International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-0035
Article

Multidimensional Data Mining Model Analysis of Financial Structure Optimization of Universities in Government Accounting Comprehensive Budget Performance Management

Qingquan Huang *

Dongguan City University, Dongguan 523419, Guangdong, China; huangqingquan@dgcu.edu.cn

Abstract: Under the background of government accounting comprehensive budget performance management, this study focuses on the optimization of financial structure of colleges and universities, and constructs a set of financial decision-making model system based on multi-dimensional data mining. By integrating high-precision financial data such as budget accounts, project expenditures, scientific research funds, tuition income and expenditure of universities. Innovative fusion of three types of data mining techniques. Multi-dimensional association rule mining reveals inter-departmental expenditure associations, cost composition laws, project hierarchy differences, etc. ID3 decision tree algorithm realizes financial risk classification and budget execution evaluation.DBSCAN density clustering divides student groups by the amount of outstanding fees and supports poor student classification. The data read/write efficiency of the proposed method is over 42bit/s, which is 200% higher than the benchmark method, and the redundant rules are reduced by 60%. And 415 strong association rules are generated (e.g., conference fee → postal fee, confidence level 0.6125, enhancement 1.98).The AUC of the three poverty categories classified by the DBSCAN model exceeds 0.98, which indicates that the model is more accurate and achieves a better classification effect. The decision tree model accurately quantifies the poverty class (0.247 gain in per capita annual household income of the root node), providing a data kernel for scientific allocation of resources.

Keywords: government accounting; budget performance management; university financial optimization; ID3 decision tree; DBSCAN clustering

1. Introduction

With the continuous development of China's higher education, the number of colleges and universities has been increasing, which is of profound significance in promoting the progress of China's higher education. However, in the construction of higher education, there is a contradiction between the state financial allocation and the financial needs of the development of higher education undertakings themselves [1-2]. In order to improve the efficiency of using funds and realize the rational allocation of public resources, universities must introduce performance management, especially the important aspect of budget performance assessment [3-5].

Comprehensive budget performance management is a performance-centered management approach that emphasizes the key role of performance management in the process of budget execution [6]. Different from the traditional budget management, the so-called comprehensive budget performance management, including the implementation of budget management in all aspects, in all processes and in all coverage, comprehensive budget performance management emphasizes more on integrating performance management into the budget and realizing the integration of the two [7-10]. In order to allocate funds more accurately, avoid the waste of resources, and improve the efficiency of the use of funds, university enterprises must optimize the comprehensive budget performance management [11-12]. Through scientific budget performance management, it ensures that the input of resources matches the strategic development, thus promoting the realization of the strategic objectives of university enterprises [13-14]. In addition, the new governmental accounting system requires university enterprises to comply with the relevant regulations and optimize the internal financial structure, so as to improve the overall management level of university enterprises [15-17]. Only in this way can we enhance the competitiveness of university enterprises and provide strong support for the sustainable development of universities.

This study is dedicated to constructing a set of university financial data processing and analysis model system based on multi-dimensional data mining, which provides a strong methodological foundation and technical support for the core research goal of optimizing the financial structure of universities. Firstly, the data foundation is consolidated, and then advanced mining techniques are utilized to reveal the intrinsic associations and patterns in budget and financial activities, which ultimately serves the intelligent financial forecasting and decision support. Based on the basic concept of association rules, the article focuses on how the technology is applied to budget data mining in colleges and universities. By constructing a multi-dimensional dataset containing units, projects, superior departments, grades, years, etc., and utilizing multi-dimensional association rule mining technology, the association of expenditure costs among different department types, the compositional association within the categories of expenditure costs, as well as the association pattern of the costs among projects at different levels are revealed. These mining results provide a data-driven decision-making basis for optimizing the financial expenditure structure, scientifically allocating unit budget amounts, and assessing budget standards and quotas. The algorithm design of the data mining-based financial decision-making system for universities further expands the application depth of the model and builds more complex decision support capabilities. At the same time, financial data clustering is accomplished through DBSCAN, and poor students' grades are identified based on the ID3 algorithm.

2. Financial Data Processing and Multi-Dimensional Mining Model Construction in Higher Education Institutions

2.1. Financial Data Processing in Higher Education

2.1.1. Data Acquisition

Data collection includes financial data and related business data. The financial system itself has accumulated many years of data, such as budget accounts, project expenditure details, scientific research projects, tuition fees, bills, etc., which are characterized by high accuracy, clear timing and clear logical relationships. Access to related data, such as information on students, research projects, procurement, contracts, assets, etc., through the Data Central Platform significantly improves the accuracy and consistency of the data, and further enriches the dimensions of the data that can be analyzed.

2.1.2. Data Processing

Data processing includes data integration and cleaning, and data visualization. The former includes data integration, cleaning and standardization, which helps to obtain streamlined data that meets the analysis objectives and provides a data basis for indicator analysis and model building. The decision support platform provides data integration, consolidation, unified storage and other functions; data cleansing removes useless or missing attributes according to set screening and filtering conditions, or completes data supplementation and correlation according to specific conditions, so as to obtain data with a clear structure and strong usability; data standardization adopts standardization technology to process data according to the analysis goals and improves the comparability of data. Based on the results of data processing, visualization technology is used to present the data results vividly.

2.2 Intelligent Prediction of Budgeting Based on Association Rule Technology

In this section, association rule mining techniques will be applied to explore the intrinsic correlations and patterns in the process of budget preparation and execution from the appeal structured data, providing insights into the intelligent prediction and structural optimization of budgets.

2.2.1. Basic Concepts of Association Rules

The following describes the concept of association rules.

Transaction: has the property of being uniquely identified and refers to the set of all things that happen to an object in a process. Transactions have their own database, which is represented as

$D = \{t_1, t_2, \ldots, t_n\}$. In a transaction there is the concept of item, which belongs to an attribute field with a certain range of values, e.g., let $I = \{I_1, I_2, \ldots, I_m\}$ be a transaction, then each element in this set is an item. Then the set consisting of any k items in this set is an item set, and M is known as the item set length, which can be denoted as $k - item$.

Rule support: In a transaction database, for rule $X \Rightarrow Y$, the number of transactions that include both Y and X accounts for the percentage of all transactions, which can also be understood as the probability of X and Y appearing in the transaction at the same time, which can also be denoted as $Support\left(X \Rightarrow Y\right)$, as shown in equation (1).

$$Support(X \Rightarrow Y) = \frac{\left\| \left\{ T \mid (X \Rightarrow Y) \subseteq T, T \subseteq D \right\} \right\|}{D} \tag{1}$$

Item set support: the probability that a given item set appears in all transaction items in a transaction, denoted as $Support(X)$, as shown in equation (2).

$$Support(X) = \frac{\left\| \left\{ d \in D \mid X \in d \right\} \right\|}{\left\| D \right\|} \tag{2}$$

The percentage of occurrences in D is called the confidence level, which can be written as, as shown in equation (3).

Minimum support: the minimum value of the determination of support in a transaction, which can be interpreted as the degree of frequency of occurrence of the itemset in the transaction, denoted as $\min \sup$. If the support of a itemset is less than the minimum support, then the itemset is called a frequent itemset, denoted as $\min \sup \leq Support(X)$. Correspondingly, if the support of an item set is greater than the minimum support, then the set is called infrequent item set, which can be written as $\min \sup > Support(X)$.

Confidence: Let X and Y are both itemsets in the transaction set D, and for the association rule $X \Rightarrow Y$, then the percentage of occurrences of X in D is called the confidence, which can be written as $Confidence\left(X \Rightarrow Y\right)$, as shown in Equation (3) shown.

$$\begin{aligned} Confidence\left(X \Rightarrow Y\right) &= \frac{Support(X \cup Y)}{Support(X)} \\ &= \frac{\left\| \left\{ T \mid (X \cup Y) \subseteq T, T \subseteq D \right\} \right\|}{\left\| \left\{ T \mid X \subseteq Y, T \subseteq D \right\} \right\|} \end{aligned} \tag{3}$$

Minimum Confidence: i.e., the determined minimum of confidence in a transaction, which can also be understood as the probability or reliability of the production of a particular formulated rule in a transaction, denoted as $\min \ conf$.

Through the elaboration of the above concepts, it can be seen that confidence means the representation of the accuracy of a particular association rule, i.e., the probability or likelihood that the item X is contained in the transaction and also the item Y. Support, on the other hand, represents the probability of the occurrence of a particular association rule, which implies that both the associated parties X and Y will occur at the same time, i.e., the lower the support, the lower the likelihood of the occurrence of such a rule, which is a concept that is of little value from a business point of view.

Association rule discovery: In a transaction D, discover all association rules that satisfy the conditions of minimum support and minimum confidence.

2.2.2. Application of Association Rule Algorithms to University Budgets

Higher education budget data mining content is relatively rich, on the one hand, to reflect the extra-budgetary revenue and expenditure and budget revenue and expenditure, on the other hand, but also reflect the general budget revenue and expenditure and fund budget revenue and expenditure. It can be said that the college budget is a complete budget. Therefore, the accuracy of the budget of each department and each program in colleges and universities plays an important role in improving the level of budget management of the whole university. Therefore, in order to make scientific and reasonable adjustments to the budget amounts and ratios of each department and lay a solid foundation for the work of the coming year, the budget data of the previous years are used as a reference, and the data mining of unreasonable budget data is emphasized according to the requirements, so as to get the unreasonable and unscientific correlation rules in the budget. On the other hand, based on the budget data mining results can also be declared for the coming year has not yet to prepare the budget for the project associated with matching, so as to facilitate the budgeting work for the project to find a scientific and reasonable budget amount or realization of the way.

In addition, the association rule data mining technology also needs further improvement, such as optimizing the anti-interference ability of random data. Although the expenditure and income of the budget in each department of the university has a certain connection, and the fluctuation of its ratio can be followed, but a small number of departments due to policy changes or other factors to produce a large fluctuation of the situation also exists, so the optimization of its data mining technology is also of great significance, to lay the foundation for an accurate and reasonable budget.

Based on the multidimensional association rule data mining algorithm, in the budget expenditure data, the corresponding datasets are established to facilitate on-line analysis. These datasets are: unit, program, parent department, tranche, and year. Corresponding dimension tables are created to facilitate the observation of the data from different dimensions. The five corresponding dimension tables are: parent department dimension, year dimension, grade dimension, project dimension, and unit dimension, and the star model for the multidimensional dataset is shown in Figure 1.

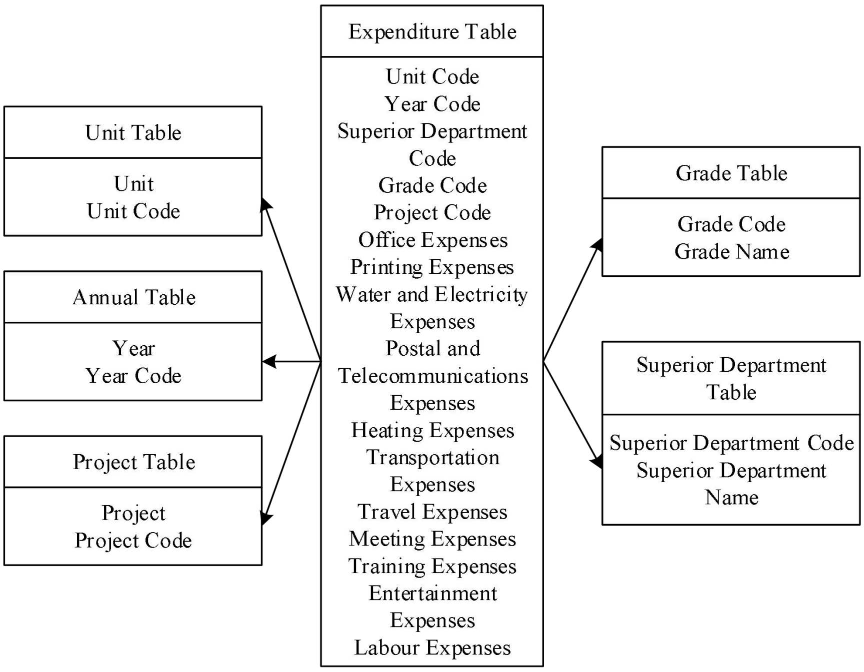

Figure 1. The star model of multi-dimensional datasets.

After on-line analysis and processing, data mining of their data using multidimensional association rule techniques shows that:

(1) There is an association between different departments. The highest amount of public expenditure cost is the comprehensive unit, followed by the general administration, and the smallest amount of expenditure is the institution.

(2) There is a correlation between the expenditure costs. In travel expenses as well as conference expenses, transportation expenses occupy a certain proportion, while in office expenses, utilities and heating expenses occupy a larger proportion.

(3) There is a correlation between different projects. In the total project cost expenditure, national-level projects are significantly higher than provincial and municipal-level projects, and the itemized expenditures in national-level projects are also higher compared with provincial and municipal-level projects.

The association rule data mining through the 2018-2024 expenditure data of a university can optimize the financial expenditure structure and improve the efficiency of budget expenditure in the following aspects.

(1) Compression of the expenditure of general meetings. From the data mining results, we can see that the substantial increase in travel expenses is due to the fact that training and meetings are held too frequently, so reducing unnecessary meetings and training will play a role in reducing financial expenditures.

(2) Focus on saving and avoiding waste. A large portion of office expenses comes from utilities and heating costs, which indicates that a few departments have more wasteful office space.

(3) Scientific and reasonable allocation of the unit's budget based on data mining. According to the 2013-2017 part of the financial budget arrangement amount, most of the departments of the budget is based on the “base” for the formulation, this formulation method is not scientific, contrary to the principle of “the right to do things + the right to finance”, resulting in the structure of the expenditure of financial funds. Unreasonable. Therefore, data mining technology should be used to guide the financial budget.

(4) Based on the data mining technology, the basic expenditures of budget units and the standards and quotas of each budget should be scientifically assessed.

2.3. Algorithm Design of Financial Decision-making System for Universities Based on Data Mining

Association rule mining reveals explicit patterns and associations in budget data, providing important clues to optimize the budget structure. In order to build a more comprehensive and intelligent financial decision support capability for universities, this section further explores more complex data mining algorithms - decision tree classification (ID3) and density clustering (DBSCAN) - to design the core algorithm module of the financial decision system for universities.

2.3.1. ID3 Algorithm

The goal of the ID3 algorithm is to generate effective classifiers in the form of decision trees from a dataset.

A decision tree is a tree structure that uses the attributes of a sample as nodes and the values of the attributes as branches.

The inputs to the algorithm are (1) a collection of data objects, (2) descriptive attributes of each object, and (3) categorization attributes of each object. Classification analysis method is also the main method of data mining. Classification mainly solves the problem of obtaining a function or model (classifier) for classification using a set of training samples. The inputs to the algorithm are: data set; data object commonality values; object characteristic values.

Classification among various analytical methods is also the main method of data mining. It first classifies the massive data set initially, selects the training samples that have been classified, then makes full use of the selected data mining algorithms to construct classification functions (i.e., classification models) on these training sample sets, and then carries out the secondary classification of the remaining large amount of data according to these models to achieve the purpose of completely classifying the massive data.

The flow of the algorithm is as follows:

First of all, initialization, initialization to set the node and data set a pair of parameters, the decision tree or sub-decision tree to be generated by the root node that becomes the node.

The dataset, which is the set of all objects of the data.

The second step is to call the iteration process from the root node, and the iteration process is as follows:

First the conditions of iteration are checked, including the constraints of iteration and the dataset. If the data object in the dataset contains only one of the values of all the categorized attributes, then the node can be used as a leaf node and can eventually be returned as a decision tree.

For the calculation of entropy value the currently used dataset is used to calculate the entropy value of each descriptive attribute.

$$H(S) = -\sum_{i=1}^{m} P(u_i) \log P(u_i) \tag{4}$$

With S as the set of training samples, then $P(ui)$ is the probability of occurrence for category i

$$P(u_i) = \frac{|u_i|}{|S|} \tag{5}$$

Calculation of information gain:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{6}$$

2.3.2. DBSCAN Algorithm

In this paper, in using DBSCAN algorithm to cluster the data through clusters, the maximum set of density connected points is obtained through the density set, the region where the data to be queried has sufficiently high density is classified into clusters and clustered points of arbitrary shape can be found in the spatial database in the presence of noise and null values.

The purpose of the experiment is to cluster students by the amount of fees owed to them and obtain clusters by different classes of fees owed. Decision makers can use the results of this analysis to get a holistic view of the impact of samples and parameters on the prediction of student fee payment, so that they can be set flexibly in the actual mining.

In this paper, we choose the amount of student fee payment as the research object, select a part of it as the training sample, and the rest as the test data.

The flow of DBSCAN algorithm analysis is shown in Figure 2.

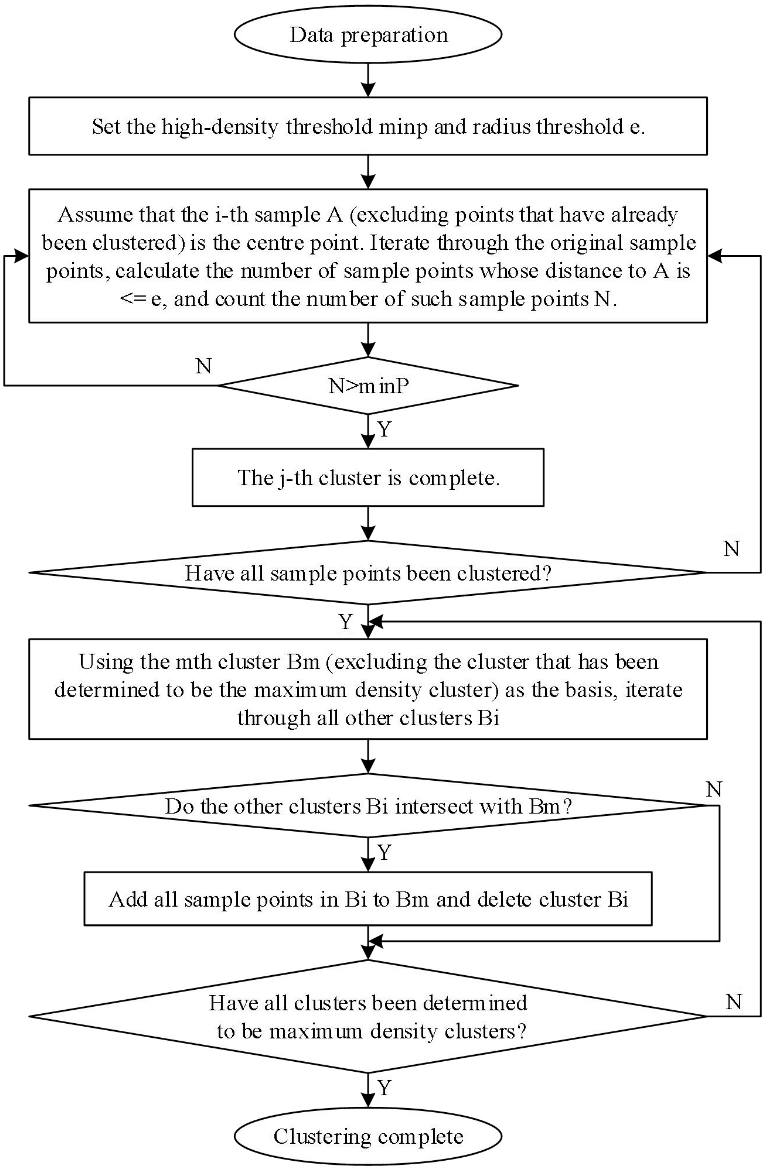

Figure 2. the process of DBSCAN algorithm analysis.

3. Performance Validation and Association Rule Mining Analysis of Intelligent Integration Methods for Higher Education Financial Data

In Chapter 2, we constructed a framework of algorithmic models for university financial data processing, budget intelligent forecasting and decision-making system based on association rules (Apriori), ID3 decision tree and DBSCAN clustering, which provides a methodological foundation for the core research objective - optimizing university financial structure. In order to verify the effectiveness and performance advantages of these models and methods in the intelligent integration of actual university financial data, this chapter will conduct a systematic experimental analysis, focusing on evaluating their read/write efficiency, algorithmic operation efficiency, and the value of the mined association rules.

3.1. Comparative Experimental Analysis of the Performance of Intelligent Integration of Financial Data in Higher Education Institutions

In order to verify whether the methods designed in this paper meet the needs of intelligent integration of university financial data, this paper analyzes the above methods experimentally. The final experimental results are presented in the form of a comparison between the intelligent integration method of university financial big data based on DMSP-OLS and NPP-VIIRS, the intelligent integration method of university financial big data based on SuperMap, and the intelligent integration method of university financial big data based on association rule mining designed in this paper.

3.1.1. Comparison of Reading/Writing Efficiency

The amount of college financial data from 10bit to 10000bit is randomly selected, and the read efficiency and write efficiency are taken as the benchmark indexes of data intelligent integration. And the read/write efficiency of the intelligent integration method of college financial big data based on DMSP-OLS and NPP-VIIRS, the read/write efficiency of the intelligent integration method of college financial big data based on SuperMap and the read/write efficiency of the intelligent integration method of college financial big data designed in this paper based on the association rules and ID3 decision tree algorithm mining are compared, and the experimental results are shown in Table 1.

Table 1. Comparison of read and write efficiencies of each algorithm.

Data volume/bit	Reading efficiency (bit/s)			Writing efficiency (bit/s)		
	DMSP-OLS&NPP-VIIRS	SuperMap	OURS	DMSP-OLS&NPP-VIIRS	SuperMap	OURS
10	17.12	45.06	47.74	45.06	98.64	156.12
50	13.17	40.18	46.37	48.32	93.43	173.07
100	13.92	37.39	47.97	59.23	89.27	208.15
500	13.01	35.39	43.51	49.30	78.32	204.84
1000	15.48	28.62	42.96	39.78	65.86	259.55
2000	16.53	24.56	45.48	42.44	61.25	235.98
5000	12.08	22.58	44.99	35.89	51.23	202.09
8000	13.66	13.71	45.59	30.37	35.33	377.44
10000	13.06	13.36	43.78	39.81	30.76	451.04

In the process of intelligent integration of college financial big data, financial data writing efficiency and reading efficiency directly affect the integration efficiency. Under the condition of the same amount of college financial data, the higher the reading efficiency and writing efficiency, the better the effect of big data intelligent integration.

Under the condition that all other conditions are the same, after using the university financial big data intelligent integration method based on DMSP-OLS and NPP-VIIRS, the read efficiency and write efficiency do not change with the increase of the university financial data volume. Among them, the read efficiency fluctuates in the range of 12bit/s~18bit/s, and the write efficiency fluctuates in the range of 30bit/s~60bit/s. It can be seen that the read and write efficiency of this method is relatively low and cannot meet the demand for intelligent integration of university financial big data.

After using the intelligent integration method of university financial big data based on SuperMap, the read efficiency and write efficiency change with the increase of the amount of university financial data. Among them, the read efficiency fluctuates in the range of 13bit/s~46bit/s, and the write efficiency fluctuates in the range of 30bit/s~99bit/s. The more the amount of financial data in colleges and universities, the lower the integration efficiency, which cannot provide large-scale data integration services for colleges and universities, and needs to be further optimized.

Using the intelligent integration method of college financial big data based on association rule mining designed in this paper, the reading efficiency and writing efficiency will not decrease with the increase of college financial data volume, and the reading efficiency is always above 42bit/s, and the writing efficiency is always above 156bit/s. It can be seen that the method can provide large-scale data integration services for colleges and universities, which plays an important role in the development of college and university finance.

3.1.2. Algorithm Runtime Comparison

This section mainly analyzes the running time of the algorithm and the number of redundant rules to verify the superiority of the algorithm in this paper. The financial data of colleges and universities in the previous section is continued to be taken, and the data volumes of 1000 and 5000 are selected respectively, and the support degree of the association rule algorithm is set to be 0.1, the minimum confidence degree is set to be 0.5, the enhancement threshold is set to be 1.3, and the threshold of the interest degree is set to be 0.1. The experimental results of the different algorithms are shown in Fig. 3.

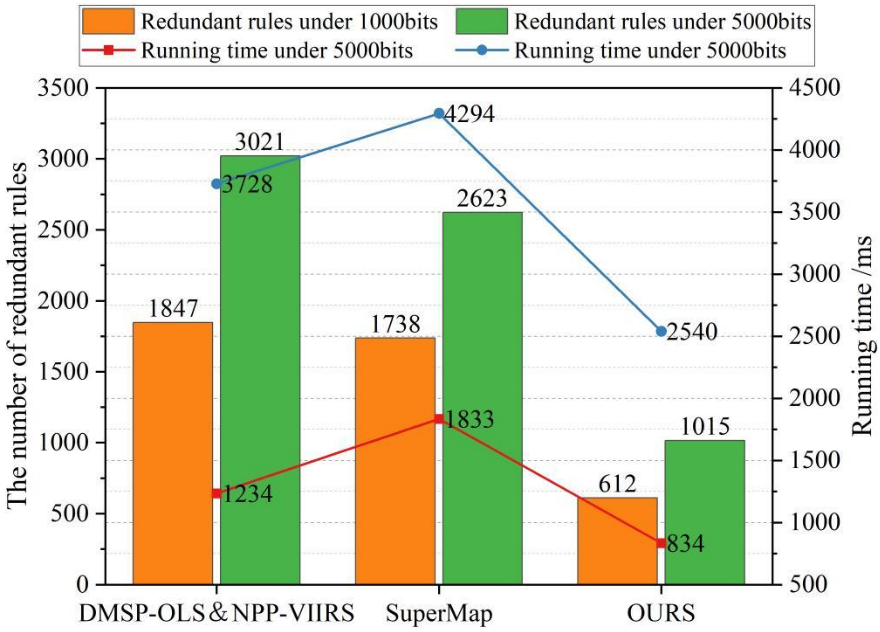

Figure 3. Comparison of algorithm redundancy rules and running time.

As can be seen in Figure 3, the association rule algorithm based on ID3 algorithm proposed in this paper can greatly reduce the number of redundant rules, and the running time of the algorithm is also shortened, which improves the efficiency of rule mining. In 1000 pieces of data, this paper's algorithm generates only 612 redundant rules, which is significantly lower than the 1847 rules of DMSP-OLS & NPP-VIIRS and the 1738 rules of SuperMap. Under 5000 pieces of data, the redundant rules of this paper's algorithm are 1015, while DMSP-OLS & NPP-VIIRS and SuperMap are as high as 3021 and 2623, respectively, which reduces the redundancy by about 60%. When processing 1000 pieces of data, the algorithm in this paper takes 834ms, which is 32% faster than the 1234ms of DMSP-OLS & NPP-VIIRS and 54% faster than the 1833ms of SuperMap. When processing 5000 pieces of data, the algorithm in this paper takes only 2540ms, which is much lower than the 3728ms of DMSP-OLS & NPP-VIIRS and the 4294ms of SuperMap, and the efficiency is improved by about 30%-40%.

3.2. Financial Linkage Rules for Higher Education Institutions

The ID3-based algorithm proposed in this paper mines a total of 415 valid rules after eliminating the irrelevant 1015 redundant rules under a data volume of 5000. Some of the rules are shown in Table 2. Some of the important descriptors about the rules are as follows: the number of incumbents (T50), office expenses (T1), printing expenses (T2), postage and electricity (T3), travel expenses (T4), conference expenses (T11), personnel expenses (T5), welfare expenses (T6), official transportation subsidies (T12), utility expenses (T7), heating expenses (T8), and training expenses (T9).

Table 2. Strong Association Rules for University Finance (Part).

Strong association rule	Confidence coefficient	Improvement coefficient	Interest coefficient
T50→T1	0.6008	1.8459	0.3969

T50→T2	0.4399	1.6451	0.3666
T50→T3	0.4973	1.8883	0.4095
T50→T4	0.5665	1.8758	0.4325
T50→T12	0.5393	1.7724	0.3695
T1→T2	0.4754	1.779	0.481
T1→T3	0.5665	1.8484	0.4331
T1→T4	0.5108	1.6904	0.3372
T11→T1	0.4522	1.6252	0.4155
T11→T3	0.6125	1.9819	0.4158
T4→T12	0.4609	1.7553	0.4886
T9→T4	0.5243	1.7682	0.4556
T7→T8	0.5479	1.5495	0.4247
T11→T9	0.4681	1.5672	0.4886
T5→T6	0.4775	1.5455	0.4094

Table 2 demonstrates 15 strong association rules mined based on the ID3 algorithm for college finance, covering 12 core factors such as the number of incumbents (T50), office expenses (T1), and travel expenses (T4). The confidence level of all the rules is higher than 0.43, indicating that the reliability of the rules is strong. The highest is T11→T3 (conference fee→postage fee, confidence level 0.6125), and the lowest is T50→T2 (number of employees→printing fee, confidence level 0.4399). All rules boosted >1 (range 1.5455~1.9819), showing a significant positive correlation between the antecedent and the consequent. The strongest correlation was T11→T3 (lift 1.9819), indicating that for every 1-unit increase in conference fees, the probability of postal fee growth increased by 98.19%. Measuring the importance of the rule, the highest is T4→T12 (travel expenses→official transportation subsidies, degree of interest 0.4886), reflecting the strong binding relationship between travel and transportation subsidies.

4. Research on the Application of Data Mining Model-Based Rating of Poor Students in Colleges and Universities

The experimental results in Chapter 3 show that the intelligent integration method of university financial data based on association rule mining and ID3 decision tree proposed in this paper is significantly better than the comparison methods in terms of read/write efficiency, running time and rule mining quality, which provides efficient and reliable technical support for optimizing the financial structure of universities. In order to further demonstrate the practical application value of the constructed data mining model in the refined management of colleges and universities, this chapter will focus on the key area of student financial aid management in colleges and universities, and use the DBSCAN clustering and ID3 decision tree model to carry out a scientific rating research on students with financial difficulties.

4.1. Rating Results of the Model

The experiment uses preprocessed data of 4000 undergraduate economically disadvantaged students from the College of Physics and Optoelectronics of a university in the class of 2024, which are divided into a training set and a test set in the ratio of 8:2.

Fig. 4 shows the ROC curve of the DBSCAN model with three classifications, when i=1, TPR_1 represents the proportion of samples with the actual grade of "general difficulty" predicted as "general difficulty", and the larger the TPR, the better. FPR_1 represents the proportion of samples with actual grades of "Particularly Difficult" and "Not Difficult" that are predicted to be "Moderately Difficult", with the smaller the FPR, the better. It can be seen from the figure that the area of the curve for each category is more than 98% (the AUC of "not difficult" is 0.991, the AUC of "average difficulty" is 0.989, and the AUC of "very difficult" is 0.997), indicating that the accuracy of the model is high and the classification effect is better.

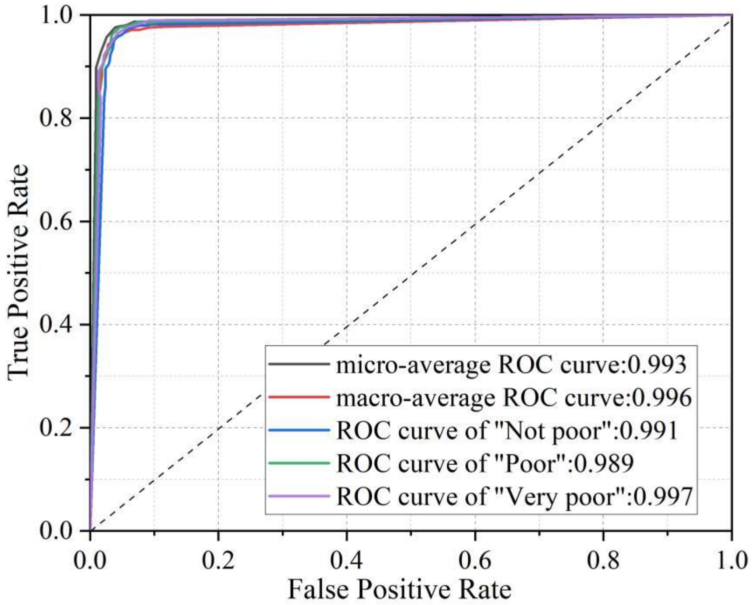

Figure 4. The ROC curve of the DBSCAN model.

4.2. ID3 Decision Tree-Based Modeling for Identifying the Grade of Poor Students

4.2.1. Classification Attribute Entropy Value and Information Gain

The ID3 decision tree algorithm was used to construct a decision tree model for the training set of poor students in the previous section. A total of 317 students from economically disadvantaged families were involved in the training set, and the classification attributes involved were A1: "whether they are orphaned or disabled", A2: "whether they are single-parent families", A3: "whether they are children of martyrs", A4: "family per capita annual income level", A5: "whether they are affected by natural disasters", A6: "whether they encounter sudden accidents", A7: "whether members are unemployed" and A8: "family debt level", and the eight categorical attributes are all from the "Questionnaire on Family Financial Difficulties of Students in Colleges and Universities". There is a correlation between the annual household income, the number of family members and the per capita annual income in the questionnaire, and the per capita annual income is selected as an indicator to reflect the economic situation of the family.

First, the student data were analyzed and the entropy of the system was calculated to be 0.957 based on the five levels in order to determine the root node of the decision tree. When the entropy of the selected system is 0.957, the entropy values of the eight attributes are calculated to obtain the corresponding information gain. Table 3 shows the entropy value and information gain corresponding to the categorized attributes of the root node.

Table 3. The entropy and information gain corresponding to the classification attribute.

Classification attribute	Entropy value	Information gain
Whether one is lonely and disabled	0.867	0.143
Whether it is a single-parent family or not	0.852	0.097
Whether they are the children of martyrs	0.956	0
The annual per capita income level of the family	0.787	0.247
Whether affected by natural disasters	0.913	0.162
Whether encountered any unexpected incidents	0.809	0.071
Whether any members unemployed	0.944	0.038
Household debt level	0.842	0.097

The highest entropy value is A3: whether you are a martyr's child or not, the entropy value is 0.956, which is close to the initial entropy of the system of 0.957, indicating that the attribute has extremely weak ability to distinguish classification, and the information gain is 0. The lowest entropy value is A4: the per capita annual income level of the family, and the entropy value is 0.787, reflecting the lowest uncertainty of this attribute and the largest contribution to the classification. The root node selection is based on A4: per capita annual income with the highest information gain (0.247) as the root node of the decision tree, which is in line with the common sense that "economic income is the core indicator of poverty".

Secondary split attributes: A5 (natural disasters) and A1 (orphaned) have a gain of >0.14, which are used as key sub-nodes to refine the poverty level of the “middle/low-income” group. Invalid attribute elimination: A3 (children of martyrs) can be ignored in the actual model to avoid overfitting due to the gain of 0.

From the data in Table 3, a bar-and-line graph is drawn to visualize the entropy value and information gain of the root node, as shown in Figure 5.

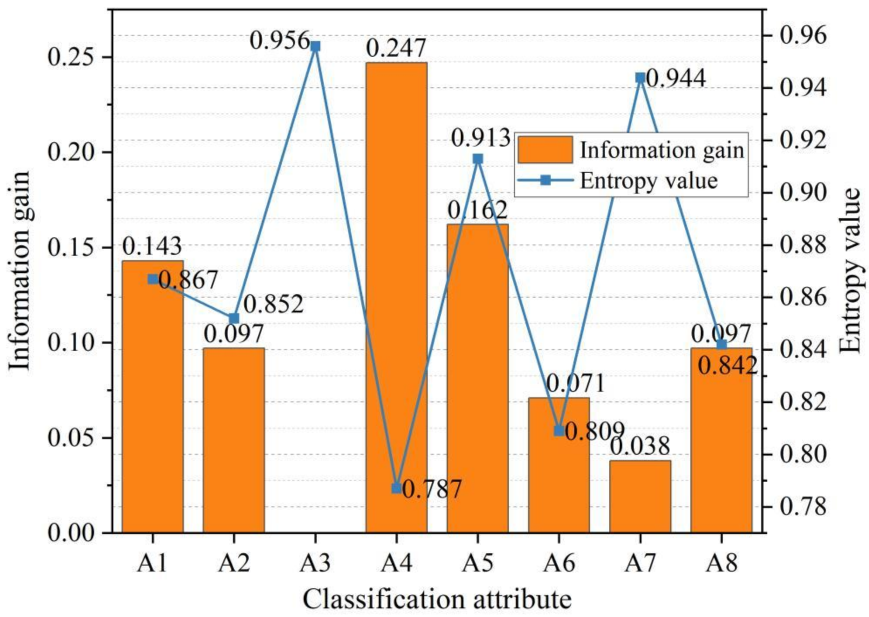

Figure 5. Root node entropy value and information gain.

4.2.2. Decision Tree Model

As can be seen from Fig. 5, “per capita annual household income level” has the highest information gain among the attributes, so this categorized attribute is the root node of the decision tree model, and thereafter each attribute value induces a branch, and the recursive operation is performed on each branch to derive the other branch nodes and obtain the final decision tree model as shown in Fig. 6.

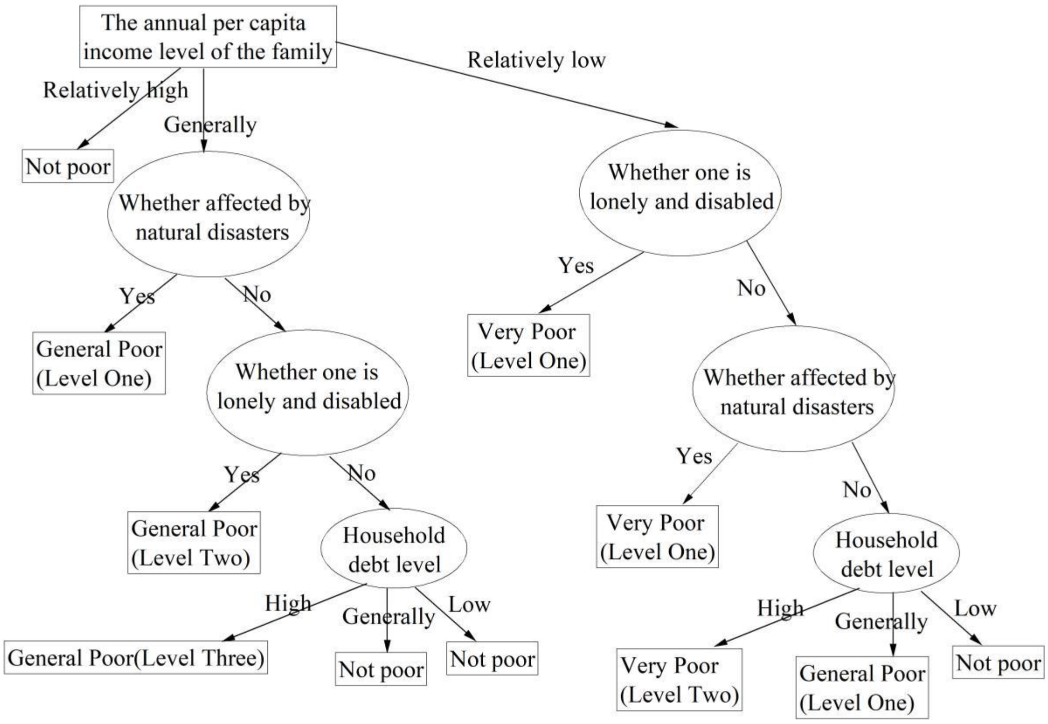

Figure 6. Decision tree model.

It can be seen from the generated decision tree that when determining the financial support level of poor students, firstly, the "per capita annual income level of the family" should be considered, and when the attribute value is "high", it usually has more of the same other attribute values, so there are fewer branches. Among them, there are more branches corresponding to the attribute values of "medium" and "low", because the proportion of the corresponding number of people in each poverty level is different, from the third level of general poverty, the second level of general poverty, the first level of general poverty to the second level of special poverty, and the first level of special poverty, the number of people has been decreasing, and the difference in the attributes involved will be greater, so a more complex branch will be formed.

The branch complexity is reflected in the middle income (A4="middle") and low income (A4="low") paths, which have more branches, which is in line with the diversity of poverty levels (general poverty level 3 → extreme poverty level 1). The attribute priority is Income> Natural Disaster> Orphaned and Disabled> Liabilities, which exactly matches the information gain ranking (A4>A5>A1>A8).

5. Conclusion

This study empirically verifies the key role of data intelligence technology in optimizing the financial structure and budget performance management of universities by constructing a multi-dimensional data mining model system, and the main conclusions are as follows

(1) The intelligent integration method based on association rules and ID3 decision tree has a stable read/write efficiency higher than 42bit/s, which is 200% higher than the traditional method.

(2) The optimized association rule mining of ID3 algorithm reduces the running time by 30%-54% (only 834ms for 1000 data), and the percentage of effective rules is increased to 79% (415/5000 rules).

(3) ID3 decision tree quantifies attribute prioritization with information gain (e.g., annual per capita household income gain value of 0.247), which significantly improves the classification accuracy (AUC>0.98 for poor student classification).

(4) Multidimensional association rules generate 415 strong rules to provide quantitative basis for budget quota formulation (e.g., T50→T1 confidence level 0.6008).

(5) The decision tree model for poor students uses income level as the root node (gain 0.247), incorporates attributes such as natural disasters and orphanhood, and has a grading error rate of <2%.

Funding

This research was supported by the Innovative Young Talent Program in Regular Institutions of Higher Learning in Guangdong Province, Research on the Innovative Development of Blockchain Industry in the Guangdong-Hong Kong-Macao Greater Bay Area (Program No.: 2019WQNCX153); Research on the Construction and Application of the Performance Evaluation System for Local Government Special Bonds under the Background of High-Quality Financial Management in Guangdong Province (Project Number: 2024WTSCX175).

References

1. Zhao, X. J. (2020). An Analysis of College Education Cost Accounting under the New Government Accounting System. Frontiers in 627 Educational Research, 3(14), 1-20.
2. Ahmed, I. E. (2019). Bridging the gap between governmental accounting education and practice. Accounting, 5(1), 21-30.
3. Fischer, M. (2018). The need for performance management systems in public higher education. Journal of Higher Education Theory and Practice, 18(6), 79.
4. Camilleri, M. A. (2021). Using the balanced scorecard as a performance management tool in higher education. Management in Education, 35(1), 10-21.
5. Tanveer, M., Karim, A. M., & Mahbub, A. (2018). Higher education institutions and the performance management. Library Philosophy and Practice (e-journal), 2183.
6. Ho, A. T. K. (2018). From performance budgeting to performance budget management: theory and practice. Public Administration Review, 78(5), 748-758.
7. Alhasnawi, M., Said, R. M., Daud, Z. M., & Muhamad, H. (2023). Budget participation and managerial performance: bridging the gap through budget goal clarity. Advances in Social Sciences Research Journal, 10(9), 187-200.
8. Bahadori, M., Taleb Nia, G., Ranjbar, M. H., & Barkhordari Ahmadi, M. (2022). Implementation of performance-based budgeting using the combined technique of Best-Worst Method (BWM) and robust optimization. Advances in Mathematical Finance and Applications, 1(1), 263.
9. Otalor, J. I., & Oti, P. A. (2017). Budget Participation, Managerial performance and Budgetary slack: The intervening role of Budget pressure and monitoring. Accounting and taxation review, 1(1), 9-37.
10. Widiawati, D., & RS, T. Y. (2019). Effect of budget participation on managerial performance mediated by job satisfaction and organizational commitment. Journal of Business Studies, 4(1), 54-66.
11. Li, H., & Guo, Y. (2022). Performance management of university financial budget execution relying on comprehensive budget management strengthening model. Advances in Multimedia, 2022(1), 4758609.
12. Liu, P. (2024). A Brief Discussion on the Three Major Modules of University Financial Management and the Current State of Budget Performance Management. Journal of Modern Business and Economics, 1(1).

13. Abdulle, M. A., Kulmie, D. A., & Hussein, M. S. (2023). Improving job performance through budget participation: Case of higher education institutions. Journal of Entrepreneurship and Business Innovation, 10(2), 38.
14. Zhang, Y. (2022). Research On Performance Management Of Project Expenditure Budget In Colleges And Universities. Financial Engineering And Risk Management, 2022, 5 (1).
15. Guo, S. (2022). Research on Multi-Objective Intelligent Optimization of Financial Resource Management and Allocation Methods in Colleges and Universities. Mathematical Problems in Engineering, 2022(1), 1906031.
16. Zhao, L., & Zhong, W. (2025). Design and optimization of financial management information system in colleges and universities under the background of big data. Journal of Computational Methods in Sciences and Engineering, 25(2), 1906-1922.
17. Vinogradova, E. B., & Kalinina, O. V. (2022). Improving University Organizational and Financial Structures: Case of Digitalization in Russia. J. Legal Ethical & Regul. Isses, 25, 1.