International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 5525-5536

https://doi.org/10.70517/ijhsa464391

Construction of a Fuzzy Logic Early Warning Model for Mental Health Management of College Students

Minghui Ma[1,*], Si Yang[2] and Weiyi Li[3]
[1] Psychological Counseling Center, Lianyungang Technical College, Lianyungang, Jiangsu, 222000, China
[2] Psychological Counseling Center, Xugou Primary School, Lianyungang, Jiangsu, 222000, China
[3] Psychological Counseling Center, Lianyungang Special Education Center, Lianyungang, Jiangsu, 222000, China
Corresponding authors: (e-mail: mmh19920515@163.com).

Abstract Mental health problems not only affect students' academic performance, but also may have a long-term impact on their personal growth and social adaptability. In this paper, a data-driven mental health management model for college students is proposed by combining fuzzy logic decision-making and risk warning mechanism. First, data mining methods are used to extract the characteristics of students' consumption behavior and Internet behavior, and feature correlation analysis is performed to discover the behavioral differences between normal and abnormal mental health student groups. Then, based on the fuzzy clustering algorithm, the students' mental health status is categorized. The experimental results show that when the number of clusters is 8, the clustering effect of the model is the best, and the error sum of squares is significantly reduced. Specifically, students' behavioral characteristics such as dietary regularity, diligence, number of shared meals, and length of time spent on the Internet are strongly associated with their mental health status. Through clustering analysis and risk prediction, the accuracy of the model reaches more than 95%, which can effectively warn of mental health risks. The study shows that the combination of students' daily behavioral data and advanced data analysis technology can provide strong support for mental health intervention in colleges and universities.

Index Terms mental health, fuzzy logic, data mining, cluster analysis, risk warning, student behavior

I. Introduction

In recent years, young college students have become an important group for the emergence of psychological crisis, they tend to have immature psychological development, unstable emotions, weak will, lack of frustration tolerance, when encountered with social adaptation, communication conflicts, emotional confusion and other problems, it is easy to trigger psychological problems [1]-[3]. Especially in the Internet era of high-speed development of network information technology, college students enjoy the convenience of technology for learning and life, but also to a certain extent suffered from new temptations and new stimuli [4], [5]. On the one hand, the development and application of Internet communication technology has brought more ways for college students to relieve psychological pressure, sort out psychological emotions, solve thought problems, and regulate emotional state [6], [7]. On the other hand, the flood of information and the collision of Chinese and Western cultures will make some college students be tempted by bad information, and they are prone to problems such as psychological lethargy, unswerving thoughts, unstable emotions, and may even embark on the road of illegal and criminal offenses as a result [8]-[10].

For this reason, colleges and universities should build a mental health management system, with all students who have established psychological files as the target of early warning, and the factors prone to psychological crisis as the basis of risk warning and intervention decision-making [11], [12]. According to the results of students' psychological profile screening and students' daily psychological dynamics, the outbreak of college students' psychological crisis can be actively prevented by means of psychological intervention [13], [14]. An effective mental health management system can effectively reduce the serious consequences of college students' psychological crisis, so as to achieve the purpose of early detection and effective intervention [15].

In this study, we first processed the data by collecting students' consumption behavior and Internet habits, and applied data cleaning and feature extraction methods. Then, based on students' behavioral data, a fuzzy clustering model is constructed to classify students' mental health status. By analyzing the behavioral characteristics of groups with different mental health status, the correlation between behavioral characteristics and mental health is further revealed. On this basis, this paper constructs a mental health risk early warning system and uses cluster analysis to predict the risk of students. Finally, the model is validated with experimental data to assess its effectiveness in practical application, aiming to provide accurate mental health management tools for colleges and universities and provide data support for the implementation of relevant interventions.

II. Mental Health Behavior Mining Model for College Students

II. A. Data processing methods

II. A. 1) Data screening

Although data mining [16] is a method for analyzing large amounts of data, it does not mean that we need to use all the collected data, because when collecting data in the early stage, there is no specific consideration of the future use of these data, but only to collect more comprehensive data as much as possible. However, the value density of these data is very low, and is not conducive to the later data analysis, so when having a specific research objective, only the data that is useful for the objective analysis needs to be selected. For the purpose of this research, the original data is the record of students' online behavior in school collected through the web log system, the key information in the record is the student id, the website accessed, the type of website accessed, and the time of accessing the website, etc. The rest of the information can be used without using, such as the log system for security purposes to design some parameters, coding, and other data that are not useful for our analysis.

II. A. 2) Data cleansing

The so-called data cleaning is used to solve the problem of poor data quality due to some accidental factors after the data screening of some useful data, the main means are the following parts:

Missing value processing: When the data set with missing data accounts for a relatively low percentage of the entire data set, and the data volume of the sample is relatively large, this situation can be dealt with by the deletion method, that is, directly discard the data items with missing values.

Outlier processing: also called outlier processing or error value processing, etc., such as a certain data item has a dimension of the value is much larger or smaller than the value of the other data items in the sample, then the data items that have the abnormal value of the dimension is called an outlier.

De-duplication: When the data values between several dimensions are the same or there is a relationship between addition and subtraction operations, they can be considered to represent the same data meaning. Removing duplicate data dimensions has a role in data thinning and model load shedding, which ensures the uniqueness as well as representativeness of the data dimensions.

Noise data processing: the measured data due to some reasons caused by the random error or variance is the so-called noise, is the interference of the data, commonly used means of split-box method and regression method. The split-box method consists of forming a small group of nearby ordered values into a “box”, and then using the mean, median, or boundaries of the data in the box to smooth out these ordered data values and make the data locally smooth.

II. A. 3) Data conversion

Data conversion processing can also be called data mapping processing, which generally has three cases. One for the text data encoding conversion, because the computer can not directly deal with text data such as the calculation of the distance between the two data operations, it is necessary to numerical text encoding, such as in the type of gender will be coded as male 1, female coded as 0, common coding methods such as unique hot code. The second is the format of the conversion, such as date data need to be converted to a uniform format type, to facilitate subsequent analysis and processing. The third is the mathematical processing of numerical data, such as when it is found that the value of a dimension encountered in the form of exponential changes, the formula (1) can be quickly converted into an exponential change in the data to facilitate the observation and analysis of small numerical data.

$$y = \log_m (x+k) \tag{1}$$

Similarly when the data changes in the form of a power function, it can be processed by opening the root of the nth power and converting it to the more easily observable small numerical data by using the formula (2), y is the converted value and x is the value obtained before the conversion.

$$y = \sqrt[n]{x} \tag{2}$$

II. A. 4) Data integration

Students' school network behavior data is recorded in the log system in the log format of the data, and some of the student's basic information data is stored in the student management system, data analysis and model training, if the data in the two systems are operated at the same time, it is cumbersome and may make mistakes, so the data integration method is needed to extract these data and save them in the same environment for processing. The main purpose of data integration is to facilitate data reading operations for the multi-dimensional characterization of samples in the subsequent data mining process, while the heterogeneous datasets required for the analysis are normalized and stored in the same database system.

II. B.Multi-dimensional Student Behavior Feature Extraction and Feature Association Analysis

II. B. 1) Student consumption feature extraction

Consumption behavior is an important part of students' daily behavior, and students' mental health status may affect their consumption behavior in school, so feature extraction was performed on students' consumption data to explore its relationship with students' mental health status.

(1) Dietary regularity. When exploring the dietary regularity, the 3 daily meals were divided into time intervals at 30-min intervals, and the number of times students consumed in each time interval was counted, and the formula of information entropy was utilized to obtain the dietary regularity of students. The opening and closing times of the 3 daily meals were asked to the relevant personnel at the cafeteria window.

The formula for calculating students' dietary regularity is:

$$S_{ER} = \frac{T_d}{D}(-\sum_{i=1}^{n} p_i \log p_i) \tag{3}$$

where D is the number of days that students use their campus cards to pay for their meals in the cafeteria, T_d is the total number of days counted, p_i is the frequency of student spending in each time interval, and n is the number of time intervals divided.

(2) Diligence. The author takes students' 1st daily campus card spending record as their 1st daily activity. Since meal consumption in the cafeteria accounts for the majority of all consumption records, the time of each student's 1st meal swipe per day was calculated and then used as a measure of the student's diligence level. From there, the original datetime format was changed to convert it to a Unix timestamp. Thus, student diligence S_{nc} is calculated as:

$$S_{DG} = \sum_{j=1}^{r} t_j / T \tag{4}$$

where T is the total number of days that students swipe their cards to make purchases, and t_j is the time that students swipe their cards for the 1st meal on the j th day.

(3) Number of people sharing meals. It is assumed that if two students are in the same class, and they swipe their cards to spend money at the cafeteria window on the same floor during meals and the interval between swipes is less than 120s, they are considered to be eating together. I counted the number of students eating together at each cafeteria window of each college on each campus, in order to prevent some students from paying by cell phone, ordering takeout or eating out on the results of the experiment, the number of times the cafeteria swipes the card to consume less than 30 records of the students will be removed in a month, and then the statistics obtained from each cafeteria window will be added up according to the corresponding students' student numbers, which will lead to the final results.

II. B. 2) Student Internet Feature Extraction

Studies have shown that there is a correlation between students' mental health status and the level of Internet addiction, so the relevant characteristics of students' Internet access were extracted to explore the relationship between students' mental health status and their Internet habits.

(1) Average duration of Internet access on weekdays and weekends. Considering that students' online habits on weekdays and weekends are different, the online characteristics of students were extracted by separating weekdays and weekends. The length of students' Internet access each time can be obtained from the dataset of students' Internet access records, and the average length of students' Internet access on weekdays or weekends:

$$S_{DI} = \sum_{i=1}^{r} I_i / T_d \tag{5}$$

where T is the number of times students go online on weekdays or weekends, I_i is the time students spend online each time, and T_d is the number of days counted.

(2) The latest time to get off the Internet on average on weekdays and weekends. The average time of students' latest offline time on weekdays or weekends refers to the half-mean of the time when students last logged out of the campus network system on weekdays or weekends, and the earliest offline time on the 2nd day will be taken as the latest offline time of the day if the students haven't logged out of the campus network before zero o'clock on that day. The time the student logged out of the campus network was converted to a Unix timestamp in the form of the latest downtime of the student's average weekday or weekend Internet access:

$$S_{LT} = \sum_{i=1}^{M} L_i / M \tag{6}$$

where M is the number of days counted and L_i is the Unix timestamp of the last time a student logged out of the campus network system each day on weekdays or weekends

(3) Average number of daily traffic used on weekdays and weekends. The formula for calculating the number of average daily traffic used by students on weekdays or weekends is:

$$S_{FM} = \sum_{i=1}^{n} F_i / d \tag{7}$$

Its traffic unit is MByte. where n is the total number of times the traffic is used, F_i is the number of traffic consumed per use of the traffic, and d is the number of days of using the traffic on the campus network.

II. B. 3) Feature correlation analysis

In order to explore the differences in related behaviors between students with mental health and psychological abnormalities, the Jenks Natural Breaks algorithm was used to label the extracted feature data. The Jenks Natural Breaks algorithm, also known as the natural breaks grading method, has the same core idea as clustering: to maximize the similarity within each group and the heterogeneity between the outer groups [17]. Then, the Apriori algorithm [18] was used to mine the behavioral feature labeling dataset of students with mental health and psychological abnormalities, respectively, and the minimum support threshold was set to 0.5 and the minimum confidence threshold was set to 0.5, and the obtained strong association rules are shown in Table 1. It can be seen from the table that students with psychological abnormalities usually have the characteristics of irregular diet, less diligence, fewer people who eat together, longer time on the Internet and more traffic, and late offline time on weekdays. However, students with normal psychology usually have a more regular diet and are more diligent.

Table 1: Strong association rules generated by Apriori algorithm

Serial number	Association rule	Support	Confidence
1	Psychological abnormal student group⇒ The diet is less regular	0.541	0.541
2	Psychological abnormal student group⇒ Not very hard-working	0.511	0.511
3	Psychological abnormal student group⇒ There are fewer people eating together	0.533	0.533
4	Psychological abnormal student group⇒ Workday Internet time is longer	0.511	0.511
5	Psychological abnormal student group⇒ Spending more time online on weekends	0.514	0.514
6	Psychological abnormal student group⇒ The average daily traffic on weekdays is more used	0.514	0.514
7	Psychological abnormal student group⇒ The average daily traffic on weekends is high	0.507	0.507
8	Psychological abnormal student group⇒ The online and offline time on weekdays is late	0.514	0.514
9	Mental normal student group⇒ The diet is regular	0.791	0.791
10	Mental normal student group⇒ More diligent	0.592	0.592

III. Data-driven decision-making model for college students' mental health

III. A. Improved data-driven fuzzy clustering based algorithm

III. A. 1) Improved fuzzy clustering algorithm models

Assuming the dataset is n, the number of categories is c, the number of categories takes the value interval of $[1, n]$, and defining $1\ h_{ik}$ as the probable degree of affiliation, which denotes the degree of affiliation corresponding to the k th object and the i th category, and takes the value interval of [1, n], the degree of affiliation of this category is correlated to the distance between the point representing this category and the clustering center, and the distance is [1, n], the smaller the distance, the greater the degree of affiliation, independent of other clustering centers. The smaller the distance, the greater the degree of affiliation, independent of other clustering centers.

Definition 2 s_{ik} is the uncertainty affiliation degree, which represents the possibility that the k th object belongs to the i th category, and takes the values of 0 and 1. Uncertainty affiliation degree has the function of memory storage, which is able to memorize the previous iteration process.

$$s_{ik} = \begin{cases} 1 & \min\limits_{1\le i\le c}\{d_{ik}\}\forall i \\ \text{Remain unchanged} & \text{Other} \end{cases} \tag{8}$$

This algorithm is defined as the precondition for uncertainty affiliation when the affiliation degrees of different clusters are all 1, which is different from other concepts of uncertainty.

Definition 3 The criterion function of the algorithm is constructed based on the parameters h_{ik}, s_{ik}, i.e., the

$$\left.\begin{aligned} J(T,P,V) &= \sum_{k=1}^{n}\sum_{i=1}^{c} h_{ik}^{r} d_{ik}^{2} + \sum_{k=1}^{n}\sum_{i=1}^{c} s_{k} d_{ik}^{2} \\ &\sum_{k=1}^{n}\sum_{i=1}^{c} h_{ik} = n \end{aligned}\right\} \tag{9}$$

where $\sum_{i=1}^{n}\sum_{i=1}^{n} h_{ik} = n$ denotes the affiliation constraints, T denotes the matrix of affiliations, P denotes the matrix of uncertainty affiliations, V denotes the matrix of clustering centers, r denotes the likelihood weight index2, d_{ik} denotes the measure sample and cluster center Euclidean distance quotient, $d_{ik} = |x_k - v_i|$.

Combining Definition 3 and the Lagrangian summation method to compute h_{ik}, the Lagrangian function is:

$$F = \sum_{k=1}^{n}\sum_{i=1}^{c} h_{ik}^{r} d_{ik}^{2} + \sum_{k=1}^{n}\sum_{i=1}^{c} s_{k} d_{ik}^{2} + \lambda\left(\sum_{k=1}^{n}\sum_{i=1}^{c} h_{ik} - n\right) \tag{10}$$

F for h_{ik} is a partial derivation, i.e.:

$$\frac{\partial F}{\partial h_{ik}} = r h_{ik}^{-1} d_{ik}^{2} + \lambda = 0 \tag{11}$$

$$h_{ik} = \left(\frac{-\lambda}{r d_{ik}^{2}}\right)^{\frac{1}{r-1}} \tag{12}$$

F is partial derivative of λ, i.e.:

$$\frac{\partial F}{\partial \lambda} = \sum_{k=1}^{n}\sum_{i=1}^{c} h_{ik} - n = 0 \tag{13}$$

Substituting h_{ik} in Eq. (13) gives:

$$\left(\frac{-\lambda}{r d_{ik}^{2}}\right)^{\frac{1}{r-1}} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{c}\left(\frac{1}{d_{ik}^{2}}\right)^{\frac{1}{r-1}}} \tag{14}$$

Substituting Eq. (14) in Eq. (11), viz:

$$h_{ik}=\left(\frac{-\lambda}{rd_{ik}^{2}}\right)^{\frac{1}{r-1}}=\left(\frac{-\lambda}{r}\right)^{\frac{1}{r-1}}\left(\frac{1}{d_{ik}^{2}}\right)^{\frac{1}{r-1}} \tag{15}$$

Solve the final possible affiliation formula as:

$$h_{ik}=\frac{n\left(\frac{1}{d_{ik}^{2}}\right)^{\frac{1}{r-1}}}{\sum_{t=1}^{n}\sum_{j=1}^{c}\left(\frac{1}{d_{ik}^{2}}\right)^{\frac{1}{r-1}}} \tag{16}$$

All data points are weighted for mean based on the parameters t_{ik},s_{ik} to solve the new clustering center equation as:

$$v_i=\frac{\sum_{k=1}^{n}(t_{ik}^{r}+s_{ik})x_i}{\sum_{k=1}^{n}(t_{ik}^{r}+s_{ik})} \tag{17}$$

III. A. 2) Construction of the evaluation indicator system

Based on the clustering results of the improved fuzzy clustering algorithm, the mental health assessment index system was constructed as shown in Table 2. College students' mental health is measured by a total of 11 indicators in 3 aspects: study, life and future development.

Table 2: Psychological health evaluation index system of college students

Primary indicator	Secondary indicator	Tertiary index	Code
Comprehensive evaluation of mental health of college students	Learning aspect	Learning motivation	A1
		Learning strategy	A2
		Learning ability	A3
		Exam psychology	A4
	Life aspect	self-cognition	A5
		Emotional control	A6
		Man-machine interaction	A7
		Will power	A8
	Future development	Career planning	A9
		Values	A10
		self-development	A11

III. B. Design process of fuzzy comprehensive assessment of college students' mental health

(1) Determine the indicator domain for the assessment object A:

$$A=\{a_1,a_2,\cdots,a_n\} \tag{18}$$

where n denotes the number of assessment indicators, the assessment system of college students' mental health adopts a two-tier structural type, the assessment object is the mental health status of college students, and the indicator thesis domain denotes the first-level indicators.

(2) Determine the assessment level thesis B:

$$B=\{b_1,b_2,\cdots,b_m\} \tag{19}$$

where m denotes the number of assessment grades, each grade is in one-to-one correspondence with each fuzzy subset, and generally m is an odd number, which can be easily utilized to assess its intermediate grade. The assessment criteria B are excellent, good, general, poor and poor.

(3) Establish fuzzy relationship matrix C. For the project to be assessed, all the assessment indicators are quantified in turn, and the fuzzy relationship matrix C is determined according to the affiliation degree $C(a_i)$ of each grade of the fuzzy subset:

$$C=\begin{bmatrix} C(a_1) \\ C(a_2) \\ \vdots \\ C(a_n) \end{bmatrix}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \tag{20}$$

where C is the fuzzy relationship from A to B, and c_{ij} denotes the j th assessment result obtained by the i th factor.

(4) Determine the fuzzy weight vector D of the assessment factors:

$$D=\{d_1,d_2,\cdots,d_n\} \tag{21}$$

For the assessment object, all assessment indicators do not have the same importance, and each factor has different influence on the overall affectivity. In fuzzy assessment, d_i denotes the affiliation degree of a_i to the fuzzy self, and the fuzzy weights are determined by the normalization method and then synthesized.

(5) The fuzzy assessment result vector E is synthesized by the synthesis operator for D and C to get different rows in C reflecting the degree of affiliation of different indicators to the fuzzy subset, and different rows are synthesized by adopting D that is to get the degree of affiliation to the fuzzy subset of all grades, and the fuzzy synthesis and assessment is modeled as:

$$E=D\cdot C=(d_1,d_2,\cdots,d_n)\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}=(e_1,e_2,\ldots,e_n) \tag{22}$$

where C denotes a fuzzy transformer for use between the indicator domain A and the assessment level domain B, with each fuzzy vector yielding the corresponding integrated assessment result E.

IV. Analysis of mental health management of students in colleges and universities

IV. A. Decision-making on the management of student mental health in higher education institutions

IV. A. 1) Trajectory data analysis

In order to verify the performance of Hadoop architecture and improvement, the experimental sample set is 10 months of student on-campus trajectory data from a university data collection system, the student trajectory data is collected at 120 points per day, and the original dataset is horizontally represented as 10 sample sets of different sizes. The first 5 trajectory data samples have small differences, and Hadoop can not reflect its advantages when processing a small number of files, but when the sample set of data volume log increases, Hadoop will be able to distribute and parallelize the large-scale student trajectory dataset, and the cleaning speed is approximately positively correlated with the cleaning volume. Example of the collection of 10 months of students' on-campus trajectory data, the largest sample set of 50,000 monitoring points, 900,000 data, data cleaning time of about 10s, its speed and processing capacity to fully meet the current and even the future period of time of the amount of on-campus trajectory data collection requirements.

Figure 1 shows the daily cleaning of 14,970 trajectory data collection points on a certain day, where the broken line is the average slope, indicating the average change trend. Because the time required for on-campus trajectory data collection is independent of the size of the student anomaly data, and Hadoop is able to handle large-scale unstructured data and classify the original data to differentiate and add timestamps, the quality of the data does not affect the cleaning efficiency of the trajectory data. To verify the efficiency of the algorithm, large-scale anomalous data are randomly generated in 24 time periods of the sample data. Through experimental verification, it is concluded that Hadoop has a strong and fast processing capability, and the cleaning time of 100,000 data is about 5723-6514ms, and the cleaning time changes are more stable for different sizes of anomalous data volume.

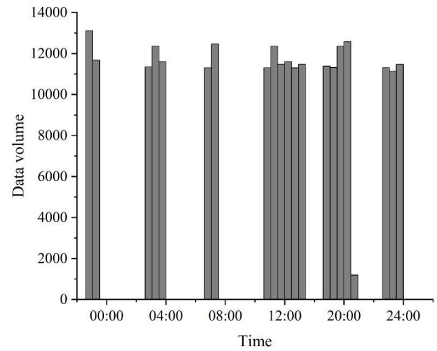


Figure 1: Trajectory data cleaning status

Figure 2 shows the clustering diagram of students' on-campus trajectory data. By collecting large-scale students' on-campus trajectory data during a period of 10 months, cleaning as well as clustering analysis are carried out, and then the results of the test of this early warning decision-making system are compared with the actual state of the students, and the conclusions are shown in Table 3. It can be seen that the results analyzed by this early warning decision-making system are basically consistent with the actual state of these students in the school, and the success rate of data prediction is close to 95%, and the error rate can be controlled within 6.7%.

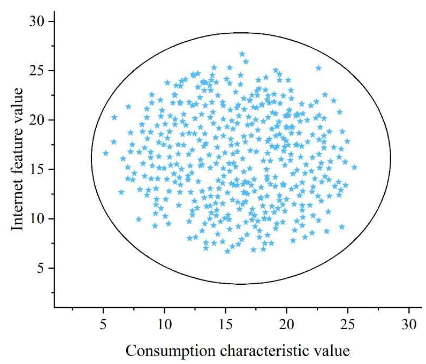


Figure 2: Campus data cluster diagram

Table 3: The system warning is compared to the actual state result

Month	Uniformity ratio /%	Remoteness	Running time /ms	Diff/%
3	95.21	0.14	55.72	90.2
4	92.5	0.11	53.23	89.4
5	93.7	0.07	62.35	91.5
6	98.5	0.06	63.45	98.4
7	93.1	0.09	58.19	92.7
8	97.5	0.07	48.11	94.6
9	96.8	0.07	56.16	94.2
10	88.4	0.15	55.30	84.8
11	99.45	0.04	60.51	99.7
12	100.0	0.01	57.62	100.0

IV. A. 2) Rules for student profiling

Student portrait tags are divided into content and weight. Labels are variable, and weights change in real-time, decaying over time. Take the student's grade record as an example: Zhang San, with a math score of 90, marks the student's performance in a certain subject. By writing student portrait rules, the label weight is calculated, and the basic weight = 90/100 = 0.9. The time decay factor is R, which decreases linearly with the extension of time D (number of days), R=1-0.05×D. Label weight = basic weight × attenuation factor. From this, it is calculated that Zhang San's math score label weight is 0.9, and the label content is the subject name "mathematics", so one of the student's labels is: mathematics, 0.9. After a week, if the decay factor changes to 0.7 and the label weight changes to 0.65, then the student's label is: Math, 0.65. When the weight of the label is continuously reduced to a certain value, such as 0.5, the label of mathematics should be "torn off" for the student, so as to better reflect the real-time nature of the label, so 0.5 is recorded as the threshold. Then use Hive rules to generate student tags and store them in the tag library.

IV. B. Early warning of students' mental health

In order to verify the validity of the researched method to predict the mental health risk of college students, a total of 80 students of grade 18 in a college of a university were selected as the research subjects. Basic information such as students' names, student numbers, attendance status, grades, etc. were collected. The data of students' names and attendance status were used as input variables for the research method. The FCM clustering algorithm was used to implement clustering on the collected student data, and the maximum number of iterations was set to 9. The results of the sum of squares of the errors of clustering at different numbers of categories are shown in Figure 3. The experimental results show that along with the improvement of the number of clusters, the error sum of squares of the clustering of this paper's method decreases, indicating that the higher the number of clusters, the better the clustering effect. When the number of clusters is 8, the trend of improving the error sum of squares of clustering is slower, indicating that the clustering algorithm has stabilized. The number of clusters for setting the mental health risk prediction of college students is 8.

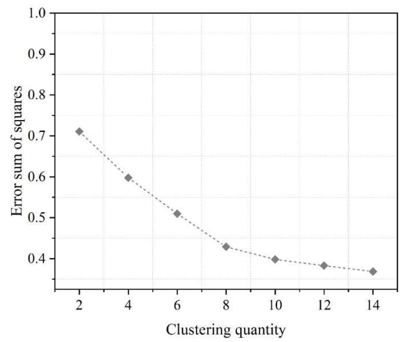


Figure 3: Error squares and results

Set the number of clusters of this paper's method as 8, and the clustering results of this paper's method are shown in Figure 4. Figure 4 clustering results can be seen, using this paper's method to implement clustering of massive college students' mental health risk prediction related data, clustering effect is better. The method of this paper can effectively cluster different categories of data, the clustering effect is not affected by the interference data, and the high-quality clustering results can improve the effectiveness of mental health risk prediction for college students.

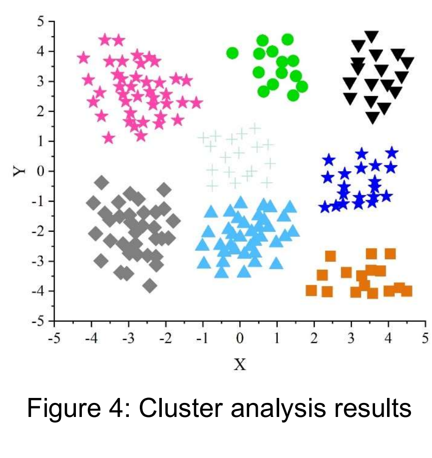
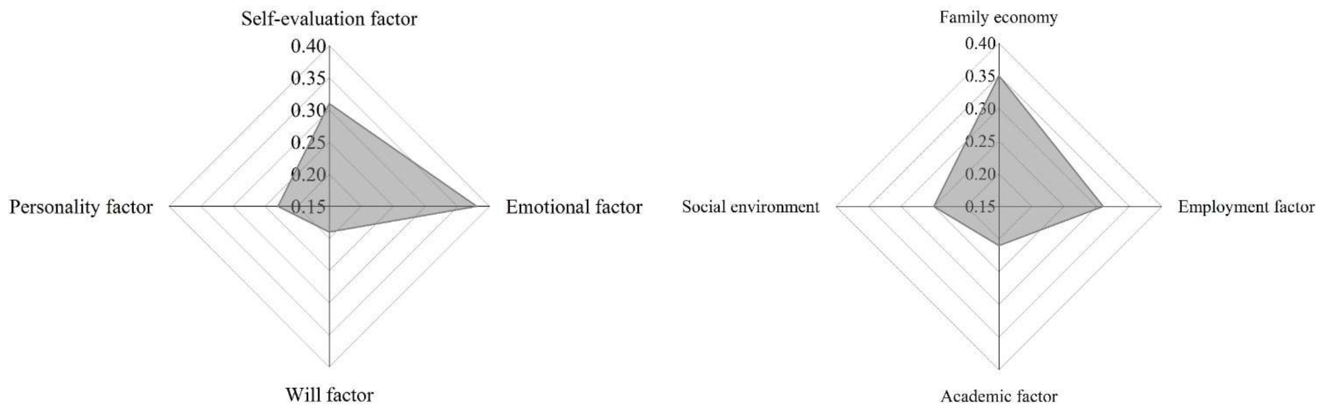


Figure 4: Cluster analysis results

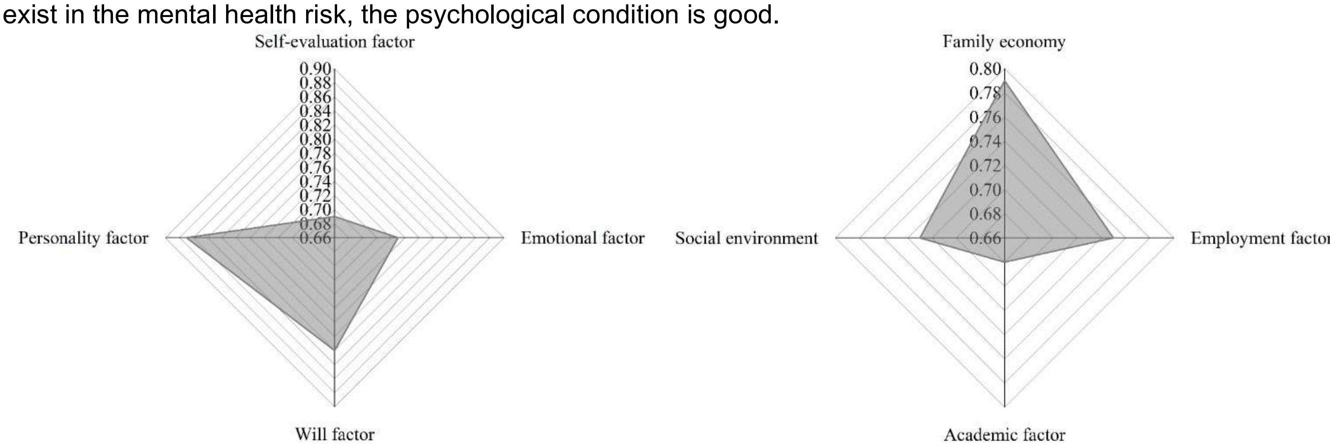
The results of using the method of this paper to obtain the weights of the secondary factors for the prediction of mental health risk of college students are shown in Figure 5. Based on the experimental results in Figure 5, it can be seen that the method of this paper can effectively obtain the weights of different factors, and utilize the obtained weights to enhance the effect of mental health risk prediction of college students.



(a) Physical signs (b) Personality interpersonal factors

Figure 5: Secondary factor weight

A student A was randomly selected from the research object, and the method of this paper was used to predict each secondary factor of the student's mental health risk, and the output interval of the prediction results was [0, 1], and the prediction results were shown in Figure 6. Figure 6 experimental results can be seen, student A's various secondary factors are above 0.7, the prediction result is completely risk-free, indicating that the student does not exist in the mental health risk, the psychological condition is good.



(a) Physical signs (b) Personality interpersonal factors

Figure 6: Student A prediction results

The final mental health risk prediction results of the four students including Student A in the study population were counted and the statistical results are shown in Figure 7. Figure 7 experimental results can be seen, student A and student C mental health risk prediction results are completely no risk; student B and student D mental health risk prediction results are slight risk and serious risk respectively. Comparing the above results with the diagnosis results of professional psychologists, the diagnostic results are completely consistent, effectively verifying that the mental health risk prediction results of the studied college students have high validity and accuracy. College administrators, teachers of Student B and Student D, and students around them need to pay strict attention to the mental health status of the above two students, and take certain interventions to avoid serious consequences for students with higher mental health risks due to their mental health status.

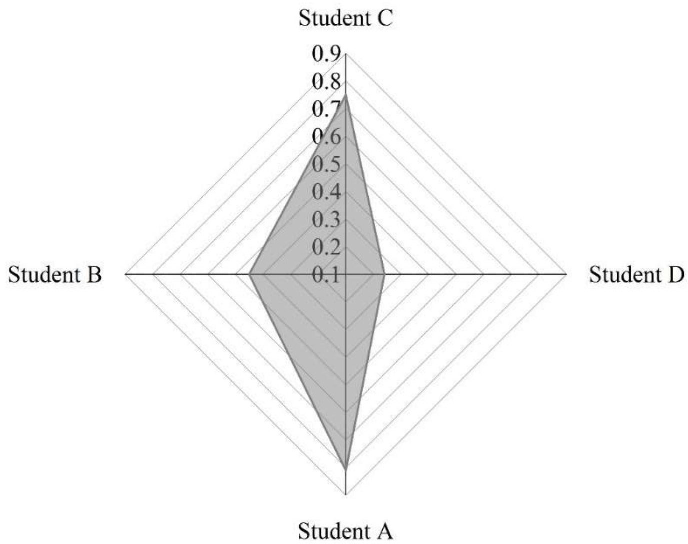


Figure 7: Mental health risk prediction results

V. Conclusion

Through the mental health management model proposed in this paper, the mental health risk of college students can be predicted more accurately. In the experiment, the mental health early warning mechanism based on the fuzzy logic decision-making model used achieved significant prediction results by analyzing the behavioral data of 80 students. The results of cluster analysis showed that the group of psychologically abnormal students showed significant differences in dietary regularity, diligence, and number of meals shared. Further analysis yielded that students' online behavior and consumption behavior were closely related to the state of mental health, and the psychologically abnormal group had longer Internet use, higher traffic consumption, and the presence of fewer shared meals. The prediction accuracy of the model is close to 95% and the error rate is controlled within 6.7%, which verifies the effectiveness of the method in the prediction of students' mental health risk. In addition, calculating the mental health risk score for each student through the fuzzy assessment method can detect potential mental health problems in time and provide a scientific basis for administrators to make accurate intervention decisions, thus effectively reducing the risk of mental health problems.

Acknowledgements

This work was supported by Philosophy and Social Sciences Research at Universities in Jiangsu Province: Research on the integration of mental health education in primary, secondary and secondary schools in Lianyungang (2023SJSZ1074).

References

[1] Campbell, F., Blank, L., Cantrell, A., Baxter, S., Blackmore, C., Dixon, J., & Goyder, E. (2022). Factors that influence mental health of university and college students in the UK: a systematic review. BMC public health, 22(1), 1778.

[2] Green, J. G., McLaughlin, K. A., Alegría, M., Costello, E. J., Gruber, M. J., Hoagwood, K., ... & Kessler, R. C. (2013). School mental health resources and adolescent mental health service use. Journal of the American Academy of Child & Adolescent Psychiatry, 52(5), 501-510.

[3] Karatekin, C. (2018). Adverse childhood experiences (ACEs), stress and mental health in college students. Stress and Health, 34(1), 36-45.

[4] Odgers, C. L., & Jensen, M. R. (2020). Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. Journal of Child Psychology and Psychiatry, 61(3), 336-348.

[5] Chaudhary, A. P., Sonar, N. S., Banerjee, M., & Yadav, S. (2021). Impact of the COVID-19 pandemic on the mental health of college students in India: cross-sectional web-based study. JMIRx med, 2(3), e28158.

[6] Harrer, M., Adam, S. H., Baumeister, H., Cuijpers, P., Karyotaki, E., Auerbach, R. P., ... & Ebert, D. D. (2019). Internet interventions for mental health in university students: A systematic review and meta-analysis. International journal of methods in psychiatric research, 28(2), e1759.

[7] Gupta, A., Khan, A. M., Rajoura, O. P., & Srivastava, S. (2018). Internet addiction and its mental health correlates among undergraduate college students of a university in North India. Journal of family medicine and primary care, 7(4), 721-727.

[8] Zhang, Y. (2023). The Influence of Internet on Adolescent Psychological Development from the Perspective of Developmental Psychology. Lecture Notes in Education Psychology and Public Media, 8, 18-23.

[9] Nopiana, N., Egie, J., & Mers, O. (2022). The Impact of Internet Addiction on Introvert Personality. World Psychology, 1(2), 73-89.

[10] Peris, M., de la Barrera, U., Schoeps, K., & Montoya-Castilla, I. (2020). Psychological risk factors that predict social networking and internet addiction in adolescents. International journal of environmental research and public health, 17(12), 4598.

[11] Liang, F., Wang, L., & Lai, L. (2023). Research on college students' psychological crisis intervention based on web crawler technology. Journal of Computational Methods in Sciences and Engineering, 23(3), 1439-1450.
[12] Li, J. (2023). Does Psychological crisis matter for college students: Role of digitalization and employment. Environmental Science and Pollution Research, 30(29), 73271-73282.
[13] Fei, W., Zhuang, Z., & Yan, S. (2020). Psychological Crisis of College Students in the Self-media Era and Its Intervention Ways. International Journal of Frontiers in Sociology, 2(9).
[14] Zhang, Z. (2024). Early warning model of adolescent mental health based on big data and machine learning. Soft Computing, 28(1), 811-828.
[15] Hill, E. D., Kashyap, P., Raffanello, E., Wang, Y., Moffitt, T. E., Caspi, A., ... & Posner, J. (2025). Prediction of mental health risk in adolescents. Nature medicine, 1-7.
[16] Yanhua Shen & Bing Gao. (2025). Prediction method of commercial customers' mental health based on data mining. International Journal of Data Mining and Bioinformatics,29(1-2),67-86.
[17] Lee Kyu Min,Lim Hyeon Seok,Cho Jae Woo,Lee Sang Kwon,Min Se Hong & Min Jeong Ki. (2019). Fire Risk Assessment of Building Use Types Using Natural Breaks (Jenks). Fire Science and Engineering,33(6),63-71.
[18] You Fu,Fang Ren & Jiantao Lin. (2025). Apriori algorithm based prediction of students' mental health risks in the context of artificial intelligence. Frontiers in Public Health,13,1533934-1533934.