A STEM Lesson in Informatics: The Importance of Technological Environment

Ivelina Stefanova
Secondary School "Neofit Rilski" – Dolna Banya, Dolna Banya, Bulgaria
ivelina.bankina@gmail.com

Malinka Ivanova
Technical University of Sofia, Sofia, Bulgaria
m_ivanova@tu-sofia.bg

Galina Bogdanova
Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, Sofia, Bulgaria
galina@math.bas.bg

STEM урок по информатика: Значение на технологичната среда

Abstract

STEM education (Science, Technology, Engineering and Mathematics) supports the acquisition of knowledge and skills by involving students in activities to explore events and processes to solve real-world problems. The teacher has an important role related to the design and implementation of lessons that include appropriate pedagogical scenarios and a contemporary technological environment, through which the set learning goals can be successfully achieved. In the field of informatics, a number of challenging problems can be addressed, aimed at data analysis, the use of artificial intelligence in a transparent and appropriate manner, and the application of machine learning algorithms for prediction or analysis. The aim of the paper is to present a design and implementation of a STEM lesson, through which students are introduced to the application of machine learning in creating, using and evaluating a predictive model. The lesson provides an opportunity to develop specific skills in programming, machine learning and data analysis, as well as soft skills, such as critical thinking, creativity and teamwork.

Keywords: *STEM Education; STEM Lesson; Machine Learning; Informatics; Programming; Technology-enhanced Learning Environment.*

Introduction

In recent years, there has been a growing emphasis an increasing focus on the potential and benefits of STEM (Science, Technology, Engineering and Mathematics) education and its effective integration across a wide variety of academic disciplines and at different educational levels [1], [2]. Recently, there has been increasing discussion about how machine learning models (ML) and artificial intelligence (AI) can be used in the STEM educational process to support and improve the required technological environment. According to McDaniel [3], the integration of technology, hands-on and AI tools into STEM education is crucial for improving student engagement and concept understanding. Using such technologies in STEM lessons contributes to gaining rich learning experience in a data-driven and personalized learning environment. AI-powered tools are the connection of theoretical concepts and hands-on applications, encouraging critical thinking and problem-solving skills. In an overview technologies for STEM education are discussed, revealing the importance of ML [4]. Machine

learning-based educational platforms and AI tools (virtual assistants, chatbots) support teachers and learners in such a way to reduce stress and increase the effectiveness of learning. The author concludes that all participants in an educational process must become familiar with ML, and this could happen through workshops. Joseph and Uzondu explore the main challenges and opportunities in integrating AI and ML in STEM education as the benefits and drawbacks are shown [5]. Among advantages are improvement of students' engagement and overall educational outcomes, personalized learning experiences, and possibilities to solve critical problems in STEM fields. However, it is important to consider privacy concerns, ethical dilemmas, the shortage of trained educators, and the lack of equipment and devices. To address the shortage of trained educators who can successfully integrate ML into STEM education, Tang et al. have created a professional development program called ML4STEM [6]. This program was tested and evaluated for effectiveness with STEM teachers and the results show that ML4STEM successfully develops understanding and skills regarding ML utilization in STEM education, as well as contributes to increased interest among middle and high school teachers in applying ML as a pedagogical tool.

All these examples point out that AI and ML are increasingly important in STEM education, and in the future, they are expected to become independent subjects. Organized educational environments are technology-based to develop skills and competencies in students that are of great importance for their future development and professional realization. The technological ecosystem that surrounds all participants in an educational process imposes greater demands on learners, such as knowledge and skills in various fields. To be successful, they must be innovative, creative, and technologically prepared. Changes in the Information Technology curriculum must address this need.

The aim of the paper is to present the meaning of technology-enhanced educational environment for implementation of a STEM lesson, through which students are introduced to the application of ML in creating, using, and evaluating a predictive model. The lesson provides an opportunity to develop specific skills in programming, ML, data analysis, creating accurate predictive models, as well as soft skills, such as critical thinking, creativity, and teamwork.

APPLICATION OF MACHINE LEARNING IN A STEM LESSON

General information

In this section, it is demonstrated how ML concepts could be presented and practiced by students from 10th grade, general education, considering the school curriculum in the learning discipline Information Technologies. A STEM lesson is designed as a part of the topic: Programming and artificial intelligence.

The expected learning outcomes of this learning discipline are as follows:

- Students to know the basic properties of algorithms (definiteness; massiveness; finiteness and efficiency; discreteness; formality; complexity).
- To know ways to represent algorithms using pseudocode and a script language program.
- To create an algorithm of the main types, presented verbally using pseudocode and as a script language program.
- To indicate the result of a short algorithmic fragment described using pseudocode and a script language program.
- To find data sets from digital data repositories.
- To process a data set using a script language.

- To understand how AI algorithms work.
- To ask questions to AI tools (AI chatbots, assistants, search engines, etc.).
- To know that sensors in digital devices and applications generate large amounts of data that are used to train ML systems.
- To explain the process of machine learning.
- To distinguish between training and test data sets in the process of machine learning.

Learning scenario and environment of a STEM lesson

The main learning objectives of the STEM lesson are:

- The learners to understand what is the purpose of ML, what is the difference between algorithms from supervised ML and unsupervised ML and how to analyze data through ML.
- The learners to obtain skills for creating a predictive model using Google Colab environment, Python libraries and AI Microsoft Copilot.

For realization of the STEM lesson, a technological-based learning environment is organized that consists of:

- Existing STEM laboratory with 26 computers for all learners with high-speed internet connection, Google Chrome web browsers installed on the computers, installed Python, installed Microsoft Office Package.
- Available access to Google Colab and AI Microsoft Copilot.
- Used data sets with real data taken from meteorological site with temperatures in Sofia, Bulgaria for one year and three years' period.
- Created an interactive presentation that contains the new knowledge in the lesson considering the following questions: What is machine learning? What are the types of machine learning and which one do we use? What is the evaluation metrics of a predictive ML model? How can we use Microsoft Copilot to create code which will help us to achieve our learning goals? How to use Google Colab to train and test the ML model?
- Video clips with real examples of how machine learning is used for solving real life problems.

The designed learning scenario is instructor-led as the teacher has the leading role in the STEM lesson. The applied educational strategy is divided into two parts: the first one is related to the guidance given by the teacher and the second one is based on collaborative work of students.

The design created of a STEM lesson focused on practice with ML predictive model is presented in Table 1.

Table 1. Created design of a STEM lesson

Elements of lesson design	Description
Learning objectives	To obtain knowledge about ML and supervised ML algorithm Random Forest To obtain skills to create and evaluate ML predictive models
Roles	Teacher, learners, group of learners, supportive staff

Elements of lesson design	Description	
Learning scenario	Instructor-led with guided instructions by a teacher, collaborative learning in groups of two students, assessment of students' knowledge and skills	
Learning environment	Computer laboratory, materials of the teacher (presentation and video), installed software, datasets	
Activities	**Type of activity**	**Description**
	Read	The teacher presents new knowledge with interactive presentations.
	Watch	A video clip, showing how machine learning is used in real life.
	Practice	The students create their own predictive model, applying what they have learned.
	Reflect	The students analyze what they learned, what was useful to them, and what could be improved.
	Collaborate	The students work in collaboration and help each other.
	Assess	The students do practical exercise.

EXPERIMENTATIONS AND RESULTS

Prepared datasets

The data is taken from the website "Reliable Prognosis" (rp5.ru) after submitting two queries for weather in Sofia, Bulgaria. This meteorological site provides free access to an archive of meteorological values. The data should be processed and checked to extract only the necessary information, which is: date, minimum temperature, maximum temperature. The raw data gives the measured temperature for the relevant date by hours, as well as precipitation, atmospheric pressure, etc. The data is initially processed manually to simplify it and leave only the necessary ones, and then it is processed and grouped by AI Microsoft Copilot, which for each date must select and leave only the minimum and maximum temperatures and add the corresponding season. Here we do not exclude the human factor. It is good to do a random check to be sure of the accuracy of the data. Two datasets are prepared considering the meteorological information in Sofia, Bulgaria: for one year (2024) and for three years (2022-2024). The records in datasets are labeled in four classes: spring, summer, autumn, winter.

Applied machine learning algorithm Random Forest

Random Forest is a popular machine learning algorithm, which is used for solving classification and forecasting tasks in a variety of domains, including in education [7], [8]. It is based on the concept of creating a collection of decision trees and accumulating their outcomes to generate the final prediction. Random subsets of data construct each decision tree [9]. The final prediction is obtained by combining the results of all trees. Random Forest is one of the most efficient and powerful algorithms in machine learning. It's often utilization is explained with several advantages like:

- High accuracy in classification and prediction.

- Random Forest uses the technique of building trees on different and independent sub-models, which reduces the problem of overfitting and improves the ability to generalize.
- Works well on large and complex data.
- Provides an assessment of the importance of characteristics
- It is easily implemented using software libraries such as SciKit-learn in Python.

The main parameters of Random Forest that students must be introduced are summarized below.

- *model* = RandomForestClassifier.
- *n_estimators*=100 - Number of trees in the forest. If there are more trees in the forest, this implies greater accuracy, but also more training time.
- *max_depth*=10 - Maximum depth of each tree. Restricting the depth prevents overfitting.
- *min_samples_split*=5 - Minimum number of examples needed for a node split.
- *max_features*='*sqrt*' - Number of features analyzed in each split.
- *bootstrap=True* - Bootstrapping is used — each tree is trained on a random sample of the data with repetitions, increasing the diversity between trees.
- *random_state*=42 - Corrects randomness to get the same results every time.
- *n_jobs*=-1 - All available processor cores are used for parallel processing, which speeds up learning.

Python code and libraries

Python code is generated through usage of Microsoft Copilot in an interactive mode through questions and answers. The main task is to create model that predict the season considering different parameters of weather in Sofia, Bulgaria. Python libraries such as *pandas*, *sklearn* and *matplotlib* are used for data processing, model training through Random Forest, visualization of results and model evaluation (Table 2). The experimentation is conducted in Google Colab.

Table 2. Python libraries

Library	Purpose
pandas	For processing of data, and for data analysis
sklearn.ensemble.RandomForestClassifier	Random Forest model – creating and training
sklearn.model_selection.train_test_split	The data is split into training and testing parts
sklearn.preprocessing.LabelEncoder	Categorical values (seasons) are encoded into numbers
sklearn.metrics	For calculation of metrics like: accuracy, classification report, confusion matrix
matplotlib.pyplot	For visualization of decision tree and confusion matrix
sklearn.tree.plot_tree	For visualization of decision tree

Evaluation metrics

The quality of classification models is performed through standard for machine learning metrics considering parameters like: Accuracy, Precision, Recall, F1 (Table 3) [10], [11]. The meaning of True Positive predictions (TP), True Negative (TN), False Positive (FP) and False Negative (FN) is presented via Table 4.

Table 3. Evaluation metrics for machine learning models

Parameter	Description
Accuracy	How many of all predicted seasons are correct? $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
Precision	Of all predicted positive seasons, how many are actually positive? $$Precision = \frac{TP}{TP + FP}$$
Recall	Of all the real positive predictions of seasons, how many are correctly predicted? $$Recall = \frac{TP}{TP + FN}$$
F1	Harmonic mean between Precision and Recall. $$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Table 4. Description of TP, TN, FP, FN

Case	Description
TP (True Positive)	The model correctly predicts a positive class.
TN (True Negative)	The model correctly predicts a negative class.
FP (False Positive)	The model (false alarm) incorrectly predicts a positive class.
FN (False Negative)	The model (missed positive value) incorrectly predicts a negative class.

Results analysis

Experiments with Python code and datasets are performed in Google Colab with different parameters of Random Forest algorithm and the datasets in order to achieve maximum accuracy. The ratio of data for training and testing is 80%/20%.

In the first experiment with the above described parameters of Random Forest and 365 records in the dataset, the achieved model accuracy ≈ 64,86%, which is not enough. This is moderate accuracy, which means the model is not completely reliable and can be improved by: adding more features, parameters optimization and using more data. The confusion matrix shows correct and incorrect seasons predictions (Fig. 1). The assessment metrics of Summer have the highest values. Autumn has the lowest precision and recall.

Accuracy: 0,64864864

	precision	recall	f1-score	support
0	0.36	0.38	0.37	13
1	0.53	0.56	0.54	18
2	0.89	1.00	0.94	16
3	0.74	0.63	0.68	27
accuracy			0.65	74

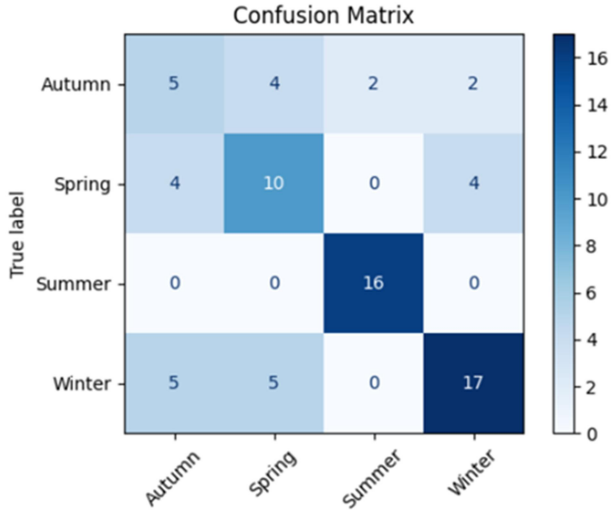

Fig. 1. Confusion matrix

Using AI Copilot for model improvement

The following activities are performed for model improvement

- The number of records is increased by using data for a three-year period (1097 records).
- Added new features to give the model more context about the seasons: mean_temp, range_temp, day_of_year.
- Used StandardScaler to make all numeric features with comparable scale. This is important for algorithms like Random Forest when combining different types of data.
- Performed Hyperparameter Optimization with GridSearchCV, with the AI Copilot testing 24 different Random Forest combinations to create maximum accuracy.

```
param_grid = {
'n_estimators': [100, 200],
'max_depth': [10, 20, None],
'min_samples_split': [2, 5],
'min_samples_leaf': [1, 2] …}
```

This allowed the model to find the best configuration instead of utilizing the first used values. The accuracy obtained of the model in the second experiment is almost 98%, which indicates accurate prediction of seasons as it is presented in Fig. 2.

Optimized Model Accuracy: 0.9773

	precision	recall	f1-score	support
Autumn	0.98	0.98	0.98	57
Spring	0.98	0.96	0.97	51
Summer	0.98	0.98	0.98	55
Winter	0.97	0.98	0.97	57
accuracy			0.98	220
macro avg	0.98	0.98	0.98	220
weighted avg	0.98	0.98	0.98	220

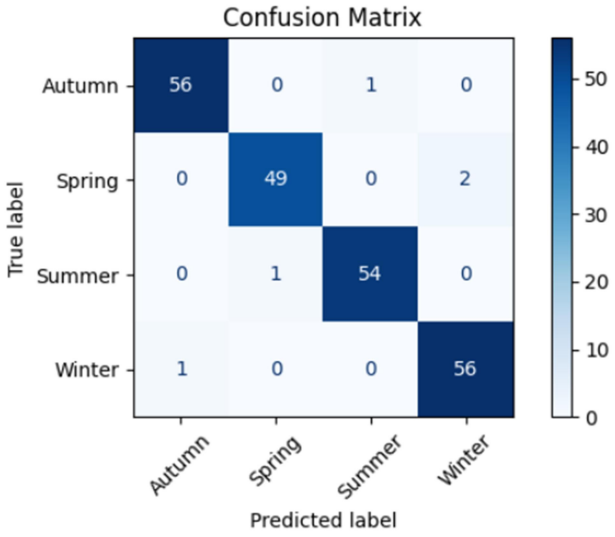

Fig. 2. Confusion matrix

CONCLUSION

The paper proposes an approach of integrating some concepts of machine learning in a STEM lesson in the discipline Information Technologies to introduce students how to prepare datasets and technological environment, how to choose parameters of Random Forest algorithm from supervised machine learning, how to evaluate the performance of the obtained models and how to interpret the results. Such initial experimentation could attract the attention of students to some scientific theories at solving real world problems.

It can be said that the technological environment is of great importance for achieving learning goals. Artificial intelligence and machine learning can be successfully adopted in STEM education and are expected to gain more attention among all participants in the educational process in the future. Also, different disciplines related to artificial intelligence and machine learning should be planned in the school curriculum. Contemporary educational environments must be technologically arranged in order to develop the necessary skills and competencies in students, which are of great importance for their future development and professional realization. The technological ecosystem that surrounds learners and educators, imposes greater requirements on them, such as knowledge and skills in many different areas.

REFERENCES

1. O. F. Marzuki et al., Innovating Education: A Comprehensive Review of STEM Education Approaches, International Journal of Academic Research in Progressive Education and Development, 13(1), 614-631, 2024. DOI: http://dx.doi.org/10.6007/IJARPED/v13-i1/20490
2. Y. Li, K. Wang, Y. Xiao, et al., Research and trends in STEM education: A systematic review of journal publications, IJ STEM Ed 7, 11, 2020. DOI: https://doi.org/10.1186/s40594-020-00207-6
3. A. M. McDaniel, Integrating Artificial Intelligence in STEM Education: Enhancing Learning Through Smart Technologies, American Journal of STEM Education: Issues and Perspectives, vol. 7, 1-10, 2025. DOI: https://doi.org/10.32674/6s0qz548
4. K. G. Fomunyam, Machine Learning and Stem Education: Challenges and Possibilities, International Journal of Difference Equations (IJDE), 17(2), 165-176, 2022. ISSN: 0973-6069, Available at: https://www.ripublication.com/ijde22/v17n2p02.pdf (last view: 18-08-2025)
5. O. B. Joseph and N. C. Uzondu, Integrating AI and Machine Learning in STEM education: Challenges and opportunities, Computer Science & IT Research Journal, 5(8), 1732-1750, 2024. DOI: https://doi.org/10.51594/csitrj.v5i8.1379
6. J. Tang, X. Zhou, X. Wan, et al., ML4STEM Professional Development Program: Enriching K-12 STEM Teaching with Machine Learning, International Journal of Artificial Intelligence in Education 33, 185-224, 2023. DOI: https://doi.org/10.1007/s40593-022-00292-4
7. M. Nachouki, E. A. Mohamed, R. Mehdi, M. A. Naaj, Student course grade prediction using the random forest algorithm: Analysis of predictors' importance, Trends in Neuroscience and Education, vol. 33, 2023. DOI: https://doi.org/10.1016/j.tine.2023.100214
8. S. A. A. Balabied, H. F. Eid, Utilizing random forest algorithm for early detection of academic underperformance in open learning environments, PeerJ Computer Science, 2023. DOI: https://doi.org/10.7717/peerj-cs.1708
9. H. A. Salman, A. Kalakech, A. Steiti, Random Forest Algorithm Overview, Babylonian Journal of Machine Learning, vol. 2024, 69-79, 2024. ISSN: 3006-5429, DOI: https://doi.org/10.58496/BJML/2024/007
10. N. S. Ahmed, M. Hikmat Sadiq, Clarify of the Random Forest Algorithm in an Educational Field, 2018 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq, 179-184, 2018. DOI: https://doi.org/10.1109/ICOASE.2018.8548804
11. Ž. Đ. Vujović, Classification Model Evaluation Metrics, (IJACSA) International Journal of Advanced Computer Science and Applications, 12(6), 2021. DOI: https://doi.org/10.14569/IJACSA.2021.0120670

Received: 23-08-2025 Accepted: 16-12-2025 Published: 29-12-2025

Cite as:
Stefanova, I., Ivanova, M., Bogdanova, G. (2025). "A STEM Lesson in Informatics: The Importance of Technological Environment", Science Series "Innovative STEM Education", volume 07, ISSN: 2683-1333, pp. 108-116, 2025. DOI: https://doi.org/10.55630/STEM.2025.0709