Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7403037, 10 pages
https://doi.org/10.1155/2022/7403037

Research Article

Research on Data Mining Combination Model Analysis and Performance Prediction Based on Students' Behavior Characteristics

Liyan Chen, Lihua Wang, and Yuxin Zhou

College of Animal Science and Technology, Inner Mongolia Minzu University, Tongliao 028000, China

Correspondence should be addressed to Lihua Wang; wanglihua7967@imun.edu.cn

Received 12 January 2022; Revised 13 February 2022; Accepted 18 February 2022; Published 23 March 2022

Academic Editor: Gengxin Sun

Using data mining technology to analyze students' behavior can effectively predict students' performance and other evaluation indicators, which is of great significance to improving the level of school information management. Aiming at the problems of imperfect information management platforms and low data analysis ability in colleges and universities, a data mining algorithm based on students' behavior characteristics is proposed. Firstly, the characteristics of the GBDT algorithm, the ANN algorithm, and the K-means algorithm are analyzed, and the three algorithms are combined to establish a combined prediction model. At the same time, five standard data sets are combined for simulation training. Then, an analysis and prediction platform based on the characteristics of students' behavior is built. Combined with the management systems such as "Campus All-in-one Card," educational administration, and library, the data collection, modeling, analysis, and mining are realized, and the evaluation index system of students' behavior is established. Finally, the data of students' consumption laws, living habits, learning, and Internet access are used to verify the combined model. The results show that compared with a single algorithm, the combined model has fast run speed and high accuracy, and the prediction results are consistent with the actual situation. The prediction platform can analyze the characteristics and laws of student behavior data, effectively predict the learning effects, grasp the dynamics of students' lives and learning in real time, and provide decision-making for school teaching management and teachers' teaching reform.

1. Introduction

The fundamental goal of a college education is to track the whole process of students' learning and life and carry out targeted teaching and management. However, with the continuous enrollment expansion of colleges and universities, the number of students is increasing, which makes the proportion of teachers and students difficult to meet the teaching requirements, thus affecting the teaching quality to a certain extent. With the development of network technology, students' behavior data such as learning, consumption, and life on campus are continuously stored in various management system platforms of the school, which has formed a relatively complete campus big data environment. Combining machine learning and big data technology to deeply mine a large number of student behavior data and analyze the hidden behavior laws can not only realize the efficient management and sharing of campus data and improve the campus information level but also timely predict the students' academic performance and teaching effects, so as to enable students to develop healthy life rules and good learning habits and supervise teachers to improve teaching methods to achieve the purpose of improving the quality of education [1, 2]. The fundamental goal of a college education is to track the whole process of students' learning and life and carry out targeted teaching and management. However, with the continuous enrollment expansion of colleges and universities, the scale of students is increasing, which makes the proportion of teachers and students difficult to meet the teaching requirements, thus affecting the teaching quality to a certain extent. With the development of network technology, students' behavior data such as learning, consumption, and life on campus are continuously stored in various management system platforms of the school, which has formed a relatively complete campus big data environment.

Using student behavior data mining and analysis has important application value and practical significance for predicting student characteristics and grades, so scholars have carried out a lot of research. Among them, Rodrigues [3] uses the clustering method to analyze students' participation in online learning to provide a reference for students' performance prediction. Elbadrawy [4] used the methods of multiple regression and matrix decomposition to predict students' course performance and achieved good results. Sweeney [5] predicts students' grades based on a factor decomposition machine algorithm, which has good prediction accuracy. Johnson [6] uses cluster analysis, decision tree analysis, and neural network algorithms to analyze the relevant factors affecting dropouts and designs a student behavior prediction system to effectively reduce the probability of students dropping out. REN [7] designed a multiple linear regression model to predict students' online learning performance, which can grasp students' learning enthusiasm in real time and effectively predict students' course performance. Berhanu [8] uses rapid miner software to preprocess students' learning data and uses a decision tree to classify them, which has achieved high accuracy. Among the research results, most of them adopt a single algorithm or improve a certain algorithm with low accuracy, and relatively few index factors are considered in classification and prediction, so they cannot fully analyze the behavioral characteristics of students.

In order to improve the effectiveness of student behavior data mining, focusing on the characteristic data of students' consumption law, living habits, learning, and scientific research abilities in the big data environment, this paper integrates GBDT, artificial neural networks (ANNs), and K-means to construct a multialgorithm combination prediction model. Based on the combination algorithm, an analysis and prediction platform based on students' behavior characteristics is established to classify and predict students' grades, so that schools and teachers can master students' life and learning dynamics in real time, guide them in time, and realize the effective management of students.

2. Data Mining Theory

2.1. Connotation of Data Mining. Generally, the process of knowledge discovery in a database is called data mining. The essence of data mining is to mine the possible valuable information that people cannot see from the surface, do not know in advance, and hidden from the massive, disordered, noisy, and random data, find out the correlation between different data, and transform it into content with certain practical significance [9, 10]· Data mining is a process of continuous cyclic optimization, which is mainly divided into data preprocessing, feature extraction, data mining, and model evaluation, as shown in Figure 1.

(1) Data preprocessing: also known as data preparation, it is a prerequisite for the accuracy of data mining results. The function is to clean and screen the collected messy and irregular data according to certain rules and process them into standard data that can be used in training and experiments. Preprocessing includes multiple processing processes such as data acquisition, data cleaning, and standardized processing to solve the problems of data loss, noise, and redundancy, so as to form a unified standard format and provide a data basis for data mining.

(2) Data mining: using a certain algorithm to extract information from the processed data to get the desired information is the core of data mining. Different mining algorithms have different methods of data extraction and processing, and the results are also different. The most appropriate and effective mining algorithm can be selected according to different data characteristics and business needs.

(3) Model evaluation: in order to evaluate whether the data mining results meet the expected requirements, the model needs to be evaluated. By solving the accuracy, precision, and recall of the model, verify the advantages and disadvantages of the model algorithm, and constantly improve and optimize it, so as to find the best model suitable for training and experiment.

2.2. Introduction to Related Algorithms

GBDT Algorithm. Gradient boosting decision trees (GBDT) are based on the Iterative Regression decision tree algorithm, which is a combination of boosting and decision tree (DT). Its core idea is to use multiple learners with poor performance to train iteratively and combine them into a learner with significantly improved performance. The residual of DT is used as the input of the next DT so that the loss caused by each iteration decreases along the negative gradient direction. Finally, the prediction results are determined according to the sum of all DTS [11, 12]. GBDT has the characteristics of short adjustment time and high prediction accuracy. Its model is shown in Figure 2.

ANN Algorithm. Artificial neural network (ANN) is an algorithm proposed by simulating the function of the human brain. It has a strong approximation function and is used to represent the mapping relationship from multi-input to single output. Its principle is shown in Figure 3.

Assuming that w_{ji} is the weight value corresponding to the input variable and the threshold of the neuron is θ_j, the output result can be expressed as [13]

$$R_j = \sum_{i=1}^{n} w_{ji} x_i - \theta_j, \quad i \neq j. \tag{1}$$

In the process of training and learning, the neural network automatically adjusts the connection threshold between neurons according to certain rules, looks for the

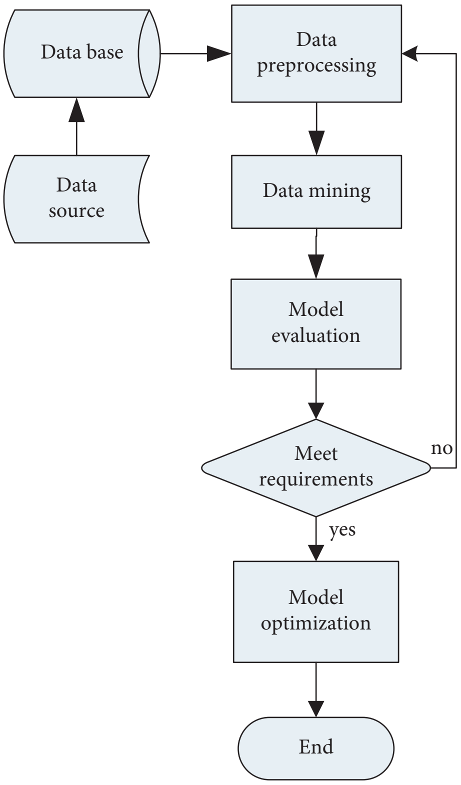

Figure 1: Data mining process.

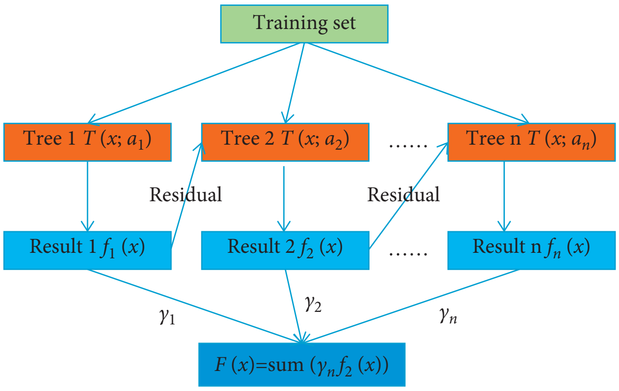

Figure 2: GBDT model.

best objective function, and can automatically adjust according to the changes in external conditions.

K-Means Algorithm. The K-means algorithm is a common unsupervised clustering algorithm. It classifies data objects based on similarity and judges the similarity between data by calculating the distance between different data points, so as to realize data classification. Before classification, the K-means algorithm randomly selects the clustering center and then continuously concentrates the data with high similarity to achieve the purpose of model convergence. The K-means algorithm has the characteristics of simple operation, strong data processing ability, fast iteration speed, and a reliable model. Its classification process is shown in Figure 4.

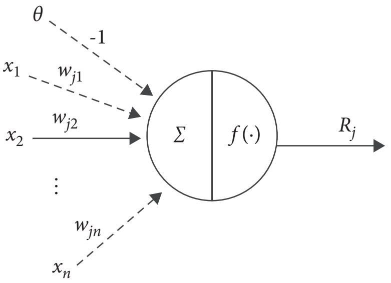

Figure 3: ANN model.

2.3. Evaluating Indicator. After the modeling of machine learning, data mining, and prediction system is completed, the effect of the model needs to be evaluated. The evaluation indexes of machine learning algorithm mainly include classification and regression. The classification evaluation indexes are established based on confusion matrix, and their rules are shown in Table 1.

In Table 1, TP indicates that the predicted true sample is actually a true sample, that is, the true sample is predicted as a true sample. TN indicates that it is predicted to be a false sample, which is also a false sample, that is, the false sample is predicted to be a false sample. FP indicates that the predicted true sample is actually a false sample, that is, the false sample is predicted as a true sample (false alarm). FN indicates that the predicted false sample is actually a true sample, that is, the true sample is predicted as a false sample (missing report). The commonly used evaluation indexes include accuracy, precision, recall, comprehensive evaluation index value (F1), etc.

Accuracy. It indicates the ratio of the number of correctly predicted samples to the total number of samples in the prediction results. The greater the ratio, the better the ideal state, *accuracy* = 1, and its expression is [14, 15]

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{2}$$

Precision. It represents the ratio of the actual true sample to the real sample. The greater the ratio, the better the ideal state, *precision* = 1, and its expression is

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3}$$

Recall. It indicates the ratio of the number of correctly retrieved samples to the total number of retrieved samples. The greater the ratio, the better the ideal state, *recall* = 1, and its expression is

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{4}$$

F_1. It represents the weighted harmonic average of *precision* and *recall*, and its expression is

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \tag{5}$$

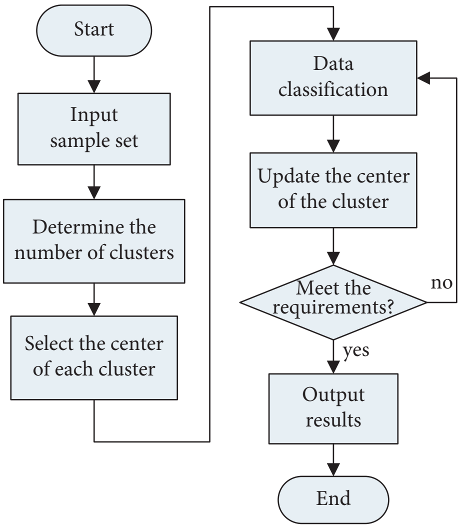
Start
Input sample set
Determine the number of clusters
Select the center of each cluster
Data classification
Update the center of the cluster
Meet the requirements?
no
yes
Output results
End

Figure 4: Process of K-means algorithm.

Table 1: Confusion matrix.

Prediction results	Negative	Positive
False	False negative (FN)	False positive (FP)
True	True negative (TN)	True positive (FN)

3. Multialgorithm Combined Data Mining Model Based on Student Behavior Characteristics

3.1. Construction of Multialgorithm Combination Model. In order to further mine students' behavior characteristics, the GBDA algorithm, the ANN algorithm, and the K-means algorithm are fused to construct a combined prediction model, and its function can be expressed as

$$\begin{aligned}(\lambda_1,\lambda_2,\lambda_3) = \sum_{i=1}^{3} [&(\lambda_1 x_i + \lambda_2 y_i + \lambda_3 z_i - x_i)^2 \\ &+ (\lambda_1 x_i + \lambda_2 y_i + \lambda_3 z_i - y_i)^2 \\ &+ (\lambda_1 x_i + \lambda_2 y_i + \lambda_3 z_i - z_i)^2 \\ &+ \eta (\lambda_1 x_i + \lambda_2 y_i + \lambda_3 z_i - 1).\end{aligned} \quad (6)$$

where x_i, y_i, z_i is the predicted value of different models; λ_1, λ_2, λ_3 is weight coefficient of the combined model; and η is coefficient of function.

Lagrange transformation is performed on formula (6), and the following formula is obtained:

$$\xi_i^* = \lambda_1^* x_i + \lambda_2^* y_i + \lambda_3^* z_i, \quad (7)$$

where ξ_i^* is predicted value of group i objects and λ_k^* is the extreme value of the weight coefficient.

The prediction steps of the combined model are as follows.

Step 1. Select the forecast sample. The data of the students' behavior characteristics are selected and preprocessed.

Step 2. Sample division. Divide the selected data samples, in which training set samples account for 70% and test set samples account for 30%.

Step 3. Modeling. Different algorithms are used to model the training set.

Step 4. Single model prediction. The first is the prediction. Predict the first mock exam sample with a single model.

Step 5. Substitute the prediction results in step 4 into formulas (4) and (5), calculate the weight coefficient, and establish a combined prediction model.

Step 6. Combined model prediction. In the combined model, the sample data in the test set are predicted and the prediction results are obtained.

The specific process is shown in Figure 5.

3.2. Simulation Analysis

Comparison of Iteration Times. GBDT algorithm, the ANN algorithm, the K-means algorithm, and the combination algorithm proposed in this paper are used for simulation analysis, and five data sets of iris, wine, glass, balance scale, and abalone in the UCI database are used for training. The data set in the UCI database is a common standard test data set, which has different properties. It can verify the effectiveness and feasibility of the composite model from multiattribute and multidimensional [16–18]. Set the mean square deviation of the results of two adjacent iterations to reach the convergence threshold of 0.0001 as the iteration termination condition. During the training process, the data sets used are tested ten times, respectively, and the average value is taken as the training result. The number of iterations of different algorithms on each data set is shown in Figure 6.

It can be seen that compared with a single algorithm, the number of iterations of the combined algorithm on different data sets is small, indicating that the combined algorithm has fast operation speed and can quickly meet the training requirements.

Run Time Comparison. The running time of different algorithms on each data set is shown in Table 2. The results show that the running time of the combined algorithm is less than that of the single algorithm on any data set, which further shows that the combined algorithm has a fast convergence time, high efficiency, and good training effect.

Comparison of Prediction Accuracy. The accuracy of different algorithms on each data set is shown in Figure 7. It can be seen that compared with a single algorithm, the clustering accuracy of the combined algorithm is high, and the accuracy on different data sets is basically the same, indicating that the algorithm has certain stability.

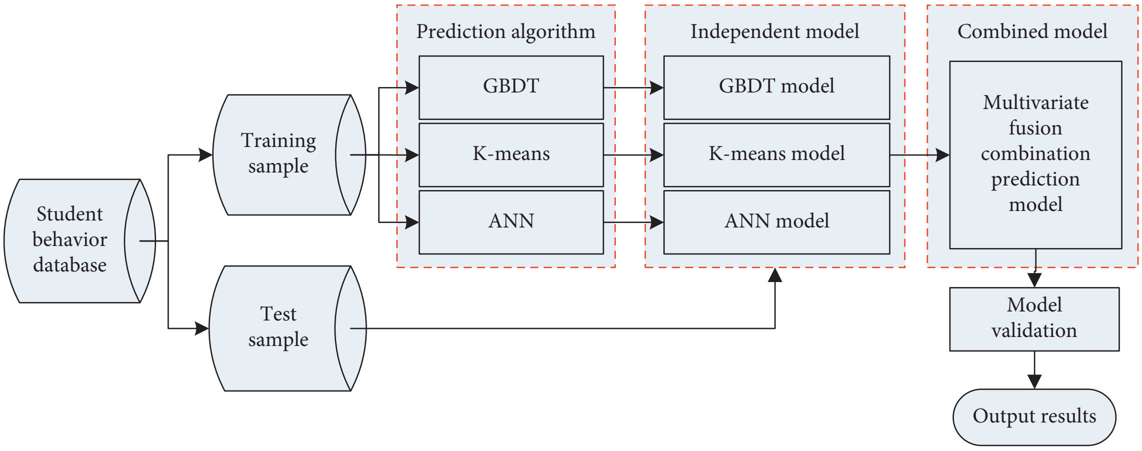

Figure 5: Prediction process of combined model.

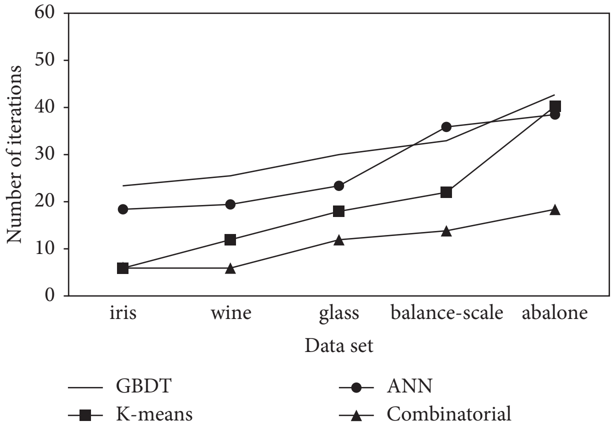

Figure 6: Iterations of different algorithms.

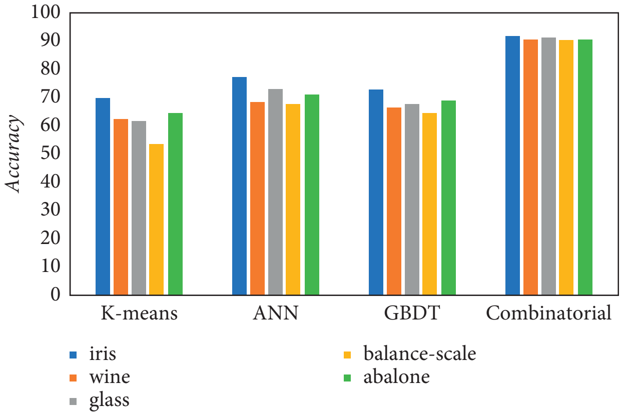

Figure 7: Accuracy comparison results.

Table 2: Running time of different algorithms on each data set.

Data set	Running time (s)			
	GBDT	ANN	K-means	Combinatorial
Iris	0.036	0.039	0.044	0.029
Wine	0.107	0.118	0.162	0.101
Glass	0.348	0.251	0.234	0.189
Balance-scale	0.225	0.246	0.214	0.196
Abalone	0.336	0.641	0.3668	0.275

Comparison of Evaluation Results. In order to verify the accuracy of classification and prediction of the combined model, combined with the classification results of the algorithm, the classification results of several algorithms are evaluated by using the values of accuracy, precision, recall, and F1. The results are shown in Table 3.

The evaluation results show that compared with other single algorithms, the combined algorithm proposed in this paper has the best prediction effect, strong generalization ability, and high prediction accuracy.

4. Construction of Analysis and Prediction Platform Based on Students' Behavior Characteristics

4.1. System Platform Architecture. According to the characteristics of student behavior data, in order to improve data processing efficiency and better apply to data mining and machine learning problems, a student behavior analysis and prediction platform is established based on a distributed parallel spark computing engine, as shown in Figure 8. Firstly, collect the behavior information of middle school students' consumption records, attendance results, course scores, and book borrowing on each management platform of the school as the data source. Then the preprocessed behavior data is stored in the student feature database. In order to ensure convenient data conversion, it is necessary to keep consistent with the data type in the relational database. Finally, cluster analysis is carried out to complete the classification, analysis, and mining of students' behavior so as to predict students' life and learning by analyzing their behavior characteristics.

Table 3: Evaluation results of different models.

Evaluating indicator	Evaluation results			
	GBDT	ANN	K-means	Combinatorial
Accuracy	0.6725	0.7764	0.8021	0.8692
Precision	0.7023	0.8679	0.7928	0.8746
Recall	0.7936	0.8256	0.8436	0.8827
F1	0.8304	0.8525	0.8211	0.8978

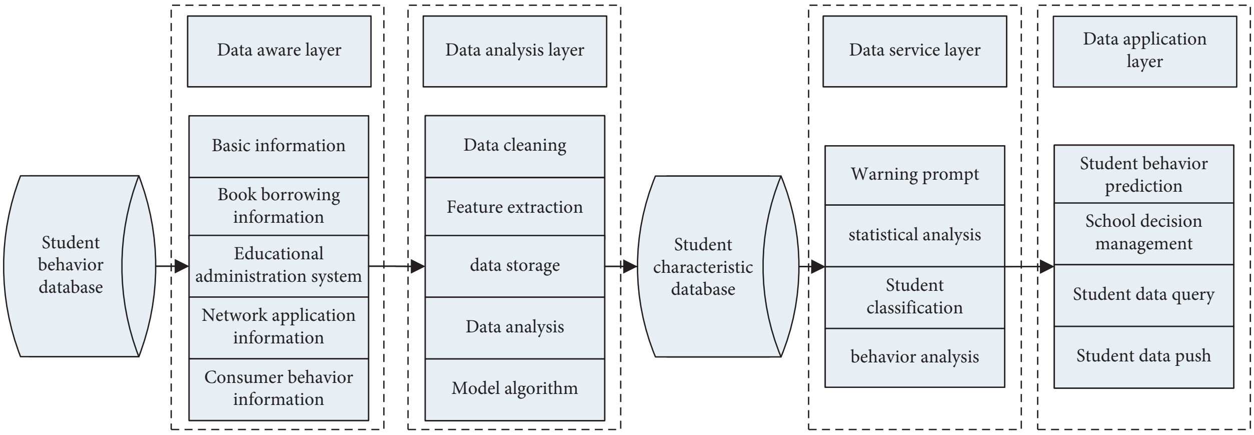

Figure 8: System platform architecture.

4.2. Student Behavior Index System. In order to realize the analysis and mining of students' school behavior, it is necessary to classify students' behavior indicators. According to the actual situation of students, this paper establishes a student behavior index system, which includes four categories: students' consumption laws, learning situations, living habits, and scientific research. Each category is refined to provide data support for the accuracy of behavior analysis results. The student behavior index system is shown in Figure 9.

5. Example Verification

5.1. Data Source and Collection. Using the all-in-one card records of 2018 students in a university and the student behavior information in the management system of relevant functional departments as the data source, collect the data of 10000 students from March 2020 to December 2020, such as classroom attendance, canteen consumption records, library borrowing, and self-study, Internet use, and physical exercise. A total of 3516000 records were collected.

5.2. Data Processing. Due to the huge amount of data and from different departments, the data storage format, data type, field standard, and other indicators are different. The collected data is often disordered and the quality is not very high, so it needs to be preprocessed. Firstly, the methods of data merging and unified type are used to clean the data and eliminate redundant and missing data. Then remove the noise in the data, merge processing, generalize processing, and realize data conversion. Finally, the data from different sources are integrated to reduce the data dimension and complexity and meet the data mining standards.

Consumption Law. Analyze the students' consumption behavior data of the "all-in-one card", extract the consumption records including dining and fetching water in the student canteen, and take the fields such as transaction time, transaction amount, and transaction type as the data feature source, so as to find out the students' consumption law and consumption level. Some data are shown in Table 4.

Learning Situation. The main purpose of mining students' behavior characteristics is to predict their effort and academic achievement by analyzing their learning situations. Extract various learning information from students and use the fields such as classroom attendance, course performance, book borrowing, learning duration, and learning habits as data feature sources to master students' learning dynamics. Some data are shown in Table 5.

Habits and Customs. Extract students' life details, and take students' work and rest time, exercise, Internet access, and other fields as data feature sources to understand students' daily living habits and laws. Some living habits data are shown in Table 6.

Scientific Research. By analyzing the students' participation in scientific research, we can judge the students' ability and quality, and extract the fields, including participation in topics, number of papers, paper level, and participation in academic forums, as the data feature source to master the students' scientific research ability, as shown in Table 7.

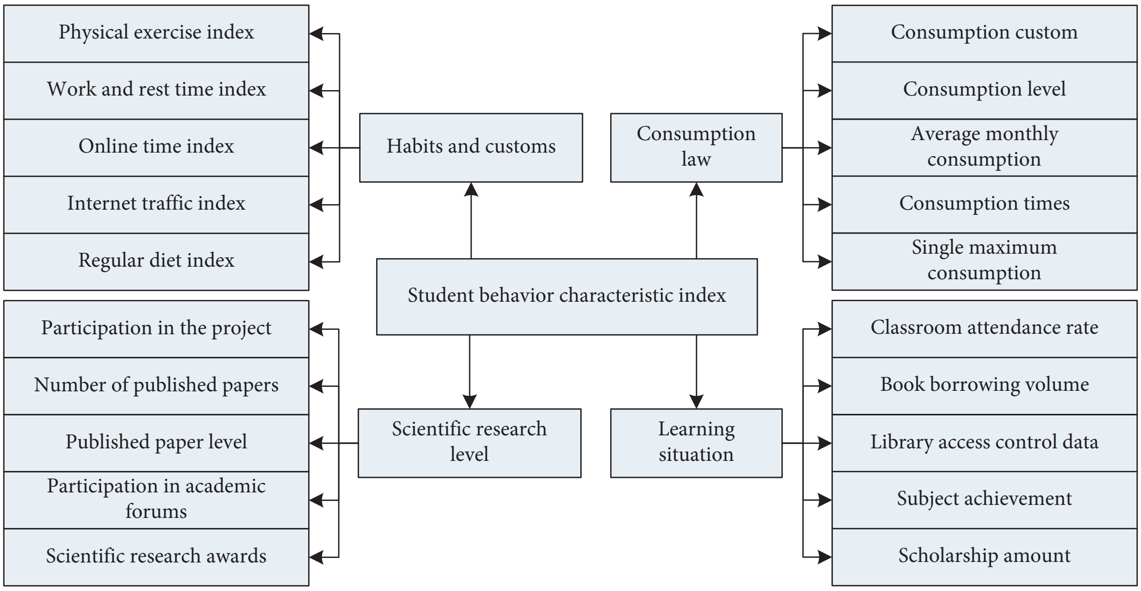

Figure 9: Student behavior index system.

Table 4: Consumption law.

Student ID	Consumption date	Consumption amount	Consumption type
20180105001	2020-03-21 07 : 10 : 15	2.0	Canteen
20180105002	2017-03-21 07 : 14 : 20	6.8	Canteen
20180105003	2017-03-22 12 : 05 : 03	7.8	Canteen
20180105001	2017-03-22 12 : 07 : 26	4.5	Canteen
20180105001	2017-03-22 17 : 45 : 25	5.0	Canteen
……	……	……	……

Table 5: Learning situation.

Student ID	Semester	Course code	Achievement	Credit
20180105001	2019–2020	1070310010	88	2
20180105002	2019–2020	1070310010	79	2
20180105003	2019–2020	1070310010	63	2
20180105001	2019–2020	1070170011	90	2.5
20180105001	2019–2020	1070170011	56	2.5
……	……	……	……	……

Table 6: Habits and customs.

Student ID	Semester	Online time (min)	Flow (MB)	IP
20180105001	2020-05-13	124	1224	10.122.1.12
20180105002	2020-05-13	168	2678	10.122.1.22
20180105003	2020-05-15	67	215	10.122.1.67
20180105001	2020-05-15	206	309	10.122.1.16
20180105001	2020-05-16	98	114	10.122.1.38
……	……	……	……	……

5.3. Student Behavior Classification and Prediction Experiment. After preprocessing, 223,5000 pieces of data meeting the test requirements are extracted according to the classification indicators. 300,000 pieces of data on consumption law, learning, living habits, and scientific research are selected to provide a data basis for students' classification and behavior prediction.

Classification of Students' Efforts. Extract information such as students' classroom attendance, number of books borrowed, library self-study time, course scores, examination pass rate, average daily online time. and number of papers. After data cleaning, integration, and conversion, they are imported into the system platform, and the effort and achievement indicators of students are subdivided by using the combination

Table 7: Scientific research.

Student ID	Year	Number of papers	Paper level
20180105001	2020	1	CNKI
20180105002	2020	0	—
20180105003	2020	2	EI
20180105001	2020	0	—
20180105001	2020	1	CNKI
……	……	……	……

Table 8: Classification of students' effort.

Category	Attendance (%)	Course passing rate (%)	Course achievement	Borrowing quantity	Library study time every day (min)	Internet time (min)	Number of papers	Student ratio (%)
1	97	100	90.25	39	245	108	2	15.15
2	98	100	89.08	43	226	88	1	17.36
3	95	98	88.16	32	179	165	1	16.56
4	87	95	82.34	12	98	175	1	18.37
5	88	92	79.18	28	125	155	0	11.89
6	84	89	75.29	8	106	231	0	7.02
7	72	86	65.82	15	112	224	0	8.96
8	65	70	58.25	7	62	269	0	4.59

model. Students can be divided into eight categories. The specific values of each indicator are shown in Table 8.

The results show that although there are differences in the overall distribution of grades, most students study hard and their grades are above average. If the average score of more than 90 points is recorded as excellent, the students who study hard and have excellent results account for 15.15% of the total. Only a very small number of students do not work hard enough, have poor results, and do not reach the qualified level. Such students account for only 4.59%. Although 8.96% of the students are qualified, they are not enthusiastic in class and have a low attendance rate. Further efforts are needed. The above classification results truly reflect the relationship between effort and achievement and show that the combination model is more effective in classifying students' behavior. Teachers can supervise students according to the classification results, so as to improve learning achievement and teaching quality.

Student Dining Forecast. The consumption data of students dining in the canteen are extracted to construct a sample set, and the combination model is used for training. Then predict the number of diners in the test set of 2020-05-08 consumption data, and compare it with the actual number of diners. The canteen is open from 6 : 30 a.m. to 8 : 00 p.m., with an interval of 15 min. The prediction results of some time periods are shown in Figure 10.

From the prediction results, the combined model used has a good effect on the prediction of the number of diners, and the prediction results tend to be consistent with the actual situation. The canteen managers can prepare the amount of food according to the prediction results, indicating that the method used is effective and reliable.

Student Achievement Prediction. According to the established student behavior index system, the data of each index are extracted, the correlation between each behavior index and students' academic performance is established, and the student performance is tested and verified by using the constructed combined prediction model. The test data scale is 10030050010002000 and 10,000 students, respectively. Compare the test results with the actual students' scores, calculate the average relative error between the predicted results and the real value, and compare with the single prediction method. The results are shown in Figures 11 and 12.

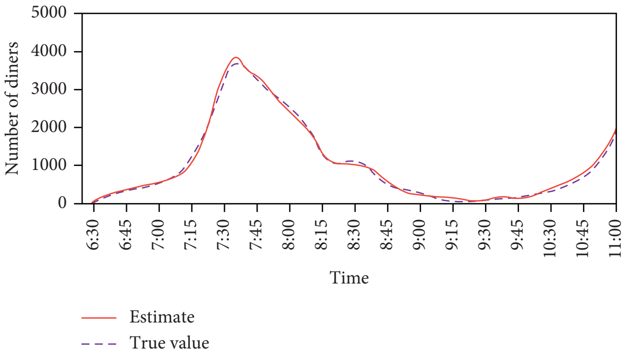

Figure 10: Prediction effect of the number of diners.

Table of analysis results: compared with a single algorithm, the student behavior prediction model based on the combination of multiple algorithms has higher prediction accuracy. The maximum absolute errors of the two groups of prediction results are 2.3% and 1.9%, respectively, which has a good prediction effect. With the increase of the number of samples, the average relative error changes little, and the prediction accuracy remains basically stable, indicating that the combined prediction model has good scalability. Among the students' behavior indicators, it has more or less impact on student's performance, but the relative error distribution

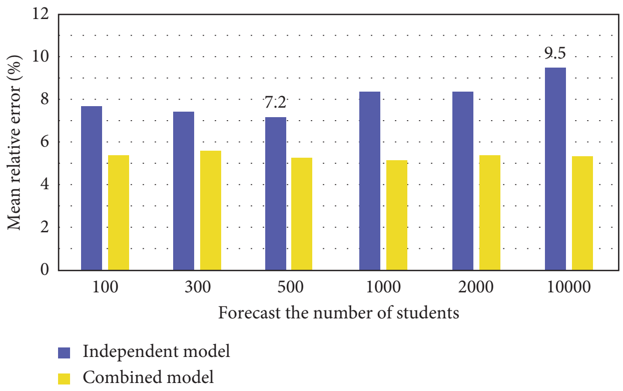

FIGURE 11: Average prediction error of different models.

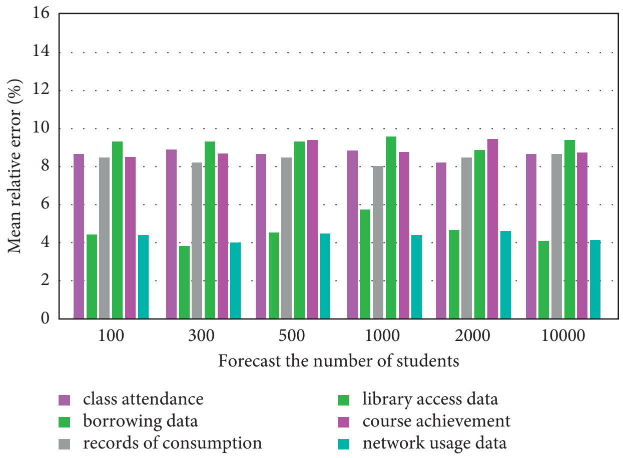

FIGURE 12: Average relative error of each index of student behavior.

on each student's behavior characteristic index is relatively uniform, which shows that the combined model has high prediction accuracy for the students behavior characteristics of each dimension and is suitable for the prediction of multidimensional students behavior.

6. Conclusion

With the development of the Internet, students' behavior data in school is constantly accumulated, which provides a data basis for students' behavior analysis and performance prediction. Aiming at the problems existing in the current student behavior analysis algorithm, this paper analyzes the student behavior indicators based on the data mining theory, so as to provide the basis for school management decision-making.

(1) Combining the GBDA algorithm, the ANN algorithm, and the K-means algorithm, a combined prediction model is constructed, and typical data sets are used for training simulation.

(2) Establish an analysis and prediction platform based on students' behavioral characteristics, combined with various school management systems, establish a student behavior evaluation system, and provide index data for model testing.

(3) Using the behavior data of students' consumption laws, living habits, learning, and Internet access, an example is verified in the proposed combination model.

(4) The results show that compared with a single algorithm, the combined model has the advantages of fast running speed, high prediction accuracy, and good scalability. It can effectively predict students' life law and academic performance according to their behavioral characteristics, and the average error is no more than 10%. The system platform is conducive to the school to timely grasp of the learning and life dynamics of students, so as to formulate targeted educational measures and provide a reference for school teaching management and teachers' teaching reform.

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 no conflicts of interest.

Acknowledgments

This work was supported by the Inner Mongolia Autonomous Region of Science and Technology innovation guiding project KCBJ201802.

References

[1] R. S. Baker, "Challenges for the future of educational data mining: the baker learning analytics prizes," *Journal of educational data mining*, vol. 11, no. 1, pp. 1–17, 2019.

[2] Y. Zhong, H. Yang, Y. Zhang, and P. Li, "Online random forests regression with memories," *Knowledge-Based Systems*, vol. 201-202, Article ID 106058, 2020.

[3] R. L. Rodrigues, J. L. C. Ramos, J. C. S. Silva, and A. S. Gomes, "Discovery engagement patterns MOOCs through cluster analysis," *IEEE Latin America Transactions*, vol. 14, no. 9, pp. 4129–4135, 2016.

[4] A. Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis, and H. Rangwala, "Predicting student performance using personalized analytics," *Computer*, vol. 49, no. 4, pp. 61–69, 2016.

[5] M. Sweeney, J. Lester, and H. Rangwala, "Next-term student grade prediction," in *Proceedings of the 2015 IEEE International Conference on Big Data (Big Data)*, pp. 970–975, Clara, CA, USA, October 2015.

[6] T. E. Johnson, E. Top, and E. Yukselturk, "Team shared mental model as a contributing factor to team performance and students' course satisfaction in blended courses," *Computers in Human Behavior*, vol. 27, no. 6, pp. 2330–2338, 2011.

[7] Z. Ren, H. Rangwala, and A. Johri, "Predicting performance on MOOC assessments using autorepression models," *Ar Xiv Preprint Ar Xiv*, vol. 1605, Article ID 02269, 2016.

[8] F. Berhanu and A. Abera, "Students' performance prediction based on their academic record," *International Journal of Computer Application*, vol. 131, no. 5, pp. 27–35, 2015.

[9] T. Hasbun, A. Araya, and J. Villalon, "Extracurricular activities as dropout prediction factors in higher education using decision trees," in *Proceedings of the IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, pp. 242–244, Austin, TX, USA, July 2016.

[10] J. Maillo, S. Ramírez, I. Triguero, and F. Herrera, "kNN-IS: an Iterative Spark-based design of the k-Nearest Neighbors classifier for big data," *Knowledge-Based Systems*, vol. 117, pp. 3–15, 2017.

[11] M. Wu, H. Zhao, X. Yan, Y. Guo, and K. Wang, "Student achievement analysis and prediction based on the whole learning process," in *Proceedings of the 2020 15th International Conference on Computer Science & Education*, pp. 123–128, Delft, The Netherlands, August 2020.

[12] M. Li, "A Study on the influence of non-intelligence factors on college students' English learning achievement based on C4.5 algorithm of decision tree," *Wireless Personal Communications*, vol. 2, no. 5, pp. 1213–1222, 2018.

[13] D. Pan, S. H. Wang, C. Y. Jin, and Y. Han, "Research on student achievement prediction based on BP neural network method," *Advances in artificial systems for medicine and education*, vol. 75, no. 5, pp. 15–23, 2021.

[14] A. Sun, T. Ji, J. Wang, and H. Liu, "Wearable mobile internet devices involved in big data solution for education," *International Journal of Embedded Systems*, vol. 8, no. 4, pp. 293–299, 2016.

[15] G. Chen, V. Kumar, and K. Yeonjoo, "Relations between student online learning behavior and academic achievement in higher education: a learning analytics approach," in *Proceedings of the International Conference on Smart Learning Environments 2014At*, Ting Kok, Hong Kong, China, December 2015.

[16] F. Deng and Z. Zhang, "Research on the construction of students' campus behavior analysis and warning management platform based on nig data," *China Educational Technology*, no. 11, pp. 60–64, 2017.

[17] N. I. Yi-kun and K. S. Liu, "Application of big data method in students' work-Taking Students' campus network behavior data mining as an example," *Studies in Ideological Education*, no. 2, pp. 152–156, 2021.

[18] J. O. Ramsay and B. W. Silverman, *Applied Functional Data X Analysis: Methods and Case Studies*, Springer, New York, NY, USA, 2002.