



## Research Article

# Research on Abnormal Behavior Prediction by Integrating Multiple Indexes of Student Behavior and Text Information in Big Data Environment

Yubiao Wang <sup>1,2</sup>, Junhao Wen <sup>1</sup>, Wei Zhou,<sup>1</sup> Quanwang Wu,<sup>3</sup> Yingchun Wei,<sup>2</sup> Heng Li,<sup>1</sup> and Bamei Tao<sup>4</sup>

<sup>1</sup>School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China

<sup>2</sup>Huxi Campus Management Center, Chongqing University, Chongqing 401331, China

<sup>3</sup>School of Computer Science, Chongqing University, Chongqing 400044, China

<sup>4</sup>CSIC Haizhuang Windpower Company Ltd., Chongqing 401122, China

Correspondence should be addressed to Junhao Wen; [wjhcqu@cqu.edu.cn](mailto:wjhcqu@cqu.edu.cn)

Received 26 August 2022; Revised 30 October 2022; Accepted 31 October 2022; Published 30 November 2022

Academic Editor: Shaohua Wan

Copyright © 2022 Yubiao Wang et al. 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.

With the wide application of information technologies such as big data, the Internet of Things, and cloud computing, college students have accumulated a large amount of personal information and daily behavior data in their daily studies and life. How to dynamically integrate multidimensional information of students to build accurate student portraits, using multi-indicator data of student behavior and comment texts, and finding out students with abnormal behavior from among many students has become an important problem to be solved. This paper proposes an abnormal behavior prediction method integrating multiple indicators of student behavior and text information (ABPM-IMISBTI) for the problem of abnormal behavior prediction of college students in the big data environment. First, given the problems of multidimensionality, timeliness, and dynamics of student behavior information fusion in the construction of student behavior portraits, by integrating students' objective tags and subjective tags, an optimized K-means algorithm based on a cloud platform environment is proposed. Second, aiming at the problem of insufficient text information analysis in the analysis of students' abnormal behavior, the ABPM-IMISBTI method is proposed to solve the prediction of students' abnormal behavior through long and short-term memory networks (LSTM) combined with student behavior multi-index data and text information. Finally, this paper takes student achievement prediction as an example for verification. The experimental results show that, compared with other prediction methods, the ABPM-IMISBTI method proposed in this paper can improve the accuracy of student behavior prediction, and then quickly determine the abnormal behavior of students, to improve the level of education management in universities and promote the development of safe campuses, smart campuses, and smart education.

## 1. Introduction

In recent years, emerging information technologies represented by big data, cloud computing, and the Internet of Things have further deepened education reform. The information revolution brought by big data is changing people's daily lives, ways of thinking, and work [1, 2]. Since the "Outline of Action for Promoting the Development of Big Data"

issued by the State Council of my country in 2015, more than 10 documents have been issued by the State Council and the Ministry of Education, making important discussions on the use of big data to promote educational reform, involving precise governance in colleges and universities and reform of teaching models. The development and utilization of educational big data have become an important part of college informatization. It is an important research direction

of big data technology in education informatization to study the behavioral laws of students' academic, psychological, and consumption in the big data environment.

At present, many information systems such as student return, behavior analysis, book borrowing, and class rate statistics have been built on university campuses, collecting a large number of students' daily studies, life, and other behavioral data, and a relatively complete campus big data environment has been formed [3]. However, with the wide application of new media, a large amount of network behavior data has been generated by college students surfing the Internet, and the formed network behavior data has multidimensional characteristics. Students are faced with a more complex and diverse social network environment, which makes student behavior more diverse. It increases the difficulty of student behavior analysis, prediction, early warning, and guidance. In colleges and universities, students with academic difficulties, psychological barriers, and life difficulties, especially those who may experience extreme events, are the groups that need the attention of college administrators. How to integrate multidimensional student behavior information, tap potential student behavior patterns, realize abnormal student behavior analysis and early warning, provide college administrators with a comprehensive understanding of students' learning and living conditions, and provide accurate services and management.

Domestic and foreign research has been carried out on student multidimensional information fusion and student abnormal behavior prediction. Luo et al. [4] studied the cross-curricular teaching method of network security and the course of big data analysis and used big data analysis technology for network log analysis. Jing et al. [5] proposed a mental health intelligence assessment method based on multisource information fusion, combined with the data description of UPI and SCL-90 and the corresponding task demand analysis, designed and realized the multisource information fusion visualization based on the Circos graph. Cao et al. [6] quantitatively analyzed the relationship pattern between student behavior and academic performance. By collecting the behavior records of undergraduates, they proposed two high-level behavioral characteristics, namely, orderliness and diligence. The experimental results showed that orderliness is a predictor of academic performance, then an important feature of performance. Yue et al. [7] aimed at the lack of multidimensional computing in the current research, analyzed learner engagement in a digital learning environment, proposed an integrated framework, and identified learning engagement from three aspects, including affective, behavioral, and cognitive states, to accurately predict the effect of student learning. Hongwei et al. [8] proposed a multisegment semantic spatiotemporal graph convolutional network (MFSTGCN) model based on the spatiotemporal graph network structure, considering the campus activity time series, hierarchical correlation, and spatial semantic feature correlation. The richness, data granularity, and data preprocessing are not considered very well. Nam and Samson [9] integrated student behavior and incoming profiles to achieve student learning prediction in higher education STEM courses. However, the model does not fully consider the profile factors

passed on by extra informative features, such as student study notes. Yao et al. [10] constructed a multitask prediction method based on a learning ranking algorithm for academic performance prediction, which considered factors such as school-term correlation and interprofessional correlation, and integrated students' similarity to predict students' performance. However, this method is not perfect for integrating the behavior information of students into the network, analyzing and mining the characteristics of each dimension of students' behavior, and then realizing real-time prediction of abnormal behavior of students.

At present, multidimensional information fusion methods focus on extracting single-dimensional features of text, ignoring the multidimensional features of semantic representation of students' behavior and do not fully consider factors such as timeliness and weight of student behavior information fusion [11–13]. Therefore, how to dynamically integrate the multidimensional information of students, consider the timeliness and dynamics of the characteristics of students and the similarity of students' behaviors, construct accurate student portraits and ideal student classification groups, and then improve the accuracy of students' abnormal behavior prediction is an urgent need in order to solve the problem. At present, the main problems faced by students' abnormal behavior prediction are as follows:

- (1) Student behavior portraits with deep integration of multidimensional information. Student behavioral portraits are based on students. By mining various campus behavioral data such as students' learning and life behaviors, students are given behavioral portrait labels [14]. With the continuous development of informatization in colleges and universities, many information systems related to student behavior have been formed. However, student behavior information is often scattered in multiple data sources and different dimensions, and a single data source and dimension only contain part of the student's behavior fragments. It is difficult to accurately mine the complete behavior patterns of students, resulting in low accuracy of students' abnormal behavior prediction and early warning [15, 16]. Meanwhile, the deep fusion of multidimensional information also has problems such as data dynamics, diversity, and complexity of feature extraction. The purpose of deep fusion of multidimensional information is to organically combine different types of information after feature extraction through several fusion methods, make full use of student feature attributes, fuse multimodal information such as images and text, and comprehensively mine student behavior rules. Therefore, according to the multidimensional information on student behavior in the campus big data environment, using the similarity data of students to conduct in-depth fusion is a challenging problem for student behavior portrait and abnormal student behavior prediction
- (2) Problems such as accuracy and backwardness in predicting abnormal behavior of students. In the era of

big data, the amount of data has increased dramatically, the types of data have varied, and the time-sensitive features contained in the data have received increasing attention [17]. The traditional student data analysis and early warning focus more on the analysis and mining of historical data in the data processing process, lack of timeliness, and the behavioral early warning mechanism still relies on the setting of thresholds, which is not in line with the development direction of educational big data [18]. Text information is an important dimension affecting student behavior analysis, and it is difficult to directly analyze it with traditional natural language processing techniques. In addition, multiple dimensions such as student class rate, book borrowing, one-card consumption, and students returning to bed are all important factors that constitute abnormal student behavior scenarios. Traditional student abnormal behavior prediction has problems such as low timeliness and accuracy. Therefore, based on the analysis of student behavior portraits and the deep integration of multidimensional information, how to realize the accurate prediction of abnormal student behavior by considering the similarity of students' behavior and text information is another problem that needs to be considered

To solve the two problems, we propose an ABPM-IMISBTI method by integrating the multidimensional information of students and constructing accurate student behavior portraits. The advantages of the proposed method are as follows: it can make full use of multidimensional and multimodal information to construct an accurate student behavior profile, which provides a basis for the prediction of abnormal student behavior; the student text behavior information is integrated into the ABPM-IMISBTI method, thereby improving the abnormal behavior of students. The accuracy of behavioral predictions.

- (1) We propose a method for constructing student behavior portraits based on the deep fusion of multidimensional information. Based on data preprocessing, all kinds of behavior data of students are clustered according to different levels and dimensions to form a portrait feature library. Cluster analysis is performed under the Spark platform, and an accurate student behavior portrait is finally constructed
- (2) We propose an ABPM-IMISBTI method, which uses a neural factorization machine and long-short-term memory neural network to solve the problem of abnormal student behavior prediction by combining multi-index data and text information of student behavior, and further improves the prediction performance, to detect abnormal behavior in time Behavioral students, to intervene and supervise students on time
- (3) We designed the analysis of behavior prediction results under the Spark platform to further verify the accuracy of the ABPM-IMISBTI method and

analyzed the relationship between students' behavior and the relationship between students' physical exercise and students' GPA scores

After this Introduction, Section 2 introduces some related works. Section 3 describes the abnormal behavior prediction method. We propose the ABPM-IMISBTI method, which combines multiple indicators of student behavior and text information. Section 4 describes experimental results and analysis; we use real data to conduct a predictive analysis of student-behavior relationships and abnormal behaviors. Conclusions and possible future works are given in Section 5.

## 2. Related Work

The work related to this paper mainly includes two aspects: (1) Student behavior portrait research; (2) student abnormal behavior research.

*2.1. Research on Student Behavior Portrait Based on Deep Fusion of Multidimensional Information.* Multidimensional data fusion is the premise of applying big data technology in specific fields, and early behavioral data fusion methods mainly focus on semantics [19] and similarity [20]. Inspired by the multisource data fusion theory, many researchers have integrated various data sources on campus, including data on motion trajectories, student learning, student consumption, and classroom attendance, and used support vector machines (SVM) and machine learning (ML) classifiers, etc. to predict student course grade [21–23]. Student multidimensional information data fusion can fuse the data generated by different data sources to mine the student behavior rules more comprehensively [24]. Due to the dynamic, complex, and multidimensional nature of student behavior data, the identification of abnormal student behavior groups cannot simply use a single-dimensional perspective. There are multiple dimensions such as the situation of returning to bed. Among them, students returning to bed can have multiple subfeatures, such as the frequency of students returning to bed early, the frequency of returning late, and the dormitory not returning at night. Only by integrating this multidimensional information can students be more accurate to determine students with abnormal behavior.

Domestic and foreign researchers have also carried out student behavior portraits based on multidimensional information fusion [25–27]. Li et al. [28] propose an adaptive Web API recommendation method that integrates multidimensional information, which can create a Web API for Mashup. Zhu et al. [29] proposed a hybrid recommendation model that combines network structural features with neural networks, as well as user interaction activities and tensor factorization. Chen et al. [30] designed a visual analysis method of college student's mental health based on multisource questionnaire data to effectively mine the connection between multisource questionnaires and reduce the uncertainty of mental health analysis, but this method involves some questionnaires. It is limited and does not consider combining students' online behaviors and integrating more

qualitative and quantitative data to analyze students' psychological problems. Liu et al. [31] obtained a low-dimensional vector representation of students through the LINE-based network embedding method to calculate the similarity between students, but this method would lose the semantic information contained in the original data and could not expand the fusion of multiple data sources. Bo et al. [32] proposed an eigenvalue extraction model based on the ID3 machine learning algorithm by establishing the characteristics and feature attribute groups of student behavior portraits, and constructed student behavior portraits based on big data. Zhang et al. [33] established a portrait of college students based on Hadoop big data processing technology, and used HDFS for data storage; we used canopy and K-means-based clustering algorithms to perform multidimensional analysis of student data. Li et al. [34] proposed a user model with five dimensions, including students' basic information, learning ability, consumption level, daily habits and interests, and preferences, and extracted the characteristic attributes of students through methods based on data collection, processing, and mining. Ding et al. [35] proposed a hybrid neural network model to mine the data of college students and build student portraits, thereby helping students' personal development and improving the quality of school teaching.

At present, multidimensional information fusion methods focus on extracting single-dimensional features of texts, ignoring the multidimensional features of semantic representation of student behavior, and do not fully consider factors such as timeliness and weight of student behavior information fusion [36]. Zhang et al. [37] proposed a student portrait construction algorithm based on K-means optimization. The canopy algorithm was used for preliminary clustering, which eliminated the uncertainty of K value selection. The sample is used as the initial center of K-means, but the algorithm only conducts preliminary analysis on the two dimensions of student consumption and learning. The scale of student data in the experiment is relatively small, and the analysis dimension of student data is not enough. Mainstream student behavior profiling methods are generally based on technologies such as machine learning, support vector machines, and supervised learning [38, 39]. However, these methods do not fully consider the security and timeliness of students' personal information, and the integration of multimodal information such as images and texts. Therefore, how to dynamically integrate the multidimensional information of students, consider the timeliness and dynamics of the characteristics of students and the similarity of students' behaviors, construct accurate student portraits and ideal student classification groups, and then improve the accuracy of students' abnormal behavior prediction is an urgent need.

*2.2. Research on Abnormal Behaviors of Students.* Deng et al. [40] used the Hadoop-based K-means clustering algorithm to cluster user behaviors, and the association rules obtained by mining were used as the preferences for users to access the campus network, thereby improving the efficiency and accuracy of student behavior analysis and achieving accurate

prediction. You et al. [41] proposed a hybrid neural network method based on a high-order attention mechanism, using generative adversarial networks to simulate students' learning behavior, mine missing data, and quickly classify students' studies, but the model considering online learning detection is not very well. Zeng et al. [42] proposed an attentive prediction model for academic abnormalities. Nie et al. [43] aimed at the problem of precise poverty alleviation in colleges and universities, based on student behavior data, combined with the time series characteristics of college data, extracted the multidimensional characteristics of students' basic information and behavior data and proposed a CW-LSTM algorithm based on deep learning theory for prediction. Yu et al. [44] analyzed the decision tree of the single classification algorithm and the random forest (RF) of the ensemble learning algorithm and constructed an online student achievement prediction model by using the RF algorithm. Xu et al. [45] proposed a novel hybrid IDA-SVR-based model to predict student performance, it is an improved decision algorithm (IDA) to optimize support vector regression (SVR). Maksimova et al. [46] combined the decision tree and rule model to establish a classification rule set and constructed a learning behavior diagnosis model combining a decision tree and deep neural network, but this model is limited by factors such as lack of data and imbalance. When predicting academic performance, the learning behavior data in the study fails to cover a variety of other factors that affect academic performance in the university study scenario. Meanwhile, with the diversification of students' online behavior, text information and images have become important factors in student behavior analysis. Traditional prediction and early warning models of students' abnormal behavior do not make full use of this information, resulting in ambiguity, uncertainty, and large errors. Therefore, it is a challenging scientific problem to consider students' real-time behavior data and student text information, improve the prediction accuracy of students' abnormal behavior, and realize the transition from post-emergency to prewarning of abnormal behavior of students.

However, in the process of data processing, the traditional analysis and early warning of abnormal student behaviors focus more on the analysis and mining of historical data, which lacks dynamics and timeliness [47]. Cao et al. [48] established a scientific student portrait evaluation index system, collected various data on college students' academic performance, normalized the collected data, determined the weight of each evaluation index through the analytic hierarchy process (AHP), and then a fuzzy evaluation model based on big data is used to evaluate various dimensions of college student's academic performance. Zhao et al. [49] integrated various data of college students' movement trajectory, consumption, social behavior, etc., used machine learning (ML) classifier support vector machine (SVM) to predict English, and analyzed the correlation between students' performance and social relations, then it was used to predict the English grades of college students. Li et al. [50] established a combined data mining model by integrating a decision tree, neural network, and naive Bayesian algorithm, and established a Spark-based college student behavior analysis and

prediction platform. Filvà et al. [51] collected and analyzed the data generated by students from scratch through behavior in scratch-based programming activities, and conducted a predictive analysis of student interest to help teachers conduct assessments. Yang et al. [52] analyzing the characteristics of campus big data, using the traditional K-means clustering algorithm, and under the mainstream Hadoop open source platform, proposed a college student behavior early warning system based on the Internet of Things and big data environment. Xie et al. [53] proposed a deep learning algorithm to evaluate college students' classroom posture, using K-means clustering (KMC) to cluster different student groups and identify the characteristics of each group. Liu et al. [54] proposed a deep learning-based method for identifying the abnormal behavior of students in classroom videos, aiming at the complex and slow process of traditional human behavior recognition. Shi et al. [55] proposed a prediction algorithm based on the BP neural network, which predicts the situation of borrowing books through course performance, and builds an early warning model of student performance. However, the data sample of this paper is not large enough, and the borrowing information is insufficiently utilized, which is not conducive to predicting students' performance. Yang et al. [56] used students' homework data to predict students' course grades in Moodle through students' procrastination behavior, but the data used for student grade prediction was not considered comprehensively, and more student activity data could be used, such as student text data, student learning resources, and network access records, etc. Meanwhile, the existing methods still have some deficiencies in introducing deep learning, integrating student behavior text information, and intelligently identifying some unknown abnormal behaviors in the network [57–59]. Many scholars have analyzed the abnormal behavior prediction and early warning of students. The comparison table of abnormal behavior student methods is shown in Table 1.

### 3. Abnormal Behavior Prediction by Integrating Multiple Indexes of Students

*3.1. Data Preprocessing.* At present, many business systems for student image collection, such as face brushing in student apartments, intersection monitoring, and classroom attendance have been built on college campuses. When searching for student trajectories, student classroom behaviors, abnormal behavior trajectories in specific areas, blacklists, and special students, these image data need to be processed in real-time and effectively. In scenarios such as public opinion analysis, it is necessary to analyze the relevant student text information. Therefore, for the data of student images and text information, this project plans to use data preprocessing based on deep learning.

Data preprocessing is an important process of data mining and knowledge discovery, and it is also the basis for extracting student behavior characteristics. In this study, data preprocessing refers to a series of data processing processes, including data collection, data cleaning, behavioral feature extraction, and data transformation. Among them,

data collection provides basic data for students' behavioral portraits, and data cleaning is to process students' basic data. Missing values and outliers in the data and student behavior feature extraction is used to obtain the data required for behavior analysis from the basic data of students. Finally, data transformation mainly eliminates the magnitude difference between different behavior data of students to establish the student behavior. The portrait model provides a unified standard.

This paper adopts a parallel processing framework to improve the efficiency of data preprocessing, adopts a non-destructive cleaning algorithm based on distributed processing framework (Hadoop), and uses Map-Reduce to process unstructured and semistructured data. For example, for student library borrowing data, structured data can be abstracted into fixed information data, such as student borrowing information, browsing information, and library book information, which can be expressed in two-dimensional form using relational databases, and unstructured data can be abstracted browsing, commenting, and page retention time when borrowing from the student's library, these data are random and nonfixed. And semistructured data is in between, it is a data model suitable for database integration; that is, suitable for describing data contained in two or more databases.

This paper adopts VGG16 for image data preprocessing and identifying abnormal images of students. Data preprocessing for text information is the same as image processing, requiring text vectorization. This paper uses the Word2Vec method to process text vectorization. Using the cleaning algorithm of the previous data preprocessing, the filtered data is stored separately according to different reasons, and the data will not be lost directly. You can clean how many pieces of data are filtered out by each cleaning rule set by the storage path cleaning query. Meanwhile, there are a large number in the case of discarded data, abnormal data can be checked, the proportion of various abnormal data can be counted, and the distribution of abnormal data in the cycle can be analyzed. Meanwhile, the daily data volume is continuously detected. After analyzing the abnormal data situation, a reasonable abnormal data threshold is set. Once it is higher than the threshold, an email can be sent for processing by calling the early warning mechanism. The preprocessing of student behavior data is shown in Figure 1.

*3.2. Construction of Student Behavior Portrait Based on Deep Integration of Multidimensional Information.* At present, the further mining and analysis of educational big data have a profound impact on the current university management and training model and even the entire education system. Establishing student behavioral portraits has become one of the important ways to apply big data to guide students' behavioral norms. The student behavior portrait in the big data environment is used to classify various behavior data of students. Therefore, the student behavior portrait is a collection of data and semantic behavior characteristics. This paper starts from the original data and obtains the objective label through statistical analysis of the original data. The objective label can be obtained by data preprocessing

TABLE 1: Comparative table of abnormal behavior student methods.

S/N	Name	Algorithm	Proposed method	Author	Year
1	MTLTR-APP	Learning ranking algorithm	A multitask prediction method based on a learning ranking algorithm, used for learning performance prediction	Yao et al. [10]	2019
2	CW-LSTM	Deep learning algorithm	It combined with the time series characteristics of college data, extracted the multidimensional characteristics of students' basic information.	Nie et al. [43]	2020
3	Bayesian	A probability-based algorithm	A method suitable for small-scale data and can handle multiclassification tasks.	Wang et al. [57]	2022
4	DNN	Learning ranking algorithm	The output layer is through hidden layer transformation, and the classification result is obtained from the output layer	Xu et al. [58]	2022
5	RF	A typical supervised learning algorithm	A random forest is used for learning performance prediction	Yu et al. [44]	2019
6	IDA-SVR	A supervised learning algorithm	An improved decision algorithm (IDA) to optimize support vector regression (SVR)	Xu et al. [45]	2022
7	ATL-BP	The recurrent neural network algorithm	A video-based transfer learning method	Ruiz et al. [59]	2022
8	GRMF	Matrix factorization algorithm	A new MF method, called graph regularized robust matrix factorization (GRMF), is based on the recent robust MF version.	Zhang et al. [60]	2020
9	CNN	A deep learning algorithm	A deep learning algorithm to evaluate college students' classroom posture.	Xie et al. [53]	2021
10	LEWM-RFA	A supervised learning algorithm	The model divides students' learning behavior data into three dimensions: Knowledge, behavior, and attitude.	Cheng et al. [61]	2018
11	Rs-YOLOv3	Deep learning algorithm	A method of student abnormal behavior recognition in classroom video based on deep learning	Liu et al. [54]	2021
12	BP neural network	A BP neural network prediction algorithm	A prediction algorithm based on BP neural network	Shi et al. [55]	2022
13	HHA	Hybrid neural network algorithm.	An attentive prediction model for academic abnormalities.	Zeng et al. [42]	2020
14	SVM	A class of supervised classification algorithms.	A method for predicting English scores based on machine learning classifiers.	Zhao et al. [49]	2020

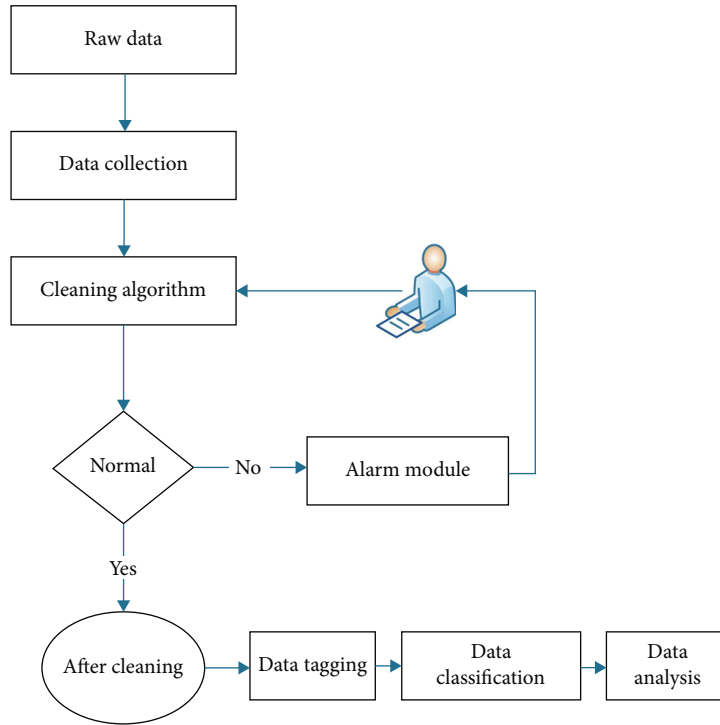


FIGURE 1: Student behavior data preprocessing.

technology, which is the real data. To better describe a student's comprehensive ability, in this paper, the researchers further model and analyze the objective labels to obtain subjective labels, such as analyzing the characteristics of students' online behavior, the characteristics of students' hybrid amount, the characteristics of students returning to bed, and the characteristics of students' course attendance from different dimensions. Data such as characteristics and student borrowing characteristics; that is, to construct student behavior characteristics portraits by describing students' study habits and living habits.

Student behavior portrait analysis refers to the process of dividing various types of behavior data of students after data preprocessing into different behavior categories according to different levels and dimensions. Taking the portrait characteristics of outstanding students and the portrait characteristics of ordinary students as an example; first, by analyzing the clustering categories of various types of behavior data of students, a data-based behavior label set is constructed to describe different types of students, and the behavior characteristics and characteristics of outstanding students are identified. The behavioral characteristics of ordinary students; secondly, we use the data-based label set analysis to obtain the situational behavioral characteristics and verify and correct various behavioral labels; finally, put the behavioral characteristics of outstanding students and ordinary students in the real student group, carry out accuracy verification, modify and improve the student behavior label according to the verification results of the real student group behavior feedback and integrate the verified student behavior characteristics into the behavior label set, which also pro-

vides a basis for the analysis of abnormal student behavior scenarios.

The traditional K-means algorithm adopts a random method for the selection of initialized clustering centers, and the quality of its clustering results is largely dependent on the selection of initialized center points, which leads to its instability. The time complexity of the cluster center algorithm is also relatively large, so it needs to be optimized and designed to further improve the work efficiency of the K-means algorithm. This paper adopts the clustering of various behaviors of students based on the optimized K-means algorithm in the cloud platform environment to construct a set of student behavior labels. The construction process of the student behavior portrait is shown in Figure 2.

The basic idea of optimizing the K-means algorithm in the cloud platform environment is to randomly select K data samples from the data set as the initial clustering center, and each iteration can correspond to a task in Map-Reduce, and each time the data samples are sent to the distance of the cluster center and the classification of the data samples are written into the Map function of Map-Reduce, and then the information of the cluster center is updated through the Reduce function. After cyclic iteration, the algorithm converges to obtain the most stable cluster center, the flow of the optimized K-means algorithm in the cloud platform environment is shown in Figure 3.

*3.3. Abnormal Behavior Prediction Integrating Multiple Indicators of Student Behavior and Text Information.* Traditional student behavior prediction methods still have some shortcomings in using multi-index data and comment texts

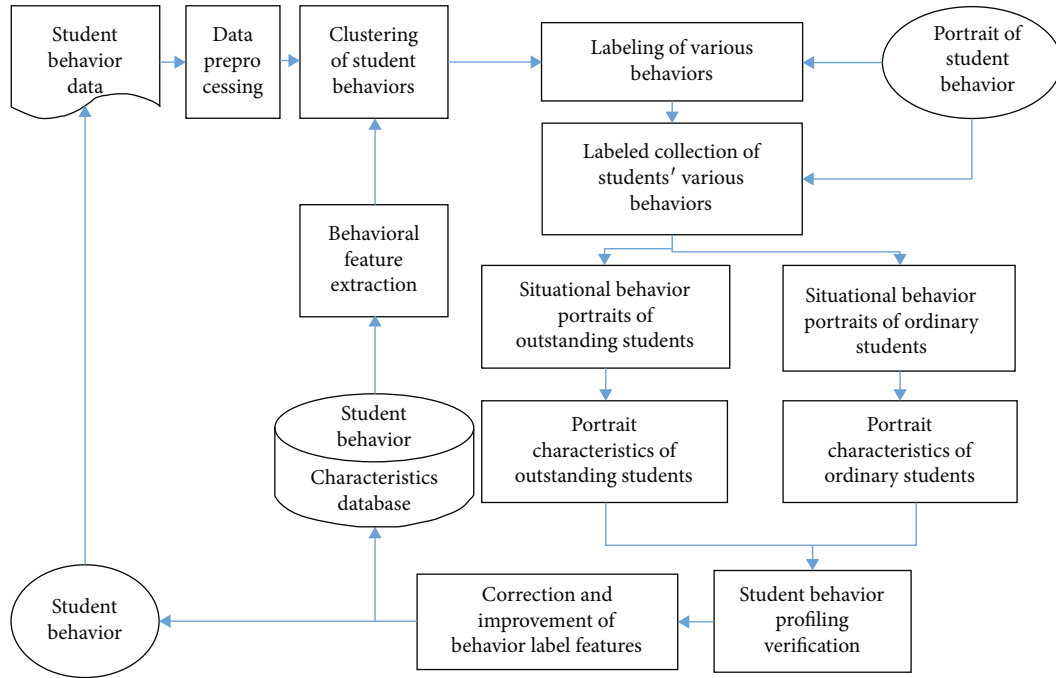


FIGURE 2: Flow chart of the student behavior portrait construction.

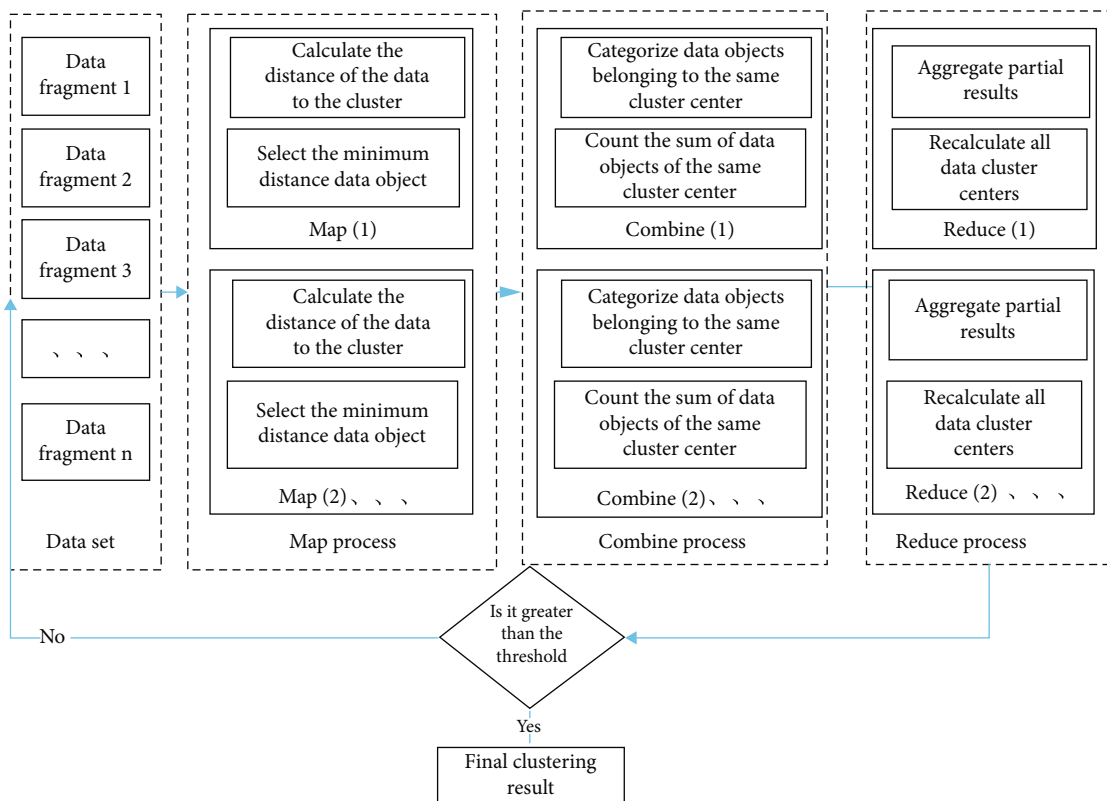


FIGURE 3: Flow chart based on the optimized K-means algorithm under cloud platform environment.

on student behavior, mining data, etc., especially when semantic analysis is required, such as student network behavior analysis, public opinion analysis, and psychological analysis. When students search and speak information such

as loans, suicides, and cults, students may have some comment information, and it is necessary to analyze the text information of such students. Due to the limited length of textual information or annotations on student behavior,



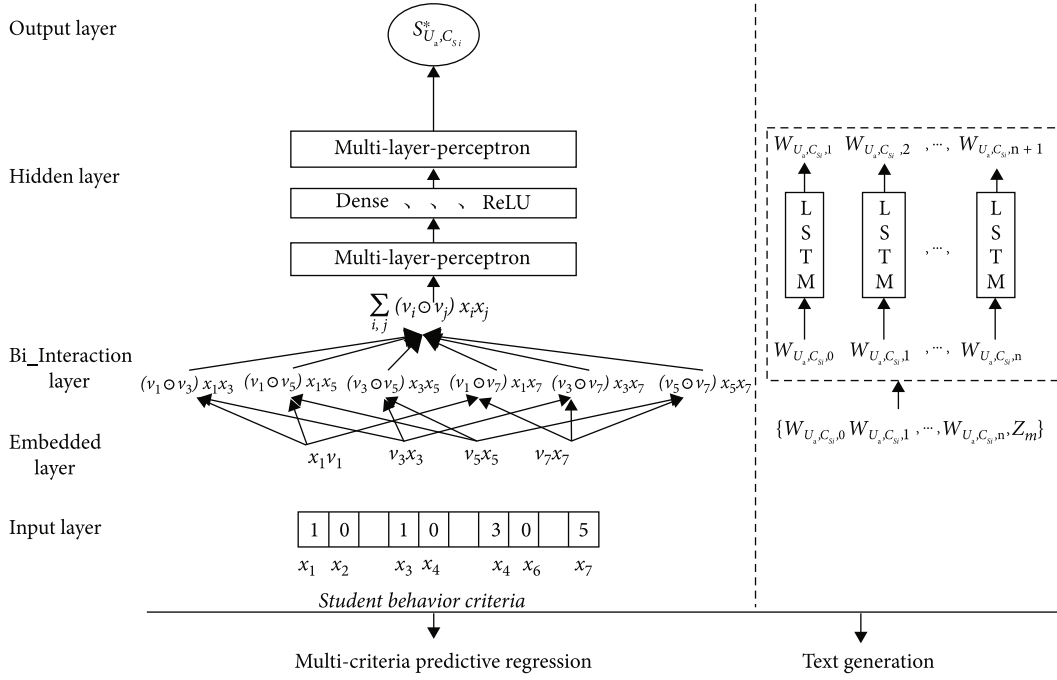


FIGURE 4: Abnormal behavior prediction combining multiple indicators of student behavior and text information.

and not all words and topics in the textual information are relevant to the analysis of student behavior, the topic factors that are usually used to obtain comments based on the distribution of latent Dirichlet topics cannot fully reflect the relevance of student behavior. The order of words is often ignored by the bag-of-words languages used. In response to these problems, this paper proposes an ABPM-IMISBTI method, which uses a neural factorization machine and LSTM to solve the problem of abnormal student behavior prediction by combining student behavior multi-index data and text information, and further improves the prediction performance.

Student behavior multi-index data and text information are different manifestations of student behavior. Therefore, this model can better reflect student behavior and improve early warning accuracy by using multiple index data and text information.  $D = \{U_a, C_{S_i}, S_{U_a, C_{S_i}}, \text{sim}_{U_a, U_j}, W_{U_a, C_{S_i}}\}$  denotes a dataset of students, certain behavioral scenarios of students, predicted values, multi-index data, and student comments,  $U_a$  denotes students,  $S_{U_a, C_{S_i}}$  denotes students' predicted values for  $C_{S_i}$ , such as student academic analysis,  $S_{U_a, C_{S_i}} = \{S_{U_a, C_{S_i}, 1}, S_{U_a, C_{S_i}, 2}, \dots, S_{U_a, C_{S_i}, m}\}$  represents a set of indicator data of student  $U_a$  to a student behavior scene  $C_{S_i}$ ,  $S_{U_a, C_{S_i}, m}$  denotes the behavior performance value on the student to student behavior scene indicator  $m$ ,  $\text{sim}_{U_a, U_j}$  denotes the student behavior similarity between student  $U_a$  and  $U_j$ ,  $W_{U_a, C_{S_i}}$  denotes student  $U_a$ 's comment on student behavior scenario  $C_{S_i}$ , and  $W_{U_a, C_{S_i}} = \{W_{U_a, C_{S_i}, 0}, W_{U_a, C_{S_i}, 1}, \dots, W_{U_a, C_{S_i}, n}\}$ ,  $W_{U_a, C_{S_i}, n}$  denotes the  $n^{\text{th}}$  word in the comment. Abnormal behavior prediction that integrates multiple indicators of student behavior and text information is to predict

a certain student behavior (such as student performance, etc.) that has not occurred based on the above-mentioned multiple indicator data, predicted value, and similarity of student behavior.

As shown in Figure 4, in the prediction stage, since the text information of students cannot be used directly; a neural decomposition machine and long short-term memory network are used to generate behavior prediction value and comment generation. Meanwhile, the model contains three components: neural multi-index regression prediction component, related behavior comment text generation component, and student behavior similarity decomposition component. The neural multi-index regression prediction component  $S_{U_a, C_{S_i}}^*$  represents the predicted value processed by the neural factorization machine. To improve the accuracy of student behavior prediction, first perform a one-hot key encoding for  $\langle \text{student, behavioral scene, index data} \rangle$  to construct the input the feature vector,  $\sum_{i,j} (v_i \odot v_j) x_i x_j$  denotes the Bi\_Interaction layer, and  $\odot$  represents the element-wise product of vectors  $v_i$  and  $v_j$ , and a neural decomposition regression model is then used to map the input feature vector to the student's predicted total score in a nonlinear transformation. In the text information generation component, this paper adopts a threshold neural network, that is, a decoder based on long and short-term memory networks, and converts the combined representation of students, behavioral scenes, and multiple indicator data into a series of words representing the comment text through LSTM. Where  $\{W_{U_a, C_{S_i}, 0}, W_{U_a, C_{S_i}, 1}, \dots, W_{U_a, C_{S_i}, n}, Z_m\}$ , where  $m$  represents the number of hidden layers,  $Z_m$  represents the high-dimensional interaction of the hidden layers, and connects the prediction component and the review generation component through the hidden layer of the shared neural factorization machine.

## 4. Experiment Study

**4.1. Data Set and Experimental Environment.** Various behavioral data of 14,800 sophomore and junior undergraduates in a university, including students returning to bed, on-campus behavioral trajectory, academic performance, campus network behavior, one-card consumption, library borrowing, and other data for behavioral analysis, were extracted about 75 dimensions to form different student behavior profiles by analyzing different scenarios, such as academic abnormalities, students returning to bed, canteen consumption abnormalities, campus network behavior abnormalities, and abnormal behaviors in specific areas. Focus on the behavior data of 612 students in the 2020 and 2021 classes of the School of Big Data and Software as the main research object.

The source of the data set of the experiment: 14,800 college students from March 2020 to March 2022, with a total of 7,202,811 records of student campus card consumption records. From March 2020 to March 2022, the library book loan record data totaled 1,468,128. From January 2020 to March 2022, there are 7,595,783 records of attendance records in educational affairs classrooms. The library access control record data from September 2019 to March 2022, a total of 1,970,764 records. From January 2019 to December 2021, through the analysis of more than 44.19 million face-swiping data of students on campus, 56,000 late return records, 147,000 nonreturn records, and 235,000 records of students returning to bed, the behavior analysis system was formed. No activity was recorded for 24 hours. From December 2019 to February 2022, the online data of students was about 43 T, the campus network behavior of 15,800 students was analyzed, and a total of 142.332 million Internet access data was obtained, including related college students' GPA, physical exercise record data, and other data.

**4.2. Evaluation Indicators.** The traditional K-means algorithm generally evaluates the effect of clustering by calculating the sum of the squares of the distances from the sample points within the class to the cluster center points. To verify the superiority of the clustering method based on density optimization in this paper, this paper adopts a more effective evaluation standard to evaluate the clustering results and comprehensively considers the intraclass similarity and interclass dissimilarity during clustering, as shown in Equation (1).  $D_{in}$  denotes the intraclass similarity, as shown in Equation (2), which is represented by the average distance between the samples in each type and the center point of the cluster to which it belongs.  $D_{out}$  denotes the interclass dissimilarity, as shown in Equation (3), which is used to measure the degree of separation between different clusters, which is represented by the average distance between cluster centers.

$$V(k) = \frac{D_{out} - D_{in}}{D_{out} + D_{in}}, \quad (1)$$

$$D_{in} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^m D(x_j, p_i), \quad (2)$$

$$D_{out} = \frac{1}{k} \sum_{i,j=1}^k d(p_i, p_j). \quad (3)$$

The clustering evaluation standard function is denoted as  $V(k)$ . It can be seen that its value range is between  $[-1, 1]$ . The closer  $V(k)$  is to 1, the higher the similarity of each sample within the class and the dissimilarity between classes. The stronger the clustering effect, the better the clustering effect. On the contrary, when  $V(k)$  is closer to -1, the clustering effect is worse, which  $D(x_j, p_i)$  represents the distance from  $x_j$  the center point  $p_i$  of the cluster to which it belongs.

To verify the accuracy of the model prediction in this paper, this paper measures the quality of the prediction results by calculating different evaluation indicators. Accuracy is an important indicator to measure, forecast, and early warning. Specific indicators include Relative Error (RE), Mean Relative Error (MRE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

$$RE = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right|. \quad (4)$$

As shown in Equation (4), RE denotes Relative Error, where  $y_i$  is the predicted value on the  $i$ -th dimension, and  $\hat{y}_i$  is the student's actual value on the  $i$ -th dimension index attribute.  $MRE = ((1/n) \sum_{j=1}^n RE_j) * 100\%$ , where  $n$  denotes the predicted number of students and RE represents the relative error of the behavior prediction result of the student  $j$ .

$$MAE = \frac{1}{m} \sum_{j=1}^m \frac{1}{n} \sum_{i=1}^n \frac{|y_{ij} - \hat{y}_{ij}|}{\hat{y}_{ij}} * 100\%, \quad (5)$$

$$RMSE = \frac{1}{m} \sum_{j=1}^m \frac{1}{n} \sqrt{\sum_{i=1}^n \left( \frac{y_{ij} - \hat{y}_{ij}}{\hat{y}_{ij}} \right)^2} * 100\%. \quad (6)$$

The calculation method MAE is shown in Equation (5), and the calculation method RMSE is shown in Equation (6), where  $m$  is the dimension of student behavior characteristics,  $n$  denotes the number of predicted students, and the student  $y_{ij}$  means that the student  $i$  is in  $j$ . The predicted value under the feature  $\hat{y}_{ij}$  represents the true value of the student  $i$  under the  $j$  feature, which further verifies the accuracy of predicting student behavior.

### 4.3. Experimental Results and Analysis

**4.3.1. Efficiency Analysis.** As shown in Figure 5, when the student density threshold is set high and the amount of student data is large, the students are divided into fewer categories, the difference between classes is small, and the aggregation effect within the class is very poor, so  $V$  takes a negative value, with the adjustment of the student density threshold, the clustering effect is better when the aggregated evaluation criteria parameters increase. When the student density is set to 300, the value of  $V$  is the largest, but when the student density drops to a certain level, although the

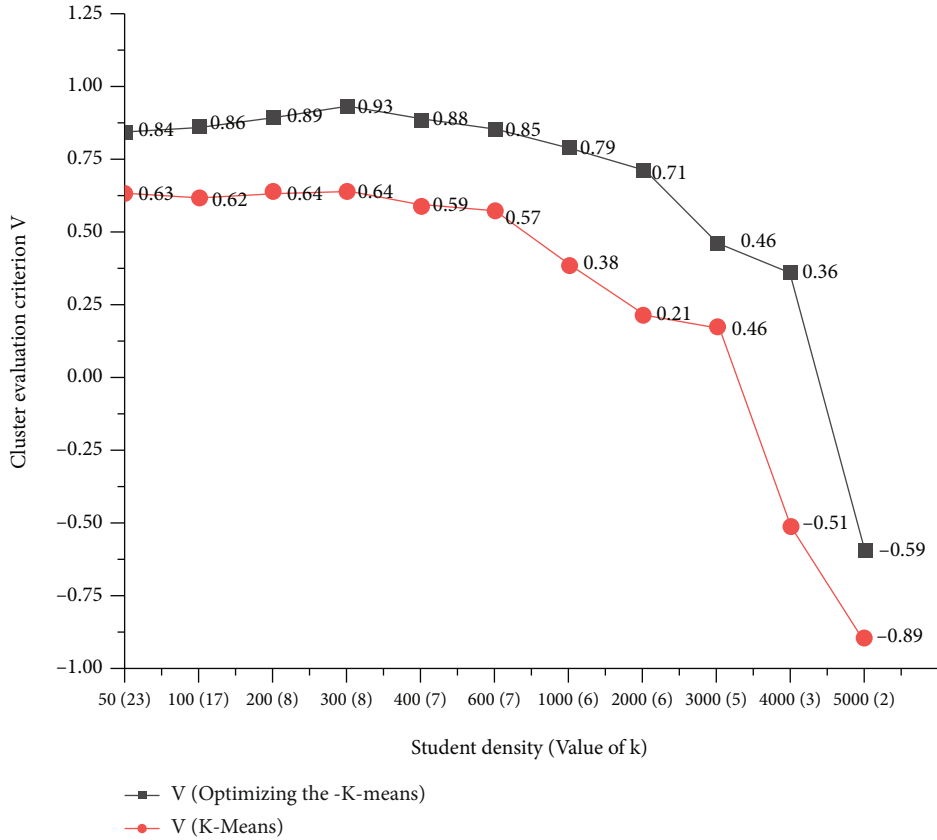


FIGURE 5: Cluster evaluation criteria results.

interclass dissimilarity decreases, the intraclass aggregation effect decreases significantly, resulting in a decrease in parameters, which is different from the traditional K-means compared with the algorithm, the clustering effect of the clustering method based on density optimization is significantly improved. This paper takes the analysis of the clustering of students' consumption patterns as an example.

As shown in Figure 5, the clustering effect of students' consumption patterns is analyzed. According to the clustering evaluation standard function  $V(k)$ , the optimized K-means algorithm is used for clustering on the Spark platform. The number of clusters is denoted as  $k$ . In this paper, when the student density is set to 300, the clustering evaluation criterion is optimal. To evaluate the clustering effect, this paper evaluates the clustering results through the evaluation standard  $V$  that reflects the similarity within the class and the dissimilarity between the classes and compares it with the traditional K-means clustering algorithm. It can be seen that this paper is based on density, because the  $V$  (Optimizing the K-means) value of the optimized K-means clustering algorithm is higher than the  $V$  (K-Means) value of the traditional K-means algorithm, so the optimized K-means clustering algorithm has achieved better results.

In this paper, aiming at the problem of student behavior segmentation, the K-means clustering algorithm based on density optimization is used for clustering analysis on the Spark cluster, and good experimental results are obtained. Meanwhile, the optimization algorithm in this paper ana-

lyzes the time efficiency of parallelization. This paper compares and analyzes the execution time of a serial on a single machine and parallel on a Spark cluster according to the number of different students. As shown in Figure 6, with the increase in the number of clusters, the running time of the K-means clustering algorithm based on density optimization on a single computer increases significantly. On a Spark cluster, the runtime increases slightly with the number of students. When the number of students is small, the running time of the parallel cluster on the Spark cluster is higher than that of the single machine because too much time is spent in the process of submitting tasks and resources, and the scheduling processing speed on the Spark cluster is not as fast as that of the single machine. However, when the amount of student data increases, the execution time required for the model does not increase very much, so the Spark cluster processing is more suitable for processing large amounts of data.

As shown in Figure 7, for the time efficiency analysis of parallelization on Spark, the parallelism of the early warning model in this paper is evaluated by analyzing the speedup ratio of different data volumes with different numbers of cluster worker nodes. This paper uses 10%, 50%, and 100% of the student data as the data set and gradually increases the number of worker nodes in the Spark cluster to train the early warning model in this paper. With the increase of worker nodes, the speedup ratio of the algorithm is improved. However, due to factors such as internode

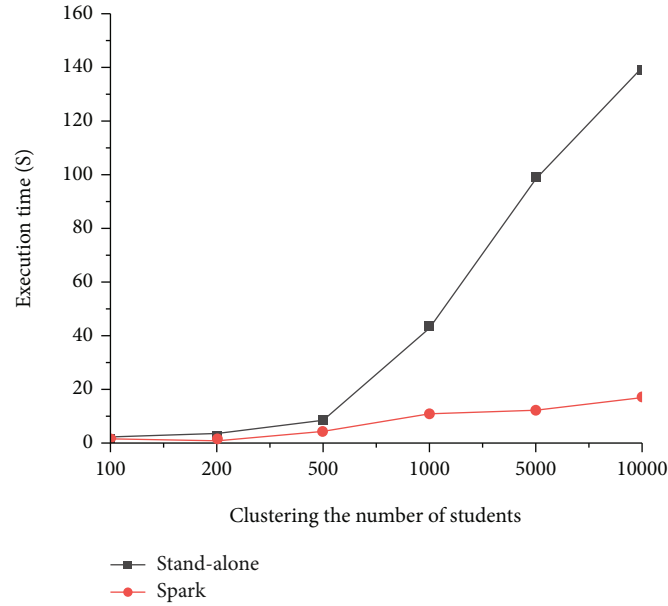


FIGURE 6: Operational efficiency analyses.

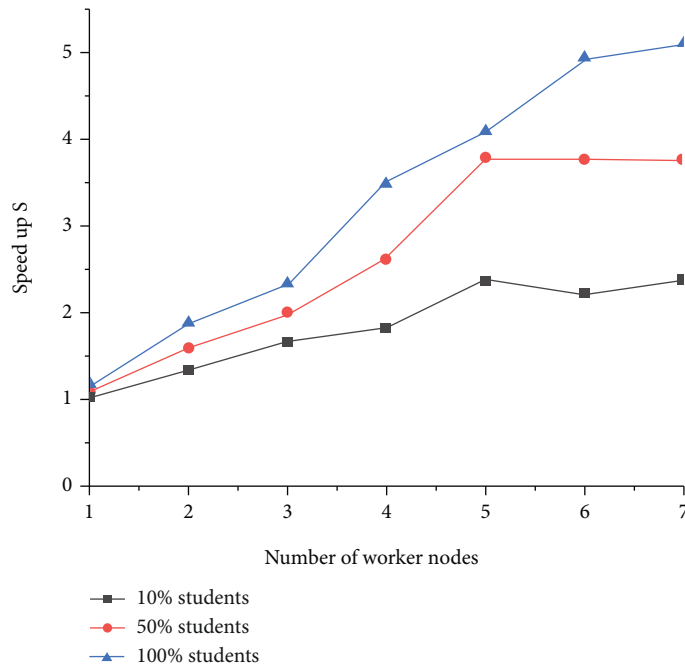


FIGURE 7: The comparison chart of the spark cluster speedup ratio.

communication and data dependence, the running time of the algorithm increases with the increase of student data. When the amount of data is small, that is, 10% of the number of students, the task submission and resource allocation of the Spark cluster and the communication between nodes take too long to reflect the advantages of parallelization. However, when the amount of data increases, especially when student data set reaches 100%, the acceleration ratio increases significantly. It can be seen that Spark has more advantages for computing large data volumes, which also

shows that Spark is more suitable for big data and distributed computing.

As shown in Figure 8, the ABPM-IMISBTI method performs an efficiency analysis. During the experiment, the values of MAE and RMSE need to be predicted at different student scales. MAE and RMSE vary with the predicted number of students, but the change is small, which can verify that the ABPM-IMISBTI method in this paper has good scalability, and the prediction accuracy will not decrease as the number of students increases. At the same time, the

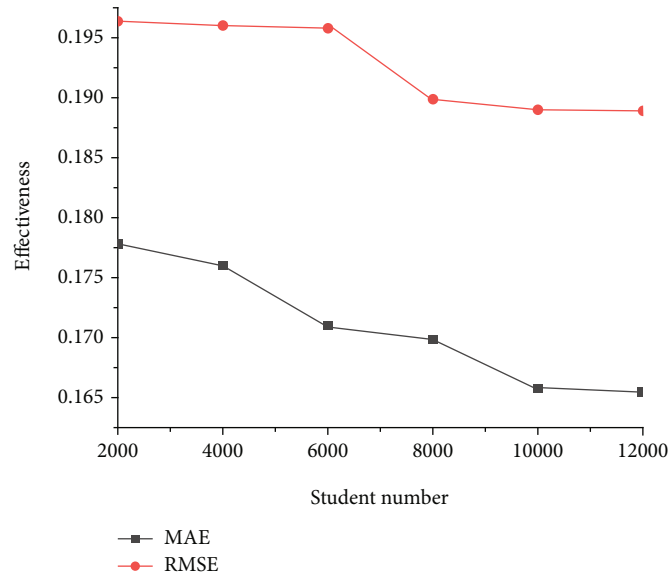


FIGURE 8: The ABPM-IMISBTI method performs an efficiency analysis.

MAE and RMSE of the predictors decrease with the increase of the student data set, indicating that the student scale has a certain influence on the accuracy of the prediction, and the data set needs a relatively large scale for better prediction, which is suitable for multidimensional prediction of student behavior.

**4.3.2. Student Behavior Analysis.** To improve the physical quality of college students, cultivate the quality of perseverance, and show the youthful vitality of college students, the university established an information system for physical exercises such as Sunshine 60 Points and morning sun plan. For example, the university organized the 2021 “Sunshine 60 Points” health walk activity, which started on April 7, 2021, and ended on May 23, 2021. A total of more than 2,000 students participated, and these records were statistically analyzed. This paper analyzes the relationship between the number of students’ physical exercise and their GPA scores.

The analysis of the relationship between students’ GPA and physical exercise is shown in Figure 8. In the Sunshine 60 Points activity, students completed three physical exercise check-in points within one hour and counted the number of check-ins. Among them, the horizontal coordinate represents the student’s GPA,  $0 < \text{GPA} < 5$ , the vertical coordinate represents the number of times a student has punched in, and the blue dot represents a certain student. As shown in Figure 9, there is no necessary relationship between students’ GPA scores and the number of physical exercises. Meanwhile, in the experiment, it was found that the number of physical exercises of junior and senior students was lower, and the number of physical exercises by freshman students was the highest. Therefore, it is necessary to educate senior students to encourage participation in activities.

**4.3.3. Student Abnormal Behavior Prediction.** This paper proposes an ABPM-IMISBTI method, which combines multiple indicators of student behavior and text information. To

verify the accuracy of the method for predicting abnormal student behavior, taking student performance prediction as an example, when students have academic abnormalities, they can quickly find there are no academically abnormal students. For student performance prediction, a prediction model is constructed by integrating study diligence indicators, including book borrowing, class rate, GPA, regular diet, study time, physical exercise, and students returning to bed. The historical scores of students are the GPA scores of students before the spring semester of 2020-2021 as the initial data, which can better reflect the students’ learning foundation, predict the students’ GPA in the fall of 2020-2021, and compare the predicted values and actual values through different methods.

The ABPM-IMISBTI method proposed in this paper is compared and analyzed with the other six methods. The first method is the MTLTR-APP method [10], which is a multi-task prediction method based on a learning ranking algorithm, used for learning performance prediction; the second method is CW-LSTM [43], which is a CW-LSTM algorithm based on deep learning theory for prediction; the third method is Bayesian [57], which is a method suitable for small-scale data and can handle multiclassification tasks. The fourth method is Deep Neural Network (DNN) [58], its feature vector reaches the output layer through hidden layer transformation, and the classification result is obtained from the output layer; the fifth method is random forest (RF) [44], random forest is used for learning performance prediction; the sixth method is IDA-SVR [45], an improved decision algorithm (IDA) to optimize support vector regression (SVR), which is a classification method that finds a regression plane so that all data of a set are closest to the plane.

In this paper, the performance of the grade prediction model is judged by comparing and analyzing the MAE and RMSE indicators between the student grade prediction results and the true value. If the values of MAE and RMSE are larger, the model is less accurate, and vice versa, the

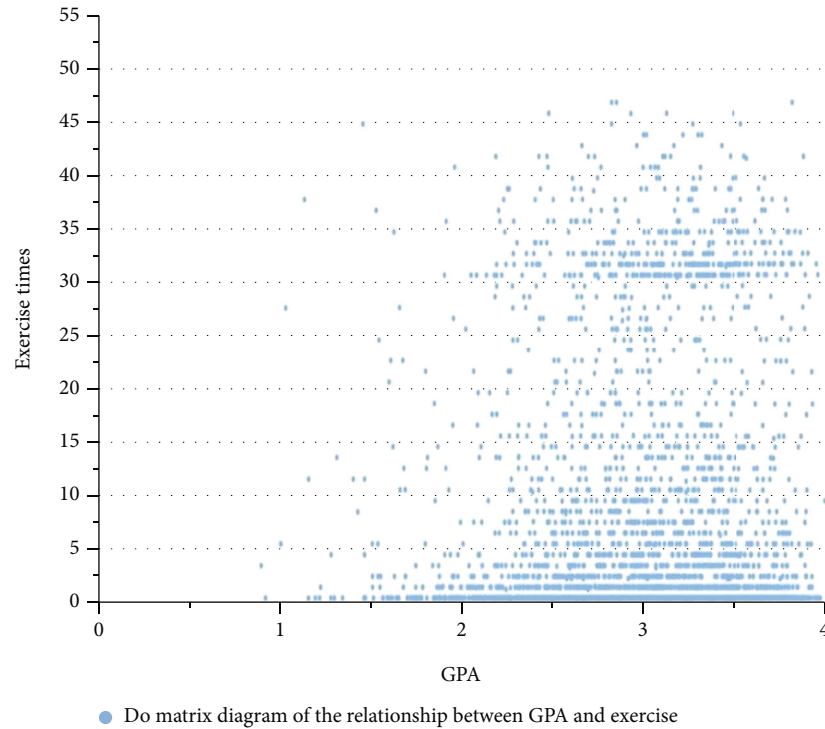


FIGURE 9: Analysis of the relationship between students' GPA and sports training.

TABLE 2: Performance comparison of different prediction methods on Big Data and Software College student dataset.

Evaluation index	ABPM-IMISBTI	MTLTR-APP	CW-LSTM	Bayesian	DNN	RF	IDA-SVR
MAE	0.1654	0.1702	0.1723	0.1798	0.1756	0.1824	0.2176
RMSE	0.1879	0.1897	0.1912	0.2134	0.1943	0.2176	0.2234

TABLE 3: Performance comparison of different prediction methods on sophomore and junior datasets.

Evaluation index	ABPM-IMISBTI	MTLTR-APP	CW-LSTM	Bayesian	DNN	RF	IDA-SVR
MAE	0.1775	0.1787	0.1812	0.1965	0.1874	0.2123	0.2254
RMSE	0.1962	0.2087	0.2101	0.2187	0.2123	0.2276	0.2356

smaller the value, the higher the accuracy. To verify the correctness of the comparative analysis model, the test data are 612 students from the School of Big Data and Software, college sophomores, and a predicted GPA of 14,800 juniors.

The performance comparisons of different prediction methods on the Big Data and Software College student datasets are shown in Table 2. The performance comparisons of different prediction methods on sophomore and junior datasets are shown in Table 3.

The ABPM-IMISBTI method in this paper integrates multidimensional information. Constructing accurate student portraits, considering students' objective labels and subjective labels, combining student behavior multi-index data, and text information to accurately predict students' abnormal behavior, and the two evaluation indicators have low errors and the best performance. MTLTR-APP is a multitask model based on matrix decomposition, which takes into account the various types of student behavior and the

similarity of students and predicts students' multitask learning ranking and academic performance, but the model does not incorporate factors such as students' classroom performance and physical exercise. This affects the accuracy of academic performance prediction. The CW-LSTM combined with the time series characteristics of college data, extracted the multidimensional characteristics of students; however, the data sample of the algorithm is small, and it does not integrate other students' classroom attendance, Internet behavior, and other data, and the accuracy rate needs to be further improved. DNN cannot use historical information to assist in the classification of students' grades, the number of parameters required for classification is too large, and the classification effect is general. Compared with the Bayesian method and the DNN method, the Bayesian classifier has a better mathematical foundation and a good classification effect, but in actual operation, the algorithm requires few parameters to run, it is less sensitive to missing

data, and the classification effect is not ideal. RF is a combined classification algorithm of ensemble learning. It is a sampling method with replacement from the original data set, and the classification effect is general. When IDA-SVR performs multiclassification predictions on grades based on study diligence indicators and living habits, it can be seen that IDA-SVR has the worst classification effect, while the overall accuracy of ABPM-IMISBTI and MTLTR-APP is relatively stable, which has a strong mathematical explanation for the data. Prediction results are compared with other prediction methods, the model has high prediction accuracy and good scalability, and the MAE and RMSE error values are the smallest. Therefore, it is feasible for this paper to use relevant data such as multiple indicators of student behavior and text information to predict student achievement. Due to the strict management of students by the relevant managers of the School of Big Data and Software, the students' class rate, study time, physical exercise, and students returning to bed are more in line with the actual situation of the students, they have higher accuracy. In summary, the model performance is ranked: ABPM-IMISBTI > MTLTR-APP > CW-LSTM > DNN > Bayesian > RF > IDA-SVR.

## 5. Conclusion

This paper proposes an ABPM-IMISBTI method. Firstly, aiming at the problems of multidimensionality, timeliness, and dynamics of student behavior information fusion in the construction of student behavior portraits, a method of constructing student behavior portraits based on the deep fusion of multidimensional information is proposed. Secondly, an abnormal behavior prediction method that integrates multiple indicators of student behavior and text information is proposed, which integrates multiple indicators of student behavior data and text information to solve the problems of accuracy and backwardness in predicting abnormal behavior of students. Finally, the real student behavior data is used to analyze the relationship between the clustering effect and student behavior, an abnormal behavior prediction model based on the Spark platform is constructed, and the student achievement prediction is used for example verification. The experimental results show that the ABPM-IMISBTI method proposed in this paper can provide early warning for students with abnormal behaviors, which is convenient for student administrators to intervene in time, thereby promoting the development of safe and smart campuses in colleges and universities.

This paper proposes the ABPM-IMISBTI method, which can solve the problems of accuracy and lag in the prediction of abnormal behavior of students. However, there are still some deficiencies in this paper. First, the number of samples needs to be further expanded. This paper studies the behavioral analysis of undergraduates. The next step is to conduct research and classification on the relevant behavioral data of master and doctoral students. Second, this paper analyzes the relationship between the number of students' physical exercise and their GPA scores, but students' scores are related to many other factors, and further quantitative anal-

ysis is needed; third, the offline information, content access, URL and other data of students' classes can be integrated. It is used for student performance prediction and personalized learning recommendation to further improve the efficiency of the algorithm. Further research is needed on these three problems.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

The work was supported by the Chongqing Natural Science Foundation of China (No. cstc2021jcyj-msxmX0515), the Chongqing Federation of Social Sciences Planning Project (No.2021NDYB110), the National Science Foundation of China (No.62072060 and No.72074036), and the Chongqing Educational Science Planning Project (No.K22YA201002).

## References

- [1] R. Zhang, W. Zhao, and Y. Wang, "Big data analytics for intelligent online education," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2815–2825, 2021.
- [2] S. Y. Zheng, S. P. Jiang, X. G. Yue, R. Pu, and B. Q. Li, "Application research of an innovative online education model in big data environment," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 8, pp. 125–138, 2019.
- [3] Q. H. Zheng, B. Dong, B. Qian et al., "The state of the art and future tendency of smart education," *Journal of Computer Research and Development*, vol. 56, no. 1, pp. 213–228, 2019.
- [4] X. Luo, C. Justice, and B. H. Sorge, "Integrated education of data analytics and information security through cross-curricular activities," in *2019 IEEE Frontiers in Education Conference (FIE)*, pp. 1–8, Covington, KY, USA, 2019.
- [5] Y. Jing, "Intelligent assessment of mental health based on multisource information fusion," *Journal of Healthcare Engineering*, vol. 2022, Article ID 7149817, 8 pages, 2022.
- [6] Y. Cao, J. Gao, D. Lian et al., "Orderliness predicts academic performance: behavioural analysis on campus lifestyle," *Journal of the Royal Society Interface*, vol. 25, no. 146, article 20180210, 2018.
- [7] J. Yue, F. Tian, K. M. Chao et al., "Recognizing multidimensional engagement of e-learners based on multi-channel data in e-learning environment," *IEEE Access*, vol. 7, pp. 149554–149567, 2019.
- [8] Y. U. Hongwei, Z. Dongbo, X. U. Wenhui, Y. U. Yaying, W. Xiaomei, and T. Yue, "Daily behavior prediction of college students on campus based on multi fragment semantic spatio-temporal graph convolutional network," *Journal of Tsinghua University (Science and Technology)*, vol. 62, no. 1, pp. 105–115, 2022.
- [9] S. J. Nam and P. Samson, "Integrating students' behavioral signals and academic profiles in early warning system," in

- Artificial Intelligence in Education. AIED 2019*, Lecture Notes in Computer Science, S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, and R. Luckin, Eds., pp. 345–357, Springer, Cham, 2019.
- [10] H. Yao, D. Lian, Y. Cao, Y. Wu, and T. Zhou, “Predicting academic performance for college students: a campus behavior perspective,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 3, pp. 1–21, 2019.
- [11] Y. Wang, J. Wen, W. Zhou, X. Wang, Q. Wu, and B. Tao, “A cloud service recommendation method based on extended multi-source information fusion,” *Concurrency and Computation: Practice and Experience*, vol. 34, no. 10, article e6826, 2022.
- [12] Y. Zhang, F. Zhang, Y. Jin, Y. Cen, V. Voronin, and S. Wan, “Local correlation ensemble with GCN based on attention features for cross-domain person Re-ID,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 2022, pp. 1–22, 2022.
- [13] J. Gou, L. Sun, B. Yu, S. Wan, W. Ou, and Z. Yi, “Multi-level attention-based sample correlations for knowledge distillation,” *IEEE Transactions on Industrial Informatics*, vol. 2022, pp. 1–11, 2022.
- [14] D. Chawla and H. Joshi, “Importance-performance map analysis to enhance the performance of attitude towards mobile wallet adoption among Indian consumer segments,” *Aslib Journal of Information Management*, vol. 73, no. 6, pp. 946–966, 2021.
- [15] W. Zheng and L. Jiang, “Algorithm analysis for big data in education based on depth learning,” *Wireless Personal Communications*, vol. 102, no. 4, pp. 3111–3119, 2018.
- [16] L. Zhao, K. Chen, J. Song et al., “Academic performance prediction based on multisource, multifeature behavioral data,” *IEEE Access*, vol. 9, pp. 5453–5465, 2019.
- [17] F. Cappa, R. Oriani, E. Peruffo, and I. McCarthy, “Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance,” *Journal of Product Innovation Management*, vol. 38, no. 1, pp. 49–67, 2021.
- [18] Z. Hu, Y.-s. Li, X.-n. Zh, and J.-g. Yu, “Location prediction model based on user behavior sequence features,” *Journal of Beijing University of Posts and Telecommunications*, vol. 42, no. 6, p. 149, 2019.
- [19] S. M. Darwish and S. K. Mohamed, “Automated essay evaluation based on fusion of fuzzy ontology and latent semantic analysis,” in *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)*. AMLTA 2019, A. Hassanien, A. Azar, T. Gaber, R. F. Bhatnagar, and M. Tolba, Eds., vol. 921 of Advances in Intelligent Systems and Computing, pp. 566–575, Springer, Cham, 2019.
- [20] X. Wang, L. Duan, D. Cui, L. Lu, Y. Dun, and R. Qin, “Search of students with similar lifestyle based on campus behavior information network,” *Jisuanji Yanjiu yu Fazhan*, vol. 57, no. 11, pp. 2442–2455, 2020.
- [21] L. S. Maurya, M. S. Hussain, and S. Singh, “Developing classifiers through machine learning algorithms for student placement prediction based on academic performance,” *Applied Artificial Intelligence*, vol. 35, no. 6, pp. 403–420, 2021.
- [22] D. Li, Q. Wu, M. Zhou, and F. Luo, “HHFS: a hybrid hierarchical feature selection method for ageing gene classification,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 2022, p. 1, 2022.
- [23] W. Huang, Y. Zhang, and S. Wan, “A sorting fuzzy min-max model in an embedded system for atrial fibrillation detection,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 18, no. 2s, pp. 1–18, 2022.
- [24] X. Li, X. Zhu, X. Zhu, Y. Ji, and X. Tang, “Student academic performance prediction using deep multi-source behavior sequential network,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 567–579, Springer, Cham, 2020.
- [25] K. Liang, J. Liu, and Y. Zhang, “The effects of non-directional online behavior on students’ learning performance: a user profile based analysis method,” *Future Internet*, vol. 13, no. 8, p. 199, 2021.
- [26] Q. Qing, T. Guo, D. Zhang, and F. Xia, “ANSWER: generating information dissemination network on campus,” in *Databases Theory and Applications. ADC 2021*, Lecture Notes in Computer Science, M. Qiao, G. Vossen, S. Wang, and L. Li, Eds., pp. 74–86, Springer, Cham, 2021.
- [27] A. Gonzalez-Nucamendi, J. Noguez, L. Neri, V. Robledo-Rella, R. M. G. Garcia-Castelán, and D. Escobar-Castillejos, “The prediction of academic performance using engineering student’s profiles,” *Computers & Electrical Engineering*, vol. 93, article 107288, 2021.
- [28] H. Li, J. Liu, B. Cao, and M. Shi, “Topic-adaptive web API recommendation method via integrating multidimensional,” *Ruan Jian Xue Bao/Journal of Software*, vol. 29, no. 11, pp. 3374–3387, 2018.
- [29] Y. Zhu, H. Lu, P. Qiu, K. Shi, J. Chambua, and Z. Niu, “Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization,” *Neurocomputing*, vol. 415, pp. 84–95, 2020.
- [30] X. Chen, M. Tong, C. Shi et al., “Visual analysis of multi-source college students’ mental health questionnaire data,” *Journal of Computer-Aided Design & Computer Graphics*, vol. 32, no. 2, pp. 181–193, 2020.
- [31] Q. Liu, Z. Huang, Z. Huang et al., “Finding similar exercises in online education systems,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1821–1830, United Kingdom, 2018.
- [32] Y. Bo, C. Wang, L. Li, and X. Lu, “Student ability portrait construction research based on Big Data,” in *2022 International Conference on Big Data, Information and Computer Network (BDICN)*, pp. 156–160, Sanya, China, 2022.
- [33] S. Zhang II, M. Shen, and Y. Yu III, “Research on student portrait system based on canopy and k-means algorithm,” in *International Conference on Computer Application and Information Security (ICCAIS 2021)*, vol. 12260, pp. 486–492, Wuhan, China, 2021.
- [34] X. Li and S. He, “Research and analysis of student portrait based on campus big data,” in *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*, pp. 23–27, Xiamen, China, 2021.
- [35] Z. Ding and X. Li, “College students’ portrait technology based on hybrid neural network,” in *Spatial Data and Intelligence. SpatialDI 2020*, X. Meng, X. Xie, Y. Yue, and Z. Ding, Eds., vol. 12567 of Lecture Notes in Computer Science, pp. 165–183, Springer, Cham, 2020.
- [36] Y. Wang, J. Wen, X. Wang, B. Tao, and W. Zhou, “A cloud service trust evaluation model based on combining weights and gray correlation analysis,” *Security and Communication Networks*, vol. 2019, Article ID 2437062, 11 pages, 2019.



- [37] S. Zhang, M. Shen, and Y. Yu, "Research on student Big Data portrait method based on improved K-means algorithm," in *2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, pp. 146–150, Guangzhou, China, 2021.
- [38] C. Ni and L. Wang, "The application of user portrait based on Big Data analysis in industrial design teaching," *E3S Web of Conferences*, vol. 179, article 02110, 2020.
- [39] Y. Jiang, Y. Zhao, and H. Xu, "Analysis and portrait of college students' online behavior habits," *Journal of Physics: Conference Series*, vol. 1302, no. 2, article 022007, 2019.
- [40] J. Deng, "College students' online behavior analysis during epidemic prevention and control," in *2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 796–799, Beihai, China, 2021.
- [41] L. You, L. Wei, L. Junzhou, J. Jian, and X. Nu, "A network user's abnormal behavior detection approach based on selective collaborative learning," *Chinese Journal of Computers*, vol. 37, no. 1, pp. 28–40, 2014.
- [42] Y. Zeng, Y. Ouyang, R. Gao, Y. Qiu, Y. Yu, and C. Wang, "HHA: an attentive prediction model for academic abnormality," *IEEE Access*, vol. 8, pp. 124755–124766, 2020.
- [43] M. Nie, Y. Zhang, H. Deng, W. Wang, H. Hie, and Z. Zhou, "Identifying poor students in universities by using basic information and behavioral data," *Journal of University of Electronic Science and Technology of China*, vol. 49, no. 5, pp. 795–800, 2020.
- [44] J. Yu, "Academic performance prediction method of online education using random forest algorithm and artificial intelligence methods," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 5, pp. 45–57, 2021.
- [45] H. Xu, "Prediction of students' performance based on the hybrid IDA-SVR model," *Complexity*, vol. 2022, Article ID 1845571, 11 pages, 2022.
- [46] N. Maksimova, A. Pentel, and O. Dunajeva, "Predicting first-year computer science students drop-out with machine learning methods: a case study," in *Educating Engineers for Future Industrial Revolutions. ICL 2020*, Advances in Intelligent Systems and Computing, M. E. Auer and T. Rüttemann, Eds., pp. 719–726, Springer, Cham, 2021.
- [47] X. Chen, "Dynamic early-warning model of college students' psychological crisis based on characteristic attribute," *Scientific Programming*, vol. 2022, Article ID 6026248, pp. 1–11, 2022.
- [48] Y. Cao, "Portrait-based academic performance evaluation of college students from the perspective of Big Data," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 4, pp. 95–106, 2021.
- [49] Y. Zhao, W. Ren, and Z. Li, "Prediction of English scores of college students based on multi-source data fusion and social behavior analysis," *Revue d'Intelligence Artificielle*, vol. 34, no. 4, pp. 465–470, 2020.
- [50] Y. Li and D. Li, "University Students' behavior characteristics analysis and prediction method based on combined data mining model," in *Proceedings of the 2020 the 3rd International Conference on Computers in Management and Business*, pp. 9–13, Tokyo, Japan, 2020.
- [51] D. A. Filvã, M. A. Forment, F. J. García-Peñalvo, D. F. Escudero, and M. J. Casañ, "Clickstream for learning analytics to assess students' behavior with Scratch," *Future Generation Computer Systems*, vol. 93, pp. 673–686, 2019.
- [52] C. Y. Yang, J. Y. Liu, and S. Huang, "Research on early warning system of college students' behavior based on Big Data environment," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3/W10, pp. 659–665, 2020.
- [53] Y. Xie, S. Zhang, and Y. Liu, "Abnormal behavior recognition in classroom pose estimation of college students based on spatiotemporal representation learning," *Traitement du Signal*, vol. 38, no. 1, pp. 89–95, 2021.
- [54] H. Liu, W. Ao, and J. Hong, "Student abnormal behavior recognition in classroom video based on deep learning," in *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering*, pp. 664–671, Xiamen, China, 2021.
- [55] C. Shi and Y. Tan, "A BP neural network-based early warning model for student performance in the context of big data," *Journal of Sensors*, vol. 2022, Article ID 2958261, 10 pages, 2022.
- [56] Y. Yang, D. Hooshyar, M. Pedaste, M. Wang, Y. M. Huang, and H. Lim, "Predicting course achievement of university students based on their procrastination behaviour on Moodle," *Soft Computing*, vol. 24, no. 24, pp. 18777–18793, 2020.
- [57] L. Wang and M. Yang, "Physical fitness evaluation of college students at the stage of physical exercise behavior based on bayesian and data mining," *Scientific Programming*, vol. 2022, Article ID 9582690, 9 pages, 2022.
- [58] R. Xu, C. Wang, and Y. A. Hsu, "A study of DNN-based media literacy and distance education management system in the context of participatory culture," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 3666274, 11 pages, 2022.
- [59] N. Ruiz, H. Yu, D. A. Alessio et al., "ATL-BP: a student engagement dataset and model for affect transfer learning for behavior prediction," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2022, p. 1, 2022.
- [60] Y. Zhang, Y. Yun, H. Dai, J. Cui, and X. Shang, "Graphs regularized robust matrix factorization and its application on student grade prediction," *Applied Sciences*, vol. 10, no. 5, article 1755, 2020.
- [61] X. Cheng, Z. Zhu, X. Liu et al., "A novel learning early-warning model based on random forest algorithm," in *Intelligent Tutoring Systems. ITS 2018*, R. Nkambou, R. Azevedo, and J. Vassileva, Eds., vol. 10858 of Lecture Notes in Computer Science, pp. 306–312, Springer, Cham, 2018.