Understanding and assembling user behaviours using features of Moodle data for eLearning usage from performance of course-student weblog

S Chayanukro1,*, M Mahmuddin2 and H Husni2
1 Suan Dusit Rajabhat University, Mueang Trang, 92130, Trang, Thailand
2 Department of Computer Science, School of Computing, Universiti Utara Malaysia, 06010, UUM Sintok, Kedah Malaysia

*songsakda@gmail.com

Abstract. In reality, students learn via eLearning (electronic online learning) system in different ways depending on their learning needs, learning behaviours as well as eLearning system policy for users. However, most learning outcome prediction models of eLearning systems are still not stable and still cannot be applied in many situations as the use of eLearning is considered to be highly dynamic. Therefore, the objective of this work is understand if eLearning system can be predicted based eLearning usage by exploiting Moodle log data. To understand it, features from web log course-student in Moodle is being considered, a number of machine learning techniques also have been applied for benchmarking in this study. The result found that the current group doesn’t give better understanding and significant groups of factors that could be able to predict the learning outcome.

1. Introduction
Nowadays, eLearning (electronic online learning) is remarkably developed. The system has tremendous changing to facilitate classroom for more efficient outcome comparing to a traditional class. This modern teaching process, in fact, applies advanced communication technologies which do not only fulfill the need of educational institutions, but it can also build the efficient communication in the digital classroom. eLearning is a widely used technology, and it can assist in providing guidance and analysing the effects of classroom management. Besides, the use of eLearning in teaching management helps to understand the impact of eLearning. It is the ubiquitous system.

The main purpose of this work is learning how to use the eLearning to predict student’s achievement. The ability to predict the behaviour of different learners could be applied to the eLearning system in order to assist both the learner and the instructor. It also develops a model for the university’s teaching and learning development as well as for those departments responsible for education management by using modern technology. In this work, the different approaches to reach a learning outcome prediction model by machine learning technique should be the more suitable for predicting tools. On the basis of the popular Moodle eLearning system of educational institutions in Thailand. Its open-source system, which is suitable for the developer to share the experiences. This work studied on eLearning system to collect log data for studying the student behaviour in order to predict the learning outcomes before the end of the course. The data and results of this finding will depend on their context.
Currently, the Learning Management System (LMS) was developed to meet the requirement of the above developing virtual classroom. Therefore, higher learning institutions are finding suitable LMS for eLearning system. There are various ways to select LMS tools in the virtual class [1]; so, open source LMS is now the most significant tools which applies to the learning environment. The study of open source LMS in eLearning has found Moodle LMS as the most outstanding eLearning system [1]. Moreover, there are also many useful features in LMSs Moodle that strive for building on quality of education and give assistance to select the necessary tools for eLearning.

2. eLearning log data and eLearning performance

Educational data is abundant due to the nature of 21st-century education, where many studies have taken place online through the advent of digital learning. The use of LMS has increased their significance to provide education without boundaries, and these systems hold a massive amount of data. Educational data mining is a process of accessing a wide range of educational data, with the objective of developing the education process through the knowledge gained from the mining process. A paper is intended to be a link to a master's degree course that has been excluded from Moodle LMS. This point may lead to a decision that can be made to improve the learning process. The results show that most students have access to the resources available during the last minute before the exam and are likely to procrastinate online submissions [2].

2.1. Moodle LMS

One of the most popular LMS nowadays is LMS Moodle which is able to support highly in daily teaching. LMS focuses on supporting ICT-based education. This structure is influenced by the teaching hypothesis and the content plan of each subject. Presently, these electronic education products focus more on the personality of the learner. The most appropriate method is to analyse its properties to determine the advantages and disadvantages of LMS using. For this purpose, they decided to use strategic SWOT analysis to present this finding as strengths and weaknesses external factors affecting the system concerning possibilities and threats [3].

By using Moodle to record user experience in eLearning systems in the learning management system, Moodle provides information exchange and communication among participants in eLearning courses and analyses the user experience to meet the teaching needs of both students and teachers in all areas. Using Moodle will make learning more accessible and more interesting. Moodle Learning Management Systems will help academics build productive online learning communities using data mining techniques. Moodle log files help the instructors to pre-process the data, predict learning strategies and summarize the website structure according to learner's interest by applying data mining techniques [4].

2.2. eLearning and student learning performance

A learning portfolio is an eLearning usage data that provides the students with a specific method for evaluating their own learning situations”. The records of the students' activities during the learning process, such as their interaction with others, assignments, test papers, personal work collections, their discussion content, and learning records are all included [5]. The structure of the database from the system illustrated below; both tables derive from open source eLearning system in Moodle. Table 1 presents the important features/fields that contain data for the mining process as mentioned in Romero et al. [6].
Table 1. Important features/fields adopted of the Moodle LMS [6].

| Field Name               | Descriptions                                      |
|--------------------------|---------------------------------------------------|
| mdl_user                 | Information about all the users.                  |
| mdl_user_students        | Information about all students.                   |
| mdl_log                  | Logs every user’s action.                         |
| mdl_assignment           | Information about each assignment.                |
| mdl_assignment_submissions| Information about assignments submitted.         |
| mdl_chat                 | Information about all chat rooms.                 |
| mdl_chat_users           | Keeps track of which users are in which chat rooms.|
| mdl_choice               | Information about all the choices.                |
| mdl_glossary             | Information about all glossaries.                 |
| mdl_survey               | Information about all surveys.                    |
| mdl_wiki                 | Information about all wikies.                     |
| mdl_forum                | Information about all forums.                     |
| mdl_forum_posts          | Stores all posts to the forums.                   |
| mdl_forum_discussions    | Stores all forums’ discussions.                   |
| mdl_message              | Stores all the current messages.                  |
| mdl_message_reads        | Stores all the read messages.                     |
| mdl_quiz                 | Information about all quizzes.                    |
| mdl_quiz_attempts        | Stores various attempts at a quiz.                |
| mdl_quiz_grades          | Stores the final quiz grade                       |
| course                   | Identification number of the course.              |
| n_assignment             | Number of assignments handed in.                  |
| n_quiz                   | Number of quizzes taken.                          |
| n_quiz_a                 | Number of quizzes passed.                         |
| n_quiz_s                 | Number of quizzes failed.                         |
| n_messages               | Number of messages sent to the chat.              |
| n_message_ap             | Number of messages sent to the teacher.           |
| n_posts                  | Number of messages sent to the forum.             |
| n_read                   | Number or forum message read.                     |
| total_time_assignment    | Total time spent on assignment.                   |
| total_time_quiz          | Total time used in quizzes.                       |
| total_time_forum         | Total time used in forum.                         |
| mark                     | Final mark the student obtained in the course      |

2.3. Log data mining

Data mining technique plays an important role as it helps to extract the information from a vast database called weblog. According to the study of Gao [7], the study of using data mining/machine learning technique to bring out behaviour pattern of the client from web log files which focus on analysing client's behaviour pattern recognition system and its application to obtain client information conveniently and automatically.

In other work from Xu and Jun [8] introduce eLearning model on the basis of Web Usage Mining (WUM) techniques, it is the research that analysed the probable rules hidden in web log and it can help personalize the design of Web content and develop web design, customer satisfaction, and user navigation via pre-fetching and caching. An intelligent and individual platform for learners can be provided by WUM which can be considered as the most popular techniques in Web data mining.

Furthermore, modelling online user behaviour [9] demonstrates the methodology for studying the relationship between activity and used time of student in using the eLearning system by used Timed Transition Automaton (TTA) algorithm for representing the output model. The finding of this study
reveals that each student behaviour shows different activities timelines rather than describing its pattern. Due to this, better understanding of the existing LMS and identifying best possible features that could represent LMS usage in the institution.

3. Implementation

3.1. Data collection and pre-processing
This study has collected data log from eLearning database date during 2012-05-25 to 2015-04-06. These three years’ log takes six semesters study period. There are 53 courses were collected during this data collection. After log cleaning, processing and matching to the available grade result there are 20 courses in total. The chosen courses were used by courses users and stored the activities log number as shown in Table 2.

| Factor ID | Factor ID | Factor ID | Descriptions         |
|-----------|-----------|-----------|----------------------|
| 1         | quiz_view | 11        | forum_view_discussion|
| 2         | quiz_view_summary | 12        | assignment_upload    |
| 3         | quiz_continue_attempt | 13        | forum_add_post       |
| 4         | quiz_close_attempt | 14        | forum_view_forum     |
| 5         | quiz_attempt | 15        | forum_delete_discussion|
| 6         | quiz_review | 16        | forum_add_discussion |
| 7         | url_view | 17        | forum_delete_post    |
| 8         | course_view | 18        | forum_unsubscribe    |
| 9         | assignment_view | 19        | forum_subscribe      |
| 10        | resource_view | 20        | forum_update_pos     |

This work requires a comprehensive data and can be used for creating sufficient models from two groups of data. Two groups of data are collected which are eLearning system web log and student learning profiles that recorded the student learning activities as students’ usage history and students’ learning database. These two parts of data will be used for extracting and learning result respectively for this research methodology.

Before that, one important issue is the source of data log (web log) need to identify for later collection. In this purpose, the study has identified two (2) main sources of data: (a) from eLearning system web log and (b) student learning database. Both of this data are collected from the same eLearning system. A students’ learning database is the students’ personal data such as first name, last name, student identification, status, grade etc. The personal data files are exported directly from student’s learning database in order to use in the data extraction process. The next process is to remove all the unwanted or meaningless records and fields/features. This can be done by removing empty field and records.

![Figure 1](image-url)  
**Figure 1.** Collecting data from two main source namely eLearning weblog and student database.

For courses sampling, the traditional courses and eLearning courses have the whole activities from the beginning to finishing point within a semester. All usage history and student results are kept in a
database and web log in the same session. Thereby, web log collected from each course within one semester is sufficient for the study used. However, this study has been collecting eLearning log from six semesters to support more data mining technique performance and more variety of e-Learning factor. For this more there are 102 features that been evaluated based Table 1, and only 20 features are considered to be the final features exploited for next step. The selection is based on the available activities of the web log. e-Learning factors provided the user's action for students and teachers activities. Each action became the data and stored in the system by related module processing. After users are done the activities the log record is stored in the log table at the same time.

For this work, we group the data into eight (8) classes data set is for the prediction process based on 8 groups classification. In this case, grade of each subject is assessed by grade A, B+, B, C+, C, D+, D, F where A is the best grade and F is lowest grade. As mentioned before, these grades are based result/final marks given by the teacher of the subject taken. This subject grade originates from impose by the university that is A >= 90, B+ >= 85, B >=75, C+ >=70, C >=60, D+ >=55, D >=50, F < 50.

3.2. Benchmarking
There six (6) machine learning algorithm are deployed for benchmarking, namely ZeroR, Naïve Bayes, Support Vector Machine (SVM) by [10], J48 inspired from [11] and Decision Table. To assist this work, we are utilizing Weka [12] platform to identify the percentage of the accuracy of each machine learning algorithm. These six algorithms were chosen to test the predictive effect as the guidelines of Ribeiro and Cardoso [13].

4. Results and findings
Table 3 shows the result of classification processing that show the prediction accuracy ratio. These results processed by six algorithms as mentioned previous section. Three data sets class scheme were gathered by course identification number and by all course grouping.

| Course ID | ZeroR | Naïve Bayes | SVM | J48 | Deci. Table | Random Tree | Ave. |
|-----------|-------|-------------|-----|-----|-------------|-------------|------|
| 101       | 20.00 | 25.00       | 30.00 | 30.00 | 30.00       | 15.00       | 25.00 |
| 107       | 39.13 | 17.39       | 34.78 | 47.82 | 39.13       | 17.39       | 32.61 |
| 12        | 30.00 | 20.00       | 10.00 | 25.00 | 20.00       | 30.00       | 22.50 |
| 126       | 56.60 | 50.94       | 64.15 | 54.71 | 66.03       | 58.00       | 58.41 |
| 127       | 32.14 | 21.42       | 28.57 | 14.28 | 32.14       | 14.28       | 23.81 |
| 128       | 33.33 | 5.50        | 33.33 | 33.33 | 27.77       | 27.77       | 27.77 |
| 146       | 52.94 | 23.52       | 52.94 | 35.29 | 52.94       | 47.05       | 44.11 |
| 149       | 31.11 | 24.44       | 17.77 | 46.66 | 40.00       | 37.77       | 32.96 |
| 165       | 24.00 | 34.00       | 34.00 | 28.00 | 32.00       | 26.00       | 29.67 |
| 178       | 31.25 | 20.83       | 25.00 | 25.00 | 31.25       | 27.00       | 26.72 |
| 190       | 36.00 | 28.00       | 32.00 | 40.00 | 36.00       | 32.00       | 34.00 |
| 191       | 50.00 | 43.33       | 50.00 | 40.00 | 53.33       | 46.66       | 47.22 |
| 195       | 50.00 | 38.88       | 50.00 | 44.44 | 50.00       | 11.11       | 40.74 |
| 204       | 0.00  | 23.07       | 23.07 | 53.84 | 7.69        | 15.38       | 20.51 |
| 206       | 45.83 | 45.83       | 50.00 | 29.16 | 45.83       | 33.33       | 41.66 |
| 28        | 40.00 | 33.33       | 40.00 | 26.66 | 26.66       | 26.66       | 32.22 |
| 29        | 34.88 | 62.79       | 37.20 | 83.72 | 79.06       | 76.74       | 62.40 |
| 33        | 30.00 | 26.66       | 33.33 | 20.00 | 30.00       | 16.66       | 26.11 |
| 34        | 22.58 | 38.70       | 29.03 | 83.87 | 64.51       | 61.29       | 50.00 |
| 36        | 61.53 | 38.46       | 61.53 | 38.46 | 61.53       | 46.15       | 51.28 |
| Ave.      | 36.07 | 31.10       | 36.84 | 40.01 | 41.57       | 33.31       | 36.48 |

| All Course | 25.40 | 18.99       | 26.20 | 23.91 | 25.40       | 21.51       | 23.57 |
Notice that, the result does not show promising group of features since the average accuracy is less than 50%. This also show that, group using based on subject taken by the user is not suitable for this work. New combination is needed to scientifically satisfy higher average accuracy.

5. Future works
This works can be extended into other classes for example into 2 classes (Pass and Fail) and 3 classes (Poor learning, good learning, and average learning) these new group can be categorised. We are also planning to group analysis using Factor Analysis for better understand.

Acknowledgments
This work is made possible due to supports and helps from our colleagues from Suan Dusit Rajabhat University and Universiti Utara Malaysia.

References
[1] Graf S and List B 2005 An evaluation of open source e-learning platforms stressing adaptation issues Proceedings - 5th IEEE International Conference on Advanced Learning Technologies, ICALT 2005 2005 pp 163–5
[2] Nkomo L M and Nat M 2016 Discovering students use of learning resources with educational data mining 13th HONET-ICT International Symposium on Smart MicroGrids for Sustainable Energy Sources Enabled by Photonics and IoT Sensors, HONET-ICT 2016 pp 98–102
[3] Nkomo L M and Nat M 2016 Discovering students use of learning resources with educational data mining 13th HONET-ICT International Symposium on Smart MicroGrids for Sustainable Energy Sources Enabled by Photonics and IoT Sensors, HONET-ICT 2016 pp 98–102
[4] Sheshasaayee A and Bee M N 2017 Evaluating user experience in moodle learning management systems IEEE International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2017 - Proceedings pp 735–8
[5] Chen C M, Li C Y, Chan T Y, Jong B S and Lin T W 2007 Diagnosis of students’ online learning portfolios Proceedings - Frontiers in Education Conference, FIE pp T3D-17-T3D-22
[6] Romero C, Ventura S and García E 2008 Data mining in course management systems: Moodle case study and tutorial Comput. Educ. 51 368–84
[7] Gao W H 2010 Research on Client behavior pattern recognition system based on Web Log Mining 2010 International Conference on Machine Learning and Cybernics, ICMLC 2010 1 pp 466–70
[8] Wu H Y, Zhu J J and Zhang X Y 2009 The explore of the web-based learning environment base on web sequential pattern mining Proceedings - 2009 International Conference on Computational Intelligence and Software Engineering, CiSE 2009 pp 1–6
[9] Milani A, Jassó J and Suriani S 2008 Soft user behavior modeling in virtual environments Proceedings - 3rd International Conference on Convergence and Hybrid Information Technology, ICCIT 2008 vol 2 pp 1182–7
[10] Cortes C and Vapnik V 1995 Support-vector networks Mach. Learn. 20 273–97
[11] Quinlan J R 1993 C4.5: Programs for Machine Learning (Morgan Kaufmann Publishers Inc.)
[12] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P and Witten I H 2009 The WEKA data mining software ACM SIGKDD Explor. Newsl. 11 10–8
[13] Ribeiro B and Cardoso A 2008 Evaluation system for e-learning with pattern mining tools Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics pp 3051–6