control, the model sets at least two samples per node to split, and each leaf node retains at least one sample to ensure the flexibility and predictive ability of the decision tree. Randomness is controlled by random seeds to ensure the repeatability of the experiment. The maximum number of features in the model will automatically select the number of features used during splitting to adapt to the characteristics of the data. In addition, the model used in this article does not have additional restrictions on the decrease of splitting purity, so the model will obtain as refined splitting conditions and splitting results as possible. Overall, the model described in this article fully utilizes the complexity and diversity of data to mine potential patterns without limiting the depth of decision trees. It is suitable for complex tasks with high feature dimensions and large data scales, and can adapt to the budget forecasting task in financial management of vocational colleges described in this article. At the same time, setting random seeds ensures the stability of the model results, and through variable importance analysis, cost and time are identified as the most important predictive features in the model, further reflecting the sensitivity of the model to key variables such as cost and time [24].

### 3.2 Multidimensional Analysis of Financial Data

The online multidimensional analysis subsystem for financial data is designed for financial big data. According to the data analysis requirements of financial managers described in this article, data needs to be extracted from various business systems, integrated, stored in a data warehouse, and then an analysis model is established based on the data warehouse for online statistical analysis and visual display [25]. The overall framework of the multidimensional analysis system is shown in Fig. 4.

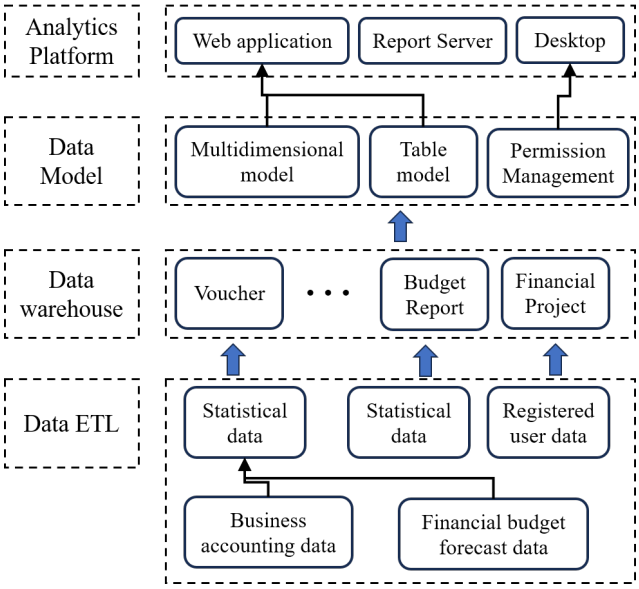



**Fig. 4.** Overall framework diagram of the database

The overall architecture adopts a multi-layer architectural style, and this subsystem consists of the following four layers:

1) The data ETL module first extracts business data into an intermediate database after processing, then performs data cleaning and transformation in the intermediate database, and finally completes the three stages of loading the data warehouse in the intermediate database. Data ETL, which includes the three functions of data extraction, transformation, and loading, is a necessary step in building a data warehouse. The first two stages are mainly completed using the "Oracle Job" mechanism, while the last stage is completed using XSIS. The business database and intermediate database use Oracle database. Business data mainly includes financial accounting business data, related business data after financial budget forecasting, such as budget reports, etc. There are also some specific business data that are not in the business database. When using them, they need to be imported into the data warehouse in the form of Excel files. In addition, to manage login users and their data access permissions, it is necessary to obtain personnel data from the unified portal system and manage user data.

2) Data warehouse uses Vertica database to store ETL fact tables and dimension tables, mainly including voucher fact tables, budget fact tables, transaction fact tables, account dimension tables, project dimension tables, indicator dimension tables, research funding dimension tables, horizontal project funding tables, experimental material funding tables and other related business dimension tables [26].

3) Data model, using SSAS, different analysis sections are configured with different instances according to data analysis needs, and multidimensional models, table models, and permission management modules are built separately.

This article defines specific mapping relationships based on actual needs, realizes the migration from financial business system data sources to target data warehouses, and implements necessary data transformation and cleaning during the migration process. Financial business data itself is complex and intricate, and different business data may come from databases of different schools and departments. This requires extracting the required data from heterogeneous data sources, cleaning, transforming, and finally loading it into the target data warehouse according to a pre-defined data model. The structure is shown in Fig. 5.

The data extraction methods are timestamp method, log table method, trigger method, and full table comparison method.

1) Timestamp method: There is a timestamp field in the business data table. When extracting, only the updated data after the last extraction time is extracted according to this timestamp. This data extraction method needs to meet the timestamp condition of the business table. However, in actual financial and budget management systems, most tables do not have a timestamp field designed, so the timestamp method is limited in the system operations described in this article.

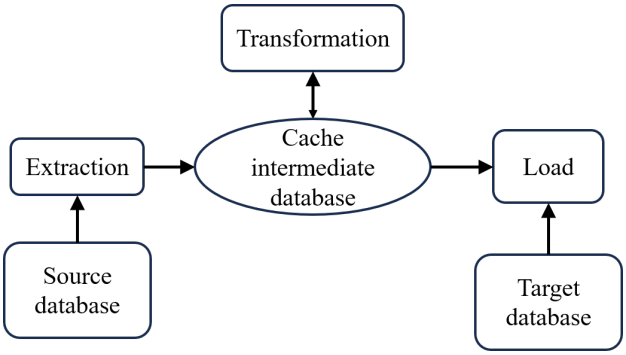



**Fig. 5.** Data conversion structure diagram

2) Log table format: The business system has a business operation log table, and when extracting data, the updated data of each related business table is extracted based on the changes in the logs of the business operation log table. This data extraction method requires the business system to have a business log table, but in reality, the critical business operations of the university financial business system are recorded in a log table. However, there are still many business operations that do not record logs, so the log table method is limited in the system operations described in this article.

3) Trigger method: There are triggers built in the business data table. When the business data table performs insert, delete, or update operations, the relevant change data is inserted into a dedicated incremental log table, and the data of the incremental log table is extracted according to the timestamp of the incremental log table or the marker of whether it has been extracted. The data of the business data table is not directly extracted. The actual situation is that establishing triggers on business tables will bring additional costs to the system, which may reduce the operational efficiency of the business system. Even if the table structure of the business table is updated, the incremental log table's table structure may not be updated in a timely manner, resulting in errors when executing insert, delete, and update operations on the business table, affecting the normal operation of the business system. Therefore, the financial business system will not establish triggers that are unrelated to business operations, and therefore cannot use this method of operation.

4) Full table comparison method: When extracting, compare the data between the business data table and the target table one by one, obtain the difference records, and update them to the target table. The amount of operation in the second half of the update phase is relatively small, but in the first half of the comparison phase, the amount of operation is huge. When the business data table is particularly large, the entire process takes much more time than directly truncating or inserting the target table. Considering the cost-effectiveness of the operation, choose this method of operation.

### 3.3 Table Structure Conversion

Table structure conversion is the standardization of field names, field types, and field lengths in business tables to achieve row column data conversion, etc. Standardize field names, field types, and field lengths [27]. The financial business system includes numerous subsystems, and different subsystems may use different field names, types, and lengths to describe the same entity. When entering the data warehouse, parallel standardization is required. The schematic diagram of the entire processing process is shown in Fig. 6.

After ETL, the raw business data is mainly formed into factual tables such as vouchers, transactions, budgets, logistics, and teaching and research achievements of faculty and staff, as well as dimension tables such as dates, subject information, subject sections, projects, aging types, student information, faculty and staff information, and classification of teaching and research achievements [28]. The specific list is shown in Table 2.

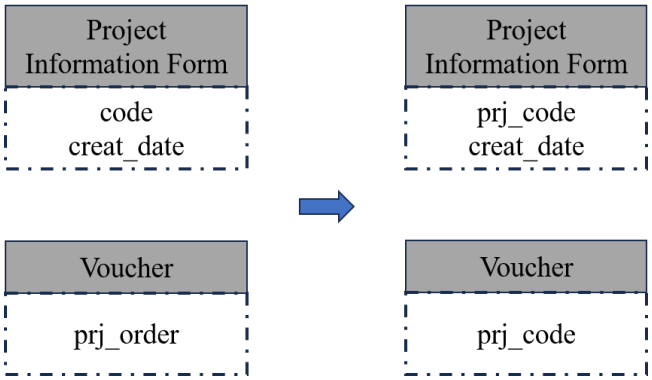



**Fig. 6.** Diagram of table structure conversion

**Table 2.** Tables in the data warehouse

| Table name | Meaning |
|---|---|
| FACT PZD ALL | Voucher entry table |
| FACT CLEARDTL | Correspondence form |
| FACT_OPROJ_PLAN | Budget expenditure table |
| FACT YY INCOME | Budget revenue table |
| FACT YB APPLY | Appointment reimbursement form |
| FACT FLOW LOG | Booking reimbursement form processing logistics log table |
| FACT_STUDENTBASIC | Student basic information fact sheet |
| FACT STUDENTDEPARTMENT | Fact Sheet of Student College Situation |
| FACT_STUDENTSUBJECT | Student professional direction fact sheet |
| FACT TEACHERSTAFFCATE | Fact sheet of faculty college |
| FACT TEACHERSTATISTIC | Basic Information and facts of faculty and staff |
| FACT ACHEVEMENTAWARD | Scientific research achievement award fact table and paper fact table |
| FACT ACHEVEMENTTHESIS | Date table |
| DIM DATE | Subject information table and subject section table |
| DIM_ SUBJ_INFO | Project list |
| DIM_SUBJ_DIM_INFO | Project category table |
| DIM PROJ_INFO | Student professional direction fact sheet |
| DIM PROJ TYPE | Fact sheet of faculty college |
| DIM PROJ_DIM_ INFO | Project section table |
| DIM ZL TYPE | Aging type table |
| DIM YY_OPROJ | Budget expenditure item table |
| DIM YY OTYPE | Budget expenditure node table |
| DIM YY OATTR | Budget expenditure section table |
| DIM_YY_IPROJ | Budget Revenue Project Table |
| DIM YY ITYPE | Budget revenue node structure table |
| DIM DEPARTMENT | College dimension table |
| DIM CATEGORY | Student category table |
| DIM SUBJECT | Professional direction dimension table |
| DIM STAFFCATEGORY | Dimension table of faculty and staff categories |
| DIM CATESCIENTIFICNEW | Dimension table of research achievement categories |
| DIM AWARDCATEGORY | Award category dimension table |
| DIM THESISCATEGORY | Table of paper category dimensions |

A data model typically consists of two levels, namely the conceptual layer (logical layer) and the physical layer. The logical data model is the data obtained by abstracting the real world from a conceptual perspective, while the physical data model focuses on the specific implementation process in a specific environment. The data model described in this section is mainly a logical data model. In the process of data creation, the data model is created according to financial analysis requirements, which is the core link of the design of the online

financial data analysis system. The creation process involves loading the basic tables in the financial data warehouse into the model library, constructing data association models according to analysis requirements, and further processing the data tables in the model to generate calculation tables, calculation columns, measurement values, etc. The forms of financial analysis data models include multidimensional models and tabular models. The multidimensional model is used for the cross analysis module of multi-dimensional basic information and financial income and expenditure data of teachers and students, while the tabular model is mainly used for the analysis module of pure financial data. This article focuses on the construction process of a financial data model: financial data analysis includes multiple topics such as financial overview and income and expenditure analysis, transaction analysis, etc., including multiple fact tables. At the same time, financial item dimension tables are used between different topics. Therefore, the overall data model will be constructed using a fact constellation model. The specific construction process is as follows:

```
  (FACT_PZD_ALL, FACT_CLEARDTL, FACT_OPROJ_PLAN, FACT_YY_INCOME, FACT_YB_APPLY,
FACT_FLOW_LOG)
  Extract dimension tables from the data again:
  (DIM_DATE, DIM_SUBJ_INFO, DIM_SUBJ_DIM_INF, DIM_PROJ_INFO, DIM_PROJ_TYPE, DIM_
PROJ_DIM_INFO, DIM_ZL_TYPE, DIM_YY_OPROJ, DIM_YY_OTYPE, DIM_YY_OATTR, DIM_YY_IPROJ,
DIM_YY_ITYPE)
```

After loading the table into the model library, establish association relationships using the fact constellation model according to the following relationship diagram. The key themes in the model include financial overview and income and expenditure analysis, current and account analysis, budget income and expenditure analysis, and appointment reimbursement situation analysis. Each theme is represented by a snowflake model, and when combined into a whole, it forms a factual constellation model. In the theme of financial overview and income and expenditure analysis, the fact table is associated with the dimension table, where the dimension table is further associated with the refined dimension table, and the dimension table is further associated with the refined dimension table, forming the snowflake model. In the subject of transaction analysis, the fact table is associated with the dimension table, and the detailed association of the dimension table is the same as the description of the financial overview and income and expenditure analysis subject, forming the snowflake model.

### 3.4 Generate Calculation Columns and Measure Values

Many indicators are required in financial analysis, and there are no direct column data in the initial default table. They need to be obtained through appropriate function processing of existing basic columns. If there are basic columns such as research funding amount and research funding balance in the fact table, but there is no income indicator column in the fact table, it can be calculated by calling the year, month, subject section, research funding amount, and funding balance columns.

A metric is a key indicator value calculated from a fact table and is a formula created for use in reports or Excel pivot tables. It can be defined based on standard aggregation functions and DAX. Metrics are business elements that can be added up, averaged, or mathematically processed in some way to obtain information. Unlike calculation columns, which do not calculate measurement values at the row level, measurement values support the numerical statistics function of Excel pivot tables, while calculation columns cannot participate in numerical statistical analysis in Excel pivot tables. Financial data analysis requires the provision of Excel pivot table analysis functionality for some users, and to support this feature, measurement values must be used. In financial analysis data models, use DAX to establish measurement values.

## 4 Simulation Experiment and Result Verification

Firstly, perform data analysis and display on the system designed in this article, according to the analysis requirements of business themes and job roles, it is divided into seven sub modules related to data analysis and display, including financial overview, revenue analysis, expenditure analysis, balance analysis, data comparison, warning monitoring, and business management. The financial overview mainly includes data analysis of the overall financial situation of the school. The modules of income analysis, expenditure analysis, balance analysis, data comparison, warning monitoring, etc. mainly include data analysis at the school level and data analysis

at the subordinate department level. The business under management mainly provides corresponding financial data analysis functions according to the school's personnel, scientific research and other business sectors. The submodule for data analysis display is shown in Fig. 7.

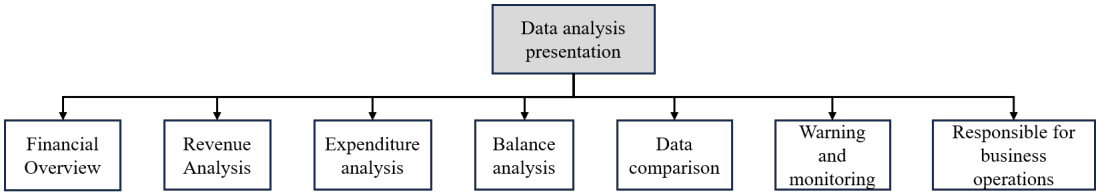



**Fig. 7.** Submodule for data analysis display

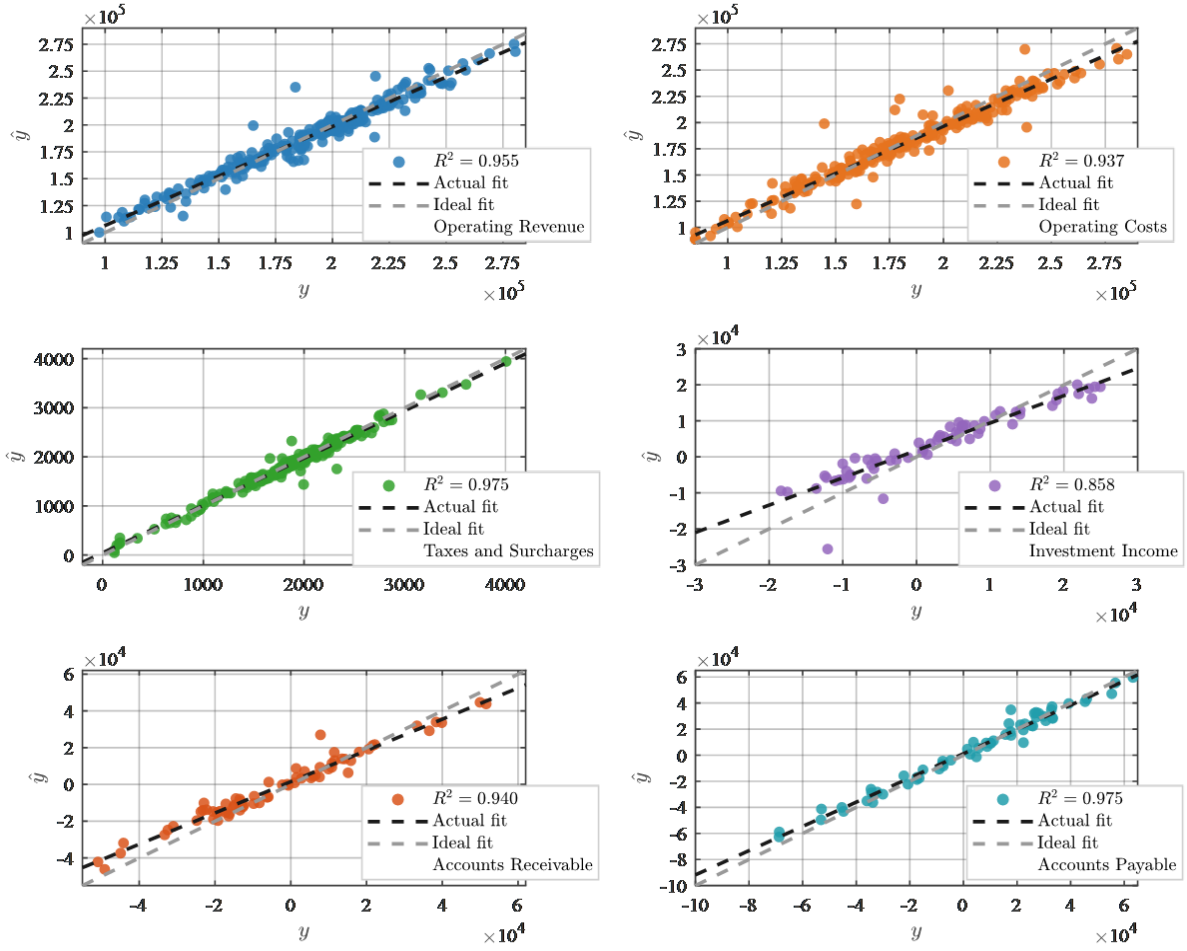



**Fig. 8.** Prediction accuracy

After determining the decision tree algorithm to construct the prediction model in this article, it is necessary to conduct experiments with the income and expenditure of vocational colleges as the target variables. The prediction accuracy results of the six experiments are shown in Fig. 8. In six experiments, the accuracy rate of operating revenue was 95.5%; The accuracy rate of operating costs is 93.7%; The accuracy rate of operating profit is 97.5%; The accuracy of net cash flow is 85.8%; The accuracy rate of accounts receivable is 94%; The accuracy rate of accounts payable is 97.5%. Although there are individual points that deviate from the theoretical

best fit line, the scatter distribution shows a certain degree of linear trend, and most of the scatter points are located near the ideal prediction line, indicating that the model's prediction results are very close to the actual values. The slopes of the fitted lines are close to the ideal prediction line, indicating that the overall prediction trend of the model is relatively close to the trend of the actual value. Except for the accuracy of net cash flow, which is 85.8%, the accuracy of all other experimental variables is between 93.7% and 97.5%, indicating that the model can explain the majority of data changes and has a good fit to the data, which can explain the majority of variance changes. Overall, the model performs well in prediction and has strong explanatory power, with some differences in performance on different datasets.

The learning curve can be used to evaluate the learning progress of a model during the training process by plotting the performance of the training and validation sets, such as loss values or accuracy, as a function of the number of training samples or iterations. In the learning curves of six experiments, the training performance of the model under different target variables was demonstrated, showing the scoring of the model on the training and validation sets as the amount of training data gradually increased. From the graph, it can be seen that the training scores of the learning curves for the six experiments are consistently close to 1.0, while all validation scores steadily improve with increasing training volume. The cross validation score for operating revenue and operating costs is 0.94; The cross validation score for operating profit is 0.96; The cross validation score for net cash flow is 0.9; The cross validation score for accounts receivable is 0.94; The cross validation score for accounts payable is 0.93. When there are 550 data points, the overall accuracy of the model approaches its peak, which means that increasing the training data can help improve the predictive ability of the model; Meanwhile, the difference between the cross validation score and the training score is relatively small, indicating that the model has strong predictive ability for new data. When the model is trained with different target variables, it exhibits high accuracy and prediction precision, indicating that the model has good generalization ability. The learning curve is shown in Fig. 9.

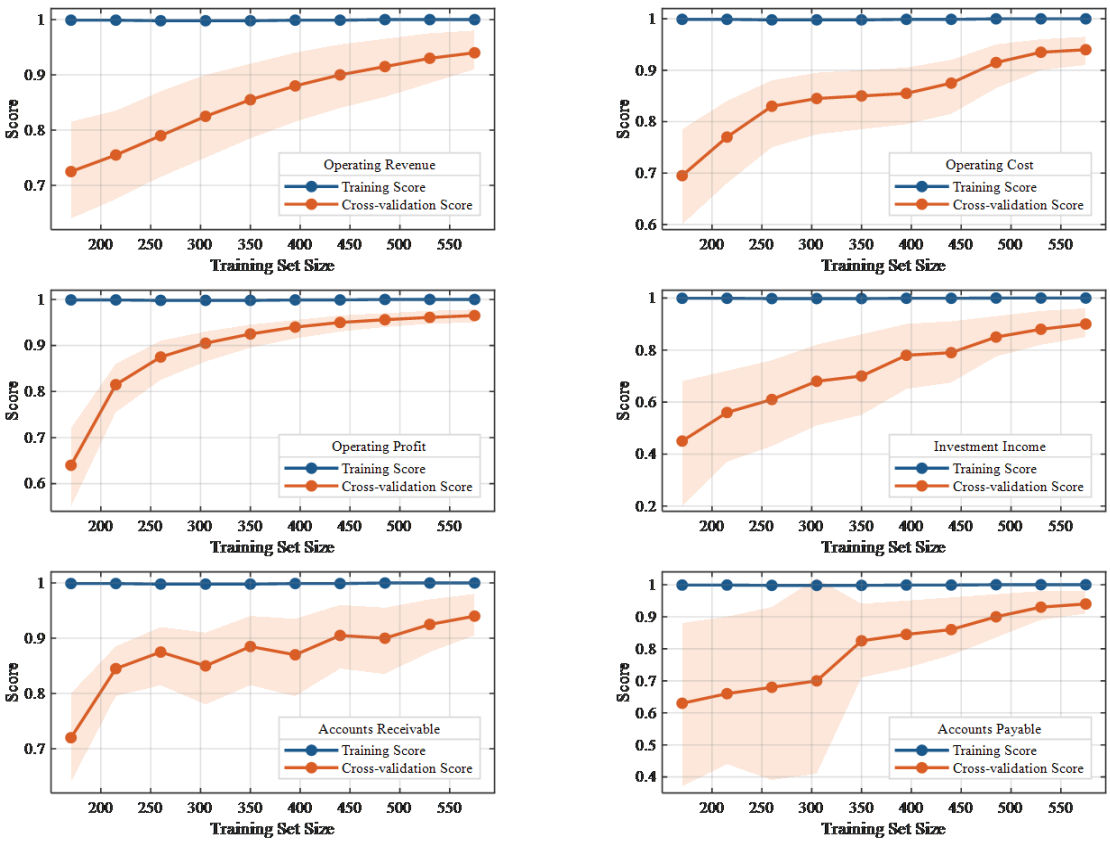



**Fig. 9.** Learning rate curve

This chapter mainly presents the system function design scheme, with a focus on scheme design. In terms of method effectiveness, the prediction curve graphs of scientific research fund receivables and payables in the system are displayed to demonstrate the prediction effect. At the same time, the comparison of the loss value or accuracy curve of the model proves the effectiveness of the method proposed in this paper.

## 5 Conclusion

This article studies and develops a university financial data analysis and intelligent personnel allocation system based on big data. The system extracts basic business data from various business systems, and after appropriate cleaning and processing, enters the financial data warehouse. Analysis models are established according to the analysis needs of management personnel in different positions, and visualized in multiple dimensions; Simultaneously conducting in-depth analysis and mining of financial logistics log data, identifying opportunities for improvement, and implementing intelligent allocation of financial personnel to assist in financial decision-making and management work.

1) The online multidimensional analysis subsystem for financial data, through the construction of financial data ETL and data warehouse, realizes the creation, publication, and visual display of multidimensional analysis charts for financial vouchers, project basic information, project budgets, and other financial data.

2) The analysis and mining subsystem of financial logistics logs identifies operational bottlenecks and improvement opportunities through statistical analysis of service process data, modeling of financial personnel capabilities, and prediction of dwell time for financial service task instances, thereby assisting in the optimization and performance improvement of financial service processes.

3) Summarized the deficiencies in financial budget forecasting in universities, and preliminarily identified potential management issues in financial budget preparation by examining the principles and processes of financial budget preparation in universities; By analyzing the differences in budget and final accounts through seven financial projects that run through the entire budget preparation process of universities, it can be confirmed through questionnaire surveys and other forms of analysis that universities face the serious management problem of insufficient accuracy in financial budget preparation. This problem is caused by rough budget preparation on the business side, insufficient efficiency of traditional preparation methods, and the lack of dynamic adjustment mechanisms for budgets. Its essence lies in the fact that traditional methodologies are difficult to cope with complex financial data relationships and lack necessary technical and organizational support. The machine learning model used in this article can effectively solve the above problems.

On the basis of existing research results, further research and improvement will be conducted on the methods of financial data analysis and budget forecasting in universities in the following three aspects. The research plan is as follows:

1) In terms of multidimensional data analysis, further expand the scope of business source data, increase the number of fact tables and dimension tables, add calculation tables, calculation columns, and measurement values, construct more complex data models, and use machine learning methods to transform financial statement data into more understandable intelligent dynamic financial reports;

2) We plan to improve the prediction of task stagnation time in the financial reimbursement process through two methods: firstly, optimizing data sampling by detecting outliers through clustering to reduce noise in the original data; The second is to apply more advanced machine learning models based on neural networks.

3) In terms of budget forecasting and evaluation, financial budget preparation is an important part of comprehensive budget management for enterprises. Based on existing research, we first continuously optimize algorithm models or explore other cutting-edge algorithms, such as hybrid models combining deep learning and time series analysis, from the perspective of model technology. In particular, we need to enhance the model's ability to respond to emergencies and external macroeconomic fluctuations; At the data level, integrate more real-time business data, unstructured information, and external data to establish a more comprehensive data ecosystem; At the application level, the prediction model is deeply integrated with the comprehensive budget management system to build an intelligent budget preparation, execution, monitoring, and analysis platform, realizing the automation and intelligence of the entire budget process, and expanding the prediction scope to finer grained cost elements, fund planning, and even combining risk management for early warning; At the organizational management level, we will continue to strengthen the construction of data-driven decision-making culture, enhance the data literacy of faculty and staff, optimize cross departmental collaboration mechanisms, and ensure the coordinated evolution of technical tools and management practices. Through these continuous efforts,

the data-driven financial forecasting ability will be transformed into a powerful support for vocational colleges to achieve refined management, optimize resource allocation, and enhance core competitiveness in complex operating environments.

## 6 Acknowledgement

Research on the Optimization and Efficiency Improvement of Financial Budget Management in Higher Vocational Colleges from the Perspective of Governance Modernization, Project No.: SD26002.

## References

[1] D. Zhang, C. Li, S.-H. Gao, Research on Pathways for Enhancing University Scientific and Technological Innovation Capabilities in the Context of Higher Education Empowering the Construction of a Scientific and Technological Powerhouse: A Fuzzy set Qualitative Comparative Analysis Based on Universities in 30 Provinces, Science and Technology Management Research (5)(2026) 67-76. https://doi.org/10.3969/j.issn.1000-7695.2026.5.008

[2] Y.-H. Ma, L.-P. Jiang, C.-Y. Kuang, Optimization Path of Innovation Performance Incentive Mechanism in Universities under the Reform of "Decentralization, Regulation, and Service" of Scientific Research Funding, Science and Technology Management Research (4)(2026) 93-100. https://doi.org/10.3969/j.issn.1000-7695.2026.4.010

[3] X.-L. Wu, J. Shen, J. Lu, Construction of the Budget Performance Evaluation Index System for Double First-class Universities Based on DEMATEL, Journal of Wuhan University of Technology (Information & Management Engineering) 48(2)(2026) 179-185. https://doi.org/10.3963/j.issn.2095-3852.2026.02.009

[4] H.-Q. Li, W. Cai, X.-L. Zhang, Research on Problems and Solutions in University Budget Performance Management, Journal of Qingdao Agricultural University (Social Science) 37(4)(2025) 49-53. https://doi.org/10.3969/J.ISSN.1674-1471.2025.04.006

[5] L. Ma, Z.-Y. Yu, L. Wang, Research on Investment in Higher Education Funding in Shaanxi Province, Journal of Yangzhou University (Higher Education Study) 29(4)(2025) 1-10. https://doi.org/10.16526/j.cnki.11-4762/tp.2022.08.019

[6] Y.-K. Yang, Y.-B. Wei, W.-Q. Wang, D.-X. Yang, F.-Y. Hong, Key Technologies for a Dual-Chain Structure-Based University Financial Reimbursement System, Journal of Applied Sciences 43(4)(2025) 617-629. https://doi.org/10.3969/j.issn.0255-8297.2025.04.005

[7] Y.-N. Wu, S.-Y. Zhao, Research on Engineering Cost Prediction Model Based on Deep Learning, Modern Engineering Technology 4(13)(2025) 1-4. https://doi.org/10.26929/j.cnki.issn.2097-1672.2025.13.001

[8] Y. Xu, Z.-J. Rao, M. Huang, F.-Y. Ma, Journal of Yunnan Agricultural University (Social Science) 17(3)(2023) 26-31. https://doi.org/10.12371/j.ynau(s).202212062

[9] C.-H. Zhang, Public Hospital Financial Budget Information Management System Based on Principal Component Analysis, Ordnance Industry Automation 43(4)(2024) 46-49+76. https://doi.org/10.7690/bgzdh.2024.04.009

[10] S.-Y. Mo, L. Hu, Y. Zhu, X.-X. Rong, C.-R. Wang, Design of Hospital Comprehensive Budget Management System Based on Improved HRP, Microcomputer Applications 42(3)(2026) 348-351. https://doi.org/10.3969/j.issn.1007-757X.2026.03.082

[11] Z.-Q. Ding, M.-D. Sun, X.-L. Wang, Research on a Financial Fraud Prediction Model for Listed Companies Based on LightGBM, West Forum on Economy and Management 36(6)(2025) 25-39. https://doi.org/10.12181/jjgl.2025.06.03

[12] J.-J. Liu, C.-R. Zheng, Y.-M. Hong, How Can Machine Learning Empower Management Research? A Domestic-Foreign Frontier Review and Future Prospects, Journal of Management World 39(9)(2023) 191-216. https://doi.org/10.19744/j.cnki.11-1235/f.2023.0106

[13] H.-L. Niu, K.-L. Xu, M.-Y. Xiong, Enterprise Performance Prediction from Perspective of Executive Governance: Empirical Evidence from Machine Learning, Operations Research and Management Science 33(12)(2024) 224-231. https://doi.org/10.12005/orms.2024.0409

[14] S.-Q. Yang, S.-L. Li, Application of the revenue and user prediction model based on machine learning-linear regression algorithm in audit projects, Wireless Internet Science and Technology 20(23)(2023) 94-98. https://doi.org/10.3969/j.issn.1672-6944.2023.23.025

[15] B.-W. Wang, Y. Huang, X.-B. Meng, T.-Y. Cao, Research on port container throughput forecasting based on secondary decomposition and LSTM optimized by algorithm, Navigation of China 48(4)(2025) 121-131. https://doi.org/10.3969/j.issn.1000-4653.2025.04.014
[16] B. Chen, X. Liu, Design of power engineering data analysis algorithm integrating Attention and improved LSTM, Electronic Design Engineering 32(24)(2024) 114-118. https://doi.org/10.14022/j.issn1674-6236.2024.24.024
[17] L. Jia, G.-W. Shi, N.-J. Jiang, X.-J. Wang, L. Li, Y. Miao, Research on oil well production prediction method based on machine learning technology and optimization algorithm, Oil Drilling & Production Technology 47(6)(2025) 757-764. https://doi.org/10.13639/j.odpt.202504037
[18] Y.-F. Xia, Y. Chen, P.-Y. Zou, A clustering semi parametric logistic learning inputation method for categorical missing data, Journal of Sichuan University (Natural Science Edition) 60(6)(2023) 32-39. https://doi.org/10.19907/j.0490-6756.2023.061001
[19] Y. Tang, P. Li, W.-H. Li, Y. Zhou, Y.-L. Feng, Electricity Theft Data Detection Based on Missing Value Imputation and Random Forest, Microcomputer Applications 42(1)(2026) 163-166. https://mall.cnki.net/eread/mall/maga/WXDY202601.html
[20] L. Ma, L.-Q. Guo, B.-J. Liu, J. Yang, Research About Interpolating Missing Data on Small Sample Trials of Equipment Based on Ensemble Learning, Computer Measurement & Control 30(8)(2022) 116-121. https://doi.org/10.16526/j.cnki.11-4762/tp.2022.08.019
[21] H.-W. Peng, Z.-B. Lin, J.-J. Yang, Improved Algorithm of Decision Tree Based on Neural Network, Computer and Network 51(2)(2025) 161-166. https://doi.org/10.20149/j.cnki.issn1008-1739.2025.02.010
[22] X.-Y. Wei, J.-P. Guo, G.-J. Li, Research on medical data governance and multi-source data fusion based on integrated learning, Electronic Design Engineering 33(3)(2025) 34-37+42. https://doi.org/10.14022/j.issn1674-6236.2025.03.007
[23] G.-J. Han, X.-R. Zeng, H.-N. Yu, Steered minimum variance algorithm based on virtual array element extension, Acoustics and Electronics Engineering (2)(2026) 19-24. https://doi.org/10.3969/j.issn.2096-2657.2026.02.04
[24] W.-W. Yang, L.-J. Zeng, Research on Big Data Classification Algorithm Based on Optimized Decision Tree Model Under Block Chain Framework, Journal of Shenyang Institute of Engineering (Natural Science) 17(4)(2021) 57-62. https://doi.org/10.13888/j.cnki.jsie(ns).2021.04.012
[25] T.-J. Jia, Y. Wang, P.-M. Li, Y.-P. Wang, Y. Wu, Design of financial information automatic classification system based on digital information technology, Electronic Design Engineering 34(1)(2026) 35_39+44. https://doi.org/10.14022/j.issn1674-6236.2026.01.007
[26] Y. Zhao, P.-F. Zou, L. Wang, N. Li, W.-J. Hao, Railway financial consolidation reporting system based on multidimensional data warehouse technology, Railway Computer Application 34(7)(2025) 29-33. https://doi.org/10.3969/j.issn.1005-8451.2025.07.06
[27] Q.-X. Wang, Y. Li, C.-G. Ding, J.-C. Xie, Research on Data Query Optimization Technology Based on Table Structure Transformation, Electronic Product Reliability and Environmental Testing 38(2)(2020) 30-34. https://doi.org/10.3969/j.issn.1672-5468.2020.02.006
[28] Y. Zhuang, Construction of Enterprise Financial Statement AnalysisModel Based on K-means Clustering Algorithm, Journal of Shanxi Normal University (Natural Science Edition) 40(1)(2026) 40-44. https://doi.org/10.16207/j.cnki.1009-4490.2026.01.007