

Retraction

Retracted: Data Analysis of College Students' Mental Health Based on Clustering Analysis Algorithm

Complexity

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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) Manipulated or compromised peer review

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.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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] Y. Chu and X. Yin, "Data Analysis of College Students' Mental Health Based on Clustering Analysis Algorithm," *Complexity*, vol. 2021, Article ID 9996146, 10 pages, 2021.

Research Article

Data Analysis of College Students' Mental Health Based on Clustering Analysis Algorithm

Yichen Chu¹ and Xiaojian Yin²

¹Department of Mathematics, Qilu Normal University, Jinan 250200, China

²Student Affairs Office, Qilu Normal University, Jinan 250200, China

Correspondence should be addressed to Yichen Chu; 20157029@qlnu.edu.cn

Received 6 March 2021; Revised 29 March 2021; Accepted 2 April 2021; Published 13 April 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Yichen Chu and Xiaojian Yin. 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.

Mental health is an important basic condition for college students to become adults. Educators gradually attach importance to strengthening the mental health education of college students. This paper makes a detailed analysis and research on college students' mental health, expounds the development and application of clustering analysis algorithm, applies the distance formula and clustering criterion function commonly used in clustering analysis, and makes a specific description of some classic algorithms of clustering analysis. Based on expounding the advantages and disadvantages of fast-clustering analysis algorithm and hierarchical clustering analysis algorithm, this paper introduces the concept of the two-step clustering algorithm, discusses the algorithm flow of clustering model in detail, and gives the algorithm flow chart. The main work of this paper is to analyze the clustering algorithm of students' mental health database formed by mental health assessment tool test, establish a data mining model, mine the database, analyze the state characteristics of different college students' mental health, and provide corresponding solutions. In order to meet the needs of the psychological management system based on the clustering analysis method, the clustering analysis algorithm is used to cluster the data. Based on the original database, this paper establishes the methods of selecting, cleaning, and transforming the data of students' psychological archives. Finally, it expounds on the application of data mining in students' psychological management system and summarizes and prospects the implementation of the system.

1. Introduction

Nowadays, the world is in an era of fierce competition. The so-called competition is essentially the competition of talents. Education shoulders the important task of cultivating high-quality talents for the 21st century [1]. As an important position to train builders and successors of the socialist cause, colleges and universities should comprehensively promote quality education, which is an important goal of their work, and mental health education is an important foundation and condition to comprehensively improve the overall quality of students. In recent years, the mental health of college students has been the general concern of the whole society, and the in-depth study of college students' mental health and the exploration of college students' psychological intervention mode are the common focus of domestic and

foreign scholars [2]. Local education departments and colleges and universities have done a lot of work in carrying out college students' mental health education. Some colleges and universities have gradually incorporated this work into the school moral education system, established special institutions for mental health education, psychological counseling, or consultation, and carried out corresponding teaching, education, scientific research, and practical activities [3]. Many health care institutions in colleges and universities have done a lot of work in carrying out mental health education, psychological or counseling for college students. However, the current situation of this work in all colleges and universities in China is not balanced, and a lot of mental health education work is not perfect. Due to the psychological barriers, the number of students' suspension, dropout, and other malignant events is increasing [4].

The university mental health database provides important information for the early prevention of students' mental health problems. In order to improve and optimize the decision-making and improve the efficiency of psychological counseling for college students, it is necessary to analyze this information timely and accurately. As an unsupervised data mining technology, cluster analysis has a broad application market. The rational use of related technologies will provide a scientific reference for the actual decision-making activities [5]. The data of the traditional university of mental health only stay in the operation of adding, deleting, modifying, and checking and do not effectively analyze the potential psychological information of the data. This paper puts forward a kind of psychological management system based on the cluster analysis method, which uses the idea of data mining to make secondary use of students' psychological data based on traditional system functions [6]. By optimizing the iterative process of cluster analysis algorithm, the valuable part of many students' psychological data is extracted, and the data model is established to provide decision-making guidance for managers; scientific management of students' mental health process can not only effectively improve the overall efficiency of psychological counseling but also play an early warning role in the prevention of risk factors.

In view of the above problems, this paper uses the cluster analysis algorithm to study. For specific mental health problems, combined with a two-step cluster analysis, we use data mining techniques to find information from these data, to provide the basis for the planning and decision-making of mental health education. The rest of this paper is organized as follows: the second section discusses the related work. The third section analyzes the clustering analysis algorithm and mainly introduces the concept of clustering analysis, similarity measurement method, criterion function, and mental health data model based on clustering analysis algorithm. The fourth section establishes the mental health data model based on the clustering analysis algorithm. The fifth section is based on the mental health data mining of college students, combined with the actual clustering analysis algorithm research, and verifies the performance of the algorithm through implementation. The sixth section summarizes the core content and main work of this paper and analyzes the main achievements and some areas that need to be improved.

2. Related Work

Clustering analysis algorithm is one of the important technologies in the field of data mining. Clustering analysis is the process of dividing the collection of physical or abstract objects into several subcategories composed of similar objects. This process requires data to be grouped according to the distance or similarity of the data itself [7]. The objects in the same cluster have great similarity, and the objects in different clusters have great dissimilarity. Clustering and classification are essentially different; clustering is to classify unknown data attributes, while classification is to classify new samples after analyzing and summarizing data

attributes [8]. At present, there are two kinds of clustering analysis: one is based on the partition, and the other is hierarchical clustering analysis. Because the two discriminant graphs of hierarchical clustering analysis are too scattered and difficult to explain, this paper uses the fast-clustering method to analyze the mental health data of college students. In recent years, cluster analysis algorithm has been widely concerned and developed. There is a lot of progress in the research of cluster analysis algorithms and applications [9].

Relevant scholars use cluster analysis technology to mine the psychological information in the psychological management system, and according to the corresponding characteristics of the psychological diseases existing in the system, they mine the association and value information rules between different psychological evaluation data from the massive students' psychological data, construct the classification model of psychological diseases, and verify that the application of the students' psychological management system by using the mining technology application of data in practice solves the problem of finding information from massive data and puts forward the methods and improvement suggestions for the construction of original students' psychological archives [10]. According to the specific spatial distribution of the data, the two-step clustering analysis algorithm is adopted, and the clustering analysis model is constructed, which provides technical support for the national universities to carry out the work related to the mental health of college students, and has a significant improvement for the mental health education of college students [11, 12].

After analyzing the application of rough set and neural network in psychological measurement, this paper uses rough set to analyze the relevant data, puts forward the fuzzy comprehensive evaluation method based on genetic algorithm, and makes a beneficial analysis on the psychological measurement data of Likert scale. It applies the decision tree mining method to the analysis of college students' mental health data, with the help of clementine 12.0 platform, using C5.0 algorithm to construct decision tree mining model, to study the factors affecting college students' mental health, using a vague set theory to construct core factor set. Applying HMM, the model predicts the psychological crisis of college students, uses the k-means clustering analysis method to analyze the test data of students' psychological health management, studies the feasibility of introducing data mining technology in the analysis of college students' psychological problems, and uses the decision tree algorithm C4.5 to analyze the data of college students' psychological problems. Through the establishment of binary logistic stepwise regression model and decision tree model used to analyze and predict the influencing factors of college students' subhealth, some scholars design and implement college students' psychological data mining model based on cart decision tree, pattern recognition network, and BP artificial neural network algorithm [13]. Using classification algorithms for psychological data mining or correlation analysis, these researches have achieved certain research results in their respective fields [14]. However, due to the complexity of mental health test data, a single classification

algorithm has some shortcomings in classification accuracy and recall rate, and the stability is not good in different data sets.

According to combinatorial optimization theory, scholars choose a decision tree as the basic classifier and construct a powerful classifier by using the ensemble learning algorithm AdaBoost and apply it to college students' mental health data mining [15]. AdaBoost is an iterative algorithm based on a weak learning theorem. Its core idea is that for a group of mental data training sample sets, different training sets can be obtained by changing the distribution probability of each sample [16]. For each training set, a basic classifier can be obtained by training, and then these basic classifiers can be combined according to different weights and get a strong classifier. With the increase of the number of iterations, the upper bound of the training error rate will gradually decrease. At the same time, it can avoid the overfitting problem in other algorithms, to effectively improve the classification accuracy of the classifier.

3. Clustering Analysis Algorithm

3.1. Basic Concepts of Clustering Analysis Algorithm. Clustering is to group data objects into multiple clusters, so that the objects in the same cluster have high similarity, while the objects in different clusters have great differences [17]. A good clustering method should produce clustering results with the following characteristics: the objects in the cluster are highly similar, while the objects between the clusters are rarely similar. Clustering analysis is a very important research topic in the field of data mining, which is different from classification. The goal of clustering is to aggregate data into different clusters according to the similarity of data without any prior knowledge so that the elements in the same cluster are as similar as possible, and the elements in different clusters are as different as possible. Therefore, it is also called unsupervised classification [18]. As a module of the data mining system, clustering analysis can be used as a separate tool to discover the deep information of data distribution in the database and can also be used as a preprocessing step of other data mining analysis algorithms.

The cluster analysis model can be described as follows: given n vectors in m -dimensional space R^m , each vector is assigned to one of the k clusters, so that the "distance" between each vector and its cluster center is the minimum. The essence of the cluster analysis problem is a globally optimal problem. Among them, m is the number of attributes of clustering samples, n is the number of samples, and k is the number of categories present by users. The mathematical model is as follows: for vectors $x_i, x_j, x_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}, \gamma_i = \{\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{in}\}$ in m -dimensional space R^m , Mahala Nobis distance has many advantages; it is not affected by dimension. Mahala Nobis's distance between two points has nothing to do with the measurement unit of original data. Mahala Nobis distance between two points calculated by standardized data and centralized data is the same.

3.2. Criterion Function. The criterion function of the clustering analysis algorithm can be understood as the constraint condition of the end of the algorithm. When the final classification result of the algorithm meets the criterion function, the algorithm exits the cycle [19]. The selection of criterion function has a direct impact on the quality of the clustering algorithm, so in order to get a better clustering effect, we must choose the appropriate criterion function. In order to measure the quality of clustering, the sum of the square error criterion function is used.

$$T_c = \frac{\sum_{\alpha=1}^n \|x^i - \gamma_i\|}{x^i - \gamma_i}, \quad (1)$$

$$\gamma_i = \frac{1}{n_i (\sum_{i=1}^n x_i)} - c,$$

where γ_i represents the mean value of the samples in the j -the class and n represents the number of samples in the class. According to the definition of the above criterion function, it is not difficult to find its value C cluster centers and samples in each cluster. The larger the value of T_c , the larger the clustering error and the worse the clustering effect [20]. On the contrary, the higher the value of T_c , the better the clustering effect. The sum of squared error criterion function is suitable for all kinds of sample distribution with dense samples and little difference in the number of samples. When the number of different types of samples varies greatly, the sum of squares criterion may sometimes be used to separate the types with more samples.

$$Q = \sum_{i=1}^c \alpha_i \beta_i + \sum_{i=1}^n w_i^*, w_i^* = \frac{1}{n_i (n_i - 1)} \cdot \sum_{x \in x_i} \|x - c\|, \quad (2)$$

where w_i^* is the mean square distance between samples within a class. For a large number of samples, it is easier to get the correct clustering results by using the weighted mean square distance sum criterion than by using the error square sum criterion, which can prevent more classes from splitting.

$$R = \frac{\sum_{i=1}^n (v_i - \tau_i)}{2} \cdot \|v_i - \tau_i\|^2, \quad (3)$$

where v_i is the sample mean vector of τ_i type and R is the mean vector of all samples. Obviously, the larger the distance between clusters and the value of criterion function, the better the separation of each type of clustering results, so the clustering quality is high. We can use the distance between clusters and the criterion to describe the distribution of the distance between clusters.

3.3. Steps of Cluster Analysis. There are three steps in clustering algorithm: feature extraction, algorithm selection, and parameter setting.

- (1) *Feature Extraction.* Attribute selection is very important, and the essence of the clustering algorithm is to divide the samples with similar characteristics into a class according to some characteristics of the

sample itself [21]. The distance between classes is mainly determined by the sample itself. Therefore, in order to make the clustering results better reflect the characteristics of the sample itself, we must first select the characteristics of the samples waiting for clustering. Some attributes closely related to the sample itself are selected as an attribute item waiting for clustering [22]. However, those classes that have nothing to do with the clustering results of samples cannot be used as a reference, which can ensure the correctness of the clustering results. Reasonable feature selection should make the clustering of similar samples smaller and the distance between different samples larger. If we choose features that are irrelevant to clustering requirements at the beginning, it will not only affect the complexity of the algorithm itself but also lead to the failure of clustering results. Sometimes, we choose too many attributes, which is not conducive to the correctness of clustering. At this time, we must reduce the dimension. Secondly, standardization is also very important. The data types may be different because the attributes we select have different meanings. Some may be numeric, some may be text, and some may be other representations [23]. Even data of the same type may have different units; in order to eliminate the incommensurability caused by different dimensions and dimensional units, the feature indexes should be standardized before clustering. These characteristics must be normalized to a dimensionless interval according to a certain utility function. At present, the utility function usually adopts the $[0, 1]$ interval method, and the standardized data is generally a standard sample matrix X , which is an $n * D$ matrix, that is, n samples, each sample contains d -dimensional characteristics, and the elements in the matrix are data between $[0, 1]$.

- (2) *The Choice of Algorithm.* The choice of algorithm will directly affect the result of clustering. According to the characteristics of the sample itself and the needs of clustering, we can choose different clustering methods. This also requires us to have a deeper understanding of many clustering methods and choose the method suitable for us [24].
- (3) *Parameter Setting.* Researchers rely on experience and domain knowledge; according to the specific application, they determine the selection of parameters. Parameter setting is often a cumbersome process because sometimes we must experiment constantly to see what kind of parameter setting can get a better clustering effect, and without the participation of domain experts, only relying on the algorithm itself, generally, cannot get satisfactory results, so we need to have a certain understanding of the field of clustering samples [25–27].

Firstly, initial clustering centers are selected according to certain principles or randomly, and then, the distance between all sample data and centers is calculated. Euclidean

distance is usually used to judge the distance between each cluster and the initial cluster center, and then, the samples are divided into k classes according to the distance, and the average value of the new class is calculated. The average value is used as the new cluster center, and the Euclidean distance is used for clustering again. Until the new cluster center does not change, it is considered that the cluster center has been obtained. Otherwise, the average value is used as the new clustering center and the above process is repeated. The basic flow chart of the clustering algorithm is shown in Figure 1.

3.4. Simulation and Fitting of Radial Basis Function Neural Network. When RBF neural network is used to approximate the noncurrent system, the form of nonlinear function is not very important to the performance of the network, and the number of hidden layer cells determines the process to which the RBF network will fit the training set data. Too many hidden layer neurons will reduce the generalization ability of the network, resulting in overfitting; if there are too few neurons and too large errors of the training set, the fitting effect is not good. Therefore, the key of RBF network modeling is the selection of hidden layer parameters, determining the connection weight between the hidden layer and output layer.

4. Mental Health Data Model Based on Clustering Analysis Algorithm

There are many factors affecting college students' mental health, and there is no direct correlation between these factors, so it is not suitable to classify the data directly, while cluster analysis has good adaptability. There is no unified standard to judge whether a college student is psychologically healthy. With the continuous change of external factors, the psychological state of college students is also constantly changing. The characteristics of each sample change with the change of time, environment, and other factors. Cluster analysis provides a fuzzy analysis method, which gathers some similar attributes to highlight the characteristics of such attributes, and can achieve an active and effective defined mechanism to a certain extent. Firstly, the attributes of students' mental health-related factors are subdivided, and then, the objective and rational judgment of students' mental health is made through cluster analysis, to establish a management mechanism with a practical reference value.

In order to realize the intelligent analysis of mental health data and effective prediction of mental health, a mental health prediction model is established. Firstly, the model reads in the training sample data, cleans and normalizes the data, and then gives initial weights to all samples; secondly, taking the vector set D composed of all sample weights as the parameter, it uses a clustering algorithm to train the sample set and obtains a group of basic classifiers; finally, it uses AdaBoost algorithm to calculate the weighted error rate of the basic classifiers and calculates to get the weight of the classifier. The weight of the classifier is used as the parameter to adjust the sample weight distribution D ,

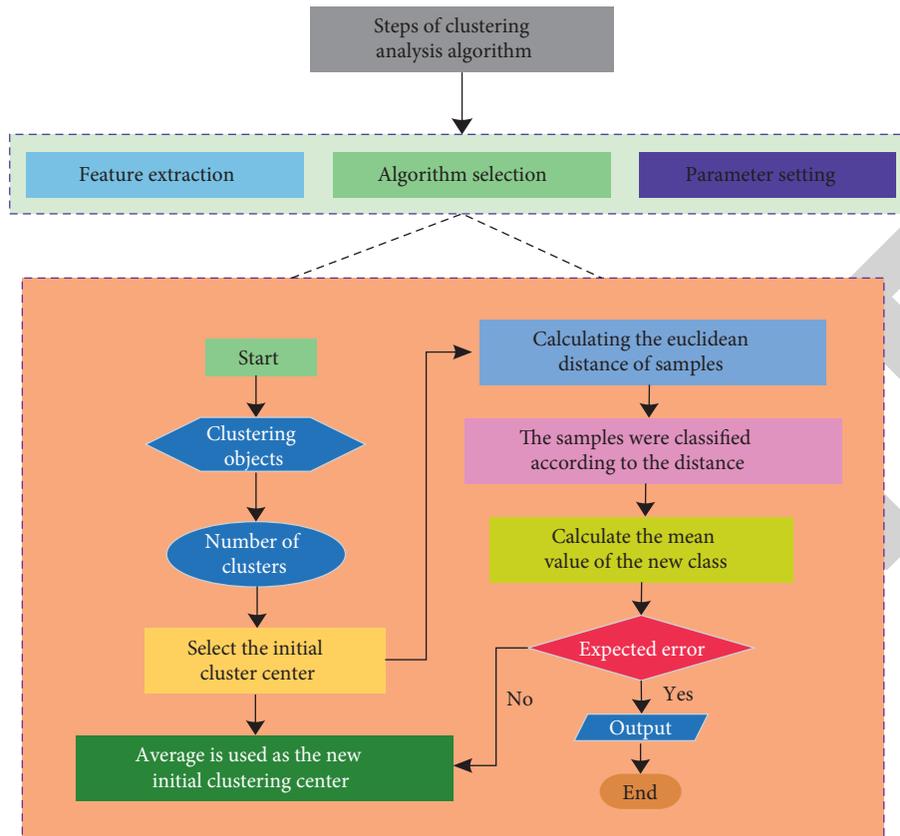


FIGURE 1: Flow chart of clustering analysis algorithm.

and the division and weight of each basic classifier are obtained after several iterations of training; these basic classifiers are linearly combined according to the AdaBoost algorithm to get the final strong classifier; the classifier is used to process the mental health data to be tested and recognize that the characteristics of mental health and effective prediction were given. The mental health prediction model is shown in Figure 2.

There are certain objective conditions for the occurrence of any psychological problems. A single objective factor itself or a combination of multiple objective factors can lead to mental health. It is precisely because the objective factors leading to mental health can be decomposed into a single factor or a combination of single factors, so it is suitable to use the two-step clustering algorithm of cluster analysis so that a series of potential attributes can be mined, which need to rely on the data of the psychological survey. To use the two-step clustering algorithm for analysis, it is necessary to first clear up the data, sort out the scores of 16 personalities in the 16PF questionnaire and the data of UPI table, then carry out the second-order factor analysis through the formula, generate the comprehensive data table of mental health, and establish the corresponding accident database and table. Finally, the relationship between the four second-order public factors and mental health is derived. The implementation process of college students' mental health analysis is first to collect the data of college students' health. After the selection and cleaning of data attributes, the useful

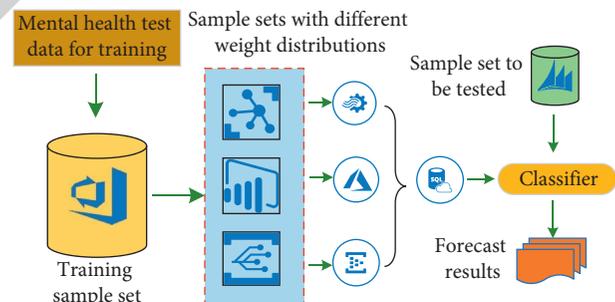


FIGURE 2: Prediction model of mental health.

data is established into a comprehensive psychological test database. The database is analyzed by using a two-step clustering algorithm. Through the analysis of clustering results, according to the prediction and analysis of college students' mental health factors, the results are summarized, that is, the evaluation of college students' mental health. The data mining flow chart of college students' mental health analysis is shown in Figure 3.

5. Experiment and Result Analysis

According to the designed college students' mental health management system and after one month of testing, it is found that there are many factors affecting college students' mental health in the specific database, and there is no direct

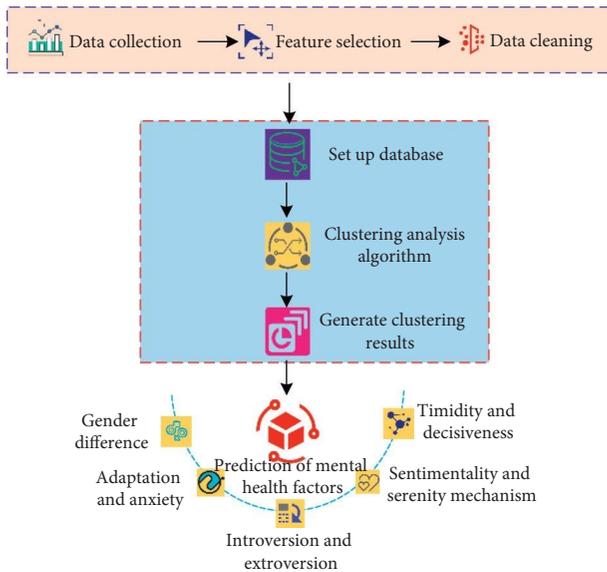


FIGURE 3: Data mining analysis chart of college students' mental health.

correlation between these factors, so it is not suitable to directly classify the data, while cluster analysis has good adaptability. With the continuous changes of external factors, the psychological state of college students is also constantly changing. The characteristics of each sample change with the changes of time, environment, and other factors. Cluster analysis provides a fuzzy analysis method, which gathers some similar attributes to highlight the characteristics of these attributes, and can achieve an active and effective defined mechanism to a certain extent. Firstly, it subdivides the attributes of the related data in the database table of the student mental health management system and then makes an objective and rational judgment on the student's mental health through the data analysis function of the system, to establish a management mechanism with a practical reference value.

5.1. Data Preprocessing. In order to achieve the validity of clustering analysis, the following preprocessing is needed for the original data. The purpose is as follows. (1) To ensure the validity of the data: ensure that the collected data is related to the content attributes of the research, and avoid the loss of relevant data and the destruction of incomplete data. (2) Remove the data noise: remove some inaccurate data or "outlier" data. (3) Unified data scale: the data is quantified to facilitate the operation of the clustering algorithm. In the preprocessing process, this paper mainly uses data filtering, transformation, protocol, and other methods to preprocess the original data.

According to the attribute code conversion table, the data set of the sample is {21, 31, 42, 51, 61, 71, 82}, which guarantees the rapid processing of data. Consider that student management is generally managed separately by departments and grades, and each grade is generally managed by a full-time counselor. Because similar majors have

the same curriculum and management mode, the analysis of the characteristics of students' related activities in the same department and grade is helpful for counselors and colleges to provide a valuable reference in the process of student management, so a counselor's institute is selected. It is reasonable and representative to cluster the data of students.

The data specification includes the following contents. (1) Conversion variables: for example, the gender of students, male (female), is converted into code. (2) Calculation variables: 16 personality values of 16PF were used to calculate the values of adaptation and anxiety, introversion and extroversion, emotion and serenity, timidity and boldness, mental health, professional achievement, creativity, and growth ability. The establishment of a comprehensive psychological test database and the table is as follows: due to the fact that the efficient application of data mining technology is a large amount of data and safety database management system, through the selection of effective data, remove the noise data and some irrelevant data in the original data, merge the attributes that can be used to establish the database, and establish the corresponding data table.

Data preprocessing found that individual or single-parent families, in all aspects compared with other students, have a certain gap, leading to their inferiority complex and lack of self-confidence; there are difficulties where they often suffer their own psychological problems. Moreover, they are not willing to take the initiative to communicate with teachers and classmates, which makes it difficult to find their psychological problems in the early stage. Although students of this kind study step by step, they do not have much communication with teachers and classmates, the learning methods are not right, and the learning efficiency is not high, which leads to their poor academic performance and aggravates their psychological pressure. In the long-term stimulation of a variety of mental health adverse factors, if they encounter some major setbacks or they were difficult to deal with problems, such students tend to make extreme behavior. Therefore, students of this kind should be the focus of counselors and class teachers. In addition, according to the variable attribute value and sample collection scheme, there are more boys, which makes it more difficult for girls to find the object to talk to when they have psychological problems.

5.2. Analysis of Clustering Results. After data preprocessing, SPSS software is used for clustering, and the data preparation interface is shown in Figure 4. The number of clusters was determined by the classification of SCL-90 and UPI. In this paper, $k=3$ was set, and the initial cluster center was randomly generated by the system. The input variables are "gender," "personality," "family income," "single-parent family," "only child," "grade point," and "attendance." The maximum number of iterations is 10, and the clustering information of each case is displayed. The initial cluster center is shown in Figure 4.

After the initial cluster center is determined, the distance between each data and the initial center is calculated and added to the new cluster; then, a new clustering center is

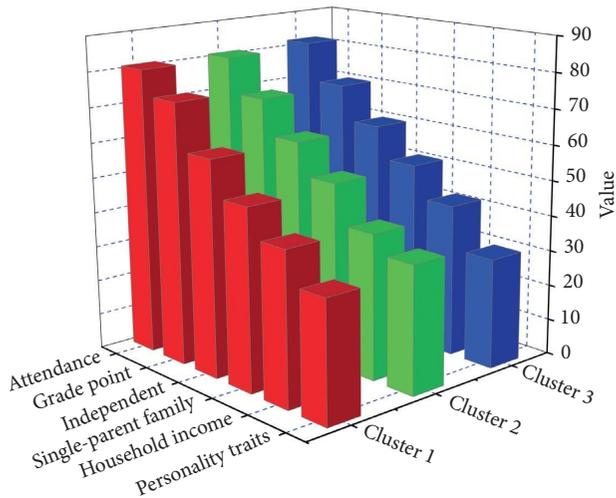


FIGURE 4: Initial cluster center.

generated. After six iterations, the center remains unchanged and the algorithm ends. The classification results and iteration history are shown in Figure 5.

Finally, the final clustering results are output by SPSS software, including 23 people in the first category, 27 people in the second category, and 50 people in the third category. The final clustering center is shown in Figure 6.

From the variable attribute value of the final cluster center, the first group of students are introverted and stubborn, most of the family income is low, some are not only children, and some are single-parent families. There is a certain gap compared with other students in all aspects, which leads to their inferiority, lack of self-confidence, and psychological problems such as the fact that they often bear difficulties silently. Moreover, they are not willing to take the initiative to communicate with teachers and classmates, which makes it difficult to find their psychological problems in the early stage. Although students of this kind study step by step, they do not have much communication with teachers and classmates, the learning methods are not right, and the learning efficiency is not high, which leads to their poor academic performance and aggravates their psychological pressure. In the long-term stimulation of a variety of mental health adverse factors, if they encounter some major setbacks or they are difficult to deal with problems, such students tend to make extreme behavior. Therefore, students of this kind should be the focus of counselors and class teachers. In addition, according to the variable attribute value and sample collection scheme, there are more boys, which makes it more difficult for girls to find the object to talk to when they have psychological problems. Therefore, female students should pay more attention to students of this kind. The second kind of students are extroverted and emotional, family conditions are generally good, academic performance is excellent, and they are often class activists, class cadres, or student union cadres. We should make good use of students of this kind and bring their positive, optimistic, and cheerful mental state to every student in the class. This group is the main group to transfer positive energy to

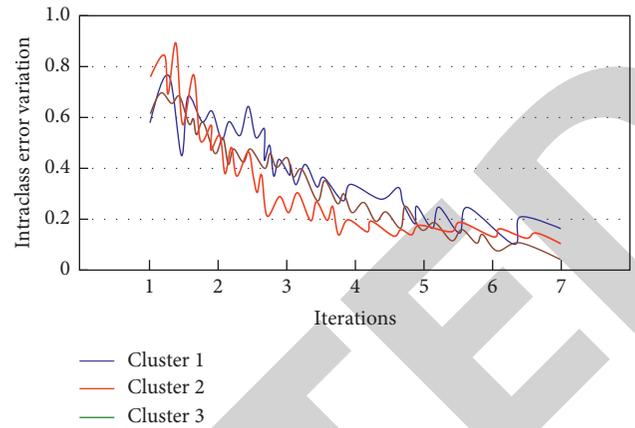


FIGURE 5: Variation of inner class error with iteration times in clustering process.

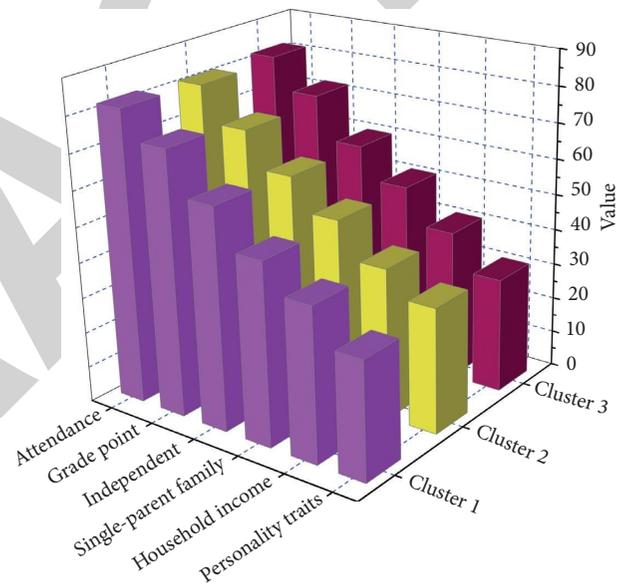


FIGURE 6: The final cluster centers.

the class. Counselors and class teachers should strengthen their guidance. The third group of students is the main group of the school, but from the variable attributes and the specific distribution of sample cases, they can be divided into two types. The first is extroverted students who seldom communicate with others. The second is introverted students but will take the initiative to find someone to talk to the students, indicating that students of this kind have a certain self-regulation ability. This group is relatively stable, but we need to establish an effective management mechanism to understand their psychological state regularly.

5.3. Difference Analysis of Each Variable. In order to show the mean value and standard deviation of each variable in each cluster clearly, we make a PivotTable. Mean represents the mean difference of the corresponding attributes of this cluster, and STD represents the variance. From the distribution of the scores of 16 personality factors, we can

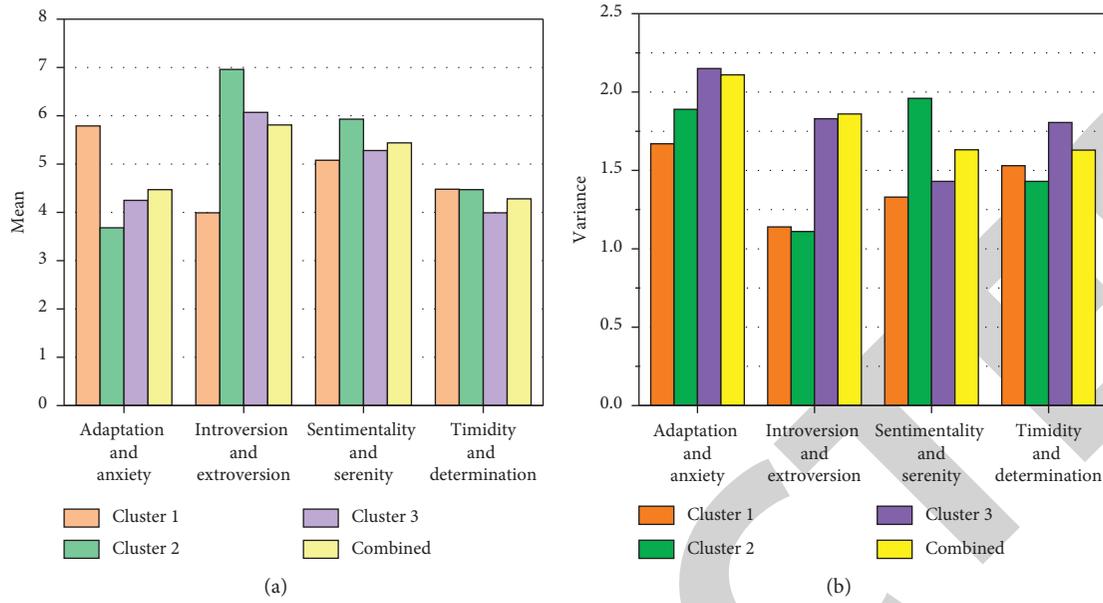


FIGURE 7: Distribution map of mean difference and variance of each variable in clustering. (a) Distribution of mean. (b) Distribution of variance.

basically think that the personality structure of the subjects is harmonious. The standard of each factor of 16PF is 10 points, 1–3 points is low, 4–7 points is medium, and 8–10 points is high. It can be seen from Figure 7 that the average of 16 personality factors of the tested group is between 4 and 7 points, indicating that the personality structure of the group is basically harmonious. Then, according to the information of men and women, the mean difference and variance are calculated, and the results are shown in Figure 7.

It can be seen from Figure 7 that p values are less than 0.01, indicating that there are significant differences between boys and girls in these four factors. From the comparison of male and female students, we can see that the structure of personality factors of female students is significantly better than that of male students, and the personality problems of male students are more prominent, mainly indifference and obstinacy, low trust, suspicion and obstinacy, and poor self-discipline.

Taking the abscissa 0 as the dividing line, the ordinate represents the clustering; the direction of the bar graph to the left indicates that the value of the corresponding variable is lower than the average level; the direction of the bar graph to the right indicates that the value of the corresponding variable is higher than the average level. It shows that the adaptability and anxiety of the third group are much higher than the average level, indicating that the adaptability of this group is relatively low; the anxiety performance is more obvious, usually easy to be excited and anxious, and often dissatisfied with their own situation. High anxiety not only reduces the efficiency of work but also affects the health of the body. The second class is much lower than the average level, which indicates that the students of this class have strong adaptability and feel satisfied. However, those with extremely low scores may lack perseverance, retreat from difficulties, and refuse to work hard, but the first kind is just right, in a relatively good state. The students of the third

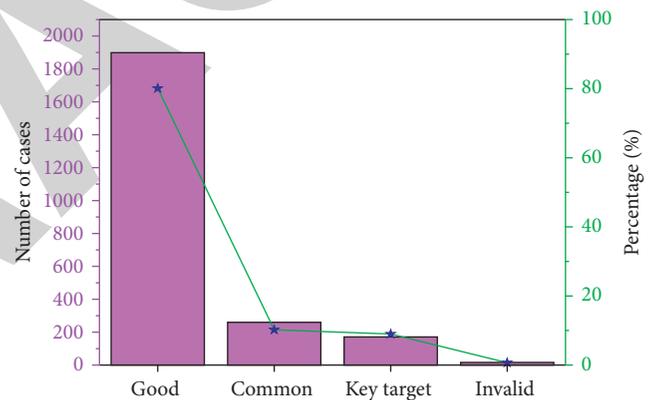


FIGURE 8: Statistical analysis of college students' psychological questionnaire.

group are more introverted, the students of the second group are relatively extroverted, and the students of the first group are the most extroverted. The students of the first class are quiet and alert, enterprising, and positive, while the students of the third class are emotional and calm and alert, which are far lower than the average, indicating that they are emotional, troubled, and often feel frustrated and discouraged. The proportion of timidity and decisiveness in the first category is high, while that in the third category is the lowest.

Compared with the previous attributes (such as introversion and extroversion, calmness, timidity, and boldness), it also plays a certain role. Therefore, it is suggested that counselors should pay attention to Freshmen's adaptability, as well as the influence of personality combined with other three factors, and make statistics on the students who participate in the questionnaire adjustment, and the specific data is shown in Figure 8.

It can be seen from Figures 8 and 9 that our conjecture is like the conclusion of the questionnaire and the result of the

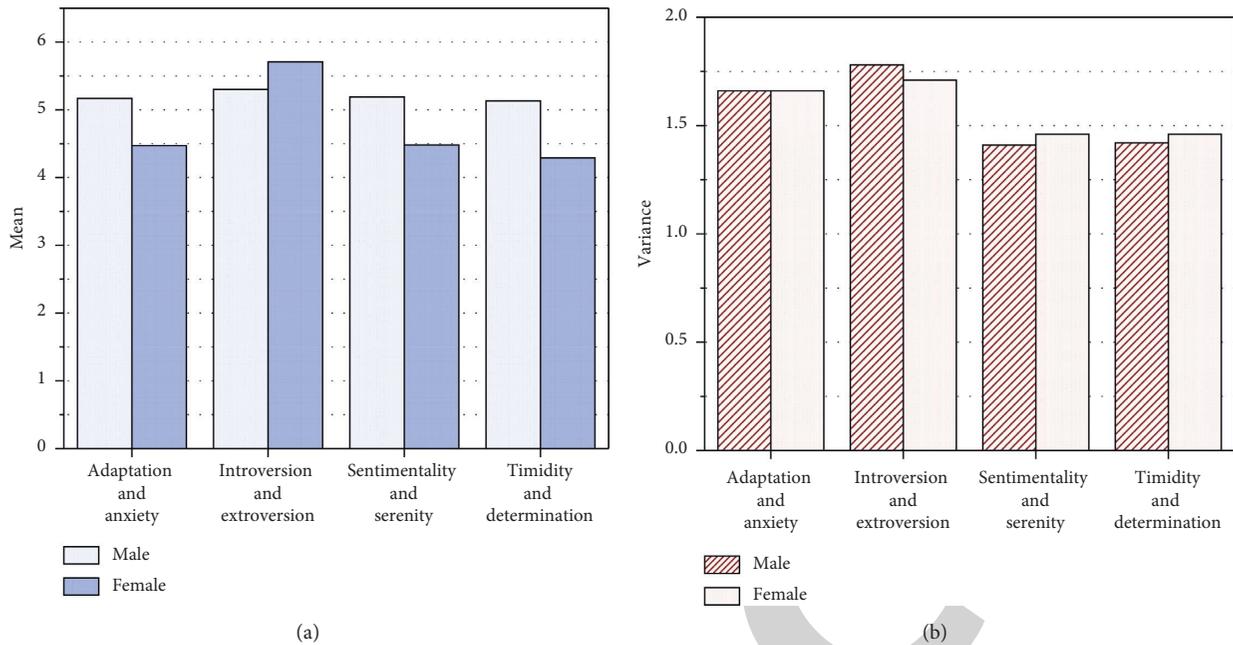


FIGURE 9: Difference chart of each variable between male and female students. (a) Distribution of mean. (b) Distribution of variance.

UPI questionnaire, which proves that the cluster analysis in the system is feasible.

6. Conclusion

Aiming at the problem of passive defines in the existing psychological management system and some college students' mental health early warning mechanism, this paper uses cluster analysis to analyze the students' psychological state actively and obtains more accurate and stable classification results. This method not only helps school counselors and student managers to provide more and better mental health services for students but also provides a certain reference value for mental health educators in colleges and universities. This paper proposes a psychological management system based on the clustering analysis method, which uses the idea of data mining based on the basic functions of the traditional system to make secondary use of student psychological data. Through the iterative process of clustering analysis algorithm optimization, a large amount of precipitation exists. The valuable part of the psychological data of the students is extracted, the data model is established, and the decision-making guidance is provided to the manager. The scientific management of the student's mental health process can not only effectively improve the overall efficiency of psychological counseling but also prevent risk factors. There should be an early warning effect of trouble before it happens. In the follow-up research, we will further improve the clustering method, extract more effective data features, and embed the data mining technology into the student management system, to improve the work efficiency of relevant managers and make up for the limitations of traditional analysis methods, to achieve the

purpose of reflecting the psychological state of students scientifically, reasonably, and quickly. Data mining is a complex process, limited to the relationship between manpower and time. This paper only does a limited discussion on the clustering analysis method in data mining, and there are many areas that need to be improved, which need further in-depth research in the future. The main points are as follows: in the two-step clustering analysis algorithm, we need to further study the influence of parameter settings on the algorithm and grasp the characteristics and rules of parameter setting to make the algorithm get the best performance. For the collection of college students' mental health data, we should further use professional knowledge to explore the influencing factors of mental health, to make the prediction more accurate. In a word, it can be used to diagnose and find the cause of mental health, but it also has some limitations. If we can recognize these limitations and pay attention to avoid the deviation of diagnosis caused by their limitations in practical use, we will be able to provide valuable information in mental health diagnosis and become an effective tool for college students' mental health diagnosis.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Z. T. Xiong, "Mental health prediction method of Higher Vocational Students Based on ensemble learning algorithm," *Software Guide*, vol. 15, no. 6, pp. 162–164, 2016.
- [2] Z. F. An and P. Zhang, "Research on data analysis and feedback system construction of college students' mental health under the background of big data," *Ideological Education Research*, vol. 12, no. 9, pp. 93–96, 2016.
- [3] F. R. Tang, "Application of big data in mental health education in public security colleges," *Journal of Guangxi Police College*, vol. 32, no. 6, pp. 100–104, 2019.
- [4] X. F. Chen, M. Tong, C. Shi et al., "Visual analysis of multi-source college students' mental health questionnaire data," *Journal of Computer Aided Design and Graphics*, vol. 32, no. 2, pp. 181–193, 2020.
- [5] Q. L. Tan and J. G. Zhang, "Discussion on the application of big data in the field of health management," *Health Research*, vol. 40, no. 1, pp. 25–27, 2020.
- [6] F. Zhao and L. Tao, "Analysis of mental health education path of medical college students in the era of big data," *Journal of Jinzhou Medical University*, vol. 18, no. 1, pp. 68–70, 2020.
- [7] P. Feng, R. Zhang, and S. Yang, "Analysis and preventive measures of college students' theft crime based on big data," *Chinese and Foreign Entrepreneurs*, vol. 15, no. 3, pp. 230–231, 2020.
- [8] Y. Zhu and S. Lin, "Construction of college students' psychological early warning system from the perspective of big data," *Coal Higher Education*, vol. 38, no. 1, pp. 60–67, 2020.
- [9] D. Cai, "Research on the innovation of dynamic monitoring system of higher vocational students' mental health based on big data," *Chinese Journal of Multimedia and Network Teaching*, vol. 18, no. 8, pp. 44–46, 2020.
- [10] H. K. Ma, "Analysis on the current situation of college students' mental health education and thinking on the path," *Journal of Jilin University of Chemical Technology*, vol. 37, no. 10, pp. 67–70, 2020.
- [11] Y. T. Li, "Research on college mental health education mode under the background of big data," *Modern Communication*, vol. 12, no. 7, pp. 163–165, 2020.
- [12] J. Yan, "Research on decision tree algorithm and its application in college students' mental health data processing," *Journal of Jiangnan University*, vol. 43, no. 4, pp. 371–375, 2015.
- [13] Y. Li, "Data collection and analysis of mental health data of university students based on internet plus," *Psychological Monthly*, vol. 11, no. 9, pp. 15–16, 2018.
- [14] J. L. Li, K. Y. Wang, and M. L. Zhang, "Internet based mental health data platform for college students: analysis based on blockchain technology," *Wireless Internet Technology*, vol. 16, no. 4, pp. 160–161, 2019.
- [15] Y. L. Fan, "Innovative research on the way of college students' mental health education under the big data environment," *Journal of Jilin Agricultural Science and Technology University*, vol. 28, no. 3, pp. 19–22, 2019.
- [16] M. M. Zhuang, "Research on the innovation of college students' mental health education from the perspective of big data," *Journal of Heilongjiang Vocational College of Ecological Engineering*, vol. 32, no. 5, pp. 104–107, 2019.
- [17] T. Wu and J. Y. Huang, "Problems of historical security model in college students' mental health management," *Information and Computer*, vol. 22, no. 7, pp. 163–165, 2016.
- [18] B. B. Huang and L. Zhang, "The value, current situation and path of establishing university students' mental health database," *China Health Industry*, vol. 14, no. 12, pp. 163–165, 2017.
- [19] X. J. Lu, D. B. Huang, and M. Q. Ye, "Research on the four-layer structure of big data for promoting students' mental health," *Computer Knowledge and Technology*, vol. 13, no. 3, pp. 12–13, 2017.
- [20] C. Q. Zhang and H. D. Meng, "Clustering analysis algorithm and application research in data mining," *Shandong Industrial Technology*, vol. 11, no. 7, pp. 159–162, 2017.
- [21] Y. S. Hou, "A new clustering algorithm based on deep learning," *Journal of Xinxiang University*, vol. 35, no. 12, pp. 21–24, 2018.
- [22] Z. F. Li and Y. Zhang, "Analysis and evaluation of clustering analysis algorithm," *Electronic Technology and Software Engineering*, vol. 13, no. 7, pp. 157–160, 2019.
- [23] M. Huang, E. Li, Y. Yuan, and J. Zheng, "Research and implementation of clustering analysis algorithm based on road network topology," *Journal of Sun Yat University*, vol. 54, no. 6, pp. 99–103, 2015.
- [24] L. Zhang and H. Y. Pan, "Research on application of clustering analysis algorithm," *Digital Technology and Application*, vol. 17, no. 10, pp. 143–145, 2016.
- [25] S. G. Zhou, "Analysis of college students' online learning behavior based on k-means clustering analysis algorithm," *Information and Computer*, vol. 32, no. 16, pp. 220–222, 2020.
- [26] E. Zhu, Y. Zhang, and P. Wen, "Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index," *Neurocomputing*, vol. 363, pp. 149–170, 2019.
- [27] O. I. Liu, M. A. B. Roslan, R. A. A. Helmi, and A. A. Ahmed, "Comparison and detection analysis of network traffic datasets using K-means clustering algorithm," *Journal of Information & Knowledge Management*, vol. 19, no. 3, p. 2050026, 2020.