Analysis of Student Behavior and Success Based on Logs in Moodle

Nikola Kadoić* and Dijana Oreški**
* Faculty of Organization and Informatics/Department of Organization, Varaždin, Croatia
** Faculty of Organization and Informatics/Department of Information Systems Development, Varaždin, Croatia
nkadoic@foi.hr, dijoresk@foi.hr

Abstract – Today, it is almost impossible to implement teaching processes without using information and communication technologies (ICT), especially in higher education. Education institutions often use learning management systems (LMS), such as Moodle, Edmodo, Canvas, Schoology, Blackboard Learn, and others. When accessing these systems with their personal account, each student's activity is recorded in a log file. Besides analyzing the raw data from log files directly, there is an option to use Moodle plugins that provide learning analytics and enable the faster analysis of students' behavior on LMS. In this paper, some of these plugins are presented. However, this paper is focused on analyzing the log files of a course implemented on the LMS Moodle at the Faculty of Organization and Informatics at the University of Zagreb. The results of the students' behavior, based on logs in Moodle, will be interpreted in terms of student success.

Keywords - LMS, Moodle, log files, student success, student behavior

I. Introduction

With the development and expanding availability of the Internet, different systems for electronic learning (e-learning) have also been developed, implemented, and used. At first, those systems, also called learning management systems (LMS), were very simple—they were basically just web pages listing the main information about study programs, courses, teachers, and contact methods. A few years later, however, more serious LMS began to be developed [1]. Developing LMS became an area of scientific research: in 2003, more than 40 academic journals were specializing in related topics [2]. LMS enabled the creation of personal profiles for students, the sharing of teaching materials, and most important, online communication among students and with teachers. LMS were mainly used to support face-to-face teaching [3].

Today, besides being a support for face-to-face teaching, LMS aid in the teaching of fully online courses and, even more extensive, fully online study programs [4]. In most cases, online teaching is related to the area of the higher education. Students can become masters in specific fields without leaving their home or physically visiting an educational institution. The whole teaching process, from enrolling in a study program, to attending classes, to teaching classes, to taking examinations, is done entirely online. In addition, some new phenomena in teaching activities have been developed as a result of LMS and online learning trends, such as Massive Online Open Courses (MOOCs) or Open Educational Resources (OER) [5]. Even though contemporary trends lean toward leaving the physical classroom entirely, in some scientific areas, that is not possible. However, very creative ways of e-learning are still being developed for such situations. They include combining LMS with web 2.0 tools [6] and even changing the teaching paradigm. One such example is a flipped classroom: students are taught theory at home and come to the classroom to work on projects with teachers and other students, and they then apply the theory they learned to solve problems [7]. The concept of the flipped classroom is opposite to that of the usual teaching process, where the basics are learned in the classroom and practical assignments are left as homework.

In Croatia, many higher education institutions use LMS to support teaching process. The most used LMS in Croatia, and specifically at the University of Zagreb, is Moodle (which is available as Merlin to all higher education institutions in Croatia, [8]). Depending on the number and intensity of the online activities that are supported by Moodle and used in practice, courses at the University of Zagreb can be categorized into one of three levels related to the application of e-learning in education [9], [10]. The highest level, Level 3, requires that all teaching materials be available online; the organization of the teaching process enables individualization within the process; and students assume an active role in terms of achieving outcomes and and other requirements. At this level, the teaching process can be conducted entirely online or be implemented as blended learning. Based on the document "Criteria for the Evaluation of Online Study Programs," prepared by the National Council for Higher Education in 2013, an online study program is a program in which a minimum of 50 percent of the courses are done online, and a course is considered to be an online course when a minimum of 50 percent of the teaching hours are done online [11].

Even though Level 3 represents very advanced e-courses in terms of e-learning, there are additional opportunities for increasing the application of e-learning in education. Current trends in Croatia have led to the implementation of a concept called learning analytics (LA) [12], [13]. When accessing an LMS with their personal account, students create a digital profile that is saved in LMS log files. The main idea of LA is to analyze the raw data in log files and generate new knowledge (often through dashboards [14]) about students' behavior. This knowledge can then be connected to students' success in order to generate ideas for new activities that will decrease the negative effects of student behavior on student success and increase the positive effects of student behavior on student success. For example, if an analysis showed that the students who accessed teaching materials only one day before the test (student behavior) had unsatisfactory test results (student success), it is necessary to create activities that will influence them to access teaching materials earlier. However, LA is about even more than this; the goal is to analyze all the data retrieved through any system (e.g., LMS systems, library systems, e-portfolio systems, or student services systems), find connections between variables in log files, and make modifications to improve the teaching process (e.g., increasing the quality of lessons, increasing student satisfaction, and decreasing the dropout rate) [15].

In this paper, we are focused on the LMS Moodle. In Section 2, we will describe the LMS Moodle in greater detail, especially its plugins that are originally oriented to LA. In Section 3, we will present some papers in which student behavior has been analyzed using LA. Finally, in Section 4, we will present the most exciting results from the log file analysis of several courses at the Faculty of Organization and Informatics of the University of Zagreb.

II. The LMS Moodle and LA plugins

As said earlier, the LMS Moodle is one of the most used LMS in practice. Moodle is organized through plugins that represent specific activities and extra features that can be added to the fundamental feature of the Moodle: courses. The most used Moodle plugins, which come with basic Moodle installation, are Assignment, Attendance, Choice, Lesson, Page, Quiz, URL, Workshop, Folder, File, Glossary, SCORM Package, Feedback, and Database [16].

Here are some of the plugins that can be used in terms of LA (Moodle 2.9): Logs, Activity, Activity Completion, Live Logs, (Quiz) Statistics, (Course) Participation, Survey, Course Overview, Course Completion Status, Progress Bar, Events List, Activity Results Block, Configurable Reports, (Gradebook) Overview, Ad-hoc Database Queries, Engagement Analytics, Course Dedication, Graph Stats, GISMO, and Graphical Interactive Student Monitoring [17]. The current version of Moodle (3.4) includes the plugin Inspire Analytics. One of the most exciting features of this plugin is a model that predicts students who are at risk of non-completion (dropping out) of a Moodle course, based on low student engagement [18].

Additionally, there is always an option to develop your own plugin or adjust an existing one. One LA plugin contributed by external contributors is the SmartKlass™ Learning Analytics Moodle. Some of the features of this plugin include the identification of students lagging behind and the identification of students for which course content is not challenging enough [19]. If heatmap analysis needed, the plugins MAV [20] and Heatmap [21] can be helpful.

There are many other plugins available online, but before installing any of them, a more in-depth analysis is recommended. Plugins are installed on the level of the server, so it might be inappropriate to install a new plugin on a specific site if only 1% of the site managers and teachers will use it. Still, the analysis of log files can always be done.

III. Related Work

Components of LMS offer various opportunities for the improvement of student learning and, accordingly, can impact students' final grades [22]. An LMS stores data about participants' actions in log files, which gives researchers information about student behavior. A log is a list of a students' events in which each line contains a time-stamp as well as one or more fields that hold information about the activity performed [23]. A Moodle log consists of the time and date it was accessed, the Internet Protocol (IP) address from which it was accessed, the name of the student, each action completed (i.e., view, add, update, or delete), the activities performed in different modules (e.g., the forum, resources, or assignment sections), and additional information about the action [23]. The stored data can be useful for data mining algorithms.

Romero, López, Luna and Ventura [24] noticed a trend toward the combined use of data mining learning techniques for the analysis of activity data. Having the data in an LMS provides various opportunities for the use of data mining methods to examine them [25]. Data mining can be useful to explore, visualize, and analyze data with the aim of identifying useful patterns in order to understand students' learning behavior and feedback. Teachers can then use this information when designing instruction and delivery. Data mining includes tasks and methods concerning statistics, visualization, clustering, classification, association rule mining, and sequential pattern mining [26]. Baker and Yacef [27] also recognized data mining as a promising field for the exploration of data from computational educational settings. Romero, Ventura, and García [28] describe the process of mining e-learning data step by step, with the guidelines about how to use data mining techniques for mining Moodle data. Kazanidis, Valsamidis, Theodosiou, and Kontogiannis [29] also proposed a platform-dependent framework for processing and analyzing data from LMS. Their framework consists of three steps: logging the data, data pre-processing, and data mining. Mainman and Rokach [30] divided data mining techniques for educational data into two groups: verification-oriented techniques (which rely on traditional statistics such as hypothesis tests, the goodness of fit, and the analysis of variance) and discovery-oriented techniques (which rely on prediction and description, such as classification, clustering, prediction, relationship mining, neural network, and web mining). Many authors have reported highly accurate predictions using different classification algorithms such as C4.5, EM, Naïve Bayes, and support vector machines [24].

Research in this area has recently become very significant. The key components of this literature review are the data and methods (tools and techniques) used to explain student behavior in LMS. A brief overview is presented in the following subsection, in which we cite the most current articles on the given topic. Cocea and Weibelzahl [31] identified learner motivation as a key factor in the quality of learning. One aspect of motivation is engagement. Thus, they analyzed 10 attributes related to pages, tests, hyperlinks (number of hyperlinks, average time), and glossary (number of times accessed, average time). Mogus, Djurdjevic, and Suvak [32] examined data (activity logs) obtained by students to detect the frequencies and priorities of students' choice of activities in the LMS. Their research aimed to determine whether students' activity logs correlate with their final marks. Observed activities included course view, assignment view, resource view, forum view, assignment upload, project upload, and final mark. The highest influences on students' marks were assignment view, course view, forum view, and resource view. The authors concluded that those four activities particularly influence learning effectiveness. The lowest number of logs were found to be the activities concerning forum view and discussion view. This suggests a need to encourage students to engage in communication within LMS. The authors' suggestion for future research includes reviewing a broader range of factors, a larger sample of students, a higher number of courses, and a greater range in the types of courses evaluated by studying students from different disciplines to investigate the influence of participants' gender.

Jovanovic, Vukicevic, Milovanovic, and Minovic [33] defined a classification model to predict whether a student would achieve excellent performance in a course. They defined clustering models that would detect groupings of students with respect to their overall performance and used a k-means clustering algorithm adapted for use over categorical data. Models were based on the following data: number of quizzes passed or failed; number of messages sent or read on the forum; total time spent on assignments, quizzes, and forum; and final mark obtained by the student in the course. Kotsiantis [34] used regression techniques to predict students' marks in a distance learning system. The stored data (virtual courses and e-learning log file) were shown to be useful in predicting marks. In 2013, Kotsiantis, Tselios, Filippidi and Komis [35] investigated ten different variables related to the students' activities: assignment_view, course_view, forum_add_post, forum_view, glossary_view, questionnaire_view, resource_view, user_view, and final course grade. Bovo, Sanchez, Heguy, and Duthen [36] grouped students by mining Moodle log data. The first objective was to define relevant clustering features. The second objective was to determine whether their students showed different learning behaviors. Their application uses data mining and machine learning methods to solve the problem of monitoring students in e-learning. Data used in the research consisted of login frequency; the last login time; time spent online; the number of lessons read and downloaded; the number of quizzes, crosswords, and assignments completed; the average grade obtained; and the number of forum topics read and created.

Recently, Gašević, Dawson, Rogers, and Gasevic [37] investigated the extent to which instructional conditions influence the prediction of academic success in a blended learning model. Variables from the LMS included information about the usage of the following Moodle features: forums, course logins, resources, assignments, book, quizzes, feedback, lessons, and chat. The data represented the number of times students used a particular feature. Compared to forums, course logins, resources, and assignments, features such as quizzes, feedback, lessons, and chat were not accessed by a substantial number of students. These were dichotomized into "accessed" and "did not access" categories. Statistical analysis included ANOVA, Chi-square test, and linear regression to explore the association between students' online interactions and students' marks.

IV. Research Description and Results

In this section, we present a case study of a Moodle course log collected from an actual class. This study examines activity logs from when students log in to the course, with the aim of detecting frequencies and finding patterns in students' choice of activities. The study aims to detect the relationship between the observed variables and students' final grades. The objective is to find out the impact of a particular activity in an LMS on a final grade. The following two research questions are posited: RQ1: To what extent are individual variables derived from log data a reliable predictor of academic success? RQ2: What is the level of similarity in student LMS usage between genders? To answer these research questions, a Moodle course log was collected after the spring semester of a blended course called "Business Decision Making". A total of 73 students registered with Moodle, and 180-minute teaching lectures and seminars were held every Wednesday.

Fig. 1 shows the distribution of logs by the overall grade students achieved in the course. The highest number of logs is achieved by the students with the highest grades, 4 and 5. To investigate if there is any correlation between specific activities on the LMS and student grades, we have performed correlation analysis.

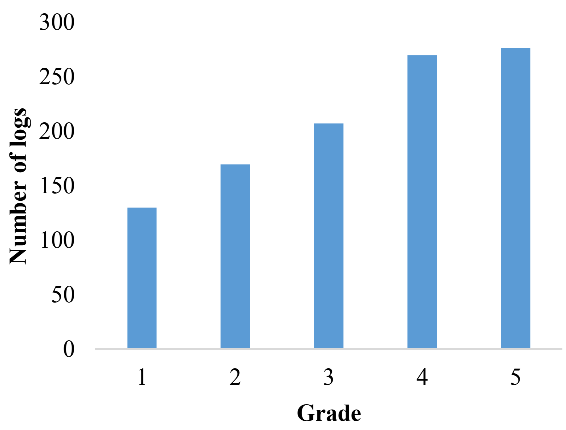

Figure 1. Frequency of logs by grade

TABLE 1. CORRELATION BETWEEN VARIABLES (*P < 0.05)

Variable	*Grade*	*File usage*	*Forum usage*	*Link usage*	*Assignment uploads*
Grade	-	0.35*	0.18	-0.01	0.08
File usage	0.35*	-	0.41*	0.05	0.19
Forum usage	0.18	0.41*	-	0.13	0.38*
Link usage	-0.01	0.05	0.13	-	0.09
Assignment uploads	0.08	0.19	0.38*	0.09	-

Table 1 examines correlations between achievement over the span of the course (measured by grades) and effort in files, forums, and link usage, as well as the assignments uploaded. The results indicate a statistically significant correlation among students' grades and the opening of files. The correlation is positive, which indicates that students with a higher frequency of file openings have higher grades. There is a lack of association between grades and other logs in the course. File opening is correlated with activities on the forum, demonstrating that students who are active in forum discussions opened files more often.

We have also examined students' activity by gender (Fig. 2). Female students have a higher number of logs than their male colleagues. Differences between genders is also visible in the average grade received. Female students have a higher average grade than male students.

According to experience of teachers of the course currently being studied, students often complete their assignments (e.g., write essays, answer questions, or study for exams) in the last moments before a deadline. The idea of next analysis is to determine the distribution of students' logs on the system on the days before the two exams.

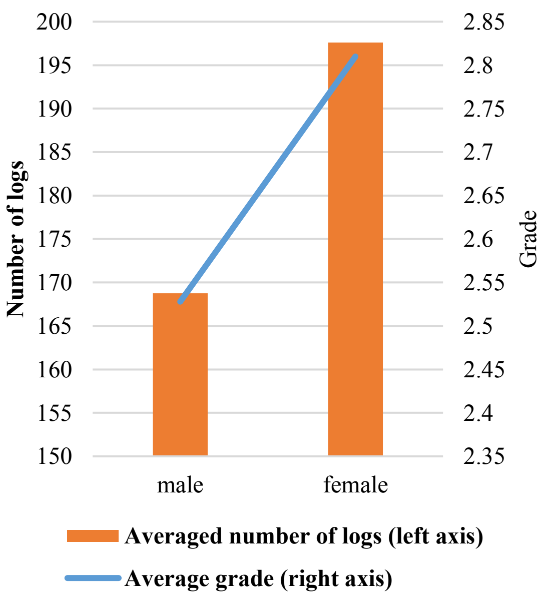

Figure 2. Analysis by gender

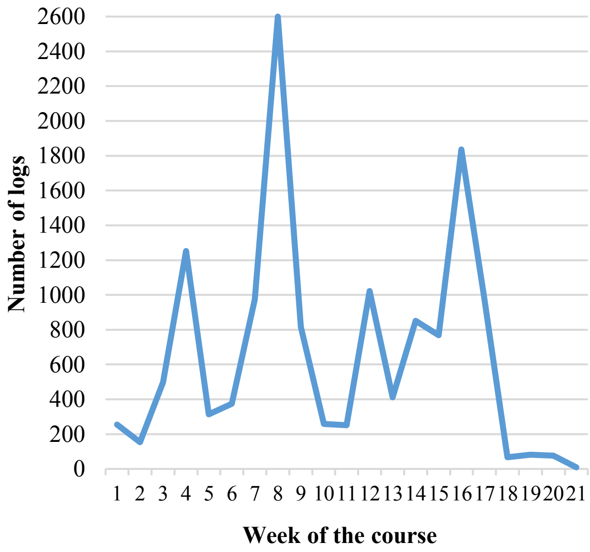

Figure 3. Distribution of logs per week

Fig. 3 presents the number of logs from the 1st to 21st week of the course. The 8th and 16th week of classes are the weeks during which tests were performed. The first test was held on 04-19-2017 (the 8th week of the course), and the second was held on 06-14-2017 (the 16th week of the course). We studied only these two weeks, and specifically, only the two days on which the tests were held. Weekly analysis showed that the highest number of logs appear on the day before test days.

Fig. 4 indicates course logs related to the days before test days. Page views occurred mostly between 17:00 and 20:00 on the day before the first test day (04-18-2017) and between 15:00 and 16:00 on the day before the second test day (06-13-2017). In both cases, there is a significant number of logs in the late hours of the day. During these times, the most students downloaded test materials and started to study. However, course activity during the late hours is more conspicuous for the first test than it is for the second.

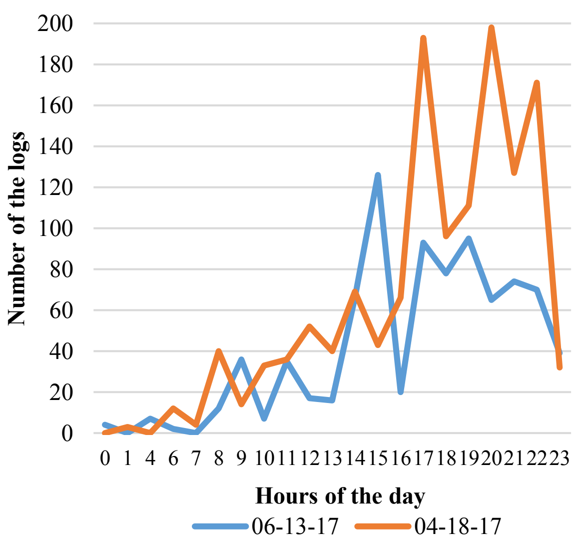

Figure 4. Distribution of logs on the days before exams

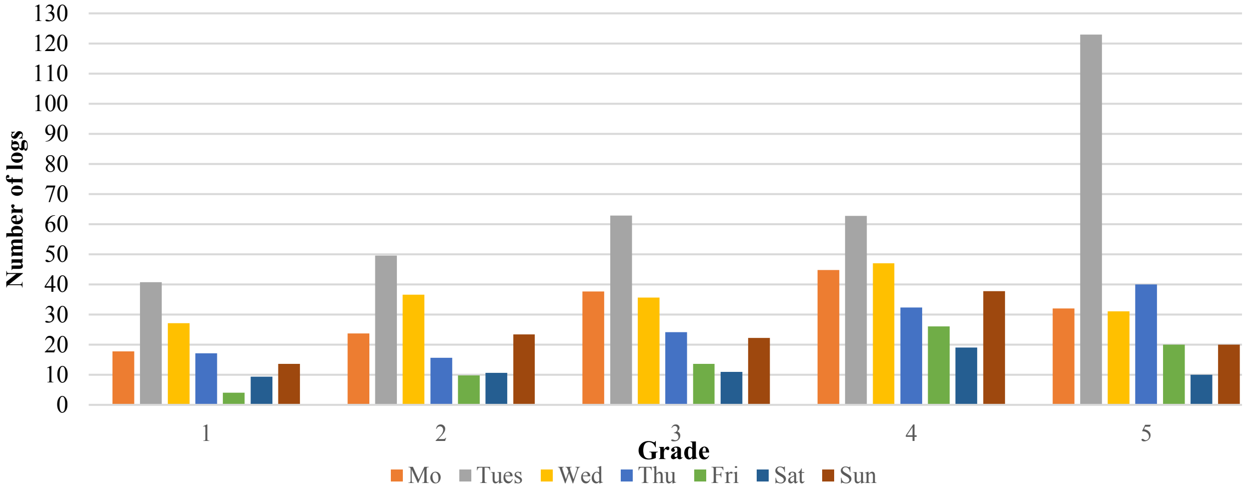

Figure 6. Distribution of logs per day in week with respect to achieved grade

Similarly, the time-focused analysis results for the whole course period level are presented in Fig. 5. It shows the hours of the day during which students were logged into the course, and the opening is concentrated from 11:00 on. During the afternoon, a number of logs persist, but this decreases in the evening. In the period after the midnight, there is no activity on the e-course. However, this does not mean that students were not active on the course—some may have performed an offline assignment after downloading it earlier.

Figure 6 presents an analysis of students' activities with respect to the grade they achieved and the day in the week. Surprisingly, for the students with the highest grade, most of the course activity was done on the day before lectures, seminars, and tests.

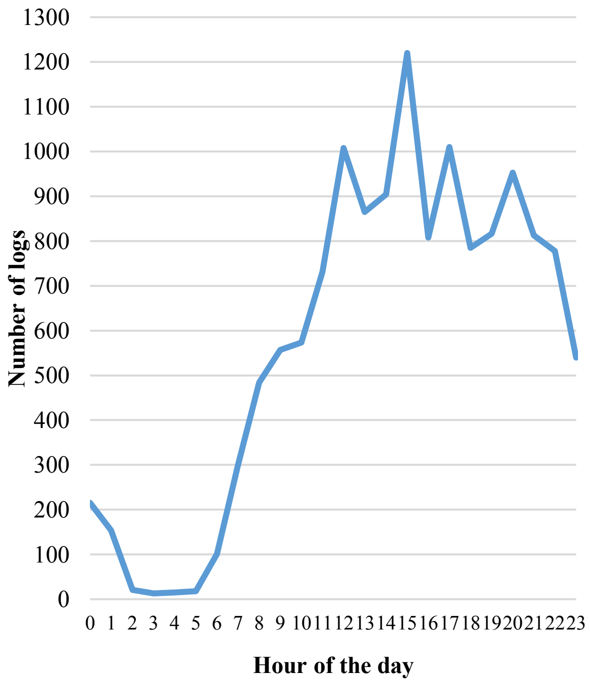

Figure 5. Distribution of logs per day hours

V. Conclusion

Our results are potentially beneficial in the early detection of students experiencing difficulties in a course. Both teachers and students benefit from this kind of research, as teachers can identify excellent students for collaboration and students find out how to give greater effort to obtain good results.

In the conducted research, the female students are more active and successful in the course than are the male students. There is a correlation between the number of logs in the e-course and the final grades. The students were most active in the test weeks and, specifically, on the day before the tests. Students can be characterized as "last-minute" students, as they perform their obligations as late as possible in terms of the deadline and are active in the late hours. However, this cannot be generalized because the research was conducted in only one course. Also, the research covered only informatics students. In future research, the analysis will be performed across several courses. Additionally, students from other disciplines, not only informatics, will be included in future research.

Acknowledgment

Croatian Science Foundation has partly supported this paper under the project Higher Decision, IP-2014-09-7854.

References

[1] H. Coates, R. James, and G. Baldwin, "A Critical Examination of the Effects of Learning Management Systems on University Teaching and Learning," *Tertiary Education and Management*, vol. 11, no. September, pp. 19–36, 2005.

[2] M. Dougiamas and P. C. Taylor, "Moodle: Using Learning Communities to Create an Open Source Course Management System."

[3] N. Sclater, "There are now more than 40 academic journals specialising in these topics." Center for Applied Research, Colorado, p. 13, 2008.

[4] J. J. Downing and J. E. Dyment, "Teacher Educators' Readiness, Preparation, and Perceptions of Preparing Preservice Teachers in a Fully Online Environment: An

Exploratory Study," *The Teacher Educator*, vol. 48, no. 2, pp. 96–109, Apr. 2013.
[5] S.-W. Kim, "MOOCs in Higher Education," in *Virtual Learning*, InTech, 2016.
[6] M. Á. Conde, F. J. García-Peñalvo, M. J. Rodríguez-Conde, M. Alier, M. J. Casany, and J. Piguillem, "An evolving Learning Management System for new educational environments using 2.0 tools," *Interactive Learning Environments*, vol. 22, no. 2, pp. 188–204, Mar. 2014.
[7] J. Bishop and M. Verleger, "Testing the flipped classroom with model-eliciting activities and video lectures in a mid-level undergraduate engineering course," *Proceedings - Frontiers in Education Conference, FIE*, pp. 161–163, 2013.
[8] T. Radobolja, S. Kučina Softić, and A. Jurjević, "1-2-3 Moodle na primjeru sustava za e-učenje Merlin." Sveučilišni računski centar, Centar za e-učenje, Zagreb, p. 26, 2016.
[9] Povjerenstvo za e-učenje Sveučilišta u Zagrebu, "Odluka o oblicima sveučilišne nastave prema razini primjene tehnologija e-učenja." Sveučilište u Zagrebu, Zagreb, p. 6, 2009.
[10] L. Pinter *et al.*, "E-learning Strategy 2007-2010." University of Zagreb, Zagreb, p. 22, 2007.
[11] N. Kadoic and N. Begičević Redep, "Methodology for online study program development - case study ePITUP," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, no. May, pp. 948–953.
[12] B. Divjak and P. Vondra, "Learning Analytics: Meeting the Needs of Students and Teachers in Pre-tertiary Education."
[13] M. Maretić and B. Divjak, "Learning Analytics for Peer-assessment: (Dis) advantages , Reliability and Implementation Learning Analytics for Assessment : State of the Art," *Journal of Information and Organizational Sciences*, vol. 41, no. 1, pp. 21–34, 2017.
[14] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos, "Learning Analytics Dashboard Applications," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1500–1509, Oct. 2013.
[15] T. Anya and E. Lias, "Learning Analytics: Definitions, Processes and Potential," 2011.
[16] Moodle, "Features - MoodleDocs," *Moodle.org*, 2017. [Online]. Available: https://docs.moodle.org/34/en/Features. [Accessed: 09-Jan-2018].
[17] Moodle, "Learning analytics - MoodleDocs," *Moodle.org*, 2015. [Online]. Available: https://docs.moodle.org/29/en/Learning_analytics. [Accessed: 09-Jan-2018].
[18] Moodle, "Analytics - MoodleDocs," *Moodle.org*, 2015. [Online]. Available: https://docs.moodle.org/34/en/Analytics. [Accessed: 09-Jan-2018].
[19] Oscar Klassdata and Angel KlassData, "Moodle plugins directory: SmartKlass^{TM} Learning Analytics Moodle," *Moodle Docs*, 2016. [Online]. Available: https://moodle.org/plugins/local_smart_klass. [Accessed: 09-Jan-2018].
[20] "The Moodle Activity Viewer (MAV) – Heatmaps of Student Activity – Damo's World," 2013. [Online]. Available: https://damos.world/2013/08/30/the-moodle-activity-viewer-mav-heatmaps-of-student-activity/. [Accessed: 09-Jan-2018].
[21] Dan Marsden, "Moodle plugins directory: Attendance," *Moodle.org*, 2016. [Online]. Available: https://moodle.org/plugins/block_heatmap. [Accessed: 09-Jan-2018].
[22] C. Pislaru and R. Mishra, "Using VLEs to support student centred learning in Control Engineering Education," no. January, pp. 457–461, 2014.
[23] C. Romero, J. R. Romero, S. Ventura, and Á. D. Á. Data, *Educational Data Mining*, vol. 524. 2014.
[24] C. Romero, M. I. López, J. M. Luna, and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," *Computers and Education*, vol. 68, pp. 458–472, 2013.
[25] S. Preidys and L. Sakalauskas, "Analysis of students' study activities in virtual learning environments using data mining methods," *Technological and Economic Development of Economy*, vol. 16, no. 1, pp. 94–108, 2010.
[26] D. Monk, "Using data mining for e-Learning decision making," *Electronic Journal of e-Learning*, vol. 3, no. 1, pp. 41–54, 2005.
[27] R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009 : A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–16, 2009.
[28] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," *Computers and Education*, vol. 51, no. 1, pp. 368–384, Aug. 2008.
[29] I. Kazanidis, S. Valsamidis, T. Theodosiou, and S. Kontogiannis, "Proposed framework for data mining in e-learning: the case of Open e-class," *Proc IADIS Applied*, pp. 254–258, 2009.
[30] L. Rokach and O. Maimon, "The Data Mining and Knowledge Discovery Handbook: A Complete Guide for Researchers and Practitioners." New York: Springer, 2005.
[31] M. Cocea and S. Weibelzahl, "Log file analysis for disengagement detection in e-Learning environments," *User Modeling and User-Adapted Interaction*, vol. 19, no. 4, pp. 341–385, 2009.
[32] A. M. Mogus, I. Djurdjevic, and N. Suvak, "The impact of student activity in a virtual learning environment on their final mark," *Active Learning in Higher Education*, vol. 13, no. 3, pp. 177–189, 2012.
[33] M. Jovanovic, M. Vukicevic, M. Milovanovic, and M. Minovic, "Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study," *International Journal of Computational Intelligence Systems*, vol. 5, no. 3, pp. 597–610, Jun. 2012.
[34] S. B. Kotsiantis, "Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades," *Artificial Intelligence Review*, vol. 37, no. 4, pp. 331–344, 2012.
[35] S. Kotsiantis, N. Tselios, A. Filippidi, and V. Komis, "Using learning analytics to identify successful learners in a blended learning course," *International Journal of Technology Enhanced Learning*, vol. 5, no. 2, p. 133, 2013.
[36] A. Bovo, S. Sanchez, O. Heguy, and Y. Duthen, "Clustering moodle data as a tool for profiling students," *2013 2nd International Conference on E-Learning and E-Technologies in Education, ICEEE 2013*, vol. 2013, no. September, pp. 121–126, 2013.
[37] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *Internet and Higher Education*, vol. 28, pp. 68–84, 2016.