World Scientific Research Journal Volume 8 Issue 12, 2022
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

Teaching Quality Evaluation of University Teachers Based on Value-Added Student Performance

Mingfei Guo, Jun Feng*, Yaguan Qian, Yuan Yuan, Wanqing Ma

Department of Sciences, Zhejiang University of Science and Technology, Hangzhou 310023, Zhejiang, China

* Corresponding author: Jun Feng (Email: gemf0411@163.com)

Abstract

Teaching quality evaluation of higher education teachers is an essential aspect of higher education evaluation. It is also the core factor that drives the development of high-quality teaching and learning in universities. How to evaluate teaching scientifically is a common concern of the current educational evaluation reform in universities. A new method of evaluating the teaching quality of university teachers based on the value-added of students' performance is proposed to address the shortcomings of scientific and objective nature and the large consumption of human and time resources of most current university evaluation work. Firstly, we analyzed the status data of university students to build a table of four features related to university course performance containing six subfeatures and selected "College English" and "Advanced Mathematics" courses as samples to predict students' course performance. Firstly, we analyzed the status data of university students to build a table of four features related to university course performance containing six subfeatures and selected "College English" and "Advanced Mathematics" courses as samples to predict students' course performance. Then, by comparing the difference between student's predicted and actual test scores, we obtained the value-added of teachers' performance due to the difference in their teaching quality. We selected this value-added to quantify the teachers' teaching quality at the class level. Through the combination of university course performance-related features table and student performance prediction and value-added evaluation, this method realizes the evaluation of university teachers' teaching quality by using student performance value-added, which provides a new method for university teachers' teaching quality evaluation with high reliability and high efficiency. It is a valuable exploration of university education evaluation reform in the digital era.

Keywords

University teacher teaching quality evaluation; value-added evaluation; performance prediction; machine learning.

1. INTRODUCTION

Teachers in higher education are responsible for cultivating a new generation of innovative talents and serving the significant needs of the country and social development. They are also the core factor in promoting the development of high-quality teaching in higher education. As higher education moves towards the stage of popularization, the teaching force in universities has expanded significantly. Building a good university faculty has become a critical link and an important goal in developing higher education teaching. The quality and standard of teaching are one of the core competencies of teachers. Differences in the quality of teachers' teaching can

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

lead to differences in the education received by recipients, causing problems such as educational inequity[1]. The urgency of scientifically and effectively evaluating the quality of teachers' teaching has become increasingly evident. Scientific evaluation of the teaching quality of university teachers is a common concern in global education evaluation and an important breakthrough point in education evaluation reform.

2. LITERATURE

There are two main methods for universities to evaluate teachers' teaching quality: subjective evaluation of teachers' teaching quality with students as the main part and subjective evaluation of teachers' teaching quality with fellow teachers as the main part[2].

The first method is based on the fact that students are the direct beneficiaries of the teacher's teaching and that it is common practice in the business field for the served to evaluate the service provided by the service provider. Thus, it is appropriate to use students as the most direct candidates for teacher-teaching evaluation in education. SET is an important measure used by universities in Europe and the United States to improve the quality of teaching and learning. SET is also used to measure teacher's overall performance and determine a title or position promotion. In most cases, SET tends to be the only criterion for evaluating teachers' teaching ability. Based on the SET, Li Wang and Na Gao[3] designed the CCSS questionnaire to evaluate the teaching quality of undergraduate courses based on students' experience by referring to two scales: the National Survey of Student Engagement (NSSE) and the China College Student Survey (CCSS). The questionnaire uses factor analysis to give good reliability to the questionnaire. Drawing on the extensive SET teaching evaluation framework, scholars such as Yuanxun Sun[4] proposed the Evaluation Indicator of Mathematical Teaching (EIMT) system for evaluating the quality of mathematics classes. However, with the widespread implementation of student evaluations, more and more voices believe that the results of student evaluations do not have sufficient credibility. They believe students' awareness, understanding, and judgment are still insufficient. The evaluation results from student evaluations do not have enough credibility. This evaluation method risks decreasing the quality of education and teaching and creating negative guidance for teachers' teaching behavior strategies. Foote[5] and other scholars found that most teachers had better SET results, not because they possessed better teaching skills, but because they handled the teacher-student relationship with their students very well. Nearly 70% of teachers felt that the more rigorously they judged student performance, the lower the teaching evaluations students gave teachers. In reality, the administration will correlate student performance with teacher teaching performance. However, student performance contains a variety of factors, and the evaluation indicators constructed with it may be diverse and without uniform standards, which may prompt teachers to adopt different teaching responses. For example, merit rates and class averages are often used as indicators to evaluate teachers' teaching standards, but this may cause teachers to subjectively ignore students who are less well-off and raise issues of inequity in the teaching process[6].

Peer-based professional evaluations to test teachers' teaching have high credibility. Based on the basic theory of fuzzy computing, Yanqing Ren and Xianfeng Yu[7] proposed a fuzzy multi-level evaluation model of teaching quality in primary and secondary schools, objectively evaluating teaching quality through the peer perspective. Jing Liu[8] addresses the problems of poor accuracy and a long time of catechism teaching quality evaluation and establishes a support vector regression-based teaching quality evaluation method for catechism courses in universities based on peer perspective for each rating of teachers. Then the evaluation results are obtained by support vector regression model. Although peer evaluation has some professional advantages, the resources required for it are costly and inefficient, making it

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

difficult to use it as a primary means of teaching evaluation. Thus, most of the peer evaluations are only auxiliary methods to evaluate the teaching quality of university teachers, which are complementary to and complete the student evaluations. In summary, although the existing teaching quality evaluation method of university teachers with student evaluation as the main body and peer evaluation as the auxiliary is commonly applied, the problems of low credibility of student evaluation and high resource cost, and low efficiency of peer evaluation are evident.

3. FEASIBILITY OF VALUE-ADDED EVALUATION

3.1. Value-added evaluation

The concept of value-added is learned from economics[9]. When it is introduced into teaching evaluation, value-added evaluation is based on the value added by students under the influence of schools and teachers to evaluate schools or teachers. Student value-added refers to the improvement of students' knowledge, abilities, and attributes and the difference between student's knowledge and skills after learning and their pre-study, i.e., the value of experience accumulated by students in educational institutions[10]. Student value-added comes from four primary sources: schools, teachers, students, and random factors. The basic formula is value-added = output value - input value.

In addition to overcoming many of the problems associated with the two aforementioned subjective evaluation approaches, value-added evaluation has the outstanding advantage that it can isolate the relationship between student value-added and complex random factors over time and then estimate the contribution of other factors to student value-added (value-added evaluation mechanism). When using value-added student performance to evaluate the quality of teachers' teaching, several other factors need to be strictly controlled to consider the teacher factor as the only contributor to value-added student performance. The improvement of students' knowledge and skills is closely related to the teacher factor, and the evaluation of teachers through the value-added of students is gradually being recognized[11].

3.2. Value-added performance

3.2.1 The value-added performance model

Value-added achievement is influenced by the school, teacher, student, and random factors. School factors include differences in school size, instructional policies, and facilities, which may cause student performance to fluctuate. The quality of teaching among teacher factors is central to student achievement. The direct beneficiaries of teachers' teaching are students, and differences in the quality of teaching are key factors in causing student performance to fall short of expectations. Student factors include student demographic features (e.g., gender, age, economic status, number of courses attended, Internet access, etc.), behavioral features (e.g., study skills, learning methods, study habits learning strategies, perceived social support, motivation, etc.), and sociodemographics (health patterns, etc.). Random factors are not discussed under the control of the value-added evaluation mechanism. The following model can represent the factors influencing value-added performance.

$$\begin{cases} P = C + T + S + \varepsilon \\ \hat{P} = C + S + \varepsilon \\ V = P - \hat{P} = T \end{cases} \tag{1}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

where P is the actual performance, and C, T, S, and ε are the contributions of school, teacher, student, and random factors to performance, respectively. $\hat{P}$ is the predicted performance, and the teacher factor is considered a constant when predicting the performance, so it is not expressed in the prediction. V is value-added to performance. It is clear from equation (1) that the teacher factor is the only contributor to value-added performance.

3.2.2 Performance prediction

Predicting student performance is the basis and prerequisite for obtaining value-added performance. The more accurate the predicted performance is, the more effective the value-added performance is, and the more representative the contribution of teacher factors to the value-added student performance, controlling for factors other than teacher factors.

In the area of performance prediction, previous research has been achieved. Alshanqiti and Namoun[12] argue that improving the accuracy of student performance prediction requires an in-depth understanding of the factors and features that influence student performance. Cruz-Jesus et al.[13] used 16 statistical features, including age, gender, attendance, and the number of courses, to predict students' academic performance by a machine learning algorithm. Using demographic features and mid-semester academic achievement, Fernandes et al.[14] used a gradient-boosting algorithm to predict student performance. The results showed that past performance and absence data were the best features for predicting performance. Musso et al.[15] proposed a machine learning performance prediction model based on learning strategies, motivation, health status, and socio-demographic features. The results indicated that learning strategies had the most significant impact on performance prediction. Bernacki et al.[16] attempted to use learning management system log records to predict performance, and he found that a behavior-based prediction model successfully predicted 75% of the learning failure population. Yağcı[17] uses machine learning algorithms to predict students' performance based on their midterm performance and departmental information, contributing to a certain extent to the prevention of unsatisfactory performance and helping to improve the quality of teaching.

Although these methods have solved the problems related to performance prediction to some extent, the evaluation indexes of the prediction models are all unsatisfactory. For example, the coefficient of determination (R Square, R^2) performed between 0.512 and 0.849, the Mean Absolute Error (MAE) all performed at 3.280 and above, and the Mean Squared Error (MSE) performed at 5.761 and above. This indicates that the model building of performance prediction still needs some improvement. Furthermore, collecting and processing such a wide variety of data takes much time and requires strong expertise and skills. Hoffit and Schyns[18] argue that collecting so much data and analyzing so many features is burdensome and unnecessary and that some features do not always give the desired predictive contribution.

Considering the above problems, this paper will use four categories (academic foundation, study habits, economic status, and region) and six sub-features (entrance performance, number of library check-outs, number of library entries, consumption times, average consumption, and student region) to predict students' performance, without using other statistical features and social data, aiming to save data resources for schools while improving the accuracy rate of performance prediction.

3.3. Feasibility

Presently, the application of value-added evaluation to teacher teaching quality evaluation is mainly at the basic education level, and the value-added evaluation at the higher education level is primarily at the school level but rarely at the teacher level. This study will use value-added student performance as a medium to apply value-added evaluation to the evaluation of the teaching quality of university teachers. The value-added performance will be obtained through

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

the difference in performance generated by each student receiving teaching from the teacher. Then the value-added performance will be used to evaluate and quantify teachers' teaching quality. It is feasible in both theory and reality.

1. Among the many value-added indicators for students, the value-added in performance is the one that best represents the degree of value-added in knowledge and competence and also highlights the professional performance of students in different academic areas after receiving teaching from teachers.

2. Doran et al.[19] argue that value-added models are viewed as a class of statistical models that explain what proportion of the change in student performance is affected by teacher teaching and that student learning gains can be measured by data on student performance over some time.

3. William L. Sanders et al.[20] applied value-added evaluation to educational evaluation at the end of the last century by using the difference between U.S. students' university entrance exam performance and graduate school entrance exam performance as school performance value-added to provide a scientific and objective evaluation of school effectiveness. The system was later incorporated into the official evaluation system to measure the effectiveness of U.S. higher education institutions.

4. Chetty et al. [21] address the question, "Is the value-added impact of teachers on student test performance a good measure of teacher quality?" This question was examined in a study of the school district and tax records of more than one million children. The findings showed that teachers with a greater positive value-added impact on performance were more likely to teach students who went on to university and were paid more.

It has been demonstrated in many scholarly articles in academia that student performance can be predicted by various types of features related to performance. The predictability of student performance provides for the capture of value-added performance.

The implementation of value-added evaluation can enhance the scientific and objective nature of the evaluation work for universities while reducing the work's difficulty and improving the evaluation's efficiency.

3.4. Purpose of the study

Based on big educational data, this study establishes a teaching quality evaluation system for university teachers based on value-added student performance by obtaining effective value-added performance through efficient and accurate performance prediction. It aims to make a more scientific, reasonable, objective, fast, and convenient evaluation of university teachers' teaching to improve the teacher recognition of evaluation results and the applicability of teaching management departments. It is essential to strengthen teaching quality assurance, enhance continuous improvement, improve teacher evaluation, and improve university governance. It is an innovation in teacher teaching quality evaluation in the era of big data.

Using value-added student performance to evaluate the teaching quality of university teachers is challenging. Many factors influence students' value-added performance in higher education. Even under the mechanism of value-added evaluation, it isn't easy to have a uniform standard for the evaluation results to be predicted in advance. Therefore, using value-added student performance to evaluate the teaching quality of university teachers requires methodological breakthroughs and innovations.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

4. DATA ANALYSIS & EXPERIMENTAL DESIGN

4.1. Data sources and features

The study uses data from the enrollment admissions, academic performance, and one-card spending of first-year students in a school in the class of 2018 and 2019. In the study, we selected two courses, "College English" and "Advanced Mathematics," as evaluation samples because they are almost mandatory courses for first-year students, the data records are relatively completer and more accurate, and the research results are more general and practical guidance. As detailed in Table 1 below, several datasets were merged, and privacy was processed appropriately.

Table 1 Some attributes of the dataset and their attribute description

Dataset	Properties	Property Description
Student Performance Dataset	Gender	Male, Female
	Credits	0~6
	Performance	0~100
	Student number	Student card number
	Examination subjects	College English, Advanced Mathematics
	…	…
Admissions slotting dataset	Admissions major	Applied mathematics, mechanical science, design science, etc.
	Candidate category	Rural freshmen/past graduates, urban freshmen/past graduates
	Region of origin	Zhejiang, Hubei, Anhui, etc.
	Total performance	0~750
	Performance by subject	0~150, Jiangsu province language and mathematics scores are 160 points
	…	…
Library management dataset	Book borrowing date	Details time
	Book return date	Details time
	Entrance time	Details time
	Departure time	Details time
	Student region	Provinces of China
	Student card number	Student card number
	…	…
Canteen campus card consumption dataset	Transaction amount	>0
	Consumption times	>0
	Student worker number	Student card numbers
	Card balance	>0
	Trading hours	September 2018 to July 2020
	…	…

4.2. Data pre-processing

Data pre-processing is a primary step in the research process, whereby raw data are processed into high-quality datasets that are easy to analyze. Data pre-processing consists of four steps: data cleaning, data integration, data reduction, and data transformation. Data cleaning mainly corrects and removes abnormal information. The dataset initially collected overall data of more than 7,000 visits, and after cleaning the abnormal data, the data became comprehensive data of 6,980 visits. Data integration organizes and merges data by combining the original data and grouping valid data with different structures and attributes. For example, the datasets in the study are grouped by student number, and the data belonging to the same

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

number in other datasets are integrated and grouped. Data reduction is a feature selection of the data, and this work will be performed in the following feature engineering. Data transformation is the standardization or discretization of data to change its form and discover new valid information. The continuous variables were normalized, and the processed data obeyed a normal distribution with the interval [0,1]. For example, each subject's results in the enrollment slotting dataset are standardized by the total value of the national subject scores, etc. We get a more analysis-friendly dataset by pre-processing, and next, we perform feature engineering on the dataset.

4.3. Feature Engineering

Feature engineering is a process that uses data-related knowledge to build and optimize relevant features so that machine learning algorithms arrive at the best performance. The feature engineering process consists of three parts: feature extraction, feature construction, and feature selection.

Feature extraction converts some original features into features with significant statistical significance. In this paper, using principal component analysis, the features of "Student source regions" were extracted from the part of each province and city into "Educationally underdeveloped regions," "Educationally developing regions," and "Educationally developed regions," which were coded with codes 1, 2 and 3 respectively. We extracted "University English" and "Advanced Mathematics" as new features from "Examination subjects" and gave them new feature names, "Advanced mathematics performance" and "English performance". The above features divided the dataset into the "College English performance dataset" and "Advanced Mathematics performance dataset". Accordingly, "English" and "Mathematics" are extracted from the " Performance by subject" feature as new features and given the new feature name "Academic foundation".

Feature construction refers to finding patterns from the original data to construct new features. Here the "Book borrow date" and "Book return date" are summed up by the "student card number" to create a new feature, "Number of library check-outs". In the same way, the "Entrance time" and "Departure time" are sorted and summed to obtain the new feature "Number of library entries". The new feature of "Average consumption" is obtained by dividing the "Transaction amount" by the "Consumption times" after adding up the "Transaction amount" according to the "Student worker number".

Feature selection selects a subset of the most statistically significant features from the feature set to achieve the effect of dimensionality reduction. The random forest algorithm was used here to rank the dataset features' importance. The Pearson correlation coefficient was used as the critical indicator; the results are shown in Fig.1.

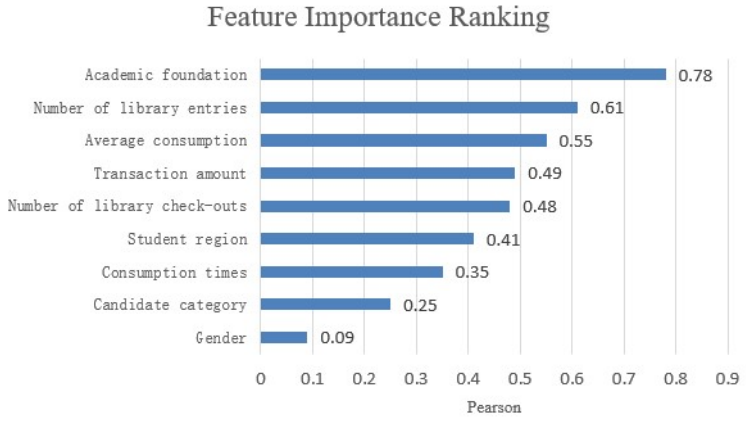

Figure 1. Feature Importance Ranking

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

4.4. Experimental design

4.4.1 Establishing university course performance-related features table

It is based on the available data and the above feature engineering as a theoretical basis to establish a table of features related to university course performance. As known from the preceding, the factors that influence value-added achievement are school factors, teacher factors, individual student factors, and random factors. Since the datasets used for the study were all taken from the same school, the effect of school factors on each student's performance can be viewed as a constant and can be left out of the discussion. Our study focuses on teacher and student factors as the core factors affecting course performance. Random factors occur mainly in isolated samples and are divorced from course grades under value-added evaluation mechanisms, so their effects are not considered. The student factors influencing course performance are complex and lengthy, and data collection is resource-intensive. We designed to collect personal factor characteristics in four category directions (academic foundation, study habits, economic status, and region) based on the importance of features in feature engineering and established a table of features related to university course performance based on student factors (Table 2).

Table 2. Characteristics of university students' performance

Property Category	Properties	Property Description
Academic foundation	Admission performance	0~1
Study habits	Number of library check-outs	≥0
	Number of library entries	≥0
Economic status	Average consumption	≥0
	Consumption times	≥0
Region	Student region	1, 2, 3

4.4.2 Prediction algorithms

The study used three regression prediction algorithms, Lasso, XGBoost, and Random Forest, to predict course performance. Lasso regression can effectively solve the problem of multicollinearity in multiple feature attributes of university student datasets. The XGBoost algorithm is highly flexible and can help us solve the problem of remaining missing values in the dataset by automatically learning the splitting direction of the missing values. Random Forest is one of the most widely used machine learning algorithms today. Its powerful performance and a high degree of randomness meet our needs for accurate student performance prediction.

(1) Lasso regression[23]

Lasso (Least absolute shrinkage and selection operator) is a compressed biased estimation algorithm with the idea of reducing the set of variables (Dimension reduction). It can compress the coefficients of variables and make some regression coefficients zero by constructing a penalty function, thus achieving the purpose of variable selection. Usually, the Lasso method has to standardize the dataset to eliminate the magnitude effect between variables.

Let the linear regression model be:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \varepsilon \quad (2)$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

Where β_0 is the constant term, $\beta_1, \beta_2, ..., \beta_m$ is the regression coefficient, and ε is the random disturbance term.

Define the Lasso estimate as follows.

$$\hat{\beta} = \arg\min_{\beta}^{2} \left\| y - \sum_{j=1}^{p} x_i \beta_i \right\|^2 + \lambda \sum_{j=1}^{p} |\beta_i| \tag{3}$$

Where λ is a non-negative regular parameter and $\lambda \sum_{j=1}^{p} |\beta_i|$ is a penalty term. Let the least squares estimate of β_j be $\hat{\beta}_j^0$, $\lambda_0 = \sum_{j=1}^{m} \left| \hat{\beta}_j^0 \right|$. When $\lambda < \lambda_0$, the Lasso absolute value estimate of the regression coefficient is smaller than the least squares absolute value estimate. Decreasing λ, the estimated values of some coefficients become small or even zero, causing the corresponding variables to be eliminated, and the selection of variables is achieved. When $\lambda \geq \lambda_0$, $\hat{\beta}_j^0$ is the global optimum of the Lasso estimate, at this point, the model will be selected into all variables and is not having a constraining effect.

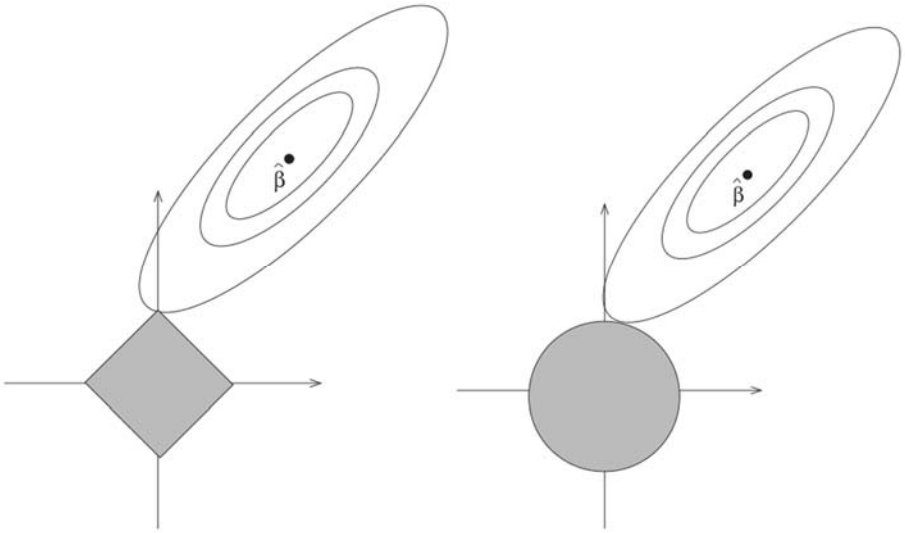

Figure 2. Comparison of Lasso estimates[23]

(2) XGBoost regression[24]

XGBoost (Extreme Gradient Boosting) is one of the boosting algorithms and is a highly scalable gradient boosting model. The Boosting algorithm aims to integrate many weak classifiers to form a robust classifier, and XGBoost is a boosting tree model. It is the integration of many tree models together to create a robust classifier. In terms of effect, XGBoost uses boosting to focus more on reducing bias to improve accuracy and uses L_2 regular terms to reduce model complexity to improve generalization ability. In terms of speed, XGBoost continuously optimizes the split point and uses parallelism, caching, and out-of-core computing in its implementation to optimize speedups. XGBoost supports different base learners, such as GBT for tree models and GBLinear for linear-based models. The custom loss function makes the loss function and the objective function decoupled. The algorithm is implemented as follows.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

A compositionally integrated model based on K trees can be represented as an additive model given a dataset.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathrm{F} \tag{4}$$

Where $\mathrm{F} = \left\{ f(\mathbf{x}) = w_{q(\mathbf{x})} \right\} \left(q : \mathbb{R}^m \to T, w \in \mathbb{R}^T \right)$ is the function space consisting of CART regression trees. q represents the structure of a tree that maps sample data to the corresponding leaf nodes, and T is the number of leaf nodes of the tree. An independent tree f_k is determined from the tree structure q and the leaf weights w. New sample data is mapped to the leaf nodes according to the tree's decision rules, and each leaf node's weight scores are summed to obtain the predicted value of that sample data. For the above additive model, XGBoost adds the L_2 regular term Ω to the GBRT objective function. The objective function of XGBoost is:

$$\mathrm{L}(\phi) = \sum_i l\left(\hat{y}_i, y_i\right) + \sum_k \Omega\left(f_k\right) \tag{5}$$

$$\Omega(f) = \gamma \mathrm{T} + \frac{1}{2}\lambda \|w\|^2 \tag{6}$$

Where the loss function l must be a differentiable convex function. The regular term Ω reduces the model complexity, and the regular term L_2 helps to smooth the learned weights and avoid overfitting. After t rounds of iterations of the forward distribution algorithm, the new objective function is finally determined.

$$\begin{aligned} \mathrm{L}^{(t)} &= \sum_{i=1}^{n} \left[g_i f_t\left(\mathbf{x}_i\right) + \frac{1}{2} h_i f_t^2\left(\mathbf{x}_i\right) \right] + \Omega\left(f_t\right) \\ &= \sum_{i=1}^{n} \frac{1}{2} h_i \left[f_t\left(x_i\right) - \left(-\frac{g_i}{h_i} \right) \right]^2 + \Omega\left(f_t\right) + \text{ constant} \end{aligned} \tag{7}$$

(3) Random forest regression[25]

Random Forest (RF) is one of the bagging algorithms, which is a model based on decision trees as a base learner to construct bagging and further introduce random attributes in the training process of decision trees. First, RF also uses CART decision trees as a weak learner. Second, RF makes improvements in the decision tree building process. For the ordinary decision tree, select an optimal feature among all the sample features on the nodes for the left and right subtrees. However, RF enhances the model's generalization ability by randomly selecting a portion of the sample features on the nodes for division. It should be noted that when the sample is unbalanced and more attention is paid to negative samples, the voting function needs further optimization. The algorithm flow is as follows.

Suppose there exists a dataset $D = \left\{ x_{i1}, x_{i2}, \ldots, x_{in}, y_i \right\} (i \in [1, m])$, and N is the number of features; at this point, a put-back sampling is performed to generate a sampling space $(m * n)^{m*n}$.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

Building base learners. Each sampling $d_j = \{x_{i1}, x_{i2}, \ldots, x_{ik}, y_i\} (i \in [1, m])$ (where $K << \mathrm{M}$) generates a decision tree and records the result of each decision tree as $h_j(x)$.

Training T times makes:

$$H(x) = \max \sum_{t=1}^{T} \phi\left(h_j(x) = y\right) \tag{8}$$

Where $\phi(x)$ is the voting algorithm.

4.4.3 Experimental method

After data pre-processing, we obtained the "College English performance dataset" and the "Advanced Mathematics performance dataset". Now both datasets are divided into two parts, 50% of the data as the training set and 50% as the test set, and the features are input according to Table 2. Training prediction models using Lasso, XGBoost, and RF, respectively. The model parameters are continuously tuned during the experiment until a prediction model with high accuracy is obtained. After the predicted result is obtained, the difference between the predicted and actual performance is the value-added performance. The teacher factor is the only contributor to this value-added performance. Finally, according to the student number and course selection information, the final quantitative value of teachers' teaching quality is obtained by summing up and averaging the course classes as a unit.

Assuming that the quantitative value of teacher teaching quality is E and the value-added of achievement is V, the formula for calculating teacher teaching quality is as follows.

$$E = \frac{1}{n} V = \frac{1}{n} \sum_{i=1}^{n} (P_i - \hat{P}_i) \tag{9}$$

where n is the course class size, P_i is the actual performance of student i, and $\hat{P}_i$ is the predicted performance of student i. The workflow is shown in Fig.3.

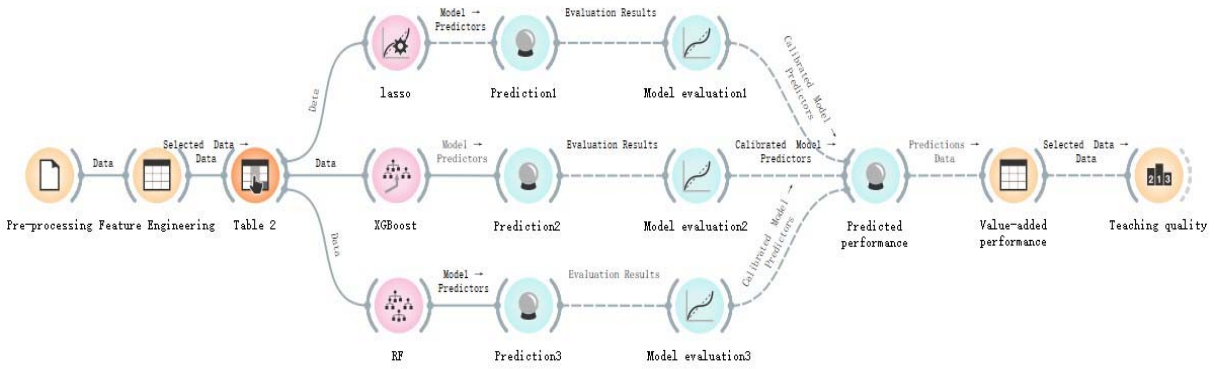

Figure 3. Workflow of performance value-added evaluation of teaching quality

The hardware environment for the experiment is Processor Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz Memory 16GB. The software environment for the experiments is Python 3.9. The parameter settings in the model are referenced below.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

Table 3. Model Parameters

Lasoo	XGBoost	RF
alpha=0.1 max_iter=1000 selection=cyclic tol=0.0001 …	max_depth=3 learning_rate=0.1 n_estimators=300 colsample_bytree=0.3863 random_state=4 …	max_depth=2 learning_rate=0.1 n_estimators=200 random_state=22 …

5. EXPERIMENTAL

5.1. Experimental results

The experiments used Lasso, XGBoost, and random forest to predict students' performance in "College English ①" and "Advanced Mathematics ②" based on Table 2. After feature selection and parameter tuning, the three models were trained and tested. To evaluate the accuracy of the performance prediction model, the study used the coefficient of determination (R2), Mean Absolute Error (MAE), and Mean Squared Error (MSE) as model evaluation indicators. The prediction model's evaluation indicators for the two courses' performance are shown in Table 4.

Table 4. Effect of the performance prediction model

Indicators	College English ①				Advanced Mathematics ②			
	Lasso	XGBoost	RF	Average results	Lasso	XGBoost	RF	Average results
R^2	0.851	0.897	0.905	0.912	0.837	0.911	0.907	0.918
MAE	3.580	2.385	2.406	2.301	3.512	2.447	2.408	2.387
MSE	4.511	1.931	2.014	1.871	4.288	2.101	1.985	1.930

According to Table 4, the fitting effect of the Lasso regression is average for predicting grades in both courses, with R^2 only 85.1% and 83.7%, and the MAE and MSE error indicators are also the highest. XGBoost was the best fit for the prediction of "Advanced Mathematics" course performance, with an R^2 of 91.1%, but the MAE and MSE error indicators were slightly higher than the random forest results. The random forest was the best fit for the "College English" course, with an R^2 of 90.5%, but the MAE and MSE error indicators were slightly higher than the XGBoost results. This situation may be caused by the random perturbations generated during the experiment. To prevent this interference, we combine the results of the two better-fitting models and take an arithmetic average to obtain an average result. The results show that the three indicators of the average results are at the current experimental optimum, and the random perturbations are controlled to some extent. The reason may be that after arithmetic averaging the results, the random space of perturbations is compressed, resulting in a minor effect of concerns on the averaged results, making the average result the current optimum.

The value-added of performance, as well as the quantitative value of teachers' teaching quality, can be obtained from equation (9), and the whole teacher teaching quality evaluation model is shown in Table 5.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

Table 5. Teaching quality evaluation model based on performance value-added

Teacher number	Student number	Predicted performance	Actual performance	Value-added performance	Teaching quality
Teacher A_m (College English)	1	75.22	73.78	-1.44	4.62
	2	78.37	76.56	-1.81	
	3	81.35	85.68	4.32	
	…	…	…	…	
	47	81.30	85.70	4.40	
	48	80.23	83.55	3.32	
	49	74.71	80.29	5.58	
…	…	…	…	…	…
Teacher A_M (College English)	1	80.75	82.98	2.23	-1.43
	2	80.30	75.41	-4.89	
	3	78.49	81.79	2.30	
	…	…	…	…	
	42	80.26	81.20	0.94	
	43	76.11	75.32	-0.79	
	44	79.46	77.94	-1.52	
…	…	…	…	…	…
Teacher A_n (Advanced Mathematics)	1	75.37	82.08	6.71	3.29
	2	77.23	86.01	8.78	
	3	78.35	80.26	1.91	
	…	…	…	…	
	58	78.67	77.65	-1.02	
	59	72.03	74.13	2.10	
	60	78.49	83.51	5.02	
…	…	…	…	…	…
Teacher A_N (Advanced Mathematics)	1	79.12	71.13	-7.99	-5.57
	2	75.78	79.11	3.33	

As seen from Table 5, the teacher teaching a course to a class and the performance prediction by the current prediction model yields the value-added performance in terms of the difference between the predicted and actual performance, contributed by the teacher factor. After summing up the average at the course level, the quantitative value of "College English" taught by teacher A_m is 4.62, which means that teacher A_m contributed an average of 4.62 points of value-added to the "College English" performance of the students in the class. The quantitative value of teacher A_M is -1.43, which means that teacher A_M contributes an average of -1.43 points of value-added to the "College English" performance of the students in the class. The quantified value of "Advanced Mathematics" taught by teacher A_n is 3.29, which means that teacher An contributed an average of 3.29 points of value-added to the student's performance in "Advanced Mathematics" in this class. The quantitative value of "Advanced Mathematics" taught by teacher A_N is -5.57, which means that teacher A_N contributed an average of -5.57 points of value-added to the student's performance in "Advanced Mathematics" in the class. This means that the teacher A_M and A_N teaching is a negative enhancement for the students. The significant variation in individual teachers' contributions to performance, in addition to differences in performance-related factors such as teaching level among teachers, maybe because individual students in the class have too weak a foundation in the courses taught and have difficulty improving under uniform teacher teaching, a situation that magnifies the impact of teachers' value-added contributions to performance.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

5.2. Comparative analysis

We used three indicators, R^2, MAE, and MSE, for comparative analysis with the experimental results of other researchers. Table 6 shows the performance prediction results of the performance prediction models with different features selected for the current dataset.

Table 6. Comparison of model effects

Model	Features	Algorithm	R^2		MAE		MSE	
			min	max	min	max	min	max
Cruz-Jesus [13]	Study period, gender, age, years of enrollment, scholarships, etc.	ANN, LR, SVM	0.512	0.811	4.334	10.690	7.451	23.250
Fernandes[14]	Classroom use environment, gender, age, student welfare, city, etc.	GBM	0.849		3.280		5.761	
Musso[15]	Learning strategies, coping strategies, cognitive factors, social support, etc.	ANN	0.807		4.631		7.010	
Bernacki[16]	Logging in the Learning Management System	LR,NB,J-48 DT,J-Rip DT	0.537	0.673	7.190	8.730	39.260	45.380
Yağcı[17]	Midterm tests, department information, teachers' information	RF, NN, SVM, LR, NB, CNN	0.699	0.746	5.986	7.031	8.917	16.311
Hoffart & Schyns[18]	nationality, research, prior education, mathematics, scholarship	LR, ANN, RF	0.704	0.841	3.820	6.313	5.790	11.013
Current model	Academic foundation, study habits, economic status, region	Lasso, XGBoost, RF	①0.851 ②0.837	①0.912 ②0.918	①2.301 ②2.387	①3.580 ②3.512	①1.871 ②1.930	①4.511 ②4.288

Table 6 shows that the prediction of the current experimental model is improved over previous studies by other researchers for the same dataset. R^2 improved by 0.063 to 0.406, MAE decreased by 0.893 to 8.389, and MSE decreased by 3.831 to 43.509, and these indicators verified the model's validity. In particular, the experimental model not only improves model effectiveness but also reduces the resource consumption required for data feature acquisition, allowing more accurate performance prediction by collecting only the necessary features according to four categories.

6. CONCLUSION & OUTLOOK

6.1. Conclusion

This study applies the value-added performance evaluation method to university teachers' teaching evaluations. It establishes a model for evaluating university teachers' teaching quality based on value-added performance, which provides a new method for university teaching evaluation. By analyzing the data of university students, the study established an innovative table of characteristics related to university course performance. It made accurate regression predictions of students' performance in "College English" and "Advanced Mathematics" courses based on only six features collected from four aspects while reducing the difficulty of data acquisition. The difference between the predicted and actual student performance is used to

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

obtain the value-added by teacher factors. The value-added is used to quantify the teacher's teaching quality at the class level.

Evaluating teachers' teaching quality by the standard of value-added students' performance can make reasonable and scientific teaching evaluation of university teachers' teaching work, change the evaluation mode of universities which used to be too much based mainly on experience and subjective evaluation, achieve a scientific and intelligent evaluation of teaching, and promote the development of education reform process. At the university level, value-added evaluation of student performance facilitates more rational management decisions and educational reform. At the teacher level, the value-added of student performance has become an important indicator of teacher effectiveness. Understanding the results of scientific evaluations allows them to adjust their teaching strategies better and improve the quality of their teaching. From the government and societal level, universities make the supervision and management of universities more transparent and more convenient for the government and the public by collecting students' value-added performance, thus enhancing university operations' transparency.

6.2. Outlook

The following aspects of the study still need to be improved and explored.

1. Due to the limitations of data and space, the quantitative values of teachers' teaching quality have not been standardized, and we will standardize the quantitative values of teachers' teaching quality using the percentage system in the subsequent study.

2. The traditional evaluation data can be compared and analyzed with the evaluation results based on value-added student performance. For example, when there is a significant difference between the conventional and value-added performance evaluation results, further analysis can be conducted on what features such teachers have and the reasons for such differences.

3. Given that most university teachers teach multiple courses, further consideration can be given to how teaching quality will be quantified in situations that affect multiple course delivery.

4. The teacher factor is a multi-feature fusion factor, not only the quality of teaching represented by one feature; if we can get the teacher information data matching with the current dataset, we will analyze the multi-feature fusion factor and quantify the "quality of teaching" feature more accurately.

REFERENCES

[1] Hanushek E A, Rivkin S G. The distribution of teacher quality and implications for policy[J]. Annual Review of Economics, 2012, 4(1):131.

[2] Jiang H. Teacher evaluation in universities from the perspective of high-quality education system [J]. Journal of Hebei Normal University/Educational Science Edition, 2022, 24(2):6.

[3] Wang L, Gao N. The construction of the teaching quality evaluation index for undergraduate courses- a perspective based on student experience [J]. Research in Higher Education of Engineering, 2021.

[4] Sun Y, Shen Y, Zhao J. Constructing and implementing an evaluation index system of mathematics classroom teaching quality (EIMT)[J]. Bulletin des Sciences Mathematics, 2021, 60(6):6.

[5] Foote D A, Harmon S K, Mayo D T. The impacts of instructional style and gender role attitude on students' evaluation of faculty[J]. Marketing Education Review, 2003, 13(2):9.

[6] Lei W, Ma H, Huang H. Teacher teaching quality evaluation based on residual decomposition technique of student achievement [J]. Journal of East China Normal University（Educational Sciences）, 2021, 39(7):8.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0012

[7] Ren Y, Yu X. A fuzzy multilevel evaluation model of classroom teaching quality in primary and secondary schools [J]. Microcomputer Applications, 2022, 38(3):5.

[8] Liu J. Research on the evaluation of teaching quality of university catechism based on support vector regression [J]. Information Technology, 2022, 46(3):6.

[9] Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance[J]. Journal of Educational & Behavioral Statistics, 2004, 29(1):11.

[10] Harvey, L. Analytic quality glossary, quality research international.[EB/OL]. (2021-12-01)[2021-12-01]. http://www.qualityresearchinternational.com/glossary//.

[11] Ouma C A. Performance of CART-based value-added model against HLM, multiple regression, and student growth percentile value-added models[J]. Dissertations & Theses - Gradworks, 2014.

[12] Alshanqiti A, Namoun A. Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification[J]. IEEE Access, 2020, 8:203827.

[13] Cruz-Jesus F, Castelli M, Oliveira T, et al. Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country[J]. Heliyon, 2020, 6(6).

[14] Fernandes E, Holanda M, Victorino M, et al. Educational data mining: Predictive analysis of the academic performance of public-school students in the capital of Brazil[J]. Journal of Business Research, 2019, 94:335.

[15] Musso M F, Hernández, Carlos Felipe Rodríguez, Cascallar E C. Predicting key educational outcomes in academic trajectories: a machine-learning approach[J]. 2020.

[16] Bernacki M L, Chavez M M, Uesbeck P M. Predicting Achievement and Providing Support before STEM Majors Begin to Fail[J]. Computers & Education, 2020, 158(6):103999.

[17] Yağcı, M. Educational data mining: prediction of students' academic performance using machine learning algorithms[J]. Smart Learn. Environ, 2022, 9(1). https://doi.org/10.1186/s40561-022-00192-z.

[18] Hoffait A S, Schyns M. Early detection of university students in potential difficulty[C]// 2017:963.

[19] Doran, H. C., Lockwood, J. R. Fitting value-added models in R[J]. Journal of Educational and Behavioral statistics,2006, 31(2):205.

[20] Sanders W L. Value-Added Assessment from Student Achievement Data: Opportunities and Hurdles[J]. Journal of Personnel Evaluation in Education, 2000, 14(4):329.

[21] Chetty R, Friedman J N, Rockoff J E. Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates[J]. American Economic Review, 2014, 104(9):2593-2632.

[22] Chetty R, Friedman J N, Rockoff J E. Measuring the impacts of teachers II: teacher value-added and student outcomes in adulthood[J]. National Bureau of Economic Research Working Paper Series, 2013, 19424.

[23] Tibshirani R J. Regression Shrinkage and Selection via the LASSO[J]. Journal of the Royal Statistical Society. Series B: Methodological, 1996, 73(1):273.

[24] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[J]. ACM, 2016.

[25] Breiman L. Random Forests[J]. Machine Learning, 2001.