International Journal of Scientific Research in Multidisciplinary Studies
Vol.7, Issue.5, pp.59-66, May (2021)

Research Paper
E-ISSN: 2454-9312
P-ISSN: 2454-6143

E-Crypto Learning and Applying Data Mining in Education Sector to Improve Learner's Performance

E.C. Mwakamo[1*], R. Kalimuthu[2]

[1]UG Student, School of Computer Science and Information Technology, DMI St. John the Baptist University, Malawi
[2]Lecturer cum Head, School of Computer Science and Information Technology, DMI St. John the Baptist University, Malawi

[*]*Corresponding Author:chriseddie98@gmail.com, Tel.: +265 996230640*

Available online at: www.isroset.org

Received 10/May/2021, Accepted: 20/May/2021, Online: 31/May/2021

Abstract— Education is being positively and innovatively enhanced through technology to create new strategies of delivering academic information both online and face to face. Online, Distance and Blended learning are the modern strategies that are and going to impact teaching approaches in academic institutions, relatively through introducing e-learning platforms which are designed as models used to collect and analyze meaningful information online for assessing, analyzing and presenting data patterns taking place among learners. This study proposes an e-learning requirement that focuses on mining student educational data with aim of exploring the behavior of learning, interaction with the platform from data generated by students in an education environment. The data is collected from the event logs or scrapped csv files of student's information for analysis to determine a model to extract their behavior and predictive patterns in an e-learning environment. The techniques applied in this study are decision tree decision making process of filtering content of data, cluster analysis performing the organization of pattern grouping in collections, and random forest for creating prediction model that closely provides accuracy of student's success. Elliptic curve and SHA-256 cryptography method is applied to ensure security per the concerns of user confidentiality both for the instructors and student when performing authentication entries into the platform and during chat interactions. Mining of student's data would help in projecting student's performance and interaction activities, where the random forest technique surpasses logistic regression and support vector machine classification methods of machine learning predictions. With this approach learning institutions will be able to classify groups of learners with aim of analyzing and directing well informed decisions to help them improve in their learning sessions.

Keywords- Educational Data Mining (EDM), Classification, Learning management system (LMS), Online learning, Learner's performance

I. INTRODUCTION

Education is the path that guides learners to create, brainstorm and showcase their various skills from a learning environment up to industry level. It provides principle unique in specific as well as general areas concerning social welfare of the students, technology, culture and industrial tutoring through internships. However, a situation like pandemics brings the question among educational sector authorities on how to devise an approach fair to specific learning environments to help learners in delivering at a limited distance and allow flexibility continuing their journey of having careers [1]. Nowadays courses offered by institutions can be centralized on leaners as a learning approach among the online distance learning, live learning and traditional face to face into a blended environment allowing students to attend anywhere and anytime.

Open Distance e-learning comes with its benefits as it improves the learning experience through provision classroom task in flipped form, where content is posted in advance before class commencement [2]. The crucial factor is to show concern on the privacy equivalence to time that trust must assured in the surrounding of learning associations. This is next option to adopt in the strategies accredited for learning in the future by introducing e-learning through all the education premises in the country.
Articles and journals reviewed in the past year imply that, open and distance learning is the alternative and current option for individuals to support them if they are interested in learning outside the physical premises. This gives the edge to the students in enrolling for the desired course by registering with the institution online and thereby participating in the course according to their convenient time as well as assigned time in line with which notifications provide such details.

This study is subject to create a learning environment suited with a few integrations that provide analysis on the educational structure of teaching and students learning, identifying and finding possible solutions in the academia.

It takes advantage of privacy and data integrity which are prior to ensure trusting in the institution system environment both for the infrastructure and the candidate are safe [3].

Data mining is a subfield of machine learning which has been put to use in healthcare, engineering, cybersecurity, advertising and more generally intelligent information systems. Its core carries out analytic, inducive and observative approach concerned with exploring and discovering knowledge by identifying patterns in large datasets. Mining student education data has emerged uniquely as research area with focus on techniques to answer question like "Define the challenges being encountered when students are learning, how do discussions speed up problem solving etc. We devised a data driven system that analyzes data from educational settings to enhance the process of learning and teaching.
The mining techniques of data succinctly used in e-learning infrastructure circle around learning paths or recommendations, resource availability, performance tasks for learner's improvement and an adaptation to learning experience.

Generally, e-learning systems describe the ability of learning through electronic means or simply the internet and are created in various models depending an institution's overview and selection of technical requirements. It comprises of several important parts namely:

- Learning management system, the major part of the system that is headed by the administrator who authorizes permits users, assigns permissions as an instructor/staff or a studentt.
- Content management system that holds all the information about files being shared such as presentations, course overviewing etc.
- Database storage where information is persisted for use during learning sessions and also extended to allow querying and manipulations

E-learning transcends in the facilitation of wide spread tasks such communication, learning under guidance as well as studying of oneself. These tasks also require protection of the information to ease the issues of man in the middle, password hijacking, inducing malicious codes among other over unauthorized threat activities over the network to prevent damage, exposure and data loss [4]. Practicing authentication methods for users upon entry is an important security method to determine confidentiality, authorization, identification and possession in terms use of the online platform.

II. RELATED WORK

E-Learning Environments

Learning environments has had a turn -around through the establishment of online course e-learning levels where various methodologies are used to complement the shortcomings of learning face to face [5]. These environments include the basic level environment in which the learning process is primary, regular with minimal intervention of technology which simply maintains the class presence with fairly high attendance, blended level which accommodates the teaching and learning combinations involved in a classroom with medium intervention of technology which involves online as well as traditional resources provided, mostly online courses and optional physical interactions.

Furthermore, these levels are modes that are required as a recommendation of the staff academic management in the university who need to be aware of the sufficiency of the structures, strategies and the support each level of development to improve the learning environment and quality education. This study takes views on the live interactions, online courses and other online activities in the blended environment.

Learning management system (LMS)

Digitalization has plunged its way into the education sector with the objectives of reducing costs of human welfare such as transportation problems, improved enrollment time by acquiring learners from various backgrounds to allow them to explore the environment of the platform and its course offers, and has also allowed to approach of blending the learning process [6]. Flipped classroom environment is an approach provided by blended learning strategies which involves content being provided to participants in advance through the internet before the commencement of the class session.

Instructors may also use other mechanisms of online information delivery such as simulations, webinars, e-book links and virtual classrooms. The learning environment is modeled with requirement of collecting information from instructors as well as learners. It is modelled in several ways that includes content management system, learning management structure and visual presentations according to the functionality provided in the platforms [7]. The activities that are involved in the environments include registry, enrollment, discussions, tests and assignment work that provide the learners the time and space to work on their academic work.

Mining Educational Data

Data warehouses that have obtained amounts of data have rapidly increased, hence there is need to include more suitable intelligent and automatic organization of much larger collections of data. The aim of data mining is to extract and discover knowledge providing meaningful forms of insight through description and predictions. Date mining that can closely be coined to prediction and the most commonly used is classification. It determines collection of data from various sources such as logs, databases, csv compiled files, undergoes processing of collected data through cleansing, transformation of variables and partitioning, applying student's performance using mining algorithms and finally, make prediction interpretations over the models that have been analyzed [8].

- Technically, Classification locates already existing features that provide new description of data and causes a better understanding of each class's section in a dataset/database.
- Classification enables preparation of models to properly describe classes of any given data which provides certain useful predictions. The different techniques for statistical exploring classification functions are; Support Vector Machine, Naive Bayes, Neural Nets, Random forest Decision Tree etc.

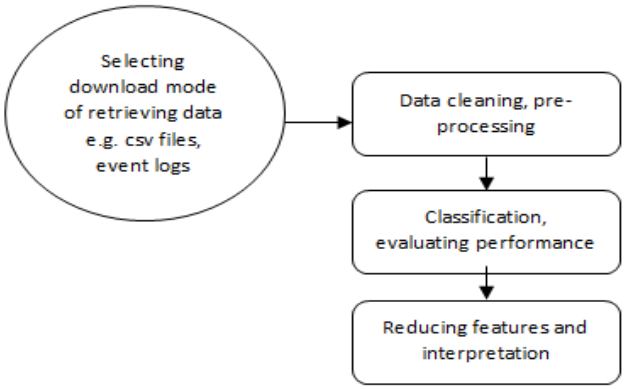

Figure 1. Circle of research in mining educational data

III. Literature Survey

Here, a survey on the already published related books and research papers is taken into account by going through the recent papers and discussing the methods that got to be used in such a paper. For this project 15 papers were taken into account, in additional to the base paper of the whole project.

S.no	Paper Name	Publication Details	Description of Method	Merit	Demerit
1.	Software Implementations and Applications of Elliptic Curve Cryptography	Wright State University CORE Scholar 2019	This paper focuses on implementation of Elliptic curve Cryptography to achieve the same level RSA has in security, with lesser key size, time efficient and full resource utilization	Faster Key management Smaller key size Achieves same level of security as RSA Time efficient and resource utilization Provides forward secrecy	Requires PC's with multicore (quad core processors) to accurately run execution time experiments Requires Standard C++ GNU Compiler of version 4.9.2
2.	Exploring the contribution of Data mining through educational research on classroom technology	University of Wollongong Research Online 2017	This a paper that briefly addresses and explains the studies of educational data mining technology, through association rules mining and fuzzy representation of problem-solving simulation in Europe and Australia.	Research can possibly be facilitated to make significant contribution to educational technology through use of data mining tools for researchers	No prior experience in organizing data for mining is a challenge for researchers. Appropriate cases are not demystified. Biased questionnaire subject answers
3.	Analysis on Academics and Mining Higher International Education data	Scholarship Journal for Teaching and Learning, 2010	This paper studies the emergence of educational data mining and academic analysis reviewing possibilities of collecting, analysing and presenting student data.	Document management and broadcast communication-oriented tools are used heavily Open Source CMSs use data mining already packaged for analysing courses	Course redesign is a time-consuming process as new curriculums are being adopted
4.	Exploring the implementation of Flipped Learning Approach by applying Data mining to discover Students' Learning Patterns in Coding Education	Symmetry Journals, 2020	This paper traces explores application of flipped learning, distance and face to face learning through educational data mining in a blended learning course in Northern Taiwan	Offers flexibility to manage time, progress and pace to learn, that face to face lacks Offers repeated learning not often offered in traditional learning	Platform efficiency in implementing self and instructor paced features High cost of maintenance and requires developers to review the system

5.	Data mining techniques and cryptography used in an E-learning system	International Research Journal of Engineering and Ttechnology, 2018	This is a paper that creates a framework of an e-learning platform, cryptography for security and data mining to execute student recommendations and cite learner's behaviour to improve course performance	Can allow students to access from anywhere and learn Only authorizes exact credentials of a person to use the e-learning services	Course and content management is a crucial for the admin to curate addition and removal of members and interaction interfacing
6.	E-Learning using the Blackboard system in Light of the Quality of Education and Cyber security	International Journal of Current Engineering and Technology 2019	This paper studies the implementation of a safe e-learning system called Blackboard by linking cybersecurity which provides many group quality and trust in the learning environment	Helps spread education for institution by breaking the barrier of racial issues Impowers solutions for students in security of confidential data and contributes transition from traditional to e-learning	Requires setup and constant monitoring of security for cyberthieves Developer needs to devise updates to curate course redesign which consumes a lot of time.
7.	A Review on Learning Analytics and Educational Data Mining as Research Areas	International journal of scientific & technology, December 2019	This paper reviews the research areas carried out in Intelligent Tutoring System, Latent Knowledge Estimation, Recommender Systems, clustering etc. to develop models for educational software	Learning analytics and educational inference model of student are key for use Provides the correlation advantage of learning performances between students	Tools are proprietary meaning it requires licenses which are expensive to get Requires researchers who have direct access to online logs
8.	Educational data mining using cluster analysis and decision tree technique: A case study	International Journal of Engineering Business Management, 2020	This paper explores techniques used in mining education platform data in Croatia which uses decision trees and clustering to analyse student behaviour in the e-learning platform event logs.	Analysing student data which eliminates, unnecessary data of lectures The values of the best resulted chosen considering branching of the trees and the approval of previously obtained clustering models of the results for interpretation	It is limited in huge databases that holds outbound dimensionality that is high.
9.	An encryption protocol using Elliptic Curve Cryptography for cloud computing environments.	International Journal of Engineering Development and Research	This paper focuses on the resolution of new generation schemes in public key cryptography which accommodate the limitations of bandwidth and power provide security adequate for all application.	The small key size leads to savings in processing power, bandwidth and memory The best curve chosen leads to intensive security hard to attack	Only fast and efficient in constrained environments
10.	Developing a learner-based management system using web 2.0	IODL&ICEM Joint Conference and Media Days, 2010	This determines importance of developing e-learning system in improving the standard functions, offers more interactive, dynamic, personal and flexible and features	Provide ease in customization of the interaction interface Rich in structural content features and functions for both lectures and learners	Constant need for updated tools to collect and connect the active parts of the learning platform

11.	Using Machine Learning to enhance Students Assessment	Moodle Application 2020	This paper reviews the role of integrating IoT, LMS and Machine learning with the aim of enhancing the assessment of learners automatically to evaluate their results using prediction model in machine learning with minimal error percentage.	IoT gives control to the institution to visibly enhance the performance levels using network utilities, sensors, big data etc. E-learning combined with machine learning and IoT brings about autonomous collection of data as well as deriving models for predicting student's progress	Host configuration is tedious job as to acquire hardware for sensing, compatibility issues and the machine learning solutions are expensive to deliver as a PC needs up to 8GB RAM Parameters of e-learning systems scale up since the experiments are done on multiple intervals which requires change in batch size.
12.	Scurrying Prevalent Issues and Trends of Distance Education in Malawi, to provide a successful model for Open and Distance Learning	Journal of Arts, Science and Commerce, 2016	This paper describes the necessity of adopting e-learning type education in Malawi through Open and Distance learning strategy. It investigates the new phenomenon within its context as path of adapting in technology of education which brings the influence of using both face to face and electronic learning.	Support services are provided to students which allow private interaction with the instructor in solving academic problems More resources are allocated in the establishment of new curriculums from systems	Limited funding and equipment's due the minimum wage given to people in Malawi Cost of maintenance is expensive such purchase of the device is at high demand
13.	Open data and technology exploring; E-mobility and Land use monitoring as e-learning examples	STS Conference, 2019	This paper combines the practical and theoretical knowledge on education-oriented matters which key for future professionals. It aims at presenting the best study cases on e-learning platforms through e-mobility and land monitoring and the advantages of Open Source platforms for e-learning.	Open source software creates ease for an institution in purchasing a premium offer which becomes proprietary E-mobility enables exploration of spatial insight about infrastructure of charging e-vehicles	Intensive computational demand leads to high cost in bandwidth, long processing time for beginners Only registered institutions with good funding are assured in participating in growth of open source evolution
14.	Using data mining and video learning analytics techniques in predicting performance of students in institutions of Higher Education.	Applied Sciences, MDPI, 2020	This study describes the empowering edge technology has brought upon higher educational institutions through use of various e-learning platforms. This paper aims at making student prediction on their performance regulated from initiating data mining and video learning analytics	Educational data mining brings about the valuable information through predictive analysis in data logs Video learning analysis enables researchers in identifying effectiveness using tools and related practices	Impacts students that have gadgets compatible with the technical requirements of the e-learning systems
15.	Improving Online Education Using Big Data Technologies	International Journal of Research in Engineering And Technology, 2020	This paper describes the framework that integrates cloud computing, Big Data and e-learning with the aim of developing sophisticated yet up scaling and on demand product that yields improvement in educational technology and	Processing is powerful since it's done in parallel The mechanism of creating LMS layer for e-learning content, big data layers that offers machine learning abilities and cloud	Big data technologies require lots of memory, speed in executing and great processing power to perform desired activities Cloud security is often done by the

			implements performance analysis on tracing learners through machine learning.	infrastructure for storage, networks and computing brings flexibility over traditional server hosting with limited resources.	develop to ensure confidentiality and ensure privacy

IV. METHODOLOGY

The objective of this study is to maintain the requirement of quality and not quantity bearing the minimum to ease difficulty in using the platform and making it more user friendly.

1. The learning registration and enrollment of participants by creating usernames, passwords and being checked with a Captcha method to verify identity.
2. Instructor Authorization: The administrator is tasked with allowing participants to be staff all privileges or a student with less privileges
3. Content and learning management API: The API's needed here included embedding video lectures through URLs to YouTube related subject channels, presentation slide files, tests assignments e-books etc.
4. SQL and NoSQL usage: the database required to store both structured, semi-structured including unstructured data.
5. Chatroom: This is a room being controlled by the instructor by posting a discussion topic through which class participants relate in issues concerning the topics.
6. Event logs/CSV files: This is a file that contains the information of activities in the platform such as all clicks in system to identify what was being done, what time and the person who doing some activity [9]. CSV files are spreadsheet type of comma spaced data which can be collected using scraping programs to extract students features such as time spent on the platform, assignment marks, test scores during interim and the end semester.

This platform is aimed at enrolling participant of all cultures to enable interaction, knowledge dispensation and understanding of origin backgrounds. Here events are emailed to participants on the commencement of class sessions, provided with an overview of the subject's contents in advance and probably some pre-discovery work. It is based on the instructor's method on giving learner a startup on knowing what to expect in the course.
The administrator as the driver of the system provides the interface of instructors to be assigned as staff with lecturing privileges and students who are there to view, communicate, collect and return assignments in the timeline specified. A chatroom is created by the instructor as required to build interactive sessions concerning tops, exercises among other things [10].

The data that is generated in this platform is transported to a file known as an event log to proceed with data mining and creating predictions on the usage. The researcher can only to acknowledge the performance of the participants depending on the models that are trained in the process and fair attendance level. With extensive attendance, interaction with the instructors and the platform will lead to more accuracy measures closely how much student can maintain grades especially for highly active students as well as giving warnings to others.

SYSTEM ARCHITECTURE

The architecture below describes components, interfaces and features of dataflow for a system to satisfy specified requirements in a given scenario. The diagram in the following section describes the overall architecture of the system.

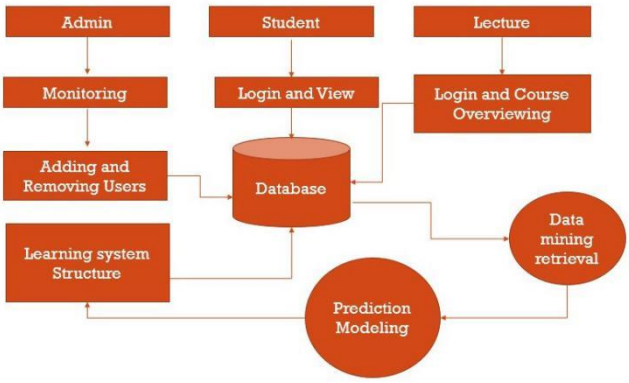

Figure 2: Overall System ARCHITECTURE

V. ALGORITHM AND TECHNIQUES

CLASSIFICATION

[11] This goal of this technique is to basically locate already existing features that provide new description of data and causes a better understanding of each class's section in a dataset/database by distinguishing labels. Classification enables the ability to prepare models to properly describe classes of any given data which provides certain useful predictions. The different techniques for statistical exploration used for a classification function such as SVM (support vector machine) [12].

DECISION TREE

This is exploratory algorithm for classification and regression that adopts an incumbent model that creates a structured tree from a predefined dataset, where each point or node represents the condition or test attributes and final leaf nodes represents the test classes or results [13]. The decision tree construction strategy aims at using the divide and conquer method where the attributes to come across here are seemingly categorical and if assured of the continuous values then a conversion to discrete values will be required before beginning the process. All samples are

initially on a single root node and the rest are created based on the attribute of condition partitioning.

- If each node belongs to the same class sample
- If there are no other portions to be sampled
- If there is no continued division to be made in the existing sample.

RANDOM FOREST

This technique is used to curb the limitations of decision trees as they may confuse or overfit the data that is being trained. Essentially it will be able to create more trees to do a better job in making predictions among classified clusters of student groups who are using the platform to their advantage and those who are not attending assigned class sessions and performing less. This algorithm will be used to acquire accuracy, speed and scalable predictions that outweigh other algorithms such as Random Forest and decision tree.

VI. RESULT

Statistical Methods used in the study	Prediction statistic percentage (%)
Decision Trees	**80**
Random Forest	**80**
K-means	**60**
SVM	**70**

Students that have more active participation on the discussions, ask questions frequently, use the system during extended hours in search of resources have a sharp accuracy of getting good marks. A lot of females need help in engaging with the system in order to increase their scores. A lot of male student have a tendency to exposure such that most of them tend to get good marks according the relationship of High, Low and Medium class target variable over assignment marks, visited resources, discussion sessions, midsemester accumulation and active times on the system.

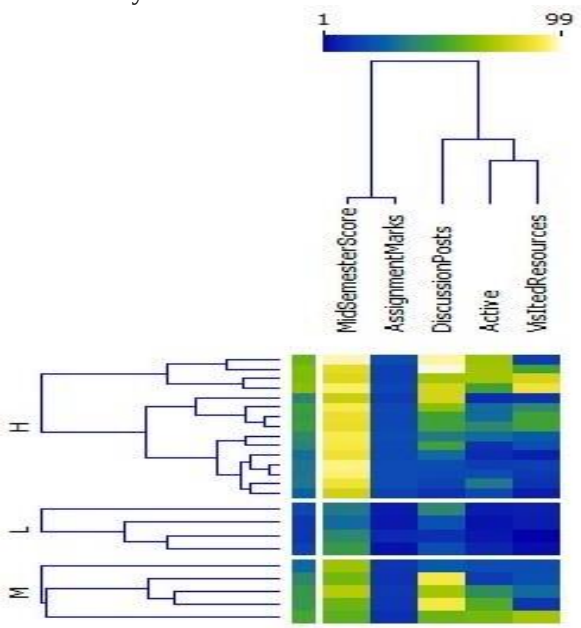

Figure 3. Heat Map [2]

VII. CONCLUSION

In conclusion data mined from educational environments can be explored to enable academic learning institutions to improve many challenges both on the side of faculty and the students. The recent popularity of mining e-learning environment has led to implementat ion of classifying groups of students from various faculties by analyzing clusters and applying decision tree for immersed behavior of student in lecturing processes [14]. The random forest, decision tree and provide interpretation of data that can effectively influence learning processes through which boosting learner's performance through analysis education which will be made better with innovations that enhance mode of teaching as well as learning to bring a future that commendable for all people [15]. Data scrapping and collection will enablee mining of student's data to create models for improved learning experience using decision trees, clustering of various group of individuals and provide recommendations obtained from the results.

REFFERENCES

[1] D.B. Vaghela, P. Sharma, "A data mining approach used in University exams of Engineering learners on the Early prediction of student's performance", International Journal of Scientific Research in Multi-disciplinary Studies, **Vol:1, Issue 5, 2020.**

[2] R. Gautam, D. Pahuja "A review on mining Educational Data", International Journal of Science and Research, **Vol.3 Issue.11, 2014**

[3] W. Chou, Elliptic Curve Cryptography and Its Applications to Mobile Devices, Department of Mathematics.

[4] M. Chibambo, "Scurrying Prevalent Issues and Trends of Distance Education in Malawi, to provide a ssuccessful Model for Open and Distance Learning", Journal of Arts, Sscience and commerce, **2016**

[5] R. Hasan, and M. U. Sattar, A. Abbas, K.l U. Sarker, "Using data mining and video learning analytics techniques in predicting performance of students in Higher Education institutions", **2020**

[6] O. S Ahmed, "Cyber security in a Blackboard system for E-Learning used in Light of Educational Quality", IJCET, **2019**

[7] P. Baepler, C. James, "Analysis on Academics and Mining Higher International Education data", Journal of scholarships for Teaching and Learning, **2010.**

[8] S. Krizanic ,"Clustering and decision analysis implemented with decision tree techniqueis used for mining Educational Data", International Research Journal of Engineering Business Management, **2020.**

[9] A. O. Mohamed, M. Omar, "A Requirements Modeling for E-Learning Management System", Iinternational Journal of Rresearch Ttechnology and Eengineering, **Vol: 7 Issue-6S2, 2019.**

[10]K. Dahdouh, A. Dakkak, "Improving Online Education Using Big Data Technologies", **2020.**

[11]A. Sankari, Dr. S. Masih and Dr. Maya Ingle, "A Review on Research Areas in Educational Learning Analytics and Data mining", International Journal of Scientific & Technology Research, **Vol. 8, Issue 12, 2019.**

[12]S. Herold, H. Meinel, G. Lorenzen-Zabel, A. "Blessings of open data and technology: E-mobility and Land use monitoring as e-learning examples", STS Conference, **2019.**

[13]A. Mousawi, K. Mershad .l Said, P. Ibrahim,"Using Machine Learning to enhance Students Assessment Moodle Application", published in May **2020**

[14]A. S. Chautan, "Enhancement of Academic decision-making using Clustering and Classification techniques in environments of

Higher Education", International Journal of Scientific Research in Computer Engineering, Vol.8, Issue.2

[15]O. Deperlioglu, A. Kocatepe, "Developing a learner-based management system using web 2.0", IODL&ICEM Joint Conference and Media Days, **2010.**

AUTHORS PROFILE

Mr. E. C. Mwakamo is currently studying Bachelor of Science in Computer Science from University of DMI St John the Baptist, Malawi. He is recent member of the ISROSET who is publishing his first research study paper in 2021.

