

FINDING THE BEST ALGORITHMS AND EFFECTIVE FACTORS IN CLASSIFICATION OF TURKISH SCIENCE STUDENT SUCCESS

Enes Filiz, Ersoy Öz

Introduction

Educational Data Mining (EDM) is a widely used methodology that overcomes big and complex educational data sets. Application of EDM unveils the information hidden in these data sets that cannot be revealed by use of the basic statistical methods that are often employed by educators in reading the data. The information revealed through EDM scrutinises the successes of students and, based on that information, helps policy-makers in the field of education form appropriate norms and policies for better education practices.

The International Association for the Evaluation of Educational Achievement (IEA) is a notable international organisation that oversees the monitoring of educational evaluation in many countries. Effective application of EDM can occur only when there are reliable data sets that can be studied. The IEA makes such data sets available to participating countries. The organisation acquires these data sets by undertaking comparative studies among the participating countries which results in the above-mentioned data sets and helps them examine the various education practices being followed in various countries and their effects there. One of the more ambitious ventures that IEA has undertaken in recent times is called Trends in International Mathematics and Science Study (TIMSS) which is administered every four years to science and mathematics students in their fourth and eighth grades. Over 60 countries from across the world are participants in TIMSS. Not only does this test reveal information about the outcome of various education norms being followed in the participating countries, it also allows the researchers to evaluate the success rates of the students in their respective countries and compare them with those from other countries (Mullis, Martin, Foy, & Arora, 2012).

Many studies in the past dealing with subjects from the field of education have utilised data that was a result of TIMSS application. These research studies have incorporated the popular and commonly used statistical methods such as factor analysis and regression but it has been seen that these methods of analyses are not apt for large and complex data sets due to their inherent limitations. EDM is therefore gradually becoming more widespread



JOURNAL
OF BALTIC
SCIENCE
EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Abstract. *Educational Data Mining (EDM) is an important tool in the field of classification of educational data that helps researchers and education planners analyse and model available educational data for specific needs such as developing educational strategies. Trends International Mathematics and Science Study (TIMSS) which is a notable study in educational area was used in this research. EDM methodology was applied to the results of TIMSS 2015 that presents data culled from eighth grade students from Turkey. The main purposes are to find the algorithms that are most appropriate for classifying the successes of students, especially in science subjects, and ascertaining the factors that lead to this success. It was found that logistic regression and support vector machines – poly kernel are the most suitable algorithms. A diverse set of features obtained by feature selection methods are “Computer Tablet Shared”, “Extra Lessons Last 12 Month”, “Extra Lessons How Many Month”, “How Far in Education Do You Expect to Go”, “Home Educational Resources”, and “Student Confident in Science” and these features are the most effective features in science success.*

Keywords: *classification algorithms, educational data mining, eighth grade, science success, TIMSS 2015.*

Enes Filiz, Ersoy Öz
Yildiz Technical University, Turkey



for use for such data sets as it does not apply the commonly used classical assumptions such as normality, linearity, and variance homogeneity (Han, Kamber, & Pei, 2012; Sinharay, 2016).

Literature Review

The concept of EDM has been gaining ground in the field of education research all over the world. It has become one of the more important approaches used to examine current methods of education, analysing them, and devising new and better techniques for the future. It aids researchers gain a better insight in the mind of students, their learning processes, and also helps comprehend available data sets better. Not only does EDM help the researchers, it also supports the students by disseminating useful information through practical approaches (Romero & Ventura, 2010). An exhaustive study of available literature and data on the subject of education was perused by Romero and Ventura in 2007 and they came to the conclusion that EDM acts as a repetitive tool for testing, hypothesis building, and improvement in student performance. Educators too gain immensely by EDM as they discern information about evaluating students that can be quite valuable. Definitions of EDM, uses of data mining in education, and future use of EDM have also been discussed in research papers on the subject (Romero & Ventura, 2010). Current literature on EDM includes the history of EDM and the changes that have taken place since its inception (Baker & Yacef, 2009), the latest developments in the field of data mining in education and how it has grown through the years (Mohamad & Tasir, 2013), and concurrent application of data mining and analytical methods (Siemens & Baker, 2012). Another important study by Peña-Ayala (2014) had a two-pronged approach. One was to notate the history of development of EDM through the years while the other part of the study included their analysis and resultant outcomes of the data mining method they employed.

As mentioned earlier, there's a diverse set of algorithms that are utilised in EDM for categorising data sets as well as analysing and predicting their outcomes. The various studies done on this subject offer numerous suggestions. In their study, Kotsiantis, Pierrakeas, and Pintelas (2004) did a comparative assessment of various EDM techniques and opined that the naive Bayes (NB) algorithm was most appropriate for developing a software support tool through result analyses. An exploration of existing studies on the subject for students from 2002 to 2015 proved that decision tree (DT), artificial neural network (ANN), k-nearest neighbors algorithm, support vector machine (SVM), and NB were most efficacious for estimating performance of the most successful students (Shahiri & Husain, 2015). A survey was conducted in 2010 to examine academic performance and was applied to both, students as well as their school principals (Ramaswami & Bhaskaran, 2010). Using the chi-square automatic prediction models, the survey offered prediction rules and proved to be better than other existing prediction models. Another study by Baradwaj and Pal (2011) suggested that data mining algorithms which used the classification method to assess student performance did the task as expected. This model using the DT method was aimed at the higher education category. Yet another study aimed at the higher education category, specifically engineering education, propounded EDM as the apt model for that segment of education. This study used ANN and DT models as prediction models to forecast engineering entrance examination data and solve the engineering education planning problem (Rajni & Malaya, 2015). Other studies led to somewhat different conclusions. For instance, another study showed that DT gave better results and interpretation of data than other methods. The procedure in this study by Martinez Abad and Chaparro Caso López (2017) suggested, based on DT techniques, that academic success factors could be isolated by statistical analysis done through data mining methods and that personal factors played an important role in academic performance. A Portuguese study utilised ANN, random forest DT and SVM to develop a student performance model for secondary school pupils which led to the conclusion that the past performance of a student and academic success had a close association; it also showed what the best prediction model was in such a scenario (Cortez & Silva, 2008). Horáková, Houška, and Dömeová (2017) found that ANN is a more accurate prediction method in comparison to the classification and regression trees for sample of 120 text fragments.

Studies like TIMSS that examine education data from many countries from various parts of the world helped in improving educational policies of the participating countries in such studies and improve their student success rate through comparison with educational norms followed by other countries. One study to gauge the educational success for TIMSS in 2011 established that a student's confidence was the determining factor for success. The LR and ANN techniques were employed to compute the prediction and classification performance (Askin & Gokalp, 2013). A considerable number of studies have been undertaken to examine TIMSS from various countries and different age groups and years. These research exercises have also used varied techniques for the purpose. For instance, the science and mathematics successes in TIMSS 1999, The Programme for International Student Assessment (PISA)



2003, and PISA 2006 data sets were scrutinised by Kiray, Gok, and Bozkir (2015) who made use of DT and clustering to reach a conclusion. Another study examined the mathematics and science successes of eighth grade students from Taiwan in TIMSS 2007 for which DT methodologies like CART, ID3, CHAID, C4.5 and Bayesian classifier, the k-nearest neighbors approach, particle swarm optimisation algorithm, and neural networks were utilised (Pai, Chen, Hung, Hung, & Chang, 2014). Yet another study dealt with TIMSS 2011 data to ascertain the success factors of Turkish and Korean students in their science and mathematics examinations. This study conducted by Topçu, Erbilgin, and Arıkan in 2016 also reviewed the consequences their findings could have on education. An analysis was undertaken to find the best possible algorithms to help classify eighth grade students from Turkey by applying NB, DT, ANN, and LR to TIMSS 2011 data (Kılıç-Depren, Askin, & Öz, 2017).

Aim of Research

This research aims to contribute the current literature of TIMSS studies which are in the concept of EDM. For this purpose, mainly two research questions are being fully addressed: (1) which EDM method is more appropriate for classifying data culled through TIMSS 2015 Turkish science data, and (2) what features (factors) influence students and lead them to success.

To find answers to the first question listed earlier (to find the most suitable EDM method), the most commonly used algorithms in EDM studies have been used. These include the NB, RepTree decision tree (DT-RepTree), random forest decision tree (DT-RF), and C4.5 decision tree (DT-C4.5) algorithms. Apart from these algorithms, ANN, LR, and SVMs consisting of three kernels: polynomial (SVM-POLY), radial basis function (SVM-RBF), and Pearson VII function-based universal (SVM-PUK) kernels are used as well. All these processes are employed under EDM to develop, analyse, and decode the data sets.

What constitutes success for students was the second question mentioned earlier. To find features that lead to successful students is imperative since there is no 'one size fits all' answers. There are numerous extraneous elements that impact education. For this EDM algorithms are applied to student performances which helps analyse the features that are more effective than others in achieving success in education and also to identify the poor performance of students in some areas (Ramaswami & Bhaskaran, 2010). Since TIMSS data sets have revealed many possible factors that may influence student performance, it is essential that the most significant of these that are likely to be instrumental in affecting student success are identified so that appropriate action is taken. This research, while keeping the number of features in the classification algorithms to the minimum and providing maximum classification performance, uses feature selection algorithms such as correlation attribute, correlation-based feature selection (CFS) subset, gain ratio, and One-R to identify factors leading to student success.

Research Methodology

Data Set

TIMSS is targeted at students in their fourth and eighth grades and is used to determine their levels in mathematics and science. It has been conducted every 4 years and in 2015, there were 7 participating countries that set the benchmark with 39 other countries that undertook the eighth grade assessment. During the first phase of TIMSS, a set of schools from these countries was chosen in a proportionate manner. In the second phase, some classes were randomly selected from the chosen schools. The survey was based on self-reported questionnaires (LaRoche, Joncas, & Foy, 2016).

For this research, the TIMSS 2015 data set was used. This assessment was administered to eighth grade students in Turkey in subjects related to science. The set consisted of 6079 students, of which 2943 were females and 3136 males. Some of the information in the data set, however, was either inaccurate or missing; therefore, those were not taken into account. Thus, the eventual numbers included in the assessment stand at 4481, of which 2273 were females and 2208 were males. Table 1 shows the 35 features and one dependent variable used in the research. The 35 features were independent variables and were considered important factors for students' success in science. As was displayed in the table, the "1st Plausible Value Science" (BSSSCI01) was a dependent variable and denoted the science success of the students. TIMSS 2015 average science score was 500 with a standard deviation of 100 (Mullis, Martin, Foy, & Arora, 2012). Thus, the score of 500 was the centrepoint and if the student scores higher than 500, the BSSSCI01 was scored as 1 or, failing which, 0.



Table 1. Student related features.

Factor Name	Description	Domain
ITSEX	Sex of Students	1= Female, 2=Male
BSBG03	OFTEN SPEAK AT HOME	1=Always, 2=Almost Always, 3=Sometimes, 4= Never
BSBG05	DIGITAL INFORMATION DEVICES	1=None, 2=1-3 devices, 3=4-6 devices, 4=7-10 devices, 5= More than 10 device
BSBG06A	COMPUTER TABLET OWN	1=Yes, 2=No
BSBG06B	COMPUTER TABLET SHARED	1=Yes, 2=No
BSBG06C	STUDY DESK	1=Yes, 2=No
BSBG06F	OWN MOBILE PHONE	1=Yes, 2=No
BSBG06G	GAMING SYSTEM	1=Yes, 2=No
BSBG06H	HEATING SYSTEMS	1=Yes, 2=No
BSBG06I	COOLING SYSTEMS	1=Yes, 2=No
BSBG06J	WASHING MACHINE	1=Yes, 2=No
BSBG06K	DISHWASHER	1=Yes, 2=No
BSBG08	HOW FAR IN EDUCATION DO YOU EXPECT TO GO	1=Finish lower secondary, 2= Finish upper secondary, 3= Finish post-secondary, non-tertiary, 4=Finish short-cycle tertiary, 5=Finish bachelor's or equivalent, 6=Finish post graduate degree
BSBG11	ABOUT HOW OFTEN ABSENT FROM SCHOOL	1=Once a week or more, 2=Once every two weeks, 3=Once a month, 4=Never or almost never
BSBG12	HOW OFTEN BREAKFAST ON SCHOOL DAYS	1=Every day, 2=Most days, 3=Sometimes, 4=Never or almost never
BSBG13A	HOW OFTEN USE COMPUTER TABLETHOME	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG13B	HOW OFTEN USE COMPUTER TABLET SCHOOL	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG13C	HOW OFTEN USE COMPUTER TABLETOTHER	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG14A	ACCESS TEXTBOOKS	1=Yes, 2=No
BSBG14B	ACCESS ASSIGNMENTS	1=Yes, 2=No
BSBG14C	COLLABORATE WITH CLASSMATES	1=Yes, 2=No
BSBG14D	COMMUNICATE WITH TEACHER	1=Yes, 2=No
BSBG14E	FIND INFO TO AID IN MATH	1=Yes, 2=No
BSBs25AB	HOW OFTEN TEACHER GIVE YOU HOMEWORK/SCIENCE	1=Every day, 2=3 or 4 times a week, 3=1 or 2 times a week, 4=Less than once a week, 5=Never
BSBS26AB	EXTRA LESSONS LAST 12 MONTH\SCIENCE	1=Yes, to excel in class, 2=Yes, to keep up in class, 3=No, 9= Omitted or Invalid
BSBS26BB	EXTRA LESSONS HOW MANY MONTH\ SCIENCE	1=Did not attend, 2=Less than 4 months, 3=4-8 months, 4=More than 8 months
BSBGHER	Home Educational Resources	
BSBGSSB	Students Sense of School Belonging	
BSBGSB	Student Bullying	
BSBGSLs	Students Like Learning Science	
BSBGESL	Engaging Teaching in Science Lessons	
BSBGSCS	Student Confident in Science	
BSBGSVS	Students Value Science	
BSDSLWP	Science Achievement Too Low for Estimation	1=Yes, 2=No



Factor Name	Description	Domain
BSDSWKHW	Weekly Time Spent on Science Homework	1=3 Hours or more, 2=More than 45 minutes but less than 3 hours, 3=45 minutes or less
BSSSCI01	1 ST PLAUSIBLE VALUE SCIENCE	0=Not successful, 1= Successful

The TIMSS 2015 questionnaire was developed to gauge a single hidden structure called a scale. The Rasch partial credit model is one of the Item Response Theory (IRT) scaling methods employed here (Masters & Wright, 1997). This research consisted of 9 scales: "Home Educational Resource" (BSBGHER), "Student Bullying" (BSBGSB), "Students' Sense of School Belonging" (BSBGSSB), "Engaging Teaching in Science Lessons" (BSBGESL), "Students Value Science" (BSBGSVS), "Student Confident in Science" (BSBGSCS), "Science Achievement Too Low for Estimation" (BSDSLOWP), "Students Like Learning Science" (BSBGSLs), and "Weekly Time Spent on Science Homework" (BSDSWKHW) variables.

k-fold Cross-validation: k-fold cross-validation is an essential method of a data mining process. Under this technique, the available data set is separated into two groups: one is used as the training set and the other as testing set through partitions such as 50%-50% or 70%-30%. The data set can also be divided into k pieces, for instance, k-1 of the set can be used for training with the rest utilised for testing. The entire data set is thus broken into pieces k times and used as testing sets. Then the mean of all the testing is computed to result in classification measures (Filiz & Öz, 2017).

Classification Algorithms

Naive Bayes Algorithm (NB): NB is considered by researchers as one of the most effective inductive learning algorithms used for data mining purposes (Zhang, 2004). It is a type of Bayesian network and requires two conditions to function optimally. One condition requires the classes to be independent of each other under certain conditions and the second one needs the variables, especially those that are likely to affect the results, to be visible (John & Langley, 1995).

Decision Tree (DT): DTs are at the forefront of classification techniques because they have an innate interpretable nature. They are necessary when one is being developed out of a given data set as it lowers the generalisation error to a minimum (Rokach & Maimon, 2005). C4.5, RF, and RepTree are the most commonly used DT algorithms. C4.5 is also known as J48 in an open source Weka application and uses information entropy to give rise to a binary DT. This is especially useful for pattern recognition problems (Quinlan, 2014). The RepTree algorithm, using regression tree logic, duplicates various trees and chooses the best one made. For tree trimming and selection from this group, the mean square error criterion is applied (Kalmegh, 2015). Under the RepTree, the DT is developed based on knowledge acquisition or by decrease in variance. At the start, it uses the values of numerical characteristics and then employs the fractional samples of C4.5 for any missing observations (Srinivasan & Mekala, 2014). The DT-RF technique uses variations of training data to develop decision trees. Parts of the original training data are replaced in a random fashion to get new versions of training data. Every tree is advanced as much as it can be with no shearing. Each tree is then allowed to create its own classification thus making a decision for most of the cases (Chen & Liu, 2005). Breiman (2001) considers this method to be far superior, as far as performance is concerned, than other algorithms of the genre.

Artificial Neural Network (ANN): The ANN model consists of three elements: the input layer, hidden layer, and output layer. It is akin to a human brain and works similarly wherein results are based on past experiences. Thus, even when working with complex non-linear situations, ANN models do not require strict hypotheses like standard statistical methods do. It employs the back propagation algorithm during the training process (Han, Kamber, & Pei, 2012).

Support Vector Machines (SVMs): SVMs create an n-dimensional hyperplane to separate the data into two categories (Haykin, 1999). Linear SVM is employed if the data has been separated in a linear fashion. If it cannot be separated linearly then non-linear SVM is applied (Alpaydin, 2004). The latter is used for



choosing the correct kernel function from an array of linear, radial basis, sigmoid, second-order multiple, polynomial, and reverse second-order kernels. The choice of kernels results in different SVMs and, hence, the outcomes are different too (Shawe-Taylor, Bartlett, Williamson, & Anthony, 1998). The right choice of kernel can impact learning capacity significantly (Varshney & Arora, 2004).

Logistic Regression (LR): LR is used to determine the association between dependent and independent variables, much like in standard regression models used. The crucial factor in this model requires the dependent variable to be continuous. In case the dependable variable has a value of 0 or 1, the binary LR is used for discernible independent variables, if any, which leads to classification of the dependent variable (Hosmer & Lemeshow, 2000).

Classification Criteria

To determine which algorithms were most effective, many classifications came into play. For the current research, the criteria used were accuracy (ACC), mean absolute error (MAE), Kappa (κ) statistic and ROC area.

Depending on the results, classifications could be termed as True positive (TP): correct positive prediction, False positive (FP): incorrect positive prediction, True negative (TN): correct negative prediction, and False negative (FN): incorrect negative prediction.

Equation 1 shows how ACC was computed. To compute ACC, the number of correct predictions were divided by the number of data sets (Donner & Klar, 1996). The κ statistic, suitable for categorical variables, measures the predictive performance of the model. It is also based on the value on the chi-square table. Equation 2 exemplifies this where p_0 and p_e clarify the link between two categorical variables. The MAE calculation is displayed in Equation 3. The MAE statistic shows the difference between the predicted and observed values, denoted as P_i and O_i , as shown in the Equation, where $P_i - O_i$ is the prediction error of the model (Willmott & Matsuura, 2005). The Receiver Operating Characteristic (ROC) curve is often used to gauge the performance of classification algorithms wherein the area under the curve shows how the classifier has fared (Bradley, 1997). This curve has the TP value on the Y axis and (1-TN) value on the X axis; the higher the ROC area value, the better the classification by the algorithm. All of these classification criteria can be computed numerically. Among these criteria, ROC Area provides visual assessment as well as numerical results about the performance of EDM algorithms. Thus, comparing different EDM algorithms' performances can be easily interpreted.

$$ACC = \frac{TP+TN}{TP+TN+FN+FP} \quad (1)$$

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (2)$$

$$MAE = n^{-1} \sum_{i=1}^n |P_i - O_i| \quad (3)$$

Feature Selection

Classification of algorithms requires close attention to be paid to the choice of features therein and only the effective ones are to be incorporated. The parsimony principle is to be followed for this and the minimum number of features in classification algorithms with no noteworthy depreciation in results are to be chosen. This rule was followed in this research and various feature selection techniques were used to obtain the relevant important features.



Cfs Subset: This method aims to find the best feature set by evaluating the feature sets with correlation. It tries to select a set of features with low correlation between them and features with high correlation with class tags (Gennari, Langley, & Fisher, 1989; Gümüşçü, Aydılek, & Taşaltın, 2016). In this method, highly correlated features (with high correlation coefficient) are excluded from the data set because uncorrelated features produce better classification (Hall, 2000).

Correlation attribute: It measures the Pearson correlation between features and outcome which is binary. In this method if the feature is measured by nominal scale then a weighted average is calculated for the overall correlation (Jiang, Meng, & Meng, 2009).

Gain Ratio: Information Gain is a feature selection method based on entropy. The information gain method tends to choose features with different values, so sometimes the results are obtained with bias. In order to reduce this bias, the Gain Ratio method is used. Split information is applied to reduce the bias. Gain ratio is the normalisation which is obtained by taken the ratio of the information gain values and split information (Karegowda, Manjunath, & Jayaram, 2010).

One-R: The One-R algorithm is employed for testing the entire data in order to form decision trees with specified rules. This technique is highly precise allowing for a suitable study of the data structure (Kabakchieva, 2013). The feature with the lowest error rate is selected with this algorithm as a result of which the features that are in the minority as far as feature value is concerned add to the error rate (Muda, Yassin, Sulaiman, & Udzir, 2011).

Application

For this research, the TIMSS 2015 result of eighth grade students in Turkey was used. The 'not-available' and 'missing' factors were excluded from the data set and the 10-fold cross validation was used to derive the training and testing data sets. The entire process was done in three steps that are listed ahead.

Step 1: As shown in Table 2, all 35 features were used to ascertain the performances of the algorithms with the classification criteria. This demonstrates the significance of feature selection by displaying the performance of algorithms and comparing with the other steps.

Step 2: This step was used using the scales which were mentioned in the data set section. By using these scales, algorithms were run and their classification performances were reported given in Table 3. The aim of this step was to determine the effectiveness of scales for classification of student science success.

Step 3: Tables 4-6 explain this step in which four feature selection methods were employed: the Cfs subset, gain ratio, correlation attribute, and One-R algorithm. These were applied to all 35 features and the most efficient factors were obtained. The analyses on these chosen factors were done again.

In this step, 4 different feature selection methods which were Cfs subset, correlation attribute, gain ratio and One-R algorithms were applied to all 35 features. Thus, most effective features which were extracted by using all feature selection methods were reported and the analyses on these chosen features were done again. With using new feature sets, the values of classification criteria for all classification algorithms were obtained and given in Table 4 – Table 6.

Weka, the Java-based open source software developed by the University of Waikato for application of data mining algorithms was utilised for this research (Frank, Hall, & Witten, 2016).

Research Results

Tables 2-6 illustrate the classification performances of algorithms based on classification criteria mentioned earlier (ACC, κ statistic, MAE, and ROC). The best algorithm in each step was considered to be the one based on the ACC criterion. Also, the figures of ROC areas illustrated the classification performances of selected algorithms as best in each step. Table 7 shows the results obtained. Other classification criteria were also used to support the final result.



Results of Step 1: The classification results of step 1 are given in Table 2. As it was seen, LR was the best algorithm according to ACC (0.738). κ statistic (0.469) and ROC area (0.820) support this result. Also, DT-RF and SVM-POLY had similar performances when compared with LR. Figure 1 shows the ROC areas of these selected algorithms.

Table 2. Classification results of step 1 (35 features).

Criteria	Classification Algorithms								
	NB	DT-C4.5	DT-RepTree	DT-RF	ANN	SVM-POLY	SVM-RBF	SVM-PUK	LR
ACC	0.712	0.668	0.696	0.732	0.682	0.734	0.724	0.722	0.738
κ statistic	0.413	0.329	0.385	0.457	0.355	0.463	0.440	0.437	0.469
MAE	0.306	0.354	0.373	0.374	0.322	0.266	0.276	0.278	0.342
ROC Area	0.789	0.661	0.739	0.811	0.749	0.730	0.718	0.717	0.820

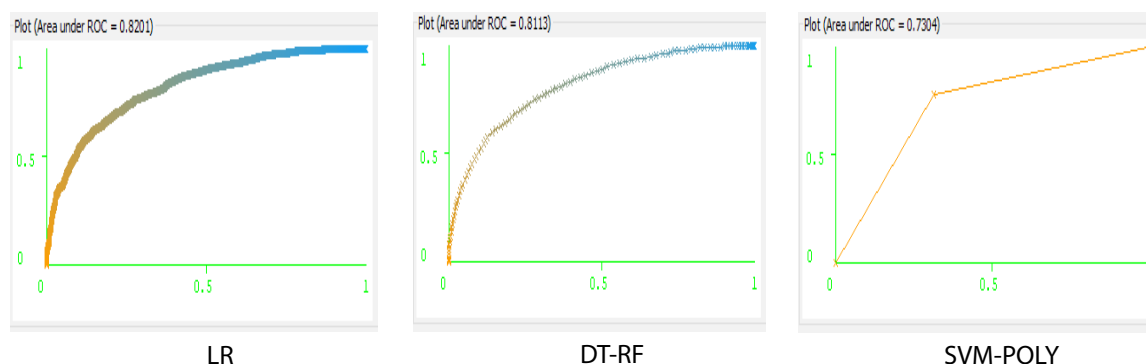


Figure 1. The ROC areas of LR, DT-RF and SVM-POLY in step 1.

Results of Step 2: The most important features were obtained by using scales. These scales were "Home Educational Resource", "Students' Sense of School Belonging", "Student Bullying", "Students Like Learning Science", "Engaging Teaching in Science Lessons", "Student Confident in Science", "Students Value Science", "Science Achievement Too Low for Estimation", "Weekly Time Spent on Science Homework". The performances of classification algorithms which were obtained by using these 9 features are given in Table 3. SVM-POLY was the best algorithm according to ACC (0.703). κ statistic (0.400) and MAE (0.298) supported this result. Also, LR and SVM-PUK had similar performances when compared with SVM-POLY. In Figure 2, ROC areas of these selected algorithms are illustrated.

Table 3. Classification results of step 2 (9 features).

Criteria	Classification Algorithms								
	NB	DT-C4.5	DT-RepTree	DT-RF	ANN	SVM-POLY	SVM-RBF	SVM-PUK	LR
ACC	0.675	0.677	0.673	0.687	0.696	0.703	0.687	0.702	0.701
κ statistic	0.348	0.348	0.341	0.370	0.388	0.400	0.359	0.397	0.394
MAE	0.383	0.389	0.394	0.388	0.381	0.298	0.313	0.298	0.391
ROC Area	0.737	0.696	0.721	0.749	0.758	0.700	0.677	0.698	0.766

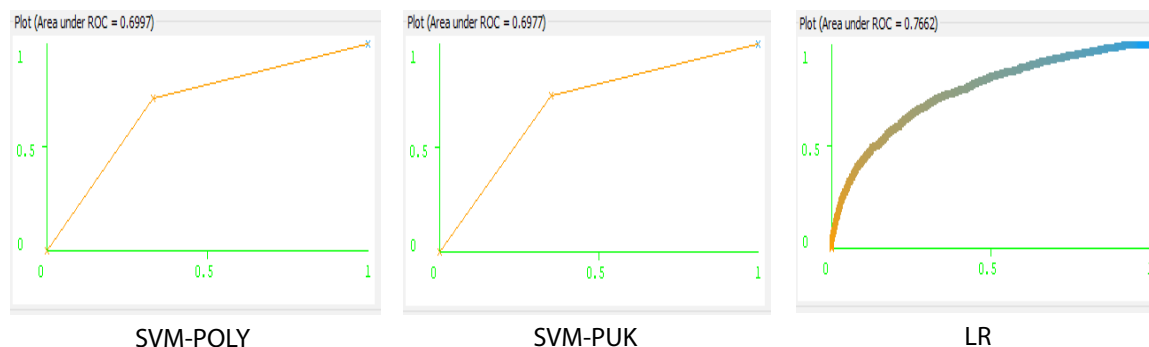


Figure 2. The ROC areas of SVM-POLY, SVM-PUK and LR in step 2.

Results of Step 3: In this step, according to four different feature selection methods, most effective features were extracted and classification performances of the algorithms were calculated.

The most important features were obtained by using Cfs subset feature selection method. These features were "Often Speak at Home", "Computer Tablet Shared", "Study Desk", "How Far in Education Do You Expect to Go", "About How Often Absent From School", "How Often Breakfast on School Days", "Communicate with Teacher", "How Often Teacher Give You Homework", "Extra Lessons Last 12 Month", "Extra Lessons How Many Month", "Home Educational Resources", "Student Confident in Science" and "Science Achievement Too Low For Estimation". The performances of classification algorithms which were obtained by using these 13 features are given in Table 4. SVM-POLY (ACC=0.734) and LR (ACC=0.733) were the best algorithms. κ statistic (0.460) and MAE (0.266) for SVM-POLY and ROC area (0.809) for LR supported this result. Figure 3 shows the ROC areas of these selected algorithms.

Table 4. Classification results of step 3 (13 features – Cfs subset method).

Criteria	Classification Algorithms								
	NB	DT-C4.5	DT-RepTree	DT-RF	ANN	SVM-POLY	SVM-RBF	SVM-PUK	LR
ACC	0.716	0.683	0.699	0.713	0.716	0.734	0.720	0.713	0.733
κ statistic	0.419	0.359	0.391	0.420	0.427	0.460	0.429	0.419	0.457
MAE	0.313	0.364	0.370	0.361	0.346	0.266	0.280	0.287	0.354
ROC Area	0.794	0.707	0.754	0.783	0.791	0.728	0.712	0.709	0.809



Figure 3. The ROC areas of SVM-POLY and LR in step 3 (Cfs subset).

When the correlation attribute and one R feature selection methods were used to extract the most important features, the same 7 features were obtained. These features were "Computer Tablet Shared", "How Far in Education



Do You Expect to Go”, “About How Often Absent From School”, “Extra Lessons Last 12 Month”, “Extra Lessons How Many Month”, “Home Educational Resources” and “Student Confident in Science”. The performances of classification algorithms which were obtained by using these 7 features are given in Table 5. SVM-POLY (ACC=0.714), SVM-PUK (ACC=0.714) and LR (ACC=0.713) were the best algorithms. κ statistic (0.421) and MAE (0.286) for SVM-POLY and SVM-PUK and ROC area (0.791) for LR supported this result. Figure 4 shows the ROC areas of SVM-POLY, SVM-PUK and LR in step 3 with correlation attribute and One-R feature selection methods.

Table 5. Classification results of step 3 (7 features – correlation attribute and One-R method).

Criteria	Classification Algorithms								
	NB	DT-C4.5	DT-RepTree	DT-RF	ANN	SVM-POLY	SVM-RBF	SVM-PUK	LR
ACC	0.708	0.700	0.703	0.673	0.709	0.714	0.710	0.714	0.713
κ statistic	0.405	0.394	0.400	0.339	0.414	0.421	0.410	0.421	0.417
MAE	0.330	0.377	0.377	0.370	0.366	0.286	0.290	0.286	0.371
ROC Area	0.782	0.733	0.756	0.733	0.784	0.710	0.704	0.710	0.791

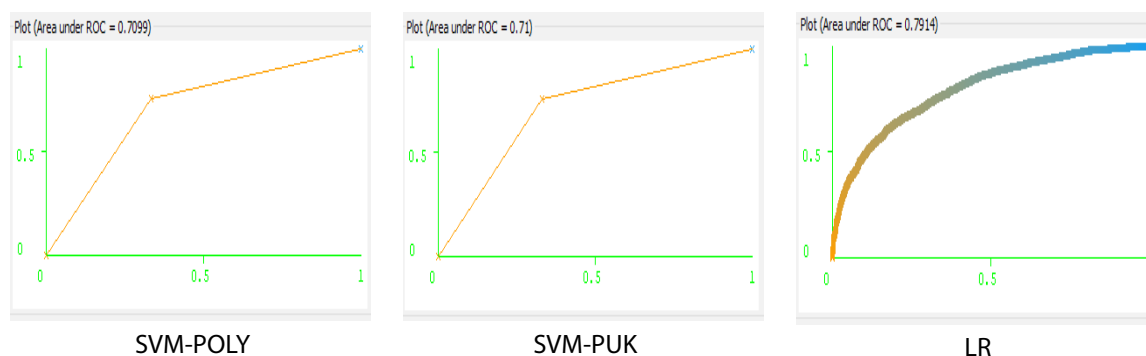


Figure 4: The ROC areas of SVM-POLY, SVM-PUK and LR in step 3 (correlation attribute and One-R).

The most important features were obtained by using gain ratio feature selection method and these features were “Often Speak at Home”, “Computer Tablet Shared”, “Washing Machine”, “How Far in Education Do You Expect to Go”, “Extra Lessons Last 12 Month”, “Extra Lessons How Many Month”, “Home Educational Resources”, “Student Confident in Science” and “Science Achievement Too Low For Estimation”. The performances of classification algorithms which were obtained by using these 9 features are given in Table 6. SVM-PUK was the best algorithm according to ACC (0.719). κ statistic (0.431) and MAE (0.281) supported this result. Also, LR, SVM-POLY and ANN had similar performances when compared with SVM-PUK. In Figure 5, ROC areas of these selected algorithms are illustrated.

Table 6. Classification results of step 3 (9 features – gain ratio method).

Criteria	Classification Algorithms								
	NB	DT-C4.5	DT-RepTree	DT-RF	ANN	SVM-POLY	SVM-RBF	SVM-PUK	LR
ACC	0.703	0.696	0.694	0.673	0.713	0.713	0.705	0.719	0.713
κ statistic	0.393	0.385	0.381	0.339	0.422	0.418	0.399	0.431	0.416
MAE	0.330	0.374	0.378	0.366	0.364	0.287	0.296	0.281	0.369
ROC Area	0.782	0.739	0.755	0.733	0.783	0.708	0.698	0.714	0.792

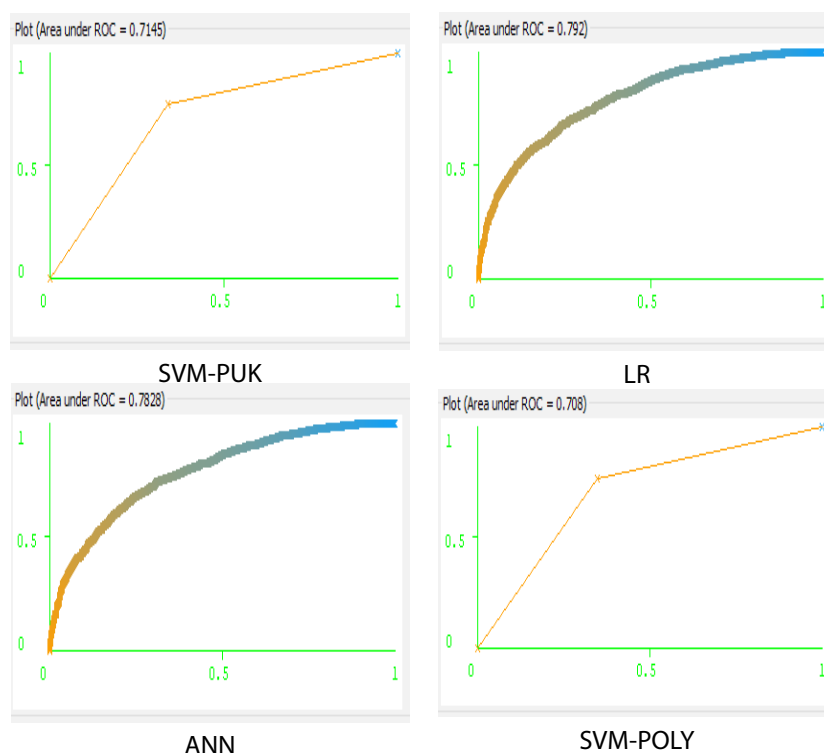


Figure 5. The ROC areas of SVM-PUK, LR, ANN and SVM-POLY and LR in step 3 (gain ratio).

The results obtained from step 1 to step 3 are summarized in Table 7. As it was seen, in all steps, LR and SVM-POLY produced the most successful classification results according to classification criteria. The classification results of SVM-PUK were very close to LR and SVM-POLY except step 1 and step 3 (Cfs subset). Also, DT-RF for step 1 and ANN for step 3 (Gain ratio) had similar performances when compared with LR and SVM-POLY. In addition to this, Table 7 shows the effective features that draw attention in the feature selection methods implemented in step 3. “Computer Tablet Shared” (BSBG06B), “How Far in Education Do You Expect to Go” (BSBG08), “Extra Lessons Last 12 Month” (BSBS26AB), “Extra Lessons How Many Month” (BSBS26BB), “Home Educational Resources” (BSBGHER) and “Student Confident in Science” (BSBGSCS) features were common in step 3. Therefore, it could be said that these features were the most important features in classifying the science success of the students.

Table 7. Summarized results of step 1 to step 3

Step	Algorithms	Features
Step 1	LR, SVM-POLY, DT-RF	All features
Step 2	SVM-POLY, LR, SVM-PUK	BSBGHER, BSBGSSB, BSBGSB, BSBGSL, BSBGESL, BSBGSCS, BSBGSVS, BSDSLOWP, BSDSWKHW
Step 3 (Cfs subset)	SVM-POLY, LR	BSBG03, BSBG06B, BSBG06C, BSBG08, BSBG11, BSBG12, BSBG14D, BSBS25AB, BSBS26AB, BSBS26BB, BSBGHER, BSBGSCS, BSDSLOWP
Step 3 (correlation attribute, One-R)	SVM-POLY, SVM-PUK, LR	BSBG06B, BSBG08, BSBG11, BSBS26AB, BSBS26BB, BSBGHER, BSBGSCS
Step 3 (Gain ratio)	SVM-PUK, LR, SVM-POLY, ANN	BSBG03, BSBG06B, BSBG06J, BSBG08, BSBS26AB, BSBS26BB, BSBGHER, BSBGSCS, BSDSLOWP

In order to investigate the performance changes of classification algorithms, one of the classification criteria could be used. In this research, ACC was chosen to show changes. As it was mentioned before, Table 2 shows the performances of classification algorithms with using all 35 features (step 1). In this step, the highest ACC belong to

LR (73.8%), SVM-POLY (73.4%) and DT-RF (73.2%). These classification accuracies were the highest values among step 1 to step 3. When the classification accuracies (according to ACC) of step 1 and other steps were compared, the following results were obtained. In step 2, with the 9 scales as effective features, ACC were found as SVM-POLY (70.3%), SVM-PUK (70.2%) and LR (70.1%). In step 3 (Cfs subset), there were 13 important features and with these features, the ACC were SVM-POLY (73.4%) and LR (73.3%). In step 3 (correlation attribute, One-R), there were 7 important features and with these features, the ACC were SVM-POLY (71.4%), SVM-PUK (71.4) and LR (71.3%). Lastly, in step 3 (Gain ratio), with 9 most effective features, the ACC were SVM-PUK (71.9%), LR (71.3%), SVM-POLY (71.3%) and ANN (71.3%).

The importance of parsimony principle was featured in step 3 (Cfs subset) where it was noticed that the values of classification accuracy of algorithms were almost the same as in step 1; this proved that reducing features led to almost the same results.

Discussion

Determining the factors that lead to success of students is an important element in education. TIMSS 2015 data of eighth grade students of science in Turkey is chosen in this research to determine some of these factors. Delineation of such factors can help in developing suitable educational policies and conditions to enhance the success rate of students. This research is an attempt to find answers to two research questions in this area: (1) which EDM algorithm(s) is/are appropriate to classify student success and, (2) what factors are extracted with using different feature selection methods in the purpose of determining the most effective ones.

LR and SVM-POLY were found the most apt algorithms for the first research question. In many existing studies, performed algorithms within EDM suggesting that LR and SVM-POLY were most appropriate for classifying student success. LR was considered by many scholars to be the best modelling technique among the various data mining methods to ascertain academic performance of students (Schreiber, 2002; Kılıç-Depren, Askin, & Öz, 2017). Another research by Delen (2010) employed ANN, SVM, C5 decision tree algorithm and LR to develop analytical models to forecast attrition among students and established that SVM gave the best results in this area (Delen, 2010). Also, it is found in this research that SVM-PUK too is not far behind LR and SVM-POLY.

The kernel selection is an important process to determine SVM classification performances. As a result of the selection of different kernel functions, different classification performances are achieved (Shawe-Taylor, Bartlett, Williamson, & Anthony, 1998). Depending on this, it was observed that SVM-RBF was not found as the best classification algorithm, but other SVMs algorithms had best classification performances in some steps. In addition, when the classification performances in all steps were examined, NB could not be found as the best performer algorithm in any step. Also, when the classification performances of three DT algorithms were examined in all steps, only DT-RF showed a successful classification performance in only first step.

As for the second question, different feature selection methods are used to determine the most important factors related with students' successes in science. The aim of using different feature selection methods is to use fewer factors without compromising the success rate of classification. Thus, the parsimony principle is achieved. The "Computer Tablet Shared", "Extra Lessons Last 12 Month", "Extra Lesson How Many Month", "How Far in Education Do You Expect to Go", "Home Educational Resources" and "Student Confident in Science" factors were seen to be common in all the feature selection algorithms used. The factors given above were also found important variables in earlier researches. The "Student Confident" factor was proven to be an important element of success in various studies (Liu & Meng, 2010; Hammouri, 2010; Askin & Gokalp, 2013; Kılıç-Depren, Askin, & Öz, 2017). Topçu, Erbilgin, and Arkan (2016) asserted in their study that those students who have ample access to educational resources achieve success which is corroborated by other studies too making "Home Educational Resources" as another important element. "Computer Environment" was found to be yet another crucial factor by Anil (2009) who examined PISA 2006 science data for Turkey. Extra time spent on science lessons outside of regular school classes was seen as another vital element in the PISA 2006 science study for Turkey (Ozer & Anil, 2011). In addition, in the research of Ogura (2006), extra time spent out of school was found an important factor on student levels of success.

Conclusions

The current research is quite different from earlier researches in the field of education. It has examined data from TIMSS 2015 in the subjects of science; this is the latest data released and has not been researched in any other



research on the subject yet. Moreover, when the earlier researches on TIMSS literature are investigated, it is seen that standard statistical techniques, (e.g., factor analysis, ANOVA) were used in clustering, prediction and regression. However, these techniques are losing favour somewhat due to their inherent precise assumptions. Data mining algorithms are becoming the technique of choice and the algorithms used in this research are the ones generally being opted for in EDM literature. Another property that is unique to this research is the importance given to data reduction when quantifying student success. It proves that only the most relevant factors need to be taken into account while classifying and all factors need not be collated. In addition to these contributions, educators and education policy makers can use the most important factors extracted in this research. In order to develop beneficial educational strategies and thus to improve the students' academic success, knowing important factors plays a vital role. Also, findings of this research are supported by earlier researches.

This research has some limitations. First of all, TIMSS study evaluates both mathematics and science successes but in this research the results are only based on TIMSS 2015 science data set. Moreover, findings of this research are based on the results of self-reported questionnaires developed by IEA. Another limitation, factors that effect students' successes are extracted for only Turkish eight grade students. Future researches in the subject can include research on mathematics results of TIMSS 2015. Many national/international studies like PISA can also be examined by the application of EDM algorithms in order to classify academic success.

References

- Alpaydın, E. (2004). *Introduction to machine learning*. Cambridge: The MIT Press.
- Anil, D. (2009). Factors effecting science achievement of science students in programme for international students' achievement (PISA) in Turkey. *Education and Science*, 34(152), 87-100.
- Askin, O. E., & Gokalp, F. (2013). 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. <http://dx.doi.org/10.1016/j.sbspro.2013.12.076>.
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Baradwaj, B. K., & Pal, S. (2011). Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, 2(6), 63-69.
- Bradley, A. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145-1159. [http://dx.doi.org/10.1016/S0031-3203\(96\)00142-2](http://dx.doi.org/10.1016/S0031-3203(96)00142-2).
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <http://dx.doi.org/10.1023/A:1010933404324>.
- Chen, X. W., & Liu, M. (2005). Prediction of protein-protein interactions using random decision forest framework. *Bioinformatics*, 21(24), 4394-4400. <https://doi.org/10.1093/bioinformatics/bti721>.
- Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In: A. Brito & J. Teixeira (Eds.), *Proceedings of 5th Annual Future Business Technology Conference*. (9-11 April 2008, Porto) (pp. 5-12). Porto, Portugal: EUROSIS.
- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4), 498-506.
- Donner, A., & Klar, N. (1996). The statistical analysis of kappa statistics in multiple samples. *Journal of Clinical Epidemiology*, 49(9), 1053-1058. [http://dx.doi.org/10.1016/0895-4356\(96\)00057-1](http://dx.doi.org/10.1016/0895-4356(96)00057-1).
- Filiz, E., & Öz, E. (2017). Classification of BIST-100 Index'Changes via machine learning methods. *Marmara University Journal of Economic & Administrative Sciences*, 39(1), 117-129. <https://doi.org/10.14780/muiibd.329913>.
- Frank, E., Hall, M. A., & Witten, I. H. (2016). *The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques* (Fourth Edition). United States: Morgan Kaufmann.
- Gennari, J. H., Langley, P., & Fisher, D. (1989). Models of incremental concept formation. *Artificial Intelligence*, 40(1-3), 11-61. [https://doi.org/10.1016/0004-3702\(89\)90046-5](https://doi.org/10.1016/0004-3702(89)90046-5).
- Gümüşçü, A., Aydılek, İ. B., & Taştaltın, R. (2016). Comparison of feature selection algorithms on microarray data classification. *Harran University Journal of Engineering*, 1(1), 1-7.
- Hall, M. A. (2000). *Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning*, The University of Waikato, Working Paper 00/8. Hamilton-New Zealand.
- Hammouri, H. (2010). Attitudinal and motivational variables related to mathematics achievement in Jordan: Findings from the Third International Mathematics and Science Study (TIMSS). *Educational Research*, 46(3), 241-257. <http://dx.doi.org/10.1080/0013188042000277313>.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concept and techniques* (3rd ed.). Burlington, MA: Morgan Kaufmann Publishers.
- Haykin, S. (1999). *Neural Networks: A comprehensive Foundation*. Upper Saddle River-N.J.: Prentice Hall International.
- Horáková, T., Houška, M., & Dömeová, L. (2017). Classification of the educational texts styles with the methods of artificial intelligence. *Journal of Baltic Science Education*, 16(3), 324-336.



- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). Hoboken-N.J.: John Wiley & Sons Incorporated.
- Jiang, F., Meng, W., & Meng, X. (2009). Selectivity estimation for exclusive query translation in deep web data integration. In: *International Conference on Database Systems for Advanced Applications* (21-23 April 2009, Brisbane, Australia) (pp. 595-600). Heidelberg-Berlin: Springer.
- John, G. H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. (18-20 August 1995, Montréal, Qué, Canada) (pp. 338-345). San Francisco-United States: Morgan Kaufmann Publishers Incorporated.
- Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61-72. <https://doi.org/10.2478/cait-2013-0006>.
- Kalmegh, S. (2015). Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news. *International Journal of Innovative Science, Engineering & Technology*, 2(2), 438-446.
- Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, 2(2), 271-277.
- Kiliç-Depren, S., Askin, Ö. E., & Öz, E. (2017). Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS. *Educational Sciences: Theory & Practice*, 17(5), 1605-1623. <https://doi.org/10.12738/estp.2017.5.0634>.
- Kiray, S. A., Gok, B., & Bozkir, A. S. (2015). Identifying the factors affecting science and mathematics achievement using data mining methods. *Journal of Education in Science, Environment and Health*, 1(1), 28-48.
- Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5), 411-426. <http://dx.doi.org/10.1080/08839510490442058>.
- LaRoche, S., Joncas, M., & Foy, P. (2016). Sample design in TIMSS 2015. Martin, M. O., Mullis, I. V. S., and Hooper, M. (Eds.). (2016). *Methods and procedures in TIMSS 2015*. Retrieved from Boston College, Chestnut Hill-United States: TIMSS & PIRLS International Study Center.
- Liu, S., & Meng, L. (2010). Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics self-concept. *Educational Psychology*, 30(6), 699-712. <http://dx.doi.org/10.1080/01443410.2010.501102>.
- Martínez Abad, F., & Chaparro Caso López, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, 28(1), 39-55. <http://dx.doi.org/10.1080/09243453.2016.1235591>.
- Masters, G. N., & Wright, B. D. (1997). The partial credit model. In M.J. van de Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory*. Berlin: Springer.
- Mohamad, S. K., & Tasir, Z. (2013). Educational data mining: A review. *Procedia-Social and Behavioral Sciences*, 97, 320-324. <http://dx.doi.org/10.1016/j.sbspro.2013.10.240>.
- Muda, Z., Yassin, W., Sulaiman, M. N., & Udzir, N. I. (2011). Intrusion detection based on k-means clustering and OneR classification. In: *7th International Conference on Information Assurance and Security (IAS)*. (5-8 December 2011, Melaka, Malaysia) (pp. 192-197). United States: The Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/ISIAS.2011.6122818>.
- Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill-United States: TIMSS & PIRLS International Study Center.
- Ogura, Y. (2006). Background to Japanese student achievement in science and mathematics. In S. J. Howie & T. Plomp (Eds.), *Content of learning mathematics and science. Lessons learned from TIMSS* (pp. 313-331). London, New York: Routledge.
- Ozer, Y., & Anil, D. (2011). Examining the factors affecting students' science and mathematics achievement with structural equation modeling. *Hacettepe University Journal of Education*, 41, 313-324.
- Quinlan, J. R. (2014). *C4.5: programs for machine learning*. San Mateo-California: Morgan Kaufmann Publishers..
- Pai, P. F., Chen, C. T., Hung, Y. M., Hung, W. Z., & Chang, Y. C. (2014). 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.
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4), 1432-1462. <http://dx.doi.org/10.1016/j.eswa.2013.08.042>.
- Rajni, J., & Malaya, D. B. (2015). Predictive analytics in a higher education context. *IT Professional*, 17(4), 24-33. <http://dx.doi.org/10.1109/MITP.2015.68>.
- Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. *International Journal of Computer Science Issues*, 7(1), 10-18. <http://dx.doi.org/10.1.1.403.8058>.
- Rokach, L., & Maimon, O. (2005). Decision trees. In: Maimon O., Rokach L. (Eds), *Data mining and knowledge discovery handbook*. Boston-MA: Springer. https://doi.org/10.1007/0-387-25465-X_9.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146. <http://dx.doi.org/10.1016/j.eswa.2006.04.005>.
- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *The Institute of Electrical and Electronics Engineers (IEEE) Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618. <http://dx.doi.org/10.1109/TSMCC.2010.2053532>.
- Schreiber, J. B. (2002). Scoring Above the International Average: A Logistic Regression Model of the TIMSS Advanced Mathematics Exam. *Multiple Linear Regression Viewpoints*, 28(1), 22-30.
- Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422. <https://doi.org/10.1016/j.procs.2015.12.157>.
- Shawe-Taylor, J., Bartlett, P. L., Williamson, R. C., & Anthony, M. (1998). Structural risk minimization over data-dependent hierarchies. *The Institute of Electrical and Electronics Engineers (IEEE) Transactions on Information Theory*, 44(5), 1926-1940. <https://doi.org/10.1109/18.705570>.



- Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In: *Proceedings of the 2nd international conference on learning analytics and knowledge*. (29 April-2 May 2012, Vancouver, British Columbia, Canada) (pp. 252-254). New York-United States: Association for Computing Machinery.
- Sinharay, S. (2016). An NCME instructional module on data mining methods for classification and regression. *Educational Measurement: Issues and Practice*, 35(3), 38-54. <https://doi.org/10.1111/emip.12115>.
- Srinivasan, D. B., & Mekala, P. (2014). Mining social networking data for classification using REPTree. *International Journal of Advance Research in Computer Science and Management Studies*, 2(10), 155-160.
- Topçu, M. S., Erbilgin, E., & Arıkan, S. (2016). Factors predicting Turkish and Korean students' science and mathematics achievement in TIMSS 2011. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(7), 1711-1737. <https://doi.org/10.12973/eurasia.2016.1530a>.
- Varshney, P. K., & Arora, M. K. (2004). *Advanced image processing techniques for remotely sensed hyperspectral data*. Berlin-Heidelberg: Springer Science & Business Media.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. <http://dx.doi.org/10.3354/cr030079>.
- Zhang, H. (2004). The optimality of naive Bayes. In: *FLAIRS Conference*. (12-14 May 2004, Miami Beach, Florida) (pp. 562-567). Menlo Park-California: The Association for the Advancement of Artificial Intelligence Press.

Received: January 17, 2019

Accepted: April 03, 2019

Enes Filiz	PhDc., Research Assistant, Department of Statistics, Faculty of Arts and Sciences, Yildiz Technical University, Davutpasa Campus, Istanbul, 34220, Turkey. E-mail: enesf@yildiz.edu.tr Website: http://avesis.yildiz.edu.tr/enesf/
Ersoy Öz	PhD., Associate Professor, Department of Statistics, Faculty of Arts and Sciences, Yildiz Technical University, Davutpasa Campus, Istanbul, 34220, Turkey. E-mail: ersoyoz@yildiz.edu.tr Website: http://avesis.yildiz.edu.tr/ersoyoz/

