

Shala Riza, Ledian; Abazi Bexheti, Lejla

Article

ML Techniques Integration in Digital Learning Platforms: Students' Dataset Statistical Analysis

ENTRENOVA - ENTERprise REsearch InNOVation

Provided in Cooperation with:

IRENET - Society for Advancing Innovation and Research in Economy, Zagreb

Suggested Citation: Shala Riza, Ledian; Abazi Bexheti, Lejla (2025) : ML Techniques Integration in Digital Learning Platforms: Students' Dataset Statistical Analysis, ENTRENOVA - ENTERprise REsearch InNOVation, ISSN 2706-4735, IRENET - Society for Advancing Innovation and Research in Economy, Zagreb, Vol. 10, Iss. 1, pp. 22-32,
<https://doi.org/10.54820/entrenova-2024-0003>

This Version is available at:

<https://hdl.handle.net/10419/317946>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc/4.0/>

ML Techniques Integration in Digital Learning Platforms: Students' Dataset Statistical Analysis

Lediana Shala Riza

South East European University, North Macedonia

Lejla Abazi Bexheti

South East European University, North Macedonia

Abstract

With the use of technology-enhanced learning platforms and an abundance of available educational data, it is possible to analyze student learning behavior and solve problems, improve the learning environment, and make data-driven decisions. A virtual learning environment effectively provide datasets for analyzing and reporting student learning, as well as its reflection and participation in their individual performances, which complements the learning analytics paradigm. This work is intended to explain the use of AI-based approaches in online learning, with a particular focus in offering a statistical approach on students VLE dataset. The study uses quantitative methodology to highlight the association between the variables in the obtained dataset. The purpose of this research is to examine the correlation and dependency of the dataset variables in order to observe the relationship between these variables and the effect that these attributes may have on students' performance in a digital learning environment. According to the findings of this study, there is a correlation between student performance and a number of different factors, such as resource (page) views, course modules, assessment type, assessment weight and sum of clicks in a VLE.

Keywords: ML techniques, digital platforms, engagement, attributes, analysis, statistics,

JEL classification: C88

Paper type: Research article

Received: 11 June 2024

Accepted: 16 August 2024

DOI: 10.54820/entrenova-2024-0003

Introduction

In recent years, due to the ever-increasing volume of data and the evolving demands of higher education, including digital education, the application of AI and machine learning techniques has grown significantly in all fields. Similar to this, there is a vast amount of data about students in digital education available on online educational information platforms. Anyway, there are a number of issues with online learning platforms, prominent among them being the absence of student interest in the different course materials and assignments (Waheed et al., 2020).

With the potential to completely transform human social interactions, artificial intelligence is a rapidly expanding field of technology. AI is increasingly being used in education to create novel teaching and learning strategies that are being tested in various environments (UNESCO, 2019). Technological developments in the field of Information and Communications (ICT) have contributed to the expansion of Massive Open Online Courses (MOOCs) in remote learning settings. Interactive content has been delivered using a variety of technologies, such as images, figures, and videos, which can inspire students to develop new cognitive abilities. MOOCs are an effective dashboard platform that top colleges have used, enabling students from all around the world to enroll in these kinds of courses (Gurmoorthi et al., 2023). On the other side, Virtual learning environments' (VLEs') explosive growth has offered the worldwide community open-access, large-scale learning possibilities, which had a huge impact on the educational landscape using MOOCs, or by using LMSs to improve particular corporate and institutional learning. These settings include, but are not exclusive to, educational establishments, such as schools and colleges (e.g., Open Universities, career training facilities (like Codecademy and Coursera), and non-traditional educational settings like online clever tutoring and academies (like Udemy and Khan Academy) mechanisms (like ASSISTments). These VLEs broaden the breadth of education and give students everywhere more flexibility and accessibility allowing students to take tests remotely, participate in online conversations, and turn in their homework conveniently (Alnasyan et al., 2024).

Recognizing students' participation in beneficial activities is crucial to lowering dropout rates in virtual learning environments (VLEs). The likelihood that a student will finish the online course and receive a high assessment score rises with increased student participation in course activities, which also makes the experiences more engaging (Hussain et al., 2018). The level of student involvement in web-based learning systems is comparatively lower than that of traditional education methods. The ability to participate in online VLE activities is used to evaluate student involvement. Considering the course is web-based, there is frequently no in-person interaction between the lecturer and the students. Given that many of these variables are not readily available in e-learning systems, it is challenging to quantify a student's engagement in web-based systems using traditional approaches (e.g., measures such as class attendance, participation in discussions, and grades). Therefore, it is a difficult task to investigate students' participation with web-based learning (Soobramoney, 2021). One of the most crucial metrics for assessing the effectiveness of the educational process is the ability to predict students' academic performance. This is because it can give administrators and teachers early warning about students who may be at risk of failing or dropping out, allowing them to intervene and support these students with tutoring, individualized learning experiences, feedback and guidance, and other forms of support (Alnasyan et al., 2024).

The framework for our research focuses on creating a better machine learning model to forecast low-engagement students in online learning settings. Our main goal in this area is to provide a statistical approach of the dataset imported from the Open

University Learning Analytics (OULAD) dataset. The study approach used to comprehend the dataset variations, tables, and their interactions is also described in this work. There are several possible uses for the dataset. It permits the evaluation of predictive models for the purpose of predicting the outcomes of student assessments and final courses, as well as the comparison of systems with additional systems created by other researchers. Course structure can be studied from a learning viewpoint using the VLE data, and the data itself can be used to assess how VLE affects learning results (Kuzilek et al., 2017).

This is how the remainder of the paper is structured: the study's problem statement, research objectives, and research questions are introduced in Section 1. Section 2 highlights the most popular machine learning algorithms for predicting student performance and engagement and also emphasizes the revolutionary potential of AI in education. A thorough explanation of the research design, methods, dataset characteristics, instruments used, and conclusions are given in Section 3. Section 4 presents the study's conclusion.

The main goal of this paper is to give an analytic approach of the variable's correlation in a massive VLE OULAD dataset. The specific objectives are:

- To identify some of the most important attributes that may impact student engagement in VLE
- To determine the feature correlation in an online open dataset
- To study variables dependency, association and effect on each other.

Research Questions

Based on the goal of the research the primary research question posed for this case study are:

RQ1: Is there a difference in the frequency of clicks between student groups pursuing different courses?

RQ2: Does gender, student level of education and the interaction between them have statistically significant effect on the sum of clicks in VLE?

RQ3: Is there any association between VLE activity type and the number of times the student interacted with the material?

RQ4: Is there any correlation between VLE module, assessment type, assessment weight and assessment score?

Related work

Transformative potential of Artificial Intelligence in education

As part of in-person, blended, or fully online learning contexts, digital education refers to "teaching and learning activities which make use of digital technology." Digital technologies include artificial intelligence (AI), which works with machines and intelligent applications to solve problems in the real world. Whereas DL is a subset of ML techniques, ML is a subset of AI that offers the capacity to automatically learn from experiences and data in order to tackle complicated issues. DL analyse (Munir et al., 2023).

By streamlining a wide range of administrative duties, teachers can spend more time interacting with students by incorporating AI into the classroom. Teachers can give more immediate feedback and assess students more quickly with the use of AI. Additionally, with the use of this technology, learners and students can enhance their critical thinking, creativity, problem-solving, and digital literacy. Ultimately, with the help of teachers, AI can tailor the learning process, resulting in enhanced academic achievement and greater flexibility in meeting a range of learning requirements

(World Economic Forum, 2024). By analyzing student data, AI algorithms may design personalized learning programs that consider the individual strengths, weaknesses, and learning preferences of each student. Students who pick up topics quickly might be sent to higher coursework or enrichment programs, while those who require more time might be given extra explanations, practice questions, or one-on-one tutoring (Cardona et al., 2023). Additionally, by implementing tailored instructions, the predictions guarantee that every student has an opportunity to reach their maximum potential and give teachers insights into the unique preferences and learning requirements of each student. Moreover, these predictions help administrators and educators make better-informed choices about the creation of curricula, the distribution of resources, evaluation, and training of teachers (Alnasyan et al., 2024).

Today, a wide range of communities have developed in close proximity to one another, sharing a common interest in the ways in which educational data can be utilized to further learning research and education:

Educational Data Mining" (EDM) refers to the use of data mining (DM) techniques to a particular kind of dataset sourced from educational settings in order to address significant educational inquiries.

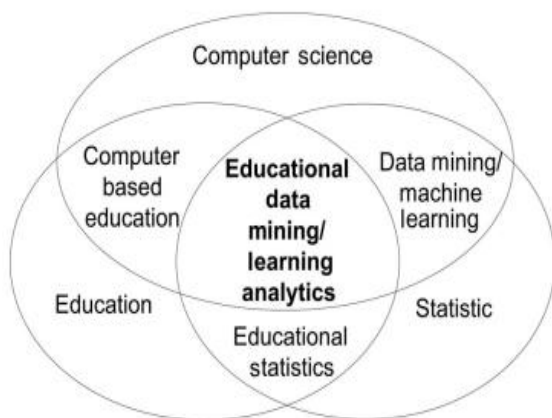
Learning Analytics (LA) is concerned with the measurement, gathering, analysis, and reporting of data on learners and their surroundings with the goal of comprehending and improving learning and the environments in which it takes place.

Academic analytics (AA) and **Institutional Analytics (IA)** deal with the gathering, analyzing, and visualizing of data related to academic program activities, including courses, degree programs, research, student fee revenue, course evaluation, resource allocation, and management, in order to produce institutional insight.

Big Data in Education (BDE) is the use of big data techniques to educational environment data, with big data being the basic connotation summed up in volume, variety, value, and velocity (Romero & Ventura, 2019).

Figure 1

Major Domains in Learning Analytics and Educational Data Mining



Source: (Kaur & Gupta, 2023)

ML Algorithms for Student Performance and Engagement Prediction

With the rising availability of big data in education, predictive analytics has gained popularity in higher education in recent years. Specifically, the majority of academic

institutions' learning management systems (LMSs) provide an abundance of data on student activities (Chen & Cui, 2020). ML algorithms are capable of producing suggestions, classifications, and predictions by training models using available data. ML is used in adaptive learning to comprehend student behavior, customize content, and modify teaching methods (Gligorea et al., 2023).

Several studies in the field of machine learning have been conducted in recent years with the goal of identifying low engagement students due to the various challenges associated with e-learning systems, the most significant of which is the lack of student motivation in various course activities. In the work of Kuzilek et al., 2017 using the Open University (OU) dataset, the authors ran many machine learning algorithms in order to predict low-engagement students. In comparison to the other investigated models, the J48, decision tree, JRIP, and gradient-boosted classifiers performed better in terms of accuracy, kappa value, and recall. The study's input variables included the number of clicks performed on the virtual learning environment (VLE) activities (*dataplus*, *forumng*, *glossary*, *oucollaborate*, *oucontent*, *resources*, *subpages*, *homepage*, and *URL*) during the first-course assessment, as well as the highest education level, final results, and assessment score. The student's level of participation in the various activities was the output variable. Additionally, Soobramoney, 2011 employed ensemble machine learning techniques to offer a better answer for the student at risk prediction problem (SAR). In this study, twenty-five different ML classifiers were used to train forty different ML predictive models, one for each week of the semester. Four classifiers—AdaBoostClassifier, LGBMClassifier, RandomForestClassifier, and XGBClassifier—were chosen as the ensemble classifier's base learners. As a result, the voting classifier ensemble technique can offer a better answer to the SAR prediction problem than the individual classifiers.

A random forest classifier was employed in conjunction with the SMOTE data-balancing technique by the authors X in order to attain high accuracy results in their investigation. By utilizing SMOTE data balancing instead of not using it, the suggested methodology showed a 5% improvement in performance (Jawad et al., 2022). The study conducted by Alnassar (2023) employed a similar methodology for data balancing techniques, but with the addition of multiple machine learning models. These models were applied to two benchmark VLE datasets, namely Open University Learning Analytics Dataset (OULAD) and Coursera. The models included Extreme Gradient Boosting (XGBoost), CATBoost, K-Nearest Neighbour (KNN), and Support Vector Classifier (SVC). Consequently, the most effective machine learning models in relation to various investigations were found to be MLP, XGBoost, and CATBoost, with 90% classification accuracy. The study effort of Palani et al, 2021) employed three unsupervised clustering algorithms—Gaussian Mixture, Hierarchical, and K-prototype—using the same dataset. Findings show that compared to the other suggested models, the K-Prototype model produced highly partitioned clusters and more accurately clustered the low-engagement students.

Methodology

In our research work we have applied a quantitative methodology which is based the causal statistical analysis. This method aims to identify the causal relationships between various variables found in the raw data. Large-scale data collection is the first step in the statistical analysis process, which identifies trends, patterns, and insights using statistics and other data analysis mechanisms. For data processing we used SPSS statistical tool and employed descriptive statistics focusing on methodologies such as cross-tabulation, frequencies; and bivariate statistics (means, correlation, ANOVA).

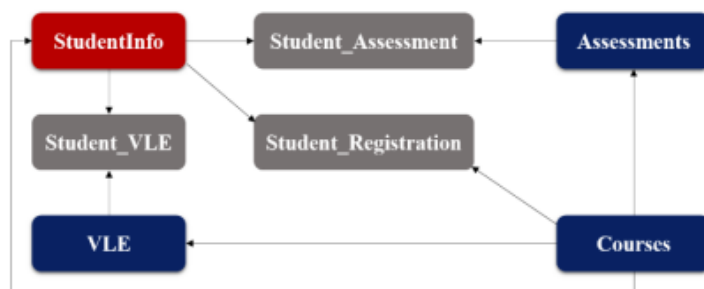
Excel is used to combine and merge tables with common columns before the data are processed and analyzed in SPSS.

Dataset

We made use of the OULA dataset, which comprises data on 32,593 students, 22 courses, assessment results, demographic data and logs of their interactions with the VLE, which are summarized daily by clicking counts (10,655,280 entries). This dataset makes it possible to evaluate predictive models for predicting the outcomes of student assessments and final course grades, as well as to compare the models with other existing state of the art ML based models (Kuzilek, Hlostá, & Zdrahal, 2017).

Figure 2

The structure of OULAD database



Source: (Hlioui, Aloui, & Gargouri, 2021)

The student's registration details, and demographic data are connected. The dataset includes logs of student interactions with VLE and the outcomes of student assessments for each triplet of student, module, and presentation. Seven modules make up the dataset: three social science modules (AAA, BBB, GGG) and four STEM (CCC, DDD, EEE, FFF) modules, while the module has a significant number of failing students. We have linked table *studentInfo* to *studentVle* using column *id_student*, assessments table links to *studentAssessment* using *id_assessment* and *vle* to *studentVle* using *id_site*.

Results

To ascertain whether there is a relationship between the number of clicks made in the VLE by students pursuing various modules (AAA, BBB, CCC, DDD, EEE, FFF, and GGG), we employed the one-way ANOVA test. We used *sum_click* as the dependent variable. One-Way ANOVA makes it possible to examine variations in the dependent variable between the several categories. Because it compares the variance, or score variability, between the several groups (which is thought to be caused by the independent variable), it is known as an analysis of variance (Bittner, 2022).

Table 1 shows that there are statistically significant differences between the groups, since the **p-value** is less than the significance level (**p < .05**). Hence, we have answered our first research question **R1** and we also have evidence to reject the null hypothesis that there is no difference in the population between the averages of each group of students attending different modules. Additionally, a larger difference between the group means is indicated by the **F-value**. It implies that there are notable differences between the groups.

Table 1

Relationship between sum of clicks in and student modules

num_of_prev_attempts					
	Sum	of	Mean	F	Sig.
	Squares	df	Square		
Between Groups	6029.981	4	1507.495	7000.096	.000
Within Groups	225813.227	1048570	.215		
Total	231843.208	1048574			

Source: Author's work

Further we will evaluate the effect of student gender and level of education as two independent factors on the sum of clicks as the dependent variable in order to address the second research question. We will also determine whether or not their interaction has a meaningful impact on the dependent variable. As seen in Tab 2, Levene's test rejects the null hypothesis thus the error variance of the *sum_click* variable is not equal across groups.

Table 2

Levene's Test of Equality of Error Variances

Dependent Variable: studentVle.sum_click			
F	df1	df2	Sig.
175.618	9	1048565	.000

Note: a. Design: Intercept + highest_education + gender + highest_education * gender

Source: Author's work

From the Test of Between-Subject Effects we can see for gender and level of education and the interaction between them there is a statistically significant difference on the sum of clicks as a dependent variable ($.000 < .05$), so all of these three factors have a significant effect on the number of clicks. The rejection of null hypothesis is reinforced from high **F** values as well ($F > 2.5$), which represents the ratio of the variance between the groups to the variance within the groups (Bittner, 2022).

Table 3

Test of Between-Subject Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	21389.413 ^a	9	2376.601	92.129	.000
Intercept	181978.470	1	181978.470	7054.389	.000
highest_education	3927.884	4	981.971	38.066	.000
gender	346.947	1	346.947	13.449	.000
highest_education * gender	10052.172	4	2513.043	97.418	.000
Error	27049297.158	1048565	25.796		
Total	39045765.000	1048575			
Corrected Total	27070686.571	1048574			

Source: Author's work

In order to answer the third research question (**RQ3**: Is there any association between VLE activity type and the number of times the student interacted with the material?), we will be combining *vle* and *studentvle* tables and perform the Eta correlation which measures the strength of a non-linear association between a nominal variable and scale variable (Liu, 2022). Here *VleActivityType* is the independent variable while *sum_click* is the dependent variable. Based in the coefficient value ($0.21 < 0.205 < 0.40$) we can conclude that there is a weak association between these two variables (Table 4).

Table 4
Directional Measures

			Value
Nominal by Interval	Eta	VLEActivityType Dependent	.268
		sum_click Dependent	.205

Source: Author's work

To find the significant value we performed the bivariate correlation test which gave a **p** value of .000 which is again lower than the reference vale (.05) so we can conclude that there exists a statistically significant weak correlation between *VLEActivityType* and *sum_click* variables.

Table 5
Bivariate correlation test

		VLEActivityType	sum_click
VLEActivityType	Pearson Correlation	1	-.161**
	Sig. (2-tailed)		.000
	N	1048575	1048575
sum_click	Pearson Correlation	-.161**	1
	Sig. (2-tailed)	.000	
	N	1048575	1048575

Note: **. Correlation is significant at the 0.01 level (2-tailed)

Source: Author's work

Finally, to detect if there is any correlation between VLE module, assessment type, assessment weight and assessment score we have performed the correlation test and the multiple linear regression in these four variables by merging *StudentAssessment* and *Assessment* tables. Predictive modelling is accomplished through the use of linear regression analysis. A bivariate model is constructed in simple linear regression to predict a response variable (*y*) from an explanatory variable (*x*). A multivariate model is created when the model in multiple linear regression is expanded to incorporate more than one explanatory variable (x_1, x_2, \dots, x_p). In social science study and practice, this is a commonly employed method (Tranmer et al., 2020).

From Table 6 we can see that all significant values are less than .05 which indicates that all these variables have a significant relationship with each other. Assessment weight is positively correlated to *AssessmentType* and negatively coorelated to *studentAssessment_score* & *CodeModule*. *studentAssessment_score* is positively correlated with *CodeModule* and negativity correlated with weight and *AssessmentType*; *CodeModule* is negatively correlated to weight and *AssessmentType* while positively correlated to *studentAssessment_score*; and finally *AssessmentType* is

positively coorelated to *weight* and negatively correlated to *studentAssessment_score* and *CodeModule*.

Table 6
Multiple linear regression test

Correlations ^b					
		weight	studentAssessme nt.score	CodeModule	Assessme ntType
weight	Pearson Correlation	1	-.166**	-.168**	.392**
	Sig. (2-tailed)		.000	.000	.000
studentAssess ment.score	Pearson Correlation	-.166**	1	.078**	-.216**
	Sig. (2-tailed)	.000		.000	.000
CodeModule	Pearson Correlation	-.168**	.078**	1	-.167**
	Sig. (2-tailed)	.000	.000		.000
AssessmentTy pe	Pearson Correlation	.392**	-.216**	-.167**	1
	Sig. (2-tailed)	.000	.000	.000	

Note: **. Correlation is significant at the 0.01 level (2-tailed). b. Listwise N=173739

Source: Author's work

Discussion

This study focuses on providing a statistical approach to data analysis over the OULAD dataset, which is one of the most used datasets in the machine learning industry. It is primarily used to predict academic performance, engagement, and success of students in virtual learning environments. We combined and connected the datasets using programs like Excel and analyzed the data by responding to the study questions using the statistical package SPSS.

By performing the ANOVA test we have shown that there is a difference in the frequency of clicks between student groups pursuing different courses. Test of Between-Subject Effects was used to determine that gender, student level of education and the interaction between them have statistically significant effect on the sum of clicks in VLE. Through Eta correlation it was concluded that there is a weak association between VLE activity type and the times the student interacted with the material. Lastly, we ran a correlation test and multiple linear regression on these four variables to see if there was any correlation between the VLE module, assessment type, assessment weight, and assessment score. The results show that there is a significant relationship between all of these variables.

Conclusion

The analysis of the OULA dataset was done with the aim of serving as an aid during the design process of an ML based model that predicts student engagement in a virtual learning environment. The design of such algorithms is certainly based on the use of software such as Weka or RapidMiner, which contain tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

Using this software, the feature selection step will be used to determine which characteristics are most crucial for the designed algorithm. However, the statistical

approach we have provided will be useful at this stage of variable selection and aid in the overall comprehension of the dataset's features, the tables' connections, the characteristic variables, and the relationships and effects of the variables on one another.

References

1. Alnassar, F. M. (2023). Predicting Student Performance on Virtual Learning Environment. London: Goldsmiths, University of London.
2. Alnasyan, B., Basher, M., & Allassafi, M. (2024). Deep Learning Techniques for Predicting Student's Academic Performance on Virtual Learning Environments: A Review. Research Square, 65. Retrieved from <https://doi.org/10.21203/rs.3.rs-3888441/v1>
3. Alnasyan, B., Basher, M., & Allassafi, M. (2024). The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review. Computers and Education: Artificial Intelligence, 6, 100231. doi:<https://doi.org/10.1016/j.caeai.2024.100231>
4. Bittner, A. (2022). Analysis-of-variance (ANOVA) Assumptions Review: Normality, Variance Equality, and Independence. The XXXIVth Annual International Occupational Ergonomics and Safety Conference , (pp. 28-33). Washington. doi:10.47461/isoes.2022_bittner
5. Cardona, M., Rodríguez, R., & Ishmael, K. (2023). Artificial Intelligence and the Future of Teaching and Learning. Washington DC: Office of Educational Technology.
6. Chen, F., & Cui, Y. (2020). Utilizing Student Time Series Behaviour in Learning Management Systems for Early Prediction of Course Performance. Journal of Learning Analytics, 7, 1-17. Retrieved from <http://dx.doi.org/10.18608/jla.2020.72.1>
7. Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., & Tudorache, P. (2023). Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. Education Sciences, 13, 1216. Retrieved from <https://doi.org/10.3390/educsci13121216>
8. Gurmoorthi, Kumar, V., Reddy, V., & Kiran, U. (2023). Student Performance Prediction in Online Courses Using Machine Learning Algorithms. International journal for recent developments in science & technology, 35-42.
9. Hlioui, F., Aloui, N., & Gargouri, F. (2021). A Withdrawal Prediction Model of At-Risk Learners Based on Behavioural Indicators. International Journal of Web-Based Learning and Teaching Technologies, 16, 32-53. doi:10.4018/IJWLTT.2021030103
10. Hussain, M., Zhu, W., Zhang, W., & Abidi, S. R. (2018). Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Score. Computational Intelligence and Neuroscience, 2018, 21. doi:10.1155/2018/6347186
11. Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. {Sustainability}, 14, 14795. Retrieved from <https://www.mdpi.com/2071-1050/14/22/14795>
12. Kaur, A., & Gupta, D. (2023). A Hybrid Classification Model for Prediction of Academic Performance of Students: An EDM Application. Emergent Converging Technologies and Biomedical Systems (pp. 59–71). Singapore: Springer Nature Singapore. Retrieved from https://doi.org/10.1007/978-981-99-2271-0_6
13. Kuzilek, J., Hlostá, M., & Zdrahal, Z. (2017). Data Descriptor: Open University Learning Analytics dataset. Scientific Data, 4, 170171. doi:10.1038/sdata.2017.171
14. Liu, X. S. (2022). Bias Correction for Eta Squared in One-Way ANOVA. Methodology, 18, 44-57. Retrieved from <https://meth.psychopen.eu/index.php/meth/article/view/7745>
15. Munir, H., Vogel, B., & Jacobsson, A. (2023). Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision. Information, 13, 203. Retrieved from <https://doi.org/10.3390/info13040203>
16. Palani, K., Styne, P., & Pathak, P. (2021). Clustering Techniques to Identify Low-engagement Student Levels. Proceedings of the 13th International Conference on Computer Supported Education (CSEDU 2021). 2, pp. 248-257. Ireland: SCITEPRESS – Science and Technology Publications.

17. Romero, C., & Ventura, S. (2019). Educational data mining and learning analytics: An updated survey. Wiley Periodicals, Inc, 1-21. doi:10.1002/widm.1355
18. Soobramoney, R. (2021). EARLY PREDICTION OF STUDENTS AT RISK IN A VIRTUAL LEARNING ENVIRONMENT USING ENSEMBLE MACHINE LEARNING TECHNIQUES. Durban: Durban University of Technology.
19. Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). Multiple Linear Regression. Cathie Marsh Institute Working Paper 2020-01. Retrieved from <https://creativecommons.org/licenses/>
20. UNESCO. (2019). Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. Paris, France: United Nations Educational, Scientific and Cultural Organization.
21. Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting Academic Performance of Students from VLE Big Data using Deep Learning. Computers in Human Behavior, 104, 106189. doi:<https://doi.org/10.1016/j.chb.2019.106189>
22. World Economic Forum. (2024). Shaping the Future of Learning: The Role of AI in Education 4.0. Geneva: World Economic Forum.

About the authors

Lediana Shala Riza is a PhD candidate at Faculty of Contemporary Sciences and Technologies, South-East European University, North Macedonia. She received her B.S. degree in Computer Engineering from University of Prishtina in 2018 and the M.S. degree in Subject Teaching with Specialization in Technology & ICT from University of Prishtina in 2021. Online assessment, e-teaching & e-learning, and educational technology are among her research interests. She has written or co-written papers on ICT and education at international conferences and journals. The author can be contacted at ls31402@seeu.edu.mk

Lejla Abazi Bexheti is an Associate Professor at the Faculty of Contemporary Sciences and Technologies at South East European University in Macedonia. She holds a PhD in Computer Science and has been part of the CST teaching staff since 2002. Her main research activity is in Learning Systems and eLearning, and she has been involved in many international projects and research activities in this area. At SEE University, she was involved in resolving the Learning Management System issue. Currently, she is Prorector for academic issues at SEEU. The author can be contacted at l.abazi@seeu.edu.mk