Academic Data Federation, Knowledge Base Construction and Heterogeneous Data Analysis For Improved Management at University Levels

Ezekiel U. Okike[1], Seamogano Mosanako[2] and Merapelo Mogorosi[3]
Department of Computer Science[1], Department of Media Studies[2] and Centre for Academic Development[3]
University of Botswana, Gaborone, Botswana

Email:**euokike@gmail.com[1],mosanakos@ub.ac.bw[2] and msmogorosi@gmail.com[3]**

Abstract

Purpose – The main objective of this study is to mine data which has been accumulated through Black Board (BB) learn and Moodle Learning Management Systems (LMSs) over the past 15 years in order to evaluate what has taken place and how the information can be useful for strategic planning at the University of Botswana.

Design/Methodology – This pilot study used three course files consisting of over 20000 instances of the data set. Three objectives and three research questions guided the study. Data mining- the non-trivial extraction of implicit, previously unknown, and potentially useful information from data is employed in this study. Using an educational data mining technology tool WEKA, massive stored data about students learning and teaching using Moodle and Blackboard (BB) Learn at the University of Botswana are pre-processed, classified, clustered, and visualized.

Findings – Valuable information about students engagement with learning resources provided by the Learning Management Systems (LMS), students study and learning pattern, lecturers engagement and interaction with students on the LMS platforms (BB and Moodle) were discovered. In all the three course files, Quiz has the highest rating of 8697.00, 8308.00, and 38530.00 respectively for course files 1, 2, 3. This is followed by Forum, Assignment and Chat. This shows that the course lecturer spends more time on Quizzes. In course 3, the lecturer concentrated efforts on Quizzes and very little time on the other resources including forum, chat and assignment (Research questions 1,2,3). This suggests that teaching and learning in this in this course is not fully engaging, and the lecturer is not utilizing the LMS resources fully. In course 2, the same conclusion was reached, although the lecturer used a bit of forum (Research questions 1,2,3). In course 1, the lecturer also gave more attention to quizzes and forums, assignments, and less of chats.

Practical Implications - This report could guide improvement on teaching and learning using learning management platforms at the university.

Keywords – Data Visualization, Knowledge Base Construction, Educational Data Mining, Learning Management Systems, WEKA tools.

INTRODUCTION

Academic data refers to all data relating to students, courses and course lecturers in an academic environment. The academic environment of universities world-wide include its immediate environment with all stakeholders and their organs both within and outside of the immediate environment. Within the immediate environment of a university data about students, courses and course lecturers are vital components of the university system for planning and administrative purposes. Therefore, vital academic data must be maintained and accessed using modern technology for teaching, learning and reporting.

The concept of data federation (also data visualization) is a means to achieving the need of mapping multiple autonomous data sources into a single federated database**.** *A federated database system is a type of meta-database*

management system (DBMS), which maps multiple autonomous database systems into a single federated database[16]. The constituent databases may be interconnected through a computer network such as an intranet or an internet if the data sources are geographically decentralized [9,10]

Traditionally, the challenges of efficient organization, storage and retrieval of relevant academic data for improved management at university levels stem from the heterogeneous nature of academic data. However, with advances in Database technology and Information Processing and Retrieval Systems, volumes of stored academic data can be appropriately federated and harnessed to aid the decision making and academic planning at university levels. Therefore, this study explores the use of Learning Management software tools such as BlackBoard Learn and Moodle and their role in providing useful academic data for strategic planning and decision making at university levels.

A. Learning Management tools in Higher Education

In today's university teaching and learning settings, all academic work is technology dependent. As such, universities invest in appropriate teaching and learning software tools generally referred to as Course Management Systems (CMS) or Learning Management software systems (LMS) for both pedagogical and administrative purposes of teaching and learning solutions at University levels [6]. As observed in [7] LMS platforms are increasingly important academic systems, in higher education . Two main LMS/CMS tools notably used by Universities are BlackBoard and Moodle. In some African Universities for instance, Moodle and Blackboard learning management systems are used to provide a uniform interface for the faculty and students' teaching and learning. Both Moodle and Blackboard platforms enable the creation and distribution of teaching materials, assessment of students, communication and collaboration among users. In general, institutions make use of learning management systems (LMS) to plan, implement, facilitate and monitor students learning [1]. In using any LMS, it is suggested that the lecturer's role be redefined to accomplish a shift in pedagogy from an instructor-centred to a learner-centred environment [2]. In this regard, the lecturer makes available all necessary learning materials and assessment records, file transfers, and discussion forum. This enhances and guarantees the potentials of the LMS which includes convenient , efficient and effective course delivery, effective assessment, communication between lecturers and students and among the students themselves. Therefore, adequate use of LMS could make teaching and learning interesting and successful. At the University of Botswana for instance, the use of BlackBoard (BB) Learn started in 2001 as WEbCT, and Moodle in 2010. Figure 1 shows the number of courses on BB between 2008 and 2017. Figure 2 shows the number of courses on Moodle between 2010 and 2017.

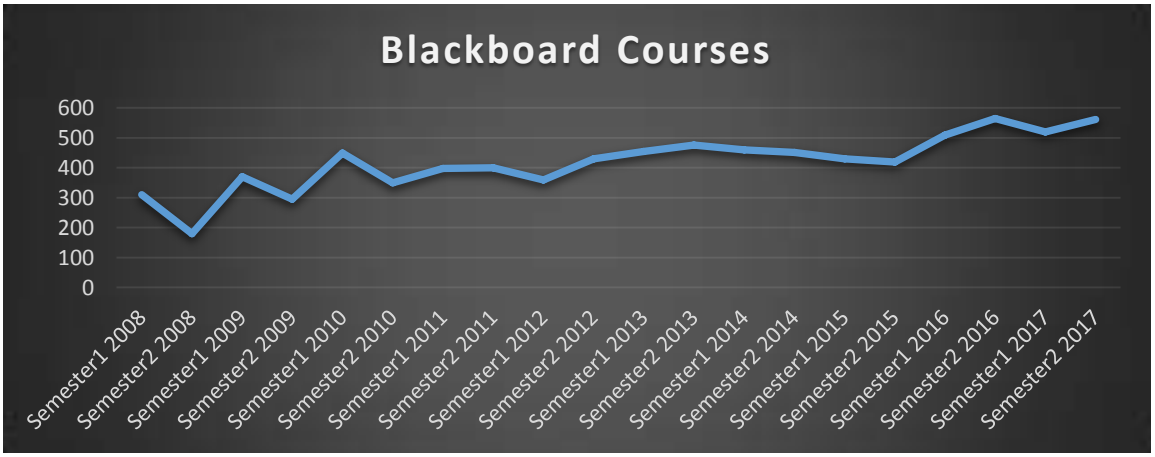


Figure 1. Distribution of courses on BlackBoard

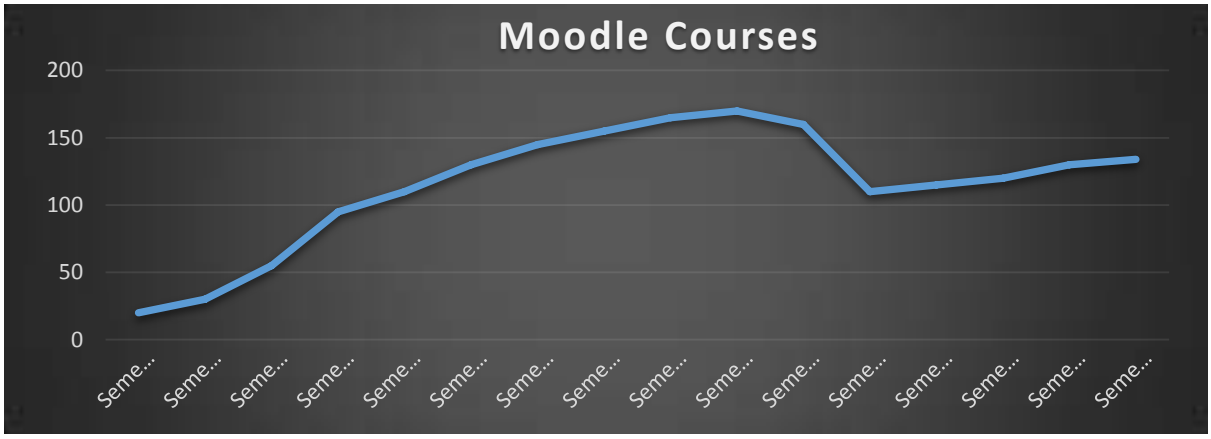


Figure 2: Distribution of courses on Moodle

B. Statement of the Problem

Universities are generally concerned with efforts to improve student and staff performance. Data about students and staff generated through the Learning Management Systems could be very useful in this regard. However, since the introduction of LMS, specifically at UB, there has been limited use of academic data generated from the LMS platform to inform decision making. As such, this study seeks to propose a model for generating and using academic data from Blackboard and Moodle Learning Management Systems to inform decision making with a view to improving learning, teaching and research at the university of Botswana.

C. Study Objective

The main objective of this study is to mine data which has been accumulated through Black Board (BB) learn and Moodle Learning Management Systems(LMSs) over the past 15 years in order to evaluate what has taken place and how the information can be useful for strategic planning at the University of Botswana. The specific objectives of this study are:

1) To investigate students learning activity patterns at the university of Botswana
2) To investigate students engagement patterns in their learning process through their use or none use of the resources provided by the LMS
3) To investigate lecturers adequate use of LMS resources in teaching and learning at the University of Botswana
4) To develop a model of Academic Data federation that could provide useful data for improved management at university level

D. Study Research Questions

This paper poses 3 vital questions namely:

1) Are UB students engaged in their Learning activities on the LMS?
2) Do academic staff actively use LMS resource in teaching and learning?
3) Does our model of academic data federation based on the LMS provide useful data for managerial decision making?

II. LITERATURE REVIEW

Learning management system (LMS) tools such as Moodle and Blackboard are popular tools among universities worldwide [1,2,3, 4, 5, 6,7, 8]. Many universities acquire these systems in the belief that they deliver on practical application to teaching and learning by enhancing students learning outcomes. As a technology, LMS is still problematic due to some perceived resistance of some academics as observed in [6]. The authors suggest that some lecturers believe that LMS technologies still lack satisfactory tools that address the social aspects of teaching and learning, and thus lacks the richness of natural social setting [7]. Nevertheless, LMS tools are very good platforms [8]. The current study applies educational data mining using LMS platforms.

III. METHODOLOGY

The WEKA tool, a collection of machine learning algorithms for data mining tasks is employed in this study [11]. Three course files containing over 20000 data sets instances concerning students teaching and learning collected using Moodle and Blackboard (BB) learn at the University of Botswana were mined

using machine learning schemes selected from the schemes menu on the WEKA tool. The selected schemes include tools for preprocessing, classification, regression, clustering and visualization. The algorithms were applied directly to each data set for each stage of preprocessing, classification, regression, clustering, and visualization invoked from the schemes menu as follows[2]:

1) a data file is selected from the File menu;
2) important attributes of the data are selected;
3) aggregates of existing attributes are created using the spreadsheet;
4) a machine learning scheme is selected from the Schemes menu;
5) results are viewed as trees, text or three-dimensional plots;
6) attribute/aggregate selections are revised;
7) the scheme is re-run on the revised data

In order to maintain format independence, data is converted to an intermediate representation - Attribute Relation File Format (ARFF). ARFF files contain blocks describing relations and their attributes, together with all the instances of the relation. They are stored as plain text for ease of manipulation. Relations are simply a single word or string naming the concept to be learned. Each attribute has a name, a data type (which must be one of enumerated, real or integer) and a value range (enumerations for nominal data, intervals for numeric data). Pre-processing tools in WEKA are called filters. Hence, the preprocess involves loading the data file, preprocessing and analyzing the attributes using the tool. Similar steps were repeated for classification, regression, clustering and visualization by selecting these options from the tools schemes menu.

IV. SIGNIFICANCE OF THE STUDY:

ACADEMIC DATA , CHALLENGES OF ORGANIZATION AND EFFICIENT RETRIEVAL FOR MANAGEMENT DECISION MAKING

Academic Data (Conceptual view).

In the academic environment massive data (big data) heterogeneously exist in various places and forms as depicted in figure 3 below. As a concept, this paper represents academic data as data emanating from the concepts shown in figure 3 below. From figure 3 aside from research and publications data, the rest are readily available through the LMS. However, the data often exist as log files in Comma Separated Value (CSV) file formats in the data bases. As such, the files appear unorganized, unformatted and meaningless as shown in figure 4.

Figure 3: Conceptual understanding of academic data.

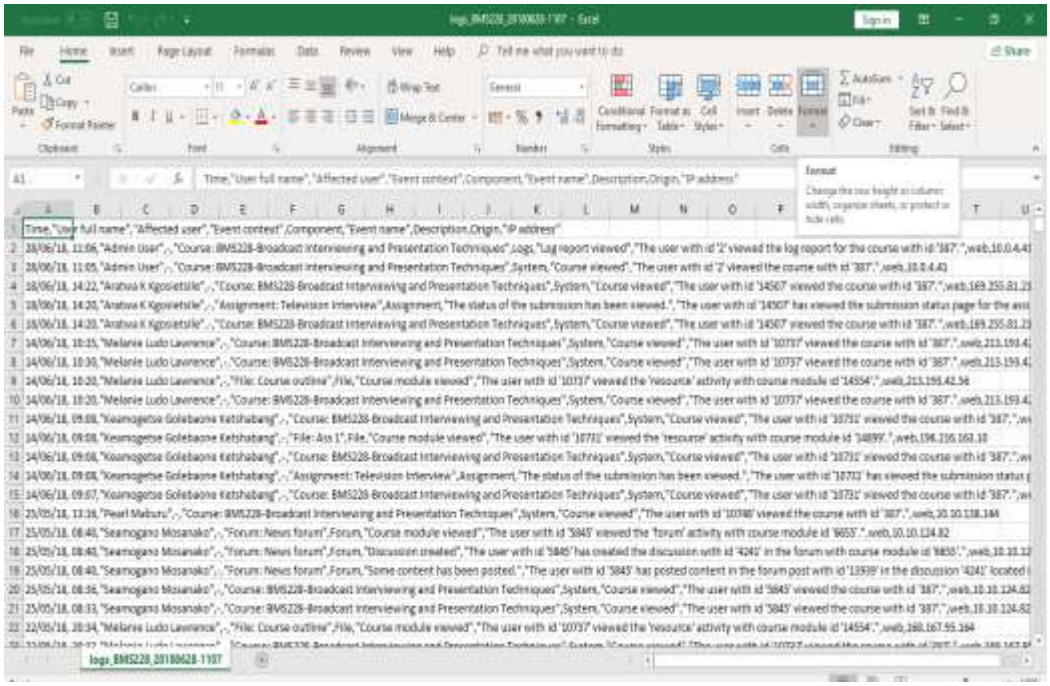


Figure 4. Example of a CSV log file

CSV log files when properly organized and mined provide useful information like other Information Retrieval Systems (IRS). An Information Retrieval System (IRS) represents users with possible information problems and the systems which provide answers to the information needs of users through appropriate retrieval mechanisms. Therefore, using the approach of this study meaningful and useful academic data can be mined and used for strategic planning and forecasting in the academic environment. This is the significance of this study.

A. Relational Data Structures for Academic Data management (Physical view)

For efficient storage and retrieval data is organized in terms of physical storage for easy access (retrieval, modifications, updates). This is often based on the relational Database model and anchored on the principles of Data Normalization resulting in the following Normal Forms (NF): First Normal Form (1NF), Second Normal Form (2NF), Third Normal Form (3NF), Boyce Codd Normal Form (BCNF), Fourth Normal Form (4NF), Fifth Normal Form (5NF) [12]. These are approaches that form the basis of efficient organization for an Information Retrieval System (IRS). The general schema of an IRS consisting of a federated database, users, user interface is shown figure 5 .

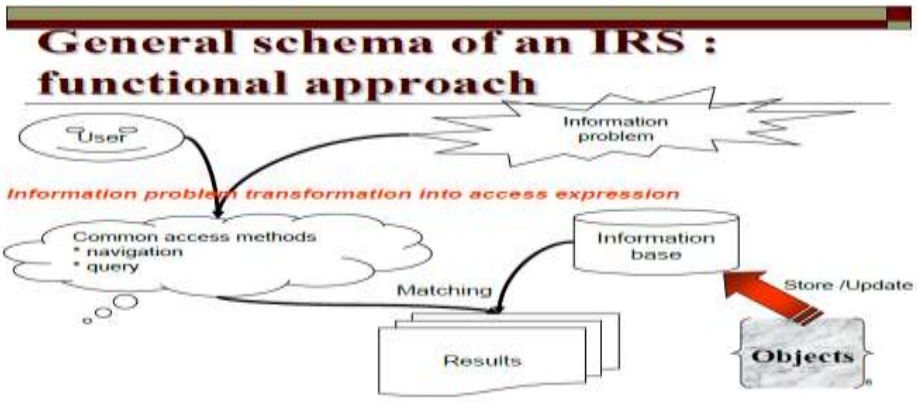


Figure 5: General Schema of an IRS
Source: Amos David [13]

V. LMS USE FOR IMPROVED PERFORMANCE:A PROPOSAL FOR THE UNIVERSITY OF BOTSWANA

The university of Botswana has about 15000 students, and academic staff around of around 1200. There are eight Faculties, fifty six (56) Departments, a School of Graduate Studies and numerous academic programmes. The University uses Blackboard Learn and Moodle as Learning Management systems (LMS) for teaching and learning in addition to the use of Academic and Student Administration System (ASAS) for course registration and keeping of other academic records of students. Data from the LMS and ASAS when mined as discussed in this study could generate useful information for improved performance at the university.

Consider the model presented in figures 6 and 7 below. Figure 6, is a typical LMS platform and its expected benefits which should enhance students performance when properly utilized.

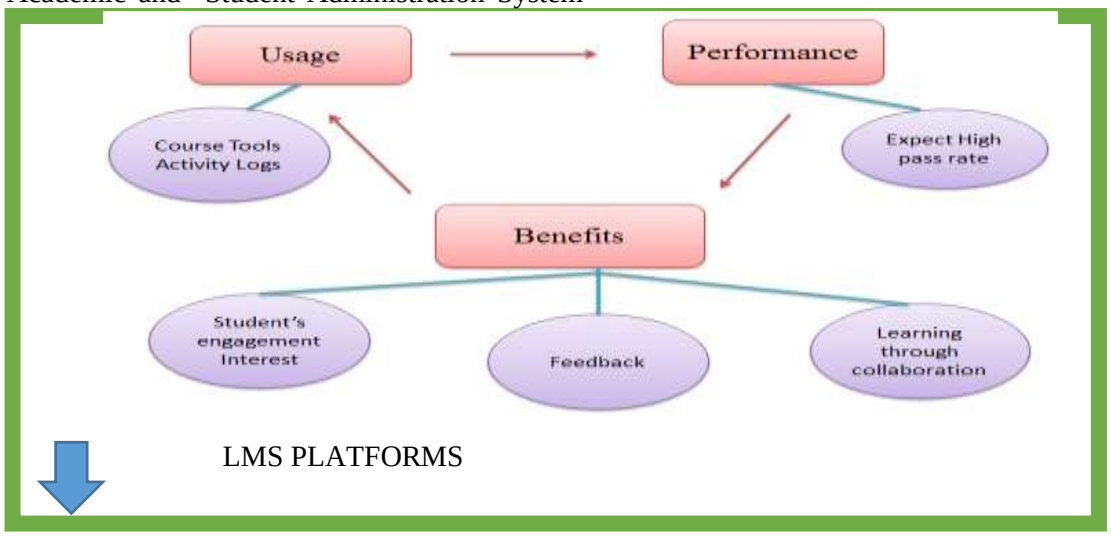


Figure 6: Benefit of LMS

Figure 7 depicts the practical realization steps using the approach of this paper. The appropriateness of the approach is demonstrated with sample test data as shown in section VI.

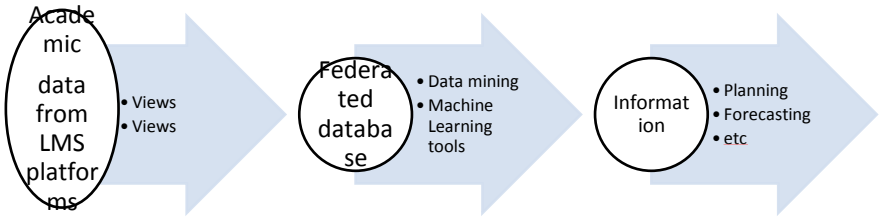


Figure 7. A Model of Academic Data Federation For improved Management at University levels

VI. RESULTS AND DISCUSSION

Results from a pilot study using the Waikato Environment for Knowledge Analysis (WEKA) tool to analyze three course files are shown in Tables 1-5. Table 1 presents the attributes of a course file from the logs. There are 8 attributes displayed on the top horizontal bar and their data type. In addition, detailed description of each attribute is also shown. Any statistical analysis may be obtained from data on the selected attributes depending on the objectives and research questions of a study. In this paper, one attribute of great concern is the "component" (attribute 5). Component in this study includes 4 teaching and learning activities that could indicate the engagement pattern of students and lecturers on the LMS resources. These activities are analyzed for the selected course files as shown in Tables 2-5.

Are UB lecturers and learners engaged on the LMS platforms? (Research Questions 1&2)

To understand the extent to which the lecturers and learners engage with the LMS, the researchers investigated the types of activities lecturers and students engage in on the LMS platforms. The teaching and learning activities indicated are Forum, Chat, Assignment, and Quiz. For the purpose of this study, the weighting of each learning activity is indicated. A comparative analysis of the engagement pattern is shown in Table 5. In all the three course files, Quiz has the highest rating of 8697.00, 8308.00, and 38530.00 respectively for course files 1, 2, 3. This is followed by Forum, Assignment and Chat. This shows that the course lecturer spends more time on Quizzes. In course 3 for instance, the lecturer concentrated efforts on Quizzes and very little time on the other resources including forum, chat and assignment. It can thus be concluded that teaching and learning in this course is not fully engaging, and the lecturer is not utilizing the LMS resources fully. In course 2, the same conclusion can be reached, although the lecturer uses a bit of forum. In course 1, the lecturer also gave more attention to quizzes, forums, assignments, and less of chats.

Table 5: Comparative analysis of teaching and learning engagement

	Forum Weight	Chat Weight	Assignment Weight	Quiz Weight
Course File 1	8509.00	459.00	5227.00	8697.00
Course File 2	4485.00	270.00	309.00	8308.00
Course File 3	8802.00	122.00	1395.00	38530.00

Although the data has been derived from a small sample size for this pilot study, it can be concluded that while courses might be available on LMS platforms, the lecturers tend to use LMS platforms for assessment purposes mostly. The reason for this is not clear from the data available in this study, but it could be that lecturers use quizzes more than other components because of the efficacy of the platforms, especially Moodle in automatically grading the students' work. However, the lack of use of other components (whereas there are more tools/resources/activities on the LMS platforms: blocks, databases, glossary, scrum,

folders, etc.) suggests a missed opportunity by lecturers to enhance teaching and learning. This finding on the limited use of LMS by lecturers is consistent with findings from other studies such as [14] and [15], who observed that the University of Botswana lecturers were rather reluctant to adopt e-leaning platforms.

Furthermore, there is need to perform statistical analysis on the data using selected machine learning algorithm could reveal more patterns, and even enable the prediction of performance for a student in a course. It is possible to predict pass rate in a course by giving the components of assignment and quizzes and applying for example J48 algorithm decision tree or C4.5 algorithm. These tools and other statistical measures are available in WEKA. Therefore the model of academic data federation based on LMS as shown in figures 6 and 7 can provide useful data which could support planning and forecasting decisions in the academic environment.

Table 1: Course File_Attributes as shown on top horizontal bar

No.	1: Time	2: User full name	3: Affected user	4: Event context	5: Component	6: Event name	7: Origin	8: IP address
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	16 M...	Admin User	-	Course: BM...	Logs	Log report v...	web	10.210.1.1...
2	16 M...	Admin User	-	Course: BM...	Logs	Log report v...	web	10.210.1.1...
3	16 M...	Admin User	-	Course: BM...	Logs	Log report v...	web	10.210.1.1...
4	16 M...	Admin User	-	Course: BM...	Logs	Log report v...	web	10.210.1.1...
5	16 M...	Admin User	-	Course: BM...	System	Course mo...	web	10.210.1.1...
6	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...
7	16 M...	Admin User	-	Course: BM...	Quiz	Course mo...	web	10.210.1.1...
8	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...
9	16 M...	Admin User	-	Course: BM...	System	Course mo...	web	10.210.1.1...
10	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...
11	16 M...	Admin User	-	Course: BM...	Quiz	Course mo...	web	10.210.1.1...
12	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...
13	16 M...	Admin User	-	Course: BM...	Forum	Course mo...	web	10.210.1.1...
14	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...
15	16 M...	Admin User	-	Course: BM...	Assignment	Course mo...	web	10.210.1.1...
16	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...
17	16 M...	Admin User	-	Course: BM...	Forum	Course mo...	web	10.210.1.1...
18	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...
19	16 M...	Admin User	-	Course: BM...	Assignment	Course mo...	web	10.210.1.1...
20	16 M...	Admin User	-	Course: BM...	System	Course vie...	web	10.210.1.1...

Table 2: Course_file_1 Result on Components

Table 3: Course_file_2 Result on Components

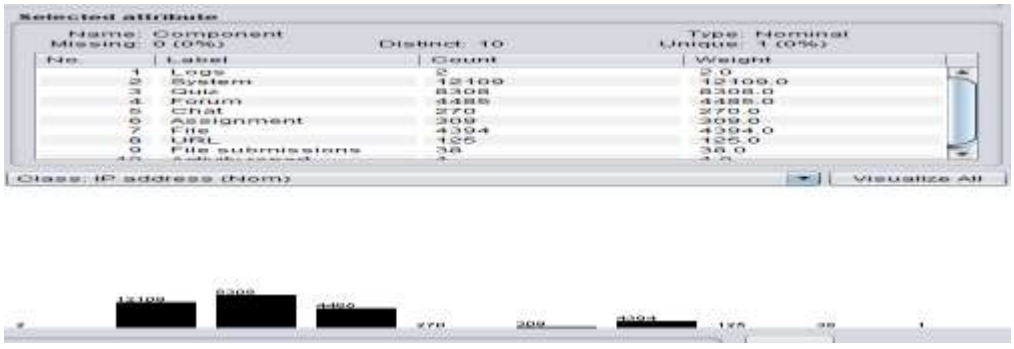
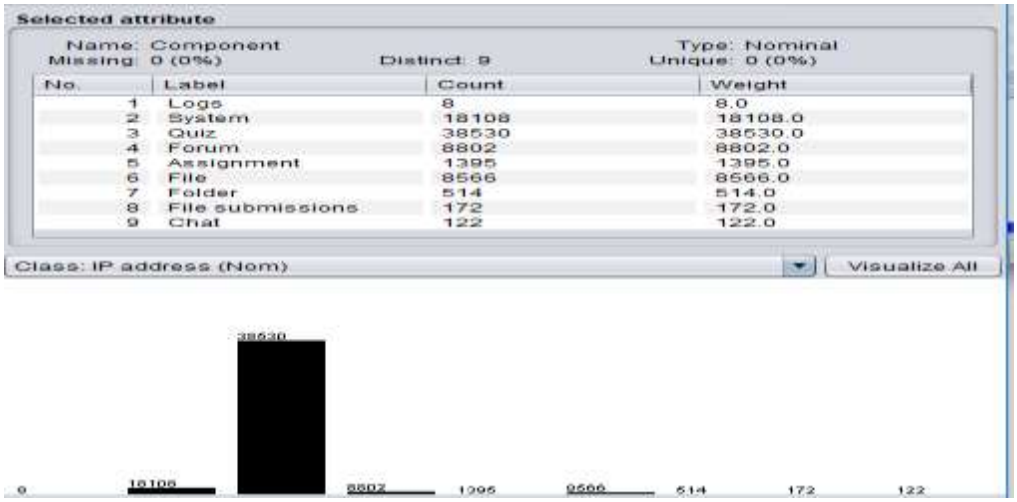


Table 4: Course_file_3 Result on Components



Does our model of academic data federation based on the LMS provide useful data for managerial decision making? (Research Question 3).

In general therefore, it seems that our model is able to help in generating data that is useful for decision making at tertiary institutions. The LMS platform provide raw data that is almost meaningless in the csv format as demonstrated in Fig 4. By mining the data using our model, it is possible to generate useful data. The available data generally relates to the time a user (student, lecturer or system administrator) logs onto the system, the activities they engage on, the components of the system they use (quiz, forum, charts, assignments etc) and the IP address of the computer used as reflected in Table 1. The findings, while preliminary, suggest that with appropriate data federation models such as ours, Universities can mine useful data to enhance teaching and learning through LMS platforms. This is even significant for the University of Botswana, since the LMS platforms have been used for the past ten years, and have generated huge data that is still to be systematically organized and interpreted. Another important implication for this finding is that other African universities intending to implement blended learning could consider the potential usefulness of the data generated by the LMS platforms.

VII. CONCLUSION

LMS platforms such as Blackboard and Moodle are very useful course management systems. They enable design, dissemination, and use of teaching and learning resources including notes, assessments, as well as enhancing collaboration among the learners and lecturers. In general, academic data generated from the LMS platform could be used to help the University of Botswana (UB) in planning and execution of its teaching and learning strategies with the view to improve decision making related to improved learning and teaching at UB.

Data mined from LMS could provide useful data for planning and forecasting in the educational environment. From the analysis on the uptake of these platforms, this study provides information suggesting that there are some lecturers who are resisting to adopt these technologies. The UB has more than 1200 courses per semester, and figures 1 & 2 (section I A) above reflect that just over half of the courses are available on the platform. This information is important in the decision making process at UB, as one of the strategic goals has been to enhance teaching and learning through blended learning that involves the use of various methods of teaching and delivery.

This study has shown that there is data in the LMS platforms. The UB expects that with blended learning, students academic performance will also be improved. However, this calls for a rrigorous testing of the correlation between the use of LMS by both the students and the lecturers and its possible impact on student performance in general. Such data, could also guide management in key decision making such as making the use of LMS platforms

mandatory, or even abolishing them if they seem to have no impact in teaching and learning.

Moever, further studies could be conducted to establish the low up take of LMS courses at UB, and the possible challenges on the lack of use of other components in the LMS platforms. Such data could be useful in guiding managerial decision making relating to possible training and contextual factors such as ensuring lecturers have provisions for time on LMS platforms.

References

[1] Al-Ajlan S. Ajlan (2012). A Comparative study between E-learning Features. Intechopoen. Pp. 192-214. Available at http://www. Intechopen.com

[2] Geoffrey, H., Andrew, D., and Jan H, W (1994). WEKA: A machine learning workbench. IEEE. Pp 357-361. Available at http://www.cs.waikato.ac.nz/~ml/publications/1, retrieved July, 2018.

[3] Kiget, Nicholas; Wayambi, G; Petrs, Ikoha A (2014). Evaluating Usability of E-learning Systems in Universities, *International Journal of Advanced computer Science and Applications* 5(8), pp. 97-102

[3] Massingham, P. (2014). An Evaluation of Knowledge Management tools: Part 1-Manageing Knowledge resources. *Journal of Knowledge Management*, 18(6), pp.1075-1100.

[4] Ozkan, S & Koseler, R (2009). Multiple-dimensional students' evaluation of e-learning systems in higher education context: An empirical investigation. *Computers & Education*, 53(4), pp. 1285-1296.

[5] Paechter, M., Maier, B & Macher, D (2010). Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction. *Computers & Education*, 54(1), pp. 222-229.

[6] Steel, C.H & Steel, C. (2009). Creativity and constraint: Understanding teacher beliefs and the use of LMS technologies. *Proceedings of ascilite Auckland*, pp. 1013-022

[7] Unal, Zafer, and Unal Asli (2011). Evaluating and Comparing the Usability of web-based course management systems. *Journal of Information Technology in education*, Vol. 10, pp. 19-38

[8] Okike, Ezekiel U. and Mogorosi, Merapelo, M (2017). Measuring The Usability Probability of Learning Management Software Using Logistic Regression Model. *Proceeding of IEEE Computing Conference 2017, London, UK*; pp. 1217-1223. ISBN (IEEE XPLORE) 987-1-5090-5443-5; ISBN (USB): 987-1-5090-5442-8

[9] Sheth and Larson (1990). Federated Database Systems for Manageing Distributed Heterogeneous and Autonomous Databases. *ACM Computing Surveys*, Vol. 22 No. 3, pp. 183-236

[10] Masood, Nayyer; Eagle, Barry (2003). Components and Federation Concept Models in a Federated Database System. Malaysian Journal of Computer Science, Vol. 16 No, 2, pp. 47-57.

[11] Witten, H. Ian and Frank, Eibe (2005). Data Mining: Practical Machine Learning Tools and Techniques. Elsevier, 2nd ed,

[12] Avison, D and Fitzgerald G (2006). Information systems development, methodologies, techniques and tools. McGraw Hill Education:Berkshire

[13] David A. (2008). Data Modelling - General schema of an IRS: Users and usage centered approach . Part of slides prepared by Amos David. University of Lorraine France. Unpublished.

[14.] N. Obasi, Isaac & Motshegwa, Baakile. (2005). The University of Botswana as a learning organization: The challenges of fostering change. Higher education in a challenging world, pp. 344-350.

[15] Moakofhi Moakofhi, Oratile Leteane, Tawona Phiri, Thato Pholele, and Perncy Sebalatlheng (2017). Challenges of introducing e-learning at Botswana University of Agriculture and Natural Resources: Lecturers' perspective, International Journal of Education and Development using Information and Communication Technology (IJEDICT), 2017, Vol. 13, Issue 2, pp. 4-20

[16] Wikipedia (2018). Federated database system. Retrieved October, 2018.