

## Research Article

# An Analytical Model of College Students' Mental Health Education Based on the Clustering Algorithm

Jing Lei 

*School of Business Guangzhou College of Technology and Business, Guangzhou 510800, Guangdong, China*

Correspondence should be addressed to Jing Lei; [leijing@gzgs.edu.cn](mailto:leijing@gzgs.edu.cn)

Received 8 July 2022; Revised 30 August 2022; Accepted 5 September 2022; Published 19 September 2022

Academic Editor: Gengxin Sun

Copyright © 2022 Jing Lei. 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 paper proposes an improved  $k$ -means clustering algorithm to analyze the mental health education of college students. It offers an improved  $k$ -means clustering algorithm with optimized centroid selection to address the problems of randomly selected class cluster centroids that lead to inconsistent algorithm results and easily fall into local optimal solutions of the traditional  $k$ -means clustering algorithm. The algorithm determines the neighborhood parameter based on the Euclidean distance between the data object and its nearest neighbor in the data set. It counts the object density based on the neighborhood parameter  $Eps$ . In the initial class cluster centroid selection phase, the algorithm randomly selects the first-class cluster centroid, and subsequent class cluster centroids are chosen based on the data object density information and the distance information between the data object and the existing class cluster centroids. The proposed improved  $k$ -means clustering algorithm and clustering validity metrics are tested using several simulated and real datasets. In this paper, the characteristics and application areas of the improved  $k$ -means clustering algorithm are sorted out, the self-determination theory related to the enhanced  $k$ -means clustering algorithm is investigated, and the behavior of the improved  $k$ -means clustering algorithm in the enhanced  $k$ -means clustering algorithm system and the octagonal behavior analysis method is also sorted out through the improved  $k$ -means clustering algorithm mental health management cases. The path of intervention in mental health education is designed through the improved  $k$ -means clustering algorithm. The intervention points are explained, including motivation discovery, mechanism setting, and component matching of the enhanced  $k$ -means clustering algorithm.

## 1. Introduction

To improve the efficiency of mental health education in schools, the organization and analysis of student mental health data are essential. This mental health analysis system uses more common data analysis methods, such as cluster analysis, to effectively analyze students' psychological information problems. The organization and analysis of student mental health data are essential [1]. The mental health analysis system uses the more commonly used data analysis methods, such as cluster analysis, to effectively analyze the problems of students' psychological information. Relying on the psychological management education system platform, combined with data mining technology, through many data analyses, the potential, and data in the implicit information, such implicit details, will provide the corresponding

reference basis for major universities, as well as practical solutions. Today, increasingly severe psychological problems have become one of the important research directions of psychology and education. Today's college students are facing unprecedented pressure [2]. How can college counselors or teachers get more information about mental health in a shorter time and better intervene in college students' mental health problems.

How to let college psychological counselors or teachers obtain more mental health data in a short time and better intervene in college students' mental health problems. Therefore, data mining technology, combined with a student psychological database and a psychological management system to analyze students' psychological and behavioral data, has a high research value and has become an important research direction for major

universities [3]. A commonly used clustering algorithm in various fields is based on the partitioning method. It is not easy to obtain valuable information from the fragmented student information by traditional methods, so the fragmented student psychological data are formed into multiple correlation infographics according to the associated attributes [4]. By using cluster analysis in the student psychological education system, we can analyze the potential value of information on student psychology and the correlation between each information factor from the vast amount of student psychometric data in a school database and provide more scientific solutions for student mental health to the general psychological teaching staff. At present, data mining methods have achieved good results in many fields. These research results provide theoretical and technical support for my psychological data from psychoeducational systems.

The clustering algorithm is a commonly used clustering algorithm based on division methods. The data mining cluster analysis research can be divided into two directions: improvement of the clustering algorithm and optimization of the clustering effectiveness index. Among them, the improvement focuses on optimally separating the data objects into the data set given by the optimal number of clusters in the data set. In contrast, the optimization of the clustering effectiveness index mainly focuses on paying attention to the problem of evaluating the quality of the clustering algorithm results after the  $k$ -means clustering algorithm is executed. Because it is simple, effective, and suitable for analyzing large data sets, the clustering algorithm uses distance as a similarity measure and divides the samples into different clusters based on the similarity [5]. In the same group, the similarity between samples is ensured to be great, and the dissimilarity between clusters is excellent. Several studies have found that the clustering algorithm has two main problems: determining the number of groups ( $k$ -value) and selecting the initial centers since the initial center. The reasonableness of the initial center point selection directly affects the quality of the clustering results of the  $k$ -means algorithm, which can be obtained from the research status of cluster analysis. The improvement of cluster center selection mainly includes minimum variance, reverse nearest neighbor search, based on the idea of division, and using dissimilarity matrix to construct Huffman tree and other methods. In the psychological management system using cluster analysis technology and psychological information mining, from the vast amount of student psychological data to mine the association between different psychological assessment data and value information rules, construct a classification model of psychological disorders and use mining technology to verify the application data of the student psychological management system. The problem of finding information, the method, and improvement suggestions of the original student psychological profile construction was proposed [6]. The probability of solving psychological diseases is greatly improved.

## 2. Related Works

Research in data mining clustering analysis can be divided into two directions: improving clustering algorithms and optimizing clustering validity metrics. In contrast, the optimization of cluster validity metrics focuses on the quality assessment of the clustering algorithm results after executing the  $k$ -means clustering algorithm [7]. Daenekindt uses a  $k$ -means algorithm to analyze the mental health of a class, complete the classification of students and offer personalized teaching strategies based on the characteristics of each type, which helps to accomplish actual teaching. In the paper, Huisman constructs the feature clustering model of learning style and mental health, uses a  $k$ -means algorithm for clustering analysis and completes the design of a student grouping system [8]. Li used  $k$ -means to analyze students' campus behaviors on mental health. The results showed that the number of trips to the library, borrowing professional books, resting time, and other habits are closely related to performance [9]. Hutchinson DM used the optimized  $k$ -means algorithm in her paper to analyze the characteristics of school behaviors such as academic performance, scholarship, and competition status of school students and came up with typical features of each type of students, which helps teachers to teach students according to their needs and manage them efficiently [10].

The first psychological clinic was established in the 1890s, which opened a new chapter in integrating psychology and education. Mental health education is provided through courses, by classifying students in different psychological states or of other races and genders, and by carrying out more targeted educational activities related to students' lives, such as setting up peer counseling stations and mental health books corners [11]. Mental health education in developed countries has professional mental health teachers and education teams, and mental health teams are trained through a series of professional training programs and standards [12]. Mental health education focuses on students' inner spiritual world and consciousness. The quality of the clustering results is not only the selection of the clustering center but also the determination. Another important indicator to measure the clustering effect is selecting the appropriate number of clusters and dividing the data to be tested reasonably. It involves students' spiritual state, faith orientation, life attitude, physical health, and related environment. Secondly, mental health education should be adjusted to treat and channel mental health problems after they are formed. It should focus more on preventing psychological difficulties and guiding psychological development toward health and positive energy. In addition, improve the teaching staff's moral quality and the professional academic level and use the teachers' natural charisma to influence and cultivate students subconsciously. The advantage of mental health education in overdeveloped countries is that it is strongly supported and promoted by the government with excellent organization and power. In contrast, most developed countries are led by each university, so various forms exist. In conclusion, although there

is a gap between mental health education in developing and developed countries, we are gradually narrowing this gap by learning from our strengths and weaknesses.

The increasing attention to the research on clustering algorithms and the development of  $k$ -means clustering algorithm techniques have encouraged scholars from major universities to research clustering algorithms. The use of cluster analysis research methods uses clustering algorithms to achieve functional analysis of students' psychological states by extracting data features in the students' primary information database and obtaining accuracy and stability. The algorithm randomizes the selection of  $k$ -class cluster centroids given the number of clusters. To avoid the problem that the algorithm quickly falls into the local optimum due to the random selection of initial class cluster centroids, Huebner proposed the semi-supervised  $k$ -means++ clustering algorithm. Yu proposed the DC- $k$ -means algorithm in 2018; based on the Canopy algorithm idea, the initial class clusters of the  $k$ -means algorithm were selected based on the distance between the data object and the existing class cluster centroids [13]. Since the twentieth century, many researchers have used the data accumulated in the psychological assessment system of colleges and universities for data mining and did a lot of related research. For example, Wenjuan Qi et al. applied association rule mining to college students' mental health assessment data. Thompson combined the improved  $k$ -means clustering algorithm with the established psychological correlation analysis system of college students [14]. This paper uses the enhanced  $k$ -means clustering algorithm to mine the correlations between nine dimensions of psychological symptoms. The mining results are analyzed to assist in the intervention and prevention of college students' psychology.

### 3. Analysis Model Design of College Students' Mental Health Education Based on the Clustering Algorithm

**3.1. Clustering Algorithm System Model Construction.** The execution of the  $k$ -means algorithm requires a given number of clusters  $k$  value, which is an essential factor in determining the good or lousy clustering results. The user subjectively selects the  $k$ -value based on experience. When clustering, it is necessary to decide the number of clusters because human subjectivity often makes the  $k$ -value deviate from the precise number of groups. There are many improved algorithms for determining the number of sets. The idea of most of the enhanced algorithms is that, firstly, the clustering evaluation index function is proposed using the free combination of inter-class distance and intra-class distance. Secondly, the  $k$ -values are calculated separately in the range of  $[2, \text{int } n]$ . The optimal clustering index value is searched, at which time the corresponding  $k$ -value is the optimal number of classifications of the data set [15]. The algorithm also tends to fall into local optimal solutions. By choosing high-quality initial centroids, the improved clustering algorithm improves the accuracy of the algorithm and the stability of the clustering

results compared with the traditional  $k$ -means clustering algorithm. The flow chart of the enhanced  $k$ -means algorithm steps is shown in Figure 1. The advantages of the  $k$ -means clustering algorithm are (1) the algorithm is fast and simple; (2) it is efficient and scalable for large data sets; (3) the time complexity is nearly linear and suitable for mining large data sets. The time complexity of the  $k$ -Means clustering algorithm is  $O$ , where  $n$  represents the number of objects in the data set,  $t$  represents the number of iterations of the algorithm, and  $k$  represents the number of clusters.

The clustering effect is not good, and it is easy to fall into the local optimum. The isolated points have a significant influence on them. And the improvement of the clustering center selection is mainly based on minimum variance, reverse nearest neighbor search, division-based idea, and Huffman tree construction using the dissimilarity matrix. We want to select the high-density data points that are far away and have data representative characteristics as the initial centroids, so this paper uses the density-based method of the improved  $k$ -means algorithm to select the initial centroids and selects the centroids one by one according to the density of the sample points and the radius of the neighborhood until  $k$  initial centroids are chosen. The proposed method of selecting the initial centers is inspired by considering the weights of the attribute indicators, combined with the idea of density to determine the initial centers, and the data set  $D$  is the new data set after weight assignment; the relevant definition is as follows: the density of the sample object Known clustering data set to be measured  $D = \{x_1, x_2, \dots, x_n\}$ , then the density of clustered objects is as follows:

$$\text{density}(x_1) = \{x \leq D | \text{dist}(x_1, x) \in R\}. \quad (1)$$

The region within the radius  $R$  is  $x_1$  the neighborhood, and  $x_1$  the Euclidean distance between the number of objects more minor than the radius of the area is recorded as  $x_1$  the density of the number of things. The more the number of objects, the greater the thickness. The selection of the neighborhood radius has a significant impact on the density estimation.

$$\text{mean}(D) = \sqrt{n(n+1)} \sum_{i,j=1} (x_i - x_j). \quad (2)$$

Here,  $\text{mean}(D)$  is the mean value of the distance between all clustering objects,  $\text{coef}R$  is the radius adjustment coefficient, referring to the experience of other researchers so that  $\text{coef}R = 0.13$  the clustering effect is relatively good currently. It is the choice of clustering centers and the determination of the number of clusters that affect the clustering results. Choosing the correct number of groups and dividing the data reasonably is another vital index to measure the clustering effect. The  $k$ -value is given in advance according to experience, which is subjective, and the clustering results are not very satisfactory. Moreover, the clustering evaluation criterion function only considers the intra-class compactness, which leads to inaccurate classification and too one-sided evaluation [16]. Many scholars believe in addressing the problem of  $k$ -value selection and the limitations of the

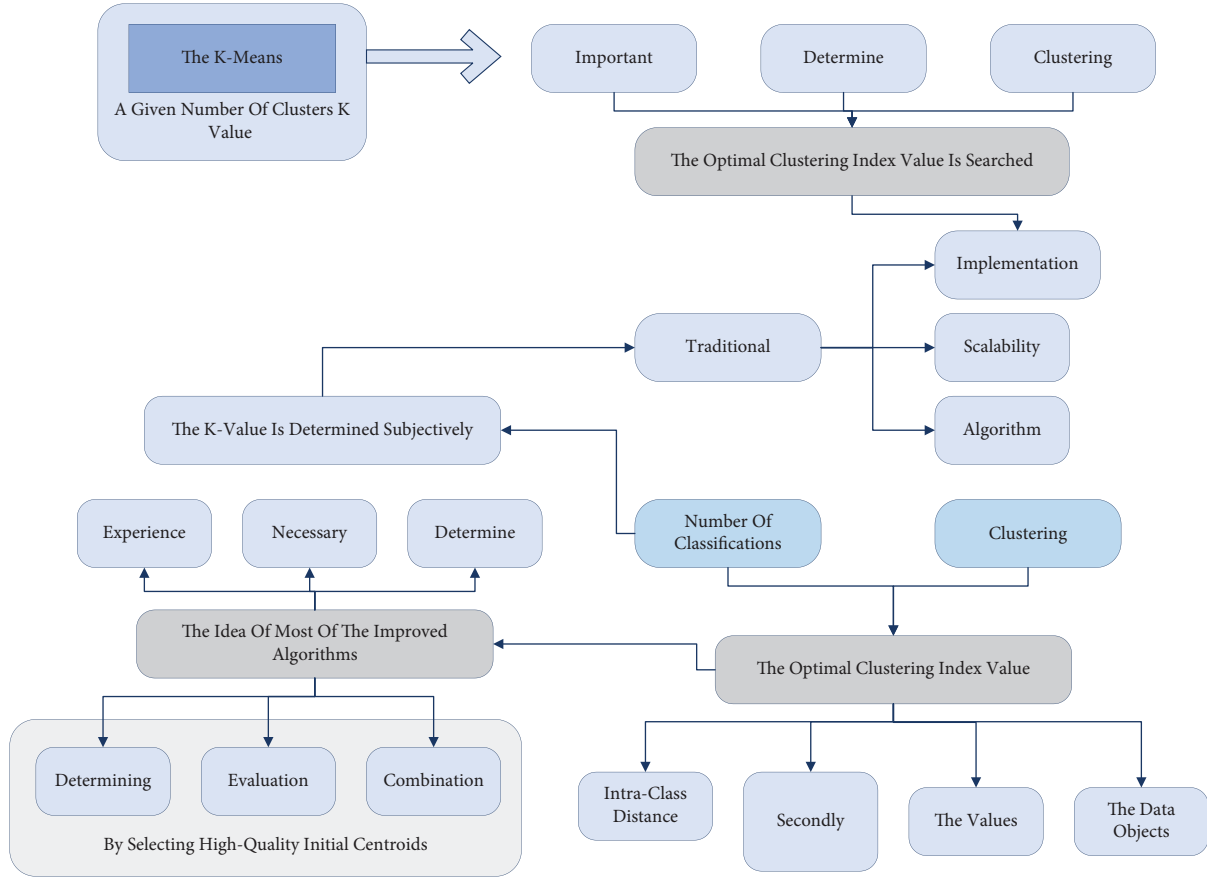


FIGURE 1: Flow chart of the improved k-means algorithm.

traditional evaluation function. Section, the individual intra-class distance of the sample data is defined as the cluster center of the course. Average: the sum of the intra-class distances of all  $k$  classes is the intra-class distance we seek. This definition can reasonably measure the intra-class similarity of the sample. The inter-class distance is the average Euclidean distance between the  $k$  centroids, and this value is a good measure of the similarity between the clustering centers (Related concept definition: intra-class distance). Suppose the data set to be  $n$ -dimensional data objects, which are into  $k$  classes. The average distance from the sample points to the cluster centers in each category is the individual intra-class distance. The average of all intra-class distances of the  $k$  classes is defined as the intra-class distance required for clustering.

$$W(k) = \frac{1}{k} \sum_k \frac{1}{n} \sum_{n_1} d(x_j, c_i). \quad (3)$$

Analyzing the clustering criterion function, we can get that the intra-class distance of the sample data in dataset  $D$  that can reflect the relationship between the sample objects and the cluster centers and the intra-class structure characteristics; the inter-class length of the sample data can remember the relationship between the clusters and the inter-class structure characteristics. The practical criterion function of clustering can better evaluate the accuracy of the

clustering effect, and the value of  $k$  is close to the optimal value  $W(k) = B(k)$ .

$$k_{\text{best}} = \min\{iawrb(K), K = 1, 2 \dots K\}. \quad (4)$$

The improved algorithm does not need to determine the value of  $k$  in advance  $\sqrt{n}$ . It lets the algorithm execute the improved k-means algorithm from a  $2 \sim in\sqrt{n}$  loop and selects the value of  $k$  corresponding to the smallest value of the criterion function at the end of the circle, which is the value of  $k$  when the clustering result is optimal. Thus, the improved algorithm can automatically determine the number of clusters according to the minimum value of the criterion function. In the initial centroid selection, the density-based idea is used to optimize the initial clustering centers' passage and improve the initial centers' quality  $T$ .

First, the weights of each attribute are calculated according to the entropy method to form a new dataset  $D$ . Input: dataset  $D$  with  $n$  data objects and  $\text{coef}R = 0.13$ . Output: the optimal clustering number of  $k$  values. (1)  $2 \sim in\sqrt{n}$ ; (2) calculate the Euclidean distance between the remaining data objects and the initial centers and divide the clusters they belong to based on the minimum Euclidean distance; (3) end of the algorithm. The improved algorithm first uses a density-based selection of initial centers to avoid the instability of random selection and obtain stable clustering results without iterating them many times. Thus, the

number of iterations is reduced compared with the original algorithm. The weights of each attribute index are considered in the clustering analysis, making the clustering more accurate. The improved clustering algorithm first selects the initial center based on density, which avoids the instability of random selection and can obtain stable clustering results without multiple iterations. Hence, the number of iterations is lower than the original algorithm. In the cluster analysis, the weight of each attribute index is considered, which makes the clustering more accurate.

**3.2. Analysis of Model Design of College Students' Mental Health Education.** The multilevel analysis model of college student's mental health management system is summarized by combining the relevant elements designed by the improved k-means clustering algorithm, as shown in Figure 2. Then the gamification strategy of college students' mental health management under different module contents is proposed according to this. College students come from other places, their growing environment is different, and their psychological state is also very different. After entering university, facing an independent living environment, and dealing with society, study, and interpersonal relationships, college students will have some mental health problems, leading to deviation in the direction of psychological development if not solved in time. Mental health education in colleges and universities can enable students to maintain a positive attitude towards life, actively adjusting their emotions and relieving their psychological pressure at the right time when they encounter difficulties. A healthy psychological state can help college students give better play to their potential and better adapt to the stress arising from independent living and studying, as well as understand the nature of society more clearly, set up life goals, and make themselves high-quality talents urgently needed by organization. The details are as follows: in the mental health data monitoring stage, an encouragement monitoring strategy oriented to achievement reward should be followed; in the mental health level assessment stage, an induction assessment strategy driven by unknown loss should be followed; in the mental health education stage, an inspirational education strategy targeting interest socialization should be followed; in the mental health intervention stage, an intensive intervention strategy with task punishment as a constraint should be followed. The author provides a specific explanation through strategy scientific and feasibility analysis through corresponding design or management cases.

The in-depth development and research results of contemporary college students' mental health education have gradually increased [17]. The clustering algorithm is used to realize the functional analysis of students' psychological states by extracting the data features in the primary student information database and getting an accurate and stable classification. Then, through the subjective analysis of school counselors and student managers, the type achieved by this algorithm proved to be of a specific reference value, which can provide more and better services to the relevant staff for the mental health management of students and

provide a new working idea and working mode. Later, we will use the clustering method to implement cognitive health management systems for college students. We can obtain more basic data about students and make data extraction more stable, reliable, efficient, and scalable. Finally, we can reflect the students' psychological status scientifically and reasonably, thus giving more information to appropriate instructors. Analyzing the existing student workflow and organizing related research activities, we summarized and contained the detailed functional requirements based on the basic needs, as shown in Figure 3.

The conversion direction in the framework of the clustering algorithm and closely around the issue of cultural appropriateness. The goal of converted mental health education for college students is to cultivate psychological growth, enhancing students' positive psychological qualities. The theory is used to guide practice, and the goal of the converted theory is psychological growth, so the construction of the working model should also take psychological growth as the goal. Under the multicultural framework, colleges and universities' mental health education model is a "growth model." The characteristics of this model are: everything is based on the premise of cultural appropriateness, paying attention to the needs of students' psychological growth, and integrating multiculturalism as a concept, method, and technology into every aspect of the mental health education model. This "growth model" differs from the previous models regarding teaching concepts, educational subjects, educational medium, and management mechanism. The rule is to achieve the splitting of samples at a given node, which is also used as an attribute selection metric [18]. Relying on the platform of the psychological management education system, the potential and implicit information in the data can be obtained. Effective solution, information gain is the splitting rule of the ID3 algorithm, which favors the selection of attributes with multiple attribute values; however, in many cases, such features with multiple attribute values usually do not have a practical categorical meaning, such as school number. The ID3 algorithm is based on Occam's Razor (using less to do just as well): the smaller the decision tree, the better it is over. The most important feature of the C4.5 algorithm is that it overcomes the disadvantage of ID3 bias towards the number of features and introduces the information gain rate as a classification criterion. The C4.5 algorithm optimizes this by changing the splitting rule to rate again. Let  $D$  be a training sample set containing class markers, and the class marker attribute has  $m$  different attribute values corresponding to mother classes  $c_i = (i = 1, 2, 3, \dots, m)$  defined as the  $c_{id}$  set of pieces of courses  $c_i$  in the training sample set  $d$ ,  $|d|$  be the number of samples in  $d$ , and  $|d_{id}|$  be the  $d_{id}$  number of samples in  $d$ . The following equation can derive the expected value required to classify the  $D$  pieces.

$$\text{Info}(D) = \sum_{i=1}^m \frac{d_i}{d} \times \log\left(\frac{d_i + 1}{\sqrt{d} - 1}\right). \quad (5)$$

Suppose the attribute is discrete, taking values and the training sample set  $D$  is partitioned by point A. In that case,



FIGURE 2: Multilevel analysis model of college student’s mental health education system.

$D$  is divided into  $v$  subsets  $\{d_1, d_2, d_3, \dots, d_v\}$ , where the samples take matters on the attribute  $A$  with values  $a_j$ . These subsets then correspond to several branches of  $D$ . This leads to the following equation:

$$\text{Info}(A) = \frac{\sqrt{d_j/d}}{\text{Info}(d_j)} \tag{6}$$

The system dynamics simulation study is completed for the total system of mental health education, the design of students’ characteristics, the subsystem of school factors, and the relationship between the different input ratio structures of the system and the changing trend of the level of mental health education for college students are compared and analyzed. By using cluster analysis in the students’ psychological education system, the potential value information of students’ psychology and the correlation between various factors are analyzed from the massive psychological measurement data of students in the school database to provide more psychological teaching workers with more information. Science, student mental health solutions through the subsystem simulation, is the actual rate of the effect of each subfactor on the story and the subsystem is calculated. The substantial impact of each influencing factor is quantitatively analyzed to provide personalized, scientific,

and reasonable teaching decisions for student management and improvement.

#### 4. Analysis of Results

**4.1. Clustering Algorithm System Model Analysis.** The category is based on the extent of similarity between models. Therefore, the clustering results reflect the samples’ inherent characteristics. In the psychological management system, cluster analysis technology is used to mine psychological information; the associations and value information rules between different psychological assessment data are mined from the massive psychological data of students. The classification model of mental illness is constructed, and the application data of students’ psychological management system in practice is verified by mining technology, the problem of finding information from massive data is solved, and methods and improvement suggestions for the construction of students’ psychological files are put forward [20]. The data is prepared in three steps: data selection, information preprocessing, and data transformation.

**4.1.1. Data Selection.** Data selection is based on a precise task direction and a particular understanding of the data itself. The selection stage requires the integration of the task

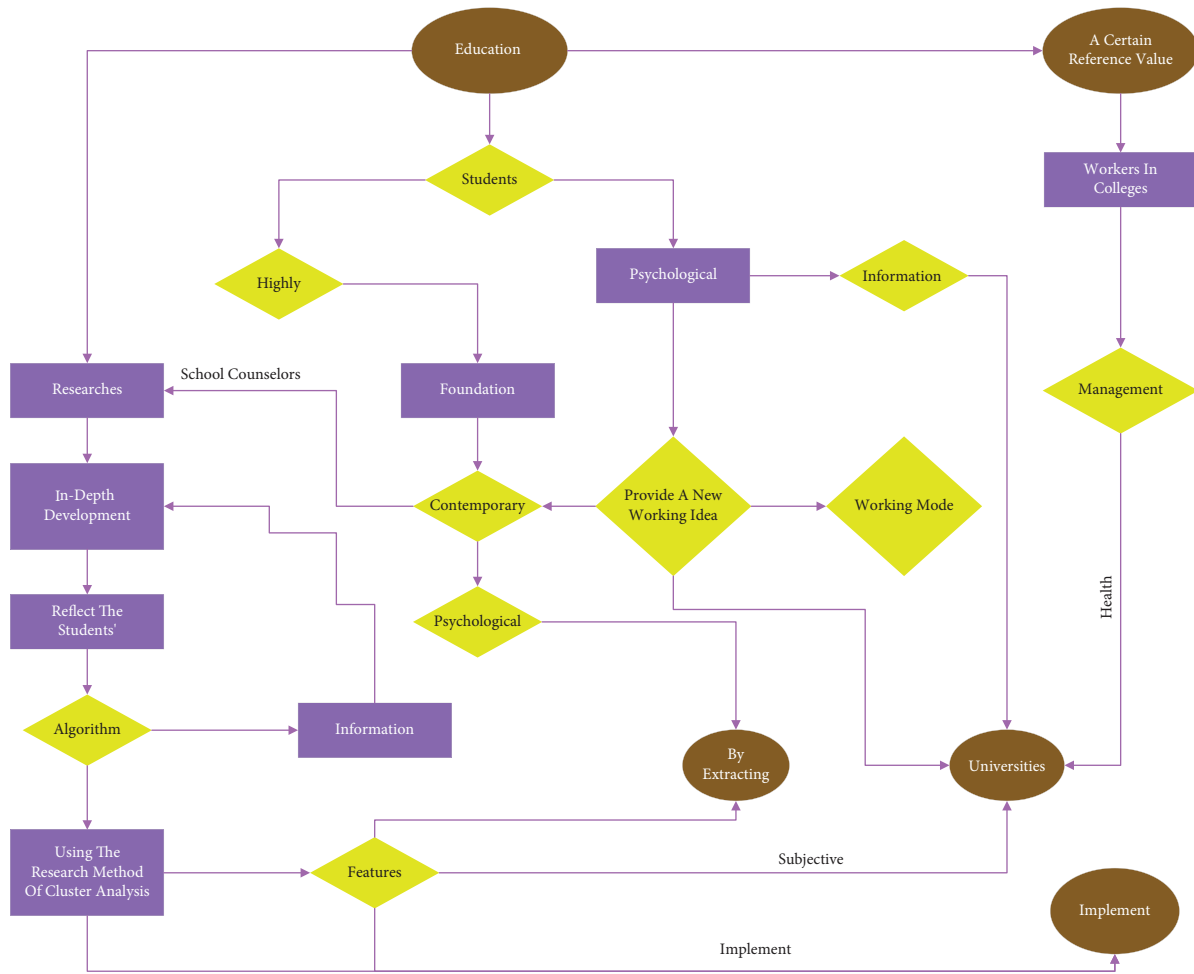


FIGURE 3: Multi-authority user management structure diagram.

object attributes. The features that are less relevant to the mining task and tend to increase the complexity are eliminated, such as student number and name, to reduce the load brought by the algorithm and to achieve the purpose of improving the system’s robustness. Finally, the remaining attribute elements are integrated. The factors that affect the students’ psychology are shown in Figure 4.

**4.1.2. Data Preprocessing.** In the mental health system, for various problems in the data set (such as normality, dichotomy, and repetitiveness), the stability of the system in the following mining analysis is further strengthened by preprocessing student data such as noise reduction, missing values, and de-weighting, which also makes the system reduce the secondary processing time of the data at a later stage, that is, the accuracy of the results is strengthened, and the robustness of the system is also enhanced. The data analysis chart of student psychological education is shown in Figure 5.

**4.1.3. Data Conversion.** Since numbers and letters tend to be more sensitive than characters in the system algorithm, some

of the data are converted after they are obtained, and the conversion is only based on the values required by the user or the system itself to define the system, which has reached the purpose of optimizing the system. In mental health systems, using the size of the ability value indicates the height of its ability in that area.

**4.1.4. Establishing Database and Corresponding Charts.** In the database management of the mental health system, the relevant data are obtained by noise reduction processing of the valid data and merging the attributes of the data created by the database, as shown in Figure 6.

The data were applied to the generated clustering model. Its accuracy was validated, the status of the remaining data with forced symptoms was known, and the classification prediction of the test set using the initially generated clustering algorithm was 83% when comparing the already existing categories with the predicted classification results; the anticipated classification results of the test dataset using the pruned clusters were compared with the known types with an accuracy of 84.6%. The accuracy of the clustering algorithm model without pruning was lower than that of the pruned model. Therefore, the

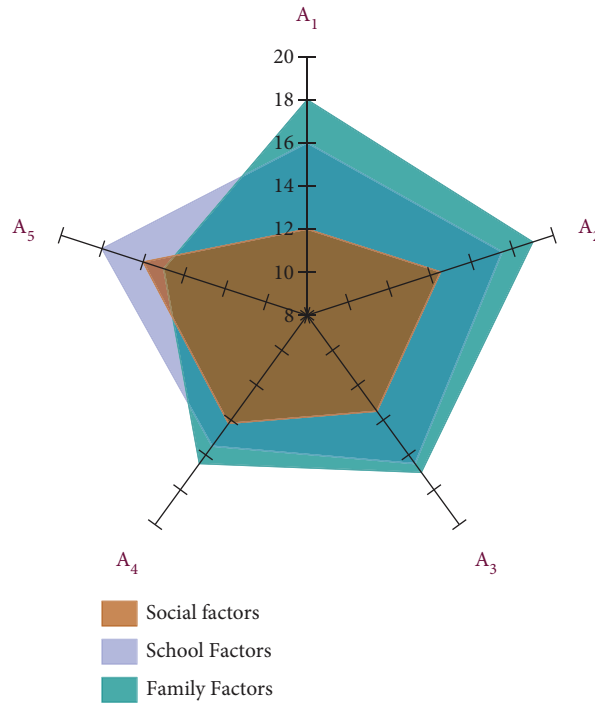


FIGURE 4: Factors affecting students' psychology.

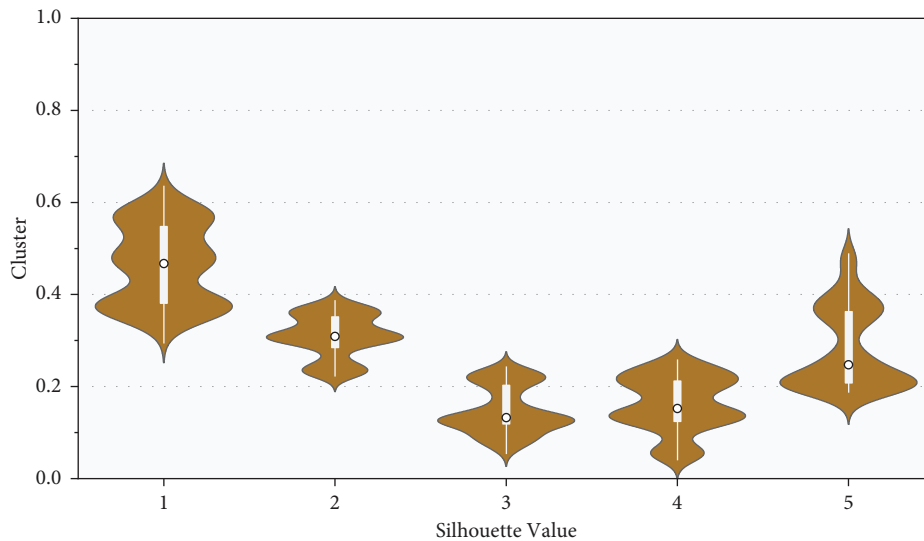


FIGURE 5: Analysis of student psychoeducational data.

classification mining of psychometric data using the ID3 clustering algorithm to construct a decision tree and pruning based on the PEP algorithm can be helpful for psychological prevention and intervention. The PEP algorithm is proposed in the C4.5 decision tree algorithm, where a subtree (with multiple leaf nodes) is replaced by a leaf node (which seems to be replaced by the root of the subtree after I studied many articles); it does not require a separate test data set than the REP pruning method, it does not require a different test data set.

4.2. Analysis of College Students' Mental Health Education Realized. DeletedIn order to better verify the effectiveness of the improved k-means clustering algorithm, in addition to the two small simulated datasets and nine small real datasets, this paper also conducts comparative experiments on three large real datasets. The dataset used in the experiments is the page-blocks dataset, which is a set of datasets for classifying page modules, the dataset contains a total of 5235 data objects, each data object contains 10 different attribute values, which are recorded as the length, width, area, and



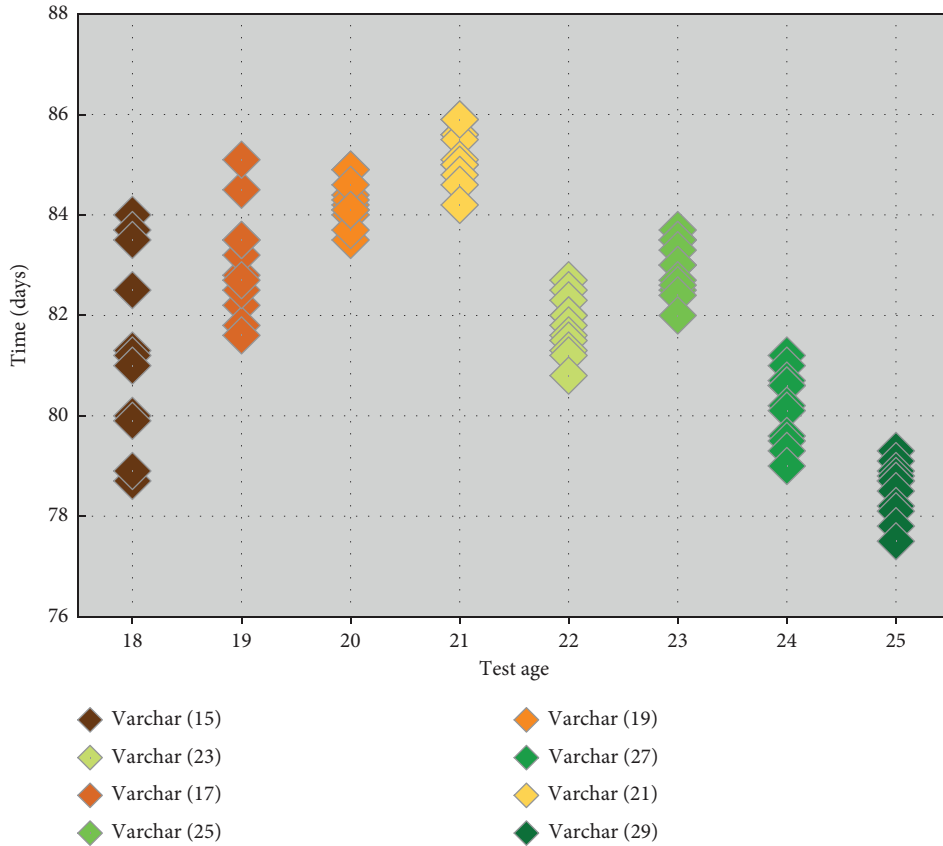


FIGURE 6: Create the database with related data.

other attribute values of page modules, the dataset is divided into four ( $KOPT = 4$ ). When performing the clustering validity evaluation task on the page-blocks dataset, only the CSI index can obtain the best number of cluster divisions. In contrast, all cluster validity indexes cannot get the best cluster divisions, and the equation is as follows:

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_1 - x_j)^2}. \quad (7)$$

In this paper, experiments were conducted to compare 14 different types of datasets, including petite simulated, small actual, and significant honest, with five standard cluster validity indicators in 3 different types of datasets. The experimental results are shown in Figure 7. The experimental results show that the newly proposed CSI indicator can get the best clustering number of 22 data sets of 3 different types; among the other five clustering validity indicators in the comparison experiment, the COP indicator and the I indicator can get the best clustering number of 20 data sets, DBI indicator can get the best clustering number of 21 data sets, DI indicator can get the best clustering number of 3 data sets, and CSP indicator can get the best clustering number of 3 data sets. The CSP indicator can obtain the best clusters for 3 of 11 small data sets.

The density-based method is first used in the clustering analysis to select the initial centers, followed by combining

the average intra-class and inter-class distances to form the criterion function. The optimal  $k$ -value is chosen according to the minimum value of the criterion function. Finally, the Iris, Wine, and Glass datasets in the UCI database were used to verify the improved algorithm's effectiveness. The results showed that the improved algorithm in this paper improved the accuracy and stability of the clustering effect. User studies, desktop studies, and fieldwork were conducted for college student mental health management. Quantitative research and qualitative research were undertaken mainly for college students. Interview research was undertaken primarily for the parents of college students. Desktop research, field study, and interview research were conducted primarily in the institutions and their personnel related to college mental health management; interview research was undertaken primarily by psychology experts. Through the above analysis, we analyzed the characteristics of college students' thoughts and behaviors on mental health management, established a typical user model of college students' group, sorted out the needs related to college students' mental health, summarized and analyzed the ways of college students' mental health management and the critical and challenging points of college students' mental health management at the present stage and thus analyzed the feasibility and opportunity points of the clustering algorithm design to intervene. To provide a reference for later research on the mental health management service model of college students

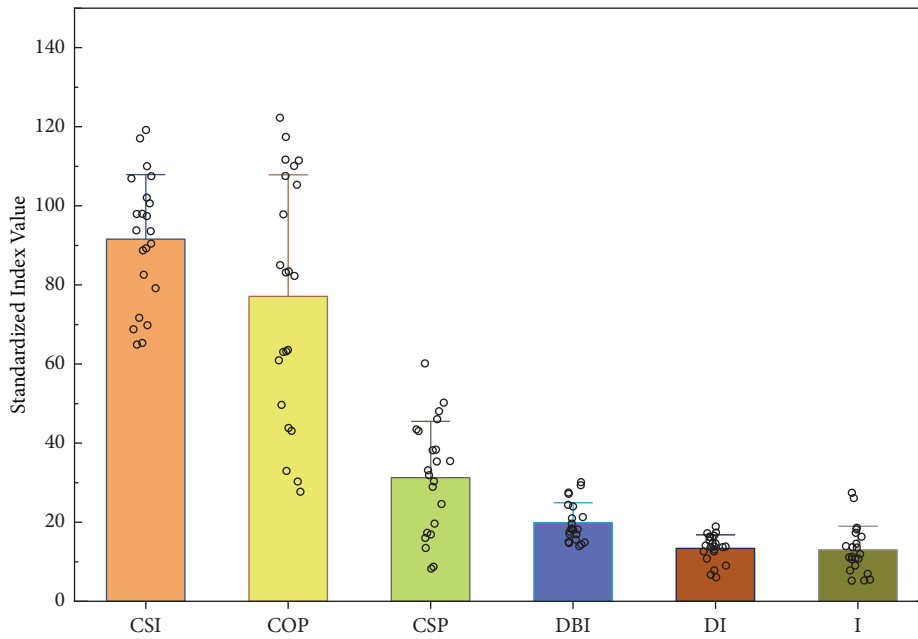


FIGURE 7: Experimental results of the clustering validity index.

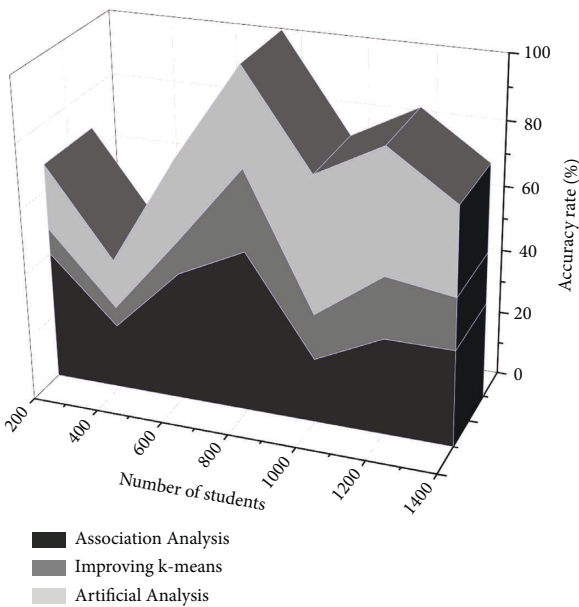


FIGURE 8: Comparison of students' psychological assessment results.

comparison chart of students' psychological assessment results is shown in Figure 8.

The unified management of college students, mental health status, weekend reviews, classes, class teachers, and other information is realized. The organic integration of these functions makes it sufficient to meet the needs of most colleges and universities for student management. The information on the top of the system puts students' mental health issues in a more critical position for colleges and universities to pay attention to the mental health issues of college students.

### 5. Conclusion

With the advent of the digital era, the Internet is gradually changing people's lifestyles, providing new opportunities and challenges for mental health education. Good mental health is essential for the overall development of college students. The author hopes to understand further the study of the problems of college student's mental health, find countermeasures and ways to improve college student's mental health, bring specific theoretical significance and practical value to the current academic and empirical research on college student's mental health and raise the wide attention of society to college students' mental health education. This study aims to innovate the college student's mental health model through the theory and method related to the improved *k*-means clustering algorithm. This study analyzes the intervention points of the enhanced *k*-means clustering algorithm design for mental health education through the preliminary theoretical research related to the theory of mental health education and the creation of the clustering algorithm; through qualitative research and quantitative research on college students' groups, as well as the research on school mental health centers, students' parents and psychology experts, it summarizes the characteristics of college students' groups, analyzes the health education of college student's needs, sorted out the ways of mental health education for college students and transformed their needs by the clustering algorithm. Finally, the content of the clustering algorithm design system for college students' mental health education is improved and realized. [19].

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares no conflicts of interest.

## Acknowledgments

This work was supported by the School of Business, Guangzhou College of Technology and Business.

## References

- [1] N. C. Jao, L. D. Robinson, P. J. Kelly, C. C. Ciecierski, and B. Hitsman, "Unhealthy behavior clustering and mental health status in United States college students," *Journal of American College Health*, vol. 67, no. 8, pp. 790–800, 2019.
- [2] M. Qin and Q. Gui, "Intervention algorithm of horse racing for students' psychological disorders based on big data," *Mobile Information Systems*, vol. 2022, Article ID 3513240, 17 pages, 2022.
- [3] J. E. Halladay, J. L. Dawdy, I. F. McNamara et al., "Mindfulness for the mental health and well-being of post-secondary students: a systematic review and meta-analysis," *Mindfulness*, vol. 10, no. 3, pp. 397–414, 2019.
- [4] Z. X. Liu, D. G. Zhang, G. Z. Luo, M. Lian, and B. Liu, "A new method of emotional analysis based on CNN-BiLSTM hybrid neural network," *Cluster Computing*, vol. 23, no. 4, pp. 2901–2913, 2020.
- [5] D. Talevi, V. Socci, M. Carai et al., *Rivista di psichiatria*, vol. 55, no. 3, pp. 137–144, 2020.
- [6] L. Brookman-Fraze, S. Roesch, C. Chlebowski, M. Baker-Ericzen, and W. Ganger, "Effectiveness of training therapists to deliver an individualized mental health intervention for children with ASD in publicly funded mental health services: a cluster randomized clinical trial," *JAMA Psychiatry*, vol. 76, no. 6, pp. 574–583, 2019.
- [7] M. Gallé-Tesson, D. Bach Johnsen, and G. Keppens, "The relationship between mental health and school absenteeism in a community sample of French secondary school students: four profiles derived from cluster analysis," *European Journal of Education and Psychology*, vol. 12, no. 1, pp. 77–90, 2019.
- [8] S. Daenekindt and J. Huisman, "Mapping the scattered field of research on higher education. A correlated topic model of 17,000 articles, 1991–2018," *Higher Education*, vol. 80, no. 3, pp. 571–587, 2020.
- [9] C. Li, Q. Wu, and Z. Liang, "Effect of poverty on mental health of children in rural China: the mediating role of social capital," *Applied Research in Quality of Life*, vol. 14, no. 1, pp. 131–153, 2019.
- [10] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychological Medicine*, vol. 49, no. 09, pp. 1426–1448, 2019.
- [11] A. Rezaei and E. Mousanezhad Jeddi, "Relationship between wisdom, perceived control of internal states, perceived stress, social intelligence, information processing styles and life satisfaction among college students," *Current Psychology*, vol. 39, no. 3, pp. 927–933, 2020.
- [12] F. J. Eiroa-Orosa and H. Garcia-Mieres, "A systematic review and meta-analysis of recovery educational interventions for mental health professionals," *Administration and Policy in Mental Health and Mental Health Services Research*, vol. 46, no. 6, pp. 724–752, 2019.
- [13] Y. Yu, C. Wan, E. S. Huebner, X. Zhao, W. Zeng, and L. Shang, "Psychometric properties of the symptom check list 90 (SCL-90) for Chinese undergraduate students," *Journal of Mental Health*, vol. 28, no. 2, pp. 213–219, 2019.
- [14] A. M. Thompson, A. J. Frey, and M. S. Kelly, "Factors influencing school social work practice: a latent profile Analysis," *School Mental Health*, vol. 11, no. 1, pp. 129–140, 2019.
- [15] K. Streimann, A. Selart, and A. Trummel, "Effectiveness of a universal, classroom-based preventive intervention (PAX GBG) in Estonia: a cluster-randomized controlled trial," *Prevention Science*, vol. 21, no. 2, pp. 234–244, 2020.
- [16] Y. Y. Lien, H. S. Lin, Y. J. Lien et al., "Challenging mental illness stigma in healthcare professionals and students: a systematic review and network meta-analysis," *Psychology & Health*, vol. 36, no. 6, pp. 669–684, 2021.
- [17] X. Zhou, C. L. Snoswell, L. E. Harding et al., "The role in reducing the mental health burden from COVID-19," *Telemedicine and E-Health*, vol. 26, no. 4, pp. 377–379, 2020.
- [18] S. Yamaguchi, J. C. Foo, A. Nishida, S. Ogawa, F. Togo, and T. Sasaki, "Mental health literacy programs for school teachers: a systematic review and narrative synthesis," *Early Intervention in Psychiatry*, vol. 14, no. 1, pp. 14–25, 2020.
- [19] A. M. Rogowska, "The relationship between demographic Variables and substance use in undergraduates," *International Journal of Mental Health and Addiction*, vol. 17, no. 6, pp. 1550–1563, 2019.
- [20] A. Tanhan and R. W. Strack, "Online photovoice to explore and advocate for Muslim biopsychosocial spiritual wellbeing and issues: e," *Current Psychology*, vol. 39, no. 6, pp. 2010–2025, 2020.