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

# A Study on the Learning Effectiveness and Behavioural Association of Predictive Analytics in English Teaching in Higher Education

Danqun Huang[1], Yilu Ouyang[1,✉]

[1] *Hunan Petrochemical Vocational Technology College, Yueyang, Hunan, 414000, China*

ABSTRACT

With the booming development of large-scale open online courses, blended teaching, which combines traditional closed teaching and online open teaching, is increasingly favored by colleges and universities. In this paper, from the perspective of blended teaching of English in colleges and universities, based on the LSTM model to predict the relevant learning data in English teaching in colleges and universities, and based on the density optimization K-mean algorithm to cluster the student subjects with different learning behaviors, and then use the Apriori algorithm to study the correlation rules of the learning effectiveness and behaviors, to provide ideas for English teaching in colleges and universities. The clustering results show that the average learning scores of the first, second and third categories of learners are 92.35, 83.57 and 64.96 respectively. The results of association rule analysis show that routinely, the more active learners are in each learning session, the greater the possibility of getting better learning outcomes. The LSTM learning prediction model Precision, Recall and F1 assessment indexes trained with 4-month behavioral data are 0.899, 0.785 and 0.833 respectively, which are all greater than the corresponding index values of SVM, MLP and RF models, and have a significant advantage in prediction effect. This study provides lessons and references for improving the effectiveness of English teaching in colleges and universities.

*Keywords:* k-means clustering, Apriori algorithm, LSTM model, association rules, learning effectiveness prediction

## 1. Introduction

With the rapid development and popularization of information technology, college education is facing many new challenges and opportunities. In the large amount of student data, how to deeply mine the student learning behavior patterns and make effective predictions has become an important issue for the reform of college education and the enhancement of student learning effects [1-4]. For this reason,

✉ Corresponding author.
*E-mail address:* babylululuface@163.com (Y. Ouyang).
Received 01 March 2024; Revised 25 May 2024; Accepted 10 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-187

researchers have actively explored the analysis and prediction model of students' learning behavior in colleges and universities based on data analysis to provide personalized learning support and accurate teaching assessment [5-8]. Behavioral analysis model of students learning English in colleges and universities is to collect and analyze the behavioral data in the process of students' learning, from which the characteristics of students' behavior in learning English are extracted and a suitable model is constructed for analysis [9-12]. According to the different sources of data, English learning behavior analysis models can be divided into two main categories: first, learning behavior models based on online teaching platforms, and second, learning behavior models based on students' English learning trajectories.

With the popularization of online education, a large number of higher education institutions have established their own online education platforms [13-15]. These platforms record various behaviors of students in the process of learning English, such as watching videos, submitting assignments, and participating in discussions. By analyzing these behavioral data, they can reveal the patterns of students' English learning, discover students' learning difficulties, and provide targeted English learning suggestions. Students' learning trajectory refers to the traces left by students in the process of learning English, which can be obtained through the records in the learning management system, the submission of students' assignments, and teachers' evaluation [16-18]. Learning trajectories can not only reflect the time students spend learning English and the use of English learning resources, but also reveal students' habits and behavioral patterns of learning English. By analyzing students' English learning trajectories, students' learning status and learning level can be inferred and personalized learning support can be provided.

The English learning behavior prediction model for higher education students refers to the establishment of mathematical models based on students' historical English learning behavior data and personal characteristics, including students' personal information, learning time records, learning resource usage, and academic performance to predict students' future English learning behaviors and academic performance through methods such as machine learning and data mining [19-23]. Discover students' learning habits, learning interests, and the achievement of students' learning goals. Potential problems and difficulties can be found in time and timely measures can be taken to solve them. Schools and teachers can provide personalized learning counseling and support based on the prediction results to improve teaching quality and student satisfaction [24].

In this study, the k-Shape clustering algorithm based on density optimization and the Apriori algorithm based on Boolean matrix are used as the research methods to carry out the correlation research of learning effectiveness and behavior in English teaching in colleges and universities, and the learning effectiveness prediction model based on LSTM is constructed, and the model is evaluated for application. The results of the study show that the constructed LSTM model predicts learning effectiveness well, learning effectiveness and behavior have significant correlation, and positive learning behaviors can get better learning effectiveness under regular circumstances. In response to the findings, this study provides relevant teaching strategy suggestions for English teaching in colleges and universities.

## 2. Data Mining and Prediction of English Language Teaching in Colleges and Universities

### *2.1. Density optimization based k-means clustering algorithm*

For the problem of distance calculation, the traditional K-means algorithm has a cumbersome calculation process. In this paper, we use the density-based method to select the initial center of mass and optimize the algorithm by reducing the number of distance calculations. After optimization, the method only needs to calculate the distance between k initial centers of mass and each other, the distance from all sample points to the first initial center of mass $\mu_1$, as well as the distance between

the sample points and the initial center of mass that is closer to them. According to this method, i.e., the samples can be divided into clusters that are closest to them, which greatly reduces the number of calculations of the distances between the samples and the initial centers of mass, and improves the efficiency of the algorithm's operation.

Suppose the data set $X=\{x_1,x_2,\ldots,x_n\}$, $x_i$ is one of the sample points, $k$ is the number of clusters, $n$ is the number of samples, and $N$ is the number. Arbitrarily select two samples $x_i, x_j$ and calculate the distance between them using Euclidean distance to get:

$$d(x_i,x_j)=\sqrt{(x_i-x_j)^2}, i=1,2,\ldots,n, j=1,2,\ldots,n \tag{1}$$

The density of any sample point $x_i$ in dataset $X$ is denoted as:

$$D(x_i)=N\{x \mid d(x,x_i)<r, x\in X\} \tag{2}$$

Calculating the average density of the dataset yields:

$$\text{avg}\, D(x_i)=\frac{\sum_{i=1}^{n} D(x_i)}{n} \tag{3}$$

Calculate the dataset density threshold, where $c$ is a constant, to get:

$$Density\ threshold = c\times avgD(x_i)=c\times\frac{\sum_{i=1}^{n} D(x_i)}{n} \tag{4}$$

The specific flow of the improved K-means algorithm is shown below:

Step1: Calculate the density $D(x_i)$, and $Density\ threshold$ of the sample point $x_i$ in dataset $X$.

Step2: Compare the sample density $D(x_i)$ with the Density threshold, if $D(x_i)$ is smaller than the Density threshold, the sample is determined as an outlier and divided into the low-density sample set $L$, and vice versa, the sample is divided into the high-density sample set $H$.

Step3: Select the sample point with the highest density in sample set $H$ as the first initial center of mass $\mu_1$ and move it out of sample set $H$.

Step4: Select the second initial center of mass $\mu_2$, where $d(\mu_1,\mu_2)\geq r$, and $D(\mu_2)=\max\{D(x) \mid x\in(X-\mu_1)\}$.

Step5: Repeat Step4 until $k$ initial centers of mass are selected.

Step6: Calculate the distance between the $k$ initial center of mass points and the distance from all sample points to $\mu_1$.

Step7: Compare the size of $2d(x_i,\mu_1)$ and $d(\mu_1,\mu_2)$. If $2d(x_i,\mu_1)\leq d(\mu_1,\mu_2)$, then $d(x_i,\mu_1)\leq d(x_i,\mu_2), x_i$ is closer to $\mu_1$, and then $x_i$ does not belong to the clustering category where $\mu_2$ is located.

Step8: Calculate the distance between the sample and the next initial center of mass that has not been compared again, select which center of mass the sample point is closer to, and eliminate the initial center of mass that is farther away from the sample point after comparison.

Step9: Repeat Step8 until the $k$th initial center of mass, and so on, until all samples are classified into the closest cluster.

### *2.2. Boolean matrix based Apriori algorithm*

Aiming at the problem that Apriori algorithm needs to scan the transaction dataset many times during the operation process, this paper proposes a Boolean matrix-based Apriori algorithm, by representing the transaction dataset into the form of Boolean matrix, so as to realize the effect of scanning the

transaction dataset once to derive the correlation rules, which effectively saves the algorithm between the operations.

The specific steps of Apriori algorithm based on Boolean matrix are as follows:

Step1: Input the minimum support degree min_sup and the most confidence degree min_conf.

Step2: Scan the transaction dataset $D$ and construct the transaction matrix $D_{m\times n}$, where each row of the matrix represents the interaction record of a user, and each column represents the situation in which that kind of item has been interacted with, and $d_{i,j}=1$ when the $i$rd transaction contains item $I_j$, and 0 otherwise. The resulting matrix $D_{m\times n}$ is shown below:

$$D_{m\times n}=\begin{bmatrix} 1 & \cdots & 0 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \cdots & 1 & \cdots & 0 \end{bmatrix} \tag{5}$$

Column $i$ of matrix $D_{m\times n}$ represents the interaction of item $I_i$, and summing over this column yields the support of item $I_i$, $Support(I_i)=\frac{\sum d_{\cdot,i}}{m}$, where $\sum d_{\cdot,i}$ represents the support count of item $I_i$ and $m$ represents the number of transactions. Delete the columns in matrix $D_{m\times n}$ whose support is less than min_supp to get frequent 1 itemset $L_1$.

Step3: Generate candidate 2-item set $C_2$ from frequent 1-item set $L_1$. Sum each row of matrix $D_{m\times n}$ and delete rows less than 2. This is because it is not possible to generate frequent 2-item sets for items with values less than 2. The support of each item $\{I_i, I_j\}$ in $C_2$ is calculated by equation (6):

$$Support(I_i \rightarrow I_j)=\frac{\sum_{i=1}^{m}(d_{t,i}\wedge d_{t,j})}{m} \tag{6}$$

$\wedge$ in Eq. (6) represents the with operation, which is used to count the number of times items $I_i$ and $I_j$ appear in the user interaction list at the same time. The item sets with support less than min_supp are removed from the candidate 2-item set $C_2$ to obtain the frequent 2-item set $L_2$.

Step4: When $k\geq 3$, find the candidate $k$-item set $C_k$ by frequent $(k-1)$-item set $L_{k-1}$. First delete every row in matrix $D_{m\times n}$ and rows less than $k$. The support of each item set in $C_k$ can be calculated by formula (7). Delete the itemsets with support less than min_supp to get frequent $k$ itemset $L_k$:

$$Support(I_i, I_j, \ldots I_{k-1})=\frac{\sum_{i=1}^{m}(d_{t,j}\wedge d_{t,j}\ldots\wedge d_{t,k-1})}{m} \tag{7}$$

The Apriori algorithm is a process of gradually finding the frequent $k$ itemset $L_k$ from the candidate $(k-1)$ itemset $C_{k-1}$. Thus the frequent $k$ itemset $L_k$ and the candidate $k-1$ itemset $C_{k-1}$ can be represented by equation (8):

$$L_k=\{c\in C_{k-1}\mid Support(c)\geq \min\_supp\} \tag{8}$$

Step5: Find candidate $(k+1)$ itemset $C_{k+1}$ from frequent $k$ itemset $L_k$, if $C_{k+1}$ is still not empty, then continue to execute Step4, otherwise execute Step6.

Step6: Output the association rule Rules that satisfy the minimum confidence condition.

### *2.3. LSTM prediction algorithm*

The Long Short-Term Memory (LSTM) network, which is capable of learning long-term dependencies, is one of the most suitable neural networks for time series prediction, and controls the circulation and loss of features through the input gate $(i_t)$, output gate $(o_t)$, and forgetting gate $(f_t)$. Its internal structure is shown in Fig. 1. In LSTM, the memorized state $C_t$ is one of the core of LSTM, which is updated at each time step. Specifically, the memory state consists of two parts, one part is the information that needs to be deleted, which is controlled by the forgetting gate: the other part is the new information that needs to be added, which is controlled by the input gate, and the final memory state $C_t$ consists of the combination of these two parts.

Where $C_{t-1}$ is the output of the previous cell, $X_t$ is the input of the current state, and $\sigma$ is the sigmoid function. First, the first step of the LSTM decides which cell state information to discard by means of a forgetting gate, after inputs $H_{t-1}$ and $X_t$, and after the sigmoid function, the inputs are a number between 0 and 1, with 0 indicating discard and 1 indicating complete retention. The formula is shown in (9):

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{9}$$

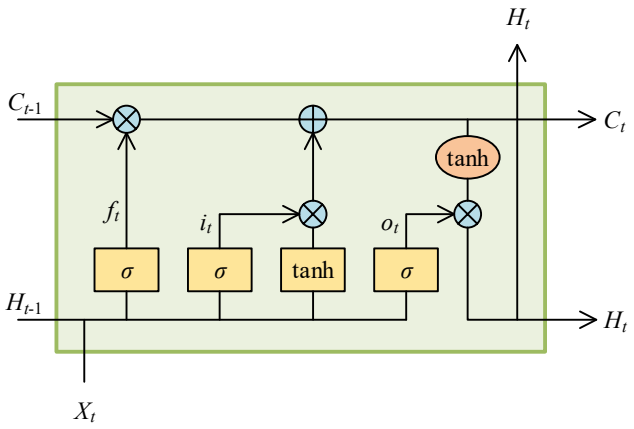


**Fig. 1.** Schematic diagram of the neuron structure of the LSTM network

After going through the oblivion gate, a decision needs to be made on what cell state should be added in, the process is shown in Eqs. (10), (11):

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{10}$$

$$\tilde{c}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{11}$$

Next it is possible to update the cell state, as shown in Equation (12):

$$C_t = f_t \times C_{t-1} + i, \tilde{c}_t \tag{12}$$

Finally, the cell state is output through the output gate, the cell state is transformed by the tanh function to keep the output value at (1, 1), and then dot-multiplied by the output value of the sigmoid function, and finally added to the hidden state. The formulas are as in (13) and (14):

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{13}$$

$$h_t = o_t \times \tanh\left(C_t\right) \tag{14}$$

The LSTM network is executed in the following steps:

Step1: First determine the structure, activation function and loss function of the LSTM network.

Step2: Map the input sequence to a fixed vector dimension representation.

Step3: Calculate the output layer prediction results and the actual labeled loss function values according to the task needs.

Step4: Calculate the gradient of the loss function with respect to the neural network parameters by back propagation algorithm and update the parameters.

Step5: Repeat the above steps until the pre-set number of iterations is reached or the termination condition is satisfied.

## 3. Evaluation of LSTM-based Predictive Model for English Learning Effectiveness in Colleges and Universities

### *3.1. Specific steps of LSTM learning effectiveness prediction modeling*

The steps of LSTM network execution are as follows:

1) First determine the structure, activation function and loss function of the LSTM network.

2) Map the input sequence to a fixed vector dimension representation.

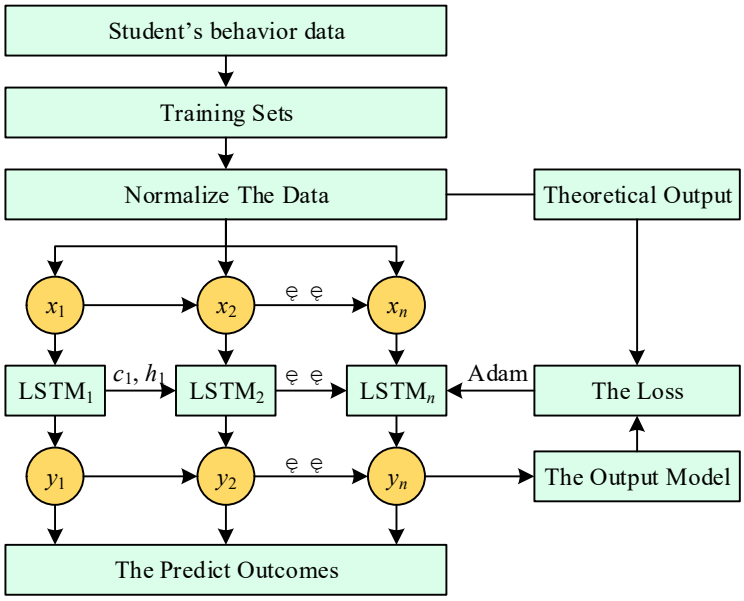


**Fig. 2.** LSTM-based learning effect prediction model

The LSTM prediction model is shown in Fig. 2, and its specific implementation process is as follows:

Step1: Divide the training set and test set. Divide the total 1386 students' experimental dataset into training set and test set, of which 75% is used as the training set to train the model to ensure better accuracy, and 25% is used as the test set to test the effect of the prediction experiment, to ensure that there is no overlap of data between the training set and the test set.

Step2: Input feature items. Sort the inputs according to temporality and forward compute the neuron output values.

Step3: Calculate the error term. Meanwhile the propagation of the error term includes two directions: reverse time propagation, and upward layer propagation.

Step4: Adjust the hyperparameters of the model. Use backward propagation to make LOSS minimized.

### 3.2. Evaluation indicators

The evaluation of experimental results for categorical prediction problems is generally done using metrics such as precision, recall, and $FI$. $Precision$, or accuracy, indicates the probability that a sample whose experimental prediction is positive is actually equally positive:

$$precision = TP / (TP + FP) \tag{15}$$

$Recall$ i.e., the recall, the proportion of samples that are true to the positive class to those that are predicted to be in the positive class, is given by the following formula:

$$recall = TP / (TP + FN) \tag{16}$$

The F1 value index is used to evaluate the classification performance of the classifier, which is used as the main index to evaluate the prediction effect in this paper, and the index level is directly proportional to the precision and recall, and the specific formula for calculating the F1 value is as follows:

$$F1 = \frac{2Precison * Recall}{Precison + Recall} \tag{17}$$

### 3.3. Comparison of experimental results analysis

This paper randomly selected five horizontal classes in a university English major as the source of data, and the study of their English is followed by a semester, collecting the relevant data. The LSTM learning effect prediction model studied in this chapter is trained in steps of months, firstly, every month as a span of increase, outputting a learning prediction result, testing the effect of LSTM affected by the size of the amount of training data on the prediction effect, followed by the entire semester data as a training set and traditional machine learning SVM, MLP, RF classification algorithms for comparative analysis, the training set and test set used are the The same, to compare the advantages and disadvantages of each classification algorithm in this learning prediction experiment.LSTM to one month as a node to predict the results of the examination is shown in Table 1.

From the experimental data in Table 1, it can be seen that the accuracy of the trained LSTM prediction model increases with the growth of the input sample period, and the F1 value reaches 0.833 by the time the data in the training set is expanded to the fourth month, but it still has not stabilized. This indicates that the prediction effect of the fully trained LSTM model will be better, and then the model prediction effect is compared horizontally with the prediction results of traditional machine learning classification algorithms.

**Table 1.** Prediction results of LSTM learning effectiveness

| Time | Precision | Recall | F1 |
|---|---|---|---|
| The first month | 0.710 | 0.656 | 0.720 |
| The second month | 0.779 | 0.708 | 0.778 |
| The third month | 0.781 | 0.837 | 0.723 |
| The fourth month | 0.899 | 0.785 | 0.833 |

As shown in Figure 3, the horizontal comparison graph of the prediction results of different models shows that with the whole semester data as the training set, the LSTM learning prediction model Precision, Recall, and F1 evaluation indexes trained with 4 months of behavioral data are 0.899, 0.785, and 0.833, respectively, which are greater than the corresponding index values of the other models, and have achieved a significant advantage. The reason for this is that after the LSTM neural network adds feature temporality, the LSTM information propagation is more orderly and the information transfer is more efficient, and the prediction effect of the prediction model is significantly improved after the model is adequately trained with the training set data of long time series.

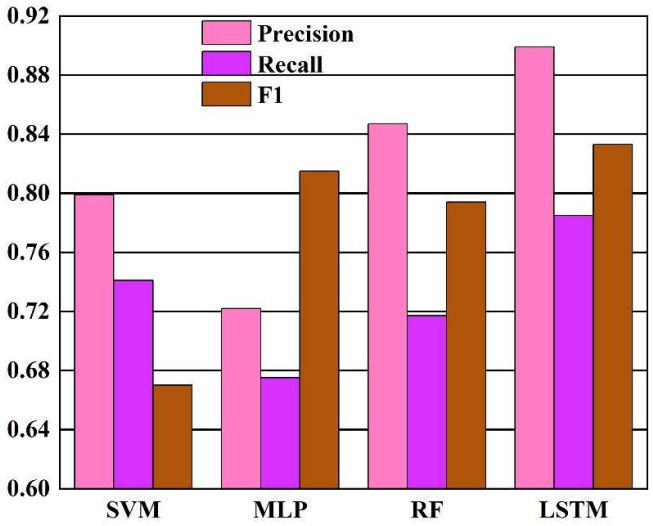


**Fig. 3.** Horizontal comparison of the predictions made by different models

The experiments show that the advantages of LSTM are more than dealing with temporal feature data with unique and convenient characteristics, which can be fully iterated with limited data, and the predicted values are used as input for backward derivation, and there is also a further improvement in the prediction effect relative to the traditional algorithms.

## 4. Research on the correlation between the effectiveness of English learning and behavior in tertiary schools

With the booming development of massively open online courses, blended teaching, which combines traditional closed teaching and online open teaching, is increasingly favored by colleges and universities. Therefore, this paper conducts a correlational study of students' learning behavior and effectiveness from the perspective of blended teaching of English in colleges and universities.

### *4.1. Learning behavior cluster analysis*

In this paper, we adopt K-means analysis method based on density optimization to filter the different learning behaviors of learners in English teaching in colleges and universities, and select eight indicators as clustering variables: online learning experience, blended learning experience, self-efficacy, homework, exercises, discussions, classroom quizzes, and classroom interactions. The clustering situation of the eight clustered behavioral indicators is shown in Table 2 below.

From Table 2, it can be seen that each behavioral element has high significance to the clustering, among which homework and exercises have the highest contribution F to the clustering, respectively 379.546 and 76.824, with significance lower than 0.001, which indicates that the one-way analysis is significant, and it can be seen that these two components can reflect the learning effect of the learners more accurately, followed by discussion, blended teaching experience and self-efficacy to the clustering. Contribution is relatively weak, while classroom quizzes and classroom interaction have the least contribution to the clustering.

**Table 2.** Clustering single factor analysis of each behavior element

| Behavioral indicators | Clustering | | Error | | F | Saliency |
|---|---|---|---|---|---|---|
| | Mean squared | df | Mean squared | df | | |
| Online Learning Experience | 0.002 | 2 | 0.121 | 279 | 8.026 | 0.018 |
| A blended learning experience | 0.682 | 2 | 0.212 | 279 | 34.418 | 0.035 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Self-efficacy | 31.695 | 2 | 24.538 | 279 | 23.184 | 0.028 |
| Homework | 14987.846 | 2 | 38.779 | 279 | 379.546 | 0.000 |
| Exercise | 3217.538 | 2 | 42.762 | 279 | 76.824 | 0.000 |
| Discussion | 69.868 | 2 | 18.653 | 279 | 35.340 | 0.014 |
| Quizzes in class | 20.534 | 2 | 66.835 | 279 | 8.058 | 0.020 |
| Classroom interaction | 1.512 | 2 | 1.589 | 279 | 12.145 | 0.037 |

In the learning behavior clustering analysis, one of the most critical influencing factors is the k-value, which sets the number of clustering results, and the setting of the k-value has a great relationship with the clustering results. In this paper, we refer to a large number of literature on the clustering study of learners, and calculate the clustering results of k-value of 2/3/4/5, respectively, and after many analyses, we finally arrive at the optimal number of categories of 3 based on the sample actuality, i.e., the final results of the experiments will be learners into three categories. The clustering results are shown in Table 3.

Table 3 shows that the learners are divided into three categories. The self-efficacy of the first category of learners is 35, which is relatively high, with strong self-discipline and learning ability, and the scores of 91, 29 and 18 for homework, exercises and discussion, respectively, are also relatively high, and their daily performance is also relatively good, indicating that this category of learners has a more active learning atmosphere in the English course. The second category of learners scored 19 points in the classroom quiz, the best among the three categories of learners, and their self-efficacy was 34, which was relatively good, indicating that this category of learners had a better attitude towards learning in this course than the other learners and had good classroom performance, and that they performed poorly in their homework and exercises online behavior, and their independent learning ability was relatively weak. The third category of learners is weaker than the other two categories in all aspects, in which the homework score is 53, which is a big gap with the other two categories, and there is also a certain gap between the exercises, discussions, and classroom quizzes and the other two categories of learners.

**Table 3.** k-means clustering results

| Behavioral indicators | Clustering | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Online Learning Experience | 2 | 2 | 2 |
| A blended learning experience | 3 | 3 | 3 |
| Self-efficacy | 35 | 34 | 32 |
| Homework | 91 | 82 | 53 |
| Exercise | 29 | 21 | 19 |
| Discussion | 18 | 17 | 15 |
| Quizzes in class | 16 | 19 | 18 |
| Classroom interaction | 2 | 2 | 2 |

In order to reflect the learning effect of all types of learners, this paper will make statistics on the average score of their various types of learning achievements, and the statistical results are shown in Table 4.

As can be seen from Table 4, there is a significant difference in the average scores of different types of learners, and the first category of learners has the highest average learning score of 92.35, which is 4.20 points higher than the total average score. The second category of learners had a mean score of 4.58 points lower than the total mean score. Category 3 learners had the lowest average score of 64.96, with their scores hovering on the edge of the passing line. In view of the above statistics, it can be seen that the effect of different types of learning behavior characteristics is positively proportional to the achievement.

**Table 4.** Average scores of various types of learners

| Type of learner | Average academic performance (out of 100) |
|---|---|
| The first type of learners | 92.35 |
| The second type of learner | 83.57 |
| The third type of learner | 64.96 |
| Total average score | 88.15 |

### *4.2. Analysis of Learning Effectiveness and Behavioral Association Rules*

In order to explore the correlation rules between learning behaviors and learning effectiveness, this study uses SPSS Statistics 23 and SPSS Modeler18 analysis tools to import and discretize the data on learning effectiveness and behaviors of online teaching of English in colleges and universities. The results of data discretization are shown in Table 5, and the letters in parentheses are the corresponding learning behavior serial numbers.

**Table 5.** Results of discretization of online learning behavior data

| Behavioral indicators | Value | Level | Behavioral indicators | Value | Level | Behavioral indicators | Value | Level |
|---|---|---|---|---|---|---|---|---|
| Number of videos (A) | [0,4] | I | Video views (E) | [0,10] | I | Number of self-test assignments participated (I) | [0,1] | I |
| | [5, 12] | II | | [11,85] | II | | [2,8] | II |
| | [13, 30] | III | | [86,455] | III | | [9,16] | III |
| | [31,inf] | IV | | [456,inf] | IV | | [17,inf] | IV |
| Login times (B) | [0,6] | I | Number of posts (F) | (inf,0] | I | Number of Posting words (J) | [inf,0] | I |
| | [7,18] | II | | [1,5] | II | | [1,2150] | II |
| | [19,36] | III | | [6,10] | III | | [2151,8750] | III |
| | [37,inf] | IV | | [11,inf] | IV | | [8751,inf] | IV |
| Learning Outcomes (C) | [0,12] | I | Self-test the number of reworks (G) | [0,1] | I | Video Viewing time (K) | [0,5.21] | I |
| | [12.01,72.95] | II | | [2,10] | II | | [5.22,115.72] | II |
| | [72.96,84] | III | | [11,27] | III | | [115.73,372.95] | III |
| | [84.01,inf] | IV | | [28,inf] | IV | | [372.96,inf] | IV |
| Average score of assignments (D) | [0,52.48] | I | Number of non-video resources used (H) | [0,4] | I | Total number of non-video resources used (L) | [0,4] | I |
| | [52.49,73.5] | II | | [5,21] | II | | [5,25] | II |
| | [73.51,inf] | III | | [22,inf] | III | | [26,inf] | III |

Import the discrete processed data into SPSS Modeler, add Apriori modeling node, and make relevant settings for Apriori node, set the preterm as each learning behavior, the postterm as learning effect, the minimum support as 10%, set the minimum rule confidence as 30%, and set the maximum number of preterm as 3. Organize the analysis results, the first nine association rules are shown in Table 6.

As shown in Table 6, the confidence level of association rules 0~4 are 56.357%, 53.166%, 43.573%, 45.319% and 93.342%, which are all greater than 40%, while the confidence level of association rules 5~9 are 37.563%, 34.258%, 33.124%, 35.621% and 32.876%, which are all less than 40%. In contrast, the support of the association rules of 0~4 and 5~9 are not much different. Therefore, they can be divided into two categories, one is conventional rules (0~4), which are more in line with common sense, and the other is non-conventional rules (5~9), which reflect some of the existence of anomalies in the learning process.

From the conventional rules, the law that can be summarized is that if the richer the learning behavior of the learner, can be active, hard, and persistent in all learning sessions, there is a greater

possibility of being able to obtain better learning results, which is consistent with our conventional experience. The unconventional rules, on the other hand, have some associations that defy our understanding.

Rule 5 shows that some learners did not end up with good learning outcomes despite their efforts to persistently learn through resource utilization, with video viewing lengths of 372 minutes or more. Rules 7 to 9 show that a considerable number of learners, although they did not watch the videos actively and persistently, and did not actively participate in the self-tests and homework sessions, but eventually achieved good learning results. These rules reflect that there is a certain degree of chance in the examination of English courses in colleges and universities, and it is speculated that it may be due to some problems such as the setting of the paper or due to the fact that there is no one to supervise the examination remotely, which leads to a part of the learners obtaining a better academic performance without systematic learning.

From these unconventional rules, video teaching does not apply to all students. Some students are still unable to achieve good teaching results even if they insist on watching video. And some students can get enough knowledge through their own learning methods, even if they don't watch video.

In addition, it can be seen from rules 5~7 that the video viewing duration does not play a big role in learning effect in comparison, so we should pay more attention to the quality of the learners when they watch the video, instead of trying to quantify the duration of the learners watching the video as a criterion for the quality of the learners' learning.

**Table 6.** Results of association rules

| ID | Association rule | Confidence degree | Support degree |
|---|---|---|---|
| 0 | A=Ⅳ, E=Ⅳ, H=Ⅲ →C=Ⅳ | 56.357% | 13.818% |
| 1 | J=Ⅳ, F=Ⅳ, D=Ⅲ →C=Ⅳ | 53.166% | 11.247% |
| 2 | A=Ⅲ, D=Ⅱ →C=Ⅲ | 43.573% | 12.654% |
| 3 | K=Ⅲ, I=Ⅲ →C=Ⅲ | 45.319% | 11.843% |
| 4 | G=Ⅰ, A=Ⅱ, D=Ⅰ →C=Ⅰ | 93.342% | 18.425% |
| 5 | K=Ⅳ, D=Ⅱ →C=Ⅱ | 37.563% | 11.639% |
| 6 | K=Ⅲ →C=Ⅳ | 34.258% | 27.021% |
| 7 | K=Ⅱ, H=Ⅱ, L=Ⅱ →C=Ⅲ | 33.124% | 11.588% |
| 8 | I=Ⅱ, H=Ⅱ, L=Ⅱ →C=Ⅲ | 35.621% | 13.529% |
| 9 | A=Ⅱ, E=Ⅱ →C=Ⅲ | 32.876% | 12.943% |

## 5. Conclusions and Suggestions for Teaching Strategies

### *5.1. Conclusion*

In this paper, from the perspective of blended teaching of English in colleges and universities, we constructed a learning effectiveness prediction model based on LSTM and evaluated the application of the model, and also mined the data related to learning effectiveness and behavior using the improved K-mean algorithm and Apriori algorithm.

The LSTM learning prediction model Precision, Recall and F1 evaluation indexes trained with 4-month behavioral data are 0.899, 0.785 and 0.833, respectively, which are greater than the corresponding index values of SVM, MLP and RF models, and have significant advantages in prediction effectiveness.

In this study, the learners were categorized into three groups, the first group of learners had strong self-discipline and learning ability, and their daily performance was also relatively good, and they were more active in the English course. The second category of learners have more excellent attitude and good classroom performance in English courses, and perform poorly in homework and exercises

online behavior, and are relatively weak in self-directed learning. Category 3 learners were weaker in all aspects relative to the other two categories. There is a significant difference in the average scores of different types of learners, with the average learning scores of the first, second and third types of learners being 92.35, 83.57 and 64.96 respectively. Different types of learning behaviors characterize their effects and grades positively.

There is a significant correlation between learning behaviors and learning outcomes. Routinely, the richer the learners' learning behaviors are, and the more they can study actively, diligently and persistently in all learning sessions, the more likely they are to achieve better learning outcomes.

### *5.2. Recommendations for Teaching Strategies:*

1) Teachers should pay attention to the reasonable distribution of online and offline teaching activities and the number of tasks when carrying out blended teaching, and should not focus on the pursuit of the richness of the teaching design and students' learning behavior data, and set up more learning tasks online and offline, which will increase the burden of students' learning and cause fatigue, thus weakening the motivation and interest of students' learning, which is not worth the loss.
2) In the actual teaching process, teachers should control the students' answering time to the relevant questions, and at the same time, according to the students' answering time to grasp the students' learning of the knowledge points, so as to give some after-school counseling to the students who have a longer answering time. For the general students answering time are longer learning content in the next class to consolidate and strengthen again.
3) In the process of online teaching, we should pay more attention to the quality of the learners watching the video, rather than to quantify the length of the learners watching the video as a standard of the quality of the learners' learning. When taking online exams, we should reasonably set the paper surface and strengthen the supervision of remote exams to ensure the authenticity of the exam results, so as to improve the effectiveness of student learning.

## Funding

2023 Approved Project of Hunan Provincial Educational Science "14th Five-Year Plan": Research on the Cultivation of Field Engineers in Equipment Manufacturing Major with the Apprenticeship System with Chinese Characteristics as the Main Form (XJK23BZY027).

## References

[1] Dong, Y. . (2022). Application of artificial intelligence software based on semantic web technology in english learning and teaching. Journal of Internet Technology(1), 23.

[2] John, C. , & Margarita, M. B. . (2021). 1023the uses and abuses of regression models: a new approach to teaching regression analysis. International Journal of Epidemiology(Supplement_1), Supplement_1.

[3] Ahmed, F. , Kim, J. K. , Khan, A. U. , Park, H. Y. , & Yeo, Y. K. . (2017). A fast converging and consistent teaching-learning-self-study algorithm for optimization : a case study of tuning of lssvm parameters for the prediction of nox emissions from a tangentially fired pulverized coal boiler. Journal of Chemical Engineering of Japan, 50(4), 273-290.

[4] Zheng, L. , Chen, C. , Liu, W. , Long, Y. , & Lu, C. . (2018). Enhancement of teaching outcome through neural prediction of the students' knowledge state. Human Brain Mapping, 39(2), 3046.

[5] Iván García-Magario, Plaza, I. , Raúl Igual, Andrés S. Lombas, & Jamali, H. . (2020). An agent-based simulator applied to teaching-learning process to predict sociometric indices in higher education. IEEE Transactions on Learning Technologies, 13(2), 246-258.

[6] Yuling, M. A. , Cui, C. , Jun, Y. U. , Guo, J. , Yang, G. , & Yin, Y. . (2020). Multi-task miml learning for

pre-course student performance prediction. Frontiers of Computer Science, 14(5).

[7] Chen, Y. . (2021). College english teaching quality evaluation system based on information fusion and optimized rbf neural network decision algorithm. Journal of Sensors.

[8] Zhen, C. . (2021). Using big data fuzzy k-means clustering and information fusion algorithm in english teaching ability evaluation. Complexity, 2021(5), 1-9.

[9] Yi, Y. , Zhang, H. , Karamti, H. , Li, S. , Chen, R. , & Yan, H. , et al. (2022). The use of genetic algorithm, multikernel learning, and least-squares support vector machine for evaluating quality of teaching. Scientific programming(Pt.5), 2022.

[10] Zhang, Y. . (2021). The realization of intelligent algorithm of knowledge point association analysis in english diagnostic practice system. Complexity, 2021(5), 1-10.

[11] Xu, C. . (2021). Pbl english micro-audio and video teaching model based on data mining algorithm. Journal of Intelligent and Fuzzy Systems(6), 1-9.

[12] Hou, J. . (2021). Online teaching quality evaluation model based on support vector machine and decision tree. Journal of Intelligent and Fuzzy Systems, 40(2), 2193-2203.

[13] Aldabbagh, G. , Alghazzawi, D. M. , Hasan, S. H. , Alhaddad, M. , Malibari, A. , & Cheng, L. . (2020). Optimal learning behavior prediction system based on cognitive style using adaptive optimization-based neural network. Complexity, 2020.

[14] Yin, H. , Wu, H. , & Tsai, S. B. . (2021). Innovative research on the construction of learner's emotional cognitive model in e-learning by big data analysis. Mathematical Problems in Engineering, 2021.

[15] Fang, C. . (2020). Intelligent online teaching system based on svm algorithm and complex network. Journal of Intelligent and Fuzzy Systems, 40(5), 1-11.

[16] Zheng, Lifen, Chen, Chuansheng, Liu, & Wenda, et al. (2018). Enhancement of teaching outcome through neural prediction of the students' knowledge state. Human brain mapping, 39(7), 3046-3057.

[17] Chen, L. , Wang, L. , & Zhou, Y. . (2022). Research on data mining combination model analysis and performance prediction based on students' behavior characteristics. Mathematical Problems in Engineering, 2022.

[18] Chu, N. , & Ma, W. . (2021). Distribution of large-scale english test scores based on data mining. Complexity, 2021.

[19] Qian, Y. , Cheng-Cin Li, Xiu-Guo Zou, Xue-Bin Feng, & Yong-Qian Ding. (2021). Research on predicting learning achievement in a flipped classroom based on moocs by big data analysis. Computer Applications in Engineering Education(4).

[20] Li, H. . (2022). Analysis of the role of hs-hkrvm analytic hierarchy process in the evaluation of english teaching quality. Mobile information systems(Pt.6), 2022.

[21] Lim, T. M. , & Yunus, M. M. . (2021). Teachers' perception towards the use of quizizz in the teaching and learning of english: a systematic review. Sustainability, 13(11), 6436.

[22] Ma, F. . (2019). An analysis of the application of cooperative learning in english teaching in the context of online education. Basic & clinical pharmacology & toxicology.(S9), 125.

[23] Wu, J. , & Chen, B. . (2020). English vocabulary online teaching based on machine learning recognition and target visual detection. Journal of Intelligent and Fuzzy Systems(5), 1-12.

[24] Hao, K. . (2020). Multimedia english teaching analysis based on deep learning speech enhancement algorithm and robust expression positioning. Journal of Intelligent and Fuzzy Systems, 39(3), 1-13.