Decoding Instructor Performance with Data Mining and Machine Learning

Sai Gopal Lanka, Namratha Addagada, Haasitha Ambati, Sumalatha Saleti
Department of Computer Science and Engineering
SRM University, Andhra Pradesh, India
<gopal_lanka,namratha_a,haasitha_ambati,sumalatha.s> @srmap.edu.in

***Abstract*—Educational Data Mining is an emerging discipline with the objective of improving educational quality and problems like instructor performance prediction. Instructor performance is being measured in this research using three datasets: the Turkiye Student Evaluation Dataset, the Poland University Dataset, and a Synthetic Dataset. Feature selection techniques like Recursive Feature Elimination and Random Forest-based selection are used. Seven classification models—Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machines (SVM), AdaBoost, Multi-Layer Perceptron (MLP) Classifier and XGBoost—are implemented on all the datasets, hence it is a 3×3 comparison. The best performing model, which is Decision Tree with SVM on the Synthetic Dataset, has an accuracy rate of 90.15%. A web application is also built with this model so that users can enter instructor information and get performance predictions with actionable recommendations.**

***Index Terms*—Instructor Performance, Educational Data Mining, Recursive Feature Elimination, Decision Tree, SVM, Web Application**

I. Introduction

The evaluation of teaching quality persists as an ongoing educational challenge. Educational Data Mining (EDM) represents an emerging discipline which implements data mining techniques and machine learning methods to enhance educational instruction while improving student achievement. The system uses student feedback together with course attributes and instructional strategies to forecast instructor performance.

The evaluation of instructor performance (IP) measures both teaching effectiveness and student management abilities and academic standard maintenance. The evaluation of instructor performance traditionally depends on student ratings combined with Model-Based Teaching (MBT) predefined teaching scores. The process of collecting detailed datasets that measure teaching variables and student engagement proves to be quite challenging.

The research project establishes dependable predictive models for instructor performance through machine learning methods while testing their performance using three datasets including the Turkiye Student Evaluation Dataset, Poland University Dataset, and a synthetically created dataset. The quality of input data improves through Recursive Feature Elimination (RFE) as a feature selection method. The analysis employs Decision Tree and Naive Bayes alongside Logistic Regression and SVM and AdaBoost and MLP Classifier and XGBoost because these models provide both interpretability and strong predictive capabilities in educational settings.

The primary contributions of this research are as follows:

- The 3×3 model-to-dataset comparison framework is established by running each machine learning method on all three datasets.
- A synthetic dataset is created to replicate the actual instructor evaluation behavior.
- Decision Tree + SVM has the highest (90.15% accuracy) on the synthetic dataset.
- A web application is developed to run the best performing model for making real time predictions of instructor performance.

The rest of this paper is organized as follows. Section II presents the literature review. Section III describes the Methodology. Section IV outlines the experimental results. Finally, Section V concludes the paper.

II. Literature Review

The authors in Paper [1] employed two of the most widely used data mining methods: stepwise regression and decision tree methods. For regression analysis, the stepwise regression method and for decision trees CHAID and CART algorithms are utilized. The stepwise regression process is one of the variable selection techniques in regression analysis. The data set contains all course sections from Fall 2004 to Summer 2009. The data gathered is based on 12 questions: To decrease number of independent variables and to find the patterns of relationship among them the factor analysis was utilized. Teachers having good prepared course syllabi, employing sat- atisfactory material, assist the students beyond lectures, mark the examinations equitably and timely receive improved appraisals. Another instructor trait; the employment status of the instructor which is not measured through the questionnaire is shown to be important.

The research paper [2] analyzed student evaluation of teaching (SET) survey data obtained from Gazi University in Turkey. The dataset consisted of 33 attributes. The authors have used Weka Ranker to rank the attributes. The 10 least important attributes have been dropped and the remaining attributes—which were the top 24—were considered for the next step. The authors used four different classification approaches, and for each type of classifier, they trained two models: one model with the top 24 attributes and the second trained and tested on all the attributes.The authors achieved their highest accuracy rate at 85% through implementing this procedure on the classifiers.

The research paper [3] analyzed a dataset which contained student evaluations of teaching (SET) surveys from Gazi University in Turkey. The weka ranker tool allowed the authors to rank 33 attributes in the dataset. The analysis eliminated the 10 least important attributes but retained the top 24 remaining attributes for further evaluation.The authors implemented four distinct classification approaches and developed two models for each type within these approaches. The first model received the top 24 attributes while the second model received all attributes for training and testing. The authors achieved their highest accuracy rate at 85%. The authors determined that eliminating the least important attributes failed to significantly enhance accuracy because the best outcome from the attribute removal procedure resulted in only a 2% accuracy boost for the classifiers.

The authors in [4] employed a two-layered system which incorporated an MLP and a decision tree. The authors used two distinct decision tree methods via ID3 and C4.5. The research data were derived from the academic unit of a university in Ondo State, South West Nigeria. The research data covered six years, from 2010 to 2015. The data comprised two categories of information: instructor information (ID, age, rank, experience, and qualification) and Course details (course ID and student satisfaction). The data contained course-related questions along with information about the instructor. The process started with attribute selection, followed by training and testing. The outcome indicated C4.5 outperformed ID3 decision trees by achieving 83 accuracy. The authors concluded that experience and rank attributes heavily influenced instructor performance but professional qualification had little impact on their performance.

The authors of paper [5] used the sudden increase in instructor demand during COVID-19 to create a predictive model of instructor performance which would aid educational institutions in their hiring process and support current instructors in improving their teaching effectiveness. The researchers developed a dataset containing more than 10,000 records from SET surveys across four semesters. The dataset included five attributes which were courtesy, punctuality, grooming, attendance and the instructor's ability to explain concepts clearly. The performance levels in the target variable extended from outstanding to very satisfactory and satisfactory and unsatisfactory and poor. The authors implemented a decision tree approach to forecast instructor performance.

The main concern of the work in [6] is predicting the performance of the instructor and studying the factors that influence students' success in order to enhance the education system quality. The most significant characteristics of the Turkey Student Evaluation records dataset were chosen, and 4 distinct classifications (J48 Decision Tree, Multilayer Perceptron, Na¨ıve Bayes, and Sequential Minimal Optimization) were run on it. The attribute selection affected the performance of classifiers, and varied results occurred in varied cases. It also noticed that performance of an instructor is principally determined by the amount of courses being taught.

The research 'Predicting Instructor Performance Using Data Mining Techniques in Higher Education' in [7] presents an innovative method for predicting instructor performance in higher education through data mining techniques. The research holds importance because it tackles the complicated process of instructor performance assessment which remains based on personal opinions and various elements such as student prejudices together with curriculum content and educational methods. The authors employed a dataset of 47 instructors and 1,676 student evaluations to predict four metrics of instructor performance: overall instructor rating, course quality, course difficulty, and student learning. The researchers applied four different data mining approaches to forecast instructor performance from the given dataset: decision tree, random forest, artificial neural network and discriminant analysis, and they evaluated their results. The C5.0 classifier achieved the highest accuracy rate of 92.3% together with precision at 94.4% and specificity at 92.5%. The authors discovered that teacher experience together with student attendance and course grade distribution proved to be significant factors affecting instructor performance.According to the authors this method serves to detect areas where teacher performance needs improvement and develop specific interventions that enhance teaching quality. The data mining research results about course grade distribution as a performance factor would lead to modifications in course design and grading standards to enhance student learning results and instructor assessment scores.This research demonstrates how data mining methods deliver valuable insights about instructor performance in higher education while supporting current initiatives to enhance teaching quality.

The researchers in the paper "Predicting Instructor Performance in Online Courses: A Multi-Criteria Approach Using Data Mining Techniques" [8] used data mining techniques to predict online instructors' performance by means of several factors.The researchers gathered data from a big online learning platform which included course ratings along with teacher ratings as well as student engagement and instructor experience and complexity of course. A model of predicting instructor performance was developed by the implement-comparison of decision trees and random forests and support vector machines.The authors assessed the performance of models using accuracy and precision and recall and F1 score measures. The highest accuracy of 86.6% was obtained using the random forest model which indicated that it had the capability of accurately predicting instructor performance.The authors conducted a sensitivity analysis to assess which criteria significantly affected the model's performance. The authors realized that course ratings along with instructor ratings were the most significant factors for anticipating instructor performance when student engagement and course complexity came as secondary predictors.This study illustrates how data mining techniques can forecast instructor success in online learning while indicating that performance assessment requires more than one criterion of evaluation.The model attained high accuracy values which reflect its capability to improve online education quality and students' learning experiences.

118 ML algorithms' performance were taken into account by Basem S. Abunasser et al. in [9], among which Extra Trees Regressor proved to be the top Machine Learning model based on accuracy, precision, recall and f1 score across a dataset that has been collected from the UCI repository. Besides this, the authors suggested a Deep Learning model comprising seven Dense layers: an input layer (33 features), five hidden layers (256,128, 64, 32, and 16 neurons), and an output layer with five classes and softmax function). This DL model yielded better results compared to the rest of the 18 algorithms in terms of accuracy, precision, recall and f1 score (98.92%).

n the paper [10], The article 'Development and validation of an observation protocol for measuring science teachers' modeling-based teaching performance' details the development and testing of a tool to assess the effectiveness of science teachers' modeling-based teaching (MBT) practices. Modeling-based teaching practices typically consist of the 5E's, which are Engage, Explore, Explain, Elaborate, and Evaluate. This essentially translates to the fact that the instructor must bring a problem which will actually involve the students to probe, followed by discussing their understanding of the problem to one another, clarify their understanding by relating that problem to new circumstances, relating those understandings to real life situations they face, and lastly the instructor must assess their understanding. The research entails developing an observation protocol on the basis of a literature review and interviews with science education professionals. The method was then piloted in real classroom settings and found to be a valid and reliable tool for assessing MBT practice. The authors suggest that this tool could be used to enhance scientific teacher training and professional development, as well as being a useful tool for scientists researching science education. Likewise, within our Synthetic Dataset, we dealt with the questions which would produce improved leads to forecasting the performance of the Instructor. We randomly populated the data for every entry and created a hybrid machine learning framework and received reviewed by Dr. Saleti Sumalatha. The subsequent step following review is data preprocessing and execution of out protocol/ framework.

In paper [11] The 'Addressing the feasibility of the teacher performance rate and accuracy scale as a treatment integrity tool' paper investigates whether a tool is effective thatused to measure the accuracy of implementation fidelity of a behavioral intervention by teachers. The tool examined is the Teacher Performance Rate and Accuracy Scale (TPRAS), used to measure to what extent teachers accurately implement a behavioral intervention with a great deal of correctness and consistency. The research demonstrated that the TPRAS was an effective and applicable instrument in measuring treatment integrity, and can be applied to quality control the implementation of behavioral interventions in schools. The authors believe that the TPRAS can be utilized in order to improve the quality of studies on the implementation of behavioral interventions in education, and assist in the professional development of teachers. Briefly, it is demonstrated that the TPRA scale is a valuable aid for the teacher supervisors since it supplies independent measures of teacher (or teaching assistant) effectiveness, provides a means whereby the teacher and the supervisor can ascertain if skill acquisition courses are being administered properly, is a permanent record of progress of both the student and the teacher, and provides a means of setting new targets. To put the same into one sentence, the authors essentially adopted the approach in which a group of people visit the classrooms and assess the performance of the instructor by collecting all their thoughts with the assistance of feedback from the students. Likewise, In synthetic dataset, we used the questions that are available in the student portal of SRMAP where students would be providing feedback about their profes-sors. The questions are posed with a group of non-teaching professionals of the academic department who keep an eye on the classroom progress and then construct the questions for which students would provide feedback for their professors. The information is generated randomly and is trained and tested with the machine learning framework which is developed (Fusion of Boosting and Recursive Feature Elimination).

To determine the most suitable characteristics for dataset preparation a synthetic dataset, [12] demonstrated best practices such that any teacher would receive optimized feedback on their performance. In [12],The paper 'Supporting students' meaningful engagement in The research investigates how epistemological messages could serve as tools to enhance student participation in scientific modeling through a case study of teaching methods that differ from each other. The study investigates the impact of two distinct teaching approaches through comparative analysis. The first approach relied on epistemological messages while the second approach did not use epistemological messages. The research examined how students participated in modeling activities. Research found Students who received epistemological instructions displayed better student engagement and better understanding of scientific modeling practices than students who did not receive this education. The authors believe this training will help students develop scientific modeling competency while also suggesting its potential use in science teaching courses to improve scientific learning quality.The paper [**?**] essentially helped us know the good questions to evaluate the instructor's performance at the primary level. This paper helped us realize that there are at least two levels of contrast teaching techniques, teacher-student interaction based questions such as "is teacher being fair and unbiased?", etc are required to evaluate the performance of an educator in a better way.

III. Methodology

A. Dataset Descriptions

To evaluate instructor performance using machine learning techniques, three distinct datasets were considered: the Turkiye Student Evaluation Dataset, the Poland University Dataset, and a synthetically generated dataset.

- **Turkiye Student Evaluation Dataset**: This dataset contains 5820 records collected from Gazi University, Turkey. Each record includes student responses to a 28-question evaluation of the instructor and the course, rated on a scale of 1 to 5.

- **Poland University Dataset**: Contains 8015 records, and each record answers to 9 student feedback questions on scale of 1-5. This dataset also includes average student evaluation scores along with instructor-specific features such as gender, number of degrees, teaching experience, and academic qualifications (e.g., master's or PhD).
- **Synthetic Dataset**: This dataset was programmatically generated to simulate real-world instructor feedback. It includes 11 evaluation questions with ratings from 1 to 5. The dataset was created to include an equal number of high, low, and mixed performance samples, and was shuffled to reduce model bias.

B. Data Preprocessing

- **Turkiye Dataset**: The analysis eliminated the unnecessary columns which included Satisfaction Index together with Q9 and Q10. The target variable received designation as *Recommend Instructor*. The dataset was partitioned into training and test sets in a ratio of 80 to 20.
- **Poland University Dataset**: The respective attributes' mean values served to impute missing data points. The correlation heatmap helped identify highly related features which were subsequently removed. The gender information was combined into one category. A new binary target variable *Result* was derived from calculating the average SET score. The solution to class imbalance problems was achieved through the implementation of SMOTE (Synthetic Minority Over-sampling Technique).
- **Synthetic Dataset**: The data received new ratings which were subsequently rearranged. The target variable originated from the assessment of ”Is a role model and lifelong mentor for me”. All values above the mean of the column received a label of 1 (satisfactory) while all other values received a label of 0 (unsatisfactory).

C. Machine Learning Models Applied

Turkiye Student Evaluation Dataset:

- *Approach 1 – Satisfaction Index*: The analysis employed multiple models which included Logistic Regression, Decision Tree, Bagging, Random Forest, AdaBoost, XGBoost, Voting Classifiers (Hard and Soft), and SVM. The highest accuracy was achieved by AdaBoost at 84.4% while the lowest accuracy was obtained by Soft Voting at 77.4%.
- *Recursive Feature Elimination (RFE)*: Random Forest reached 86.9% accuracy when using Logistic Regression with Stratified K-Fold Cross-Validation (10 splits, 3 repeats) but Decision Tree reached the lowest accuracy at 80.6%.
- *Approach 2 – Clustering*: The analysis employed K-Means and Agglomerative Clustering because of previous research findings. The Elbow and Silhouette methods showed that 2 clusters represented the best solution. The predicted cluster labels served as the new targets. The RFE and cross-validation pipeline was used to train classifiers.

Poland University Dataset: The system evaluated student performance data from previous semesters together with instructor qualifications and teaching experience. The analysis used four different models which were Logistic Regression and Decision Tree and Naive Bayes and Random Forest. Random Forest with RFE yielded 66% accuracy. The stacking-based ensemble model which used Decision Tree and Random Forest and KNN and XGBoost in the first layer and Logistic Regression as the meta-learner in the second layer reached 78.5% accuracy.

Synthetic Dataset:

A new approach was developed which merged boosting techniques with dynamic recursive feature elimination methods. The first model received training to produce predictions which became the new feature named ”First Prediction”. A second model was applied to the expanded dataset. R-squared values from linear regressions between independent variables and the dependent variable determined recursive feature elimination through iterative removal of low-contributing features.

Models Used:

- Decision Tree
- Logistic Regression
- Support Vector Machine (SVM)
- AdaBoost Classifier
- Multi-Layer Perceptron (MLP) Classifier
- Gradient Boosting Classifier
- XGBoost Classifier

49 model combinations were tested in total, and the hybrid model involving Decision Tree and SVM on the synthetic dataset attained the best accuracy of 90.15

In total, each of the methodologies outlined above was used individually on all three datasets, thus totaling **nine run of evaluation**. This cross-testing allows for an overall comparison of how various machine learning approaches work on datasets of different structure, complexity, and feature representation.

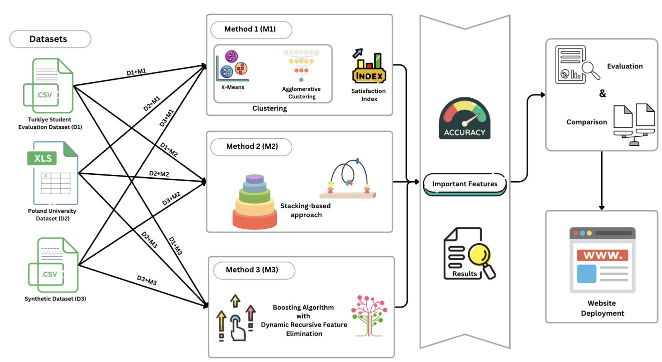

Fig. 1. Proposed methodology workflow

Figure 1, The workflow diagram in provides visual representation of the process which shows how each methodology applies to every dataset and how results are evaluated to determine the optimal model.

IV. Results & Discussion

A. Accuracies Obtained on Turkiye Student Evaluation Dataset

The Turkiye dataset was analyzed by two methods: Satisfaction Index, Clustering (K-Means Clustering, and Agglomerative Clustering). The best accuracy in the Satisfaction Index approach using individual models was obtained by AdaBoost Classifier (84.4%), while Random Forest Classifier combined with RFE and K-Fold CV gave the highest accuracy overall of 86.9%. In the clustering approaches, Logistic Regression and Voting Classifier performed best in both individual and RFE cases (Table IV-A).

Classifier	Satisfaction Index		K-Means Clustering		Agglomerative Clustering	
	Ind. Model	RFE + CV	Ind. Model	RFE + CV	Ind. Model	RFE + CV
Logistic Regression	81.7	81.5	**99.4**	98.5	95.4	98.5
Decision Tree Classifier	77.4	80.6	94.8	95.1	92.5	95.1
Bagging Classifier	80.7	84.7	96.8	96.5	95.1	96.5
Random Forest Classifier	77.4	**86.9**	94.8	97.7	92.5	97.7
AdaBoost Classifier	**84.4**	83.0	97.4	97.8	94.8	97.8
XGBoost Classifier	83.6	86.1	98.1	97.6	96.4	97.6
Support Vector Machine	82.9	82.2	99.0	98.4	95.4	98.4
Voting Classifier (Soft)	83.7	–	98.8	–	96.1	–
Voting Classifier (Hard)	82.6	–	98.9	–	**96.5**	–

Note: Bold values indicate the highest accuracy achieved in each approach.

B. Accuracies Obtained by Applying Turkiye Student Evaluation Dataset Methodologies on Poland Dataset

Turkiye Student Evaluation Dataset methodologies were applied to the Poland dataset using the same three approaches. As shown in Table IV-B, Decision Tree, Bagging, and XGBoost achieved perfect or near-perfect scores (up to **100%**). The effectiveness of these classifiers showcases strong generalizability of clustering-based methods across domains.

Classifier	Satisfaction Index		K-Means Clustering		Agglomerative Clustering	
	Ind. Model	RFE + CV	Ind. Model	RFE + CV	Ind. Model	RFE + CV
Logistic Regression	71.31	71.39	98.71	98.33	94.89	94.12
Decision Tree Classifier	97.30	97.98	**100.00**	**100.00**	**100.00**	**100.00**
Bagging Classifier	98.00	98.48	**100.00**	**100.00**	99.83	**100.00**
Random Forest Classifier	98.96	**99.14**	**100.00**	**100.00**	**100.00**	**100.00**
AdaBoost Classifier	76.55	74.61	98.46	98.85	94.18	94.93
XGBoost Classifier	98.79	98.43	**100.00**	**100.00**	99.83	**100.00**
Support Vector Machine	54.93	65.56	98.09	97.98	96.17	96.31
Voting Classifier (Soft)	98.25	98.59	**100.00**	**100.00**	99.83	99.98

C. Accuracies Obtained by Applying Turkiye Student Evaluation Dataset Methodologies on Synthetic Dataset

When applied to the synthetic dataset, the Turkiye Student Evaluation Dataset-derived models achieved surprisingly high accuracy. Table IV-C shows that the Soft and Hard Voting Classifiers and K-Means clustering yielded accuracies above 98%, with Random Forest reaching up to 97.7% even after RFE. These results suggest the synthetic dataset shares structural similarities with the Turkiye dataset.

Classifier	Satisfaction Index		K-Means Clustering		Agglomerative Clustering	
	Ind. Model	RFE + CV	Ind. Model	RFE + CV	Ind. Model	RFE + CV
Logistic Regression	50.63	50.61	50.97	50.61	50.97	50.61
Decision Tree Classifier	50.07	49.51	50.97	49.61	50.97	49.56
Random Forest Classifier	48.73	49.48	50.97	49.45	50.97	49.10
Bagging Classifier	48.23	–	50.97	–	50.97	–
AdaBoost Classifier	49.40	50.94	50.97	50.94	50.97	50.94
XGBoost Classifier	48.67	49.65	50.97	49.65	50.97	49.65
Support Vector Machine	50.97	50.95	50.97	50.95	50.97	50.95
Voting Classifier (Soft)	48.80	–	50.97	–	50.97	–
Voting Classifier (Hard)	48.60	–	50.97	–	50.97	–

D. Accuracies Obtained on Poland Dataset

For the Poland dataset, as shown in Table IV-D, Bagging Classifier performed best among standard classifiers (67%), while the stacking approach outperformed all with an accuracy of 78.5%. Random Forest with RFE achieved 66.6%, showing slight improvements with feature selection.

Classifier	Performance (Accuracy)	
		`n_splits = 10, n_repeats = 3`
Logistic Regression	56	61
Decision Tree Classifier	63	64
Bagging Classifier	**67**	66
Random Forest Classifier	66.5	**66.6**
AdaBoost Classifier	58	61
XGBoost Classifier	54	66
SVM	54	59
Voting Classifier	64	–
Stacking Classifier	**78.5**	–

E. Accuracies Obtained by Applying Poland Dataset Methodologies on Turkiye Student Evaluation Dataset

Poland-derived models were evaluated on the Turkiye dataset. As shown in Table IV-E, ensemble models like Bagging, XGBoost, and Random Forest achieved **100%** accuracy. Even the Stacking Classifier and Adaboost performed exceptionally well. Only SVM and Logistic Regression slightly underperformed compared to ensemble models.

Classifier	Model Accuracy (%)
Decision Tree Classifier	100.00
XGBoost Classifier	100.00
Bagging Classifier	100.00
Random Forest Classifier	100.00
Adaboost Classifier	100.00
Stacking Classifier	100.00
Voting Classifier	100.00
Logistic Regression	98.92
SVM	98.38

F. Accuracies Obtained by Applying Poland Dataset Methodologies on Synthetic Dataset

Models trained using Poland methodologies, when applied to the synthetic dataset, showed significantly lower performance Table IV-F. Most classifiers scored around 48–51%, suggesting poor generalizability of Poland-specific approaches to a synthetic setting with artificial label structure.

Classifier	Model Accuracy (%)
Logistic Regression	50.63
Decision Tree	50.03
Random Forest	48.93
Bagging Classifier	49.43
Adaboost Classifier	49.40
XGBoost	48.67
SVM	50.97
Voting Classifier (Soft Voting)	47.93
Voting Classifier (Hard Voting)	48.53
Stacking Classifier	**51.10**

G. Accuracies Obtained on Synthetic Dataset

The Synthetic Dataset received evaluation through a boosting-based approach that combined Recursive Feature Elimination (RFE) with dynamic selection. The highest accuracy of **90.15%** emerged from the combination of Decision

Tree + SVM according to Table IV-G. The results in Table IV-G show the top 10 model pairs which demonstrate how sequential modeling together with RFE improves prediction performance.

S.No	First Model	Acc. 1	Acc. 1 + RFE	Second Model	Acc. 2	Acc. 2 + RFE
2	Decision Tree	67.25	67.73	SVM	**90.15**	**90.15**
3	Decision Tree	67.46	67.66	Adaboost Classifier	90.04	90.04
4	Decision Tree	67.23	67.37	MLP Classifier	89.90	89.90
1	Decision Tree	67.59	67.59	Logistic Regression	89.85	89.85
5	Decision Tree	67.37	67.73	Gradient Boosting	89.69	89.69
6	Decision Tree	67.84	67.84	XGBoost	88.38	88.38
23	Adaboost Classifier	72.05	72.12	SVM	73.89	73.89
30	MLP Classifier	71.46	72.51	SVM	74.62	74.62
37	Gradient Boosting	72.01	72.01	SVM	74.12	74.12
38	Gradient Boosting	72.01	72.01	Adaboost Classifier	74.03	74.03

H. Accuracies Obtained by Applying Synthetic Dataset Methodologies on Turkiye Student Evaluation Dataset

Table IV-H shows the top 10 model combinations where synthetic dataset methodologies were applied on the Turkiye dataset. Most combinations achieved **100% accuracy**, indicating excellent generalizability of synthetic-trained models to real-world educational data.

S.No	First Model	Acc. 1	Acc. 1 + RFE	Second Model	Acc. 2	Acc. 2 + RFE
0	Decision Tree	100.00	100.00	Decision Tree	100.00	100.00
1	Decision Tree	100.00	100.00	Logistic Regression	100.00	100.00
2	Decision Tree	100.00	100.00	SVM	100.00	100.00
3	Decision Tree	100.00	100.00	Adaboost Classifier	100.00	100.00
4	Decision Tree	100.00	100.00	MLP Classifier	100.00	100.00
5	Decision Tree	100.00	100.00	Gradient Boosting	100.00	100.00
6	Decision Tree	100.00	100.00	XGBoost	100.00	100.00
21	Adaboost Classifier	100.00	100.00	Decision Tree	100.00	100.00
24	Adaboost Classifier	100.00	100.00	Adaboost Classifier	100.00	100.00
40	Gradient Boosting	100.00	100.00	Gradient Boosting	100.00	100.00

I. Accuracies Obtained by Applying Synthetic Dataset Methodologies on Poland Dataset

The results of applying models trained with synthetic data to the Poland dataset are presented in Table IV-I. XGBoost, Decision Tree, and Gradient Boosting models achieved several combinations above **99% accuracy**. This shows that synthetic methodologies transfer well even across different institutional datasets.

S.No	First Model	Acc. 1	Acc. 1 + RFE	Second Model	Acc. 2	Acc. 2 + RFE
42	XGBoost	98.21	98.71	Decision Tree	99.29	99.29
48	XGBoost	98.21	98.71	XGBoost	99.04	98.67
13	Logistic Regression	62.87	62.74	XGBoost	98.34	98.71
20	SVM	63.58	63.28	XGBoost	97.71	97.75
27	Adaboost Classifier	67.73	67.73	XGBoost	97.96	97.96
34	MLP Classifier	83.37	–	XGBoost	98.67	98.05
41	Gradient Boosting	78.00	77.34	XGBoost	97.30	96.59
5	Decision Tree	99.29	99.21	Gradient Boosting	99.29	99.29
2	Decision Tree	99.29	99.21	SVM	99.29	99.29
3	Decision Tree	99.29	99.21	Adaboost Classifier	99.29	99.29

V. Conclusion

This research aims to enhance educational quality by assessing instructor performance based on student feedback, focusing on instructor-specific attributes like experience, preparedness, and fairness, along with course-related factors. Unlike traditional CGPA-based methods, this study adopts a data-driven approach. The research utilized three different datasets consisting of the Poland University Dataset and the Turkiye Student Evaluation Dataset and a synthetic dataset developed by the authors to model academic situations.

A total 9 outputs are obtained by applying machine learning models and feature selection techniques to test each model across all datasets. The highest accuracies achieved were 78% (Poland), 99% (Turkiye), and 90.15% (Synthetic). The study chose to deploy the Decision Tree + Support Vector Machine (SVM) combination because of the synthetic dataset's controlled design and high accuracy.

A web application was developed to display these results, allowing users to search by professor name or ID to view ratings and performance details. The platform enables students to evaluate instructors who teach the same course and facilitates academic decisions through data analysis. The research shows how machine learning technology effectively improves teaching evaluation practices in higher education institutions.

References

[1] S. Mardikyan and B. Badur, "Analyzing teaching performance of instructors using data mining techniques," *Informatics in Education*, vol. 10, no. 2, pp. 245–257, 2011.

[2] A. M. Ahmed, A. Rizaner, and A. H. Ulusoy, "Using data mining to predict instructor performance," *Procedia Computer Science*, vol. 102, pp. 137–142, 2016.

[3] R. Ahuja and S. C. Sharma, "Exploiting machine learning and feature selection algorithms to predict instructor performance in higher education," *Journal of Information Science & Engineering*, vol. 37, no. 5, 2021. [Online]. Available:

[4] M. O. Asanbe, A. O. Osofisan, and W. F. William, "Teachers' performance evaluation in higher educational institution using data mining technique," *International Journal of Applied Information System (IJAIS)*, vol. 10, pp. 10–15, 2016.

[5] B. Benosa and M. R. R. Caguiat, "Prescriptive faculty performance analysis: A case at the onset of covid-19 pandemic," in *Proc. IEEE TENCON*, 2021, pp. 839–844. [Online]. Available:

[6] A. M. Ahmed, A. Rizaner, and A. H. Ulusoy, "Using data mining to predict instructor performance," *Procedia Computer Science*, vol. 102, pp. 137–142, 2016.

[7] M. Agaoglu, "Predicting instructor performance using data mining techniques in higher education," *IEEE Access*, vol. 4, pp. 2379–2387, 2016.

[8] J. R. McConnell, B. G. Bruster, and V. B. Smith, "Predicting teacher effectiveness based on various preservice factors: Implications for higher education and the evaluation of teacher preparation," *International Journal of Educational Reform*, vol. 28, no. 1, pp. 63–78, 2019.

[9] B. S. Abunasser, M. R. J. AL-Hiealy, A. M. Barhoom, A. R. Almasri, and S. S. Abu-Naser, "Prediction of instructor performance using machine and deep learning techniques," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2022.

[10] F. Shi, L. Wang, X. Liu, and M. H. Chiu, "Development and validation of an observation protocol for measuring science teachers' modeling-based teaching performance," *Journal of Research in Science Teaching*, vol. 58, no. 9, pp. 1359–1388, 2021.

[11] K. S. Hranchuk and M. J. Williams, "Addressing the feasibility of the teacher performance rate and accuracy scale as a treatment integrity tool," *Behavioral Interventions*, vol. 36, no. 2, pp. 355–368, 2021.

[12] L. Ke and C. V. Schwarz, "Supporting students' meaningful engagement in scientific modeling through epistemological messages: A case study of contrasting teaching approaches," *Journal of Research in Science Teaching*, vol. 58, no. 3, pp. 335–365, 2021.

[13] A. Almasri, "Experiences in mining educational data to analyze teacher's performance: A case study with high educational teachers," 2017.

[14] R. Odegua, "An empirical study of ensemble techniques (bagging, boosting and stacking)," in *Proc. Deep Learn. IndabaXAt Conf.*, 2019.

[15] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[16] G. Gunduz and E. Fokoue, "Uci machine learning repository," 2013, irvine, CA: University of California, School of Information and Computer Science.