In the new situation of modernization, there is an influx of diverse social thinking. At the same time, coupled with the influence of COVID-19, which has swept the world, the ideological and psychological space of college students has been greatly impacted. In this context, the ideological and psychological health of college students is an important value for the education of college students. To be specific, as an important place for cultivating college students, colleges and universities should pay attention to students’ thoughts and ideas from their hearts. In addition, colleges and universities should give full play to the role of educational psychology in colleges and universities and actively promote the synergistic development of all educational sectors in schools, so as to promote the realization of the goal of education in the new era. In recent years, a series of mental health problems such as anxiety, depression, low self-esteem, and interpersonal sensitivity have become frequent among college students, and some have even developed suicidal ideation. This has a very serious negative impact on individuals, families, and society. Therefore, if the mental health problems of college students can be detected early, the relevant school departments and counselors can provide timely and targeted help to such students. At the same time, these at-risk students can receive early treatment, thus reducing the harm. As a result, it is quite valuable to find an effective method to identify students with mental health problems. Traditionally, researchers have used questionnaires to survey students about their mental health. However, this approach has the disadvantage of being easily concealed and inefficient. In recent years, researchers have begun to use weblogs to identify students with mental health problems, but this approach still has shortcomings. First of all, they still use questionnaires to obtain labels. In addition, students’ psychological activities may not only be reflected in their online behavior but also in their other daily behaviors. Big data in higher education plays a crucial role in analyzing and identifying students with psychological abnormalities. As a result, this research mainly extracts the behavioral characteristics of students by cleaning and transforming a large amount of disorganized student school data based on the educational data collected from school cards, academic affairs systems, access control systems, and related business systems. What is more, this study further analyzes the differences in behavioral characteristics between normal and abnormal students through hypothesis testing and finally establishes a model to identify abnormal students and evaluate the results.

1. Introduction

With the rapid development of information technology, artificial intelligence and big data are influencing various industries [1–3]. Nowadays, digitalization and informatization in colleges and universities have been promoted and improved, which has increased the amount of data in research results, students’ daily behavior, academic performance, and network usage [4]. In this context, nowadays, there are relatively complete application systems among different departments of the university, such as academic application systems, access control systems, one-card data systems, and library circulation systems. Each of these systems stores a large amount of student data [5]. However, these data are only stored in the database and are not exploited. Therefore, one of the most important issues is that how to use a large amount of data effectively, analyze the information in the data through data mining technology,
make the data into useful information, and accelerate the level of information construction in universities [6]. At the same time, with the arrival of big data, data mining technology has also flourished. Since the 20th century, data mining technology has been widely used in various industries [7]. In the field of education, many universities at home and abroad have started to use data mining technology to analyze educational data in order to understand students’ various dynamics in time.

While China has achieved rapid economic development, the mental health of the population also needs attention [8]. To be specific, numerous psychologically related disorders are present in the population. These mental illnesses have many negative effects and are pervasive and harmful in nature [9]. As a result, mental health has become an important public health issue in our country. At the same time, with the rapid development of the knowledge economy and the increase in the popularity of higher education, the number of college students is increasing, and the number of students with psychological problems is also increasing [10, 11]. In recent years, there have been many cases of university students’ psychological problems affecting their studies, dropping out of school, and even committing suicide. College students, as part of the high level of talent, have been considered to be outstanding members of society and are expected to have strong psychological qualities [12]. However, the reality is very disappointing. Specifically, the stresses of academics, social relationships, and employment have led to many tragedies that have left college students mentally exhausted [13]. As a result, the timely identification of students with psychological abnormalities has become one of the most essential and challenging issues for universities [14].

Mental health problems can have serious detrimental effects. For the individual, mental health problems can have a number of negative effects on the individual [15]. More seriously, some mental health problems may make individuals less adaptable to society and may even pose a serious threat to their physical health [16]. In addition, mental health problems can be a psychological burden to the family and an economic burden to the family. For society, mental health problems can disrupt public order, consume social resources, and increase the burden of disease on society [17].

As a result, mental health problems account for the majority of the negative effects of all illness-related hazards [18]. However, due to a lack of mental health knowledge and discriminatory attitudes toward the mentally ill, a large number of people with mental health problems hide the truth when answering the questionnaires [26]. This can lead to a large discrepancy between the test results and the real situation. Last but not least, the mental health status of each student changes over time, usually lasting half a month or a month. As a result, if researchers want to know the mental health status afterward, they need to take the test again, which is both labor-intensive and costly.

The outbreak of COVID-19 has affected all walks of life, and the teaching format of schools has changed from offline to online. Therefore, colleges and universities have become important places for the prevention and control of the epidemic, which makes it more difficult for teachers to grasp the psychological state of mind of students [28]. This urgently requires the ideological and political education work in colleges and universities to insist on the combination of educating the mind and educating people. To be specific, colleges and universities should not only do a good job in ideological and political education but also pay attention to cultivating a sound personality and great psychological quality of college students [29]. As a result, under the background of a diversified society, college students face many problems such as academic pressure, employment pressure, interpersonal relationship trouble, and wrong thinking at home and abroad. In addition, many pressures make the desire to be valued and to have an equal communication status while having a strong sense of self [30].
Psychological counseling is a scientific psychological treatment method and also an important supplement to ideological and political education methods, which plays an important role in improving the educational effect and cultivating a positive and optimistic mindset of college students. In addition, ideological and political education in colleges and universities should meet the inherent demand of improving its effectiveness [31]. Modern ideological and political education should not only adhere to Marxism as the theoretical basis but also draw on the excellent achievements of different disciplines. The application of psychological guidance to ideological and political education in colleges and universities is not only an expansion and innovation of ideological and political education methods but also an effective reference for the new problems in ideological and political work [32]. In practice, ideological and political education should continuously improve its service function and highlight its social function [33].

With the advent of the Internet era, digital campuses have flourished. In this context, the campus life and learning data of college students are gradually recorded in the student management system. These massive data contain a huge amount of information and provide a better opportunity to understand students’ behaviors. This study considers that the psychological status of college students is reflected by their behavioral status in school and is recorded in the relevant system. As a result, the purpose of this study is to identify students with mental health problems by analyzing the massive amount of data on campus and mining the behaviors related to mental health, so that machine learning algorithms can be effectively used to identify students with mental health problems. Furthermore, with the use of information-based big data, counselors can focus on these students based on the identified results and provide them with psychological guidance.

2. Technology Related to Big Data

2.1. Data Mining. With the rapid development of computer and network technologies, many human activities are no longer limited by time and space, such as telecommuting and information sharing. Increasingly sophisticated database technologies and pervasive data applications have led to the exponential growth of human data. In this context, people have started to explore the potential information from these massive data, resulting in a variety of data mining techniques. Among them, educational data mining is a new technology created to make full use of the huge data streams that emerge from the operation of digital campuses. To be specific, this technology integrates, classifies, and refines a large amount of data through the comprehensive application of various data mining technologies, thus making it very valuable in the ever-evolving education. As a result, educational data mining refers to a multidirectional and cross-cutting research area that encompasses three main disciplines: education, computer science, and statistics. At the same time, their combination has given rise to research areas similar to or related to educational data mining, such as computer-based education, machine learning, data mining, and learning analytics.

As shown in Figure 1, data mining can be generally divided into three phases: data collection and preprocessing, data analysis and mining, and evaluation and visualization of results. The first stage is data collection and preprocessing. This process mainly includes the collection of target data, the establishment of a data warehouse, data cleaning, data integration, statute and transformation, and other operations. To be specific, data collection and preprocessing are the foundation and key to the whole data mining process. In other words, the quality of data can determine the effect of data analysis and mining. By establishing a unified data warehouse, noise and missing data can be removed or filled with certain rules. Data integration is the unified management of data to facilitate subsequent use. Data preprocessing is a tedious and heavy workload part, but it also lays a solid foundation and support for the subsequent mining analysis. The second stage is the analysis and mining of data. This process is the core part of the whole data mining process. After using the preprocessed data, the data can be extracted, filtered, and modeled by using certain algorithms according to the data mining objectives. The third stage is the evaluation and visualization of the results. Finally, the mining model is evaluated to measure whether the model fulfills the expected tasks and objectives. The results of data mining are analyzed and filtered through certain visualization techniques, and useful results are selected and presented.

2.2. Random Forest Algorithm. The random forest algorithm mainly applies a decision tree as the base classifier, combined with the construction of Bagging, and incorporates the selection of random attributes. As a result, the detailed workflow of the random forest algorithm is illustrated in Figure 2. To be specific, the random forest algorithm can decide the class of samples by voting, and each base classifier has one vote. After that, the category of the sample can be determined based on the principle of majority rule. The randomness of the random forest is reflected in the randomness of attribute selection. Specifically, for each node of the base classifier, k attributes are first randomly selected from the set of attributes of the current node to form a subset. Then, the best splitting attribute is computed from the subset according to the attribute selection algorithm, instead of finding the best attribute from all attributes of the current node as in a traditional decision tree. As a result, the introduction of the parameter k reflects the randomness of the random forest algorithm. The random forest algorithm can effectively solve the problem of overfitting easily in decision trees. When dealing with high-dimensional data, it is not necessary to perform dimensionality reduction and feature selection before training because of the random property. In addition, since each tree is trained independently, parallel training can be realized, thus speeding up the training speed. At the same time, the error of the dataset can be balanced according to the error generated by each tree. However, when the noise of the dataset is high, overfitting may occur.
2.3. Description of the Hypothesis Test. Hypothesis testing is a common method used in statistics to determine whether a random variable is consistent with a scientific hypothesis. Its main idea is the small-probability counterfactual method. To be specific, small probability generally refers to the probability of occurrence of 1% or 5% in a single experiment. The counterfactual method is to first formulate a hypothesis and then test it by a statistical method of sampling studies. In fact, the theory and method of determining the validity of the hypothesis make up the hypothesis testing. Figure 3 shows the confidence level and rejection domain of the hypothesis test.

2.4. Logistic Regression Model. Logistic regression is the transformation of a linear model into a nonlinear model to predict the actual test values and output as much as possible for a given data set. Although it is called regression, the actual solution is a classification model. Compared with other algorithms, logistic regression is simple, interpretable, and fast to train. As a result, this algorithm is mainly suitable for linear classification problems. The main idea is to fit the decision boundary as closely as possible and output the predicted values. The main idea of logistic regression is to transform the results of linear regression into the...
classification of predicted values by mapping them to functions. As a result, the sigmoid function as illustrated in Figure 4 is usually chosen as the mapping function, and its expression is

\[ g(z) = \frac{1}{1 + e^{-z}}, \]

where \( z \) refers to the input value.

The model classification results can be optimized by the loss function of the model. When solving for the parameters in the model, the loss function is minimized by using the gradient descent method. The logistic regression classification model is simple to implement, computationally small, and easy to understand and implement. However, when the feature space is large, the classification performance is affected.

3. Case Study of Basic Educational Psychology

3.1. Algorithm Framework. The overall framework of the educational psychological problem identification algorithm based on big data technology is shown in Figure 5. The whole algorithm process is divided into three parts, including data acquisition and preprocessing, feature extraction, and model training and recognition. In the data acquisition and preprocessing stage, four data sources are acquired, namely, weblogs, access control data, achievement data, and consumption data. In the feature extraction stage, relevant features such as students’ online patterns and abnormal consumption scores can be extracted from the four data sources. In the model training and identification stage, the optimal classifier is selected among five common classification algorithms.

With the rise of digital campuses, more and more student behavior data are stored. These data have two characteristics, one is a large amount of data and the other is the complexity and diversity. So far, although various student management systems have been established in various universities in China and abroad, the data collected by these systems are still not well utilized. As a result, it is quite necessary to understand and analyze these data and establish relevant data models.

3.2. Student Behavior Characteristics. In the mining of educational data in colleges and universities, most scholars tend to focus only on the characteristics related to students’ performance. This research focuses on extracting a comprehensive set of behavioral characteristics of students from the Big Five personality traits theory. Based on the original data, this study further extracts and quantifies the characteristics of students’ behaviors, aiming at extracting the characteristics that can reflect the students’ behaviors in school more intuitively. In addition, based on the basic data of students, the indexes can be established by referring to psychology and pedagogy in order to extract the characteristics of students. In addition, this research constructs a better behavioral profile to measure students’ performance in school, so that we can detect abnormal students from their daily behavior.

Indeed, students with psychological abnormalities manifest in different ways in the relatively open and free environment of the university. Most students focus on the Big Five personalities of openness, responsibility, extraversion, agreeableness, and neuroticism (Figure 6). However, due to limitations in school data, this paper can only analyze certain traits to compare and contrast the differences between psychological abnormalities and normal students in order to better identify students with possible psychological problems from a large number of students promptly.

3.3. Social Relationship Mining. Traditional discovery of buddy networks is usually done by hypothesis testing or clustering. However, hypothesis testing requires the assumption that cooccurrence between students is independent, while clustering results are strongly influenced by the cluster centroids. Therefore, the main idea of this study is that the number of cooccurrences between two students is proportional to the probability of two students becoming friends. In other words, this study uses association rules to identify friends. On campus, it is assumed that if two students are friends, they will often be in the same place together, such as eating or spending money together, going to the library together, and so on. This is consistent with the pattern of friendships between college students. As shown in Figure 7, if students X and Y are friends, then students X and Y will be seen together many times.

As information technology and databases continue to evolve and mature, the size of data is exploding. There are more and more unbalanced data in various industries. However, since most classification algorithms are based on global balanced data, the prediction results of most unbalanced data are not promising. Therefore, it is necessary to equalize the data. Currently, there are two main types of unbalanced data processing. The first one is to resample the data from the data level to eliminate the majority samples or expand the minority samples. This can change the distribution of the original data samples and thus improve the imbalance of the data. The other is an improvement of the classification model algorithm. Specifically, the original
The traditional way of a questionnaire survey on mental health problems has the problems of easy concealment and small scale. In recent years, a method of identifying mental health problems based on Internet logs has emerged. This method compensates for the shortcomings of questionnaires, but students’ behaviors on campus are diverse. To be specific, Internet behavior is only a part of students’...
behaviors, and it is not enough to reflect all of their psychological activities. In addition, the method of identifying mental health problems based on Internet logs still uses questionnaires to obtain labels. Therefore, the labels are still unreliable. In this paper, we propose a model for mining students’ social relationships in school by preprocessing a large amount of messy data and identifying students with psychological abnormalities. Based on this model, this research can extract other behavioral characteristics of students in school and analyze the difference between the performance of psychologically abnormal students and normal students in school. Furthermore, based on the preliminary data processing, a model for identifying students with psychological abnormalities was established to identify students with psychological abnormalities from a large amount of data.

However, the work done and the proposed method in this paper have the following areas for improvement in future research. In the model for identifying students with psychological abnormalities developed in this paper, due to the highly uneven data, the use of dichotomous classification requires data equalization to achieve better identification results. The one-category model is simple, but it does not contribute to the accuracy of the model. Therefore, further research is needed to establish a more convenient and accurate model for identifying students with psychological abnormalities based on big data from university education.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The research was supported by 2021 National Social Science Fund General Project "Research on the Coupling Mechanism of Cultural Revitalization of Traditional Villages in Yunnan, Guizhou, and Guangxi and High-Quality Development of Rural Tourism" (21BMZ073); 2021 Guangxi University Young and Middle-Aged Teachers’ Basic Research Ability Improvement Project “Research on the Influence Mechanism and Improvement Path of Farmers’ Happiness in the Background of Rural Revitalization” (2021KY0825); 2018 Guangxi Zhuang Autonomous Region Philosophy and Social Science Planning Project “Guangxi Rural Culture in the Process of Tourism Urbanization: Research on the Coordinated Development of Ecology and Tourism” (18CGL001); and 2021 Research Project on the Theory and Practice of Ideological and Political Education for College Students in Guangxi “Research on the Influence Mechanism and Improvement Path of Guangxi College Counselors’ Happiness in the New Era” (2021LSZ089).

References


[31] B. Cheng, J. Huang, Z. Guo, J. Li, and H. Chen, “Towards sustainable construction through better construction and demolition waste management practices: a SWOT analysis of