International Conference on Computer, Engineering, Law, Education and Management 2017
[ICCELEM 2017]

ISBN	978-81-933584-7-4		**VOL**	01
Website	www.iccelem.com		**eMail**	iccelem@asdf.res.in
Received	30-May-2017		**Accepted**	25-July-2017
Article ID	ICCELEM001		**eAID**	ICCELEM.2017.001

Embedding the Multiple Linear Regression Model to Monitor Student Performance in the Flexible Digital Learning Environment

Benilda Eleonor V Comendador[1]
[1]Polytechnic University of the Philippines, Anonas St., Sta. Mesa, Manila, Philippines

Abstract-*This study focused on providing a Decision Support System (DSS) that integrates Multiple Linear Regression (MLR) model to monitor student performance in the flexible digital learning environment. The author carried out series of experiments in order to evaluate the performance and usefulness of the generated models. MLR was adopted in the development of the Learning Analytics Decision Support System. The developed application predicts the performance of the university portal users which may help the Distance Education (DE) students succeed in the blended learning approach being provided by the DE educators.*

I. Introduction

Nowadays, instructional technologies are on transition from mobile learning to ubiquitous learning wherein educational materials are accessible anytime, anywhere in any form (text, video and audio) to all educational stakeholders via eLearning platforms. With this innovation, students and enrollment in Distance Education (DE) courses became attractive for most learners. Conversely, the report on course drop out and failure rates is more incessantly increasing in this mode of learning. In this light, the academe need to develop tools and methods that will explore data coming from the eLearning software which can support teachers and students to take action based on the evaluation of educational data.

Policy makers and administrators should include analytics, user modeling, user profiling and clustering, domain modeling, relationship mining and data visualization to unveil outcome-oriented actionable insights from specific learning behaviors [1]. Consequently, some educational institutions (e.g. University of the Philippines Open University, Mindanao State University, California State University, Monash University in Australia) were implementing Learning Management System (LMS) to manage the courses offered in the Internet [3]. In 2015, eLearning industry reported that 74% of the companies currently use LMS and Virtual classroom, webcasting and video broadcasting [4]. Some of the most popular open source LMS includes Edmodo and Modular Object-Oriented Dynamic Learning Environment (Moodle) [8]. Nevertheless, based on the studies, MOODLE was the most recommended LMS because the administration and control can be handled by the institution to do further analytics such as tracking web logs of students for further monitoring of their progress and other activities [2]. According to Romero et. al., the application of data mining in e-learning is not much different than any other application area [9]. However, there are some important issues that make data mining in e-learning different than in the others such as (1) data; (2) objective and (3) techniques. In other web-based systems the data used is normally a simple web server access log, but in e-learning there is much more information available about the student's interaction such as details on each online assessment task (assignment, quiz, discussion forum and chat) [6]. Conversely, Picciano suggested that LMS's should provide constant monitoring of student activity whether there are responses, postings on a discussion board, accesses of reading material, completions of quizzes, or some other assessment. Thus, university should analyze the data gathered from the LMS users which are stored in the web server to discover knowledge that will enhance the students' online experience [7]. ECAR Working Group (2015) suggested that educators can tap more dynamic data produced from a range of instructional technologies (such as LMS event log data, electronic gradebook data, attendance data, library data, etc.) for learning analytics which when combined with traditional measures—allow for a more nuanced and personalized analysis [5]. Some of the challenges of implementing data mining and learning analytics, include high cost of collection, storage, development of algorithms, interoperable administrative and learning systems (systems/data types). As such, their report recommends researchers to combine the data types, with acceptable validity, privacy, and ethical standards applied, for improved predictive power [1].

Apparently, in one of the student workshops conducted by the University of Lincoln, the students cited some ideas on capabilities they would like to see in a learning analytics application. They included notifications on grades and progress toward objectives; the ability to give immediate feedback to lecturers and professors in order to improve the course; and reading list functionality that presents metrics on how students engage with the texts [10].

Currently, little research has been conducted that focus on university portal learning analytics that will lead to the prediction of academic performance of the DE students. Thus, this study focused on providing a Decision Support System (DSS) that integrates Multiple Linear Regression (MLR) model of data mining for portal providers and users to analyze and predict the performance of distance students scientifically.

II. Framework of the Study

Fig.1 illustrates the framework of the study. It consists of four (4) major phases 1) development of a student performance predictive model. 2) testing and implementation of the predictive model, 3) development of the decision support system to identify at risk-students for early intervention and attrition prevention and 4) evaluation of the developed software by the respondents in terms of the following: functionality, usability, reliability and portability of the output.

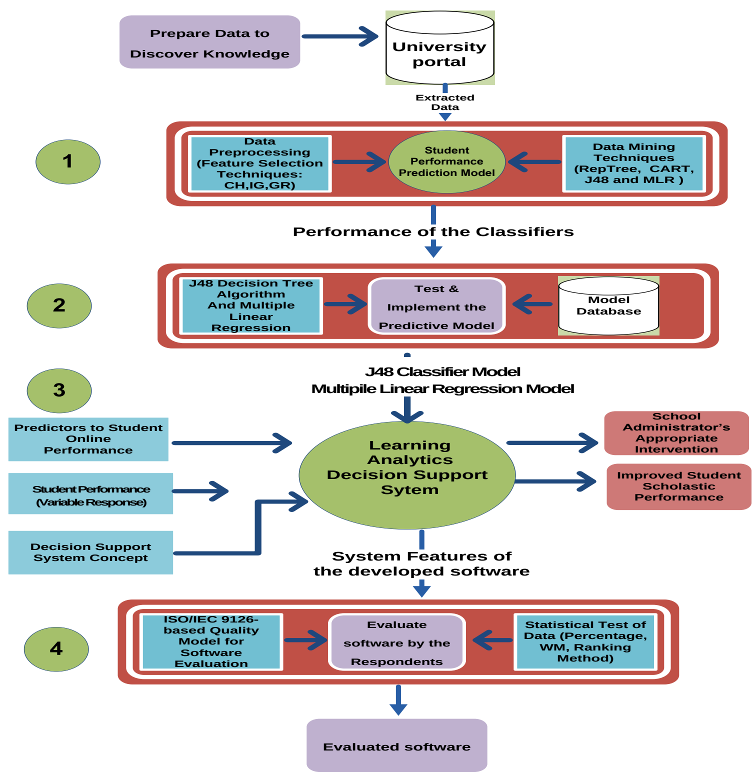

Figure 1. Framework of the Study

During the first phase, the author utilized the data sets from the university Academic Institutions Management Systems (AIMS) and the students' usage history while accessing the Polytechnic University of the Philippines (PUP) eMabini Learning Portal. After which, the author performed steps to preprocess the data then converted the extracted data into the required format by

Waikato Environment for Knowledge Analysis (WEKA) tool. Also, to generate the students' performance prediction model, the author conducted series of experiments to evaluate the appropriate classification for predicting students' final rating based on their usage data in the university portal. During the next phase: testing and implementation of the model, the author identified interesting rules and patterns for decision-making. The author repeated the data mining and the pattern analysis process if she considered that the results were not remarkable. For the third phase of the study, the predictors to Student Online Performance, the data mining techniques and decision support system concept were integrated to develop the software as shown on Fig.1.

The independent variables in this study were the records of student's performance in online assessment tasks posted by the course specialists in the university portal. The expected output was the developed student performance model with MLR. With this model, it may give students considerable time and opportunity for early interventions to improve his scholastic performance and for the Distance Education providers to lessen the drop-out rate. During the fourth phase of the study, the developed software was evaluated by the respondents in terms of functionality, usability, reliability and portability of the output based on the ISO 9126 Quality Model for Software Evaluation. The author used the statistical tools such as Weighted Mean, Ranking Method and Percentage to summarize and analyze the respondents' evaluation on the developed software.

III. Results and Discussions

A. The Multiple Linear Regression Model

The original database for MLR was divided into two using the 80:20 rule - the training and validation dataset which consisted of 19 instances and the test dataset which consisted of 7 instances. Equation 3.0 shows the MLR equation generated using Weka in a 10-fold cross validation and a confidence factor of 0.25 without pre-processing of attributes. This was named as MLR Model A in this study.

MLR Model A in equation form is:

$$\text{Final Rating} = 0.2082 \times \text{Assign\_1} + 0.1987 \times \text{Assign\_2} + 0.205 \times \text{QUIZ\_1} + 0.1963 \times \text{QUIZ\_2} + 0.1986 * \text{Assign\_3} + -0.4605 \tag{3.0}$$

Fig. 2 described the attributes selection output. It can be gleaned that Log_Count got the highest influence followed by Mat_Access_Count, Activity_Rating and Final_Grade.

=== Run information ===

Evaluator: weka.attributeSelection.CfsSubsetEval

Search: weka.attributeSelection.BestFirst -D 1 -N 5

Relation: LMS_Prediction_Merged_v6-weka.filters.unsupervised.attribute.Remove-R1-8,10,12,14,16-17,19-20

Instances: 248 ; Attributes: 5

Log_count, Mat_Access_Count, Activity_Ratng, Exam_Points, Final_Grade

Evaluation mode: 10-fold cross-validation

=== Attribute selection 10 fold cross-validation seed: 1 ===

number of folds (%) attribute

10(100 %) 1 Log_Count; 1(10 %) 2 Mat_Access_Count;

0(0 %) 3 Activity_Rating; 10(100 %) 5 Final_Grade

Figure 2. Attribute Selection Output for MLR

From the processing of attributes making use of the same validation and confidence factor, the MLR model generated is shown in Equation 4.0. This was named as MLR Model B in this study.

MLR Model B in equation form is:

$$\text{Exam\_Points} = -1 \times \text{Activity\_Rating} + 2 \times \text{Final\_Grade} + 0 \tag{4.0}$$

Table I shows the error measures of the two MLR models generated with and without preprocessing of attributes. It compares the fitting of the models as to the differences between the observed values and the model's predicted values. The correlation coefficient of MLR Model A is higher than Model B. On the other hand, MAE and RMSE of MLR Model B was greater than MLR Model A's.

Table I Error Measures of the MLR Models

	MLR Model A	MLR Model B
Correlation Coefficient	0.9999	0.9188
Mean Absolute Error (MAE)	0.1046	8.9507
Root Mean Squared Error (RMSE)	0.1383	11.6134

Fig. 3 examines the difference between the performance of the two MLR model as far as evaluation on test set is concerned. It could be gleaned from the figure that MLR Model B had a higher MAE and RMSE than that of the MLR Model A.

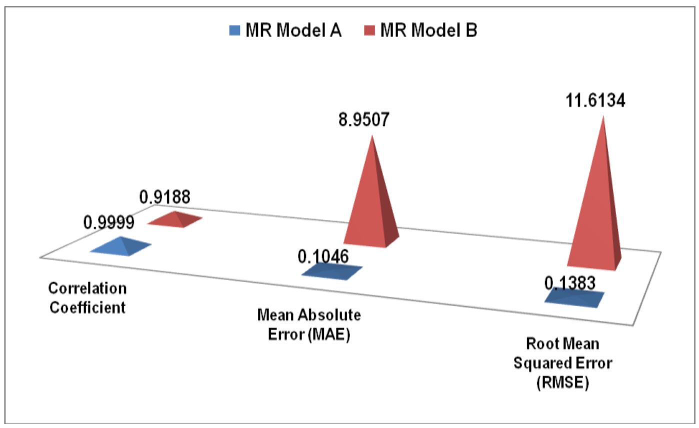

Figure 3. Performance of the MLR Models based from Evaluation on Test Set

B. Evaluation of the Developed Software by the Respondents

As the developed software will be used for wider DE learners and providers, the author asked select active portal users to evaluate the system in terms of functionality, usability, reliability and portability of the output. After exploring the demonstration, the author encouraged them to fill-out the online survey form.

There were one hundred ninety-one (191) online portal users who filled-out the survey form. Below is the highlight of the information that were gathered from the participants.

Table II described the summary of the evaluation of the developed software as perceived by the respondents. The ratings in this table indicates that the respondents rated the developed software as "Moderately Acceptable" which signifies that it is functional, usable, reliable and portable for the DE stakeholders.

Table ii. Summary of the evaluation of developed software by the respondents

Parameters	Total Weighted Mean	Interpretation
Functionality	4.43	Moderately Acceptable
Usability	4.34	Moderately Acceptable
Reliability	429	Moderately Acceptable
Portability	4.32	Moderately Acceptable
Overall Mean	4.34	Moderately Acceptable

In terms of system functionality, the respondents found that the system can shut out access from people who are not part of the course. It provides online submission of assignments where the result can be evaluated by the professor then recorded in the database. This is an indication of an effective integration management that should be maintained by the system. In terms of system usability, the system rated as "Moderately Acceptable" of most of the respondents. This indicates the versatility of the developed software when it comes to usability as it gives opportunity to the users to recapitulate, retrieve and interact with the system whenever the user desires using his available technology gadget. In terms of reliability, the respondents rated all questions as "Moderately Acceptable". It signifies that the developed system provides confidentiality in each user account and it has more secured delivery and distribution of information to its intended users. These indicate that it generates notifications from authorized users only; provides consistent result and response correctly when encountered failure. In terms of portability it is rated as "Moderately Acceptable". This indicates that the system can be accessed from one gadget (smart phone, tablet, laptop and personal computer) to another. Thus, it allows the users (OUS Administrator, course specialists and learners) to view course materials and online assessment tasks with one login password in any computer or mobile device which promotes higher system flexibility to the DE learners who were predominantly part-time students but fulltime employees. The ratings on this parameter exhibit the portability of the system which also promotes robustness of the developed model for predicting the online performance of the DE learners.

IV. Conclusion

The results demonstrated that the generated MLR model can be harnessed to develop the Learning Analytics Decision Support System which may provide powerful educational tool that can analyze and predict the performance of the learners in the flexible digital learning environment. During the testing and simulation of real institutional data, the developed software displayed the same output with that of the two reliable application programs the Microsoft Excel and WEKA. The respondents rated the developed software as "Moderately Acceptable" with overall mean of 4.31 which signifies that it is functional, usable, reliable and portable for the Distance Education (DE) stakeholders. For future research, the author may concentrate on greater number of instances using the other variables and may explore other data mining algorithms.

References

1. Bienkowski, M., Feng, M., & Means, B. 2012. Center for technology in learning sri international, U.S. Department of Education, Office of Educational Technology. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. 1-77.
2. Cavus, N. and Zabadi, T. 2014. A comparison of open source learning management systems, Elsevier Ltd. Procedia - Social and Behavioral Sciences 143, 521 – 526.
3. Comendador, B.E.V.and Guillo, A. C., 2014. A community cloud-based course management system using platform as a service (paas) for higher educational institutions, International Journal of Information and Education Technology, (IJIET), ISSN: 2010-3689 International Association of Computer Science and Information Community Press, 4
4. eLearning Industry 2015. Retrieved from Elearning Statistics and Facts For 2015 You Need to Know, on January 13, 2017
5. ECAR Working Group 2015. The predictive learning analytics revolution leveraging learning data for student success, Educase Center for Analysis and Research. Retrieved from http://net.educause.edu/ir/library/pdf/ewg1510.pdf
6. Jailia, M. & Tyagi A., 2013. Data mining: a prediction for performance improvement in online learning systems, International Journal of Advanced Research in Computer Science and Software Engineering, 3(7), 628-635.
7. Picciano, A. G., 2014. Big data and learning analytics in blended learning environments: benefits and concerns, International Journal of Artificial Intelligence and Interactive Multimedia, 2(7), 35-43.
8. Red, E. R., Cancino, G. C. C., Hanrath, M. J. W. V., & Ricardo, J. G. E., 2014. Analysis of the offline and online data of students using clustering and classification techniques, International Journal of the Computer, the Internet and Management, 22(3).
9. Romero, C., Ventura, S., & García,E. 2010. Data mining in course management systems: Moodle case study and tutorial, Computers & Education, 51(1), 368-384.
10. Sclater, N., 2015. "What do students want from a learning analytics app?"