

## Retraction

# Retracted: Optimization of Ideological and Political Education Management Strategies under $k$ -Means Algorithm in Big Data Environment

### Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] H. Wang, "Optimization of Ideological and Political Education Management Strategies under  $k$ -Means Algorithm in Big Data Environment," *Security and Communication Networks*, vol. 2022, Article ID 6120230, 14 pages, 2022.

## Research Article

# Optimization of Ideological and Political Education Management Strategies under k-Means Algorithm in Big Data Environment

Hongyan Wang 

*School of Automotive Engineering, Henan Polytechnic Institute, Nanyang, Henan 473000, China*

Correspondence should be addressed to Hongyan Wang; 2013012@hnpi.edu.cn

Received 9 February 2022; Revised 21 March 2022; Accepted 28 March 2022; Published 23 April 2022

Academic Editor: Chin-Ling Chen

Copyright © 2022 Hongyan Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The development of information technology has promoted the reform of ideological and political education in colleges and universities. A large amount of data has been accumulated in the education management database, and the information implied by these data can provide scientific guidance for the optimization of educational management strategies in colleges and universities. The relationship and characteristics of big data in ideological and political education in colleges and universities were expounded, and the feasibility of applying big data technology to ideological and political education was analyzed. The k-means algorithm was selected for cluster analysis, and its process and principles were expounded. As the traditional k-means clustering algorithm has low data processing efficiency and large deviation of the results, the algorithm was optimized by controlling the iterative method of the algorithm. Besides, the ideological and political education management under the optimized k-means algorithm was established. The work assessment quantitative scale in the management of ideological and political education was adopted as the data source, and the optimized k-means algorithm was used to carry out cluster analysis. The results show that management attitude was scored as 0.634, the management ability was 0.6092, the management effect was 0.6082, and the management method was 0.5792. It was indicated that all the scores were above the middle for greater than 0.5, suggesting that the overall management level was above the middle, which was relatively good. The optimized k-means-based ideological and political education management strategy model can analyze the current educational management status of colleges and universities more accurately. It can also provide scientific guidance for colleges and universities to conduct teaching management reasonably and scientifically according to the data analysis results. The optimized k-means algorithm was compared with the traditional algorithm, from which the optimized algorithm was obviously better than the traditional algorithm in terms of clustering effect and operation stability.

## 1. Introduction

With the continuous development of the electronic information industry, big data technology has also played a great role in all walks of life [1, 2]. It not only changes the way people live and learn, but also updates people's thinking constantly. In colleges and universities, the development of information technology has also promoted the reform of teaching methods and the optimization of teaching management strategies. Colleges and universities have gradually entered the era of large-scale data mining (DM) and data application [3]. How to use new media and new technologies to strengthen and innovate the ideological and political education is an important and realistic topic for colleges and universities in the new era [4, 5]. In the education and

teaching management, the informatization of the teaching system and the education management system has led to the generation of massive amounts of data in various educational systems. There is a lot of information within these data. The application of DM technology in the education management system is of important theoretical significance and practical application value [6, 7]. Many scholars have conducted related in-depth works on the massive information contained in big data in a variety of educational systems.

Ji et al. (2020) [8] explored educational DM, with special attention to education big data mining algorithms. First, the relevant elements of educational DM were analyzed, and big data technology was introduced according to the needs of educational data application. Then, the commonly used

education big data mining algorithms and their applications were also introduced, and finally, the development trend of education big data mining algorithms was discussed. Alsuwaiket et al. (2020) [9] expressed the extracted knowledge in a way ensuring accurate and reliable results. The student attendance data collected from the education system were cleaned up, to eliminate any randomness and noise. Then, various attributes were explored to highlight the most important attributes that the affected actual attendance of students. With the attributes selected, an equation was derived to measure the credibility of student attendance. The credibility of the newly developed measurement was also evaluated to check its consistency. Finally, the modules were classified in line with the strength of the attendance credibility value using the J48 DM classification technology. The result showed that the credibility value obtained by the derived equation gave an accurate, credible, and true index of the student attendance rate. The accurate classification of the modules was also performed regarding the credibility of the student attendance rate on these modules. Tan and Lin (2021) [10] proposed a new predictive model to detect the technical aspects of teaching and e-learning using DM in the virtual education system. The association rule mining and supervision technology were applied for factor detection in the virtual education system. The experimental results showed that the proposed prediction model satisfied the accuracy, precision, and recall factor in predicting student teaching and online learning behaviours in the virtual education system. Fang and Lu (2021) [11] analyzed learners' behaviour data in the learning process through the learner model, and three characteristics of learners' cognitive ability, knowledge level, and learning preference were extracted. The preference model was constructed by using the ontology, the semantic relationship among knowledge was better understood, and the interest of students in learning was discovered. Black et al. (2021) [12] analyzed the academic performance and behaviour of some engineering students and collected data from score tables and other related factors. The final model for two datasets was constructed under decision trees and naive Bayes algorithms, and the model could be used to predict the performance of students accurately.

From the current research, all kinds of education, whether psychological education, medical education, or learning model, all contain a lot of data mining and analysis. The management of ideological and political education in colleges and universities also contains a lot of data. Therefore, the ideological and political education of university A in China is taken as an example, the relationship and characteristics of big data in the ideological and political education of universities are first described, and the feasibility of the application of big data technology in the ideological and political education of universities is analyzed. Secondly, the clustering analysis is introduced in data mining, the process and principle of clustering analysis are described, and the k-means algorithm is selected in the partition method for clustering analysis. On this basis, the management strategy of ideological and political education based on the k-means algorithm is established. The Work Assessment Quantification Table in ideological and political

education management in colleges and universities is taken as the data source, and the k-means algorithm is applied to cluster analysis.

## 2. Theoretical Basis

### 2.1. *Big Data and Ideological and Political Education in Colleges and Universities*

**2.1.1. Big Data.** As network technology develops rapidly and the network infrastructure has been improved continuously, data have become the most important method of information storage and information transmission in modern society. Especially with the rapid development of computer technology, the Internet of Things and the Internet have also been applied successively, which will undoubtedly generate countless data. Therefore, modern society has also entered the era of big data [13, 14]. Abundant information is hidden behind the data; the hidden information can be mined through DM and data analysis, thereby providing great guidance for work, research, and even social development. Firstly, the amount of big data is extremely huge, and the data are integrated. In modern society, everyone can make full use of data technology in life, and everyone is a producer of big data. The exchange of data follows new technologies. After the communication among people is increased for their work, research, and life, the use will also inevitably increase. With data development, the accuracy required by data is getting higher and higher. People can participate in the collection and mining of large amounts of data according to their own needs, and combine qualitative and quantitative analyses to explore the development of society and nature. Secondly, there are many types of data, which mainly include the structured data and unstructured data. People are moving into unstructured data, careers, working lives, research, and unstructured models, because most of the data are tied to human activity. Thirdly, the data processing block determines the processing speed of the data. The growth speed of the data is exponential. Therefore, it is necessary to process the data in a timely and effective manner. The data changes dynamically. If the value contained in the data is not processed and calculated timely, the data may lose its original functions, and big data itself becomes useless [15].

**2.1.2. Characteristics of Ideological and Political Education in the Era of Big Data.** Big data has become a keyword in modern society. For the ideological and political education in colleges and universities, there are a mass of data. It is necessary to make full use of data and share data to make ideological and political education more scientific [16, 17]. Therefore, ideological and political education in the big data age mainly has the following characteristics. (1) The main body of ideological and political education is surrounded by big data information. With the global diversification, different cultures and social systems of various countries can be displayed through big data. Different countries have different ideas, and the opposite sex have different cultural backgrounds and adopt specific solutions to different problems. In addition, due to the increasing dependence of

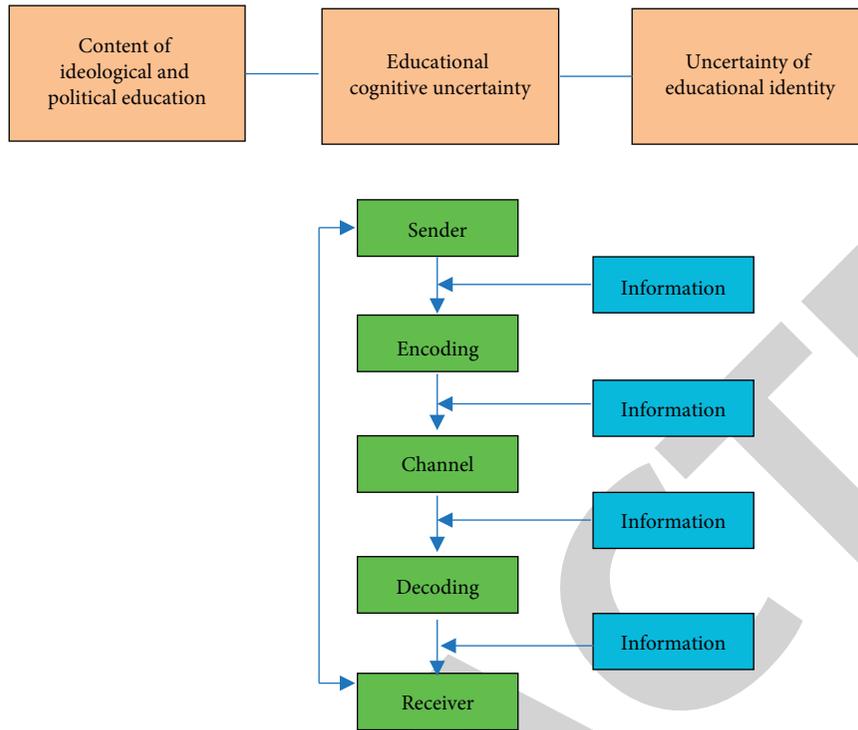


FIGURE 1: Information essence and information flow module.

personal life on data, the massive overloading information will have a deconstructive effect on the core ideas in China, impacting mainstream values and beliefs. (2) The object of ideological and political education is affected by big data information [18]. In ideological and political education, the most obvious word is indoctrination. The transmission of some ruling ideas will help to construct core values in society; but in the era of big data, the amount of logical information that can be contacted by the objects may be at the same level with that by educators. Educators are no longer the only indoctrination subjects [19]. Some communication platforms and mainstream media influence the minds of the educated. These data are not all positive. Thus, if there is no theoretical guidance, institutional constraints, and ethical guidance, it may cause confidential disclosure and chaos when using big data [20]. (3) The process of ideological and political education is disturbed by the dissemination of big data information [21, 22]. However, under the big data model, the huge amount of information and the characteristic of high dispersion can increase the uncertainty of information easily. There is too much valueless information increasing. In the ideological and political education, the information flow can also refer to the general information flow model, which is shown in Figure 1 [23].

2.1.3. *The Relationship between Big Data and Ideological and Political Education.* The relationship between big data and ideological and political education in colleges and universities is shown in Figure 2.

It could be observed from Figure 2 that in the ideological and political education in colleges and universities, the

functions of big data mainly included information collection and screening, scientific prediction and judgment, and personalized education and guidance. Big data contains a lot of information. Some valuable information and negative information are usually mixed. Therefore, in college education, educators must collect data and screen data to filter out valuable information. The harmful or even undesirable information should be resisted. Educators could use big data technology to collect registration information and Internet activities of online users, and filter out the identities of university network users. Then, the core values and the character education of college students could be promoted and strengthened. Big data can be used to predict the network behaviour of college students and explore the law of ideological development of college students. In this way, problems can be found and effective ideological and political education can be carried out. In addition, in a diversified environment, the ideological and political status of different college students can be summarized through the conclusion of data visualization, and it becomes a quantitative and visual form. Thus, the conclusions could be drawn accurately, to conduct ideological and political education concretely for better guidance [24]. Big data technology was also applied to develop personalized hidden education for college students.

2.2. *Clustering Technology in Big Data Mining.* DM [25] refers to extracting knowledge that people are interested in from massive data. These information and knowledge are usually implicit and contain a lot of potential information. The mined information and knowledge can be expressed as

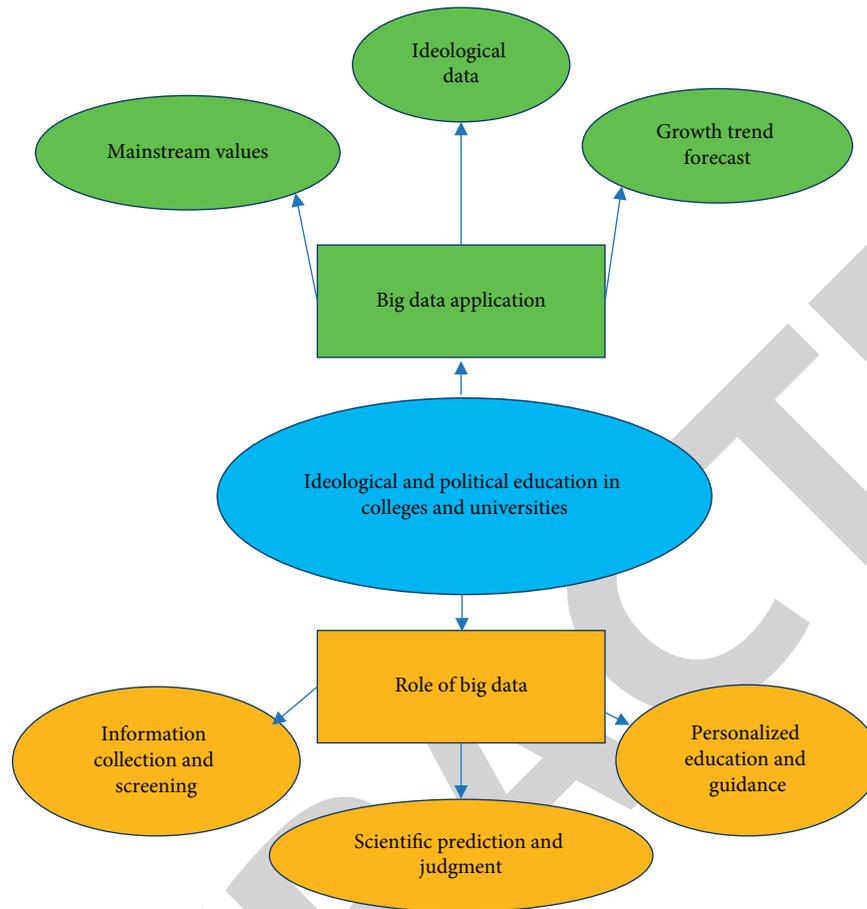


FIGURE 2: The relationship between big data and ideological and political education in colleges and universities.

concepts, rules, laws, and other visualized forms. DM is the process to seek for optimal decision support of the module in the plentiful information. Data obtained by DM can be classified into structured data and unstructured data regarding different objects [26, 27]. DM is a multidisciplinary technology. It is not just a simple data query, but is to dig out the hidden knowledge and information in the massive data from a low level to a high level. After further processing, the mined data can be directly provided to decision-makers to assist the decision-making process. Otherwise, experts can modify their existing body of knowledge. Or, it can be viewed as a knowledge storage structure that transforms new knowledge into an application system, such as expert systems and rule bases. Clustering technology is important in DM, and the application of clustering technology is quite extensive [28].

**2.2.1. Cluster Analysis of the *k*-Means Algorithm.** Cluster analysis is a process of dividing and grouping datasets of fictitious or physical data objects [29]. Its objective is to make similar data objects form a data set. During the process, a generated group of data objects is called a cluster, which is a collection of data objects. In a cluster, objects have a high similarity with each other. While in different clusters, there are large differences among objects. The application fields of cluster analysis are shown in Figure 3 [30].

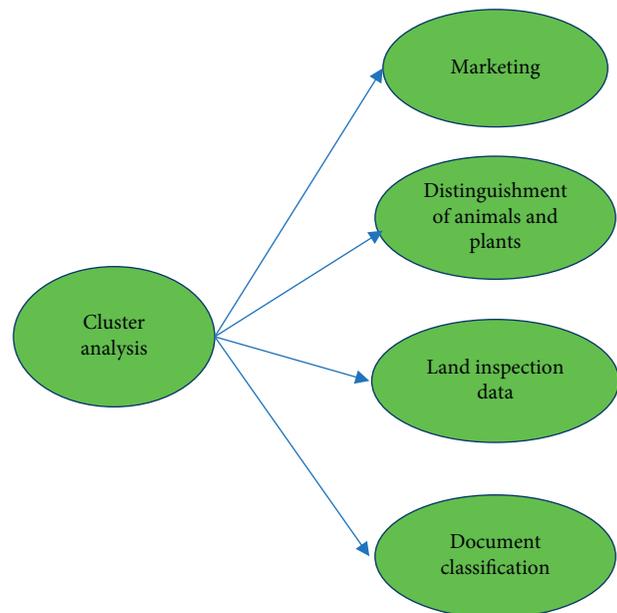


FIGURE 3: Application areas of cluster analysis.

It can be found from Figure 3 that cluster analysis has a wide range of applications. The in-depth analysis and mining can be conducted of data from different industries, to guide

practice better. Now, the massive data sets are processed with the large and complex data warehouses, and DM puts forward higher requirements on the computing power of clustering algorithms.

**2.2.2. Data Types of the *k*-Means Algorithm.** In this section, more data types used in the clustering process are further discussed, thereby further analyzing the preprocessing of these data types [31, 32]. It was assumed that there were  $n$  data objects to be processed in the clustering issue.  $n$  objects represented different types, such as people, cars, materials, and so on. Generally, there are two types of data that appeared in cluster analysis.

**2.2.3. Data Matrix.**  $P$  variables were selected to describe  $n$  objects. For example, price, use, shelf life, and quality were used to describe the properties of the objects. Then, these data were measured with the interval scale to obtain an  $n \times p$  matrix as shown in the following:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}. \quad (1)$$

**2.2.4. Dissimilarity Matrix.** The dissimilarity matrix was used to place the similarity between two of  $n$  objects. Its specific expression can be described as an  $n \times n$  matrix, which is shown in the following equation:

$$\begin{pmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ d(n,1) & d(n,2) & \cdots & \cdots & 0 \end{pmatrix}. \quad (2)$$

In the above matrix,  $d(i,j)$  was the specific quantified form of the dissimilarity between two objects  $i$  and  $j$ . Generally,  $d(i,j) \geq 0$ . When  $d(i,j)$  approached 0, it meant the similarity between the objects  $i$  and  $j$  was high. When  $d(i,j)$  approached the maximum value, it indicated that the dissimilarity between  $i$  and  $j$  was high. Usually, there were different measurement methods for the dissimilarity  $d(i,j)$ . At present, a lot of clustering analyses are based on the dissimilarity matrix, but in some cases, the formation of data is described by the data matrix, so the data matrix should be converted into the dissimilarity matrix before clustering algorithm analysis. The next step is discussing the measurement method of dissimilarity  $d(i,j)$ , usually using the following variable types.

a. Interval scale measurement. The degree of dissimilarity (or similarity) between objects was described by interval scale variables, which was measured by the distance between the objects. The most classic method of calculating

TABLE 1: Condition dependency table.

$i/j$	1	0	Sums
1	$q$	$r$	$q+r$
0	$s$	$t$	$s+t$
Sums	$q+s$	$r+t$	$P$

the distance is Euclidean distance, which is defined as follows:

$$d(i, j) = \sqrt{|x_{j1} - x_{i1}|^2 + |x_{j2} - x_{i2}|^2 + \cdots + |x_{jn} - x_{in}|^2}. \quad (3)$$

Here,  $i = (x_{i1}, x_{i2}, \cdots, x_{in})$  and  $j = (x_{j1}, x_{j2}, \cdots, x_{jn})$  were  $n$ -dimensional data objects. Another distance measurement method is Manhattan (or city block) distance, which is defined as

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \cdots + |x_{in} - x_{jn}|. \quad (4)$$

Both Equations (3) and (4) needed to meet the following conditions. When  $d(i, j) \geq 0$ , both the distances were non-negative values. When  $d(i, j) = 0$ , the distances were 0. When  $d(i, j) = d(j, i)$ , the function between the two distances had a symmetry  $d(i, j) \leq d(i, h) + d(h, j)$  suggested that the direct distance from object  $i$  to object  $j$  was less than or equal to the other distance to object  $j$  through object  $h$ .

b Binary variable. Each variable had two states, which were represented by 0 and 1. 0 means it is empty, and 1 means it is present. It was assumed that the weight of any binary variable was the same, and then, a  $2 \times 2$  conditional dependency table could be obtained, which is shown in Table 1.

In Table 1,  $q$  represents the number of binary variables where the object  $i$  was 1, and  $r$  represents the number of binary variables where the object  $i$  was 1 when the object  $j$  was 0.  $s$  was that where the data object  $i$  was 0 and the data object  $j$  was 1, and  $t$  was that where both the object  $i$  and object  $j$  took 0.  $P$  was the total number of binary variables, and  $p = q + r + s + t$ .

A simple matching correlation coefficient was used to describe the dissimilarity between object  $i$  and object  $j$ , which could be defined as

$$d(i, j) = \frac{(r + s)}{(q + r + s + t)}. \quad (5)$$

If the meanings included in the two states of 0 and 1 were not equally important, the binary variable was called an asymmetric binary variable. In such a case, the Jaccard coefficient was the rating method for measuring the dissimilarity of the asymmetric binary variable. It is expressed as

$$d(i, j) = \frac{(r + s)}{(q + r + s)}. \quad (6)$$

Categorical variables are further extensions of binary variables. For instance, the categorical variable blood type has four possible state values, namely, O-type, A-type, B-type, and AB-type. The categorical variables  $i$  and  $j$  could be calculated with dissimilarity between them, which is expressed as

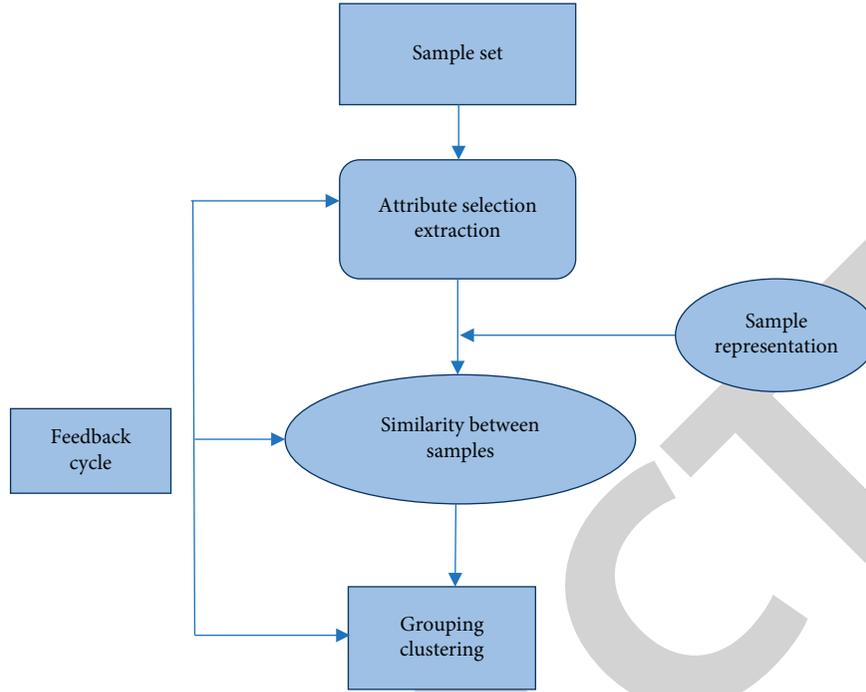


FIGURE 4: Process of cluster analysis.

$$d(i, j) = \frac{p - m}{p}. \quad (7)$$

Here,  $m$  is the number of matches and  $p$  is the number of all variables.

### 2.3. Processes of Cluster Analysis and k-Means Algorithm

**2.3.1. Cluster Analysis Process.** Figure 4 shows the flow of the cluster analysis.

It was found from Figure 4 that the input end of cluster analysis was a sample set composed of data that needed to be processed. The sample set data were preprocessed and classified, and then, the data attributes were extracted. The clusters are grouped according to the similarity of the samples. Cluster analysis was usually a cyclic process, and finally, a sample cluster was obtained. In the clustering algorithm, there are many partition methods. If the number of data objects to be processed was set as  $n$ , the partition method firstly adopted is the objective function minimization. With repeated iterations, the  $n$  objects are divided into  $k$  blocks. Each block represents a cluster, and  $k$  is much smaller than  $n$ . There are many partition methods, among which the k-means algorithm was the most commonly used.

**2.3.2. k-Means Algorithm.** The number of data objects to be processed was  $n$ . The division method was the strategy of minimizing the objective function. Through repeated iteration, the  $n$  objects were divided into  $k$  blocks, each block represented a cluster, and  $k \ll n$  was satisfied. In other words, the  $k$  divisions must meet the following two conditions: (1) each cluster had at least one object. (2) Each

object corresponded to a cluster. k-Means is a commonly used division method, so the algorithm was focused on [33].

The k-means algorithm was to calculate the mean value of the objects in each cluster, which was just the centroid in the cluster. The processing steps of k-means were as follows.  $k$  objects were arbitrarily selected as the preset  $k$  centroid, and then, the centroid of each new cluster was calculated. The above steps were repeated until the criterion function converged to minimize the objective function. The commonly used criterion function was the squared error criterion function, which is shown in the following equation:

$$E = \sum_{i=1}^t \sum_{p \in c_j} |p - m_j|^2. \quad (8)$$

Here,  $E$  represents the sum of the squared errors of all objects in the space;  $p$  denotes all points in the dataset, used to represent the data objects.  $m_j$  represents the centroid of the cluster  $c_j$ , where both  $p$  and  $m_i$  are multidimensional. The k-means algorithm flow is shown in Figure 5; it aimed at making the  $E$  value smaller and smaller. When the final clusters were compact but distinguishable with each other, it indicated that the algorithm performed well. Especially when the number of objects in the dataset was large, the efficiency of the k-means algorithm was better, because its complexity could be calculated as  $o(nkt)$ , in which  $n$  represents the number of objects in the dataset and  $t$  is the number of iterations. In general,  $k$  was much smaller than  $n$ , and  $t$  was also much smaller than  $n$ . The clustering stopped when a local optimum was obtained. The following was an example to illustrate the k-means algorithm.

5 points  $\{x_1, x_2, x_3, x_4, x_5\}$  were described by two-dimensional coordinates, which represent the two-dimensional samples of the cluster analysis  $x_1 = (0, 4)$ ,  $x_2 = (0, 0)$ ,  $x_3 = (3, 0)$ ,  $x_4 = (4, 0)$ , and  $x_5 = (5, 3)$ . The number of initial clusters  $k$

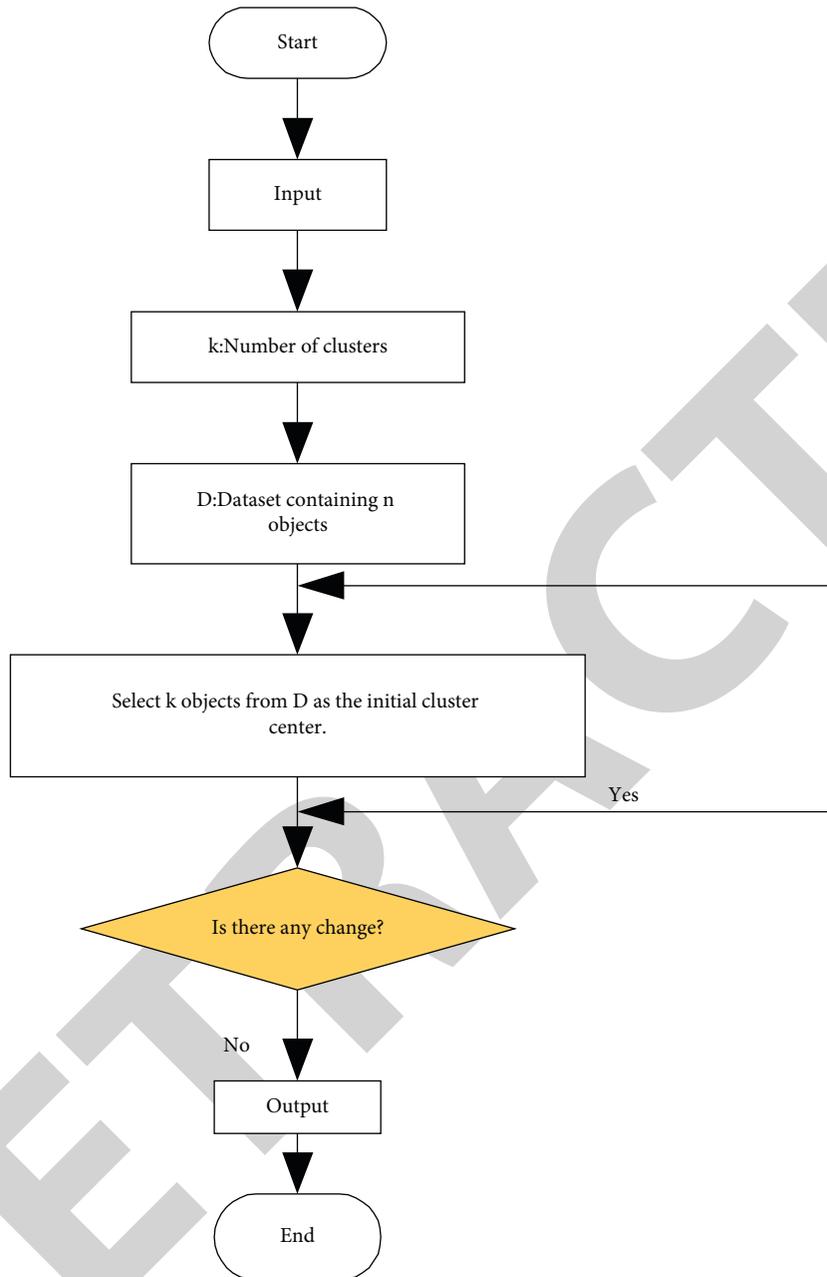


FIGURE 5: Flowchart of k-means algorithm.

was 2. Then, the k-means algorithm was carried out to analyze the following.

*Step 1.*  $k$  objects were arbitrarily selected from the given data samples and taken as the initial cluster centers. The centers  $M_1$  and  $M_2$  of these two clusters are expressed in

$$\begin{aligned} M_1 &= x_1 \\ &= (0, 4). \end{aligned} \quad (9)$$

$$\begin{aligned} M_2 &= x_2 \\ &= (0, 0). \end{aligned} \quad (10)$$

*Step 2.* For the rest of the data objects, the distances from the cluster centers  $M_1$  and  $M_2$  to them were computed using Euclidean distances. Each object was also reassigned to the closest cluster.

For  $x_3$ , the following is obtained:

$$d(M_1, x_3) = \left( (0 - 3)^2 + (4 - 0)^2 \right)^{\frac{1}{2}} = 5 \quad (11)$$

$$d(M_2, x_3) = \left( (0 - 3)^2 + (0 - 0)^2 \right)^{\frac{1}{2}} = 3 \Rightarrow x_3 \in c_2.$$

For  $x_4$ , the distances are computed in the following:

$$d(M_1, x_4) = \left( (0-4)^2 + (4-0)^2 \right)^{\frac{1}{2}} = 5.7 \quad (12)$$

$$d(M_2, x_4) = \left( (0-4)^2 + (0-0)^2 \right)^{\frac{1}{2}} = 4 = \Rightarrow x_4 \in c_2.$$

For  $x_5$ , the distances are calculated in the following equation:

$$d(M_1, x_5) = \left( (0-5)^2 + (4-3)^2 \right)^{\frac{1}{2}} = 5.1 \quad (13)$$

$$d(M_2, x_5) = \left( (0-5)^2 + (0-3)^2 \right)^{\frac{1}{2}} = 5.8 = \Rightarrow x_5 \in c_2.$$

They were updated, and the new clusters  $c_1 = \{x_1, x_5\}$  and  $c_2 = \{x_2, x_3, x_4\}$  were obtained.

The squared error criterion was calculated, and the mean value of the objects in each cluster is worked out by the following equation:

$$\begin{aligned} E_1 &= \left[ (0-0)^2 + (4-4)^2 \right] + \left[ (0-5)^2 + (4-3)^2 \right] = 26 \\ E_2 &= \left[ (0-0)^2 + (0-0)^2 \right] + \left[ (0-3)^2 + (0-0)^2 \right] \\ &\quad + \left[ (0-4)^2 + (0-0)^2 \right] = 25. \end{aligned} \quad (14)$$

The overall squared error was expressed as  $E = E_1 + E_2 = 51$ .

*Step 3.* The new centroid was calculated.

$$M_1 = \left( \frac{(0+5)}{2}, \frac{(4+3)}{2} \right) = (2.5, 3.5) \quad (15)$$

$$M_2 = \left( \frac{(0+3+4)}{3}, \frac{(0+0+0)}{3} \right) = (2.33, 0).$$

Steps 1 and 2 were carried out repeatedly. The distances between each point and the new centroid  $M_1$  and  $M_2$  in this step are calculated in the following equations.

For  $x_1$ , the equation for the distances is shown as follows:

$$d(M_1, x_1) = \left( (2.5-0)^2 + (3.5-4)^2 \right)^{\frac{1}{2}} = 2.55$$

$$d(M_2, x_1) = \left( (2.33-0)^2 + (0-4)^2 \right)^{\frac{1}{2}} = 4.63 = \Rightarrow x_1 \in c_1. \quad (16)$$

For  $x_2$ , the distances are computed as shown in the following:

$$d(M_1, x_2) = \left( (2.5-0)^2 + (3.5-0)^2 \right)^{\frac{1}{2}} = 4.3$$

$$d(M_2, x_2) = \left( (2.33-0)^2 + (0-0)^2 \right)^{\frac{1}{2}} = 2.33 = \Rightarrow x_2 \in c_2. \quad (17)$$

For  $x_3$ , the distances are obtained in the following equation:

$$d(M_1, x_3) = \left( (2.5-3)^2 + (3.5-0)^2 \right)^{\frac{1}{2}} = 3.8$$

$$d(M_2, x_3) = \left( (2.33-3)^2 + (0-0)^2 \right)^{\frac{1}{2}} = 0.67 = \Rightarrow x_3 \in c_2. \quad (18)$$

For  $x_4$ , the following equation calculates the distances:

$$d(M_1, x_4) = \left( (2.5-4)^2 + (3.5-0)^2 \right)^{\frac{1}{2}} = 3.8$$

$$d(M_2, x_4) = \left( (2.33-4)^2 + (0-0)^2 \right)^{\frac{1}{2}} = 1.67 = \Rightarrow x_4 \in c_2. \quad (19)$$

For  $x_5$ , the following equation computes the distances:

$$d(M_1, x_5) = \left( (2.5-5)^2 + (3.5-3)^2 \right)^{\frac{1}{2}} = 2.55$$

$$d(M_2, x_5) = \left( (2.33-5)^2 + (0-3)^2 \right)^{\frac{1}{2}} = 4.02 = \Rightarrow x_5 \in c_1. \quad (20)$$

The new clusters  $c_1 = \{x_1, x_5\}$  and  $c_2 = \{x_2, x_3, x_4\}$  of this step were obtained as these above were updated. The centroids were  $M_1(2.5, 3.5)$  and  $M_2(2.33, 0)$ .

The corresponding variance and squared error are expressed in the following equation:

$$\begin{aligned} E_1 &= \left[ (0-2.5)^2 + (4-3.5)^2 \right] + \left[ (2.5-5)^2 + (2-3)^2 \right] \\ &= 13.75 \end{aligned}$$

$$\begin{aligned} E_2 &= \left[ (0-2.33)^2 + (0-0)^2 \right] + \left[ (3-2.33)^2 + (0-0)^2 \right] \\ &\quad + \left[ (4-2.33)^2 + (0-0)^2 \right] = 8.67 \end{aligned} \quad (21)$$

The overall squared error in this step was described as  $E = E_1 + E_2 = 22.42$ .

It could be clearly shown from the above steps that after one iteration, the overall error decreased a lot from 51 to 22.42. If repeated iterations were continued, the same conclusion would be given. The reason was that all the samples would be divided into the same cluster; as they would not be redivided, the algorithm ended and it will not change any longer.

*2.4. Optimization and Improvement of k-Means Algorithm.* Since the selection of the cluster center directly affected the results of the clustering, and average error of many clusters was relatively large, the local optimum in the clustering process occurred. Therefore, in the process of each iteration, the weight of each data object was regulated, the cluster

center was recalculated, and the threshold was adjusted. In addition, the method of weighting was introduced to measure the size of the weight assigned to the cluster by controlling the error value. Through the weighted penalty method, the data objects farther from the cluster center in the cluster with larger average error would be reclassified to the adjacent cluster with smaller average error. In this iterative process, the boundaries of the clusters were continuously determined by weighted thresholds, until the data objects were classified into the correct clusters. The penalty factor and the memory factor were introduced, and the penalty control of the cluster was realized through the changes of these two factors. The memory factor was initialized by the influence of the weight of the last iteration on the weight of the cluster in this iteration. Then, the penalty factor was introduced to change the cluster with no weight change in the last iteration, so as to improve the stability of the algorithm. Finally, the clustering results could converge to the global optimum successfully. The specific optimization method was described as below.

The dataset to be clustered was defined. At the end of the  $t-1$ -th iteration, the  $k$  cluster centers are expressed as

$$C = \{C_m^{t-1}\}_m^{t-1}. \quad (22)$$

After the  $t-1$ -th iteration, the  $k$  clusters are represented as follows:

$$V = \{V_m^{t-1}\}_{m-1}^k. \quad (23)$$

The error value of the cluster is denoted as follows:

$$f = \{f_m^{t-1}\}_{m-1}^k. \quad (24)$$

Then, the penalty factor  $P_k^t$  of cluster  $k$  at the  $t$ -th iteration was defined by the iteration result of the  $t-1$ -th iteration. It was expressed in the following equation:

$$P_k^t = \frac{f}{1 + \exp(V - C)}. \quad (25)$$

The memory factor was introduced into the weight update process, and the cluster weight calculation result of the last iteration was substituted. As the memory factor was set as  $\alpha$ , the weight of cluster  $k$  at the  $t$ -th iteration was defined in the following equation after being improved:

$$W_k^t = \alpha P_k^t + (1 - \alpha). \quad (26)$$

### 3. Path Exploration for Big Data to Promote Ideological and Political Education

*3.1. Opportunities and Challenges of Ideological and Political Education in Big Data.* The statistics has been made on the characteristics of big data. The data of ideological and political education came from the 2021 quantification scale for the work assessment of counselors in a university in China. The general characteristics of big data are shown in Figure 6.

Figure 6 shows that the amount of big data was huge, the data were integrated, and there are many types of data. The

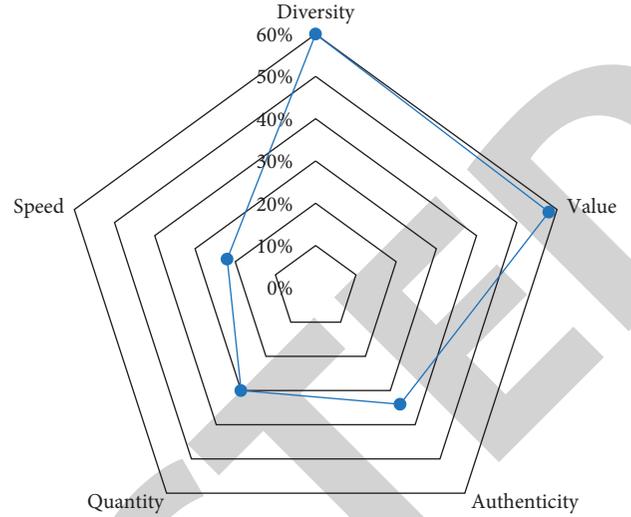


FIGURE 6: Feature attention of big data.

data mainly included unstructured data, which was in a very important position, regardless of the total amount and the reliability of the analysis source.

*3.1.1. Opportunities of Big Data in Ideological and Political Education.* Big data has opened the vision of the subjects of ideological and political education in colleges and universities, changes the traditional education mode, and makes the boundary between the objects and the subjects of education no longer strictly clear. It also impacts the thinking of the educational subjects and breaks the boundary of space. Through high-tech sharing of data, the convenience of technology is widely used in ideological and political education, which truly realized the integration of resources, differentiated analysis, and specific practical teaching. Secondly, big data can improve the quality of ideological and political education objects in colleges and universities. Data expand the knowledge of college students, and students can satisfy their curiosity through data. Students conduct DM, data collection, data analysis, and induction, forming their knowledge structure. Thirdly, big data has updated the dissemination way of ideological and political education in colleges and universities. In traditional information transmission, letters were the main form of communication. However, due to time and space limitations, the information was not transmitted timely. The resource sharing of big data is completely free of time and space constraints. Students can receive school news, regulations, policies, notices, and information freely, and can also strengthen their interaction with the colleges.

*3.1.2. Challenges of Big Data in Ideological and Political Education.* Big data has brought about tremendous changes in ideological and political education, but meanwhile, there are some challenges inevitably. The first is information asymmetry, which indicates differences in the information and structure of the two parties receiving the information. Secondly, big data threatens personal privacy and freedom.

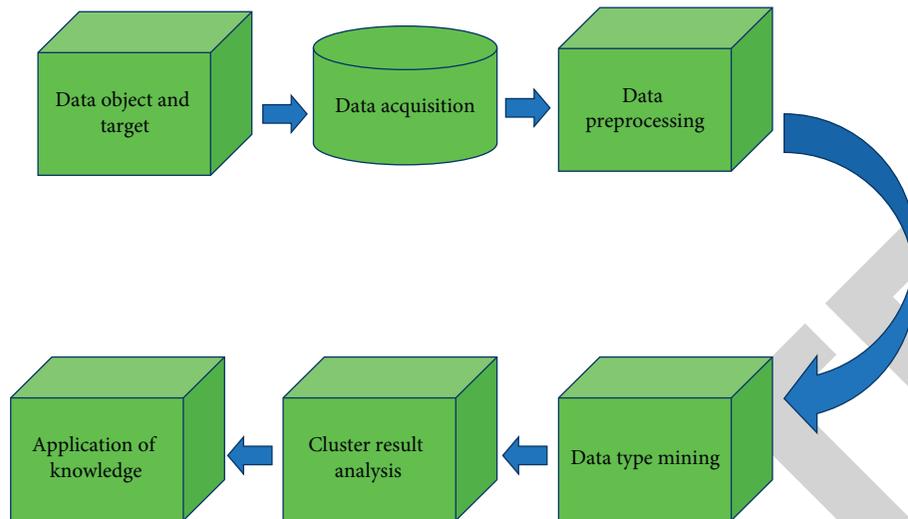


FIGURE 7: Clustering DM process.

Some shopping software would record and count the shopping habits of viewers, and some social software would keep a large amount of personal private information in the background. Thirdly, the digital divide has caused the information gap to widen. At last, culture is left behind science and technology.

### 3.2. Construction of Ideological and Political Education Management Strategy under Optimized *k*-Means Algorithm.

The application of clustering technology in ideological and political education was mainly explored. The counselor work assessment quantitative scale was adopted to conduct DM and analysis. Various data were preprocessed, and finally, cluster analysis was performed through the *k*-means algorithm.

#### 3.2.1. Ideological and Political Education Management.

The continuous deepening of the ideological and political education, as well as the expanding enrollment and other policies of major colleges and universities, has a greater impact on the smooth implementation of college teaching plans. Therefore, the issue between management and learning has always been a difficulty. Therefore, colleges and universities established a team of student counselors. It is very urgent for the current college management to construct a high-efficiency and well-controlled decision supporting system. The system consists of a plenty of data in the education management, but these data only stay at the low-level stage. That's to say, it is only a simple query and simple classification of the data, and the information contained in the data is not mined. Thus, the counselor work assessment quantization scale was taken to dig deeper into the information.

**3.2.2. Solution.** As the counselor work assessment scale was adopted, cluster analysis was applied to perform DM. The answers to some questions were obtained, such as “How is

the management level of college counselors?”, “What loopholes exist in management?”, and “Which aspects of management are not in place?”. The DM process is shown in Figure 7.

From Figure 7, it was suggested that DM for the counselor work assessment scale was divided into 6 steps. First, the data object and its target needed to be clarified. Mining data blindly usually failed to get the desired results instead of some invalid and useless information. Data collection was also the most onerous part. Some data could be obtained directly, and the other data needed to be obtained through investigation. In data preprocessing, because not all collected data could be used, there was a lot of redundant information in the data, and the data needed to be processed into the types available for the model. Data clustering mining was to classify the data, as the data were divided into different groups according to the data models. After clustering result analysis, and finally, the application of knowledge was performed.

**3.2.3. Ideological and Political Management Plan under *k*-Means.** According to Figure 7, a series of tasks such as data collection were carried out:

- (1) Determination of the objects and goals of DM. The data in the counselor work assessment quantitative scale of university A in 2021 were collected. The scale included a series of questions such as “How is the management level of college counselors?”, “What are the loopholes in management?”, “What in the management are not in place?”.
- (2) The counselor work assessment scales in 2021 were collected from the Academic Affairs Office of University A.
- (3) For data preprocessing, the collected data needed to be converted. With the four attributes of management attitude, management ability, management method, and management effect, the data in the work

assessment scales were reorganized and merged. Then, the four attributes were quantified, and the five grades of the evaluation rating are sorted in line with excellent, good, qualified, poor, and awful grades. The grades were projected to the interval [0, 1] using the mapping method. The quantized values of the five grades were set to be 1, 0.75, 0.5, 0.25, and 0, respectively. The values of the four attributes were measured by the arithmetic mean of the items included in them, which are described as follows.

Management attitude = (insistence on standards + harmonious relationship with students + decent talks + objectivity)/4.

Management ability = (accurate grasp of the situation of poor students + accurate grasp of the situation of special student groups + strict education and investigation of students who violate disciplines + competent work + earnest organization for students to do good work)/5.

Management method = (class and dormitory visits more than 3 times a week + insistence on checking the hygiene of the student dormitory once a week + prepared talks with the students during the school year + active participation in and check of the morning running of the students + student scholarships in place)/5.

Management effect = (active participation in or hosting of meetings + understanding of the situation after class + comments and suggestions)/3.

3.2.4. *Algorithm Implementation.* Python was used for further implement of the k-means algorithm. It was supposed that there was a dataset  $D = \{X_1, \dots, X_n\}$ , which aimed at finding  $k$  clusters  $\{C_1, \dots, C_A\}$ . The algorithm steps are shown in Figure 8.

The algorithm flow chart is shown in Figure 9.

#### 4. Result Analysis

The commonly used five clustering algorithms—k-means algorithm, clustering using representative (CURE) algorithm, balanced iterative reducing and clustering using hierarchies (BIRCH) algorithm, density-based spatial clustering of applications with noise (DBSCAN) algorithm, and statistical information grid (STNG) algorithm—are compared. The comparison focuses on six aspects, namely, the scalability of the algorithm, the detection of arbitrary shape clustering, the ability to process noisy data, the sensitivity to the order of input objects, the multidimension, and the efficiency of the algorithm. The comparison results are shown in Table 2.

Since clustering technology is adopted in many application fields of databases, and different application layers have different requirements for clustering algorithms. Table 2 can be used as a reference for selecting clustering algorithms.

In the clustering technology of small- and medium-scale databases, the partition method can get local optimal solution. The partition method has more advantages than other clustering algorithms in terms of easy understanding,

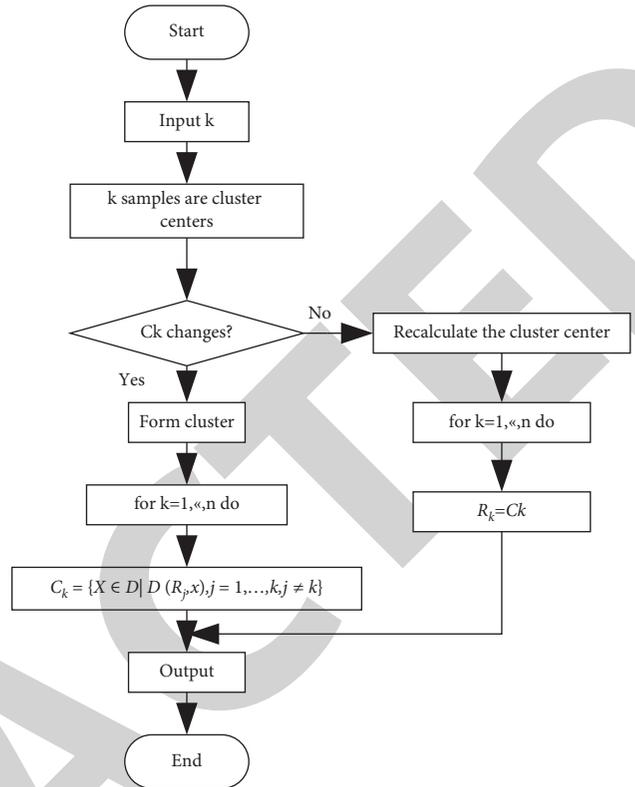


FIGURE 8: Algorithm steps.

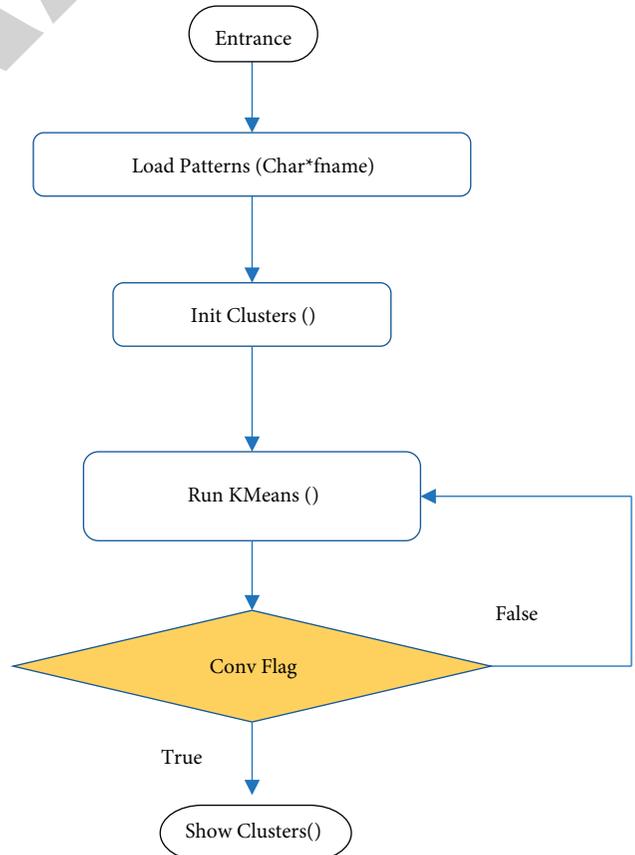


FIGURE 9: Flow chart of the algorithm.

TABLE 2: Comparison results of clustering algorithms.

Algorithm	Scalability	Detection of arbitrary shape clustering	Ability to process noisy data	Sensitivity to the order of input objects	Multidimension
k-Means algorithm	Great	Convex or globose	Sensitive	Sensitive	Medium
CURE	Poor	Arbitrary shape	Insensitive	Sensitive	Good
BIRCH	Good	Convex or globose	Medium	Medium	Good
DBSCAN	Very good	Arbitrary shape	Insensitive	Sensitive	Medium
STNG	Great	Arbitrary shape	Insensitive	Insensitive	Good

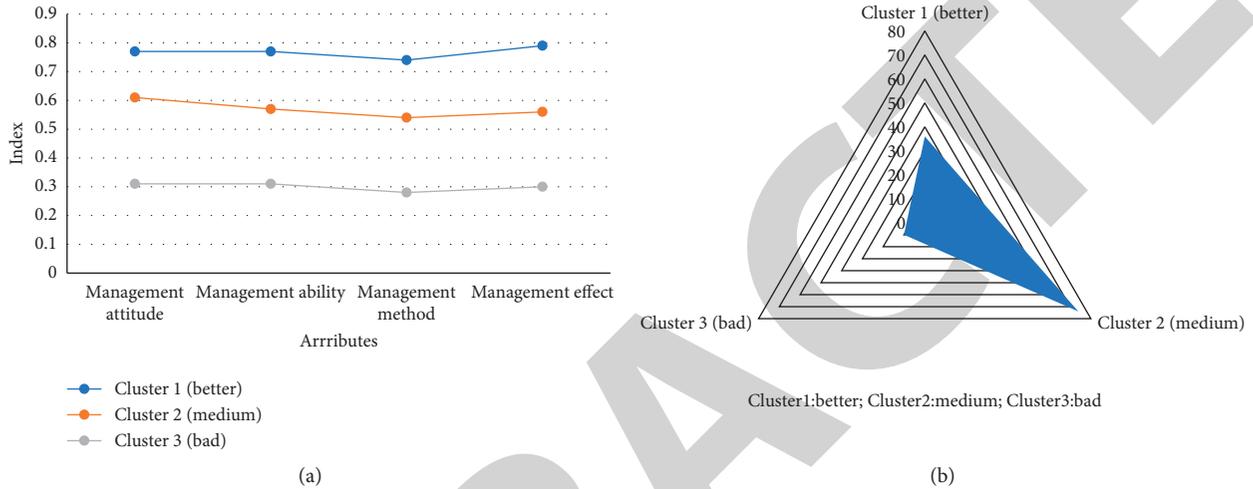


FIGURE 10: (a) Clustering results and (b) sample size.

easy training, easy implementation, and wide use. After the comparison of algorithms, the k-means algorithm is finally selected to achieve the classification of sample data.

The optimized k-means algorithm was applied to classify the sample data, and the three standard samples representing good, medium, and poor ones were compared, which is shown in Figure 10.

From Figure 10, the final proportional distribution range of the included data samples for each cluster was calculated. The proportional distribution range of cluster 1 (good) was 30%, that of cluster 2 (medium) was 62%, and that of cluster 3 (poor) was 8%. Given the results of DM again, the centroids of the three evaluation grades were carefully considered and were compared with the defined standard samples. Most of them were improved, except that the attribute value of the first cluster management method was 0.74, which was lower than the 0.75 of the defined standard sample. The overall score of other clusters was all increased. To get the overall score of each single attribute, weighting was performed due to the unequal sample sizes. The weighting coefficient was the proportion of the samples of the three grades assessed by clustering in the total samples, which were 30%, 62%, and 8%, respectively. The results are shown in Figure 11.

It was found from Figure 11 that the overall scores were ranked from high to low as management attitude (0.634), management ability (0.6092), management effect (0.6082), and management method (0.5792). It could be concluded

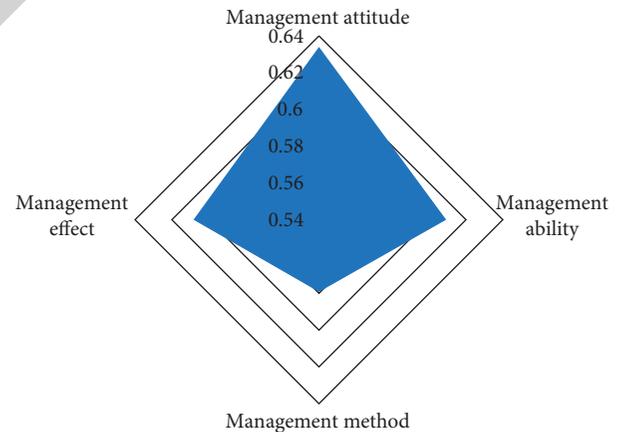


FIGURE 11: The overall scores of the four attributes.

that all the scores were in the upper-middle range as they were greater than 0.5, indicating that the overall management was in an upper-middle level. Thus, the management was relatively good. In the management of students, it was necessary to further strengthen the improvement of management methods, fully understand the actual ideological status of students, understand the life and learning of students actively, and improve the scholarship granting system.

The performance of the optimized k-means algorithm and the traditional algorithm was compared and analyzed. For the verification of the effectiveness and feasibility of the

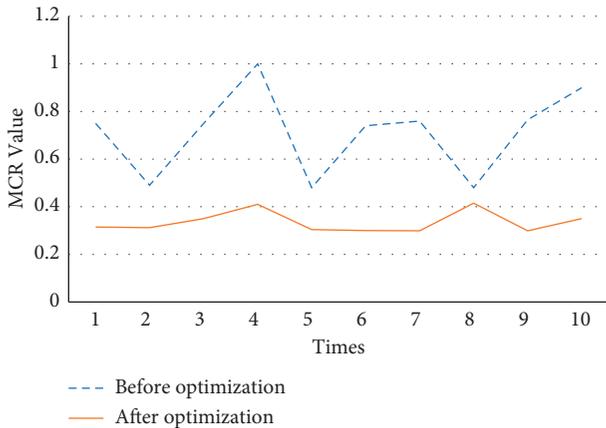


FIGURE 12: Comparison of MCR index values before and after k-means optimization.

improved k-means algorithm, Gaussian function was used to construct random data, and 600 two-dimensional data points were chosen as artificial datasets. The scale parameters were set as  $\mu = 0.33$ ,  $\mu = 0.87$ , and  $\mu = 1.6$ . On this basis, the MCR (M indicates for the dataset, C the cluster, and R the Euclidean space) index was adopted to determine the effect of the artificial data clustering experiment. The MCR index was the ratio of the sum of the improved Euclidean distances to the amount of clustered data. The smaller the value of the MCR index, the higher the effectiveness of the k-means algorithm used. The simulation clustering was performed for 10 times continuously, and the obtained MCR index values are shown in Figure 12.

From the comparison in Figure 12, it was suggested that for the rough k-means clustering of different multidimensional cluster centers, the improved algorithm had a higher accuracy and a smaller average Euclidean distance than the common rough k-means algorithm. It was easy to classify the data into correct clusters, and the improved algorithm promotes the stability and feasibility of the algorithm to a certain extent.

## 5. Conclusion

College education has been into the era of big data. In the management of college education and teaching, there are massive data in various educational systems because of the informatization of the teaching system and the informatization of the educational management system. The traditional k-means clustering algorithm had a low efficiency in data processing with large deviation of the results; thus, the algorithm was optimized. The optimization model for the management strategy of ideological and political education in colleges and universities was also constructed under the optimized k-means algorithm in the context of big data. The work assessment quantitative scale in the management of ideological and political education in colleges and universities was taken as the data source, and the optimized k-means algorithm was used for clustering analysis. The clustering of the work assessment quantitative scale for the counselors was then realized. The students' comprehensively

quantitative score reflected the counselors' work effectiveness directly and was to verify the clustering conclusion of the counselors' work assessment quantification model. It provided scientific guidance for the development of educational work according to the cluster analysis results. Compared with the traditional algorithm, the optimized algorithm was obviously better in terms of clustering effect and operation stability.

The traditional analysis method is based on the calculation of absolute scores, which has some defects in the objectivity and accuracy of evaluation results. It is unfair to evaluate counselors according to the traditional analysis method, and it cannot evaluate the management effect of counselors effectively and properly. Hence, cluster analysis is introduced into the application of ideological and political education management in colleges and universities. The application of data mining technology to automatic data analysis and mining of many evaluation data to find useful information can effectively overcome the defects and deficiencies of traditional analysis methods.

## Data Availability

The labeled dataset used to support the findings of this study is available from the author upon request.

## Conflicts of Interest

The author declares no conflicts of interest.

## References

- [1] S. Tu and M. Zhang, "Research on planning and design of settlement from cities to rural areas based on big data technology," *Soft Computing*, pp. 1–12, 2021.
- [2] K. Rybicka, "Usage of big data technology in controlling," *Research in World Economy*, vol. 10, no. 4, 92 pages, 2019.
- [3] G. Du, Z. Liu, and H. Lu, "Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment," *Journal of Computational and Applied Mathematics*, vol. 386, no. 12, Article ID 113260, 2021.
- [4] C. Hetrick, C. M. Wilson, E. Reece, and M. O. Hanna, "Organizing for urban education in the new public square: using social media to advance critical literacy and activism," *The Urban Review*, vol. 52, no. 1, pp. 26–46, 2020.
- [5] A. Dave, "Review of the textbook and the lecture: education in the age of new media by norm friesen," *The Liminal: Interdisciplinary Journal of Technology in Education*, vol. 1, no. 1, p. 10, 2019.
- [6] F. F. Bbosa, J. Nabukenya, P. Nabende, and R. Wesonga, "On the goodness of fit of parametric and non-parametric data mining techniques: the case of malaria incidence thresholds in Uganda," *Health Technology*, vol. 11, no. 4, pp. 929–940, 2021.
- [7] H. Asami, M. Golabi, and M. Albaji, "Simulation of the biochemical and chemical oxygen demand and total suspended solids in wastewater treatment plants: data-mining approach," *Journal of Cleaner Production*, vol. 296, no. 2–4, Article ID 126533, 2021.
- [8] L. Ji, X. Zhang, and L. Zhang, "Research on the algorithm of education data mining based on big data," in *Proceedings of the 2020 IEEE 2nd International Conference on Computer*

- Science and Educational Informatization (CSEI)*, Xinxiang, China, June 2020.
- [9] M. Alsuwaiket, C. Dawson, and F. Batmaz, "Measuring the credibility of student attendance data in higher education for data mining," *International Journal of Information and Education Technology*, vol. 8, no. 2, pp. 121–127, 2020.
  - [10] C. Tan and J. Lin, "A new QoE-based prediction model for evaluating virtual education systems with COVID-19 side effects using data mining," *Soft Computing*, vol. 4, 2021.
  - [11] C. Fang and Q. Lu, "Personalized recommendation model of high-quality education resources for college students based on data mining," *Complexity*, vol. 2021, Article ID 9935973, 2021.
  - [12] E. W. Black, S. R. Buchs, and B. Garbas, "Using data mining for the early identification of struggling learners in physician assistant education," *Journal of Physician Assistant Education*, vol. 32, no. 1, pp. 38–42, 2021.
  - [13] Y. Xiang and G. Yamamoto, "A data mining approach to investigate the carbon nanotubes mechanical properties via high-throughput molecular simulation," *Materials Science Forum*, vol. 1023, pp. 29–36, 2021.
  - [14] A. Ishaq, S. Sadiq, M. Umer et al., "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE Access*, vol. 9, no. 4, Article ID 39707, 2021.
  - [15] M. Tang and H. Liao, "Multi-attribute large-scale group decision making with data mining and subgroup leaders: an application to the development of the circular economy," *Technological Forecasting and Social Change*, vol. 167, no. 6, Article ID 120719, 2021.
  - [16] Y. Zhang, J. Wang, K. Liang, B. Ling, and X. Li, "Practice and Exploration of curriculum ideological and political education in the construction of online teaching teams in medical universities," *Advances in Applied Sociology*, vol. 11, no. 4, pp. 194–199, 2021.
  - [17] P. Wang, "Realization of the effectiveness of discourse in ideological and political education," *World Scientific Research Journal*, vol. 6, no. 4, pp. 308–311, 2020.
  - [18] X. Wang, "Factors affecting the management of college students in ideological and political education in the information age," *Journal of Physics: Conference Series*, vol. 1744, no. 4, Article ID 042170, 2021.
  - [19] J. Deng, "Research on effectiveness of colleges ideological and political education in the age of big data," *Journal of Physics: Conference Series*, vol. 1744, no. 3, Article ID 032058, 2021.
  - [20] L. Chen, P. Zhi, J. Min, and L. Guan, "[Thinking on integration of ideological and political education into Human Parasitology teaching]," *Zhongguo xue xi chong bing fang zhi za zhi = Chinese journal of schistosomiasis control*, vol. 31, no. 4, pp. 431–433, 2019.
  - [21] X. Li, "Current situation and measures to improve the ideological and political education among students in higher vocational institutions," *Journal of Contemporary Educational Research*, vol. 5, no. 5, pp. 107–110, 2021.
  - [22] Y. Li, "Research on the practice of implicit ideological and political education in tertiary institutions," *Region - Educational Research and Reviews*, vol. 3, no. 2, pp. 22–25, 2021.
  - [23] M. Luo, "A probe into the teaching of preschool education from the perspective of 'course for ideological and political education'," *Vocational Education*, vol. 10, no. 1, pp. 40–43, 2021.
  - [24] Y. Wang, "Research on the innovation of college students' ideological and political education under the new media," *Advances in Education*, vol. 11, no. 4, pp. 1131–1135, 2021.
  - [25] T. Yang, L. Zhang, T. Kim, Y. Hong, D. Zhang, and Q. Peng, "A large-scale comparison of artificial intelligence and data mining (AI&DM) techniques in simulating reservoir releases over the upper Colorado region," *Journal of Hydrology*, vol. 602, no. 6, Article ID 126723, 2021.
  - [26] S. Ramanathan and M. Ramasundaram, "Accurate computation: COVID-19 rRT-PCR positive test dataset using stages classification through textual big data mining with machine learning," *The Journal of Supercomputing*, vol. 10, pp. 1–15, 2021.
  - [27] A. Maqsood, D. Oslebo, K. Corzine, L. Parsa, and Y. Ma, "STFT cluster Analysis for DC pulsed load monitoring and fault detection on naval shipboard power systems," *IEEE Transactions on Transportation Electrification*, vol. 99, p. 1, 2020.
  - [28] S. Qin, L. Vodovotz, R. Zamora et al., "Association between inflammatory pathways and phenotypes of pulmonary dysfunction using cluster Analysis in persons living with HIV and HIV-uninfected individuals," *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 83, no. 2, pp. 189–196, 2020.
  - [29] J. Jang and D. B. Hitchcock, "Model-based cluster Analysis of democracies," *Journal of Data Science*, vol. 10, no. 2, pp. 297–319, 2021.
  - [30] C. Wang, G. Si, C. Zhang, A. Cao, and I. Canbulat, "Location error based seismic cluster analysis and its application to burst damage assessment in underground coal mines," *International Journal of Rock Mechanics and Mining Sciences*, vol. 143, no. 1, Article ID 104784, 2021.
  - [31] H. Fan, V. H. Bennetts, E. Schaffernicht, and A. J. Lilienthal, "A cluster analysis approach based on exploiting density peaks for gas discrimination with electronic noses in open environments," *Sensors and Actuators B: Chemical*, vol. 259, pp. 183–203, 2018.
  - [32] X. A. Bi, Q. Shu, S. Qi, and X. Qian, "Random support vector machine cluster analysis of resting-state fMRI in Alzheimer's disease," *PLoS One*, vol. 13, no. 3, Article ID 0194479, 2018.
  - [33] S. Krishnakumar and K. Manivannan, "Effective segmentation and classification of brain tumor using rough K means algorithm and multi kernel SVM in MR images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 6751–6760, 2021.