IJEEC

International Journal of Electrical Engineering and Computing
Vol. 9, No. 1 (2025)

Original research paper
UDC 681.586:612.087-057.874
DOI 10.7251/IJEEC2501020M

Two and Three Dimensional Support Vector Algorithm for Students' Classification

Marina Milićević

[1] Faculty of Production and Management Trebinje, University of East Sarajevo, Bosnia and Herzegovina

E-mail address: marina.milicevic@ues.fpm.rs.ba

Abstract— **In an era where technology with its development is constantly reshaping the field of professional practice, it is evident that educating engineers nowadays is a challenging task. This is supported by the increasingly popular teaching methods in which traditional learning methods are combined with modern technologies, all with the aim of greater motivation for learning and work among the students. The subject of research in this paper is specifically one teaching method, adapted for teaching at technical faculties, which uses machine learning methods (support vector method) in its background. The introduced method recognizes the importance of assessing students' abilities and knowledge because a good classification of students allows the teacher to plan the lesson more easily and choose the approach they believe will give the best results.**

Keywords - support vectors method; machine learning algorithm, classification of students;

I. INTRODUCTION

One of the challenges faced by all organizations today, including universities, is the adoption of innovations. The situation is particularly interesting when talking about technological innovations in the teaching process, which, although challenging, have a wide range of benefits both for universities and for students. By integrating modern methods, such as artificial intelligence and machine learning, into the traditional processes at university, very useful information can be obtained that will guide activities of university in the direction of achieving the projected goals.

Predicting the academic success of students and their classification is, generally speaking, a set of methods applied to collect data and obtain information about the academic success of students, which most often include some of the machine learning algorithms. The increasing computerization of universities has led to the fact that today there are loads of data about students in universities. A lot of important information for universities is hidden within that data, and getting that information is not an easy task and certainly involves artificial intelligence and machine learning methods. Some of the useful information can be the dependence of a student's academic success on the grade point average in high school or the structure of students who achieve the best (worst) results. Better utilization of available data leads to better knowledge and academic success of students, which ultimately results in a better image and position of a university. The daily efforts as such to raise the level of quality in education to a higher level have led to the more intensive use of machine learning algorithms in data mining at universities (EDM - Educational Data Mining).

There is a large number of different machine learning algorithms developed and adapted for specific tasks, and what they all have in common is that they are based on the theory of functional analysis and statistics. The existence of data and patterns within are the two main prerequisites for the application of machine learning, and learning itself can be divided into supervised and unsupervised learning. The first category proved to be the one that gives better results, mostly because in this method the input data is labeled, i.e. in addition to the set of input data, we also have a set of labels that represent the output values. With unsupervised learning, it is not known in advance what the output from the system will look like because we are working with unlabeled data. In addition to these two groups, reinforcement learning is often encountered in practice - a special type of machine learning in which the agent learns from interaction with the environment using the trial and error method.

Supervised machine learning algorithms can have different output values, thus we can classify them into two groups: (1) classification algorithms whose output values are from a set with a finite number of elements and each instance of a problem

---

This paper is a revised and expanded version of the paper presented at the XXIV International Symposium INFOTEH-JAHORINA 2025

that is observed will take one of these values at the output, that is, it will be classified into one of the output classes and (2) regression algorithms whose output is a single, real number, and their goal is to find a function based on the input values that will give the corresponding output values.

Various metrics are used to determine the success rate of the selected algorithm, such as the mean square for regression algorithms. In classification algorithms, precision is most often used - the percentage of correctly classified data. In situations where the number of examples of one class is significantly higher than the number of examples of the other class (unbalanced data) in machine learning classification algorithms, we use metrics such as F1 score, cross-entropy and area under the ROC curve.

The main problem of machine learning is overfitting. Situations when the results of the machine learning algorithm are not satisfactory are solved by adding new attributes and by creating a more complex hypothesis, we try to get an improvement during the training of the algorithm. However, there is a danger of overfitting because complex functions might easily record mappings between input and output data during training, but also, fail to notice useful patterns and regularities among the data that could then be applied in the general case (on completely new data), and not only on the data from the algorithm's training set. If the function is too simple for us, we fall into the opposite problem (underfitting), because simple functions cannot learn complex relationships between input and output data.

This paper presents a machine learning model for predicting a student's academic success in one subject. Data on the academic success of first-year students were collected on the Industrial Management and Industrial Engineering in Energetics study programs at the University of East Sarajevo during a period of 4 years (from 2017 to 2021), and the faculty's information system was used as a means of data collection. Since it is related to the study program of a technical profile, the subject from the first year of studies that is considered as the most important and the mastery of which is crucial for the further course of study for any student is mathematics.

The main goal of the research is to be able to assess in the middle of the semester whether or not the student will pass mathematics at the end of the semester, using data from the university's information system. Since mathematics is a basic discipline that permeates and connects many other areas, especially areas of technical orientation, it is assumed that data on student success in technical and practical courses can be used to predict student success in mathematics. The following data were highlighted as important data: average grades from high school, success in laboratory exercises during the semester and the average number of points in technical disciplines in the middle of the semester.

The machine learning algorithm used in the work is the Support Vector Machine (SVM) method, and as we solved the classification problem, the two defined classes have been indicated by F (the student will not pass the course) and P (the student will pass the course).

II. A REVIEW OF LITERATURE

With the rapid development of technology and the increasing amount of data collected at universities, various machine learning methods have started to be intensively applied for prediction of academic success of students. However, various researchers have used different techniques and attributes (academic, demographic, sociological, etc.) for prediction and classification. Below is a brief overview of some of the studies that are of significance to researchers given in this paper.

A reliable overview for future researchers on the topic of applying artificial neural networks in the analysis and collection of data about students at universities (EDM) is given in [1]. Attention is directed towards processing and collection of data, and then use of neural networks for EDM purposes by the authors is analyzed. The most common types of neural networks are recurrent neural networks, then one-layer and multilayer feedforward neural networks, while supervised learning is the most common type of algorithm used, with mean square error appearing as the most common method for assessing the quality of the network, or for the cost function minimization.

Decision trees are a machine learning model that can be applied to classification and regression tasks as well. The authors of the study [2] presented two decision tree models in which, from a large number of attributes for each student, only those that significantly affect prediction and classification were selected, while the dependent variable in the trees was varied. In one model, the average grade was used as the dependent variable, and in the other, the time required to finish the study.

In the paper [3], a decision tree was used to classify students depending on their success in mathematics courses, where the students were classified into three classes: pass – students who pass the course, fail – students who will not pass the course, and conditional – students who are on the border between these two classes. Experimental results showed that the accuracy of the model was 72%. Another study on the application of decision trees to classify students into two classes depending on the degree of success in creating a specific 3D model is given in the paper [4].

One of the most popular algorithms of classical machine learning is the support vector machine learning (SVM) method, which is basically a non-probabilistic binary classifier. Its use has been particularly well in the problem of student classification. In the paper [5] this method was used to predict whether a student will drop out or graduate from university, while in [6] the SVM method was used to predict student success using data available from social media.

SVM algorithm for predicting students' performance is widely used mostly due to its good performance. A systematic review on application SVM algorithm in a field of education is [7]. In this research authors aim to determine which factor affects students' performance most and what kind of machine learning algorithms is mostly used. In the study [8] SVM algorithm was compared to different machine learning algorithms with the aim of investigating the relationship between students' preadmission academic profile and their final academic performance.

It is important to note that the applicability of machine learning algorithms is very broad and includes disciplines ranging from energy, sociological studies to extracting semantics and text processing [9], [10] and [11].

III. METHODOLOGY

A. Support Vector Method

The support vector method (SVM) is a type of algorithm most commonly used for classification problems, although it can be used for regression problems as well. The support vector method is a non-probabilistic (the output from the model is only a decision), binary (divides the data into one of two classes), linear (extensions are possible) classifier. The goal of the SVM method is to find the optimal hyperplane that maximally separates the data points of one class from another, that is, that has the maximal margin. The dimension of the hyperplane depends on the number of features: if the number of features is equal to two, then the hyperplane is a line, while for three features, the hyperplane is a plane of dimension 2.

The hyperplane is determined as the maximum distance of the data from the decision limit - in this way, only the data closest to it (support vectors) influence the determination of the decision limit, while those that are far from it have no influence on the optimization process. In the case of linearly separable binary classification, the model represents the separation hyperplane equation. If *n* is the number of features used in the model, then the hypothesis is of the form:

$$h(x) = W^T X + b = 0,$$

where *b* represents *bias* and *W* is a normal vector on hyperplane. If examples of one class are indicated with $y = 1$, and examples of another class with$y = -1$, the equations of the hyperplanes that rest to the classes defined in this way are:

$$W^T X + b = 1 \text{ i } W^T X + b = -1.$$

The width of the margins defined in this way is given by the formula $\frac{2}{||W||}$, so the optimization problem is now defined as the maximization of the expression $\frac{2}{||W||}$, i.e. minimization $||W||$.

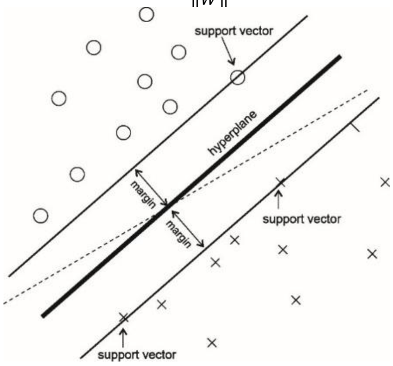

Figure 1. SVM model hyperplane illustration.

Insisting on linear separation of classes can lead to overfitting the model to the data, and instead data is allowed to enter the margin or to cross over to the wrong side of the hyperplane (to some extent, of course). The margin in which this is allowed is called a soft margin.

As the performance criteria, this paper used confusion matrix and precision. A matrix used to specify how well a classification method performs is called a confusion matrix. It is an m-dimensional square matrix, where m is the number of classes. Elements on a matrix diagonal are the correctly classified number of instances. The confusion matrix consists of four numbers that are used to define the metrics of the classification algorithm:

- TP (True positive): predicted to be positive and the actual value is also positive;
- FP (False positive): predicted to be positive but the actual value is negative;
- TN (True negative): predicted to be negative and the actual value is also negative;
- FN (False negative): predicted to be negative but the actual value is positive.

These parameters are frequently used for calculating other measures like accuracy, precision and error rate.

One of the simplest and most widely applied metrics in classification problems is accuracy. It is defined as the ratio of correctly predicted instances to the total instances in the dataset. In terms of mathematics, it may be expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

Performance criterion called precision is a classification metric that indicate the accuracy of positive predictions. It measures the proportion of instances the model predicted as positive that were true positives. It is the number that ranges from 0 to 1 and it is mathematically defined as follows:

$$Precision = \frac{TP}{TP + FP}.$$

Error rate measures the quantity of inaccurate predictions and classifications, which enhances accuracy. It is frequently given as a percentage and is computed by subtracting the accuracy from one:

$$Error\ rate = 1 - Accuracy.$$

B. Data

Data collection and preparation was the first step in constructing the model. 50 students enrolled in the first year of industrial engineering were observed over a 4-year period. Data preparation included the elimination of unnecessary attributes (personal data, elimination of success on general courses such as foreign language).

The data were divided into two groups: data for model training (80%) and data for model testing (20%). Three attributes were identified as potentially significant for the research objective: high school grade point average (GPA), performance in lab exercises during the semester, and performance in technical disciplines from the first year at midterms. During the experiment, the last two attributes were used for the two-dimensional support vector algorithm, while by adding a third attribute, high school grade point average, the students were classified in a three-dimensional algorithm.

Table I provides an analysis of the selected attributes for the model with a domain for each attribute. The dependent attribute, i.e. that we predict is the student's success in the mathematics course taken at the end of the academic year.

TABLE 1 SIGNIFICANT ATTRIBUTES WITH DESCRIPTION

Attribute	Description	Min	Max	Mean	Domain
GPA	Grade point average	2.50	4.78	3.61	Od 2.5 do 5.00
Lab exercises	Lab exercises success	0.00	20.00	13.89	Number from 0 do 20
Midterms	Average point number from technical disciplines at midterms	6.25	24.5	17.85	Number from 0 to 25
Dependent attribute	Success from the Math course	-	-	-	F, P

IV. RESULTS

The model from this study was developed and the experiment was conducted using the MATLAB programming environment. A set of 50 instances of the collected data was divided in a ratio of 80:20 for training and testing. During binary classification we conducted, the dependent variable "success in the observed subject", which is basically expressed through the number of points obtained, was converted into two classes: for the number of points greater than 60, class P, and for the number of points less than 60, class F.

Two support vector models for student classification were developed in the paper. In the first, two-dimensional model, the success of students in the subject of mathematics was observed in relation to two attributes: success in lab exercises during the semester and the average number of points in technical disciplines at the middle of the semester. The separation hyperplane in this case is a line. By adding a third attribute, the average grade point average from high school, the dimension of the model increased by one, so the hyperplane from this model (Figure 8) is actually a plane.

A. Two-dimensional SVM algorithm

Experimental results have shown that the accuracy of the model on both the validation and testing data is very high - 90%.

Figure 1 shows a scatter plot of the data in relation to the two observed features, where:

- the red dot represents the correctly classified class P;
- the red x represents the incorrectly classified class P;
- the blue dot represents the correctly classified class F;
- the blue x represents the incorrectly classified class F.

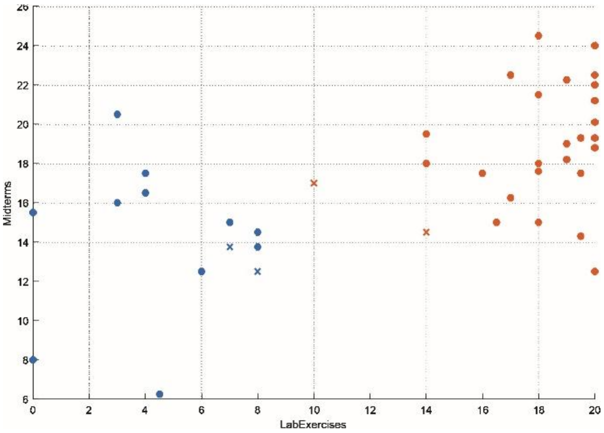

Figure 2. Scatter plot of model data

From the confusion matrix for training data shown in Table II (left), we see that out of a total of 27 instances belonging to class P, 25 of them are correctly classified, while 2 of them are wrongly classified as class F. Of 13 instances belonging to class F, 11 of them are correctly classified, while 2 of them are wrongly classified as class P. A similar ratio of correctly and wrongly classified instances applies to the test data (table II right).

TABLE II CONFUSION MATRIX ON TRAINING (LEFT) AND TEST DATA (RIGHT)

Predicted class

Actual class	F	P
F	11	2
P	2	25

Predicted class

Actual class	F	P
F	1	1
P	0	8

Figure 3 shows a separation hyperplane with upper and lower soft margins and support vectors.

For a hyperplane defined in this way, we have the following characteristics: bias: *b=-4.4020* and linear coefficients of prediction (beta): *0.2175* and *0.1209*.

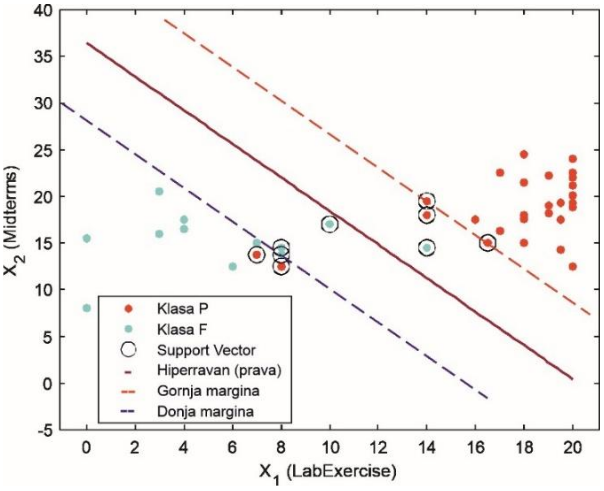

Figure 3. Separation hyperplane with soft margin support vectors

B. Three - dimensional SVM algorithm

Experimental outcomes showed that the accuracy of the model on the training data (40 instances) is – 92.5%. Figures 4, 5 and 6 show scatter plots of the data against the observed features, where circles correspond to a correctly classified instance of one class, while x indicates a misclassification.

Scatter plot is a graphical representation of the relationship between multiple variables. It is a useful tool in machine learning for visualizing the correlation between features in a dataset. They display the scatter plot of each pair of features in a dataset.

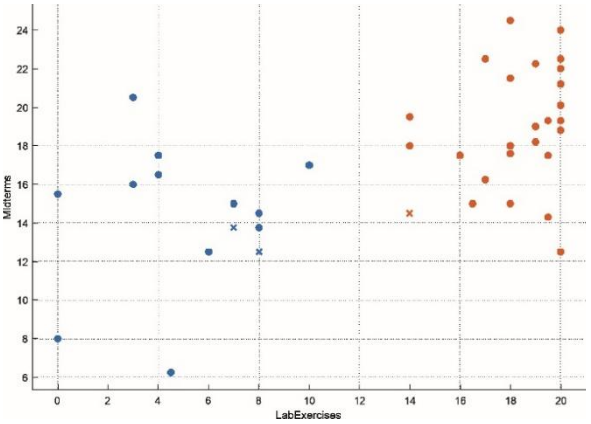

Figure 4. Scatter plot for Midterms and Lab Exercises attributes

The figures show different levels of connection between the attributes. Figure 4 shows a correlation between efficacy in lab exercises during the semester and efficacy in technical disciplines from the first year at midterms. Students who achieved higher success on one of these attributes generally achieved higher success on the other, and vice versa.

A significant correlation in the same direction is also observed in figure 5, where we see a positive scatterplot between the attributes high school grade point average (GPA) and performance in lab exercises during the semester.

A somewhat smaller, but still positive correlation (trending upwards from left to right can be seen in figure 6, between the variable high school grade point average (GPA) and performance in technical disciplines from the first year at midterms.

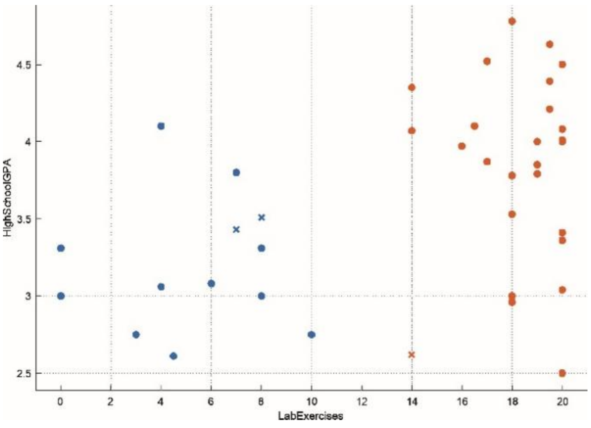

Figure 5. Scatter plot for High School GPA and Lab Exercises attributes

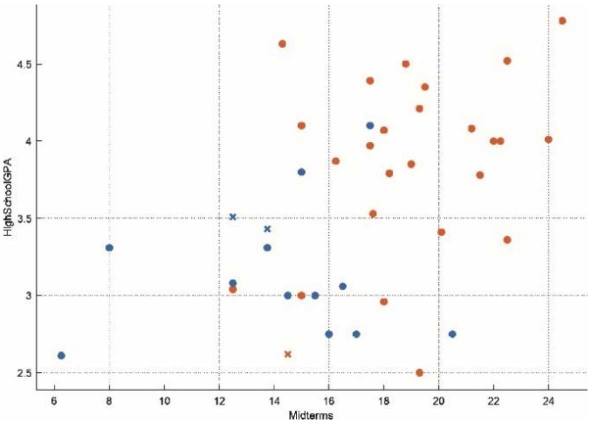

Figure 6. Scatter plot for High School GPA and Midterms attributes

A graphical figure called the Receiver Operator Characteristic (ROC) curve is used to demonstrate how well binary classifiers can diagnose problems. A trade-off between true positive and false positive data is offered by the ROC. For the three-dimensional support vector machine algorithm's, the area under the Receiver Operating Characteristics curve is shown in Figure 7.

The larger the area under the ROC curve are, the machine learning algorithm is better in distinguishing given classes – classes F and P in this study. AUC for the ideal value is 1. Three dimensional SVM model in this study gain area under ROC curve 0.97.

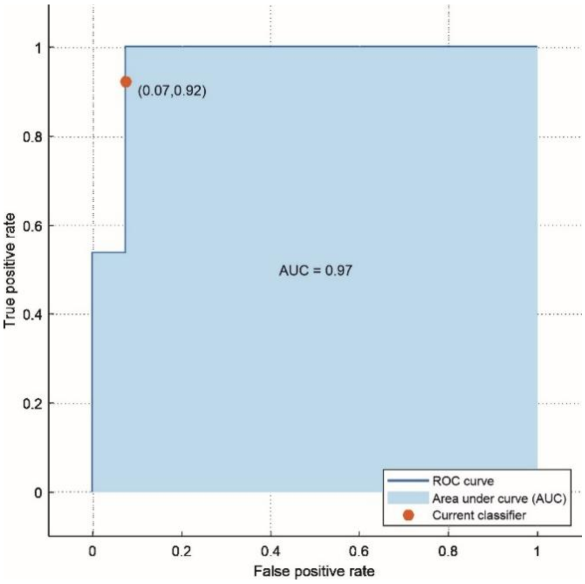

Figure 7. Area under the ROC curve

The confusion matrix for the three-dimensional support vector algorithm is given in Table III. The left side of Table III shows the correct and incorrect classifications of students in numbers, and the right side in percentages, where:

TPR – True Positive Rates;

FNR – False Negative Rates.

The training set consisted of a total of 40 examples, 13 examples of class F and 27 examples of class P. 1 example of class F was misclassified, which the system estimated to belong to class P. Out of a total of 27 examples of class P, 25 were correctly classified, and 2 were misclassified as members of class F.

These results demonstrate the ability of the algorithm to learn with high accuracy based on input attributes to correctly classify students depending on their performance in the subject of mathematics.

TABLE III CONFUSION MATRIX FOR THREE-DIMENSIONAL SVM

Predicted class

		F	P	TPR	FNR
Actual class	F	12	1	92,3%	7,7%
	P	2	25	92,6%	7,4%

On the Figure 8, the 3-dimensional visualization of binary students' classification and performance of SVM for the three attributes is shown. Hyperplane (bias = -0.0674) that separate two classes are also given. Red circles represent instances of class F, blue circles represent elements of class P. Samples that are close to hyperplane are support vectors and they are represented as rounded circles (red or blue).

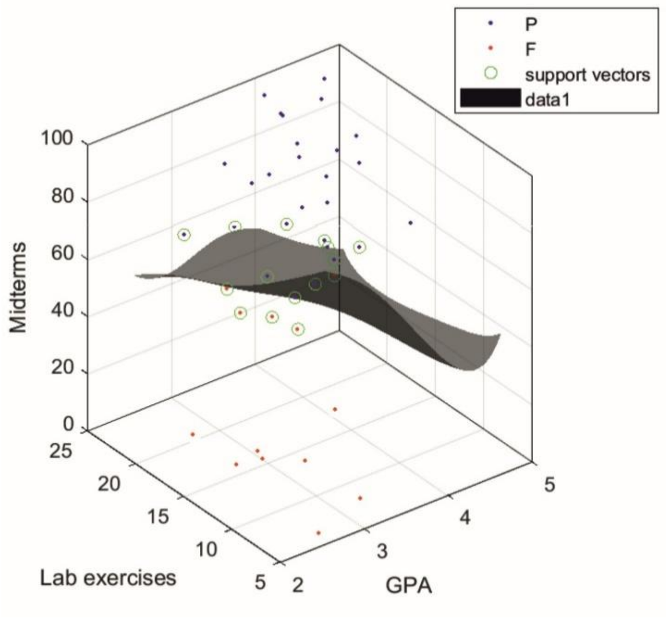

Figure 8. Hyperplane in three-dimensional space

V. CONCLUSION

The result of the research in the paper is the proposed SVM model in two and three dimensions for classification of first-year students of technical faculties depending on their success in mathematics as a basic discipline. The models presented in the paper are one of the most popular machine learning algorithms - support vector methods. Various factors affect the academic success of students, but in this paper, data available from the university information system are used.

Based on the research conducted in the paper, the following conclusions can be summarized:

(1) The support vector model presented in the paper can be used to classify students into two classes based on the predicted success of the student in the observed subject.

(2) The study has also provided insight into the fact that data on students stored in university information systems can be used to predict the academic success of students.

(3) As presented in the model, there is a strong correlation between the success of students in technical subjects and the success of students in mathematics.

The research presented in the paper can be expanded in several directions. First, it is possible to increase the dimension and observe a larger number of attributes. In addition to that, various machine learning algorithms can be combined to create a hybrid model for classifying and predicting success of students.

REFERENCES

[1] E. Okewu, P. Adewole, S. Misra, R. Maskeliunas, and R. Damasevicius, "Artificial neural networks for educational data mining in higher education: a systematic literature review," Appl. Artif. Intell. 35 (2021), no. 13, 983–1021.

[2] M. Dragičević, M. P. Bach, and V. Simičević, "Improving university operations with data mining: predicting student performance," Int. J. Econ. Manag. Eng. 8 (2014), no. 4, 1101–1106.

[3] M. Milićević, B. Marinović, and Lj. Jeftić, "Machine learning methods as auxiliary tool for effective mathematics teaching," Comput. Appl. Eng. Educ. (2024), e22787.

[4] M. Miljanović, M. Milićević, M. Jokanović Đajić, and S. Dostinić, "FabLab: Encouraging the Creativity of Secondary School Students of Various Professional Orientations," Revista De Education, Spain, 2024.

[5] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. Abbas, and J. S. Alowibdi, "Predicting Student Performance using Advanced Learning Analytics," In Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion), Republic and Canton of Geneva, 2017.

[6] Benablo, C. I. P., Sarte, E. T., Dormido, J. M. D., and Palaoag, T. "Higher education Student's academic performance analysis through predictive analytics," In Proceedings of the 2018 7th International Conference on Software and Computer Applications—ICSCA 2018. New York.

[7] M. Kumar and Y. K. Salal, "Systematic Review of Predicting Student's Performance in Academics," Int. J. Eng. Adv. Technol., vol. 8, no. 3, (2019), 54–61.

[8] S. Oloruntuba and J. L. Akinode, "Student academic performance prediction using support vector machine," IJESRT Int. J. Eng. Sci. Res. Technol., May, 2019.

[9] M. Milicevic and B. Marinovic, "Machine learning methods in forecasting solar photovoltaic energy production," Therm. Sci. 28 (2024), 479–488.

[10] A. Onan, "An ensemble scheme based on language function analysis and feature engineering for text genre classification," J. Inform. Sci. 44 (2018), no. 1, 28–47.

[11] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," Expert Syst. Appl. 57 (2016), 232–247.

Marina Milićević is an assistant professor of Computer Science at the University of East Sarajevo, Faculty of Production and Management Trebinje with the main research field in interactive theorem proving in geometry. She is also engaged in the application of machine learning tools in modeling and forecasting processes. She has a PhD in Applied Mathematics from the University of Novi Sad (2020).