

Design of Decision Support System for Distance Learning Based on Data Mining

Chen Shao*, Xiaochen Chen

Shandong Management University; Jinan Shandong 250357 China

*Corresponding Author.

Abstract

Since the existing system cannot assess the quality of distance education, the credibility and efficiency of decision support results are low, and the stability of the system is also poor. Combined with data mining technology, a remote learning decision support system based on data mining is designed and proposed. First, the actual situation of each university is analyzed, the system is designed in combination with the B/S architecture, and the various components of the system are described in detail. Then, the C4.5 algorithm of the decision tree algorithm is used to establish a distance teaching quality evaluation model, and the corresponding classification rules are extracted to effectively realize the teaching quality evaluation. Finally, a simulation test is carried out. The experimental results show that the designed system can comprehensively improve the stability and execution efficiency of the system, enhance the credibility of decision support results, and have certain practical applicability.

Keywords: Data mining, distance learning, decision support system, decision tree algorithm

I. Introduction

There are a large number of students' performance test data in colleges and universities every year [1,2]. At present, the existing system can only backup, query and statistics the data, and does not comprehensively analyze the problems existing in the teaching process of teachers and students. At this stage, it is urgent to study a more effective and fast support system for teaching decision-makers. Relevant experts have given some high research results, in the system reform at the same time, reference [3] set up a decision support system to effectively solve the auxiliary judicial discrimination, and combined with text mining and automatic classification technology, a decision support system was set up. Reference [4] designed a decision support system for the purpose of improving the decision-making level, which mainly included data layer, business layer and presentation layer. In reference [5], the intelligent decision support system was designed by distributed physical structure and intelligent agent technology.

Although the above three systems have achieved satisfactory research results at this stage, they failed to evaluate the quality of distance education, resulting in the decline of the credibility and implementation efficiency of decision support results, and the poor stability of the system. Therefore, combined with data mining technology, this paper proposes a decision support system for distance learning based on data mining. The system is designed in combination with the B/S architecture and the various components of the system are described in detail. Then, through the C4.5 algorithm of decision tree algorithm, a distance teaching quality evaluation model was established, and the corresponding classification rules were extracted, which effectively realized the teaching quality evaluation. The simulation results show that the designed system can improve the stability of the system, at the same time, it can also increase the execution efficiency of the system and the credibility of decision support results.

II. Decision Support System for Distance Learning Based on Data Mining

In the development process of decision support system for distance learning, it pays more attention to students' achievements, makes full use of all valuable information, and formulates system objectives. It mainly selects B/S three-tier system for structure division, as shown in Figure 1.

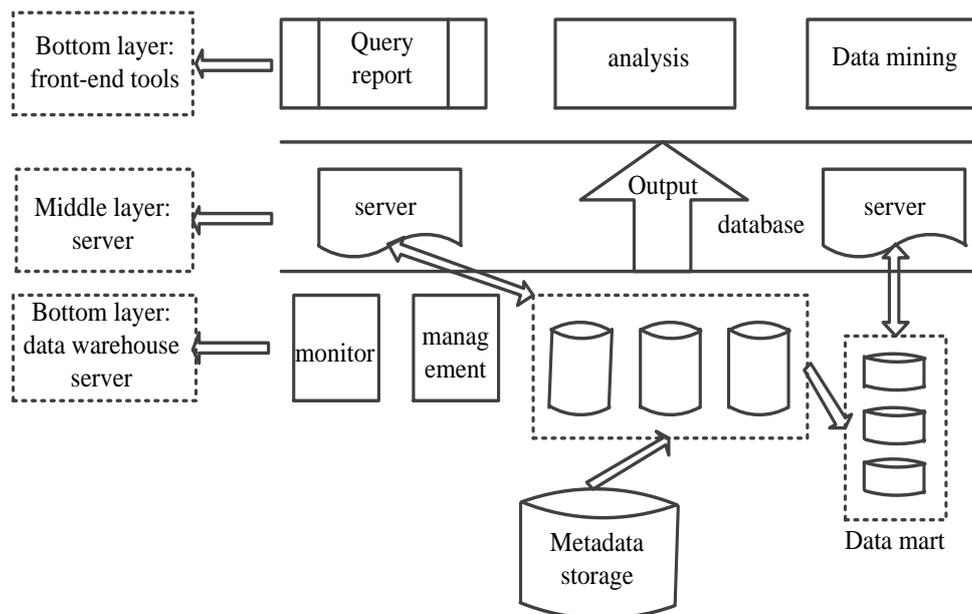


Fig 1: B/S tertiary structure

2.1 Functional requirement analysis

(1) In recent years, China's higher vocational colleges have been deeply valued by the state, and a large number of support policies have been adopted and implemented by various colleges and universities. The training goal of each university is the same, strengthen students' comprehensive quality, establish personality charm. Various colleges and universities pay attention to the cultivation of students' basic skills, at the same time, they need to cultivate students in a targeted way, requiring all students to develop the quality of hard work, and at the same time, they have the courage to bear hardships and the spirit of dedication in the process of employment.

(2) Analyzing the social needs of different colleges and universities, each college needs to set up a variety of courses to cooperate with teachers to achieve the ultimate teaching purpose [6]. In order to have a more in-depth and comprehensive understanding of the learning situation and comprehensive quality development of different students since entering the university, it is necessary to carry out course acceptance within the set cycle. While increasing the communication between students and teachers, it can also provide some theoretical support for decision-makers to make decisions.

(3) The design of the system should be not only comprehensive but also complete. The information query and modification functions of the system are mainly to meet the needs of teachers to modify, register and query the initial scores of students. Compared with the traditional system, the designed system can improve the shortcomings of the traditional system, enhance the speed and accuracy of data processing, and finally send the statistical results to the destination in time.

(4) The system should also have a preliminary decision-making function, which can provide certain decision-making basis for the leaders and managers of various colleges and universities. At the same time, it also has a very important guiding significance to complete the teaching system of various colleges and universities and improve the teaching methods. In addition, the system can also describe the learning situation of different students in school in detail [7,8], which plays an auxiliary role for school decision-makers to make scientific decisions.

(5) The system also needs to have the ability of simple error detection, error correction and fault tolerance. Because the system is an open system, it should also have scalability and expansibility. It can ensure the stable operation of the system and realize data update and maintenance.

(6) In the process of system development, it is needed to pay attention to the data and the security of the system, so as to avoid the devastating failure of the data that cannot be recovered when the system fails.[9]

(7) The interface is friendly and easy to operate, which is convenient for users and managers.

2.2 System design

The system organization chart is given by using Figure 2.

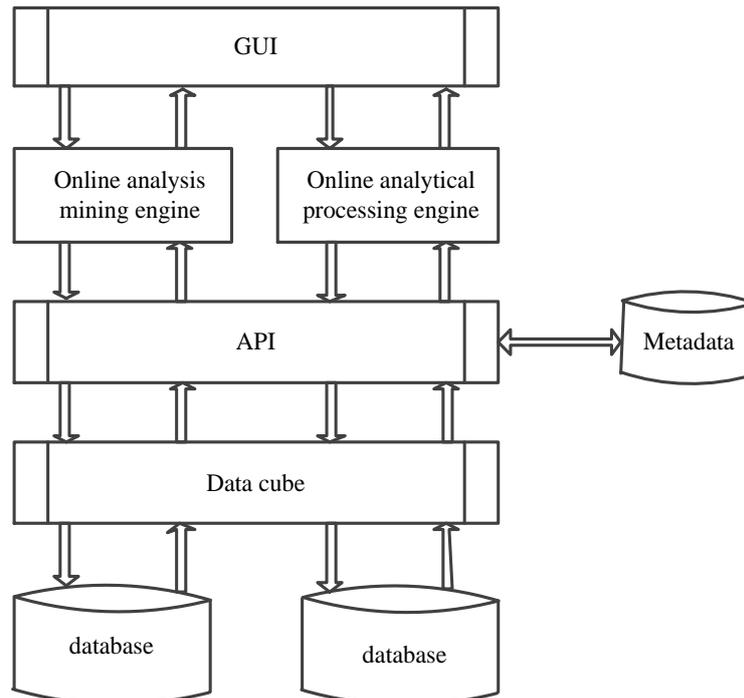


Fig 2: The structure of decision support system for distance learning based on data mining

It can be seen from Figure 2 that the remote teaching decision support system finally shows the user a good graphical interface, which is convenient for the user to master and understand the data. When the data in the database after integration and filtering and other related operations, it can effectively simplify the data analysis process, and promote users to deeply mine the valuable information hidden in the information, and finally present it through the display interface.

Figure 3 is used to give the design diagram of the decision support system for distance learning based on data mining, the students' test scores of various colleges and universities in the past years are collected, to make statistics according to the scores of different subjects in different colleges and departments [10,11]. After the completion of the statistics, all the data will be transmitted to the decision support system for distance learning. After the prediction analysis, the final prediction results will be obtained to provide decision-making basis for decision makers.

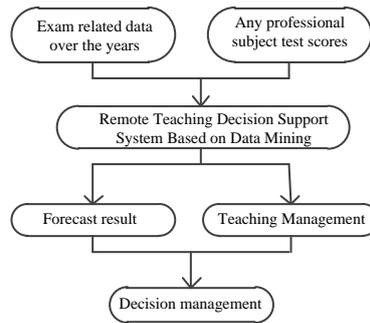


Fig 3: The design of decision support system for distance learning based on data mining

The structure diagram of decision support system for distance learning based on data mining is given in Figure 4.

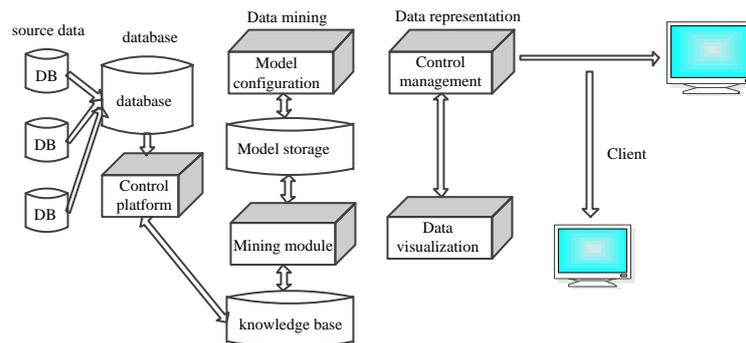


Fig 4: Structure diagram of decision support system for distance learning

2.3 System function division

The system is mainly composed of user login management module, database management module and other modules. The data mining and evaluation module can be subdivided [12], including student performance analysis and evaluation module, etc. The specific schematic diagram is shown in Figure 5.

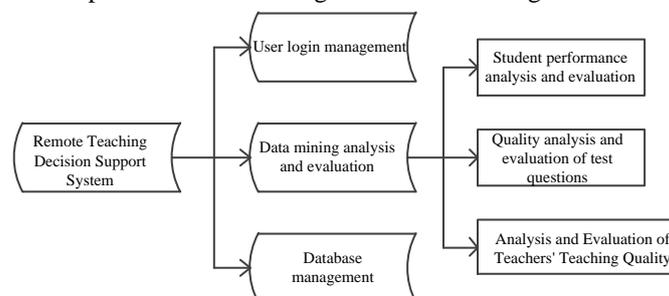


Fig 5: Schematic diagram of system function module

(1) User login management module:

The whole module is mainly responsible for password and user name modification, and user selection according to permissions.

(2) Database management module:

It includes the operations of deleting and adding various types of tables, mainly including the following tables:

(a) Student table:

It is mainly responsible for the entry of each student's student number, name, major and other basic information. It can add and delete timely, and inform the establishment of the corresponding database [13,14].

(b) Teacher list:

It is mainly responsible for the input of teachers' basic information.

(c) Curriculum:

It is mainly responsible for the operation of course number and course type.

(d) Results questionnaire:

It is mainly responsible for the statistics of students' scores and total scores in various subjects.

① Student achievement analysis and evaluation sub module:

It is mainly responsible for the analysis of the examination results of students in each class and discipline, using SPSS Clementine tool for statistical analysis, and the final results are displayed in the form of tree or chart; the evaluation results mainly display the average score and ranking of students in the interface, and the specific functions are shown in Figure 6

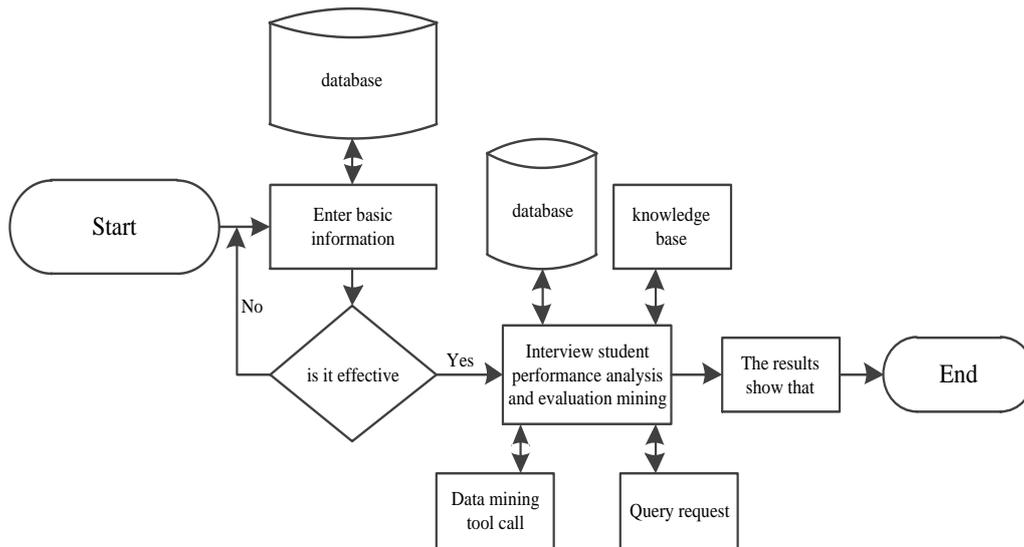


Fig 6: Student achievement analysis and evaluation sub module

② Test paper quality analysis and evaluation sub module:

Test paper quality analysis mainly displays the final statistical results in the form of tree or column through SPSS Clementine tool. The main display contents include: subject name, teacher name, test time and assessment method, etc. The specific functions are shown in Figure 7 [15].

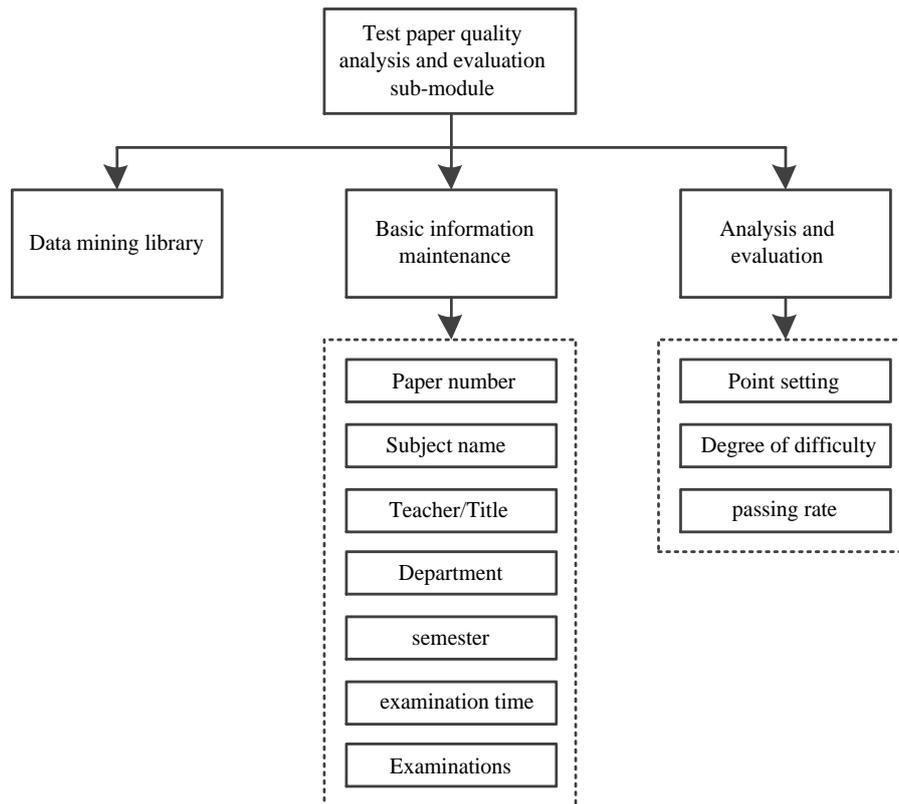


Fig 7: Test paper quality analysis and evaluation sub module

(3) The analysis module of teachers' teaching quality is as follows

The whole module mainly analyzes the teaching quality of teachers through SPSS Clementine tool, and the final analysis results are displayed through the tree or column structure, mainly including the prediction of teaching quality level and teaching effect and other related information. The specific functions are shown in Figure 8:

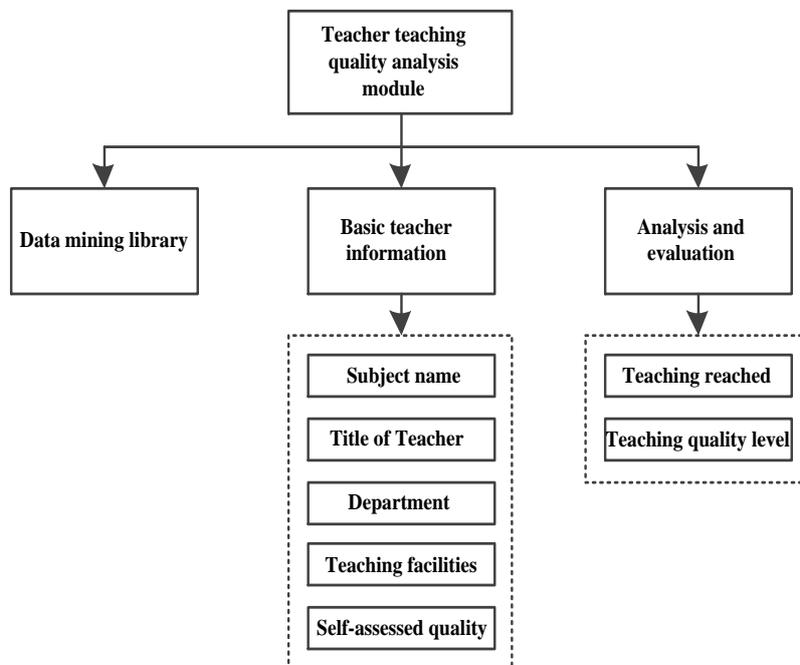


Fig 8: Analysis module of teachers' teaching quality

III. Distance Teaching Quality Evaluation Model Based on Decision Tree

Through C4.5 algorithm of decision tree, the preprocessed teacher evaluation information is taken as the training sample data set, the decision tree model of distance teaching quality evaluation is established, and the corresponding classification rules are extracted, the specific operation steps are as follows:

Firstly, based on the analysis of academic achievement, data mining technology is used to analyze the functions of the system, as shown in Figure 9

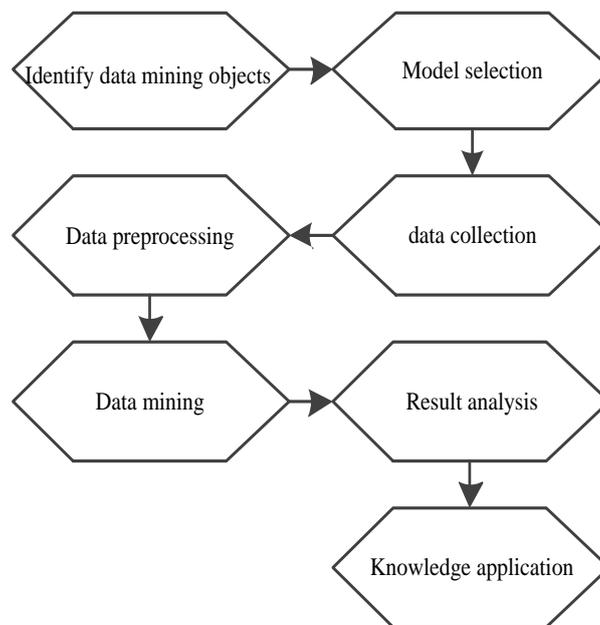


Fig 9: Flow chart of data mining

(1) The information entropy of classification attribute in training set is as follows:

Firstly, the formula is used to calculate the analysis information entropy of the training set, and then the information gain rate of different attributes is calculated through the information entropy. In the training data set, the total number of samples needs to be set, and the specific calculation formula is given as follows:

$$I(S_1, S_2, S_3, S_4) = I(23, 25, 23, 9) \quad (1)$$

(2) The information entropy of subsets divided by different attributes is calculated:

A certain number of teaching age attribute values are selected to calculate the information entropy of the corresponding subset of each attribute value, the specific calculation formula is:

$$I(S_{11}, S_{12}, S_{13}, S_{14}) = I(7, 6, 3, 0) \quad (2)$$

(3) The corresponding information expectation information is calculated through the teaching age attribute. The expected information of teaching age attribute is calculated by formula (3), that is:

$$E = \frac{16}{80} I(S_{11}, S_{12}, S_{13}, S_{14}) + \frac{41}{80} I(S_{21}, S_{22}, S_{23}, S_{24}) + \frac{23}{80} I(S_{31}, S_{32}, S_{33}, S_{34}) \quad (3)$$

(4) The information gain is calculated by the teaching age attribute, that is:

$$Gain = I(S_1, S_2, S_3, S_4) - E \quad (4)$$

(5) According to the teaching age attribute, the corresponding splitting information is calculated:

$$SplitInfo = \frac{16}{80} \log^2 \frac{16}{80} - \frac{41}{80} \log^2 \frac{41}{80} - \frac{23}{80} \log^2 \frac{23}{80} \quad (5)$$

(6) The corresponding information gain is calculated by the teaching age attribute, that is:

$$Gainratio = \frac{Gain}{SplitInfo} \quad (6)$$

(7) Through the above calculation steps, the information gain rates of nine attributes such as degree, professional title and basic skills are calculated respectively:

$$Gainratio(x) = \frac{Gain(x)}{SplitInfo(x)} \quad (7)$$

$$Gainratio(z) = \frac{Gain(z)}{SplitInfo(z)} \quad (8)$$

$$Gainratio(j) = \frac{Gain(j)}{SplitInfo(j)} \quad (9)$$

$$Gainratio(s) = \frac{Gain(s)}{SplitInfo(s)} \quad (10)$$

$$Gainratio(f) = \frac{Gain(f)}{SplitInfo(f)} \quad (11)$$

$$Gainratio(g) = \frac{Gain(g)}{SplitInfo(g)} \quad (12)$$

$$Gainratio(q) = \frac{Gain(q)}{SplitInfo(q)} \quad (13)$$

$$Gainratio(c) = \frac{Gain(c)}{SplitInfo(c)} \quad (14)$$

In the above formula, x is the degree, z is the professional title, j is the basic skill, s is the teaching content, f is the teaching method, g is the teaching effect, q is the emotion attitude values, c is the student foundation.

(8) The root node of the decision tree is established

Through C4.5 algorithm, it is the best classification to select the "student foundation" as the root node of the decision tree. The following takes "student foundation" as the test attribute to build a decision tree for a lesson. The root node of the decision tree is marked as "student foundation".

(9) Repeat the above steps for each branch subset in the decision tree to determine each node of the branch until all attributes are divided.

A decision tree is established based on the operation results. Because the decision tree is too large to be cleared and displayed in a graph, only a small part is shown below, as shown in Figure 10.

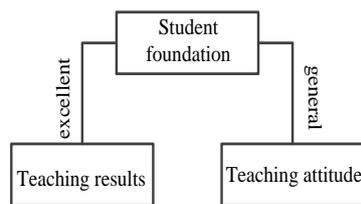


Fig 10: the root node of the decision tree for distance teaching quality evaluation

By analyzing the decision tree, it can be seen that under different rules, the distribution of the number of people in different levels of distance teaching quality and performance will lead to the final decision-making error, so it is needed to modify the decision tree. Add statements in each algorithm, and add and record qualified statistical data in the leaf node.

Among them, direct extraction of classification rules is the biggest advantage of decision tree. The following is mainly from the decision tree model to obtain the relevant factors that affect the quality evaluation of distance education. Through the analysis, it can be seen that the information gain rate of students' foundation and professional title is obviously better than other test indexes, so they are several key evaluation indexes to measure the quality of teaching. In the actual teaching process, different teaching methods and attitudes are adopted according to the different foundation of students. If the students' foundation is excellent, teachers should pay more attention to the use of teaching methods; if the students' foundation is general, teachers should pay more attention to the classroom teaching content while teaching learning methods.

IV. Simulation Experiment

In order to verify the comprehensive effectiveness of the proposed decision support system for distance learning based on data mining, the simulation test is carried out on the experimental platform of windows 764-bit system, CPU 2.5GHz, RAM4.0GB and MATLAB R2017b.

4.1 System stability / (%):

The following experiments test the stability of different teaching decision support systems. Only the stable operation of the system can ensure the smooth completion of distance teaching decision support. The specific experimental comparison results are given in Figure 11.

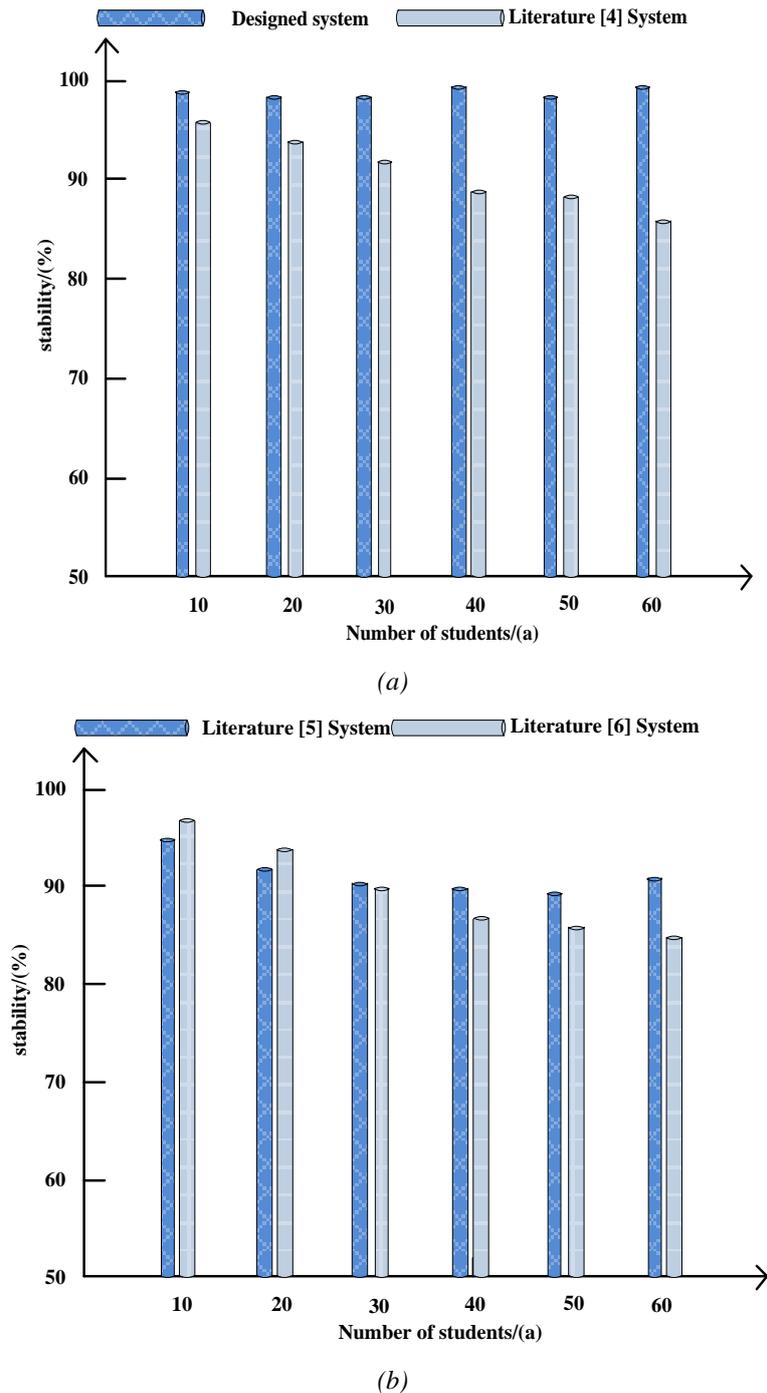
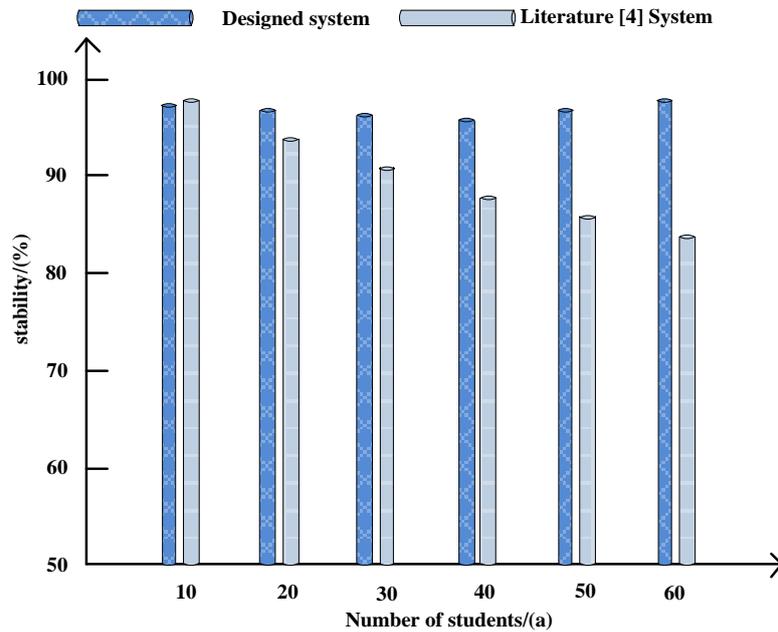


Fig 11: Stability comparison results of different systems

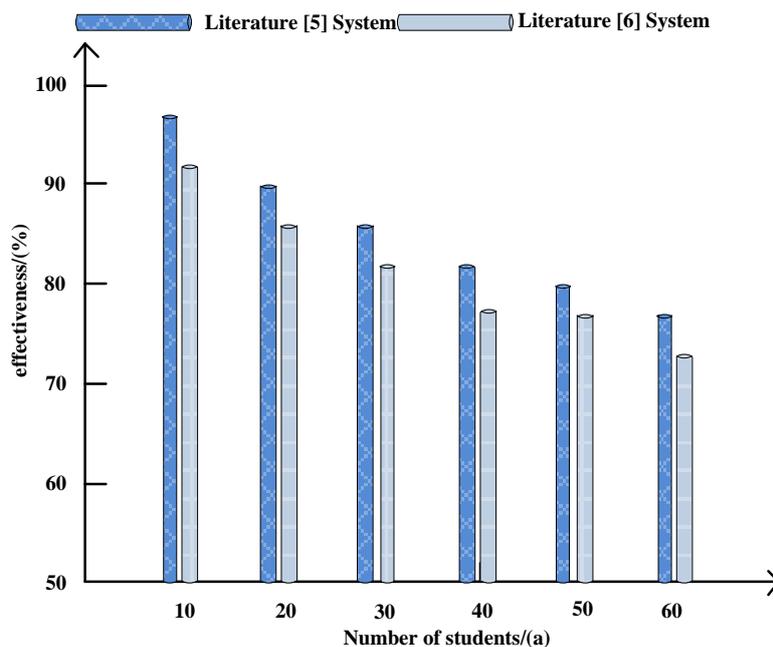
Analysis of the experimental data in Figure 11 shows that with the continuous increase of the number of students, the stability of each system has changed significantly, but compared with the other three systems, the stability of the designed system is in a relatively stable state. The main reason is that C4.5 algorithm is added to the designed system to evaluate the quality of distance education, which can provide decision support for system decision-makers, improve the stability of the system, and is obviously better than the other three systems.

4.2 Execution efficiency / (%):

The speed of system execution has always been a topic of wide concern of sociologists. The following will take the execution efficiency as a test index to compare the execution efficiency of four different decision support systems. The specific experimental comparison results are shown in Figure 12.



(a)



(b)

Fig 12: Comparison of execution efficiency of different systems

Analysis of the experimental data in Figure 12 shows that the implementation efficiency of different systems will be affected by the number of students. With the continuous increase of the number of students, the implementation efficiency of the system shows a linear downward trend. Due to the design of the system in the process of practical application, the establishment of a distance teaching quality evaluation model, the whole face of the teacher's

teaching quality evaluation, for different students are given different teaching programs, effectively reduce unnecessary time, promote the implementation efficiency of the whole system has been comprehensively improved, and significantly better than the other three systems.

4.3 Credibility of decision support results / (%):

In order to comprehensively verify the comprehensive effectiveness of the designed system, the following experimental tests compare the credibility of decision support results of different systems. The specific experimental results are shown in Table 1:

Table 1 Comparison of credibility of decision support results of different systems

Number of experiments /(times)	Credibility of decision support results / (%)			
	The designed system	System in reference [3]	System in reference [4]	System in reference [5]
5	98.36	95.55	96.74	97.55
10	97.52	94.54	94.25	95.74
15	98.74	92.29	94.35	94.51
20	97.25	90.36	93.44	93.85
25	99.33	91.14	93.66	92.69
30	98.45	90.24	92.41	92.75
35	97.20	88.56	91.96	91.65
40	98.37	87.29	90.71	90.33
45	97.21	86.54	90.36	90.45
50	98.87	85.84	89.00	89.27
55	97.36	84.96	88.36	88.35
60	97.14	83.36	87.41	86.21

By analyzing the experimental data in Table 1, it can be seen that the credibility of the decision support results of the designed system is significantly higher than that of the other three kinds of decision support systems. The main reason is that the designed system establishes a distance teaching quality evaluation model combined with the decision data, comprehensively understands the learning state of different students, gives the corresponding teaching scheme in time, comprehensively improves the teaching quality of teachers, and promotes that the decision support scheme provided by the whole system has high credibility.

V. Conclusion

Aiming at a series of problems existing in the traditional system, combined with data mining technology, this paper designs and proposes a decision support system for distance learning based on data mining. The main research results are as follows:

(1) Combined with B/S architecture, the whole system is divided into five modules, including three large modules and two sub modules. The functions of each module are analyzed and introduced in detail.

(2) The decision tree algorithm is applied to the quality evaluation of distance teaching, and the evaluation model is established to further realize the quality evaluation of distance teaching. Meanwhile, the main factors affecting the quality of teaching are obtained.

(3) Through the specific simulation test, the effectiveness and superiority of the designed system are fully verified.

References

- [1] Aken. K.V, Strekowski. L, Patiny. L, "EcoScale, a semi-quantitative tool to select an organic preparation based on economical and ecological parameters," *Beilstein journal of organic chemistry*, vol. 2, no. 1, pp. 3-5, 2018.
- [2] Mandal. N, Mitra. S, Bandyopadhyay. D, "Paper-Sensors for Point-of-Care Monitoring of Drinking Water Quality," *IEEE Sensors Journal*, no. 99, pp. 11-19, 2019.
- [3] G.X. Li, J. Zhang, L. Qi, et al, "Research on Decision Support System for Automatic Brine Mining in Salt Lake Chemical Industry," *Science and Technology Management Research*, vol. 37, no. 08, pp. 230-236, 2017.
- [4] Q. Zhu, K.Z. Wei, L.L. Ding, et al, "Design of a decision support system for court judgments based on text mining and automatic classification," *China Management Science*, vol. 26, no. 01, pp. 170-178, 2018.
- [5] T. Dai, "Design and implementation of health decision support system," *Chinese Health Policy Research*, vol. 9, no. 12, pp. 1-1, 2016.
- [6] Julaluk. N, Nguyen. M.P, Chamindie. P, et al, "Development of Paper-Based Analytical Devices for Minimizing the Viscosity Effect in Human Saliva," *Theranostics*, vol. 8, no. 14, pp. 3797-3807, 2018.
- [7] Afrifah. K.A, Osei. L, Ofosu. S, "Suitability of Four Varieties of Cocos Nucifera Husk in Ghana for Pulp and Paper Production," *Journal of Natural Fibers*, no. 3, pp. 1-8, 2021.
- [8] Mako. T.L, Levine. M, "Design, Implementation, and Evaluation of Paper-Based Devices for the Detection of Acetaminophen and Phenacetin in an Advanced Undergraduate Laboratory," *Journal of Chemical Education*, vol. 96, no. 8, pp. 1719-1726, 2019.
- [9] X.Y. Liu, P. Wang, C.H. Wang, et al, "Development and application of a GIS-based water environment management decision support system for the Yongjiang River Basin in Nanning City," *Hydropower Energy Science*, vol. 34, no. 04, pp. 170-173, 2016.
- [10] Z.H. Chen, J.Z. Li, "Intelligent Semantic Collection and Intelligent Decision Support System for Coal Mine Safety Hazards," *Industry and Mine Automation*, vol. 44, no. 11, pp. 11-17, 2018.
- [11] J. Zhao, P. Zhang, G.G. Zhang, "Rapid decision support system for on-site repair and processing of runner chamber," *Modern Manufacturing Engineering*, no. 11, pp. 146-151, 2016.
- [12] D.M. Qian, A.N. Luo, Y.F. Gu, "Research on the Design of Provincial Educational Science Decision Service Support System under the Background of Big Data," *Educational Development Research*, vol. 038, no. 05, pp. 68-74, 2018.
- [13] Jafarpour. B, Abidi. S.R, Van. W.W, et al, "Execution-time integration of clinical practice guidelines to provide decision support for comorbid conditions Artificial intelligence in medicine," vol. 94, no. 3, pp. 117-137, 2019.
- [14] Garg. R, Kumar. R, Garg. S, "MADM-Based Parametric Selection and Ranking of E-Learning Websites Using Fuzzy COPRAS," *IEEE Transactions on Education*, vol. 62, no. 1, pp. 11-18, 2019.
- [15] Wada. R, Takeuchi. J, Nakamura. T, et al, "Clinical Decision Support System with Renal Dose Adjustment Did Not Improve Subsequent Renal and Hepatic Function among Inpatients: The Japan Adverse Drug Event Study," *Applied Clinical Informatics*, vol. 11, no. 5, pp. 846-856, 2020.