Retraction

Retracted: The Prediction of English Online Network Performance Based on the XGBoost Algorithm

Mobile Information Systems

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Chen, "The Prediction of English Online Network Performance Based on the XGBoost Algorithm," *Mobile Information Systems*, vol. 2022, Article ID 9263736, 11 pages, 2022.

Research Article

The Prediction of English Online Network Performance Based on the XGBoost Algorithm

Xiaohua Chen

Zhanjiang University of Science and Technology, Zhanjiang 524000, Guangdong, China

Correspondence should be addressed to Xiaohua Chen; chenxiaohua@zjkju.edu.cn

The demand of C language informatization for English learning has increased greatly, and strengthening the application of informatization in teaching has become a current trend, while deep learning algorithms have been applied in various tasks due to their obvious advantages. In this paper, an English score prediction method based on the XGBoost algorithm is proposed. In order to verify the effectiveness of the model in English score prediction, the principle of the XGBoost algorithm is firstly analyzed as a basis. The English test scores of a university for 2019–2021 were used as the basic data source, and the output probabilities of the proposed model were used to compare the results under different conditions. The experimental results show that the predicted scores are basically consistent with the actual scores. From a scientific point of view, the ability to predict unknown data with low error suggests that it enables students and teachers to identify the underlying factors that make it difficult for students to answer questions. Understanding these causes is useful for designing high-quality courses and lesson plans.

1. Introduction

With the development of the current application of informatization in teaching, it has become a current trend to strengthen the application of informatization in teaching [1]. For example, Wang et al. [2] applied data mining algorithms to the evaluation of English teachers' abilities, so as to obtain the necessary abilities in English teachers' teaching, others applied the data mining algorithm to the adult English test and obtained some rules for passing the adult English test. With the application of data mining, its drawbacks have also begun to be exposed, such as the amount of data to be mined is small and the mining accuracy is not high [3]. In this regard, people began to try to apply the XGBoost algorithm to the English subject field. For example, the real scene data of the large-scale unified test of spoken English as the basis and used the XGBoost algorithm model to recognize the speech and then conduct the evaluation. The results show that the method has strong robustness in dealing with noise; Wang et al. [4] applied the XGBoost algorithm to English writing, processed the text by natural processing language, and used the XGBoost algorithm to mine, thereby improving

English writing ability. It can be seen from the above research that deep learning algorithms are widely used in English research [5].

XGBoost is the abbreviation of eXtreme Gradient Boosting [6]. On the basis of massively parallel boosted tree, XGBoost has played an important role as its extension tool and has become the best speed in the current open source boosted tree toolkit. The fastest algorithm, which is more than 10 times faster than common toolkits [7], is widely used on the crowdsourcing platform Kaggle, and a large number of contestants use it in various data mining competitions. The XGBoost algorithm is used in the winning plan of several Kaggle competitions. In the practical application of real business, XGBoost's portability is widely used and still retains local. In addition, there are other efficiency improvement methods, so that it can achieve good results under the huge scale of the industry [8]. XGBoost implements the general tree boosting algorithm by engineering, the gradient boosting decision tree (GBDT) is the representative of the tree boosting algorithm, and it is also called MART (Multiple Additive Regression Tree) [9].

After obtaining high-quality data, this paper analyzes the original attributes of the data and combines relevant knowledge in the professional field to design and generate high-quality features, which will have a direct impact on the model prediction effect [10]. Therefore, the work of data preprocessing and feature construction will occupy 90% of all the time of the system. Next, use the XGBoost algorithm to model the data, adjust its parameters and then fuse the above three models to generate a high-accuracy prediction model. The feasibility of predicting students' answering results by using historical data is possible, and the prediction accuracy is extremely high [11]. From a scientific point of view, the ability of predicting unknown data with low error has demonstrated that it enables students and teachers to discover the underlying factors of students' difficulty in answering questions. Understanding such reasons can be greatly beneficial for designing high-quality courses and lesson plans [12].

2. Related Work

Educational data mining technology has been weak at the beginning; through years of research and innovation by researchers, this technology has made great progress and has been popularized in the field of foreign education [13]. In 2007, an article published by a scholar described the broad application space of machine learning methods in the field of education, and that article was widely recognized by the academic community [14]. Li et al. [15], in order to find the law of improving students' performance, used rough set theory to combine with problems in the field of education and found a method that can dominate students' motivation. Li et al. [16] analyzed the records of the log system of the education platform and generated a model that can predict the students' performance through the students' behavior and record other scholars in order to infer the newly enrolled graduate students. The candidates' future academic performance was analyzed and modeled on their undergraduate grades, and their GGPA1 was successfully predicted. From the above content, it can be seen that foreign countries have a profound background on the performance prediction system, and the development speed is very fast. Foreign scholars have effectively combined data mining technology with the field of education and developed many excellent technologies and projects.

In my country, the field of data mining entered the field of education relatively late, and it is still insufficient compared with foreign development. However, my country is catching up at this stage, and there are many excellent scientific research and engineering personnel who have made achievements in this field. Jia et al. [17] conducted data modeling for the training of students in this major and finally used the C4.5 algorithm to successfully find the laws of student behavior hidden behind the data. Shang et al. [18] used the K-mean algorithm in the analysis of student' test scores and applied it to the computer grading module, so that educators could obtain the mastery of students' scores and reduce the difficulty of work. In order to find the influencing factors of students' English test, scholar used the

decision tree algorithm to analyze test data and found rule sets in data mining, which provided method guidance for the improvement of the teaching effect [19]. There are still many unsolved problems waiting for these outstanding talents in our country to be studied and solved.

3. Improved Algorithm Model

XGBoost can efficiently construct augmented trees and run in parallel. There are two kinds of augmented trees in XGBoost, regression trees and classification trees [20]. Optimizing the value of the objective function is the core of XGBoost. Here, the objective function is used as an example to introduce the theory. The objective function is shown in the following equation [21]:

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} l\left(\left(y_i, \widehat{y}^{(t-1)}\right) + f_t(x_i)\right) + \Omega(f_i) + C, \quad (1)$$

where $\widehat{y}^{(t-1)}$ denotes the retention of the model predictions from the previous $t-1$ round, as shown in the following equation [22]:

$$g_i = \partial_{\widehat{y}^{(t-1)}} l\left(y_i, \widehat{y}^{(t-1)}\right) h_i = \partial^2_{\widehat{y}^{(t-1)}} l\left(y_i, \widehat{y}^{(t-1)}\right). \quad (2)$$

It can be changed to the form of the following equation [23]:

$$\text{Obj}^{(t)} \approx \sum_{i=1}^{n} \left[l\left(y_i, \widehat{y}^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_i^2(x_i)\right] + \Omega(f_i) + C. \quad (3)$$

When the model is trained, the objective function can be expressed by [23]

$$\text{Obj}^{(t)} = \sum_{j=1}^{t} \left[\left(\sum_{i \in I_j} g_i\right) w_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda\right) w_j^2\right] + \gamma T. \quad (4)$$

Defining the equation as

$$G_j = \sum_{i \in I_j} g_i, H_j = \sum_{i \in I_j} h_j. \quad (5)$$

Bringing (5) into (4), we get the following equation:

$$\text{Obj}^{(t)} = \sum_{j=1}^{t} \left[G_j w_j + \frac{1}{2}\left(H_j + \lambda\right) w_j^2\right] + \gamma T = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T. \quad (6)$$

The optimal combination of parameters is finally substituted into the XGBoost algorithm to improve the prediction performance [24]. After constructing the GS-XGBoost model, multi-step prediction is performed, and the model is applied and then the prediction results are compared with the original XGBoost model, GBDT model, and SVM model, and finally the model is validated according to the evaluation index. The specific experimental flow chart is shown in Figure 1.

The prediction performance evaluation indexes of the prediction model were compared with the experimental results using three evaluation indexes: mean square error
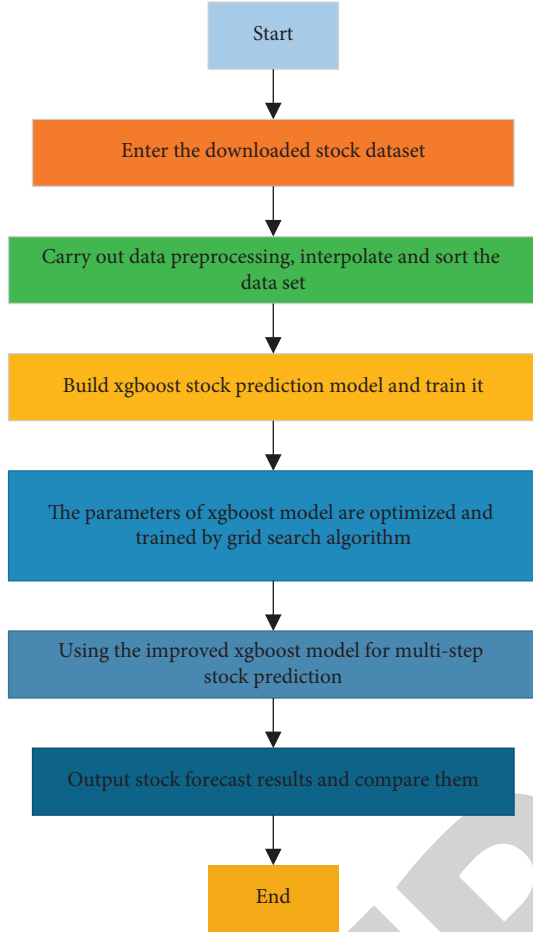
Figure 1: Experimental flow chart.

(MSE), root mean square error (RMSE), and mean absolute error (MAE). (1) Mean square error is the average of the minimization of the sum of squared errors (SSE) [26] cost function in the process of fitting the linear regression model. The better the prediction effect, the closer the value is to 0, and vice versa, the farther the value is from 0. Its calculation formula is shown in the following equation [27]:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y^{(i)} - \widehat{y}^{(i)} \right)^2. \tag{7}$$

Here, y is the true value and $\widehat{y}$ is the predicted value.

(2) The formula for calculating the root mean square error is shown in the following equation [28]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y^{(i)} - \widehat{y}^{(i)} \right)^2}. \tag{8}$$

(3) The formula for calculating the average absolute error is shown in the following equation [29]:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y^{(i)} - \widehat{y}^{(i)} \right|. \tag{9}$$

The main steps of the improved PrefixSpan algorithm (called the Im_PrefixSpan algorithm in this paper) are as follows: Step 1: Scan the sequence database S once, find all the 1st order sequences and count them. If the support of a sequence of order 1 is less than the broad value, the sequence is divided into two, and its left and right subsequences are put back into the sequence database, and the original sequence is deleted from the sequence database. For each $L(L > 1)$ prefix, only the first item of the sequence in the suffix database is scanned for counting. If the support count is lower than the queue value, the sequence corresponding to the first item is removed from the suffix database and the expansion of the first item is stopped. Step 2: Combine the first items that satisfy the support count and the current prefixes to obtain some new prefixes. Step 3: Make $L = L + 1$, scan the current suffix database and construct the corresponding suffix database with the new prefixes. Return to step 3 until the suffix database is empty.

In this genetic algorithm, the genetic operators use constant crossover probabilities as well as variation probabilities. This is more effective for simple optimal problems but has disadvantages for complex problems. The disadvantage is that it can lead to early "premature" and slow convergence, and the final result can easily fall into local optimum. The crossover probability and variation probability can be changed in time, and the linear function is used for adaptive change adjustment, which can effectively solve the problems of premature maturity. In Srinvas' improved algorithm, the equations of the crossover rate and the variation rate are [30]

$$P_c = \begin{cases} \dfrac{k_1 \left(f_{\max} - f' \right)}{f_{\max} - f_{\text{avg}}} & f' \geq f_{\text{avg}}, \\[2mm] k_2 & \text{other.} \end{cases} \tag{10}$$

$$P_m = \begin{cases} \dfrac{k_3 \left(f_{\max} - f \right)}{f_{\max} - f_{\text{avg}}} & f \geq f_{\text{avg}}, \\[2mm] k_4 & \text{other.} \end{cases} \tag{11}$$

Here, $f_{\max}$ and f_{avg} are the maximum fitness of the individuals in the population and the average fitness of all individuals, respectively, and f is the fitness of the individuals in the population that are about to mutate. From the intuition of (10) and (11), the variation rate and the crossover rate are adjusted for linear processing and are no longer fixed. When the fitness of an individual calculated from the fitness function is lower than the average fitness, it means that the solution represented by the individual is less effective, and then a larger evolution is performed according to the idea of the genetic algorithm, i.e., a larger crossover rate and variation rate are used. If the fitness of the individuals in the population is higher, then linear adjustment is performed according to equations (11) and (12).

The above improvements to the variance and crossover rates can significantly improve the ability of the model to find the best solution. However, there is a problem when f is equal to $f_{\max}$, at this time, according to (1) and (2), both P_m and P_c are 0, resulting in the genetic algorithm.

At the early stage of the calculation, individuals with high fitness in the population can only undergo smaller changes and are easily trapped in a local optimum. Therefore, to address this problem, this paper establishes an improved linear adaptive genetic algorithm (ILAGA), which further optimizes the crossover rate and variation rate in the genetic operator.

The following optimization is performed for P_c and P_m, and the crossover rate and the variation rate are calculated according to (12) and (13).

$$P_c = \begin{cases} cI - \dfrac{(c1 - c2)\left(f' - f_{avg}\right)}{f_{max} - f_{ar}} & f' \geq f_{avg}, \\ \\ cI & \text{other.} \end{cases} \quad (12)$$

$$P_m = \begin{cases} mI - \dfrac{(mI - m2)\left(f - f_{av}\right)}{f_{max} - f_{avg}} & f \geq f_{avg}, \\ \\ mI & \text{other.} \end{cases} \quad (13)$$

The flow chart of the improved linear adaptive genetic algorithm is shown in Figure 2. The detailed steps of the improved linear adaptive genetic algorithm based on the crossover rate and the variation rate are as follows:

Step 1: Encoding: After determining the set of parameters of the actual problem, some form of encoding is performed for the variables to be solved, and the encoding should reflect the solution space of the problem.

Step 2: Initial population generation: GA starts the selection of generations with these N string structure data as the initial point.

Step 3: According to the optimization objective of the actual problem, determine the objective function and fitness function of the problem, such as in regression, you can use RMSE as the objective function, and the inverse of RMSE as the function to calculate the fitness.

Step 4: Adaptation calculation: Substitute the individuals in the population into the objective function and fitness function of the optimization problem and calculate the fitness value of each individual. If the optimization index of the problem is satisfied or the maximum number of selected generations is reached, the solution of the problem is output, otherwise, the genetic operation of the chromosome (step 5-step 6) is continued and the population is upgraded [31–35].

Step 5: Crossover operation: The better individuals selected in the fifth step are crossed over in a certain way to produce new individuals to make the population more diverse.

Step 6: Mutation operation: Mutation is performed on some of the chromosomes after the crossover operation, i.e., some gene values of individual strings in the population are changed to further expand the diversity of the population.

4. Realization of English Test Score Prediction

XGBoost algorithm is used to predict the grades of middle school students in the college English skills training system, as shown in Figure 3.

4.1. Data Extraction and Preprocessing.
This research mainly extracts student information from the college English skill training system and selects the data of first- and second-grade students in the four academic years from 2019 to 2021 in the spring and autumn, respectively. CSV format is selected as the final type of file to store tabular data in plain text [23, 24].

4.2. Feature Selection.
Through feature processing, this experiment identified 18 important features for mining students' grade prediction, such as student number, name, gender, answering time, question type, and other 18 dimensions. Table 1 below lists some of the characteristics and data.

4.3. Prediction Results and Analysis.
In this study, the data mining regression method XGBoost algorithm model is used, and the relevant data of students' English exams in two academic years in the English skills training system are used as training data. After streamlining, the prediction model of students' grades in the college English test is finally constructed, and the prediction of students' grades is realized, so that the scores that are close to the real grades of students are obtained. The model uses 18 features as the final forming factors of XGBoost and constructs a decision tree with 6 lessons, the minimum sample leaf node is 6, and the maximum depth is 5. Table 2 shows actual and predicted scores of the data, of which the full score is 50 points.

The experimental results are evaluated by MAE. The smaller the value, the better it is. Finally, the MAE of all data sets is 0.7, and 79.86% of the data errors are 0. That is, the prediction accuracy is 79.86%. Comparing the actual score and the predicted score curve, it is found that the two curves are very similar, indicating that the predicted score is very close to the real score.

According to the English test scores of 120 samples predicted by the established prediction model equation, the predicted scores of these 120 students in the English test are compared with their actual scores. The predicted scores of some samples are almost exactly the same as the actual scores, which further shows that the predicted scores are basically reasonable and reliable and also shows the accuracy of the prediction model. Of course, there may be deviations in the predictions of individual students, and there may be many reasons for this, but basically they are all caused by objective reasons or very special subjective reasons, as shown in Figure 4.

Figure 5 shows that the average total English score of students in 2019–2021 is between 366.43 and 434.74. Except for the 2019 students, the total scores of the students in other grades are lower, indicating that the students' overall English
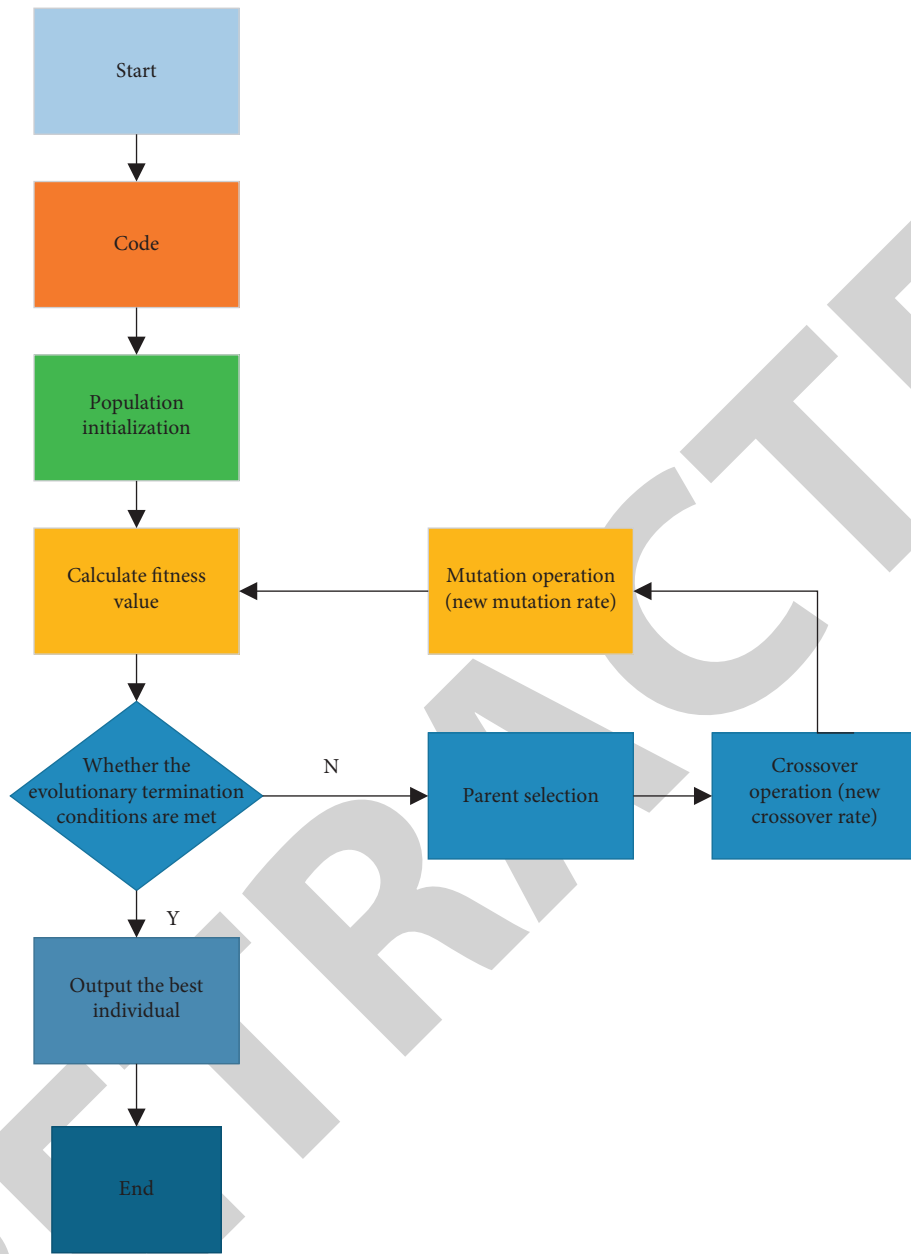
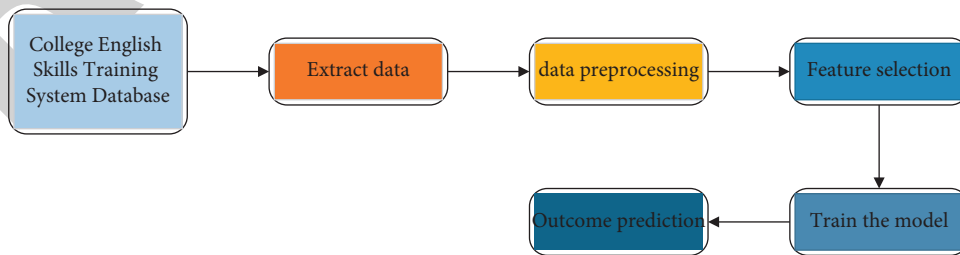FIGURE 2: Flow chart of the improved linear adaptive genetic algorithm.

FIGURE 3: The performance prediction process.

level is not high. Considering that students with scores below 220 did not participate in the statistics, the actual average grade for all students was lower. However, in the past two years, the total score has shown a rapid upward trend. Over three years, the overall score has improved by 18%. The average annual increase was 4.4%. There are many reasons for this. In addition to the continuous improvement of the cultural quality of the school's students in recent years, the

Table 1: Some characteristics of performance prediction.

Serial number	Gender	Class	Question type	Answer time	Self-rating	Question score	Passing score
1	Male	1	SAQ	301	8.6	7	6
2	Female	2	TF1	286	9.1	8	6
3	Male	1	SC	254	7.8	8	6
4	Female	3	SAQ	311	8.3	7.5	6
5	89.76	2	SW	292	6.6	6.5	6
6	Female	3	TF2	288	8.1	7	6
7	Male	4	SC	276	7.9	7.0	6
8	Female	4	SW	272	7.3	6.0	6

Table 2: Comparison of actual grades and predicted grades.

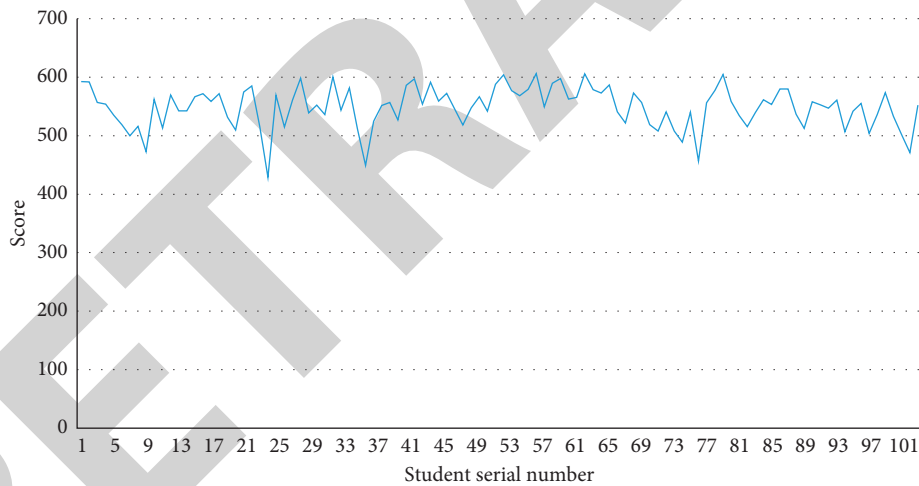| Serial number	Real grade	Predicted grades
1	41	43
2	40.5	44
3	33	35.5
4	37	35
5	32	33
6	41	41
7	36	36
8	43	38
9	38	38
10	41.5	39

Figure 4: Comparison of English scores and predicted scores.

students pay more attention to English learning and gradually adapt to new topics, and the conditions have improved significantly after the school welcomes evaluation and promotes construction. At the same time, it also shows that the level of college English teaching has continued to improve in the past two years.

The scores of listening and writing are steadily improving every year, and the listening of the 2021 students is 40% higher than that of the 2019 students. In recent years, this may be related to the obvious improvement of school pronunciation facilities and the establishment of bilingual courses in some courses of various majors. Writing results may be abnormal due to changes in assessment objectives and other reasons.

Students' reading scores have improved in recent years, especially for the class of 2021, but not yet. This is related to the high proportion of English reading and the importance that students attach to it, and it is more likely to be related to the improvement of students' reading ability due to the bilingual courses offered by schools in the past two years. The results of the comprehensive test showed a slow upward trend. This question type includes translation questions, cloze, grammar, and vocabulary. Due to the variety of question types and low proportions, teachers and students have neglected the question types.

Judging from the predicted results of the results, in the next two to three years under the condition that the examination mode, teaching form and the quality of students
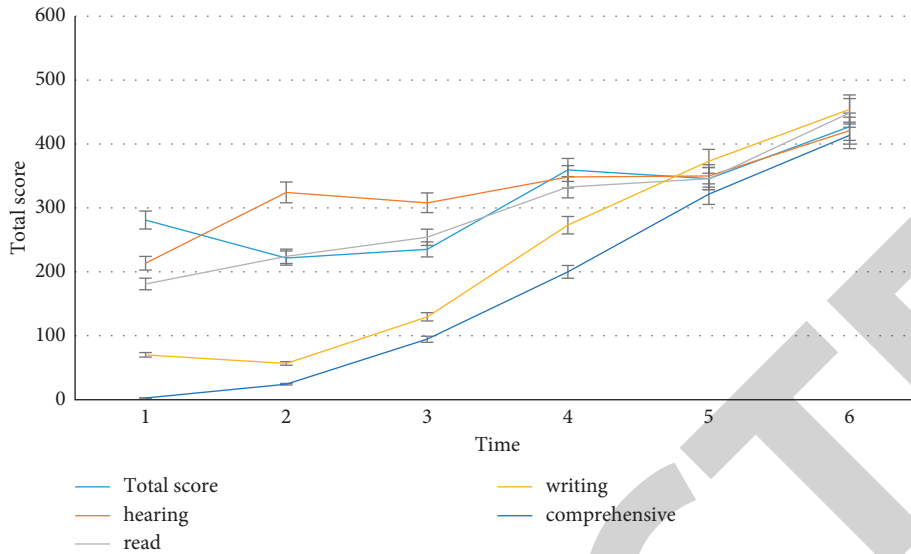
FIGURE 5: The English percentile score of the second-grade students in 2019~2021.
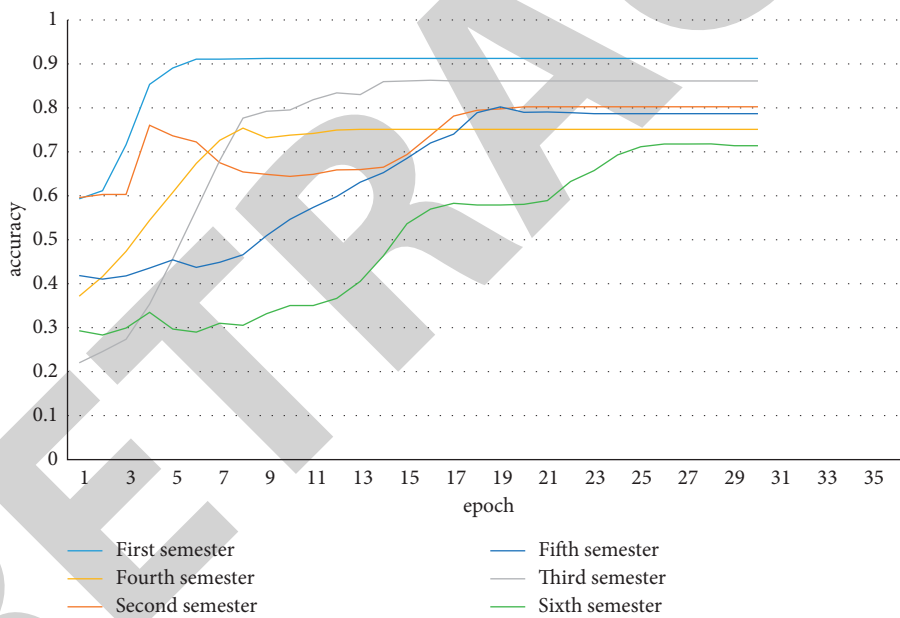

FIGURE 6: Student achievement prediction.

remain unchanged, and the total score of the candidates will have a significant improvement. The performance of the single test will also improve, but there are still some unstable factors, which should attract the attention of teachers and strengthen some targeted training for students.

The prediction accuracy obtained after 100 iterations is shown in Figure 6. It can be seen that the prediction accuracy rate of each semester fluctuates from 70% to 100%. Among them, due to the small number of samples in the eighth semester, the prediction accuracy rate reaches 100% after 10 iterations. The sixth semester, which had the lowest accuracy rate, also reached 70%.

In this experiment, 320 random records are randomly generated as the training set, and the remaining 80 records

are used as the verification set. The XGBoost algorithm adopts a dynamic learning rate, and the predicted value is obtained by inverse normalization. Figure 7 shows the comparison between the actual grades and the predicted values. The red dots represent the actual grades and predicted values of the students. 1 (black dots) represents the predicted grades using only student's grades as the dependent variable. The predicted value is 2, which is expressed as the predicted value of student performance and behavior as dependent variables. It can be seen from the figure that the predicted value of students' grades and behavioral information as dependent variables is closer to the real grades, indicating that the students' performance behavioral information is reasonable considering the factors that affect students' performance and is in line with previous expectations.
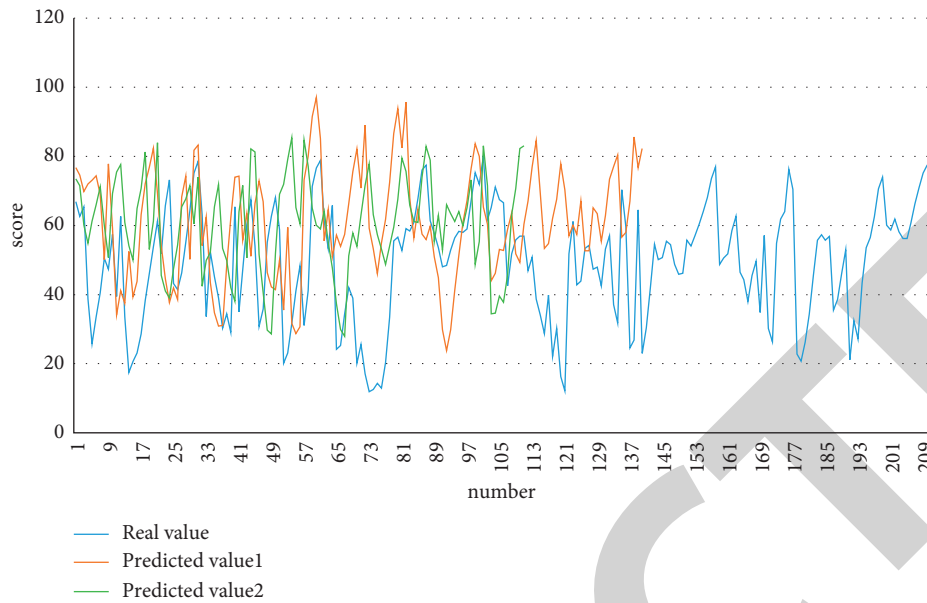
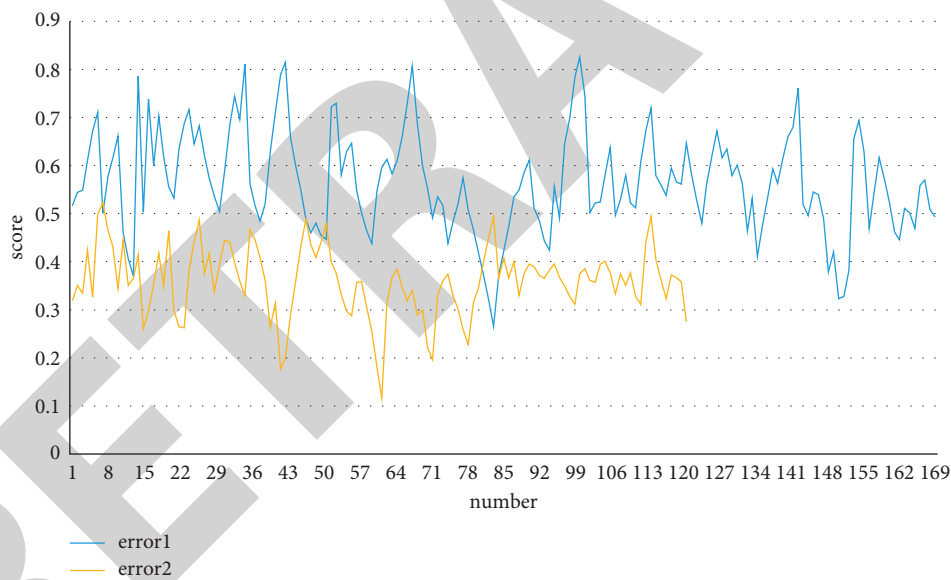Figure 7: Results of grade prediction.

Figure 8: Relative error.

Relative errors of the comparative experiments are shown in Figure 8. The red line (error 1) represents the relative error that only considers student's grades into account, and the black line (error 2) represents the relative error of the prediction that takes both the student's grades and behavioral information into account. As can be seen from such figure, most of the black lines are below the red line and only a small amount is above the red line, which indicates the relative error of prediction is lower than only considering the relative error of students' grades. Experimental results show that behavior of students has a certain influence on grades. Schools should not only pay attention to students' previous grades but also focus on the daily

behavior of students. It is significance to cultivate good study and living habits.

Figure 9 shows the coefficient of determination that only consider student's grades. The black line and the red line represent the actual value and the predicted value, respectively. From Figure 10, it can be seen that the coefficient of determination is about 0.79424; Figure 10 shows students' grades and the determination coefficient is considered for all behavioral information. The blue line and the red line are same as Figure 9, and the determination coefficient is about 0.99843. It can be seen that adding the student's behavioral information to the affecting factors, the performance of students greatly improves the determination coefficient. The
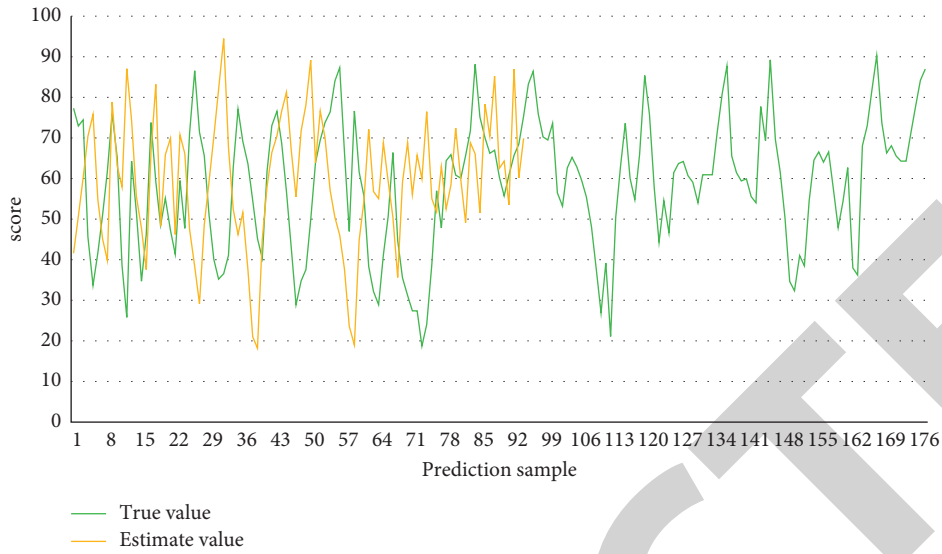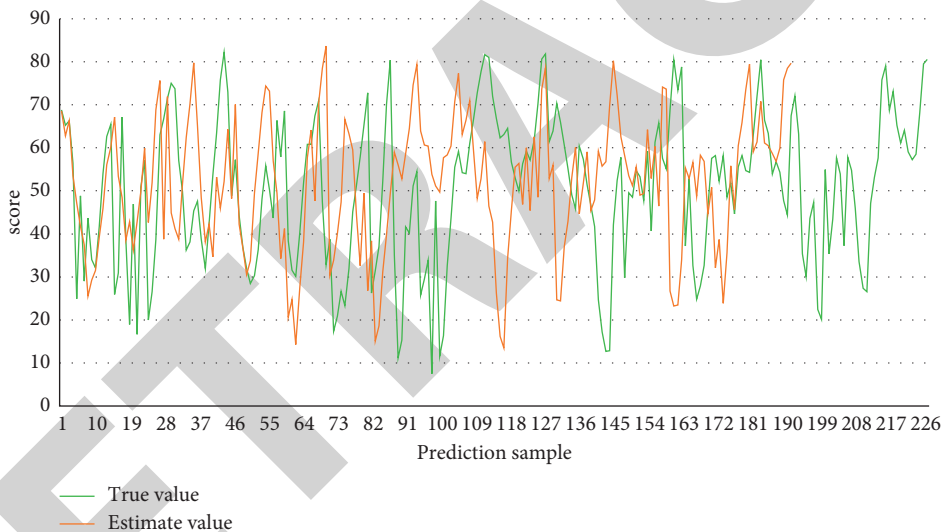
Figure 9: Determination coefficient 1.

Figure 10: Determination coefficient 2.

behavioral information of students has a significant effect on the final grades of students.

The above results show that the XGBoost algorithm model can predict college English test scores. Through data mining technology, it analyzes and evaluates students' test scores and extracts the degree of students' mastery of English knowledge in teaching process and targeted teaching.

5. Conclusion

In this paper, we use grades data in the English skill training system and use XGBoost model to predict students' grades. Experimental results prove that data mining technology has the accuracy and feasibility in English grade prediction. It is useful to improve the efficiency of the education field by utilizing such technology. Furthermore, the use of data mining technology will definitely change the traditional

education in some way. Prediction of college English grades is helpful for English learning of students, and it can also help teachers to better analyze test results. [25].

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001.

[2] S. N. Fallah, M. Ganjkhani, S. Shamshirband, and K. W. Chau, "Computational intelligence on short-term load forecasting: a methodological overview," *Energies*, vol. 12, no. 3, p. 393, 2019.

[3] M. Jin, X. Zhou, Z. M. Zhang, and M. M. Tentzeris, "Short-term power load forecasting using grey correlation contest modeling," *Expert Systems with Applications*, vol. 39, no. 1, pp. 773–779, 2012.

[4] B. Wang, B. Leng, and X. Zhang, "Overview of the application of support vector machine in short-term load forecasting," *Journal of Electric Power Systems and Automation*, vol. 23, no. 4, pp. 115–121, 2011.

[5] H. Xie, W. Jiangping, and H. Liu, "Parameter selection and optimization method of support vector machine model in short-term load forecasting," *Chinese Journal of Electrical Engineering*, vol. 26, no. 22, pp. 17–22, 2006.

[6] Li Jin, J. Liu, and J. Wang, "Medium and long-term load forecasting method using support vector machine and simulated annealing algorithm," *Chinese Journal of Electrical Engineering*, vol. 31, no. 16, pp. 63–66, 2011.

[7] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2531–2539, 2010.

[8] Z. Zou, Y. Sun, and Z. Zhang, "Short-term load forecasting based on ant colony optimization algorithm recurrent neural network," *Power Grid Technology*, vol. 29, no. 3, pp. 59–63, 2005.

[9] P. Kohlhepp, H. Harb, H. Wolisz, S. Waczowicz, D. Müller, and V. Hagenmeyer, "Large-scale grid integration of residential thermal energy storages as demand-side flexibility resource: a review of international field studies," *Renewable and Sustainable Energy Reviews*, vol. 101, pp. 527–547, 2019.

[10] Li Zuo, B. Zhou, and L. Nan, "Large-scale grid integration of residential thermal energy storages as demand-side flexibility resource: a review of international field studies," *Renewable and Sustainable Energy Reviews*, vol. 101, no. 3, pp. 527–547, 2019.

[11] S. Ge, O. Jia, and H. Liu, "Short-term power load forecasting under the condition of real-time electricity price based on genetic grey neural network model," *Power Grid Technology*, vol. 36, no. 1, pp. 224–229, 2012.

[12] H. Z. Li, S. Guo, C. J. Li, and Jq Sun, "A hybrid annual power load fore-casting model based on generalized regression neural network with fruit fly optimization algorithm," *Knowledge-Based Systems*, vol. 37, no. 2, pp. 378–387, 2013.

[13] D. Wang and Z. Sun, "Power user-side big data analysis and parallel load forecasting," *Chinese Journal of Electrical Engineering*, vol. 35, no. 3, pp. 527–537, 2015.

[14] J. Corbett, K. Wardle, and C. Chen, "Toward a sustainable modern electricity grid: the effects of smart metering and program investments on demand-side management performance in the US electricity sector 2009–2012," *IEEE Transactions on Engineering Management*, vol. 65, no. 2, pp. 252–263, 2018.

[15] C. Li, X. Zheng, Z. Yang, and L. Kuang, "Predicting short-term electricity demand by combining the advantages of ARMA and XGBoost in fog computing environment," *Wireless Communications and Mobile Computing*, pp. 1–18, 2018.

[16] Y. Li, Z. Wang, Y. Zhou, and X. Han, "The improvement and application of XGboost method based on the Bayesian optimization," *Journal of Guangdong University of Technology*, vol. 35, no. 1, pp. 23–28, 2018.

[17] R. Jia, X. Ran, J. Wu, C. Dai, and Z. Qi, "Airport passenger flow prediction based on XGBoost algorithm," *Journal of Civil Aviation*, vol. 2, no. 6, pp. 35–39, 2018.

[18] Q. Shang and Q. Shi, "CUI BingMou A study on passenger volume forecast at the ticketing counter based on gray-Markov forecast mode," *Railway Transport and Economy*, vol. 41, no. 1, pp. 85–89, 2019.

[19] C. Wu, S. Guo, and C. Li, "Research on the discretization algorithm of association classification based on Gaussian mixture model," *Journal of Chinese Computer Systems*, vol. 39, no. 4, pp. 732–737, 2018.

[20] Z. M. Fadlullah, D. M. Quan, N. Kato, and I. Stojmenovic, "GTES: an optimized game-theoretic demand-side management scheme for smart grid," *IEEE Systems journal*, vol. 8, no. 2, pp. 588–597, 2013.

[21] F. Shi, Y. Yao, Y. Bin, C. H. Zheng, and J. Xia, "Computational identification of deleterious synonymous variants in human genomes using a feature-based approach," *BMC Medical Genomics*, vol. 12, no. S1, pp. 12–88, 2019.

[22] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense modeling," *Applied Soft Computing*, vol. 75, pp. 323–332, 2019.

[23] Z. H. Zhou, *Machine Learning*, Tsinghua University press, Beijing, 2016.

[24] A. Zakeri and A. Hokmabadi, "Efficient feature selection method using real-valued grasshopper optimization algorithm," *Expert Systems with Applications*, vol. 119, pp. 61–72, 2019.

[25] L. Pang, J. Wang, L. Zhao, C. Wang, and H. Zhan, "A novel protein subcellular localization method with CNN-XGBoost model for alzheimer's disease," *Frontiers in Genetics*, vol. 9, pp. 751–757, 2018.

[26] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection," *Information Processing & Management*, vol. 59, no. 2, Article ID 102844, 2022.

[27] R. Ali, M. H. Siddiqi, and S. Lee, "Rough set-based approaches for discretization: a compact review," *Artificial Intelligence Review*, vol. 44, no. 2, pp. 235–263, 2015.

[28] R. Ali, S. Lee, and T. C. Chung, "Accurate multi-criteria decision making methodology for recommending machine learning algorithm," *Expert Systems with Applications*, vol. 71, pp. 257–278, 2017.

[29] G. Cai, Y. Fang, J. Wen, S. Mumtaz, Y. Song, and V. Frascolla, "Multi-carrier M-ary DCSK system with code index modulation: an efficient solution for chaotic communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 6, pp. 1375–1386, Oct, 2019.

[30] K. Chandra, A. S. Marcano, S. Mumtaz, R. V. Prasad, and H. L. Christiansen, "Unveiling capacity gains in ultradense networks: using mm-wave NOMA," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 75–83, June 2018.

[31] S. Palanisamy, B. Thangaraju, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and F. Alassery, "A novel approach of design and analysis of a hexagonal fractal antenna array (HFAA) for next-generation wireless communication," *Energies*, vol. 14, no. 19, p. 6204, 2021.

[32] S. Nagi Alsubari, S. N. Deshmukh, A. Abdullah Alqarni et al., "Data analytics for the identification of fake reviews using supervised learning," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 3189–3204, 2022.

[33] A. Radwan, M. F. Domingues, and J. Rodriguez, "Mobile caching-enabled small-cells for delay-tolerant e-Health apps," in *Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 103–108, IEEE, Paris, France, May 2017.

[34] F. B. Saghezchi, A. Radwan, J. Rodriguez, and T. Dagiuklas, "Coalition formation game toward green mobile terminals in heterogeneous wireless networks," *IEEE Wireless Communications*, vol. 20, no. 5, pp. 85–91, 2013.

[35] Q. Liu, C. Liu, and Y. Wang, "etc. Integrating external dictionary knowledge in conference scenarios the field of personalized machine translation method," *Journal of Chinese Informatics*, vol. 33, no. 10, pp. 31–37, 2019.