BİTLİS EREN ÜNİVERSİTESİ
FEN BİLİMLERİ DERGİSİ

ISSN: 2147-3129 / e-ISSN: 2147-3188

Article Type	: Research Article	**Year**	: 2025
Received	: February 10, 2025	**Volume**	: 14
Revised	: March 28, 2025	**Issue**	: 2
Accepted	: June 25, 2025	**Pages**	: 1041-1059
DOI	: 10.17798/bitlisfen.1636812		

COMPARISON OF THE PERFORMANCE OF GRADIENT BOOSTING AND EXTREME GRADIENT BOOSTING METHODS IN CLASSIFYING TIMMS SCIENCE ACHIEVEMENT

Özlem BEZEK GÜRE [1]

[1] Batman University, Department of Medical Services and Techniques - Medical Documentation and Secretariat Programme, Batman, Türkiye, obezekgure@gmail.com

ABSTRACT

This study aims to compare the classification performance of machine learning methods Gradient Boosting (GB) and Extreme Gradient Boosting (XGBoost). The Trends in International Mathematics and Science Study 2019 (TIMSS 2019) science data set was used in the study. The dataset consists of data collected from a total of 2565 students, 1309 of whom are girls (51%) and 1256 (49%) are boys. A Python-based program was used for data analysis. In the study, Area Under the Curve (AUC), accuracy, precision, recall, F1 score, Matthews correlation coefficient (MCC), and training time were used as performance indicators. The study revealed that hyperparameter tuning had a positive impact on the performance of both methods. The analysis results show that the GB method was more successful compared to the XGBoost method in all performance measures except for training time. According to the GB method, 'student confidence in science' was identified as the most influential factor in science achievement, while the XGBoost method highlighted 'home educational resources' as the most significant predictor.

Keywords: Timms, Science, Gradient boosting, Extreme gradient boosting, Machine learning.

1 INTRODUCTION

Machine Learning (ML) is a family of methods that enable computers to make successful predictions by learning from past experiences. The success of these methods depends on the structure of the data and the performance of the algorithms used [1]. ML integrates several fields including artificial intelligence, information technology, and statistics [2]. These methods are used to uncover hidden relationships in datasets, find meaningful connections, and make predictions. ML has developed rapidly in parallel with the advancement of computer technologies [3]. ML algorithms are used in various sectors including finance, entertainment, health, engineering, and education. In the context of education, countries participate in large-scale exams to observe educational outcomes in science, mathematics, and reading and to compare their outcomes internationally [4]. Evidence suggests that countries showing progress in mathematics and science are likely to achieve success in scientific and technological fields, which contributes to economic growth and cultural enrichment [5]. Therefore, large-scale studies evaluating student achievements are considered important [6]. These studies not only set national standards but also provide feedback to students and parents, and aim to guide teachers in their profession. Furthermore, research findings are used to enhance international collaborations and make comparisons concerning educational issues [7]. The TIMMS study, which is conducted every four years, involves educational research in mathematics and science for fourth and eighth grades [8]. Along with student tests, surveys are conducted with students, teachers, and school administrators to collect data on factors that may affect student success [9-10]. The existing literature includes numerous studies using classical statistical methods with the mentioned data [5,11-17]. On the other hand, there are also studies using machine learning methods [18-26]. In addition to TIMMS data, various machine learning applications on datasets from educational research have also been observed [25, 27-34]. This study aims to compare the classification performance of the boosting methods, GB and XGBoost, using TIMMS 2019 eighth-grade science dataset. The literature lacks comparative studies utilizing these two methods together for classification purposes in the educational context.

1.1 Related Studies

There are frequent applications of classical statistical techniques using TIMMS data in the relevant literature [12,13-16,35-41]. However, the number of studies employing machine learning methods for this purpose remains limited. Bezek Güre [24] analyzed the performance of the Ensemble methods Adaboost and Bagging using the TIMMS 2019 mathematics dataset under various conditions. Similarly, Filiz and Öz [21] utilized the K nearest neighbors algorithm (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Artificial Neural Networks (ANN), Decision Tree (DT), and Logistic Regression (LR) on mathematics data. Askin and Gokalp [18] implemented LR and ANN methods, while Depren et al. [19] added Bayesian networks and DT to their methodology. Additionally, Akdemir applied the CHAID decision tree method. Pai et al. [41] employed DT and Bayesian classifiers, KNN, Radial Basis Function Neural Networks (RBFNN), and Particle Swarm Optimization (PSO) to analyze mathematics and science achievement. Similarly, Şevgin and Eranıl [25] utilized the Random Forest (RF) method to identify factors associated with school performance. On the other hand, Aydoğan and Gelbal [9] applied the CART method, and Filiz and Öz [22] deployed NB, DT, ANN, Polynomial SVM (SVM-POLY), and LR to determine factors affecting science achievement.

Only a few studies in the educational literature have employed the methods used in the present study in combination. For instance, in the study conducted by Susanto and Utami [42], these methods were used for regression purposes to predict PISA 2022 reading skills. On the other hand, Asselman et al. [43] used Random Forest, Adaboost, and XGBoost methods in their study to predict student performance. Saidani et al. [44] sought to predict student employment using XGBoost, Category Boosting (CatBoost), and Light Gradient Boosted Machine (LGBM). Meanwhile, studies involving educational data often employ the XGBoost method, as seen in the work by Wahyuningsih et al. [45] to predict student success using XGBoost and Linear Regression. Similarly, Liu, Chen, and Liu [46] identified factors influencing reading skills using XGBoost, Support Vector Regression (SVR), and Random Forest Regression (RFR). In the study conducted by Jeganathan et al. [47], Extra Trees and linear regression methods were also employed in addition to these methods.

Additionally, Cao, Zhang, and Tin [48] utilized XGBoost to assess the impact of scientific literacy on reading. Woo and Kim [49] aimed to determine learning orientations based on gender using XGBoost, and Yan [50] attempted to predict student achievement using XGBoost, RF, Lasso, Elastic Net, SVM, and DT. Zopluoğlu [51] aimed to detect cheating in exams using XGBoost, while Çakıt and Dağdeviren [52] predicted university selection scenarios using XGBoost among other machine learning algorithms. Similarly, Nirmala [53] used RF and XGBoost to predict students' graduation statuses, and Guang-yu and Geng [54] analyzed university students' performance and behaviors using XGBoost. Similarly, in the study by Ridwan, Priyatno, and Ningsih [55], school dropout and academic achievement were predicted using the XGBoost method. In a different context, Şevgin and Uçar [56] utilized XGBoost to predict organizational commitment among school principals.

The aforementioned methods have also been applied in fields outside education. For instance, Bentéjac, Csörgő, and Martínez-Muñoz [57] conducted a comparative analysis of gradient boosting methods (GB, XGBoost, LightGBM, and CatBoost) focusing on speed, accuracy, generalization performance, and hyperparameter configuration. Şahin [58] used GB, XGBoost, and RF for landslide susceptibility mapping. In another study, Şahin [59] employed GB, CatBoost, XGBoost, LightGBM, and RF for the same purpose. Demir and Şahin [60] predicted soil liquefaction using GB, XGBoost, and Adaboost. Sibindi, Mwangi, and Waititu [61] applied Adaboost, GB, XGBoost, and LGBM along with LGBM-XGBoost to predict house prices.

2 MATERIAL AND METHOD

The TIMMS 2019 assessment included 39 countries at the eighth-grade level. Turkey, among these countries, participated with 4,077 students from 181 schools at this level [8]. Prior to analysis in the current study, data cleaning and data transformation processes were conducted. After removing categorical variables, data collected from a total of 2,565 students—1,309 girls (51%) and 1,256 boys (49%)—were were retained for analysis. The science scores of the students were clustered into three groups: low, medium, and high. The dependent variable was the students' science achievement status, with independent variables including gender, educational resources at home, sense of belonging, peer bullying, love for science, clarity of science teaching, student confidence in science, value given to science by the student, self-efficacy in computer usage, number of study supports at home, parental education level, having extra lessons in the last year, time devoted to studying for science class, frequency of homework assignment by teachers, absenteeism, educational aspirations, ownership of a mobile phone, internet access, having one's own room, having a study desk, and owning a computer/tablet. The dataset used in the study was obtained from the official TIMSS website: https://timss2019.org/international-database/.

Descriptive statistics for the independent variables used in the study are provided in Table 1.

Table 1. Descriptive Statistics for the Independent Variables.

Predictive Variables	Sub-categories	%
Gender	Female	51
	Male	49
Parents' Highest Education Level	University or Higher	14,8
	Post-secondary but not University	7,5
	Upper Secondary	27,8
	Lower Secondary	29,2
	Some Primary, Lower Secondary or No School	16,7
	Don't Know	4
Computer ownership	Yes	75,6
	No	24,4
Possession of a desk	Yes	79,8
	No	20,2
Possession of own room	Yes	65,2
	No	34,8
Internet availability status	Yes	73,3
	No	26,7
Own mobile phone	Yes	55,8
	No	44,2
Home Educational Resources	Many Resources	9
	Some Resources	60
	Few Resources	31
About how often absent from school	Once a week	4,1
	Once every two weeks	5,4
	Once a month	13,1
	Once every two months	15,5
	Never or almost never	61,9
How far in education do you expect to go	Finish <Lower secondary education	1,8
	Finish <Upper secondary education	4,6
	Finish <Post-secondary, non-tertiary education	4,2
	Finish <Short-cycle tertiary education	3,5
	Finish <Bachelor’s or equivalent level	46,8
	Finish <Postgraduate degree: Master’s	39,1
How many minutes spent on homework	My teacher never gives me homework in…	6,6
	1–15 minutes	27,5
	16–30 minutes	37,6
	31–60 minutes	20,2
	61–90 minutes	4,9
	More than 90 minutes	3,2
How often teacher give you homework	Every day	6,6
	3 or 4 times a week	27,1
	1 or 2 times a week	36,8
	Less than once a week	20,9
	Never	6,6

Table 1 (Continued). Descriptive Statistics for the Independent Variables.

Predictive Variables	Sub-categories	%
Extra lessons last 12 month	Yes, to excel in class	46,9
	Yes, to keep up in class	12,9
	No	40,1
Students Sense of School Belonging	High Sense of School Belonging	54,5
	Some Sense of School Belonging	36,9
	Little Sense of School Belonging	8,6
Student Bullying	Never or Almost Never	73,5
	About Monthly	23,2
	About Weekly	3,3
Students Like Learning Science	Very Much Like Learning Science	54,1
	Somewhat Like Learning Science	36,8
	Do Not Like Learning Science	9,2
Instructional Clarity in Science Lessons	High Clarity of Instruction	70,8
	Moderate Clarity of Instruction	21,4
	Low Clarity of Instruction	7,8
Student Confident in Science	Very Confident in Science	40,7
	Somewhat Confident in Science	36,4
	Not Confident in Science	22,9
Students Value Science	Strongly Value Science	46,0
	Somewhat Value Science	39,2
	Do Not Value Science	14,7

2.1 Gradient Boosting

Among machine learning methods, the Gradient Boosting (GB) method is used in regression and classification problems. Originating from boosting techniques and developed by Friedman, this method is also known as Gradient Boosted Trees [62]. GB is a powerful algorithm that delivers successful results in many fields [63], although it is prone to overfitting [57]. The method tries to create strong learners by combining the results of many predictors. In GB, fixed-sized decision trees are typically used as the base learners [64]. The model is built incrementally by minimizing the expected value of a specified loss function [65]. The model emphasizes more on the misclassified observations by learning from errors. In essence, GB applies a gradient-based optimization process to reduce the overall error of the ensemble by determining the optimal contribution of each weak learner to the final prediction [66].

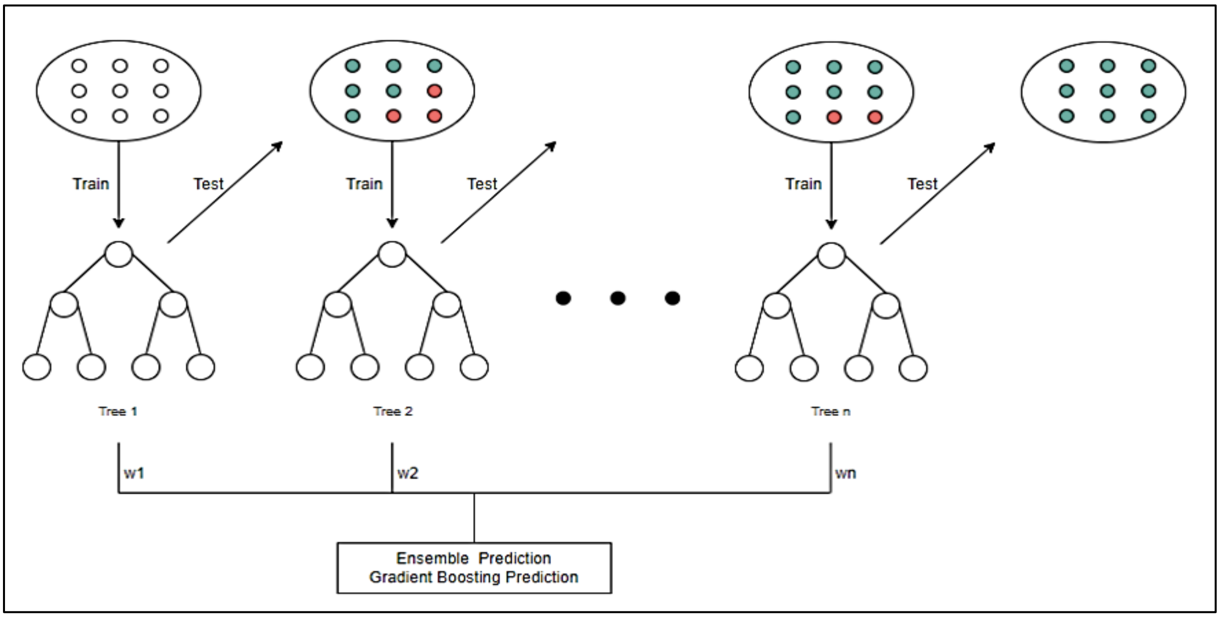

Figure 1. Structure of Gradient Boosting [67].

Gradient Boosting Algorithm

$$F_0(x) = argmin_\rho \sum_{i=1}^{N} L\,(y_i\,,\rho) \tag{1}$$

For $m = 1$ to M do:

$$\tilde{y}_i = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}, i = 1, N \tag{2}$$

$$a_m = \arg min_{a,\beta} \sum_{i=1}^{N} [\tilde{y}_i - \beta h(x_i; a)]^2 \tag{3}$$

$$\rho_m = argmin_p \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \rho h(\mathbf{x}_i;\ a_m)) \tag{4}$$

$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \rho_m h(\boldsymbol{x}; a_m) \qquad [68] \tag{5}$$

2.2 Extreme Gradient Boosting Method

The XGBoost method, developed by Chen and Guestrin in 2016, is also utilized for solving classification and regression problems. It has been shown to produce successful results, particularly when working with large datasets [56]. XGBoost is known for its resistance to overfitting and its user-friendly interface [69]. Additionally, it features a high degree of flexibility and scalability. Similar to the Gradient Boosting method, it aims to create stronger models by combining numerous models. It does this by focusing on weak learners and combining their prediction results [56]. In this method, decision trees are sequentially constructed. The structure of each tree is optimized, and the importance levels of variables are calculated accordingly [70]. The method operates on the principle of reducing the prediction error of the previous tree. To accomplish this, the algorithm utilizes a derived gradient to update the weights [45]. The method requires numerous prediction parameters. The success of the model depends on the best combination of parameters [71].

$$\hat{y_i} = \sum_{m=1}^{M} f_m (x_i), f_m \epsilon\ F \tag{6}$$

where m is the number of trees, and, F represents the base model of the trees.

$$L = \sum_{i} \text{l}(\hat{y_i}, y_i) + \sum_{m} \Omega(f_m) \tag{7}$$

Here, l is the loss function that measures the error between the estimated and true values, and, Ω is a regularization function used to prevent overfitting.

$$\Omega(f) = \gamma\, T + \frac{1}{2}\lambda\|\omega\|^2 \tag{8}$$

where ω, denotes the weights and T the number of leaves in each tree.

$$Gain = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} + \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma \tag{9}$$

In this formulation, a split is made if the information gain exceeds the threshold value γ [72].

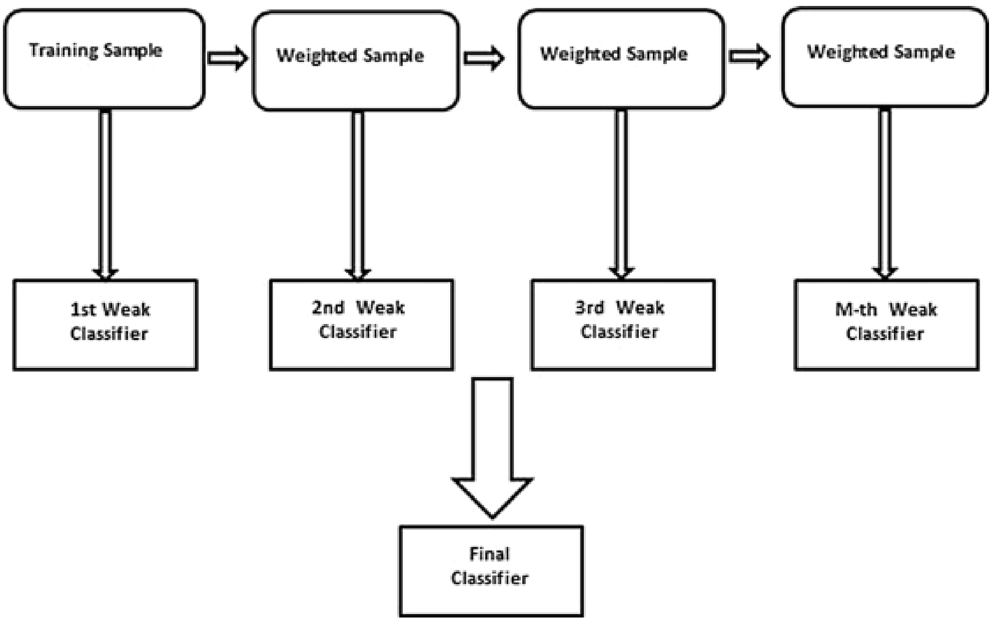

Figure 2. XGBoost structure [73].

3 RESULTS AND DISCUSSION

Before beginning the analyses, a multicollinearity test was conducted to determine if there was any multicollinearity issue among the variables. It is generally accepted that multicollinearity exists when the Variance Inflation Factor (VIF) exceeds 10 and tolerance values fall below 0.1 [74]. According to the analysis results, VIF values ranged from 1.060 to 2.48, while tolerance values varied between 0.412 and 0.944. These results indicate that multicollinearity is not a concern in this study.

To assess the performance of the methods, the following evaluation metrics were used: Area Under the Receiver Operating Characteristic Curve (AUC), accuracy, precision, recall, F1 score, Matthews Correlation Coefficient (MCC), and training time.

AUC: It is a measure that indicates the accuracy of the model. There is a direct relationship between the size of the area under the curve and the classification performance of the model [33].

MCC: It is a measure that indicates the relationship between the predicted class and the actual class. The measure ranges between -1 and +1. Correctly classified predictions are indicated with +1, while incorrectly classified predictions are marked with -1 [75-76].

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \tag{10}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$F1\ score = 2x\ \frac{(Precision\ x\ \ Recall)}{(Precision +\ \ Recall)} \tag{14}$$

Before the analysis of the methods, 70% of the dataset was used for training and the remaining 30% for testing. For the analyses, Orange, a free Python-based software, was utilized.

Table 2. Performance Criteria of Models (Default hyper-parameters).

Model	AUC	CA	F1	Precision	Recall	MCC	Training
GB	0,74	0,56	0,56	0,56	0,56	0,34	15,85
XGBoost	0,72	0,54	0,54	0,54	0,54	0,31	7,70

Subsequently, the performance of the methods was tested using parameters such as the number of trees, learning rate, and maximum tree depth. The best results were obtained when the number of trees was set to 100, the learning rate to 0.3, and the maximum depth to 6. To avoid overfitting, a 20-fold cross-validation method was applied. The analyses were performed on an Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.71GHz processor. The results are presented in Table 3.

Table 3. Performance Criteria of Models (Optimum hyper-parameters).

Model	AUC	CA	F1	Precision	Recall	MCC	Training time
GB	0,90	0,77	0,77	0,77	0,77	0,65	62,41
XGBoost	0,90	0,75	0,75	0,75	0,75	0,63	6,61

Table 3 indicates that the GB method outperforms the XGBoost method in terms of classification performance across all evaluation metrics. However, it has been determined that the XGBoost method is faster in terms of training time.

A comparison of Table 2 and Table 3 reveals that hyperparameter tuning significantly improved the models’ performance.

Table 4. Importance levels of factors affecting science achievement according to the GB method.

Number	Variables	Significance Level
1	Student confident in science	9,3
2	About how often absent from school	5,9
3	Home educational resources	4,0
4	How far in education do you expect to go	2,5
5	Parents' highest education level	2,4
6	Students sense of school belonging	2,3
7	Number of home study supports	2,0
8	How often teacher give you homework\science	1,9
9	Extra lessons last 12 month	1,9
10	How many minutes spent on homework	1,8
11	Instructional clarity in science lessons	1,7
12	Self-efficacy for computer use	1,5
13	Gender	1,5
14	Students value science	1,3
15	Home possess\own mobile phone	1,2
16	Students like learning science	1,1
17	Student bullying	1,1
18	Home possess\study desk	1,0
19	Home possess\internet connection	0,9
20	Home possess\computer tablet	0,9
21	Home possess\own room	0,7

As seen in Table 4, the most influential variable in predicting science achievement using the GB method is “student confidence in science.” It is followed by, in order, 'About how often absent from school ', 'Home educational resources ', ' How far in education do you expect to go ' and ' Parents' highest education level ' variables.

Table 5. Importance levels of factors affecting science achievement according to the XGBoost method.

Number	Variables	Significance Level
1	Home educational resources	7,0
2	Student confident in science	5,7
3	About how often absent from school	2,9
4	Parents' highest education level	2,5
5	How far in education do you expect to go	2,5
6	How often teacher give you homework\science	2,1
7	Home possess\study desk	1,8
8	Extra lessons last 12 month	1,7
9	Home possess\internet connection	1,6
10	Number of home study supports	1,6
11	Self-efficacy for computer use	1,6
12	Instructional clarity in science lessons	1,5
13	Student bullying	1,5
14	Home possess\computer tablet	1,4
15	How many minutes spent on homework	1,3
16	Students sense of school belonging	1,3
17	Home possess\own room	1,3
18	Students value science	1,2
19	Students like learning science	1,2
20	Home possess\own mobile phone	1,2
21	Gender	1,1

According to Table 5, “home educational resources” is the most influential variable in the XGBoost model, followed by “student confidence in science,” “school absenteeism,” and “educational aspirations.”

The aim of this study was to identify the factors affecting eighth-grade science achievement in TIMSS 2019 using the boosting algorithms Gradient Boosting (GB) and Extreme Gradient Boosting (XGBoost), and to compare the classification performance of these two methods. The results indicate that the GB algorithm outperformed XGBoost across all performance metrics, except for training time.

Furthermore, according to the GB method, the most significant variable influencing science achievement was “student confidence in science,” whereas for the XGBoost method, it was “home educational resources.” Following GB, other key variables included school absenteeism, home educational resources, educational aspirations, and parents’ highest education level. In the case of XGBoost, after “home educational resources,” the next most influential variables were “student confidence in science,” “school absenteeism,” “parents’ highest education level,” and “educational aspirations.” These findings demonstrate that both models identified the same top five variables, albeit in different orders.

In contrast to the present study, research conducted by Aksu and Doğan [77] and Filiz and Öz [22] identified information and communication technologies as the most significant factor affecting science achievement. A study by Uğuz, Şahin, and Yılmaz [78], which utilized different educational datasets, also found a positive relationship between the use of computer technologies and science achievement. Additionally, several studies in the literature have highlighted students’ self-efficacy perception as a key factor [5, 9, 33]. Consistent with the current findings, Zeybekoğlu and Koğar [79] reported that the number of books at home was the most influential factor, while Anıl [27] found that the father’s education level played the most significant role.

Although previous studies using TIMSS data have employed various machine learning methods, none have used GB and XGBoost in combination as in the current study. For example, Filiz and Öz [22] applied NB, DT, ANN, SVM, and LR to determine factors affecting science achievement, finding that LR and Polynomial SVM performed best. In another study, Filiz and Öz [21] evaluated factors influencing mathematics achievement using KNN, NB, SVM, ANN, DT, and LR, concluding that LR yielded the most effective results. Similarly, Bezek Güre [24] compared the performance of the ensemble methods Adaboost and Bagging using the TIMSS 2019 mathematics dataset, and while no significant difference was found regarding the number of variables, Bagging was found to be superior when comparing different sample sizes. A study by Askin and Gokalp [18] comparing LR and ANN concluded that ANN performed better, whereas Depren et al. [19] found LR to be more successful.

In the present study, the GB method achieved higher classification performance in terms of accuracy (0.77), F1 score (0.77), precision (0.77), recall (0.77), and MCC (0.65), whereas XGBoost demonstrated faster model training time. These findings are in line with similar results reported in the literature.

The study initially evaluated model performance using default hyperparameters and then tested optimized configurations, including the number of trees, learning rate, and maximum tree depth. The best outcomes were obtained with 100 trees, a learning rate of 0.3, and a maximum tree depth of 6. These findings emphasize the importance of hyperparameter tuning in boosting methods, as supported by Bentéjac, Csörgő, and Martínez-Muñoz [57].

Moreover, to enhance the generalizability of the results, this study employed 5-fold, 10-fold, and 20-fold cross-validation techniques, with the highest performance observed in the 20-fold cross-validation. Consistently, Bezek Güre [76] also highlighted the superior performance of 20-fold cross-validation in his study. Contrary to the present findings, Susanto and Utami [42], who utilized the same methods to predict PISA 2022 reading proficiency, reported that XGBoost outperformed the alternative models.

Furthermore, studies utilizing educational data frequently employ the XGBoost method, which is one of the algorithms used in the present study. Asselman et al. [43], in their effort to predict student performance, employed RF, Adaboost, and XGBoost methods. They reported that XGBoost achieved superior classification performance with accuracy rates of 78.75% on the ASSISTments dataset, 74.96% on the Andes dataset, and 72.00% on the Simulated dataset. Saidani et al. [44] used XGBoost, CatBoost, and LGBM to predict student employment and found that LGBM yielded more accurate predictions. Wahyuningsih et al. [45] used XGBoost and Linear Regression to predict student success, achieving an accuracy rate of 68%. Similarly, Liu, Chen, and Liu [46] used XGBoost, SVR, and RFR to identify factors affecting reading skills and reported that XGBoost made the most accurate predictions. Likewise, Ridwan, Priyatno, and Ningsih [55], in their study on predicting school dropout and academic achievement using the XGBoost method, achieved 88% precision and an F1 score of 81%, indicating strong predictive performance. Yan [50] attempted to predict student success using XGBoost, RF, Lasso, Elastic Net, SVM, and DT, and found that XGBoost outperformed the other methods. Cao, Zhang, and Tin [48], using XGBoost to examine the impact of scientific literacy on reading, concluded that the algorithm is well-suited for handling multivariate datasets. Meanwhile, Woo and Kim [49] applied XGBoost to determine learning orientations based on gender, reporting classification accuracies ranging from 76.97% to 81.88%. Zopluoğlu [51], in his study on exam fraud detection using XGBoost, demonstrated its effectiveness in classification tasks. Çakıt and Dağdeviren [52], in their research on predicting university selection scenarios using multiple machine learning algorithms, also concluded that XGBoost performed better than the other methods. Similarly, Nirmala [53], in a study predicting students' graduation statuses using RF and XGBoost, found XGBoost to yield more accurate predictions. Guang-yu and Geng [54] analyzed university students' performance and behaviors using XGBoost, reporting a correct classification rate of 73%.

The application of these methods is not limited to education; they have also been employed in other fields. Similar to the present study, Bentéjac, Csörgő, and Martínez-Muñoz [57] conducted a comparative analysis of gradient boosting methods (GB, XGBoost, LGBM, and CatBoost), evaluating them in terms of speed, accuracy, generalization performance, and hyperparameter configuration. Their findings emphasized the critical role of hyperparameter tuning. Şahin [58] used GB, XGBoost, and RF for landslide susceptibility mapping and found that XGBoost performed most successfully. In a subsequent study on the same topic, Şahin [59] used GB, CatBoost, XGBoost, LGBM, and RF, concluding that CatBoost yielded superior predictive accuracy. Likewise, Demir and Şahin [60] employed GB, XGBoost, and Adaboost to predict soil liquefaction, with XGBoost emerging as the most successful method. Sibindi, Mwangi, and Waititu [61] utilized Adaboost, GB, XGBoost, and LGBM—as well as a hybrid LGBM–XGBoost model—for predicting house prices, concluding that the hybrid method provided the best predictions.

As previously noted, the current study found that the GB method outperformed XGBoost in terms of classification performance. However, some studies in the literature have reported different findings [58, 60]. It has also been determined in this study that the XGBoost method requires less training time. In contrast, Bentéjac, Csörgő, and Martínez-Muñoz [57] and Ke et al. [80] found that LGBM is faster than XGBoost. Similarly, Wen et al. [81] stated that their proposed algorithm outperforms XGBoost in terms of computational speed.

4 CONCLUSION AND SUGGESTIONS

This study aimed to compare the predictive performances of the tree-based methods Gradient Boosting (GB) and Extreme Gradient Boosting (XGBoost) using a large-scale dataset from the TIMSS assessment. Both algorithms were tested under default configurations as well as with a variety of hyperparameter settings. Their performances were evaluated using 5-, 10-, and 20-fold cross-validation techniques. The optimal results were obtained with 100 trees, a learning rate of 0.3, a maximum tree depth of 6, and 20-fold cross-validation.

According to the findings, the GB algorithm demonstrated better classification performance than the XGBoost method. These results highlight the critical importance of hyperparameter optimization in assessing model effectiveness. Future research is encouraged to investigate the performance of these algorithms under alternative parameter configurations and cross-validation strategies.

The current study is limited to the Turkish sample and eighth-grade science data. The TIMSS survey is administered at both the fourth- and eighth-grade levels, collecting data not only from students but also from teachers, families, and school administrators. This provides an opportunity to examine a broader set of factors that influence student achievement. Future research could explore data collected from these additional sources.

Additionally, studies could be conducted using data from countries other than Türkiye, or by utilizing other large-scale educational datasets such as PISA, ABIDE, or PIRLS. Moreover, further research could be carried out using datasets from different educational domains. These studies may also benefit from the application of diverse machine learning techniques.

Acknowledgements

No support was received from any individuals, institutions, or organizations in the conduct of this study.

Statement of Research and Publication Ethics

This study was conducted in accordance with the principles of research and publication ethics. The dataset used in the present study was sourced from the publicly accessible https://timss2019.org/international-database/ ; therefore, the research did not necessitate ethical committee approval."

Artificial Intelligence (AI) Contribution Statement

Artificial intelligence (AI) tools were used in a limited capacity solely to polish minor aspects of language and expression in the manuscript. No AI tools were employed for writing, data analysis, or content generation. All core content, including conceptualization, analysis, and interpretation, was produced entirely by the author.

REFERENCES

[1] Sarker, I. H. “Machine learning: Algorithms, real-world applications and research directions”, *SN computer science*, vol.2(3), 160, pp. 1-21,2021.

[2] Nasteski, V.” An overview of the supervised machine learning methods”. Horizons. b, 4(51-62), 56, 2017.

[3] Baştanlar, Y. & Özuysal, M. “Introduction to machine learning”. miRNomics: *MicroRNA biology and computational analysis*, 105-128, 2014.

[4] Chang, Y.C. & Bangsri, A. “Thai Students’ Perceived Teacher Support on Their Reading Ability: Mediating Effects of Self-Efficacy and Sense of School Belonging”. International Journal of Educational Methodology. 6(2), 435 – 446, 2020.

[5] Juan, A., Hannan, S. & Namome, C. “I believe I can do science: Self-efficacy and science achievement of Grade 9 students in South Africa”. South African Journal of Science, 114(7-8), 48-54, 2018.

[6] Eser, M. T. & Çobanoğlu Aktan, D. “Educational data mining: The analysis of the factors affecting science instruction by clustering analysis”. International Journal of Educational Methodology, 7(3), 487-500, 2021

[7] Torney-Purta, J. & Amadeo, J. A. “International large-scale assessments: Challenges in reporting and potentials for secondary analysis”. Research in Comparative and International Education, 8(3), 248- 258, 2013.

[8] Turkey Ministry of National Education (TMNE). ” TIMSS 2015 national math and science preliminary report 4th and 8th grades”. Ankara: MEB: Measurement. General Directorate of Evaluation and Examination Services, 2020.

[9] Aydogan, I. & Gelbal, S. “Determination of the characteristics predicting science achievement through the classification and regression tree (cart) method: The case of TIMSS 2015 Turkey”. Egitim ve Bilim-Education and Science, 47(209), 2022.

[10] Hooper, M., Mullis, I. V. S. & Martin, M. O. “TIMSS 2015 contex questionnaire framework”. I. V. S. Mullis and M. O. Martin (Ed.), TIMSS 2015 assessment frameworks içinde (s. 61-82). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, 2013.

[11] Akyüz, G. “The effects of student and school factors on mathematics achievement in TIMSS 2011”, Education and Science, Vol. 39, No 172, 2014.

[12] Akyuz, G. & Berberoglu, G. “Teacher and Classroom Characteristics and Their Relations to Mathematics Achievement of the Students in the TIMSS”. New Horizons in Education, 58(1), 77-95, 2010.

[13] Yalcin, S., Demirtasli, R.N., Dibek, M. I. & , Yavuz, H. C. “The Effect of Teacher and Student Characteristics on TIMSS 2011 Mathematics Achievement of Fourth-and Eighth-Grade Students in Turkey”. International Journal of Progressive Education, 13(3), 79-94, 2017.

[14] Kilic, S. & Askin, Ö. E. “Parental influence on students’ mathematics achievement: The comparative study of Turkey and best performer countries in TIMSS 2011”. Procedia-social and behavioral sciences, 106, 2000-2007, 2013.

[15] Sandoval-Hernández, A. & Białowolski, P. “Factors and conditions promoting academic resilience: a TIMSS-based analysis of five Asian education systems”. Asia Pacific Education Review, 17(3), 511-520, 2016.

[16] Topçu, M. S., Erbilgin, E. & Arıkan, S. “Factors predicting Turkish and Korean students' science and mathematics achievement in TIMSS 2011”. Eurasia Journal of Mathematics, Science & Technology Education, 12(7), 1711-1737 2016.

[17] Nilsen, T., Kaarstein, H. & Lehre, A. C. “Trend analyses of TIMSS 2015 and 2019: school factors related to declining performance in mathematics”. Large-scale Assessments in Education, 10(1), 1-19, 2022.

[18] Askin, O. E. & Gokalp, F. “Comparing the predictive and classification performances of logistic regression and neural networks: a case study on Timss 2011”. Procedia-Social and Behavioral Sciences, 106, 667-676, 2013.

[19] Depren, S. K., Aşkın, Ö. E. & Öz, E. "Identifying the classification performances of educational data mining methods: A case study for TIMSS". Educational Sciences: Theory & Practice, 17(5), 2017.

[20] Siemssen, A. M. "Using Data Mining to Model Student Achievement on the 4 th Grade TIMSS 2015 Mathematics Assessment: A Five Nation Study". The University of Texas at El Paso, 2018.

[21] Filiz, E. & Öz, E. "Educational data mining methods for TIMSS 2015 mathematics success: Turkey case". Sigma Journal of Engineering and Natural Sciences, 38(2), 963-977, 2020.

[22] Filiz, E. & Öz, E. "Finding the Best Algorithms and Effective Factors in Classification of Turkish Science Student Success". Journal of Baltic Science Education 18, 239-253, 2019.

[23] Topal, K. H. "Variable Selection via the Adaptive Elastic Net: Mathematics Success of the Students in Singapore and Turkey". Journal of Applied Microeconometrics (JAME). 1(1), 40-54, 2021,

[24] Bezek Güre, Ö. "Investigation of ensemble methods in terms of statistics: TIMMS 2019 example. Neural Computing and Applications, 35(32), 23507-23520, 2023.

[25] Şevgin, H. & Eranıl, A. K. "Investigation of Turkish Students' School Engagement through Random Forest Methods Applied to TIMSS 2019: A Problem of School Psychology". International Journal of Psychology and Educational Studies, 10(4), 896-909, 2023.

[26] Wang, F., King, R. B. & Leung, S. O. "Why do East Asian students do so well in mathematics? A machine learning study". International Journal of Science and Mathematics Education, 21(3), 691-711, 2023.

[27] Anıl, D." Uluslararası öğrenci başarılarını değerlendirme programı (PISA)'nda Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler". Eğitim ve Bilim, 34(152), 2010.

[28] Bernardo, A. B., Cordel, M. O., Lapinid, M. R. C., Teves, J. M. M., Yap, S. A. &, Chua, U. C. "Contrasting profiles of low-performing mathematics students in public and private schools in the Philippines: insights from machine learning". Journal of Intelligence, 10(3), 61, 2022.

[29] Güre, Ö. B., Kayri, M. & Erdoğan, F. "Analysis of Factors Effecting PISA 2015 Mathematics Literacy via Educational Data Mining". Education & Science/Egitim ve Bilim, 45(202), 2020.

[30] Immekus, J. C., Jeong, T. S. & Yoo, J. E. "Machine learning procedures for predictor variable selection for schoolwork-related anxiety: evidence from PISA 2015 mathematics, reading, and science assessments". Large-scale Assessments in Education, 10(1), 30, 2022.

[31] Karaboğa, H. A., Akogul, S. & Demir, I. "Classification of Students' Mathematical Literacy Score Using Educational Data Mining: PISA 2015 Turkey Application". Cumhuriyet Science Journal, 43(3), 543-549, 2022.

[32] Lezhnina, O. & Kismihók, G. "Combining statistical and machine learning methods to explore German students' attitudes towards ICT in PISA". International Journal of Research & Method in Education, 45(2), 180-199, 2022.

[33] Şevgin, H. & Önen, E. "MARS ve BRT Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması: ABİDE-2016 Örneği". Eğitim ve Bilim, 47(211), 2022.

[34] Şevgin, H., Güre, Ö. B. & Kayri, M. "An Analysis of MARS and Logistic Regression Methods in Educational Data Mining in Light of Some Performance Indicators". International Journal of Research in Teacher Education (IJRTE), 14(3), 2023.

[35] Berberoglu, G., Celebi, O., Ozdemir, E., Uysal, E. & Yayan, B. "Factors Effecting Achievement Level of Turkish Students in The Third International Mathematics And Science Study (TIMSS)". Journal of Educational Sciences & Practices, 2(3), 2003.

[36] Ceylan, E. & Berberoğlu, G. "Öğrencilerin fen başarısını açıklayan etmenler: Bir modelleme çalışması". Eğitim ve Bilim, 32(144), 36-48, 2010.

[37] Pektaş, M. "Uluslararası matematik ve fen bilimleri eğilimleri çalışması (TIMSS 2007) Türkiye örnekleminde fen bilimleri başarısını etkileyen bazı değişkenlerin incelenmesi "(Master's thesis, Sosyal Bilimler Enstitüsü), 2010.

[38] Abazaoğlu, İ. & Taşar, M. F. "Fen bilgisi öğretmen özelliklerinin öğrenci fen başarısı ile ilişkisi: TIMSS 2011 verilerine göre bir durum analizi (Singapur, Güney Kore, Japonya, İngiltere, Türkiye)". İlköğretim online, 15(3), 2016.

[39] Kahraman, N. "Öğrenci katılımı ve TIMSS 2011 fen başarısı arasındaki ilişkinin sınıflara göre karşılaştırılması". Eğitim ve Bilim, 39(172), 2014.

[40] Ölçüoğlu, R. & Çetin, S. "TIMSS 2011 sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenlerin bölgelere göre incelenmesi". Journal of Measurement and Evaluation in Education and Psychology, 7(1), 202-220, 2016.

[41] Pai, P. F., Chen, C. T., Hung, Y. M., Hung, W. Z. & Chang, Y. C. "A group decision classifier with particle swarm optimization and decision tree for analyzing achievements in mathematics and science". Neural Computing and Applications, 25(7-8), 2011-2023, 2014.

[42] Susanto, H., & Utami, E. Comparative analysis to predict reading literacy based on Pisa 2022 using gradient boosted decision trees and extreme gradient boosting. G-Tech: Jurnal Teknologi Terapan, 9(1), 390-399, 2025.

[43] Asselman, A., Khaldi, M., & Aammou, S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), 3360-3379, 2023.

[44] Saidani, O., Menzli, L. J., Ksibi, A., Alturki, N. & Alluhaidan, A. S. "Predicting student employability through the internship context using gradient boosting models". IEEE Access, 10, 46472-46489, 2022.

[45] Wahyuningsih, T., Iriani, A., Purnomo, H.D. & Sembiring, I. "Predicting students' success level in an examination using advanced linear regression and extreme gradient boosting". Computer Science and Information Technologies, 5(1), 23-31, 2024.

[46] Liu, H., Chen, X. & Liu, X." Factors influencing secondary school students' reading literacy: An analysis based on XGBoost and SHAP methods". Frontiers in Psychology, 13, 948612, 2022.

[47] Jeganathan, S., Lakshminarayanan, A. R., Ramachandran, N., & Tunze, G. B. Predicting Academic Performance of Immigrant Students Using XGBoost Regressor. International Journal of Information Technology and Web Engineering (IJITWE), 17(1), 1-19. https://doi.org/10.4018/IJITWE.304052, 2022.

[48] Cao, C., Zhang, T. & Xin, T. "The effect of reading engagement on scientific literacy–an analysis based on the XGBoost method". Frontiers in Psychology, 15, 1329724, 2024.

[49] Woo, H. & Kim, J.M. "Impacts of learning orientation on the modeling of programming using feature selection and XGBOOST: A gender-focused analysis". Applied Sciences, 12(10), 4922, 2022.

[50] Yan, K. "Student performance prediction using XGBoost method from a macro perspective". In 2021 2nd International Conference on Computing and Data Science (CDS) (pp. 453-459). IEEE, 2021.

[51] Zopluoglu, C. "Detecting examinees with item preknowledge in large-scale testing using extreme gradient boosting (XGBoost)". Educational and psychological measurement, 79(5), 931-961, 2019.

[52] Çakıt, E. & Dağdeviren, M. "Predicting the percentage of student placement: A comparative study of machine learning algorithms". Education and Information Technologies, 27(1), 997-1022, 2022.

[53] Nirmala, I., Wijayanto, H. & Notodiputro, K.A. "Prediction of Undergraduate Student's Study Completion Status Using MissForest Imputation in Random Forest and XGBoost Models". ComTech: Computer, Mathematics and Engineering Applications, 13(1), 53-62, 2022.

[54] Guang-yu, L. & Geng, H. "The behavior analysis and achievement prediction research of college students based on XGBoost gradient lifting decision tree algorithm". In Proceedings of the 2019 7th International Conference on Information and Education Technology (pp. 289-294), 2019.

[55] Ridwan, A., Priyatno, A. M., & Ningsih, L. (2024). Predict Students' Dropout and Academic Success with XGBoost. Journal of Education and Computer Applications, 1(2), 1-8.

[56] Şevgin, H. & Uçar, R. "An Investigation into the Variables Affecting School Principals' Burnout through XGBoost Method". Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, 5(2), 420-439, 2023.

[57] Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. "A comparative analysis of gradient boosting algorithms". Artificial Intelligence Review, 54, 1937-1967, 2021.

[58] Sahin, E. K. "Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest". SN Applied Sciences, 2(7), 1308, 2020.

[59] Sahin, E. K. "Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping". Geocarto International, 37(9), 2441-2465, 2022.

[60] Demir, S. & Sahin, E. K. “An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost”. Neural Computing and Applications, 35(4), 3173-3190, 2023.

[61] Sibindi, R., Mwangi, R. W. & Waititu, A. G. “A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices”. Engineering Reports, 5(4), e12599, 2023.

[62] Yangın, G. “XGboost ve Karar Ağacı tabanlı algoritmaların diyabet veri setleri üzerine uygulaması” (Master's thesis, Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü), 2019.

[63] Natekin, A. & Knoll, A. “Gradient boosting machines, a tutorial”. Frontiers in neurorobotics, 7, 21, 2013.

[64] Biau, G., Cadre, B. & Rouvière, L. “Accelerated gradient boosting”. Machine learning, 108, 971-992, 2019.

[65] Zhang, Y. & Haghani, A. “A gradient boosting method to improve travel time prediction”. Transportation Research Part C: Emerging Technologies, 58, 308-324, 2015.

[66] Islam, S. & Amin, S. H. “Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques”. Journal of Big Data, 7(1), 65, 2020.

[67] Son, J. & Yang, S. “A new approach to machine learning model development for prediction of concrete fatigue life under uniaxial compression”. Applied Sciences, 12(19), 9766, 2022.

[68] Friedman, J. H. “Greedy function approximation: a gradient boosting machine”. Annals of statistics, 1189-1232, 2001.

[69] Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F. & El-Shafie, A. “Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia”. Ain Shams Engineering Journal, 12(2), 1545-1556, 2021.

[70] Chen, T. & Guestrin, C. “Xgboost: A scalable tree boosting system”. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794), 2016.

[71] Carmona, P., Climent, F. &, Momparler, A. “Predicting failure in the US banking sector: An extreme gradient boosting approach”. International Review of Economics & Finance, 61, 304-323, 2019.

[72] Zhao, Z., Yang, W., Zhai, Y., Liang, Y. & Zhao, Y. “Identify DNA-binding proteins through the extreme gradient boosting algorithm”. Frontiers in Genetics, 12, 821996, 2022.

[73] Sikander, R., Ghulam, A. & Ali, F. “XGB-DrugPred: computational prediction of druggable proteins using eXtreme gradient boosting and optimized features set”. Scientific reports, 12(1), 5505, 2022.

[74] Keller, P. S., El-Sheikh, M., Granger, D.A. & Buckhalt, J. A. “Interactions between salivary cortisol and alphaamylase as predictors of children’s cognitive functioning and academic performance”. Physiology & Behavior, 105, 987-995, 2012.

[75] Akgül, G., Çelik, A.A., Aydın, Z.E. & Öztürk, Z.K. “Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı”, Bilişim Teknolojileri Dergisi, vol.13, no.3, pp.255-268, 2020.

[76] Bezek Güre, Ö. “Classification of Liver Disorders Diagnosis using Naïve Bayes Method”. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 13(1), 153-160, 2024.

[77] Aksu, G. & Dogan, N. “Comparison of Decision Trees Used in Data Mining= Veri madenciliginde kullanilan karar agaçlarinin karsilastirilmasi”. Pegem Journal of Education and Instruction, 9(4), 1183-1208, 2019.

[78] Uğuz, E., Şahin, S. & Yılmaz, R. “PISA 2018 fen bilimleri puanlarının değerlendirilmesinde eğitsel veri madenciliğinin kullanımı”. Bilgi ve İletişim Teknolojileri Dergisi, 3(2), 212-227. 2021

[79] Zeybekoğlu, Ş. & Koğar, H. “Investigation of variables explaining science literacy in PISA 2015 Turkey sample”. Journal of Measurement and Evaluation in Education and Psychology, 13(2), 145-163, 2022.

[80] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W. ... & Liu, T. Y. “Lightgbm: A highly efficient gradient boosting decision tree". Advances in neural information processing systems, 30, 2017.

[81] Wen, Z., He, B., Kotagiri, R., Lu, S. & Shi, J. Efficient gradient boosted decision tree training on GPUs. In 2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS) (pp. 234-243). IEEE, 2018.