International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 5498-5510

https://doi.org/10.70517/ijhsa463475

Bayesian network-based prediction of students' mental health status and the design of college interventions

Jing Zhao[1,*], Gang Wang[2], Yongning Qian[3] and Yifan Xue[4]
[1] School of Accounting, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China
[2] The Office of Student Affairs, College Student Employment Guidance Center, School of Innovation and Entrepreneurship, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China
[3] The School of Humanities and Arts, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China
[4] The Academic Affairs Office, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China
Corresponding authors: (e-mail: 18700024444@163.com).

Abstract Currently, the mental health problems of students in colleges and universities are becoming more and more prominent, this study constructed a Bayesian network-based prediction model of college students' mental health status and designed corresponding interventions. In terms of methodology, firstly, multi-dimensional student behavioral features were represented and extracted, including consumption features (dietary regularity, diligence, number of meals shared) and Internet access features (length of time on the Internet, downtime, traffic use); secondly, the Jenks Natural Breaks algorithm was used to label the featured data, and the Apriori algorithm was used to mine the behavioral association rules of the psychologically healthy and psychologically abnormal students; then a Bayesian network model was built to predict the mental health status of students and design interventions accordingly. Then a Bayesian network model is established to predict students' mental health status, and the results are compared with those of decision tree, support vector machine, and boosting algorithm. The results show that the Bayesian network prediction model has the best performance, with an accuracy of 0.9415, a recall of 0.9387, and an F1 value of 0.9389, which are higher than those of the other three algorithmic models in the anxiety binary classification experiment; and in the anxiety multiclassification prediction experiment, the Bayesian network model has an Fmacro value of 0.8549, and an Fmicro value of 0.8814, which are also better than the other models. The study also found that the group of psychologically abnormal students is usually characterized by less regular diet, less diligence, fewer number of people sharing meals, longer time on the Internet and more traffic use, and later time off the Internet on weekdays. The Bayesian network prediction model constructed in this study has high accuracy in predicting the mental health status of college students, which can provide technical support for mental health monitoring and precise intervention in colleges and universities.

Index Terms Bayesian network, student mental health, state prediction, behavioral characteristics, feature extraction, intervention measures

I. Introduction

According to statistics released on the official website of the Ministry of Education, the total number of students enrolled in colleges and universities in 2024 has exceeded 47.63 million. According to the rate of 0.6% of the incidence of mental illness in clinical population, the number of college students with mental illness is expected to be more than 280,000 [1]. College students are in their prime, in the period of plucking and pregnancy, their physiology and psychology tend to be mature, and they are open-minded, independent, and courageous to compete. However, due to the influence of learning environment, interpersonal relationship, love and emotional frustration, employment pressure and other factors, the mental health problems of college students are becoming more and more prominent, especially depression, anxiety, psychological disorders and even suicidal tendency are not uncommon among college students, which is a concern to all walks of life [2]-[5]. Some researchers have proposed that the mental health status of the human population can be seriously affected after a major disaster [6]. The outbreak of 2019 novel coronavirus infection in early 2020 was a major public health emergency. After the new coronavirus epidemic, the situation of mental health monitoring and guidance and control work in colleges and universities has become even more severe, with a large student base, climbing abnormality rates, and a teacher-student ratio of 1:400 or even below the reality, the current stage of mental health work is still characterized by a large workload, fatigue work, and low efficiency, which urgently needs to be explored as a new path to solve the problem [7]-[10].

Currently, mental health work in colleges and universities mainly uses psychological scales as a measurement tool to identify key populations [11]. Scale measurement scores reflect characteristics such as susceptibility to anxiety, depression, and a tendency toward stronger emotional reactions [12]. Although the scale and timeliness of using psychometric scales to measure the mental health of populations is remarkable, its inherent shortcomings have reached a basic consensus in the field [13]. The first is the debate over reliability and validity, the second is the non-objectivity of the measurements in the context of feelings of morbidity and shame, and the last is the inequality between the cost of the measurements and the utilization of the measurements [14]. Moreover, psychological counseling in colleges and universities follow the principle of student voluntarism, so this work is prone to work blind spots. Therefore, predicting and intervening in the mental health status of college students is a problem that needs to be solved urgently by the current society and colleges and universities.

Mental health prediction means based on artificial intelligence algorithms can assist and replace manual labor to a certain extent. Literature [15] used logistic regression method and R4.2.3 software to analyze the factors affecting the mental health status of college students and and design a mental health risk prediction nomogram model, respectively. Literature [16] applied the improved seagull optimization algorithm to improve the back-propagation neural network to produce a prediction model of mental health status in Dasher province, which has better reliability compared to the logistic regression model. Literature [17] proposed XGBoost, Random Forest, Decision Tree and Logistic Regression prediction models with excellent performance in identifying whether college students suffer from anxiety and depression. Literature [18] obtained data related to the mental health status of college students through data mining techniques and used neural network algorithms to assess, predict and correlate them. Literature [19] provided three radial basis function neural network-based risk prediction models for suicidal ideation with an accuracy rate higher than 90%, and pointed out that the presence of past suicidal intent and poor sleep quality are important early warning factors for suicidal ideation. However, the mental health status is difficult to be predicted accurately due to complex interaction behavior, causality, computational complexity, logistic regression, and other machine learning and deep learning performed by neural networks.

Among the proposed psychological interventions, literature [20] revealed the effectiveness of various psychological interventions such as cognitive-behavioral therapy, interpersonal therapy, and localized interventions to improve depression among college students. Whereas, literature [21] compared the effectiveness of web-based cognitive behavioral therapy in psychotherapeutic interventions in guided and self-directed formats, as well as against conventional interventions, yielding the result of guided > self-directed > conventional interventions. Literature [22] designed an online mental health and study skills support system, which was evaluated for its ease of use, professionalism, and efficiency by application, and the mental health training techniques in it could enhance students' thinking and self-confidence, which could help in online interventions for students' mental health problems. Literature [23] developed a fuzzy augmented predictive neural system incorporating a neural network with backpropagation and deep fuzzy, which can be used to assess the mental health problems of college students and provide personalized intervention plans. Literature [24] describes a PDNN (Photonic Deep Neural Network)-supported mental health prediction model for college students and a methodology for developing a psychological intervention system, and the intervention system helps to improve psychological problems such as anxiety and depression. These interventions have some applicability, but psychological problems are often formed over a long period of evolution and are accompanied by dynamic fluctuations of various uncertainties, and it is clear that the above methods do not have such an exploratory function. Bayesian network (BN) is an effective probabilistic graphical model that represents variables with nodes, dependencies between variables with directed edges, and the strength of relationships between variables with probabilities [25].The main advantage of BN is that it can handle uncertainty, complexity, and incomplete information. Literature [26] assessed the mental state of students using probabilistic deep belief BN, and the algorithm used in the study had higher classification accuracy and assessment results compared to convolutional neural networks and deep neural networks. Literature [27] revealed depression and anxiety co-morbid symptoms through network analysis and clump penetration, and combined with BN to reveal the causal relationship and network structure between the two.

According to statistics released on the official website of the Ministry of Education, the total number of students enrolled in colleges and universities in 2024 has exceeded 47.63 million. According to the 0.6% incidence rate of mental illness in the clinical population, the number of college students with mental illness is expected to exceed 280,000 people. College students are in their prime, in the period of plucking and pregnancy, their physiology and psychology tend to be mature, their thoughts are open and independent, and they have the courage to compete. However, due to the influence of learning environment, interpersonal relationship, love and emotional frustration, employment pressure and other factors, the mental health problems of college students are becoming more and more prominent, especially depression, anxiety, psychological disorders and even suicidal tendency are not uncommon among college students, which is a concern for all walks of life. Some researchers have proposed that the mental health status of the human population will be seriously affected after a major disaster.The outbreak of the 2019 novel coronavirus infection epidemic in early 2020 is a major public health emergency. After the new coronavirus epidemic, the situation of mental health monitoring and guidance and control work in colleges and universities has become even more severe, with the reality of a large student base, climbing abnormality rates, and a teacher-student ratio of 1:400 or even less, the current stage of mental health work is still characterized by a large amount of workload, fatigue work, and low efficiency, which needs to be explored and solved by new paths. Currently, mental health work in colleges and universities mainly adopts psychological scale as a measurement tool to determine the key population. Scale measurement scores reflect characteristics such as easy anxiety, depression and stronger tendency to emotional reactions. Although the scale and timeliness of using psychological scales to measure the mental health of populations is remarkable, its inherent shortcomings have reached a basic consensus in the field. The first is the debate over reliability and validity, the second is the non-objectivity of the measurements in the context of feelings of morbidity and shame, and the last is the inequality between the cost of the measurements and the utilization of the measurements. Moreover, psychological counseling in colleges and universities follow the principle of student voluntarism, so this work is prone to work blind spots. Therefore, predicting and intervening in the mental health status of college students is an important problem that needs to be solved by society and colleges and universities at present.

This study proposes a Bayesian network-based model for predicting students' mental health status. Firstly, we constructed a framework for students' psychological modeling and extracted students' behavioral characteristics from multiple dimensions, including consumption characteristics such as dietary regularity, diligence, and the number of shared meals, and Internet characteristics such as Internet duration, downtime, and traffic use. Then the feature data are labeled by Jenks Natural Breaks algorithm, and Apriori algorithm is applied to mine the association rules of behavioral features of mental health and mental abnormal student groups. Based on the association relationship between behavioral features and mental health status, a Bayesian network model was established to predict students' mental health status. In order to verify the effectiveness of the model, the Bayesian network model is compared with the decision tree, support vector machine, boosting algorithm and other models in the comparison experiments, and the performance of the model is evaluated by the indexes such as accuracy, recall and F1 value. Finally, based on the prediction results, precise interventions for students with different mental health status are proposed.

II. Mental health state prediction model construction

II. A.Framework for psychological modeling of students

For the background and current situation of mental health screening of college students at this stage, this paper proposes a mental health state prediction model framework as shown in Figure 1, college students' behavioral characteristics mining and modeling analysis framework, the core content of the framework is the use of machine learning algorithms to extract behavioral characteristics of the data for learning modeling, the mental health state of college students to predict, and based on the prediction results to make accurate interventions predictions [28].

II. B.Multi-dimensional Student Behavioral Feature Representation and Extraction

II. B. 1) Digital representation of behavioral characteristics

Since only quantitative data operations can be performed in computers, the collected behavioral features need to be mapped to numerical types. In this paper, the collected behavioral features are proposed to be mapped into a behavioral feature vector scheme for data format unification. Firstly, the definition of e.g. expression is given:

$$F=\left(f_1,f_2,f_3,\cdots,f_n\right)F^{\alpha}=\left(f_1^{\alpha},f_2^{\alpha},f_3^{\alpha},\cdots,f_m^{\alpha}\right) \tag{1}$$

where the set F^{α} denotes all behavioral features collected from current college students; the set F^{α} denotes the behavioral features included for a particular student a; n denotes the number of all behavioral features; and m denotes the number of behavioral features for a particular individual. These behavioral features are mapped into a feature vector:

$$V=\left(v_1,v_2,v_3,\cdots,v_n\right) \tag{2}$$

According to the definition of equation (2), a particular behavioral feature vector such as $V=\{0,1,0,0,1,0,1,\cdots\}$, 1 means that the feature represented by that position is present in the current student, while 0 means that the feature is not present in the current student [29].

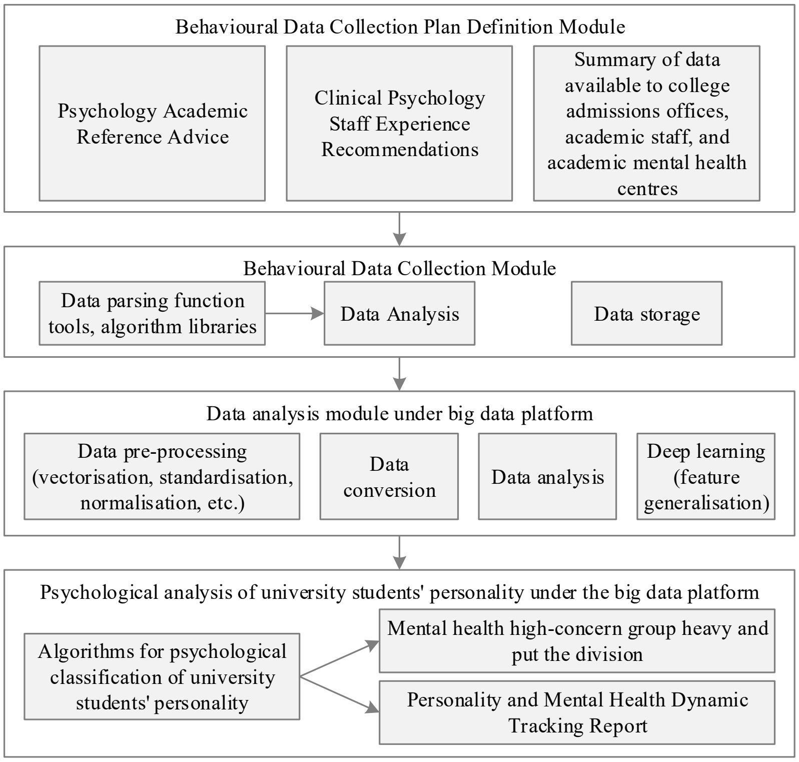
Behavioural Data Collection Plan Definition Module
Psychology Academic Reference Advice
Clinical Psychology Staff Experience Recommendations
Summary of data available to college admissions offices, academic staff, and academic mental health centres
Behavioural Data Collection Module
Data parsing function tools, algorithm libraries
Data Analysis
Data storage
Data analysis module under big data platform
Data pre-processing (vectorisation, standardisation, normalisation, etc.)
Data conversion
Data analysis
Deep learning (feature generalisation)
Psychological analysis of university students' personality under the big data platform
Algorithms for psychological classification of university students' personality
Mental health high-concern group heavy and put the division
Personality and Mental Health Dynamic Tracking Report

Figure 1: Psychological modeling analysis framework

II. B. 2) 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 we extracted the features of students' consumption data to explore its relationship with students' mental health status.

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

The formula for calculating students' dietary regularity is:

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

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, the student diligence S_{DC} is calculated as:

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

where T is the total number of days the student swiped his/her card, and t_j is the time when the student swiped his/her card for the 1st meal on the j day.

3) The number of people who share a meal. It is assumed that if two students are in the same class, and they swipe their cards for consumption 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. The author counted the number of students eating together in each cafeteria window of each college on each campus, in order to prevent some students from paying by cell phone, ordering takeaway or eating out on the results of the experiment, the number of students who

swiped their cards for consumption in the cafeteria in one month was less than 30 records were removed, and then the statistics obtained from each cafeteria window were added according to the students' student numbers, and then the final results were obtained. The process of counting the number of students who eat with their classmates in a cafeteria window is as follows.

Input: Student consumption data set C, cafeteria window number N and a list of cafeteria window numbers corresponding to N on that level $Nlist$.

Output: a list of the number of all students who ate at that cafeteria window with their classmates.

1) Get the class list $major\_class$ based on the student consumption data set C.

2) Loop through the class list to get the students stu_major of a class and get the $month$ of consumption of the class based on the date of consumption.

3) Loop over the list of $month$ to get the consumption record stu_month of the students of the class in a particular month.

4) Loop over stu_month to get a student's consumption record stu_one in the current month.

5) Determine whether the length of stu_one is greater than or equal to 30, if it is greater than or equal to 30, loop traversing stu_one to get the consumption data of the cafeteria window number is equal to N, converting the time of its consumption to a Unix timestamp, and depositing it into the $timestmp\_1$ list.

6) Obtain the consumption data set of the other students of the class other than the student in the cafeteria window $Nlist$ of the level corresponding to the cafeteria window N, convert the time in the obtained consumption data set into the timestamp list $timestmp\_2$.

7) Loop through the $timestmp\_1$ list and the $timestmp\_2$ list, if the difference between the timestamps in $timestmp\_1$ and the timestamps in $timestmp\_2$ is less than or equal to 120, then the counter $count$ is increased by 1.

8) Deposit $count$ into the list $consume\_num$ of the number of students who have eaten with their fellow classmates at the cafeteria window N and set $count$ to 0. Go to step 3) until all loops have been executed.

II. B. 3) 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 time spent online on weekdays and weekends

Considering that students' online habits are different on weekdays and weekends, we extracted students' online characteristics by separating weekdays and weekends. The length of each student's online time can be obtained from the data set of students' online records, and the average length of students' online time on weekdays or weekends:

$$S_{UT}=\sum_{i=1}^{T}I_i/T_d \tag{5}$$

Where T is the number of times students go online on weekdays or weekends, I_i is the amount of 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 weekdays and weekends on average. The average offline time on weekdays and weekends refers to the average of the time when students log out of the campus network system for the last time on weekdays and weekends. If students have not logged out before zero o'clock on that day, the earliest offline time on the 2nd day will be taken as the latest offline time on that day. The time the student logged out of the campus network is converted to a Unix timestamp in the form of the average weekday or weekend latest offline time of the student's 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 a weekday or weekend.

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. 4) Characteristic correlation analysis

In order to investigate the differences between the mental health and psychological abnormal student groups in related behaviors, the Jenks Natural Breaks algorithm is first applied to label the extracted feature data, which is also known as the natural breaks grading method, and the core idea is the same as that of clustering: to maximize the similarity of each group, and maximize the differences of each group, then the Apriori algorithm is applied to mine the behavioral feature labeled data sets of mental health and psychological abnormal student groups respectively, and set the minimum support threshold as 0.5 and the minimum confidence level as 0.5, and set the minimum support threshold as 0.5. The core idea is the same as clustering: maximize the similarity within each group, and maximize the dissimilarity between external groups, and then apply Apriori algorithm to mine the labeled dataset of behavioral characteristics of mental health and psychological abnormalities students' groups respectively, and set the minimum support threshold to 0.5 and the minimum confidence threshold to 0.5, and the resulting strong association rules are shown in Table 1. From the table, it can be seen that psychologically abnormal student groups are usually characterized by less regular diet, less diligence, fewer number of shared meals, longer time on the Internet and higher number of traffic, and later time off the Internet on weekdays. On the other hand, the psychologically normal student group is usually characterized by eating more regularly and being more diligent.

Table 1: Strong association rules generated by Apriori algorithm

Serial number	Association rule	Support	Confidence
1	The psychological abnormal student group is less regular	0.542	0.542
2	The psychological abnormal student group is less diligent	0.511	0.511
3	The number of students in the psychological abnormal student group was less large	0.536	0.536
4	The psychological abnormal student group is long longer on the Internet	0.514	0.514
5	The psychological abnormal student group has a longer time on the weekend	0.516	0.516
6	The daily traffic of the psychological abnormal student group is more commonly used	0.517	0.517
7	The daily traffic of the psychological abnormal student group is more used	0.509	0.509
8	The psychological abnormal student group is late in the working day	0.516	0.516
9	The psychological normal student group is more regular	0.798	0.798
10	The psychological normal student group is more diligent	0.596	0.596

II. C.Bayesian Network Modeling for Predicting Students' Mental Health Status

II. C. 1) Bayesian network modeling

Bayesian networks, also known as belief networks, a model for expressing probabilistic relationships of random variables, is an important branch in the development of graph theory. It consists of two main parts, the directed acyclic graph (DAG) and the conditional probability table, where the directed edges represent the interdependencies between the nodes, while the conditional probability table (CPT) describes the strength of the influences between the nodes in relation to the strength of each node's dependence on its parent node. In each belief network graph, the direction pointed by the arrow is the child node, its other end is the parent node, and the one without a parent node is called the root node. Bayesian network in the form of probability to complete the transmission of uncertainty information and reasoning, in the knowledge reasoning process can realize the key factor analysis, reverse inference and other functions [30].

If the Bayesian network structure S consists of the set of node variables $V\left(V=\{V_1,V_2,\cdots,V_n\}\right)$ and the set of directed edges $E\left(E=\{V_iV_j \mid V_i,V_j \in V\}\right)$, then the dependency or causality between variables V_i,V_j is reflected by the directed edges E. And for each node $X_i \in V$ is accompanied by a conditional probability distribution table $P\left(X_i \middle| pa(X_i)\right)$, where $pa(X_i)$ corresponds to X_i of the set of parent nodes. The formula for the joint probability distribution is obtained by combining the conditional independence between nodes and the chaining rules of probability as follows:

$$P(X_1,X_2,\cdots,X_n)=\prod_{i=1}^{n}P\left(X_i \middle| pa(X_i)\right) \tag{8}$$

where if $pa(X_i)$ is the empty set, then $P\left(X_i \middle| pa(X_i)\right)$ denotes the prior probability $p(X_i)$.

II. C. 2) Bayesian network structure learning

Bayesian network structure is simple and easy to understand, clear hierarchy, and in the form of probability to complete the information uncertainty inference. As one of the machine learning methods is also able to make full use of historical data to synthesize expert empirical opinions to set model parameters, which is an ideal tool for risk prediction and risk modeling.

II. C. 3) Bayesian network parameter learning

In order to obtain the input probabilities of the state nodes X_i for the simple Bayesian network with final node states of "High" and "Low", it is necessary to defuzzify the seven states of the expert linguistic fuzzy set, which are transformed into the corresponding probability values of the two states [31].

Fuzzy set theory is an extension of classical set theory. Compared with other fuzzy numbers, triangular fuzzy numbers have the advantages of simplicity and clarity, and convenient operation. In this paper, the triangular fuzzy number is chosen as the affiliation function, and the affiliation function based on the triangular fuzzy number will be described to the expert language [32].

Fuzzy number refers to the uncertain value in the measurement process, which includes the fuzzy upper limit, lower limit, and intermediate value. In this study, it is assumed that each state node of postgraduate students' mental health risk is a fuzzy measure, the affiliation function is A, the value of affiliation at 1 is m, the upper limit and lower limit are a and b, respectively, and $b-a$ denotes the degree of fuzzy of the node, and the higher the value is, the higher the degree of fuzzy is. Then the whole risk system affiliation set A is denoted as $A \cong (a,m,b)$ and the affiliation function is:

$$\tilde{A}(x)=\begin{cases}0 & x<a \text{ Or } x>b \\ \dfrac{x-a}{m-a} & a \le x \le m \\ \dfrac{b-x}{b-m} & m \le x \le b\end{cases} \tag{9}$$

In order to obtain the a priori probability value of each state node, the triangular fuzzy numbers corresponding to the questionnaire data are processed, which mainly includes homogenization, defuzzification and normalization.

(1) Homogenization

Homogenization applies the method of arithmetic averaging to arithmetically average the results of the acquired measures. Adopting this process mainly cleans the data in the training set, removes individual outliers and invalid values, so that the fuzzy probability of each node can converge to a reasonable range. The specific formula is:

$$\tilde{P}=\frac{p_1+p_2+\cdots+p_n}{n}=\left(\tilde{a},\tilde{m},\tilde{b}\right) \tag{10}$$

The triangular fuzzy numbers of X_i are calculated by homogenization, respectively, and the result after node homogenization follows the form of fuzzy set consisting of upper, lower and intermediate values.

(2) Defuzzification and normalization

In the process of averaging calculation, the result calculated after the averaging of each state node consists of fuzzy data, which needs to be processed to determine the probability value in order to get the probability measure result of the Bayesian network. The triangular fuzzy number after taking the mean value is defuzzified by the method of average area to get the exact probability value P' of the node. The specific formula is expressed as:

$$P'=\frac{\tilde{a}+2\tilde{m}+\tilde{b}}{4} \tag{11}$$

The probability value of the corresponding state of each node has been obtained after the completion of homogenization and defuzzification, but in order to satisfy the condition that the sum of the probability value of the occurrence of each state of the node is 1, i.e., normalization, the probability value of the state of each node will be normalized to obtain the a priori probability value of the node of each state. The normalization process is shown in equation (12):

$$P_i=\frac{P'}{\sum_{i=1}^{2}P'} \tag{12}$$

According to the above principle, the defuzzification and normalization processes are completed sequentially by the triangular fuzzy number mean statistics.

The implied nodes, except the state nodes, can be calculated according to the Bayesian formula, which is given as:

$$P\left(Y_1,Y_2,\cdots,Y_5,H\right)=\frac{\prod_j P\left(Y_j \mid parent\left(Y_j\right)\right)\prod_i P\left(X_i \mid parent\left(X_j\right)\right)}{\prod_i P\left(X_i\right)} \tag{13}$$

III. Experimental program and experimental results

III. A. Experimental data and program

III. A. 1) Data preparation

This experiment uses a sample dataset consisting of two parts for model training, i.e., labeled data and feature dimension data. The labeled data were first processed, including labeling the mental state information of the anxiety dimension. Data with anxiety symptoms on the student scale were categorized as label “1”, while data without anxiety symptoms were categorized as label “0”. Next, the feature vector data were processed, including the influencing factors with a greater degree of association with anxiety symptoms obtained in Chapter 4 as feature vectors, i.e., depressive symptoms, obsessive-compulsive disorder (OCD) symptoms, age, and sensitivity to interpersonal relationships. Finally, the data on students' labeled mental states were combined with the data on the features affecting their mental states to create a sample data set for model training.

III. A. 2) Experimental steps

This section tries a variety of different algorithmic models, including decision tree, support vector machine, gradient boosting decision tree, etc., using different evaluation indexes to measure the performance of the model and select the optimal prediction model for training, the process of constructing the prediction model of mental health status is shown in Figure 2.

Step 1: Input the psychological assessment data information after the student data processing and the classification algorithm, the assessment data information includes the student basic data and scale assessment data.

Step 2: Based on the results of logistic regression analysis of students' basic attribute data, construct static feature vector S. Based on the ISM-FP-Growth algorithm to mine the psychological dimensions affecting mental health status, construct dynamic feature vector D. Based on the results of logistic regression analysis of students' basic attribute data, construct dynamic feature vector D. Based on the results of logistic regression analysis of students' basic attribute data, construct dynamic feature vector D.

Step 3: According to different experiments, construct the label vector y=(D, label) from the psychological assessment data according to the method in the data preparation, e.g., the label vector yls for binary classification experiments, and the label vector ymul for multiclassification experiments.

Step 4: Connect the static feature vector S and the dynamic feature vector D and the label vector y into a data matrix Data according to the ID field, and divide the data set D into a training set Train and a test set Test according to the ratio of 7:3 of the data set.

Step 5: The training set Train is used as the input to the classification algorithm, and the best classification prediction model is determined after training.

III. B. Selection of Student Behavioral Characteristics

Mental health-related data collected in this study included a total of 37 variables including general demographic indicators, sleep, exercise, personality and stress. Recursive random forests were used to select features, and the subsequent predictive models were learned based on the final selection results. In order to find the optimal number of features, cross-validation was used along with recursive random forests to obtain the accuracy scores of different subsets of features and select the set of features with the best scores. During the process, the number of cross-validation was set to 5 and finally 20 features were selected as the result of feature selection for RFECV. The relationship between the number of features and the fifty-fold cross-validation score is shown in Fig. 3.

As can be seen from the figure, the score is in an overall upward trend as the number of features increases. When the number of features is 20, the curve appears to peak at 0.8127, while thereafter the scores fluctuate as the number of features increases, but the effects are all lower than when the number of features is 20. The 20 features thus selected are: emotional stress, self-acceptance, life feelings, neuroticism, sleep quality, career choice stress, self-evaluation, interpersonal stress, relationship stress, academic stress, school environment stress, sense of seeking meaning, life attitude, life goals, agreeableness, extraversion, health stress, family stress, openness, and frustration stress. The variables obtained after feature selection were all continuous type variables.

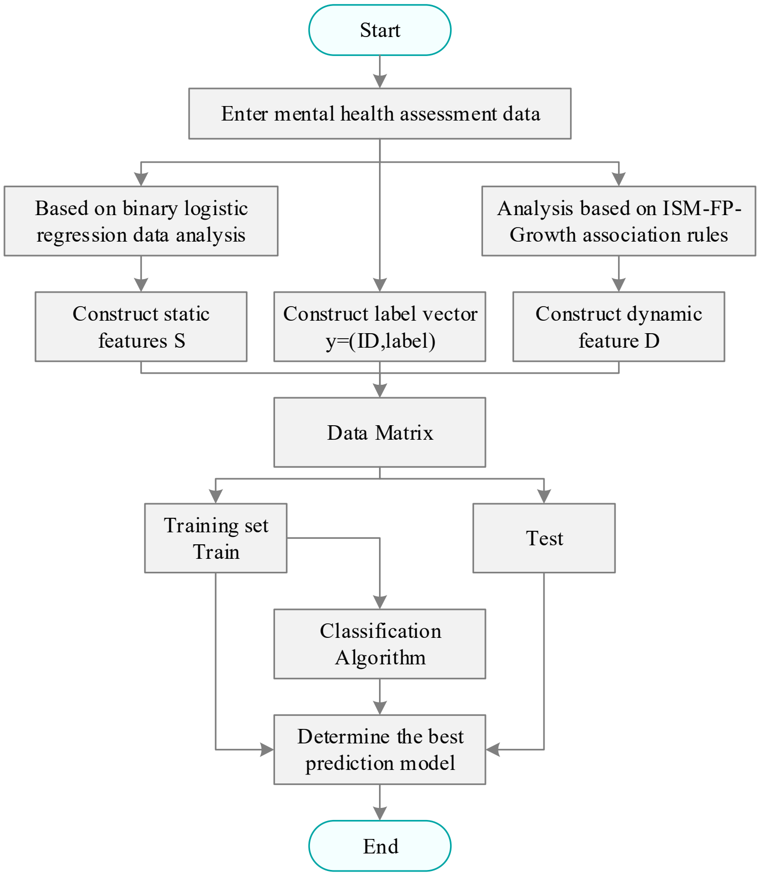
Start
Enter mental health assessment data
Based on binary logistic regression data analysis
Analysis based on ISM-FP-Growth association rules
Construct static features S
Construct label vector y=(ID,label)
Construct dynamic feature D
Data Matrix
Training set Train
Test
Classification Algorithm
Determine the best prediction model
End

Figure 2: mental health state prediction model construction process

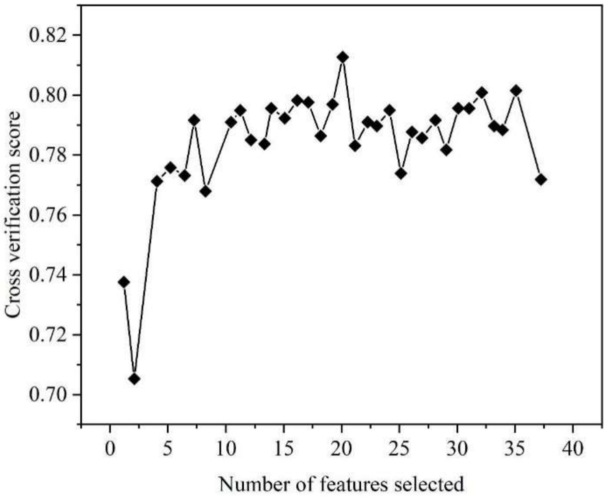

Figure 3: The number of features and the cross-verification scoring room

III. C. Analysis of the results of the prediction of students' mental health status

In this paper, the performance of the classification algorithm model is evaluated using the confusion matrix, and the evaluation metrics include accuracy, recall and F1 value. Before the specific experiments, it is necessary to determine the number of influencing factors in the frequent itemsets of the strong association rules mined in this paper, d, which is taken as d = 2, 3, 4, 5, in the same algorithmic model. Decision Tree Algorithm DT, Support Vector Machine SVM, Boosting Algorithm AdaBoost, and Bayesian Network (BN) under the anxiety binary classification performance results are shown in Figure 4. First of all, according to the experimental results, it can be seen that the number of different mental health state influencing factors has a certain impact on the feature vector, and each model is affected, but each model F1 value can reach more than 0.6, which shows that the model features constructed through the previous section are effective influencing factors of the user's mental health state.

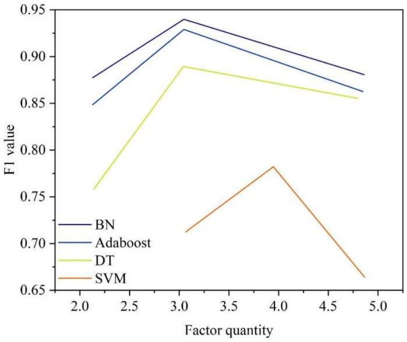

Figure 4: F1 values of the different factors of anxiety

The evaluation results are shown in Table 2, among the four algorithmic models, the optimal value of the number of influencing factors is 3, in which the BN model performs better, and its F1 value can reach 0.9389, only followed by the Adaboost model, whose size of F1 value is only second to the BN prediction model, and the performance of the remaining DT prediction model and the SVM prediction model is slightly lower than the performance of the previous BN algorithm and Adaboost algorithm models.

Table 2: Evaluation indexes of each model of anxiety dichotomy

Algorithm model	Factor quantity	P value	R value	F1 value
DT	2	0.7528	0.7734	0.7645
	3	0.9015	0.8884	0.9024
	4	0.8814	0.9002	0.9013
	5	0.8536	0.9748	0.8573
SVM	2	-	-	-
	3	0.7235	0.7074	0.7158
	4	0.7802	0.7759	0.7786
	5	0.6635	0.6848	0.6693
Adaboost	2	0.8574	0.8574	0.8547
	3	0.9245	0.9214	0.9208
	4	0.9047	0.9028	0.9004
	5	0.8836	0.9074	0.8725
BN	2	0.8804	0.8897	0.8847
	3	**0.9415**	**0.9387**	**0.9389**
	4	0.9247	0.9226	0.9217
	5	0.8948	0.9028	0.8982

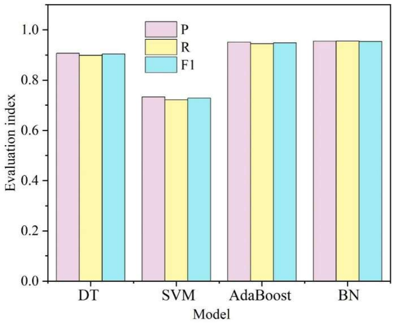

Figure 5: Accuracy, recall rate and f1 value of different models

Again, the prediction results are shown in Figure 5, among the four algorithmic prediction models, except for the SVM prediction model, which performs slightly lower than the other three prediction models, the three prediction models of DT, Adaboost, and BN can all reach more than 0.9 in terms of accuracy, more than 0.86 in terms of recall, and more than 0.89 in terms of F1 value. Secondly, among the four algorithmic models, compared with the other algorithmic prediction models, the BN prediction model can reach an accuracy of 0.9548, a recall rate of 0.9552, and an F1 value of 0.9533, which is still the best performance of the BN model.

The experimental results are shown in Table 3. Through the analysis of the above experimental results, it is shown that the students' anxious psychological state is predictable, and there exists a significant correlation with the students' psychological assessment data and basic information, and by analyzing the correlation rules, we can dig out the factors affecting the mental health state and their correlation degree, and ultimately predict the changes of the students' psychological state.

Table 3: Anxiety multifaceted model evaluation index

Algorithm model	Factor quantity	Fmacro	Fmicro
DT	3	0.7652	0.8036
SVM	3	0.6824	0.7142
Adaboost	3	0.8183	0.8536
BN	3	**0.8549**	**0.8814**

To further improve the accuracy of the model prediction results, the experimental feature matrix was linked to the anxiety multiclassification labels and the dataset was partitioned based on cross-validation. To ensure the uniqueness of the experimental variables, the multiclassification prediction model continues to use the four algorithms of the previous biclassification, i.e., Decision Tree DT, Support Vector Machine SVM, Adaboost, and BN algorithms, to train the model and test the data, and the experimental results of the multiclassification prediction model obtained are shown in Figure 6.

According to the experimental comparison results in the figure, the performance of the prediction model for anxiety multicategorization is similar to that of the bicategorization prediction model, and the prediction performance of the two integrated learning models is better than that of the other prediction models, in which the BN model still has the best performance, and the Fmacro can reach 0.8504, and the Fmicro can reach 0.8595. The difference is that, since the multicategorization problem is more than the bicategorization problem complex, so the performance in each subcategory is slightly worse than the dichotomous results, but in terms of the overall results, the multicategorization results can still reach the significant level, which again confirms the conclusion of the previous dichotomous experimental results, indicating that the prediction model can predict anxiety level at a finer level.

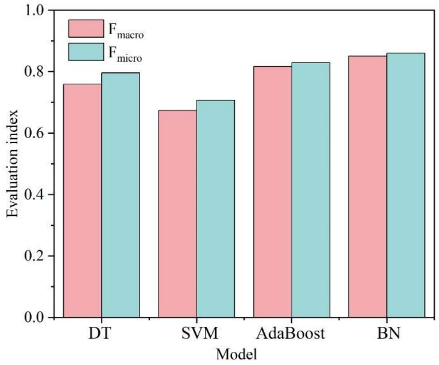

Figure 6: Fmacro and Fmicro of different models

IV. Conclusion

This study constructed a Bayesian network-based prediction model of students' mental health status, which realized the effective prediction of students' mental health status. The study first identified 20 key features through feature selection, including emotional stress, self-acceptance, life feelings, neuroticism, and sleep quality, which were significantly associated with students' mental health status. In the algorithm model comparison experiment, the optimal number of influencing factors for all four algorithms (decision tree, support vector machine, boosting algorithm, and Bayesian network) is 3, indicating that complex mental health status can be effectively predicted by a small number of key factors.

In the anxiety binary classification experiment, the Bayesian network prediction model performed the best, with an accuracy of 0.9548, a recall of 0.9552, and an F1 value of 0.9533, which was significantly better than the other three algorithmic models. This result proves that students' anxious psychological state is predictable and significantly associated with students' psychological assessment data and basic information. In the more complex anxiety multiclassification prediction experiments, the Bayesian network model still maintains the lead, with its Fmacro value reaching 0.8504 and its Fmicro value reaching 0.8595, which further confirms that the model can predict students' anxiety levels at a finer level.

The strong association rules mined by the Apriori algorithm revealed that the psychologically abnormal student group usually showed the characteristics of less regular diet, less diligence, fewer number of people sharing meals, longer time on the Internet and higher number of traffic used, and later time off the Internet on weekdays; while the psychologically normal student group usually showed the characteristics of more regular diet and more diligence. The association between these behavioral characteristics and mental health status provides an important basis for mental health intervention in colleges and universities.

The Bayesian network prediction model constructed in this study has high prediction accuracy, which can provide technical support for universities to carry out mental health monitoring and accurate intervention, help solve the current problems faced by the work of mental health in universities, and improve the efficiency and accuracy of work.

References

[1] De Graaf, R., Bijl, R. V., Ravelli, A., Smit, F., & Vollebergh, W. A. M. (2002). Predictors of first incidence of DSM‐III‐R psychiatric disorders in the general population: findings from the Netherlands Mental Health Survey and Incidence Study. Acta Psychiatrica Scandinavica, 106(4), 303-313.

[2] Quilon, A., & Kurniawan, Y. (2023). Online learning environment and mental health among university students. Bedan Research Journal, 8(1), 259-284.

[3] Sun, Y. (2023). The relationship between college students' interpersonal relationship and mental health: Multiple mediating effect of safety awareness and college planning. Psychology Research and Behavior Management, 261-270.

[4] Ritter, L. J., Hilliard, T., & Knox, D. (2022). "Lovesick": mental health and romantic relationships among college students. International journal of environmental research and public health, 20(1), 641.

[5] Liu, B. (2022). Research on the Relationship Between College Students' Mental Health and Employment Based on Data Mining. International Journal of Information Systems in the Service Sector (IJISSS), 14(3), 1-17.

[6] Saeed, S. A., & Gargano, S. P. (2022). Natural disasters and mental health. International review of psychiatry, 34(1), 16-25.

[7] Huang, Y., Su, X., Si, M., Xiao, W., Wang, H., Wang, W., ... & Qiao, Y. (2021). The impacts of coping style and perceived social support on the mental health of undergraduate students during the early phases of the COVID-19 pandemic in China: a multicenter survey. BMC psychiatry, 21, 1-12.

[8] Barker, R., Hartwell, G., Bonell, C., Egan, M., Lock, K., & Viner, R. M. (2022). Research priorities for mental health in schools in the wake of COVID-19. J Epidemiol Community Health, 76(5), 448-450.

[9] Zapata-Ospina, J. P., Patiño-Lugo, D. F., Vélez, C. M., Campos-Ortiz, S., Madrid-Martínez, P., Pemberthy-Quintero, S., ... & Vélez-Marín, V. M. (2021). Mental health interventions for college and university students during the COVID-19 pandemic: a critical synthesis of the literature. Revista Colombiana de psiquiatria (English ed.), 50(3), 199-213.

[10] Wagner, B., Snoubar, Y., & Mahdi, Y. S. (2023). Access and efficacy of university mental health services during the COVID-19 pandemic. Frontiers in Public Health, 11, 1269010.

[11] Xiaoyi, F. A. N. G., Xiaojiao, Y. U. A. N., Wei, H. U., Linyuan, D. E. N. G., & Xiuyun, L. I. N. (2018). The development of college students mental health screening scale. Studies of Psychology and Behavior, 16(1), 111.

[12] Dunstan, D. A., Scott, N., & Todd, A. K. (2017). Screening for anxiety and depression: reassessing the utility of the Zung scales. BMC psychiatry, 17, 1-8.

[13] Bruno, F., Mautone, A., Ait Ali, D., Fassima, A., Khabbache, H., & Rizzo, A. (2025). Evaluating Facebook scales: A systematic review of the psychological assessment tools. Adv Med Psychol Public Health, 2(3), 142-156.

[14] Trindade, I. A., Ferreira, C., & Pinto‐Gouveia, J. (2017). Chronic illness‐related shame: Development of a new scale and novel approach for IBD patients' depressive symptomatology. Clinical Psychology & Psychotherapy, 24(1), 255-263.

[15] Mao, X. L., & Chen, H. M. (2023). Investigation of contemporary college students' mental health status and construction of a risk prediction model. World Journal of Psychiatry, 13(8), 573.

[16] Zhang, P., Han, W., & Liu, Q. (2024). Research on Predicting the Mental Health of College Students with Prediction Models based on Big Data Technology. IEIE Transactions on Smart Processing & Computing, 13(4), 393-401.

[17] Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., ... & Du, X. (2025). Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students. Journal of Counseling & Development, 103(1), 110-125.

[18] Pei, J. (2022). Prediction and analysis of contemporary college students' mental health based on neural network. Computational Intelligence and Neuroscience, 2022(1), 7284197.

[19] Liao, S., Wang, Y., Zhou, X., Zhao, Q., Li, X., Guo, W., ... & Qiu, P. (2022). Prediction of suicidal ideation among Chinese college students based on radial basis function neural network. Frontiers in public health, 10, 1042218.

[20] Fu, Z., Zhou, S., Burger, H., Bockting, C. L., & Williams, A. D. (2020). Psychological interventions for depression in Chinese university students: a systematic review and meta-analysis. Journal of Affective Disorders, 262, 440-450.

[21] Benjet, C., Zainal, N. H., Albor, Y., Alvis-Barranco, L., Carrasco Tapia, N., Contreras-Ibáñez, C. C., ... & Kessler, R. C. (2025). The Effect of Predicted Compliance With a Web-Based Intervention for Anxiety and Depression Among Latin American University Students: Randomized Controlled Trial. JMIR mental health, 12, e64251.

[22] Papadatou-Pastou, M., Campbell-Thompson, L., Barley, E., Haddad, M., Lafarge, C., McKeown, E., ... & Tzotzoli, P. (2019). Exploring the feasibility and acceptability of the contents, design, and functionalities of an online intervention promoting mental health, wellbeing, and study skills in Higher Education students. International Journal of Mental Health Systems, 13, 1-15.

[23] Ji, X. (2024). Research on Mental Health Assessment and Intervention Methods for College Students based on Big Data Analysis. Scalable Computing: Practice and Experience, 25(6), 4702-4711.

[24] Zhao, J., Gangqian Wang, N. Y., & Xue, Y. (2025). Design of a neural network model-based system for predicting students' mental health status and psychological intervention in colleges and universities. J. COMBIN. MATH. COMBIN. COMPUT, 127, 485-506.

[25] Reichenberg, R. (2018). Dynamic Bayesian networks in educational measurement: Reviewing and advancing the state of the field. Applied Measurement in Education, 31(4), 335-350.

[26] He, Y. M. (2024). Online Assessment of Mental Health Micromedia for College Students Incorporating Bayesian Network Algorithm. International Journal of Maritime Engineering, 1(1), 83-96.

[27] Wang, Y., Li, Z., & Cao, X. (2024). Investigating the network structure and causal relationships among bridge symptoms of comorbid depression and anxiety: a Bayesian network analysis. Journal of Clinical Psychology, 80(6), 1271-1285.

[28] Xuexing Du,Jennifer Crodelle,Victor James Barranca,Songting Li,Yunzhu Shi,Shangbang Gao & Douglas Zhou. (2025) .Biophysical modeling and experimental analysis of the dynamics of C. elegans body-wall muscle cells. PLoS computational biology,21(1),e1012318.

[29] Youngok Choi & Sue Yeon Syn. (2016) .Characteristics of tagging behavior in digitized humanities online collections. Journal of the Association for Information Science and Technology,67(5),1089-1104.

[30] Yangfan Zhou,Jianchun Fan,Baoqian Dai,Shengnan Wu,Rujun Wang,Xinwei Yin... & Xiaofeng Zhang. (2025) .Risk analysis of urban low-pressure natural gas networks based on hybrid dynamic Bayesian networks. Journal of Loss Prevention in the Process Industries,96,105649-105649.

[31] Florian van Daalen,Lianne Ippel,Andre Dekker & Inigo Bermejo. (2024) .VertiBayes: learning Bayesian network parameters from vertically partitioned data with missing values. Complex & Intelligent Systems,10(4),5317-5329.

[32] Shing Chung Ngan. (2025) .A concrete extension principle for fuzzy set theory. Expert Systems With Applications,280,127328-127328.