ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 326

3rd International Conference on Current Issues in Education (ICCIE 2018)

Can We Predict Student Learning Performance from LMS data? A Classification Approach

Ashish Dutt
Department of Information Systems
Faculty of Computer Science & Information Technology (FSKTM)
University of Malaya
Kuala Lumpur, Malaysia
ashishdutt@yahoo.com.my

Maizatul Akmar Ismail
Department of Information Systems
Faculty of Computer Science & Information Technology (FSKTM)
University of Malaya
Kuala Lumpur, Malaysia
maizatul@um.edu.my

***Abstract*—The Learning Management System (LMS) is a common occurrence in most educational institutions. This system is a software application helping the educator in administration, facilitation, and tracking of course content to the learner. Educators have always been interested in understanding student interaction with systems like LMS. Such a system generates a plethora of data in a various form such as student performance on the individual course, activities, student behaviors, etc. The most prominent solutions involve performing dimensionality reduction technique to improve classifier accuracy and reducing the fewer error rates. Therefore, this study utilizes feature selection as a dimensionality reduction technique. The multiclass data were handled using the Learning Vector Quantization (LVQ) algorithm to identify significant predictors and thereby reducing the biased result. The efficiency of feature selection technique is evaluated with five different classifiers such as Linear Discriminate Analysis (LDA), Classification and Regression Tree (CART), k-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF). The performance of the classifier is evaluated using the kappa statistics and confusion matrix. Our extensive experimental results show that RF classifier produces optimum kappa statistic (85 %) with LVQ.**

Keywords*—*learning management system; classification; student performance; kappa statistic

I. INTRODUCTION

Learning Management System (LMS) is an e-learning tool that connects teachers with students beyond the traditional classroom for effective learning and activities [1]. The Higher Education Institutions (HEIs) adoption rate to LMS is gradually increasing [2] [3]. Maor and Volet concluded in their study [4], that interactivity with an LMS is an essential medium and constitutes a critical factor in online learning.

According to Nichani [5], Kaplan & Leiserson [6] and Itmazi et al. [7], an LMS is also known as "Content Management System (CMS) and Learning Content Management System (LCMS)". Generally speaking, an LMS is a self-contained system consisting of educational resources to direct the teaching and learning process by retaining, pursuing, and utilizing student interaction records within the LMS; a CMS gives resources for the design, administration, and distribution of teaching and learning resources on web-based platforms; and an LCMS merges the components of LMS with CMS. Although LMS and LCMS often overlap in functionalities, they differ on their primary purpose: LMS is inclusive of learner requirements, learner performance planning, and management, while LCMS deals only with the management of learning content [8]. It should also be mentioned that there does not exist a unified classification of such systems [9]. For example, some researchers like Depow [10] and Kennedy [11] have classified the Moodle system [12] as LMS, while others like Cole [13] classify it as CMS. In this study, we have adopted the term LMS in a wider context, that is the LMS is a self-contained catering to both student and teacher to enhance learning and teaching experience.

The data generated in educational settings such as LMS, traditional classroom-based interactions, etc. are utilized in understanding the student academic performance, and this process is termed as Educational Data Mining (EDM). It is relatively a new research area where methods are studied, developed, and tested to enhance the standard of teaching and learning. To this effect, in literature, there exist several notable reviews on the application of algorithms on the educational dataset to derive patterns (unsupervised approach) [14] or predictions (supervised approach) [15].

In this study, we are utilizing the supervised approach to predict student performance from interaction within an LMS, which is an emerging topic of interest among educators. This problem defines intricate activities to be observed and analyzed, suggesting resources to students favoring an effective learning process. The problem can be devised such that: student may complete several

interactions within a course helping them to strengthen the concepts acquired in class. On completion of the course, the students appear for the assessment. The students obtaining grades over a predefined threshold are labeled as 'pass' and those lower the threshold are labeled as 'fail'. Basis of this concept, the problem is to predict student performance dependent on the interactions with the LMS. The principal idea of this study is to determine if supervised algorithms can solve this problem efficiently.

This study is guided by two research questions, namely;

a. What are the factors that enrich the learning experience for students interacting with an LMS?

b. To predict the student academic performance from historical data obtained from LMS.

The remainder of this paper is organized as follows. In Section 2, we discuss the research method including the dataset description, exploratory data analysis, feature selection, and machine learning techniques. In Section 3, we have discussed the analysis and results obtained. Finally, Section 4 concludes the study.

II. RESEARCH METHOD

A. Dataset Description

The experiment was carried out using a real-world educational dataset on Kalboard 360 LMS [16]. The LMS system provided the user's synchronous access to its educational resources from a variety of devices. The dataset consisted of 480 learners in 16 features (12 categorical and 4 continuous) with no missing values and one response variable. The features were classified into three main divisions; (a) Demographic features that consisted of gender and nationality of the student, (b) Academic status attributes such as educational stage, grade level, and section, and (c) Behavioral attributes such as raised hand in class, visited resources, parents answering the school survey, and parents school satisfaction. The response variable had three levels namely low (values between 0-69), middle (values between 70-89), and high (values between 90-100). None of the continuous predictors suffered from near zero variance. Nor did we find any evidence of high correlation among the continuous predictors. In Table 1 and Table 2 we show the summary statistic for the continuous and categorical variables.

TABLE I. SUMMARY STATISTIC FOR CONTINUOUS VARIABLES

	Raise Hand	Visit Resources	View Announcement	Discussion
Minimum	0.0	0.0	0.0	1.0
1st Quartile	15.8	20.0	14.0	20.0
Median	50.0	65.0	33.0	39.0
Mean	46.8	54.8	37.9	43.3
3rd Quartile	75.0	84.0	58.0	70.0

TABLE II. SUMMARY STATISTIC FOR CATEGORICAL VARIABLES

Categorical Variable	Level	Frequency
Gender	Male	305
	Female	175
Nationality	Kuwait	179
	Jordan	172
	Palestine	28
	Iraq	22
	Lebanon	17
Birthplace	Kuwait	180
	Jordan	176
	Iraq	22
	Lebanon	19
	Saudi Arabia	16
Semester	First	245
	Second	235
Relation	Father	283
	Mother	197
Parent answering survey	Yes	270
	No	210
Parent satisfied with the school	Good	292
	Bad	188
Absent days	Above 7 days	191
	Under 7 days	289
Subject	IT	95
	French	65
	Arabic	59
	Science	51
	English	45
StageID	High School	33
	Lower level	199
	Middle school	248
GradeID	G-02	147
	G-04	48
	G-06	32

	G-07	101
	G-08	116
SectionID	A	283
	B	167
	C	30
Class (response variable)	High	142
	Low	127
	Medium	211

B. *Exploratory Data Analysis*

With respect to the first research question, we have formulated three hypotheses to understand student interaction with the LMS.

Hypothesis 1: Student interaction with an LMS and raising hands in class is dependent on the guardian.

In Figure 1, we show that the exploratory relationship between student interactions with LMS is dependent on the guardian. The male student with the father as a guardian raises more hands in class as compared to a male student with the mother as a guardian, and such students are also high performers. The girls with the mother as a guardian raise more hands in class as compared to girls with the father as a guardian. Such girls are high achievers when compared to boys with the mother as guardian. This is in sharp contrast to students (both boy and girl), who do not raise hands in class and are also of low performing students.

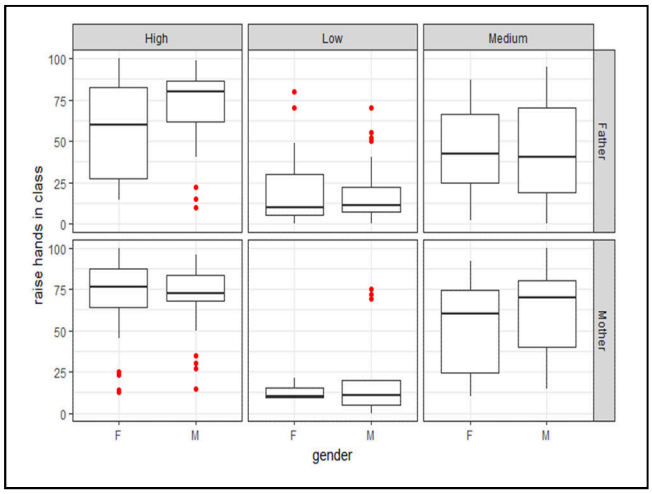

Fig. 1. Relationship between raising hands in class and guardian

Hypothesis 2: High performing girls actively participate in discussions as compared to high performing boys and its dependency on the guardian

Continuing further in Figure 2, it is interesting to note that high achieving girls with the father as a guardian are slightly more active in the discussion when compared to high achieving boys. This phenomenon is in sharp contrast to low achieving girls with the father as a guardian, more active in discussions when compared to the high achieving girls. Meanwhile, the low achieving girls with the mother as a guardian are the least active in discussions, compared with either medium or high achieving girls.

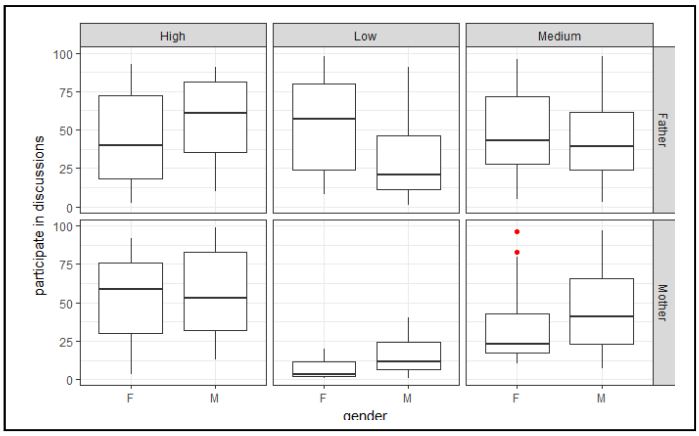

Fig. 2. Relationship between participation in discussion, gender, and guardian

Hypothesis 3: Visiting resources and raising hands in class are precursors to scoring low marks

Thereafter in Figure 3, we have visualized the relationship between visiting resources and raising hands in class for students obtaining high to medium level scores. It is interesting to note that there are students who visit resources and raise hands in the class but have low-level scores. Perhaps the reason for their low performance can be attributed to their absenteeism rate, which is higher than the students with middle-level scores; or, another reason for low-level scores could be viewing announcements, that is lower than students who have gained high and medium level scores.

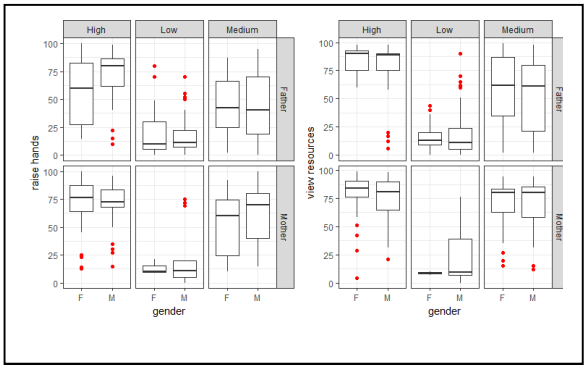

Fig. 3. Relationship between visiting resources and raising hands in class

1) *Feature Selection*

The process of feature selection is an imperative task to extract features that are free from high

correlation, biases, and unwanted noise, and to aid in accurate modeling. Initially, a control function was set up with a *k*-folds (10 folds) cross-validation as the controlling parameters. We then processed the dataset in a training function, where we applied the *Learning Vector Quantization (LVQ)* algorithm proposed by Kohonen [17]. LVQ is a supervised learning model. This model is a type of artificial neural network. We chose LVQ to deal with a multi-class dependent variable. Also, LVQ can support both binary and multi-class classification problems. Another reason to choose LVQ was that it is nonparametric, meaning that it does not rely on the method inference it is comparing. In this model, the goal is to determine a set of patterns that best represent each class. The information processing objective of the LVQ algorithm is to prepare a set of pattern (or model) vectors in the expanse of the detected input data instances and to use these vectors to classify unseen examples. In an iterative cycle, the vectors from the training sample occurring most frequently are selected until no more similarities are found. The repetition of this process results in the distribution of pattern vectors in training data instances, which approximates the underlying distribution of instances from the test set.

2) *Machine Learning Techniques*

In this study, we have applied an ensemble of classification methods to improve the accuracy of prediction. The methods applied were Classification and Regression Trees (CART), Linear Discriminant Analysis (LDA), k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), and Random Forest (RF). A tree is a directed graph consisting of the main root node and several branched sub-nodes. The common algorithm choices for prediction tasks are classification and regression. These are popular alternatives to regression (where the predictor variable is continuous), discriminant analysis, and other procedures based on algebraic models. A CART model combines both classification (requiring a categorical predictor variable) and regression trees. One of the common dimensionality reduction techniques is Linear Discriminant Analysis (LDA). This algorithm was proposed by Ronald A. Fisher in 1936 [18]. It works by determining the variables with maximum variance in a lower dimensional space to avoid overfitting. The original algorithm proposed by Fisher would work only for a 2-class level classification problem. It was further extended to handle multi-class level problems by C.R. Rao in 1948 [19]. The k-NN is a non-parametric method used for classification. In this algorithm, the training sample vectors undergo a voting process such that the vectors with maximum votes are selected. Commonly, the Euclidean distance is chosen to measure the distance between adjoining vectors. An SVM is a classifier that works by creating a set of hyperplanes. The vectors are segregated on these hyperplanes. The algorithm then searches for an optimum separation within the training vectors and on the hyperplanes. SVM is commonly used for outlier detection, regression and classification tasks. There have been relevant improvements in the classification accuracy. Popular classifiers like decision tree, random forests, have improved the classification accuracy manifold. For instance, RF is an ensemble of random vectors that control the growth of each tree in an ensemble. The generalization error for a forest closes to a limit as the number of trees in the forest becomes large. This error also depends upon the strength of each tree and the maximum correlation between them.

III. Analysis and Results

The experimentation was carried out in R programming language for data analysis, model building, and visualization [18]. As discussed in the previous section, the feature selection was carried out in the dataset using LVQ from the class package in R [19]. After setting the cutoff mark at 65%, we found eleven predictors (*visited resources, raised hands in class, student absent days, announcement views, parent answering the survey, parent relationship with the student, parent-school satisfaction rate, discussion in class, and student gender*) fulfilling this criterion. For subsequent analysis, we removed the six non-contributing predictors from the dataset. Figure 4 shows the feature importance plot.

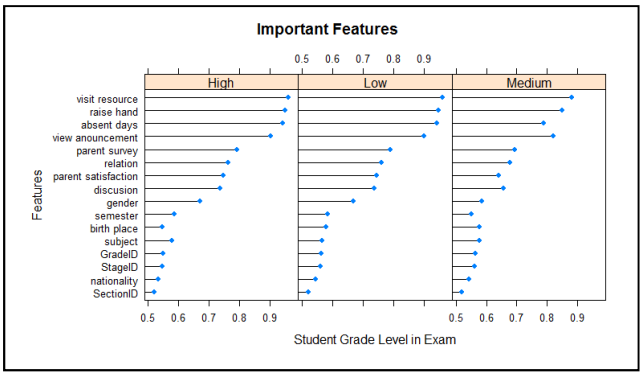

Fig. 4. Feature importance plot

The dataset was randomly split into a training set (70%) and test set (30%). The model validation was performed in the training set utilizing 10-folds cross-validation. The five machine learning algorithms discussed in the previous section were implemented on the eight-predictor variable. From the results, it is observed that RF produces a higher accuracy of 90%, when compared to SVM (81%), LDA (80%), CART (75%), and kNN (71%). "In a classification problem, the *kappa* statistic is a metric that compares an observed accuracy with an expected accuracy (random chance)" [20]. This metric helps in evaluating single or multiple classifier algorithms. It also considers the possibility of a chance event occurring by random. Therefore, it is more accurate than simply using accuracy for a metric. Essentially it

is a measure of how closely the vectors or data points are placed by the classifier. It also helps in checking the classifier performance as well as comparison of several kappa models. However, in literature, there does not exist a standard meaning of the kappa statistic. Thus, Landis & Koch [21] "considers 0-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1 as almost perfect". Fleiss [22], "considers kappa > 0.75 as excellent, 0.40-0.75 as fair to good, and < 0.40 as poor". Needless to state that both scales are somewhat optional. In this paper, we have followed the kappa interpretation given by Landis & Koch. Thus, the random forest classifier has the highest kappa=0.85 as compared to other classifiers. The accuracy and kappa statistic for the different algorithms are shown in Figure 5.

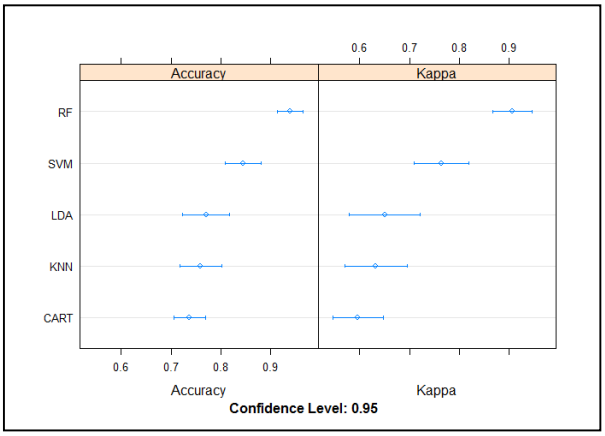

Fig. 5. Accuracy and Kappa metrics for the classifiers

Similar research to our study is undertaken by [16],[23]. In [16] the researchers have applied an ensemble of classification algorithms (decision tree, naïve bayes, and random forest) on student behavioral features (like visiting resources, participating in discussions, raising hands, etc.). The researchers applied bagging and boosting to improve the results. Similarly, the researchers in [23] have shown no evidence of data preprocessing and feature selection techniques used. However, in our experiments, we have obtained better results when we applied the LVQ algorithm for feature selection to deal with a multiclass problem (see section 2.3 for details). In Table 3, we show the comparative results.

TABLE III. COMPARATIVE RESULTS

Algorithms	[16]		[23]		Our Results	
	Accuracy (%)	F-measure	Accuracy (%)	F-measure	Accuracy (%)	Kappa
J48	75.8	75.9	75.8	75.9	-	-
Naïve Bayes	67	67.1	67.7	67.1	-	-
ANN	79.1	79.1	-	-	-	-
LDA	-	-	-	-	80	70
CART	-	-	-	-	75	62
KNN	-	-	-	-	71	56
SVM	-	-	-	-	81	71
RF	-	-	76.6	76.6	90	85

IV. CONCLUSION

With the modernization of instruments to facilitate educational resources, there has been an urgent requirement to determine the student performance in interacting with these instruments. One of these learning instruments used by educational institutions is an LMS. In this study, we have addressed the multilevel classification problem to extract important feature using LVQ. The feature importance indicates features such as *"visiting resources and raising hands in class"* are prominent variables for students obtaining high to medium level scores. It is interesting to note that there are still some students who visit resources and raise hands in the class but have low-level scores. Perhaps the reason for their low performance can be attributed to their absenteeism rate, which is higher than the students with middle-level scores; or, another reason for low-level scores could be viewing announcements, that is lower than students who have gained high and medium level scores. Another interesting pattern for parents satisfied with schools; the students with low-level scores have a relatively higher number of parents satisfied with school when compared to those with medium level scores.

Thereafter, we have compared the accuracy of prediction on five classifiers such as LDA, CART, KNN, SVM, and RF. The RF classifier gives optimum accuracy (90%) and kappa metric (85%) over other algorithms. Finally, to conclude this study, we have laid ample emphasis on initial data preprocessing and feature selection. This is a vital step and cannot be overlooked in any data analytical activity. The implication of unsupervised learning at the data preprocessing stage and its effect on classification performance and multiclass balancing are left as future work.

REFERENCES

[1] N. A. Adzharuddin and L. H. Ling, "Learning Management System (LMS) among University Students: Does It Work? f," *International Journal of e-Education, e-Business, e-Management and e-Learning,* vol. 3, p. 248, 2013.

[2] M. S. Kerr, K. Rynearson, and M. C. Kerr, "Student characteristics for online learning success," *The Internet and Higher Education,* vol. 9, pp. 91-105, 2006.

[3] J. Larreamendy-Joerns and G. Leinhardt, "Going the distance with online education," *Review of educational research,* vol. 76, pp. 567-605, 2006.

[4] D. Maor and S. Volet, "Interactivity in professional online learning: A review of research-based studies," *Australasian journal of educational technology,* vol. 23, 2007.
[5] M. Nichani. (2001). *LCMS= LMS+CMS - how does this affect the learner? And the instructional designer?*
[6] E. Kaplan-Leierson, "ASTD's Learning circuits glossary," *Retrieved on 16th September* 2002.
[7] J. A. Itmazi, M. G. Megías, P. Paderewski, and F. L. G. Vela, "A comparison and evaluation of open source learning management systems," *IADIS AC,* vol. 1, 2005.
[8] L. Greenberg, "LMS and LCMS: What's the Difference," *Learning Circuits,* vol. 31, 2002.
[9] R. K. Ellis, "LCMS roundup," *Learning Circuits,* vol. 8, 2001.
[10] J. Depow, "Open source software: Two learning management systems," *The International Review of Research in Open and Distributed Learning,* vol. 4, 2003.
[11] D. M. Kennedy, "Student managed to learn management systems: Teachers as designers," in *EdMedia: World Conference on Educational Media and Technology*, 2005, pp. 3172-3178.
[12] M. Dougiamas and P. Taylor, "Moodle: Using learning communities to create an open source course management system," 2003.
[13] J. Cole and H. Foster, *Using Moodle: Teaching with the popular open source course management system*: " O'Reilly Media, Inc.", 2007.
[14] A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review of Educational Data Mining," *IEEE Access,* vol. 5, pp. 15991-16005, 2017.
[15] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *JEDM-Journal of Educational Data Mining,* vol. 1, pp. 3-17, 2009.
[16] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Preprocessing and analyzing educational dataset using X-API for improving student's performance," in *Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on*, 2015, pp. 1-5.
[17] T. Kohonen, "Learning vector quantization," in *Self-Organizing Maps*, ed: Springer, 1995, pp. 175-189.
[18] R. Ihaka and R. Gentleman, "R: a language for data analysis and graphics," *Journal of computational and graphical statistics,* vol. 5, pp. 299-314, 1996.
[19] W. N. Venables and B. D. Ripley, *Modern applied statistics with S-PLUS*: Springer Science & Business Media, 2013.
[20] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement,* vol. 20, pp. 37-46, 1960.
[21] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics,* pp. 159-174, 1977.
[22] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin,* vol. 76, p. 378, 1971.
[23] B. Kapur, N. Ahluwalia, and R. Sathyaraj, "Comparative Study on Marks Prediction using Data Mining and Classification Algorithms," *International Journal of Advanced Research in Computer Science,* vol. 8, 2017.