



INTERNATIONAL CONFERENCE INFORMATION AND COMMUNICATION TECHNOLOGIES IN BUSINESS AND EDUCATION



CONFERENCE PROCEEDINGS



Publishing house "Science and economics"
University of Economics - Varna

International Conference
INFORMATION AND COMMUNICATION
TECHNOLOGIES IN
BUSINESS AND EDUCATION

Conference proceedings

**INFORMATION AND COMMUNICATION
TECHNOLOGIES IN
BUSINESS AND EDUCATION**

**Proceedings of the International Conference
dedicated to the
50th anniversary
of the Department of Informatics**

2019

University publishing house “Science and economics”
University of Economics – Varna

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Permission for use must always be obtained from University publishing house “Science and economics”. Please contact izdatel@ue-varna.bg

Published reports have not been edited or adjusted. The authors are solely responsible for the content, originality and errors of their own fault.

With the support of ADASTRA, 411 Marketing Business Services.

© University publishing house “Science and economics”,
University of Economics – Varna, 2019.

ISBN 978-954-21-1004-0

PROGRAMME COMMITTEE

Abdel-Badeeh M. Salem, Egypt

Aijaz A. Shaikh, Finland

Aleksandar Dimov, Bulgaria

Alexander Feldman, USA

Alexander Sergeev, Russia

Antoanela Naaji, Romania

Avram Eskenazi, Bulgaria

Dariusz Mrozek, Poland

Dimitrina Polimirova, Bulgaria

Dmitriy Berg, Russia

Elena Nechita, Romania

Emil Denchev, Bulgaria

Georgi Dimitrov, Bulgaria

Ivan Popchev, Bulgaria

Jeanne Schreurs, Belgium

Julian Vasilev, Bulgaria

Kameliya Stefanova, Bulgaria

Krasimir Shishmanov, Bulgaria

Maciej Czaplewski, Poland

Maciej Pondel, Poland

Maksim Medvedev, Russia

Maria Hristova, Bulgaria

Marian Cristescu, Romania

Marina Medvedeva, Russia

Michael Voskoglou, Greece

Nadezhda Filipova, Bulgaria

Nikola Nikolov, Ireland

Niv Ahituv, Israel

Pavel Petrov, Bulgaria

Rami Malkawi, Jordan

Sara Paiva, Portugal

Silvia Parusheva, Bulgaria

Snezhana Sulova, Bulgaria

Todorka Atanasova, Bulgaria

Valentin Kisimov, Bulgaria
Valery Trofimov, Russia
Veselin Popov, Bulgaria
Violeta Kraeva, Bulgaria
Vladimir Sulov, Bulgaria
Vladimir Zanev, USA
Zdzislaw Polkowski, Poland

ORGANIZING COMMITTEE

Chairman: Prof. Vladimir Sulov, PhD

Members:

Prof. Avram Eskenazi, PhD
Prof. Julian Vasilev, PhD
Assoc. Prof. Todorka Atanasova, PhD
Assoc. Prof. Silvia Parusheva, PhD
Assoc. Prof. Snezhana Sulova, PhD
Assoc. Prof. Pavel Petrov, PhD

Secretary: Tijen Talib

FOREWORD

These proceedings contain the papers of the International Conference “Information and Communication Technologies in Business and Education” which took place at the University of Economics – Varna, Bulgaria, 18 October 2019.

The international scientific conference is dedicated to the **50th anniversary of the Department of Informatics at the University of Economics – Varna**. The conference is also dedicated to the 100th anniversary of the University. The included papers describe recent scientific and practical developments in the field of information and communication technologies, information systems, and their applications in business and education.

The papers in the Proceedings are peer reviewed and are checked for plagiarism.

C O N T E N T S

1. Kamelia Stefanova, Dorina Kabakchieva Challenges and Perspectives of Digital Transformation	13
2. Julian Vasilev, Marian Cristescu Approaches for information sharing from manufacturing logistics with downstream supply chain partners	24
3. Abdel-Badeeh M. Salem Computational Intelligence in Smart Education and Learning	30
4. Marian Niedźwiedziński Proposal of the Multi-aspect Analysis of ICT's Needs in Management	41
5. Krasimir Shishmanov Digital Transformation in Bulgaria	52
6. Tatyana Makarchuk, Valery Trofimov, Svetlana Demchenko Modeling the life cycle of e-learning course using Moodle Cloud LMS	62
7. Vanya Lazarova Exploring the functionality of Bulgaria eGovernment site	72
8. Veselin Popov, Petya Emilova Big data and information security	83
9. Luben Boyanov Approaches for enhancing digitalization and digital transformation in supply chain management	91

10. Monika Tsaneva	
Challenges of GDPR compliance in consumer financing companies	103
11. Todorka Atanasova	
Main factors influencing digitization in construction companies	116
12. Silvia Parusheva	
Digitalization and digital transformation in construction – benefits and challenges	126
13. Snezhana Sulova	
The Usage of Data Lake for Business Intelligence Data Analysis	135
14. Pavel Petrov	
Web security technologies used in banks of Estonia, Latvia and Lithuania	145
15. Tanka Milkova	
ABC analysis of inventory in MS Excel	155
16. Ivan Kuyumdzhiev, Radka Nacheva	
Choosing Storage Devices with Consideration of Backup and Restore Performance in MS SQL SERVER	166
17. Yanka Aleksandrova	
Predicting Students Performance in Moodle Platforms Using Machine Learning Algorithms	177
18. Latinka Todoranova, Bonimir Penchev	
M-learning Applications	188
19. Boris Bankov	
The Impact of Social Media on Video Game Communities and the Gaming Industry	198

20. Velina Koleva, Svetoslav Ivanov	
Social media and the recruitment of IT professionals in Bulgarian IT companies	209
21. Deyan Mihaylov	
Modeling and Simulation of Some Functions of Two Independent Random Variables	225
22. Oumar Sy	
Using Association Rules as Semantics for Domain Ontologies	233
23. Ina Stanoeva	
Social media possibilities for improving arts marketing performance	247
24. Iskren Tairov, Vladislav Vasilev	
Perspectives on mobile devices adoption in healthcare sector	257
25. Angelin Lalev	
Deep Neural Networks for Detection of Credit Card Fraud	263
26. Miglena Stoyanova	
Priorities for Digital Transformation in Property Management	275
27. Mihail Radev	
Organizational Variants of IT Department at the University	284
28. Mariya Armyanova	
IoT problems and design patterns which are appropriate to solve them	291
29. Yavor Christov	
Corporate Big Data and Hybrid Integration Platforms	306

30. Ivan Belev	
Making financial data readable using Inline XBRL	314
31. Ivan Belev	
eXtensible Business Reporting Language (XBRL) – a technical overview	322
32. Anna Sobczak	
Entrepreneurship in the education of university students	331
33. Mihaela Markova	
E-commerce in the process of digital transformation: preparedness, requirements and technological tendencies	343
34. Petya Petrova	
Basic Soft Computing Methods In User Profile Modeling	353
35. Petar Dimitrov	
Methods for storing authentication data – a historical review	369
36. Svetoslav Ivanov	
Development stages of starting software company, problems and approaches for software development	382
37. Antonio Hadzhikolev	
The role of the smart city concept in the process of urban transformation	397
38. Oumar Sy	
Ontology-based materialized views for reverse engineering: Toward a set of conceptual guidelines	410

PREDICTING STUDENTS PERFORMANCE IN MOODLE PLATFORMS USING MACHINE LEARNING ALGORITHMS

Yanka Aleksandrova¹

*¹ University of Economics Varna,
Varna, Bulgaria, yalexandrova@ue-varna.bg*

Abstract

E-learning platforms have become a widely used and advanced media to enhance the educational process. They bring benefits to all participants in this process – teachers, students and administration – in several different areas like teaching, learning, communication and sharing. The paper focuses on the application of machine learning algorithms for predicting students' performance based on their interaction with the e-learning platforms. The research hypothesis is that the success or failure on e-learn courses could be predicted using data from activity logs. To support the hypothesis several machine learning algorithms have been performed, such as logistic regression, random forest, gradient boosting decision trees (xgboost) and neural network. The results indicate that all algorithms perform the classification task satisfactory with accuracy above 0.84. The comparison of the evaluation metrics reveals a better performance for neural network and gradient descent boosting trees compared to logistic regression and random forest. The experiments have been performed using R programming language.

***Keywords:** machine learning, e-learn, Moodle, students' performance, gradient boosting, random forest, R language.*

INTRODUCTION

The application of machine learning algorithms in educational electronic platforms has been a research topic for many scientists. Nespereira, Elhariri et al. (Nespereira, et al., 2016) compared the performance of random forest and support vector machines to predict students' tendency to pass/fail as a relationship with their past course interactions with Learning Management Systems (LMS). A software framework has been proposed by Olive, Huynh et al. (Olive, et al., 2018) with implemented predictive model to identify students at risk of abandoning a course. In another research these authors (Olive, et al., 2019) presented a case study model which predicts students at risk of dropping out a Moodle course. The average accuracy achieved with a neural network was 88.81%.

Conijn, Snijders, Kleingeld and Uwe (Conijn, et al., 2017) analysed data from 17 blended Moodle courses to predict student performance from LMS variables using multi-level and standard regressions. Their results showed that predictive modelling strongly vary across courses and suggest low portability of the models across courses. They conclude that a more specific theoretical argumentation is needed to complement the LMS data.

Gamie et al. (Gamie, et al., 2019) developed a model with new features grouped in dimensions according to their relativeness to the teaching style and students' activities on an e-learn system. The prediction analysis has been performed with the best fitting classifier reaching accuracy of 87%. Other authors (Hussain, et al., 2018) implemented machine learning classification and clustering techniques to detect the low-performance students prior the examination. Their experimental results identified that the higher accuracy in identifying the inactive students could be achieved by fuzzy unordered rule induction algorithm (FURIA). K-means clustering could be also implemented to separate active from inactive and poorly performed users.

Nguyen et al. (Nguyen, et al., 2018) proposed a forecast model based on students' interaction with e-learn systems. They tested their model in the Moodle LMS system and achieved an accuracy of over 50% with 75% of students showing outcomes close to the predicted results.

1. METHODOLOGY

The research hypothesis is stated as follows: the success or failure of e-learn courses could be predicted using data from activity logs. For the purpose of the research the students' performance is defined as a success or failure on the final exam after the completion of the course. In order to support the hypothesis several research stages have been defined as depicted on figure 1. The Moodle based platform (e-learn.uevarna.bg) has been used for blended learning with all the learning resources provided only through the platform. There are two main limitations of the current research. The first one is regarding the stated hypothesis which only explores the success or failure of the exam. The second limitation is connected to the input dataset. It contains relatively small

number of observations (112) from only two bachelor programs – “Informatics” and “Business Information Systems”.

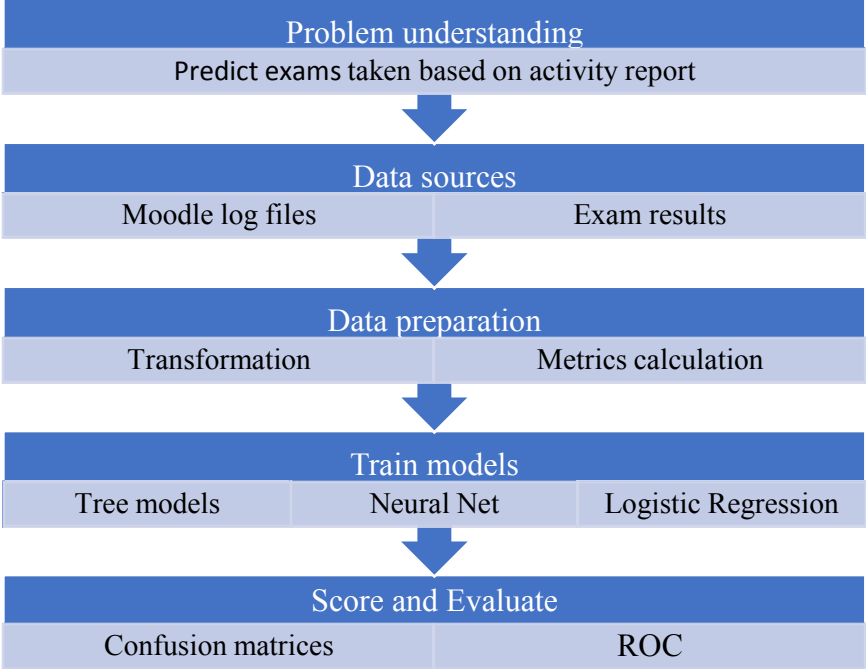


Figure 1. Stages of the research process

The problem is defined according to the research hypothesis. Data sources include activity log from two completed courses in a Moodle based platform and exam results. The structure of the activity log file is as follows: activity log (time, user full name, affected user, event context, component, event name, description, origin, IP address). Data from the log file has been transformed and summarized at a student level to form the structure of the dataset shown on table 1. In order to eliminate differences between courses due to course structure and number of resources provided, some metrics have been calculated as a ratio to the average values. These include tasks, files viewed and files uploaded.

The definition of the students’ performance as a success or failure leads respectively to the chosen algorithms for binary classification –

logistic regression, tree models and neural network. The models have been fitted using cross-validation due to the relatively small number of observations. The trained models have been scored and evaluated using performance metrics as accuracy, sensitivity, specificity, etc.

Table 1

Dataset structure

Variable	Type	Description
Students ID	integer	sequential number
Gender	category	2 levels - F and M
Program	category	currently with two states
Enrolment days after start of semester	integer	[Enrolment date] - [Semester start date]
Days from first to last session	integer	[Date of last session] - [Date of first session]
Unique days count	integer	Count distinct of date
Actions taken	integer	Actions count
Average actions per day	num	[Actions count] / Avg([Actions count for the course])
Actions taken from university network (ratio)	num	[Actions count taken from university]/[Actions count]
Actions taken from outside university network (ratio)	num	[Actions count taken from outside the university]/[Actions count]
Tasks count compared to average	num	[Task actions]/Avg([Task actions])
Files viewed compared to average	num	[Files viewed actions]/Avg([Files viewed actions])
Files uploaded compared to average	num	[Files uploaded actions]/Avg([Files uploaded actions])
Exam taken	category, target	2 levels - 0 (not taken) and 1 (exam taken)

2. PRELIMINARY ANALYSIS

The dataset contains 112 examples of students completed a course of “Design of Information Systems” from two bachelor programs – “Informatics” (69) and “Business Information Systems” (43). The target variable (exam_taken) is relatively equally distributed with 52.68% from

the students successfully passed the exam and 47.32% failed. As a preliminary analysis several boxplots charts of some numerical variables by target have been drawn as depicted on figure 2.

The boxplot charts suggest that some of the variables could be better predictors as there are clear differences in their distribution among the different groups. Such variables are number of actions taken, unique days, tasks from average and files viewed by average. At the same time there is no clear difference in the distribution of actions per day and days from start of the semester to the enrolment date.

The target distribution between the two categorical variables – gender and program is shown using stacked bar chart (figure 3 and figure 4). The target variable is unequally distributed in the two gender groups with 66% from female students successfully passed the exams related to 41.94% from the male students. The students from the two programs performed almost similarly with a slightly greater percentage of students from Informatics taken the exam (56.52%) compared to those from “Business Information Systems” (46.51%).

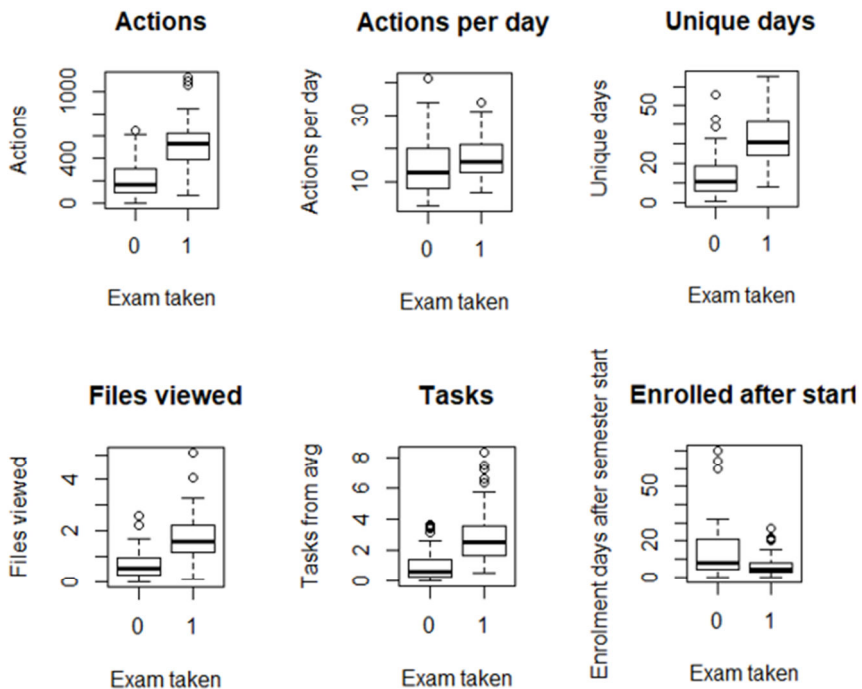


Figure 2. Boxplots of some numeric variables by target

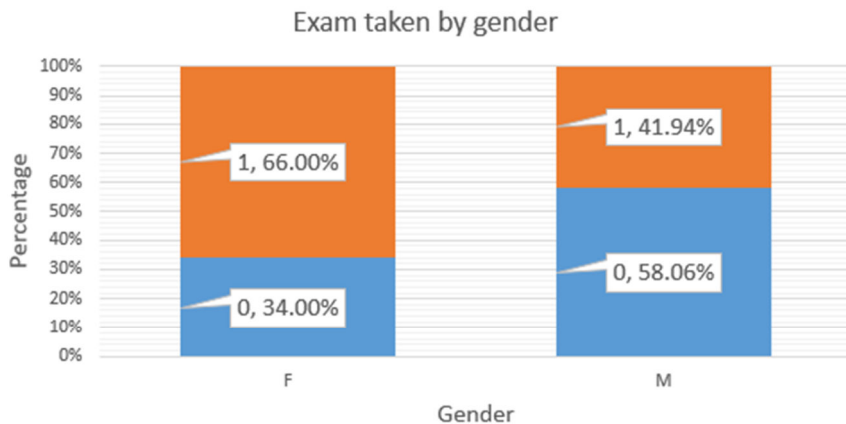


Figure 3. Exam taken by gender (percentage)

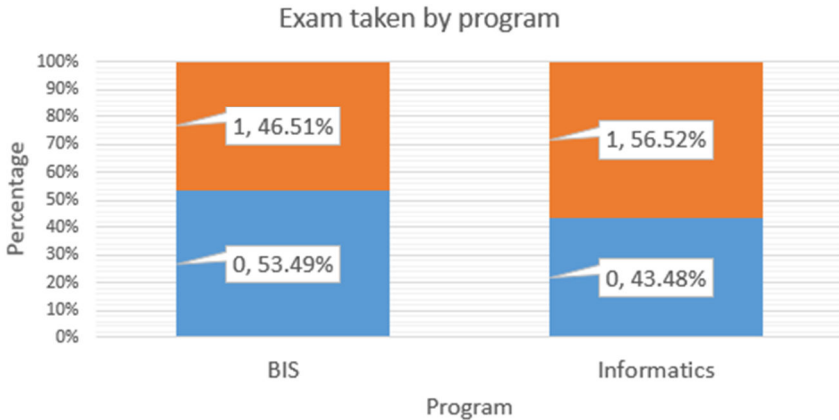


Figure 4. Exam taken by program

3. RESULTS AND FINDINGS

First, a logistic regression has been performed to set the benchmark level for model performance. The variable importance is shown on table 2 with significant variables ($p < 0.05$) flagged by “*”. The most important variables identified by the model are “files uploaded from avg”, “actions uni” and “files viewed from avg”. The logistic regression achieved accuracy of 0.8750 with specificity of 1.0000 and sensitivity of 0.7647.

Table 2

Logistic regression – variable importance

Variable	Overall	Significant
files_uploaded_from_avg	2.8990461	*
actions_uni	1.9134639	*
files_viewed_from_avg	1.7894404	*
programInformatics	1.4277667	
tasks_from_avg	1.3805153	
days_unique	1.1362720	
days_frst_last	0.6005918	
actions	0.4044596	
avg_actions_per_day	0.1968169	
genderM	0.1596336	
enr_days_after_sem_start	0.1322345	

The second applied machine learning algorithm was random forest. The model has been fitted with a 10-fold cross-validation. The parameters have been tuned to number of trees = 2500 and mtry=4. The variable importance derived from random forest algorithm is shown on figure 5. According to MeanDecreaseAccuracy and MeanDecreaseGini the most significant variables are “tasks from avg”, “actions” and “files uploaded from avg”. The model achieved an accuracy of 0.8438, sensitivity 0.8824 and specificity 0.8000.

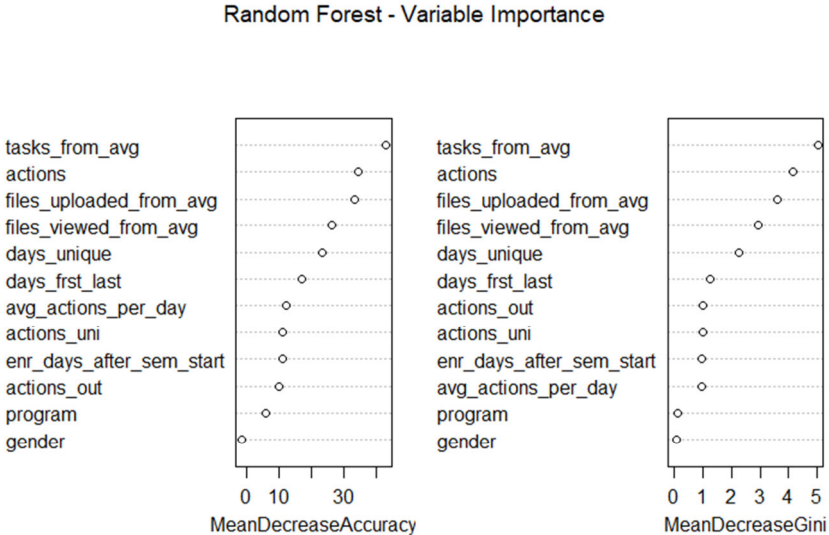


Figure 5. Variable importance derived from Random Forest model

The gradient descent boosting trees based on xgboost package and neural net (12-7-1) performed better than the first two algorithms achieving both accuracy of 0.9062. The xgboost model was hypertuned to nrounds=3, max_depth=20, eta=0.1, gamma=0.2, colsample_bytree=0.6.

Table 3

Evaluation results of applied machine learning algorithms¹

Metric	Random Forest	Logistic Regression	Neural Net	XG Boost
Accuracy	0.8438	0.8750	0.9062	0.9062
95% CI	(0.6721, 0.9472)	(0.7101, 0.9649)	(0.7498, 0.9802)	(0.7498, 0.9802)
No Information Rate	0.5312	0.5312	0.5312	0.5312
p-value [ACC > NIR]	0.000216	4.15E-05	6.19E-06	6.19E-06
Kappa	0.6850	0.7529	0.8110	0.8125
Mcnemar's Test P-Value	1.0000	0.1336	1.0000	1.0000
Sensitivity	0.8824	0.7647	0.9412	0.8840
Specificity	0.8000	1.0000	0.8667	0.9333
Pos Pred Value	0.8333	1.0000	0.8889	0.9375
Neg Pred Value	0.8571	0.7895	0.9286	0.8750
Prevalence	0.5312	0.5312	0.5312	0.5312
Detection Rate	0.4688	0.4062	0.5000	0.4688
Detection Prevalence	0.5625	0.4062	0.5625	0.5000
Balanced Accuracy	0.8412	0.8824	0.9039	0.9078

CONCLUSION

The research results confirm the hypothesis that the success or failure on exams could be predicted using machine learning algorithms on data extracted from students' interaction with e-learning platforms. All from the applied algorithms show satisfactory accuracy with xgboost and neural net with accuracy of 0.9062. Some of the significant variables identified by the models include number of actions taken, tasks from average and files uploaded from average.

¹ Positive class = 1

Related future works would focus on implementing machine learning algorithms on bigger datasets derived from activity log files. Another dimension of future improvement is to apply regression machine learning algorithms to predict the grade mark or exam points. Machine learning algorithms could be also used to predict students with greater probability of failure long before the exam date which could optimize the educational process.

REFERENCES

1. CONIJN, R., SNIJDERS, C., & KLEINGED, A. (2017). *Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS*. IEEE Transactions on Learning Technologies.
2. GAMIE, E., EL-SEOUD, M., SALAMA, M., & HUSSEIN, W. (2019). *Multi-dimensional analysis to predict students' grades in higher education*. International Journal of Emerging Technologies in Learning, 14(2), 4-15. doi:10.3991/ijet.v14i02.9905
3. HUSSAIN, S., HUSSAIN M., ZHANG, W., ZHU, W., THEODOROU, P., & ABIDI, S. (2018). *Mining moodle data to detect the inactive and low-performance students during the moodle course*. 2nd International Conference on Big Data Research, ICBDR 2018, (pp. 133-140). doi:10.1145/3291801.3291828
4. NESPEREIRA, C., ELHARIR, E., EL-BENDARY, N., & VILAS, A. (2016). *Machine learning based classification approach for predicting students performance in blended learning*. Advances in Intelligent Systems and Computing, 407, 47-56. doi:10.1007/978-3-319-26690-9_5
5. NGUYEN, V., NGUYEN, Q., & NGUYEN, V. (2018). *A model to forecast learning outcomes for students in blended learning courses based on learning analytics*. 2nd International Conference on E-Society, E-Education and E-Technology, ICSET 2018, (pp. 35-41). doi:10.1145/3268808.3268827
6. OLIVE, D., HUYNH, D., REYNOLDS, M., DOUGIAMAS, M., & WIESE, D. (2018). *A supervised learning framework for learning management systems*. First International Conference on Data Science, E-learning and Information Systems. doi:10.1145/3279996.3280014

7. OLIVE, D., HUYNH, D., REYNOLDS, M., DOUGIAMAS, M., & WIESE, D. (2019). *A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC*. Journal of Computing in Higher Education. doi:10.1007/s12528-019-09230-1