International Journal of Instruction *January 2026 ● Vol.19, No.1*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 367-386

Article submission code: 20250515084750
Received: 15/05/2025
Revision: 09/08/2025
Accepted: 21/08/2025
OnlineFirst: 03/10/2025

STEM Workshops and Students' Interest in Mathematics, Physics, and Computer Science: Machine Learning Approach

Aleksandar Milenković
University of Kragujevac, Faculty of Science, Serbia, *aleksandar.milenkovic@pmf.kg.ac.rs*

Dragutin Ostojić
University of Kragujevac, Faculty of Science, Serbia, *dragutin.ostojic@pmf.kg.ac.rs*

Dalibor Rajković
University of Kragujevac, Faculty of Science, Serbia, *d.rajkovicrs@gmail.com*

Milan Milikić
University of Kragujevac, Faculty of Education, Serbia, *milikic.milan@yahoo.com*

Although STEM education is present in many countries and various aspects of this concept are being extensively researched, many educational systems still rely heavily on subject-specific learning. This study aims to examine the attitudes of students aged 14-17 regarding whether their participation in STEM workshops integrating content from mathematics, physics, and computer science contributes to an increased interest in studying these individual subjects. More specifically, we sought to determine whether the degree of increased interest in learning these STEM subjects could be predicted based on a set of dependent variables, including students' sociodemographic data and their attitudes toward the different aspects of the conducted STEM workshop. The data was analyzed using 18 ML classifiers, with outlier removal methods applied to the five models that yielded the best results. The best-performing model was the decision tree with the IF automatic outlier removing technique, achieving an accuracy of 0.94. The key factors contributing to students' increased interest in learning mathematics, physics, and computer science were primarily level of their engagement in the STEM workshop, beliefs about newly acquired knowledge, age, prior experience with STEM workshops, and current grade in physics.

Keywords: STEM workshop, students' interest, STEM subjects, data mining, decision tree

INTRODUCTION

The goal of STEM education is to develop students' knowledge and skills in science, technology, engineering, and mathematics. As such, it is the subject of numerous studies. These studies explore various approaches to STEM education at all educational

Citation: Milenković, A., Ostojić, D., Rajković, D., & Milikić, M. (2026). STEM workshops and students' interest in mathematics, physics, and computer science: Machine learning approach. *International Journal of Instruction*, *19*(1), 367-386.

levels from preschool to post-doctoral studies (Granovskiy, 2018). The current research also considers the education of different groups of students, aiming to enhance the achievements of students learning through STEM approaches compared to those studying specific scientific disciplines. Additionally, research investigates students' interest and motivation in STEM subjects, differences in students' success in STEM education, and more. However, despite the widespread adoption of this principle in the curricula of various educational institutions, this approach to teaching and learning has not yet taken root to the desired extent. For instance, in the Republic of Serbia, a subject-based approach to education remains deeply rooted, with teachers attempting to establish interdisciplinary connections to some extent (Milenković & Momčilović, 2025) which has also previously led to positive teaching and learning outcomes in other countries (Djudin, 2023).

The former research has shown that the STEM approach to education boosts students' interest in mathematics and natural sciences (McDonald, 2016). However, students generally understand the importance of STEM subjects in everyday life, find them interesting, but challenging; they are commonly perceived as disciplines that not everyone can manage.

Machine learning serves as a powerful tool for determining various models, and identifying relationships between different variables. It has broad applications across mathematics, computer science, natural sciences, medicine, economy, and social sciences. Some machine learning methods are convenient for interpretation and can be quite intuitive, such as decision trees. Considering that students in higher grades tend to lose interest in mathematics and other natural sciences, the goal of this research was to use machine learning to determine how students perceive STEM workshops. Specifically, the aim was to identify the indicators of active participation in such workshops and determine whether there is a positive impact on students' interest in STEM subjects, particularly mathematics, physics, and computer science. The responses would be beneficial for those planning and implementing STEM workshops in that they can help them gain deeper insight into the aspects which need most attention while creating and conducting these workshops to enhance students' interest in STEM subjects.

Theoretical Framework

STEM Education

The concept of STEM education emerged in the USA in 2009, to provide students with STEM skills that will prepare them for the future (Morgan et al., 2022). This approach first entered the scene as an extension of the inquiry-based learning approach (Deák et al., 2021). To encourage students to improve their skills in STEM subjects and prepare them for work relying on technology, the growing emphasis is being placed on the development of a STEM approach to teaching (Tawbush et al., 2020).

There have been numerous attempts to define the concept of STEM education. Aguilera & Ortiz-Revilla (2021) mention several different definitions of the STEM approach found in the literature. For instance, one definition states that STEM education is an interdisciplinary approach to learning in which rigorous scientific concepts are linked with real-world lessons, while students simultaneously apply science, technology, engineering, and mathematics in contexts that connect school, community, work, and global business (Tsupros et al., 2009 in Mohr Schroeder et al., 2015). The STEM approach to education has also been defined as the integration of two or more disciplines when solving real-life problems (Bozkurt Altan & Tan, 2020). It has been acknowledged that this approach to education allows students to think and propose solutions to everyday life problems with minimal support from teachers (Sang & Simpson, 2019). Finally, it has been asserted that this educational approach encourages student creativity, allowing them to think and learn experientially (Sirajudin et al. 2021).

Despite the importance attached to STEM education as an innovative teaching approach, the results of the TIMSS and PISA studies have shown that students' STEM skills are not at an adequate level and that they are not improving from cycle to cycle (Roungos et al., 2020). Some tests in mathematics and natural sciences show that in recent years, students from OECD countries have shown a significant decline in interest in STEM fields and the skills associated with them (Jeffries et al., 2020). Also, numerous studies have shown that students' interest in STEM fields has decreased, and they are moving away from these fields (Habig & Gupta, 2021; Hiğde & Aktamiş, 2022). All these findings have led numerous countries to take measures to prevent the negative trend and to determine what has led to the decline in students' motivation for STEM fields.

Students' Attitudes and Motivation Towards STEM

With the transition from elementary to secondary school, many students decide not to continue studying science (King et al., 2015, see Mateos-Núñez & Martínez-Borreguero, 2023). Although they have positive attitudes toward learning mathematics and science in elementary school (Martínez-Borreguero et al., 2020), their interests decline as they move to secondary school (Marbà-Tallada & Márquez, 2010). High-school students exhibit exceptionally negative attitudes toward natural science classes (physics, chemistry) (Dávila-Acedo et al., 2021). Reportedly, not all students are motivated to actively engage in mathematics or science classes or activities, and since studies have indicated a connection between students' interests and achievements in mathematics and other natural sciences (Japashov et al., 2022; Ing, 2014), the question arises how to boost students' motivation and interest in these STEM fields (McDonald, 2016).

STEM-based activities are grounded in real-life problems. Through these activities, students develop skills in experimentation, design, data collection, analysis, and drawing conclusions, as well as connecting formal knowledge with natural events. Hiğde and Aktamiş (2022) examined the effects of STEM activities on students' motivation, interest, and attitudes toward STEM education with a sample of 22 students aged 13 and 14. The findings after an eight-week experimental program that included the implementation of STEM activities show an increase in motivation and the development of positive attitudes for STEM disciplines and skills related to creativity, peer collaboration, critical thinking, and problem-solving within the experimental group. Their results are in agreement with the previous studies that confirm that STEM activities increase student motivation (English, 2017; Sudarsono et al., 2022).

One analysis of the students' attitudes toward STEM education was conducted with 170 students aged 7 to 14 who participated in STEM education (Timur et al., 2020). The data were collected through a five-point Likert-type STEM Attitude Scale (Faber et al., 2013) and the interviews focusing on students' attitudes and opinions toward STEM education. The results show that the activities involved in the STEM workshops lead to an improvement in students' attitudes toward STEM education with no significant difference between the genders. The responses obtained through the interviews indicate that students were not sufficiently informed about STEM education, but that learning through practical activities in STEM workshops positively affected their attitudes. Almost all students expressed satisfaction for having attended the workshops. They confessed that they felt comfortable and acquired a lot of new knowledge. Most students stated that they understood the content in the fields of mathematics, science, and computer science better when it was presented through STEM workshops (Timur et al., 2020).

Another study was dedicated to the impact of STEM workshops on students' emotions and attitudes (Mateos-Núñez, 2020). Specifically, with a sample of 256 elementary school students aged 10 to 12, the authors conducted a quasi-experimental study to investigate which emotions (curiosity, fun, confidence, satisfaction, boredom, worry, anger) students predominantly display during STEM workshops and how the workshops influence the formation of their attitudes toward STEM education. The results show that the experimental group exhibited a higher percentage of positive emotions compared to the control group. For example, nearly 90% of students in the experimental group found learning during the STEM workshop fun, compared to 60% of students who found learning the same content through traditional methods fun. Almost two-thirds expressed satisfaction with the experimental STEM approach, compared to just over one-third of the control group. Every fifth student in the control group stated that the lesson was boring, whereas only 3.6% of students who learned through the STEM workshop reported feeling bored. The differences in response distribution between the two groups were statistically significant in terms of positive emotions such as fun, satisfaction, and confidence. The results of the same study also showed that the implementation of the STEM workshop with the experimental group generates positive attitudes among students compared to the traditional teaching method applied in the control group. An identical percentage of students in the experimental group (as much as 99.3%) stated that they liked the STEM workshop and that they successfully mastered the instructional content (the percentage of positive responses to these two questions in the control group ranged between 85% and 90%). Interestingly, 97.1% of students in the experimental group stated that they would like to do more of these practical activities in science and math classes, while as many as 73.2% of them expressed a preference for seeing practical examples rather than just theoretical concepts, which they were exposed to in traditional teaching. Additionally, 95.7% of students in the experimental group stated that they believe they will remember the contents they have learned more easily thanks to the practical workshop. Finally, when it comes to the practical application of acquired knowledge, compared to less than half (46.4%) of students in the control group who believed they could independently make a model of their own related to the content covered in the lessons, more than two-thirds (68.1%) of those who participated in the STEM workshop agreed with the statement that they could make the model by themselves without help (Mateos-Núñez et al., 2020).

When it comes to interest in pursuing a career in STEM, it has also previously been confirmed that this interest is positively correlated with students' attitudes toward STEM, as well as with their motivation to study science (Razali, 2021).

Since different studies have confirmed that STEM activities increase student motivation for STEM subjects (English, 2017; Kanadli, 2019), it is necessary to deepen the understanding of the factors that influence the increase in student motivation for pursuing STEM fields and implement the necessary educational reforms aimed at achieving more favorable student outcomes.

Data Mining

Data mining is the process of obtaining useful and relevant information from large data sets. This technique uses various data analysis methods to identify patterns, trends, and unknown relationships among data (Fayyad et al., 1996). The goal of data mining is to extract useful information that can be used to make informed decisions, predict future events, or optimize business processes (Han et al., 2022). Key steps in data mining include data collection, data transformation and cleaning, selection of appropriate algorithms for analysis, model training, and evaluation of results (Witten et al., 2005).

In an educational context, especially in the annexes of STEM projects, data mining can provide deeper insights into the attitudes and behaviors of project participants. For example, data mining can analyze survey responses, identifying factors that influence participants' choice to engage in STEM projects. Using various machine learning methods, participants can be classified according to characteristics such as interest, previous experience, prior knowledge, and motivation. This allows educators and project authors to understand what motivates participants to engage in STEM fields and tailor project activities to their interests and needs. Data mining makes the survey analysis process more efficient, providing relevant information that supports the improvement of STEM education and student engagement (Romero & Ventura, 2013). Artificial intelligence has been used to predict students' STEM attitudes, where strong and positive correlations were observed between the predicted values obtained through a model based on combining fuzzy logic and artificial neural network and the actual values provided by middle school students (Göktepe Körpeoğlu & Göktepe Yıldız, 2024). Machine learning methods have only recently begun to be applied in STEM education research. Although, according to the findings of a systematic literature review conducted by Ismail and Yusof (2023), these methods were present in only 16 out of 105 scientific papers, the authors identified a trend of increasing use of these methods since 2018. Additionally, machine learning methods have been used for predicting students' academic performance in STEM education (Abdrakhmanov et al., 2024) as well as for predicting college STEM major selection (Chang et al., 2023).

Decision Trees

A decision tree is a graphical model in the field of machine learning that is used for data analysis and classification (Breiman et al., 2017). This technique is reliable and often used because of its simplicity and easy interpretation. The basic idea is to present the decomposition of the problem into a series of conditions that are applied to the data.

The tree consists of nodes and branches, where each node represents a specific test on a data attribute, and the branches indicate possible test outcomes. This structure allows data to be divided and directed in a way that leads to a final output or decision. Using a decision tree, it is possible to categorize or predict the class or value of new data instances.

The process of building a tree involves selecting the best attribute for splitting the data and continuing to split until the condition for completing the structure is reached. Algorithms such as ID3, C4.5, and CART are often used in this context (Quinlan, 1986). There is also the concept of ensemble methods, such as Random Forest, which uses multiple trees to better guide classification and regression (Liaw & Wiener, 2002). Although decision trees allow for a simple drawing of the decision structure, it is important to take care of the transitional adaptation of the model to the data, so it is necessary to use appropriate regularization and validation techniques.

METHOD

Within the framework of the K1 projects implemented in 2021/2022 under the call of the Center for the Promotion of Science of the Republic of Serbia, two projects were accepted, and their co-authors are also co-authors of this manuscript. These projects are ”Treasure Hunt” and ”Mission (Im)possible.” Both projects involved the STEM approach with emphasis on content from mathematics, physics, and computer science. The content was selected either to extend beyond the regular curriculum (for students to gain new knowledge) or to represent a practical application of knowledge acquired through formal education, exposing students to unfamiliar and unconventional problems. The content from science (especially physics), technology, engineering and mathematics intertwined, creating an interesting and enjoyable STEM experience for students.

Activities within the first project, ”Treasure Hunt” provided students with a creative and engaging introduction to the laws of geometric optics, the principle of laser operation, and other electronic components. Project activities were related to a series of interesting puzzles, brain teasers, and real-life problems with mathematical and physical backgrounds. By applying the laws of geometric optics (laws of light reflection), students adjusted the angles of mirrors to guide laser light to a photoresistor after multiple reflections. The change in voltage on the photoresistor, detected by Arduino, opened an electronic lock on a chest containing a message written in invisible ink, becoming visible due to exposure to light of a specific wavelength. After opening the bottle and deciphering the message, students solved mathematical and physical problems, puzzles, and riddles. This project was designed for students aged 14 and 15.

The second project, ”Mission (Im)possible”, is a STEM project designed for students aged 16 and 17. The work with the students occurred through several phases. In the introductory phase, students were familiarized with the laws of optics, and the operation and connection of sensors, lasers and other components. In this phase, students were introduced to basic concepts and ideas from cryptography, its historical application, and the use of computers to (de)encrypt messages. In one workshop, students were divided into two teams. Each team selected one member to be the representative (captain) of the team. Members of one team left the given space, while members of the other team, with the help of educators and their instructions, first encrypted a message of a certain length, using knowledge from cryptography and a written program, recorded it on paper, and left it in the designated place. Then, they assembled an alarm system by connecting sensors, lasers, and other electronic components at various locations within the given space to make it more challenging for the opposing team to reach the message that needs deciphering. After the installation, members of the second team were called back in, where they had to skilfully avoid lasers, other sensors, motion detectors, and infrared radiation, based on the acquired knowledge of their operation, using a piece of black fabric, an auxiliary laser, and a mirror to deceive the alarm system and successfully reach the paper. If, during movement, they intersected the laser beam or were detected by some sensor, an alarm was activated, and the opposing team gained a point. When they reached the requested message, they deciphered it using knowledge acquired during lectures on cryptography and instructions for deciphering, using an appropriate computer program. Through the analysis of feedback, team members were required to decipher the original message. After the first team had done their assignment, the teams switched their roles. Thus, activities within both projects were designed to engage students through play and collaboration.

When organising the workshops, the focus was placed on an active learning methodology in the context of practical STEM workshops, an approach that has already been successfully applied in practice to date (Mateos-Núñez et al., 2020, Martínez-Borreguero et al., 2020). In both projects, the implementation team first briefly introduced the students to the topics they would be working on. Students received worksheets for writing, and short PPT presentations were held to familiarize them in detail with all the activities and tasks they were required to complete through collaboration with their peers. After that, the described workshop activities were carried out (as part of the “Treasure Hunt” and “Mission (Im)possible” projects). Upon completion of the activities, the students received certificates of appreciation for their participation. After the workshops were conducted, students were surveyed. In addition to general data (related to gender, age, previous experience participating in similar workshops, grades in mathematics, physics, and computer science), students expressed their degree of agreement with statements on a five-point Likert scale. The statements related to whether students were engaged in workshop activities, whether they were motivated, their educational significance, the level of their collaboration and peer learning, and whether they believed that participating in the workshop increased their interest in mathematics, physics, and computer science.

The data collected through the student survey were compiled into a unified database. After that, the data were analysed using machine learning methods and techniques in a Python environment.

Participants

The data set has 208 data tuples. The research sample consists of 101 students from the last two grades of primary school, aged 14 and 15, and 107 students from the first two grades of high school, aged 16 and 17. Younger students participated in workshops within the project "Treasure Hunt" while older students participated in workshops within the project "Mission (Im)possible." Workshops for both projects were conducted in Kragujevac (central Serbia) and Užice (western Serbia) in autumn 2022, in collaboration with the Center for the Promotion of Science of the Republic of Serbia.

Regarding the gender representation of students, the sample consists of 105 female students and 103 male students. Students achieved different grades in mathematics (4 students with a grade 1, 46 with a grade 2, 38 with a grade 3, 55 with a grade 4, and 65 with a grade 5), physics (2 students with a grade 1, 30 with a grade 2, 35 with a grade 3, 60 with a grade 4, and 81 with a grade 5), and computer science (15 students with a grade 2, 11 with a grade 3, 97 with a grade 4, and 85 with a grade 5). It refers to the current grades that students had in the specified subjects at the time when the STEM workshops were conducted. The distribution of grades among the sampled students closely aligns with the distribution of grades in the specified subjects at the national level.

Concerning whether students had previous experience participating in similar workshops, 122 students stated they had no prior experience, while 86 students reported having some similar experience in their formal or informal education. This data corresponds to the current situation regarding interdisciplinary connections between the content of various subjects.

Aim of the Study

The aim of the study is to predict the level of students' agreement that their active participation in the described STEM workshops influenced their increased interest in studying subjects covered the most in these workshops as part of their formal education — mathematics, physics, and computer science.

To gain insight into which predictors influence students' perceptions of increased interest in learning mathematics, physics, and computer science, machine learning methods and techniques were applied.

FINDINGS

Classification model requires training data for model learning, and unseen test data for measuring its performance in real situations. For the task of choosing the best classification model, two important factors must be considered. First, stochastic-based models will be used, and because the results may vary for different random seeds, they must be tested for more than one time, and consider the average value and standard deviation. Second, the data set is small, which must be exploited maximally. Because of these two reasons, the Cross-Validation K-Fold method was chosen for measuring the performance of selected models.

As far as models are concerned, there was an attempt to cover a wide range of available and well-implemented classifiers. The following models were selected: Logistic Regression (aka logit, MaxEnt) Classifier (LRC) (Yu et al., 2011), K-Nearest Neighbors Classifier (KNN) (Cover & Hart, 1967), C-Support Vector Classification with linear (SVMl) and rbf kernel (SVMr) (Chang & Lin, 2011), Gaussian process classification based on Laplace approximation (GPC) (Seeger, 2004), Decision Tree Classifier (DTC) (Breiman et al., 2017), Random Forest Classifier (RFC) (Breiman, 2001), Multi-layer Perceptron classifier (MLP) (Glorot & Bengio, 2010), AdaBoost Classifier (ABC) (Zhu et al., 2009), Gaussian Naive Bayes (GNB) (Chan et al., 1982), Quadratic Discriminant Analysis (QDA) (Tharwat, 2016), Regularized Linear Model with Stochastic Gradient Descent Classifier (SGD) (Bottou, 2012), Passive Aggressive Classifier (PAC) (Crammer et al., 2006), Classifier using Ridge regression (RDC) (He et al., 2014), Naive Bayes Classifier for Multinomial Models (MNB) (Manning, 2009), Complement Naive Bayes Classifier (CNB) (Rennie et al., 2003), Naive Bayes Classifier for Multivariate Bernoulli Models (BNB) (Manning, 2009), Naive Bayes Classifier for Categorical Features (FNB) (Hsu et al., 2008). All selected models are implemented in the scikit-learn 1.2.0 library (Pedregosa et al., 2011). Default parameter values were used for all applied classifiers.

Results for Raw Data

The first experiment results are represented in Table 1. Cross-validation with 10 folds is used, and accuracy is in percent for test sets. The fields in the table represent the mean value for those 10 folds, with the standard deviation displayed. The bolded field represents the highest scores.

The first experiment (Table 1) shows a big variance in results for different classification models, as well as different standard deviation values.

Table 1
Accuracy for selected classifiers for raw data set

Classifier	Mean accuracy[1]	σ accuracy[2]	Classifier	Mean accuracy	σ accuracy
LRC	0.75	0.08	KNN	0.63	0.13
SVM_l	0.78	0.07	SVMr	0.76	0.09
GPC	**0.89**	0.12	DTC	**0.88**	0.10
RFC	**0.92**	0.09	MLP	0.67	0.15
ABC	**0.80**	0.05	GNB	0.68	0.09
QDA	**0.88**	0.07	SGD	0.69	0.10
PAC	0.61	0.10	RDC	0.70	0.06
MNB	0.63	0.13	CNB	0.61	0.10
BNB	0.48	0.09	FNB	0.71	0.06

[1] Mean accuracy value for 10 cross-validation folds.
[2] Standard accuracy deviation for 10 cross-validation folds.

The classifier with the highest mean accuracy is RFC with a value of 0.92 and a standard deviation of 0.09. Models with 0.8 accuracy and higher, sorted in nonincreasing order are: RFC, GPC, DTC, QDA and ABC, where RFC, DTC and ABC are based on Decision Tree models, and GPC and QDA are based on Gaussian Distribution. The most stable model is ABC with a standard deviation of 0.05.

Results for Data Prepared With Outlier Detectors

An attempt to further improve the results was to apply automatic outlier-removing methods to raw data. The selected outlier removing methods were: IsolationForest (IF) (Liu et al., 2012), EllipticEnvelope (EE) (Rousseeuw & Driessen, 1999), LocalOutlierFactor (LO) (Breunig et al., 2000) and OneClassSVM (OC) (Li et al., 2003), all from scikit-learn 1.2.0 library. Default parameter values were used for all applied automatic outlier removing methods except the amount of contamination of the data set for IF, EE and LO, and the upper bound on the fraction of training errors and a lower bound of the fraction of support vectors for OC which is set to 0.05.

Sizes of reduced data sets are shown in Table 2. Results of the second experiment are represented in Table 3 for the raw data with applied automatic outlier removers. From the second experiment it can be concluded that applying automatic outlier-removing methods to raw data has been beneficial for the performance of some classification models, but some combinations of outlier detectors and classifiers can have a negative impact. As a result, it is difficult to make general conclusions.

Table 2
Data set sizes after applied different outlier detectors

Outlier detector	Data set size
/	208
IF	197
EE	197
LO	197
OC	194

Note: Parameters for the amount of data to be removed are the same (0.05), but they are just indicators, and the size of the resulting data set may vary.

However, the new top-performing model is DTC with 0.94 accuracy and standard deviation of 0.06 with the IF automatic outlier removing technique (Table 3).

Table 3
Accuracy for selected classifiers and applied outlier detectors on raw data set

Classifier	Outlier detector	*Mean accuracy*[3]	*σ accuracy*[4]
RFC	IF	0.93	0.06
	EE	0.92	0.08
	LO	0.92	0.08
	OC	0.92	0.08
GPC	IF	0.91	0.09
	EE	0.90	0.10
	LO	0.90	0.10
	OC	0.90	0.10
DTC	IF	**0.94**	0.06
	EE	0.90	0.08
	LO	0.89	0.09
	OC	0.92	0.07
QDA	IF	0.86	0.07
	EE	0.88	0.08
	LO	0.87	0.09
	OC	0.83	0.10
ABC	IF	0.76	0.09
	EE	0.76	0.09
	LO	0.78	0.06
	OC	0.75	0.07

Also, RFC and GPC have improved results with all applied outlier detectors, but QDA and ABC have the same or worse results in all cases.

Results of the Decision Tree Classifier

Based on the previously obtained results, the Decision Tree Classifier (DTC) combined with the Isolation Forest (IF) outlier removal method has the highest Mean Accuracy value for 10 cross-validation folds. Therefore, we will present the results obtained using this machine learning technique.

First, it was observed that no student expressed either absolute disagreement or partial disagreement with the statement that participation in the STEM workshops "Treasure Hunt" and "Mission (Im)possible" positively influenced their interest in learning mathematics, physics, and computer science. At the top of the decision tree, when dividing students based on the dependent variable, the first highlighted variable is i3. This variable represents the level of agreement among upper primary and lower high-school students with the statement: *I was interested in actively participating during the workshop*. Thus, increasing students' interest and active participation in STEM workshop positively impacts their willingness to further explore mathematics, physics, and computer science as part of their formal education (Figure 1). Again, for students who were undecided or partially agreed with the level of agreement on the dependent

[3] Mean accuracy value for 10 cross-validation folds.
[4] Standard accuracy deviation for 10 cross-validation folds.

variable, their classification depends on whether they partially agreed, were undecided, or disagreed with the statement that they were interested in actively participating in the workshop. For students who fully agreed with the statement of the dependent variable, the value of the dependent variable is positively correlated with statement i6, which concerns the educational significance of the workshop: *I acquired new knowledge in mathematics, physics, and computer science during the workshop*.

Furthermore, in the hierarchy, for students who were undecided about whether their interest in learning individual subjects increased after the workshops, it is important to consider: which workshop they attended; whether they have prior experience participating in similar STEM workshops; and their level of agreement with statement i14 - *I believe that the activities I participated in during the workshop are important for my education*.

Whether students who moderately agree with the statement that is considered the dependent variable will remain in this category or shift into one of the other two categories (being undecided or fully agreeing with the statement) is most influenced by: their grade in physics, their level of agreement with statement i5 – *The workshop was fun,* and their level of agreement with statement i11 – *I believe that collaboration during group work is extremely important for the successful implementation of workshops of this type.*

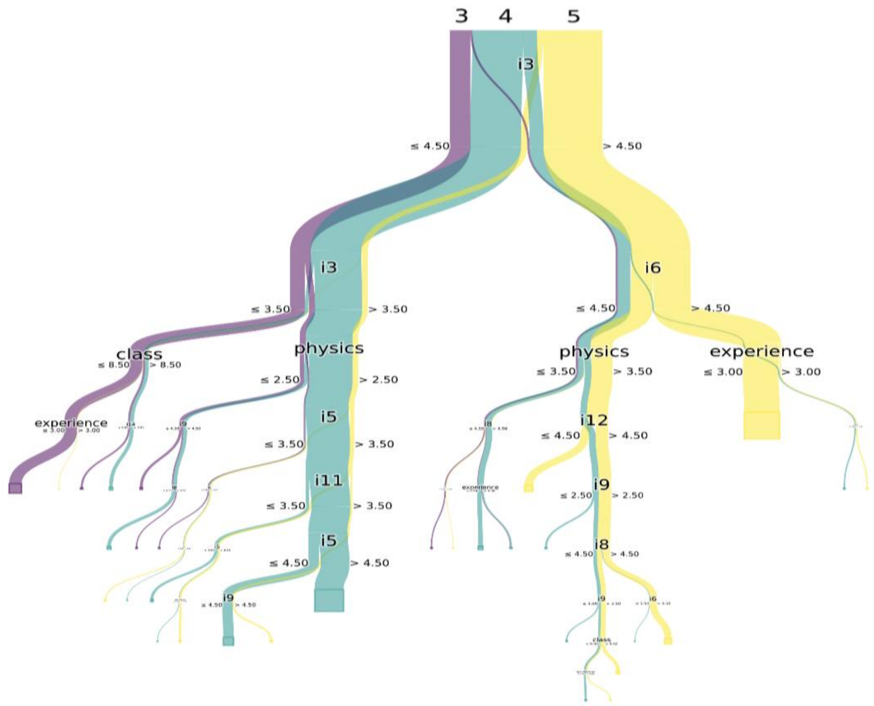

Figure 1
Resulting decision tree model

In the group of students who mostly fully agree with the statement that their participation in STEM workshops increased their interest in learning mathematics, physics, and computer science, significant influencing factors include their grade in physics, prior experience in participating in STEM workshops, and their level of agreement with the following statements i12 – *During the workshop, I was fully engaged in the activities*; i9 – *I felt connected to my peers in the group during the workshop*; i8 – *I believe that I can learn more in this way compared to traditional teaching methods* (Figure 1).

Interestingly, among the most significant predictors of increased interest in learning mathematics, physics, and computer science in more detail after participating in STEM workshops, the following factors do not appear: gender, average grade in mathematics, average grade in computer science, students' level of agreement with whether they would willingly participate in a STEM workshop again, whether they believe the instructional content was appropriate for their age group, and whether they felt capable of completing all the activities expected of them.

DISCUSSION AND CONCLUSION

With this research, we aimed to shed light on the factors that influence the extent to which students' interest in learning physics, mathematics, and computer science increases after actively participating in STEM workshops that integrate content from these subjects. First, it must be highlighted that, in general, students responded at a high level, stating that the practical, interdisciplinary presentation of content—through collaboration in an environment that includes elements of both play and competition—positively influenced their interest in studying mathematics, physics, and computer science within individual subjects seriously. Practically speaking, no student responded negatively to this question, with only a small number giving neutral responses, and the majority either partially or fully agreeing with this statement. An earlier study showed that students generally have a positive attitude towards STEM activities and that STEM activities have a significantly positive influence on students' attitudes towards and interest in science (Şimşek, 2019). Like our results, the findings of the study conducted by Mohd Shahali et al. (2017) show that exposing early secondary-school students to integrated STEM education has a positive impact on their level of interest in STEM subjects and related careers. Previous research also suggests that STEM activities can be implemented to improve students' science process skills, motivation, and attitudes about STEM education (Hiğde & Aktamış, 2022). The conclusions of a meta-analysis of 26 studies investigating the effects of integrating STEM into science and physics education on students' attitudes toward science show that the application of STEM approaches has a strong influence on students' attitudes towards learning individual STEM subjects (Hikmah et al., 2025). Furthermore, the authors emphasise that STEM integration has the greatest impact in terms of educational attainment when applied in primary and secondary school — the very age group targeted by the STEM workshops in this study.

Regarding the application of machine learning methods, our findings indicate that the decision tree with the IF automatic outlier removing technique proved to be the most reliable, and therefore, its results were interpreted. Machine learning methods have also proven to be reliable in recently conducted studies on STEM education (Abdrakhmanov et al., 2024; Göktepe Körpeoğlu & Göktepe Yıldız, 2024). In sixteen studies included in the systematic literature review by Ismail and Yusof (2023), the decision tree technique was used in as many as six studies, while in two of them, the decision tree achieved the best results.

Among the factors whose influence was examined on increasing students' interest in studying the STEM subjects in greater depth, the most prominent was the students' individual level of interest in actively participating in STEM workshops. In achieving outcomes such as students' active engagement, interest, and high achievement in STEM fields, a high level of motivation and peer collaboration are crucial (Fiorella et al., 2021; Salsa et al., 2022). Saleh et al. (2019) highlight student interest as one of the important factors that later influences career choice in STEM fields, while Franks and Capraro (2019) identify the lack of interest in pursuing STEM education as a result of students not envisioning themselves as scientists. Mateos-Núñez et al. (2020) found that more than 99% of students who participated in STEM workshops expressed that they liked the workshop they attended, and the positive emotions they displayed after the STEM workshops (fun, curiosity, satisfaction, confidence).

The next factor was the educational character of the workshop, that is, students' perception that they were acquiring new knowledge in the subjects supposed to be presented to them in a more engaging manner. The results of the study by Timur et al. (2020) confirm that students understood the content in the fields of mathematics, science, and computer science better when it was presented through STEM workshops. Furthermore, the conclusion of the same study aligns with our findings that satisfaction with attending STEM workshops does not depend on gender.

The next most influential factor is students' age (grade), which, in this specific study, can also be linked to the nature of the workshops themselves. Namely, older students exhibit a greater increase in interest in studying individual subjects, and it is worth noting that the workshops designed for them placed a stronger emphasis on programming and applying technological devices. In various studies, the impact of students' age on their interest in STEM subjects varies. For example, in the research on STEM interests conducted by Japashov et al. (2022), it is shown that 11th-grade students demonstrate greater interest in studying mathematics compared to 8th- and 9th-grade students, while they do not show increased interest in studying natural sciences and technology. In contrast, Koyunlu Ünlü and Dökme (2020) indicate that the interest of students from different grades differs only in learning natural sciences, but not in mathematics or other STEM subjects.

When it comes to academic achievement, students' success and knowledge of physics stood out the most, as higher achievement levels were associated with a greater increase in interest in studying individual school subjects. Japashov et al. (2022) point out a strong positive correlation between students' interest in science and mathematics and their achievements in physics. The results of the study conducted by Franco (2025) also indicate a significant positive correlation between students' interest and motivation for studying and their performance in mathematics. However, in our results, the mathematics grade does not appear among the most important factors when analyzing the independent variables. Additionally, Koyunlu Ünlü and Dökme (2020) demonstrate that there is a statistically significant connection between students' grades and their interest in science and mathematics, but that students' grades do not influence their interest in learning technology and engineering. Among the independent factors, prior experience also played a role—students who had previously participated in similar workshops showed a greater increase in interest in mathematics, physics, and computer science. The positive impact of STEM workshops can be linked to the results of a previous study (Mohr-Schroeder et al., 2014) in which students who had hands-on experiences at summer STEM camps expressed a greater interest in STEM fields after completing the camp compared to their responses before the camp began.

Based on the results of our study, we can conclude that in educational systems in which STEM subjects are studied separately, with little interdisciplinary connection, there should be a greater emphasis on providing students with opportunities to participate in STEM workshops that integrate different scientific disciplines, whether within formal or informal education. In planning and implementing such STEM workshops, educators should ensure that they are engaging and that students are motivated to participate actively, with special attention given to their educational value—that is, students should have a clear sense that they are acquiring concrete knowledge from various subjects during the workshops. Additionally, the frequency of such workshops and students' experience in collaborating with peers in STEM activities should further enhance their interest in studying mathematics, physics, and computer science.

LIMITATIONS

This study has its limitations. The limitation in the number of participants must be highlighted first. A similar analysis should be conducted with a larger number of students from different educational systems to allow for the generalization of our findings. Additionally, this study focused on a STEM workshop emphasizing content from physics, mathematics, and computer science. The topics from other natural sciences, such as biology and chemistry, were not explored. In addition, the impact of these STEM workshops on individual subjects was not examined. In future, STEM workshops that integrate knowledge from mathematics, multiple natural sciences (physics, biology, chemistry), and computer science, with elements of engineering and the appropriate use of technology should be designed. It should be examined which aspects of these carefully planned STEM workshops contribute to increasing students' interest in studying individual STEM school subjects via machine learning approach.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia through the Agreements No. 451-03-65/2024-03/200122 and No. 451-03-65/2024-03/200140.

REFERENCES

Abdrakhmanov, R., Zhaxanova, A., Karatayeva, M., Niyazova, G. Z., Berkimbayev, K., & Tuimebayev, A. (2024). Development of a Framework for Predicting Students' Academic Performance in STEM Education using Machine Learning Methods. *International Journal of Advanced Computer Science & Applications*, *15*(1).

Aguilera, D., & Ortiz-Revilla, J. (2021). STEM vs. STEAM Education and Student Creativity: A Systematic Literature Review. *Education Sciences*, *11*(7), 331.

Bottou, L. (2012). Stochastic gradient descent tricks. In *Neural networks: tricks of the trade: second edition* (pp. 421-436). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bozkurt Altan, E., & Tan, S. (2021). Concepts of creativity in design based learning in STEM education. *International Journal of Technology and Design Education*, *31*(3), 503–529. https://doi.org/10.1007/s10798-020-09569-y

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93-104).

Campbell, S. L., & Gear, C. W. (1995). The index of general nonlinear DAEs. *Numerische Mathematik*, *72*(2), 173-196.

Chan, T. F., Golub, G. H., & LeVeque, R. J. (1982). Updating formulae and a pairwise algorithm for computing sample variances. In *COMPSTAT 1982 5th Symposium held at Toulouse 1982: Part I: Proceedings in Computational Statistics* (pp. 30-41). Physica-Verlag HD.

Chang, C. N., Lin, S., Kwok, O. M., & Saw, G. K. (2023). Predicting STEM major choice: A machine learning classification and regression tree approach. *Journal for STEM Education Research*, *6*(2), 358-374.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 1-27.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory, 13*(1), 21-27.

Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., & Singer, Y. (2006). Online passive-aggressive algorithms. *Journal of Machine Learning Research*, *7*(Mar), 551-585.

Dávila-Acedo M. A., Airado-Rodríguez, D, Cañada-Cañada, F., & Sánchez-Martín, J. (2021). Detailed Emotional Profile of Secondary Education Students Toward Learning Physics and Chemistry. *Front. Psychol.*, *12*, 659009.

Deák, C., Kumar, B., Szabó, I., Nagy, G., & Szentesi, S. (2021). Evolution of new approaches in pedagogy and STEM with inquiry-based learning and post-pandemic scenarios. *Education Sciences*, *11*(7), 319. https://doi.org/10.3390/educsci11070319

Djudin, T. (2023). Transferring of Mathematics Knowledge into the Physics Learning to Promote Students' Problem-Solving Skills. *International Journal of Instruction*, *16*(4).

English, L. D. (2017). Advancing Elementary and Middle School STEM Education. *International Journal of Science and Mathematics Education*, *15*(S1), *5-24.*

Faber, M., Unfried, A., Wiebe, E. N., Corn, J., Townsend, L. W., & Collins, T. L. (2013). Student attitudes toward STEM: The development of upper elementary school and middle/high school student surveys. *In the Proceedings of the 120th American Society of Engineering Education Conference* (pp.1-26).

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine, 17*(3), 37-37.

Fiorella, L., Yoon, S. Y., Atit, K., Power, J. R., Panther, G., Sorby, S., Uttal, D. H., & Veurink, N. (2021). Validation of the mathematics motivation questionnaire (MMQ) for secondary school students. *IJ STEM Ed*, *8*, 52.

Franco, C. M. (2025). The influence of interest, motivation, and learning style on grade 6 pupils' mathematics performance. *International Journal of Instruction*, *18*(4), 249-268.

Franks, A. D., & Capraro, M. M. (2019). Motivated for STEM: Developing an Understanding of Highly Motivated Students' Self-Concept in STEM Education. *Science Educator*, *27*(1), 60-69.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). JMLR Workshop and Conference Proceedings.

Göktepe Körpeoğlu, S., & Göktepe Yıldız, S. (2024). Using artificial intelligence to predict students' STEM attitudes: an adaptive neural-network-based fuzzy logic model. *International Journal of Science Education*, *46*(10), 1001-1026.

Granovskiy, B. (2018). *Science, technology, engineering, and mathematics (STEM) education: An Overview*. (Report No. R45223). CRS Report for Congress.

Habig, B., & Gupta, P. (2021). Authentic STEM research, practices of science, and interest development in an informal science education program. *International Journal of STEM Education*, *8*(1), 1-18. https://doi.org/10.1186/s40594-021-00314-y

Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan Kaufmann.

He, J., Ding, L., Jiang, L., & Ma, L. (2014). Kernel ridge regression classification. In *2014 International Joint Conference on Neural Networks (IJCNN)* (pp. 2263-2267). IEEE.

Hiğde, E., & Aktamış, H. (2022). The effects of STEM activities on students' STEM career interests, motivation, science process skills, science achievement and views. *Thinking Skills and Creativity*, *43*, 101000. https://doi.org/10.1016/j.tsc.2022.101000

Hikmah, N., Siregar, I. H., Mufit, F., Lufri, L., & Andromeda, A. (2025). The Effect of STEM Integration in Science and Physics Learning on Students' Scientific Attitudes: A Meta-Analysis. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, *15*(1).

Hsu, C. C., Huang, Y. P., & Chang, K. W. (2008). Extended Naive Bayes classifier for mixed data. *Expert Systems with Applications*, *35*(3), 1080-1083.

Ing, M. (2014), Gender differences in the influence of early perceived parental support on student mathematics and science achievement and STEM career attainment. *International Journal of Science and Mathematics Education, 12*, 1221-1239. https://doi.org/10.1007/s10763-013-9447-3

Ismail, N., & Yusof, U. K. (2023). A systematic literature review: Recent techniques of predicting STEM stream students. *Computers and Education: Artificial Intelligence*, *5*, 100141.

Japashov, N., Naushabekov, Z., Ongarbayev, S., Postiglione, A., & Balta N. (2022). STEM career interest of Kazakhstani middle and high school students. Education Sciences, 12, 397-414. https://doi.org/10.3390/educsci12060397

Jeffries, D., Curtis, D. D., & Conner, L. N. (2020). Student factors influencing STEM subject choice in year 12: A structural equation model using PISA/LSAY data. *International Journal of Science and Mathematics Education*, *18*, 441-461.

Kanadli, S. (2019). A Meta-Summary of Qualitative Findings about STEM Education. *International Journal of Instruction*, *12*(1), 959-976. https://doi.org/10.29333/iji.2019.12162a

Koyunlu Ünlü, Z., & Dökme, İ. (2020). Multivariate assessment of middle school students' interest in STEM career: A profile from Turkey. *Research in Science Education*, *50*(3), 1217-1231.

Li, K. L., Huang, H. K., Tian, S. F., & Xu, W. (2003). Improving one-class SVM for anomaly detection. In *Proceedings of the 2003 international conference on machine learning and cybernetics (IEEE Cat. No. 03EX693)* (Vol. 5, pp. 3077-3081). IEEE.

Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R news*, *2*(3), 18-22.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *6*(1), 1-39.

Manning, C. D. (2009). *An introduction to information retrieval*.

Marbà-Tallada, A., & Márquez, C. (2010). What do students say about their science class? A cross-study from the sixth grade (primary school) to the fourth grade (secondary school). *Enseñanza de las Ciencias*, *28*(1), 19-30.

Martínez-Borreguero, G., Mateos-Núñez, M., & Luis Naranjo-Correa, F. (2020). Implementation and Didactic Validation of STEM Experiences in Primary Education: Analysis of the Cognitive and Affective Dimension. In *Theorizing STEM Education in the 21st Century*. IntechOpen.

Martínez-Borreguero, G., Maestre-Jiménez, J., Mateos-Núñez, M., & Naranjo-Correa, F. L. (2020). Analysis of Environmental Awareness, Emotions and Level of Self-Efficacy of Teachers in Training within the Framework of Waste for the Achievement of Sustainable Development. *Sustainability*, *12*(6), 2563.

Mateos-Núñez, M. & Martínez-Borreguero, G., Naranjo-Correa, F. L. (2020). Learning Science in Primary Education with STEM Workshops: Analysis of Teaching Effectiveness from a Cognitive and Emotional Perspective. *Sustainability*, *12*, 3095.

Mateos-Núñez, M. & Martínez-Borreguero, G. (2023). Design and Implementation of Practical Workshops for Teaching STEM Content: Analysis of Cognitive and Emotional Variables. In B. Cavadas & N. Branco (Eds.), *Handbook of Research on Interdisciplinarity Between Science and Mathematics in Education* (pp. 131-148). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-6684-5765-8.ch006

McDonald, C. V. (2016). STEM Education: A review of the contribution of the disciplines of science, technology, engineering and mathematics. *Science Education International*, *27*(4), 530-569.

Milenković, A., & Momčilović, N. (2025). Interdisciplinary Connection between Mathematics and Physics in a Grammar School for Gifted Mathematics Students. *Mathematics and Informatics*, *LXVIII* (2). https://doi.org/10.53656/math2025-2-5-icb

Mohd Shahali, E. H., Halim, L., Rasul, M. S., Osman, K., & Zulkifeli, M. A. (2017). STEM Learning through Engineering Design: Impact on Middle Secondary Students' Interest towards STEM. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(5), 1189-1211. https://doi.org/10.12973/eurasia.2017.00667a

Mohr-Schroeder, M. J., Jackson, C., Miller, M., Walcott, B., Little, D. L., Speler, L., ... & Schroeder, D. C. (2014). Developing Middle School Students' Interests in STEM via Summer Learning Experiences: See Blue STEM Camp. *School Science and Mathematics*, *114*(6), 291-301.

Mohr-Schroeder, M. J., Cavalcanti, M., & Blyman, K. (2015). Stem Education: Understanding the Changing Landscape. In: Sahin, A. (eds.) *A Practice-based Model of STEM Teaching*. SensePublishers, Rotterdam.

Morgan, A., Smaldone, D., Selin, S., Deng, J., & Holmes, M. (2022). Adding relevancy to STEM interest through adventure education: A mixed methods study. *Interdisciplinary Journal of Environmental and Science Education*, *18*(4), e2294.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Quinlan, J.R.(1986). Induction of decision trees. *Machine learning 1*, 81-106.

Razali, F. (2021). Exploring Crucial Factors of an Interest in STEM Career Model among Secondary School Students. *International Journal of Instruction*, *14*(2), 385-404.

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 616-623).

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, *3*(1), 12-27.

Roungos, G., Kalloniatis, C., & Matsinos, V. (2020). STEAM education in Europe and the PISA test. *Scientific Educational Journal*, *8*(3), 177-187.

Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*(3), 212-223.

Saleh, S., Ashari, Z. M., Kosnin, A. M., & Rahmani, A. S. (2019). A systematic literature review on the roles of interest and motivation in STEM education. 2019 In *Proceedings of the IEEE International Conference on Engineering, Technology and Education,* Yogyakarta, Indonesia (pp. 1-6).

Salsa, F. J., Sari, R. T., Muhar, N., & Gusmaweti, G. (2022). The relationship between motivation and learning outcomes of biology subject through distance learning. *International Journal of STEM Education for Sustainability*, *2*(2), 140-147.

Sang, W., & Simpson, A. (2019). The Maker Movement: A Global Movement for Educational Change. *Int J of Sci and Math Educ*, *17*(Suppl 1), 65-83.

Seeger, M. (2004). Gaussian processes for machine learning. *International journal of neural systems, 14*(2), 69-106.

Sirajudin, N., Suratno, J., & Pamuti. (2021). Developing creativity through STEM education. *Journal of Physics: Conference Series*, *1806*(1), 012211.

Şimşek, F. (2019). Effect of STEM activities on students' scientific process skills, science interest, attitude and student opinions. *Turkish Journal of Computer and Mathematics Education*, *10*(3), 654-679.

Sudarsono., Kartono., Mulyono., & Mariani, S. (2022). The effect of STEM model based on Bima's local cultural on problem solving ability. *International Journal of Instruction*, *15*(2), 83-96. https://doi.org/10.29333/iji.2022.1525a

Tawbush, R. L., Stanley, S. D., Campbell, T. G., & Webb, M. A. (2020). International comparison of K-12 STEM teaching practices. *Journal of Research in Innovative Teaching & Learning*, *13*(1), 115-128. https://doi.org/10.1108/JRIT-01-2020-0004

Tharwat, A. (2016). Linear vs. quadratic discriminant analysis classifier: a tutorial. *International Journal of Applied Pattern Recognition*, *3*(2), 145-180.

Timur, S., Timur, B., Yalçınkaya-Önder, E., & Küçük, D. (2020). Attitudes of the students attending out-of-school STEM workshops towards STEM education. *Journal of Theoretical Educational Science*, *13*(2), 334-351.

Tsupros, N., Kohler, R., & Hallinen, J. (2009). *STEM education: A project to identify the missing components*. Pennsylvania: Intermediate Unit 1: Center for STEM Education and Leonard Gelfand Center for Service Learning and Outreach, Carnegie Mellon University.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & Data, M. (2005). Practical machine learning tools and techniques. In *Data mining 2*(4), 403-413. Amsterdam, The Netherlands: Elsevier.

Yu, H. F., Huang, F. L., & Lin, C. J. (2011). Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, *85*, 41-75.

Zhu, J., Zou, H., Rosset, S., & Hastie, T. (2009). Multi-class adaboost. *Statistics and its Interface*, *2*(3), 349-360.