Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 14 (2026) pp. 1-15

Article

Research on the model of integrating craftsmanship into mental health education in colleges and universities based on group intelligence optimization under the perspective of three-whole-education

Wei Zhang [1,*] and Yan Ren [1]

[1] Shijiazhuang Insttute Of Railway Technology，Shijiazhuang, Hebei 050004, China
* Correspondence author: peacelove_wei@163.com

Abstract: Under the guidance of the concept of "three-pronged education", which is all-embracing, all-process education and all-round education, mental health education has become the focus of college education. In this paper, the index system of mental health assessment of college students is constructed, and the K-means clustering algorithm based on the improved bat algorithm is introduced. The ant colony algorithm and other optimization algorithms are combined serially to optimize the association rule mining, and the mental health education model of colleges and universities based on group intelligence optimization is constructed. Taking the students of a college as an example, the model is utilized to study the students' mental health problems. Relying on the characteristics of educational data, the specific manifestations of students with mental disorders are explored. Through cluster analysis and association rule mining of psychological scale data, the four main psychological problems that exist and their internal correlations are obtained. The results showed that the four main mental health problems among students were adaptation and anxiety, introversion and extroversion, emotionality and serenity and alertness, and timidity and boldness. The correlation rule {Adaptation and Anxiety}→{Affective and Tranquil Alertness} has the highest confidence level of 0.9132, and the support level of {Adaptation and Anxiety}→ {Introversion and Extroversion} and {Adaptation and Anxiety}→{Weakness and Boldness} are 0.1984 and 0.1375, respectively. The results show that one of the important psychological problems that cause the other three major psychological problems among the four major psychological problems is Adaptation and Anxiety, which is a significant factor for the how colleges and universities can provide students with mental health guidance has certain practical significance.

Keywords: mental health education; data mining; cluster analysis; association rule mining; improved bat algorithm; ant colony algorithm

1. Introduction

The goal of mental health education in colleges and universities is to improve the psychological quality of college students, cultivate the excellent qualities of optimism, positivity, resilience, responsibility, and honesty, and promote the overall development and healthy growth of college students. Through the development of mental health education, it enhances college students' self-knowledge and emotion management ability, improves interpersonal and communication skills, helps college students better cope with stress and frustration, and strengthens their psychological resilience [1-4]. And craftsmanship is a spirit of pursuing excellence and striving for perfection. It originated from the ancient craftsmen, who carved every detail with their heart, striving to make every piece of work the best. Nowadays, craftsmanship has become a kind of value in our time, which is not only a kind of professional spirit, but also a kind of life attitude. Craftsmanship is an important part of modern education, and it is the core spirit of cultivating talents, as well as the aspect that must be emphasized in the construction of modern education system [5-8]. The concept of “three-round education” refers to the comprehensive development of the individual's thought, character, intellect, physical and aesthetic education, which is one of the basic tasks of education. Under this concept, psychology education in higher education should focus on cultivating students' artisanal spirit, promoting the formation and development of students' artisanal spirit through the discipline system, practical experience, and evaluation and assessment, cultivating more innovative talents with artisanal spirit, and promoting the comprehensive development of an innovative country [9-12].

Chen,N.[13] suggests that integrating the artisan spirit into the ideological and political education in colleges and universities is an effective educational innovation strategy, which guides the students' thoughts and qualities by means of the educational significance of the existence of the artisan spirit to help students form good professional qualities and spirits. Hu,W. et al.[14] point out that the prevalence of affective disorders among college students has seriously affected the development of modern education. It reveals that the situation of students can be effectively improved by integrating the concept of civic politics and the spirit of craftsmanship to ensure the smooth development of students' life and learning. Li,L. et al[15] describe the connotation of civic education and the spirit of craftsmanship, reveal the problems existing in the civic education of colleges and universities, examine the effective combination of the cultivation of the spirit of craftsmanship and the ideological-political education of the curriculum, and put forward the effective strategies for its advancement. Wang,L. ...[16] analyzes the role of craftsmanship in Civic and Political Education. It shows that by utilizing the system transformation teaching mode can effectively cultivate the positive psychology of students, promote the development of teaching and improve the comprehensive quality of students. Yao, L.[17] reveals that the application of craftsmanship in innovation and entrepreneurship in colleges and universities shows a good effect, and explores the improvement of the innovation and entrepreneurship ability of college students by using craftsmanship as an entry point. The above studies have explored the application of craftsmanship in ideological education, innovation and entrepreneurship in colleges and universities, revealing its positive impact on students' comprehensive quality and teaching quality, but as of now, academics have not yet emphasized the application of craftsmanship in mental health education.

This paper adopts hierarchical analysis method to construct the index system of mental health assessment for college students, and feature construction and extraction based on data mining technology. The improved bat algorithm is proposed, and the improved bat algorithm is combined with K-means algorithm to optimize the clustering iteration. Combine the ant colony algorithm with other optimization algorithms serially, use the serial hybrid algorithm based on the ant colony algorithm for association rule mining, and propose the mental health education model of colleges and universities based on group intelligence optimization. Taking the data of a college mental health education center as an example, the model constructed in this paper is applied to analyze students' mental health. By extracting the features of the education data, the performance of students with mental disorders in terms of consumption and achievement is explored. Relying on the clustering weight comparison, four main psychological problems are obtained. Through the association rule technique, the connection between the four psychological problems is mined.

2. Construction of student mental health model based on data mining

2.1. Feature selection

Based on psychological and sociological theories, a hierarchical analysis was used to construct an indicator system for assessing the mental health of college students, which covers three core indicators at the feature level and is subdivided into nine specific assessment indicators at the base level. The static and dynamic characteristics of the students involved in the construction of the student mental health model based on data mining are divided into a three-layer assessment index system after reference, which are the personal factor level, the family factor level, and the school factor level. The specific realization of the basic layer indicators is shown in Table 1.

Table 1. Characteristic index system

Target layer	Feature layer	Base layer
Characteristic index system	Personal factor	Physical condition Networking (social intimacy) Personality orientation
	Family factor	Family financial situation History of growth Nature of household registration
	School factor	Academic performance Degree of diligence in school Participation in collective activities

2.2. Feature construction and application

2.2.1. Algorithm design

Based on consecutive months of students' school behavior records, this paper explores the synchronous action patterns of similar or closely related students during specific time periods through behavioral trajectory data mining techniques. The association rule mining algorithm is used to discover groups of students who often act together at the same time and place, which is used to assess the similarity of behavioral trajectories among students. By counting the frequency of co-occurrence of any two students, it is transformed into a relationship strength index, i.e., the higher the number of co-occurrences, the higher the behavioral similarity or relationship closeness between them.

Accordingly, this paper constructs a social graph based on individual relationship index for each student, and further integrates it to form a social network system for all students in the school. In this network, students exist as independent nodes, and the lengths or weights of the connecting lines (edges) between them represent the relationship indices between students. In addition, under the social network framework, even if two students are not directly connected, the similarity or relationship index between them can be indirectly inferred by sharing other nodes.

2.2.2. Integrated applications

The data is cleaned and counted to create a data model. Compare the status of a student by calculating the position of each normal distribution of a student in the data model, i.e., standardized scores, using the data on medical visits as an example:

For the frequency of medical attendance p_A there is a total group of students $P=\left[p_1, p_2, \ldots, p_{n-1}, p_n\right]$, it is clear that the group of data conforms to a normal distribution, calculate the mean and variance or standard deviation of this distribution, respectively, and then use the probability density function to calculate that its Z-Score.

Its probability density is given in equation (1) below:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{(x-\mu)^2}{2\sigma_0^2}\right\} \tag{1}$$

For the frequency of medical visits, mean μ and standard deviation S Eq. (2) were calculated:

$$\overline{X}_{medical}=\frac{\sum_{i-1}^{n} p_i}{n} \tag{2}$$

$\overline{X}_{medical}$ represents the sample size of the number of visits, and n is the number of that sample, as in equation (3) below.

$$S_{medical}=\sqrt{\frac{\sum\left(p_i-\overline{X}_{medical}\right)^2}{n-1}} \tag{3}$$

where, $S_{medical}$ denotes the standard deviation of this sample, which is calculated as the standardized score in equation (4) below:

$$Z_{medical} = \frac{X_{medical} - \overline{X}_{medical}}{S_{medical}} \tag{4}$$

Here, $Z_{medical}$ represents the standardized score, and $X_{medical}$ is the frequency of the student's medical visits p_A.

The same steps can be applied to the sports test scores and sports app punch card data because both of them fulfill the conditions of normal distribution.

Finally, the mean of the Z-scores of the three factors was calculated as in equation (5) below:

$$Average\ Z - score = \frac{Z_{medical} + Z_{test} + Z_{app}}{3} \tag{5}$$

This mean Z-score represents the relative position of the three factors in a normal distribution. In a normal distribution, a data point with a mean of 0 and a standard deviation of 1 has a Z-score of 0, indicating that the data point is in the center of the distribution. Based on the magnitude of the absolute value of the mean Z-score, the location of these factors relative to the center of the normal distribution can be determined. This score can then be used for the prediction and judgment of students' mental health to support the data for the subsequent use of the model.

2.3. Feature extraction

2.3.1. Consumption characteristics

Several studies have found that people with mental health problems can develop eating disorders, especially those with depression. Based on this fact, we can analyze the eating patterns of students by extracting their consumption records in the cafeteria through the names of the stores. We pay special attention to students' breakfast/lunch/dinner daily patterns. In this study, we set the breakfast time period from 6:00 to 9:00, the lunch time period from 11:00 to 13:00, and the dinner time period from 17:00 to 19:30. Since there are often multiple records in each meal, we take the first swipe time in a meal as the meal time for that meal. For example, if there are 3 records in a breakfast and they occur at 7:20, 7:21 and 7:22, then we use 7:20 as the breakfast time. The regularity of a behavior can be considered repeatable and will be measured by the entropy of the probability of the behavior occurring in a given time interval. Assuming that there is n time interval $T = \{t_1, t_2, t_3, \ldots, t_n\}$, for any given student, the probability $P_v(T = t_i)$ that behavior $v \in V = \{breakfast, lunch, dinner\}$ occurs within time interval t_i is calculated as in equation (6), where $n_v(t_i)$ denotes the frequency with which behavior v occurs within time interval t_i. Entropy is then calculated with the formula (7), assuming that for each of the three behaviors of breakfast/lunch/dinner, each time interval spans half an hour.

$$P_v(T = t_i) = \frac{n_v(t_i)}{\sum_{i=1}^{n} n_v(t_i)} \tag{6}$$

$$E_v = -\sum_{i=1}^{n} P_v(T = t_i) \log P_v(T = t_i) \tag{7}$$

The entropy value of morning/afternoon/dinner can be calculated according to equation (7). At the same time, it can be seen from Eq. (7) that the smaller the value of a behavior is, the more concentrated its probability distribution is over time intervals, and the higher its regularity is.

2.3.2. Internet characteristics

Internet behavior is strongly correlated with psychological traits such as personality, mood, and depression. The Internet era has changed people's way of life, many people like to make remarks on the Internet platform that reflect their current psychological state, or when they feel unwell, they first go online to check the relevant symptoms, and these online behaviors will be written into the database, and we can mine the online characteristics of students based on their browsing history. We establish an online time sequence for each student according to the order of access time, and mine the students' online patterns from the online time sequence. In recent years, one-dimensional convolutional neural networks (1D-CNN) have been widely used for time series feature extraction, such as time series analysis of sensor data, analysis of audio signal sequences, and also for natural language tasks. In this study, we utilize 1D-CNN to extract students' surfing patterns from surfing time series.

In this study, we designed the overall framework of 1D-CNN as shown in Fig. 1, with five neural network layers. Among them, the first and third neural network layers are convolutional layers, in the training of the model, the input of each student is a sequence of Internet access, i.e., a one-dimensional vector, which is padded with zeros due to the inconsistent length of the Internet access sequence. The operation of the convolutional layer is shown in Equation (8) where x_j^l is the feature map corresponding to the j rd convolutional kernel of layer l, x_i^{l-1} denotes the input while k_{ij}^l denotes the j th convolutional kernel of layer l, b_j^l denotes the bias corresponding to the j th kernel, $f(\cdot)$ denotes the activation function and * denotes the multiplication operation.

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{8}$$

The second and fourth neural network layers use a pooling layer, and the pooling method is divided into average pooling and maximum pooling, and the pooling method we use in this study is maximum pooling. The operation of the pooling layer is shown in Equation (9), where x_j^l, w_j^l, b_j^l, and $down\left(x_j^l\right)$ denote the input, weight matrix, bias, and downsampling function, respectively.

$$\hat{x}_j^l = f\left(w_j^l\right) down\left(x_j^l\right) + b_j^l \tag{9}$$

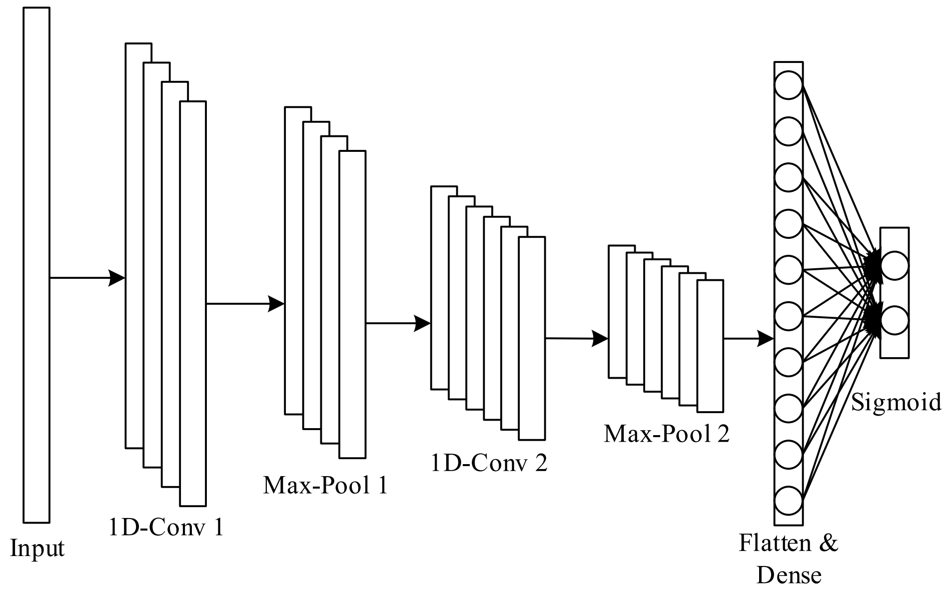


Figure 1. 1D-CNN structure

The fifth layer uses a fully connected layer and the operation of the fully connected layer is shown in Equation (10).

$$s^{pat} = f\left(w^{(5)} * x^4 + b^{(5)}\right) \tag{10}$$

For the parameter settings, the number of convolution kernels for the two convolutional layers were 16 and 32. reLU was used as the activation function and adam as the optimization algorithm. In addition, to prevent overfitting, we used three dropout layers with parameters of 0.15, 0.15 and 0.5. In the model training phase, we selected 70 positive samples and 70 negative samples to train the model. In the feature extraction stage, all the experimental samples are fed into the trained 1D-CNN model, and finally the result s^{pat} of the fully connected layer is output as the internet access pattern.

Insomnia is not only a symptom of depression, it is also an important factor in causing depression. Symptoms expressed by depressed patients at the time of consultation usually include insomnia. Depression, anxiety and obsessive-compulsive disorder are common in patients with insomnia, while patients with insomnia with prolonged sleep onset may exhibit somatization symptoms, and for shorter sleep onset is usually associated with obsessive-compulsive disorder and anxiety. Whether using the campus network during the day or in the middle of the night, students' access requests are recorded in the server, and we can determine whether a student is insomniac or not according to the access time, if a student has access records between 1:00 and 5:00 in the middle of the night on a certain day, we consider that the student is insomniac on that night, and if there are no access records during this period, then there is no insomnia, and generally long term insomnia is most likely to be caused by diseases, therefore, in the The probability of insomnia over a period of time is more indicative of the student's psychological state. We used equation (11) to calculate the insomnia probability of students, where T denotes the number of days, P_i denotes the insomnia probability of student i on day T, and t_i denotes the number of insomnia days of student i on day T. In this section of the web logs 2 characteristics were obtained: internet access pattern, insomnia rate.

$$P_i = \frac{t_i}{T} \tag{11}$$

3. Analysis techniques based on group intelligence optimization

3.1. K-means clustering based on improved bat algorithm

3.1.1. Batting Algorithm Improvement

The bat optimization algorithm is a bionic group intelligence optimization algorithm derived from the simulation of real-life bat echolocation characteristics and its movement behavior. In the search process, bats use the change of acoustic signal frequency to control the direction and range of their movement and update the speed and position in real time; the instant adjustment of pulse emissivity and acoustic loudness can effectively realize the search for the optimal solution of the problem by individual bats, at the initial stage of the algorithm, the pulse emissivity of the individual bats is relatively small, and the acoustic loudness is large, with the deepening of the search process, the number of algorithm generations is superimposed, and the pulse emissivity gradually increases. With the deepening of the optimization process, the number of generations of the algorithm is constantly superimposed, the pulse emissivity gradually increases, and the loudness decreases; due to the bat's mobile characteristics, the individual bats are constantly scanned and positioned on the target during the search process under the control of the above values, and the problem is solved in the end. The basic principles of the bat optimization algorithm are summarized and described as follows:

(1) The unique echolocation system of bats can effectively determine the differences between obstacles or food sources during the search process;

(2) Bats fly in a randomized direction by adjusting the loudness and frequency during the prey search process based on rate v_i;

(3) The impulse loudness of individual bats changes with the environment and decreases with the number of iterations in the range [A_{min}, A_{max}] during the search process.

The above criterion is a generalization of the ideal rules of the bat optimization algorithm, in the real life application, due to various subjective or objective factors, it is difficult to achieve the ideal conditions, so it is usually used to describe the relevant approximation, such as defining the range of the frequency f to be between $\left[f_{\min}, f_{\max}\right]$, and stipulating that the choice of the pulse wavelength range is $\left[\lambda_{\min}, \lambda_{\max}\right]$. Through the instantaneous adjustment of the pulse frequency, it is possible to adjust the bat search range effectively, so that the bat can efficiently search for optimality in the solution space. By adjusting the pulse frequency on the fly, the bat can effectively adjust the search range of the bat, so that the bat can carry out efficient optimization search in the solution space. In summary, assuming that the dimension of the solution space of the problem to be solved is D, the formula for updating the velocity and position of bat individual i can be expressed as follows:

$$v_i^t = v_i^{t-1} + \left(x_i^t - x^*\right) * f_i \tag{12}$$

$$x_i^t = x_i^{t-1} + v_i^t \tag{13}$$

where v_i^t and x_i^t represent the speed and position of the bat in the t-generation search process; v_i^{t-1} and x_i^{t-1} correspond to the speed and position of the bat in the $t-1$-generation search process, and x^* represents the global optimal individual position in the current optimization search process. By adjusting the pulse frequency, the motion performance of individual bats in the search process can be adjusted instantly, using f_i to represent the pulse frequency of bats i, and defining its mathematical expression as:

$$f_i = f_{\min} + \left(f_{\max} - f_{\min}\right) \cdot \beta \tag{14}$$

where β is a random vector in the range of $[0,1]$ and the initial frequency f_i is randomly assigned in the range of $\left[f_{\min}, f_{\max}\right]$ according to the specific problem. When entering the local optimization search, the position of the bat individual is updated in such a way that it becomes:

$$x_{new} = x_{old} + \varepsilon A^t \tag{15}$$

where ε represents a random number in the range of $[-1,1]$ and $A^t = \left\langle A_i^t \right\rangle$ represents the average loudness set of all bat individuals in the current iteration.

As the iterative optimization process of the algorithm progresses, the impulse emissivity r_i of the bat individuals increases and the impulse loudness A_i decreases, and their update equations can be expressed as follows:

$$A_i^t = \alpha A_i^{t-1} r_i^t = r_i^0 \left[1 - e^{-\gamma t}\right] \tag{16}$$

where A_i^t and A_i^{t-1} denote the pulse loudness values of i the bats during the t and $t-1$ generations of the search for optimization, and $\alpha \in [0,1]$ denotes the decay coefficient of the pulse loudness during the search for optimization; r_i^0 represents the maximum pulse emissivity of the bats i and r_i^t denotes the pulse emissivity of the bats i during the t generations of the search for optimization, and γ denotes the coefficient of increase of the pulse emissivity.

3.1.2. Improved Bat Algorithm with K-means Algorithm

The final clustering results obtained by the traditional K-means clustering analysis algorithm are less stable, and the clustering quality is not ideal. Using the advantages of the group intelligent optimization algorithm, the method mode of combining the intelligent optimization algorithm with the classical K-means clustering algorithm can effectively achieve the improvement of the clustering accuracy and make the clustering results present high-quality performance.

In summary the main steps of the $K-means$ clustering algorithm (LWBA-Kmeans) based on the improved bat algorithm proposed in this paper are summarized and described as follows:

(1) Input the sample data to be tested, and set the parameters of the algorithm initially: the size of the bat population N, the maximum number of iterations of the algorithm $Max\ Cycle$, the pulse emissivity r, the initial value of loudness A, the range of frequency variation $\left[f_{\min}, f_{\max}\right]$, the variation factor of loudness and pulse emissivity α, γ, the number of clusters of the clusters K, the threshold value L, etc.

(2) Population initialization: K clustering centers were randomly selected and used as the initial positions of bat populations, and the positions corresponding to the current optimal bat individuals were defined at the same time;

(3) Perform one clustering division by $K-means$ clustering algorithm, where each individual bat represents a division, and calculate the fitness value based on the objective function formula;

(4) Generate a new position using the position update formula and evaluate the resulting new position, if $rand > r_i$ exists or is out of the range of the threshold L, generate a new position based on the current optimal position by means of random perturbation and replace the old and the new position after the evaluation calculation;

(5) If $rand > A_i$ and $f_{new} > f_i$ are present, move the position of the bat to the updated position, and if the evaluated characteristics of the updated individual bat are better than the current population optimum, perform the replacement operation and regulate the pulse emissivity r and sound loudness A of the bat according to the formula;

(6) Calculate the fitness value of the updated population, update the clustering center, and perform the $K-means$ clustering operation again;

(7) If the current number of iterations reaches the pre-set maximum number of iterations or meets the termination conditions of the algorithm, the algorithm is terminated and the final clustering results are output; conversely, skip to step (4).

3.2. Association Rule Mining Based on Serial Hybrid Ant Colony Optimization Algorithm

In this paper, the ant colony algorithm is improved by serial combination of ant colony algorithm and other optimization algorithms. Serial combination of algorithms is to add other optimization algorithms in the process of a single algorithm, so as to obtain the advantages of other optimization algorithms, to complement the single algorithm has the shortcomings of slow convergence or the ability to jump out of the local optimum, the main way to achieve this is to divide the algorithm's iterative process into several segments, and use a different algorithm for each segment of the iteration. In the whole algorithm process, the population number of each algorithm is the same, and the total number of iterations is the sum of the iterations of each algorithm. In the whole process of the algorithm, when the number of iterations of Algorithm I is completed, the best population obtained as the initial population of the new algorithm to start the new algorithm's iteration is used to maintain the effect of Algorithm I.

The specific association rule mining method flow of serial hybrid algorithm based on ant colony algorithm is as follows, in which the iteration order of a single algorithm in the hybrid algorithm process is not certain due to the hybrid strategy, in order to express the convenience, the algorithm that iterates first is called Algorithm I, and then named as Algorithm II, Algorithm III and so on, according to the order of iteration at a time.

Step 1: Scan the dataset, count the number of attribute items, and transform the transactions in the dataset into 0-1 matrices for storage;

Step 2: Initialize the population individuals of Algorithm I. All individuals are randomly initialized in the solution space according to the coding rules, while setting all the parameters needed for Algorithm I. The total number of iterations T, determining the number of iterations for Algorithm I T_1, and the number of iterations for Algorithm II T_2, of which $T = T_1 + T_2$, the minimum support $Min\_Support$, and the minimum confidence $Min\_Confidence$;

Step 3: For all the individuals in the population to carry out the calculation of the fitness function, according to the results of the calculation and the update rules of Algorithm I to carry out the updating of the individuals in the population, for the updated individuals through the sigmoid function for 0-1 mapping processing;

Step 4: Perform rule evaluation for the individuals in the population, and store the conforming individual rules and rule-related support and confidence into the HashSet and HashMap collections;

Step 5: Determine whether the number of iterations of Algorithm 1 is reached T_1, if not repeat steps three and four. Instead execute step six;

Step 6: Take the population obtained from the iteration of Algorithm I as the initial population of Algorithm II, initialize and set the other parameters required for Algorithm II, and start the iteration of Algorithm II;

Step 7: For the individuals in the population, calculate the fitness function value, update the population and parameters according to the population and parameter updating rules of Algorithm II, and process the 0-1 mapping through the sigmoid function for the updated individuals;

Step 8: perform rule evaluation for the individuals in the population, and store the conforming individual rules and rule-related support and confidence into the HashSet and HashMap collections;

Step 9: Determine whether the number of iterations of Algorithm II T_2 is reached, if not repeat steps 7 and 8, and vice versa exit the loop. Output the stored rule ensemble and the support and confidence of the relevant rules into an Excel table, the statistical algorithm execution time and the number of rules searched.

Where the use of HashSet to store the population to meet the requirements of the rules of the individual is the use of HashSet data structure can not be stored in the characteristics of repeated elements to repeat the associated rules screening. At the same time the use of HashMap to store the rules and associated support and confidence, key-value pairs are stored in the form of memory.

4. Application of student mental health model based on group intelligence optimization

The data used in this study consists of two parts: psychometric scale data and educational data.

Psychological scales are effective tools for collecting data related to students' psychological status.

Most colleges and universities will launch psychological assessment activities at the time of new student enrollment and the beginning of a new semester to collect a large amount of data from students' psychological assessment questionnaires in order to understand the latest psychological status of students. Therefore, the collection of questionnaire data is feasible and convenient. The questionnaire data used in this paper comes from the mental health education center of a university, involving 1731 students.

In the process of studying at school, students interact with various business systems of the school, generating a large amount of educational data that can reflect students' individual characteristics and behavioral features. Campus card consumption data, student performance data, access control data, campus network usage data, etc. all belong to this kind of data. This paper starts from the reality of students in colleges and universities, taking into account that the dormitory access control exists in the case of one person swiping the card and many people pass, and the library is not the only choice for students to study, most students will choose to study in the dormitory or the teaching building, so this study does not use access control data. This paper mainly uses the consumption data exported from the campus card system of a university and the historical performance data of students in school to analyze students' academic performance and living habits, and to extract students' behavioral characteristics. Due to the large amount of data involved in this paper, before analyzing the data, this paper first preprocesses the data, cleanses and normalizes the data according to the actual needs, and desensitizes the data.

The psychological scale used in the study of this paper is the University Student Personality Inventory (UPI), and those with a total score of 25 (including 25) or more are categorized as Category A, those with a total score between 20-24 are categorized as Category B, and those with a total score of less than 20 are categorized as Category C.

4.1. Data cleansing

4.1.1. Student consumption data cleansing

There are some records in the consumption data where certain fields are missing, such as merchant number, transaction amount, and so on. For this kind of data records, the ignore tuple strategy is adopted to eliminate them. In addition, there are extra spaces before and after the data content in some fields, which may affect the later research. In this paper, the data types of all fields in the data are standardized, and for the fields of string type, all the data are traversed to clear the extra spaces before and after.

This consumption data has a total of 52 fields, the fields contain a large number of redundant fields that are not relevant to this study or repeat the meaning, introducing a large amount of noise. In order to facilitate the development of the study, this paper reduces the fields of the consumption data and selects five necessary fields: account number, merchant number, transaction time, transaction amount and transaction location. Among them, the account number and student number become one-to-one mapping relationship, which can correspond to the consumption students through the account number; the merchant number and transaction time can be used to analyze the student's school activity area and time; the transaction amount and transaction location can be used to analyze the student's consumption habits. Examples of the filtered data characteristics are shown in Table 2, where the first 12 digits of the order number record the transaction time of the order. Note that the data in the table are not real data and have been desensitized to protect students' privacy.

Since the research in this paper will use the questionnaire data and student consumption data in combination, the two research objects need to be consistent, so it is necessary to intersect the student lists in the two parts of the data, so as to filter part of the consumption records in the student consumption data.

Table 2. The format of campus consumption data

Order number	Account number	Merchant number	Amount of transaction
20240415091541...	23019625467	1032864	-17.00
20240415081238...	23052355411	1018933	-3.50
20240415201722...	23044332086	1010843	-14.00
20240415135521...	23065643324	1035892	-9.00
20240415165023...	23048326222	1053971	-6.50
20240415061123...	23096241475	1043522	-25.00
20240416211542...	23081286533	1052321	-30.00
20240416123548...	23017885632	1063354	-9.50
20240416171183...	23012786529	1064211	-13.50
20240416073118...	23018086598	1084246	-42.00

4.1.2. Data cleansing of student performance

A total of 14 fields were included in the student achievement data. From the research objectives, five of these fields were selected for research and analysis, including student number, course number, course grade, scoring method, and course credit. There is no missing data involved in these five fields. The form of student achievement data is shown in Table 3.

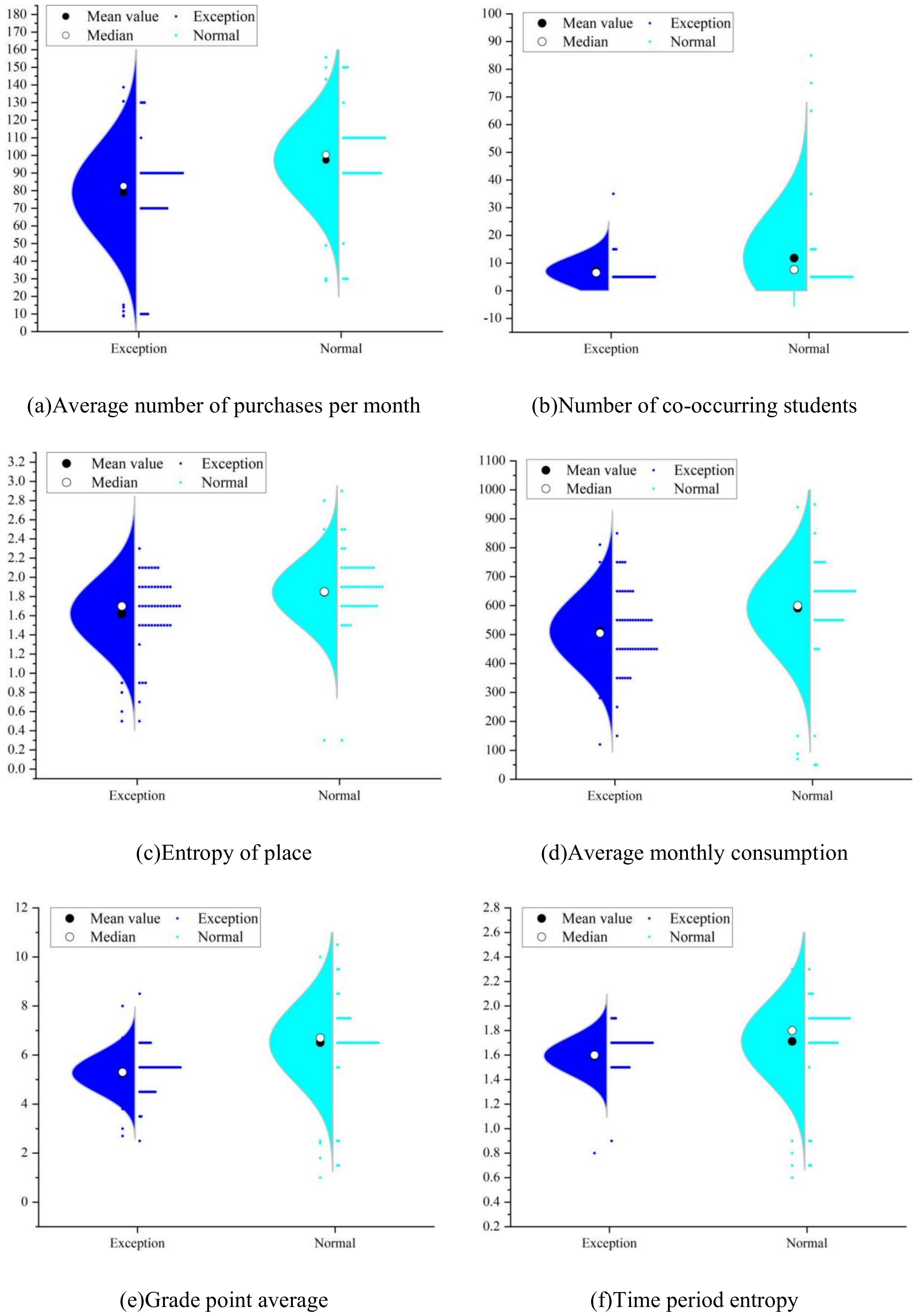
Table 3. The form of student achievement data

Student ID number	Course number	Grade	Scoring method	Course credit
00001	100230	92	System of percent	3
00002	100235	Fail	Two level system	2
00003	100310	Pass	Two level system	2
00004	100425	78	System of percent	4
00005	100325	Excellent	Five level system	2
00006	100240	Good	Five level system	2
00007	100260	Pass	Two level system	1
00008	100320	Fail	Two level system	1
00009	100455	Excellent	Five level system	3
00010	100625	89	System of percent	4

Since the training and testing of the model needs to ensure that the data in each section corresponds to the same group of students, the student performance data also needs to be aligned with the student consumption data and the student psychometric data, i.e., further filtering out some of the data.

4.2. Feature extraction

In this section, education data are analyzed and transformed according to data characteristics in order to extract data features containing representative information for subsequent model inputs. Five consumption features are extracted from the student consumption data: average number of times per month, average monthly consumption, location entropy, time period entropy, and number of co-occurring students, and the average number of credits is extracted from the student performance data. The data distribution of the six extracted features is shown in Figure 2. According to Figure 2, the overall distribution of different features in the abnormal and normal samples can be observed.



(a)Average number of purchases per month

(b)Number of co-occurring students

(c)Entropy of place

(d)Average monthly consumption

(e)Grade point average

(f)Time period entropy

Figure 2. Distribution of educational data characteristics

From Fig. 2(a) and (d), it can be found that for both the average number of monthly consumption and the amount of consumption, the abnormal samples as a whole are slightly smaller than the normal samples, which is in line with the popular cognition that students suffering from psychological disorders are prone to falling into depression, teetotaling, and losing the regularity of their lives.

The location entropy or time period entropy is either extremely large or extremely small, both of which indicate the possibility of abnormal behavior of students. However, from Fig. 2 (c) and (f), it can be found that the overall distributions of these two characteristics of abnormal students are slightly smaller than that of normal students. This may be due to the fact that students suffering from psychological disorders reduce their school activities due to depression.

From Figure 2(b), it can be found that the number of co-occurring relationships with abnormal students is smaller than that of normal students, which is in line with the information revealed by the overall distribution of the other characteristics, that is, abnormal students tend to reduce their school activities, and social activities are also reduced, and then the students who are co-occurring with abnormal students are likely to have a similar behavioral pattern with abnormal students, and they are also the target of attention.

From Fig. 2(e), it can be found that the average GPA of abnormal students is lower than that of normal students as a whole, which indicates that abnormal students may not be able to complete their studies well due to the deterioration of their mental health status, or they may have psychological disorders due to the high pressure of their studies and their inability to keep up with their courses.

4.3. Cluster analysis

The data from the UPI table was organized and the data was analyzed by clustering using the LWBA-Kmeans algorithm proposed in this paper. Firstly, the classification is determined, in this case the optimal number determined is 3 classes and the frequency of each class is shown in Table 4 Cluster Distribution Table.

Table 4. Clustering distribution

		N	% of Combined	% of Total
Cluster	1	1238	71.64%	71.52%
Cluster	2	362	20.95%	20.91%
Cluster	3	128	7.41%	7.40%
	Combined	1728	100.0%	99.83%
Excluded Cases		3		0.17%
Total		1731		100%

As can be seen from the above table, out of the 1731 cases of data in this case, 3 data records were not involved in the analysis due to missing data for one or more variables. Of the remaining 1728 cases, 1238 cases were in cluster 1 (cluster1), 362 cases were in cluster 2 (cluster2), and the other 128 cases were in cluster 3 (cluster3).

The cluster spectrum of the cluster analysis is shown in Figure 3, and it is easy to see that most of the students belonged to cluster 1, followed by cluster 2, and the least was in cluster 3.

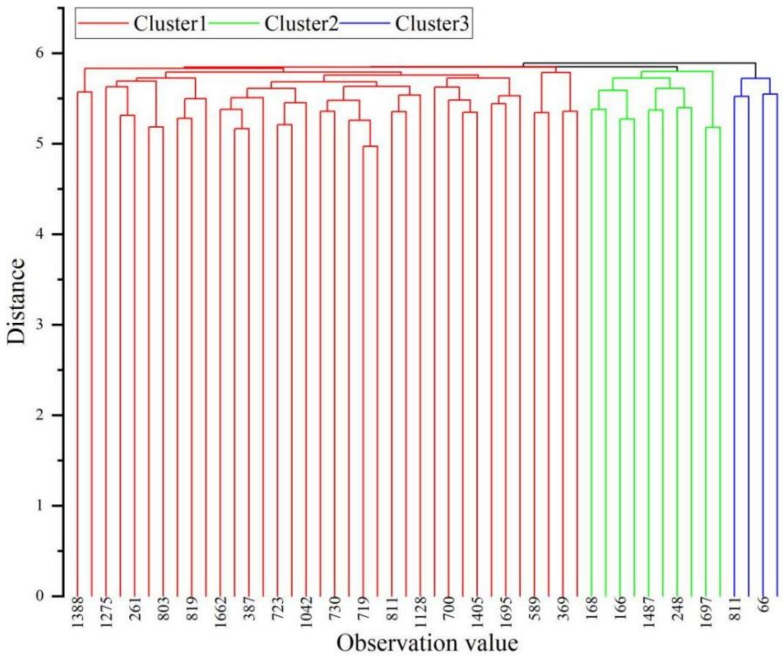


Figure 3. Cluster lineage diagram

In order to have a clearer presentation of the mean and standard deviation of each variable in each cluster, the distribution of the mean and variance values of the four psychological problem variables of the cluster, namely, Adaptation and Anxiety, Introversion and Extroversion, Emotionality and Serenity and Resourcefulness, and Timidness and Boldness, are statistically represented in this paper.

According to the statistical data, with the horizontal coordinate 0 as the dividing line and the vertical coordinate indicating the clusters, the direction of the bars to the left indicates that the values of the corresponding variables are below average; the direction of the bars to the right indicates that the values of the corresponding variables are above average. The weights of the four variables in each cluster are shown in Figure 4.

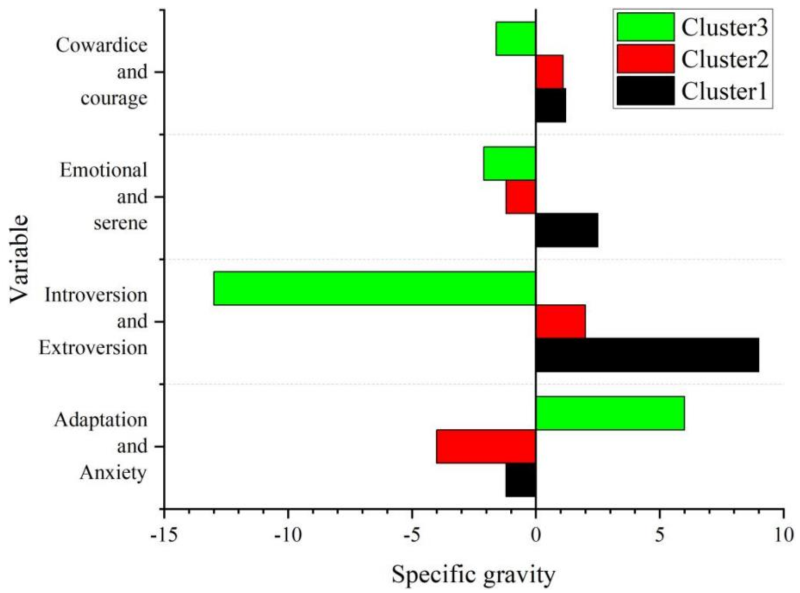


Figure 4. Proportion of four variables in each cluster

As can be seen from Figure 4, in the dimension of Adaptation and Anxiety, the proportion of the third category is higher than the average of 6%, which indicates that the adaptability of this category of students is relatively low, the anxiety is more obvious, usually easy to be agitated, anxious, and often feel dissatisfied with their own situation, and a high degree of anxiety not only reduces the efficiency of work, but also affects the health of the body. The second category is lower than the average of 4%, indicating that the students in this category are more adaptable and feel satisfied, but the extreme low-scorers may lack perseverance, know everything is difficult and refuse to work hard and hard. The first category, on the other hand, is just right and in relatively good shape.

On the introversion vs. extroversion dimension, Category 1 has an above-average 9% share and is the most extroverted, compared to the more introverted students in Category 3 and the relatively extroverted students in Category 2.

In the dimension of Emotional and Safe and Alert, the percentage of category 1 is above average by 2.5%, which indicates that students in category 1 are safe and alert, enterprising and positive. On the other hand, students in the third category are much lower than the average in terms of Emotionality and Serenity and Alertness, indicating that they are emotionally rich, emotionally disturbed and restless, and usually feel frustrated and discouraged.

In the dimension of timidity and boldness, the first category has the highest proportion, while the third category has the lowest, with a difference of 2.8 percentage points.

4.4. Association Rule Mining

The data on the variables of psychological problems were scanned using a serial hybrid algorithm based on ant colony algorithm to obtain the set of items with weighted support higher than or equal to 0.08, which were ranked according to the magnitude of the weighted support, and it can be found in Table 5 that in descending order of severity they are Adaptation vs. Anxiety, Introversion vs. Extraversion, Emotional vs. Serenity and Alertness, and Timidity vs. Courageousness.

Table 5. Table of frequent itemsets for common psychological problems

Serial number	Degree of support	Itemset
1	0.7419	{Adaptation and Anxiety}
2	0.6351	{Introversion and Extroversion}
3	0.2512	{Emotional and serene}
4	0.3167	{Cowardice and courage}
5	0.1084	{Adaptation and Anxiety,Emotional and serene}
6	0.0835	{Adaptation and Anxiety,Cowardice and courage}

In order to find the connection between various psychological problems, we produced rules with confidence level above 0.65 and used the confidence level as a benchmark for ranking, and the strong association rules for common psychological problems are shown in Table 6. It can be seen that {Adaptation and Anxiety}→{Emotionality and Serenity and Alertness} is the rule with the highest confidence level, with a confidence level of 0.9132, which indicates that 91.32% of the people who are anxious and maladaptive will be disturbed by their emotional distress, and this rule has the second highest support, with a support level of 0.1927, which indicates that the probability of it occurring is high. {Adaptation and Anxiety} → {Scowardice and Boldness}, with a weighted support of 0.1375, is the rule with higher confidence in the combination of psychological problems with higher support, and this rule has a higher likelihood of occurrence. Meanwhile, {Adaptation and Anxiety}→{Introversion and Extroversion} has the highest support level of 0.1984, and our analysis can conclude that one of the important psychological problems that contribute to the other three major psychological problems among the four major psychological problems is Adaptation and Anxiety.

Table 6. Strong association rules for common psychological problems

Serial number	Cause itemset	The resulting itemset	Weighted support	Weighted confidence
1	{Adaptation and Anxiety}	{Emotional and serene}	0.1927	0.9132
2	{Adaptation and Anxiety}	{Cowardice and courage}	0.1375	0.8807
3	{Adaptation and Anxiety}	{Introversion and Extroversion}	0.1984	0.7213
4	{Introversion and Extroversion}	{Emotional and serene}	0.0903	0.8105
5	{Emotional and serene}	{Introversion and Extroversion}	0.1224	0.6592

5. Conclusion

This paper constructs a mental health education model for colleges and universities based on group intelligence optimization, and takes the data of a college mental health education center as an example for the application analysis of this model.

The K-means clustering algorithm based on the improved bat algorithm is used to cluster and analyze the psychological scale data, and the results show that in terms of adaptation and anxiety, the third category accounts for 6% above average. In terms of introversion vs. extroversion, the first category has a higher than average percentage of 9% and is the most extroverted, compared to the third category where the students are more introverted. In terms of Emotional vs. Safe and Alert, Category 1 has an above average share of 2.5%, while Category 3 students are far below average in terms of Emotional vs. Safe and Alert. In terms of timidity and boldness, the first category has the highest share, while the third category has the lowest share, with a difference of 2.8 percentage points.

The four main psychological problems were correlated using a serial hybrid algorithm based on the ant colony algorithm. {Adaptation and Anxiety} → {Emotionality and Tranquil Resourcefulness} with a confidence level of 0.9132 and a support level of 0.1927. {Adaptation and Anxiety} → {Timidity and Boldness}, which had a weighted support of 0.1375, and {Adaptation and Anxiety} → {Introversion and Extroversion}, which had the highest support of 0.1984. The results showed that one of the important psychological problems that contributed to the other three of the four major psychological problems was adaptation and anxiety, as manifested by low adaptation and high anxiety.

References

1. Chan, V., Moore, J., Derenne, J., & Fuchs, D. C. (2019). Transitional age youth and college mental health. Child and Adolescent Psychiatric Clinics, 28(3), 363-375.
2. Turetsky, K. M., & Sanderson, C. A. (2018). Comparing educational interventions: Correcting misperceived norms improves college students' mental health attitudes. Journal of Applied Social Psychology, 48(1), 46-55.
3. Cornish, P. A., Berry, G., Benton, S., Barros-Gomes, P., Johnson, D., Ginsburg, R., ... & Romano, V. (2017). Meeting the mental health needs of today's college student: Reinventing services through Stepped Care 2.0. Psychological services, 14(4), 428.
4. Lei, X., Liu, C., & Jiang, H. (2021). Mental health of college students and associated factors in Hubei of China. PLoS One, 16(7), e0254183.
5. Zhang, K. Y., & Cerdin, J. L. (2020). The Chinese apprenticeship model: the spirit of craftsmanship. The success of apprenticeships: Views of stakeholders on training and learning, 3, 187-192.
6. Shaobo, M., Hong, S., Peng, W., & Ling, Z. (2023). A Review of Craftsmanship Spirit Research in the New Era. In West Forum on Economy and Management (Vol. 34, No. 6, pp. 79-89).
7. Haijian, L., Ren, X., & Ran, L. (2016). The Spirit of Craftsmanship and Industrial Civilization. China Economist, 11(4), 68.
8. Xu, E. (2018). Hecheng Li: sticks with the spirit of craftsmanship. Journal of Thoracic Disease, 10(Suppl 11), S1290.
9. Miseliunaite, B., Kliziene, I., & Cibulskas, G. (2022). Can holistic education solve the world's problems: A systematic literature review. Sustainability, 14(15), 9737.
10. Mukhamedovich, S. U., Haydarovich, A. A., Kholmuratovna, E. O., & Toshpulatovich, Y. B. (2023). The Question of Raising Young People in The Spirit of Craftsmanship. Journal of Advanced Zoology, 44.
11. Luo, Q., Wang, C., & Zhao, Y. (2020, June). Cultivation strategy of college students' craftsman spirit from the perspective of artificial intelligence. In Journal of Physics: Conference Series (Vol. 1575, No. 1, p. 012173). IOP Publishing.
12. Zhang, Z. (2021). Cultivation of craftsman spirit in employment and entrepreneurship education in colleges. Journal of Frontiers in Educational Research, 3(4), 94-97.
13. Chen, N. (2023). Research On Vocational Ideological and Political Education Based on Craftsman Spirit. International Journal of Vocational and Technical Education, 3(1), 7.
14. Hu, W., Zhang, X., & Liu, Y. (2022). THE GUIDING ROLE OF THE INTEGRATION OF CRAFTSMAN SPIRIT AND IDEOLOGICAL AND POLITICAL EDUCATION OF COLLEGE STUDENTS ON STUDENTS'POSITIVE PSYCHOLOGY. Psychiatria Danubina, 34(suppl 4), 795-795.
15. Li, L., Weiqi, S., Bin, Z., Feng, G., Suqin, Z., Xingxiang, G., & Kun, L. (2021, June). Exploration on the Effective Integration of Craftsman Spirit and Curriculum Ideology and Politics Education. In 2021 2nd International Conference on Mental Health and Humanities Education (ICMHHE 2021) (pp. 335-338). Atlantis Press.
16. Wang, L. (2022). THE GUIDING EFFECT OF CRAFTSMAN SPIRIT ON STUDENTS'VALUE COGNITION PSYCHOLOGY IN IDEOLOGICAL AND POLITICAL EDUCATION. Psychiatria Danubina, 34(suppl 4), 500-500.
17. Yao, L. (2020). Research on the Improvement of College Students' Innovation and Entrepreneurship Ability Based on the Cultivation of Craftsman Spirit. Frontiers in Educational Research, 3(14).