J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 3907--3925

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Construction of a Machine Learning-based Model for Analyzing English Teaching and Learner Behavior

Ranran Yan[1, ✉]

[1] *Jiangxi University of Technology, Nanchang, Jiangxi, 330098, China*

ABSTRACT

Behavior of different types of English learners tends to follow different patterns and characteristics, and the analysis of behavioral data is one of the directions for improving English learning and teaching. This study designs a set of behavioral analysis methods based on machine learning for English teaching and learners. The learners' behaviors are firstly operated with feature extraction and quantification, and the behavioral data are clustered by using the systematic clustering method (HCM) to improve the SOM model. 1DCNN is used to process the learning time-series data and enhance the data mining and performance prediction ability by BiLSTM and attention mechanism, respectively. This paper distinguishes five categories of English learners, such as excellent, diligent, average, procrastinating and negative, and filters out the factors that are highly correlated with English performance, such as the download of learning resources and the number of times of teaching viewing. Comparison experiments show that the ACC of this paper's achievement prediction model = 0.53, which is better than other comparison methods. Therefore, the idea of this paper based on machine learning methods to analyze the behavior of English teaching and learners has feasibility.

Keywords: 1DCNN, BiLSTM, attention mechanism, machine learning, HCM-SOM, behavior analysis

1. Introduction

Learned behavior, broadly speaking, refers to the behaviors that animals acquire through life experience and learning due to genetic reasons and also through the influence of the environment in which they live [1-2]. And in psychology, learning behavior is described as "the activities produced by an organism under the influence of various internal and external stimuli" [3]. At present, the study of learning behavior analysis is a hot topic in English teaching, especially in the context of the current teaching reform, the analysis of learning behavior is crucial for the reform of English digital teaching. However, learning behavior as an abstract concept, it is difficult for teachers to clearly use it as a criterion to guide students, and how to apply the results of the study of learning behavior in teaching has been the focus of research [4-6].

✉ Corresponding author.
E-mail address: sara_yan1118@163.com (R. Yan).
Received 20 February 2024; Revised 25 April 2024; Accepted 20 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-221

As a multidisciplinary cross-discipline, machine learning is mainly through the implementation of classifiers on existing big data, or algorithm training to achieve self-improvement, so that the decision function to find out, so as to predict the unknown samples more accurately and quickly, and slowly realize the goal of replacing manual work by it [7-10]. Machine learning techniques that are capable of storing and processing huge amounts of data are valuable resources and useful tools that mark another revolution in technology and can enhance the decision making, insight and discovery of the user [11-14]. With the help of machine learning technology, students' learning status and results in a certain period of time are faithfully reflected, thus helping English educators to implement more effective work under the premise of following the laws of student development and learning, implementing personalized parenting work, and achieving the goal of comprehensive and holistic development of students [15-19]. In addition, machine learning technology can also lay a solid foundation for the further improvement of the level of work of educational managers, and give students an important impact on the current and future learning and growth process [20-22].

In recent years, there have been scholars and related institutions to study students' learning behaviors and propose some effective mining and analysis methods. Literature [23] discusses the role of cluster analysis statistical methods in the field of education, takes the student samples administered by open-ended questionnaires as the research object, and proposes an unsupervised classification method of student behaviors, which classifies students with common characteristics into the same group, and provides support for the subsequent educational planning. Literature [24] uses cluster analysis technology to classify student behavior data in online English courses, and by extracting keywords of behavioral patterns in different learning scenarios and combining them with expert analysis, it provides insights into the classroom learning behavior management model in virtual learning environments. Literature [25] proposes an unsupervised integrated clustering framework to extract features from student behavior data from both statistical and entropy perspectives, and combines density-based spatial clustering with noise and k-means algorithms to summarize behavioral patterns from behavioral features, which provides data support for educational administrators to carry out services. Literature [26] shows that the educational process is an activity that matches actual student behaviors with classroom-designed behaviors, and proposes a profile-based clustering evolution analysis in which students can change their own learning behaviors by observing migration patterns between behavioral categories, which is of practical value in curriculum development as well as student assessment. Literature [27] input students' behavioral data into the adaptive K-means algorithm, which can obtain common features related to students' learning from the clustering results, providing a basis for the optimization of the teaching management mode and the improvement of students' learning ability. Literature [28] combines three-layer feed-forward neural network and proportional conjugate gradient algorithm to construct a student learning behavior feature extraction model, and conducts correlation analysis between students' online learning behavior features and course grades, which helps teachers to understand the real-time learning status of the students, and gives timely help to students with learning difficulties. Literature [29] pointed out that the learning style analysis method in the form of questionnaire survey is less accurate due to the lack of students' preference awareness, and proposed the online learning style recognition method of Felder and Silverman model, which was experimentally shown to have high performance in the automatic detection and prediction of students' learning styles. Literature [30] analyzed the application of tree-based models and different types of neural networks in the field of educational analytics, and the emergence of virtual classrooms provides massive student behavioral data for machine learning models, which helps to improve the quality of teaching and learning by constructing a performance model to find the key factors affecting students' learning performance. Literature [31] analyzes student behavior data by means of K-means clustering algorithm and Spearman correlation analysis technique, which helps to understand the situational state of students in the teaching process in order to develop an online learning system suitable for the characteristics of students, and achieves a better learning effect in teaching activities.

In order to model and analyze the behavior of English learners, this paper firstly extracts and quantitatively represents the features of English learning behavior. The HCM-SOM model is proposed, and the results obtained from SOM are merged using the systematic clustering method to improve the clustering quality of behavioral features. According to the clustering results, based on the machine learning method, a predictive analysis model of students' English performance is proposed on the basis of acquiring students' time-series behavioral data. One-dimensional CNN is used to improve the efficiency of time-series data processing, and then BiLSTM is used to enhance the backward and forward data correlation mining ability of the model, and the attention mechanism is used to screen out the effective and significant behavioral features to improve the accuracy of performance prediction. Examples of learner behavioral clustering are presented to test the behavioral features that affect English performance, and finally the performance prediction effect of this model is validated.

2. Feature Extraction and Correlation Analysis of English Learning Behavior

2.1. Extraction and Quantification of Learning Behavioral Features

2.1.1. Learning behavior feature extraction. According to the theory of behavioral science, learners use online English learning platforms for learning out of their own needs, and behavioral science focuses on observable and measurable episodic behavioral activities. In the process of online English learning, learners' operations can be observed and quantified, so the behavioral analysis of online English learning takes online operations as a breakthrough, and researches them by observing, describing and refining learners' learning operations.

Learners learning English online on the learning platform produce a series of different behaviors, which lead to different learning effects, in which the subject of the behavior, the object of the behavior, the behavioral environment, and the results of the behavior have universality, but the behavioral tools and means and the behavioral process adopted by each learner have uniqueness. Therefore, according to the behavioral tools and means and the behavioral process in the six elements, online English learning behaviors are divided into five categories: course access, video viewing, homework performance, usual performance and interaction. A total of 15 features were extracted based on theories such as behavioral science and behaviorist psychology, as shown in Table 1.

Table 1. Characteristics of English learning behavior

Feature category	Behavior characteristics
Course access	1 Check-in rate
	2 Active days
	3 Access duration
	4 Learning resources download
Video viewing	5 Video watching frequency
	6 Repeat watching frequency
	7 Video watching speed
	8 Viewing duration
Assignment performance	9 Assignment submission
	10 Assignment repetition
	11 Assignment submission moment
Interaction	12 Post depth
	13 The length of the post content
Daily performance	14 Assignment score
	15 Chapter test results

2.1.2. Quantifying Learning Behavioral Characteristics. The sign-in rate in the course access situation can reflect the most basic learning situation of the learners:

$$A_i = \frac{\sum_{j=1}^{k} m_{ij}}{\sum_{j=1}^{k} n_{ij}} (i = 1, 2, \ldots s; j = 1, 2, \ldots k) \tag{1}$$

Where A_i is the attendance rate of the ind student, m_{ij} is the number of times the ith student attended for the jth time, let n_{ij} be the number of times the ith student was not absent for the j th time, k is the total number of times the teacher took attendance, and s is the total number of students, and the learner who has reached 80% of the sign-in rate is recorded as 1, and those who have not reached it are recorded as 0.

The number of active days can reflect whether the learner is logging in to the course or not. In general, the number of days that the course lasts is the number of days that the learner should be online and active, which is set to a, and the plural of a_0 is taken as the threshold value. The actual number of days the learner is active is recorded as A, and if $a_0 \geq a$, the expression is:

$$\text{Active days} = \begin{cases} \text{High activity} & A > a_0 \\ \text{Medium Activity} & a < A \leq a_0 \\ \text{Low activity} & A < a \end{cases} \tag{2}$$

If $a_0 < a$, then the expression is:

$$\text{Active days} = \begin{cases} \text{High activity} & A \geq a \\ \text{Medium Activity} & a_0 \leq A < a \\ \text{Low activity} & A < a_0 \end{cases} \tag{3}$$

The length of visit reflects the extent to which the learner has studied the course and mastered the learning objectives. The median was used as the threshold, with visits greater than the median considered long, those equal to the median considered medium, and those less than the median considered short.

The number of downloads of learning resources reflects learners' consolidation of classroom content as well as their learning conscientiousness. The number of downloads is sorted from low to high, and the median is taken as the threshold, which indicates a medium level of the number of downloads of learning resources, and will not be affected by the extremely large and extremely small values. For the number of downloads, the median is more appropriate. A number of downloads greater than or equal to the median is labeled as “many downloads”, while a number less than the median is labeled as “few downloads”. Considering that most learners may not like to post back, the number of course updates is n as the standard value, and the number greater than or equal to n is regarded as a high level of attention to the forum and is labeled as 1, otherwise it is labeled as 0.

In the video viewing situation, the viewing frequency can reflect the learner's learning commitment, in general, the total number of videos is taken as the minimum number of times the learner should watch, set to t, take all the learners to watch the frequency of the plural t_0, if there are more than one plural, the median is used as a candidate for the threshold is set to $t0'$. The learner's actual viewing frequency is recorded as T, if $t_0 \geq t$, the expression is:

$$\text{Viewing frequency} = \begin{cases} \text{Many} & T > t_0 \\ \text{Normal} & t < T \leq t_0 \\ \text{Less} & T < t \end{cases} \tag{4}$$

If $t_0 < t$, the expression is:

$$\text{Viewing frequency} = \begin{cases} \text{Many} & T \geq t \\ \text{Normal} & t_0 \leq T < t \\ \text{Less} & T < t_0 \end{cases} \tag{5}$$

The threshold for repeat viewing frequency is the plural. Taking the multitude can exclude the number of repetitions caused by learners clicking on the video randomly or by mistake, and better reflect the concentrated trend of learners' repeated viewing. Repeat frequency greater than the threshold is recognized as multiple repetitions, equal to the threshold is recognized as normal repetitions, and less than the threshold is recognized as fewer repetitions.

Watching time can reflect the learner's mastery of the course, set the total length of the video as l, that is, the length of time the learner should watch, take the plural l_0, the actual length of time the learner watches is L, if $l_0 \geq l$, the expression is:

$$\text{Viewing duration} = \begin{cases} \text{Long} & L > l_0 \\ \text{Normal} & l \leq L \leq l_0 \\ \text{Short} & L < l \end{cases} \tag{6}$$

If $l_0 < l$, the expression is:

$$\text{Viewing duration} = \begin{cases} \text{Long} & L \geq l_0 \\ \text{Normal} & l_0 < L \leq l \\ \text{Short} & L < l_0 \end{cases} \tag{7}$$

In assignment performance, the threshold of whether the assignment is submitted or not is recorded as m, and m indicates the number of assignments assigned by the teacher, i.e., the number of assignments that the learner should complete. The number of times a learner submits an assignment that reaches the threshold, i.e., every assignment is completed and submitted, will be recorded as excellent, those that reach 70% of the threshold will be recorded as average, and the rest will be recorded as failing.

Assignments can be submitted repeatedly on the online learning platform, and the reasons for repeated submissions include errors in assignments that need to be revised and submitted again, improvement of previous assignments, errors in submission of assignments, and submission of assignments in response to a new request from the instructor. There are negative factors such as learners' sloppiness, lack of seriousness, plagiarized assignments, etc., and it is difficult to distinguish the specific reasons, so the feature threshold of whether the assignment is repeated is set to 1, and if the learner has repeatedly submitted the assignment, it will be recorded as 1, otherwise it will be recorded as 0.

Assignment submission time can reflect the learner's self-efficacy, set the assignment submission period for d day, take the middle value of d days d' (such as $d=10$, then $d'=5$, rounded up when there is a remainder), the assignment submission time ascending order, the learner's assignment submission time is recorded as Job, due to the online learning platform learner's learning time is different, so take the median of Job_{sub}, if $Job_{sub} < d'$, the expression is:

$$\text{Job submit time} = \begin{cases} \text{Positivity} & Job \le Job_{sub} \\ \text{Normal} & Job_{sub} < Job \le d \\ \text{Not active} & Job > d' \end{cases} \tag{8}$$

If $Jobb_{sub} > d'$, then the expression is:

$$\text{Job submit time} = \begin{cases} \text{Positivity} & Job \le d' \\ \text{Normal} & d' < Job \le Job \\ \text{Not active} & Job > Job_{sub} \end{cases} \tag{9}$$

In usual performance, chapter test grades are scored out of 100, with a score greater than or equal to 90 being excellent, average in the [80, 89] range, passing in the [60, 79] range, and failing in the less than 60 range. Assuming that there is a total of x chapter chapter quiz, the plurality of each chapter quiz grade is taken to determine the overall chapter quiz grade.

Assignment results can indicate the learner's attitude toward the course, set the total number of assignments for v, the teacher of the learner's homework submitted for assessment, the assessment grade is divided into “good and bad”, if the percentage system, is greater than or equal to 80 for excellent, [60, 79] interval for good, less than 60 for poor, take the plural of all the assignment grades The overall assignment grade for the learner is taken, and if the grades are exactly evenly divided or there are multiple plurals, the assignment grade is recorded as good.

In the interactive case, the total post content length threshold is the median of the post content contribution rate, and the post content contribution rate formula is as follows:

$$C_{ij} = \frac{f_{ij}}{\sum_{j=1}^{n} f_{ij}} (i = 1, 2, \ldots, m; j = 1, 2, \ldots n) \tag{10}$$

where C_{ij} is the contribution rate of the words of the j rd post of the i nd student to the total individual posts, and f_{ij} is the number of words of the j th post of the ith student.

The post depth threshold is the median post count contribution rate, and the post count contribution rate is calculated as follows:

$$C_i = \frac{\sum_{j=1}^{n} f_{ij}}{\sum_{i=1}^{m} (\sum_{j=1}^{m} f_{ij})} (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n) \tag{11}$$

where C_i is the contribution of the number of posts or replies of the ind student to the total number of posts or replies of all students, f_{ij} is the number of jth posts of the ith student, m is the number of students, and n is the number of posts.

2.2. Behavioral features HCM-SOM clustering

The SOM model used in this paper, i.e., self-organizing mapping model, is a kind of neural network model, which can be used as data dimensionality reduction and clustering [32]. The data-processed student behavioral descriptors are input into the input layer of the SOM model, and the steps of the algorithm are shown below.

Step 1 Initialize the network. Initialize the network weights: the number of neurons in the input layer, the number of neurons in the output layer n,m and the maximum number of cycles T.

Step 2 Input training samples and normalize vectors. X is the input vector, which is calculated as follows:

$$\hat{X} = \frac{X}{\| X \|} = \left[\frac{x_1}{\sqrt{\sum_{j=1}^{n} x_j^2}}, \cdots, \frac{x_n}{\sqrt{\sum_{j=1}^{n} x_j^2}} \right]^T \tag{12}$$

Calculate the distance d_j from the competing layer neuron j:

$$d_j = \sqrt{\sum_{i=1}^{m} (x_i - w_{ij})^2} \tag{13}$$

Step 3 Calculate the distance between the input vectors and the neuron to find the winning neuron c. The distance is smaller when the area of the points between the vectors is maximized, as shown in the following formula:

$$\| \hat{X} - \hat{W}_j \| = \min_{j \in \{1,2,\ldots,m\}} \{ \| \hat{X} - \hat{W}_J \| \} \tag{14}$$

$$\| \hat{X} - \hat{W}_J \| = \sqrt{2(1 - W_J^T \hat{X})} \tag{15}$$

$$\hat{W}_{J^*}^T \hat{X} = \max_{j \in \{1,2,\ldots,m\}} (\hat{W}_J^T \hat{X}) \tag{16}$$

Step 4 The network outputs and adjusts the weight coefficients of the winning neuron c with its domain, η for the learning rate, t for the iteration period, and continues the training process if η does not decay to zero:

$$W_{ij}(t+1) = W_{ij} t + \eta (X_i - W_{ij}) \tag{17}$$

Systematic clustering method (HCM), is a kind of hierarchical clustering method, which is used in this paper to optimize the SOM model, and the main idea is to use systematic clustering to merge classes of the results obtained from the SOM, and to validate the clustering results so as to improve the quality of the clustering [33]. The specific steps of the process of HCM-SOM are shown in Fig. 1.

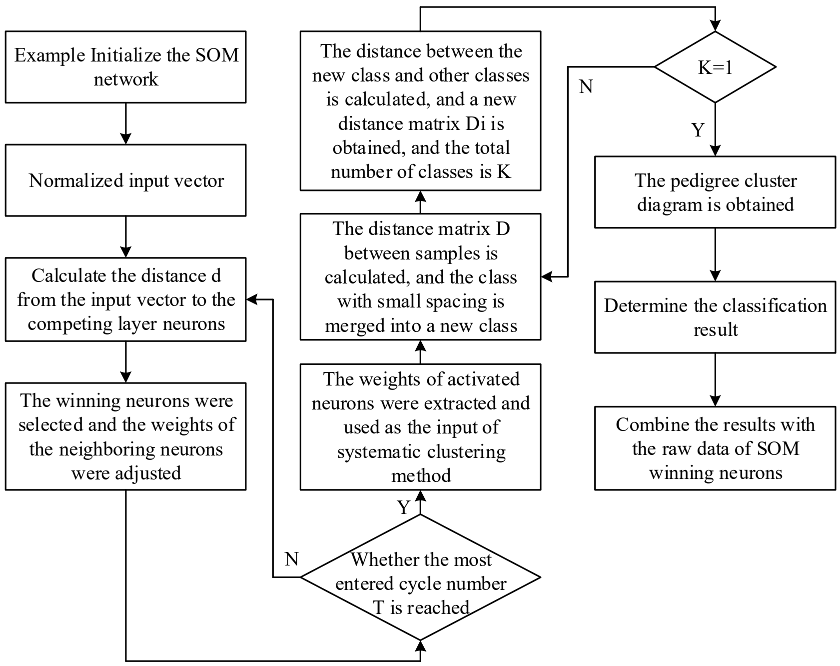

Fig. 1. Flowchart of SOM improved by system clustering method

2.3. *Correlation analysis*

Correlation analysis is a method used to analyze two or more correlated variables to determine the strength of the relationship between the variables and to determine the strength of the correlation between two variable factors. The Pearson correlation test in correlation analysis used in this study is explained.

Let the distribution function of the binary aggregate (X,Y) be the variances of $F(X,Y)$ are $var(X)$ and $var(Y)$, respectively, the overall covariance is $cov(X,Y)$, and the overall correlation coefficient is defined as:

$$\rho_{xx} = \frac{cov(XY)}{\sqrt{var(X)} \cdot \sqrt{var(Y)}} \tag{18}$$

Let $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ be an independent sample selected from each of the binary aggregates (X,Y). The correlation coefficient of the samples can be calculated:

$$r_{XY} = \frac{S_{XY}^2}{\sqrt{S_X^2} \cdot \sqrt{S_Y^2}} \tag{19}$$

where S_X^2 and S_Y^2 are the variances of samples X and Y, respectively, and S_{XY}^2 is the covariance of sample XY. In the usual case, r_{XY} computed from the samples is not zero, even when the random variables X,Y are independent. Therefore when $\rho_{XY} = 0$, it is not practically meaningful to use r_{XY} to measure the correlation of X,Y, so hypothesis testing is needed:

$$\begin{gathered} H_0 : \rho_{XY} = 0 \\ H_1 : \rho_{XY} \neq 0 \end{gathered} \tag{20}$$

It can be shown that the statistic when (X,Y) is a binary normal totality and when H_0 is true:

$$t = \frac{r_{XY}\sqrt{n-2}}{1-r_{XY}^2} \tag{21}$$

obeys a t-distribution with degree of freedom $n-2$.

The correlation of the data X, Y can be tested by utilizing the property that the statistic t obeys a t distribution with $n-2$ degrees of freedom. Since the correlation coefficient r_{XY} is known as the Pearson correlation coefficient, the test is also known as the Pearson correlation test.

3. English performance analysis model based on CNN-BiLSTM-Attention

According to the different types of learning behaviors, they were counted separately, thus obtaining the overall student-learning behavior data. Separately, the student's behavioral data was partitioned into behavioral data according to weekly time units. From this, the student's time-series data was obtained and the model labeled values were academic final grade y, $y \in \{0,1,2,3\}$, 0 means that this student dropped the course, 1 means that this student did not drop but did not pass the course in the end, 2 means that this student passed the course, and 3 means that this student ended up with a good grade.

3.1. Model for predictive analysis of English performance

The purpose of the machine learning based learning method investigated in this paper is to be able to be by the students' learning behaviors and to obtain potential behavioral features that can predict students' academic performance. Different from the existing methods for predicting students' performance, this study argues that we get the cumulative click data in weekly units, thus obtaining the students' time-series behavioral data, and proposes a new model for predicting students' performance CNN-BiLSTM-Attention The model consists of three important components: 1D-CNN, BiLSTM, and Attention. the existing studies have shown that 1D-CNN has strong ability in feature extraction in time-series data processing, and the introduction of CNN can enhance the prediction ability of the whole model.BiLSTM has the function of forward and backward data relevance mining.BiLSTM network is introduced to mine the temporal and periodic features embedded in the data. It is based on the principle of giving different levels of attention to different features, i.e., using different weights to increase or decrease the importance of features to the overall data. The workflow of the model is shown in Figure 2.

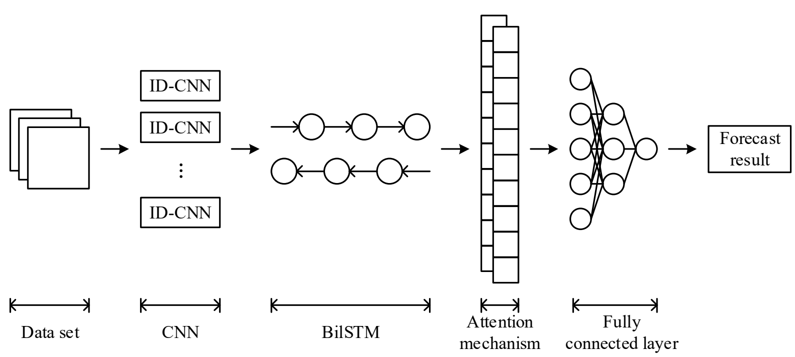

Fig. 2. Model frame

In this model. First, a lightweight 1D-CNN structure is used, which can significantly reduce the network complexity. Each 1D-CNN network processes only one week of temporal behavioral information. This approach enhances the predictive power of the model while ensuring a concise network structure. The study categorized the overall behavioral data of the students on a weekly basis into 32 weeks of behavioral data. Then the 32 weeks of behavioral data were fed into 32 1D-CNN networks in parallel, from which high-dimensional features of the behavioral data were obtained, and then 1D-CNN outputs were accessed in a BiLSTM network, which obtained weekly time-series behavioral correlation features, and finally, the BiLSTM network output attention mechanism as well as the fully connected layer were processed to obtain the final results.

3.2. Convolutional neural network feature extraction

Convolutional Neural Networks (CNNs) are a special class of artificial neural networks that are different from other models of neural networks in that they are mainly characterized by convolutional arithmetic operations. So CNNs have outstanding performance in many areas of applications, especially image related tasks such as, image classification etc., semantic segmentation of images, image retrieval etc., object detection and other computer vision problems. CNNs have a pivotal role in the field of computer vision and CNNs can also be applied to time series. The difference is that the convolutional kernel applied to an image is two dimensional and the convolutional kernel applied to a time series has one dimensional properties, i.e., a one-dimensional convolutional neural network, referred to as a 1D-CNN [34]. The CNN performs convolutional operations and pooling operations on the input data to capture implicit features. Then the extracted features are integrated and sent to the fully connected layer. Finally a partial activation function is used to introduce nonlinear output in the neuron.

Convolutional layers are an important part of CNN. Each convolutional layer has a number of convolutional kernels that are convolved to the input information to capture the hidden features and form the feature map. The feature map is used by a nonlinear activation function to produce an output at the convolutional layer. The convolutional layer can be expressed as:

$$m_i = f(w_i * x_i + b_i) \tag{22}$$

Inside the formula w_i denotes a weight matrix, $*$ denotes dot product operation, b_i denotes a bias vector, f denotes an activation function, there is more than one CNN activation function, generally ReLU function is used as an activation function, m_i is a feature map. Mathematically ReLU is defined as:

$$m_i = f(k_i) = \max(0, k_i) \tag{23}$$

Inside the formula k_i is the element of the feature map in the convolution operation.

Pooling operations have the effect of reducing the dimensionality of the feature map, avoiding overfitting, and so on. Among them, maximum pooling is the most common. By calculating the maximum value in a prescribed region on the feature map based on the equation:

$$\mu(m_i, m_{i-1}) = \max(m_i, m_{i-1}) \tag{24}$$

$$c_i = \mu(m_i, m_{i-1}) + \eta_i \tag{25}$$

Inside the formula μ is the sampling function in the pooling layer. η_i is the offset, and c_i is the pooling layer output. Finally the features obtained by convolution and pooling operations are transferred to the fully connected layer and then the final output vector is computed.

$$y_i = f(v_i c_i + \lambda_i) \tag{26}$$

In the formula y_i denotes the final output vector, λ_i is the offset and v_i is the weight matrix.

3.3. BiLSTM network

LSTM is a kind of RNN.LSTM is well suited for temporal data modeling due to its design characteristics, such as BiLSTM, which consists of a combination of forward LSTM and backward LSTM.Each LSTM cell has a total of three gating structures, namely, forgetting gate and output gate.

Each LSTM cell has a forgetting gate, input gate and output gate a total of 3 gating structures. each gating structure of the LSTM cell, the output of the hidden layer, the transfer process of the cell state, for example, formula (27) to formula (32) [35]:

Inside the forgetting gate Eq:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{27}$$

Enter the formula inside the door:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{28}$$

$$\tilde{c}_t = \tanh\,(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{29}$$

$$C_t = f_t^* C_{t-1} + i_t * \tilde{c}_t \tag{30}$$

Output gate inside formula:

$$o_t = \sigma(W_0[h_{t-1}, x_t] + b_0) \tag{31}$$

$$h_t = o_t * \tanh(C_t) \tag{32}$$

where x_t denotes the input timing data. f_t, i_t, o_t denotes the outputs of the forgetting, input and output gates, respectively. W_f, W_i, W_0 is the weight matrix of the three gates, and b_f, b_i, b_c is the corresponding three bias cells. σ is the sigmoid function. tanh is the hyperbolic tangent function. * denotes the inner product operation. $\tilde{c}_t$ denotes the candidate vectors created through the tanh layer, w_c, b_c corresponds to the weight matrix and bias cells of that layer, respectively. C_t is the cell state. h_t denotes the hidden state.

However, LSTM only feeds the information in the forward sequence into the neural network for prediction, and it is difficult to perceive the backward content of the data during the training of the model. The emergence of Bidirectional Long Short-Term Memory Network (BiLSTM) technique solves the problem of insufficient attention to backward information. By bi-directional, it means, there are forward LSTM units, backward LSTM modules set up within the said BiLSTM, each LSTM unit conforms to the construction of the LSTM as described above and the two units of forward and backward shift are independent of each other.

3.4. Attention mechanisms

Attention mechanism is like the visual processing module of human eye, human eye will notice the more interested images in the picture when viewing the picture, attention mechanism emphasizes the significance of these features on the prediction result by giving higher weight value to the important features. Attention mechanism has good predictive effect on machine learning and better processing for text categorization and language recognition. The Attention Mechanism module encodes a small number of features with good results, and when predicting grades in this study, the Attention Mechanism module tightens the core features and gives a greater proportion of important features to improve the prediction accuracy.

The Attention Mechanism utilizes the input feature values to help the model filter out valid and significant features. To filter out important features, the attention module combines the attention score module with the attention focus module to filter out significant features. For the hidden layer state of the neural network, different weight vectors are assigned to the features based on their influence on the final prediction results. In this study, important student behavioral features were given more weight based on the effect of different student time-series behavioral data on performance prediction.

The input to the Attention layer is the hidden layer state h_t computed by the BiLSTM layer, m_t represents the value of the probability distribution of attention determined by h_t, a_t is the value of the attention weights, and the output of the Attention layer at the moment of t is s_t, as specified in Eq:

$$m_t = s_{t-1} \times h_t^{'} \tag{33}$$

$$a_t = \frac{exp(m_t)}{\Sigma_{j=1}^{t} exp(m_j)} \tag{34}$$

$$s_t = \sum_{i=1}^{t} a_i h_i^{'} \tag{35}$$

The fully connected layer is chosen as the output layer and sigmoid is chosen as the activation function to output the predicted value of the score at the $t+1$ st moment, denoted as $y^{'}$, which can be expressed as:

$$y' = sigmoid(\omega_{st} + b) \tag{36}$$

Where: ω is the weight matrix. b is the bias.

4. Analysis and Prediction Results of English Learning Behavior

4.1. Cluster Analysis of Learning Behavioral Characteristics

In this study, the behavioral data of 100 students from a university's business school were collected from March to June 2023, totaling more than 500,000 records of online English learning behaviors on the SuperStar Learning Channel platform, and the students' personal information was encrypted.

4.1.1. Descriptive statistics of clustering results. In order to mine and analyze students' online English learning behavioral features, this study uses the proposed HCM-SOM clustering algorithm to cluster and analyze 15 behavioral features after the above feature selection. Five clusters are obtained, and the results of some behavioral features are shown in Table 2, and Figure 3 shows the statistics of clustering results.

The clustering results of 100 students are in five categories, and the number of cluster 5 has the highest percentage (42%). Observing the data characteristics of different clusters, it is found that Cluster 4 students have the most active days and Cluster 2 students have the least active days, and the more active days indicate that students are more active in online English learning. Cluster 5 students' behavioral characteristic values of the number of instructional video viewings, video playback multiplier speed, and video viewing duration are better than other students, indicating that Cluster 5 students are more active in completing the video task and use their class time to complete the video task. On the other hand, all behavioral eigenvalues of Cluster 2 students, except for the eigenvalue of completing daily performance, were lower than those of other students, indicating that Cluster 2 students were less active in all online English learning behaviors, and needed to be improved through personalized advice to ultimately improve their learning outcomes.

Table 2. Partial behavior characteristics result

Grade	Sign-in rate	Active days	Video watching frequency	Assignment submission moment	Chapter test results	...
B	1	High	Many	Positive	Excellent	...
A	1	High	Many	Positive	Excellent	...
C	1	Medium	Normal	Average	Normal	...
B	1	Medium	Many	Positive	Excellent	...
A	1	High	Many	Positive	Excellent	...
D	1	Low	Few	Passive	Failed	...
D	0	Low	Few	Passive	Passed	...
...	...	...	...	...	...	...

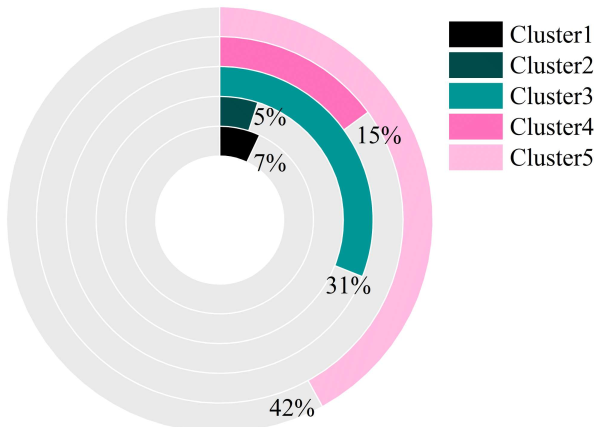

Fig. 3. Clustering results

4.1.2. Visualization of Learner Behavioral Characteristics. In order to better analyze the learning effects of students in different clusters, this paper presents the learning effects, and Figure 4 shows the analysis of the learning effects of each cluster. It can be seen that the upper limit, average value and lower limit of the learning effect of each cluster are decreasing in the order of Cluster 4, Cluster 5, Cluster 3, Cluster 1 and Cluster 2. Therefore, the online English learning groups in this paper are categorized into the following five types: excellent students (Cluster 4), diligent students (Cluster 5), average students (Cluster 3), procrastinating students (Cluster 1), and negative students (Cluster 2).

1) The excellent type of students accounted for 15% of the total number of students and their learning outcomes were higher than the other types. The values of several behavioral characteristics of this type of students are higher than those of other types, which shows that the online English learning behaviors of excellent type students are positive and active. Excellent students have good self-discipline, can actively and positively complete various course learning tasks, and obtain excellent learning results through their own efforts in learning. For the excellent type students, the network learning resources should be enriched, the depth of learning should be broadened, and the students should be guided to a wider and deeper level of development.

2) The learning effect of diligent students is only lower than that of excellent students, and several behavioral characteristics are also better. It can be seen that diligent students can make full use of the learning resources on the Internet and actively participate in online English learning, so as to obtain better learning effect. For diligent students, they should strengthen pre-study before class and review after class, make full use of the time to study in class to increase their knowledge reserve, and actively participate in teacher-student interactions to accumulate learning experience so as to improve their academic performance.

3) Students of the average type have moderate learning effects and perform well in many online English learning behaviors. The behavioral characteristics of this type of students, such as participation in discussions, number of active days, and video completion rate, are lower than those of diligent students, which indicates that average students have poor learning initiative, are not strongly motivated, and lack communication with teachers and classmates. For general type students, they should improve their independent learning by actively participating in group collaborative learning, teacher supervision and homework counseling to enhance their interest in learning, improve their learning ability, and then improve their learning effect.

4) The learning effect of procrastinating students is slightly higher than that of negative students, and the value of learning behavior characteristics is in the middle. For procrastinating students, they should be reminded in time to complete the course learning tasks, rationalize the study time, and improve the learning enthusiasm and initiative, enhance the independent learning ability, and strengthen the learning of basic knowledge related to the course. Teachers should encourage this type of students to actively participate in the study of display courses, avoid delay in completing learning tasks as much as possible, and improve their academic performance to some extent.

5) The number of negative students is small, only 5% of students. Such students have poor learning results, rank last in quiz scores, hardly participate in discussions, and have far fewer active days and check-in rates than other students. For negative students, they can improve their learning motivation through self-motivation, make personalized learning plans, optimize the learning level of advantageous subjects, improve the learning ability of disadvantageous subjects, and strengthen communication with teachers and students, improve learning initiative, complete learning tasks by heart, and gradually improve their learning performance and complete their studies successfully.

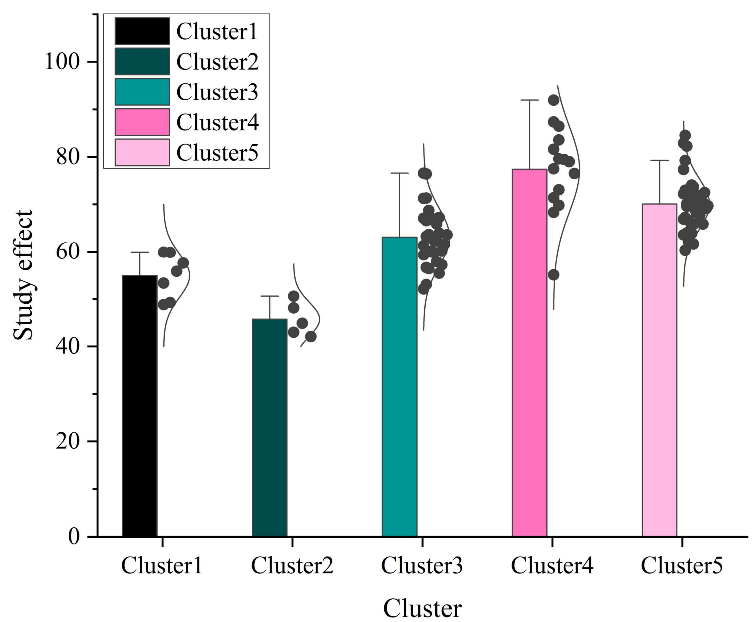

Fig. 4. Analysis of the learning effects of various clusters

4.2. English performance correlation analysis

4.2.1. Analysis of Factors Affecting English Achievement. English learning achievement is an important content of teaching concern, in order to more intuitively analyze the factors affecting achievement, this paper selects a few more typical characteristic attributes as analytical references: cumulative video viewing time, cumulative number of video viewing times, video viewing progress, and number of course interactions. In order to explore the relationship between the above four variables and grades, scatter plots of each variable and grades are drawn, and the relationship between the four variables and grades are shown in Fig. 5(a)~(d), respectively.

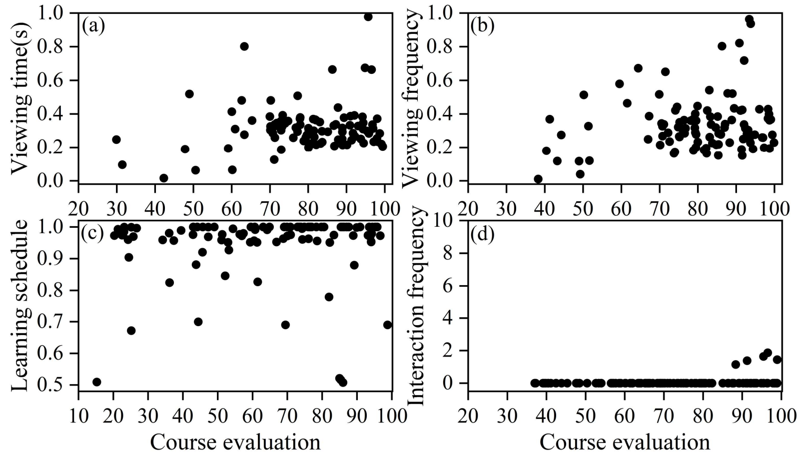

Fig. 5. Scatter diagram of variable and grade

The descriptive statistics of the four characteristics of cumulative video viewing hours, cumulative video viewing times, video viewing progress and number of course interactions are shown in Table 3. It can be seen that the distribution of students' video viewing hours is more centralized, with some fluctuations of a smaller magnitude. The number of video viewing times fluctuates to a greater extent, with a distribution from 100 to 550. Learning progress of a majority of students have completed the progress requirements, there are a small number of students did not finish watching the course, which is distributed more dispersed. The number of course interactions was mostly zero except for a few values, which means that student activity in the forums during the course was low, and most students did not ask or answer questions in the forums.

Table 3. Descriptive statistics of various indicators

	N	Minimum	Maximum	Mean	Se.	Sd.
Interaction frequency	100	0.0	10.0	0.062	0.0381	0.6212
Video viewing schedule	100	0.48	1.00	0.9846	0.00389	0.06798
Video viewing duration	100	38358	182721	80358.37	645.397	11091.736
Video viewing frequency	100	102	548	237.39	3.352	56.338
Course evaluation	100	32.2548	100	86.338	0.7388	12.4821

4.2.2. Visualization of the correlation coefficient matrix. The degree of linear correlation between each learning behavior characteristic and achievement is analyzed through correlation coefficient matrix analysis and heat map visualization, and the heat map of correlation analysis is shown in Figure 6.

From the figure, it is found that check-in rate, active days, login duration, learning resources download, and viewing times are highly correlated with English grades. Video multiplier, video viewing duration, homework submission, homework repetition, homework submission time, chapter test scores, post length and depth are moderately correlated with English achievement. Whereas, the number of video repetitions and posting length had low correlation with English achievement.

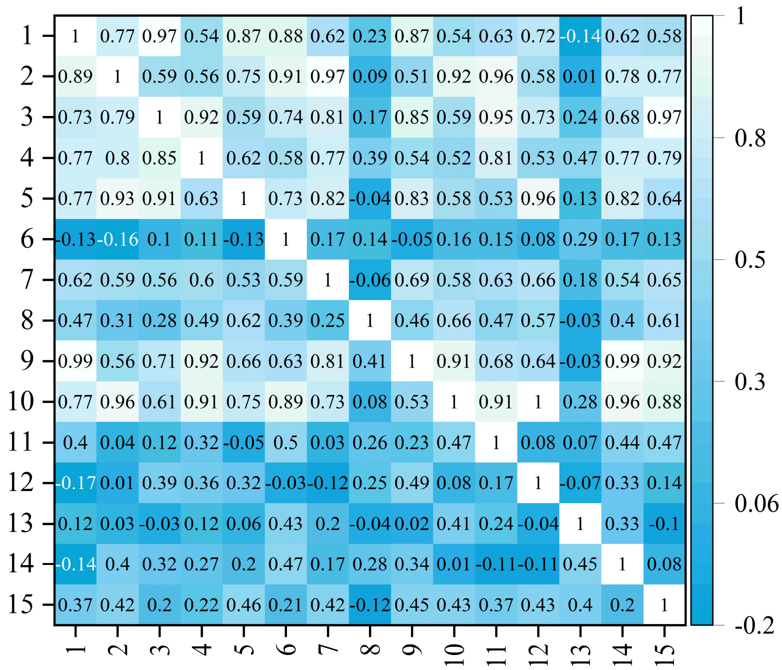

Fig. 6. Correlation analysis heat diagram

4.3. Experimental analysis of machine learning models

Table 4 shows the experimental results of machine learning algorithms. By studying the prediction of users' online English learning effect, this paper analyzes the prediction effect of the multi-classification model with the value of [0,10] in the machine learning model, and compares the prediction effect of the CNN-BiLSTM-Attention machine learning model, which has a good effect in the performance and prediction of online English learning effect. In the model construction of machine learning, CNN-BiLSTM-Attention can effectively classify the learning effect of users according to the relevant features of online English learning, so as to improve the prediction effect of users' online English learning, through the comparison of the effect in the model found that the CNN-BiLSTM-Attention has been improved in terms of the accuracy rate to reach 0.53. At the same time, in the calculation of Deviate, CNN-BiLSTM-Attention also shows better model effect, CNN-BiLSTM-Attention effectiveness prediction discrete, indicating that in the prediction process of the model, the online English learning effectiveness prediction data has better convergence in data deviation, so CNN-BiLSTM-Attention model is suitable to be used for the analysis of online English learning behaviors. English learning behavior analysis.

Table 4. The experimental results of the deep learning model algorithm

	ACC	Pre	Recall	F1	MCC	Deviate	MeanAE	EVS
RNN	0.050	0.013	0.114	0.011	-0.103	13.959	3.404	-0.532
GRU	0.286	0.046	0.167	0.086	0.000	5.539	1.712	-0.015
LSTM	0.075	0.007	0.160	0.022	0.004	2.723	1.498	0.014
FCN	0.119	0.015	0.171	0.038	-0.003	4.322	1.826	-0.016
ResNet	0.260	0.095	0.219	0.110	0.099	3.632	1.498	-0.248
ResCNN	0.068	0.042	0.187	0.029	0.003	2.803	1.519	-0.024
TCN	0.079	0.020	0.171	0.024	0.003	2.740	1.504	-0.002
XceptionTime	0.140	0.197	0.167	0.042	0.036	4.325	1.827	0.016
OS-CNN	0.055	0.010	0.167	0.006	-0.006	13.456	3.296	-0.006
LSTM-FCN	0.295	0.039	0.166	0.089	-0.004	5.546	1.708	0.002
CNN-BiLSTM-Attention	0.530	0.132	0.170	0.150	0.166	2.762	0.987	-0.120

5. Conclusion

In this paper, English learning behavior features are extracted, quantified and clustered to operate, and an English performance prediction model is proposed to predict English performance based on learning behaviors. Using an online learning platform to collect English learning behavior data, 15 behavioral features were selected for clustering to distinguish students of excellent, diligent, average, procrastinating and negative types, and to propose corresponding learning suggestions. The Pearson correlation coefficients (PCCs) between the behavioral features were calculated, and it was found that the number of active days, login hours, learning resources downloads, and the number of viewings were highly correlated with English performance. The ACC value of the achievement prediction model in this paper is 0.53, which is higher than that of other models and shows good prediction performance. Overall, this study completes the analysis of learner behavior based on machine learning and lays the foundation for improving English teaching.

References

[1] Latif, G., Abdelhamid, S. E., Fawagreh, K. S., Brahim, G. B., & Alghazo, R. (2023). machine learning in higher education: students' performance assessment considering online activity logs. IEEE access, 11, 69586-69600.

[2] Paquette, L., & Baker, R. S. (2019). Comparing machine learning to knowledge engineering for student behavior modeling: a case study in gaming the system. Interactive Learning Environments, 27(5-6), 585-597.

[3] Li, M. (2021). Recognition of Psychological Characteristics of Students' Behavior Based on Improved Machine Learning. Journal of Sensors, 2021(1), 8135942.

[4] Bundasak, S. (2020). Student behavior analysis affecting learning achievement of information technology and computer science students. International Journal of Machine Learning and Computing, 10(2).

[5] Memon, M. M., Hashmani, M. A., Jameel, S. M., Junejo, S., & Raza, K. (2020). Analysis of student procrastinatory behavior in virtual learning environments using machine learning. Journal of Hunan University Natural Sciences, 47(10).

[6] Burman, I., Som, S., & Hossain, S. A. (2021). Model for Analyzing Psychological Parameters Recommending Student Learning Behaviour using Machine Learning. Int. J. Com. Dig. Sys, 10(1).

[7] Karthikeyan, R., Satheesbabu, S., & Gokulakrishnan, P. (2021). Machine learning based student performance analysis system. Information Technology in Industry, 9(1), 1174-1181.

[8] Adane, M. D., Deku, J. K., & Asare, E. K. (2023). Performance Analysis of Machine Learning Algorithms in Prediction of Student Academic Performance. Journal of Advances in Mathematics and Computer Science, 38(5), 74-86.

[9] Orji, F., & Vassileva, J. (2020, September). Using machine learning to explore the relation between student engagement and student performance. In 2020 24th International Conference Information Visualisation (IV) (pp. 480-485). IEEE.

[10] Su, Y. (2021, September). Student Behavior Based on Information Technology and Machine Learning. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 2619-2623).

[11] Luo, Y., Han, X., & Zhang, C. (2024). Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses. Asia Pacific Education Review, 25(2), 267-285.

[12] Ilic, M. P., Paun, D., Popovic Ševic, N., Hadžic, A., & Jianu, A. (2021). Needs and Performance Analysis for

Changes in Higher Education and Implementation of Artificial Intelligence, Machine Learning, and Extended Reality. Education Sciences, 11, 568.

[13] Masood, M. F., Khan, A., Hussain, F., Shaukat, A., Zeb, B., & Ullah, R. M. K. (2019, November). Towards the selection of best machine learning model for student performance analysis and prediction. In 2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI) (pp. 12-17). IEEE.

[14] Raihan, M., Islam, M. T., Ghosh, P., Angon, J. H., Hassan, M. M., & Farzana, F. (2020, January). A machine learning approach to identify the correlation and association among the students' educational behavior. In Proceedings of the International Conference on Computing Advancements (pp. 1-6).

[15] Kumar, M., Singh, A. J., & Handa, D. (2019). Performance analysis of students using machine learning & data mining approach. International Journal of Engineering and Advanced Technology (IJEAT), 8(3).

[16] Shi, Y., Sun, F., Zuo, H., & Peng, F. (2023). Analysis of learning behavior characteristics and prediction of learning effect for improving college students' information literacy based on machine learning. IEEE Access, 11, 50447-50461.

[17] Wang, C., Chen, J., Xie, Z., & Zou, J. (2024, July). Research on education big data for student's academic performance analysis based on machine learning. In Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Education Digitalization and Computer Science (pp. 223-227).

[18] Palkhiwala, S., Shah, M., & Shah, M. (2022). Analysis of Machine learning algorithms for predicting a student's grade. Journal of Data, Information and Management, 4(3), 329-341.

[19] Ciolacu, M., Tehrani, A. F., Beer, R., & Popp, H. (2017, October). Education 4.0—Fostering student's performance with machine learning methods. In 2017 IEEE 23rd international symposium for design and technology in electronic packaging (SIITME) (pp. 438-443). IEEE.

[20] Ning, S. (2022). Analysis of college students' behavior based on machine learning and big data technology. In The 2021 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy: SPIoT-2021 Volume 1 (pp. 145-151). Springer International Publishing.

[21] Toti, D., Capuano, N., Campos, F., Dantas, M., Neves, F., & Caballé, S. (2021). Detection of student engagement in e-learning systems based on semantic analysis and machine learning. In Advances on P2P, Parallel, Grid, Cloud and Internet Computing: Proceedings of the 15th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2020) 15 (pp. 211-223). Springer International Publishing.

[22] Goldberg, P., Sümer, Ö., Stürmer, K., Wagner, W., Göllner, R., Gerjets, P., ... & Trautwein, U. (2021). Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. Educational Psychology Review, 33, 27-49.

[23] Battaglia, O. R., Di Paola, B., & Fazio, C. (2016). A new approach to investigate students' behavior by using cluster analysis as an unsupervised methodology in the field of education. Applied Mathematics, 7(15), 1649-1673.

[24] Labrović, J. A., Petrović, N., Anđelković, J., & Meršnik, M. (2023). Patterns of behavioral engagement in an online english language course: cluster analysis. Journal of Computing in Higher Education, 1-26.

[25] Li, X., Zhang, Y., Cheng, H., Zhou, F., & Yin, B. (2021). An unsupervised ensemble clustering approach for the analysis of student behavioral patterns. Ieee Access, 9, 7076-7091.

[26] Priyambada, S. A., Er, M., Yahya, B. N., & Usagawa, T. (2021). Profile-based cluster evolution analysis: Identification of migration patterns for understanding student learning behavior. IEEE Access, 9, 101718-101728.

[27] Shen, X., & Yuan, C. (2021). A college student behavior analysis and management method based on machine learning technology. Wireless Communications and Mobile Computing, 2021(1), 3126347.

[28] Yan, N., & Au, O. T. S. (2019). Online learning behavior analysis based on machine learning. Asian association of open universities journal, 14(2), 97-106.

[29] Mehenaoui, Z., Lafifi, Y., & Zemmouri, L. (2022). Learning behavior analysis to identify learner's learning style based on machine learning techniques. Journal of Universal Computer Science, 28(11), 1193.

[30] Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Rodríguez, S., & Corchado, J. M. (2019). Students performance analysis based on machine learning techniques. In Learning Technology for Education Challenges: 8th International Workshop, LTEC 2019, Zamora, Spain, July 15–18, 2019, Proceedings 8 (pp. 428-438). Springer International Publishing.

[31] Purwoningsih, T., Santoso, H. B., & Hasibuan, Z. A. (2019, October). Online Learners' behaviors detection using exploratory data analysis and machine learning approach. In 2019 fourth international conference on informatics and computing (ICIC) (pp. 1-8). IEEE.

[32] Liu Jinxin & Li Kexin. (2022). Research on an Improved SOM Model for Damage Identification of Concrete Structures. Applied Sciences(9),4152-4152.

[33] Kun Zheng,Chaolin Wang,Yu Zhao,Jing Bi,Yong Niu & Shuailong Lian. (2024). A dominant frequency-based hierarchical clustering method for evaluating the mode-II fracture mechanisms of shale using the acoustic emission technique. Theoretical and Applied Fracture Mechanics(PA),104644-104644.

[34] Vinay Priy Mishra,Yogendra Narain Singh,Feroz Khan & Malay Kishore Dutta. (2025). SeqDPI: A 1D-CNN approach for predicting binding affinity of kinase inhibitors.. Journal of computational chemistry(1),e27518.

[35] Ci Yusheng,Gao Xueyi,Li Haowen,Fai Yuen Kum & Wu Lina. (2025). Forecasting Freeway Traffic Volumes with Adverse Weather via a CNN-BiLSTM-Attention Model. Journal of Transportation Engineering, Part A: Systems(2).