Journal of Modern Technology and Engineering
Vol.5, No.3, 2020, pp.271-282

ONLINE COURSE SUCCESS PREDICTION OF STUDENTS WITH MACHINE LEARNING METHODS

Emrah Emirtekin[1*], **Melike Karatay**[2], **Tarık Kışla**[2]

[1]Yaşar University, İzmir, Turkey
[2]Ege University, İzmir, Turkey

Abstract. Modeling the success of students according to the variables obtained from the online learning environment and student information system; It is important to identify students who tend not to attend the course, to predict the possible failure of students at the end of the semester and to provide the student with the right support. These models can be used to classify students' achievements and give individual or group-specific feedback, taking into account the interaction levels of students in online learning environments. In addition, it will be beneficial for the instructor who offers the online course to see the general level of the students and to organize the education process. In this study, students' course success in online courses was estimated by using different machine learning methods by using their interactions on the learning management system and demographic information in the student information system. Predictions were made with Naïve Bayes, Logistic Regression, k-NN (k-Nearest Neighborhood), SVM (Support Vector Machines), Artificial Neural Networks and Random Forest machine learning models. In the classification of students' end-of-term achievements, k-NN has been the machine learning model with the highest accuracy with 98.5.

Keywords: online learning, machine learning, education.
AMS Subject Classification: 68T07, 68T01.
Corresponding author: Emrah Emirtekin, Yaşar University, Bornova, İzmir, Turkey, Tel.: +905056065260, e-mail: *emrah.emirtekin@yasar.edu.tr*
Received: 13 July 2020; Accepted: 2 October 2020; Published: 24 December 2020.

1 Introduction

The rapid development of technology has closely affected many areas. One of these areas is distance education, the importance of which is more understood along with epidemics. Students have the opportunity to participate in their classes online or in a blend, regardless of time and place, through distance education. These courses, which are conducted online, are usually conducted over a Learning Management System (LMS). LMSs are one of the systems widely used in both online courses and blended courses, in the presentation of course contents within a certain plan, in communication and interaction between the student and the instructor, and in the measurement and evaluation processes of the student. It is also used as a system where data on students' course-related activities are kept and reported later. It uses data recorded in various sources such as determining the academic performance of educational institutions and organizations, predicting students' general or course-based success, determining the factors affecting their success. With these data, observing the academic or course-related performance of students is critical in terms of improving / shaping education in the process. Today, both Machine Learning (ML) and Deep Learning (DL) methods are frequently used in predicting the success of students. In the literature, there are many national scientific studies conducted by researchers on the prediction of course and academic success in the field of education. These studies are presented in Table 1 together with their performance values. In the studies, it has been determined that machine learning models are widely used. This study is important in terms of supporting the studies in this field in the literature and creating a generalizable and dynamic model with a higher performance value than existing studies that can be used in online courses. The models that provide feedback to both the student and the lecturer during the semester will contribute to the student's self-assessment and the instructor's ability to observe the education process of the students individually and collectively. In this study, it was aimed to develop and compare ML models based on the data of the student's demographic and course information on the Student Information System (SIS) and the course activities on LMS for a period in the evaluation of students' online course success at the end of the term. In this study, students' course success in online courses was estimated by using different machine learning methods, using their interactions on the learning management system and demographic information in the student information system. The data used within the scope of the study consists of the interactions of 4470 students studying in different courses conducted online at Yaşar University in the fall semester of 2019-2020 on the learning management system and demographic information on the student information system.

2 Literature

2.1 Machine Learning

Machine learning can be defined as software modeling of self-learning systems by making meaningful inferences from data or experiences with some mathematical and statistical operations (Alpaydın, 2020). Today, many machine learning models are used in order to find solutions to different problems. Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machines (SVM), k-Nearest Neighbor (k-NN), Artificial Neural Networks (ANN) and Random Forest (RF) can be shown as examples (Murat, 2017; Aluko et al., 2016; Corrigan & Smeaton, 2017; Kotsiantis, 2012; Slim et al., 2014). Several different learning methods are used during the training of these models. These are supervised learning, unsupervised learning and reinforcement learning. Supervised learning is a machine learning method that aims to obtain a function that covers all of these data by making use of a data set that has been observed and whose outputs are known (labeled) (Nizam & Akın, 2014). In other words, supervised learning is a method of generating a function from training data in solving a problem. In this method, it is aimed to estimate the output with the inputs in the data set. Unsupervised learning, on the other hand, is a machine learning method used to discover connections that are not clearly visible on the raw data set. In this method, inputs do not have an output. So what we want to predict is not tagged on the data set. It is a method that tries to learn only through inputs. It is generally used in the solution of clustering problems (Rana & Garg, 2016). Reinforcement learning is a machine learning method that has been used in many different disciplines, especially computer science, and has shown great interest in recent years. In reinforcement learning, the agent (learner) learns a desired behavior through trial and error. The learner is not trained with a certain data set. Learning takes place through reward and punishment method (Sasakawa et al., 2008; Özyer, 2016). Various mathematical and statistical calculations are used to solve a problem in the PLC models developed in line with these methods. Sometimes it can be a classification problem and sometimes a clustering problem that is desired to be solved with models. It is very important in terms of the usability of the model that the model presented can solve the problems with the least calculation process and the least resource. At this stage, computer science is used a lot.

2.2 Studies on Prediction of Student Success

Predicting student success using Support Vector Machines (SVM), Decision Trees (DT), Naive Bayes (NB), Artificial Neural Networks (ANN), Regression, Logistic Regression (LR), Random Forest (RF) and other machine learning methods Detailed information about the studies, methods, attributes and success situations are presented below.

Table 1: Previous Studies

Study	Sample	Method	Type	Success
Güner ve Çomak (2010)	434	ML	Course	%86
Şengür ve Tekin (2013)	127	ML	academic	%93
Turhan, Kurt ve Engin (2013)	111	ML	Course	%93
Akçapınar (2014)	76	ML	Course	%86
Aydın ve Özkul (2015)	180554	ML	academic	%82
Turhan, Kurt ve Engin (2013)	111	ML	Course	%93
Güldal ve Çakıcı (2017)	70	ML	Course	%71
Gök (2017)	1492	ML	academic	%63
Aydemir (2017)	1387	ML	academic	-
Özkan, Koçoğlu ve Erol (2018)	395	DL	Course	%84
Aydoğan ve Karcı (2018)	200	ML	academic	%94
Aydoğan ve Zırhlıoğlu (2018)	657	ML	Course	%97
Altun, Kayıkçı ve Irmak (2019)	578	DL	Course	%94
Aydemir (2019)	3789	ML	Course	%81
Duygu ve Aracı (2019)	300	ML	academic	%84

ML: *Machine Learning*, **DL**: *Deep Learning*

Güner and Çomak (2010) estimated the success of 434 students who were enrolled in engineering faculty in Mathematics I course by using the data of mathematics, Turkish and science results in the university entrance exam and their high school graduation success scores. In the study, 289 student data were used for education and 145 student data were used for the test. In the study, in which the SVM model was used, it was estimated that the students will pass the relevant course with an accuracy of 86 percent. Şengür and Tekin (2013) predicted the graduation grades of their students with ANN and Decision Trees. The qualifications of the research are the 49 course passing grades received by 127 students studying between 2007 and 2011. In the research, first estimation was made by using the scores of the 24 courses they took in the first two years and then the scores of the (38) courses they took in the first 3 years. In the study, 101 samples were reserved for training and 26 samples for testing, and tanh was used as an activation function. The estimation was made with an accuracy rate of 74 percent in the decision trees model and 93 percent in the ANN. In their study, Turhan et al (2013) estimated the scores of 111 students, 56 percent male and 44 percent female, who were studying at the Faculty of Medicine, by means of Artificial Neural Networks (ANN) and regression analysis. In this study, the scores of three midterm exams, the type of high school they graduated from, gender, family income, harmony with family, harmony with friends, physical environment and accommodation features were used to predict the final exam. In the multi-layer model they created, there are nine input variables, 27 neurons, seven hidden layers and an output layer layer. Conjugate gradient descent algorithm was used in training the network. After 48 hours of network development, they got the least error in the 65th epok. As a result of the research, it was stated that ANN ($R = .93$) has higher estimation than regression analysis ($R = .85$).

Akçapınar (2014) conducted all extracurricular activities such as writing reflection, participating in discussions, and tracking resources, in addition to the 14-week courses she conducted face-to-face with second year university students, through an online learning environment. 28 attributes have been removed for 76 students using this environment. Decision tree algorithms such as ID3, C4.5, CART were used for feature selection. After scaling the data, 10 features selected according to the gini index were determined, and student performance was estimated with 86 percent accuracy with the CN2 model. Aydın and Özkul (2015) estimated student performance with data from different sources on students studying at Anadolu University Open Education System in order to contribute to the planning of the distance education system. The data collected within the scope of the research consists of 11 features and 180.554 samples. 50 percent of this data was used as training and 50 percent as a test. The attributes used within the scope of the research are: course name, e-Book, e-Audio book, e-Exercise, e-TV, e-Exam, amount of course purchase, average of the exam, age, number of correct answers in mock exams, success rate in mock exams. C5.0 decision tree, LR, ANN, CHAID decision tree, CRT decision tree and QUEST models were used to train the data. The highest accuracy rate was C5.0 with 82.14 percent. In their study, Turhan et al (2013) estimated the scores of 111 students, 56 percent of whom were males and 44 percent females, who were studying at the Faculty of Medicine, by ANN and regression analysis. In this study, the scores of three midterm exams, the type of high school they graduated from, gender, family income, harmony with family, harmony with friends, physical environment and accommodation features were used to predict the final exam. In the multi-layer model they created, there are nine input variables, 27 neurons, seven hidden layers and an output layer layer. Conjugate gradient descent algorithm was used in training the network. After 48 hours of network development, they got the least error in the 65th epok. As a result of the research, it was stated that ANN (93 percent) had higher estimation than regression analysis (85 percent). On the learning management system, Güldal and Çakıcı (2017) used the attributes of gender, number of homework submissions, number of days they attended the course, the number of lectures, the number of clicks on the lecture, and the success status to predict the academic success of 70 students who interacted with the course on the learning management system. NB, DT and KNN models were used to classify students. The highest accuracy rate was NB with 71.2 percent. Gök (2017) predicted the academic achievement of students in Turkish, Mathematics, and general success at the end of the semester with machine learning models with 24 attributes and 1492 samples based on family, demographic and school information. It used the Correlation-based Feature Subset Evaluation method while choosing the attribute. Regression, RF, Linear SVM and RTF. SVM models were used in the study, with 63.8 percent, RF gave the most successful prediction. Özkan et al (2018) used the "student performance" data set of 395 students with 33 attributes in the UCI Machine Learning Repository in their study, in which they predicted student performance with artificial intelligence learning methods. The data set were randomly allocated for 90 percent training and 10 percent for testing. C5, Boosted C5, Regression Trees, SVM, LR, RF and Deep Learning algorithms were used in data training. In the deep learning model, three hidden layers are used except the input and output layers. There were 256 nerve cells in each secret layer, and relu was preferred as the activation function. In the classification models, C5.0 gave 86 percent accuracy, Boosted C5 82 percent, SVM 78 percent, Logistic Regression 82 percent, RF 84 percent and Deep Learning 87 percent. Aydoğan and Zırhlıoğlu (2018) showed 657 university students' academic achievement at the end of the third year with 17 attributes (faculty, department, age, gender, high school type, diploma grade, OSYM (Student Selection and Placement Centre) score, score type, mother's education status, father's education status, status, family monthly income, province, affectionate settlement in the department, 2nd grade term grade), 54 percent of the data were allocated for education, 29.1 percent for testing and 16.9 percent for validity (holdout). There are a total of 103 cells in the model consisting of input, output and intermediate layer. Sigmoid was used as the activation function. The classification accuracy rate of the study was found to be 97.2 percent. In their study, Altun et al (2019) estimated the graduation grades of 578 students studying at the Faculty of Education Classroom Teaching using multiple linear regression analysis and ANN. Gender, marital status, age and midterm exam scores of the courses taken in the first semester of the first year were used as attributes in the study. In their study based on the average absolute error value, they provided 94.3 percent successful prediction with multiple linear regression and 94.43 percent with ANN. The ANN model consists of 10 input variables, 2 hidden layers (four neurons in layer 1, three neurons in layer 2) and an output layer. Sigmoid function is used as activation function. The lowest error rate of 0.02 was observed as the learning rate. Cross validation was used in the training and testing process of the model. Aydemir (2019) used different learning models in the Weka program in his study, where he predicted the grades of 3789 university students to pass the Foreign Language II course. He estimated the course passing scores of the students with 12 different subjects (education type, faculty, department, program, program type, instructor and title, the type of student entry to the program, entrance score and entry ranking, and the grade average of the previous term). Among 19 different classification methods, Bagging (M5P) is the most successful model with an accuracy of 0.81. Emotion and Tool (2019) used 300 students with 24 attributes from the data set sharing site called UCI Machine Learning Repository in their study where they classified the academic performance of students with decision tree algorithms. Among the decision trees algorithms, J48 gave the most successful result with 84.73 percent accuracy. In the study, attributes such as gender, marital status, family income, place of residence, number of friends, family size, working hours, course attendance, travel time to school, debt belonging to the previous period were used.

3 Method

3.1 Data set

In this section, information is given about the data set to be used in the study. Within the scope of the research, the data consists of the interactions of 4470 students studying in the online courses conducted at Yaşar University in the fall semester of 2019-2020, and the demographic information on the SIS. The data set contains 17 features, 11 of which are numeric and 6 are categorical. The attribute to be predicted is the letterCode. There are 11 different classes (A, A-, B +, B, B-, C +, C, C-, D +, D, F) in the letterCode attribute. Information on other features are presented in detail in Table 2. 80 percent of the data set was used for training and 20 percent for testing.

3.2 Data Pre-Processing

In this section, pre-processing for the features used in ML models are listed as follows;

Data obtained from different sources (LMS, SIS) were combined.

The missing data of the numerical features on the data set are filled with the average values of the related features.

Data belonging to categorical features are marked as other. Outlier samples in the data set were determined by using one class SVM (one class SVM). For extreme values, Nu parameter, which specifies the upper and lower limits, was selected as 10 percent, and kernel coefficient as 0.05. At the end of this process, 456 samples were removed from the data set.

Whether data is normally distributed or not is known as a factor affecting the operation of some algorithms. If the data is leaning to the right or left, the performance of the model may be affected. It has been scaled by standardization to ensure the normal distribution of numerical features on the data set.

ML learning models do not work directly on categorical data. Categorical features must be transformed in order to perform the calculations. After these attributes are digitized with

Table 2: Attributes

Resources	Attributes	Type	Description
LMS	Homework	Numeric	homework submission
LMS	CourseContent	Numeric	Visiting course pages
LMS	EducationalMaterials	Numeric	Download education materials
LMS	syllabus	Numeric	Visiting the syllabus
LMS	TestResult	Numeric	Test or quiz
LMS	Visit	Numeric	Visit count
LMS	video	Numeric	watching course videos
LMS	CourseCode	Categorical	Course code
LMS	CourseLanguage	Categorical	Course Language
SIS	Gender	Numeric	gender
SIS	Age	Numeric	age
SIS	Faculty	Categorical	Faculty
SIS	Department	Categorical	Department
SIS	Level	Categorical	Class
SIS	LetterCode	Categorical	Final letter grade
SIS	Ects	Numeric	Work load
SIS	CourseCount	Numeric	the number of courses taken in the term

the Label Encoder, they are represented as binary with One Hot Encoding. As a result of this process, 17 features in the data set increased to 80.

Since the attribute to be predicted (letterCode) has more than two classes, this attribute is also digitized and represented as binary in order to perform operations in the output layer.

The data set to be used in ML models is divided into two groups as test and training data. The model was trained with training data, and the success of the model was evaluated with test data. 20 percent of the data set was used for testing and 80 percent for training.

Stratified sampling was used to represent classes equally in both training and test data sets.

3.3 Machine Learning Models

In this section, information about the ml models used in the research is presented.

3.3.1 Naïve Bayes

It is a ML method created on the basis of Bayes' Theorem. It is a parametric classification model that gives the highest probability for a group of observed variables (Rish, 2001). For an input vector $x = (x1, x2, ..., xn)$ where C is a class, $P(F|C) = x1, x2, .., xn|A$ class C that makes the probability $C)$ the greatest is sought.

3.3.2 Logistic Regression

LR ensures that the relationship between the other attributes (independent variables) and the target attribute (dependent variable) in case the target attribute is a categorical variable. Classes of dependent variables are predicted rather than continuous output values. If the desired attribute has two categories, it is expressed as binary LR, if it has more than two ordered categories, ordinal LR, and if it has more than two unordered categories, it is expressed as multinomial LR (Karabulut and Alper, 2011).

3.3.3 kNN

k-NN is a non-parametric classification method because it does not have any probabilistic assumptions. Classification process is made according to distance calculation. The distance of an instance to which class it is desired to be found is calculated by calculating the distance from the other samples, ranked in descending order, and the process is done by taking the observations with the smallest distance value. Majority voting or weight voting methods are used (Taşcı and Onan, 2016). In this model, k is determined as 10. Manhattan is used for distance calculation, and distance is used for weight.

3.3.4 Support Vector Machines

SVMs were developed by Cortes and Vapnik (1995) and are mainly used for separating data belonging to two classes and for pattern recognition. Decision limits are determined for the classification process. In other words, hyper planes are created. It is used as a supervised learning model based on statistical learning theory. They are effective in higher dimensional spaces when the number of dimensions is greater than the sample size. It is divided into linear and nonlinear SVM. Cost value was taken as 1 and epsilon value as 0.1.

3.3.5 Random Forest

RF is a decision tree-based classification algorithm. It is an algorithm that can perform successful classification operations where a forest is created using many decision tree structures. For an example, classification is carried out on each tree created in the forest. The class in the forest is determined by voting (Pal, 2005). After combining the results of multiple decision trees, more reliable results can be obtained by making a single decision. RF does not have an overfitting problem. In order to create the tree structure, the number of instances and trees to be used in each node must be determined. 10 trees are used in the forest in this model.

3.3.6 Artificial Neural Networks

ANN is a model of ML based on human nerve cells. This model has a structure similar to brain nerve cells, which is called the perceptron. In Figure 1, there are inputs of a sensor, a cost function processing these inputs and a probabilistic value activation function as a result of the system. The training process of showing the first instance on the data set to the network starts. A new value is obtained by multiplying each input value with the random weights given in the model. These values are passed through the activation function and an output is produced. The forward propagation process of the network is completed by calculating all output values. After this process, the back propagation process begins. The purpose of back propagation is to optimize the weights to reduce the total error so that the network can learn better. One time the forward propagation and back propagation process occur is called an epoch. The number of entries to train in an epoch is determined by bacth size. ANNs find wide application areas in many different disciplines today. In addition, it is very successful in solving complex problems successfully (Ergezer, Dikmen,and Özdemir, 2003). This model has been trained in 500 iterations using 450 neurons, ReLU activation function, adam optimization algorithm, 0.0001 as learning rate.

3.4 Implementation

Python programming language was used in the creation and testing of ML models. The scikit-learn library, which is frequently used in the pre-processing of data, creating and visualizing ML models, was used.

4 Findings

In this section, performance of the ML models are presented. Classification performances of ML models are given by Sieve diagram. However, first of all, brief information was given about the

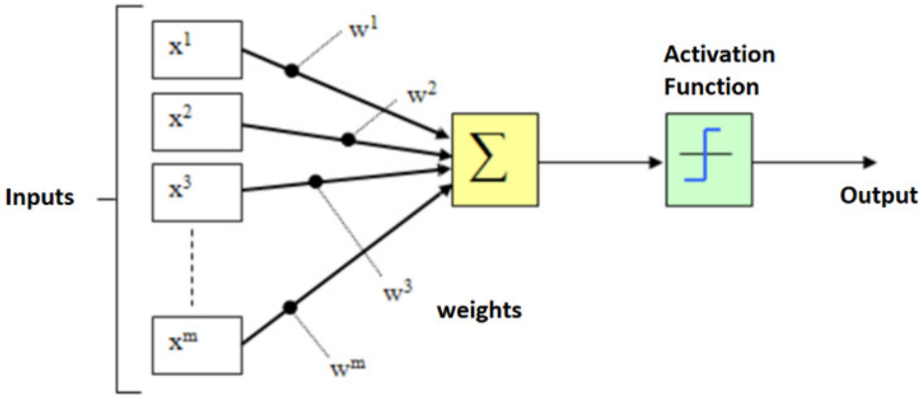

Figure 1: Basic structure of a sensor belonging to ANN, (https://en.m.wikipedia.org/wiki/)

basic concepts used in evaluation in order to better evaluate the performance situations. These concepts are accuracy, precision, recall, F-Score (F1), ROC Curve (Receiver Operator Curve) and AUC (Area Under Curve). Confusion Matrix (Table 3) is used to evaluate the success of ML models. The Confusion Matrix (CM) consists of calculating correctly classified entries and incorrectly classified entries. The rows in the CM indicate the real class of the samples in the test set, and the columns indicate the classification estimates of the ML models.

Table 3: Confusion Matrix

		Predicted Class	
		Positive	Negative
Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

4.1 Evalution Criteria

4.1.1 Accuracy

The accuracy is obtained by dividing the classified samples (TN + TP) by the total number of samples (FN + TN + TP + FP).

4.1.2 Precision

recision is obtained by dividing correctly classified positive samples (TP) by the number of samples whose class was positively predicted (FP + TP).

4.1.3 Recall

Recall is obtained by dividing correctly classified positive samples (TP) by the total number of correctly classified samples (FP + TP).

4.1.4 F-Score

There is an opposite relationship between sensitivity and sensitivity. Increasing the value of one can decrease the value of the other. The F-score, which is the harmonic mean of both cases, is used to obtain more precise and sensitive results.

4.1.5 ROC-Curve

Accuracy, precision and recall on some data sets do not always give accurate results. Such data sets usually contain imbalanced data. ROC-Curve is used in the evaluation of models in data sets containing such data.

4.1.6 AUC

It is known as the area under the ROC-Curve. The size of the area under the curve is directly proportional to the classification success of ML models. In other words, the larger the area indicate the higher the success of the model.

4.2 Performance of Models

Table 4 shows the performance values of ML models over the evaluation criteria.

Table 4: Performance of ML Models

Model	AUC	Accuracy	F1	Precision	Recall
kNN	0.998	0.985	0.985	0.986	0.985
ANN	0.997	0.974	0.974	0.975	0.974
RF	0.990	0.948	0.947	0.948	0.948
SVM	0.896	0.509	0.469	0.558	0.509
Naive Bayes	0.779	0.363	0.356	0.362	0.363
LR	0.764	0.350	0.300	0.281	0.350

4.3 Performance Diagrams of Models

Sieve diagram was used to show the performance levels of ML models in classification. In kNN, the distinction between classification of BC models visually towards LR is shown in figure 2.

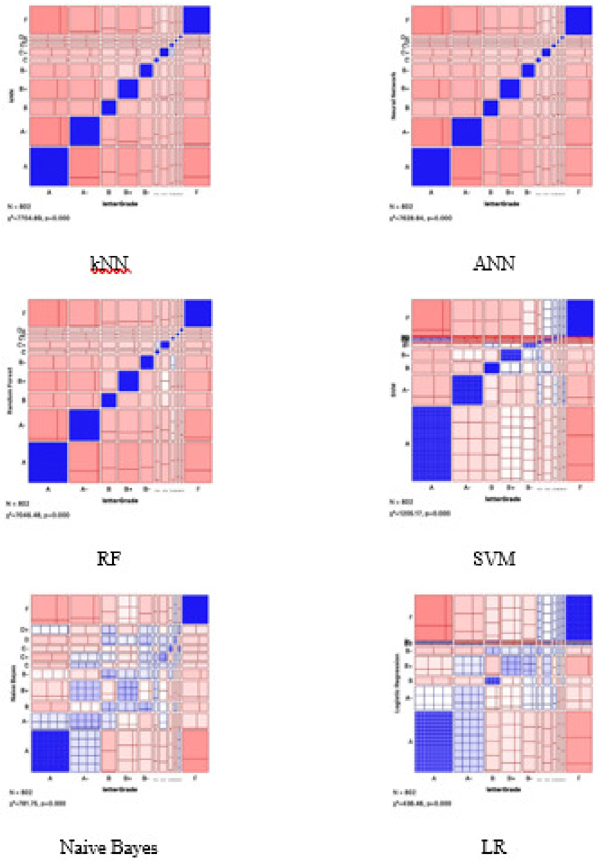

Figure 2: Sieve diagram for ML models

5 Conclusion

In this study, demographic and course information obtained from students' LMS course activities and student information system were used as data set. By using this data set, students' online course success was classified with different ML models. Six different ML models (NN, ANN, RF, SVM, Naive Bayes and LR) were used in the classification of students' course achievements. KNN is the ML model that has the highest accuracy (98.5 percent) in the classification of student achievements. Although ANN (97.4 percent) and RF (94.8) performed with lower accuracy than the kNN model, they gave successful results in classifying students' success. Compared with SVM (50.9 percent), Naive Bayes (36.3 percent) and LR (35 percent) other ML models, the accuracy was very low in classification according to accuracy. In the literature, it has been observed that the accuracy criterion does not always give good results in evaluating the performance of an ML model (Huang & Ling, 2005). Therefore, it would not be correct to evaluate the success of an ML model just by looking at the accuracy criterion. In the literature, it has been suggested to use the AUC criterion, which expresses the area under the ROC-Curve, to evaluate the performance of ML models (Wu & Flach, 2005). When the models were evaluated with the AUC criteria, they were classified with success rates of kNN: 99.8 percent, ANN: 99.7 percent, RF: 99 percent, SVM: 89.6 percent, Naive Bayes: 77.9 percent, LR: 76.4 percent. When the ML models used in the study were evaluated with AUC criteria, it was observed that the classification performance among the models did not change. However, we can say that SVM, NaiveBayes and LR models, which show lower performance according to the accuracy criterion, perform at an acceptable level when evaluated according to AUC criteria. When the ML models used in this study are compared with the national studies in the literature (Table 1), it is concluded that kNN and

ANN show higher performance in the classification of students' success in terms of accuracy and AUC criteria. The data set used in the study is a very general and accessible data set. For this reason, the method applied in the study will be beneficial for educational institutions to be able to observe the performance status of students individually or collectively, to be able to plan for learning deficiencies, to be able to evaluate students.

References

Akçapınar, G. (2014). A Data Mining Approach To Students' Academic Performance Modeling In Online Learning Environment Based On Their Interaction Data, Doctoral dissertation (in Turkish).

Alpaydin, E. (2020). Introduction to machine learning. MIT press.

Altun, M., Kayıkçı, K., & Irmak, S. (2019). Estimation of Graduation Grades of Primary Education Students by Using Regression Analysis and Artificial Neural Networks. *e-International Journal of Educational Research, 10*(3), 29-43 (in Turkish).

Aluko, R.O., Adenuga, O.A., Kukoyi, P.O., Soyingbe, A.A., & Oyedeji, J.O. (2016). Predicting the academic success of architecture students by pre-enrolment requirement: using machine-learning techniques. *Construction Economics and Building, 16*(4), 86.

Aydemir, E. (2019). Forecasting of The Course Learning Notes by Data Mining Methods. *European Journal of Science and Technology,* 15, 70-76 (in Turkish).

Aydın, S., & Özkul, A. E. (2015). Data Mining and an Application in Anadolu University Open Education System (in Turkish). *Journal of Research in Education and Teaching, 4*(3), 36-44.

Aydoğan, İ., & Zırhlıoğlu, G. (2018). Estimation of Student Successes By Artificial Neural Networks (in Turkish). *Van Yuzuncu Yil University Journal of Education, 15*(1), 577-610.

Aydoğan, M., & Karcı, A. (2018) Analysis of Vocational School Students' Success Performance Using Machine Learning Methods. 2nd *International Symposium on Multidisciplinary Studies and Innovative Technologies* (in Turkish).

Corrigan, O., & Smeaton, A.F. (2017, September). A course agnostic approach to predicting student success from VLE log data using recurrent neural networks. In *European Conference on Technology Enhanced Learning* (pp. 545-548). Springer, Cham.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Duygu, B., & Arıcı, N. (2019) Classification of Students' Academic Performance Data with Decision Tree Algorithms. (*3nd International Symposium on Innovative Approaches in Scientific Studies* (in Turkish).

Ergezer, H., Dikmen, M., & Özdemir, E. (2003). Artificial neural networks and recognition systems. *PiVOLKA, 2*(6), 14-17 (in Turkish).

Gök, M. (2017). Predicting Academic Achievement with Machine Learning Methods. *Part C: Design and Technology, 5*(3), 139-148 (in Turkish).

Güldal, H., & Çakıcı, Y. (2017). Analysis of Course Management System Software Users' Interactions Using Classification Algorithms. *Journal of Graduate School of Social Sciences, 21*(4), 1355-1367 (in Turkish).

Güner, N., & Çomak, E. (2011). Predicting Performance of First Year Engineering Students in Calculus by Using Support Vector Machines. *Pamukkale University Journal of Engineering Sciences, 17*(2), 87-96 (in Turkish).

Huang, J., & Ling, C.X. (2005). Using AUC and accuracy in evaluating learning algorithms. I*EEE Transactions on knowledge and Data Engineering, 17*(3), 299-310.

Karabulut, E., & Alpar, R. (2011). *Logistic Regression, Applied Multivariate Statistical Methods.* Detay Publishing Ankara.

Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artificial Intelligence Review, 37*(4), 331-344.

Murat, G. Ö. K. (2017). Prediction of Academic Success with Machine Learning Methods. *Gazi University Journal of Science Part C: Design and Technology, 5*(3), 139-148.

Nizam, H., & Akın, S.S. (2014). Comparison of the performances of balanced and unbalanced data sets in emotion analysis with machine learning in social media. XIX. Internet Conference of Turkey (in Turkish).

Özkan, Y., Koçoğlu, F. Ö., & Erol, Ç. S. (2018). Prediction of Student Performance By Deep Learning Algorithm. Preface of The Editors, 136 (in Turkish).

Özyer, Y. (2016). Teaching different tasks to robots with reinforcement learning. Doctoral dissertation (in Turkish).

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217-222.

Quadri, M.M., & Kalyankar, N.V. (2010). Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology.*

Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 3(22), 41-46.

Sasakawa, T., Hu, J., & Hirasawa, K. (2008). A brainlike learning system with supervised, unsupervised, and reinforcement learning. *Electrical Engineering in Japan, 162*(1), 32-39.

Şengür, D., & Tekin, A. (2013). Prediction of Student's Grade Point Average by Using the Data Mining Methods. *International Journal of Informatics Technologies, 6*(3), 7-16 (in Turkish).

Slim, A., Heileman, G.L., Kozlick, J., & Abdallah, C.T. (2014, December). Predicting student success based on prior performance. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, (pp. 410-415). IEEE.

Taşcı, E., & Onan, A. (2016). The Investigation of Performance Effects of K-Nearest Neighbor Algorithm Parameters on Classification. *Academic Informatics* (in Turkish).

Turhan, K., Kurt, B., & Engin, Y.Z. (2013). Estimation of Student Success with Artificial Neural Networks. *Education and Science, 38*(170) (in Turkish).

Wu, S., & Flach, P. (2005, August). A scored AUC metric for classifier evaluation and selection. In *Second Workshop on ROC Analysis in ML*, Bonn, Germany.