

# Educational Data Mining for Monitoring and Improving Academic Performance at University Levels

Ezekiel U Okike<sup>1</sup>, Merapelo Mogorosi<sup>2</sup>  
Department of Computer Science  
University of Botswana, Gaborone, Botswana

**Abstract**—This study applied Educational Data Mining on 712 sample of logs extracted from Moodle Learning Management System (LMS) at an African University in order to measure students and staff patterns of use of the LMS resources and hence determine if the quantity of participation measured in the amount of time spent on the use of LMS resources improved academic performance of students. Data collected from Moodle LMS was preprocessed and analyzed using machine learning algorithms of clustering, classification and visualization from WEKA system tools. The dataset consisted of Course tools (Quiz, Assignment, Chat, Forum, URL, Folder and Files), Lecturer and Student usage of the tools. Furthermore, SPSS was used to obtain a matrix for coefficients of correlations for course tools, tests and final grade. Correlation analysis was done to verify if students use of course tools had impact on student's academic performance. Findings indicated the pattern of usage for course1 as Quiz (38358), System (17910), Forum (8663), File (8566), Assignment (1235), Folder (514), File Submission (172), and Chat (37); Course2 as System (11920), Quiz (8208), Forum (4476), File (4394), Assignment (257), Chat (247), URL (125), and File Submission (38); Course3 as System (2622), File (1022), Folder (570), Forum (258), and URL (2). Overall, evaluating the correlation between the use of LMS resources and students' performance, findings indicated there is significant relationship between the use of LMS resources and students' academic performance at 0.01 level of significant. The findings are useful for strategic academic planning purpose with LMS data at the university.

**Keywords**—Educational data mining; learning management systems; Weka system tools; improved academic performance

## I. INTRODUCTION

Data Mining (DM) and Machine Learning (ML) as sub-disciplines of computer science provide powerful tools for knowledge discovery from massive data sets [1,2]. As a process of discovering patterns in data a DM process must be automatic or semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage [3]. The concept of Educational Data mining (EDM) is defined in this paper as the application of data mining to derive meaningful patterns from educational system repositories which in turn could be used to improve teaching and learning experiences. One such educational teaching and learning system tool is the learning management system (LMS). LMS tools are computerized teaching and learning platforms for creating contents, delivering courses to

learners and managing courses in teaching and learning environments [4,5].

### A. Statement of the Problem

Many African universities have invested in learning management systems. Over the years, massive data have also been accumulated through the LMS, but which has not been appropriately mined in order to provide information useful for strategic decisions at university levels. The aim of this paper is to demonstrate with empirical evidence the usefulness of EDM in discovering hidden but useful patterns from teaching and learning data accumulated through the LMS. Several studies for example [4, 5, 6, 7, 8, 9] have also shown the central place of LMSs in teaching and learning, although, only a few addressed the need for EDM from logs of LMS's data. The specific objectives of this study were:

The specific objectives of this study were:

- 1) To measure students and staff use of LMS resources in teaching and learning at university level.
- 2) To evaluate the correlation between potential use of LMS resources and students' performances.

### B. Research Questions

The following research questions were investigated in the study:

- 1) What were the recognizable patterns in teaching and learning from Moodle LMS logs for the sampled data set?
- 2) Is there any correlation between potential use of LMS and prevailing students' performance at the university?

The rest of this paper counts of three sections, and a list of references. Section two presents a review of literature relevant to the study. Section three presents the empirical study and the methodology. Section four presents the results and discussion of the study.

## II. LITERATURE REVIEW

The application of Learning management systems and their effectiveness in higher education have been widely discussed in the literature [6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17]. A common approach to EDM modeling is a combination of sequential processes which includes data collection from LMSs, data preprocessing, data mining and analysis, preprocessing, result generation, and application (see Fig. 1).

A successful educational data mining (EDM) which produced satisfactory results at the University of Cordoba, Spain was reported in [9]. The EDM approach enabled the discovery of new rules of association which were used to improve the design of online courses at the university.

Similarly, strategies to mine data from activity logs found on Moodle LMS was investigated in [6]. Applying data mining and using simple statistics to analyze the logs the author recommended the use of Access watch analogue and Web start applications to infer student's attitudes to learning and for predicting examination scores through multiple regressions. Fig. 2 shows a data mining model using LMS.

Following a data cycle approach, Fig. 3 demonstrates the usefulness of data mining in the context of this paper. Every decision-making process is based on a data transformed into information culminating in a decision being made. The cycle begins with identifying the problem, collecting and storing data using appropriate tools, preprocessing data, mining data, and discovering new knowledge from data which provides necessary feedback from the system for future activities. Fig. 3 portrays the Data Cycle in knowledge discovery using data mining.

A brief explanation of each stage of the DCKD include:

- 1) *Problem definition*: An initial definition of the problem, or the mission, or the purpose, for which data is required.
- 2) *Identifying data sources*: Understanding what data are pertinent, and where they can be located.
- 3) *Data collection and storing*: Retrieval of data from various sources and storing them in an accessible location.
- 4) *Data mining*: Selection of relevant data out of the Big Data using appropriate DM and ML tools.
- 5) *Knowledge discovery*: New knowledge discovered and presented to decision makers (Classified, Clustered and Visualized).
- 6) *Learning and decision-making*: The final stage that is the purpose of the data cycle. The results are displayed to the decision makers and decisions are taken.
- 7) *Feedback for further cycles*: This stage is not always necessary. However, very often, the need to make a certain decision is repetitive, so the customer (the decision maker) can affect the usefulness and the effectiveness of the cycle by forwarding comments and changes.

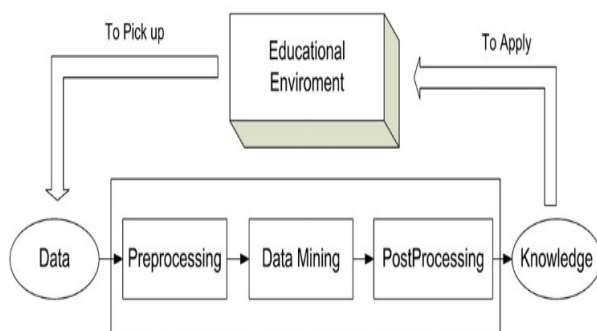


Fig. 1. EDM Processes Source: (García et al., 2011).

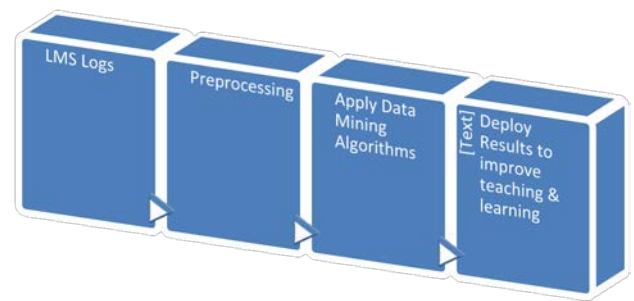


Fig. 2. Mining LMS Data.

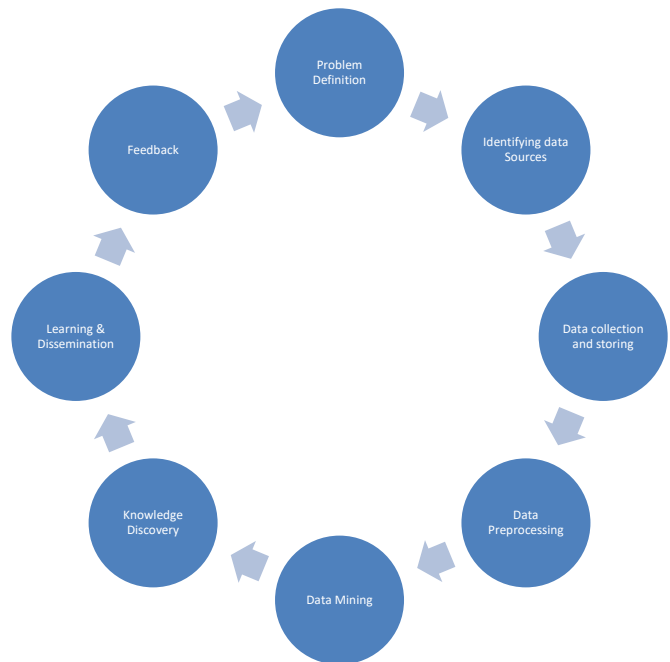


Fig. 3. Data Cycle in Knowledge Discovery (DCKD) using Data Mining.

### III. EMPIRICAL STUDY

#### A. Methodology

A sample of 712 Moodle data logs accumulated using Moodle Learning Management System were used in the study. The sample comprised three courses selected randomly from three faculties. The logs comprised of students, academic staff, and course data in the format: user full name, description, time, components, affected user, event context, event name, origin and IP address. The following criteria were observed in the sample selection:

- 1) Permission was obtained to conduct the study from which the data was obtained.
- 2) The experience of a lecturer offering a course on Moodle LMS was considered.
- 3) A selected course was offered in the first semester.
- 4) A selected course was taught by a lecturer who showed enthusiasm in online activities.

### B. Data Collection Instrument

The logs were downloaded as .csv files, prepared and preprocessed using a Waikato Environment for Knowledge Analysis (WEKA) tool.

### C. Data Preparation and Preprocessing

The preparation and preprocessing consisted of data extraction, cleaning, aggregation/ integration, filtering and transformation. Data was mined using machine learning schemes selected from the WEKA tool. The selected schemes included tools for pre-processing, classification, clustering and visualization. The algorithms were applied directly to each dataset for each stage of preprocessing, classification, clustering, and visualization invoked from the schemes menu as follows:

- 1) A data file is selected from the file menu.
- 2) Important attributes of the data are selected.
- 3) Aggregates of existing attributes were created using the spreadsheet.
- 4) A machine learning scheme was selected from the Schemes menu.
- 5) Results were viewed as trees, text or three- dimensional plots.

- 6) Attribute/aggregate selections were revised.
- 7) The scheme was re-run on the revised data.

Furthermore, in order to maintain format independence, the data was converted to an intermediate representation - Attribute Relation File Format (ARFF). Data logs of semester1 of the academic year 2017/18 were downloaded, extracted and used for the study. The extraction process is shown in Fig. 4.

Data filtering was done through the WEKA system filtering tool targeting the attributes required for use in the selected machine learning algorithms ZeroR and J48.

### D. Data Mining and Analysis using Machine Learning Algorithms ZERO R and J48

J48 and ZeroR machine learning algorithms were applied on the data sets. The justification for the choice of J48 and ZeroR algorithms in this study was due to their perceived usefulness in classification, clustering and visualization [18, 19] J48 algorithm uses unsupervised learning to form clusters of a data set, while ZeroR algorithm uses supervised learning. Both classifiers differentiate each case according to some set criteria.

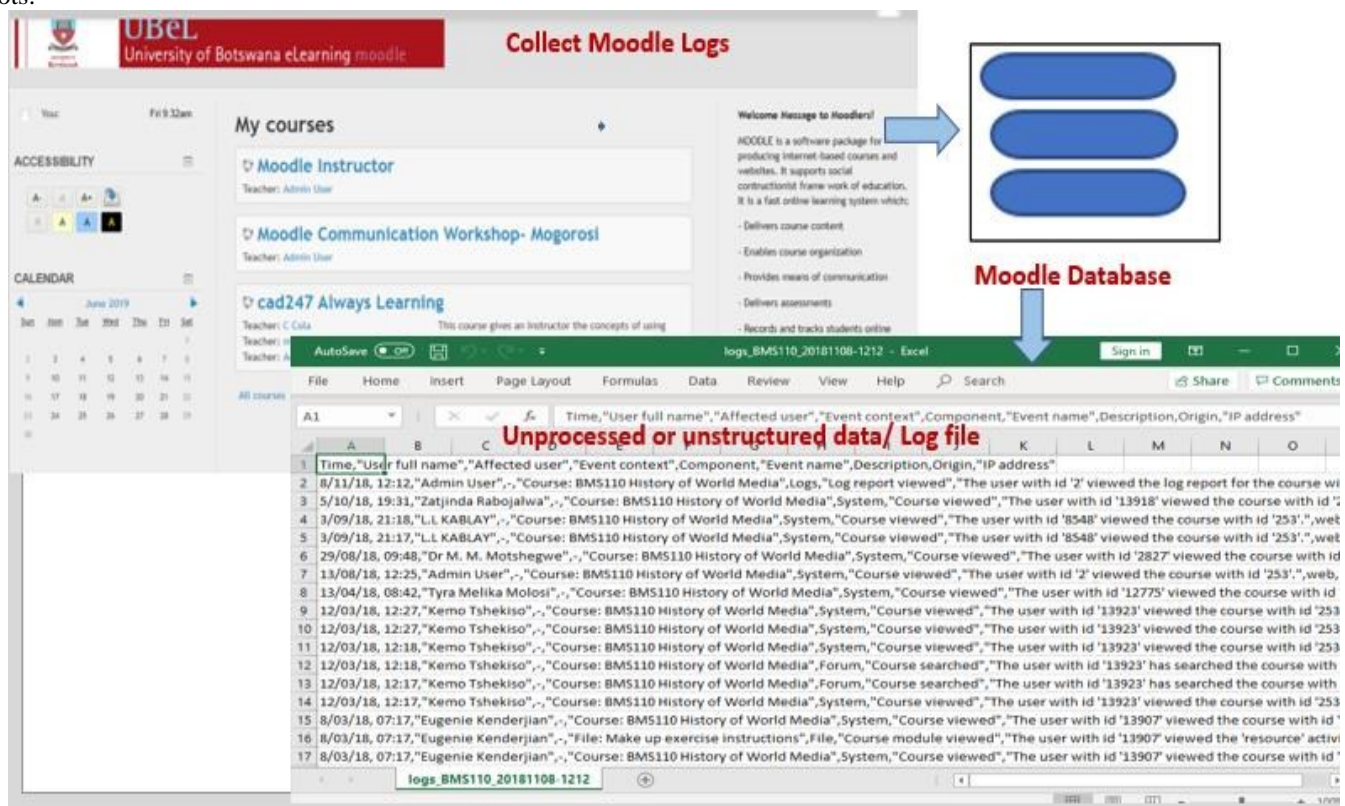


Fig. 4. Extracting Data from Moodle LMS for Data Mining.

## V. RESULTS AND DISCUSSION

### A. Classification with ZeroR and J48 Algorithms

Fig. 5 shows the results of using ZeroR to predict Course1. The prediction class value was Quiz which took 0.09 seconds to build the model. The number of correctly classified instances was 38530 (50.553%) while the number of incorrectly classified instances was 37687 (49.447%).

The mean absolute error and the root mean squared error for these predictions were 0.147 and 0.2711 respectively. There were five attributes (User Full Name, Event context, Components, Event Name and Origin) and the focus was on the Component attribute which explains how the course activities and resources were used during the semester. The accessibility of the logs of the course by a lecturer and the students was 76217 which is shown as the total number of instances. The number of correctly classified instances was 38530 (50.556%) and incorrectly classified instances was 37687 (49.4%). It should be noted that when an algorithm is based on probability, there is the risk of type 1 errors (false positives) and type 2 error (false negatives). This accounts for the noticeable mean absolute error, and the root mean squared error above. The detailed accuracy by class is shown below in the confusion matrix (Table I) which shows the activity tools with their accessibility values.

Table II, Table III, and Fig. 6 show the results of predicting activities in course1 using J48 algorithm.

From Table II the statistical summary of the results shows 75463 (99%) as number of correctly classified instances and 754 (0.98%) as incorrectly classified instances. The mean absolute error is 0.0028, root mean squared error is 0.039, and relative absolute error is 1.91%. The errors are minimal. The detailed accuracy by class is shown in the confusion matrix (Table III).

As a machine learning algorithm, J48 uses the decision tree technique. When applied to the LMS logs on course activities and resources, the pattern of activities classification shows that Quiz (column c) has the highest activity at 38358, followed by System (column b) at 17910, Forum (column d) at 8663, File (column f) at 8566, and Assignment (Column e) at 1238, Folder (column g) at 514, File submission (column h) at 172 and chat (column i) at 37.

Table IV, and Table V show the result of prediction of course 2 and its confusion matrix in classification using J48 algorithm.

From Table V the number of correctly classified instances is 29765 (99%), while the number of incorrectly classified instances is 276 (0.91%). The mean absolute error was 0.0026, root mean squared error was 0.0387, and relative absolute error was 1.84%. The errors are minimal. The detailed accuracy by class is shown below in Table VI.

Table V above shows the confusion matrix of all classes. The pattern shows that System column b) has the highest activity at 11920, followed by Quiz (column c) at 8308, Forum (column d) at 4476, File (column g) at 4398, Assignment (column f) at 257, Chat (column e) at 247, URL

(column h) at 125 and File submission (column i) at 38. Similarly, for ZeroR the prediction of course2 in classification is shown in Fig. 6, Table VI, and Table VII. The same explanation offered in J48 upholds except that the prediction identified the system as predictor as shown in Fig. 6 and Table VI.

The results of ZeroR predicted class value were Quiz (38530), System (18108), Forum (8802), File (8566), Assignment (1395), Folder (514), File submission (172), and Chat (122). ZeroR algorithm predicted Quiz tool to be the highest. This explains the pattern of how the lecturer taught the class. The access of the course (System:18108) led learners to use the quiz tool (38530) more often because they had to discuss (Forum 8802) the concept of the given topic and read notes (File:8566). The learners also had to work on the given assignments (1395) for the lecturer to check whether they had understood the concepts of the topic and submit (172) back their assignments for marking. Detailed accuracy by class is shown in Table VII, while the confusion matrix is shown in Table VIII.

The report on Confusion Matrix above (Table VIII) shows the activity tools with their accessibility values. The results of ZeroR predicted class value as System (12109), Quiz (8308), Forum (4485), File (4394), Assignment (309), URL (125), Chat (125) and File submission (38). ZeroR algorithm predicted System tool to be the highest followed by Quiz, Forum, File, Assignment, URL, and Chat. This explains the pattern of how the lecturer taught the class. The learners accessed the course (System:12109) more often which led them to use the quiz tool. The Table XI above and Table XII shows the results of using ZeroR to predict Course3. The figure has five attributes (User Full Name, Event context, Components, Event Name and Origin) and the focus was on the Component attribute which explains how the course activities and resources were used during the semester. The prediction class value was System which took 0 seconds to build the model. The number of correctly classified instances was 2622 (58.5007%) while the number of incorrectly classified instances was 1860 (41.4993%). The accessibility or the logs of the course by the lecturer and the students was 4482 which is shown as the total number of instances. The ZeroR algorithm is based on probability, hence the likely presence of errors. The Mean Absolute and Root Mean Squared Errors are minimal in the range of 0.1676 and 0.2894 respectively.

Table XII presents a detailed accuracy by class and Confusion Matrix in classification for ZeroR

From Table IX the results above have 5 attributes and focus was on component attribute which explains how the course activities and resources were used during the semester. The course activities were System (2622),

File (1022), Folder (570), Forum (258) and URL (2). This suggests that the Lecturer used the System to create Files, Forum and Folder for the course. Folders were probably used to add notes for students. Files might have been used to create teaching materials like notes while Forum might have been used to create discussion topics.

```

Scheme:          weka.classifiers.rules.ZeroR
Relation:        bms110_seml_2015-16-weka.filters.unsupervised.attribute.Remove-R1,3,8
Instances:       76217
Attributes:      5
                 User full name
                 Event context
                 Component
                 Event name
                 Origin
Test mode:       10-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: Quiz

Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      38530           50.553 %
Incorrectly Classified Instances    37687           49.447 %
Kappa statistic                     0
Mean absolute error                 0.147
Root mean squared error             0.2711
Relative absolute error              100 %
Root relative squared error          100 %
Total Number of Instances           76217
    
```

Fig. 5. Prediction of Course1 in Classification – ZeroR.

TABLE I. CONFUSION MATRIX- COURSE1

A	B	C	D	E	F	G	H	I	Classified as
0	0	8	0	0	0	0	0	0	A = Logs
0	0	18108	0	0	0	0	0	0	B=System
0	0	38530	0	0	0	0	0	0	C=Quiz
0	0	8802	0	0	0	0	0	0	D=Forum
0	0	1395	0	0	0	0	0	0	E=Assignment
0	0	8566	0	0	0	0	0	0	F=File
0	0	514	0	0	0	0	0	0	G=Folder
0	0	172	0	0	0	0	0	0	H=File submission
0	0	122	0	0	0	0	0	0	I=Chat

TABLE II. PREDICTION OF COURSE 1 IN CLASSIFICATION USING J48 ALGORITHM

Item	Value	
Correctly classified instances	75463	99.0107%
Incorrectly classified instances	754	0.9893 %
Kappa statistic	0.9851	
Mean absolute error	0.0028	
Root mean squared error	0.039	
Relative absolute error	1.9191%	
Root Relative squared error	14.3847%	
Total number of instances	76217	

TABLE III. CONFUSION MATRIX COURSE 1 USING J48 ALGORITHM

A	B	C	D	E	F	G	H	I	Classified as
8	0	0	0	0	0	0	0	0	A = Logs
0	17910	47	111	23	15	2	0	0	B=System
0	46	38358	100	26	0	0	0	0	C=Quiz
0	52	36	8663	35	16	0	0	0	D=Forum
0	30	42	88	1235	0	0	0	0	E=Assignment
0	0	0	0	0	8566	0	0	0	F=File
0	0	0	0	0	0	514	0	0	G=Folder
0	0	0	0	0	0	0	172	0	H=File submission
0	10	0	2	1	72	0	0	37	I=Chat

=== Run information ===

```

Scheme:      weka.classifiers.rules.ZeroR
Relation:    bns201_seml_2014-15-weka.filters.unsupervised.attribute.Remove-R1,3,8
Instances:   30041
Attributes:  5
             User full name
             Event context
             Component
             Event name
             Origin
Test mode:   10-fold cross-validation
    
```

=== Classifier model (full training set) ===

ZeroR predicts class value: System

Time taken to build model: 0.06 seconds

Fig. 6. Prediction of Course 2 in Classification using ZeroR.

TABLE IV. PREDICTION OF COURSE 2 IN CLASSIFICATION USING J48 ALGORITHM

Item	Value	
Correctly classified instances	29765	98.0813%
Incorrectly classified instances	276	0.9187 %
Kappa statistic	0.9872	
Mean absolute error	0.0026	
Root mean squared error	0.0387	
Relative absolute error	1.8404%	
Root Relative squared error	14.46437%	
Total number of instances	30041	

TABLE VI. CONFUSION MATRIX COURSE2 – J48

A	B	C	D	E	F	G	H	I	J	Classified as
0	2	0	0	0	0	0	0	0	0	A = Logs
0	11920	115	41	6	4	22	1	0	0	B=System
0	0	8308	0	0	0	0	0	0	0	C=Quiz
0	1	8	4476	0	0	0	0	0	0	D=Forum
0	0	4	0	247	0	0	0	0	0	E=Chat
0	0	0	0	0	257	0	0	0	0	F=Assignment
0	0	0	0	0	0	4394	0	0	0	G=File
0	0	0	0	0	0	0	125	0	0	H=URL
0	0	0	0	0	0	0	0	38	0	I=File submission
0	1	0	0	0	0	0	0	0	0	J=Activity report

TABLE VII. PREDICTION OF COURSE 2 IN CLASSIFICATION USING ZEROR ALGORITHM

Item	Value	
Correctly classified instances	12109	40.30807%
Incorrectly classified instances	17932	59.6918 %
Kappa statistic	0	
Mean absolute error	0.1434	
Root mean squared error	0.2678	
Relative absolute error	100%	
Root Relative squared error	100%	
Total number of instances	30041	

TABLE VIII. CONFUSION MATRIX- COURSE2

A	B	C	D	E	F	G	H	I	J	Classified as
0	2	0	0	0	0	0	0	0	0	A = Logs
0	12109	0	0	0	0	0	0	0	0	B=System
0	8308	0	0	0	0	0	0	0	0	C=Quiz
0	4485	0	0	0	0	0	0	0	0	D=Forum
0	270	0	0	0	0	0	0	0	0	E=Chat
0	309	0	0	0	0	0	0	0	0	F=Assignment
0	4394	0	0	0	0	0	0	0	0	G=File
0	125	0	0	0	0	0	0	0	0	H=URL
0	38	0	0	0	0	0	0	0	0	I=File submission
0	1	0	0	0	0	0	0	0	0	J=Activity report

TABLE IX. PREDICTION OF COURSE 3 IN CLASSIFICATION USING ZEROR ALGORITHM

Item	Value	
Correctly classified instances	2622	58.5007%
Incorrectly classified instances	1860	41.4993 %
Kappa statistic	0	
Mean absolute error	0.1676	
Root mean squared error	0.2894	
Relative absolute error	100%	
Root Relative squared error	100%	
Total number of instances	4482	

TABLE X. CONFUSION MATRIX- COURSE2

A	B	C	D	E	F	G	H	I	J	Classified as
0	6	0	0	0	0	0	0	0	0	A = Logs
0	2622	0	0	0	0	0	0	0	0	B=System
0	1022	0	0	0	0	0	0	0	0	C=File
0	258	0	0	0	0	0	0	0	0	D=Forum
0	570	0	0	0	0	0	0	0	0	E=Folder
0	2	0	0	0	0	0	0	0	0	F=Activity report
0	2	0	0	0	0	0	0	0	0	G=URL

Similarly, Fig. 7 shows the J48 pruned tree and the event context is course3. The summary of classification (Table X) shows correctly classified instances as 4482 (100%). The root mean absolute error is 0, the root mean squared error is 0, and the relative absolute error is also 0. Table XI shows the confusion matrix where the leading diagonal displays numbers which represent interaction with the system and the total of these numbers represent the logging behavior of students and the lecturer(s). Therefore, the algorithm J48 predicts System as the major classifier.

Fig. 7 shows prediction of course3 activities in classification using J48 with 5 attributes. Table X shows the summary of the results reported correctly classified instances as 4482 (100%), and 0 (0%) as incorrectly classified instances. The mean absolute error was 0, root mean squared error was 0, and relative absolute error was 0%. The errors are minimal.

Table XI shows 7 classes (Logs, System, File, Forum, Folder, Activity report and URL) in the confusion matrix for Course3. The pattern shows that System (column b) has the highest activity at 2622, followed by File (column c) at 1022, Folder (column e) at 570, Forum (column d) at 258 and, URL (column g) at 2.

#### B. Pattern of Usage Discuss

1) *Research question 1:* With reference to research question1, the data analysis revealed a mixed picture. There is a substantial use of the quiz tool (about 70%) for assessment purposes such as gauging the level of achievement of instructional objectives. It is worth noting that another facility which was also used substantially was the Resources tools (File, Folder and URL) which were mainly used for posting notes and communication between lecturer and students. They can be considered as an entry point for lecturers to make teaching and learning digital. In all, it appears that the pattern

of usage identified above are complemented by blended learning where both traditional and technology-based approaches are mixed depending on the instructional goals.

2) *Research question 2:* With reference to research question 2 patterns were already revealed in the results presented. It was clearly observed that student's login into their Moodle portal to check for new course content (system use). Usually, students would have been alerted by the lecturer on the existence of new contents. Students could download the course content into their personal devices to read and to discuss amongst themselves. However, they could also do limited discussions using forums created by the lecturer.. Students might spend minimal time online when the course contents were not engaging. Furthermore, students might have exhibited, "the student syndrome", where there was a rush to do assignments just before the due date. These activities were discernible from the students' activity logs and their dates of submissions.

On the side of lecturers, the typical pattern of usage suggested, posting of notes using the resources tool and creating forums. In addition, it was observed that lecturers made good use of quizzes, assignments, chats and URLs.

In terms of correlation between potential use of LMS and prevailing students' performance (Research question 2).

Fig. 8 and Table VII, Table XIII and Table XIV show student's interaction with LMS and performance at final exam.

Fig. 8 shows that Files, Quiz, Test1, Test2 and Final Exams contributed to final grade. The correlation is explained in Tables XII, XIII and XIV for course 1, course 2 and course 3 respectively. The tables only reflect correlating components at 0.01 level of significance.



```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    bns305_seml_2014-15-weka.filters.unsupervised.attribute.Remove-R1,3,8
Instances:   4482
Attributes:  5
              User full name
              Event context
              Component
              Event name
              Origin
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

```

Fig. 7. Prediction of Course3 in Classification – J48.

TABLE XI. SUMMARY OF CLASSIFICATION

Item	Value	
Correctly classified instances	4482	100%
Incorrectly classified instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0 %	
Root Relative squared error	0%	
Total number of instances	4482	

TABLE XII. CONFUSION MATRIX USING J48- COURSE3

A	B	C	D	E	F	G	Classified as
6	0	0	0	0	0	0	A = Logs
0	2622	0	0	0	0	0	B=System
0	0	1022	0	0	0	0	C=File
0	0	0	258	0	0	0	D=Forum
0	0	0	0	570	0	0	E=Folder
0	0	0	0	0	2	0	F=Activity report
0	0	0	0	0	0	2	G=URL

TABLE XIII. SUMMARY OF SIGNIFICANT CORRELATIONS IN COURSE 1

Tool	Files	Folder	Quiz	Final Exam	Test1	Test2	Final Grade
Files		.390	.402				
Folder	.390		.869	.180			.171
Quiz	.402	.869		.207	.122	.133	.202
Final Exam		.180	.207			.452	.798
Test1				.417		.334	.557
Test2				.452	.334		.627
Final Grade		.171	.202	.798	.557	.627	

Correlation is significant at the 0.01 level (2-tailed)

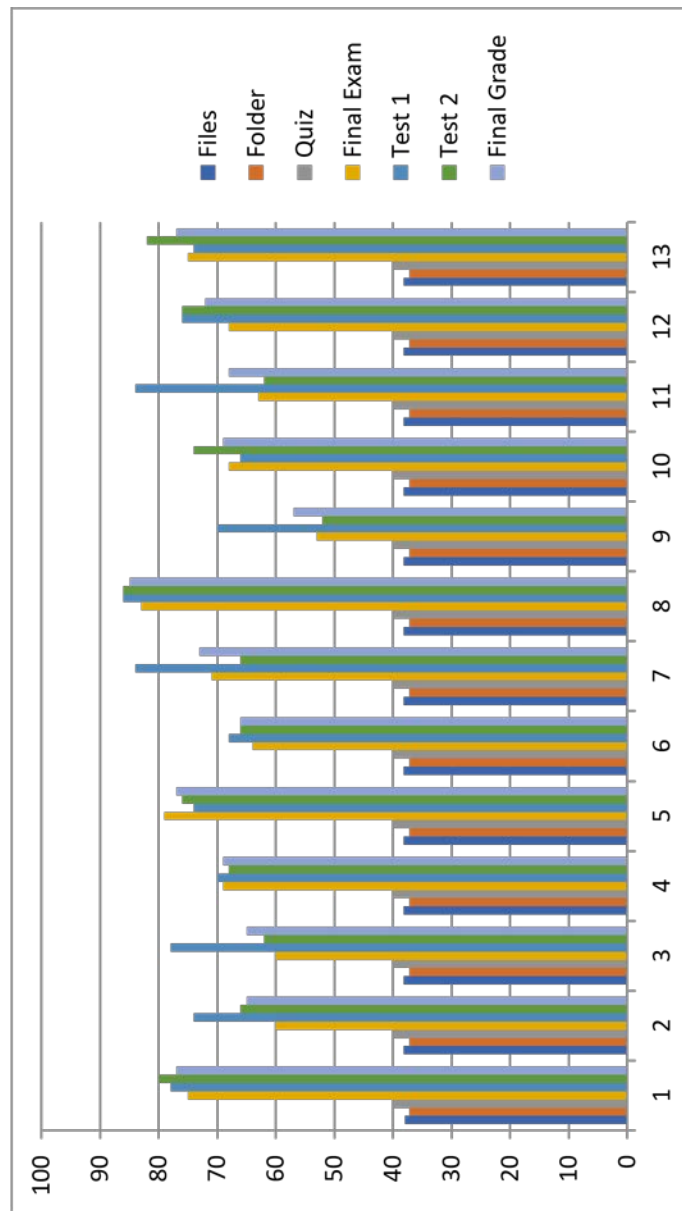


Fig. 8. Data Analysis based on Correlation Matrix of Course Tools and Students Academic Performance.

TABLE XIV. SUMMARY OF SIGNIFICANT CORRELATIONS IN COURSE 2

Tool	Files	Folder	Quiz	Assignment	Forum	Final Exam	Test1	Test2	Final Grade
Files		0.342	0.607	0.675	0.540				0.232
Folder	0.342		0.859	0.816	-0.726	0.288		0.200	0.372
Quiz	0.607	0.859		0.901	-0.786	0.254			0.388
Assignment	0.675	0.816	0.901		-0.885	0.302	0.212	0.241	0.438
Forum	-0.540	-0.726	-0.786	-0.885		-0.281	-0.226	-0.224	0.419
Final Exam		0.288	0.254	-0.302	-0.281		0.465	0.563	0.761
Test1				0.212	-0.226	0.465		0.416	0.582
Test2		0.200		0.241	-0.224	0.563	0.416		0.564
Final Grade	0.232	0.372	0.388	0.438	-0.419	0.731	0.582	0.564	

Correlation is significant at the 0.01 level (2-tailed)

TABLE XV. SUMMARY OF SIGNIFICANT CORRELATION OF STUDENTS PERFORMANCE – COURSE3

Tool	Files	Folder	Quiz	Assignment	Forum	Final Exam	Test1	Test2	Final Grade
Files		0.394	0.406	0.540	0.513				0.144
Folder	0.394		0.869	0.858	-0.671	0.184		0.145	0.276
Quiz	0.406	0.869		0.847	-0.724	0.209	0.144	0.156	0.286
Assignment	0.540	0.858	0.847		-0.742	0.216	0.148	0.174	0.298
Forum	-0.513	-0.671	-0.727	-0.742		-0.232	-0.211	-0.202	-0.331
Final Exam		0.184	0.209	0.216	-0.232		0.414	0.432	0.731

Correlation is significant at the 0.01 level (2-tailed)

### C. Analysis of Correlations

In Course1 (Table XII) the correlation between students' interactions with LMS and their performance were observed as follows:

There is significant relationship at 0.01 level between Files and Folder (0.390), Files and Quiz (0.402); Quiz and Files (0.402), Quiz and Folder (0.869), Quiz and Final Exam (0.207), Quiz and Final Grade (0.202); Final Exam and Folder (0.180), Final Exam and Quiz (0.207), Final Exam and Test1 (0.417); Final Exam and Test2 (0.452), Final Exam and Final Grade (0.709); Test1 and Final Exam (0.417), Test1 and Test2 (0.334), Test1 and Final Grade (0.557); Test2 and Final Exam (0.452), Test2 and Test1 (0.334), Test2 and Final Grade (0.627); Final Grade and Folder (0.171), Final Grade and Quiz (0.202); Final Grade and Final Exam (0.798), Final Grade and Test1 (0.557), Final Grade and Test2 (0.627); Therefore, the various components influenced final grade. Hence, there is significant relationship between use of LMS tools and students' performance. Since Pearson Coefficient value is greater than 0.01, we accept null hypothesis (research question 2).

In course 2 (Table XIII). There is significant relationship at 0.01 level between Files and Folder (0.342), Files and Quiz (0.607), Files and Assignment (0.675), Files and Forum (-0.540) and Files and Final Grade (0.232); Folder and Files (0.342), Folder and Quiz (0.859), Folder and Assignment (0.816), Folder and Forum (-0.726), Folder and Final Exam (0.288), Folder and Test2 (0.200) and Folder and Final Grade (0.372); Quiz and Files (0.607), Quiz and Folder (0.859), Quiz and Assignment (0.901), Quiz and Forum (-0.786), Quiz and Final Exam (0.254) and Quiz and Final Grade (0.388); Assignments and Files (0.675), Assignments and Folder (0.816), Assignments and Quiz (0.901), Assignments and Forum (-0.885), Assignments and Final Exam (0.302), Assignments and Test1 (0.212), Assignments and Test2 (0.241) and Assignments and Final grade (0.438); Forum and Files (-0.540), Forum and Folder (-0.726), Forum and Quiz (-0.786), Forum and Assignment (-0.885), Forum and Final Exam (-0.281), Forum and Test1 (-0.226), Forum and Test2 (-0.224), Forum and Final Grade (-0.419); Final Exam and Folder (0.288), Final Exam and Quiz (0.254), Final Exam and Assignment (0.302), Final Exam and Forum (-0.281), Final Exam and Test1 (0.465), Final Exam and Test2 (0.563), Final Exam and Final Grade (0.761); Test1 and Assignment (0.212), Test1 and Forum (-0.226), Test1 and Final Exam (0.465), Test1 and Test2 (0.416), Test1 and Final Grade (0.582); Test2 and Folder (0.200), Test2 and Assignment (0.241), Test2 and

Forum (-0.224), Test2 and Final Exam (0.563), Test2 and Test1 (0.416), Test2 and Final Grade (0.564); Final Grade and Files (0.232), Final Grade and Folder (0.372), Final Grade and Quiz (0.388), Final Grade and Assignment (0.438), Final Grade and Forum (-0.419), Final Grade and Final Grade (0.761), Final Grade and Test1 (0.582), Final Grade and Test2 (0.564); Therefore, the various components influence final grade. Since Pearson Coefficient value is greater than 0.01, we accept null hypothesis and conclude that there is significant relationship between use of LMS and prevailing students' performance. The results show that the Files, Folder, Quiz, Assignment, Forum, Final Exam, Test1 and Test2 all had an impact on academic performance significantly. In this case the lecturer used the tools (research question2).

In course 3 (Table XIV) There is significant relationship at 0.01 level between Files and Folder (0.394), Files and Quiz (0.406), Files and Assignment (0.540), Files and Forum (-0.513) and Files and Final Grade (0.144); Folder and Files (0.394), Folder and Quiz (0.869), Folder and Assignment (0.858), Folder and Forum (-0.671), Folder and Final Exam (0.184), Folder and Test2 (0.145) and Folder and Final Grade (0.276); Quiz and Files (0.406), Quiz and Folder (0.869), Quiz and Assignment (0.847), Quiz and Forum (-0.727), Quiz and Final Exam (0.209), Quiz and Test1 (0.144), Quiz and Test2 (0.156) and Quiz and Final Grade (0.286); Assignments and Files (0.540), Assignments and Folder (0.858), Assignments and Quiz (0.847), Assignments and Forum (-0.742), Assignments and Final Exam (0.216), Assignments and Test1 (0.148), Assignments and Test2 (0.174) and Assignments and Final Grade (0.298); Forum and Files (-0.513), Forum and Folder (0.671), Forum and Quiz (-0.727), Forum and Assignment (-0.742), Forum and Final Exam (-0.232), Forum and Test1 (-0.211), Forum and Test2 (-0.202), Forum and Final Grade (-0.331); Final Exam and Folder (0.184), Final Exam and Quiz (0.209), Final Exam and Assignment (0.216), Final Exam and Forum (-0.232), Final Exam and Test1 (0.414), Final Exam and Test2 (0.432), Final Exam and Final Grade (0.731). Therefore, the various components influence final grade. Hence there is significant relationship between use of LMS tools and student's performance. Since Pearson Coefficient value is greater than 0.01, we accept null hypothesis and conclude that there is significant relationship between use of LMS and prevailing students' performance.

### VI. CONCLUSION

In conclusion, using our approach of educational data mining and LMS, it was possible to monitor students and staff use of the LMS resources at the university. In addition it was

obvious from correlation analysis that students use of the resources could affect academic performance. In the future, it is suggested that this study be conducted with several courses from different disciplines in order to determine if academic disciplines affect students' performance based on the same tools.

#### REFERENCES

- [1] G. Mariscal, O. Marban and C. Fernandez, "A survey of data mining and knowledge discovery process models and methodologies," *The Knowledge Engineering Review*, Vol. 25:2, 2010, pp. 137-166.
- [2] O. Marban, G. Mariscal, and J. Segovia, "A data mining and knowledge discovery process model," *Open Access Data Base www.intechweb.org*, 2009, pp. 1-16.
- [3] H. I. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Technology*, 2nd ed, New York:Elsevier, 2005.
- [4] E. U. Okike, M. Mosanako, and M. Mogorosi, "Academic data Federation, Knowledge Base Construction and Heterogeneous Data Analysis for Improved management at University levels," *Journal of Applied Information Science and Technology*, vol. 11, no. 2, 2018, pp. 21-30.
- [5] S. N. Hamade, "Students Perceptions of Learning Management Systems in a University Environment : Yahoo Groups Vs BlackBoard, Ninth International Conference on Information Technology, New generations 2012, pp. 594-599.
- [6] Y. B. Kurata R. M. L. P. Bano, M. C. T. Marcelo, "Effectiveness of Learning Management System Application in the Learnability of Tertiary Students in an Undergraduate Engineering Program in the Philippines". In: Andre T. (eds) *Advances in Human Factors in Training, Education, and Learning Sciences. AHFE 2017. Advances in Intelligent Systems and Computing*, vol 596, 2018. Springer, Cham. [https://doi.org/10.1007/978-3-319-60018-5\\_15](https://doi.org/10.1007/978-3-319-60018-5_15).
- [7] Y. Ghlay, Effectiveness of Learning Management Systems in Higher Education: Views of Lecturers with Different Levels of Activity in LMS, " *Journal of Online Higher Education*, vol. 3, no. 2, 2019 pp. 29-50.
- [8] C. Romero, S. Ventura and E. Garcia, "Data Mining in Course Management Systems:Moodle case Study and Tutorial," *Computers & Education*, vol. 51, 2008, pp. 368-384.
- [9] E. Garcia, C. Romero, S. Ventura, and C. de Castro, "A collaborative Educational Association Rule Mining tool," *The Internet and Higher Education*, vol. 14, no. 2, 2011, pp. 77-88.
- [10] A. K. Alhazmi and A. A. Rahman, "Why LMS failed to support student learning in higher education institutions," *2012 IEEE Symposium on E-Learning, E-Management and E-Services*, Kuala Lumpur, 2012, pp. 1-5, doi: 10.1109/IS3e.2012.6414943.
- [11] R. Babo, and A. Azevedo, *Higher Education Institutions and Learning Management Systems: Adoption and Standardization*, IGI Global, 2012.
- [12] J. G. Boticario, and O. C. Santos, "Issues in Developing Adaptive Learning Management Systems for Higher Education Institutions," Available <https://core.ac.uk/reader/55533720>. Retrieved 2 Oct. 2020.
- [13] N. Darko-Adjei, "Students Perceptions and Use of the Sakai Learning Management System in the University of Ghana. Available <http://ugspace.ug.edu.gh/handle/123456789/26847>.
- [14] H. Coates, R. James, & G. A. Baldwin, "Critical Examination Of The Effects Of Learning Management Systems On University Teaching And Learning," *Tert Educ Manag* 11, 19-36 (2005). <https://doi.org/10.1007/s11233-004-3567-9>
- [15] L. V. Ngeze, "Learning Management Systems in Higher Learning Institutions in Tanzania: Analysis of Students' Attitudes and challenges towards the use of UDOM LMS in teaching and learning at the University of Dodoma," 2016. Available at <https://pdfs.semanticscholar.org/2c36/89101b4a64c85c25f3179c5a95e50f8a719a.pdf> retrieved 2 Oct. 2020.
- [16] M. F. Paulsen, "Experiences with Learning Management Systems in 113 European Institutions," *Journal of Educational Technology & Society*, vol. 6, No. 4, 2003, pp.134-148
- [17] N. Fathima, D. M. Shannon, and M. Ross, Expanding The Technology Acceptance Model (TAM) to Examine Faculty Use of Learning Management Systems (LMSs) In Higher Education Institutions, " *Journal of Online Learning and Teaching*, vol. 11, no. 2, June, 2015, pp. 210.
- [18] V. Mhetre, "Classification based Data mining Algorithms to Predict slow average and fast learners in educational systems using Weka," *ICCM*, 2017, pp. 475-479.
- [19] J. Talukdar, S. K. Kalita, "Detection of Breast Cancer using Data mining tool (Weka)," *International Journal of Scientific & Engineering Research*, vol. 6, no. 11, 2015, pp. 1124-1128.