Research Article

Data Mining Algorithm for College Students’ Mental Health Questionnaire Based on Semisupervised Deep Learning Method

Maoning Li

Xi’an Jiaotong University City College, Shaanxi, Xi’an 710018, China

Correspondence should be addressed to Maoning Li; limaoning@stu.xjtu.edu.cn

Received 18 May 2022; Revised 20 June 2022; Accepted 23 June 2022; Published 1 August 2022

Academic Editor: Lianhui Li

Copyright © 2022 Maoning Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, there are many cases of college students who have psychological problems affecting their studies, dropping out of school, and even committing suicide. College students, as a part of high-level talents, have always been regarded as outstanding members of society. They default to have strong psychological qualities, but the reality is disappointing. Various pressures such as academics, social relations, and employment make college students the mental exhaustion has led to many bloody tragedies. The timely detection of psychologically abnormal students has become one of the most concerned and thorny issues in major universities. By constructing a mental health state perception model for college students and optimizing the model parameters, it can be seen that the $f$ score of the internal and external tendency model has increased by 3.3%, the $f$ score of the depression binary model has increased by 2.5%, and the anxiety binary model of the $f$ score has increased by 2.5%. The score increased by 8%. The established model has an obvious effect and can quickly analyze the difference between the behaviors of psychologically abnormal students and normal students in school and also provide a management decision-making basis for college student managers and psychological counselors.

1. Introduction

The frequent dropout and suicide incidents in recent years have proved that many students have more or less psychological problems, and these psychological problems have aroused widespread concern in society. At the same time, at all stages of life, people may have psychological problems due to various complex reasons. Early detection of these psychological problems can play a huge role in protecting personal safety. Because of the rapid development of disciplines such as computer science and mathematics, technologies such as deep learning, data mining, and big data are also making rapid progress and are increasingly integrated into people’s daily lives. More and more algorithms are emerging in the field of computer science to solve certain problems and in many ways outperform traditional methods. For example, traditionally, people use indicators such as degree centrality and edge betweenness to analyze social networks, which can better describe some characteristics of social networks but usually only express part of the information in social networks and may contain more noise, and network representation learning based on deep learning solves this problem very well, which usually captures more information in the network for more detailed analysis and identification. This research uses network science, deep learning, network representation learning, other technologies, and students’ social network data to identify students who are more likely to have psychological problems. Accurate identification can give families and schools the opportunity to intervene in students as early as possible and prescribe the right medicine to solve students’ problems. Helping psychologists provide psychological counseling can largely avoid the deterioration of students’ psychological problems; reduce school dropout, suicide, and other incidents; and reduce social tragedies. At the same time, a new way to efficiently utilize multiview network data is proposed; more potential social information between views is preserved; and the problem of label imbalance that is common in such data is proposed. It is proposed to identify students’ psychological problems through deep learning technology,
which also provides a direction for the identification and research of students’ psychological problems in the future [1–7].

2. Related Work

In recent years, many experts and scholars have realized the importance of educational big data. They have used data mining technology to analyze and mine the data generated by students in school and have achieved remarkable results. They have put the data mining results into education and teaching. A person’s psychological characteristics are often expressed through daily life behaviors and routines. Therefore, recently, researchers have begun to try to dig out information that can reflect their mental health status from the daily behavior data of college students. Several studies have shown a strong relationship between mental health status and online behavior. DongNie et al. explored the relationship between search behavior and personality traits and further attempted to determine how search behavior can be used to identify personality. They collected two data sets: one from a questionnaire on 16 personality factors and the other from web access logs from Internet gateways. By calculating the correlation coefficient between individuals’ search behavior and personality, some interesting patterns were found; there are several specific behaviors that have a strong correlation with personality, such as directory index search, knowledge search, dwell time, keyword usage, click habits, and so on. Through regression analysis, most personality dimensions can be predicted by search engine behavior. AngLi et al. propose an algorithm for predicting mental health problems through network usage behavior. They recruited 102 college students and used the SCL-90 questionnaire to conduct a psychological survey. Through the results of the questionnaire, they obtained the mental health level of college students (10 dimensions) and conducted statistical analysis on the online behavior of college students. Based on web usage behavior, a computational model for predicting the scores of each dimension of SCL-90 is established. The results show that the fluctuation range of the Pearson correlation coefficient between the predicted score and the actual score of each dimension is between 0.49 and 0.65, and the fluctuation range of the relative absolute error is between 75% and 89%. Zangane and Hariri et al. explore the role of emotional factors in doctoral students’ online information retrieval. Their study sample, 50 PhD students, aggregated information by observing user facial expression records, Morae software log files, and pre- and postsearch questionnaires. The findings suggest that there is a significant relationship between emotional expression and the individual characteristics of searchers. Searcher satisfaction with search results, Internet search frequency, search experience, interest in search tasks, and familiarity with similar searches were associated with increased happiness. An examination of user emotions during search shows that users with happy emotions spend a lot of time searching for and viewing search solutions. ChangyeZhu and BaobinLi et al. proposed a new method to detect depression through time-frequency analysis of network behavior. They recruited 728 graduate students, obtained their depression scores through the Zung self-rating depression scale (SDS), and then collected digital records of their online behavior. Through time-frequency analysis, they built a classification model to distinguish between high and low SDS groups, and a predictive model that more accurately identified the mental state of the depressed group. The experimental results show that both the classification model and the prediction model can better reflect the change in mental health, and the time-frequency feature can better reflect the change in mental health. The research at this stage is mainly limited to the behavior analysis of a certain type of specific group and the psychological analysis and interpretation of the influencing factors. However, the continuous update and iteration of data mining technology have provided a great boost to the research work in psychology. Driven by data mining technology, the construction of university informatization platforms has become more perfect. The use of mathematical models in machine learning makes it possible to predict the psychological state of students. Based on the main research contents of the above research scholars, the analysis of college education data has always been a research hotspot in the field of data mining. Through the data mining of students’ in-school education, the researchers use data mining technology and analysis theory to draw conclusions about students’ academic performance, social relations, and poor students and provide data support and data support for the construction of college informatization and smart campuses. The theoretical basis also provides valuable guidance and suggestions for college administrators [8–14].

3. Related Technical Methods

3.1. Data Processing Technology

3.1.1. Data Filtering. Although data mining is a method for massive data analysis, it does not mean that we need to use all the collected data because, when collecting data in the early stage, we did not specifically consider the use of these data in the future. However, the value density of these data is very low, which is not conducive to later data analysis. Therefore, when there are specific research goals, only the data that are useful for target analysis need to be selected. For the research purpose of this study, the original data are the records of students’ online behavior in school collected through the network system. The key information in the records are the student ID, the website visited, the type of website visited, the time of visiting the website, and so on. The rest of the information cannot be used, such as some parameters and codes designed by the log system to ensure security, which are useless for our analysis.

3.1.2. Data Cleaning. Data cleaning is mainly to solve the problem of poor data quality caused by some accidental factors for some useful data after data screening. The main means are as follows:

Missing value processing: when the data set with missing data accounts for a relatively low proportion of the whole
data set and the data volume of the sample is relatively large, in this case, it can be processed through the deletion method, that is, the data items with missing values can be directly discarded. Another common processing method is the filling method. This method is used to fill in the data according to the average value of the data near the dimension where the missing value is located when the data volume itself is not particularly large and there are many missing samples.

Outlier processing: it is also called outlier processing or error value processing. If the value of one dimension in a data item is far greater than or less than the value of other data items in the sample, the data item with the outlier is called an outlier. In the case of outlier data, it cannot be discarded directly. It is necessary to analyze it to judge whether it is reasonable and then decide on the processing strategy. For example, in real life, the age range of people is greater than 0 and less than 150. When the value in the data item exceeds this range, it can be regarded as an abnormal value. This method is a simple analysis method.

When the data is considered to have the same meaning as the addition and subtraction of several dimensions, they can represent the same meaning. Eliminating duplicate data dimensions plays a certain role in data downsizing and model burden reduction. It ensures the uniqueness and representativeness of data dimensions.

Noise data processing: the random error or variance of the measured data caused by some reasons is the so-called noise, which is the interference to the data. The commonly used methods are the box division method and regression method. The box division method forms a small group of nearby ordered values, namely “box,” and then smooths these ordered data values with the mean, median, or boundary of the data in the box to make these data locally smooth. The regression rule is to use a regression function to fit these noisy data and play the role of smoothing data denoising [15].

3.1.3. Data Conversion. Data conversion processing can also be called data mapping processing, which generally has three cases. One is the encoding conversion of text data. Since the computer cannot directly process text data, such as calculating the distance between two data, it is necessary to numerically encode the text. For example, in the gender type, the male is coded as 1, and the female is coded as 0; common encoding methods include one-hot encoding and so on. The second is format conversion. For example, date data need to be converted into a unified format type to facilitate subsequent analysis and processing. The third is the mathematical processing of numerical data. For example, when it is found that the numerical value of a certain dimension encountered changes in the form of an exponential, the exponentially changing data can be quickly converted into decimal numerical data, which is convenient for observation and analysis through the following formula [16]:

\[ y = \log_{\text{base}}(x + k). \] (1)

In the same way, when the data changes in the form of a power function, it can be processed by the method of opening the \( n \)th power, and it can be converted into small numerical data that is easier to observe by formula (2), where \( y \) is the converted value and \( x \) is the value obtained before conversion.

\[ y = \sqrt[n]{x}. \] (2)

3.1.4. Data Integration. The student’s on-campus network behavior data are recorded in the log system and saved in log format, while some basic information data of the student are saved in the student management system. During data analysis and model training, it is cumbersome and error-prone to operate the data in each system, so a data integration method is needed to extract these data and save them in the same environment for processing.

3.2. Data Analysis Technology. The biggest difference between logistic regression analysis and linear regression analysis is that the data types of the variables \( Y \) analyzed are different. Logistic regression analysis can analyze discrete categorical data, while linear regression analysis can only analyze continuous data types [17].

3.2.1. Logistic Regression Distribution

Definition 1 (Logistics). Let \( x \) be a continuous random variable, when \( x \) has the following distribution function and density function; then \( x \) is said to obey the logistic regression distribution.

\[ F(x) = P(X \leq x) = \frac{1}{1 + e^{-(x - \mu)/\sigma}} \] (3)

\[ f(x) = F'(X \leq x) = \frac{e^{-(x - \mu)/\sigma}}{[1 + e^{-(x - \mu)/\sigma}]^2}. \]

The distribution function and density function plots are shown in Figure 1.

3.2.2. Binary Logistic Regression Analysis. Binomial logistic regression is suitable for dealing with the situation when the value of the dependent variable \( y \) has only two categories, which is essentially a prediction model of classification probability. The value of the variable \( x \) is a real number. When the value of the dependent variable \( y \) is 1, the probability model is shown in equations 4 and 5, and when it is 0, it is shown in the following equations:

\[ P(Y = 1|x) = \frac{\exp(\omega \cdot x + b)}{1 + \exp(\omega \cdot x + b)} \] (4)

\[ P(Y = 0|x) = \frac{1}{1 + \exp(\omega \cdot x + b)} \] (5)

where \( x \in \mathbb{R}^n \) is the input value of the model, \( y \in \{0, 1\} \) is the output value of the model, \( \omega \in \mathbb{R}^n \) is the parameter, \( \omega \) is called the weight vector, \( b \) is called the bias, and \( \cdot \) is the inner
product of $\omega$ and $x$. Assuming a given input vector $x = (x_1, x_2, x_3, \cdots, x_n)$, the logistic regression analysis model can separately calculate $p(y = 0|x_i)$ and $p(y = 1|x_i)$ for each dimension $x$ of the input vector value. The concept of event probability and logarithm is introduced here. Let $P$ be the probability of occurrence of an event, and $1 - P$ corresponds to the probability that it does not occur. The probability of an event can be expressed by the ratio between the two, that is, $P/(1 - P)$; on this basis, the logarithmic probability of the event can be obtained by taking the logarithm. The formula is shown in (7). The odds can be given by the following equation [18]:

$$
\text{logit}(p) = \log \frac{p}{1 - p},
$$

and

$$
\text{logit}(p) = \log \frac{p(Y = 1|x)}{1 - p(Y = 1|x)} = \omega \cdot x + b. \tag{7}
$$

It can be seen from formula (7) that when the value of the dependent variable $y$ is 1, the calculation formula of the logarithmic probability of logistic regression analysis is actually a linear function. The logistic regression analysis is essentially fitting this linear function so that this linear function can distinguish the two categories of the original data as much as possible. The larger the value of the linear function, the greater the log probability of the logistic regression, the closer the classification type is to class 1, and vice versa; it is closer to class 0.

3.3. Data Mining Technology

3.3.1. Classification Algorithm. The essence of a classification algorithm is to train a classifier on a labeled data set so that it can divide a new data set. The process of evaluating the results of the classification algorithm is inseparable from the existence of a confusion matrix, which is a computing tool often used in classification algorithms. Figure 2 shows the composition of the confusion matrix [19]. Then the basic elements of classification problems are as follows: training data, that is, the sample data set used to learn the model; feature, that is, the attribute used to describe data and the basis of classification; model, that is, the external framework of classifier; algorithm, that is, the method of constructing classification rules; and evaluation, that is, the final evaluation of the effect of the model. The purpose of the classification algorithm is to mine the hidden rules in label data, so as to divide the data set in feature dimension space.

There are four parameters in the confusion matrix, which are as follows:

- **TP** (true positive): true examples, which refer to the positive tuples correctly classified by the classifier
- **TN** (true negative): true negative, which refers to the negative tuples correctly classified by the classifier
- **FP** (false positive): false positive, which refers to negative tuples that are misclassified as positive tuples by the classifier
- **FN** (false negative): false negative, which refers to the evaluation index of the positive tuple classification model that is misclassified as a negative tuple by the classifier can be calculated by the four parameters of the above confusion matrix:

Accuracy rate: also known as overall recognition rate, it generally measures how well the classification model can correctly identify various data sets:

$$
\text{accuracy} = \frac{TP + TN}{P + N} \tag{8}
$$

Precision: it reflects the proportion of correct classification in the classification results of the classification model for each category, that is, the accuracy of the model when judging each category.

$$
\text{precision} = \frac{TP}{TP + FP} \tag{9}
$$
Recall rate: also known as sensitivity, it can reflect the sensitivity of the classification model to each category of data set, that is, the proportion of a certain type of data that can be correctly identified by the model.

\[
\text{recall} = \frac{TP}{TP + FN} = \text{sensitivity.} \tag{10}
\]

\[F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \tag{11}\]

### 3.3.2. Integration Algorithm

There is no perfect algorithm model in the field of data mining. The idea of an integrated algorithm is to combine different types of algorithm models through a certain strategy, so as to improve the overall model’s ability to classify data sets. There are two main strategies for ensemble learning algorithms: bagging and boosting. The main idea of the bagging strategy is to combine the results of each base classifier and then determine the final classification result of the overall model by voting, which can effectively increase the stability of the classification. When training the base classifier, a part of the data of the sample is extracted by the method of bootstrap to construct the training data set of the base classifier. This kind of “incomplete learning” is to reduce the difference between each base classifier.

### 4. Feature Construction of a Mental Health State Perception Model Based on Deep Learning

#### 4.1. Data Preprocessing

The network log data mainly comes from a dedicated network log collection server. Through the user’s application to access the network, the link data accessed by the user is collected, so as to obtain the user’s network log information. The main content of the log information is: “A record of a user accessing a certain network type at a certain point in time.” A sample of log information is shown in Table 1 [20].

| Property name   | Numerical value |
|-----------------|-----------------|
| Student ID      | 2021*********** |
| Gender          | Male            |
| Age             | Twenty-one      |
| Point in time   | 2021-11-15, 00:00:00 |
| Website name    | Taobao          |
| Site type       | Shopping        |

Compared with the website type, the attribute of the website name is too careful and narrow, and it obviously belongs to the category of shopping. There is no need to distinguish the two. Therefore, when extracting feature dimensions from log information, the website name is not a required item. For the point-in-time information, in order to facilitate the processing of data at a point in time, it is divided into two parts: the part of year-month-day is used as “date,” and the part of hour:minute:second is used as “time.” The finally extracted feature dimensions are student ID, gender, age, date, time, and website type; the format is shown in Table 2 [21].

#### 4.2. Static Variable Analysis Based on Binary Logistic Regression

As shown in Table 2, there are two static variables in this research, gender variable and age variable. Static variables are property variables that are basically unchanged or unchanged for a long time. Firstly, the univariate binary logistic regression analysis was performed on the gender variable and the age variable. The dependent variable was divided into two groups: the extroverted “1” and the introverted “0” were the one group, and the depression “1” and the asymptomatic “0” were the one group.

Next, a multivariate binary logistic regression analysis was performed with the combination of gender and age as the dependent variable, and the results are shown in Table 3. It can be seen from the results in Table 3 that the results of logistic regression are the same as those of univariate analysis. Combining the results of the two analyses, it can be concluded that the gender and age factors are not statistically significant for the psychological state indicators and can be ignored.

#### 4.3. Feature Construction Based on Information Entropy

This study proposes two concepts for the design and construction of features. These two concepts are regularity...
of surfing behavior and degree of dependence of surfing behavior. The regularity of surfing behavior is a measure of the regularity of students’ surfing behavior of visiting different types of web pages within a period of time. The design of this regularity is based on the concept of Shannon’s information entropy. According to the information entropy, the order and purity of a dataset can be measured. Therefore, combined with the theory of information entropy, a method for calculating the regularity of students’ online behavior is designed. Method. For example, for online shopping behavior, let Shopping Regularity be SR, set shopping times to different intervals, and the shopping frequency interval is \([0, 5], [6, 11], [12, 25], [26, \ldots]\), according to the number of times of shopping per day to determine which interval it belongs to, and finally we will get the frequency distribution corresponding to these n intervals as \(C = \{C_1, C_2, \ldots, C_n\}\), and the probability corresponding to each interval is \(p_i\), and the calculation formula is as follows:

\[
p_i = \frac{C_i}{S},
\]

\[
i = 1, 2, \ldots, n,
\]

\[
S = \sum_{i=1}^{n} C_i.
\]

Then the calculation formula of shopping regularity SR is

\[
SR = \sum_{i=1}^{n} p_i \log p_i.
\]

The method of interval probability is also used to calculate the degree of dependence on online behavior. As shown above, the number of visits of a certain type of Internet access in the day is divided into intervals, and then the frequency of occurrence of this type in which interval is calculated in the statistical period. If the interval is the highest, then it is determined that the interval is dependent on this type of network behavior. If the interval is a low-order partition, the degree of dependence is light, and if it is a high-order partition, the degree of dependence is higher. This research transforms mild dependence into number “1,” moderate dependence into number “2,” and high dependence as “3.” When defining the interval division of the degree of dependence, two cases are considered. When accessing data of social platform type, such as when processing microblog data, it is divided into two types: browsing microblogs and publishing microblogs. The interval can be divided into \([0, 30]\) times as a low degree of dependence interval, \([31, 60]\) times as a medium degree of dependence interval, and more than 60 times as a high degree of dependence interval. The dependence degree interval for publishing microblogs can be divided into \([0, 10]\) times as a low dependence degree interval, \([10, 19]\) times as a medium dependence degree interval, and more than 20 times as a high dependence degree interval. The feature dimensions of the final constructed sample data set are shown in Table 4.

4.4. Feature Selection Based on Genetic Algorithm. After removing the uncorrelated static variables, the regularity and degree of dependence of various network types and the existence of the student ID are left in the sample data set. For each pair of labels, not all types of network behavior data are helpful for model training, and redundant data participating in the training will reduce the accuracy of the model. Therefore, this study uses the adaptive iterative ability of the genetic algorithm to perform feature selection and uses it to perform feature selection according to different label states. The iterative graph of the feature dimension obtained by the genetic algorithm is as follows: the horizontal axis is the number of feature combinations, and the vertical axis is the fitness function score, as shown in Figure 3.

When based on the internal and external trend tags, the extracted feature dimensions are 8, and the effect is the best.
These feature dimensions are WeChat regularity, WeChat dependence, WeChat posting regularity, Weibo regularity, video Watching regularity, video viewing regulation dependence degree, reading regularity, and reading dependence degree.

Based on the label of depression or not, the effect is best when the extracted feature dimensions are 5. These feature dimensions are shopping dependence, listening to music, WeChat dependence, map website regularity, and game dependence. Based on the anxiety label, the extracted feature dimension is 5, and the effect is the best. These feature dimensions are short video website dependence degree, information website dependence degree, music listening regularity, game regularity, and shopping regularity.

5. Experiment and Result Analysis

5.1. Experimental Environment. The experimental environment and related parameters are shown in Table 5.

5.2. Data Preparation. The data used in this experiment comes from two parts, which are divided into two parts: label data and feature dimension data. The processing of these two parts of data is as follows: label data processing: when labeling the psychological state information of internal and external tendency, the data with external tendency score are classified as label “1,” and the data with internal tendency score are classified as label “0.” For the pair of labels with or without depressive symptoms, the score 4 is taken as the threshold. The data less than the threshold are classified as the label with no depressive symptoms “0,” and the data

Table 4: Sample data set example table.

| Student ID | Gender | Age  | Shopping regularity | Shopping dependence | Browsing Weibo regularity |
|------------|--------|------|---------------------|----------------------|--------------------------|
| 1          | 1      | Twenty-one | 2.35                | 2                    | 3.25                     |

Table 5: Experimental environment and configuration.

| Lab environment | Environment configuration |
|-----------------|---------------------------|
| Operating system| Windows 10                |
| CPU             | Intel(R) Core(TM) i5-6500 3.2 GHz |
| RAM             | 32G                        |
| Programming language | Python                |
| Programming tools | Jupyter notebook         |

Figure 3: Iterative diagram of genetic algorithm.
greater than the threshold are classified as the label with depressive symptoms “1.” Similarly, for the treatment of the pair of labels with or without anxiety symptoms, the score 4 is also used as the threshold. The data less than the threshold are classified as the label without anxiety symptoms “0,” and the data greater than the threshold are classified as the label with anxiety symptoms “1.” Examples of specific sample data sets are shown in Table 6.

5.3. Comparison and Analysis of Model Experiment Results. The optimal model of the three mental state models is the random forest model, which does have a strong role in the field of classification. The parameter adjustment and optimization of the random forest model is mainly carried out through the grid search method. The parameter adjustment of the random forest mainly involves the following parameters: (1) n_estimators: the maximum number of iterations or the number of weak learners. If n_estimators is too small, it is easy to underfit, and if n_estimators is too large, overfitting will occur, so a suitable value for n_estimators is very important. (2) min_samples_split: the minimum number of samples required for internal node subdivision 3, min_samples_leaf: the minimum number of samples for leaf nodes 4, and max_depth: the maximum depth of the decision tree. Taking the GA-RF model of the two-category depression as an example, the process of parameter adjustment is as follows: in the first step, take n_estimators as the variable; the initial value is 10, and the interval is 10; and the output result is shown in Figure 4.

In the second step, (min_samples_split, max_depth) is used as the parameter combination; the starting value of min_samples_split is set to 100, and each change is 200; and the starting value of max_depth is set to 3, and each change is 2, and the output result is shown in Figure 5.

In the third step, take (min_samples_split, min_samples_leaf) as the parameter combination; the starting value of min_samples_split is set to 20, and each change is 10; and the starting value of min_samples_leaf is set to 60, and the output results are shown in Figure 6.

After comprehensively considering the above three steps, the optimal parameter combination of the depression model is obtained: n_estimators = 50, min_samples_split = 150, max_depth = 7, and min_samples_leaf = 25. Follow this step for the optimal parameters of the other two models and bring these parameters into the model. The f value of the model is shown in Table 7.

It can be seen from Table 7 that after the optimization of the model parameters, the f score of the model has been improved. For the internal and external tendency model, the f score increased from 0.765 to 0.79, an increase of 3.3%. For the depression binary model, the score of F increased from the original 0.81 to 0.83, an increase of 2.5%. For the anxiety Table 6: Sample data set example table.

| Student ID | Shopping regularity | Shopping dependence | Weibo regularity | Label 1 | Label 2 | Label 3 |
|-----------|---------------------|---------------------|------------------|---------|---------|---------|
| 001       | 3.25                | 2                   | 2.35             | 0       | 1       | 1       |
binary classification model, the score of $f$ increased from the original 0.75 to 0.81, an increase of 8%.

6. Conclusion

Nowadays, the topic of students’ mental health has attracted more and more attention from society. For example, the incidents of college students committing crimes and committing suicide caused by the abnormal psychology of college students have also often caused heated discussions in public opinion. At present, most of the students have insufficient understanding of mental illness or even have an attitude of neglecting and not paying attention, so these students with mental abnormalities cannot be found and treated effectively in time. These students and conducting interventions are a top priority in student management efforts. With the development of data mining technology, the construction of the data analysis model has been solved for us in terms of model analysis. This paper is based on the research on the psychological state prediction model that is used to capture students’ psychological state information based on the students’ online data collected on the university campus and the psychological assessment scale indicators. This research is based on the deep learning theory. The model is constructed, analyzed, and adjusted, aiming to grasp the psychological state information of students more comprehensively and accurately through the network behavior data of students in school. The results of the model experiments show that the $f$ scores of the three models have improved. The models used in this study are only classification models in the field of machine learning, and the more popular deep learning models are not used. The next step will use deep learning. The network structure model of the aspect is used for model experiments to compare the operation of the two models.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

The author would like to thank 2018 Shaanxi Provincial Social Science Fund Project, Research and Practice of Research Evaluation System of Higher Vocational Colleges Based on Need Level Theory (no. 2018Q09).

| Model                  | n_estimators | min_samples_split | max_depth | min_samples_leaf | F1 value |
|------------------------|--------------|--------------------|-----------|------------------|----------|
| Internal and external tendencies | 70           | 300                | 13        | 50               | 0.79     |
| Depression             | 50           | 150                | 7         | 25               | 0.83     |
| Anxiety                | 55           | 100                | 7         | 30               | 0.81     |

Table 7: Model output values under optimal parameters.

References

[1] J. Park, D. S. Lee, and H. Shablack, “When perceptions defy reality: the relationships between depression and actual and perceived Facebook social support,” *Journal of Affective Disorders*, vol. 200, pp. 37–44, 2016.

[2] L. A. Jelenchick, J. C. Eickhoff, and M. A. Moreno, “Facebook depression?” social networking site use and depression in Older adolescents[1],” *Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, vol. 52, no. 1, pp. 128–130, 2013.

[3] K. C. Fernandez, C. A. Levinson, and T. L. Rodebaugh, “Profiling: predicting social anxiety from facebook profiles,” *Social Psychological & Personality Science*, vol. 3, no. 6, pp. 706–713, 2012.

[4] Y. Bengü and über Ahmet, “The Relationship between Internet Addiction, Social Anxiety, Impulsivity, Self-Esteem, and Depression in a Sample of Turkish Undergraduate-Medical-Students,” *Psychiatry-Research*, vol. 1, Article ID S0165178118306127, 2018.

[5] C.-C. Lin, “The relationships among gratitude, self-esteem, depression, and suicidal ideation among undergraduate students,” *Scandinavian Journal of Psychology*, vol. 56, no. 6, pp. 700–707, 2015.

[6] S. D. Amarasuriya, A. F. Jorm, and N. J. Reavley, “Predicting intentions to seek help for depression among undergraduates in Sri Lanka,” *BMC Psychiatry*, vol. 18, no. 1, p. 122, 2018.

[7] K. e. ZHANG, “Four Changes of Modern Universities from the Perspective of "4V" of Big Data,” *Canadian Social Science*, vol. 11, 2015.

[8] M. I. Jordan and D. E. Rumelhart, “Forward models: supervised learning with a distal teacher[1],” *Cognitive Science*, vol. 16, no. 3, 1992.

[9] T. Dietterich, “Approximate statistical tests for comparing supervised classification learning algorithms,” *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, 1998.

[10] S. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” *Informatica*, vol. 31, 2007.

[11] H. B. Barlow, “Unsupervised learning,” *Neural Computation*, vol. 1, no. 3, pp. 295–311, 1989.

[12] T. Hofmann, “Unsupervised learning by probabilistic latent semantic analysis,” *Machine Learning*, vol. 42, no. 1-2, pp. 177–196, 2001.

[13] X. Zhu and A. B. Goldberg, “Introduction to semi-supervised learning,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 3, no. 1, p. 130, 2009.

[14] M. F. A. Hady and F. Schwenker, “Semi-supervised learning,” *Intelligent Systems Reference Library*, vol. 49, no. 2, pp. 215–239, 2013.

[15] A. G. B. Rs Sutton, “Reinforcement Learning,” *A Bradford book*, vol. 15, no. 7, pp. 665–685, 1998.

[16] L. K. Pack, M. L. Littman, and A. W. Moore, “Reinforcement learning: an introduction,” *Ieee Transactions on Neural Networks*, vol. 16, no. 1, pp. 285–286, 2005.
[17] T. Joachims, “Making large-scale SVM learning practical,”
Advances in Kernel Methods Support Vector Learning, 1998.
[18] S. R. Safavian and D. Landgrebe, “A survey of decision tree
classifier methodology,” IEEE Transactions on Systems Man &
Cybernetics, vol. 21, no. 3, pp. 660–674, 1991.
[19] V. Vladimir and Y. Svetnik, “Random Forest A Classification
and Regression Tool for Compound Classification and QSAR
Modeling,” Journal of Chemical Information & Modeling,
vol. 43, 2003.
[20] G. Ratsch, “Soft margins for adaboost,” Machine Learning,
vol. 42, no. 3, pp. 287–320, 2001.
[21] M. Collins, R. E. Schapire, and Y. Singer, “Logistic regression,
AdaBoost and bregman distances,” Machine Learning, vol. 48,
no. 1/2/3, pp. 253–285, 2002.