Cumhuriyet Sci. J., 43(3) (2022) 543-549
DOI: https://doi.org/10.17776/csj.1136733

Cumhuriyet Science Journal

| csj.cumhuriyet.edu.tr | Founded: 2002 ISSN: 2587-2680 e-ISSN: 2587-246X Publisher: Sivas Cumhuriyet University

Classification of Students' Mathematical Literacy Score Using Educational Data Mining: PISA 2015 Turkey Application

Hasan Aykut Karaboğa [1,a], Serkan Akogul [2,b,*], Ibrahim Demir [3,c]

[1] *Department of Educational Science, Faculty of Education, Amasya University, Amasya, Türkiye*
[2] *Department of Statistics, Faculty of Sciences, Pamukkale University, Denizli, Türkiye*
[3] *Turkish Statistical Institute, Ankara, Türkiye.*
**Corresponding author*

Research Article

History
Received: 27/06/2022
Accepted: 12/09/2022

ABSTRACT

PISA 2015 mathematical literacy score of Turkey is 420 while the average score of all countries is 461. It is understood that; Turkish students' PISA 2015 mathematical literacy score was lower than the average. The basic reasons for the below average score need to be truly examined and developmental activities should be revealed. The aim of this study is to classify students according to the factors affecting their mathematical literacy score and to reveal the effects of these factors in classification.The data of the study is obtained from 5895 students who participated in PISA 2015. In this study, we used Random Forest, Naïve Bayes Classifier, Logistic Regression, Decision Tree Algorithm and Discriminant Analysis as classifiers. According to the results, Random Forest method produced more accurate scores than other methods with 76.32% accuracy. We also calculated the correct classification rate and determined the factors that positively and negatively affect the classification with discriminant analysis. According to the discriminant analysis home possessions, information and computer technology resources at home and students' expected occupational status were the most positive effective variables on mathematical literacy score. On the other hand, family wealth possessions, student-related factors affecting school climate and anxiety have negative effect on mathematical literacy score.

Keywords: Classification, Educational data mining, PISA 2015, Mathematics education, Discriminant analysis.

[a] *h.aykut.karaboga@amasya.edu.tr* *https://orcid.org/0000-0001-8877-3267*
[b] *sakogul@pau.edu.tr* *https://orcid.org/0000-0002-0346-4308*
[c] *idemir@gmail.com* *https://orcid.org/0000-0002-2734-4116*

Introduction

Countries continue their existence with educational investments. Because high qualified people ensure keeping up with industrial competitions, developments and global changes in the world. The protection of the interests of the countries depends on such highly-educated citizens. In order to reveal the effectiveness of education, it is needed to be evaluated objectively at national and international level. For this reason, according to certain evaluations like PISA, it is necessary to determine educational level of country and to take measures to eliminate deficiencies. As an OECD member, Turkey regularly participates in the PISA program to assess the quality of education [1].

Mathematical literacy, science literacy, reading literacy, computer skills, students' motivations, opinions about themselves, the school environment and their families are collected in the PISA exam. Mathematical literacy focuses on measuring the capacity to use, formulate and interpret mathematics. While Turkey's PISA 2015 mathematical literacy score is 420, the average score of OECD countries is 490. The results of Turkey's PISA average mathematical literacy score between 2003 and 2015 are given in Table 1. It is seen that PISA 2015 score is the lowest. The basic dynamics of this decrease should be correctly identified, examined and prevented. For this purpose, we classified students' mathematical literacy score with the factors that affecting success using Educational Data Mining (EDM) techniques and discriminant analysis.

Table 1. PISA Turkey ranking - average score of mathematics

Years	2003	2006	2009	2012	2015
All Country Average	489	484	465	470	461
OECD Average	500	494	495	494	490
Turkey Average	423	424	445	448	420
Number of Participating Countries	41	57	65	65	70
Turkey Ranking	35	43	43	44	49

Overall, the EDM shows us an evaluation process with advantages for educational development and assessment. Researchers prefer EDM techniques to classical statistical methods due to the increase in data type and amount [2]. Because data mining techniques are easy to use and the number of methods is high according to different data types. Also these methods generate efficient results faster than classical methods. The most commonly used data mining methods in education are clustering, classification, regression and association analysis [3]. In addition, against to classical statistical techniques the absence of constraints such as normality, covariance, linearity and normal distribution is also give advantage to the EDM [3]. However, since the methods have certain advantages over each other in some conditions, they should be used together or the comparative results should be examined too. In this study, taking this into account, different classification techniques were compared. Our research aims to contribute the EDM centred literature of PISA studies. For this purpose, mainly two research questions are addressed: (1) which factors have positive or negative impact of students mathematical success, (2) and which EDM method or discriminant analysis is more appropriate for classifying data for PISA 2015 Turkish mathematical dataset.

This study differs from other studies in the literature on PISA in terms of the EDM methods and variables used [4-7]. Random Forest [4,7], Discriminant Analysis [8], Decision Tree [5,7,9], Logistic Regression [10], Naïve Bayes [11,12] studies are encountered in the literature. However, no article was found using these methods and the data set mentioned in the article. In this respect, it is thought that the study will contribute to the literature.

Literature

In the literature, the EDM is used in many different applications. Slater et al. [13] reviewed 40 research tools for the EDM and learning analytics (LA). Reviewers pointed out that it is not a suitable tool for end-to-end the EDM and the LA analysis. Therefore, the combination of the EDM tools is more useful for complex analysis.

In a different study, Devasia, Vinushree and Hegde [14] conducted an experiment to predict students' performance with proposed web based application. According to the paper Naïve Bayes technique produced more accurate results from the other techniques. To get faster and more efficient accreditation process Tastimur, Karakose, and Akin [15] performed an IT-based accreditation model for engineering education. They suggested 10 criterions and used Genetic Algorithm method to train k-Nearest Neighbor classifier. Different methods have been developed for classification based on variable and/or model selection [16]. Agaoglu [17] appraised, instructors' performance with seven different classification methods. He used support vector machines, C5.0, CART, Discriminant Analysis and Artificial Neural Network methods to analyse 2850 course evaluation scores. C5.0 classifier gave the best classification result.

Shahiri and Husain [9] reviewed literature systematically to find most successive students' performance prediction methods from 2002 to 2015. They found that the most used prediction methods are Neural Network, Decision Tree, Support Vector Machine, k-Nearest Neighbor and Naïve Bayes respectively. Osmanbegović and Suljić [12] compared data mining applications for student success prediction at the University of Tuzla. Developing learners' creativity was the most important part of web based learning system.

One of the EMD application study for higher education prediction in Turkey is conducted by Tekin [18]. In that study, it was aimed to prevent the students who determined to give up the school. Artificial Neural Networks, Support Vector Machines and Extreme Learning Machine methods used to predict students' graduation grades. Using PISA and TIMMS data Kiray, Gok and Bozkir [19] identified the affective variables on mathematics and science with decision trees and clustering algorithms. In their study, they found out reading and problem solving skills affect mathematical achievement and, so on science achievement is affective variable on mathematical achievement.

Aksu and Güzeller [5], classified PISA 2012 mathematical literacy scores of Turkish students with CHAID method. As a result of their study, it was determined that attitude towards the course, perception of self-efficacy and anxiety were important variables on mathematical literacy score. Dolu [20] examined science performance and economic, social and cultural status index (ESCS) relation for PISA 2015 Turkey survey with Hierarchical Linear Models. As a result of her study, she found that ESCS had a low positive effect on science achievement. Aksu and Keceoglu [10] compared prediction results of mathematical success with logistic regression, CHAID and REPTree methods.

Gure et al. [21] used Multilayer Perceptron and Random Forest methods to estimate PISA 2015 mathematical literacy score. As a result of the study, it was stated that the Random Forest algorithm produced more successful results. Toprak and Gelbal [8] compared classification performances of artificial neural networks, decision trees and discriminant analysis at PISA 2012 mathematical literacy score for different sample sizes. They used all student data for analysis with 17 mathematical success related variables. Koyuncu and Gelbal [11] tested performance of Naïve Bayes, k-Nearest Neighbor, Neural Network, and Logistic Regression methods under different sample size conditions for PISA 2012 dataset. As it's seen in literature, different EDM methods give different results [10]. In this study, analyzes were made with some of the algorithms that produced the most successful results, considering the superiority of the methods to classify PISA 2015 mathematical literacy success of the Turkish students.

Materials and Methods

As indicated by Romero and Ventura [3], the "Educational Data Mining" term was first introduced in 2005. EDM is combination of education, statistics and computer sciences [3]. We can describe data mining models in two ways; predictive or descriptive. Predictive models contain Prediction, Classification, Time Series Analysis and Regression. Descriptive models generally used for summarizing the data, clustering, discovering Association Rules and Sequence Discovery [3,22]. The EDM applications are expending in traditional and computer-based education. These applications are helpful to education designer and pedagogues [23].

The EDM is a lodestar for educators and managers to obtain educational expert knowledge about learning systems and student behaviours. It is crucial that defining the problem and converting the data to a suitable form for curing an educational problem [23]. So, there is an immense opportunity to judge all pedagogical paradigms and educational approaches with the EDM applications [24].

On detecting factor related performance analysis, compared to other data analytic techniques the EDM give detailed and more efficient results. Furthermore, in human sciences these results adapt to needs better than other techniques. It is often stated that evaluating the results by using a combination of several data mining methods is healthier instead of using data mining algorithms alone [18]. Because depend on the data structure and data size, the EDM methods have some advantages [23].

For this reason, weakness of the data mining techniques should be decrease with comparative studies to classical methods [7]. The EDM applications are not separated by sharp lines [25]. But in our literature, it seems to be a deficiency in this subject. One of the purposes of our paper is contributing the usage of these powerful methods in our literature.

Research Sample

Academic success is a complex issue because, academic achievement are composite of a variety of family-related variables, school-related variables, personal variables and social or environmental variables. Family-related variables include socioeconomic and sociocultural variables, parents' education and occupation, parents' support, family structure, and parents' relation in school. School-related variables are relevant with school assessment, teacher support and assessment, learning opportunities, class size, and schools' social and cultural support. Social variables based on student' living era and schools' social environment [7].

Personal factors like school related variables and demographic variables are indicative on students' academic achievement [7]. Mathematic skills are not only important for high school performance, but also determinative of undergraduate success [26]. Socio-demographic variables, studying attitudes and previous achievements have positive effect on success [12]. A lot of studies have shown that demographic variables, past academic achievements, family income are effective variables on academic achievement [27]. Other highly correlated factor with student performance is qualification of parents [14].

As a result, mathematical achievement is closely related to these internal and external factors. By identifying the factors that lead to mathematical achievement or failure, and by removing negative situations, the overall academic achievement will turn into motivation with an increasing effect. In this study, we selected variables from the PISA 2015 dataset given in Table 2.

Table 2. Mathematical Achievement Related Factors

#	Variable Name	Explanation of variables
1	BSMJ	Students' Expected Occupational Status
2	HISEI	Highest Occupational Status of Parents
3	OUTHOURS	Out-of-School Study Time Per Week
4	MMINS	Mathematics Learning Time (minutes per week)
5	TMINS	Total Learning Time (minutes per week)
6	BELONG	Sense of Belonging to School
7	ANXTEST	Test Anxiety
8	MOTIVAT	Students' Achievement Motivation
9	CPSVALUE	Value of Co-Operation
10	EMOSUPS	Parents Emotional Support
11	CULTPOSS	Cultural possessions at Home
12	HEDRES	Home Educational Resources
13	HOMEPOS	Home Possessions
14	ICTRES	Information and Computer Technology Resources
15	WEALTH	Family Wealth Possessions
16	ESCS	Index of Economic, Social and Cultural Status
17	EDUSHORT	Shortage of Educational Material
18	STAFFSHORT	Shortage of Educational Staff
19	STUBEHA	Student-related Factors Affecting School Climate
20	TEACHBEHA	Teacher-related Factors Affecting School Climate
21	STRATIO	Student Teacher Ratio

The research sample is 5895 Turkish students who participated in the PISA 2015 Program. Indices were chosen instead of variables affecting the mathematical literacy score. Some selected variables have positive effect on mathematical literacy score, while some factors have negative effect on mathematical literacy score. The target variable was coded with (0= not successful,1= successful). On this variable 0 means under average mathematical literacy score and 1 means above average mathematical literacy score.

Classification Algorithms

In this study, we used C4.5, Logistic Regression, Naïve Bayes and Random Forest, which is the most widely used EDM methods in the literature. In addition, discriminant analysis was used to examine the factors affecting success.

Discriminant Analysis is one of the statistical technique that give canonical functions for classification of cases into two or more mutually exclusive groups or scores about two or more variables. Furthermore, Discriminant Analysis help to detect most powerful discriminators' characteristics named discriminating variables. In summary Discriminant Analysis find linear combination of group discriminators, membership prediction of new cases with discrimination functions, evaluating the group differences based on variables [28].

Classification with Discriminant Analysis has two main concepts. Firstly, to differentiate classes using canonical discriminant functions or discriminating variables. Secondly, predicting group membership of the future observations. For each group classification function is a linear combination of variables [28].

C4.5 is a type of classification algorithm which is a member of decision tree family. This algorithm is one of the most popular machine learning method. This method was developed by Quinlan [29] to overcome some deficiencies of ID3 algorithm. C4.5 algorithm handles both continuous and discrete data.

Relationships between dependent and independent variables can be investigated using Logistic Regression. The most important assumption in standard regression is that the dependent variable must be continuous. If the dependent variable takes the value 0 or 1, binary Logistic Regression should be applied to predict or classify the dependent variable of the observable independent variables [30].

Naïve Bayes is in the supervised learning subclass of machine learning. In other words, it is clear which class the sample data in the data set belongs to. This statistical method is based on calculating the conditional probability of the effect of each attribute on the outcome. Naïve Bayes has stands out as one of the most efficient and effective inductive learning algorithms for machine learning methods and data mining [2,22].

In the Random Forest method, many decision trees are created using different variations of a training data. New versions of the training data are obtained by randomly selecting a sample from the original training dataset by displacement. Every tree in the forest should be advanced to the greatest imaginable level without pruning. To classify a new test substance, each tree in the forest is allowed to make a classification decision. As a result, the classification decision is made for the majority among these situations [2]. In addition, it is a method that gives better results compared to its corresponding algorithms [5].

Performance Criteria

Various criteria are used to evaluate the results in machine learning studies. In this study, some commonly used criteria were taken into account as comparison criteria. These classification measures were determined as accuracy (ACC), F-measure, Kappa and mean absolute error (MAE) statistics, respectively.

ACC is determined by the ratio of all correctly classified samples in the model to the total number of samples. The ACC formula is given in equation (1).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Here, depending on the results, it is expressed as True positive (TP): true positive estimate, False positive (FP): false positive estimate, True negative (TN): true negative estimate, and False negative (FN): false negative estimate [22].

The F-measure, also called the F-score, is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'. The F- measure is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall. A perfect model has an F- measure of 1. Mathematical definition of the F- measure is given in equation (2).

$$F - measure = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = \frac{2 * TP}{2 * TP + FP + FN} \tag{2}$$

Kappa statistics (κ) investigates the predictive performance of a classification model. It is a convenient statistic to measure the evaluation of categorical variables. It is also a value based on the chi-square table [31]. Those whose κ value is close to 1 are closer to the solution. P_o and P_e show the relationship between two categorical variables [32].

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{3}$$

Mean absolute error (MAE) statistics help to reveal the differences between predicted and observed values of a model [33]. The MAE calculates the average of the absolute differences between the predicted and observed values. Here, the MAE statistic is calculated as follows to show the predicted and observed values of P_i and O_i [33].

$$MAE = n^{-1} \sum_{i=1}^{n} |P_i - O_i| \tag{4}$$

Accuracy and Kappa, which are used extensively in EDM studies, should be calculated too instead of using only f-measure to show effectiveness [34]. Table 5 shows the classification performances of algorithms based on the previously mentioned classification criteria (ACC, F-Measure, κ statistic and MAE). The best algorithm was accepted based on the ACC criterion. Other classification criteria were also used to support the final result.

Results

Discriminant Analysis was first applied in the study. At the end of the analysis, the students were classified according to their mathematical literacy score. Classification rates according to discriminant analysis are given in Table 3.

Table 3. Classification Results of Discriminant Analysis

Category	Total	Predicted Group Membership	
		Unsuccessful	Successful
Unsuccessful	3245 (100%)	2361 (72.8%)	884 (27.2%)
Successful	2650 (100%)	835 (31.5%)	1815 (68.5%)

According to the results given in the table, the accuracy of our model was found to be 70.8%. In addition, the weights of the factors for the separation function are as in Table 4. According to these weights, the most effective variables are understood.

Table 4. Standardized Canonical Discriminant Function Coefficients

No	Variable Name	Score	No	Variable Name	Score
1	HOMEPOS	0.955	1	WEALTH	-0.592
2	BSMJ	0.336	2	STUBEHA	-0.396
3	ICTRES	0.335	3	ANXTEST	-0.247
4	HISEI	0.287	4	OUTHOURS	-0.210
5	MOTIVAT	0.231	5	CULTPOSS	-0.172
6	MMINS	0.169	6	HEDRES	-0.172
7	STRATIO	0.146	7	STAFFSHORT	-0.172
8	TEACHBEHA	0.126	8	ESCS	-0.140
9	TMINS	0.058	9	EDUSHORT	-0.117
10	CPSVALUE	0.041			
11	BELONG	0.040			
12	EMOSUPS	0.036			

As its seen in Table 4, BSMJ, HISEI, MMINS, TMINS, BELONG, MOTIVAT, CPSVALUE, EMOSUPS, HOMEPOS, ICTRES, TEACHBEHA, STRATIO have positive effect on mathematical literacy score. However, OUTHOURS, ANXTEST, CULTPOSS, HEDRES, WEALTH, ESCS, EDUSHORT, STAFFSHORT and STUBEHA have negative effect on mathematical literacy score.

According to these results, home possessions the most positive effective variable on mathematical literacy score. Second effective variable was ICT resources at home. Also 'students' expected occupational status' and 'highest occupational status of parents' had positive effect on mathematical literacy score. On the other hand, 'family wealth possessions', 'student-related factors affecting school climate' and 'test anxiety' had negative effect on mathematical literacy score.

Table 5. Comparison of Classification Achievements of the Methods

Classifier	ACC	F-Measure	κ statistic	MAE
Random Forest	76.57%	0.764	0.521	0.348
C4.5 Classifier	71.23%	0.712	0.416	0.317
Logistic Regression	71.09%	0.709	0.411	0.378
Discriminant Analysis	70.80%	0.708	0.412	0.294
Naïve Bayes	68.36%	0.684	0.365	0.332

As it was seen, Random Forest was the best algorithm according to ACC (76.57%), F-Measure (0.764), κ statistic (0.521) and MAE (0.348). According to the results, Random Forest method produced more accurate scores than other EDM methods.

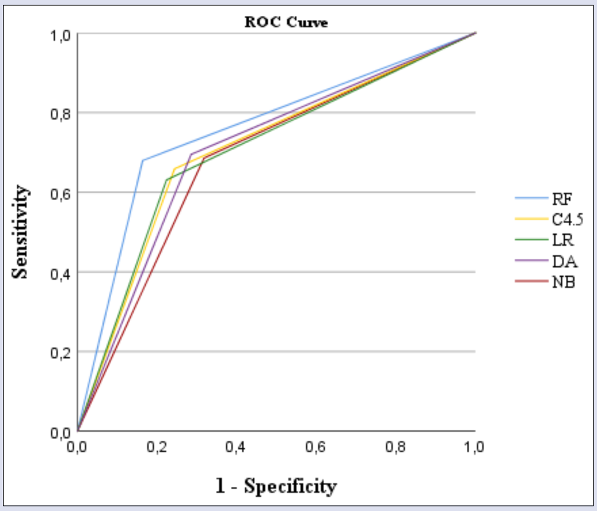

Figure 1. ROC Curves of the Methods

RF, C4.5, LR, DA, NB have compared with ROC area. As it seen, RF has the greatest ROC area than other methods (AUC_{RF} =0.758; $AUC_{C4.5}$ =0.707; AUC_{LR} =0.704; AUC_{DA} =0.705; AUC_{NB} =0.684). As can be seen from the ROC curve, the difference between the methods is not very high. However, it is still seen that EDM methods produce slightly better results than Discriminant Analysis.

Discussion, Conclusion and Recommendations

Actual topics of EDM are prediction, clustering, outlier detection, relationship mining, social network analysis, process mining, text mining and data refinement for human judgement, discovery with models, knowledge tracing and nonnegative matrix factorization [3]. On detecting factor related performance analysis, compared to other data analytic techniques EDM techniques give detailed and more efficient results. So, in human sciences these results adapt to needs better than other techniques. But weakness of the data mining techniques should be decrease with comparative studies with classical methods [7].

In our study, Discriminant Analysis from classical methods and Random Forest, Decision Tree, Logistic Regression and Naïve Bayes algorithms from data mining methods were compared. The aim of this study is to determining the factors affecting PISA 2015 mathematical literacy score by using Discriminant Analysis and comparing the classification capabilities of this method with Random Forest, C4.5, Logistic Regression, and Naïve Bayes algorithms. As a result of the analysis, random forest method was found to be the most successful classification method in PISA 2015 Turkey data.

When the results are examined, it is seen that the 'Home Possessions' variable has a positive effect on mathematics achievement [6,20]. The 'Highest Occupation Status of Parents' variable also has a positive effect on student success. Highly educated parents' children are more successful than other students [6]. In addition, 'Students' Expected Occupation Status' also has a positive effect on success. It can be assumed that this variable also increases the motivation of the student and high motivation increases success [6]. ICT Resources, which can be used for the course, have a positive effect on academic success [35]. Also, mathematics learning time increase one of the most influential variables on mathematics achievement is the mathematical achievement [6,36]. Furthermore, 'Student Teacher Ratio' and 'Teacher-related Factors Affecting School Climate' [37] have positive effects on mathematical achievement.

On the other hand, unlike many studies in the literature [20] 'family wealth possessions', 'cultural possessions at home', 'home educational resources' and 'Index of Economic, Social and Cultural Status' variables has a negative effect on mathematical literacy score. Accordingly, family wealth has not a positive effect on success. As it is known, the negative student behaviour decreases the academic achievement. For this reason, the student-related factors affecting school climate variable is also one of the variables that negatively affect mathematical literacy score [38]. In addition, the anxiety has also negative effect on academic achievement [5,6]. 'Out-of-School study time' is another variable that negatively affects academic achievement. Working hard outside of school has not increased mathematical literacy score. This may be due to the anxiety-enhancing effect and the fact that students get bored and lose motivation. Also it is known that lack of material have negative effects on achievement [39]. In this study, 'Shortage of Educational Staff' and 'Shortage of Educational Material' variables, similar to the literature, had a negative effect on mathematics achievement. It can be assumed that the results obtained may be due to the research sample. For this, work can be repeated with different countries.

As a result, family wealthy is not an indicator of academic achievement. On the other hand, the education level of the family and the professional expectations of the student had positive effects on achievement. Students' expectations can motivate them and make it easier for them to achieve. ICT resources have made a positive contribution to achievement as a means of finding solutions for academic problems. In our education system, students can get higher scores by creating multimedia-supported learning environments with projects such as the FATIH project. These projects eliminate educational inequality and enriching the learning environment for all learning types [40]. In addition, motivated students contribute to their mathematical literacy score with the study time they allocate to the mathematics lesson. In addition, it is expected that reducing student anxiety and eliminating the lack of educational materials, which are obstacles to success, will contribute to mathematics achievement.

This research has some limitations. The results of this research are based only on the PISA 2015 Turkey data set and only mathematics achievement was used in this study. Another limitation is that this dataset measures 15-year-old students' ability to use their mathematical knowledge and skills to cope with real-life challenges. The study can be repeated in comparison with the mathematics achievements of different countries.

Conflicts of interest

There are no conflicts of interest in this work.

References

[1] Taş U.E., Arici Ö., Ozarkan H.B., Özgürlük B., PISA 2015 ulusal raporu, *Ankara: Milli Eğitim Bakanlığı,* (2016).

[2] Witten I.H., Frank E., Data mining: practical machine learning tools and techniques with Java implementations, *Acm Sigmod Record*, 31(1) (2002) 76-77.

[3] Romero C., Ventura S., Educational Data Mining: A Review of the State of the Art, *IEEE Trans. Syst., Man, Cybern. C*, 40(6) (2010) 601–618.

[4] Aksu G., Doğan N., Veri madenciliğinde kullanılan öğrenme yöntemlerinin farklı koşullar altında karşılaştırılması, *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 51(3) (2018) 71-100.

[5] Aksu G., Güzeller C.O., Classification of PISA 2012 mathematical literacy scores using Decision-Tree Method: Turkey sampling. *Egitim ve Bilim*, 41(185) (2016) 101–122.

[6] Dos Santos R.A., Paulista C.R., da Hora, H.R.M., Education Data Mining on PISA 2015 Best Ranked Countries: What Makes the Students go Well, *Technology, Knowledge and Learning*, (2021) 1-32.

[7] Martínez Abad F., Chaparro Caso López A.A., Data-mining techniques in detecting factors linked to academic achievement, *School Effectiveness and School Improvement*, 28(1) (2017) 39–55.

[8] Toprak E., Gelbal S., Comparison of Classification Performances of Mathematics Achievement at PISA 2012 with the Artificial Neural Network, Decision Trees and Discriminant Analysis, *International Journal of Assessment Tools in Education*, 7(4) (2020) 773-799.

[9] Shahiri A.M., Husain W., Rashid N.A., A Review on Predicting Student's Performance Using Data Mining Techniques, *Procedia Computer Science*, 72 (2015) 414–422.

[10] Aksu G., Keceoglu C.R., Comparison of Results Obtained from Logistic Regression, CHAID Analysis and Decision Tree Methods, *EJER*, 19 (84) (2019) 1–20.

[11] Koyuncu İ., Gelbal S., Comparison of Data Mining Classification Algorithms on Educational Data under Different Conditions, *Journal of Measurement and Evaluation in Education and Psychology*, 11(4) (2020) 325-345.

[12] Osmanbegovic E., Suljic M., Data Mining Approach for Predicting Student Performance, *Economic Review: Journal of Economics and Business*, 10(1) (2012) 3–12.

[13] Slater S., Joksimović S., Kovanovic V., Baker R.S., Gasevic D., Tools for Educational Data Mining: A Review, *Journal of Educational and Behavioral Statistics*, 42(1) (2017) 85–106.

[14] Devasia T., Vinushree T.P., Hegde V., Prediction of students performance using Educational Data Mining. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), (2016) 91-95.

[15] Tastimur C., Karakose M., Akin E., Improvement of relative accreditation methods based on data mining and artificial intelligence for higher education, *15th International Conference on Information Technology Based Higher Education and Training (ITHET)*, (2016) 1–7.

[16] Gogebakan M., A Novel Approach for Gaussian Mixture Model Clustering Based on Soft Computing Method, *IEEE Access*, 9 (2021) 159987–160003.

[17] Agaoglu M., Predicting Instructor Performance Using Data Mining Techniques in Higher Education, *IEEE Access*, 4(1) (2016) 2379–2387.

[18] Tekin A., Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach, *EJER*, 14(54) (2014) 207–226.

[19] Kiray S.A., Gok B., Bozkir A.S., Identifying the Factors Affecting Science and Mathematics Achievement Using Data Mining Methods, *JESEH*, 1(1) (2015) 28.

[20] Dolu A., Sosyoekonomik Faktörlerin Eğitim Performansı Üzerine Etkisi: PISA 2015 Türkiye Örneği, *Yönetim ve Ekonomi Araştırmaları Dergisi*, 18(2) (2020) 41-58.

[21] 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).

[22] Tan P.N., Steinbach M., Kumar V., *Introduction to data mining*, Second edition, Global edition, New York, (2020).

[23] Romero C., Ventura S., Data mining in education, *WIREs Data Mining and Knowledge Discovery*, 3(1) (2013) 12–27.

[24] Peña-Ayala A., Educational data mining: A survey and a data mining-based analysis of recent works, *Expert Systems with Applications*, 41(4) (2014) 1432–1462.

[25] Bakhshinategh B., Zaiane O.R., ElAtia S., Ipperciel D., Educational data mining applications and tasks: A survey of the last 10 years, *Educ Inf Technol*, 23(1) (2018) 537–553.

[26] Chong S., Mak M., Loh W.M., Data-mining applications with the admission data of adult learners in higher education: a pilot study, *IJMIE*, 10(2) (2016) 131.

[27] Mishra T., Kumar D., Gupta S., Students' Performance and Employability Prediction through Data Mining: A Survey, *Indian Journal of Science and Technology*, 10(24) (2017) 1-6.

[28] Izenman A.J., *Modern multivariate statistical techniques: regression, classification, and manifold learning*. New York, Springer, (2008).

[29] Quinlan J.R., *C4.5: Programs for Machine Learning*. San Mateo, California: Morgan Kaufmann Publishers, (1993).

[30] Hosmer Jr D.W., Lemeshow S., Sturdivant R.X., *Applied logistic regression*, John Wiley & Sons, 398 (2013).

[31] Donner A. Klar N., The statistical analysis of kappa statistics in multiple samples, *Journal of Clinical Epidemiology*, 49(9) (1996) 1053–1058.

[32] Turanoglu-Bekar E., Ulutagay G., Kantarcı-Savas S., Classification of Thyroid Disease by Using Data Mining Models: A Comparison of Decision Tree Algorithms, *The Oxford Journal of Intelligent Decision and Data Science*, 2016(2) (2016) 13–28.

[33] Willmott C. Matsuura K., Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Clim. Res.*, 30 (2015) 79–82.

[34] Costa E.B., Fonseca B., Santana M.A., de Araújo F.F., Rego J., Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses, *Computers in Human Behavior*, 73 (2017) 247–256.

[35] Skryabin M., Zhang J., Liu L., Zhang D., How the ICT development level and usage influence student achievement in reading, mathematics, and science, *Computers & Education*, 85 (2015) 49–58.

[36] Martínez-Abad F., Gamazo A., Rodríguez-Conde M.-J., Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment, *Studies in Educational Evaluation*, 66 (2020) 100875.

[37] Topçu M.S., Arıkan S., Erbilgin E., Turkish Students' Science Performance and Related Factors in PISA 2006 and 2009, *Aust. Educ. Res.*, 42 (1) 117–132.

[38] Altun A., Kalkan Ö.K., Cross-national study on students and school factors affecting science literacy, *Educational Studies*, 47(4) (2021) 403-421.

[39] Arends F., Winnaar L., Mosimege M., Teacher classroom practices and Mathematics performance in South African schools: A reflection on TIMSS 2011, *South African Journal of Education*, 37(3) (2017).

[40] Zhao Y., Lu Z., Study on the Application of Multimedia Network Teaching Platform in College Physical Education Teaching, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 9(4) (2016) 193–202.