MoodleMiner: Data Mining Analysis Tool for Moodle Learning Management System

MoodleMiner: Moodle Öğrenme Yönetim Sistemi için Veri Madenciliği Analiz Aracı

Gökhan AKÇAPINAR, *Hacettepe University, Department of Computer Education & Instructional Technology, gokhana@hacettepe.edu.tr,* ORCID: 0000-0002-0742-1612
Alper BAYAZIT, *Yeditepe University, Department of Computer Education & Instructional Technology alper.bayazit@yeditepe.edu.tr,* ORCID: 0000-0003-4369-587X

Abstract: The purpose of this study is to develop a tool by which non-experts can carry out basic data mining analyses on logs they obtained via the Moodle learning management system. The study also includes findings obtained by applying the developed tool on a data set from a real course. The developed tool automatically extracts features regarding student interactions with the learning system by using their click-stream data, and analyzes these data by using the data mining libraries available in the R programming language. The tool has enabled users who do not have any expertise in data mining or programming to automatically carry out data mining analyses. The information generated by the tool will help researchers and educators alike in grouping students by their interaction levels, determining at-risk students, monitoring students' interaction levels, and identifying important features that impact students' academic performances. The data processed by the tool can also be exported to be used in various other analyses.

Keywords: Educational data mining, learning analytics, Moodle, R, learning management system, log analysis tool

Öz (Abstract in Turkish): Bu çalışmanın amacı, uzman olmayan kişiler için Moodle öğrenme yönetim sisteminden elde ettikleri log kayıtları üzerinde temel veri madenciliği analizleri yapabilecekleri bir araç geliştirmektedir. Aynı zamanda geliştirilen aracın gerçek bir derse ilişkin veri seti üzerinde uygulanması sonucu elde edilen bulgulara da yer verilmiştir. Geliştirilen araç, öğrencilerin tıklama verilerinden öğrenme sistemi ile etkileşimlerine yönelik değişkenleri otomatik olarak oluşturmakta ve arka planda R programlama dilinde yer alan veri madenciliği kütüphanelerini kullanarak bu verileri analiz etmektedir. Araç sayesinde veri madenciliği ya da programlama konusunda uzman olmayan kullanıcıların veri madenciliği analizlerini otomatik olarak gerçekleştirmesi olanaklı hale gelmiştir. Araç tarafından üretilen bilgiler, öğrencilerin etkileşim düzeylerine göre gruplara ayrılmasında, risk altındaki öğrencilerin belirlenmesinde, öğrencilerin etkileşim düzeylerinin izlenmesinde ve öğrencilerin akademik performansları üzerinde etkisi olan önemli değişkenlerin belirlenmesinde araştırmacılara ve eğiticilere yardımcı olacaktır. Aynı zamanda araç tarafından işlenen veriler farklı analizlerde kullanılmak üzere dışarıya aktarılabilmektedir.

Anahtar Sözcükler (Keywords in Turkish): Eğitsel veri madenciliği, öğrenme analitikleri, Moodle, R, öğrenme yönetim sistemi, log analiz aracı

INTRODUCTION

Data mining is used in numerous fields for the purpose of discovering hidden patterns and making predictions by utilizing the available data. There is an ever-increasing number of successful implementations in the field of education. The increasing amount of stored data on students learning processes and the tools used in analyses becoming more widely-utilized are among its most significant reasons. Its main areas of use in education can be listed as follows: predicting the academic performances of students, developing early warning systems for at-risk students, grouping students with similar attributes, or developing educational recommender systems.

Data stored in the databases of online learning environments constitute a significant portion of the data resources used in educational data mining studies (Peña-Ayala, 2014). However, generating learning-related features from these records that are kept in learning record stores or relational databases, and carrying out data mining analyses on this data is a time-consuming and challenging process for non-experts (Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Chatti et al., 2014).

The purpose of this study is to develop a tool (MoodleMiner) by which users who are inexperienced in programming and data mining can carry out basic data mining analyses on log records they obtained via the Moodle learning management system. In this way, educators will be able to obtain data-driven information on the status of the learning environment and students, while researches will be able to seek answers for research questions regarding online learning. The tool to be developed will carry out the following operations:

- Preprocessing on log data regarding learning processes and generating features that reflect student interactions.
- Carrying out the data mining analyses predefined on the created data set and reporting the results.

The tool will allow automatic analysis of log records, which in turn will facilitate carrying out data mining analyses. Since the tool to be developed is aimed to cater to numerous researchers and educational institutions, the Moodle learning management system was selected as the online learning environment. Moodle is one of the most widely used online learning environments by educational institutions. According to the available data, it is used by over 138 million registered users in 230 countries[1]. Furthermore, it is distributed for free with an open source code. It offers educators effective tools for providing course materials to students and organizing online learning activities.

Analysis Tools Developed Previously for Moodle

Various free and paid tools have been developed for the purpose of analyzing Moodle data. Some of these tools only carry out statistics and visualization, some have been developed exclusively for a specific analysis such as social network analysis, while some have been produced to carry out certain data mining analyses. These tools can be divided into two categories with regards to their access to Moodle data: 1-Tools that have been usually developed as a desktop application and operate independently from Moodle. 2-Tools that have been developed as Moodle add-ons and are integrated into Moodle.

Desktop tools were the first to be developed among the educational data mining and learning analytics tools. One of such tools, CoSyLMSAnalytics was developed to provide students with feedback, and track and assess their progress individually and as a group (Retalis, Papasalouros, Psaromiligkos, Siscos, & Kargidis, 2006). The purpose of the tool is to mainly generate reports and uncover patterns based on students' browsing behavior on Moodle environment. Another such tool, ViMoodle, is a visual analysis tool developed for Moodle. It shows social networks by applying clustering and association rule mining algorithms on log

---

[1] https://moodle.net/stats/

records in the database related to exams, activities, forum and resources, and visualize success through graphs. Thus, it facilitates the extraction of certain information from education and learning processes, and the tracking of this information (Aguilar, Therón, & Peñalvo, 2008). Another similar tool, CIECoF (Continuous Improvement of E-Learning Courses Framework), offers recommendation to educators for better organization of Moodle courses. The tool, which applies association rules on student data, was developed to find, share and recommend the most suitable paths to boost course effectiveness (García, Romero, Ventura, & De Castro, 2009). Meerkat-ED is a tool that analyzes student contributions on Moodle discussion forums. It offers information on the examination and interpretation of contributions on online discussion forum by applying social network analysis on the data (Rabbany, Takaffoli, & Zaïane, 2011). The tool named MDM, on the other hand, applies data mining methods such as classification, association rules etc. on exported log records on the Moodle environment (Perez, Romero, & Ventura, 2010). DRAL, another desktop tool, aims to identify what Moodle activities are necessary for a student to pass the course. It makes success/fail predictions by using the classification method (Zafra, Romero, & Ventura, 2013).

One of the tools that works integrated with the Moodle system, GISMO, is an interactive student monitoring and tracking system. It uses the tracking data of the online courses taught on the Moodle system. The tool allows educators an in-detail tracking of their courses via graphical presentations. These visuals are based on behavioral, cognitive or social data of students, and also include student activities, participation, and learning outputs (Mazza & Milani, 2004). SNAPP is another Moodle add-on similar to GISMO. It has been developed for the visualization of students' contributions in Moodle classes and the relations on discussion forums by means of social network analysis (Bakharia & Dawson, 2011). The purpose of the MOClog tool is to carry out numerical analysis on large scale log data, and to report it to the user with visual support (Mazza, Bettoni, Faré, & Mazzola, 2012). E-Learning WebMiner aims to examine behavioral and navigational patterns in student data for virtual classes. It utilizes association rule mining and clustering algorithms (Zorrilla & García-Saiz, 2013). CLVA, a Moodle add-on, integrates learning analytics into standard Moodle reports. It was developed with the aim to provide both educators and learners with multiple data structures and analytic methods in a single interface (Drăgulescu, Bucos, & Vasiu, 2015). Intelliboard.net is a commercial analytic and reporting tool. It aims to report all Moodle data under a single interface, along with graphs and visuals in a printable format. Although lacking an analysis feature, it allows visualization of log data and student tracking (IntelliBoard, 2015). MEAP is an add-on that is used for tracking students' in-course activities over Moodle. It has been developed to identify the students that are likely to fail, based on student behaviors (Liu, Froissard, Richards, & Atif, 2015). Analytics Graphs is an add-on developed to examine students' access to resources available on the Moodle environment. It serves three purposes; distribution of student scores with bar charts, reporting which students have accessed the resources, visualization of which students have uploaded their assignments in time or late, and identifying which students have accessed their courses and the related course resources every week (Singh, 2015).

To sum up, a significant portion of the developed tools serve to visualize the data obtained from the Moodle environment and report the basic usage statistics, while the tools that have been developed for data mining and analytics purposes appear to have limited features. Unlike the available tools, the tool that has been developed within the scope of this study carries out analyses through data mining libraries in R programming language. By doing so, it offers advance features to users, while providing a convenient infrastructure for the addition of new features.

R is a programming language and a free software development environment with open source code that was developed for data processing, calculation and visualizations (R Core Team, 2017). It features a large variety of statistical and machine learning methods and offers powerful data visualization tools. It allows users to write their own codes, thus making it possible for them to add extra features. By means of the developed tool, users will be able to harness the analysis and visualization power of the R programming language without having to write complex R codes.

DEVELOPMENT PROCESS

The developed tool operates in accordance with the process shown in Figure 1. It uses the logs exported from the Moodle learning management system as input and carries out data mining analyses on these logs. It generates visual and textual reports of the obtained results for the user. The details regarding each component of the process are given below.

FIGURE 1. *Data mining process*

Moodle Interaction Records

All sorts of student interactions on the Moodle environment (e.g. view, delete, create, update) are recorded in a database. For example; interactions such as a student opening a discussion forum, reading and replying to the messages there, accessing course-related resources etc. are all recorded. A full list of the information available in these records is provided in Table 1.

Table 1. *Structure of moodle log*

No	Field	Description
1	Time	The time activity is carried out
2	Username	Name - surname of the user carrying out the activity
3	Affected user	The name of the related user if the activity concerns another user
4	Event context	The course where the activity was carried out
5	Component	The course in which the activity was carried out
6	Event name	The type of activity (e.g. view, delete, create, update)
7	Description	Detail information regarding the activity
8	Source	The source of the activity
9	IP Address	The network address (IP) of the device on which the activity was carried out

With the help of the logging system, records can be kept on when, from which IP address, in which course component, and what type of activity was carried out. The Moodle learning management system in its database stores more than 50 different events regarding the learning activities. These records can be viewed as pivot tables within Moodle, or they can be exported. The developed tool uses these records as input. The Moodle learning management system allows administrators and course instructors to access and export these records in the following formats: comma-separated values (.csv), JavaScript object notation (.json), or MS Excel (.xlsx). The developed tool runs in CSV file format.

Preprocessing

In data mining studies, preprocessing of the data is important before moving on to the analysis stage. The aim of preprocessing is to enhance the data quality and extract ideal features for the selected analysis (Hämäläinen & Vinni, 2010). It is also important to obtain comprehensible results. In the developed tool, preprocessing is carried out automatically. The operations carried out in this stage can be briefly listed as session identification, feature extraction, data merging, and data transformation.

The log data exported by Moodle does not contain users' session information, but includes the information on the date and time of the activity. In this step of preprocessing,

session identification is carried out by using date and time information. After login identification, the records related to the same student are aggregated, and features to be used for analysis purposes are created. A list of these features with their descriptions is provided in Table 2.

Table 2. *Features created during preprocessing with their descriptions*

No	Feature	Description
1	Session	The number of sessions by the student
2	Time	The total amount of time the student has spent on the Moodle environment
3	Day	The number of different days the student has logged on to the environment
4	Total Action	The number of total activities
5	Forum View	The number of discussion forum views
6	Assignment View	The number of assignment views
7	Resource View	The number of course resource views
8	Forum Create	The number of posts on the discussion forum
9	Assignment Create	The number of assignment creations

To be able to carry out prediction analyses, the scores regarding students' success in the course must be added into the data set. This data must be provided by the user as it is not included in the Moodle logs. After feature extraction, an empty column is added to the table so that users can enter the score data. In the final preprocessing stage, data transformation is applied on the performance data provided by the user, and students are divided into two groups, low performers (Score <= 50) and high performers (Score > 50). The division process is carried out based on normalized student scores. Percentile rank normalization is used for normalization.

Data Mining Analyses

In this stage, predefined data mining analyses are applied on the features obtained after preprocessing. Data mining analyses are carried out using R programming language. R codes are automatically generated by the tool, and the outputs obtained by the execution of these codes on the R software are reported to the user. Any version of R software installed on the user's system would suffice for carrying out these analyses. The initially developed version of the tool is able to carry out classification and clustering analyses. These analyses were selected since they are the most preferred data mining methods in educational data mining studies (Peña-Ayala, 2014; C. Romero & Ventura, 2010). These methods constitute 69% of the data mining methods adopted in educational data mining studies (Peña-Ayala, 2014). Detailed information on these two analyses and in what kind of educational problems they can be used as a solution are stated below.

Classification Analysis: Classification analysis is one of the most commonly used methods in data mining studies (R. Baker & Yacef, 2009). The purpose of the classification analysis is to predict the value of the target feature on a categorical level by using another group of features. The developed tool uses classification analysis with the aim to predict students' academic performances (low or high) by drawing on their interaction data on the Moodle learning system. Another area the classification analysis is used is the identification of features that are important for predicting the target feature. In educational settings, classification algorithms are used for academic performance prediction, student behavior prediction (Ryan S.J.d. Baker, 2007), and the prediction of students who are highly likely to fail the course (Kotsiantis, 2009).

The developed tool includes 12 classification algorithms that are frequently used in educational data mining studies (LogitBoost, knn, rf, svmRadial, J48, rpart, rpart2, glm, naive_bayes, nnet, svmLinear, gbm). Apart from this, a cross validation method has been used to increase the generalizability of the obtained prediction models. Classification algorithms implemented in the tool also use parameter optimization to obtain the best results for each algorithm. The parameters that can be set by the user are classification algorithm and the

number of fold to be used for cross validation (k). Parameter optimization results, confusion matrix, performance metrics and feature importance analysis results are reported as output. Researchers can utilize the tool to compare the classification performance of different algorithms, or investigate which features are important in predicting student success.

Cluster Analysis: The purpose of the clustering method is to group together the items that show similarities in certain aspects. In clustering analyses, distance metrics are usually utilized to identify the items that have similar attributes (Cristobal Romero & Ventura, 2013). One of the important points in clustering analysis is the process of deciding how many separate clusters the data will be grouped into. If there is prior knowledge regarding how many separate clusters the data will be grouped into, this can be set as the number of clusters, or the optimum number of clusters can be found out by means of algorithms (R. S. J. d Baker, 2010). In a data that is grouped into accurate number of clusters, the items in the same cluster are expected to have more similarities with those in the same cluster, and fewer similarities with those in other clusters (Ryan Baker & Siemens, 2014). In educational environments, clustering methods are used for grouping students (Beal, Qu, & Lee, 2006) or student behaviors (Amershi & Conati, 2007).

In the tool, k-Means algorithm was preferred for clustering, which is widely used in educational data mining studies. The number of clusters can be either entered manually by the user or set automatically by the tool. Gap statistics (Tibshirani, Walther, & Hastie, 2001) was used to automatically set the optimum number of clusters. The results can be used for group formation or to group similar students based on their activities.

The tool was developed using C#.Net programming language, and data mining analyses were carried out through R programming language (R Core Team, 2017). In the clustering analysis, *Cluster* package was used for determining the optimum cluster number, while *factoextra* and *ggplot* were used for the visualization of results. Classification analyses were carried out through *caret* package.

CASE STUDY

Data Collection

The developed tool was tested through the real usage data obtained from the Moodle learning management system. The testing data was obtained from Database Management Systems course, which was run in a blended learning environment. 74 sophomore students were enrolled in the course, 6 of whom dropped the course before the end of the term. 37.751 rows log data from the remaining 68 students was exported from the Moodle environment for analysis purposes. The course was run in a blended learning model as face-to-face and online. For the face-to-face part, the students participated in 2-hour-long laboratory activities in addition to 3 hours of theoretical courses. Online activities (e.g. assignments, discussions etc.) were fully carried out over Moodle. The end-of-term scores of the students were calculated according to the scores they had obtained from the two mid-term exams and one final exam.

Data Preprocessing

In the preprocessing stage, data was created related to the nine features reflecting student interactions with Moodle. An academic performance feature was added manually to the data during the preprocessing stage. For the classification analysis, students were divided into two groups, low performers (Score <= 50) and high performers (Score > 50), in accordance with final score feature. Descriptive statistics related to the features obtained as a result of the preprocessing are displayed in Table 3. In order to assess if there exist any significant difference between the groups, an independent sample t-test analysis was carried out (see Table 3 for results).

Table 3. *Descriptive statistics and the results of the t-test analysis*

	High Performers (n = 34)			Low Performers (n = 34)			t test	
	$\overline{X}$	ss	median	$\overline{X}$	ss	median	t	p
Session	99.7	42.6	92.0	65.4	40.4	52.5	3.41	0.01
Time	764	495	628	392	262	339.5	3.87	0,001
Day	49.1	15.2	50.0	37.4	20.0	30.5	2.71	0.01
TotalAction	680	312	590	430	284	313.5	3.45	0,001
ForumView	209	179	153	85.5	112	37.5	3.42	0.01
AssignmentView	139	45.1	135	99.2	67.5	75.5	2.84	0.01
ResourceView	69.0	28.2	60.5	58.5	29.2	53.5	1.51	0.5
ForumCreate	10.8	18.2	2	1.91	4.02	0	2.77	0.01
AssignmentCreate	33.6	7.34	34.5	22.9	8.81	22	5.46	0,001
Final Score	74.1	10.2	74	40.2	14.1	45.5	11.31	0,001

An examination of the data in Table 3 reveals that all features considered, high performer students have statistically significantly higher scores compared to low performer students. These students logged in more frequently to and spent more time on the environment, had more views of discussion forum, contributed more to the discussions, and submitted more assignments. The end-of-term scores of high performer students are also significantly higher than those of low performer students.

Classification Analysis Results

Results of the classification analysis performed via the tool are displayed in Figure 2. The tool generates as output the parameter optimization results (Figure 2-a), confusion matrix, and performance metrics that were obtained as a result of cross validation that display the performance of the model (Figure 2-b), feature importance table (Figure 2-c), and the feature importance graph (Figure 2-d).
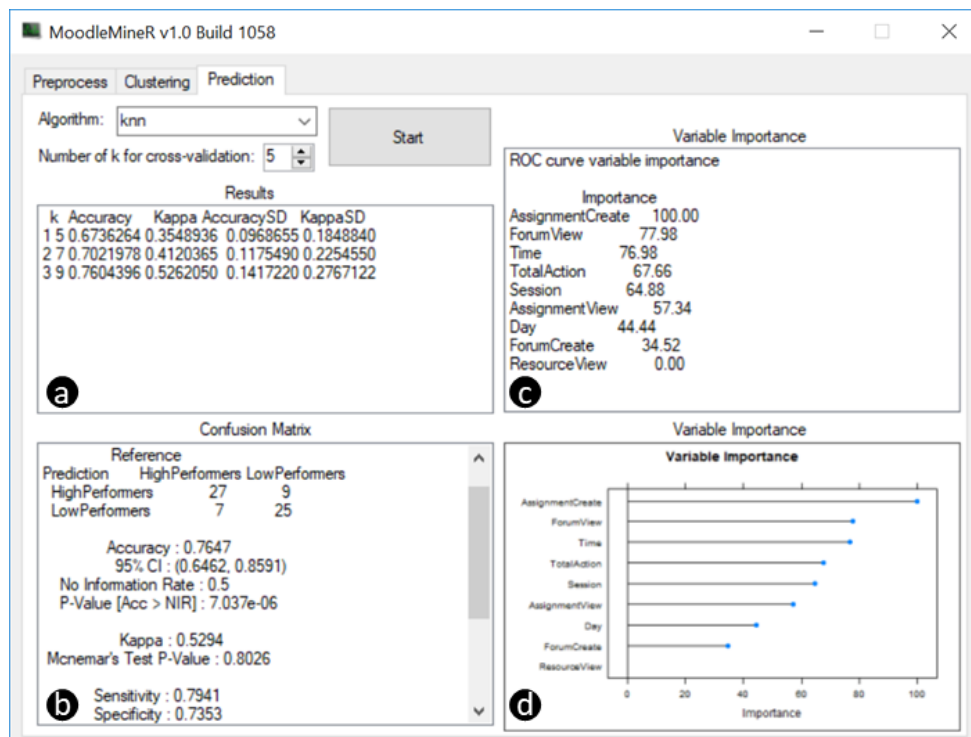

FIGURE 2. *Classification analysis results*

As a classification algorithm kNN was chosen and number of k was set as 5 for the cross validation. When the results given in Figure 2 are examined, it can be seen that the kNN achieved its best performance when the k is set as 9. In that case, the algorithm classified a total of 76.04% of the students accurately (Accuracy). An examination of the confusion matrix obtained from the cross validation in Figure 2-b reveals that the algorithm accurately classified 27 of the 34 high performer students and 25 of the 34 low performer students. In that sense, the prediction model performs similarly in separating high and low performers. An examination of the feature importance on Figure 2-c and Figure 2-d, on the other hand, reveals that the most important feature in separating low and high performers is *AssignmentCreate*. This feature is followed by *ForumView* and *Time* features, respectively. In other words, students' assignment submission numbers, discussion platform views, and time spent on the environment are all important features in predicting their academic success.

Clustering Analysis Results

Results of the clustering analysis performed via the tool are displayed in Figure 3. The number of clusters for the present data was determined automatically by the tool. When the results in Figure 3-a and Figure 3-b related to the determination of the optimum number of clusters are examined, it is seen that the data can be ideally divided into 3 groups. Figure 3-c shows the cluster centers, while Figure 3-d includes the cluster means of each cluster in terms of the features used in clustering. Since the data was standardized during the clustering process, the data on the graph shows the standard values. An examination of the cluster means reveals that the resulting clusters can be named as low-, medium-, and high-active students.
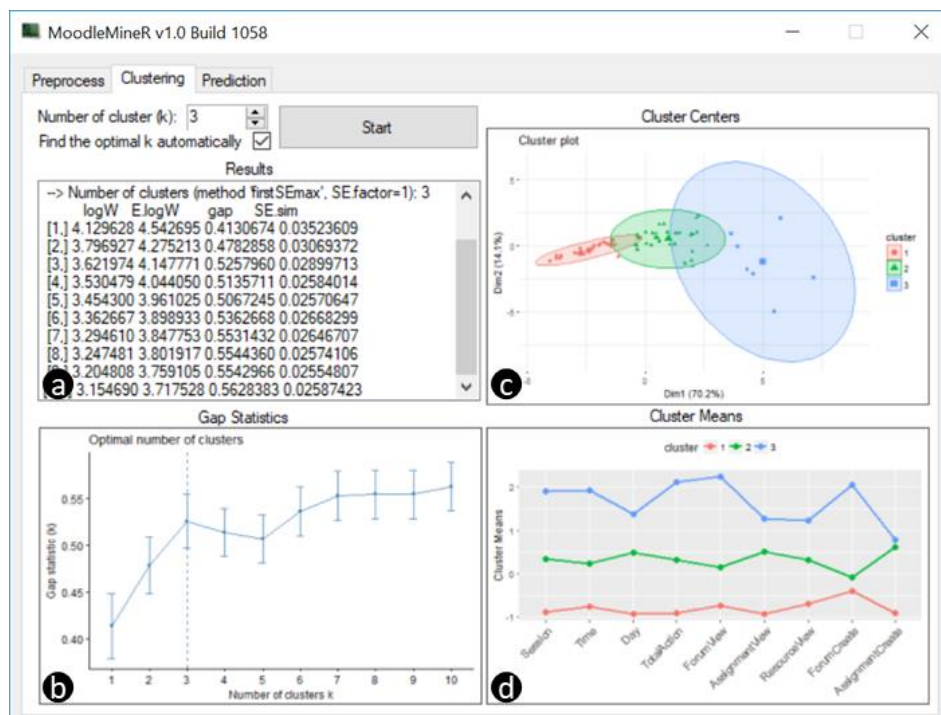

FIGURE 3. *Clustering analysis results*

DISCUSSION and CONCLUSION

In this study, a tool was developed by which non-experts can carry out basic data mining analyses on the data sets obtained over the Moodle learning management system. The tool automatically processes the log records in the form of click-stream, converting them into meaningful features, and performs classification and clustering analyses on these features. The

target group of the tool is the individuals who do not have any expertise in data mining. For this reason, the tool was planned to offer a minimum number of options to the user in analysis screens, and perform other adjustments automatically in the background. Furthermore, the users are given the opportunity to benefit from the features of the R programming language without writing any R code.

The tool includes clustering and classification analyses that are widely used in educational data mining studies. A single algorithm was used for clustering, while 12 algorithms were used for classification. The users are offered features such as automatic cluster number settings in clustering analyses, and performing parameter optimization in classification analyses. The tool also provides the users with tables and graphs that can be used for reporting the results. For researchers who wish to perform more advanced analyses, the tool was designed to export the data it processes.

The case study has proven that the tool can successfully perform classification and clustering analyses on a Moodle data set. The information generated by the tool will help researchers and educators alike in grouping students by their interaction levels, determining at-risk students, monitoring students' interaction levels, and identifying important features that impact students' academic performances.

During preprocessing, the tool reduces a high number of log records (activity sequences) down to a single line of data for each student that displays their general usage. In future versions of the tool, we plan to add methods that will allow the analysis of students' usage patterns from sequential data. Also, with the addition of alternative algorithms for available analyses, a variety of data mining methods will be integrated into the tool.

REFERENCES

Aguilar, D. A. G., Therón, R., & Peñalvo, F. G. (2008). *Understanding educational relationships in Moodle with ViMoodle.* Paper presented at the Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on.

Amershi, S., & Conati, C. (2007). *Unsupervised and supervised machine learning in user modeling for intelligent learning environments*. Paper presented at the Proceedings of the 12th international conference on Intelligent user interfaces, Honolulu, Hawaii, USA.

Baker, R., & Siemens, G. (2014). Educational Data Mining and Learning Analytics. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2 ed., pp. 253-272). Cambridge: Cambridge University Press.

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *J. Educ. DataMining, 1*(1), 3-17.

Baker, R. S. J. d. (2007). *Modeling and understanding students' off-task behavior in intelligent tutoring systems*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, California, USA.

Baker, R. S. J. d. (2010). Data Mining. In *International Encyclopedia of Education (Third Edition)* (pp. 112-118). Oxford: Elsevier.

Bakharia, A., & Dawson, S. (2011). *SNAPP: a bird's-eye view of temporal participant interaction.* Paper presented at the Proceedings of the 1st international conference on learning analytics and knowledge.

Beal, C. R., Qu, L., & Lee, H. (2006). *Classifying learner engagement through integration of multiple data sources*. Paper presented at the Proceedings of the 21st national conference on Artificial intelligence - Volume 1, Boston, Massachusetts.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *Int. J. Technol. Enhanc. Learn., 4*(5/6), 318-331. doi:10.1504/ijtel.2012.051815

Chatti, M. A., Lukarov, V., Thüs, H., Muslim, A., Yousef, A. M. F., Wahid, U., . . . Schroeder, U. (2014). Learning Analytics: Challenges and Future Research Directions. *e-learning and education : eleed*(10).

Drăgulescu, B., Bucos, M., & Vasiu, R. (2015). *CVLA: integrating multiple analytics techniques in a custom moodle report.* Paper presented at the International Conference on Information and Software Technologies.

García, E., Romero, C., Ventura, S., & De Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modeling and User-Adapted Interaction, 19*(1-2), 99-132.

Hämäläinen, W., & Vinni, M. (2010). Classifiers for Educational Data Mining. In *Handbook of Educational Data Mining* (pp. 57-74): CRC Press.

IntelliBoard. (2015). IntelliBoard.net. Retrieved from http://www.intelliboard.net/

Kotsiantis, S. (2009). Educational data mining: A case study for predicting dropout-prone students. *Int. J. Knowl. Eng. Soft Data Paradigm., 1*(2), 101-111. doi:10.1504/ijkesdp.2009.022718

Liu, D. Y.-T., Froissard, J.-C., Richards, D., & Atif, A. (2015). An enhanced learning analytics plugin for Moodle: student engagement and personalised intervention.

Mazza, R., Bettoni, M., Faré, M., & Mazzola, L. (2012). Moclog–monitoring online courses with log data.

Mazza, R., & Milani, C. (2004). *Gismo: a graphical interactive student monitoring tool for course management systems.* Paper presented at the International Conference on Technology Enhanced Learning, Milan.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications, 41*(4, Part 1), 1432-1462. doi:10.1016/j.eswa.2013.08.042

Perez, R. P., Romero, C., & Ventura, S. (2010). *A Java desktop tool for mining Moodle data.* Paper presented at the Educational Data Mining 2011.

R Core Team. (2017). R: A language and environment for statistical computing: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rabbany, R., Takaffoli, M., & Zaïane, O. R. (2011). *Analyzing participation of students in online courses using social network analysis techniques.* Paper presented at the Proceedings of educational data mining.

Retalis, S., Papasalouros, A., Psaromiligkos, Y., Siscos, S., & Kargidis, T. (2006). *Towards networked learning analytics—A concept and a tool.* Paper presented at the Proceedings of the fifth international conference on networked learning.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 40*(6), 601-618. doi:10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12-27. doi:10.1002/widm.1075

Singh, J. (2015). New Block: Analytics graphs. Retrieved from http://www.moodleworld.com/new-block-analytics-graphs/

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*(2), 411-423. doi:doi:10.1111/1467-9868.00293

Zafra, A., Romero, C., & Ventura, S. (2013). DRAL: a tool for discovering relevant e-activities for learners. *Knowledge and information systems, 36*(1), 211-250.

Zorrilla, M., & García-Saiz, D. (2013). A service oriented architecture to provide data mining services for non-expert data miners. *Decision Support Systems, 55*(1), 399-411.