Hindawi
Journal of Mathematics
Volume 2024, Article ID 9783245, 1 page
https://doi.org/10.1155/2024/9783245

Retraction

Retracted: Multidimensional State Data Reduction and Evaluation of College Students' Mental Health Based on SVM

Journal of Mathematics

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

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Peiqing, "Multidimensional State Data Reduction and Evaluation of College Students' Mental Health Based on SVM," *Journal of Mathematics*, vol. 2022, Article ID 4961203, 11 pages, 2022.

Hindawi
Journal of Mathematics
Volume 2022, Article ID 4961203, 11 pages
https://doi.org/10.1155/2022/4961203

Research Article

Multidimensional State Data Reduction and Evaluation of College Students' Mental Health Based on SVM

Han Peiqing

Henan Institute of Economics and Trade, Zhengzhou, Henan 450000, China

Correspondence should be addressed to Han Peiqing; han.peiqing@outlook.com

Received 23 December 2021; Revised 9 January 2022; Accepted 15 January 2022; Published 23 February 2022

Academic Editor: Naeem Jan

In response to the shortcomings of the traditional methods for evaluating the mental health status of college students in terms of computational complexity and low accuracy, a method for evaluating the mental health status of college students based on data reduction and support vector machines was proposed. A model experiment containing internal and external personality tendency classification, anxiety, and depression dichotomy was designed using logistic regression analysis, information entropy, and SVM algorithm to construct the feature dimensions of the network behavior data, combined with the labeled data of mental state to derive the sample data set for model experiments. In the experimental process, to reflect the difference in the effect of different models, various types of mathematical models were constructed for horizontal comparison; at the same time, to reflect the influence of the parameters of the same type of model, different combinations of parameters were constructed using a grid search algorithm to vertically compare the difference in the effect. The average accuracy of the dichotomous model for anxiety and depression in the sample of 1433 students was 0.80 or higher. The experiments show that the method of predicting students' psychological status through their online behavioral data is feasible, and the mathematical classification model can be used to grasp students' psychological status in real time and to warn students with abnormal psychological status, thus helping school counselors to intervene and prevent them promptly.

1. Introduction

With the increasing competition in society, college students are facing multiple pressures in employment, life, study, and emotion, which lead to the frequent occurrence of college students' mental health problems and directly affect the stability of campus life and learning environment; therefore, it is of great practical significance and theoretical value to evaluate the mental health of college students [1]. At present, the methods for evaluating the mental health status of college students are mainly based on traditional machine learning algorithms, such as decision trees (DT) and feedforward neural networks (BPNN), which have the disadvantages of great computational effort and low accuracy rate [2]. The university campus life is a critical period for students' rapid psychological development and maturity and is a key step in the formation of healthy psychology; school educators and parents should pay attention to the guidance and education of students during this period to help them shape healthy psychology. Unsatisfactory schooling has a number of detrimental repercussions on pupils' psyche, as well as their subsequent incapacity to integrate effectively into social life. The goal of this research is to gather information regarding students' psychological well-being without putting them under undue stress so that psychologists can perform their jobs more effectively. This paper investigates the relationship between students' psychology and data and builds the feature vector associated with psychology using behavioral data, including mining students' social relationship characteristics using behavioral data and proving the scientificity of the method using students' social sensitivity data, constructing regularity correlation index features using the theory of information entropy and penalty factor, and measuring students' social sensitivity data.

In today's era of big data, students, as the majority of the campus population, produce the most data, the most

valuable of which is data records of students' online behavior, which are generated by students of their own volition, can reflect differences between students to the greatest extent, and have laid a solid data foundation for this topic's research [4]. The traditional psychological method of data collection is based on the concept of random sampling, and its analysis method is based on hypothesis testing, both of which have drawbacks that affect the accuracy of the results, whereas big data psychology uses the full volume and full data analysis method, which completely squeezes the value contained in the data and greatly improves the accuracy of the experimental results. In this paper, based on the online behavioral data generated by school college students on the Internet and the basic attribute data of students, combined with the content of the MBTI assessment scale, PHQ-9 depression screening scale, and GAD-7 anxiety screening scale, the correlation between students' behavioral data and their psychological states is explored, and the data mining method is used to construct the models of internal and external personality tendency classification, depression state classification, and anxiety classification [5]. The data mining method is used to construct a classification model for personality internal and external tendencies, depression, and anxiety, to predict the psychological state of students based on their behavioral data at school, to warn students with psychological abnormalities, and to assist psychologists in better psychological intervention.

With the digitization of colleges and universities, the degree of information technology in each college and university is getting higher and higher, and everything is gradually being data-driven. Every instantaneous behavior of students on campus is transformed into a data record, and these data records lay the foundation for the study of this topic. The psychological research approach based on big data technology has its significant superiority compared with traditional psychological research. Furthermore, the big data research approach analyzes all data from the entire dataset, avoiding the sampling error that may arise from the traditional research method of sampling samples from the entire dataset, then generalizing to the entire dataset, collecting data directly from the real environment, and avoiding the experimental error caused by the traditional research method of overcontrolling the experimental conditions, which results in the generation of behaviors that are not representative of the real environment. In terms of the efficiency and scale of data processing, it can relieve the pressure of traditional research methods that require a lot of time and workforce; in terms of time validity, it can instantaneously collect, analyze, and predict real-time psychological states, avoiding the bias of analysis results caused by the lag of data collection of traditional research methods. This paper analyzes the correlation and degree of correlation between students' behaviors and their psychological states based on the behavioral data and scale data generated by students at school, constructs the features associated with students' psychological states by combining students' static attributes, and builds a prediction model of students' psychological states through machine learning methods. The goal is to forecast students' present psychological states by collecting real-time data on their school behavior, to warn students with aberrant psychological states, and to aid campus psychology professors and specialists in intervening with kids who may have psychological issues.

2. Current Status of Research

Based on summarizing previous studies, a model based on robust multitask learning was established using the Big Five personality assessment scale as an indicator to achieve accurate classification prediction of personality variables of active Weibo users [6]. As for the research on the prediction of students' depression, Barenholtz et al. established a regression model to analyze the severity of students' depression by using the number of posts, likes, and likes for others as the characteristic dimensions of students' microblog comments in a week, and the results showed that people with heavy depression had a higher number of dynamic posts, but a lower number of likes for others [7]. In a study on student suicide risk prediction, Kumari et al. used keywords extracted from text data published by microblog users for data analysis and used MLP models to achieve a predictive assessment of suicide risk based on microblog text data [8]. However, one of the obvious drawbacks of these studies is that they only focus on users who have social network accounts, such as Sina Weibo, and not every student uses Sina Weibo; in fact, according to the survey, the average number of active users of Sina Weibo is only 4 to 5 [9]. There are no studies that show a psychological difference between Weibo users and non-Twitter users. Therefore, data from social networks such as Weibo are not sufficient to predict the psychology of all school students. The higher the psychological indicator of gratitude, the lower the value of the corresponding indicator of depression [10]. The cross-sectional study demonstrated that the quality of sleep of college students is related to their depression level, the quality of sleep and depression level can interact with each other, and long-term poor sleep quality significantly increases the occurrence of depression. A study of a sample of more than 2,000 individuals through a mediated moderation model verified that peer friendships influence the psychological aspects of depression in students and that friendship support negatively predicts depression. However, these aspects of the study only made psychological causal explanations and did not do so to the extent of predicting psychological states [11].

Anomaly detection techniques aim to automatically identify those observations in a large set that are valuable or whose behavior is different from the expected one. Anomaly detection is one of the important applications in data mining technology and has specific applications in many practical production's lives, such as credit card fraud, industrial damage detection, and image detection. Through in-depth research, scholars at home and abroad have proposed many anomaly detection algorithms with high feasibility, which has laid a solid foundation for further research on anomaly detection [12]. Nevertheless, the basic mathematical models are still very useful and have eventually been adapted to many computational scenarios. Current research tools are divided into nonparametric and parametric approaches: the

nonparametric approach does not require the assumption of knowledge of any parameters and uses nonparametric techniques to estimate the density of the distribution, for example, histograms and Parzen window estimation; the parametric approach requires the assumption that normal data are generated based on parametric distributions, and it requires these parameters from training samples, for example, outlier detection methods based on normal distributions [13]. Neural networks, which may be classified into single-classification neural networks and multiclassification neural networks [14], are an important field of nonlinear modeling approaches. Multiclassification neural networks, such as multilayer perceptron, neural tree, and others, use data from multiple classifications to train the model and then input test data into the model, which the network interprets as normal or abnormal; single-classification neural networks, such as Replicator Neural Networks (RNNs), use a function (like a step function) to transform the sample into N discrete variables for sample clustering.

The mental health status data of college students collected according to the conventional model assessment guidelines and the symptom self-assessment scale SCL-90 are high-dimensional datasets, so there are drawbacks of huge computation and redundant correlations between data features when conducting mental health assessment of college students. With the extensive research of popular learning algorithms, this method has been widely applied to image retrieval, text classification, face recognition, and plant leaf recognition as data feature extraction and dimensionality reduction. A mental health assessment approach based on a local linear embedding algorithm (LLE) and support vector machine (SVM) is presented to increase the accuracy of college students' mental health status evaluation. LLE-SVM may successfully increase the accuracy rate of college students' mental health status assessment when compared to SVM, BPNN, and DT.

3. SVM Analysis of Multidimensional State Data Reduction and Evaluation of College Students' Mental Health

3.1. SVM Multidimensional State Data Reduction Evaluation Design. The evaluation of the college students' mental health status is essentially a nonlinear classification problem. Since the characteristics of everyone's mental state data are multidimensional and these characteristics involve many nonlinear factors and have the characteristics of multilevel, multivariable, nonlinear, and strong coupling, it is difficult to describe quantitatively by traditional mathematical models or methods [15]. To improve the accuracy of the evaluation of the mental health status of college students, it is very necessary to establish a more scientific and reasonable evaluation model of the mental health status of college students. In this study, nine dimensions of psychoticism, paranoia, hostility, terror, anxiety, depression, obsessive-compulsive symptoms, interpersonal sensitivity, and somatization were used as input vectors of the LLE-SVM model, and the mental health status of college students was divided into healthy, mildly unhealthy, and unhealthy as output vectors of the LLE-SVM model to establish the evaluation model of college student's mental health status based on LLE-SVM. The evaluation model is shown in Figure 1.

In this study, the data of nine dimensions, including psychoticism, paranoia, hostility, terror, anxiety, depression, obsessive-compulsive symptoms, interpersonal sensitivity, and somatization, were used as the input vectors of the LLE-SVM model, and the mental health status of college students was divided into healthy, mildly unhealthy, and unhealthy as the output vectors of the LLE-SVM model to establish a mental health status evaluation model of college students based on the LLE-SVM. The process of the evaluation algorithm based on LLE and SVM can be described in detail [16]. The data of mental health characteristics of college students were collected: SCL-90 was issued to collect the data of mental health characteristics of college students, which included nine dimensions of psychoticism, paranoia, hostility, terror, anxiety, depression, obsessive-compulsive symptoms, interpersonal sensitivity, and somatization. The LLE algorithm was used to reduce the dimensionality of college students' mental health characteristics data to reduce the computational effort. The reduced dimensional data were divided into training samples and test samples, and the training samples were used to build the LLE-SVM model for evaluating the mental health status of college students, in which the reduced dimensional data were used as the input of the SVM and the mental health status of college students was used as the output of the SVM. The test sample data were used to verify the effectiveness of the LLE-SVM model for evaluating the mental health status of college students.

Machine learning can be used to discover dependencies between inputs and outputs by learning from training samples to make predictions about unknown inputs. This can usually be expressed as follows: there is some dependency between the variables y and x. And, it is based on the independently distributed sample data sources.

$$S = \left(x_1^2, y_1^2\right), \left(x_2^2, y_2^2\right), \ldots, \left(x_l^2, y_l^2\right). \tag{1}$$

The goal of machine learning is to minimize the desired risk by training with sufficient training samples. In practice, the set of samples available for training is limited and does not achieve the minimization of the expected risk [17]. According to equation (1), it can be found that the final composition of the expected risk is influenced by various aspects such as the error function and the joint probability distribution; therefore, for the classification judgment problem involved in this paper, the empirical risk minimization principle is used.

$$R_{\text{emp}}(\alpha) \approx \frac{1}{l} \sum_{i=1}^{l} L^2\left(y_i^2, f\left(x_i, \alpha\right)\right). \tag{2}$$

The theoretical use of empirical risk to approximate the expected risk assumes that the sample size l is infinitely large and that equations (1) and (2) expressed by mathematical calculation are equivalent when the sample size is infinite.

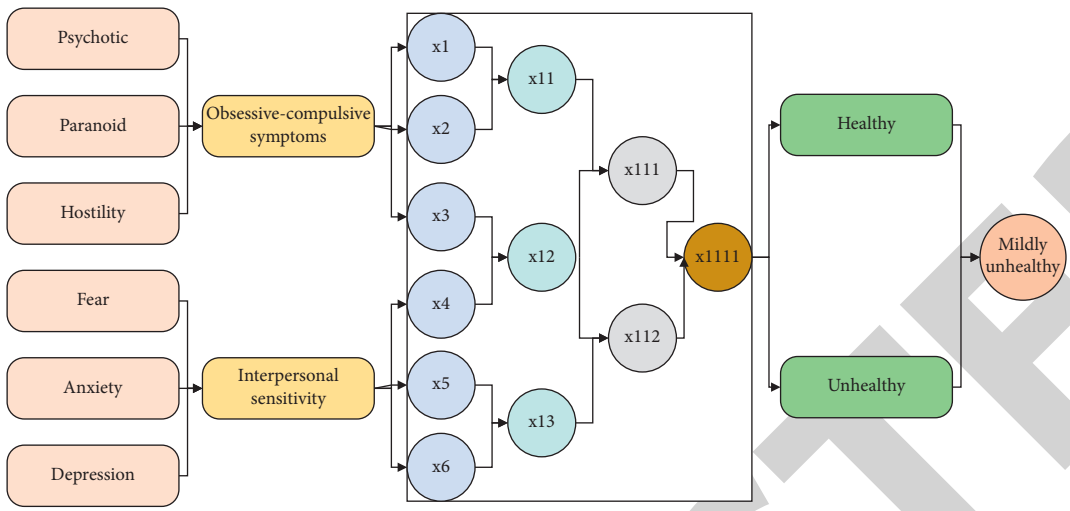

Figure 1: Mental health state evaluation model.

Because the training samples are limited in real life, the method of using such an approximation instead is not fully proven for the time being, so there are some errors between the empirical risk and the expected risk when considering practical problems where the training samples are far less than an infinite magnitude.

According to the foregoing, using the linearly divisible support vector machine classification technique, the linear decision function may be immediately generated in the feature space for linearly divisible data to be processed. However, building a nonlinear decision function directly in the feature space for linearly indistinguishable data is too difficult to execute, and it may also lead to a "dimensional catastrophe" when the dimensionality of the data features is enormous. It can be concluded from the statistical learning theory that the inductive ability of the learning system is independent of the dimensionality of the data; therefore, the data in the low-dimensional input space can be mapped to the high-dimensional space, the classification hyperplane can be found in the high-dimensional space to classify the data, and the decision function can be solved according to the classification hyperplane. The SVM algorithm is based on the idea of first mapping the low-dimensional data samples to the high-dimensional space by the kernel function and then finding the hyperplane that can classify the data samples reasonably in the high-dimensional feature space to achieve the classification of the data.

$$\begin{aligned} k\left(x_i, x_j\right) &= \varphi\left(x_i^2\right)\varphi\left(y_i^2\right), \\ Q(a) &= \sum_{i=1}^{l} a_i - L^2\left(y_i^2, f\left(x_i, \alpha\right)\right), \\ f(x) &= \sin\left(\sum_{i=1}^{l} a_i - L^2\left(y_i^2, f\left(x_i, b\right)\right)\right). \end{aligned} \tag{3}$$

Based on the network behavior of the malicious code detection system, the detection data come from the various types of operational data collected from the network, and the network data have the characteristics of large data volume and strong uncertainty. On the one hand, a large number of various types of data provide the detection system with a lot of valuable information, but on the other hand, there are many duplicate redundant data and features in the network data, the existence of these invalid data not only adds a burden to the storage of the detection system, increasing the cost of hardware storage equipment, but also increases the computational burden of the detection system, affecting the monitoring of real time and, more seriously, the existence of interference data. There is also the possibility of affecting the accuracy of the final detection. Therefore, feature extraction is to select relatively important features from a large amount of network data through some evaluation criteria, reduce the overall feature dimension, thus reducing the complexity of model computation, and improve the training time and detection time while ensuring the detection accuracy.

$$\begin{aligned} x_i^2 &= \frac{x_i - x_{\min}}{x_{\min} + x_{\max}} * (U + L) - L, \\ S &= \frac{1}{n_1} \sum_{i=1}^{l} f\left(a_i - L^2\left(y_i^2, f\left(x_i, b\right)\right)\right). \end{aligned} \tag{4}$$

Data cleaning is an essential technique for resolving data quality issues caused by external influences, such as those mentioned previously. Missing data are filled in, errors are detected, duplicates are filtered out, and consistency is checked, among other things. Filling methods include zero-value filling, mean value filling, and filling according to the probability distribution; for example, the missing of some age fields in the basic information can be filled by the mean value of the whole group; error detection means verifying some data records in some fields in the data content by the rules and the data missing because a certain reason need to be filled by a certain data association. Consistency check is necessary to ensure the consistency of data between multiple data sources when a data field has multiple data sources. For example, the college and major in the basic information of students and the college and major in the data of students' grades should ensure the consistency of data. Meanwhile, the

consistency and the inconsistent data should be corrected by rules or handled manually, as shown in Figure 2.

Each university will establish various business departments according to different functional divisions. Each department's business systems are relatively independent and connected, such as grades and course selection data from the Registrar's Office, consumption, and access control data from the Logistics Department [18]. Therefore, before mining multidimensional student data, it is necessary to physically integrate the data from different data sources with different storage media and different data formats, a process called data integration. Data sources can be divided into online data and offline data according to the form: online data can be obtained directly through the online form of data, such as database data and interface data; offline data is usually obtained in an offline manner, such as excel tables and text files. At present, most of the data in universities are online data, such as consumption, access control, course selection, and grade data. Only very few data are stored in an offline way, such as student registration data when enrolling and class-based information statistics summary data. Data analysis is the interpretation of key information contained in or between data through visual data graphs or metrics such as correlation coefficients. The purpose of data analysis is to get a sufficiently accurate picture of the distribution of the data and the correlation between them, which helps in feature engineering to construct feature vectors with labels characterizing the data.

3.2. Experimental Design for Multidimensional State Evaluation of College Students' Mental Health. The process of model application is to classify unknown samples by the constructed model, and the model is applied to predict sample data with unknown labels only if the accuracy of the model is within an acceptable range. The estimated accuracy of a model is usually assessed using accuracy, where the prediction accuracy of an individual sample is judged by comparing the known labels of the test sample with the results predicted by the model, with the accuracy being the percentage of the test data correctly classified by the model. The test dataset should be independent of the training dataset to avoid model overfitting problems. The model may achieve excellent accuracy when categorizing imbalanced sample class labels by explicitly predicting the test sample as the majority sample class [19]. As a result, it is also important to think about the model's recall, which refers to the model's check-all rate or the proportion of positive cases in the test data that the model correctly predicts. In most classification issues, the F1 value, which is the summed average of the model's accuracy and recall, is used to determine the model's quality.

Although significant progress has been achieved in theoretical research and social services, the building of a mental health care system still lags behind and is unable to satisfy the demands of social growth. As China's comprehensive strength has grown, several institutions have progressively begun to construct a flawless psychological care system to promote students' mental health. At present, all colleges and universities have established mental health centers to provide professional psychological counseling for college teachers and students, and some colleges and universities will regularly conduct psychological assessment activities to comprehensively measure students' current psychological status and provide timely guidance and intervention for students with psychological abnormalities. This chapter further explains the correlation between the psychological assessment dimensions based on the data of the student psychological scale.

To verify the validity of the LLE-SVM for evaluating the mental health status of college students, the SCL-90 data of college students' mental health symptoms of a school entering college in 2020 were selected for the study, and the data of mental health characteristics of each college student were nine dimensions of indicators such as psychoticism, paranoia, hostility, terror, anxiety, depression, obsessive-compulsive symptoms, interpersonal sensitivity, and somatization. The data were categorized as unhealthy. The mental health status of college students was divided into three states: unhealthy, mildly unhealthy, and healthy, and the distribution of the three sample data is shown in Figure 3.

Data preprocessing is a very important part of the data mining process, but data preprocessing is very time- and effort-consuming. Experience shows that if the data preprocessing is done well, then it saves effort in building the model. Because the existence of a high number of noisy data, redundant data, missing data, and so on in the dataset influences the prediction results, we must preprocess the data. There are a variety of reasons for incomplete data in the dataset; for example, the student may have misplaced a card and failed to replace it on time, resulting in missing data for that period, or there may be a mechanical failure resulting in data loss and incorrect entry; all of these anomalies result in a large amount of redundant data in the database, and due to the presence of these incorrect data and vacant data, preprocessing is essential. For the processing of missing values, either direct deletion or interpolation of missing values can be performed. Direct deletion causes partial loss of information and loss of some important data with missing values, especially when the data are small, which is more harmful to the sample; missing value interpolation can be divided into mean interpolation, median interpolation, arbitrary value interpolation, and model interpolation. Missing values may also be a kind of information in some cases and can be captured by adding a feature to capture whether it is missing or not (i.e., adding a column of features, where 1 means it is a missing value NaN and 0 means it is not a missing value), and certain students in the dataset have missing values in some consumption patterns, as shown in Table 1.

For the missing values, different values need to be filled in using different values depending on the nature of each feature. Cartoon student consumption data have much vacant information, and only these missing data need to be populated with 0. Because the public database did not contain the required peripheral physiological signals for the previous studies in this section, it was not possible to conduct a mixed physiological signal emotion classification

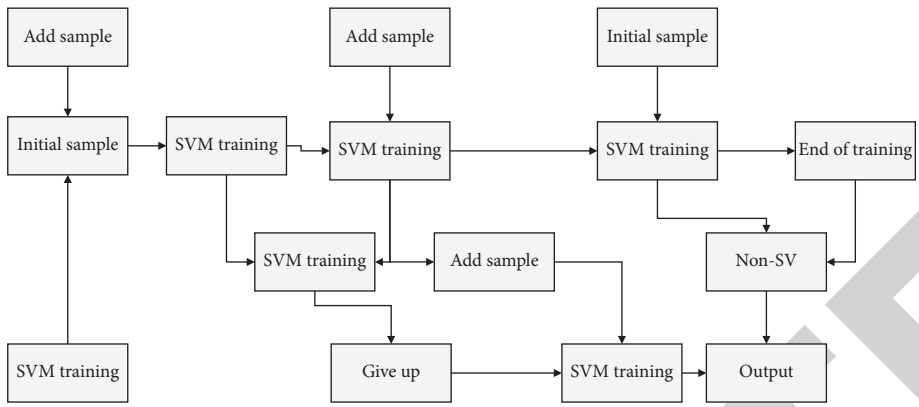

Figure 2: Batch SVM training process.

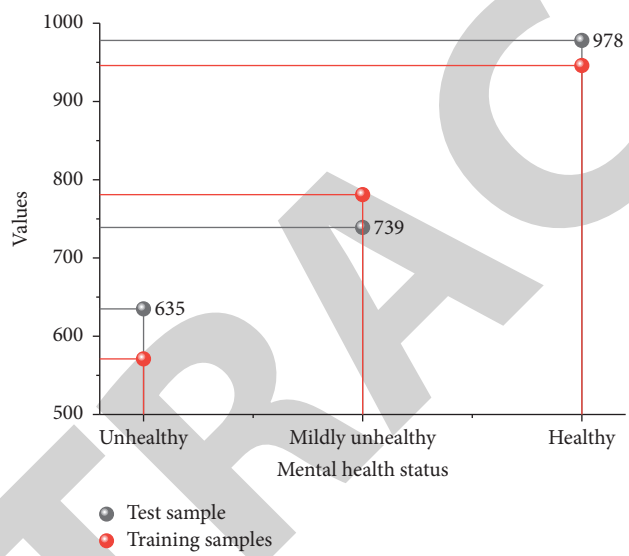

Figure 3: Distribution of training and test sample data.

Table 1: Dataset.

Student ID	1	2	3	4	5
Library	11.52	1.94	12.66	3.57	14.35
Boiling water	7.54	7.59	7.26	6.25	13.35
Academic affairs office	13.18	7.86	12.78	4.46	4.96
Print center	3.08	6.73	1.71	13.12	14.65
School clinic	2.37	2.08	12.34	4.93	9.75
School bus	12.36	12.5	4.89	1.65	2.72
Laundry room	14.67	8.23	8.72	9.29	12.25

study, and the superiority of the spatiotemporal features was facilitated by the algorithmic analysis of the spatiotemporal features in the self-constructed dataset validation. Therefore, in this paper, we first conduct experiments on the mixed physiological signal acquisition of emotions to obtain experimental data and then validate the spatiotemporal features on the EEG signals in the collected experimental data.

The selection of emotional elicitation material is the first step of the whole experiment, and this experiment chose the audiovisual stimulus as the elicitation material, that is, the music + dynamic pictures as the whole elicitation material, giving the subjects visual stimulation from both visual and auditory dimensions. These two approaches require the selection of moving pictures and background music, which is also crucial because it is necessary to consider that different people respond differently to music and pictures, and this experiment requires the selection of background music and moving pictures that can effectively evoke most of the subjects, as well as considering factors such as the main groups of recruited subjects and their living environment. Considering that the subjects were mainly students of the university, music and pictures were selected to stimulate

young students aged 20 to 24 years. Before the formal experiment, some students were selected for evaluation. After the music and pictures were completely selected, the experiment was conducted by playing 14 dynamic pictures for one minute for each emotion induced. The experiment was designed to collect physiological signals from a total of 60 subjects with an average age of about 24 years. When the subjects were recruited before the start of the experiment, the exact procedure and the purpose of the experiment needed to be explained clearly to the subjects, and those who were willing to participate in the experiment were required to sign an informed consent form. EEG, ECG, EDA, and RSP signals were collected during the experiment. The EEG acquisition equipment used was the Enobio system. The ProComp Infiniti device was used for peripheral physiological signals. The EEG signal was acquired at 500 Hz, the ECG signal was acquired at 2048 Hz, and the EDA signal and the RSP signal were acquired at 1024 Hz. The material was played back in a way that gave the subject a more visual stimulus, and an ASUS computer and a Holly 75-inch display were used to play the stimulus material. The experimental flow is shown in Figure 4.

Before the start of the experiment, it is necessary to explain the purpose of the experiment and the whole process to the subject. After the subject is informed, he/she needs to fill in the relevant information form and sign the informed consent form. The equipment will also need to be worn, and the equipment will need to be commissioned. The room must be darkened before the official experiment starts so that the emotions may be created more effectively. The collector must be aware of the signals acquired during the experiment, noting any noticeable anomalous signals or if the equipment is malfunctioning in order to assess whether the data has to be recollected [20]. The experiment started with the official admission into the physiological signal elicitation acquisition phase, and it went like this: the patient stayed quiet throughout the two-minute blackout; during this time, the subject's basal signal was captured. The stimulus material was presented after 2 minutes of basal signals had been recorded. There were four different sorts of content, each of which was played for one minute, followed by a two-minute black screen to allow the subject to rest and recuperate.

4. Analysis of Results

4.1. SVM Multidimensional State Data Dimensionality Reduction Results. The mental state perception model is to perceive students' mental state information based on their online behavioral data for psychological prediction. The output of the model is the categorization of the mental state information, that is, the classification of students' mental states according to the online behavioral data, the form of classification is binary, and the content is two kinds of student's mental health status, healthy and unhealthy, so the mental state perception model essentially belongs to the category of binary classification model in supervised learning. A preselection-elimination mechanism is used to construct the mental state perception model. Preselection means that several different types of classification models with classification ability are selected first, given the same training data set for training, and finally, the model with the best output is selected as the final perception model based on the average results of multiple experiments, and the rest of the models are eliminated. The reason for choosing this method to construct models is that the network behavior data used in this paper is a kind of data that has not been studied and applied, so some models need to be preselected to compare and analyze the advantages and disadvantages of different models' performance for this kind of data before finally determining the perceptual models.

The data used in this experiment came from two parts, respectively, the psychological status data obtained from the questionnaire platform and the web access log data obtained from the log server database; all the data have strict control over the privacy of students not to be leaked, such as psychological status data only have the student ID data, no student's name, major, class, and other information, so it only can be treated as a unique identification of the existence. The same applies to the weblog data, where the data collected are limited to the type of website visited by a student ID, the name of the website, and so on and do not relate to the specific content of the visit, as shown in Figure 5.

As can be seen in Figure 5, the P and R values corresponding to each model and the F score on the fold are marked, and all models except the SVM model have an F score above 0.6, and the optimal GA-RF model has an F score reaching 0.765, indicating that the model trained by the feature dimension data of the network behavior can achieve the judgment of the population's internal and external tendencies. Second, compared to the results of the model without the extraction of features by the genetic algorithm, all values of the model group with the genetic algorithm have improved, indicating that the data of these eight feature dimensions extracted by the genetic algorithm are more representative compared to the data without processing when based on the internal and external tendency labels, which are the regularity of WeChat, the degree of WeChat dependency, the regularity of WeChat posting friend regularity, microblog regularity, video viewing regularity, video viewing gauge dependency degree, reading regularity, and reading dependency degree.

By further statistical analysis of the data of these 5 dimensions, it was found that the depressed people have lower dependence on shopping and web, while the dependence on listening to music, regularity of using map websites, and dependence on games have higher values because the depressed people are cold not only to interpersonal relationships but also to life, so they are not interested in buying new things and maintaining. Instead, they are more dependent on music and games than others. For a normal person, there are more or less a few days in a month-long cycle when he or she goes out with friends to have fun, and then he or she will use the map. The use of the map-type website is regular, as it is not used every day, as shown in Figure 6.

The main purpose is to construct, analyze, and tune the models of three different kinds of students' mental health status, aiming to be able to grasp the information of students' mental status more comprehensively and accurately through

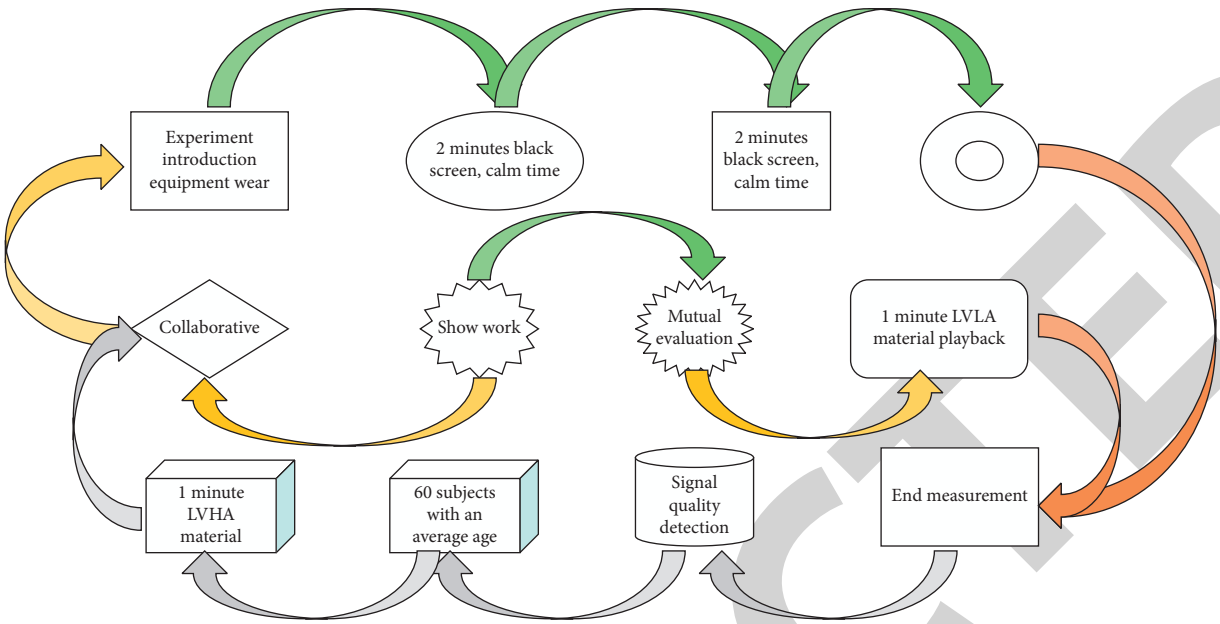

Figure 4: Experimental flowchart.

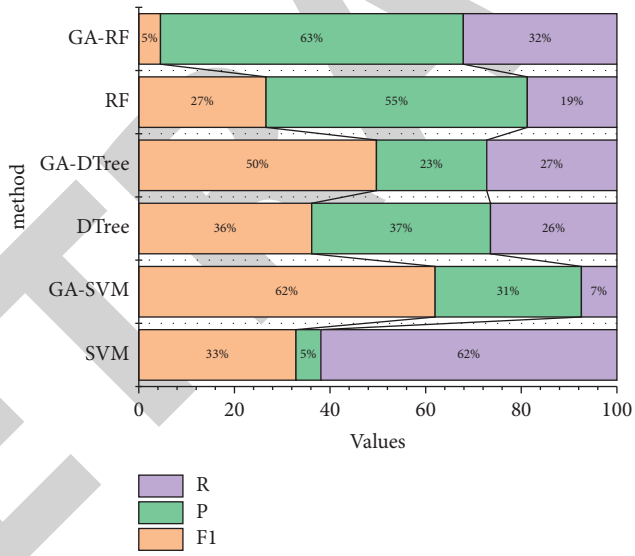

Figure 5: Evaluation of different classification models for internal and external tendencies.

their online behavior data at school. In this chapter, the labeling and feature dimension parts of the sample data set are processed in conjunction with Chapter 4 so that they can be used for the subsequent model training, and then the evaluation indexes of the classification model are selected. Firstly, when building the model, a series of classification models are designed for experiments by combining the contents of Chapter 3, and the results of multiple models are briefly analyzed by selecting the model with the best output for the parameter optimization work. After a series of steps down, the best model algorithm is selected in the case of horizontal comparison of different classification models, and the best model algorithm is selected in the case of vertical comparison of different optimal combinations of parameters and different combinations of parameters, and finally the good output of the model of the perceived mental health status of students was ensured.

4.2. Results of Multidimensional State Evaluation of College Students' Mental Health. As shown in Figure 7, the results are binary logistic regression results for univariate and multivariate factors, and the depression subscale of the symptom self-rating scale yields results as continuous-type values indicating different levels of depression, so it is also necessary to use multiple linear.

The goal of the student psychological state perception model is to be able to predict student psychology in real time

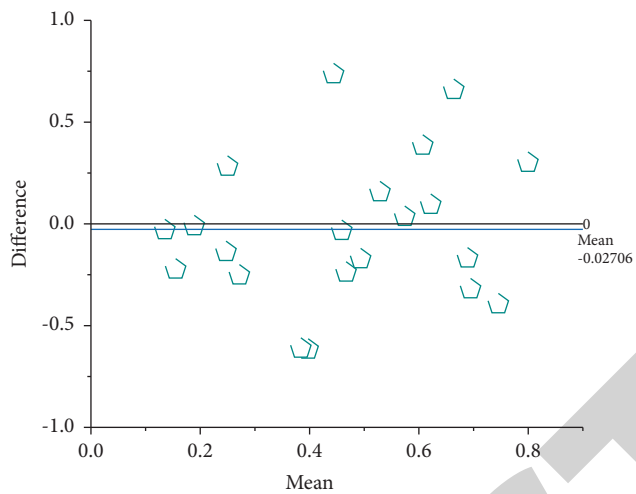

Figure 6: F1 values for a different number of iterations.

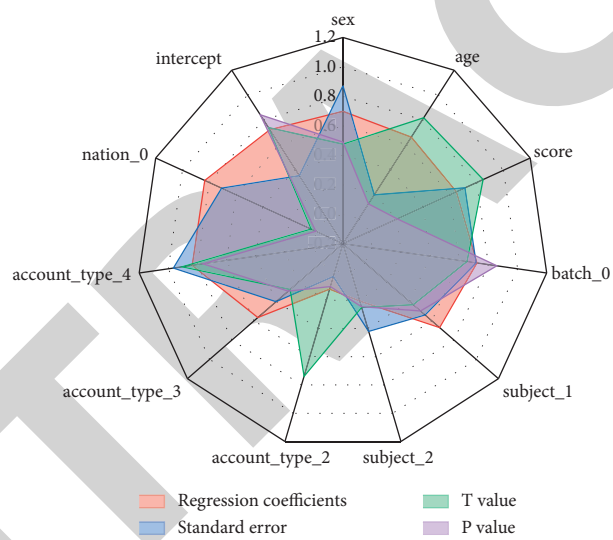

Figure 7: Regression analysis.

based on the behavioral data generated by students at school, so this section aims to construct dynamic features in three major areas from the behavioral data generated by students at school, including student consumption features, student behavior features, and social relationship features. The behavioral data generated by students at school include daily consumption data, containing the time, place, amount of consumption, and so on; library access data, containing access status, time, and so on; and course selection data, containing the semester of course selection, course type, and so on. The human psychological state has stability within a certain period; that is, no human psychological state will change much over a while, but to throw out the possible influence of expost behavior on the exante state, this topic only selects the behavioral data within the period from the entrance of new students to the test, which is also beneficial to the model in the application of the demand on the data period, as shown in Table 2.

The LLE algorithm involves two parameters: the embedding dimension a and the nearest neighbor parameter K. The magnitude of these two parameters directly affects the effect of college students' mental health status evaluation. When various K and d values are used as values, the parameters of SVM are set as penalty parameters, radial basis kernel function parameters, and the accuracy rate of college students' mental health status rating. The greatest accuracy percentage for assessing the mental health status of college students was 96.5 percent. The correlation between each factor and depression status was analyzed using multifactor binary logistic regression; finally, the scores of the depression subscale of the original symptom self-assessment scale were used as the target, that is, the magnitude of depression; and whether each factor was an influential factor in the degree of depression was investigated using a multivariate logistic regression mode.

Table 2: Behavioral data used in the student psychological state perception model.

Type	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8
$d = 2$	43	29	19	28	16	11
$d = 3$	21	10	26	26	22	31
$d = 4$	38	10	16	11	43	28
$d = 5$	15	22	47	13	24	41
$d = 6$	45	12	47	23	45	12
$d = 7$	21	17	33	27	30	48
$d = 8$	25	24	15	20	44	35

5. Conclusion

In the early stage, through reading and detailed analysis of a large amount of relevant literature, we have a certain understanding of the current status of research on a mental state based on network data at home and abroad, and by comparing the advantages and disadvantages of different research methods, it is clear that the current research on mental state mainly focuses on the analysis of users on social network platforms; while there are fewer methods suitable for the analysis of mental health in a more closed environment like school students, in this context, this paper proposes the concept of a mental state perception model based on school students' online behavior data. The process of knowledge discovery is also the process of discovering the value of data, and data mining is the key link in the process of knowledge discovery. By analyzing and studying each step of the overall process of knowledge discovery, we have a grasp of the overall process for the subsequent establishment of the mental state perception model. It is clarified that there needs to be a process of data preparation stage before data mining and a process of result expression and interpretation stage after data mining. An improved SVM incremental learning algorithm that effectively utilizes the results of the above preselection algorithm and combines KKT conditional judgments and error-push strategies is proposed. In the case of using the above data preselection method, a new SVM incremental learning algorithm combining the traditional KKT conditional and error-push strategies is proposed. This algorithm improves the learning speed by effectively using the results of data preselection; by adding the relevant error score vectors to the added vectors, the detection accuracy is improved relative to the traditional incremental learning.

Data Availability

The data used to support the findings of this study are included within the paper.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] G.-H. Dong, Z. Wang, H. Dong et al., "More stringent criteria are needed for diagnosing internet gaming disorder: evidence from regional brain features and whole-brain functional connectivity multivariate pattern analyses," *Journal of Behavioral Addictions*, vol. 9, no. 3, pp. 642–653, 2020.

[2] Y. A. Suh and M.-S. Yim, "A worker's fitness-for-duty status identification based on biosignals to reduce human error in nuclear power plants," *Nuclear Technology*, vol. 206, no. 12, pp. 1840–1860, 2020.

[3] J. Wu, P. Guo, Y. Cheng, H. Zhu, X.-B. Wang, and X. Shao, "Ensemble generalized multiclass support-vector-machine-based health evaluation of complex degradation systems," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 5, pp. 2230–2240, 2020.

[4] A. Baghdadi, Y. Aribi, R. Fourati, N. Halouani, P. Siarry, and A. Alimi, "Psychological stimulation for anxious states detection based on EEG-related features," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8519–8533, 2021.

[5] J. S. Rahman, T. Gedeon, S. Caldwell, R. Jones, and Z. Jin, "Towards effective music therapy for mental health care using machine learning tools: human affective reasoning and music genres," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 11, no. 1, pp. 5–20, 2021.

[6] K. Zhao, Y. Ding, Y. Han et al., "Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer's disease: diagnosis, longitudinal progress and biological basis," *Science Bulletin*, vol. 65, no. 13, pp. 1103–1113, 2020.

[7] E. Barenholtz, N. D. Fitzgerald, and W. E. Hahn, "Machine-learning approaches to substance-abuse research: emerging trends and their implications," *Current Opinion in Psychiatry*, vol. 33, no. 4, pp. 334–342, 2020.

[8] P. Kumari, P. Gupta, A. K. Piyoosh, B. Tyagi, and P. Kumar, "COVID 19: impact on mental health of graduating and post graduating students," *Journal of Statistics and Management Systems*, vol. 24, no. 1, pp. 67–79, 2021.

[9] D. Dodell-Feder, L. M. Tully, E. Dudek, and C. I. Hooker, "The representation of mental state information in schizophrenia and first-degree relatives: a multivariate pattern analysis of fMRI data," *Social Cognitive and Affective Neuroscience*, vol. 16, no. 6, pp. 608–620, 2021.

[10] P. Mohan and I. Paramasivam, "Feature reduction using SVM-RFE technique to detect autism spectrum disorder," *Evolutionary Intelligence*, vol. 14, no. 2, pp. 989–997, 2021.

[11] H. Akbari, M. T. Sadiq, and A. U. Rehman, "Classification of normal and depressed EEG signals based on centered

correntropy of rhythms in empirical wavelet transform domain," *Health Information Science and Systems*, vol. 9, no. 1, pp. 1–15, 2021.

[12] J. Shin and M. J. Gierl, "More efficient processes for creating automated essay scoring frameworks: a demonstration of two algorithms," *Language Testing*, vol. 38, no. 2, pp. 247–272, 2021.

[13] J. S. Pimentel, R. Ospina, and A. Ara, "Learning time acceleration in support vector regression: a case study in educational data mining," *Stats*, vol. 4, no. 3, pp. 682–700, 2021.

[14] B. ElOuassif, A. Idri, M. Hosni, and A. Abran, "Classification techniques in breast cancer diagnosis: a systematic literature review," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 9, no. 1, pp. 50–77, 2021.

[15] S. Gupta, M. Kaur, S. Lakra, and Y. Dixit, "A comparative theoretical and empirical analysis of machine learning algorithms," *Webology*, vol. 17, no. 1, pp. 377–397, 2020.

[16] J. G. Rodríguez-Ruiz, C. E. Galván-Tejada, and S. Vázquez-Reyes, "Classification of depressive episodes using nighttime data; a multivariate and univariate analysis," *Programming and Computer Software*, vol. 46, no. 8, pp. 689–698, 2020.

[17] S. Yang, H. Xiong, K. Xu, L. Wang, J. Bian, and Z. Sun, "Improving covariance-regularized discriminant analysis for EHR-based predictive analytics of diseases," *Applied Intelligence*, vol. 51, no. 1, pp. 377–395, 2021.

[18] W. Hu, G. Huang, L. Li, L. Zhang, Z. Zhang, and Z. Liang, "Video-triggered EEG-emotion public databases and current methods: a survey," *Brain Science Advances*, vol. 6, no. 3, pp. 255–287, 2020.

[19] Y. Zheng, X. Zhao, and L. Yao, "Mixture kernel density estimation and remedied correlation matrix on the EEG-based copula model for the assessment of visual discomfort," *Cognitive Computation*, vol. 12, no. 6, pp. 1130–1143, 2020.

[20] N. Zemmal, N. Azizi, M. Sellami et al., "Particle swarm optimization based swarm intelligence for active learning improvement: application on medical data classification," *Cognitive Computation*, vol. 12, no. 5, pp. 991–1010, 2020.