Article

An Intelligent Boosting and Decision-Tree-Regression-Based Score Prediction (BDTR-SP) Method in the Reform of Tertiary Education Teaching

Ling Zhu [1], Guangyu Liu [2,*], Shuang Lv [1], Dongjie Chen [1], Zhihong Chen [1] and Xiang Li [1]

[1] School of Artificial Intelligence and Information Management, Zhejiang University of Finance and Economics, Hangzhou 310012, China; zhuling@zufe.edu.cn (L.Z.); lvshuang@zufe.edu.cn (S.L.); dj233@zufe.edu.cn (D.C.); kwatb@zufe.edu.cn (Z.C.); dkioumu@zufe.edu.cn (X.L.)

[2] Key Laboratory of IoT and Information Fusion Technology, School of Automation, Hangzhou Dianzi University, Hangzhou 310005, China

[*] Correspondence: g.liu@hdu.edu.cn

Abstract: The reform of tertiary education teaching promotes teachers to adjust timely teaching plans based on students' learning feedback in order to improve teaching performance. Thefore, learning score prediction is a key issue in process of the reform of tertiary education teaching. With the development of information and management technologies, a lot of teaching data are generated as the scale of online and offline education expands. However, a teacher or educator does not have a comprehensive dataset in practice, which challenges his/her ability to predict the students' learning performance from the individual's viewpoint. How to overcome the drawbacks of small samples is an open issue. To this end, it is desirable that an effective artificial intelligent tool is designed to help teachers or educators predict students' scores well. We propose a boosting and decision-tree-regression-based score prediction (BDTR-SP) model, which relies on an ensemble learning structure with base learners of decision tree regression (DTR) to improve the prediction accuracy. Experiments on small samples are conducted to examine the important features that affect students' scores. The results show that the proposed model has advantages over its peer in terms of prediction correctness. Moreover, the predicted results are consistent with the actual facts implied in the original dataset. The proposed BDTR-SP method aids teachers and students to predict students' performance in the on-going courses in order to adjust the teaching and learning strategies, plans and practices in advance, enhancing the teaching and learning quality. Therefore, the integration of information technology and artificial intelligence into teaching and learning practices is able to push forward the reform of tertiary education teaching.

Keywords: teaching reform; small samples; score prediction; ensemble learning; decision tree regression

1. Introduction

Recently, various teaching and learning tools based on the cutting-edge information technology, such as MOOC, online practice platforms, and smart teaching assistant APPs, have been invented to feed back different data during the learning process to educators, teachers and students who will extract useful information from the data to reform the on-going teaching strategies, plans and practices. As the scale of online and offline education continues to grow, the issue of ensuring and improving the quality of education has become increasingly prominent. Score prediction has been a concern of academics as an important criterion for judging students' learning conditions in the education realm [1–3]. Since the prevalence of the COVID-19 pandemic, the development of online learning platforms in universities has become increasingly sophisticated, enabling online platforms to provide massive amounts of data for score prediction. At the same time, the continuous development

of information management systems in universities has made it convenient and easy for teachers to collect and analyze student behavior. This will help educators to extract important factors affecting scores from the systems so that they can obtain early feedback on student learning and optimize educational decisions [4].

It is well known that deficiencies in learning behavior are often considered one of the major causes of academic failure. Assuming that a student has poor completion of classwork, it may indicate a problem with their learning attitude or habits [5–7]. If we can identify these problems through score prediction to optimize educational decisions in time, it will help to avoid the risk of some students being in inferior position or neglected in the learning process. Therefore, it is necessary to look for important factors that influence scores to obtain information about students and to develop methods for predicting scores. This will enable educators to obtain feedback on student scores in advance using predictive methods and to make better teaching decisions in time for the overall development of students.

The main data-mining techniques commonly used within the education realm include classification, cluster analysis, association analysis and visual analysis. Techniques in the area of score prediction are studied and improved accordingly in this paper. The relevance of various factors in teaching and learning is analyzed to infer the future learning status and trends of students. Using data-mining tools to find the deeper content embedded in the data, educators can easily visualize the learning environment of students and make clear decisions [8]. The value of educational data in education can be realized by combining the experience of previous generations with the law presented by modern education to provide a more scientific basis for educational reform [9]. Our purpose is to guide teachers to correct teaching plans for students' differentiation by mining and analyzing high-dimensional small-sample educational data and predicting students' scores. How to better deal with high-dimensional small sample data is a problem we are currently facing. If the problem is solved by a new and effective tool (e.g., an artificial intelligent method), then the tool enables teachers and students to anticipate whether students could have achieved the expected teaching goals and generates timely academic warning for students, which helps the reform of teaching and learning strategies, plans and practices.

2. Related Work

2.1. Study of Education Data Dimensions

When dealing with educational data with a large number of dimensions, we inevitably encounter problems, such as missing data or inaccurate information. Wongvorachan T. et al. used random oversampling (ROS), random undersampling (RUS), and the combination of the synthetic minority oversampling technique for nominal and continuous (SMOTE-NC) and RUS as a hybrid resampling technique [10]. Fei Lin et al. used trusted population data to recover abnormal values, while using a heap model framework combined with random sampling to compensate for the shortcomings of single sampling [11]. Qizeng Zhang explored oversampling techniques using discretization methods and minority sample synthesis to improve accuracy [12]. Xinyu Cao et al. used the KNN model to correct the data [13]. Based on the data obtained, Zhuoxuan Jiang et al. conducted a correlation analysis between the eigenvalues and the scores and selected the eigenvalues with high correlation as input for the experiment [14]. On the other hand, Wei Xu et al. selected data features from multiple perspectives, including not only the students' online learning data, but also an analysis of the learning status in class, while using data from manual observation and multiple data combinations [15]. Dan Song et al. used data from previous classes of the same major, as well as learning data from other related courses [16]. Catarina et al. analyzed a large amount of relevant literature to analyze the corresponding models through the econometric depth of the literature, identified features used in the experiment and classified them [17]. Ali Yağci added data from a population survey questionnaire to the input features [18]. Similarly, Jiaxin You et al. combined multiple data features, such as the students' psychological cognitive level, prior knowledge acquisition, etc. [19]. It is

evident that the reasonableness of the selected features has a great impact on the model training results. However, the above-mentioned studies only processed the input data. Filtering the data may improve the correctness of data with only a small number of samples, but single anomalies variables can have a huge impact on the results. Equally, increasing the dimensionality of the variables, though increasing the stability, also introduces many confounding factors that interfere with the subsequent training of the model.

2.2. Prediction Model of Education Field

Along with the rapid development in the realm of machine learning, there are currently many models for score prediction. Mingzhi C. et al. integrated the theory of educational psychology and proposed an enhanced learning and forgetting model for contextual knowledge tracking (LFEKT), which enriched the feature representation of exercises by combining difficulty levels, and considered how students' reaction behavior affected the knowledge state [20]. Mengying Li et al. proposed a two-way attention model considering the influence of selected and unselected features, respectively. However, the number of selected features is small and less practical for multi-dimensional student behavior data nowadays [21]. Yangyang Luo et al. found that the model with incremental learning was more advantageous by comparing it with batch learning. However, this experiment was only for the random forest algorithm [22]. Xi Chen et al. used correlation information to construct a knowledge graph and proposed a method that fuses the knowledge graph and collaborative filtering to predict students' correlation scores, although the method requires a large amount of historical data as input features, and its performance would be greatly reduced if the data samples were small [23]. Yang Zhang et al. used a deep learning graph self-encoder model for prediction, which automatically extracts features without human intervention and predicts grades without a large amount of prior knowledge, but the efficiency of the method is greatly reduced when faced with a small base of input features [24]. Zijian Chen et al. nested an ensemble learning training model on top of the original classical model, but in this study, only three categories of academic performance prediction results were classified, and the prediction accuracy was low [25]. Emre Toprak et al. compared artificial neural networks, decision trees and linear discriminant methods, and analyzed the advantages and disadvantages of each method for different data types [26]. Pratya et al. combined artificial neural networks, decision tree algorithms and Bayesian algorithms to construct prediction models and validate the models using confusion matrices, accuracy, precision, recall, etc. [27]. Lykourentzou et al. used multiple feedforward neural networks to dynamically predict student grades over the course of a semester and classify the results into groups for post-learning, although all are predictive models of student academic level grades, only predicting the grades obtained at the end of the course, with a relatively coarse classification [28].

2.3. High-Dimensional Small Sample Data Processing

Zhijiang Peng used statistical-learning- and machine-learning-based methods to fill in missing data, and used multiple data-resampling methods to tackle imbalance data [29]. Botao Wei used Bayesian learning to estimate the posterior distribution using the prior distribution of parameters and a small number of samples and proposes a Bayesian variational generative adversarial network model for generating small sample data [30]. Jingyuan Huang combined prior knowledge to determine the nature of the dataset, used feature-selection-based classification methods for classification for gene data containing a large amount of noise features and investigated several feature selection methods and classification methods [31]. Jing Zhang proposed an ensemble feature selection algorithm ECGS-RG based on correlation for feature selection. This method can generate multiple effective feature subsets to make up for the lack of information of a single subset and improve the stability of feature selection [32]. Zhe Wang et al. presented a personality prediction method based on particle swarm optimization (PSO) and the synthetic minority oversampling technique+Tomek Link (SMOTETomek) resampling (PSO-SMOTETomek), which, apart from

the effective SMOTETomek resampling of data samples, was able to execute PSO feature optimization for each set of feature combinations [33]. Dazi Li et al. presented an advanced prediction model for the small sample analysis called extreme gradient boosting (XGBoost) based on nearest neighbor interpolation (NNI) and the synthetic minority oversampling technique (SMOTE) [34]. Through the analysis of recent research on small sample datasets, we find that existing research mainly focuses on dealing with small sample image datasets and high-dimensional small sample datasets. The former mainly performs data augmentation, while the latter mainly performs feature selection. The model explored in this paper is mainly provided for teachers to use. The selected multi-dimensional small-sample datasets in the field of education do not have high sample-size requirements. Therefore, machine learning algorithms are selected to improve the model training performance by combining multiple "base learners".

3. A Challenging Prediction Problem

Learning score prediction is an important research topic in the reform of tertiary education teaching. With the continuous expansion of online and offline education, multi-dimensional teaching data have emerged. How to find potential learning patterns related to various factors that affect scores in a course is a common concern in the educational and academic community. However, the dataset available to a teacher or educator is usually limited. Indeed, learning performance prediction based on the small sample dataset is an open topic because it is practically important to the teachers or educators who are confronted with challenges regarding how to anticipate and improve their classes' performance. Therefore, we propose a problem to deal with small samples with a new artificial intelligence model.

4. Materials and Methods

4.1. Definition of the Dataset

We collect learning data from a group of students in a computer science course at a university in China.

Inspired by the influential study of some attributes in predicting student learning performance, we select some attributes from the dataset as shown in Table 1 in relation to the students' characteristics, academic performance and learning behaviors in the selected course of university. Let us discuss the selection of these attributes. Online test in online practice platform mainly involves teachers breaking down the teaching content into multi-level experimental and practical sub-projects based on the curriculum syllabus where each sub-project is relatively independent and interrelated with the other ones. Online video completion refers to the students' completion in the learning of MOOC. The scores in the online test i(25 in total) and the online video completion indicate how well students have completed their online assignments after a class, which reflects their self-study ability and their mastery of the content. The higher the corresponding value, the better the students' mastery of the course at the moment. Studies have shown that students with higher attendance rates have better academic performance [35]. We adopt a smart teaching assitant APP (namely, "Rain Classroom") to record students' attendance rates and in-class test scores in real-time. We take the in-class test score as an auxiliary evaluation method of the attendance rate. If students come to class and study hard, their accuracy rate will be higher; if students come to class but do not listen, their accuracy rate will be in the middle. The students who skipped class had the lowest accuracy in the in-class test. Therefore, attendance rates and in-class test scores can reflect students' concentration and activeness in learning. The attendance rate and the score on an in-class test indicate the students' learning status in the class and the students' concentration and motivation in learning. The in-class test is an online test posted in a cloud software, which is an effective way to check students' learning status and efficiency in class. The origin and the scores in mathematics and English show the students' abilities, assumed to be relevant to the course. Mathematics is an important foundation for computer science such that students who

are good at mathematics are generally considered to be able to achieve better results in computer science courses. English is the basis for mainstream programming languages and active projects such that a good foundation in English will have a significant impact on the learning performance of a student in their computer science courses. The results in mathematics and English in the entrance exam represent the mastery level of these two fundamental subjects. Here, the origin variable is not a numerical number. It is a categorical number that reflects the different incoming preparation or the abilities assumed to be relevant to the course of the students. Due to different educational resources and regulations in different regions, the origin of a student may have an impact on his/her learning performance.

Table 1. Attributes for dataset.

Attribute	Description	Range
Gender	Gender of the student	{Male, Female}
Student number	Identification of the student	String
Online test score $i \in \mathcal{I}^+$	An online test score	[0, 100]
In-class test score	An in-class test score	[0, 100]
Attendance rate	Students' attendance rate	[0, 1]
Origin	The provinces or metropolises where students come from	Int
Initial score in Mathematics	The student's entrance score in Mathematics	[0, 100]
Initial score in English	The student's entrance score in English	[0, 100]
Final score	The student's final score in the selected course of the university	[0, 100]
Online video completion	The student's completion in the learning of MOOC	[0, 1]

4.2. Correlation Analysis of Variables

　　We define a list of variables in Table 2 for a range of statistical measures.

Table 2. Variables in statistics.

Variables	Note
SS_{xy}	Sum of product of x and y
SS_{xx}, SS_{yy}	Sum of product of x, y
x_i, y_i	Eigenvalues (model input variables, such as test scores, place of origin, etc.)
$\bar{x}$, $\bar{y}$	The average value of x, y
r_{xy}	The correlation coefficient of x and y
$\hat{y}_i$	The corresponding value on the trend line of x_i
n	The amount of samples.

　　These statistical measures are used to demonstrate the correlation between different variables quantitatively. The correlation between the two variables is measured by

$$SS_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{1}$$

which reflects whether the directions of the change of x and the change of y are the same, while the speed of change of SS_{xy} also reflects whether the change in y about x is sensitive.

　　It is noted that SOP is easily influenced by the size of the samples and the unit of the values. So, we adjust the sum of the product by the averaging and normalizing tools

according to the sample size, leading to a correlation coefficient. The correlation coefficient is calculated by

$$r_{xy} = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} \, . \tag{2}$$

The sum of squares due to error is used in statistics to express the dispersion given by

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

which can be used for measuring the dispersion of each sample point to the trend line.

If the values of x are the same, the further away from the trend line the y values are, the more uncertain the relationship between x and y is, and the greater the randomness is. The way to compare different dispersion values is not sufficient because the size of the samples also affects the final calculation. It is necessary to use the residual mean squared error to express the average dispersion:

$$MSE = \frac{1}{n-2} \cdot SSE \, . \tag{4}$$

The above formulae help create a heat map for a visual inspection of the correlation between the attributes:

$$R^2 = 1 - \frac{SSE}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

The R^2 is an indicator for evaluating the model, with values ranging from 0 to 1. The closer is to 1, the better the performance.

4.3. The Method of BDTR-SP

We propose an intelligent boosting and decision-tree-regression-based score prediction (BDTR-SP) method for predicting final scores of a course from the existing factors.

To overcome the drawback of limited experimental data, we adopt not only some prediction model but also an ensemble learning method to deal with the scarce sample problems and improve the prediction accuracy [36]. Ensemble learning completes a learning task by constructing and combining multiple learners, which are referred to as "base learners". Our model selects the decision tree regression method as the base learner, which is based on the decision-making tree. Each branch node makes judgments on the attributes of the dataset, and the result is obtained by the leaf nodes traveling based on the judgments of each attribute. This way, a decision tree with a strong generalization ability is generated. The data have multi-dimensional characteristics, where each dimension has different weights for the final prediction results. Then, the decision tree regression method is applied to AdaBoost. By adjusting the data distribution and assigning weights based on the previous round of training results, AdaBoost can alleviate the problem of multiple dimensions in the dataset and uneven data distribution to a certain extent [37].

The above method is summarized as the boosting and decision-tree-regression-based score prediction model (BDTR-SP). The proposed method of BDTR-SP is introduced as follows (see Figure 1):

- Firstly, we divide the dataset into the training set and the test set in a ratio of 7:3, and set the initial weight of each sample in the original training as the same as the other ones.
- Secondly, we set the ensemble learner to consist of M base learners, divide the original training set into K datasets, and allocate them to all base learners.
- Thirdly, we set the decision tree as the base learner in the ensemble learner and use the base learner to train the training set.
- Fourthly, after the training is completed, we use the base learner to predict the score data of the students in the original training set.

- Fifthly, for samples with correct classification prediction, it indicates that the sample has a relatively small impact on performance prediction to reduce the weight of the sample. On the contrary, for samples with incorrect classification prediction, as they have a significant impact on the results of performance prediction, we increase the weight of the sample so as to give it a greater probability of being selected as new training data.
- Sixth, a new classifier is constructed through the above steps and trained iteratively to predict erroneous data and yield better performance.
- Finally, after the training is completed, the ensemble learner adjusts the weight of the classifier, increases the weight of the classifier with small classification error, and synthesizes the prediction results of the learner through weighted adjustment.

Figure 1. A schematic structure of BDTR-SP.

5. Experimental Results and Analysis

5.1. Dataset and Features

To make the data sufficient and comprehensive, we establish a dataset based on data collected from course resources, student information resources and online education resources. The course resources include the attendance rates and the in-class test scores. The student information resources include the scores in mathematics and English in the entrance exams of the university in terms of different origins (e.g., provinces and metropolises). The online education resources include the the online video completion results and the scores in the online tests. The students' math scores and English scores and origin are known before the students enroll in the course, and other attributes can be obtained in real time after the teacher finishes each class. Therefore, the teaching plan can be adjusted in time according to the real-time data so that students can make substantial corrections. Removing the data of students who cannot participate in the final exam, finally, we obtain a valid dataset with 31 variables.

Figure 2 shows that a raw dataset is constructed through a number of input variables with different statistical indexes. The dataset has some features. For example, the students' scores in mathematics and English and scores in final exam are basically normally distributed, but the distribution of online test scores tends to be monotonic such that 75% of scores are 100, which differs significantly from the predictions.

	Attendance rate	Online video completion	Math score	English score	Score of final exam	Score of online test 1	Score of online test 2	Score of online test 3	...
Count	52.00	52.00	52.00	52.00	52.00	52.00	52.00	52.00	...
Mean	0.99	50.51	80.23	73.94	76.52	99.68	97.12	78.00	...
Std	0.04	33.75	6.00	7.01	10.60	2.31	12.23	37.38	...
Min	0.83	0.00	62.00	61.33	42.00	83.33	33.33	0.00	...
Max	1.00	98.20	86.00	92.00	94.00	100.00	100.00	81.25	...
25%	1.00	17.75	69.83	75.33	70.00	100.00	100.00	100.00	...
50%	1.00	61.40	73.33	81.33	78.50	100.00	100.00	100.00	...
75%	1.00	78.63	78.67	84.17	83.00	100.00	100.00	100.00	...

Figure 2. An illustration of input variables with different statistical indexes to establish a raw dataset.

Figure 3 shows the monotonic distribution of the attendance rate variable and reveals all possible values within its range. However, according to school regulations, if the students are absent three times, they cannot take the final exam. Therefore, we exclude the data of these students and obtain the dataset shown in Figure 2 after filtering and preprocessing.
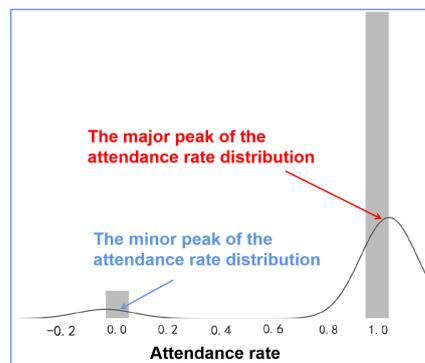

Figure 3. The monotonic distribution of attendance rate.

Combined with the statistical analysis of the attendance rate in Figure 2, we find that due to the exclusion of three absences, the standard deviation of this variable is very small and almost consistent, so it has little effect on the experiment and little significance. Therefore, we delete it in subsequent experiments.

Figure 4 shows that the correlation distribution between the scores in mathematics does not affect the scores in the final exam. Students with good scores in mathematics and English often achieve bad results in the course learning and vice versa. We deduce that a basic programming course does not need advanced mathematical knowledge such that advantages in mathematical skills have little effect on the course learning. Additionally, we regard the learning environment in a high school as self-dominated learning, to which some students are more accustomed, while the environment in a university is teacher-dominated learning. The learning mode in university is the self-dominated mode. Students are responsible for their own learning, consciously determine learning objectives, choose the learning methods, standardize the learning process, and evaluate the learning results. The learning mode in middle school is the teacher-dominated mode, i.e., learning objectives, plans, progress, process and evaluation are mainly formulated by teachers according to curriculum plans and standards. So, some students may not maintain their advantage all the way. Actually, the same observation can be made on the correlation between the scores in English and the scores in the final exam.

Figure 5 shows a heat map due to the correlation coefficients of different variables. It is found that the scores in mathematics are less related to the scores of the final exam because the correlation coefficients are below 0.1. This is somewhat different from our previous intuitions that the foundation in mathematics and English would have some effects on the score of the final exam.
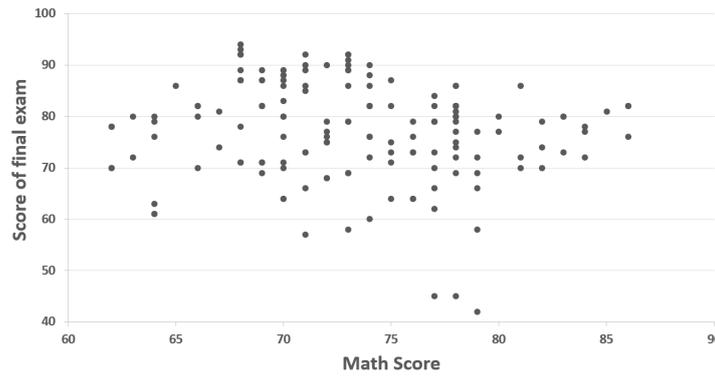
Figure 4. The chaotic relationship between the scores in mathematics and the scores in final exam.
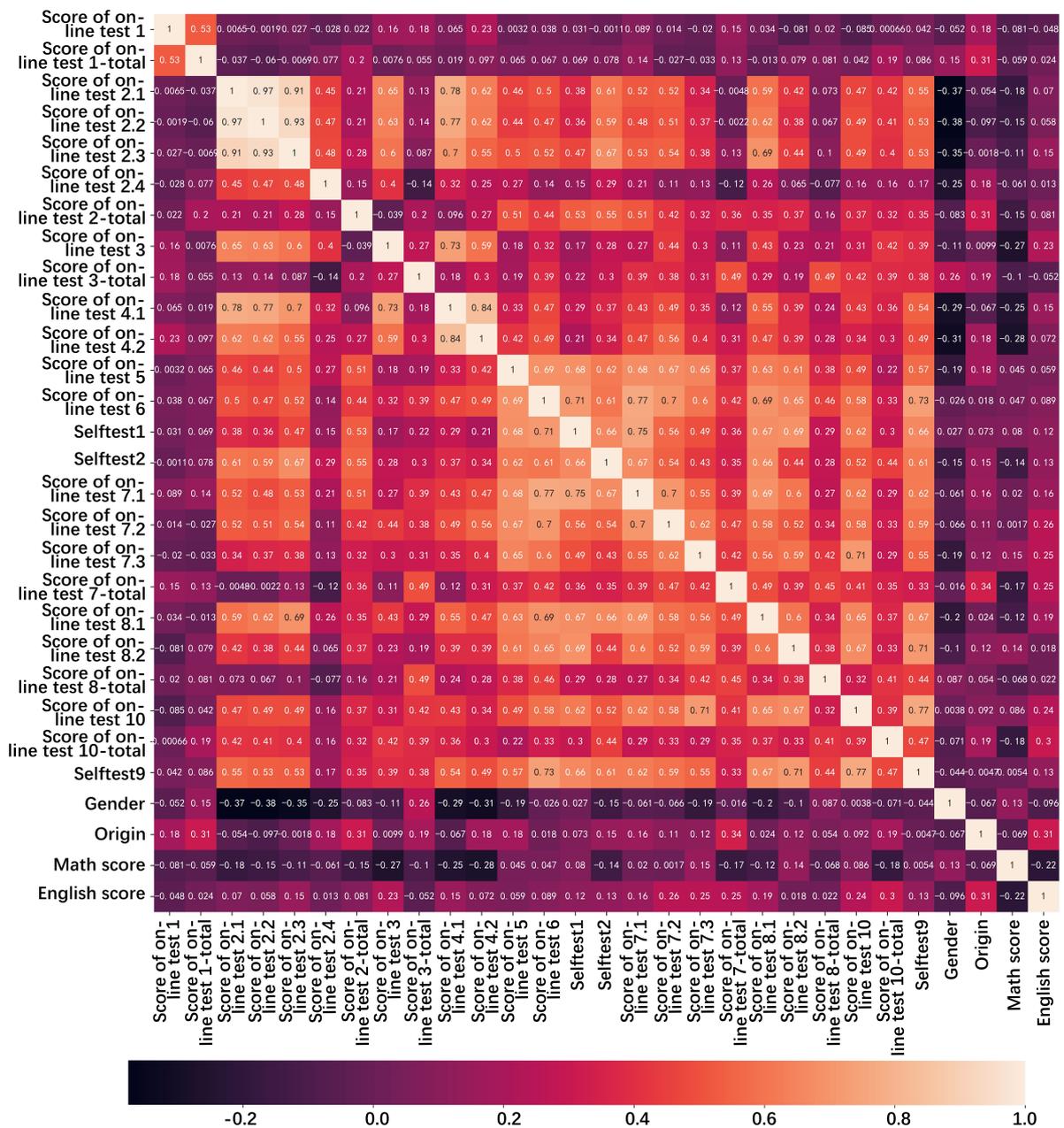

Figure 5. A heat map of correlation coefficients.

Table 3 shows the statistical results of the correlation variables between the scores in mathematics and those in the final exam based on the formulae in Section 4.2. The observation is consistent with the findings in Figure 5 such that the scores in mathematics and those in the final exam have little relationship.

Table 3. Correlation indexes.

SS_{xy}	r_{xy}	SSE	MSE
239.194	0.0485	10190.42	156.7751

5.2. Experiment Setting

The base learner is the most important factor affecting the ensemble learning. So, we compare the proposed method of BDTR-SP with a single model in terms of the decision tree regressor (DTR) and another ensemble model with Gaussian naïve Bayes (GNB). We also use random forest as a comparison. R^2 is used as the model indicator for comparing. Through a preliminary analysis, we arrive at the conclusion that the optimal base learner is the decision tree, whose depth is set to 5, and the number of base learners is set to 150. Additionally, we set the learning rate to 0.1 upon a linear loss function. In this way, we apply the proposed model of BDTR-SP for the dataset to derive the input–output relationship.
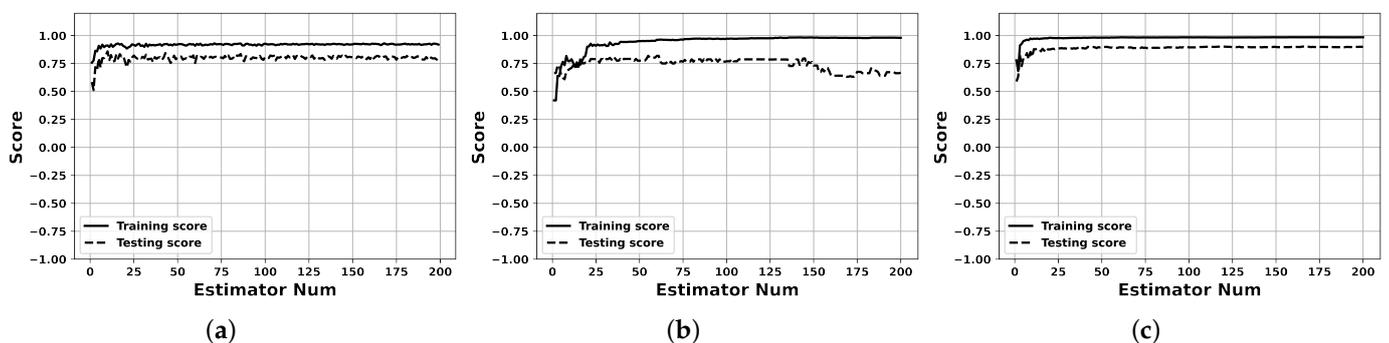
5.3. Experiment Results

Table 4 shows that the single learner prediction model does not give good prediction results, and the ensemble regression prediction model can effectively improve the accuracy.

Table 4. Comparison of model prediction results.

Model	R^2
Single model with DTR	0.3012
Random Forest	0.6221
Ensemble model with GNB	0.6572
BDTR-SP	0.9324

Figure 6 shows that, compared with the random forest and Gaussian naïve Bayes ensemble models, the curve of the decision tree ensemble model is more stable and the accuracy is higher.

(a)　　　　　　　　　　**(b)**　　　　　　　　　　**(c)**

Figure 6. The training results of random forest and ensemble models with DTR and GNB, respectively. (**a**) Random forest (max_depth = 5); (**b**) Gaussian naïve Bayes; (**c**) Decision tree (max_depth = 5).

Table 5 shows the accuracy due to different models when the scores in mathematics are a factor. It is shown that the proposed model of BDTR-SP has a higher score than the random forest and ensemble model with GNB with or without the scores in mathematics. Therefore, the given method has an advantage in predicting the scores in final exam over

its peer. Meanwhile, there is not a big difference between the predication results of the cases with math scores and the results of those without math scores. This observation is consistent with that shown in Figure 4, where the scores in mathematics do not have any direct correlation with the score in the final exam.

Table 5. The R^2 with or without the scores in mathematics.

Differences in Attributes	Random Forest	Ensemble Model with GNB	BDTR-SP
Cases with math scores	0.6221	0.7249	0.8676
Cases without math scores	0.6051	0.7400	0.8800

Table 6 compares the prediction performance of different models with or without students' origins. Again, it is found that the accuracy of BDTR-SP is higher than that of random forest and the ensemble model with GNB. These models give a conclusion that different origins of students have only a little impact on the final exam.

Table 6. The R^2 with or without students' origins.

Differences in Attributes	Random Forest	Ensemble Model with GNB	BDTR-SP
Cases with origins	0.6221	0.7484	0.8521
Cases without origins	0.6077	0.6848	0.8290

Through the analysis of the heat map, we find that the influence of the students' math scores and origin on the introductory courses of students' programming learning needs to be explored, while the influence of other attributes on students' programming learning has a significant positive correlation.

With a comparison of Tables 5 and 6, the math score decreases the accuracy, while adding the origin increases the accuracy. Students' origins are more influential for the final exam than the students' math scores. In general, both attributes have little impact on the final exam.

It is emphasized that the observations yielded by the specific dataset do not reflect the general trends of different universities. The importance of our study is to provide a useful tool of BDTR-SP, which leads to higher prediction accuracy in general.

6. Conclusions

The reform of tertiary education teaching has adopted the cutting-edge information technology tools including MOOC, online practice platforms, and smart teaching assistant APPs which can help an individual teacher timely collect the learning data from different aspects in his or her classes. Due to the importance of learning score prediction in the reform of tertiary education teaching and the limitation of the small dataset for the individual teacher or educator, we proposed an effective artificial intelligent tool to help predict students' scores well, namely, a boosting and decision-tree-regression-based score prediction (BDTR-SP) model. The tool relies on an ensemble learning structure with base learners of decision tree regression (DTR) to improve the accuracy of the learning performance of different groups of students. Experiments on small samples are conducted to examine the important features that affect students' scores. Our method is more practical than previous methods for the use of small sample datasets in the education realm, and teachers can apply it better. Moreover, the predicted results are consistent with the actual facts implied in the original dataset. It is a promising tool to help teachers and educators predict better students' scores in relation to their learning behaviors and previous performance. It reflects that the proposed BDTR-SP method is able to assist teachers predict students' future course performance in their classes. Based on the prediction results, these teachers are able to adjust their teaching strategies, plans and executing processes in advance, which

helps improve their teaching quality in the impending classes. Meanwhile, the learning performance prediction provides timely the academic warnings to some students at the risk and urges them to strengthen their knowledge learning and practice learning. Therefore, the integration of information technology, artificial intelligence and tertiary education has not only promoted the innovation of educational modernization but also built an intelligent learning environment. It gives impetus to the reform of tertiay education teaching.

References

1. Le, Q.; Hu, H.; Xiong, Y. Research on Students' Learning Behavior in Smart Classroom Teaching Mode. *China Educ. Technol. Equip.* **2020**, *483*, 39–41.
2. Chen, X. Analysis of Student Behavior in Data Mining Based on Decision Tree. *Intell. Comput. Appl.* **2020**, *10*, 127–130.
3. Yalan, D.; Kuanming, Z. Analysis and Countermeasures of Online Education Quality Problems under the Background of "Internet+". *Comput. Knowl. Technol.* **2022**, *18*, 134–136.
4. Liu, H.; Zhu, Y.; Zang, T.; Xu, Y.; Yu, J.; Tang, F. Jointly Modeling Heterogeneous Student Behaviors and Interactions among Multiple Prediction Tasks. *ACM Trans. Knowl. Discov. Data* **2021**, *16*, 16. [CrossRef]
5. Zhao, Y.; Li, H.; Yu, Y. A correlational study of the learners online learning behavior and their academic achievements. *J. Shijiazhuang Univ. Appl.* **2022**, *34*, 1–8.
6. Öğütcü, S. Student Performance Score Prediction Using Artificial Neural Network with the Support of Exploratory Factor Analysis and Clustering. In Proceedings of the 2020 IEEE Third International Conference on Data Stream Mining and Processing (DSMP), Lviv, Ukraine, 21– 25 August 2020; pp. 90–95.
7. Li, F. An Empirical Study on Influences of College Students' Learning Behavior Input on Academic Performance—From the Perspective of Intermediary Effect of Curriculum Cognition Goal and Learning or Weariness of Learning. *J. Chang. Univ.* **2022**, *32*, 26–36.
8. Li, T.; Fu, G. An Overall View of the Educational Data Mining Domain. *Mod. Educ. Technol.* **2010**, *20*, 21–25.
9. Jin, Y.; Chen, H. Value Orientations, Realistic Limitations and Effective Strategies of Big Data Mining for Education. *Theory Pract. Educ.* **2021**, *41*, 3–8.
10. Tarid, W.; Surina, H.; Okan, B. A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information* **2023**, *14*, 54.
11. Fei, L.; Zhan, Z. Learning Results of Large Scale Network Courses Based on Downsampled Heap Models Predicting. *Comput. Appl. Softw.* **2018**, *35*, 131–137.
12. Zhang, Q.; Dai, H. Student Performance Prediction Model Based on Data Preprocessing Technology. *J. Hubei Univ. Sci.* **2019**, *41*, 8.
13. Cao, X.; Cao, W.; Li, Z.; Sun, J. Prediction Method of College Students' Scores toward Uncertain Missing Data. *Mod. Electron. Tech.* **2018**, *41*, 145–149.
14. Jiang, Z.; Zhang, Y.; Li, X. Learning Behavior Analysis and Prediction Based on MOOC Data. *J. Comput. Res. Dev.* **2015**, *52*, 614–628.
15. Xu, W.; Li, W.; Zhan, X.; Xu, X.; Chen, X. Research on the Performance Prediction Methods Based on Classroom Learning Behavior. *Mod. Educ. Technol.* **2021**, *31*, 104–111.
16. Song, D.; Liu, D.; Feng, X. Course Performance Prediction and Course Early Warning Research Based on Multi-source Data Analysis. *Res. High. Educ. Eng.* **2020**, *1*, 89–194.

17. de Oliveira, C.F.; Rolland, S.S.; João, F.M.; Fernando, M. How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education: A Systematic Literature Review. *Big Data Cogn. Comput.* **2021**, *5*, 64. [CrossRef]

18. Ali, Y.; Çevik M. Prediction of academic achievements of vocational and technical high school (VTS) students in science courses through artificial neural networks (comparison of Turkey and Malaysia). *Educ. Inf. Technol.* **2019**, *24*, 2147–2167.

19. You, J.; Sun, Z. Predictions of and interventions in university students' academic achievement on cloud-based learning platforms. *Chin. J. Distance Educ.* **2016**, *9*, 14–20.

20. Chen, M.; Bian, K.; He, Y.; Li, Z.; Zheng, H. Enhanced Learning and Forgetting Behavior for Contextual Knowledge Tracing. *Information* **2023**, *14*, 168. [CrossRef]

21. Li, M.; Wang, X.; Ruan, S.; Zhang, K.; Liu, Q. Student Performance Prediction Model Based on Two-Way Attention Mechanism. *J. Comput. Res. Dev.* **2020**, *57*, 17–29.

22. Luo, Y.; Hang, X. A Model for Predicting Student Performance in Hybrid Courses Based on Incremental Learning Algorithm. *E-Educ. Res.* **2021**, *42*, 83–90.

23. Chen, X.; Mei, G.; Zhang, J.; Xu, W. Student Grade Prediction Method Based on Knowledge Graph and Collaborative Filtering. *J. Comput. Appl.* **2020**, *40*, 595–601.

24. Zhang, Y.; Lu, M.; Zheng, Y.; Li, H. Student Grade Prediction Based on Graph Auto-Encoder Model. *Comput. Eng. Appl.* **2021**, *57*, 251–257.

25. Chen, Z.; Zhu, X. Research on Prediction Model of Online Learners' Academic Achievement Based on Educational Data Mining. *China Educ. Technol.* **2017**, *12*, 75–81.

26. Toprak, E.; Gelbal, S. Comparison of Classification Performances of Mathematics Achievement at PISA 2012 with the Artificial Neural Network, Decision Trees and Discriminant Analysis. *Int. J. Assess. Tools Educ.* **2020**, *7*, 773–799. [CrossRef]

27. Nuankaew, P.; Nuankaew, W.; Teeraputon, D.; Phanniphong, K.; Bussaman, S. Prediction Model of Student Achievement in Business Computer Disciplines. *Int. J. Emerg. Technol. Learn. (iJET)* **2020**, *15*, 160–181. [CrossRef]

28. Lykourentzou, I.; Giannoukos, I.; Mpardis, G.; Nikolopoulos, V.; Loumos, V. Early and dynamic student achievement prediction in e-learning courses using neural networks. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 372–380. [CrossRef]

29. Peng, Z. Research on Feature Analysis Technology for Small Sample Data. Diploma Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2021.

30. Wei, B. Research on Classification of High-dimensional Small Sample Data based on Bayesian Deep Learning. Diploma Thesis, North China Electric Power University, Beijing, China, 2020.

31. Huang, J. Comparative Study on Classification Methods of Two High-Dimensional and Small Sample Data. Diploma Thesis, Huazhong University of Science and Technology, Wuhan, China, 2018.

32. Zhang, J. Classification and Feature Selection on High-dimensional and Small-Sampling Data. Diploma Thesis, Hefei University of Technology, Hefei, China, 2014.

33. Wang, Z.H.E.; Wu, C.; Zheng, K.; Niu, X.; Wang, X. SMOTETomek-Based Resampling for Personality Recognition. *IEEE Access* **2019**, *7*, 129678–129689. [CrossRef]

34. Li, D.; Liu, J.; Liu, J. NNI-SMOTE-XGBoost: A Novel Small Sample Analysis Method for Properties Prediction of Polymer Materials. *Macromol. Theory Simul.* **2021**, *30*, 2100010. [CrossRef]

35. Dey, I. Class attendance and academic performance: A subgroup analysis. *Int. Rev. Econ. Educ.* **2018**, *28*, 29–40. [CrossRef]

36. Xu, P.; Ji, X.; Li, M.; Lu, W. Small data machine learning in materials science. *npj Comput. Mater.* **2023**, *9*, 42. [CrossRef]

37. Huang, Z. Research on the Method of Predicting Students' Dropout Behavior in MOOC. Master's Thesis, South China University of Technology, Guangzhou, China, 2021.