Paradigm Academic Press
Innovation in Science and Technology
ISSN 2788-7030
NOV. 2023 VOL.2, NO.6

The Design of College Students' Mental Health Analysis System Based on Human-Computer Interaction

Xinyuan Hu[1#] & Naijia Ye[1#]

[1] University of Sussex, Sussex House, Brighton, BN1 9RH, United Kingdom

Correspondence: Xinyuan Hu, University of Sussex, Sussex House, Brighton, BN1 9RH, United Kingdom.

[#] These authors shared first authorships.

doi:10.56397/IST.2023.11.06

Abstract

This paper presents a machine learning-based system for analyzing the mental health of college students. The system utilizes data mining techniques to analyze and process psychological data, enabling personalized mental health assessment and guidance. Firstly, the paper introduces the basic concepts and steps of data mining, as well as the system architecture of the data warehouse. Then, it discusses the methods of incorporating clustering, anomaly mining algorithms, and association rules into the analysis of psychological data in college students. Next, the paper provides a detailed description of the overall structure and workflow of the machine learning-based system for analyzing the mental health of college students. The mental health assessment model utilizes evaluation criteria determination and weight assignment methods. Finally, the accuracy and effectiveness of the system are validated through performance testing. This system provides college students with scientifically feasible mental health assessment and guidance, which has significant practical implications for addressing mental health issues among college students.

Keywords: artificial intelligence, mental health of college students, analyzing system

1. Introduction

In recent years, the mental health problems of college students have increasingly attracted widespread attention in society. With the rapid development of information technology, artificial intelligence has gradually penetrated various fields, providing new possibilities for addressing mental health issues among college students. Artificial intelligence-based mental health analysis systems have the ability to utilize data mining techniques to uncover underlying patterns and knowledge within psychological data of college students. Based on individual characteristics and needs, these systems can provide personalized assessment and guidance. This paper aims to design an artificial intelligence-based system for analyzing the mental health of college students to meet their psychological needs. Firstly, the paper lays a foundation by discussing the basic concepts and steps of data mining, which are essential for the subsequent system design. Secondly, it explores the applications of clustering, anomaly mining algorithms, and association rules in the analysis of psychological data among college students, extracting valuable mental health information. Finally, through performance testing, the accuracy and effectiveness of the system in assessing and guiding the mental health of college students are validated. The results of this research will provide scientifically feasible solutions to address mental health issues among college students, offering them personalized psychological support and assistance. Additionally, it will provide new insights and methods for the application of artificial intelligence in the field of mental health (Wang T & Park J., 2021).

2. Overview of Data Mining for Artificial Intelligence

2.1 Basic Concepts of Data Mining

2.1.1 Definition of Data Mining

Data mining is a process of discovering underlying patterns, relationships, and knowledge from large-scale datasets. By employing techniques such as statistics, machine learning, and pattern recognition, data mining can reveal hidden information within data and uncover useful trends and patterns. The objective of data mining is to extract meaningful knowledge from massive data in an automated manner, which can support decision-making, problem-solving, and future predictions. Data mining finds widespread applications in various fields, such as market research, financial risk analysis, medical diagnosis, and social media analytics. Through data mining, we can unearth valuable insights that are hidden within data, providing better decision-making and actionable guidance to individuals. The definition of data mining encompasses not only the discovery of patterns and knowledge within data but also emphasizes the automation and practicality of the process (Li D., 2022).

2.1.2 Basic Steps in Data Mining

The basic steps of data mining can vary according to different domains and data processing needs, giving flexibility and room for play. In a narrow sense, data mining is one of the key aspects of knowledge discovery, usually as an important step in the knowledge discovery process (KDD). The KDD process generally consists of three parts: data organization, data mining, and interpretation and evaluation of the results. Its specific steps are shown in Figure 1:

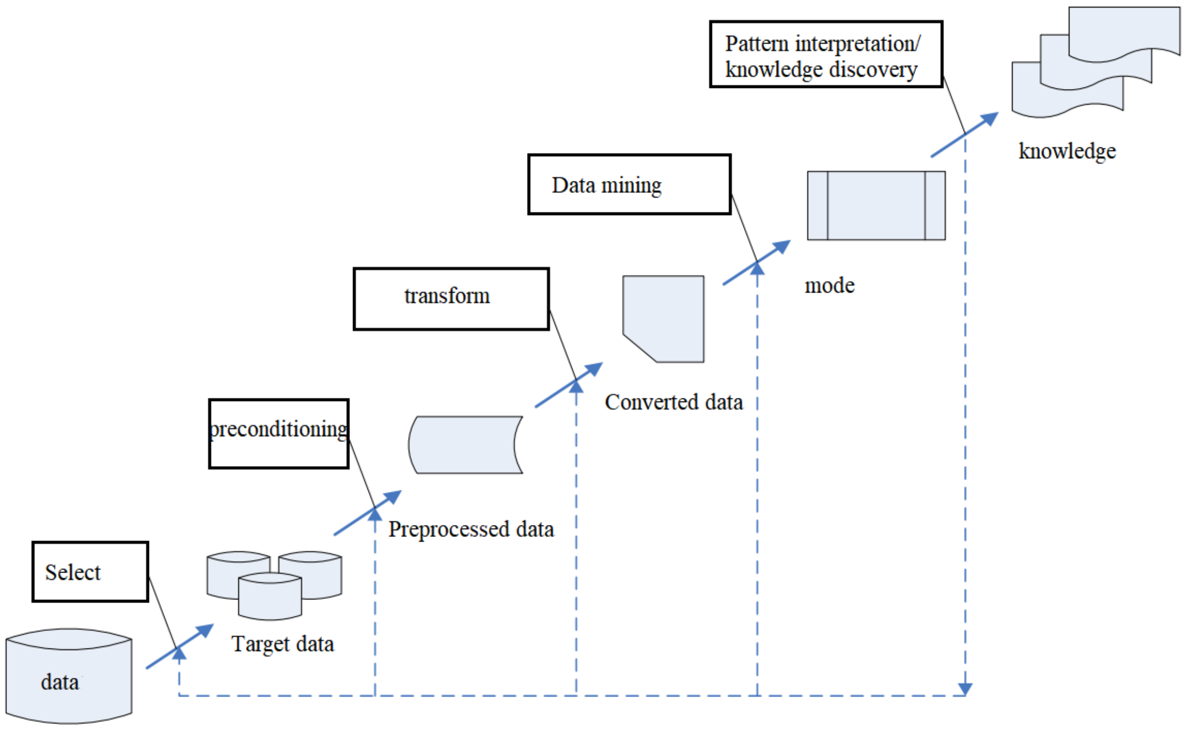

Figure 1. KDD Steps

Firstly, data preprocessing is a necessary step in the KDD process. It involves data acquisition, data cleaning, data integration, and data transformation processes. Data acquisition involves data collection and consolidation, which may involve multiple data sources and different data types. The goal of data cleaning is to remove noise, handle missing values and outliers, and ensure data consistency and accuracy. Data integration involves consolidating data from different sources, eliminating duplicates and redundancies. Data transformation may include operations like feature selection, feature extraction, data normalization, and data transformation to better fit the requirements of mining algorithms. Secondly, data mining is the core step in the KDD process, which includes data exploration, model construction, and model evaluation processes. The data exploration stage discovers patterns and trends in data through methods like visualization, statistical analysis, and clustering. The model construction stage involves selecting appropriate mining algorithms such as classification, clustering, and association rules to build models and perform modeling and training on the data. The model evaluation stage validates and examines the predictive performance and generalization ability of the models to ensure their reliability and effectiveness. Lastly, result interpretation and evaluation are the final steps in the KDD process (Wang C., 2023). It involves interpreting, understanding, and evaluating the data mining results. The interpretation stage focuses on explaining and understanding the results, deriving insights and conclusions about the data. The evaluation stage assesses the effectiveness and practicality of the mining results based on the opinions of domain experts and practical application requirements. Additionally, the mining process can be reviewed and summarized to provide feedback and improvements for the next round of the KDD process.

In conclusion, the basic steps of data mining include data preprocessing, data mining, and result interpretation and evaluation. Each step plays an important role, interconnected and collectively driving the process of knowledge discovery. The success of data mining depends on careful planning and implementation of each step, as well as effective interpretation and evaluation of the results (Du C, Liu C, Balamurugan P, et al., 2021).

2.2 System Architecture of the Data Warehouse

A complete data warehousing system typically consists of data sources, data warehouse, data mart, OLAP server, and front-end analysis tools. These components can be divided into three hierarchical layers: the warehouse data server layer, the OLAP server layer, and the front-end client layer. The data warehouse structure diagram is shown in Figure 2.

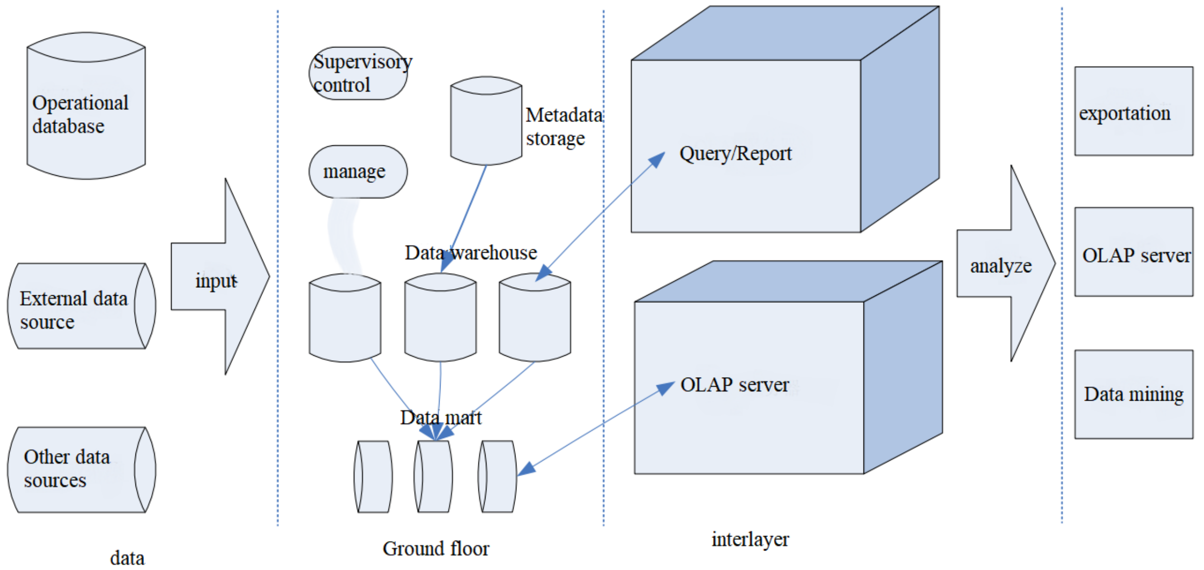

Figure 2. Data warehouse structure

At the bottom layer, the warehouse data server typically uses a relational database system for data storage and management. The relational database system can provide stable and reliable data storage and support transaction processing and query operations. Data sources extract raw data from different sources through the ETL (Extract, Transform, Load) process and store it in the data warehouse after undergoing cleaning, transformation, and loading operations.

The middle layer is the OLAP (Online Analytical Processing) server layer, which is the core component of the data warehousing system. Typically, the OLAP server can be implemented using either a relational OLAP model or a multidimensional OLAP model. The relational OLAP model is based on a relational database and uses SQL query language for multidimensional analysis. The multidimensional OLAP model revolves around multidimensional data cubes and improves query performance by precomputing and storing aggregated data. The OLAP server provides fast data analysis and query capabilities, supporting user-friendly operations such as data slicing, drilling, and pivoting (Rezapour M & Elmshaeuser S K., 2022).

The top layer Is the front-end client layer, which includes query and reporting tools, analysis tools, and data mining tools. These tools assist users in data analysis, report generation, and decision support through intuitive user interfaces. Query and reporting tools provide basic data querying and reporting functions, analysis tools support more complex data analysis and visualization, while data mining tools help users discover patterns and trends hidden in the data.

3. Combining Data Mining Algorithms in Psychological Data Analysis for College Students

3.1 Clustering Method

3.1.1 Segmentation Methods

The segmentation method is a commonly used data clustering technique for dividing a dataset containing n objects into k clusters. Each cluster contains at least one more object and each object can belong to only one cluster. The method is implemented by following steps:

(1) Set the parameter k, which represents the number of clusters. Determine the goal of partitioning the dataset into k clusters based on the application scenario and requirements.

(2) Perform an initial simple partition. Allocate the objects in the dataset to the initial k clusters according to predefined initial partitioning rules. Common initial partitioning methods include random partitioning and distance-based or similarity-based partitioning (Nag A, Das A, Sil R, et al., 2022).

(3) Use iterative relocation techniques to adjust and optimize each cluster. By comparing the similarity between objects and their current and other clusters, decide whether to reassign objects to more similar clusters. This process is repeated iteratively until the membership of each cluster no longer changes, reaching a convergence state.

(4) After the iterations, obtain the final partitioning result. At this point, the objects within each cluster have high similarity, while there is a significant difference between objects in different clusters. This aligns with the ultimate goal of clustering, which is to achieve tight intra-cluster cohesion and distinctiveness among different clusters.

3.1.2 Hierarchical Methodology

Hierarchical clustering (shown in Figure 3) is a commonly used clustering method, which mainly includes cohesive hierarchical clustering and split hierarchical clustering. These two methods differ in the decomposition strategy of clustering.

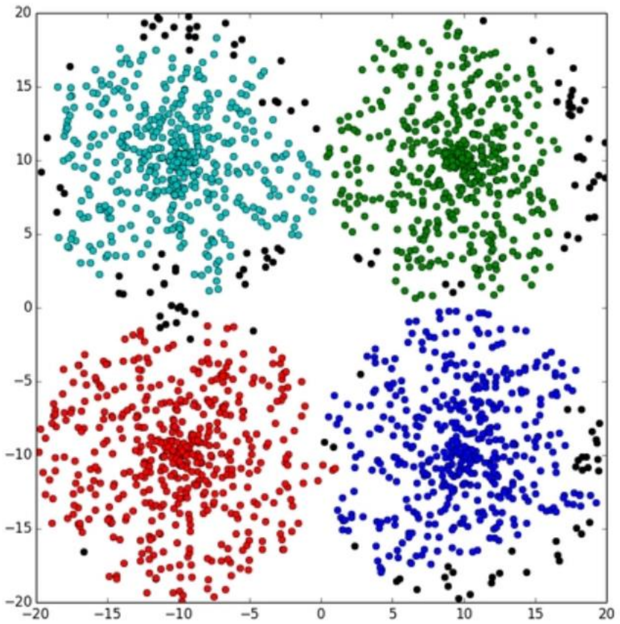

Figure 3. Hierarchical clustering approach

Agglomerative hierarchical clustering is a bottom-up hierarchical decomposition strategy. It starts by initially assigning each object to a separate cluster, and then gradually merges these clusters using an appropriate algorithm until certain conditions are met. The merging process is typically based on the similarity or distance measurement between clusters, selecting the clusters with the highest similarity to merge. This step-by-step merging process forms a hierarchical structure, ultimately resulting in a complete clustering result. Divisive hierarchical clustering, on the other hand, is a top-down hierarchical decomposition strategy. It starts by initially assigning all objects to one cluster, and then iteratively splits this cluster into smaller subclusters using an appropriate algorithm until certain conditions are met. The splitting process is usually based on the intra-cluster dissimilarity or the size of the clusters, selecting the cluster to be split. Through continuous splitting, a hierarchical clustering structure is obtained (Li F, Gu L & Xu H., 2022).

3.1.3 Grid-Based Approach

Grid-based methods are a commonly used clustering algorithm that partitions the data space into grid cells for clustering. This approach divides the dataset into a regular grid structure, simplifying the clustering process and improving efficiency. Here are several representative grid-based clustering algorithms:

(1) STING algorithm (Statistical Information Grid): The STING algorithm uses grid-based multi-resolution clustering techniques, dividing the spatial region into a specified number of multi-level rectangular units. The granularity of the grid can impact the final clustering result, so it is necessary to select an appropriate granularity for better clustering performance.

(2) WaveCluster algorithm (Clustering using Wavelet Transform): The WaveCluster algorithm is a grid-based and density-based clustering method. It uses wavelet transform to cluster grids at different levels of granularity, making it capable of handling clusters of different shapes. Additionally, the WaveCluster algorithm is not affected by outliers, maintaining good clustering results.

(3) CLIQUE algorithm (Clustering in High-dimensional Subspaces): The CLIQUE algorithm is suitable for high-dimensional datasets and demonstrates good clustering performance. It discovers the data distribution of the entire space by distinguishing sparse and dense regions in space, and identifies the highest-dimensional subspaces. The CLIQUE algorithm can automatically adapt to different subspaces and perform efficient clustering analysis (Budler L C, Gosak L & Stiglic G., 2023).

3.2 Anomaly Mining Algorithms

Anomaly detection algorithms are a class of techniques used to identify and discover anomalous behavior, outliers, or abnormal patterns. In many real-world applications, anomalous data often carries significant importance, including data anomalies, network attacks, financial fraud, and more. Here are several common anomaly detection algorithms:

(1) Statistical-based methods: These methods assume that normal data follows a certain statistical distribution and establish a statistical model for normal data based on the assumption. The deviation between actual data and the model is then compared to determine if it is an anomaly. Common statistical methods include z-score, box plots, etc.

(2) Cluster-based methods: These methods use clustering techniques to partition the data into different clusters, and then determine if a data point is an anomaly by calculating its distance or similarity to its cluster. If the distance between a data point and its assigned cluster is large or the similarity is low, it may be considered an anomaly. Common clustering methods include k-means clustering, DBSCAN, etc.

(3) Density-based methods: These methods assume that normal data is distributed in high-density areas, while anomalous data is distributed in low-density areas. Anomalies are determined by calculating the density of data points in their vicinity. LOF (Local Outlier Factor) and OPTICS (Ordering Points to Identify the Clustering Structure) are commonly used density-based anomaly detection algorithms.

(4) Machine learning-based methods: These methods utilize machine learning techniques to train models for identifying normal and anomalous data. Common algorithms include Support Vector Machines (SVM), Random Forests, Neural Networks, etc. By training the models, these methods are able to extract potential anomalous patterns from a large number of features.

(5) Time series-based methods: These methods are specifically designed to handle anomalies in time series data. They detect anomalies by analyzing the trends, periodicity, and outliers in the time series. Common time series anomaly detection algorithms include ARIMA models, exponential smoothing, and outlier detection algorithms. (Mengi M & Malhotra D., 2022)

4. Artificial Intelligence-Based Mental Health Analysis System Design for College Students

4.1 Overall System Structure

The college student psychological health analysis system is a system consisting of five core modules, including data collection, data preprocessing, anomaly detection, association rules, and psychological testing. The system typically adopts a three-tier architecture, including the database layer, server layer, and client layer. The database layer is used to store and manage data, the server layer handles logic and algorithms, and the client layer provides the user interface. This architecture ensures system stability and scalability while providing a convenient user experience. The data collection module is used to gather multi-source psychological health data, the data preprocessing module cleans and transforms raw data, the anomaly detection module identifies abnormal behavior, the association rules module discovers data relationships, and the psychological testing module evaluates the user’s psychological state. Through the collaborative work of these modules, the system can provide accurate analysis and support for the psychological health of college students. Figure 4 shows the three-tier architectural diagram of the college student psychological health intelligent analysis system.

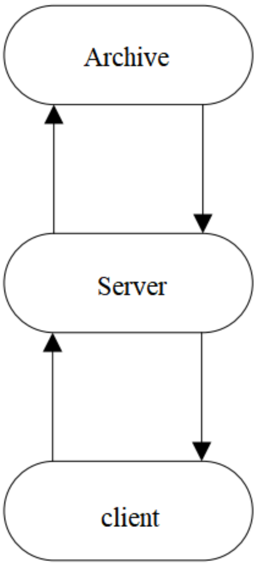

Figure 4. Three-layer structural relationship diagram of the intelligent analysis system for college students' mental health

4.2 System Workflow

The workflow of the college students' mental health analysis system can be divided into the foundation layer, data layer, analysis layer, decision-making layer and application layer. Its system workflow is shown in Figure 5:

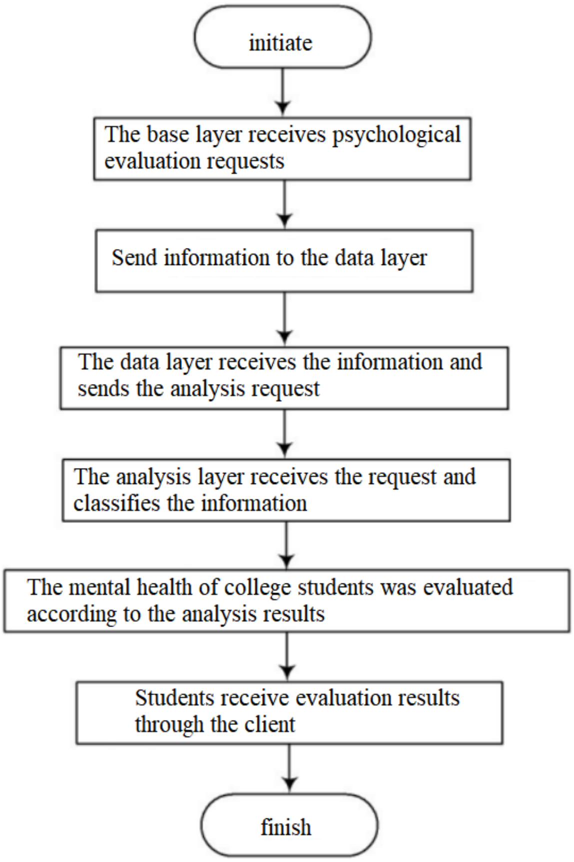

Figure 5. System workflow

4.3 Mental Health Evaluation Model

4.3.1 Determination of Evaluation Indicators

In the process of evaluating college students' psychological health, this paper chooses the SCL-9 Psychological Health Measurement Scale as the assessment standard and selects indicators such as personality (G1), willpower (G2), emotion (G3), depression (G4), anxiety (G5), and psychotic symptoms (G6) for evaluation. The target layer is psychological health assessment, represented by the symbol K. To establish a mathematical model, this paper collected psychological test data from a university's students over the past three years and selected 50 samples for statistical analysis. The statistical results are shown in Table 1, based on the scoring method of the SCL-90 scale.

Table 1. SCL-90 Scale Scores

index	Sample number					Mean value	variance
	1	2	3	...	50		
G1	1.44	1.78	1.43	...	1.19	1.52	0.58
G2	1.41	1.71	1.81	...	3.31	1.85	0.69
G3	1.67	1.11	1.44	...	2.89	1.93	0.69
G4	1.65	1.11	1.42	...	2.19	1.83	0.76
G5	1.11	1.82	1.25	...	2.25	1.56	0.55
G6	1.44	1.91	1.61	...	2.44	1.78	0.65

The statistical results in Table 1 will provide a basis for further analysis and discussion in this study in order to gain a deeper understanding of the characteristics and problems of college students' mental health and to provide a scientific basis for related decision-making and intervention (Klos M C, Escoredo M, Joerin A, et al., 2021).

4.3.2 Setting Evaluation Indicator Weights

(1) Modeling the judgment matrix

1) In constructing the judgment matrix model, we used the 1 to 9 scale and the inverse determination method to compare a set of indicators two by two to determine the relationship of importance between them. Based on the comparison results, we constructed a judgment matrix, and the specific construction results are shown in Table 2.

Table 2. Judgement matrix

	G1	G2	G3	G4	G5	G6
G1	1	3	5	7	9	11
G2	1/3	1	3	5	7	9
G3	1/5	1/3	1	3	5	7
G4	1/7	1/5	1/3	1	3	5
G5	1/9	1/7	1/5	1/3	1	3
G6	1/11	1/9	1/7	1/5	1/3	1

By using a scale of 1 to 9, we can compare each indicator with the others and determine the relative importance between them using inverse measurements. This approach allows us to quantify the level of importance and translate it into numerical values for analysis. The results of the judgment matrix in Table 2 will provide us with a visual reference to better understand the relationship of importance between indicators and to inform the subsequent modeling and decision-making process.

2) Consistency test

In performing the consistency test, we first find the eigenroot solution of the judgment matrix Z through equation (1) and normalize it to get the importance weight values between the indicators corresponding to the same level and other indicators. Then, we carry out the consistency test on the judgment matrix, which is calculated according to equation (1).

$$ZW = \lambda_{\max} W \tag{1}$$

For consistency test, we first calculate the critical indicator value CI for each matrix and find the random consistency indicator RI at the same time. By calculating, we can get the random consistency CR as shown in equation (2).

$$CR = \frac{CI}{RI} \tag{2}$$

When the CR value is less than or equal to 0.1, it indicates that the hierarchical single sort structure has a more appropriate consistency. When the CR value is greater than 0.1, it is necessary to reacquire the index value of the matrix, as shown in Equation (3).

$$\begin{cases} \lambda_{\max} = \sum_{i=1}^{n} \frac{(ZW)_i}{(ZW_i)} \\ CI = \frac{R_{\max} - n}{n-1} \\ CR = \frac{CI}{RI} \end{cases} \tag{3}$$

Through the consistency test, we can ensure that the judgment matrix has better consistency and reliability, so as to provide accurate weighting and prioritization for the subsequent multi-layer judgment and decision-making process. This allows for more effective decision making and resource allocation to achieve the desired goals and results.

3) Identify mental health evaluation models

In order to determine the mental health evaluation model, this paper adopts a multi-objective linear weighted summation comprehensive evaluation model. The model can help us obtain the weight of each indicator and calculate the comprehensive evaluation value and assessment value of the target layer. The specific representation is shown in equation (4).

$$\begin{cases} K = \sum_{i=1}^{n} D_i W_i \\ D_i = \sum_{i=1}^{n} E_i r_{ij} \end{cases} \tag{4}$$

In the multi-objective linear weighted summation comprehensive evaluation model, the weight of each index plays a key role. The determination of weights is usually based on expert evaluation, statistical analysis and other methods. Through these methods, we are able to quantify the importance of each indicator and weight them according to the actual needs. Taken together, this multi-objective linear weighted summation comprehensive evaluation model provides a comprehensive and accurate method to assess mental health status. It not only takes into account the differences and weights of each indicator, but also is able to give the comprehensive evaluation value and assessment value of the target level, which helps decision makers to better understand and manage mental health issues. Equation (4) specifically represents the calculation of this model, which provides a scientific basis for mental health evaluation.

5. Conclusion

This paper designs an analysis system for college students' psychological health based on artificial intelligence technology. Through the application of data mining algorithms, the system can discover potential patterns and knowledge from a large amount of psychological data, providing personalized psychological health assessment and guidance for college students. After performance testing, the system has demonstrated good accuracy and effectiveness. In summary, this research provides innovative solutions for practical issues in the field of college students' psychological health, with important practical value and application prospects. Further research and practice will contribute to the development and improvement of college students' psychological health work.

References

Budler L C, Gosak L, Stiglic G., (2023). Review of artificial intelligence-based question-answering systems in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13*(2), e1487.

Du C, Liu C, Balamurugan P, et al., (2021). Deep learning-based mental health monitoring scheme for college students using convolutional neural network. *International Journal on Artificial Intelligence Tools, 30*(06n08), 2140014.

Klos M C, Escoredo M, Joerin A, et al., (2021). Artificial intelligence-based chatbot for anxiety and depression in university students: pilot randomized controlled trial. *JMIR formative research*, *5*(8), e20678.

Li D., (2022). Artificial Intelligence Based Face Image Recognition Research on the Impact of Campus Security System Design on Audience Psychology. *Psychiatria Danubina, 34*(suppl 4), 601-601.

Li F, Gu L, Xu H., (2022). The Mining Method of Ideological and Political Elements in University Public Mental Health Courses Based on Artificial Intelligence Technology. *Journal of Environmental and Public Health*, 2022.

Mengi M, Malhotra D., (2022). A systematic literature review on traditional to artificial intelligence based socio-behavioral disorders diagnosis in India: Challenges and future perspectives. *Applied Soft Computing*, 109633.

Nag A, Das A, Sil R, et al., (2022). Application of artificial intelligence in mental health // International Conference on Intelligent Systems Design and Applications. Cham: Springer Nature Switzerland, 128-141.

Rezapour M, Elmshaeuser S K., (2022). Artificial intelligence-based analytics for impacts of COVID-19 and online learning on college students' mental health. *PLoS One, 17*(11), e0276767.

Wang C., (2023). Application of MPP database and artificial intelligence system in online evaluation of college students' mental health. *Preventive Medicine*, 107580.

Wang T, Park J., (2021). Design and implementation of intelligent sports training system for college students' mental health education. *Frontiers in psychology, 12*, 634978.