Research on the Application of Improved Decision Tree Algorithm based on Information Entropy in the Financial Management of Colleges and Universities

Huirong Zhao
School of Economics and Management, Jiangsu College of Engineering and Technology, Nantong, 226000, China

***Abstract*—In the era of information technology, the work relies on information technology to generate a huge amount of data and information. Among them, the financial data information of universities is growing exponentially, and the manual method of organizing data and extracting key information can no longer meet the requirements of financial data management of universities. Taking the financial management of higher education institutions as an example, it is difficult to grasp the progress of financial budget execution with frequent and complicated daily expenditure and income problems, and then it is difficult to execute correct decisions in the management. The study uses information entropy as the decision basis of decision tree in the financial management of higher education institutions. The higher the value of information entropy generated in financial management, the higher the prediction accuracy of the decision tree. The metric calculation method is introduced to obtain the information entropy as well as the information gain rate to predict the likelihood of problematic events. The study validates the performance of the improved decision tree with a dataset that achieves a maximum accuracy of 95% in the experiment. With the higher prediction accuracy, for the university financial management system, a decision tree for financial warning is established and the link between the current month's financial expenditure and the warning mechanism is analyzed, and finally the two common decision tree algorithms, (Iterative Dichotomiser3，ID3) ID3 and (Classification and regression tree，CART) CART, are compared with the algorithm proposed in the study. The mean square error and the sum of squared error metrics are used to conclude that the algorithm proposed in the study has better performance. By improving the existing decision tree algorithm, the study proposes a decision tree model based on information entropy, which aims to help decision makers to quickly and accurately distill relevant data and make correct decisions in a large amount of information data for more rational financial management.**

Keywords*—Information entropy; financial management; decision tree; information gain rate; C4.5 algorithm; early warning structure

I. Introduction

Financial management in colleges and universities is one of the core management works of higher education institutions, and also an important guarantee for normal daily work of institutions. As the enrollment scale of colleges and universities is expanding, the financial data of colleges and universities also usher in an explosive growth. How to mine these data information and fully apply the useful information in the data for effective college financial management and financial decision making is the focus of current intelligent algorithms in the field of college management research [1]. Since the society has entered the information age, computer technology and information technology have been developed rapidly. The development of computer information technology has brought a large amount of data information. The large amount of data can provide more reliable basis in extracting key information, but at the same time the redundant information mixed with it also brings redundant workload.

In the present time when the intelligent technology is popular, the information extraction technology is more in the form of automation, so the construction and improvement of financial information system has become the key research content in the management of universities. With the continuous development and improvement of financial management information systems in universities, the managers' use of traditional financial statements and manual decision-making solutions can no longer meet the needs of university financial management informatization, not to mention the development goals of data and information in the intelligent information era [2]. With the increasing requirements of university management for management effectiveness, reasonable decision quality and decision speed become crucial. The use of advanced technology and intelligent decision-making tools can make more effective management decisions and solve various problems in the financial management of colleges and universities. The financial information system of colleges and universities has a large amount of financial information. The financial information contains many important data, and the relevant information extracted from the data can be used as the judgment basis for the financial management decision of colleges and universities. The current information management of university finance is far from meeting the development requirements at this stage, and the management of finance only stays on the surface basis, focusing only on the accounting situation of financial data and the supervision of data. With the speed of digital campus construction, the needs of university development and financial data are accumulated in large quantities, and the managers need not only to achieve the most basic financial management, but also to control the risks of university development to the minimum through financial management. Therefore, using intelligent algorithms for management decisions can not only provide new management directions for university financial management, but also predict other types of financial problems.

At present, many data processing techniques and intelligent algorithms have been applied to financial management and risk decision making in universities. For example, data warehouse technology, data mining technology, data analysis technology, etc. The study introduces digital and information technology into the financial management of universities to get the improved algorithm of decision tree about information entropy. The improved algorithm can reasonably use a large amount of data generated in the financial system of universities, extract key information for the sustainable development of universities, provide a scientific theoretical basis for the construction of universities, and then control the risks in the financial management of universities for the purpose of financial budget progress management as well as financial early warning management [3]. The research content firstly uses information entropy as the decision basis of decision tree in financial management of universities, and proposes the decision tree algorithm based on information entropy and analyzes it. The higher the value of information entropy generated in financial management indicates the higher the prediction accuracy of the decision tree. Then, the metric is introduced, and the information entropy and the information gain rate are calculated, and a model based on the improved decision tree algorithm is proposed for the application of college financial management, and the likelihood of problematic events is predicted using the model. Finally, the performance of the constructed improved decision tree model is verified by the data set.

II. Related Work

The construction of higher education institutions is the basis of social progress and development. In order to ensure the normal operation and development of colleges and universities, many scholars have conducted relevant research on the management of colleges and universities. Priya E (Priya E 20210 [4] proposed a breast cancer diagnosis model based on deep segmentation of residual network, and applied it in the initial detection stage of breast cancer. In order to detect and classify breast images, the model adopts decision tree classifier for operation. It is proved by relevant experiments that the model has excellent classification performance, and the accuracy reaches 98.86%. Al C (Al C 2021) [5] classifies the students' tendency to choose TVET after completing their studies through multiple decision tree models, and constructs different decision tree algorithm types. The experimental results show that the decision tree has a low misclassification rate in TVET classification. Researchers such as Mao L (Mao L et al. 2021) [6] conducted an evaluation and analysis of entrepreneurship education in higher education institutions, improved and optimized the algorithm on the basis of decision tree and fuzzy algorithm, and constructed an efficient entrepreneurship education evaluation system. Through empirical research, the evaluation system established in the study has a certain degree of effect in practical application. S Ziweritin team (S Ziweritin et al. 2021) [7] through the simulation experiment, the decision tree is trained and tested, and the accuracy of the model reaches 99.86%. Mabuni D and Babu S A (Mabuni D and Babu S A 2021) [8] found that the classification effects obtained in smaller datasets were arranged to be reliable, and therefore performed cross-validation with multiple random folds on the same dataset. The verification method is also tested experimentally, and the experimental results show that the classification accuracy of the decision tree with prime number folding is higher than that of the existing decision tree classification.

Ahmad N (Ahmad N et al. 2019) [9] collected public university leaders' views on the overall financial management and sustainability of institutions through questionnaires. According to the survey results, the leaders interviewed believe that the key to the sustainability of financial management is to make use of university resources. The financial management of the school takes the basic activities of the school as the management, which is of great significance to the financial management of public universities. Researchers such as Asandimitra N (Asandimitra N et al. 2020) [10] designed a conclusive causal relationship study, conducted multiple regression analysis on the research results through relevant statistical software, and found that teachers' financial management behavior has an important relationship with teachers' own financial knowledge and financial literacy. Cultivating teachers' financial self-efficacy is of great help to school financial management. The Mohammad M team (Mohammad M et al. 2021) [11] used the mean and standard deviation of the collected data as a measure of the importance of the research question in their research on the quality and impact of AI on classified hotel financial statements. Through the survey results, it is found that the hotel application AI performs linear regression analysis on financial information, analyzes the important factors affecting financial management, and plays an important role in the sustainability of financial management. Malikov VV scholars (Malikov V V 2020) [12] constructed the algorithm of financial management system in enterprise management. The system first determines the overall financial goals of the enterprise, allocates the corresponding financial management subsystems, and formulates corresponding financial management rules. In practical application, the algorithm of the system successfully avoids the problem of cash gap and optimizes the relationship between creditors. Researchers such as UJ A (UJ A et al. 2020) [13] have developed some data classification models for financial management decision-making, which include the feature selection stage and data classification stage of ant colony feature optimization. Through the validation of multiple algorithms in the same dataset, the proposed model has better performance than other algorithms.

In summary, decision trees are widely used in classification, forecasting and other fields. Financial management needs to make predictive behaviors for the future. However, there are few and immature studies on the combination of the two. Therefore, the research combines the two. A relatively complete research is conducted to help colleges and universities become more scientific in financial management.

III. Application and Improvement of Decision Tree Algorithm in Financial Management

A. Improvement Analysis of Decision Tree Algorithm based on Information Entropy

Decision tree algorithm can classify data purposefully, and find and extract valuable information from the data, thus helping decision makers to provide a choice path [14]. Based on the above advantages, decision tree algorithm is widely used in prediction models. In data mining technology, the good tree structure formed by the decision tree can highly summarize the sample data and correctly identify the old samples or different types of samples. Compared with other classification algorithms, decision tree has a classification rule that is easier to understand. In the classification task, the algorithm runs faster and has higher classification accuracy.

Common decision tree algorithms include ID3 algorithm, C4.5 algorithm and CART algorithm. In the late 1970s, J Ross Quinlan proposed the ID3 algorithm, which aims to reduce the depth of the decision tree and simplify the operation steps of the model. However, ID3 algorithm ignores the study of the number of leaves while reducing the depth of the tree. Therefore, C4.5 algorithm is proposed to improve ID3 algorithm. As a decision tree algorithm based on information entropy, C4.5 algorithm has the advantage that it can use the information gain rate to replace the information entropy as the selection standard, avoiding the defects of ID3 algorithm. As an improved decision tree algorithm, the basic process of the algorithm is: first, select the training set as the root node, and select the appropriate criteria to select the splitting attributes. Secondly, according to the difference of split attribute values, the training set is divided into several subsets and used as the first sub node of the root node. Then, treat these sub nodes as root nodes and repeat the above steps until the final classification conditions are met.

In practical application, when there are many data items, more complex decision trees will be generated, which will lead to uneven data and even loss of attributes, affecting the operation efficiency and classification accuracy of the algorithm [15]. In view of the shortcomings of decision tree algorithm, C4.5 algorithm is proposed to be used in the decision tree. C4.5 algorithm is an improved decision tree algorithm, which mainly operates on the basis of information entropy. When dealing with continuous lost data, tree pruning technology is introduced. This technology not only simplifies decision-making, but also improves classification accuracy. In addition, the research attempts to introduce a measurement method into information entropy. This method can be used for data reference at the same time when classifying, and finally an improved decision tree algorithm based on measurement is obtained. The flow chart of the improved algorithm's decision tree is shown in Fig. 1.

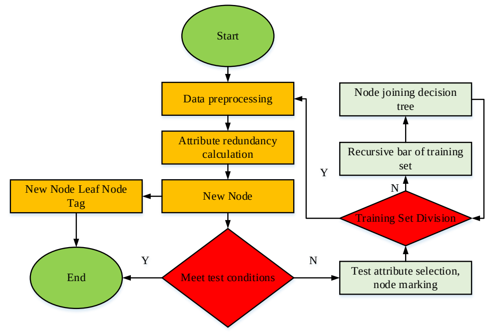


Fig. 1. Decision tree flow chart of improved algorithm

The decision tree model established by introducing the measurement method is essentially a recursive algorithm, which forms the branch conditions of the decision tree and adopts the threshold method. Assuming that $T=\{t_i\}$ the set of $A=\{a_i\}$ samples is taken as the feature space of the samples, the total number of samples is expressed by c, and the attributes of the feature space are expressed by m. The power space of the feature space is represented by $B=\{b|b\in A\}$, the misclassification threshold in the model is represented by β_e, and the cross misclassification threshold is represented by β_c. For the superset, it includes samples of multiple categories. If the sample attribute set threshold value in the superset x is the smallest, and the distance from the other category values in the typical Mahalanobis distance is the smallest, the sample will be classified into this category. Assuming that the sample is misclassified to other samples, and the proportion is greater than the cross misclassification, the two samples will be considered to be the same class and continue to be classified in the lower layer. The classification strategy adopted when the above attribute set is not found is the best classification situation of the model. When all samples are classified or can no longer be classified, the model classification process ends. In the model, the randomness of information entropy is used as a standard to select model samples to test attributes [16]. In the process of building a decision tree, an attribute is required to divide the data, so that the data class values in the child nodes are basically the same. When the data class values in the node are unevenly distributed, it means that the entropy of the node is small. The definition of entropy is shown in formula (1).

$$Entropy(p_1,p_2,\cdots p_n)=p_1\log_{2p1}-p_2\log_{2p2}-\cdots-p_n\log_{2pn} \quad (1)$$

In formula (1), "-" indicates that the logarithm of the score is negative; p represents the fraction in the definition of entropy. In fact, the entropy is positive, as shown in formula (2) in general.

$$Info\left\{\left[C_1, C_2, \cdots, C_n\right]\right\} = entropy\left[E_s\right] \tag{2}$$

In formula (2), it is the expression of the training set E_s; it is used to represent the samples in the training set C. All training sets are grouped, and the spatial features are used to $entropy\left[E_s\right]$ reduce the value. Assume that the training set is set with a new expected amount of information, and its expression is shown in formula (3).

$$new\_entropy(E_s, A) = \sum i \notin value_{(A)} = \left[\left|E_s\right| / \left|E_s\right|\right] entropy\left[E_s\right] \tag{3}$$

The reduced expectation value of the training set is the information gain of the feature space relative to the training set. If it is beneficial to the training set, the larger attribute of the information gain needs to be removed. The specific expression is shown in formula (4).

$$Gain\left[E_s, A\right] = entropy\left[E_s\right] - new\_entropy(E_s, A) \tag{4}$$

Equation (4) $Gain\left[E_s, A\right]$ is the expression of information gain. Information entropy is the average amount of information obtained after excluding redundant information, which is used to represent the degree of confusion of information. The calculation formula of information entropy is shown in Equation (5).

$$I(S_1, S_2, \cdots, S_m) = -\sum_{i=1}^{m} P_{i \log_2 (P_i)} \tag{5}$$

In formula (5), it is used to represent the sample set S, which is used to represent P_i the probability of any sample in $P_i = S_i / S$, and C_i. The measure of the amount of information is the size of the information gain. When all attributes are classified, the value of the information gain is first calculated, and then the node selected for the construction of the decision tree is determined. big. Therefore, the feature space branching obtains the information gain as shown in Equation (6).

$$Gain(A) = I(s_1, s_2, \cdots, s_m) - E(A) \tag{6}$$

In formula (6), it $I(s_1, s_2, \cdots, s_m)$ represents the expected information entropy of a given sample, and the calculation method of sample expectation is shown in formula (7).

$$E(A) = -\sum_{j=1}^{v} \frac{s_{1j} + \cdots + s_{mj}}{s} I(s_{1j}, s_{2j}, \cdots, s_{mj}) \tag{7}$$

When calculating the information gain, it mainly depends on the attribute value, but in special cases, such calculation is also meaningless. Therefore, the division of nodes plays an important role in the construction of decision trees. The C4.5 algorithm optimizes the node division by means of the information gain rate, that is, the ratio between the information gain and the information entropy. When comparing, the amount of information of sample attributes is not the total amount. When the information gain is divided into subsets, there will be a certain degree of deviation in the value of its variables, which can be reduced by formula (8).

$$SplitInfo(S, v) = \sum_{i=1}^{m} \frac{\left|s_1\right|}{\left|s\right|} \times \log_2 \frac{\left|s_1\right|}{\left|s\right|} \tag{8}$$

Therefore, the gain rate expression is specifically as shown in Equation (9).

$$GrainRatio = -\frac{Grain(S, v)}{SplitInfo(S, v)} \tag{9}$$

By introducing information entropy and measurement methods, the C4.5 algorithm assumes a data set T in which different categories are represented by sets C, selects an attribute to divide the data set into multiple subsets, and the values of the attributes are all equal. If they are not the same, the dataset is divided into the same number of subsets. The probability of category selection occurring is shown in formula (10).

$$P(C_i) = \left|C_j\right| / \left|T\right| \tag{10}$$

Because the instances of the subset all take the value of the attribute, the probability of the occurrence of the attribute value and the probability of having the category condition are shown in formula (11).

$$\begin{cases} P(v_i) = T_i / T \\ P(C_j \left| V_j \right.) = c_{jv} / \left|T_i\right| \end{cases} \tag{11}$$

In formula (11), the probability of occurrence $P(C_j \left| V_j \right.)$ with C_j category conditions is expressed, and the probability calculation formula can be obtained to further obtain the calculation of category information entropy and category condition information entropy, as shown in formula (12).

$$\begin{cases} Info(C) = -\sum_j p(C_j) \log p(C_j) = -\sum_{j=1}^{k} \frac{P(C_i)}{|T|} \log \frac{P(C_i)}{|T|} = Info(T) \\ Info(\frac{C}{V}) = -\sum_j p(v_j) \sum_1 p(\frac{C_j}{v_j}) \log p(\frac{C_j}{v_j}) = -\sum_{i=1}^{n} \frac{|T_i|}{|V_i|} Info(T_i) = Info(T) \end{cases} \tag{12}$$

Finally, the information entropy of the attribute is obtained according to formula (8), as shown in formula (13).

$$Info(V) = -\sum_i p(v_i) \log p(v_i) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \log \frac{|T_i|}{|T|} = split\_Info(v) \tag{13}$$

The attribute information gain rate is obtained according to formula (9), and the specific expression is shown in formula (14).

$$gai\_ration(V) = n\frac{gain(v)}{split\_Info(v)} \quad (14)$$

B. The Application of the Improved Decision Tree Algorithm in the Financial Management Model of Colleges and Universities

The above research method introduces the decision tree algorithm based on information entropy, and obtains the specific calculation method of information gain rate. This calculation method is very important in practical application. The size of the decision tree can be judged through the calculation of information entropy. Part 2.2 will give an overall description of the financial management of colleges and universities, and apply the decision tree to it. In college financial management, college budget management and budget performance management play an important role. The budget management of colleges and universities can make detailed use plans for financial income and use according to financial and teaching contents [17]. The flow chart of budget performance management is shown in Fig. 2.

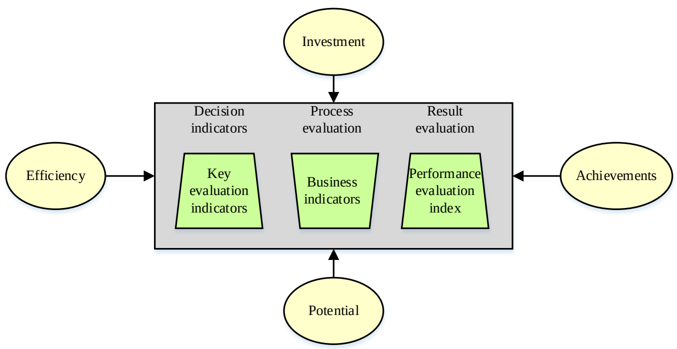


Fig. 2. Flow chart of budget performance management

The data in the process of decision tree calculation needs to be processed, so it is necessary to select a more appropriate algorithm to build the decision tree model. After the decision tree is built, it is pruned, that is, the decision tree is optimized. The main purpose is to eliminate isolated nodes and noise, cut meaningless branches, and make the decision tree have the ability to correctly classify [18-20]. The pruning idea in the algorithm firstly sets the node of the tree as the number t of $n(t)$ samples that reach the node; to $e(t)$ represent the number of samples that can reach the node but does not belong to the node; the ratio of error samples at the node is expressed as $r(t)$. The probability of repeating the experiment multiple times is the probability of occurrence $e(t)$, and the confidence interval of the wrong sample ratio can also be obtained. The confidence interval is expressed in CF, and the value of the confidence interval can control the degree of pruning. The higher the value of the confidence interval, the lower the degree of pruning. In C4.5 algorithm, the default value of confidence interval is 0.25, and the samples obey binomial distribution. The calculation process of the decision tree algorithm is integrated and calculated in the Clementine platform. The decision tree 4.5 algorithm is used to extract, process and mine the database data, and then establish a decision tree of financial budget to achieve early warning analysis of financial budget. The overall framework of early warning analysis is shown in Fig. 3.

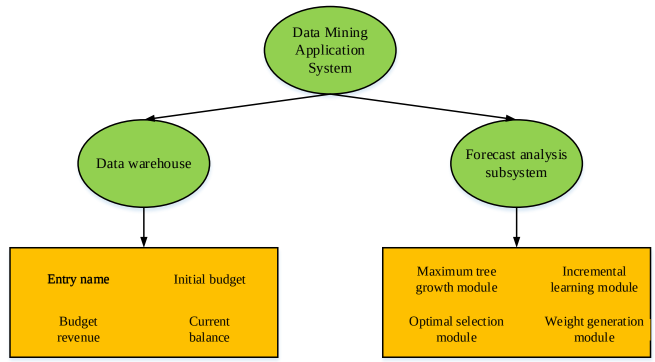


Fig. 3. Overall framework of early warning analysis

The data warehouse in Fig. 3 includes a lot of information, such as information of various departments of the university, opening balance, income and expenditure data, balance data, and so on. There are many problems in many data, such as clutter, duplication and incompleteness, so cleaning the data is a necessary step of the system. For data processing, there are methods of filling and deleting. When multiple attributes are missing in the sample, and the number of missing attribute values is more than one-tenth of the total number of attributes, the sample data will be deleted; when the mean values of the sample attributes are the same, then fill it. In the financial management early warning analysis sub-structure system, it also includes a maximum tree growth module, an incremental learning module, a pruning optimization module, and a weight generation module. Its early warning program first selects the said information from the database for preprocessing, and saves the data of the financial budget report to the data warehouse; the data warehouse is an organization method for storing data. Then, the data is classified by the C4.5 algorithms to obtain new knowledge; the new knowledge generated includes the weight generation module, which uses the weight distribution method to assign corresponding weights to each node on the decision tree value, and an incremental learning module. The incremental learning module combines incremental learning with the C4.5 algorithms, and uses incremental data and learning methods to modify the generated decision tree; output to help decision makers make corresponding countermeasures. The details are shown in Fig. 4.

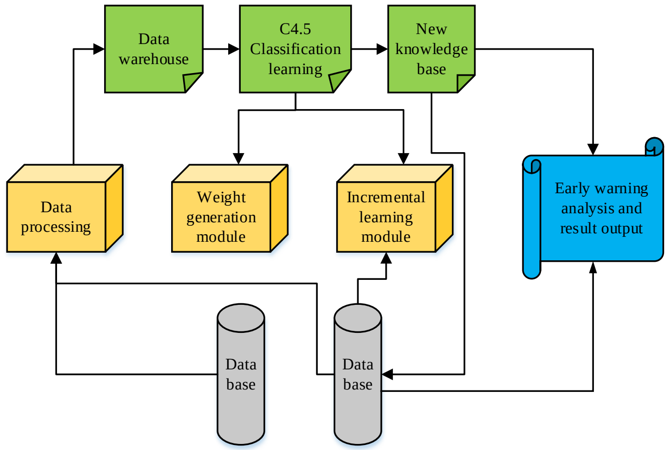


Fig. 4. Data flow chart of early warning analysis system

The construction of financial budget data warehouse is established from the three-level model of determining data source, data preprocessing and data warehouse. The data sources are all from the financial budget database of universities, and the data warehouse is studied according to the theme of decision tree; data preprocessing cleans the data, eliminates noise and irrelevant attributes, and converts the data; the three-level model of the data warehouse is established. The three-level model of the data warehouse, the first is the conceptual model, which determines the boundary of the financial budget model, extracts the data information required by the financial budget, determines the main performance indicators of the model, and defines the relevant data dimensions. The second is the logical model. The logical model is designed according to the data table division, granularity, financial budget execution progress and relational model. The third is the establishment of the physical model, the process of designing the physical database with the data organization form stored in the data warehouse as the main body.

IV. Financial Management Application Analysis of Improved Decision Tree Algorithm Based on Information Entropy

A. Analysis of the Application Effect of Financial Management based on Improved Algorithm

The research selects representative universities as the experimental object, and analyzes the data according to the accumulated data of the financial department of universities. In order to make the data mining information have the characteristics of high quality and high informatization, the selected samples should be small and representative. Table I shows the experimental data of mining 2022 college financial data after preprocessing.

TABLE I. College 2021 Financial Data Preprocessing Results

Project	Contrast ratio	Uniformity	Relevance	Entropy
T1	379.0191	19.3456	-0.0001	199.2554
	369.2883	19.2463	0.0000	209.5737
	249.2672	19.6178	-0.0001	179.8338
	489.4266	29.6728	0.0000	239.8373
T2	29.3567	9.2728	-0.0038	39.2628
	19.3638	9.2627	-0.0038	39.2727
	19.7378	9.2727	-0.3385	39.8262
	19.3665	9.2636	-0.4859	39.6528
T3	1549.2669	0.2627	-0.5049	129.3565
	1702.6229	0.2739	-0.4860	119.3743
	2489.2529	0.7283	-0.5050	209.2728
	1289.2536	0.2738	-0.4960	109.2727
T4	1039.2738	1.2738	-0.4950	18.7279
	501.2738	2.3730	-0.4951	18.2829
	203.3637	3.3738	-0.4489	8.1930
	929.2727	3.3829	-0.4960	10.8274
T5	1039.2627	1.2737	-0.1111	29.2833
	282.2719	1.1828	-0.5050	36.8189
	282.2728	0.8119	-0.5105	31.8279
	397.8192	2.2829	-0.1111	31.2828

The research uses the information entropy reduction of the C4.5 algorithms to determine the optimal branch variable and segmentation threshold of the decision tree. In the world of information data, the size of the entropy value can represent how much information is transmitted. If the entropy value is higher, it means that the transmission information is more; on the contrary, the transmission information is less, the information entropy value is smaller. In financial management, it can be divided by payment method to indicate the entropy value of the event. In ordinary budget indicators, the payment methods are divided into two types: direct payment and authorized payment, and the two payment methods are guaranteed to have equal probability of opportunity, which makes it difficult to predict the result of the next financial payment method, and the two results are mutually exclusive. Independently, the entropy in this form is one bit. When the budget index amount is relatively large and the university only allows direct payment, the entropy value of the event division method is 0. Algorithm C4.5 takes the decreasing speed of information entropy as the basis for judging the progress of budget execution, and the decision tree for getting early warning of budget execution progress is shown in Fig. 4.

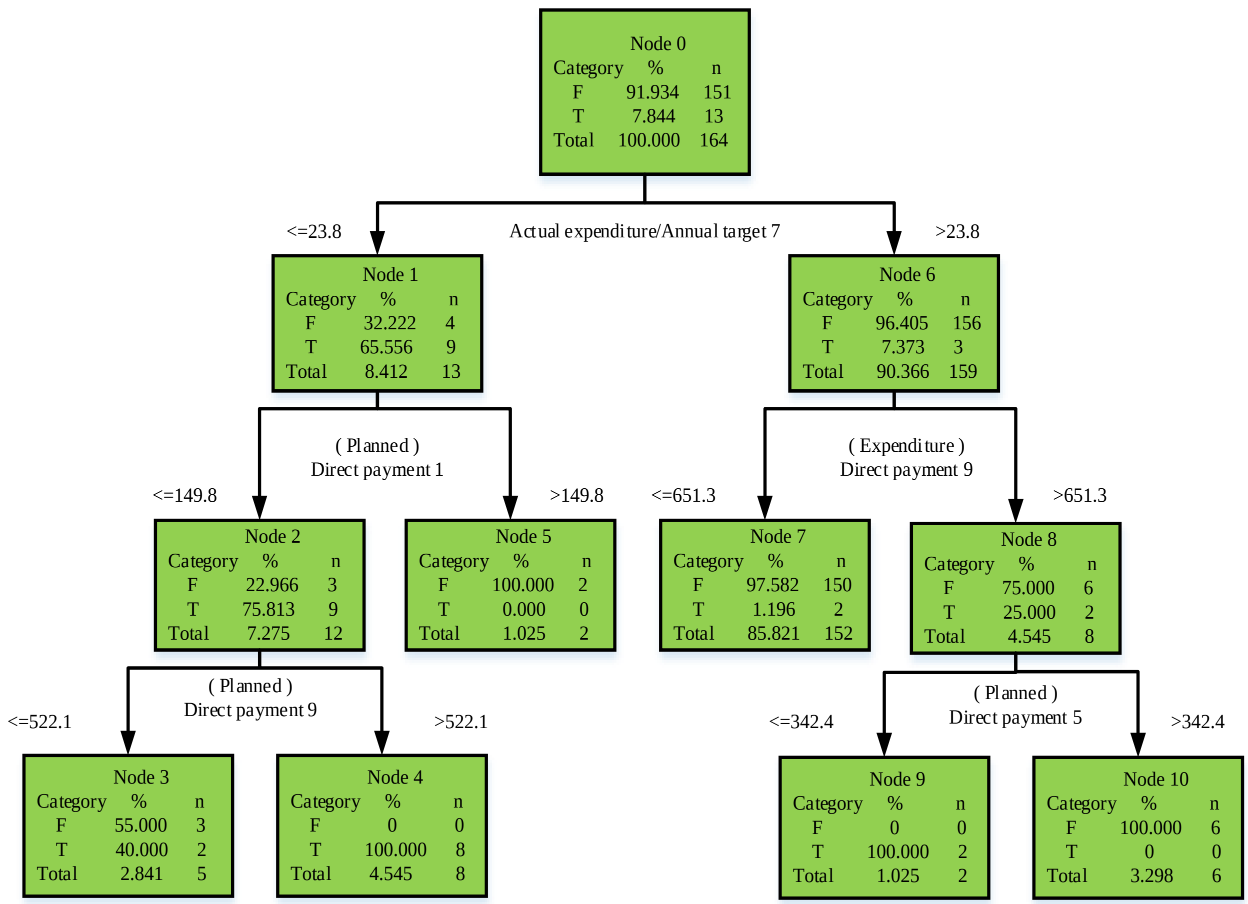


Fig. 5. Decision tree for budget implementation progress alert

In Fig. 5, 23.8% is the cut-off point for the payment method of the decision tree. A Situation: if the budget implementation progress of the project in July is not more than 23.8%, and the direct payment amount in January is not less than 150000 yuan, no budget is required; B Situation: For the same project, if the budget implementation progress is no more than 23.8%, the direct payment amount in January is less than 150000 yuan, and the authorized payment amount in September is no less than 522000 yuan. The decision tree needs to give an alert; In case b, when the authorized payment amount in September is less than 522000 yuan, it is necessary to judge according to the authorized payment amount in January. If it is not less than 8800 yuan, there is no need for early warning. On the contrary, if it is less than 8800 yuan, there is a need for early warning. By analogy, we can pay more attention to the financial budget of the current month according to the results and budget implementation progress. The study obtains an early warning decision tree about the execution progress of the budget, and puts the data in the three test sets into the decision tree. The purpose is to compare the prediction result with the real value to verify the prediction accuracy of the decision tree in financial management. A decision tree was built on some data in test set 2, and the budgets of 100 projects were predicted and analyzed. In the judgment of the decision tree, the predicted results of 95 projects were the same as the actual results, and the predicted results of only five projects were the same as the actual results. Therefore, the decision tree built on the basis of information entropy has a higher prediction effect. Different test sets are used to verify the prediction accuracy of decision tree, as shown in Fig. 6.

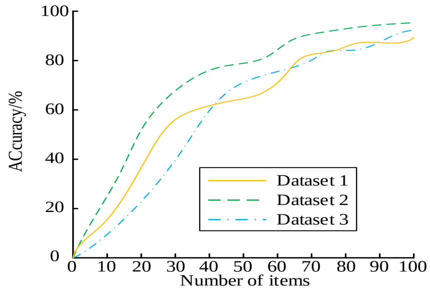


Fig. 6. Verification of prediction accuracy of decision tree with different test sets

Through the verification of the prediction accuracy of the decision tree from three data sets, it is found that the decision tree based on information entropy has a higher prediction accuracy, which verifies that the decision tree based on information entropy is applicable to a variety of data. Therefore, the analysis results of the decision tree constructed in Fig. 5, which is used in colleges and universities, are reliable.

B. Performance Comparison of Improved Decision Tree Algorithm based on Information Entropy

The experiment uses the ID3 algorithm, the CART algorithm and the improved C4.5 algorithms to compare to verify the superiority of the algorithm performance. The performance experiment of the algorithm uses the relevant error indicators as the judgment basis. The mean square error results of the three algorithms in the model are shown in Fig. 6.

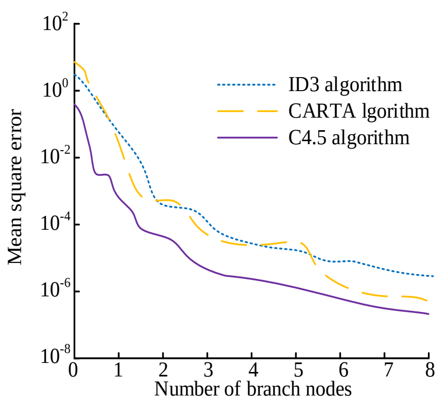


Fig. 7. Mean square error results of three algorithms in the model

From Fig. 7 it is seen that the mean square error of the C4.5 algorithms is lower among the three models, and when the number of branch nodes is small, the mean square error of the algorithm is still higher, and the mean square error gradually decreases with the increase of the number of nodes, after a certain number of nodes, the mean square error value begins to converge, and the change range is small.

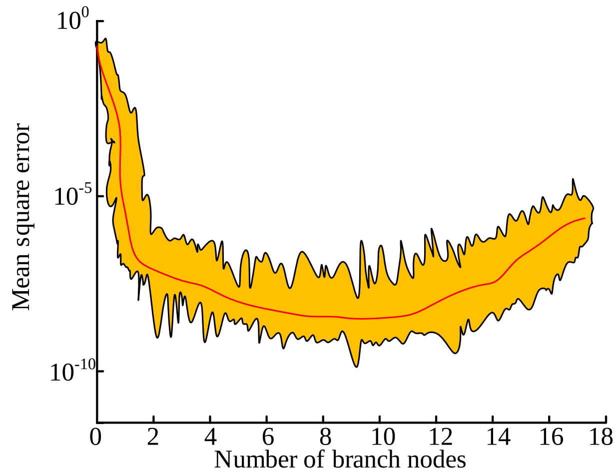
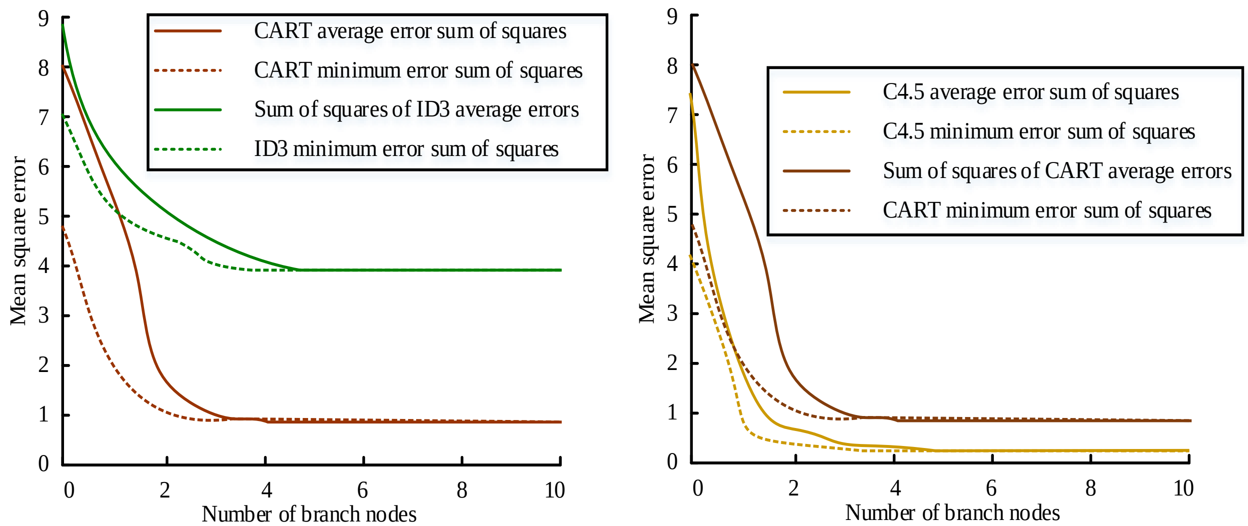


Fig. 8. Mean square error result of C4.5 algorithm

Fig. 8 shows the mean square error result of C4.5 algorithms. From Fig. 8 it is seen that when the number of nodes in the decision tree is small, its mean square error is large, and the increase in the number of nodes makes the mean square error of the decision tree change significantly. When the number of nodes is about three, the mean square error begins to converge gradually, and the variation range is small; but when the number of nodes is about 12, the mean square error increases gradually. It shows that the increasing number of nodes will lead to the increase of algorithm operation, and the decrease of its efficiency and prediction accuracy. The algorithm is further compared. The error square of the algorithm is shown in Fig. 9. The sum of the error squares of the two algorithms is further compared.

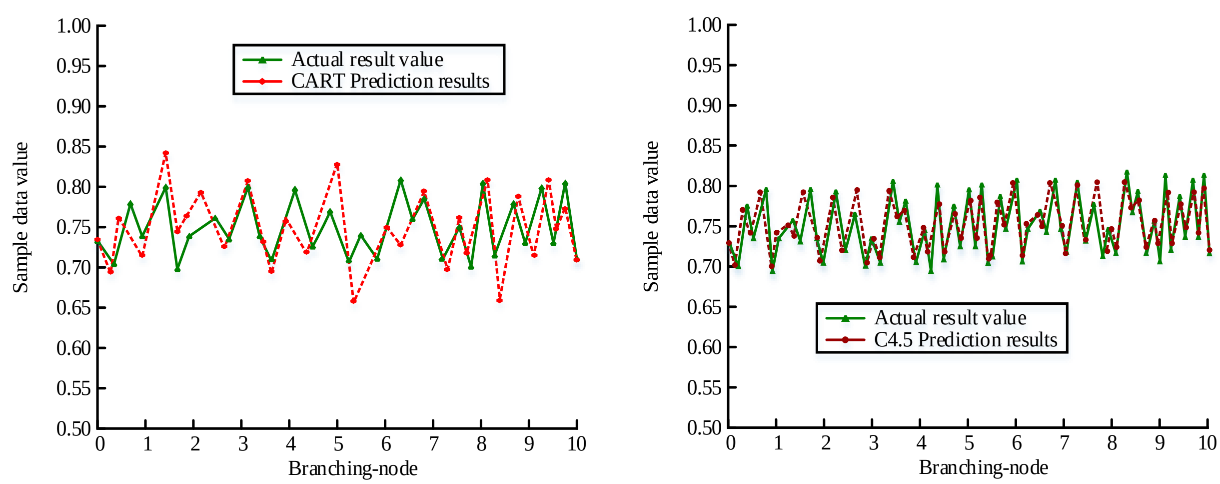
Fig. 9 shows the sum of error squares of ID3 algorithm and CART algorithm, and the sum of error squares of CART algorithm and C4.5 algorithms. It can be seen from Fig. 9(a) that the sum of error squares of CART algorithm is smaller than that of ID3 algorithm as a whole. When the number of branch nodes is 1-3, the error square changes greatly, and the sum of error squares changes from eight to about one. Since the number of nodes is four, the convergence speed of the average error square and the minimum error square of the algorithm decreases significantly, and the convergence tends to be smooth. The sum of squares of the CART algorithm converges to about 0.9. From Fig. 9 (b), it can be concluded that the sum of squares of errors of C4.5 algorithms is smaller than CART algorithm as a whole. When the number of branch nodes is no more than three, the error square of C4.5 algorithm changes greatly, and the convergence speed is fast. The sum of error squares changes from 7.5 to about 0.5. When the number of self-nodes is no less than four, the convergence speed of the average error square and the minimum error square of the algorithm decreases significantly, and the convergence tends to be smooth. The sum of squares of C4.5 algorithm converges to about 0.17.



(a) The result of the sum of squares of errors of ID3 algorithm and CART algorithm

(b) Sum of squares of errors of CART algorithm and C4.5 algorithm

Fig. 9. The result of the sum of squares of errors of ID3 algorithm and CART algorithm, and the result of the sum of squares of errors of CART algorithm and C4.5 algorithm



(a) Prediction results of CART algorithm on test samples

(b) C4.5 Prediction results of algorithm on test samples

Fig. 10. Prediction results of CART algorithm on test samples and C4.5 algorithm on test samples

Fig. 10 shows the prediction results of CART algorithm on test samples and C4.5 algorithm on test samples. It can be seen from Fig. 10(a) that there is a certain degree of difference between the predicted results of CART algorithm and the actual results. The actual data results fluctuate between 0.70 and 0.80, while the predicted results of CART algorithm vary between 0.70 and 0.85. It can be seen from Fig. 10(b) that the improved prediction results have a high coincidence with the actual results. Therefore, the improved C4.5 algorithm has a high accuracy, and there is no significant error between the prediction results and the actual results.

Fig. 11 is a comparison chart of prediction model accuracy under three different algorithms. In Fig. 11, the predicted average value of CART model is 89.96%. The predicted average value of ID3 model is 92.56%. Decision tree algorithm C4.5 based on information entropy has a prediction accuracy of 95.1% in the model. By comparing the accuracy values of the prediction models under three different algorithms, it can be seen that the model proposed in this study has better prediction accuracy, so it can be used in actual prediction.

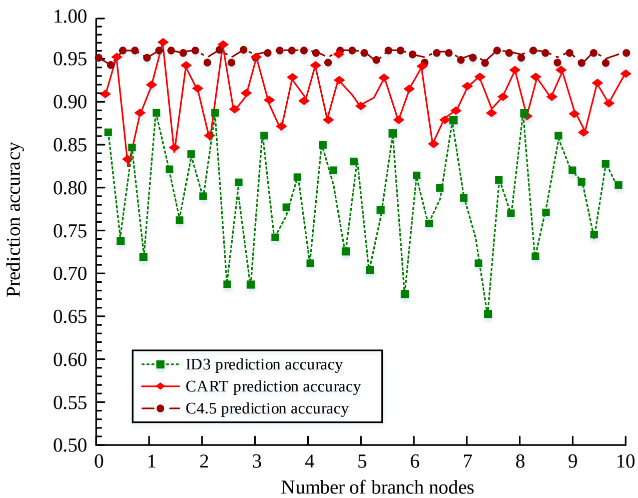


Fig. 11. Comparison chart of prediction model accuracy under three different algorithms

In conclusion, the improved decision tree algorithm model built by this research can have better prediction accuracy. Applying this model to the analysis of financial data in colleges and universities can obtain the financial implementation progress. The improved decision tree algorithm model can give an early warning according to the monthly financial income and expenditure of colleges and universities, so as to remind managers of the financial situation of colleges and universities in the current month.

V. Conclusion

The development of information industry has brought great impact to society, making information everywhere. The financial management in higher education institutions, due to the tedious work, generates a lot of information and is difficult to sort out, which affects the leaders' decision on the follow-up work. Therefore, in order to extract relevant data from redundant information data and facilitate making correct decisions, a decision tree algorithm based on information entropy is proposed. The algorithm uses entropy value and information gain rate to predict the probability of event occurrence, which can achieve different early warning effects for different events. The research verifies the prediction performance of the decision tree through the test set. In the three test sets, the precision measurement of the algorithm prediction under the support of information entropy reaches 95%. Then, the current financial data of colleges and universities are analyzed, and the financial implementation progress is obtained. Colleges and universities should be cautious about their financial expenditure in the month, otherwise the decision tree will issue a financial warning. Finally, to verify the performance of the algorithm, compared with ID3 algorithm and CART algorithm, the decision tree algorithm based on information entropy performs best by comparing the mean square error, the sum of squares of errors and the accuracy.

VI. Future Work

Although the C4.5 algorithm in the traditional decision tree algorithm has been improved in this research, the decision tree algorithm based on information entropy is proposed and analyzed with information entropy as the decision basis of the decision tree. At the same time, the measurement method is introduced to calculate the information entropy and information gain rate, and then an application model of university financial management based on the improved decision tree algorithm is proposed. However, there are still shortcomings in this research. The calculation of the information gain rate of the decision tree established on the basis of information entropy is more complex, which increases the difficulty of the algorithm. In the experiment, there are few sample data, resulting in some errors in accuracy. In the subsequent research, we can find a more simple calculation method and add more data samples from more colleges to improve the algorithm efficiency and prediction accuracy.

References

[1] B.J. Jang, L.H. Kim, K. Y.Kim, “The Effects of the Creating of Shared Value of Hospitality Industry Companies on Corporate Trust and Non-financial Management Performance: Focused on Travel Agencies.” Journal of Tourism and Leisure Research, Vol. 33, No. 1, pp. 197-216, 2021.

[2] S. Cohen, F. Manes-Rossi, I. Brusca, E. Caperchione. “Guest editorial Happy endings and successful stories in public sector financial management: a lesson drawing perspective.” International Journal of Public Sector Management, Vol. 34, NO. 4, pp. 393- 406, 2021.

[3] Z. Zhu, J. Liu, Q. Fan, Y. Liu, Y. Wu, J. Zhou, F. He. “Comparative analysis and research on composition of mulberry-sourced medicinal materials by fingerprint division total statistical moment method and information entropy.” China journal of Chinese materia medica, Vol. 46, No.10, pp. 2547-2555, 2021.

[4] E. Priya, “Resnet Based Feature Extraction with Decision Tree Classifier for Classificaton of Mammogram Images.” Turkish Journal of Computer and Mathematics Education (TURCOMAT), Vol. 12, NO. 2, pp. 1147-1153, 2021.

[5] C. Al, “Application of Decision Tree in Classifying Secondary School Students’ Tendencies to Choose TVET in Malaysia.” Turkish Journal of Computer and Mathematics Education (TURCOMAT), Vol. 12, NO. 3, pp. 3002-3012, 2021.

[6] L. Mao, W. Zhang, “Analysis of entrepreneurship education in colleges and based on improved decision tree algorithm and fuzzy mathematics.” Journal of Intelligent and Fuzzy Systems, Vol. 40, NO. 2, pp. 2095-2107, 2021.

[7] S. Ziweritin, I.A. Ibiam, T.A. Oyeniran, G.E. Okol. “KNN and Decision Tree Model to Predict Values in Amount of One Pound Table.” International Journal on Computer Science and Engineering, Vol. 9, NO. 7, pp. 17-21, 2021.

[8] D. Mabuni, S. A. Babu, “High Accurate and a Variant of k-fold Cross Validation Technique for Predicting the Decision Tree Classifier Accuracy.” International Journal of Innovative Technology and Exploring Engineering, Vol. 10, no. 2, pp. 105-110, 2021.

[9] N. Ahmad, S. Ismail, S. A.Siraj, “Financial sustainability of Malaysian public universities: officers’ perceptions.” The International Journal of Educational Management, Vol. 33, No. 2, pp. 317-334, 2019.

[10] N.Asandimitra, A. Kautsar, “The Influence Of Financial Information, Financial Self Efficacy, And Emotional Intelligence To Financial Management Behavior Of Female Lecturer.” Humanities & Social Sciences Reviews, Vol. 7, No. 6, pp. 1112-1124, 2020.

[11] M. Mohammad, A. Saleh, O.A.A. Jawabreh, R.A. Om. "Artificial Intelligence (Ai) And The Impact of Enhancing the Consistency and Interpretation of Financial Statement in The Classified Hotels in Aqaba, Jordan." Academy of Strategic Management Journal, Vol. 20, No. S(3), pp. 1-18, 2021.

[12] V. V. Malikov, "The Algorithm for Building a Financial Management System at the Enterprise." Business Inform, Vol. 3, No. 506, pp. 344-348, 2020.

[13] A. UJ, E. NMB, C. KS, D. SKL, "Financial crisis prediction model using ant colony optimization - ScienceDirect." International Journal of Information Management, Vol. 50, pp. 538-556, 2020.

[14] J. Hou, "Online teaching quality evaluation model based on support vector machine and decision tree." Journal of Intelligent and Fuzzy Systems, Vol. 40, No. 2, pp. 2193-2203, 2021.

[15] M.N. Rabbani, A. Yusuf, D. Rolliawati, "Komparasi Model Prediksi Daftar Ulang Calon Mahasiswa Baru Menggunakan Metode Decision Tree Dan Adaboost." Jurnal Sisfokom (Sistem Informasi dan Komputer), Vol. 10, No. 1, pp. 18-24, 2021.

[16] D.V. Krishnaveni, S.A. Kumar, J. Sudarvel, P. E. aswaran. "An Intelligent Sale Prediction Classification of Jute Products Using Decision Tree Clustering & Statistical Model." Türk Fizyoterapi ve Rehabilitasyon Dergisi/Turkish Journal of Physiotherapy and Rehabilitation, Vol. 32, NO. 3, pp. 1033-1039, 2021.

[17] N. Kolenda, N. Matviichuk, L. Cherchyk, H. Nakonechna, K. Amit. "Strategic Approach to The Management of Financial Security of Enterprise." Financial and Credit Activity Problems of Theory and Practice, Vol. 4, No. 35, pp. 79-86, 2020.

[18] I. Podhorska, J. Vrbka, G. Lazaroiu, M. Kovacova. "Innovations in Financial Management: Recursive Prediction Model Based on Decision Trees." Marketing and Management of Innovations Vol. 3, pp. 276-292, 2020.

[19] J. Tyndall, "Prairie and tree planting tool—PT2 (1.0): a conservation decision support tool for Iowa, USA." Agroforestry Systems, Vol. 96, No. 1, pp. 49-64, 2022.

[20] E.B. Hicham, K. Salah-Ddine, "Financial News Analysis for Moroccan Stock Trend Predictions." Test Engineering and Management, Vol. 82, No. 1, pp. 1712-1717, 2020.