

# Data Mining Models for Online Education Management

Li Hongxia  
School of Continuing Education  
Jilin University  
Changchun, Jilin, 130000, China  
[hit6828197@163.com](mailto:hit6828197@163.com)



**ABSTRACT:** *With the rapid development of internet technology, distance education has become an important way of education. However, remote education management has many problems, such as controlling students' learning progress, analyzing students' interactivity, and evaluating teaching quality. To effectively address these issues, this article proposes constructing an intelligent model for remote education management based on data mining algorithms. This article first introduces the application of data mining algorithms in remote education management. Through data mining technology, valuable information can be extracted from a large amount of data, which helps teachers better understand students' learning status and improve teaching quality. At the same time, data mining can also analyze students' learning behavior and provide personalized learning suggestions and guidance.*

**Keywords:** Data Mining Algorithm, Distance Education Management, Intelligent Mode

**Received:** 10 March 2023, Revised 5 July 2023, Accepted 18 July 2023

**DOI:** 10.6025/jic/2023/14/3/61-68

**Copyright:** with authors

## 1. Introduction

Throughout the reform of China's education management system, we can see that its basic development direction is carried out in the following aspects [1]. In the early stage of education reform, the educational management system will transfer the high gravity to the low centre of gravity. Through the decentralization of management power, the management power of the central government and the central education management department is dispersed to the government and school institutions in various places [2]. By reforming this mechanism, our education management system can be closer to the actual teaching. After such a reform, the transfer of knowledge authority has no substantive impact on the overall education system and is consistent with our original system [3]. The original complex education was transferred from the central government and the educational administration to the local government and educational administration departments.

Through the gradual decentralization of the efforts, the teaching enthusiasm of local governments and educational institutions has been greatly improved, and the educational institutions in various places can adapt to the development and differentiation

of the society in time to change the educational methods. The collection of information is also more efficient, which improves the efficiency of education management in disguise [4]. After the power is issued, the school has the autonomy of enrollment, which has a great role in promoting the development of colleges and universities. According to the characteristics of the school, the school has made the corresponding adjustment between the subject and the speciality, forming a unique way of running a school, which makes the schools of our country show a multiform and polymorphic form of education [5].

## **2. State of the Art**

In transferring the original system to the outside, it is necessary to set up the teaching pattern of the past government, which is the main body of government running and the society outside the school together [6]. It can make a combination of the teaching of the school and the needs of the society, and the learning content will be more suited to the needs of the society so that the society will widely welcome the students after graduation. The education of the functions and powers of the original government as a significant social activity must conform to the needs of society [7]. The original ducks' education in China needs to be changed accordingly. The quality of education proposed in our country is influenced by the trend of international education, and it will be promoted with the unique Chinese culture.

After a series of educational reforms, the goal we need to pursue now is to provide more quality education [8]. To expand the quality of teaching resources and improve the overall national quality of our country, we need to build a sound and energetic education system, strengthen the construction of educational institutions, and form a teaching management mode with a special socialist system [9]. In this process, we must always adhere to the people-oriented quality education [10]. The core role of education is to solve the problem of training talents, strengthen the all-around development of students and the construction of quality education.

## **3. Methodology**

### **3.1. Pre-Preparation and Functional Classification of Data Mining algorithms**

The computing functions of data mining algorithms in different computing fields are different. In general, the calculation of data mining algorithms can deal with the calculation in most fields. So far, we have used data mining algorithms for computing the following functions: The first is classification. Data mining algorithms can deeply dig out the differences between data and then summarize different data types according to certain rules. This is the most primitive computing function of a data mining algorithm and one of the most commonly used computing functions. Another form of computation combines data mining algorithms with association rules for computing research. This calculation is usually used to calculate the connection between events, and one event may lead to another event, which is used to predict future events better. There is also a data mining algorithm for clustering calculation, seeking data between the rules and the classification of their properties. The data mining algorithm has a predictive function; through the analysis of the law of events, the trend of events are predicted. Data mining has a certain relationship and relationship. The calculation of data mining algorithms is a computer algorithm that integrates these connections and relationships and magnifies their relations. The calculation of this algorithm is achieved through the calculation of this algorithm. In addition, the different computing functions of data mining algorithms mean that the data mining algorithms are different in computation form and calculation steps. At the bottom, we will also make a deep analysis.

In general, the calculation of the data mining algorithm is mainly to calculate the patterns and rules in data. The computation of data mining algorithms is a form used to describe relations in the initial calculation. Through data mining algorithm calculation and research, the relationship between data is classified and sorted out. This calculation form later evolved into a new form of calculation: predictive calculation. The calculation of predictive properties is based on analysing the relationship between data, and then the calculated data set is the basis of the calculation. At present, all data mining algorithms are calculated through these two algorithms. The data mining algorithms used in this paper are no exception, but we combine the prediction and description of two kinds of computing content in this paper. The calculation form is more complex. We need to combine the calculation steps and the form of calculation and analyze the common formula of the data mining algorithm.

The research of pattern discovery is divided into six main directions. The first is association rules. Association rules represent a relationship or relationship between data. The second is a sequential pattern, which makes the association rules mode research calculate the feasibility and the possible time. The third is the classification model. This calculation mode needs to establish a mathematical calculation model, and the data can be reasonably partitioned through the calculation of mathematical

models. The fourth is clustering mode clustering mode, which means simply dividing data into different types. The fifth is the regression model; the general situation of the regression model processing makes the discrete data difficult, and the last model makes the deviation analysis. This data mining algorithm aims to find and analyze the error data of the data set. This calculation is widely used in our daily calculations. The calculation is also studied, using the following Figure 1 to analyze the calculation steps of the data mining algorithm used.

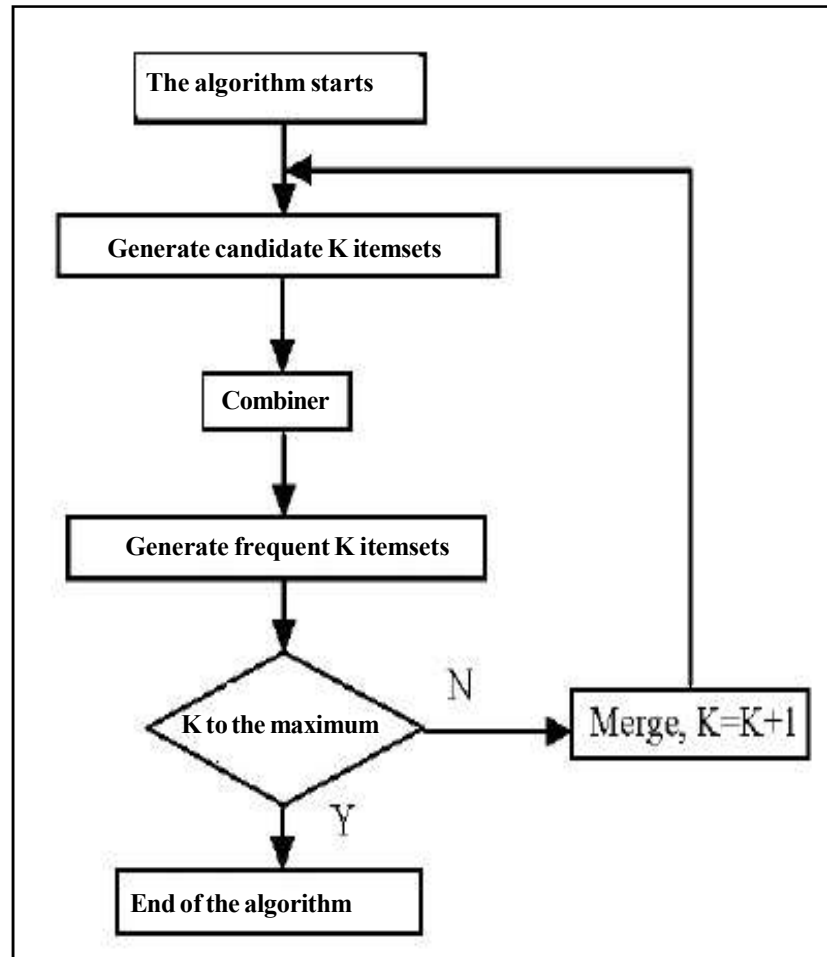


Figure 1. The calculation steps of the data mining algorithm used in this article

### 3.2. Formula Based on Data Mining Algorithm and Intelligent Model Construction in Distance Education Management

Data mining algorithms have various forms of computation, but in general, the most commonly used algorithms are the decision tree algorithm, K nearest neighbour algorithm and binary algorithm, three kinds. Before we study the calculation formula of the data mining algorithm, the computing power of the three algorithms needs to be analyzed and the data mining algorithm is selected, suitable for using this paper and optimizing it. Table 1 for analysis and comparison is used to compare the computing power of the three algorithms.

- It can be seen from the previous table that the decision tree classification algorithm performs well on the learning speed of the training set, the speed of classification, and the processing of discrete, binary and continuous data, but it is still deficient in the accuracy and the noise data and incremental learning. The K nearest neighbour algorithm has a prominent performance in the learning speed of the training set and overfitting and incremental learning. Still, the speed of classification is slow, and the accuracy needs to be improved. The binary algorithm performs well in accuracy, classification speed and dependency property measurement, but it converges slowly when machine learning is performed on a training set.

Project	Computer algorithm		
	Decision tree	K nearest neighbour	Binary
accuracy	excellent	good	good
Classification speed	good	medium	good
Leakage measurement	excellent	excellent	good
Noise data processing	good	medium	medium
Incremental learning	good	good	excellent
Model parameter processing	excellent	medium	good
Measurement of unrelated attributes	medium	excellent	medium
Measurement of redundant attributes	good	medium	excellent

Table 1. Comparing the Computational Capabilities of the Three Algorithms

- It is concluded that the three classification algorithms have their advantages and disadvantages. In the actual process, the selection of the classification algorithms should be carried out according to the actual data and the needs of their own.
- Through our research on the content of this paper, we find that the computation of the decision tree algorithm is more suitable for the whole calculation of this paper. This algorithm is more in line with the requirements of this paper, so a total of data mining algorithms is used to calculate the decision tree algorithm; through our further calculation, the decision tree algorithm is found.
- The C4.5 algorithm is more appropriate for calculation. In the following part, the C4.5 algorithm was analyzed.
- The C4.5 algorithm is also used to construct the decision tree separator and inherit the advantages of the ID3 algorithm. To deal with continuous descriptive attributes, the continuous attributes need to be discretized.
- The minimum value of the continuity attribute is assigned to *MIN*; the maximum value of the continuous attribute is assigned to *MAX*;

(1) *N* is divided into [*MIN and MAX*] intervals, and each breakpoint *B* is obtained. The expression of breakpoint is:

$$B = MIN + \frac{MAX - MIN}{N} * i \quad (1)$$

The average amount of information is used to assess uncertainty, that is, information entropy.

$$S = -\sum_i (P_i * \lg(P_i)) \quad (2)$$

The amount of information used to predict the category of a record in *S* is calculated according to the following formula:

$$Info(S) = Info(S_p, S_n) = - \left( \frac{x}{x+y} * \lg \frac{x}{x+y} + \frac{y}{x+y} * \lg \frac{y}{x+y} \right) \quad (3)$$

The amount of information in the classification of the class in the subclass is:

$$Info(D, S) = \sum_i^k \frac{x_i + y}{x+y} * Info(S_p, S_n) \tag{4}$$

The information increment of nodes is greater than the information increment of all nodes. The information increment of variable  $D$  is:

$$Gain(D) = Info(S) - Info(A,S) \tag{5}$$

The following is the general definition of information gain.

$$Gain(D) = \sum_i^n \left( \frac{|S_i|}{S_1} \right) * Info(S_i) \tag{6}$$

In this way, the calculation formula of the data mining algorithm used is over. We have studied the calculation steps of data mining algorithms. This form of calculation can help us to carry out the calculation and analysis of this paper more quickly. After studying the calculation model of the data mining algorithm, the calculation formula based on the data mining algorithm and the intelligent model in distance education management can be established. The calculation of this model is mainly based on the calculation of the data mining algorithm, so the whole process of data processing is mainly based on the data mining algorithm, which can be analyzed and studied by the intelligent model data processing situation map in the remote education management of the lower figure 2.

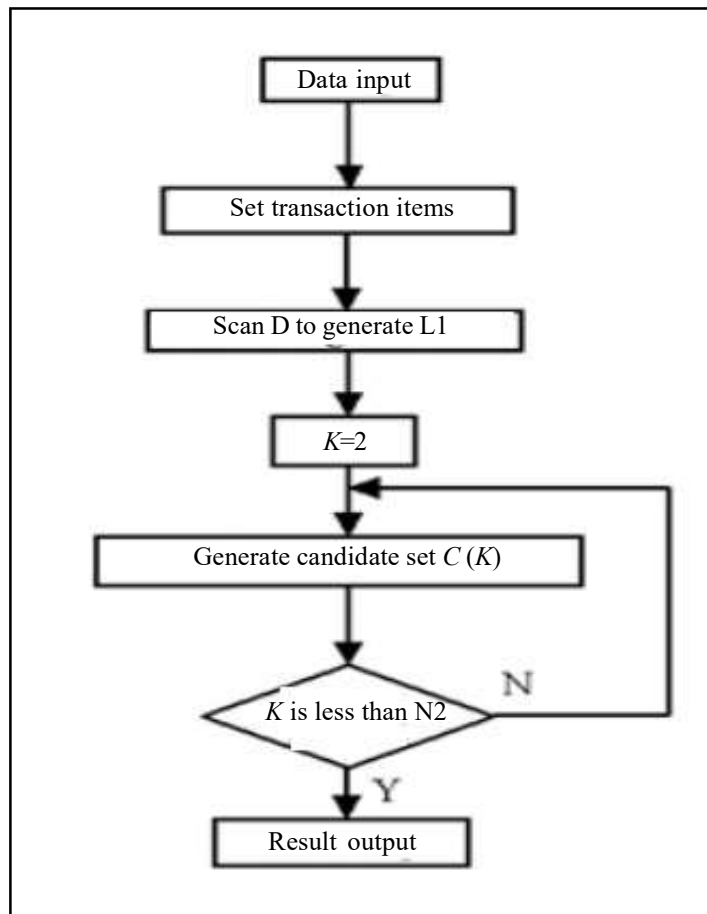


Figure 2. Data processing situation of intelligent model in distance education management

#### 4. Result Analysis and Discussion

In this paper, the calculation process and calculation steps of data mining are analyzed. In addition, the calculation formula of data mining is analyzed and calculated, and a computer model is established. However, to ensure the accuracy of the optimized data mining algorithm and the feasibility of the algorithm. Careful analysis and testing of the algorithm also need to be carried out before putting the algorithm into operation. This paper adopts A comparative experiment to test and analyse data mining. First, we research the data mining process, and then the traditional algorithm with the optimized algorithm is compared. According to the calculation steps of data mining, a set of project sets is set up to carry out the calculation and analysis, and the set is calculated and analyzed according to the calculation model established in the previous article. The results of the collection are shown in Table 2 as follows.

Candidate Set										
Items	A2, Z3	A2, E1	A2, E2	A3, Z3	A3, E1	A3, E2	A4, Z3	A4, E1	A4, E2	Z3, E2
Support%	1	3	2	10	2	9	4	3	3	2.5
Confidence%	10	20	20	30.8	16	60	53.3	9.2	60	12.5
large project set										
Items	A3, Z3		A3, Z2		A4, E3		A4, E2		Z3, E2	
Support%	20		9		4		3		10	
Confidence%	30.8		60		53.3		60		33.3	

Table 2. Association Rules Data Mining Algorithm Secondary Calculation Results

Through the calculation in our previous table, we analyze according to our specific data analysis method, and the preliminary calculation results in our table are correct and can be easily proven. The calculation results meet our requirements and meet the accuracy of our calculation. But this is not enough. We need to calculate the data further. Our final result is shown in Table 3.

The candidate set of items			The candidate set of items	
Items	A3, Z3, E2	A4, Z3, E2	Items	A3, Z3, E2
Support	8%	45.7%	Support	8%
Confidence	2%	40%	Confidence	45.7%

Table 3. Association Rules Data Mining Algorithm Final Calculation Results

The last 3 is the final result of our optimized data mining algorithm. According to the above table, the optimized algorithm can not only calculate the results we need but also calculate several sets of results at the same time, and the consequences of each group are correct and conform to our requirements for the use of the calculation results. The confidence degree and support of the three sets of experiments calculated are correct; the confidence of the A4, Z3, and E2 groups was 40% and the confidence of A3, Z3, and E2 groups of 45.7% satisfied our use requirements.

The above test proves that the optimized data mining algorithm can correctly calculate the results we need, but we still have no accurate grasp of the computational efficiency. This requires us to carry out further testing and analysis. Two groups of experiments have been established, and the traditional and optimized data mining algorithms have been compared. By comparing the computation time, the efficiency of our optimized algorithm can be proved. the 8~24 dimensions were divided into five experimental groups for calculation and analysis. Figure 3 below is a contrast diagram for our experimental results.

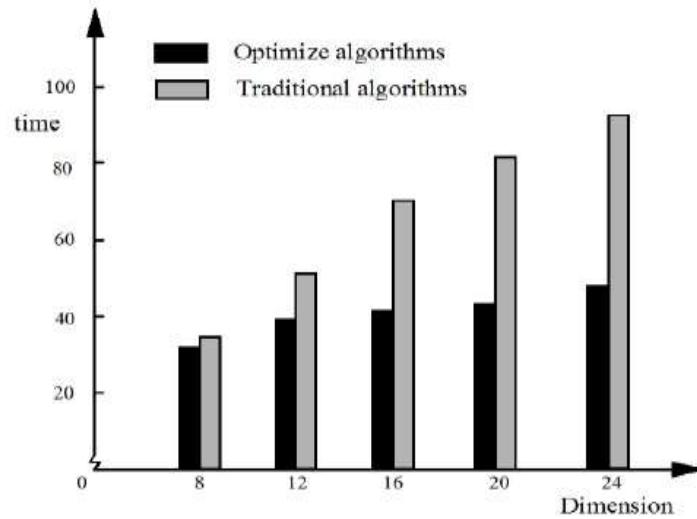


Figure 3. Contrast test test time comparison table

The computation time of the two algorithms above is tested and analyzed. The computation time of the two algorithms will increase with the computation dimension, but the increase time of the two algorithms is very different. The calculation time of the traditional algorithm increases faster. However, the computation time of our optimized algorithm is relatively small. The computation time of the two algorithms increases with the dimension increase, increasing from the first 5 seconds to the last 40 seconds. This proves that the computational efficiency of our optimized algorithm is more than that of the traditional algorithm. Finally, aiming at the task data between 100 and 400, the response time of data mining algorithm is tested. The test results are illustrated as follows:

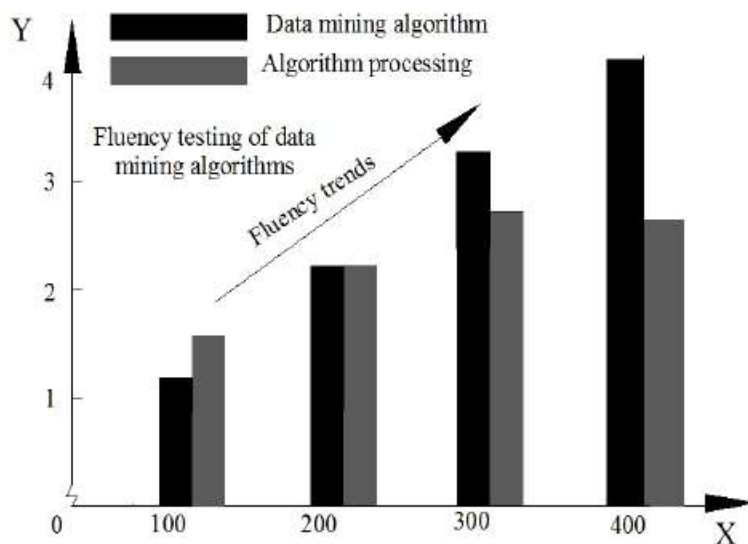


Figure 4. Optimization of processing speed for data mining algorithms

As seen from the above picture, the processing time fluctuated correspondingly with the increasing amount of information. On the premise of not adding data algorithm, the whole processing function falls into the carton. The data processing function of the whole platform has a large lag, and the server runs smoothly under the premise of less or easier data processing. After joining the algorithm, the overall data processing performance is significantly improved.

## 5. Conclusion

With the increasing social information, the era of big data has arrived. Facing the challenges of the new era, PE teaching and deepening reform are being carried out to study this reform trend. This paper used a data mining algorithm to conduct in-depth research on the development trend of public physical education teaching management in Colleges and universities. Through data mining algorithms, powerful computing power can help us solve the problems we encounter. In addition, our detailed test of the optimized data mining algorithm shows that the optimized data mining algorithm can accurately calculate the results we need. Several results can be calculated at one time. In addition, we tested the computational efficiency of the algorithm. The test results show that our algorithm has good computational efficiency. With the increase of the computing dimension, the calculation time of the two algorithms in the 24 dimension is more than 40 seconds. Although we have achieved great success in optimising the algorithm, we need to study the algorithm further and study more excellent algorithms for everyone to use.

## References

- [1] Mohanty, A. K. Senapati, M. R., Lenka, S. K. (2016). Retraction Note to: An improved data mining technique for classification and detection of breast cancer from mammograms. *Neural Computing and Applications*, 27(1), 249-249.
- [2] Robson, B., Boray, S. (2016). Data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple choice medical licensing examinations. *Computers in Biology & Medicine*, 73, 71.
- [3] Hong, H., Pourghasemi, H. R., and Pourtaghi, Z. S. (2016). Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology*, 259, 105-118.
- [4] Rathod, R. R., Garg, R. D. (2016). Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data. *International Journal of Electrical Power and Energy Systems*, 78, 368-374.
- [5] Bandaru, S., Ng, A. H. C., Deb, K. (2017). Data mining methods for knowledge discovery in multi-objective optimization: Part B - New developments and applications. *Expert Systems with Applications*, 70, 119-138.
- [6] Giabbanelli, P. J., Jean, A. (2016). Identifying small groups of foods that can predict achievement of key dietary recommendations: data mining of the UK National Diet and Nutrition Survey, 2008–12. *Public Health Nutrition*, 19(9), 1.
- [7] Wu, W., Peng, M. (2017). A Data Mining Approach Combining K-Means Clustering With Bagging Neural Network for Short-Term Wind Power Forecasting. *IEEE Internet of Things Journal*, 4(4), 979-986.
- [8] Tayefi, M., Tajfard, M., Saffar, S., et al. (2017). hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Computer Methods and Programs in Biomedicine*, 141, 105-109.
- [9] Chen, W., Pourghasemi, H. R., Naghibi, S. A. (2017). Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms. *Bulletin of Engineering Geology and the Environment*, 1-19.
- [10] Alborzi, M., Khanbabaee, M. (2016). Using data mining and neural networks techniques to propose a new hybrid customer behavior analysis and credit scoring model in banking services based on a developed RFM analysis method. *International Journal of Business Information Systems*, 23(1), 1.