Unsupervised Machine Learning Method for the Analysing of Students' Activities in E-Learning

*Alhaji Audu Goni and Abubakar Mohammed
Department of Computer Science, Mai Idris Alooma Polytechnic Geidam, Yobe State.
***Corresponding author: Aagoni2016@gmai.com +2348039519380**

Abstract:-

The advancement of machine learning algorithms has made it possible for applying some of the unsupervised algorithms in clustering the activities of online students to analyse their learning behaviours as well as the implication of their learning behaviour to their final results. In this paper, students' e-learning activities were automatically grouped into three (3) clusters using K-means clustering algorithms. Students actions were obtained from log files of their activities in Moodle Learning Management System (LMS). The aim is to group the students based on their similarities in actions on the LMS. Learning behaviors of students in each cluster were analysed and a correlation between each learning behaviour on student success or failure in their academic performance (Final Results) was investigated. The three clusters were labelled as ClusterA, Cluster B and ClusterC. The analysis shows that students in ClusterA have outperformed their counterpart parts with an average score of 91.12% in their final results, this group have higher login in the selected activities of the LMS, ClusterB with less number of interactions than ClusterA have average score of 75.65%. Finally, students in ClusterC have the least number of interactions with the LMS have failed the course with an average socre of 36.57%. The research shows that, students who post, read and respond in Forum activities perform higher than those who do not. The work discovered that Forum activity has significant factor on student's course success, however, this activity has lesss weight compare to other activities such as Assignment and CourseView. The research suggests that weight should be allocated to Forum activities to encourage students' participations.

Keywords— ***E-Learning, Educational Data Mining, K-Means, Clustering,***

A. INTRODUCTION

Electronic mode of performing business and other transactions has dominated every face of human endeavour. Education sector is not an exception. Most of the developed and of course the developing countries have adopted the use of ELearning (also written as E-Learning) as means of knowledge sharing (Jamil, 2017). The Learning Management system is the platform that supports the delivery of learning materials and also provides everything needed for knowledge delivery, these tasks that are provided by the LMS are called the E-Learning activities. E-Learning has not replaced Face-to-Face method of learning but it complements it. There are various learning management systems in use, notable ones are the Modular Object Oriented Dynamic Learning Environment (Moodle), the A-Tutor, the BlackBoard, and so on. Among these LMSs, Moodle is the most commonly used LMS especially in developing countries. Moodle LMS provides access to the courses and facilitates communications between students and tutors or among the students.

Log files are used to store students' activities in the LMS, these activities can be mined using Data Mining methods to study the learning behaviour of the students. Educational data mining is the application of data mining processes that is concerned with development of methods for discovering the uniqueness of data that come from educational database, and use the methods to for understanding learners and their learning skills (Bara, 2018).

This research is concerned with students' involvement in eLearning. The researchers studied students' learning behaviors and analyses the effects of their learning performance upon carrying out the e-learning activities. It also discovers which activity has more significance over the others. The research uses data obtained from students' log files in Moodle LMS, the students are from Computer Science undergraduate taking a Data Structure Course at Faculty of Computing, Universiti Teknologi Malaysia. The research is aimed to investigate the performance of K-Means clustering technique to obtain clusters of students who have similar way of interacting with the e-learning environment, and analyze the effects of their learning behaviours on their learning performance using their final semester result.

Advancement in machine learning has encouraged researcher the effectiveness of web-based education (e learning) system (Bara, 2018). The e-learning system allows students from anywhere and at any time to carryout different learning activities such as viewing the learning materials, carrying out assignments, Forums participations and so on (Hansen et al., n.d.). These activities usually take palace via Learning Management System (LMS), which provides access to the courses and facilitates communications between students and tutors or among the students. All students' actions are stored in log files via the LMS. The students' actions are characterized by meaningful learning which show the students' learning behaviour.

Clustering is considered to be one of the useful techniques in data science. It is a great technique to find cluster structure in a data set that is characterized by the greatest similarity within the same cluster and the greatest dissimilarity between different clusters (Sinaga & Yang, 2020).

B. LITERATURE REVIEW

Learning Management System (LMS) is the platform that supports the management of Elearning activities. Among the functions of the LMS are the provision of learning materials to students, giving the students the ability to access the learning materials, interact with teachers and students, submit assignment and so on (Broadbent, 2016). Fig.1 shows how LMS works.

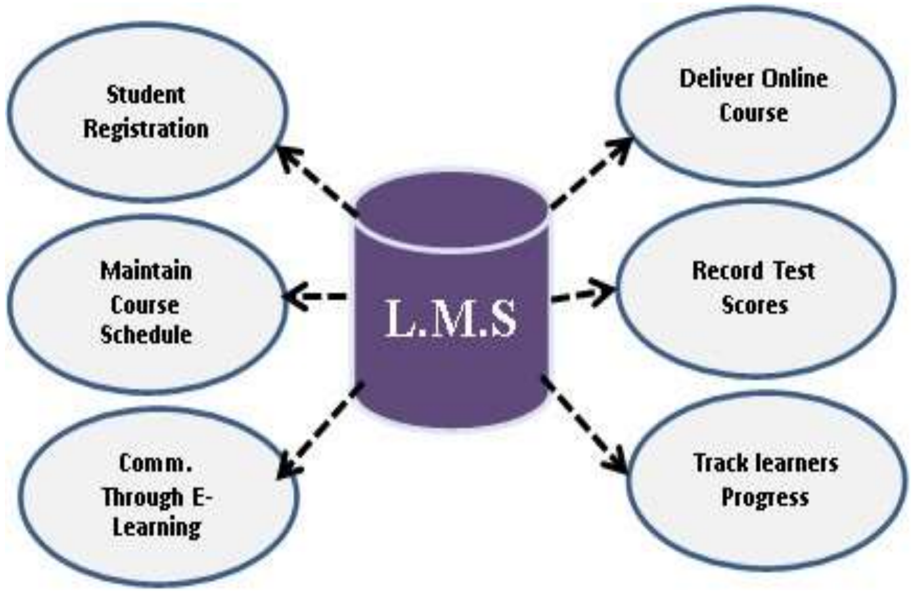

Figure 1Some Benefits of Learning Management System (Bara, 2018)

e-Learning Management System (eLMS) is also called Course Management Systems (CMS), and the two terms may be used interchangeably. Many universities have already adopted CMS and some others are planning on introducing CMS software such as Moodle or Blackboard to support their learning operations. However, currently, there has not been much research to explore the influence of Power (electricity) on the use of CMS. To this end, it is vital to learn its perceived usefulness from the perspectives of students. The results of this study could help universities and by extension, higher institutions make better investment decisions and help instructors in using this technology more effectively. Additionally, it can help designers of course management software to improve the learning tools and get higher satisfaction level in the learning environment (Nicholas-omoregbe, 2017).

Educational Data Mining (EDM) is an emerging discipline, concerned with development of methods for exploring the uniqueness of data that come from educational database, and use those methods to better understand learners and their learning settings (Ratnapala et al., n.d.). These methods include classification, clustering, prediction and visualization. While the tools used in EDM include Waikato Environment for Knowledge Analysis (WEKA), Rapid Miner, MATLAB, etc.

Educational data mining fits various research works in eLearning such identifying students' learning styles**,** prediction of students' performance as in (Tick, 2018) and determining the situation, characteristics and phenomena that affect students' performance. Reference (Hansen et al., n.d.) used educational data mining technique to improve the comprehensibility of the learning materials, students were clustered based on their behavioural usage of the e-learning, however, the researchers did not state which activities is more significant than the other.

Fig. 2 depicts the steps of mining student activities in EDM.

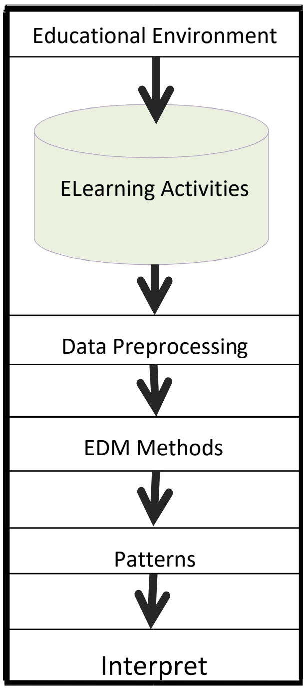

Figure 2Educational Data Mining Steps

C. K-MEANS CLUSTERING TECHNIQUES

In this steps k-means clustering algorithm was applied to the proposed data and get valuable information, k- means is an old and most widely used clustering algorithm by MacQueen in 1967 (Bhise et al., 2013)

Algorithm of Basic K-means Algorithm:

1. Select K points as the initial centroids.

2. Repeat.

3. From K- cluster by assigning all points to the closest centroids.

4. Recomputed the centroid of each cluster.

5. Until The centroids don't change.

D. METHODOLOGY

This section describes the methods and procedures followed to achieve the objectives of the study. Online students' datasets were first collected from student log records on Moodle learning management system.

Data Preprocessing

The raw data collected from the Moodle LMS need to be preprocessed to be produce an accurate result during the analysis. Unwanted feature of the data was removed. The major tasks involved in data preprocessing include data cleaning, data integration, data transformation data reduction and data discretization. Figure 3 shows example of raw data collected from Moodle LMS.

3	Course	Time	IP addres	User ID	Actio	Information
4	SCSJ201:	####	10.60.84.:	stud1	assig	Submission statement accepted by user stud1
5	SCSJ201:	####	10.60.87.:	stud1	assig	Submission status: Submitted for grading. The n
6	SCSJ201:	####	10.60.84.:	stud1	assig	Submission status: Submitted for grading. The n
7	SCSJ201:	####	10.60.84.:	stud1	assig	Submission status: Submitted for grading. The n
8	SCSJ201:	####	10.60.100	stud1	assig	Submission status: Submitted for grading. The n
9	SCSJ201:	####	10.60.84.:	stud1	assig	Submission status: Submitted for grading. The n
10	SCSJ201:	####	10.60.84.:	stud1	assig	Submission status: Draft (not submitted). The nu
11	SCSJ201:	####	10.60.86.:	stud1	assig	Submission status: Submitted for grading. The n
12	SCSJ201:	####	10.60.98.:	stud1	assig	Submission status: Submitted for grading. The n
13	SCSJ201:	####	10.60.98.:	stud1	assig	Submission status: Submitted for grading. The n
14	SCSJ201:	####	10.60.85.6	stud1	assig	Submission status: Submitted for grading. The n
15	SCSJ201:	####	10.60.84.:	stud1	assig	Submission status: Submitted for grading. The n

Figure 3 Moodle Sample Raw Data

The overall process of data preparation is shown in Fig.4 (Kularbphettong & Tongsiri, 2012).

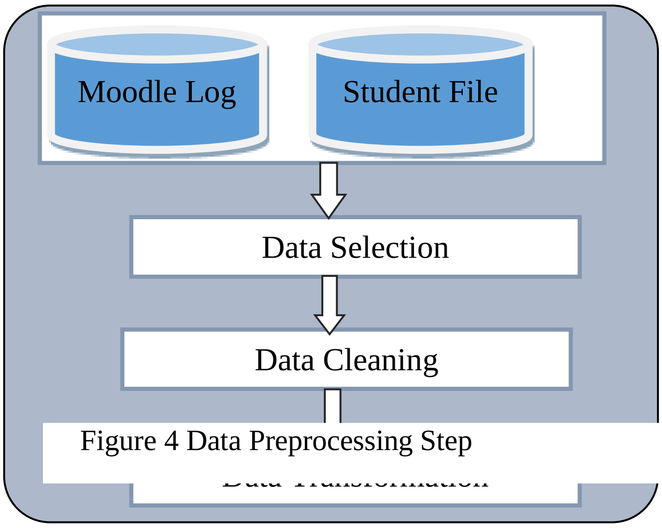

Figure 4 Data Preprocessing Step

Data Cleaning the process of removing unwanted data; these include noisy data, missing values and identifying outliers. After cleaning the data, the input variables were well defined as shown in Fig. 5. The idea of data cleaning

StudentId	IP Adress	Course_View	Resource_View	Assign_View	Assign_Submit	Forum_View	Forum_Discuss	total
1	10.60.86.118	1	2	0	1	0	0	4
	10.60.85.62	5	9	4	1	0	0	19
	10.60.106.208	1	0	0	0	0	0	1
	10.60.86.90	9	13	0	0	0	0	22
	10.60.98.31	7	14	10	2	0	0	33
	10.60.86.93	1	1	0	0	0	0	2
	10.60.86.182	1	1	0	0	0	0	2
	10.60.86.115	4	5	4	1	1	1	16
	161.139.102.98	1	0	0	0	0	0	1
	10.60.84.132	1	1	5	3	2	1	13
	10.60.84.240	4	1	3	1	0	0	9
	161.139.101.20	1	1	0	0	0	0	2
	10.60.85.229	2	1	0	0	0	0	0
	10.60.85.219	1	4	0	0	0	0	5
	10.60.86.165	4	48	0	0	0	0	52
	10.60.86.95	5	10	0	0	0	0	15

Figure 5 Sample of Cleaned Data

is to improve the quality of the data; names of students were removed and their Matric Numbers were encrypted for privacy purposes.

Data Normalization

The clean data was normalized such that every entry will have equal recognition by the data mining technique and to make the variables have equal priority and chance during the experiment (Romero et al., 2008). The Min-Max normalization was used in this project.

	A	B	C	D	E	F
1	CoursV	ResosV	AssignV	AssignS	ForumV	ForumD
2	0.11	0.66	0.10	0.36	0.11	0.06
3	0	0.00	0.01	0.00	0.00	0.00
4	0.19	0.69	0.21	0.71	0.11	0.03
5	0.80	0.84	0.68	0.50	0.00	0.00
6	0.05	0.30	0.00	0.07	0.00	0.00
7	0.47	0.48	0.20	0.21	0.11	0.13
8	0.27	0.29	0.11	0.29	0.04	0.00
9	0.98	0.86	0.44	0.79	0.11	0.09
10	0.34	0.28	0.16	0.14	0.07	0.06
11	0.32	0.62	0.48	0.64	0.22	0.47
12	0.57	0.40	0.48	0.50	0.04	0.13
13	0.24	0.19	0.26	0.86	0.00	0.00
14	0.23	0.31	0.17	0.43	0.22	0.38
15	0.35	0.19	0.24	0.57	0.04	0.09
16	0.47	0.31	0.52	0.43	0.07	0.00

Figure 6 Sample of Normalized Data

S/No	Activity	Student Actions	Description	Weights
1	Forum	View, Add discussion	The student participates in forum discussion, the student can view forum and/or add discussion in the forum	5
2	Assignment	View, Submit View submitted assignment	This is a task that is given by the lecturer; the students would download the question(s), do the assignment and upload it for online submission	3
3	Resource	View	These are files or links that support learning, lecturers add resource to their course.	2
4	Course	View	This is where the lecturers add learning materials and organize them to suit the learning interesting.	3

E. EXPERIMENT AND RESULT

a. The Dataset

The dataset was collected from the Universiti Sains Malaysia (USM) Moodle LMS log records. It contains the activities of Computer Science undergraduate students from Faculty of Computing, Universiti Sains Malaysia during the period of first semester 2020/2021 session. The activities of 104 regular students were monitored from the beginning of the semester to the end. The semester lasted for fourteen weeks. The data was downloaded in excel format. The interface is the medium that is used to carry out the data collection from the log records.

b. Training the K-Means

The network was then trained using the data at hand and the result obtained is shown in Fig. 7. The result shows three clusters; each cluster contains instances associated to it.

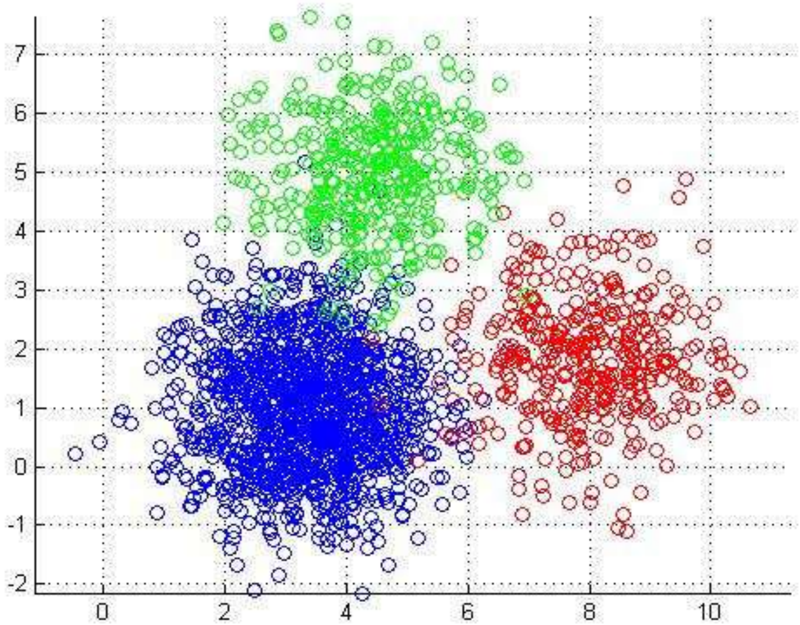

Figure 7 Three (Clusters) of Students According to their Similarity in Learning Behaviour

c. Experimental Result

From the experimental result, the students are grouped into three clusters. Each cluster contains students with similarity in interaction with the e-learning environment.

ClusterA contains students who have the highest number of actions thus; they are termed as Very Active Students. Students in ClusterB have more actions than those in ClusterC and they are termed as Active Students. ClusterC students have lowest actions hits and are termed Non-Active, this categorization is based on students' interactions and in line with the work of (Yin et al., 2015) categorization. On the other hand and based on students' course success, Clusters are categorized as High Learners (Cluater1) with the highest marks 91.12% in their Final Scores, Medium Learners (ClusterB) with Final Score marks 75.65% and Low Learners (ClusterB) with the lowest Final Score marks of 36.57. This categorization is in line with (Ratnapala et al., 2014). Fig. 8 shows the graphical representation of Actions hits means and the Final Score Means of each Cluster.

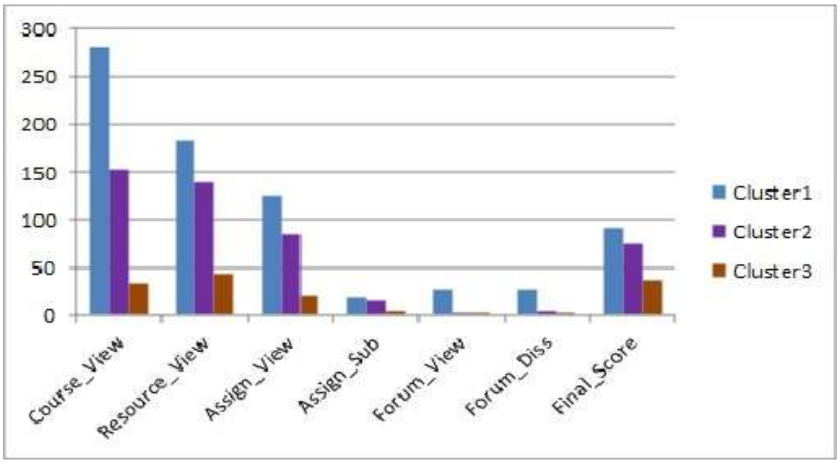

Figure 8 Showing the Activities of Each Cluster and Their Final Marks

F. RESULT FINDINGS

From the experiments carried out, it was learnt that, students' actions on e-learning activities differ according to their learning behaviours. As a result, the students are grouped into clusters such that each cluster contains students with similarities in learning behavior. The research discovered three categories of students, those that are Very Active (ClusterA), students of this group have higher number of interactions with the E-learning especially participation in Forum activities and they have the highest grades. Active (ClusterB), students of this group have average number of interactions with the ELearning and have average final grades, those that are Non-Active (ClusterC) students of this group have least number of interactions with the E-learning and they all failed the course.

It can therefore be concluded that, students' learning behaviour contributes to their learning performance. It was also learnt that, the E-learning activities differ in contributing to student's learning performance. Students who participate more in Forum (ClusterA) have high chance to outperform the student who do not participate more in Forum. Likewise, students who failed to submit their assignments regularly have chance to fail the course.

It was learnt that participation in Forum activities by students is optional and there is no mark allocated to it, this is why students feel reluctant to participate. However, if marks can be allocated to Forum participation, students can give it a priority and this would improve their learning performance.

G. CONCLUSION

In this research, the implementation of K-Means Clustering Method to group students according to their similarities in interaction with the E-learning was explained. Three clusters were obtained and are termed based on students' learning success as High Learners (ClusterA) with Final Score of 91.12%, Medium Learners (ClusterB) with Final Score of 75.65% and Low Learners (ClusterC) with Final Score of 36.57% in line with (Ratnapala et al., n.d.), as Very Active Students (ClusterA), Active Students (ClusterB) and Non-Active Students (ClusterC) in line with (Bara, 2018) categorization. The researchers also conducted analysis on the correlation between students' learning behaviors and their course success (performance). The result showed that, students that fall in both 'High Learners' and 'Very Active' group

(ClusterA) emerged to be best in their course success. The Final grade is A+ or an excellent pass for this group. ClusterB contains students that are Active and High Learners, and passed with an A grade. Finally, ClusterC students emerged to be both Low Active and Low Learners with an F grade which means they failed the course.

H. REFERENCES

Bara, M. W. (2018). *Self-Organizing Map Clustering Method for the Analysis of E-Learning Activities*.

Bhise, R. B., Thorat, S. S., & Supekar, A. K. (2013). *" Importance of Data Mining in Higher Education System ." 6*(6), 18–21.

Broadbent, J. (2016). *Academic success is about self-efficacy rather than frequency of use of the learning management system*. *32*(4), 38–49.

Hansen, C., Hansen, C., Hjuler, N., Alstrup, S., & Lioma, C. (n.d.). *Sequence Modelling For Analysing Student Interaction with Educational Systems*.

Jamil, L. S. (2017). *ASSESSING THE BEHAVIOURAL INTENTION OF STUDENTS TOWARDS LEARNING MANAGEMENT SYSTEM , THROUGH TECHNOLOGY ACCEPTANCE MODEL - CASE OF IRAQI UNIVERSITIES*. *95*(16), 3825–3840.

Kularbphettong, K., & Tongsiri, C. (2012). Mining Educational Data to Analyze the Student Motivation Behavior. *World Academy of Science, Engineering and Technology*, *6*(8), 1036–1040.

Nicholas-omoregbe, O. S. (2017). *E-LEARNING MANAGEMENT SYSTEM : A CASE OF SELECTED PRIVATE UNIVERSITIES IN NIGERIA*. *April*, 106–121.

Ratnapala, I. P., Ragel, R. G., Deegalla, S., Centre, C., & Lanka, S. (n.d.). *Students Behavioural Analysis in an Online Learning Environment Using Data Mining*.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, *51*(1), 368–384. https://doi.org/10.1016/j.compedu.2007.05.016

Sinaga, K. P., & Yang, M. (2020). *Unsupervised K-Means Clustering Algorithm*. *8*. https://doi.org/10.1109/ACCESS.2020.2988796

Tick, A. (2018). *Research on the Digital Learning and E-learning Behaviour and Habits of the Early Z Generation*. 33–38.