



MOBILE USER ENGLISH LEARNING PATTERN RECOGNITION MODEL BASED ON INTEGRATED LEARNING

QIAN ZHANG* AND YUNLAN REN†

Abstract. The research in the field of mobile assisted language learning has a long history. It basically follows the route from theory to application practice, but there are few process studies on learners' individual language skills learning behavior based on mobile platform data. This research takes vocabulary learning as the starting point, and constructs a mobile user English learning pattern recognition model with improved Stacking integration algorithm. The purpose of this study is to identify different learning modes by analyzing the learning behavior and learning data of mobile users, and to provide personalized learning suggestions for users. The evaluation goal of this study is the accuracy and robustness of the mobile users' English learning pattern recognition model, and the accuracy is the classification accuracy of the model for different learning patterns. Robustness is the stability and consistency of the model in different situations. In order to evaluate the integrated English learning mode of mobile users, the study first divides the collected learning data set into training set and test set. In this step, the method used in the study is cross-validation, which aims to reduce the difference of evaluation results caused by different data sets. For the relevant features in mobile users' learning data, the features extracted by the research include learning behavior and learning progress, which can accurately reflect the learning mode. Then, the integrated learning method is used to train the model, and the best parameter combination is selected through the training set. Finally, the study uses the test set to evaluate the trained model, and calculates the accuracy and recall index of the model on the test set. Through this evaluation method, the evaluation results of the integrated learning pattern recognition model in terms of accuracy and robustness are obtained, and reference is provided for the improvement of the model. The model proposed in this study is suitable for a large number of user data. Because the learning behavior of users is influenced by personal habits, there are limitations in obtaining enough high-quality data, so the labeling of data is subjective. It is still a challenge to select the most representative features of user learning behavior extracted from the model. The feature selection method can lead to different results, and the process requires a lot of human intervention. The experiment conducted mining and analysis on user learning behavior data of a domestic English vocabulary learning APP. Compared with the confusion matrix of the traditional Stacking model, the improved Stacking model has a stronger ability to distinguish user learning patterns. According to the formula, the accuracy of the improved Stacking model is 91.29%; The accuracy of traditional Stacking model is 90.71%. The ROC curve of the improved Stacking model is smoother than the three single models. Its AUC value is 0.85, which is the same as that of XGBoost. The function is also improved compared with the traditional Stacking model, Logistic Regression (LR) and Random Forest (RF) model. Therefore, the Stacking integrated model owns the best forecast performance and can accurately predict the long-term learning mode of users. In this study, the English learning patterns of mobile users are identified by the method of integrated learning, so that the prediction results of multiple basic learners can be integrated, and the complementarity between different learners can be effectively dealt with, thus improving the generalization ability of the model. This model aims at identifying the English learning patterns of mobile users, and can accurately identify the learning patterns of users by analyzing their learning behaviors on mobile devices. This is of great significance for personalized English learning recommendation. The English learning pattern recognition model for mobile users proposed in this study can identify users' learning patterns by analyzing their learning behaviors, thus providing personalized learning support and suggestions. In the process of digital manufacturing, they can learn from the idea of learning pattern recognition model and identify the patterns and laws in the production process by analyzing production data and process parameters, so as to optimize the production process and improve production efficiency.

Key words: Mobile assisted language learning; Learning behavior analysis; Integrated learning; Learning mode; Stacking

1. Introduction. Currently, under the advancement of mobile technique and the popularity of smart terminals, mobile devices such as mobile phones have become more widely applied in people's daily life. Meanwhile, the number of mobile phone holders has also continued to rise [1]. In this context, the popularization of smart phones in education means that the way of knowledge transmission and reception has undergone great changes. Mobile learning, born with the integration of technology and education, is characterized by personalization,

*College Of Foreign Languages, Inner Mongolia Agricultural University, Hohhot, 010018, China

†College Of Foreign Languages, Inner Mongolia Agricultural University, Hohhot, 010018, China (Corresponding author, ange120230307@163.com)

interactivity and accessibility. Therefore, this way of learning is attracting much attention [2, 3]. At present, the research on mobile learning has covered many levels from theory to practice. It involves interactive technology, learning mode, and methods of integrating with traditional classroom teaching in mobile learning. Among them, language learning is one of the important fields in mobile learning research [4]. The academic research on the application of mobile technology in language learning has undergone nearly 20 years of development, and has formed a rich research content and relatively independent research field, namely, Mobile Assisted Language Learning (MALL) [5]. MALL has not only received widespread attention and attention in academic circles, but also aroused great interest in the commercial market. Major Internet and online education enterprises have launched a wide range of language learning applications, involving vocabulary, reading, listening, translation and other language skills, to satisfy the users' learning needs in all aspect [6]. Language APPs, especially word APPs, are at the top of the list of online education product traffic with tens of millions of independent users all year round [7]. Meanwhile, the user behavior data recorded by these applications can help academic researchers understand the learning status of learners in the process of using APP from an objective perspective and at a micro level [8]. However, researchers generally prefer to use questionnaires and experimental design to obtain data. This cannot mine more valuable information from the big data. In addition, there are few studies on the development characteristics and change rules of user's learning behavior at present [9]. Therefore, from the perspective of data mining, this study uses integrated learning technology to identify mobile users' English learning patterns. The proposed model can comprehensively consider the characteristics of users' behavior and learning content, and comprehensively analyze and learn these characteristics. Compared with the traditional learning mode based on desktop computer, this method has the characteristics of portability and personalization, and can better adapt to the learning habits of mobile users. Because the mobile users' English learning pattern recognition model performs well on the training set, and there are individual differences in users' learning patterns, it can be shown that the method has strong generalization ability. For the performance of users in the new learning scene, the generalization ability of the model is further improved. When the model is faced with other design methods other than the reference architecture, the model is a complex integrated model, so the explanation of the model is effectively discussed. In order to provide an effective design method of mobile users' English learning pattern recognition model, this paper studies the collection and extraction of relevant data generated by mobile users in the process of English learning, and then preprocesses the extracted features. For the specific learning needs, the model is trained by using the preprocessed data. After evaluating the trained model in an independent test data set, this paper studies the optimization of the model, and applies the optimized model to the actual English learning scene, so as to identify the user's learning mode in real time. In addition, the research also focuses on the privacy protection and data security of the mobile users' English learning pattern recognition model, so as to ensure that the users' personal information is reasonably protected.

2. Related work. Recently, integrated learning has received extensive attention in the data mining. Many researchers use it as a tool for disease prediction and detection. Pei et al. proposed a learning modeling framework integrating genetic information to predict early treatment reaction of antidepressants for severe depression at baseline. The results showed that the performance of the suggested method was improved compared with the single-level model; The accuracy of imaging data and genetic data was improved from 0.61 to 0.86. The integrated learning framework of genetic features increased the sensitivity from 0.78 to 0.87 [10]. Vogelstein et al. applied the integrated learning model to the early detection of Alzheimer's disease in the elderly. The results indicated that the detection accuracy was 94.64%; The comparison with other models also shows that the model are the best [11]. To achieve accurate classification of rock burst intensity, Zhang et al. suggested an integrated machine learning method. The results showed that the accuracy of rockburst classification obtained by the classifier set was improved by 15.4%. In addition, the importance of the prediction variables obtained from the classifier set showed that the most unstable variable for rock burst was elastic energy index [12]. Hong et al. paid attention to the process monitoring of industrial activities and proposed an integrated process monitoring method. The experiment proved the effectiveness of the method. The research compared the performance of this method with other methods, and the results also proved that this scheme had better performance [13]. Some researchers also noticed the use of integrated learning technology in education. Nuankaew et al. proposed a higher education curriculum recommendation model. The results showed that the accuracy

of the model developed using majority voting technology is the highest, reaching 91.65% [14].

In addition, the Stacking integration algorithm is popular as well. Tan et al. focused on the development of power load forecasting technology. To carry out load forecasting, an experimental load forecasting method combining support vector regression and superposition is proposed. A series of comparison algorithms are introduced. The comparison results verify that the method has high prediction accuracy and generalization ability, and is superior to the comparison algorithm [15]. Meharie et al. applied Stacking integrated machine learning algorithm to predict highway construction cost. The results showed that the model could predict accurately. The comparison results of the models showed that the Stacking integrated model was superior to the single model in all performance indicators. For the RMSE value, the accuracy of the results produced by the Stacking integrated cost model was 86.8%, 87.8% and 5.6% higher than that of linear regression, vector machine support and neural network models respectively [16]. Acuna et al. proposed a genetic algorithm with improved efficiency. The research results showed that compared with several advanced genetic algorithms with multimodal functions, the convergence speed of this improved genetic algorithm performed better. The research results also tested the effectiveness of the model, which could be applied to the two-dimensional common reflector superposition problem [18]. Youssef et al. paid attention to the dropout of MOOC learners and developed a prediction model to classify them. This method applied feature selection method and integrated machine learning algorithm, and took several similar algorithms as the basis for comparison. Several performances of the method were evaluated by experiments. The comparison results demonstrated that this algorithm has excellent performance and its prediction accuracy reaches 98.6% [17].

To sum up, the integration algorithm, especially the Stacking integration algorithm, has been more and more widely used in many fields. However, the application of this kind of algorithm in the field of education is relatively limited, and few people pay attention to its application in data mining of language learning APP. In addition, the existing research mainly applies it to the fields of curriculum recommendation and performance prediction; However, there are few studies on the development characteristics and change rules of user's learning behavior. Therefore, the research starts with English vocabulary learning and uses Stacking integration algorithm to study mobile English learning APP. The experiment deeply digs and analyzes user learning behavior data. The purpose of this experiment is to mine the learning mode of user groups from the memorization data through clustering and modeling. The experiment takes the learning state representation attribute as the explanatory variable, and then realizes the prediction and recognition of the learning mode.

3. Mobile user English learning pattern recognition model based on integrated learning.

3.1. Establishment of mobile user English learning pattern recognition model. The English learning pattern recognition model built by the research institute consist of data preprocessing, feature screening, model training and output results, as shown in Figure 2.1. Before model training, data preprocessing and feature filtering are required. When obtaining data, due to various subjective and objective reasons, a large number of missing values and noise values may be generated; These data are very unfavorable to the training of the algorithm model [19]. Therefore, it is necessary to convert the data into standard data that can be recognized by the model in the data preprocessing stage. This can ensure the validity and precision of the model in the subsequent modeling process.

Data preprocessing includes missing values, outliers and data normalization. The missing value refers to the incomplete data caused by the lack of some data information in the original data [20]. Research the deletion of features with high and low missing values; For features with a moderate proportion of missing values, the missing feature is used as the target label; In this study, the RF model is trained by using non-missing features as variables. Abnormal value refers to the individual value in the sample that significantly deviates from its numerical range. In the study, the univariate outlier detection method is used to draw the boxplot of characteristic variables, as shown in Figure 3.2.

The distribution of variables can be intuitively understood according to the box diagram. The experiment can further deal with the abnormal value of the characteristic variable by combining the meaning and type of the variable. Data normalization is to accelerate the convergence. The Min-Max method is applied in the study

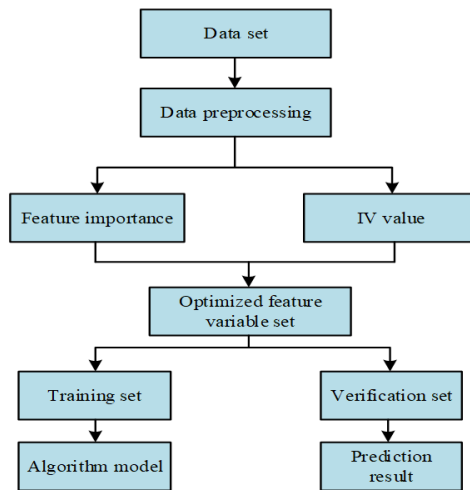


Fig. 3.1: Model building framework

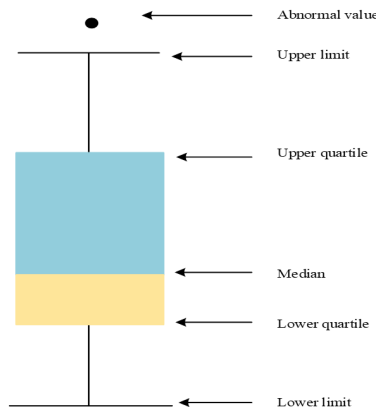


Fig. 3.2: Example of the Box-plot

to convert the original value of the variable into [0,1] standardized value, which is defined as formula 3.1.

$$x_j^i = \frac{v_j^i - \min_{1 \leq i \leq M}(v_j^i)}{\max_{1 \leq i \leq M}(v_j^i) - \min_{1 \leq i \leq M}(v_j^i)} \tag{3.1}$$

In formula 3.1, suppose x_j^i denotes the normalized value of the i -th sample on the j -th characteristic variable; v_j^i indicates the value of the i -th user on the j -th characteristic variable; M indicates the total number of samples.

Feature selection is an essential part, whose purpose is to seek the optimal feature subset. On the one hand, feature selection reduces the dimension of features and enhances the training efficiency and accuracy of the model. On the other hand, it alleviates the problem of over-fitting and has a significant impact on advancing the generalization ability of the model. This study design a multi-dimensional feature selection method, which can comprehensively consider feature importance and IV value, and score feature variables from the perspective of model and statistics. Then weaken the influence of data distribution on feature selection, and select features with strong prediction ability. The research uses Gradient Lifting Decision Tree (GBDT) to calculate feature

importance. Definition $X = \{x_1, x_2, x_3, \dots, x_n\}$ is the set of characteristic variables after processing missing values and outliers. Input into the GBDT model to obtain the square loss caused by node splitting after fitting the data. Take the square loss value as the importance score of each feature $C_j, j \in \{1, 2, 3, \dots, n\}$.

IV is applied to describe the significance contribution of each feature, and the feature set can be sorted according to the importance of the feature. The calculation of IV value depends on WOE (Weight of Evidence) value. The calculation of WOE value needs to group the data to make the feature discrete. In the experiment, the strategy of frequency splitting is adopted for continuous variables; Specify the threshold for other variables to implement the unpacking policy. The WOE value of each group can be calculated after unpacking, and the formula definition is shown in formula 3.2

$$WOE = \ln \left(\frac{Bad_i}{Bad_T} \div \frac{Good_i}{Good_T} \right) = \ln \left(\frac{Bad_i}{Bad_T} \right) - \ln \left(\frac{Good_i}{Good_T} \right) \tag{3.2}$$

In formula 3.2, taking the number of words memorized as an example, Bad_i indicates that the number of words memorized per day in the current variable group label is more than 50; i indicates that the label is the total number of words recited per day that does not exceed 50. indicates that in group i , the label is the number of samples with total recited words exceeding 1000; $Good_T$ Indicates the total number of labels with more than 1000 samples. WOE transform makes the feature not only have numerical attributes, but also reflect the weight of grouping. After obtaining the WOE value of the sub-container, formula 3.3 indicates the calculation of IV value.

$$IV_i = \left(\frac{Bad_i}{Bad_T} - \frac{Good_i}{Good_T} \right) \cdot WOE_i = \left(\frac{Bad_i}{Bad_T} - \frac{Good_i}{Good_T} \right) \cdot \ln \left(\frac{\frac{Bad_i}{Bad_T}}{\frac{Good_i}{Good_T}} \right) \tag{3.3}$$

The IV value ensures that the result is non-negative based on the WOE value. According to the IV value of the variable in each group, the calculation of the IV value of the entire variable is shown in formula 3.4

$$IV = \sum_{i=1}^n IV_i \tag{3.4}$$

The larger the IV value, the greater the role of this variable in distinguishing user learning patterns and the stronger the predictive ability. Calculate the IV value of each characteristic variable and define it as $D_j, j \in \{1, 2, 3, \dots, n\}$. To eliminate the influence of numerical measurement, it is necessary to substitute the feature importance C_j and feature IV value D_j into formula (2.4) for Min-Max normalization. Get the feature importance score $S(C_j)$ and IV value score $S(D_j)$ of feature j . Add the two to get the comprehensive score F_j of the variable. The research adopts the cumulative contribution of comprehensive score greater than 85% as the feature screening criteria, that is, in the data set $X = x_1, x_2, x_3, \dots, x_n$ composed of n features, there is a feature subset composed of k features. This subset makes F in X' formula 3.5(2.5) hold when it is greater than 85%; X' is the optimized feature variable set.

$$S(x_j) = \frac{x_j - \min_{1 \leq j \leq n}(x_j)}{\max_{1 \leq j \leq n}(x_j) - \min_{1 \leq j \leq n}(x_j)} \tag{3.5}$$

$$F = \frac{\sum_{j=1}^k F_j}{\sum_{j=1}^n F_j} \tag{3.6}$$

4. Stacking integration algorithm and its optimization. Stacking is an integrated learning algorithm that adopts a hierarchical model fusion strategy. It usually consists of a two-layer structure, as shown in Figure 4.1. The first layer consists of multiple base learners; The second layer consists of a meta-learner. The traditional Stacking algorithm fuses multiple weak learners through a two-layer layout to form a strong learner. However, the multiple base learners obtained by K-fold cross validation are different due to different training

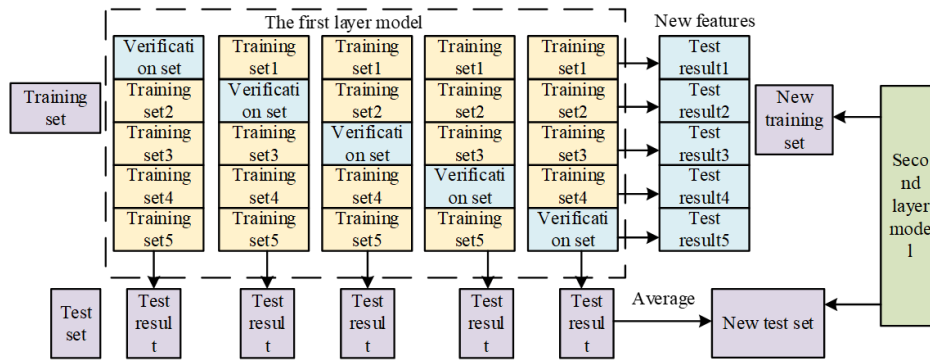


Fig. 4.1: Schematic diagram of traditional Stacking algorithm training

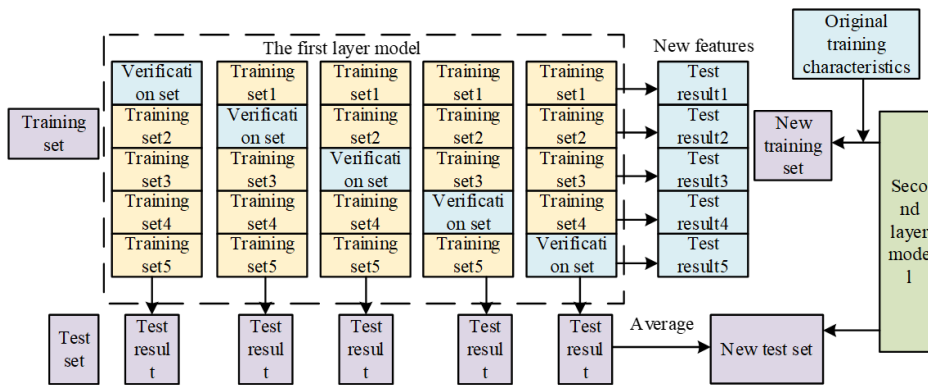


Fig. 4.2: Schematic diagram of traditional Stacking algorithm training

sets. There may be differences in the prediction performance, and it may make the base learners with good performance cover up the base learners with poor performance.

Under the current research background, when building the Stacking learning pattern recognition model, base learners' ability to distinguish data sets is particularly important. Therefore, higher weight should be given to the features produced by the base learners with stronger discrimination ability. This study uses KS value measurement model to distinguish the strength of the ability. When generating the training set and test set of the second layer through K-fold cross validation of the first layer learner, the KS value of each fold model to the verification set is calculated; The KS value is used as the weight value to weight the prediction results of the verification set and the test set, and then form the second layer of data set that has been differentiated and weighted. Meanwhile, to prevent over-fitting, the experiment expanded the second layer input features, and combined the original features with the differentiated weighted features as the input of the second layer meta-learning machine. This method can also enhance the generalization ability. Its principle is displayed in Figure 4.2.

For the selection of improved Stacking integration algorithm model, the first-level base learner selects the integration model XGBoost and RF with strong prediction ability. The second layer selects the LR model with simple structure and strong explanatory power. Because the selected model has many super parameters, the above learner's parameters can be optimized by grid search. A set of parameter combinations with the highest AUC value of the above three models under the 50% cross-validation were determined. Figure 4.3 illustrates the total structure.

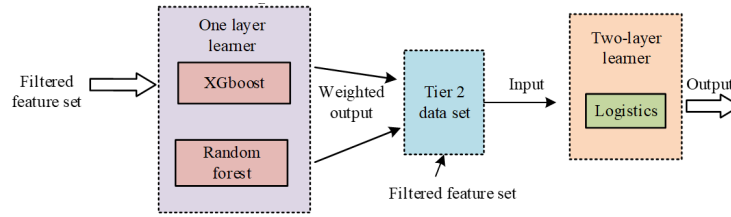


Fig. 4.3: Structure of the personal credit scoring model

The research briefly introduces the specific methods used. XGBoost algorithm is an integrated tree model for distributed implementation. It owns the characteristics of fast running velocity, high prediction precision and not easy to over-fit. XGBoost is composed of multiple CART trees. The latter tree will fit the residuals of all the previous trees, and finally output the cumulative sum of the prediction results of all the trees. In the training stage, the latter CART tree completes the training based on the previous CART tree. Suppose the sample data set is (x_i, y_j) , where $x \in R^m, y \in R$; x belongs to feature vector; y belongs to the label value. Define XGBoost model as shown in formula 4.1.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \tag{4.1}$$

In formula 4.2, \hat{y}_i represents the predicted value of the i sample; K indicates the number of current CART trees; x_i represents the characteristic vector of the i -th sample; f_k represents the predicted value of the i -th sample on the k -th tree. The current sample's output is the sum of all the predicted values of the CART tree. The weight vector w and mapping relationship q of leaf nodes in the CART tree are defined as formula 4.2 .

$$f_i(x) = w_q(x) \tag{4.2}$$

In formula 4.4, w denotes a one-dimensional vector with length T , representing the weight of each leaf node of the tree q ; q denotes the structure of a tree. The target function generated by each tree in XGBoost is defined as formula 4.3.

$$\text{Obj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \tag{4.3}$$

In formula 4.5, the former term denotes the loss function, and the latter term denotes the complexity of the tree. The number of current leaf nodes and L2 normal form constitute the complexity of the tree, as defined in formula 4.4.

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \tag{4.4}$$

The objective function of the i -th tree is shown in formula 4.5.

$$\text{Obj}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_i) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)} + f_t(x_i)) + \Omega(f_i) + \text{constant} \tag{4.5}$$

Split the regular items. Because the structure of the previous t tree has been decided, the complexity of the previous $t - 1$ tree can be expressed as a constant *constant* ; So in the above formula, only $f_t(x_i)$ is a variable,

and the rest are known quantities that can be calculated. Carry out second-order Taylor expansion of formula 4.5 to obtain the result as shown in formula 4.6.

$$\begin{cases} Obj^{(t)} \approx \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_i) + \text{constant} \\ g_i = \partial_{\hat{y}_{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \\ h_i = \partial_{\hat{y}_{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)}) \end{cases} \quad (4.6)$$

In formula 4.6, g_i and h_i are the first and second order partial derivatives of the loss function to $y^{(t-1)}$; Then divide all samples x_i with the same category as the x_i leaf node into the same sample set, and the mathematical expression is shown in formula 4.7.

$$I_j = \{i \mid q(x_i) = j\} \quad (4.7)$$

Substitute the definition formula and complexity formula of the above tree into formula 4.8 to obtain formula 4.8.

$$Obj^{(t)} \cong \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_i) = \sum_{j=1}^T [\sum_{i \in I_j} g_i w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] = \gamma T \quad (4.8)$$

In formula 4.8, $\sum_{i \in I_j} g_i$ and $\sum_{i \in I_j} h_i$ are equivalent to the total of the first and second partial derivatives of the samples contained in the leaf node j , which are two constants. Therefore, the above equation is a univariate quadratic equation. Wherein, w_j is an independent variable; $Obj^{(t)}$ is a dependent variable. According to the maximum formula, the leaf node j is optimal w_j^* as shown in formula 4.9.

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{-2 \times \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda)} = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (4.9)$$

After obtaining the optimal leaf node w_j^* , the optimal objective function can be obtained for the fixed tree structure, as shown in equation 4.10.

$$Obj^{(s)} = -\frac{1}{2} \sum_{j=1}^T \left(\frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \right) \gamma T \quad (4.10)$$

The value of the above formula is negatively correlated with the prediction ability of the model; The smaller the value is, the stronger the prediction ability of the model is. GBDT is the representative model of integration method, and its principle is shown in Figure 4.4.

GBDT will sum the gain generated by node splitting of CART tree, calculate its maximum correlation and minimum redundancy, so as to filter features of features. Let the importance of feature j be J_j^2 , as shown in formula 4.11.

$$\hat{J}_j^2 = \frac{1}{N} \sum_{n=1}^N \hat{J}_j^2(T_n) \quad (4.11)$$

In formula 4.11, N is the number of regression trees. Assuming $J_j^2(T_n)$ is the importance of the feature j in a tree, the calculation process is written in formula 4.12.

$$\hat{J}_j^2(T_n) = \sum_{t=1}^{L-1} \hat{h}_t^2 l(v_t = j) \quad (4.12)$$

In formula 4.12, $L - 1$ is the number of non-leaf nodes contained in the CART tree; v_t and h_t^2 represent the characteristics associated with node t and the loss of the node in the splitting process.

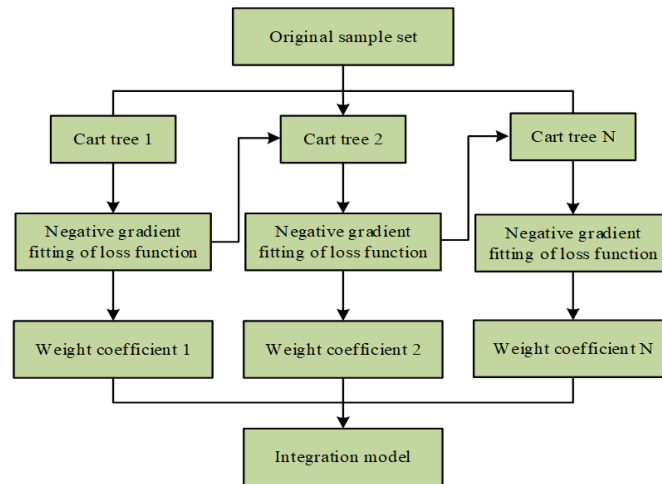


Fig. 4.4: Schematic diagram of GBDT

The RF uses the method of sampling back to randomly select data and features. It consists of multiple CART decision trees in structure. Each decision tree is regarded as a weak learner. Assume that the total number of samples in the original dataset D is n . It contains M characteristics. The number of decision trees needed to build a RF model is k . The calculation of RF contains the extraction of training subset, the decision tree's construction, the RF's generation, the classification and prediction test set and the output of final results. The prediction result is calculated as shown in formula 4.13

$$T(x) = \arg \max \sum_{i=1}^k I(t_i(x) = y) \quad (4.13)$$

In formula 4.13, i is the predicted result of the $t_i(x)$ test set. The specific indicators used in this study include AUC value, ROC curve and accuracy. AUC measures the relationship between the sensitivity of mobile users' English learning and the false positive rate under different thresholds, and its value ranges from 0 to 1, and 0.5 indicates that the working state of the recognition model at this time is random guess. The closer the AUC value is to 1, the better the performance of the model, which is a commonly used index to evaluate the performance of the model. ROC curve is a graphical representation method to evaluate the performance of the model. It draws the curve by changing the threshold of mobile users' English learning status. On the ROC curve, each point represents a different point in a specific learning mode. Each point on the curve corresponds to a different performance of the English learning recognition model, and the area under the curve represents the overall performance of the model. By observing ROC curve, we can choose the best threshold to achieve the best performance of mobile users' English learning pattern recognition model. In order to evaluate the effectiveness of low-cost digital manufacturing system, this study first draws the cost-benefit of the traditional manufacturing system according to ROC curve, and compares the cost and return of the two systems; Then calculate their differences in production efficiency. At the same time, it is judged whether the production efficiency of the manufacturing system reaches the standard through the production capacity indexes of the two systems. For the effectiveness of low-cost digital manufacturing system in quality control, this paper compares the product quality and defects of the two systems, and then evaluates the control ability of low-cost digital manufacturing system. When the low-cost digital manufacturing system is superior to the traditional system in terms of production flexibility, it shows that the low-cost digital manufacturing system can provide higher degrees of freedom. In addition, the study also evaluates the environmental sustainability of low-cost digital manufacturing systems. In this step, the energy consumption difference between the two systems is compared.

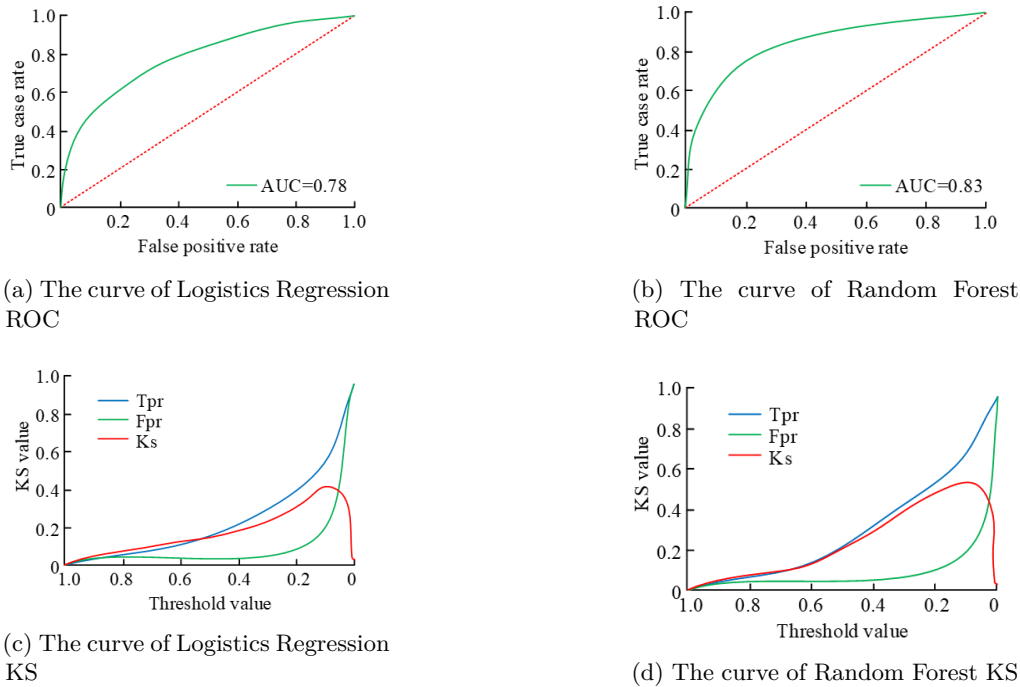


Fig. 5.1: ROC, AUC and KS values of Logistics regression model and RF model

5. Result analysis.

5.1. Model prediction performance analysis. The research carries out experimental analysis on the learning recognition model. The experimental evaluation indexes include AUC value, KS value and accuracy. The optimized data set is divided into training set and test set. Wherein 70% as training set; 30% as test set. The training set is used as the input of LR model and RF model. In the training process, the grid search method is applied to adjust the parameters. In this study, the source of data is online learning platform. This data set contains various behaviors and interactive data of users in the process of learning English. The purpose of this study is to identify different learning modes by analyzing the learning behavior and learning data of mobile users. Firstly, the collected learning data set is divided into training set and test set, and then the integrated learning method is used to select the best parameter combination of training set. Finally, the test set is used to evaluate the trained model. Through this evaluation method, the experimental results of model improvement are obtained. Feature importance and IV value are commonly used feature selection methods to evaluate the predictive ability of user behavior characteristics to English learning patterns. The importance of features measures the contribution of features in the model, while the IV value measures the prediction ability of features to target variables. When using feature importance and IV value for feature selection, the study evaluates the predictive ability of user behavior characteristics to English learning patterns according to the size of feature importance and IV value.

In this study, the test set will be predicted, and the ROC curve, AUC value and KS curve of the Golden Sine (GS) model, Proportional Integral Derivative (PID) model, LR model and RF model will be compared with them, as shown in Figure 5.1. The AUC value of the model under the combination of the processed training set and the optimal parameters is 0.78. Figure 5.1b) shows that only when the test set threshold reaches 0.1 to 0.2 can the model have good discrimination; That is, within this range, KS value reaches the maximum value, which is 0.427. Combined with the confusion matrix data, the prediction result of LR model is low. The classification accuracy of the output results of a large number of samples is low. Therefore, when the threshold

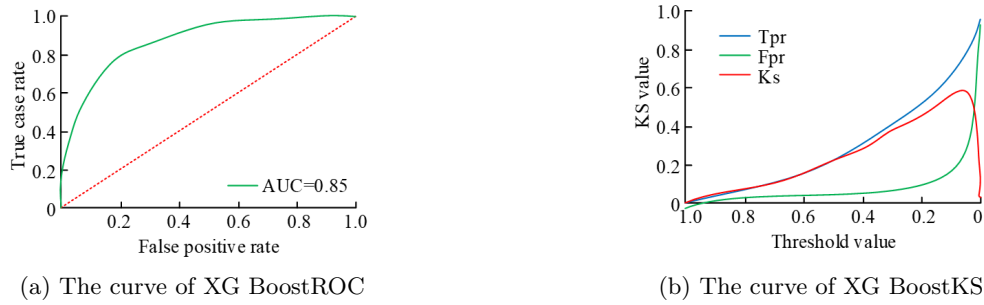


Fig. 5.2: ROC, AUC and KS values of XGBoost

Table 5.1: Confusion matrix for Stacking

Model	Correctly predict the result											
	Login days			Daily average login times			Recitation amount			Daily average reciting vocabulary		
	<30	30-60	60<	<3	3-5	5<	<200	200-500	500<	<20	20-50	50<
Traditional Stacking	35657	52397	12508	65144	28779	9541	41212	58194	9629	71758	37697	5914
Improved Stacking	35897	52482	12585	65235	28946	9652	41256	58426	9714	71912	37764	5976

set by the model is greater than 0.2, the judgment result of the model is not reliable. ROC curve, AUC value and KS value after RF model test are shown in Fig. 5.1c and 5.1d. The model's AUC value after parameter adjustment and data processing is 0.83; The KS value is 0.534, and the probability threshold of the maximum KS is between 0.05 and 0.15. The ROC curve and KS value also show that the stochastic forest model performs better discrimination than the Logistic model.

As shown in Figure 5.2, the model's AUC value is 0.85 and the KS value is 0.553. When the threshold reaches the range of 0 to 0.2, the KS value reaches the maximum. Compared with LR model and RF model, XGBoost model has better performance in classifying samples. In addition, the sample prediction probability distribution of XGBoost is more uniform than that of LR model. The distribution results are also more reasonable, indicating that the model has better discrimination.

Conduct grid search on the Stacking model, and input the training set into the two kinds of Stacking models for training. The experiment predicts the test set, and the resulting confusion matrix is shown in Table 5.1. In the improved Stacking model, 37764 samples were correctly identified on the samples with more than 20 but not more than 50 memorized words per day; In the case that the number of words recited per day is less than 20, 71912 samples were correctly identified; On the samples with more than 50 words memorized daily, it correctly identified 5976. Compared with the confusion matrix of the traditional Stacking model, the improved Stacking model has a stronger ability to distinguish user learning patterns. According to the formula, the accuracy of the improved Stacking model is 91.29%; The accuracy of traditional Stacking model is 90.71%.

ROC curve and AUC value are shown in Figure 5.3. The ROC curve of the improved Stacking model is smoother than that of the three single models; Its AUC value is 0.85, which is the same as XGBoost; The improved Stacking model has better performance than the traditional Stacking model, LR and RF model. The KS value of the model is 0.583; When the threshold is about 0.1, KS gets the maximum value.

5.2. Comparative analysis of models. LR, RF, XGBoost and traditional Stacking models were selected as the baseline models for the performance evaluation of the optimization model; The indicators for comparison include AUC, KS and accuracy.

As shown in Table 5.2, the AUC value, KS value and precision of the improved Stacking model built in the study are 0.85, 0.582 and 91.29% respectively. The improved Stacking model has the highest prediction

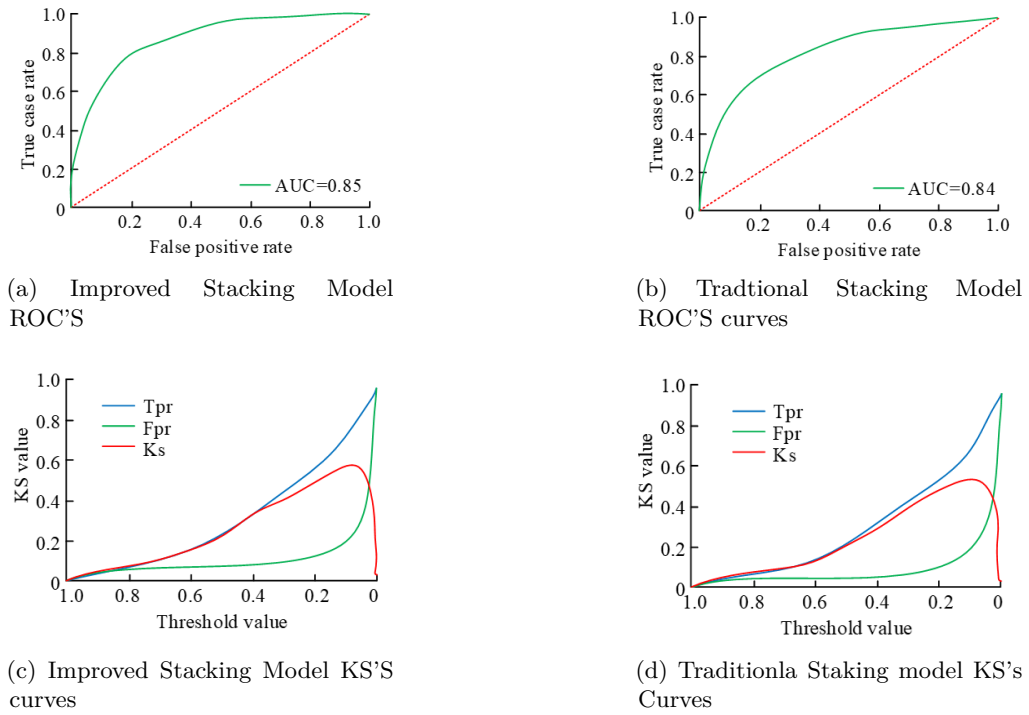


Fig. 5.3: Comparison between improved Stacking model and traditional Stacking model

Table 5.2: Comparison of integrated model evaluation indicators

Model	AUC	KS	Accuracy rate
LR	0.78	0.426	89.46%
RF	0.83	0.533	90.04%
XGBoost	0.85	0.553	90.11%
Traditional Stacking	0.84	0.556	90.71%
Improved Stacking	0.85	0.582	91.29%

accuracy; Compared with the single model, it has improved to different degrees. On the AUC value of the indicator, the improved Stacking model and XGBoost model are the same; Compared with traditional Stacking model, LR and RF model, it has been greatly improved. In terms of the key indicator KS value, the improved Stacking model also has a significant improvement compared with other models. From the comprehensive index data, compared with the traditional Stacking model and the single model, the improved Stacking model built for user learning mode has better discrimination and performance. This verifies the effectiveness of the model in the field of online English education APP user classification.

To verify the effectiveness of the multi-dimensional feature screening strategy, the experiment input the pre screening data into the single model, the traditional Stacking model and the improved Stacking model for training respectively. In the experiment, the AUC value, KS value and accuracy rate are also applied. Table 5.3 demonstrates the final results. From Table 5.3, the AUC, KS and precision of the data set before screening have decreased compared with the data after screening after training by inputting the data set before screening into Logistics, RF, XGBoost, Traditional Stacking Model and Improved Stacking Model. In particular, the key indicators KS value and AUC value have significantly decreased. This shows that the feature filtering strategy

Table 5.3: Pre-feature screening model indicators

Model	AUC	KS	Accuracy rate
LR	0.76	0.404	87.75%
RF	0.80	0.515	89.87%
XGBoost	0.81	0.534	89.16%
Traditional Stacking	0.81	0.530	89.71%
Improved Stacking	0.82	0.552	90.34%

proposed by the research can be used to optimize the recognition of user learning patterns [21, 22, 23], and it has a good effect on improving the model discrimination and accuracy.

6. Conclusion. Recently, learning behavior analysis based on mobile platform data has attracted more and more attention. This study uses feature importance and IV value to grade feature variables layer by layer. Meanwhile, the experiment adopts the integrated learning method, and improves the Stacking integration algorithm with the differentiation weighting and feature extension combined with the recognition of learning patterns. The experiment compared the effect of the model trained by the data set before and after screening, and found that the AUC value, KS value and prediction accuracy of the model built after the screening strategy designed by the research institute were improved. The results demonstrated that the strategy is efficient in improving the performance of the credit scoring model. The comprehensive experimental results indicate that compared with the single model and the traditional Stacking model, the improved Stacking model has advancement in AUC value, KS value and prediction accuracy; The AUC value is 0.85, the KS value is 0.582, and the accuracy rate is 91.29%. Therefore, this method is more effective than the previous methods and can provide reference for mobile users' English learning pattern recognition. However, there are few characteristic variables screened out in the current study. Further mining is needed in the future to obtain more accurate analysis models. This study is only aimed at online users' English learning pattern recognition, and the research on using social networks and social learning theory is still insufficient. Incorporating social factors into mobile users' English learning pattern recognition model can analyze users' interactive behavior in social networks and the influence of learning communities, so as to promote cooperative learning and knowledge sharing among users. This study is also of great significance to the study of mobile users' English learning mode, and will be carried out gradually in future research.

Fundings. The research is supported by: The Ministry of Education's Industry-school Cooperative Education Project in 2022: The construction of blended College English teaching online system based on Unipus, Subject Number 220606441100534.

REFERENCES

- [1] Pantelimon, F., Georgescu, T. & BŞ, P. The impact of mobile e-commerce on gdp: A comparative analysis between romania and germany and how covid-19 influences the e-commerce activity worldwide. *Informatica Economica*. **24**, 27-41 (2020)
- [2] Papadakis, S., Kalogiannakis, M. & Zaranis, N. Teaching mathematics with mobile devices and the Realistic Mathematical Education (RME) approach in kindergarten. *Advances In Mobile Learning Educational Research*. **1**, 5-18 (2021)
- [3] Metaverse, Z. & China, I. InMetaverse–METAVERSE 2022: 18th International Conference, Held as Part of the Services Conference Federation, SCF 2022, Honolulu, HI, USA, December 10–14, Cham: Springer Nature Switzerland.2022. *Proceedings*. **12**, 103-110 (2022)
- [4] Önal, N., Çevik, K. & Şenol, V. The effect of SOS table learning environment on mobile learning tools acceptance, motivation and mobile learning attitude in English language learning. *Interactive Learning Environments*. **30**, 834-847 (2022)
- [5] Hoang, Q., Pham, T., Dang, Q. & Nguyen, T. Factors influencing Vietnamese teenagers' intention to use mobile devices for English language learning. *In*. **12**, 230-245 (0)
- [6] Alam, A. Platform Utilising Blockchain Technology for eLearning and Online Education for Open Sharing of Academic Proficiency and Progress Records. *InSmart Data Intelligence: Proceedings Of ICSMDISingapore: Nature Singapore.2022,8(18)*. pp. 307-320 (2022)
- [7] Haldar, M., Abdool, M., Ramanathan, P., Xu, T., Yang, S., Duan, H., Zhang, Q., Barrow-Williams, N., Turnbull, B., Collins, B. & Legrand, T. Applying deep learning to airbnb search. *Inproceedings Of The 25th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining7(25)*. pp. 1927-1935 (2019)

- [8] Lee, C., Tzeng, J., Huang, N. & YS., S. Prediction of student performance in massive open online courses using deep learning system based on learning behaviors. *Educational Technology & Society*. **24**, 130-146 (2021)
- [9] Shorfuzzaman, M., Hossain, M., Nazir, A., Muhammad, G. & Alamri, A. Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. *Computers In Human Behavior*. **3**, 578-588 (2019)
- [10] Pei, C., Sun, Y., Zhu, J., Wang, X., Zhang, Y., Zhang, S., Yao, Z. & Lu, Q. Ensemble learning for early-response prediction of antidepressant treatment in major depressive disorder. *Journal Of Magnetic Resonance Imaging*. **52**, 161-171 (2020)
- [11] Vogelstein, L., Brady, C. & Hall, R. Reenacting mathematical concepts found in large-scale dance performance can provide both material and method for ensemble learning. *ZDM*. **5**, 331-346 (2019)
- [12] Zhang, J., Wang, Y., Sun, Y. & Li, G. Strength of ensemble learning in multiclass classification of rockburst intensity. *International Journal For Numerical And Analytical Methods In Geomechanics*. **44**, 1833-1853 (2020)
- [13] Hong, H., Jiang, C., Peng, X. & Zhong, W. Concurrent monitoring strategy for static and dynamic deviations based on selective ensemble learning using slow feature analysis. *Industrial & Engineering Chemistry Research*. **59**, 4620-4635 (2020)
- [14] Nuankaew, W., Bussaman, S. & Nuankaew, P. Evolutionary Feature Weighting Optimization and Majority Voting Ensemble Learning for Curriculum Recommendation in the Higher Education. *In Multi-disciplinary Trends In Artificial Intelligence: 15th International Conference, MIWAI 2022, Virtual Event, November, 2022, Proceedings. Cham: International Publishing 11(10):14-25*. pp. 17-19 (2022)
- [15] Tan, Z., Zhang, J., He, Y., Zhang, Y., Xiong, G. & Liu, Y. Short-term load forecasting based on integration of SVR and stacking. *IEEE Access*. **12**, 27719-22772 (2020)
- [16] Meharie, M., Mengesha, W., Gariy, Z. & RN., M. Application of stacking ensemble machine learning algorithm in predicting the cost of highway construction projects. *Engineering, Construction And Architectural Management*. **29**, 2836-2853 (2022)
- [17] Acuna, Y. & Sun, Y. An efficiency-improved genetic algorithm and its application on multimodal functions and a 2D common reflection surface stacking problem. *Geophysical Prospecting*. **68**, 1189-1210 (2020)
- [18] Youssef, M., Mohammed, S., Hamada, E. & Wafaa BF., A. predictive approach based on efficient feature selection and learning algorithms' competition: Case of learners' dropout in MOOCs. *Education And Information Technologies Nov.* **24** pp. 3591-3618 (2019)
- [19] Youssef, M., Mohammed, S., Hamada, E. & Wafaa BF., A. predictive approach based on efficient feature selection and learning algorithms' competition: Case of learners' dropout in MOOCs. *Education And Information Technologies*. **11**, 3591-3618 (2019)
- [20] Saeed, S., Haron, H., Jhanjhi, N., Naqvi, M., Alhumyani, H. & Masud, M. Improve correlation matrix of discrete fourier transformation technique for finding the missing values of mri images. *Mathematical Biosciences And Engineering*. **19**, 9039-9059 (2022)
- [21] Lee, S., Abdullah, A. & Jhanjhi, N. A review on honeypot-based botnet detection models for smart factory. *International Journal Of Advanced Computer Science And Applications*. **11** (2020)
- [22] Azeem, M., Ullah, A., Ashraf, H., Jhanjhi, N., Humayun, M., Aljahdali, S. & Tabbakh, T. Fog-oriented secure and lightweight data aggregation in iomt. *IEEE Access*. **9** pp. 111072-111082 (2021)
- [23] Gaur, L., Solanki, A., Wamba, S. & Jhanjhi, N. Advanced AI techniques and applications in bioinformatics. (CRC Press, 2021)

Edited by: Mudasir Mohd

Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions

Received: Nov 8, 2023

Accepted: Jan 10, 2024