SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 5799-5813
Open Access. Online & Print

DOI: 10.5281/zenodo.19400861

TOWARD A FRAMEWORK FOR PREDICTING ASSESSMENT OF TEACHER BY STUDENTS

Ricardo Ordoñez-Avila[1], Nelson Salgado Reyes[2*], Jorge Luis Veloz Zambrano[3], Jaime Meza[1], Sebastián Ventura[4] and Ramiro Pilaluisa Quinatoa[5]

[1]Departamento de Sistemas Computacionales, Universidad Técnica de Manabí (UTM), Portoviejo, 130105, Ecuador
ermenson.ordonez@utm.edu.ec
https://orcid.org/0000-0003-2583-2076
jaime.meza@utm.edu.ec
https://orcid.org/0000-0003-4960-7957
[2]Facultad de Hábitat, Infraestructura y Creatividad, Carrera Sistemas de Información, Pontificia Universidad Católica del Ecuador (PUCE), Quito, 170129, Ecuador
nesalgado@puce.edu.ec
https://orcid.org/0000-0001-8908-7613
[3]Departamento de Tecnologías de Información, Universidad Técnica de Manabí (UTM), Portoviejo, 130105, Ecuador
jorge.veloz@utm.edu.ec
https://orcid.org/0000-0002-9001-4478
[4]Departamento de Informática y Análisis Numérico, Universidad de Córdoba (UCO), Córdoba, 14071, España
sventura@uco.es
https://orcid.org/0000-0003-4216-6378
[5]Facultad de Ingeniería y Ciencias Aplicadas, Carrera de Ingeniería Civil, Universidad Central del Ecuador (UCE), Quito, 170521, Ecuador
jpilaluisa@uce.edu.ec
https://orcid.org/0000-0001-6949-350X

Received: 20/02/2026
Accepted: 30/03/2026

Corresponding Author: Nelson Salgado Reyes (nesalgado@puce.edu.ec)

ABSTRACT

Higher education has progressively incorporated teacher evaluation processes based on student perception, reflecting the complexity of measuring teaching practice with student data. The objective of this study was to construct a predictive model for teacher evaluation based on grade data and interaction in online courses. The methodology was based on CRISP-DM, utilizing Moodle records from five university courses, with a process that involved integration, cleaning, transformation, modeling, optimization, and validation using error metrics and complexity criteria. Nonlinear regression models based on Gradient Boosting, Random Forest, and XGBoost tree ensembles demonstrated the best performance compared to MLP Regressor, SVM Regressor, AdaBoost, Decision Tree, Extra Trees, and KNN. Gradient Boosting outperformed the others with coefficients of R2 = 0.63, MAE = 0.51, RMSE = 0.72, and AIC and BIC complexity values below 5.9, after recursive feature

elimination with cross-validation and hyperparameter optimization. The student's autonomous score and teacher interactions in the virtual classroom were the most influential predictors. A framework is proposed that includes explainability techniques, such as SHAP, for interpreting the results of teacher evaluation predictions. Future research could expand the set of variables and data, integrate deep learning approaches, and apply this framework in different instructional designs and universities.

KEYWORDS: Educational data mining, prediction, teacher evaluation, student performance, autonomous score, higher education, online modality, learning management system.

1. INTRODUCTION

The quality of higher education is a focus of attention in society and an ongoing management process aimed at maintaining academic standards and promoting continuous improvement (Akhtar et al., 2025). In Latin America, in addition to coverage, quality is one of the most studied variables in higher education, establishing indicators such as internet use, the availability of local research, and teacher training as support for university competitiveness (López-Leyva, 2020; Rama & Cevallos, 2016). Due to this competitiveness, many universities are adopting competency-based education models that guide teachers' interventions in course analysis and pedagogical improvement, resulting in improved outcomes in current teaching (Brauer, 2021; Malhotra et al., 2023; Vargas et al., 2024). As a result, teachers face the ongoing challenge of refining their teaching practices based on the results obtained in teacher evaluations.

Questionnaires administered to students are commonly used as a source for teacher evaluation, a process known as hetero-evaluation (Noriega et al., 2018; Yépez et al., 2023). Due to its complex nature, this form of evaluation has sparked considerable interest among researchers in identifying the most important aspects of student evaluation and the characteristics at play in different scenarios and contexts. Carpenter et al. (2020) mention that specific characteristics of teachers, students, and subjects may be associated with biased student assessments.

The bias that may arise in the use of questionnaires to evaluate teachers is widely discussed; however, the usefulness of considering variables such as student performance and motivation to learn, which are linked to teaching practice (Castro Morera et al., 2024; Schiefele, 2017), is also recognized. Specifically, student study, effort, and autonomy are indicators of their interest and motivation, which are partially dependent on the didactic organization of the subject (Ordoñez-Avila et al., 2025; Robertson & Padesky, 2020; Roksa et al., 2017). Jones et al. (2024) mention that teacher and course evaluations may be similar; they also suggest that student interest reflected in coursework and teaching methods may be the strongest predictors of students' evaluations of their teachers.

Both online and face-to-face education have implemented student evaluation of teachers (Ordoñez-Avila et al., 2023; Vélez-Loor, 2024). Efforts are being made to mine data generated by learning management systems for the analysis of teacher evaluation (Chen et al., 2025; Fan & Cheng, 2025; L. Wang, 2024). Data on student grades and interactions can help represent their academic performance and make inferences about the evaluation of teaching practice (Le et al., 2020; Maraza-Quispe et al., 2021; Ordoñez-Avila et al., 2025).

This study proposes a framework for predicting teacher evaluation by their students in online teaching, mining data from virtual learning environments at a higher education institution in Ecuador. To this purpose, the CRISP-DM methodology has been applied, utilizing educational data mining as a widely explored field that enables the analysis of data from various sources. The Moodle learning management system was used to extract information related to student academic performance, as represented by their grades, as well as their interaction with the platform, as reflected in specific characteristics of interest. This information was then used to predict students' perceptions of the teacher's teaching.

This framework can provide support in identifying improvements to design online teaching strategies, focusing on the conceptual relationships between teaching practice, student performance, and interest, while also exploring interactions on the platform, particularly analyzing data associated with performance. In this context, this research seeks to answer the following questions: Q1. What educational data mining techniques are applied in the prediction of teacher evaluation? Q2. What student performance data can explain teacher evaluation? Q3. How can the prediction of teacher evaluation based on student performance be validated and explained?

1.1. Modeling teacher evaluation with data mining

Supervised and unsupervised approaches have been employed in data mining for teacher evaluation. Classification techniques, including decision trees, support vector machines, artificial neural networks, and discriminant analysis, are used to build models based on student feedback. The C5.0 classifier has shown the best performance in terms of accuracy, precision, and specificity (Agaoglu, 2016). Methods such as Naive Bayes, logistic regression, Random Forest, and K-Nearest Neighbor have also been used to classify teacher performance based on student evaluations, along with clustering approaches (Babanli & Babanli, 2024; Ordoñez-Avila et al., 2023).

Regression techniques, such as linear and multiple regression, have also been employed to analyze the relationship between various factors and teacher evaluation results, aiming to identify how specific attributes influence overall scores (Shah et al., 2022; Sowmiya & Kalaiselvi, 2020).

Within the data analyzed in teacher evaluation, the analysis has focused on aspects such as teacher attitude, teaching content and methods, academic qualifications, classroom management, access to the virtual environment, stimulation of student motivation, classroom performance, and subsequent guidance (Demir, 2016; Peng et al., 2013; van Dijk et al., 2019; A. Yang & Yu, 2022).

2. METHODS

The methodology of this study is based on CRISP-DM (Cross-Industry Standard Process for Data Mining) used in educational data mining work (Guanin-Fajardo et al., 2024; Staneviciene et al., 2024). The stages of understanding the business and the data, data preparation, modeling, evaluation, and finally deployment were organized with a proposed framework for teacher evaluation based on student feedback.

2.1. Understanding The Business and the Data

Data was identified and collected by defining its origin and selecting variables relevant to student performance and teacher evaluation. The Moodle virtual learning environment of the Technical University of Manabí (UTM) served as the primary source, complemented by hetero-evaluation data from the Center for Evaluation and Quality Assurance. During this phase, the data was explored and described to assess its structure and quality. The main tasks included collecting records from Moodle's virtual classrooms and integrating the different sources. The dataset originated from Moodle version 3.11.6 and later, and contained information on grades and teacher–student interactions in UTM's online courses.

Python 3.8 and libraries such as pandas, numpy, and functools were used to manipulate the data. Extensive data integration work was carried out, including combining queries, appending, and transposing data. Each LMS log file and virtual classroom grade has a common field for the course code or identifier.

Data integration was carried out in three stages. The first stage focused on identifying characteristics of interest related to student performance, combining grade data and Moodle logs. Individual grade files contained between 30 and more than 300 instances per course, depending on the number of students enrolled in different subjects and courses. On the other hand, each raw event log file contained more than 200,000 instances in an unstructured format. Table 10 presents the details of the majors and courses, as well as measures of central tendency to differentiate the number of instances produced in each major.

Table 1: Instances by degree programs and courses in data integration

Degree program	Courses	Instances per course			T. Instances produced
		Average	Minimum	Maximum	
Law	41	215	76	436	9044
Economy	36	99	24	230	3946
Early Childhood Education	27	130	37	248	5843
Psychology	39	196	61	408	7838
Tourism	39	70	20	257	3250

The second stage of data integration focused on hetero-evaluation. Teacher interactions on Moodle were extracted, and all data from all courses were combined with student and teacher information. In the third stage of data integration, individual files of teacher interactions per course were combined with the median value of student performance. Finally, a final consolidated data set was generated that integrates student performance characteristics and interactions, along with teacher interactions and hetero-evaluation.

2.2. Data preparation

During the data preparation phase, normality, correlation, and multicollinearity tests were performed, along with cleaning and transformation processes. Derived variables were constructed, and the final set for modeling was consolidated. Initially, column headers were renamed, missing values were identified, outliers were verified, and measures of central tendency were analyzed. Normality was assessed using the Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling tests. Correlation was assessed using Spearman's coefficient for non-normal distributions, and multicollinearity was assessed using the variance inflation factor (VIF). Finally, integer values were transformed into numeric values to apply nonlinear regression techniques.

Subsequently, the construction of derived data was analyzed after the data integration stage. One of the important variables is the autonomous score (S_AUT), which explains the relationship with student performance (P_STU).

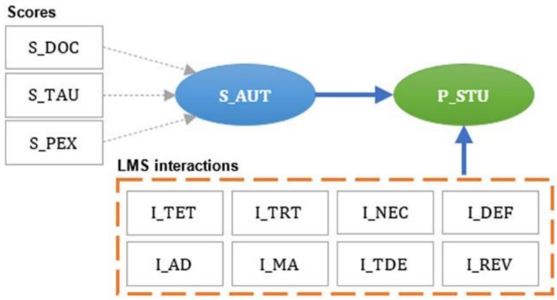

Figure 1. Causal representation of the autonomous score in student performance

Figure 1 summarizes the assessment of the autonomous score variable (P_AUT), together with related variables, in a causal representation diagram within student performance (P_STU). The observable variable of autonomous score was labeled based on student grades in an individual context. The teaching component grade (S_DOC), autonomous work grade (S_TAU), and experimentation and practical work grade (S_PEX) were added to the autonomous score, which was used to evaluate their relationship with student behavior in the virtual classroom. To this end, variables such as the task submission rate (I_TET), the task performance rate (I_TRT), failure to take the supplementary exam (I_NEC), the duration of the final exam in minutes (I_DEF), the median number of days accessed per month (I_ADM), the median number of accesses per day (I_MAD), download rate (I_TDE), and observed task feedback (I_REV) are listed as candidate predictors of the autonomous score and, consequently, of student performance.

Finally, with the final dataset, normalization and scaling tests were performed using the StandardScaler, MinMaxScaler, and RobustScaler methods. Table 2 shows the list of variables in the final dataset, describing their use for the analysis of student performance, hetero-evaluation, or both approaches.

Table 2. Fields selected according to the analysis and experimentation approach

Column	Data type	Analysis and experimentation approach
autonomous_score	Numeric	Student performance / Hetero-evaluation
task_submission_rate	Numerical	Student performance
test_ submission_rate	Numerical	Student performance
no_supplementary_exam	Logical	Student performance
final_exam_duration	Numerical	Student performance / Hetero-evaluation
feedback_viewed	Logical	Student performance
download_rate	Numerical	Student performance
Median_daily_access	Numeric	Student performance
median_days_per_month	Numerical	Student performance / Hetero-evaluation
total_course	Numerical	Hetero-evaluation
exam	Numerical	Hetero-evaluation
Num_st_supp_exam	Numerical	Hetero-evaluation
t_heteroevaluation	Numerical	Hetero-evaluation
t_median_days_per_month	Numeric	Hetero-evaluation
t_folder_viewed	Numeric	Hetero-evaluation
t_test_viewed	Numeric	Hetero-evaluation
st_excellent	Numerical	Hetero-evaluation
st_very_good	Numerical	Hetero-evaluation
st_good	Numerical	Hetero-evaluation
st_unapproved	Numerical	Hetero-evaluation

2.3. *Modeling*

During the modeling stage, test design, model construction, and optimization were carried out. In an initial modeling framework, data on student grades and interactions in the Moodle virtual classroom were analyzed. Initially, various regression techniques were tested, including MLP Regressor, SVM Regressor, Gradient Boosting, XGBoost, Random Forest, AdaBoost, Decision Tree, Extra Trees, KNN, and Linear Regressor. These algorithms were evaluated in a preliminary framework that allowed their performance to be filtered, enabling the differentiation of results between linear and nonlinear methods, and defining the models to be used in the final tests.

The modeling of teacher hetero-evaluation focused on tests with predictors based on student and teacher data, including student performance and their interactions in the classroom, as well as course results, which served as teacher predictors, in addition to their interactions in the virtual classroom.

To optimize the models, feature selection techniques such as recursive feature elimination with cross-validation (RFEcv), principal component analysis (PCA), and the least absolute shrinkage and selection operator with cross-validation (LASSOcv) were evaluated. Regarding hyperparameter optimization, grid search and randomized search techniques were tested.

2.4. Evaluation

Model evaluation metrics were defined for performance analysis and model complexity. The Shapley additive explanations (SHAP) method was also employed to explore potential feature combinations and their impact on prediction.

Root mean square error (RMSE) was used to measure the difference between the values predicted by the model and the observed values, calculated as the square root of the average of the squares of the differences between the predictions and the actual values (See Equation 1). In this study, it was used as the primary evaluation metric given its usefulness in nonparametric models.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (1)$$

Where y_i are the values predicted by the model, y_i represents the actual values, and n is the number of observations.

Mean Absolute Error (MAE) was used to measure the magnitude of the error between the predictions and the observed values (See Equation 2).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (2)$$

Where y_i are the predicted values, y_i are the actual values, and n is the total number of observations.

The coefficient of determination (R2) was used to indicate the proportion of variability in the dependent variable that the model explains (See Equation 3). Given the experiments with nonlinear models, this study utilized R-squared for informational and decision-making purposes in conjunction with the RMSE and MAE metrics.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (3)$$

Where y_i represents the values predicted by the model, y_i are the actual values, $\bar{y}$ is the mean of the observed values, and n is the total number of observations.

The Akaike Information Criterion (AIC) was used to evaluate the balance between the quality of a model's fit and its complexity; this metric penalizes the inclusion of unnecessary parameters (Kubara & Kopczewska, 2024). The AIC provides an asymptotically unbiased estimate of the mean prediction error (Kullback–Leibler divergence), which makes it helpful in selecting models that generalize well to new data (Cavanaugh & Neath, 2019). Equation 4 shows AIC formal definition.

$$AIC = -2\log(L) + 2k \qquad (4)$$

Where, L is the maximum value of the likelihood function, at this point, the mean square error (MSE) was used as a representation of likelihood, and k is the number of model parameters.

The Bayesian Information Criterion (BIC) also combines model fit with a complexity penalty, but does so from a Bayesian perspective (See Equation 5). Unlike the AIC, the BIC imposes a greater penalty on model complexity as the sample size increases, giving the best score to more parsimonious models. Its objective is to identify the model with the highest probability of being selected as the actual model (Cavanaugh & Neath, 2019).

$$BIC = -2\log(L) + k\log(n) \qquad (5)$$

Where, L is the maximum value of the likelihood function; at this point, the mean square error (MSE) was used as a representation of likelihood; k is the number of model parameters, and n is the sample size.

To apply AIC and BIC in this context, it was necessary to adapt the traditional formulas using the mean square error (MSE) metric. On the other hand, from a predictive perspective, AIC seeks to select the model that minimizes the expected prediction error for new data.

Deployment of teacher evaluation based on students

Figure 2 summarizes the framework applied for predicting teacher hetero-evaluation in online education environments, based on student data. This is structured in four stages: (1) research, supported by literature reviews focused on techniques and data used in teacher evaluation; (2) data collection from LMS systems, including interaction records and grades; (3) iterative modeling, which includes data selection, preprocessing, and transformation, the application of tree ensemble algorithms (Gradient Boosting, Random Forest, and XGBoost), hyperparameter optimization using GridSearchCV, and validation with performance metrics (R^2, MAE, RMSE) and complexity metrics (AIC, BIC); and (4) explanation and interpretation of the models, using explainability techniques such as SHAP to facilitate visualization, inferences, and the formulation of conclusions.

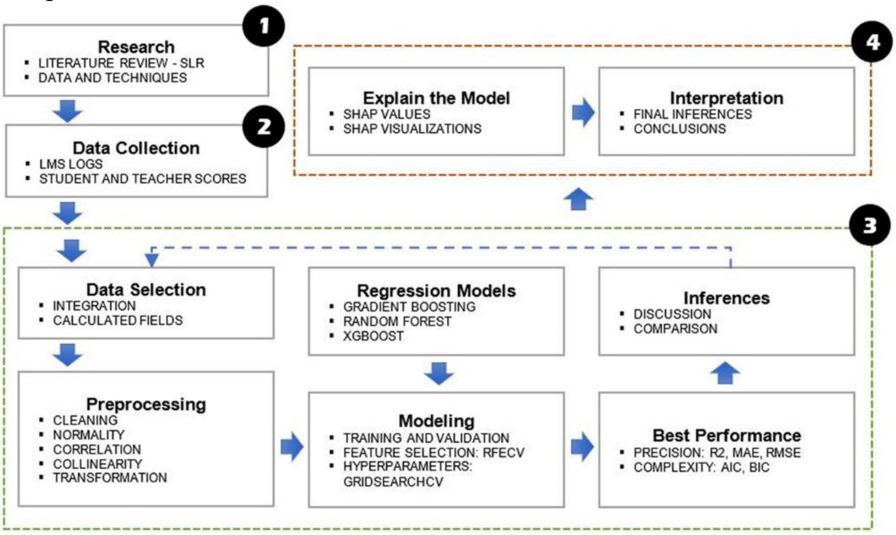

Figure 2. Framework for predicting teacher evaluation based on students, using LMS data

3. RESULTS

Nearly 30,000 instances were used to explore student performance in initial experiments. Data from 182 virtual classrooms in five online undergraduate programs were then consolidated for analysis of teacher evaluation based on student data.

Complex behaviors were identified in the data with a non-normal distribution. Figure 3 shows the distribution of the hetero-evaluation response variable. Frequency analysis identified an apparent positive asymmetry (left skew), evidenced by a significant concentration of values in the upper range, specifically between 98 and 100 points. This behavior suggests a high density of cases with maximum scores.

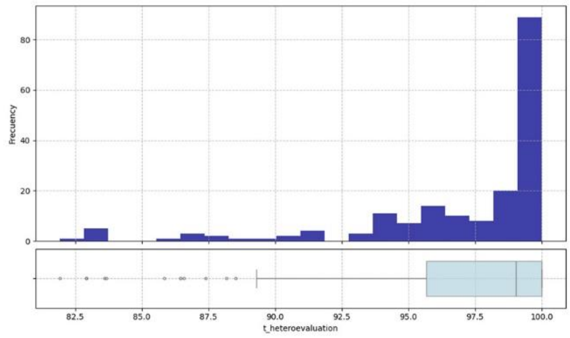

Figure 3. Distribution of the response variable, heteroevaluation

The data were normalized due to the diversity of scales in predictors and response variables. The Robust Scaler performed better than MinMax and Standard. Figure 4 shows the correlation results, using Spearman's coefficient to evaluate the monotonic relationship between the predictors and the dependent variable, heteroevaluation. The coefficients obtained vary between negative and positive values, reflecting different degrees and directions of nonparametric association. In sequence, the 90/10 training and testing ratio was selected for this experiment because it provided better results, given the number of instances per course analyzed.

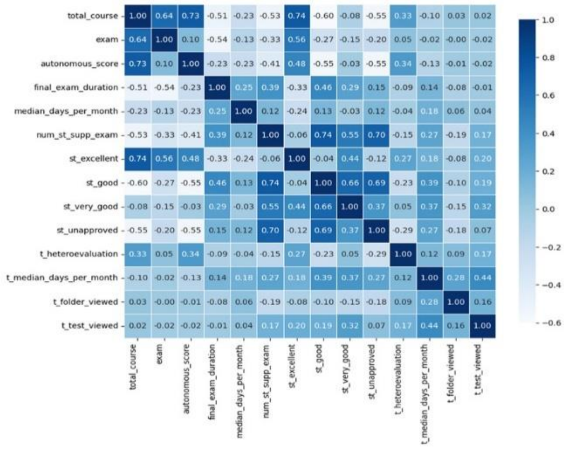

Figure 4. Distribution of the response variable, hetero-evaluation

Figure 5 supports the selection of Gradient Boosting, Random Forest, and XGBoost, which showed stable performance in the R2, MAE, and RMSE indicators, as well as in the AIC and BIC complexity metrics. These models were chosen in the preliminary experiments for further testing and final comparisons. Gradient Boosting achieved the best overall performance in this phase. In contrast, MLP Regressor, SVR, and KNN performed poorly, with high RMSE, AIC, and BIC values. Overall, Gradient Boosting and XGBoost stood out for combining a good predictive fit with adequate complexity assessment, considering all candidate predictors.

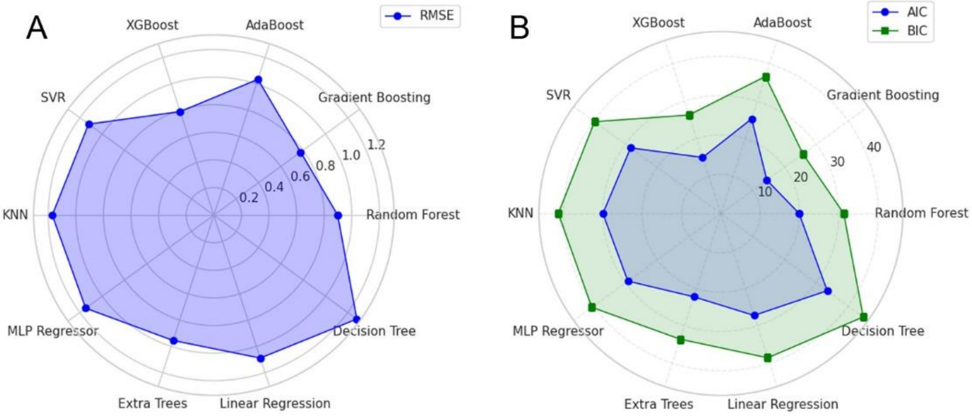

Figure 5. Preliminary performance of candidate models with RMSE (A) and AIC / BIC (B)

Recursive feature elimination with cross-validation (RFEcv) was identified as the most suitable method for variable selection in the analysis of student and teacher data in virtual environments. GridSearchCV with 10-fold cross-validation was applied for hyperparameter optimization. The hyperparameters {'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100} were evaluated in iterations between 4 and 12 features. Table 3 presents the results of RFEcv and hyperparameter optimization, with Gradient Boosting as the best-performing model.

Table 3: RFEcv results and hyperparameters in the Gradient Boosting model.

#	Selector Score R2	R2		MAE		RMSE		AIC		BIC	
		RFE	HYP	RFE	HYP	RFE	HYP	RFE	HYP	RFE	HYP
4	0.88	0.28	0.30	0.65	0.63	0.99	0.98	7.77	7,36	11,6	11,1
5	0.90	0.62	0.55	0.49	0.52	0.73	0.79	-2.09	1.03	2,64	5,76
*6	0.92	0.70	0.63	0.47	0.51	0.65	0.72	-4.54	-0.65	1.13	5,01
7	0.91	0.65	0.51	0.51	0.60	0.69	0.82	0.13	6,60	6,74	13,2
8	0.91	0.59	0.47	0.55	0.59	0.76	0.85	5.30	9,98	12,9	17,5
9	0.94	0.47	0.42	0.61	0.63	0.85	0.89	11.9	13,7	20,5	22,2
10	0.94	0.49	0.43	0.59	0.63	0.84	0.89	13.2	15,5	22,7	24,9
11	0.93	0.50	0.44	0.60	0.62	0.83	0.88	14,9	17,1	25,3	27,5
12	0.94	0.56	0.50	0.57	0.60	0.78	0.83	14.4	17,1	25,8	28,4

(*) Optimal number of variables with the best performance, RFE (Recursive Feature Elimination), HYP (Hyperparameters)

With the optimized models, explainability techniques were applied to facilitate the interpretation of results and inferences in the different contexts analyzed. Considering the cultural diversity of higher education and the particularities of instructional design in each institution, these techniques contributed to a better understanding of the findings.

Figure 6 presents the average feature importance derived from SHAP values in the Gradient Boosting model. The student's autonomous score was the most influential predictor of teacher evaluation, with an average impact of 0.25 units. It was followed by the teacher interaction variables t_questionnaire_viewed and t_median_accesses_days-month, both of which contributed 0.12, indicating a link between teacher activity in the virtual classroom and the perception of teaching practice. Additional variables related to student performance, such as st_excellent, st_good, and the number of supplementary exams, had minor effects but also contributed to the evaluation process.

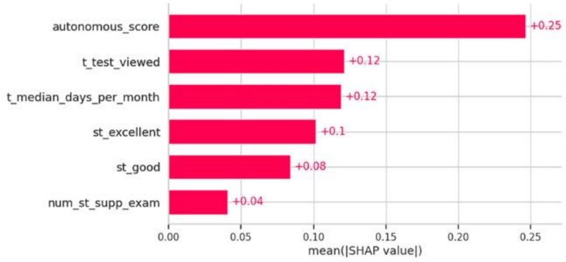

Figure 6. Average importance of characteristics in Gradient Boosting

Figure 7 displays the individual distribution of SHAP values in the Gradient Boosting model, illustrating how each feature influences the prediction of hetero-evaluation. The autonomous score shows the widest dispersion, with adverse effects on high predictions when values are low (blue) and positive effects when values are high. The variables t_cuestionario_visto and t_mediana_accesos_dias-mes follow a similar trend, with predictions increasing as their values rise, underscoring the role of teacher interaction in the virtual classroom. This visualization supports the interpretation of both the relative importance of predictors and the direction and magnitude of their effects.

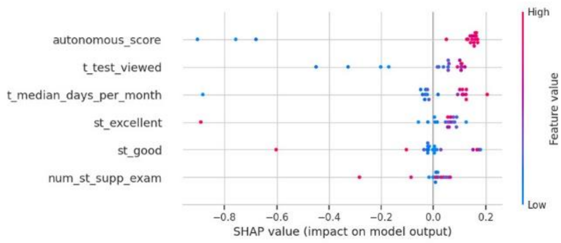

Figure 7. Distribution and impact of characteristics in Gradient Boosting

4. DISCUSSION

In response to the research question: What are the educational data mining techniques applied in teacher evaluation prediction? Regarding data mining techniques, specifically prediction algorithms, most of the work has focused on classification, as opposed to the problems analyzed using regression. In the context of classification, domains such as online education satisfaction analysis have employed logistic regression classifiers, K-nearest neighbors, support vector machines, and Naive Bayes (Chethan & Vinay, 2021; Poushy et al., 2022).

Neural networks have also been applied in the evaluation of higher education teachers, analyzing data from various actors, including students, peers, and administrators (Gao, 2025; Gu, 2022; G. Li et al., 2020). With the participation of experts and all stakeholders, neural networks have been combined with genetic algorithms to explain teacher evaluation in schools and universities (H. Zhang et al., 2021; Y. Zhang & Yang, 2022), including student performance data (Sowmiya & Kalaiselvi, 2021).

Other classification algorithms used are: Extra Trees Regressor, Random Forest Classifier, Logistic Regression, CV LGBM Classifier, Gradient Boosting Classifier, MLP Classifier, Extra Tree Classifier, Decision Tree Classifier, Label Propagation Gaussian NB, K Neighbors Classifier, SVM, Linear Discriminant Analysis, and Nearest Centroid (Abunasser et al., 2022; Agaoglu, 2016; Ahuja & Sharma, 2021; Esmaeel, 2020; D. Li, 2024). Algorithms such as Gradient Boosting Regressor and AdaBoost Regressor were reported in these studies, indicating a higher proportion of classification methods used and providing an opportunity to explore regression techniques for teacher evaluation analysis.

Decision trees have been reported as techniques used for teacher evaluation analysis (Hou, 2022; J. Sun, 2021), including in combination with other techniques such as fuzzy clustering (J. Wang, 2022). Likewise, optimization techniques such as the Dragonfly Algorithm have been used to improve classification accuracy (Y. Wang, 2022).

Regression techniques have been used to a lesser extent than classification techniques for predicting teacher evaluation. Ordinary least squares regression, for example, has been used, without employing student data, for teacher evaluation (Ng, 2025). Likewise, multiple linear regression and symbolic regression have been used in the context of teacher evaluation (Lino et al., 2016; L. Yang, 2021).

In this research, the Random Forest, Gradient Boosting, and XGBoost algorithms were used within a regression analysis. These techniques have not been widely explored, but recent studies confirm the improved performance of Gradient Boosting, for example, compared to other models such as SVM, K-Nearest Neighbor, Multi-layer Perceptron Neural Network, and Random Forest (J. Guo et al., 2025; Ibourk et al., 2023).

Gradient Boosting, Random Forest, and decision tree models have been applied to classification and regression tasks, incorporating grid search to optimize hyperparameters (Fawy et al., 2023). Random Forest enables the capture of nonlinear relationships by combining multiple trees that stabilize predictions (Geng & Zhai, 2023). Gradient Boosting and XGBoost build trees sequentially to minimize the loss function; the latter adds L1 and L2 regularization, which reduces overfitting and may be suitable in high-dimensional contexts. Both methods contribute to the handling of nonlinearity and offer greater stability in the face of outliers (Bentéjac et al., 2021).

In relation to the question of what student performance data can explain teacher hetero-evaluation, student grades are primarily identified (Paquibut et al., 2025; G. Sun, 2022; Xu & Liu, 2021). These include exam scores and final course grades (W. W. Guo, 2010; Hou, 2022). Likewise, some studies have analyzed the correlation between student grades and teaching evaluations (Luna et al., 2012; Mao et al., 2019). In general, the literature suggests considering academic performance as an input in teacher evaluation processes (Griffin, 2016).

In online higher education, especially in learning management systems such as Moodle, data on access, assignment submissions, social interactions, exams, grades, time spent, number of courses enrolled in, and activities completed have been used to analyze student academic performance (Maraza-Quispe et al., 2021; Quispe et al., 2021). The number of visits to the virtual classroom (Hernández-García et al., 2024; Le et al., 2020) and grades obtained in previous semesters (Park & Yoo, 2021) have also been taken into consideration. These variables, linked to activity and interaction in virtual environments, together with grades, have been used to estimate course performance, the results of which are discussed as indicators of teaching practice.

This study analyzed data on student interactions and grades. In the case of interactions, classroom access was examined in different categories to explore its relationship with hetero-evaluation. In terms of grades, the autonomous score, exam grade, and final course grade were considered, which were evaluated as predictors through correlation tests with the results of hetero-evaluation. In this way, both interactions and grades were integrated as performance indicators in the final models.

Finally, to answer the question, How can teacher evaluation prediction based on student performance be validated and explained?, classification tasks have used metrics such as accuracy, precision, sensitivity, and specificity, focusing on the analysis of correct and incorrect predictions. In the case of regression, validation has used residual error analysis with metrics such as RMSE, MSE, and MAE.

This study employed error metrics, including RMSE, MSE, and MAE, in conjunction with R^2 as an indicator of the variability explained by the data. Adjusted R^2 was not used, as the objective was to interpret the error directly; this metric is more appropriate for linear models based on ordinary least squares. For the nonlinear models used, additional criteria, such as AIC and BIC, were incorporated, allowing predictive performance to be balanced with complexity penalties.

In this context, AIC was applied to prioritize predictive power, while BIC offered an approach oriented toward parsimony and model interpretation, penalizing the inclusion of predictors more rigorously (Arando et al., 2021; Chicco et al., 2021; Karch, 2020). Finally, to facilitate the interpretation of the models, this research employed the Shapley method, which enabled the identification of variables both globally and locally. Other studies have combined this approach with the LIME method to expand the explanation of predictions (Salih et al., 2025; Vij & Nanjundan, 2022).

4.1. Limitations and future work

One limitation of this study is that the analysis was conducted using data from only one higher education institution, due to the difficulty in accessing and processing sensitive information. Likewise, demographic and socioeconomic variables of the students were not included, focusing exclusively on data related to the virtual classroom and interaction with the course. In this regard, it was decided not to apply imputation methods to preserve the originality of the data.

The data was requested by official letter to the Evaluation and Quality Assurance and Information Technology departments of the Technical University of Manabí. The analyzed data correspond to measurements and categories focused on grades and interactions in virtual classrooms. Consequently, to ensure the security and privacy of the names and confidential information of students and teachers, these were coded with labels during preprocessing for anonymous data processing, thereby maintaining the relationship of common fields and facilitating pattern analysis on the data.

Future research could consider the balance of scores in hetero-evaluation, apply the model in other virtual environments, expand the dataset, or integrate deep learning techniques, including hybrid approaches that combine feature selection with neural networks.

New variables could also be incorporated, such as the rate of material downloads or participation in forums, which would provide additional indicators of student engagement. These adjustments enable the exploration of various pedagogical practices and enrich the databases with complementary dimensions for further research.

5. CONCLUSIONS

Data from the Moodle LMS for online courses were used, considering grades and interactions between students and teachers in each course. Synthetic data were not used to avoid biases that would limit generalization, recognizing the unbalanced nature of the information. We relied on ensemble models based on weak trees, which, through iterative adjustments of the residual error, enable improved predictive power and the recognition of rare but significant patterns.

The variables analyzed constitute an initial proposal of predictors applicable in other contexts. The use of the median instead of the mean in specific functions was highlighted to improve the correlation between predictors and the response variable, especially in the presence of outliers and asymmetric distributions. The models allowed us to explore the prediction of students' autonomous scores based on interaction patterns in the LMS, identifying the median number of accesses per day per month as a more balanced predictor compared to the median number of daily accesses.

The optimal fit of the models was conditioned by the nature of the data and the configuration of the virtual classroom. In open environments, where task grading and material downloads are not mandatory, models must adapt to unstructured and diverse interactions. Recursive feature elimination with cross-validation (RFEcv) allowed for a more accurate selection of predictor variables compared to methods such as PCA and LASSOcv. At the same time, parameter estimation facilitated the detection of the best fit.

Gradient Boosting and XGBoost showed the most consistent performance values, although with higher computational demands compared to simpler models. Ten-fold cross-validation proved more efficient than Leave-One-Out, as it reduced computational cost. Hyperparameter tuning did not significantly modify the performance metrics (R^2, MAE, and RMSE), but it allowed for balancing the importance of predictors in the final model. Finally, the results obtained using explainability methods, such as Shapley, identified the student's autonomous score as the predictor with the most significant influence on teacher evaluation.

Future implementations of robust models such as XGBoost or Gradient Boosting in learning management systems could facilitate adaptation to new data and behaviors, considering the diversity of instructional designs, and adjust predictions to support teaching and learning processes.

Author Contributions: Conceptualizing: R.O.A., J.M, S.V; methodology, R.O.A, N.S.R, J.L.V.Z, S.V; formal analysis, R.O.A, N.S.R, J.L.V.Z; investigation, R.O.A, N.S.R, J.L.V.Z, R.P.Q; data curation, R.O.A, N.S.R, J.L.V.Z; writing—original draft preparation, R.O.A; writing—review and editing, R.O.A, N.S.R, J.L.V.Z, R.P.Q; visualization, R.O.A, J.L.V.Z, N.S.R; supervision, J.M, S.V; project administration, R.O.A. All authors have read and agreed to the published version of the manuscript

REFERENCES

Abunasser, B. S., Al-Hiealy, M. R. J., Barhoom, A. M., Almasri, A. R., & Abu-Naser, S. S. (2022). Prediction of Instructor Performance using Machine and Deep Learning Techniques. International Journal of Advanced Computer Science and Applications, 13(7), 78–83. https://doi.org/10.14569/IJACSA.2022.0130711

Agaoglu, M. (2016). Predicting Instructor Performance Using Data Mining Techniques in Higher Education. IEEE Access, 4, 2379–2387. https://doi.org/10.1109/ACCESS.2016.2568756

Ahuja, R., & Sharma, S. C. (2021). Stacking and voting ensemble methods fusion to evaluate instructor performance in higher education. International Journal of Information Technology (Singapore), 13(5), 1721–1731. https://doi.org/10.1007/S41870-021-00729-4

Akhtar, S., Bais, P., Bahadur, P. S., & Dwivedi, M. (2025). Worldwide drifts in quality management and accreditation in higher education. Global Perspectives on Quality Management and Accreditation in Higher Education, 415–432. https://doi.org/10.4018/979-8-3693-9481-6.CH017

Arando, A., González-Ariza, A., Lupi, T. M., Nogales, S., León, J. M., Navas-González, F. J., Delgado, J. V., & Camacho, M. E. (2021). Comparison of non-linear models to describe the growth in the Andalusian turkey breed. Italian Journal of Animal Science, 20(1), 1156–1167. https://doi.org/10.1080/1828051X.2021.1950054

Babanli, M., & Babanli, J. M. (2024). Ranking Faculties in Universities by Using Fuzzy Hierarchical Approach. Lecture Notes in Networks and Systems, 718 LNNS, 290–299. https://doi.org/10.1007/978-3-031-51521-7_37

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54(3), 1937–1967. https://doi.org/10.1007/S10462-020-09896-5

Brauer, S. (2021). Towards competence-oriented higher education: a systematic literature review of the different perspectives on successful exit profiles. Education + Training, 63(9), 1363–1390. https://doi.org/10.1108/ET-07-2020-0216

Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students' (mis)judgments of learning and teaching effectiveness. Journal of Applied Research in Memory and Cognition, 9(2), 137–151. https://doi.org/10.1016/J.JARMAC.2019.12.009

Castro Morera, M., Navarro-Asencio, E., & González Barbera, C. (2024). Are student evaluations of university teaching biased? Revista de Educacion, 1(404), 169–199. https://doi.org/10.4438/1988-592X-RE-2024-404-621

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. Wiley Interdisciplinary Reviews: Computational Statistics, 11(3), e1460. https://doi.org/10.1002/WICS.1460;SUBPAGE:STRING:ACCESS

Chen, M. S., Hsu, T. C., & Tu, Y. F. (2025). The use of learning analytics and educational data mining to analyze teachers' teaching in online environments: a systematic literature review. Interactive Learning Environments, 33(5), 3317–3332. https://doi.org/10.1080/10494820.2024.2443070

Chethan, G. S., & Vinay, S. (2021). Analytical Framework for Binarized Response for Enhancing Knowledge Delivery System. International Journal of Advanced Computer Science and Applications, 12(11), 507–516. https://doi.org/10.14569/IJACSA.2021.0121157

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, 1–24. https://doi.org/10.7717/PEERJ-CS.623

Demir, S. (2016). An Analysis of Pre-Service Teachers' Attitudes and Opinions Regarding the Teaching Profession via Q-Methodology. European Journal of Contemporary Education, 17(3), 295–310. https://doi.org/10.13187/EJCED.2016.17.295

Esmaeel, H. R. (2020). Analysis of classification learning algorithms. Indonesian Journal of Electrical Engineering and Computer Science, 17(2), 1029–1039. https://doi.org/10.11591/IJEECS.V17.I2.PP1029-1039

Fan, S., & Cheng, L. (2025). A Model of Big Data Analytics Applied to Teacher Evaluation Based on Enhanced Data Quality. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) , 15422 LNCS, 123–135. https://doi.org/10.1007/978-3-031-77088-3_9

Fawy, M., Hamza, M., Aziz, R. A. El, & Yafi, E. (2023). Towards Optimization of Math Pedagogical Techniques and Behaviors Using Decision and Ensemble Trees. 2023 4th International Conference on Artificial Intelligence, Robotics and Control, AIRC 2023, 120–129. https://doi.org/10.1109/AIRC57904.2023.10303052

Gao, Y. (2025). Deep learning-based strategies for evaluating and enhancing university teaching quality. Computers and Education: Artificial Intelligence, 8. https://doi.org/10.1016/J.CAEAI.2025.100362

Geng, J., & Zhai, Y. (2023). Research and Application of Prediction Model Based on Random Forest Algorithm. 2023 IEEE 3rd International Conference on Data Science and Computer Application, ICDSCA 2023, 935–938. https://doi.org/10.1109/ICDSCA59871.2023.10393439

Griffin, B. W. (2016). Perceived autonomy support, intrinsic motivation, and student ratings of instruction. Studies in Educational Evaluation, 51, 116–125. https://doi.org/10.1016/J.STUEDUC.2016.10.007

Gu, Q. (2022). Research on Teaching Quality Evaluation Model of Higher Education Teachers Based on BP Neural Network and Random Matrix. Mathematical Problems in Engineering, 2022. https://doi.org/10.1155/2022/5088853

Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). Predicting Academic Success of College Students Using Machine Learning Techniques. Data, 9(4). https://doi.org/10.3390/DATA9040060

Guo, J., Zheng, B., Zhang, L., Li, B., Guo, H., & Chen, M. (2025). Decision Tree with Light Gradient Boosting Algorithm for University Teacher Performance Evaluation. 3rd International Conference on Integrated Circuits and Communication Systems, ICICACS 2025. https://doi.org/10.1109/ICICACS65178.2025.10968311

Guo, W. W. (2010). Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction. Expert Systems with Applications, 37(4), 3358–3365. https://doi.org/10.1016/J.ESWA.2009.10.014

Hernández-García, Á., Cuenca-Enrique, C., Del-Río-Carazo, L., & Iglesias-Pradas, S. (2024). Exploring the relationship between LMS interactions and academic performance: A Learning Cycle approach. Computers in Human Behavior, 155. https://doi.org/10.1016/J.CHB.2024.108183

Hou, R. (2022). Application of Decision Tree Algorithm Based on Data Mining in English Teaching Evaluation. Wireless Communications and Mobile Computing, 2022. https://doi.org/10.1155/2022/5717895

Ibourk, A., Hnini, K., & Ouaadi, I. (2023). Analysis of the Pedagogical Effectiveness of Teacher Qualification Cycle in Morocco: A Machine Learning Model Approach. Lecture Notes in Networks and Systems, 637 LNNS, 344–353. https://doi.org/10.1007/978-3-031-26384-2_30

Jones, B. D., Topuz, K., & Sahbaz, S. (2024). Predicting undergraduate student evaluations of teaching using probabilistic machine learning: The importance of motivational climate. Studies in Educational Evaluation, 81, 101353. https://doi.org/10.1016/J.STUEDUC.2024.101353

Karch, J. (2020). Improving on adjusted R-squared. Collabra: Psychology, 6(1). https://doi.org/10.1525/COLLABRA.343

Kubara, M., & Kopczewska, K. (2024). Akaike information criterion in choosing the optimal k-nearest neighbours of the spatial weight matrix. Spatial Economic Analysis, 19(1), 73–91. https://doi.org/10.1080/17421772.2023.2176539;WGROUP:STRING:PUBLICATION

Le, M. D., Nguyen, H. H., Nguyen, D. L., & Nguyen, V. A. (2020a). How to Forecast the Students' Learning Outcomes Based on Factors of Interactive Activities in a Blended Learning Course. ACM International Conference Proceeding Series, 11–15. https://doi.org/10.1145/3404709.3404711

Le, M. D., Nguyen, H. H., Nguyen, D. L., & Nguyen, V. A. (2020b). How to Forecast the Students' Learning Outcomes Based on Factors of Interactive Activities in a Blended Learning Course. ACM International Conference Proceeding Series, 11–15. https://doi.org/10.1145/3404709.3404711

Li, D. (2024). An interactive teaching evaluation system for preschool education in universities based on machine learning algorithm. Computers in Human Behavior, 157. https://doi.org/10.1016/J.CHB.2024.108211

Li, G., Xiang, L., Yu, Z., & Li, H. (2020). Intelligent evaluation of teaching based on multi-networks integration. International Journal of Cognitive Computing in Engineering, 1, 9–17. https://doi.org/10.1016/J.IJCCE.2020.07.001

Lino, A., Rocha, Á., & Sizo, A. (2016). Virtual teaching and learning environments: Automatic evaluation with symbolic regression. Journal of Intelligent and Fuzzy Systems, 31(4), 2061–2072. https://doi.org/10.3233/JIFS-169045

López-Leyva, S. (2020). Strengths and weaknesses of Latin American higher education for global competitiveness. Formacion Universitaria, 13(5), 165–176. https://doi.org/10.4067/S0718-50062020000500165

Luna, J. M., Romero, J. R., & Ventura, S. (2012). Design and behavior study of a grammar-guided genetic programming algorithm for mining association rules. Knowledge and Information Systems, 32(1), 53–76. https://doi.org/10.1007/S10115-011-0419-Z

Malhotra, R., Massoudi, M., & Jindal, R. (2023). Shifting from traditional engineering education towards competency-based approach: The most recommended approach-review. Education and Information Technologies, 28(7), 9081–9111. https://doi.org/10.1007/S10639-022-11568-6/METRICS

Mao, Y., Zhu, Y., Zhang, S., Zhang, D., Zhang, F., & Fan, X. (2019). Detecting interest-factor influenced abnormal evaluation of teaching via multimodal embedding and priori knowledge based neural network. Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019, 1201–1209. https://doi.org/10.1109/ISPA-BDCLOUD-SUSTAINCOM-SOCIALCOM48970.2019.00171

Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., & Apaza-Huanca, J. M. (2021). Predictive Model of Student Academic Performance from LMS data based on Learning Analytics. ACM International Conference Proceeding Series, 13–19. https://doi.org/10.1145/3498765.3498768

Ng, K. (2025). Identifying high-quality teachers. Education Economics, 33(1), 93–120. https://doi.org/10.1080/09645292.2023.2286425

Noriega, J. Á. V., Castro, G. B., González, N. G. C., & Figueroa, F. L. M. (2018). Model for self-evaluation and hetero-evaluation of teaching practice in Normal Schools. Educacao e Pesquisa, 44. https://doi.org/10.1590/S1678-4634201844170360

Ordoñez-Avila, R., Meza, J., & Ventura, S. (2025). Mining autonomous student patterns score on LMS within online higher education. PeerJ Computer Science, 11, e2855. https://doi.org/10.7717/PEERJ-CS.2855/SUPP-4

Ordoñez-Avila, R., Salgado Reyes, N., Meza, J., & Ventura, S. (2023). Data mining techniques for predicting teacher evaluation in higher education: A systematic literature review. Heliyon, 9(3), e13939. https://doi.org/10.1016/J.HELIYON.2023.E13939

Paquibut, R. Y., Al-Jahwari, N., & Ram, N. (2025). Sustaining Teaching–Learning Effectiveness: Evaluating Course Evaluation. Studies in Systems, Decision and Control, 601, 503–517. https://doi.org/10.1007/978-3-031-92240-4_46

Park, H. S., & Yoo, S. J. (2021). Early Dropout Prediction in Online Learning of University using Machine Learning. International Journal on Informatics Visualization, 5(4), 347–353.

https://doi.org/10.30630/JOIV.5.4.732
Peng, H., Zeng, X., Yang, Q., & Wan, Z. (2013). Study of elements of developmental classroom teaching evaluation. Lecture Notes in Electrical Engineering, 225 LNEE(VOL. 3), 363–371. https://doi.org/10.1007/978-3-642-35470-0_44
Poushy, L. H., Bhuiyan, S. A., Parvin, M., Hossain, R. A., Moon, N. N., Nooder, J., & Mahbuba, A. (2022). Satisfaction prediction of online education in COVID-19 situation using data mining techniques: Bangladesh perspective. International Journal of Electrical and Computer Engineering, 12(5), 5553–5561. https://doi.org/10.11591/IJECE.V12I5.PP5553-5561
Quispe, B. M., Apfata, J. E. N., Figueroa, R. C. Q., & Solis, M. A. V. (2021). Design proposal of a personalized Dashboard to optimize teaching-learning in Virtual Learning Environments. ACM International Conference Proceeding Series, 77–84. https://doi.org/10.1145/3498765.3498777
Rama, C., & Cevallos, M. (2016). Nuevas dinámicas de la regionalización universitaria en América Latina. Magis, 8(17), 99–134. https://doi.org/10.11144/JAVERIANA.M8-17.NDRU
Robertson, D. A., & Padesky, C. J. (2020). Keeping Students Interested: Interest-Based Instruction as a Tool to Engage. Reading Teacher, 73(5), 575–586. https://doi.org/10.1002/TRTR.1880
Roksa, J., Trolian, T. L., Blaich, C., & Wise, K. (2017). Facilitating academic performance in college: understanding the role of clear and organized instruction. Higher Education, 74(2), 283–300. https://doi.org/10.1007/S10734-016-0048-2
Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. Advanced Intelligent Systems, 7(1). https://doi.org/10.1002/AISY.202400304
Schiefele, U. (2017). Classroom management and mastery-oriented instruction as mediators of the effects of teacher motivation on student motivation. Teaching and Teacher Education, 64, 115–126. https://doi.org/10.1016/J.TATE.2017.02.004
Shah, A. A., Uddin, M., Shah, R. U., & Adeel, M. (2022). Development of the Best-fit Models to Evaluate Academic Performance of the University Teachers: A Survey of Gomal University Khyber Pakhtunkhwa Pakistan. Interchange, 53(3–4), 505–528. https://doi.org/10.1007/S10780-022-09469-1
Sowmiya, J., & Kalaiselvi, K. (2020). Instructor performance evaluation through machine learning algorithms. Advances in Intelligent Systems and Computing, 1108 AISC, 751–767. https://doi.org/10.1007/978-3-030-37218-7_84
Sowmiya, J., & Kalaiselvi, K. (2021). Modelling higher education environment based on knowledge system transfer between instructor and learners using genetic algorithm. International Journal of System of Systems Engineering, 11(3–4), 268–283. https://doi.org/10.1504/IJSSE.2021.121458
Staneviciene, E., Gudoniene, D., Punys, V., & Kukstys, A. (2024). A Case Study on the Data Mining-Based Prediction of Students' Performance for Effective and Sustainable E-Learning. Sustainability (Switzerland), 16(23). https://doi.org/10.3390/SU162310442
Sun, G. (2022). Application of GA-BP Neural Network in Online Education Quality Evaluation in Colleges and Universities. Mobile Information Systems, 2022. https://doi.org/10.1155/2022/7846247
Sun, J. (2021). Decision Tree Classification Algorithm in College PE Teachers' Score Analysis. Advances in Intelligent Systems and Computing, 1303, 780–786. https://doi.org/10.1007/978-981-33-4572-0_112
van Dijk, W., Gage, N. A., & Grasley-Boy, N. (2019). The relation between classroom management and mathematics achievement: A multilevel structural equation model. Psychology in the Schools, 56(7), 1173–1186. https://doi.org/10.1002/PITS.22254
Vargas, H., Heradio, R., Farias, G., Lei, Z., & De La Torre, L. (2024). A Pragmatic Framework for Assessing Learning Outcomes in Competency-Based Courses. IEEE TRANSACTIONS ON EDUCATION, 67(2). https://doi.org/10.1109/TE.2023.3347273
Vélez-Loor, J. M. (2024). Teacher Evaluation Models for Online Teaching. Revista Electronica Educare, 28(3). https://doi.org/10.15359/REE.28-3.18487
Vij, A., & Nanjundan, P. (2022). Comparing Strategies for Post-Hoc Explanations in Machine Learning Models. Lecture Notes on Data Engineering and Communications Technologies, 68, 585–592. https://doi.org/10.1007/978-981-16-1866-6_41
Wang, J. (2022). Application of C4.5 Decision Tree Algorithm for Evaluating the College Music Education. Mobile Information Systems, 2022. https://doi.org/10.1155/2022/7442352
Wang, L. (2024). Classroom Teaching Evaluation Based on Data Mining Technology. Lecture Notes on Data

Engineering and Communications Technologies, 197, 409–421. https://doi.org/10.1007/978-981-97-1979-2_36

Wang, Y. (2022). Design of Chinese Teaching Evaluation System for International Students under the Background of Data Mining. Security and Communication Networks, 2022. https://doi.org/10.1155/2022/6556956

Xu, X., & Liu, F. (2021). Optimization of Online Education and Teaching Evaluation System Based on GA-BP Neural Network. Computational Intelligence and Neuroscience, 2021. https://doi.org/10.1155/2021/8785127

Yang, A., & Yu, S. (2022). Research on Teaching Evaluation System Based on Machine Learning. Mobile Information Systems, 2022. https://doi.org/10.1155/2022/9255064

Yang, L. (2021). Research on quantitative evaluation method of teachers based on multiple linear regression. Proceedings - 2021 13th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2021, 858–862. https://doi.org/10.1109/ICMTMA52658.2021.00196

Yépez, A., Torres, C., Ramayo, Y., & Morales, C. (2023). Student Feedback on Evaluation and Assessment Processes in Higher Education. Journal of Higher Education Theory and Practice, 23(5), 237–247. https://doi.org/10.33423/JHETP.V23I5.5969

Zhang, H., Xiao, B., Li, J., & Hou, M. (2021). An improved genetic algorithm and neural network-based evaluation model of classroom teaching quality in colleges and universities. Wireless Communications and Mobile Computing, 2021. https://doi.org/10.1155/2021/2602385

Zhang, Y., & Yang, S. (2022). The application of genetic algorithm and fuzzy evaluation method in Teachers' performance appraisal. Proceedings - 2022 International Conference on Artificial Intelligence of Things and Crowdsensing, AIoTCs 2022, 578–582. https://doi.org/10.1109/AIOTCS58181.2022.00123