ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

Estuscience – Se, 2024, 25 [2] pp. 262-277, DOI: 10.18038/estubtda.1394885

RESEARCH ARTICLE

CLASSIFICATION OF STUDENTS' ACADEMIC SUCCESS USING ENSEMBLE LEARNING AND ATTRIBUTE SELECTION

Derya ÇINAR [1] Sevcan YILMAZ GÜNDÜZ [2,*]

[1] Computer Engineering Department, Engineering Faculty Eskişehir Technical University, Eskişehir, Turkey
cinar1983@gmail.com - 0009-0005-4729-2172

[2] Computer Engineering Department, Engineering Faculty Eskişehir Technical University, Eskişehir, Turkey
sevcany@eskisehir.edu.tr - 0000-0002-1736-9942

Abstract

Students' success in high school plays an important role in shaping their lives, as it also affects their success in university placement. It is very important to be able to predict this situation so that in case of failure, precautions can be taken, and a solution can be produced. If success situations and failure can be predicted, success can be increased and stabilized with encouragement and support. In this study, students' academic performances were tried to be estimated with the datasets prepared with secondary school students in Portugal. The datasets include students' answers about the factors thought to affect their success-failure and their grades. The wide use and efficiency of machine learning algorithms have also affected studies on predicting student success. Different algorithms have been applied using different methods in the datasets and the correct prediction rate was tried to be maximized. Experiments were carried out using the 10-fold cross validation method. Deep learning, multilayer perceptrons, simple logistic regression, decision table, one rule, iterative classifier optimizer, logistic model tree and fuzzy unordered rule induction algorithm have been used to predict the student academic success. These algorithms have been tested with the classical and bagging methods. The experiments also tested the efficiency of the algorithms in predicting student success by selecting features and comparing the results.

Keywords

Machine learning,
Classification,
Students' academic success

Time Scale of Article

Received :23 November 2023
Accepted : 18 April 2024
Online date :28 June 2024

1. INTRODUCTION

Predicting students' grades is important to guide students correctly [1]. To increase students' academic success, it is effective to identify the factors that reduce success, to eliminate these factors, or even if this is not possible, to reduce their effects. However, these factors are not easy to determine. The reason is that there are many factors that affect students' academic performance [2]. Another reason that makes this process difficult is the presence of unmeasurable factors. Student success may depend on socio-economic and physical conditions. It can be measured by factors such as the society in which one lives, family life, school conditions, and financial means. Individual differences such as talent, interest, and intelligence, are factors that affect success [3] and they cannot be measured easily. Economic conditions, the number of siblings, the education levels, and professions of family members are also factors affecting success [4].

It is possible to predict success with the survey answers in the datasets used to predict student success and the grades taken from school systems and to find out which of these data is effective with artificial intelligence methods. A lot of different data about students can be kept through surveys and records. The type and size of these data may vary [5]. Working with big data is more convenient for making inferences and predictions. Machine learning algorithms, which have become popular in many fields recently, can be used to predict student success. Using student data with machine learning algorithms, students' success is tried to be predicted [6]. Depending on the type of dataset and function, machine

*Corresponding Author: sevcany@eskisehir.edu.tr

learning techniques may change. Some of these techniques include clustering, classification, and feature selection. To make the classification results more efficient, in addition to simple methods, compound machine methods that use several methods together can also be used [7]. The dataset on which each algorithm works efficiently varies depending on the content of the set.

In this study, simple machine learning techniques and bagging ensemble learning techniques were used to predict students' academic achievements. In the study, two open-source datasets, prepared with secondary school student data in Portugal, containing socio-economic conditions as well as three grade information from mathematics and Portuguese courses, were used. The tests were carried out with eight different algorithms by using classical and bagging methods. Initially all the features were involved and then some of them were selected. The algorithms used are deep learning (DL), multilayer perceptrons (MLP), simple logistic regression (SL), decision table (DT), one-rule (One-R), iterative classifier optimizer (ICO), logistic model tree (LMT) and fuzzy unordered rule induction algorithms (FURIA). In this study, unlike many other studies, experiments were deepened by selecting features and using ensemble learning algorithms together. Using various feature selection algorithms and comparing the results to identify common features with the highest contribution has significantly enriched the literature in this field. Consequently, the attributes that play the most important role in predicting student success were identified. This information provides valuable insights for future research and practical applications in education. Another important feature that distinguishes it from other studies is its attempt to enhance various classification algorithms using the bagging method and comparing the results of bagging methods with classical classification techniques. This study represents a comprehensive investigation that incorporates both classical classification methods, feature selection techniques, and the bagging method.

2. LITERATURE OVERVIEW

For educational sciences, being able to predict student success classes is an important and necessary choice for the scope of education. There are various studies conducted in this field and some of them will be discussed in this section.

Cortez and Silva [8] compared machine learning techniques with two different datasets containing real-world pieces containing 33 different attributes and three exam scores in commonly seen locations in two different high schools in Portugal. To deepen the study, they have created systems that predict students' academic success by transmitting exam grades. Cortez and Silva [8] carried out their work in the experimental R programming environment. They used the RMiner Library, an open-source library, and DL methods such as random forest (RF), neural networks (NN), support vector machines (SVM), decision trees (DTrees) and naive estimator (NV) training. In their study, it was seen that previous life success situations were highly effective, but other relevant characteristics were also effective. To obtain accurate predictions in their study, 200 values were obtained by performing 10-fold cross validation 20 times for the derivative. Test findings of all data in ten times with 10% of the data. The dataset with the highest prediction success of the system was determined with the experiment including two exam grades, and the most successful development was determined as the NV.

V. Vijayalakshmi and K. Venkatachalapathy [9] trained the model with the R program using different algorithms for student performance prediction and stated that the algorithm gave the most successful results was the deep neural network (DNN) with an accuracy rate of 84%. They used many algorithms such as DTrees, SVM, Naive Bayes (NB), RF, k nearest neighbors (k-NN), DNN in their study. Tosunoğlu et al. [10] conducted a review of articles in the field of educational sciences with machine learning methods. They analyzed 201 articles published between 2015 and 2020. They found that the most used algorithm in the articles was the DTrees algorithm with a rate of 22.5%. Güvenç et al. [11] tried different machine learning methods to predict the academic success of students taking the Introduction to Information Systems Engineering course, but they could not achieve successful results due to the lack of sample dataset of 71 students. Hence, to increase the sample, they increased the dataset to 300 student records with the Synthetic Minority Over-sampling Technique (SMOTE) and increased the correct prediction success of the most successful algorithm from 60% to 97.87% with the k-NN.

In their study, Salameh et al. [12] carried out the 2-stage (pass-fail) classification in the MATLAB program and used the enhanced binary genetic algorithm (EBGA) and binary genetic algorithm (BGA), which are feature selection algorithms. They stated that the NB algorithm and the EBGA feature selection method were the most successful methods with a rate of 87%, using k-NN, DTrees, NB, SVM and the hidden Latent Dirichlet allocation algorithm.

In their study, Quy et al. [13] worked on a 2-stage (pass-fail) classification with Portuguese course notes and tried the Adaboost boosting method as well as the classical methods. As a result, they achieved a success rate of 95%.

Özkan et al. [14] made a 2-stage (pass-fail) classification with the mathematics dataset in their study to find the most effective attribute. As a result of their study in the R program, they stated that the C5.0 algorithm was successful with a rate of 86% among different algorithms. In their study with a 2-stage mathematics course dataset, Başer et al. [15] carried out classification studies with various algorithms in the WEKA program and reached a 92.15% accuracy rate with the One-R algorithm. Çınar and Yılmaz Gündüz [16] worked with a 2-stage mathematics course dataset. In this study, the LMT algorithm gave the best result. Yavuzarslan and Erol [17] tried to predict student success using different algorithms in their study with 10-week log records of 93 students, but since the dataset did not work efficiently, they renewed the study by creating new dataset derivatives with the synthetic minority sample increase method, upsampling and class sampling methods. They achieved over 80% success in their study with these newly created dataset variants.

Kayalı and Buyrukoğlu [18] conducted a total of 24 different experiments with 2 datasets, using comprehensive forward selection, analysis of variance, embedded feature selection methods and DTrees, RF, and exchange of support machines to predict academic success in their country. As a result of the experiments, it was observed that the 84% accuracy rate obtained by using all attributes in the RF algorithm in the first dataset was increased to 86% by selecting attributes with advanced selection techniques. They stated that in the second dataset, the accuracy rate was increased from 82% to 90% by using all the features with the RF algorithm and SVM algorithms.

Bentaleb and Abouchabaka [19] stated in their study with datasets prepared with the notes of 3-grade mathematics and Portuguese courses, the most successful algorithm with an accuracy rate of 86% among DTrees, RF and SVM algorithms is the RF algorithm in the Python program.

Bozkurt Keser and Aghalarova [20] used a new hybrid ensemble learning algorithm (HELA) in their study. In this study, they used 3 different boosting algorithms and different combinations of these algorithms as input to the HELA. In their study, they used two and five stage classification datasets containing secondary school students' grades and other information about mathematics and Portuguese courses in Portugal. According to the results of their study, the HELA is more successful than the simple boosting algorithms.

This study offers a comprehensive analysis that distinguishes itself from other studies by combining classical algorithms, bagging methods, and feature selection. Also, this study demonstrates the contributions of feature selection in the educational dataset to the success of classical and bagging algorithms. In our study, two datasets containing mathematics and Portuguese courses of secondary school students in Portugal, which are frequently used in the literature [8,14,15,20,21,22,23], were studied. Initially, classification was made using all features with eight different algorithms. These algorithms are DL, MLP, SL, ICO, DT, FURIA, One-R, and LMT algorithms. The same classification algorithms were then used along with bagging ensemble learning method. Finally, by using various feature selection algorithms, the most effective features were selected for the datasets used. All experiments were repeated by using these features. At first, the results were compared according to classification algorithms, ensemble learning methods and feature selection. Then, the results were compared with studies using the same dataset in the literature.

3. METHODS

In this section, the main classification algorithms used will be explained one by one with their general features. Then, the feature selection algorithms used in the experiments will be discussed. Finally, the bagging ensemble learning algorithm will be explained.

3.1. Classification Algorithms

In this section, the classification algorithms used in this study will be explained.

3.1.1. Deep learning

DL is an artificial intelligence and machine learning system that learns by imitating the way humans learn information. It is a data science that makes modeling using statistics and making predictions [24]. Most of the data studied in the DL system is structured and labeled. This is because it requires outputs as well as inputs to train the system [25].

DeepLearning4J library was used in this study. DeepLearning4J is a DL algorithm library used for classification and regression with MLP [26].

3.1.2. Multiplayer perceptron

MLP is an important basis of artificial NN [27]. It is one of the supervised learning algorithms. Both input and output data are given to the system and the model is created with the learning algorithm [28]. In the model, there are hidden layers between the input and output layers. It transmits the data coming from the input layer to the intermediate layers. The number of intermediate layers varies depending on the problem. Data from one layer is transmitted to the next intermediate layer. The number of outputs in the system is the number of objects in the output layer [29]. After training, the learned model is used to test new data.

3.1.3. Simple logistic regression

SL is a statistical method used to predict a binary outcome in a dataset. It is aimed to find a value for the dependent variable by analysing the links between the independent variables in the dataset. A definitive judgment is made with this value [30]. More than one input variable can be used in SL. By comparing the values read from these variables with previous results, the result is produced by calculating the probabilities for new inputs [31].

3.1.4. Iterative classifier optimizer

ICO algorithms use predictions obtained from neighboring nodes for each step [32]. In this algorithm, the errors found after classifying the first record in the dataset are used as feedback and the inputs are changed in this way. In this way, all outputs become improved inputs of the next process. This structure can be described as a brain neural network. The system is trained by repeating the records in the network [33]. This classifier selects the best number of repetitions for an algorithm that performs cross-validation [34].

3.1.5. Decision table

DT is the decision algorithm used for cases with more than one condition. The table contains conditions in rows and columns and actions in the intersection places. It is a rule-logic scheme programmed in a table form. New rules can be added by adding a line or column. One action can be reached with more than one condition or can be directed with more than one action. According to the actions, the algorithm trained and makes classification [35, 36]. DT is simple and can be easily interpreted. It is easy to improve. Effective derivatives can be created, and the application area is wide to test. Even very complex data can easily turn into a DTrees [37].

3.1.6. Fuzzy unordered rule induction algorithm

FURIA uses fuzzy rule or rule stretching approach. It is used in classifications that contain non-sharp boundaries, that is, if there are different classes between two classes that do not have clear boundaries. It is appropriate to use it for classifications such as good-fair-bad, very good-good- satisfactory -pass-fail in a data set. It is more advantageous than strict classification rules. It works efficiently when there are no sharp boundaries between classes, that is, in classifications with smooth boundaries, which is the basis of fuzzy logic. During classification, the results are converted into sharp boundaries [38]. In this study, FURIA was used in the WEKA program as the FURIA.

3.1.7. One rule

One-R works like a single-stage DTrees. It has a simple working structure. It tries to select the feature that best predicts the target class and infers rules from error data. That is, it takes the best of the discriminative featuress, the one that gives the least erroneous result, and sets rules accordingly [39]. Considering the datasets, it takes the feature that gives the least error in the result classification and creates a rule between the result and the feature. It classifies all records according to this rule. It produces rules such as " Successful if absent less than 10 days " [40]. It can be evaluated as a single-level DTrees.

3.1.8. Logistic model tree

LMT algorithm is an algorithm created by combining DTrees induction and logistic regression algorithms [41]. It is a supervised classification model. This method is based on the tree structure previously found as a model. A piecewise linear regression model is provided from DTrees with constants in their leaves [42]. In the logistics part, a model is created at each node using the logitboost algorithm. The knot is broken using certain methods. The beginning of each stage is based on the result at the parent node. In the final stage, the tree is pruned. Classification is performed according to this tree model.

3.2. Feature Selection Algorithms

The second stage of the experiments was carried out by determining the most effective features. Feature selection was made using the feature selection algorithm in the WEKA program. In this algorithm, the entire set was used for training purposes. Correlation-based feature subset selection, gain ratio attribute evaluation, information gain attribute evaluation, correlation attribute evaluation, relief attribute evaluation, symmetrical unsert attribute evaluation, and One-R attribute evaluation algorihms were used in this study [43]. Best first, ranker and greedy stepwise are search methods for these evaluation algorithms. While some methods work with every algorithm, some of them work with certain algorithms. Feature selection was made by trying all of them to find suitable algorithms. The features found for each algorithm were scored and ranked, and the most effective ones were selected. Different features were found to be efficient for each dataset and classification group, and experiments were repeated with these features.

Although choosing attributes may seem ineffective in some experiments, it affects the completion time of the experiments and shows which attribute is effective for each course [44, 45].

3.3. Bagging Ensemble Learning Method

Bagging algorithm is an ensemble learning method. While operations are performed with a single algorithm in classical machine learning methods, algorithms can be applied together or sequentially in the bagging method. Bagging is one of the heterogeneous ensemble learning models that combine training data from data separated into training and testing by selecting different parameters in a single algorithm or using different learning algorithms [46]. It combines different algorithms while training the training set with different samples and with this method many results are obtained. The majority of these classification results are taken as output [47].

4. DATASET

In this study, two datasets containing similar data were studied, including mathematics and Portuguese courses of secondary school students in Portugal [8]. The datasets include students' answers to survey questions that meet the criteria which affect their success, their grades, and a success class. The first dataset (Set 1) contains student information, mathematics course grades and success class, and the second dataset (Set 2) contains student information, Portuguese course grades and success class.

Set 1 includes information of a total of 395 students (208 women and 187 men) for the mathematics course and Set 2 includes information of a total of 649 students (383 women and 266 men) for the Portuguese course. There are 33 attributes in the datasets. Further details are presented in [48].

The datasets include information such as family information, student time use, success, or failure in previous years, as well as two exam grades with evaluation scores between 0 and 20 and the final exam grade. The final exam was used to predict student success. In Portugal, class success is determined by the final exam grade.

Table 1. Binary classification system [8]

Country	Pass	Fail
Portugal	10-20	0-9

The binary success criterion based on the exam score was created as shown in Table 1 and the 5-class success criterion based on the exam score was created as shown in Table 2, similar to other studies [8].

Table 2. 5-class classification system [8]

Country	Excellent	Good	Satisfactory	Sufficient	Fail
Portugal	16-20	14-15	12-13	10-11	0-9

5. EXPERIMENTS

In this study, the results were compared using eight different algorithms for mathematics and Portuguese courses datasets, and it was aimed to determine which algorithm would be more efficient in datasets containing such student information and to predict student success levels at the highest rate. The experiments were conducted in the WEKA program, utilizing the 10-fold cross-validation method on the datasets. The computer employed for this purpose featured an i7 processor, 8 GB RAM, and a 64-bit operating system. Version 3.8.6 of the WEKA program was utilized. Classical methods, along with 8 different algorithms as described previously, were applied to both variants of the experimental sets. Additionally, bagging ensemble learning method has also been tried with the same algorithms. The algorithms used in the experiments were applied to these two datasets, and the most efficient features were selected and used again to improve the results. Feature selection was made through the WEKA program and the most effective features were tried to be selected.

5.1. Results of Feature Selection Algorithms

In this study, different feature selection algorithms have been utilized. The obtained results were aggregated to enhance the model's performance and attain a more robust outcome. We can examine the process of finding common features identified by different feature selection algorithms in four steps. In the first step, feature selection algorithms were applied to the given dataset, and results were obtained. In the second step, the outcomes of the feature selection algorithms were assessed. At this stage, the top 7-8 features with the highest gain from each feature selection algorithm's ranking were recorded. For some datasets, the correlation-based feature subset selection algorithm could only select five features. In the third step, common features obtained from the applied methods were identified. Common features are those consistently chosen under different algorithms, enhancing the model's performance. In the final step, classification algorithms were trained with the common obtained features, and the results were evaluated. The features obtained using the feature determination method employed in this study are listed in Tables 3 and 4 for Mathematics and Portuguese language courses, respectively.

Table 3. Results of Feature Selection Algorithms for Mathematics Dataset

Feature Selection Algorithm	Selected Attributes (binary classification dataset)	Selected Attributes (5 class classification dataset)
Correlation-based feature subset selection	G2, G1, failures, goout, higher, Mjob, guardian	Sex, failure, G1, G2
Gain ratio attribute evaluation	G2, G1, failures, higher, goout, schoolsup, guardian	G2, G1, failure, higher, schoolsup, Mjob, address, Fjob
Information gain attribute evaluation	G2, G1, failures, goout, higher, Mjob, Guardian	G2, G1, failure, Mjob, schoolsup, Fjob, higher
Correlation attribute evaluation	G2, G1, failures, goout, age, higher, Medu, Fedu	G2, G1, failures, medu, goout, age, higher
Relief attribute evaluation	G2, G1, failures, sex, Mjob, paid, schoolsup	G2, G1, Mjob, sex, failures, medu, goout
Symmetrical unsert attribute evaluation	G2, G1, failures, goout, higher, schoolsup, guardian	G2, G1, failures, higher, gout, Mjob, Fjob
One-R	G2, G1, failures, higher, sex, studytime, Fjob, reason	G2, G1, failures, absences, goout, paid, Fjob
Common attributes used in this paper	G2, G1, failures, higher, goout	G2, G1, failures, Mjob, goout

Table 4. Results of Feature Selection Algorithms for Portuguese Course Dataset

Feature Selection Algorithm	Selected Attributes (binary classification)	Selected Attributes (5 class classification)
Correlation-based feature subset selection	G1, G2, failures, higher, school	studytime, failures, schoolsup, paid, activities, internet, G1, G2
Gain ratio attribute evaluation	G1, G2, failures, higher, school, Fedu, studytime	G2, G1, failures, higher, school, medu, schoolsup, studytime
Information gain attribute evaluation	G1, G2, failures, higher, school, Fedu, reason	G2, G1, failures, higher, school, Mjob, Medu, studytime
Correlation attribute evaluation	G1, G2, failures, higher, school, studytime, Fedu	G2, G1, failures, higher, medu, school, studytime, paid
Relief attribute evaluation	School, G1, G2, failures, higher, reason, address, sex	G2, G1, school, higher, sex, failures, medu, paid, studytime
Symmetrical unsert attribute evaluation	G2, G1, failures, higher, school, Fedu, studytime	G2, G1, failures, higher, school, medu, studytime, mjob, reason, fedu, paid
One-R	G2, G1, failure, age, Fjob, guardian, traveltime	G2, G1, Medu, Fedu, traveltime, paid, failures, studytime
Common attributes used in this paper	G1, G2, failures, higher, school	G1, G2, failures, studytime, paid

5.2. Mathematics Dataset Results

In our first test, the mathematics course dataset was used. A binary and 5-class classification was made in the dataset created by subtracting the student's final grade. First, the accuracy rate, precision, sensitivity/recall, and the F-1 value were calculated with the selected classification algorithms using all the features. Then, the experiments were repeated using the bagging for all algorithms. Later, various feature selection algorithms in the WEKA program were used. The common features in the results of these algorithms were selected. Finally, all experiments were repeated using these selected features. Binary classification experiment results for the mathematics dataset are given in Table 5, and 5-class classification results are given in Table 6.

Table 5. Mathematics Course Dataset Binary Classification Results

Methods	Algorithms	All Features				Selected Features			
		Acc. Rate %	Prec.	Sens.	F-1	Acc. Rate %	Prec.	Sens.	F-1
Classic Methods	**DL**	87,595	0,899	0,876	0,889	89,114	0,906	0,891	0,893
	MLP	87,342	0,873	0,873	0,873	88,101	0,880	0,881	0,880
	SL	92,152	0,922	0,922	0,922	91,899	0,924	0,919	0,920
	ICO	91,646	0,920	0,916	0,917	91,139	0,914	0,911	0,912
	DT	90,886	0,913	0,909	0,910	91,392	0,919	0,914	0,915
	FURIA	87,848	0,878	0,878	0,878	91,392	0,916	0,914	0,915
	One-R	91,899	0,924	0,919	0,920	91,899	0,924	0,919	0,920
	LMT	**92,658**	**0,927**	**0,927**	**0,927**	91,646	0,921	0,916	0,918
Bagging Methods	**DL**	86,835	0,897	0,868	0,872	88,101	0,897	0,881	0,884
	MLP	88,354	0,883	0,884	0,883	89,620	0,896	0,896	0,896
	SL	92,152	0,921	0,922	0,921	90,633	0,908	0,906	0,907
	ICO	91,392	0,918	0,914	0,915	91,392	0,918	0,914	0,915
	DT	91,899	0,922	0,919	0,920	91,646	0,921	0,916	0,918
	FURIA	91,899	0,920	0,919	0,919	91,899	0,920	0,919	0,919
	One-R	91,899	0,924	0,919	0,920	91,899	0,924	0,919	0,920
	LMT	91,646	0,916	0,916	0,916	90,127	0,901	0,901	0,901

According to Table 5, binary classification resulted in a high accuracy rate. The main reason for this is that the number of predicted classes is small and the boundary determining success is sharper than the 5-class classification. The best result obtained in this study is the LMT algorithm using classical methods and all features. The fact that there are only two classes to be predicted and the score difference is large is effective in achieving the 92.66% accuracy rate with the LMT. The influence of the features helped to sharply distinguish the branches in the tree structure. The selected attributes mentioned above were effective on these tree branches. Although similar values were obtained in the experimental results, the small number of classes enabled simple classical learning algorithms to be more successful than bagging algorithms.

The most important features in the mathematics dataset for binary classification are selected as follows: G2 (grade 2), G1 (grade 1), failures, mother's occupation, desire to pursue higher education, and going out. When examining the selected features in this experiment, it is evident that academic success indicators such as midterm grades and past failures, along with attributes influencing study habits such as aspirations for higher education and time spent going out, are effective in classification. When examining the experimental results with classical methods, we observed that reducing the number of attributes increased the success of DL, MLP, DT, and FURIA algorithms. However, there was a decrease in the success of SL, ICO, and LMT algorithms. The success of the One-R algorithm remained constant. Considering the bagging method, while DL and MLP gave more successful results with fewer attributes, the success rate with LMT and DT decreased slightly, and the success did not change in other algorithms. The reason for the increase in success in DL and MLP is that when the number of features decreases, the parameters that need to be learned in the model decrease and as a result, learning becomes easier.

Table 6. Mathematics Course Dataset 5-class Classification Results

Methods	Algorithms	All Features				Selected Features			
		Acc. Rate %	Prec.	Sens.	F-1	Acc. Rate %	Prec.	Sens.	F-1
Classic Methods	DL	52,658	0,517	0,527	0,505	53,165	0,513	0,532	0,474
	MLP	53,165	0,523	0,532	0,526	71,392	0,714	0,714	0,714
	SL	71,646	0,716	0,716	0,715	72,658	0,727	0,727	0,725
	ICO	77,975	0,785	0,780	0,777	76,962	0,773	0,770	0,768
	DT	77,722	0,782	0,777	0,775	**78,481**	**0,791**	**0,785**	**0,782**
	FURIA	75,190	0,753	0,752	0,752	77,468	0,778	0,775	0,772
	One-R	**78,481**	**0,791**	**0,785**	**0,782**	**78,481**	**0,791**	**0,785**	**0,782**
	LMT	71,646	0,716	0,716	0,715	71,899	0,721	0,719	0,719
Bagging Methods	DL	50,380	0,772	0,772	0,771	55,443	0,548	0,554	0,503
	MLP	55,696	0,541	0,557	0,547	73,418	0,736	0,734	0,735
	SL	71,899	0,721	0,719	0,718	73,924	0,743	0,739	0,739
	ICO	77,215	0,772	0,772	0,771	76,203	0,764	0,762	0,760
	DT	77,975	0,784	0,780	0,777	78,228	0,788	0,782	0,780
	FURIA	77,215	0,772	0,772	0,772	77,468	0,777	0,775	0,774
	One-R	**78,481**	**0,791**	**0,785**	**0,782**	**78,481**	**0,791**	**0,785**	**0,782**
	LMT	74,177	0,746	0,742	0,741	70,127	0,698	0,701	0,699

As seen in Table 6, the One-R algorithm was the best classifier algorithm with an accuracy rate of 78.481% for this dataset. DT reaches the same rate by selecting attributes using the classical method. With feature selection, the complexity of the model was reduced, and results were achieved faster [44, 45]. The selected features for this experiment are G2, G1, failures, mjob, and goout. One of the most significant factors affecting students' study discipline, the time spent going out, is inversely proportional to success. The prediction speed of the system increased when fewer features were used. Factors affecting coursework and midterm exam grades also affect mathematics course success. Again, the mother's profession is effective in academic success since it affects study habits. When the number of attributes decreased in classical methods, the success of ICO decreased, the success of One-R remained constant, and the success of all other algorithms increased. The increase or decrease in the success of algorithms in feature selection with the bagging method is similar to classical methods. However, the difference lies in the fact that when bagging was applied to the LMT algorithm specifically, there was a decrease in success in feature selection.

In experiments conducted with this dataset, binary classification in two different classification schemes has a higher accuracy rate than 5-class classification because the separation between classes is greater. Making five different class predictions in the same score range (0-20) is more difficult than making two different class predictions. Since the correlations between achievement classes in mathematics are less than in other lectures, the correct class prediction is lower.

Relevant studies using the same dataset and the results of these studies are given in Table 7. The success values of the relevant studies in Tables 7 and 10 have been directly taken from articles implementing the method. Although some studies may use the same algorithms, the way experiments are conducted (percentage split or cross-validation) and parameter adjustments can vary. This has also led to different outcomes in terms of success metrics. As seen in Table 7, two different classification levels were used to predict mathematics course success. In the binary classification, Bozkurt Keser and Aghalarova [20] obtained the best predictions with an accuracy rate of 96.6% with the HELA method. In the 5-class classification, this study has achieved the second-best ranking, with an accuracy rate of 78.5% with the One-R and DT algorithms.

Table 7. Studies Related on Mathematics Course Dataset

Used Dataset	Paper	Motivation	Used Program	Used Algorithms	Results (AC)
Maths, Binary	Özkan et al. [14]	Classify and find the effective attributes	R	C5.0, DL, RT, BoostedC5, LR, SVM, RF	86 % C5.0
Maths, Binary	Başer et al. [15]	Classification	WEKA	ICO, One-R, LogitBoost, YSA,	92.2 % One-R
Maths, Binary	Ünal [21]	Classification	WEKA	NB, RF, DT	93.67 % RF
Maths, Binary	Bozkurt Keser & Aghalarova [20]	Classification	Python	HELA, XGboost, LightGBM, GB,	96,6 % HELA
Maths, Binary	Cortez & Silva[8]	Classification	R	NV, NN, RF, SVM, DT,	78,5 % NV
Maths, Binary	Our study	Classification	WEKA	DL4MLP, MLP, SL, ICO, One-R FURIA, DT	92.7 % LMT
Maths, 5-class	Kızılkaya & Oğuzlar [22]	Between gender and success	-	C4.5, SVM, FFANN, KELM	KELM
Maths, 5-class	Felicia & Ferren [23]	Classification	RapidMiner	GLM, RF, NB	48.6 % NB
Maths, 5-class	Bozkurt Keser & Aghalarova [20]	Classification	Python	HELA, XGboost, LightGBM, GB,	78.2 % HELA
Maths, 5-class	Ünal [21]	Classification	WEKA	NB, RF, DT	79.5 % J48
Maths, 5-class	Our study	Classification	WEKA	DL4MLP, MLP, SL, ICO, One-R FURIA, DT	78,5 % One-R, DT

5.3. Portuguese Course Data Set Results

In this section, binary and 5-class classification was conducted with this dataset containing grades from the Portuguese course. The experiments performed on the mathematics data set were repeated for this data set as well. Since attributes such as student grades are different in this data set, the selected attributes have also changed because of the feature selection. The experimental results for the binary classification are given in Table 8, and the five-class experiment results are given in Table 9.

Table 8. Portegues Lesson Dataset Binary Classification Results

Methods	Algorithms	All Features				Selected Features			
		Acc. Rate %	Prec.	Sens.	F-1	Acc. Rate %	Prec.	Sens.	F-1
Classic Methods	**DL**	90,909	0,910	0,909	0,910	91,834	0,919	0,918	0,919
	MLP	89,676	0,894	0,897	0,895	93,374	0,932	0,934	0,932
	SL	93,220	0,929	0,932	0,929	93,220	0,929	0,932	0,930
	ICO	93,374	0,932	0,934	0,932	92,912	0,927	0,929	0,927
	DT	92,604	0,923	0,926	0,922	93,220	0,930	0,932	0,928
	FURIA	93,220	0,929	0,932	0,929	92,296	0,919	0,923	0,919
	One-R	**93,683**	**0,935**	**0,937**	**0,933**	**93,683**	**0,935**	**0,937**	**0,933**
	LMT	**93,683**	**0,935**	**0,937**	**0,933**	93,220	0,929	0,932	0,930
Bagging Methods	**DL**	91,063	0,909	0,911	0,910	92,142	0,920	0,921	0,921
	MLP	91,063	0,905	0,911	0,906	92,912	0,927	0,929	0,927
	SL	92,296	0,920	0,923	0,921	93,529	0,933	0,935	0,933
	ICO	**93,683**	**0,935**	**0,937**	**0,935**	93,374	0,932	0,934	0,933
	DT	93,220	0,930	0,932	0,928	92,604	0,922	0,926	0,922
	FURIA	93,220	0,930	0,932	0,930	93,220	0,930	0,932	0,931
	One-R	**93,683**	**0,935**	**0,937**	**0,933**	**93,683**	**0,935**	**0,937**	**0,933**
	LMT	92,758	0,925	0,928	0,926	92,296	0,920	0,923	0,921

In the binary classification, predictions made for the Portuguese course and the accuracy rates of the algorithm results were found to be close to each other and at high values. The students' success in the Portuguese course is related to environmental factors. In this experiment, the selected features are G1 (grade 1), G2 (grade 2), failures, higher education (higher), and school. Upon examination of these attributes, it is observed that the factors influencing Portuguese course grades generally include a student's past academic performance and their future aspirations. The selection of the school feature is related to the social structure of schools, whether they are in rural or urban areas, the family structures of students attending the school, and the academic performance levels of students, all of which contribute to the formation of school culture. School culture, in turn, influences students' everyday language usage and even the language of instruction, thereby indirectly affecting the success of language courses. When feature selection is applied to the dataset, it is observed that the success of DL, MLP, and DT algorithms increases, while the success of SL and One-R algorithms remains the same, and the success of ICO, LMT, and FURIA decreases for classical algorithms. When the effect of feature selection is examined in the bagging method, it is observed that for DL, MLP, and SL algorithms, the success has increased, remained the same for FURIA and One-R, and decreased for ICO, DT, and LMT.

In 5-class classification, higher success was achieved compared to the mathematics course data set. In this dataset, the One-R algorithm gave the highest result. In the classical and bagging methods, the single rule algorithm reached the highest accuracy rate of 76.733% with both all attributes and selected attributes. When we look at the result of the feature selection algorithm for the Portuguese course dataset, the student’s study time and taking private lessons, their grades, and past failures, which are also effective in other experiments, are effective attributes.

Table 9. Portegues Lesson Data Set 5-Class Classification Results

Methods	Algorithms	All Features				Selected Features			
		Acc. Rate %	Prec.	Sens.	F-1	Acc. Rate %	Prec.	Sens.	F-1
Classic Methods	**DL**	52,851	0,528	0,529	0,526	60,247	0,580	0,602	0,577
	MLP	57,165	0,573	0,572	0,572	73,190	0,731	0,732	0,729
	SL	71,341	0,715	0,713	0,713	72,265	0,724	0,723	0,720
	ICO	74,114	0,740	0,741	0,733	75,655	0,758	0,757	0,752
	DT	74,114	0,740	0,741	0,733	76,425	0,777	0,764	0,760
	FURIA	71,341	0,711	0,713	0,708	74,422	0,745	0,744	0,737
	One-R	**76,733**	**0,782**	**0,767**	**0,764**	**76,733**	**0,782**	**0,767**	**0,764**
	LMT	71,649	0,718	0,716	0,716	72,265	0,724	0,723	0,720
Bagging Methods	**DL**	52,697	0,525	0,527	0,522	60,555	0,593	0,606	0,584
	MLP	58,089	0,579	0,581	0,578	72,727	0,730	0,727	0,725
	SL	72,111	0,722	0,721	0,720	72,265	0,723	0,723	0,721
	ICO	75,963	0,760	0,760	0,756	74,576	0,745	0,746	0,741
	DT	76,425	0,776	0,764	0,760	76,425	0,777	0,764	0,760
	FURIA	71,341	0,715	0,713	0,710	75,193	0,760	0,752	0,748
	One-R	**76,733**	**0,782**	**0,767**	**0,764**	**76,733**	**0,782**	**0,767**	**0,764**
	LMT	71,957	0,720	0,720	0,719	68,875	0,689	0,689	0,688

In binary classification, while socio-cultural environment-related attributes are effective in similar experiments, in this experiment, attributes indicating the student's study habits have proven to be effective in determining achievement levels. In classical methods, while the success remains the same for the One-R algorithm in feature selection, the success has increased with fewer attributes for all other methods. However, in the bagging method, feature selection has resulted in a decrease for ICO and LMT, remained the same for DT and One-R, and increased for all other methods.

Studies on the Portuguese and their comparative results are shown in Table 10. As seen in Table 10, although the results are close to each other in experiments conducted with different algorithms and different programs, the results vary in binary and 5-class classification. According to the experiment results, two different classification levels have been used to predict success in Portuguese course. In the 5-class classification, Bozkurt Keser and Aghalarova [20] achieved the highest accuracy of 78.5% using the HELA method, while in the 5- class classification, this study obtained the best predictions with a 93.7% accuracy rate using the One-R, ICO, and LMT algorithms.

Table 10. Studies Related on Portegues Course Data Set

Used Dataset	Paper	Motivation	Used Program	Used Algorithms	Results (AC)
Portugues, Binary	Ünal [21]	Classification	WEKA	NB, RF, DT	93.2 % RF
Portugues, Binary	Bozkurt Keser & Aghalarova [20]	Classification	Python	HELA, XGboost, LightGBM, GB	91,3 % HELA
Portugues, Binary	Cortez & Silva [8]	Classification	R	NV, NN, RF, SVM, DT	93,0 % DT
Portugues, Binary	This study	Classification	WEKA	DL4MLP, MLP, SL, ICO, One-R FURIA, DT	93.7 % One-R, ICO, LMT
Portugues, 5-class	Ünal [21]	Classification	WEKA	NB, RF, DT	77.2 % RF
Portugues, 5-class	Bozkurt Keser & Aghalarova [20]	Classification	Python	HELA, XGboost, LightGBM, GB,	78.5% HELA
Portugues, 5-class	Cortez & Silva [8]	Classification	R	NV, NN, RF, SVM, DT,	76,1 % DT
Portugues, 5-class	This study	Classification	WEKA	DL4MLP, MLP, SL, ICO, One-R FURIA, DT	76.5 % One-R

6. CONCLUSION

In this study, it was aimed to predict the student's mathematics and Portuguese language achievement in binary and 5-class classification, and experiments were carried out for this purpose. When we look at the results in general, predicting the binary classification is more successful than the 5-class classification. As the number of classes decreases, the prediction rate increases. Students' information in the datasets is more effective in determining pass-fail situations. It is more difficult to process student data and identify classes with very little difference in scores between good and very good.

The content of the courses varies in terms of making predictions. Since the effective attributes for the mathematics course and the Portuguese course are different, the successful algorithms and the accuracy rates they find are also different. The reason for this is that the factors affecting the Portuguese course are the social environment, living space and cultural environment of the student, while for the mathematics course, more academic factors such as the socio-economic level of the family, the professions of the family members, the student's desire and preparation for higher education are effective. A student’s success is based on success in previous years and family factors. The reason for this depends on the continuity of student success, the stability of success levels throughout the semester, and their educational perspectives and study habits come from the social environment they live in. Considering all these situations, student success depends on many criteria of families.

In this study, by comparing experiments conducted with different algorithms and determining effective attributes, it is aimed to predict student success with the highest accuracy and take the necessary precautions in advance, as well as to be able to make the necessary guidance by obtaining information about the criteria that affect student success.

ACKNOWLEDGEMENTS

This study was supported by Eskişehir Technical University within the scope of project number 21LÖT098.

CONFLICT OF INTEREST

The authors stated that there are no conflicts of interest regarding the publication of this article.

CRediT AUTHOR STATEMENT

Derya Çınar: Writing – Original Draft, Software, Conceptualization, Investigation, Validation, Formal analysis. **Sevcan Yılmaz Gündüz:** Supervision, Writing – Review & Editing, Conceptualization, Methodology.

REFERENCES

[1] Rao K, Rao M, Ramesh B. Predicting learning behavior of students using classification techniques. International Journal of Computer Applications, 2016; 139(7).

[2] Al-Radaideh QA, Al-Shawakfa E, Al-Najjar M. Mining student data using decision trees. In: International Arab Conference on Information Technology; January 2006; Yarmouk University, Jordan.

[3] Şama E, Tarım K. Teachers' attitudes and behaviors towards students perceived as unsuccessful. Türk Eğitim Bilimleri Dergisi, Kış 2007; 5(1):135-154.

[4] Sezer Ö. Some demographic characteristics of the repeating students and the opinions of the students and the teachers about repetition. (article in Turkish, summary in English). İnönü Üniversitesi Eğitim Fakültesi Dergisi, Güz 2007; 8(14): 31–48.

[5] Yadav SK, Pal S. Data Mining: A prediction for performance improvement of engineering students using classification. World of Computer Science and Information Technology Journal, 2012; 2(2): 51-56.

[6] Gadhavi M, Patel D. Student final grade prediction based on linear regression. Indian Journal of Computer Science and Engineering, 2017; 8(3): 274-279.

[7] Erdoğan Z, Namlı Ö, Akarsu C. Öğrenci başarısının bileşik makine öğrenme teknikleri kullanılarak tahmini. In: International Symposium on Industry 4.0 and Applications, October 2017; Karabük, Turkey. pp.150-155.

[8] Cortez P, Silva A. Using data mining to predict secondary school student performance. In: 5th Annual Future Business Technology Conference; 2008; Porto: pp. 5-12.

[9] Vijayalakshmi V, Venkatachalapathy K. Comparison of predicting student's performance using machine learning algorithms. Intelligent Systems and Applications, 2019; 11(12): 34-45.

[10] Tosunoğlu E, Yılmaz, R, Özeren E, Sağlam Z. Machine learning in education: a study on current trends in researchs. (article in Turkish, summary in English). Journal of Ahmet Keleşoğlu Education Faculty, 2021; 3(2): 178-199.

[11] Güvenç E, Sakal M, Çetin G, Özkaraca O. Classification of students' course qualifications using machine learning techniques (article in Turkish, summary in English). Düzce Üniversitesi Bilim ve Teknoloji Dergisi 2022; 10(3): 1359-1371.

[12] Salameh Shreem S, Turabieh H, Al Azwar S, Baothman F. Enhanced binary genetic algorithm as a feature selection to predict student performance. Soft Computing, 2022; 26(1):1-13.

[13] Quy TL, Nguyen TH, Friege G, Ntoutsi E. Evaluation of group fairness measures in student performance prediction problems. Communications in Computer and Information Science, January 2023.

[14] Özkan Y, Önay Koçoğlu F, Selçukcan Erol Ç. Prediction of student performance by deep learning algorithm (article in Turkish, summary in English). In: 7th International Conference on Innovations in Learning for the Future, 2018; İstanbul University, İstanbul pp.136-145.

[15] Başer SH, Hökelekli O, Adem K. Predicting the performance of students studying in secondary education using data mining methods (article in Turkish, summary in English). Bilgisayar Bilimleri ve Teknolojileri Dergisi, 2020; 1(1): 22-27.

[16] Çınar D, Yılmaz Gündüz S. Prediction of secondary school students' mathematics success with machine learning. In:13th International Istanbul Scientific Research Congress on Life, Engineering, and Applied Sciences, Nisan 2023. İstanbul:pp.94-104

[17] Yavuzarslan M, Erol Ç. Using learning management system logs to predict undergraduate students' scademic performance. Bilişim Teknolojileri Dergisi 2022; 15(2): 199-207.

[18] Kayalı S, Buyrukoğlu S. Classification of academic performance of students with the implementation of machine learning algorithms (article in Turkish, summary in English). In: 2nd International Conference on Educational Technology and Online Learning*;* Ağustos 2022. Balıkesir, pp.330-336

[19] Bentaleb A, Abouchabaka J. Ensemble learning for mining educational data. Journal of Theoretical and Applied Information Technology 2022: 100(9): 2715-2722.

[20] Bozkurt Keser S,Aghalarova S. HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students. Education and Information Technologies 2022; 27: 4521–4552.

[21] Ünal F. Data mining for student performance prediction in education.In: Derya Birant,editor. Data Mining-Methods, Applications and Systems. e-book. 2020; pp. 423-432.

[22] Kızılkaya Y M, Oğuzlar A. Comparison of some supervised learning algorithms r programming language (article in Turkish, summary in English). Karadeniz Uluslararası Bilimsel Dergi, 2018; 37: 90-98.

[23] Felicia, Ferren. Exploring secondary school performance by using machine learning algorithms. Journal of Educational Analytics, 2022; 1(1): 41-60.

[24] Krizhevsky A, Sutskever I, Hinton EG. ImageNet classification with deep convolutional neural networks. In: F. Pereira and C.J. Burges and L. Bottou and K.Q. Weinberger,editors. Advances in Neural Information Processing Systems. 2012. pp. 1090–1098.

[25] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521:436-444

[26] Raiko T, Valpola H, Lecun, Y. Deep learning made easier by linear transformations in perceptrons. In: Fifteenth International Conference on Artificial Intelligence and Statistics 2012; 22:924-932.

[27] MW Gardner, SR Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric Environment 1998; 32: 2627-2636

[28] EB Baum, On the capabilities of multilayer perceptrons. Journal of Complexity 1988; 4: 193-215

[29] Bayır F. Artificial neural networks and application on forecasting. MSc İstanbul University, İstanbul, Turkey, 2006.

[30] Diaz-Quijano FA. A simple method for estimating relative risk using logistic regression. BMC Medical Research Methodology 2012; 12: 1-6.

[31] Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. Statistics in Medicine 1998;17(14):1623-1634.

[32] Manikandan G, Vasudev A, Balasubramanian A. A survey to identify an efficient classification algorithm for heart disease prediction. International Journal of Pure and Applied Mathematics 2018; 119(12): 13337-13345.

[33] Fan S, Huang B. Training iterative collective classifiers with back-propagation In: 12th International Workshop on Mining and Learning with Graphs, San Francisco, USA.

[34] Vural T. Microwave brain mass lesion diagnosis by using data mining. MSc, Yıldız Technical University, İstanbul, Turkey, 2017.

[35] Vanthienen J. A more general comparison of the decision table and tree: a response. DTEW Research Report 9224, 1992.

[36] Kohavi R. The power of decision tables. In: Machine Learning: European Conference on Machine Learning;2005; 912; Springer, Berlin, Heidelberg: pp.174–189.

[37] Huysmans J. Dejaeger K. Mues C. Vanthienen J. Baesens B. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models, Decision Support Systems; 2011;(51)1: 141-154.

[38] Hühn J, Hüllermeier E. FURIA: an algorithm for unordered fuzzy rule induction. Data Min Knowl Disc 2009; 19: 293–319.

[39] Olaru C. Wehenkel L. A complete fuzzy decision tree technique, Fuzzy Sets and Systems, 2003; 138(1): 221-254

[40] Al-diabat M. Arabic text categorization using classification rule mining. Applied Mathematical Sciences 2012; 6(81): 4033-4046.

[41] Karadağ M. Comparison of performances of decision trees and logistic regression analysis by a simulation study. MSc, Trakya University, Edirne, Turkey, 2014.

[42] Landwehr N. Logistic model trees. University of Freiburg, Freiburg, Germany, 2003.

[43] Witten IH., et al. Practical machine learning tools and techniques. In: Data mining. Amsterdam, The Netherlands: Elsevier, 2005. p. 403-413.

[44] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research, 2003; 3: 1157-1182.

[45] Dash M, Liu H. Feature selection for classification. Intelligent Data Analysis, 1997; 1: 131-156.

[46] Wang Z, Wang Y, Srinivasan R. A novel ensemble learning approach to support building energy use prediction. Energy and Buildings 2018; 159:109-122.

[47] Onan A. A clustering based classifier ensemble approach to corporate bankruptcy prediction (article in Turkish, summary in English). Alphanumeric journal; 2018; 6(2); 365 – 376.

[48] Çınar D. Prediction of secondary school students' success with machine learning. MSc, Eskisehir Technical University, Eskisehir, Turkey, 2023.