

## ПОДОБРЯВАНЕ ЕФЕКТИВНОСТТА НА УЧЕНЕ ЧРЕЗ УЧАСТИЕ В ОНЛАЙН СЕМИНАР: АНАЛИЗ ЧРЕЗ МАШИННО ОБУЧЕНИЕ

### LEARNING PERFORMANCE IMPROVEMENT THROUGH PARTICIPATION IN ONLINE SEMINAR: MACHINE LEARNING ANALYSIS

**Malinka Ivanova**

*Technical University of Sofia, Faculty of Applied Mathematics and Informatics,  
Department of Informatics,  
E-mail: m\_ivanova@tu-sofia.bg*

#### **Abstract**

*Learning performance is related to students' learning activities during a learning process. Their learning behavior could lead to successful course accomplishment or not, to better or worse final marks. Seminar practices have their impact on development of some students' competences like: topics analysis, discussion and presentation and the planned tasks concern learning performance. In online environment, the seminars could be organized in the form of different learning scenarios and it depends on the functional and technical features of the organized educational environment as well as on the course goal.*

*In this paper an investigation and analysis of students' participation in online seminars is conducted with aim to understand the dependence between their learning performance, online tasks realization and final results. eLearning informatics gives possibilities for usage contemporary methods for research and learning analytics as one of them is machine learning. Machine learning algorithms are utilized to group students according to their learning behavior and final outcome. The created analytical models could be in support of educators and students to improve their educational activities. The accuracy of machine learning algorithms is evaluated to find the best model according to collected data during one semester.*

**Keywords:** eLearning informatics; learning analytics; learning performance; online seminar; competences; machine learning.

#### **1. ВЪВЕДЕНИЕ**

Алгоритми от машинното обучение се прилагат към набори от данни, за да се извлече определено познание, да се открият зависимости или тенденции, чиито анализ да подпомогне решаването на конкретна задача или вземането на решение. Понастоящем, машинното обучение се използва и в областта на технологично-ориентираните образователни среди и в частност на системите за електронно обучение, тъй като те разполагат с разнообразни инструменти за събиране на информация, свързана с поведението и постиженията на студентите. Тази информация може да бъде допълнително обработена за целите на прогнозния анализ и получените резултати могат да посочат възможни подходи за подобряване или оптимизиране на учебните дейности.

Ефективността на учене се свързва с редица фактори, касаещи качеството на изпълнение на учебните дейности, времето, отделено за подготовка и изпълнение на тези задачи, както и получения краен резултат. Самото понятие *ефективност на учене* е

получило различни интерпретации в научната литература, като например често се определя с полученото знание и постигнатия успех за определен период от време съобразно предварително дефинирани учебни цели, но също така включва и социоекономически аспекти, допринасящи за развитие на компетентностния модел на студента и неговата бъдеща професионална реализация [1, 2, 3]. Сред факторите, оказващи влияние върху ефективността на учене могат да се посочат както личностите характеристики на студента, функционалността на образователната среда, така и интерфейлната връзка между тях. Когато учебен процес, който обикновено е комбиниран, т.е. провежда се в учебна зала и само част от дейностите са подпомогнати от електронно обучение, се пренесе изцяло онлайн, тогава са необходими и адаптирани версии на учебните сценарии в зависимост от функционалността на използваната учебна среда.

**Целта на доклада** е да представи подход за анализ на учебната ефективност на студентите при участието им в онлайн семинар чрез прилагане на алгоритми за машинно обучение, както и да представи модели за прогнозиране на учебната ефективност съобразно участието на студентите.

## **2. ПРИЛАГАНЕ НА МАШИННО ОБУЧЕНИЕ ПРИ ПРОГНОЗИРАНЕ НА УЧЕБНАТА ЕФЕКТИВНОСТ**

Прогнозирането на учебната ефективност е тема, към която понастоящем се отчита, че се проявява засилен интерес от страна на научната общност, тъй като задълбоченото изследване може да подпомогне преодоляването на различни предизвикателства, свързани с конкретен учебен процес или с образователната среда като цяло. Машинното обучение се оказва подходящ подход за намиране на проблемни фактори и за откриване на най-добрите решения. Затова, в тази секция е представен преглед и анализ на публикувани научни резултати, свързани с прилагане на машинното обучение за прогнозиране на учебната ефективност.

Ојајуни и др. [4] въз основа на исторически данни за постиженията на студентите, правят прогноза за тяхната учебна ефективност чрез прилагане на алгоритми от контролирано машинно обучение, като Random Forest, Decision Tree, Logistic Regression, Support Vector Machines, Deep Learning и други. Резултатите показват, че академичната ефективност е свързана и с редица демографски и социални фактори. Предложеният подход ще даде възможност на преподавателите да идентифицират пропуските при учене и да откриват на ранен етап студенти с лоша учебна ефективност.

Друга публикация с автори Rimadana и др. [5] представя създадени модели за прогнозиране на учебната ефективност на студентите чрез данни от анкети относно техните умения за разпределяне на времето за изпълнение на учебни дейности. Резултатите показват, че прогнозният модел, създаден чрез алгоритъма Support Vector Machine с 80% точност може да прогнозира учебната ефективност.

Alharthi [6] прогнозира учебната ефективност на студенти, вземайки под внимание данните от студентската информационна система, което включва и успеха от средното образование и предварителните изпити. Използвани са алгоритмите Random Forest,

Decision Tree, Naïve Bayes, K-nearest neighbors, Support Vector Machine, като с най-висока точност е установено да работи Random Forest.

Naren и др. [7] изследват положителното и отрицателното влияние на Интернет върху учебната ефективност на студентите, като се съобразяват с данни от log файлове и поведението в Интернет. Към данните са приложени алгоритмите Decision Tree, Support Vector Machine и Neural Networks, като студентите са групирани според тяхната учебна ефективност и получените оценки. Авторите твърдят, че предложеният подход е с висока практическа стойност за подобряване на уменията на студентите, свързани с управление на изпълняваните учебни дейности.

Kumar и Sharma [8] с цел да подобрят учебен процес, изследват и прогнозируют учебната ефективност на чуждестранни студенти чрез машинно обучение, като използват алгоритмите: Random Forest, Logistic Regression, Naïve Bayes и CART. Като важни предиктори за учебната ефективност са посочени: посещаемост на учебните занятия и включване в учебен процес на студента със закъснение. Посочени са и други влияещи фактори като географския регион, нивото на развитие на икономиката и др.

Poudyal и др. [9] прогнозируют учебната ефективност, като най-напред върху набора с данни прилагат алгоритми за намаляване на размеростта (Principal Component Analysis и Linear Discriminant Analysis) и след това алгоритми от контролирано машинно обучение (Decision Tree, K-Nearest Neighbors и Logistic Regression). Данните включват изпитните резултати на студентите. Получени са прогнози с допустима точност за определяне на учебната им ефективност.

Nti и др. [10] изследват влиянието на социалните медии и социалните мрежи върху учебната ефективност на студентите, като показват, че степента на използване на социалните медии, включително и за учебни цели, определено влияе върху учебната ефективност. Чрез алгоритмите Decision Tree и Random Forest е създаден прогнозен модел, чрез който се определя зависимостта между използваемост на социални медии и учебна ефективност.

Martínez-Rodríguez и др. [11] показват значението на поведенчески и персонални характеристики на студентите за тяхната учебна ефективност, като акцентират проучването си върху навиците за учене и даване на самооценка за собствените им качества и ограничения. Оказва се, че приложеният алгоритъм от контролирано машинно обучение Multiple Linear Regression не е ефективен и затова е използван и алгоритъм за клъстериране K-means от неконтролирано машинно обучение. Създадени са профили на студентите съобразно тяхната учебна ефективност, като са се оформили три групи: с ниска, средна и висока учебна ефективност.

Тези проучвания показват използване на различни източници на данни, прилагане на съответни алгоритми за машинно обучение и възможност за прогнозиране на учебната ефективност с висока точност, което е обобщено чрез фиг. 1.



*Фиг. 1. Прогнозиране на учебната ефективност чрез машинно обучение*

### 3. ОНЛАЙН СЕМИНАР И УЧЕБНА ЕФЕКТИВНОСТ

Семинарните упражнения като част от учебната програма дават възможност на студентите да бъдат активни участници в учебен процес и да подобрят учебната си ефективност чрез участие в дискусии и изпълнение на малки изследователски задачи. Организирането на семинари в голяма степен зависи от преподавателя, от това какви инструкции и учебни ресурси е предоставил на студентите, каква електронна среда използва, какъв учебен сценарий е приложил и какъв компетентностен профил иска да постигне. Негов [12] разглежда най-добрите практики за организиране на семинар и препоръчва предварителна подготовка на студентите за провеждане на активна дискусия, която да ангажира и мотивира. В друга статия, Негов и Palfreyman [13] разглеждат ролята на семинара за развитие на аналитичното мислене при студентите. При провеждане на онлайн семинар, също така, трябва да се вземе под внимание и функционалността, която предоставя електронната среда за реализиране на един или друг учебен сценарий.

#### 3.1. Онлайн учебна среда и участници

Семинарните упражнения са проведени в рамките на един семестър със студенти бакалаври с профил, ориентиран към Информатика и софтуерни науки. Приложени са подходите *учене чрез изследване* и *учене чрез анализ и обобщение*, което е висша степен на прилагане на познанието според таксономията на Блум [14, 15, 16].

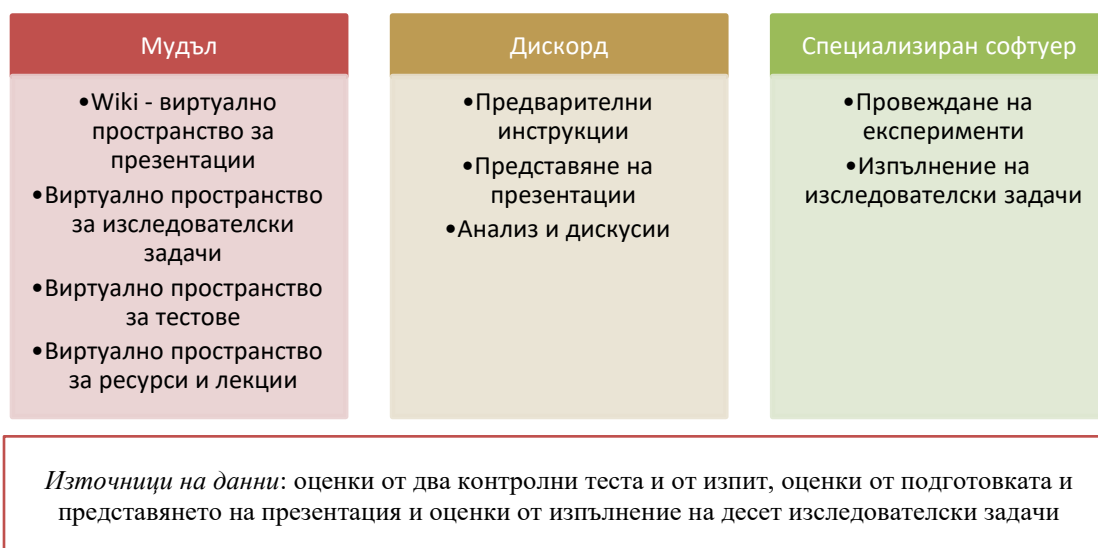
Онлайн семинарите са организирани в системата за електронно обучение Мудъл и комуникационната платформа Дискорд, като включват следните дейности:

- (1) предварително проучване и подготовка на презентация по зададена тема, като всеки студент индивидуално докладва и
- (2) изпълнение на малки изследователски задачи чрез специализиран софтуер.

В курса е приложено текущо и обобщаващо оценяване. Текущото е основано на подготовка на презентации и провеждане на дискусии, както и изпълнение на малки изследователски задачи, а обобщаващото оценяване включва изпълнение на два контролни онлайн теста по време на семестъра, като крайната оценка е формирана от оценката от тези контролни тестове и изпитната оценка, като и текущата работа на студентите през семестъра също е взета под внимание.

Темите за презентации са зададени от преподавателя в началото на семестъра чрез организиране на Wiki виртуално пространство в Мудъл. Готовите презентации се прикачват в Мудъл и се оценяват след тяхното представяне. Индивидуално изпълнените изследователски задачи също се прикачват във виртуално пространство в Мудъл и се оценяват. Лекции, литературни източници и инструкции също се намират в Мудъл.

На фиг. 2 е представен обобщен модел, който показва организираната онлайн учебна среда и източниците на данни, използвани за създаване на прогнозни модели, касаещи учебната ефективност на студентите. Източниците на данни са: оценките от двата контролни теста и от изпита, оценките от подготовката и представянето на презентацията и оценките от изпълнение на десет изследователски задачи.



**Фиг. 2.** Организирана учебна среда за семинар и източници на данни

### 3.2. Анализ на данните чрез машинно обучение

Събраните данни от текущото и обобщаващото оценяване са подготвени в два отделни .csv файла с цел създаване на два прогнозни модела, показващи учебната ефективност

съобразно данните от текущото и обобщаващото оценяване. При липсващи данни са приложени следните два подхода: (1) при наличие на повече от 2/3 празни полета в един запис, той е премахнат от набора с данни, а (2) при наличие на по-малко от 2/3 празни полета са използвани техники за запълване: *Последната стойност е нанесена в следващото празно поле* (Last Observation Carried Forward) и *Следващата стойност е нанесена в предното празно поле* (Next Observation Carried Backward).

Записите са етикетирани, като класовете са два: студенти с отлична учебна ефективност (група А) и студенти с много добра учебна ефективност (група Б). Записите за студентите, които са с лоша учебна ефективност очевидно вече не се намират в набора с данни, тъй като при тях се оказва, че 2/3 и повече от полетата са празни.

Няколко алгоритъма от контролирано машинно обучение Decision Tree, Naïve Bayes, k-NN и Deep Learning са обучени въз основа на така получените набори с данни, като отношението обучаващи/тестващи данни е 70%/30%. Избрани са алгоритми, имащи различен принцип на действие, като тяхната работоспособност е оценена чрез така известните за машинното обучение параметри: точност, абсолютна грешка, относителна грешка и средно-квадратична грешка.

В таблица 1 са показани резултатите от прилагане на алгоритмите от контролирано машинно обучение, които решават класификационна задача, а именно прогнозируют групата на студента А или Б в зависимост от текущото оценяване. С най-висока точност 93% се характеризират алгоритмите Deep Learning и Naïve Bayes. В таблица 2 се намират резултатите от прилагане на алгоритмите върху данни от обобщаващо оценяване. Вижда се, че с най-висока точност 96,67% се характеризира алгоритъмът Deep Learning, следван от Decision Tree, чиято точност е 93.33%. На фиг. 3 са показани графично резултатите от прилагане на четирите алгоритъма и може да се види разликата между оригиналните данни за учебната ефективност на студентите от двете групи и прогнозираните данни при текущото оценяване.

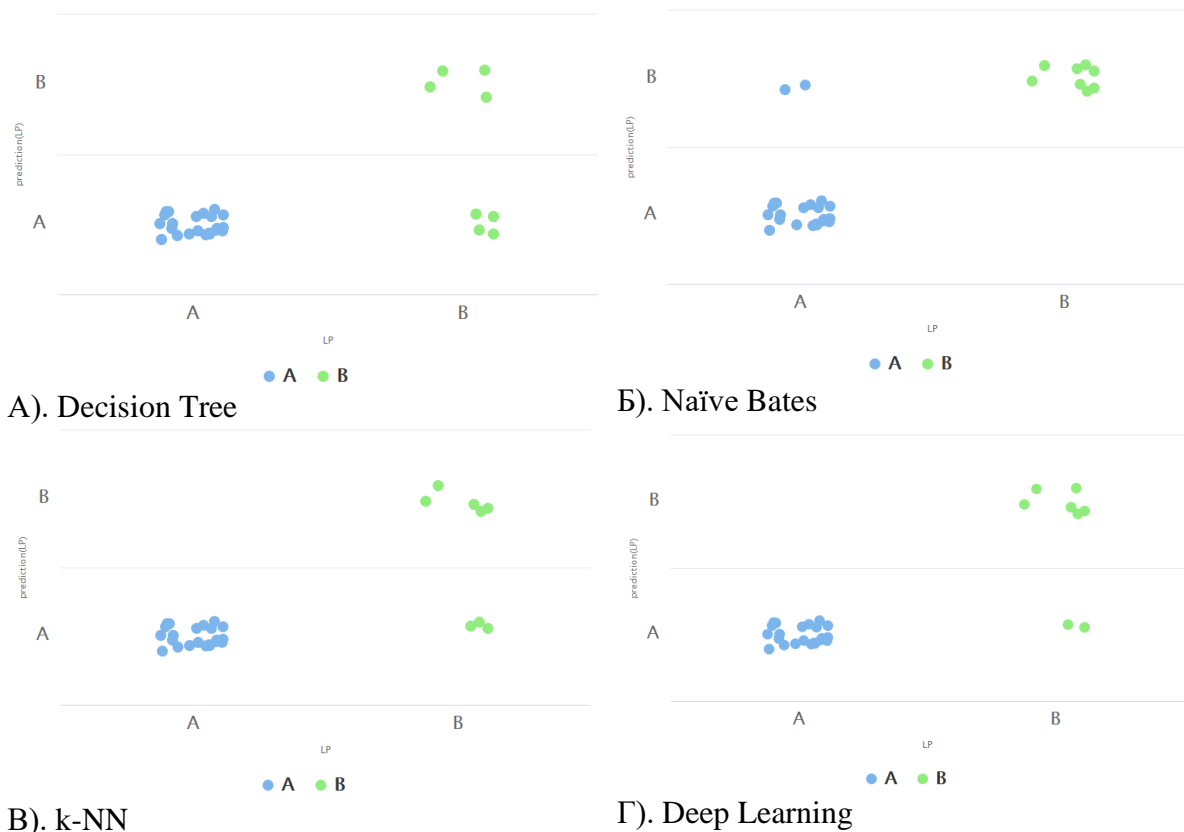
**Таблица 1.** Сравнение на алгоритми за машинно обучение при текущо оценяване

Параметър/ Алгоритъм	Decision Tree	Naïve Bayes	k-NN	Deep Learning
Точност	86.67%	93.33%	90.00%	93.33%
Абсолютна грешка	0.133 +/- 0.340	0.100 +/- 0.254	0.122 +/- 0.208	0.053 +/- 0.103
Относителна грешка	13.33% +/- 33.99%	9.98% +/- 25.36%	12.17% +/- 20.81%	5.33% +/- 10.29%
Средно-квадратична грешка	0.365	0.273	0.241	0.116

**Таблица 2.** Сравнение на алгоритми за машинно обучение при обобщаващо оценяване

Параметър/ Алгоритъм	Decision Tree	Naïve Bayes	k-NN	Deep Learning
Точност	93.33%	86.67%	83.33%	96.67%
Абсолютна грешка	0.092 +/- 0.230	0.175 +/- 0.249	0.137 +/- 0.216	0.094 +/- 0.163
Относителна грешка	9.23% +/- 23.02%	17.53% +/- 24.93%	13.70% +/- 21.61%	9.36% +/- 16.30%

<b>Средно-квадратична грешка</b>	0.248	0.305	0.256	0.188
----------------------------------	-------	-------	-------	-------



**Фиг. 3.** Оригинални и прогнозирани групи А и Б съобразно учебната ефективност (LP-Learning Performance) при обобщаващо оценяване

Таблица 3 съдържа сравнение на алгоритмите, които учат от данни за текущото и обобщаващото оценяване. Вижда се, че най-подходящи за тази задача са Decision Tree и Deep Learning.

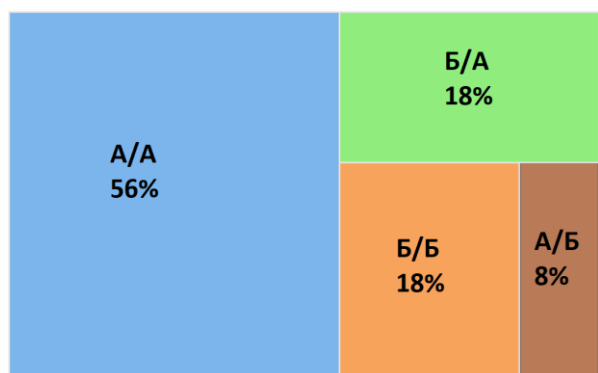
**Таблица 3.** Сравнение на алгоритми за машинно обучение при текущо и обобщаващо оценяване

Параметър/ Алгоритъм	Decision Tree	Naïve Bayes	k-NN	Deep Learning
<b>Точност</b>	96.67%	90.00%	73.33%	96.67%
<b>Абсолютна грешка</b>	0.047 +/- 0.173	0.096 +/- 0.203	0.256 +/- 0.280	0.167 +/- 0.199
<b>Относителна грешка</b>	4.65% +/- 17.31%	9.63% +/- 20.26%	25.60% +/- 27.97%	16.73% +/- 19.86%

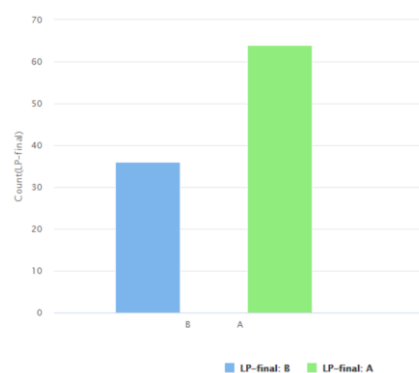
Средно-квадратична грешка	0.179	0.224	0.379	0.260
---------------------------	-------	-------	-------	-------

### 3.3. Анализ на резултатите

На фиг. 4а е показана карта с групите на студентите съобразно тяхната учебна ефективност. Първата буква посочва групата на студентите съобразно обобщаващото оценяване, а втората буква тяхната група, като се вземе под внимание текущото оценяване. Оказва се, че 56% от тях са класифицирани в група А, както според обобщаващото, така и според текущото оценяване. Това означава, че 56% от студентите имат много високи постижения, както при изпълнение на текущите учебни дейности в онлайн семинара, така при изпълнение на контролните тестове и изпитния тест. 18% от студентите се характеризират с много добри резултати от текущото и обобщаващото оценяване. Други 18% от тях са включени в група Б според обобщаващото оценяване и в група А съобразно постиженията им от текущото оценяване. 8% от студентите са показали различни резултати след анализ на обобщаващото и текущото им оценяване и затова са включени също в различни групи. И така, 64% от студентите може да се прогнозира с точност 96.67% при прилагане на алгоритмите Decision Tree и Deep Learning, че са част от група А и 36% от група Б (фиг. 4б), ако резултатите на студентите от обобщаващото оценяване се вземат с по-голяма тежест.



а). Карта на студентите по групи съобразно тяхната учебна ефективност



б). Прогноза на групата на студентите

**Фиг. 4.** Прогнозиране на учебната ефективност на студентите

## ЗАКЛЮЧЕНИЕ

Направеният анализ показва значението на активното участие на студентите в онлайн семинари в рамките на един семестър за успешното завършване на курса и постигане на по-добри резултати. Онлайн семинарът влияе както върху текущата оценка, така и върху обобщаващата. Показана е връзката между учебната ефективност, активното участие в онлайн семинар и постиженията на студентите, като те са групирани в: (1) група А –



студенти с отлична учебна ефективност и (2) група Б – студенти с много добра учебна ефективност. Една малка част от студентите при текущото оценяване попадат в една група, а при обобщаващото оценяване в друга. Това показва предпочитанията на студента към едни или други учебни и изпитни дейности, както и че те са изпълнени с различна учебна ефективност. Извършено е и крайно класифициране на студентите по групи (А или Б), което отчита цялостното изпълнение на всички дейности от студентите в рамките на курса.

Анализът е извършен въз основа на прилагане на алгоритми за машинно обучение, което се оказва подходящ подход за решаване на класификационна задача с два класа. С най-висока точност са моделите, получени след изучаване на данните от алгоритмите Deep Learning и Decision Tree.

Събраните данни и анализът на учебната ефективност могат да се използват за прогнозни цели и за насочване както на студента, така и на преподавателя за оптимизиране на съответните дейности в учебен процес.

### Благодарности

Изследването е подпомогнато от Фонд „Научни изследвания“ по проект „Моделиране и изследване на интелигентни системи за обучение и сензорни мрежи (ИСОСеМ)“, договор КП-06-Н47/4 от 26.11.2020.

### ЛИТЕРАТУРА

- [1] H. Hao, X. Geng, L. Chen, A. Shimada and M. Yamada, (2021). „Learning Analytics of the Relationships among Knowledge Constructions, Self-regulated Learning, and Learning Performance“, *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, 290-297, doi: 10.1109/TALE52509.2021.9678920.
- [2] C. Wijesooriya, J. Heales, F. Rohde, (2019). „Evaluation of virtual learning environment characteristics in learning performance IC4E '19“, *10th International Conference on E-Education, E-Business, E-Management and E-Learning*, 259–263, <https://doi.org/10.1145/3306500.3306535>.
- [3] Z. Liu, M. Fan and B. Wang, (2021). „Analysis and Research on the Influencing Factors of Learners' Learning Performance in Blended Learning Environment“, *2021 International Symposium on Educational Technology (ISET)*, 165-169, doi: 10.1109/ISET52350.2021.00042.
- [4] O. Ojajuni, et al., (2021). „Predicting Student Academic Performance Using Machine Learning, Computational Science and Its Applications – ICCSA 2021“, *Lecture Notes in Computer Science*, vol. 12957, Springer, Cham. [https://doi.org/10.1007/978-3-030-87013-3\\_36](https://doi.org/10.1007/978-3-030-87013-3_36).
- [5] M. R. Rimadana, S. S. Kusumawardani, P. I. Santosa and M. S. F. Erwianda, (2019). „Predicting Student Academic Performance using Machine Learning and Time Management Skill Data“, *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 511-515, doi: 10.1109/ISRITI48646.2019.9034585.
- [6] H. Alharthi, (2021). „Machine Learning Techniques to Predict Academic Performance of Health Sciences Students“, *2021 20th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 33-36, doi: 10.1109/DCABES52998.2021.00015.
- [7] J. Naren et al., (2021). „A Machine Learning Model for Predicting Academic Performance of Students through Internet Usage“, In: P. K. Mallick, A. K. Bhoi, GS. Chae, K. Kalita (Eds.), *Advances in Electronics, Communication and Computing, Lecture Notes in Electrical Engineering*, vol. 709, Springer, Singapore, [https://doi.org/10.1007/978-981-15-8752-8\\_32](https://doi.org/10.1007/978-981-15-8752-8_32).
- [8] P. Kumar and M. Sharma, (2020). „Predicting Academic Performance of International Students Using Machine Learning Techniques and Human Interpretable Explanations Using LIME - Case Study of an Indian University“.

- In: A. Khanna, D. Gupta, S. Bhattacharyya, V. Snasel, J. Platos, A. Hassanien (Eds.), International Conference on Innovative Computing and Communications, *Advances in Intelligent Systems and Computing*, vol. 1087, Springer, Singapore, [https://doi.org/10.1007/978-981-15-1286-5\\_25](https://doi.org/10.1007/978-981-15-1286-5_25).
- [9] S. Poudyal, M. Nagahi, M. Nagahisarchoghaei and G. Ghanbari, (2020). „Machine Learning Techniques for Determining Students' Academic Performance: A Sustainable Development Case for Engineering Education“, *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 920-924, doi: 10.1109/DASA51403.2020.9317178.
- [10] I. K. Nti, et al. (2022). „Prediction of social media effects on students' academic performance using Machine Learning Algorithms (MLAs)“, *J. Comput. Educ.* 9, 195–223, <https://doi.org/10.1007/s40692-021-00201-z>.
- [11] R. A. Martínez-Rodríguez, O. Alvarez-Xochihua, O. D. Mejia Victoria, A. Jordán Arámburo and J. Á. González Fraga, (2019). „Use of Machine Learning to Measure the Influence of Behavioral and Personality Factors on Academic Performance of Higher Education Students“, *IEEE Latin America Transactions*, 17(4), 633-641, doi: 10.1109/TLA.2019.8891928.
- [12] M. Heron, (2018). „Pedagogic practices to support international students in seminar discussions“, *Higher Education Research & Development*, DOI: 10.1080/07294360.2018.1512954.
- [13] M. Heron and D. M. Palfreyman, (2021). „Exploring Higher-Order Thinking in Higher Education Seminar Talk“, *College Teaching*, <https://doi.org/10.1080/87567555.2021.2018397>.
- [14] Bloom's Taxonomy Revised, <https://www.utica.edu/academic/Assessment/new/Bloom%20tx%20revised%20combined.pdf>.
- [15] M. Dorodchi, N. Dehbozorgi and T. K. Frevert, (2017). "I wish I could rank my exam's challenge level!": An algorithm of Bloom's taxonomy in teaching CS1“, *2017 IEEE Frontiers in Education Conference (FIE)*, 1-5, doi: 10.1109/FIE.2017.8190523.
- [16] S. Huilan et al., (2020). „Educational management in Critical Thinking Training Based on Bloom's Taxonomy and SOLO Taxonomy“, *2020 International Conference on Information Science and Education (ICISE-IE)*, 518-521, doi: 10.1109/ICISE51755.2020.00116.