International Journal of Computer Applications Technology and Research
Volume 10–Issue 08, 198-207, 2021, ISSN:-2319–8656

The Role of Machine Learning and Data Mining Techniques in Predicting Students' Academic Performance

Dr. Aliyu Y. Rufai
Department of Computer Science, Kebbi State University of Science and technology. Aleiro, kebbi State, Nigera

Dr.Hassan U. Suru
Department of Computer Science, Kebbi State University of Science and technology. Aleiro, kebbi State, Nigeria

James Afrifa
Department of Computer Science, Adamawa State Polytechnic, Yola, Nigeria

Abstracts: The advancement in Information Technology makes it easier and cheaper to collect large amounts of data, but if this data is not further analyzed, it remains only huge amounts of data. These large amounts of data set have motivated research and development in various fields to extract meaningful information with a view of analyzing it to solve complex problem. With new methods and techniques, data can be analyze and be of great advantage. Data mining and machine learning are two computing disciplines that enable analysis of large data sets using different techniques. This paper gave an overview of several applications using these disciplines in education, with focus on student's academic performance prediction. Early prediction of students' performance is useful in taking early action of improving learning outcome. The perfect methods for this are machine learning and data mining. This paper also discusses special use of data mining in education, called educational data mining. Educational Data Mining (EDM) uses different methods and techniques from machine learning, statistics, data mining and data analysis, to analyze data collected during teaching and learning. The goal of this paper is to introduce the role of machine learning and data mining in predicting student's academic performance and to present its applications and benefits

Keywords: Data, Educational Data Mining, Machine Learning, Data Mining, Education, Students' academic performance,

1. INTRODUCTION

Education plays a very vital role in human resource development and also in the society at large (Pinheiro, Wangenge-ouma, Balbaceusky & Cai, 2015). Student's academic performance determines the quality of education in a higher institution of learning. Student's performance plays an essential role in producing good quality graduates that in future will be responsible for the country's development. The performance of the student is one of the most important aspects of every educational institution. Predicting student's academic performance is one of the most important steps towards efficient education while understanding student performance is essential for the establishment of a student centric learning environment. The definition of student's academic success is based on individual perceptions and sometimes most often misused within educational research. However, the study of York, Gibson & Rankin (2015) suggested that the definition of academic success is made up of six components:

(1) Academic achievement, which is nearly entirely measured with course grades and grade point average (GPA), (2) Satisfaction, which is often captured either by course evaluation or institutional surveys, (3) Persistence, which is measured by retention between particular years of college and degree attainment rates, (4) Acquisition of skills and competencies, which can be measured by assignments and course evaluations, (5) Attainment of

learning objectives, which can also be measured by assignments and course evaluations, and finally (6) Career success, which can be determined by job attainment rates, promotion histories, career satisfaction and professional goal attainment.

Another crucial requirement for maximizing students' success is the identification of the factors that affects students' academic performance. Identification of these factors could assist in achieving the highest level of quality education (Yassein, Gaffer, Helali & Mohomed, 2017). These factors could be useful in decision making concerning student's academic performance.

Poortman & Schildkamp (2016) Suggested that Data – base decision can contribute to increased student learning and achievements.

With advancement in Information Technology (IT) development and lower prices, tertiary institutions start to collect a huge amount of data about their students. These data can be further analyzed with data mining methods and techniques.

Data mining tools, methods and techniques, allow you to analyze data and find hidden patterns and information. Data mining is said to be the most robust methodology mostly used for assessing useful information from the data warehouse (Salloum, El-emran, Addallah & Shaalan, 2017). Data miming can improve decision making by predicting hidden information through extraction method (Salloum, Mhamdi, Aliemran & Shaalan, 2017) . Data mining is used to detect patterns and relationships in data to improve decision-making processes. Data mining is an interdisciplinary field that brings together techniques from statistics, artificial intelligence, neural networks, database systems, machine learning, pattern recognition, data visualization, knowledge acquisition and information theory (Arunachalam & Velmurugan, 2018) (Sumathi & Sivanandam, 2013). The practice of Data Mining methods applied to educational data is known as Educational Data Mining (EDM) (Baker & Yacef 2009). It is an interdisciplinary area that brings together techniques from statistics, artificial intelligence, database systems, machine learning, pattern recognition, data visualization, knowledge acquisition and information theory (Sumathi and Sivanandam, 2006) to find useful patterns and, thus, help understand students' behavior and how they learn It is drawn from a variety of domains, including Data Mining, Machine Learning, psychometrics and other areas of statistics, information visualization, and computational modeling (Romero and Ventura 2007). The education system has become more balanced due to the improved mining application (Romero, Ventura & Garcia, 2008)

2. EDUCATIONAL DATA MINING (EDM)

Educational institutions collects huge amount of data and stores them. The amount of data collected and stored are growing too big and educational data analysis could not be performed manually anymore. One of the elements of data mining is educational data mining (EDM), the key focus of which is on developing models for extracting hidden knowledge from the student's data, which if use will enhance student's academic performance. Educational data mining may also be considered as a new model that is part of the prevailing education system, which is able to generate positive interaction with different parts of the system. This will enable it to eventually attain the objective of enhancing teaching (Iie-Haiyan, Biac & Yuan, 2017). The goal of EDM is to improve the educational process and to explain educational strategies for better decision making (Silva & Fonseca, 2017)

Educational Data Mining (EDM) is defined as the application of techniques of traditional data mining to educational data analysis, with the objective of obtaining solutions to problems in the field of education (Baker & Yacef, 2009). There are certain EDM applications that include the formulation of e-learning systems (Lara, Lizcano, Martinez, Pazos & Riera, 2014), clustering educational data (Chakrabary, Chakma & Mukherjee, 2016) as well as making student performance predictions(Chauhan, Shah, Karn, & Dalal,2019). Several kinds of techniques are presently quite popular in educational data mining, which are part of the following categories: sequential pattern, clustering, prediction, classification, machine learning models and association rule analysis.

2.1 Educational Data Mining Process

There are series of phases involved in the EDM, however, the EDM process has four main basic phases. The first among them is the Problem definition Phase in which a specific problem is translated into a data mining problem. This phase formulated the project goal and objectives, as well as the main research questions. The second phase of the EDM process is the data preparation and gathering phase, and it is most time-consuming phase. It can take up to 80 % of all analysis time. Data quality is a major challenge in data mining (Blake and Mangiameli, 2011). In this phase, source data must be identified, cleaned and formatted in prespecified format. Once that is done, the next phase is the Modeling and Evaluating phase in which the parameters are set to optimal values and different modeling techniques are selected and applied. The last phase is the deployment phase in which the results of data mining are organized and presented through graphs and reports.

3. MACHINE LEARNING AND DATA MINING

This study focused on the role of Machine learning and Data mining in predicting student's performance. ML and DM are mostly confusing because of similar meanings, so they have a meaningful similarity. Founder of machine learning, Arthur Samuel, describes it as an area of research which gives the ability to learn without being explicitly programmed. ML interacts with learning the pattern recognition and computational learning theory in artificial intelligence. ML is older than DM. In the recent days, the term data mining is extra popular than its sibling machine learning which can be the reason for some scholars to actually highlight their study for data mining than machine learning, so in this study, machine learning and data mining are discussed together. Every DM and ML process involves six main steps. In the First, there is a Business Understanding in which a specific problem is translated into a data mining problem. The second step is Data Understanding, which starts with data collected from all applicable data sources. In this step, data load and data integration are done. Data visualization tools are often used in this step to explore the properties of the data. The most important step is Data Preparation, and it can take enormous amounts of time depending on the amount of data analyzed and the number of data sources (Blake and Mangiameli, 2011). Data quality is a major challenge in data mining. The final data set must be cleaned, formatted and constructed into a specific form. In the Modelling and Evaluation step, mathematical models are used to find patterns in the data using sophisticated data tools and parameters are calibrated to optimal values. The last step is Deployment in which the results of data mining are presented (Oracle, 2019).

Data mining and machine learning is basically classified into supervised, un-supervised and reinforcement approaches.

Supervised Learning: whereby the machine is provided with labeled data for both input as well as expected output during its training and the supervised learning algorithm generates a mapping function that can identify the expected output for a given input. The training process continues till the algorithm reaches the preferred level of accuracy. One of the standardized goals of supervised learning is to make the computer learn a classification system; therefore, it is commonly used to solve classification problems. For example, the machine could be trained to classify a spam e-mail from a valid e-mail, already being used by Google for Gmail spam filtering. Nearest neighbor, Naïve Bayes, Decision Trees, Linear Regression, Support Vector Machines and Neural Networks are a few of the most common algorithms that are included under this category.

Unsupervised Learning: whereby the machine is provided with unlabeled and unclassified input dataset and the unsupervised learning algorithm generates a function to identify hidden structures in the given dataset as per the patterns, similarities and differences that exist among data without any former training. There is no assessment of the level of accuracy of the structure identified by the machine. One of the major focuses of unsupervised learning algorithms could be clustering and association problems. A few of the commonly used unsupervised learning algorithms are k-means algorithm for clustering and Apriori algorithm for association problems.

Reinforcement Learning: the machine is exposed to an environment where it takes decisions on a trial and error basis and learns from its own actions and past experiences. For every correct decision the machine receives a reward feedback from the environment that acts like a reinforcement signal and the information about the rewarded state-action pair is stored. Later on the machine iterates the rewarded behavior whenever faced with a similar situation. Reinforcement learning algorithms have their usage in domains where strategic decision making is the key to success like Self Driving Cars. Few of the most commonly used reinforcement learning algorithms are Q-Learning and Markov Decision processes (Puget, 2016).

4. DATA MINING AND MACHINE LEARNING TECHNIQUES FOR PREDICTING STUDENT'S PERFORMANCE.

The ability for an institution to predict results might help students develop a good understanding of how well or bad they would perform in a course and then can take steps accordingly.

Early prediction of students' performance is useful in taking early action of improving learning outcome. Predicting a student's performance from past academic data is one of the most popular applications of educational data mining and, therefore, it is a valuable source of information that can be used to improve students' performance (Buenaño-Fernández, Gil & Luján-Mora, 2019).

This section discussed the various techniques in DM and ML that are used in predicting students' performance. There are many various DM and ML d methods to build a predictive model for student's performance. The most commonly used method is classification. Among classification algorithms are Neural Networks, Decision Trees, Naïve Bayes, Support Vector Machine and K-Nearest Neighbor (Shahiri and Husain, 2015).

4.1 Classification Algorithm

Classification is a data mining technique that segments data in a collection to target categories or classes. It helps in analyzing data and predicting outcomes. The goal of classification is to accurately predict the target class or each case in the data. The classifier training algorithm uses pre-classified examples for determining the set of parameters required for classification (Oracle, 2019). In the educational sector, this technique is often used for classifying students based on some characteristics such as age, gender, grades, knowledge, academic achievements, motivation, behavior, demographic or geographic characteristics, etc

4.2 Decision tree

A decision tree is a decision support tool that uses a tree-shaped graph or model for classification. It is a supervised learning method. Each internal node represents a test on an attribute, each branch represents an outcome of the test, and each leaf node represents a class which is a decision after computing all attributes. The paths from the root to leaf are classification rules (Vidal et al, 2014). Their greatest advantage is stability and easy interpretation. Because of their simplicity, they are suitable for solving a different kind of problems in a broad range of industries such as financial, business, healthcare, education, energy, engineering, pharmaceutical, law, etc

4.3 Artificial Neural Networks.

Artificial neural network inspired by the human brain includes huge number of neurons with every neuron having an input and output with an activation function. Commonly, neural network is based on the layer approach ;the first layer is the input layer, the last layer is the output layer, and the other layers are known as hidden layer

4.4. Naive Bayes.

It is a simple probabilistic classifier [33]based on the Bayes theorem. It gives a way to calculate the future probability P(c–x), from P(c),P(x), and P(x–c). It assumes that the result of the value of the predictor (x) on the given class (c) is liberated than the other predictors. The theory is known as "class conditional independence."

4.5 Support Vector Machine.

SVM is as supervised ML algorithm to help for classification/regression problems. It based on searching an unraveling hyper plane in the feature space in two classes in such a way that the space between the hyper plane and the nearest data point of every class is maximized. This method depends upon reduced classification threat somewhat that is an ideal classification. SVM has good reputation for its generalization ability and binary classifier, and multiclass classification is understood by making an SVM for every pair of class.

4.6 Regression Analysis

Prediction refers to calculated assumptions for certain events made based on available processed data. Regression technique can be used for prediction, to model the relationship between one or more independent variables and dependent variables. Independent variables are attributes already known and response variables are what we want to predict (Bhatnagar, 2013). It has many applications in business planning, trend analysis, financial forecasting, time series prediction, trend analysis,etc. In the educational sector, it is used for the prediction of students' academic performance, prediction of enrolled students, prediction of the final grade, prediction of drop-outs, etc.

5. TOOLS FOR DM AND ML

There are four free and most famous software including RStudio, RapidMiner, Pandas, StatsModels, and Scikit-learn (python libraries) and Weka. RStudio is an IDE for analytical computing and graphics. It is a core programming language known as Rlanguage. It has console, code highlighting editor for the direct lunching code with utility for history, debugging, and so on. RStudio has different edition including free of cost, business use, and web application platform. After Rfunctionality, many high-level scripting languages like python are used by students R especially for sample regression and correlation analysis. RapidMiner is a good tool which provides the platform for data mining, machine learning, text mining, predictive analysis, and soon. It can create statistical workflow getting data from many data sources. RapidMiner is used for business, commercial, research, education, and training purposes. It is compatible with the stages of data mining like data preprocessing, creating model, graphical result, and so on. It is an excellent educational tool in the field of data science. Nowadays, Python is the most popular multipurpose programming language. It has packages, libraries, and framework for data analysis and visualization, which are related tasks. Pandas, Stats Models, and Scikit-learn are the three open-source libraries. It can be best to use all three libraries in IPython. “Pandas” is an open-source library that gives high performance easy to use data analysis tools for the Python language; “Statsmodel” is also the Python library which facilitates programmer to data explore, estimate analytical model, and perform statistical test.“Scikit-learn” is another open-source machine learning library for Python. It has many classifications, regression, random forest, K-means, and so on. It is a highly useful library to explore data and estimate analytical model and test. Weka is used to implement algorithms related to data mining and machine learning based on Java. It has tools for data preprocessing, regression analysis, clustering, classification methods, association rules, and so on. Weka is free to use and has user-friendly GUI (Bilal, Wang, & Zain, 2018).

6. RELATED WORK

The need to predicting student academic performance has become a crucial factor in improving the quality of education; assist the students’ academic performance as well as providing the teachers more options when training their students. Educational data mining is a young research area which is becoming increasingly popular due to its potential. Educational data can be used to assist instructors, to improve curriculums, to understand students' behavior, to improve teaching process, to improve e-learning systems, to identify reasons for dropping out, to support decision making and improve students learning behavior (Romero & Ventura, 2010). In recent times, there are many works been published related to this subject matter.

Anderson, Boodhwani & Baker (2019) conducted a study to predict graduation at a public R1 university using linear support vector machines, decision trees, logistic regression, and stochastic gradient descent binary classifiers. The study used a data set of over 14,000 students from six Fall cohorts, containing 104 features, drawn from pre-existing university data. The prediction accuracy was as follows; decision tree 0.786, linear SVM 0.801, logistic regression, 0.810 and stochastic gradient descent binary classifiers 0.824.

According Sedkaoui & Khelfaoui (2019), an analysis of the relationship between big data and educational environments has been presented. The work focuses on the different methods, techniques, tools, and big data algorithms that can be used in the educational context in order to understand the benefits and impact that can cause in the teaching and learning process. The discussion generated in their study suggests that the incorporation of an approach based on big data is of crucial importance. This approach can contribute significantly in the improvement of the learning process, for its implementation must be correctly aligned with the learning needs and the educational strategies.

Devasia et al. (2016) examine various data mining techniques for the prediction of students' performance. They used all student admission details, course details, subject details, student marks details, attendance details and student's academic history as input. The results of their paper show that Naive Bayesian algorithm is more accurate than other methods like Regression, Decision Tree, Neural networks etc., for comparison and prediction.

Oancea, Dragoescu, & Ciucu (2013) used the classification power of a neural network to predict the students' results measured by the grade point average in the first year of study. The input variables were:type of the study program (part-time or full-time education),gender, high-school graduation average, age and difference in years from the moment the student graduates high-school until he/she enrolls at university. The goal of their research was to help university management in order to take early action to avoid the phenomenon of leaving education.

7. The Role/Application of DM and ML

The role and applications of DM and ML are numerous. There are many research papers and studies regarding the use and applications of ML and DL in predicting student's academic performance education. The most common use of the two terms include: improving the process of studying, improving course completion, supporting students in course selection, students' profiling, finding problems leading to dropping out, students' targeting, curriculum development, predicting student's performance and as a support for decision-making at student enrolment.

Predicting a student's performance from past academic data is one of the most popular applications of educational data mining and, therefore, it is a valuable source of information that can be used to improve students' performance (Buenaño-Fernández, Gil & Luján-Mora, 2019). This is where Data Mining and Machine Learning come in. Machine Learning is a set of techniques that gives computers the ability to learn without the intervention of human programming (Navamani & Kannammal, 2015).

The application of machine learning techniques to predicting students' performance, based on their background information and their in-term performance has proved to be a helpful tool for foreseeing poor and good performances in various levels of education (Soni, Kumar, Kaur & Hemavath, 2018). Machine learning offers an advantage over traditional forms of statistical analysis, placing emphasis on predictive performance over provable theoretical properties and priori super-population assumptions. Thereby tutors are enabled to timely help the weakest ones, but also, to promote the strongest thus improving learning.

Yukselturk, Ozekes & Türel (2014) experimented with four algorithms: k-nearest neighbors, decision trees, the naive Bayes classifier and artificial neural networks, to classify students who dropped out of school. The data set was collected by administering an online test for 189 students enrolled in 2007 to 2009. The machine learning algorithms were trained and tested using the 10-fold cross-validation technique. The best results were obtained using 3-nearest neighbors and decision trees, with an accuracy of 87% and 79.7% respectively. These results were useful since they permitted prediction of student dropout in the online program data set. Finally, the authors concluded that data mining methods might help to predict different reasons why students decide to drop out before finishing their study programs.

The application of machine learning techniques to tackled problem was introduced by Kakavand, Mokfi & Tarokh (2014), with the purpose of predicting student loyalty using decision trees. The authors researched the external factors that may generate loyalty, in order to identify students who have decided to continue studying, and thus the university may invest in them and increase its educational quality. The experiments were performed using a data set of 135 instances for training, 33 for testing and 35 for validation, with 14 attributes per instance (gender, age, and income, among others). The best result was obtained using the CARTdecision tree algorithm,with 94% accuracy.

In another research, Data Mining: A prediction for performance improvement using classification Bhardwaj and Pal (2012) used Bayes classification for the construction of a prediction model to identify the difference between high learners and slow learners.

Kovačić (2012) examined the socio-demographic variables (age, gender, ethnicity, education, work status, and disability) and study environment (course program and course block) that may help in identifying successful and unsuccessful students. This research concluded that classifying students based on pre-enrolment information can help to identify students at-risk of dropping the course and suggest advising and mentoring programs to make them successful.

Maqsood (2013) stated that data mining can be used to report and analyze the data that can help in preparing marketing strategies for targeted students.

Kumar and Chadha (2011) presented an empirical study of the applications of data mining techniques in higher education in which they tried to identify the potential areas in which data mining techniques could be applied. They concluded that potential applications are: organization of syllabus, predicting the registration of students in an educational program, predicting student performance, detecting cheating in the online examination and identifying abnormal or erroneous values and used data mining techniques are: association analysis, classification, prediction, clustering and outlier analysis.

Kumar & Pal (2011) used decision trees to extract a set of academic characteristics to assess students' performance. The data set consisted of 50 examples from students of the computer applications department at the VBS Purvanchal University, India. The characteristics considered included grades in previous semesters, seminar performance, general proficiency performance, and attendance. The knowledge extracted and represented by the decision tree enabled the authors to obtain if-then rules to classify the students. With this work, the authors predicted students' end-of-semester performance and identified students who needed special attention to reduce the failure rate.

Data mining was used by Ranjan & Khalil (2008) with two main objectives: 1) planning a course for education management through new data mining applications and to explore the effects on probable changes in the recruitment and admission processes and guiding courses; and 2) ensuring quality evaluations, student performance, courses and tasks. Also, they were interested in finding patterns in how students interact with others, how the admission process is carried out, and the mechanisms in counseling to choose courses. The framework was tested to find what types of courses are interesting to certain types of students. In addition, the proposed framework had three major processes that usually occur in all management institutions, namely admissions (planning, evaluation and registration), counseling, and allocation of specialization subjects. The results were presented as a conceptual framework to adopt data mining in management in institutions. The authors concluded that data mining is useful for predicting the success of educational programs, and also understanding learning styles in order to promote proactivity in students. Oladokun, Adebanjo and Charles-Owaba (2008) proposed a neural network model to predict the performance of possible candidates for admission to university. Their model was based on a multilayer perceptron topology, with family history, age, and score of the entrance exam taken as input variables, among others. They reported about 74% accuracy in their results, but recommended conducting an oral interview to obtain more information to supply to their model.

8. CONCLUSION

This study demonstrates the role of machine learning and data mining techniques for predicting student's academic performance. The study highlighted the papers that define the use of multiple Machine Learning and Data Mining techniques for predicting student's academic performance.

Data Mining and Machine Learning within the academic context is applicable, allowing educational institutions to better allocate human and material resources, manage student performance and improve the effectiveness of performance throughout students 'education. Machine learning and Data mining techniques were developed to automatically discover hidden knowledge and recognize patterns from data. Educational data mining can be used for classifying and predicting students' performance, dropouts as well as teachers' performance.

. In this paper, we presented the benefits and applications of machine learning and data mining techniques in many educational areas. The main goal of the paper is to reveal the high potential of educational data mining applications, techniques, tools and its role in predicting student's academic performance.

9. ACKNOWLEDMENTS

Our thanks goes to everyone who in one way or the other has contributed to the success of this paper.

10. REFERENCES

1) Anderson, H., Boodhwani, A., & Baker, R. (2019). Predicting graduation at a public R1 university. In Proceedings of the 9th International Learning Analytics and Knowledge Conference.
2) Arunachalam, A.S., & Velmurugan, T. (2018). Analyzing student performance using evolutionaryartificial neural network algorithm. Int. J. Eng. Technol.7(2.26), 67–73
3) Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. JEDM| Journal of Educational Data Mining, 1(1), 3-17
4) Bhardwaj, B. K., Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. Retrived 25.5.2019. from https://arxiv.org/pdf/1201.3418.pdf
5) Bhatnagar, V. (Ed.). (2013). Data mining in dynamic social networks and fuzzy systems. IGI global
6) Bilal, A., Wang, J., & Zain, A. (2018). Role of Machine Learning and Data Mining in Internet Security: Standing State with Future Directions Journal of Computer Networks andCommunications Volume 2018, Article ID 6383145, 10 pageshttps://doi.org/10.1155/2018/6383145
7) Blake, R., & Mangiameli, P. (2011), "The effects and interactions of data quality and problem complexity on classification", Data Inform, pp. 160-175
8) Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. *Sustainability*, *11*(10), 2833-2851.
9) Chakraborty, B., Chakma, K., & Mukherjee, A.(2016). A density-based clustering algorithm andexperiments on student dataset with noises using Rough set theory. In: 2016 IEEEInternational Conference on Engineering and Technology (ICETECH), pp. 431–436

10) Chauhan, N., Shah, K., Karn, D., Dalal, J. (2019). Prediction of student's performance usingmachine learning . SSRN 3370802
11) Devasia, T., Vinushree, T. P., Hegde, V. (2016). „Prediction of students performance using Educational Data Mining". In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE). pp. 91-95.
12) Jie, W., Hai-yan, L., Biao, C., & Yuan, Z. (2017). Application of educational data mining on analysisof students'online learning behavior. In: 2017 2nd International Conference on Image,Vision and Computing (ICIVC), pp. 1011–1015
13) Kakavand, S., Mokfi, T., & Jafar, M. (2014). Prediction the loyal student using decision tree algorithms. International Journal of Information and Communication Technology Research, 4(1), 32-37
14) Kovacic, Z. (2012). Predicting student success by mining enrolment data. Research in Higher Education, 15
15) Kumar, D.N., Raju, K.S. and Sathish, T. (2004). River Flow Forecasting Using Recurrent
    a. Neural Networks. Water Resources Management, Kluner Academic, 18:143 – 161
16) Kumar, B., & Pal, S. (2011). Mining educational data to analyze students' performance. International Journalof Advanced Computer Science and Applications, 2(6), 63-60.
17) Lara, J.A., Lizcano, D., Martínez, M.A., Pazos, J., & Riera, T. (2014). A system for knowledgediscovery in e-learning environments within the European Higher Education Area-Application to student data from Open University of Madrid, UDIMA. Comput. Educ.72,23–36
18) Maqsood, A. M. (2013). Customer Relationship Management in B-Schools: An Overview. International Journal of Computer Sciences and Management Research, 2(4), 2108-2119.
19) Navamani, J.; Kannammal, A. (2015) Predicting performance of schools by applying data mining techniques on public examination results. *Res. J. Applied Science of Engineering and Technology.* ***2015****, 9, 262–271.*
20) Oancea, B., Dragoescu, R., Ciucu, S. (2013). „Predicting students' results in higher education using a neural network", available at: at https://mpra.ub.uni-muenchen.de/72041/
21) Oladokun, V. O., Adebanjo, A. T., & Charles-Owaba, O. E. (2008). Predicting students' academic performance using artificial neural network: a case study of an engineering course. The Pacific Journal of Science and Technology, 9(1), 72-79
22) Oracle (2019), "Data Mining Concepts", available at: http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#DMCON002
23) Pinheiro, R., Wangenge-Ouma, G., Balbachevsky, E., & Cai, Y. (2015). The role of higher education in society and the changing institutionalized features in higher education. In The Palgrave international handbook of higher education policy and governance (pp. 225–242). London: Palgrave Macmillan UK.
24) Poortman, C. L., and K. Schildkamp. (2016)."Solving Student Achievement Focused Problems witha Data Use Intervention for Teachers."Teaching and Teacher Education60: 425–433.
25) Puget, J. F. (2016). *What is Machine Learning? (IT Best Kept Secret is Optimization).* Retrieved March 12, 2021, from IBM Developer Works: https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en.
26) Ranjan, J., & Khalil, S. (2008). Conceptual framework of data mining process in management education in India: an institutional

perspective. Information Technology Journal, 7(1), 16-23

27) Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618

28) Romero, C., Ventura, S., & García, E.(2008). Data mining in course management systems: moodle case study and tutorial. Comput. Educ.51(1), 368–384

29) Salloum, S.A., Al-Emran, M., & Shaalan, K (2017). Mining social media text: extracting knowledge from Facebook. Int. J. Comput. Digit. Syst.6(2), 73–81

30) Salloum, S.A., Mhamdi, C., Al-Emran, M., & Shaalan, K. (2017). Analysis and classification of Arabic newspapers'Facebook pages using text mining techniques. Int. J. Inf. Technol. Lang.Stud.1(2), 8–17

31) Sedkaoui, S.; Khelfaoui, M. (2019) Understand, develop and enhance the learning process with big data. Inf. Discov. Deliv. 2019, 47, 2–16.

32) Shahiri, A. M., Husain, W. (2015). „A review on predicting student's performance using data mining techniques." Procedia Computer Science. No. 72, pp. 414-422

33) Silva, C., Fonseca, J. (2017). Educational Data Mining: a literature review. In Europe and MENA Cooperation Advances in Information and Communication Technologies (pp. 87-94).

34) Soni, A., Kumar, V., Kaur, R., & Hemavath, D. (2018). Predicting student performance using data mining techniques. *International Journal of Pure and applied Mathematics*, *119*(12), 221-227.

35) Yassein, N. A., Gaffer, R., Helali, M., & Mohomad, S. B. (2017). Citation: Predicting student academic performance in KSA using data mining techniques. Journal of Information Technology and Software Engineering,7(5), 213.

36) York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success: Practical Assess-ment. Research & Evaluation,20(5), 462

37) Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). Predicting dropout student: an application of data mining methods in an online education program. European Journal of Open, Distance and eLearning, 17(1), 118-133

International Journal of Computer Applications Technology and Research
Volume 10–Issue 08, 208-212, 2021, ISSN:-2319–8656

A Study on the Impact of COVID-19 Pandemic in African Universities

Elijah Isanda Omwenga
University of Kabianga
Kericho, Kenya

Peter M. F. Mbithi
University of Nairobi
Nairobi, Kenya

John N. Muthama
University of Nairobi
Nairobi, Kenya

John Mwarabu Chone
University of Kabianga
Kericho, Kenya

Abstract: This paper presents a study on how University students, lecturers, administrators and managers perceive the impacts of the of COVID-19 crisis on various aspects of University teaching and learning in Africa, and particularly in Kenya. The sudden closure of campuses as a preventive measure to community transmission shifted face-to-face classes to virtual learning modes. With a sample of 1,236 University students and staff from 18 universities in Kenya and Nigeria, the study reveals that amid the worldwide lockdown and transition to online learning, expectation by 44% of the students were met. Students were mainly concerned about issues to do with internet connectivity, computing devices and electrical power. In addition, utilising e-Learning tools and platforms for effective student engagement posed limitations of accessibility and affordability for many students. The teaching staff on the other hand were mainly concerned with access to the teaching resources, conducting online teaching, capacity to handle the online mode of teaching, devices and eContent development. The pandemic has exposed the shortcomings of the current higher education system and the need for enhanced policy formulation and implementation on digital infrastructure to adapt to the rapidly changing education ecosystem of the world. In the post-pandemic situation, the use of eLearning and virtual education may become an integral part of the higher education system. Key factors influencing students' satisfaction with the role of their University are also identified as internet access, quality of e-content and e-content development. Policymakers, stakeholders and higher education institutions in Africa may benefit from these findings while formulating policy recommendations and strategies to support University teaching and learning during this and any future pandemics. Universities need to plan the post-pandemic education and research strategies to ensure student learning outcomes and standards of educational quality.

Keywords: online; teaching; learning; eLearning; equity; quality; COVID-19; impact

1. INTRODUCTION

Coronavirus disease-2019 (COVID-19), a highly transmittable and pathogenic viral infection caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a rapidly spreading pandemic. The coronavirus disease 2019 (COVID-19) outbreak continues to evolve in the African continent since it was first detected in Algeria on 25 February 2020 (Marelli et al., 2021). As of May 2021, Africa had a total number of 51,000 confirmed cases.

The COVID-19 pandemic has already cost African universities substantial but yet undocumented amount of money. It has cost UK, US and Australian universities an estimated £790 million, US$ 14·6 billion and AUS$ 3.8 billion, respectively. The shutdown has meant that the higher education sector has suffered loss as regards to accommodation, catering, and conference income (Aristovnik et al., 2020; Selvi & Veilatchi, 2021).

Universities worldwide, including those in Africa, have had to quickly scale up online teaching, with concomitant unexpected expenditure. They have had to fund for staff salaries and research projects. Yet, sizeable as the losses are for the current 2020/2021 academic year, they could easily be dwarfed by those expected next year. The economic downturn will force thousands of youngsters to defer entering University. More than 50 million Africans lost their jobs in April 2020 alone. Prospective students might also be put off by the physical distancing requirements that are likely to prevail on University campuses for the foreseeable future. Much will depend on the dynamics of the pandemic.

The shift to online learning looks set to continue at least until the advent of a successful vaccine for COVID-19. Three key issues of importance, in the context of Kenya's vision 2030, Africa's Agenda 2063 and SDG 4, include: Affordable higher education; Increased access to higher education; and Quality education. The disruptive nature of COVID-19 pandemic has a possible significant threat to these three issues.

This research therefore investigated the impact of COVID-19 on University teaching and Learning. In all the institutions where the research was conducted, measures were taken to prevent COVID 19 infections during the beginning of the pandemic. Thus, face-to-face courses were suspended, and then the blended learning of digital and classroom teaching was largely adopted. The laboratories took place in small groups in face-to-face mode while keeping the social distance. Some examinations were conducted face-to-face in small groups, respecting prevention measures; others were conducted virtually. This study attempted to answer questions such as:

(a) How does quality of- and access to- University education relate to online teaching in the crisis conditions?

(b) Are there different opinions among students, faculty and senior management regarding the advantages and disadvantages of online learning?

2. MATERIALS AND METHODS

This section describes the study participants and procedure, measures and sampling, data analysis approach and the ethical considerations made.

2.1 Study Participants and Procedure

The target population comprised University students who were at least 18 years old, lecturers, administrators, and managers. The respondents in the target populations were recruited by convenience sampling facilitated by advertising on University communication systems. The online

questionnaire was designed in English. The web-based survey was launched via the open-source web application on 14 June 2021 and remained open until 2 July 2021.

2.2 Measures and Sampling

The data was obtained through a web-based comprehensive questionnaire composed of ten (10) mainly closed-ended questions, covering socio-demographic, and other characteristics as well as various components of University education, such as cost of teaching learning, quality of e-content, access to internet, quality of curriculum implementation, the roles and measures of institutions, as well as personal reflections on COVID-19. The questionnaire was divided into three broad sections. The first section comprised three (3) questions on the socio-demographic and academic characteristics of the students, e.g., country and institution of study, level and field of study, citizenship, age, and gender. The second section asked students about the implementation of curriculum in their respective universities, as well as their own performance and expectations. This was followed by a segment covering the infrastructure and skills for studying from home, including one open ended question on what needs to be done differently.

The theoretical approach of the study was implicit, using the transformative learning theory. There was the qualitative part characterized by open ended items in the questionnaire. The quantitative part of the survey constituted the other component. Assessment of the online teaching–learning processes were done through the qualitative methods, while the online teaching–learning outcomes were assessed by quantitative methods (Popa et al., 2020; Zhu & Liu, 2020).

Purposive sampling was employed. A total of 1,236 responses from 18 universities from Kenya and Nigeria were obtained during the study period spanning June to July 2021. In addition, a survey was conducted in June 2021, targeting Open Distance and electronic Learning (ODeL) and ICT Directors of the eleven (11) universities participating. Exploratory research design was therefore adopted.

2.3 Data Analysis

The data preparation, aggregation, and cleaning process were performed in MS-Excel. We report the students' gender, citizenship, status, field of study, quality of e-content, cost of teaching and learning, policy issues, and what need to be done differently.

To test statistical hypotheses, Excel was used. The results of descriptive and testing the hypotheses are reported as follows:

(a) Information on students, lecturers, administrators and managers reported the highest mean value of the analysed aspect; the range (difference between the highest and the lowest mean across all groups); and the significance of the differences.

(b) To analyse which factors influence student satisfaction with the role of their University, descriptive statistical analysis was conducted.

2.4 Ethical Considerations

All participants were informed about the details of the study. Study participation was anonymous and voluntary, and respondents could withdraw from the study without any consequences. For data-protection reasons, the online survey was open to people aged 18 or over and enrolled in a higher education institution. Only the researchers had access to the research data. The procedures of this study complied with the provisions of the Declaration of Helsinki regarding research on human participants. The National Commission for Science and Technology (NACOSTI) approved this study, through license no. NACOSTI/P/21/11161. NACOSTI is the body charged with the responsibility of regulating research and Quality Assurance in the Science, Technology and Innovation sector in Kenya.

3. RESULTS

The listing of institutions represented in the study population are shown in Table 1 below. Majority of the sample of 1,236 were Kenyan respondents with 10% of them being from private universities.

University	**Count**	**% of Total Respondents**
Chuka University	1	0.1
Daystar University	106	8.6
Dedan Kimathi University of Technology	35	2.8
Garissa University	26	2.1
Gretsa Universtiy	1	0.1
Inoorero University	1	0.1
Jomo Kenyatta University of Agriculture and Technology	277	22.4
Kenyatta University	406	32.8
Kibabii University	17	1.4
Kisii University	1	0.1
Pwani University	17	1.4
South Eastern Kenya University	48	3.9
Technical University of Mombasa	1	0.1
United States International University	3	0.2
University of Ibadan	1	0.1
University of Kabianga	80	6.5
University of Nairobi	212	17.2
Zetech University	2	0.2
Grand Total	**1236**	**100**

On the other hand, Figure 3.1 below indicates that students constituted the majority of the respondents comprising 67% followed by Lecturers 19.6%, Administrators 9.5% and lastly Management comprising 2.9% of the total respondents. This represented the study population well given that students are the majority of any given University community.

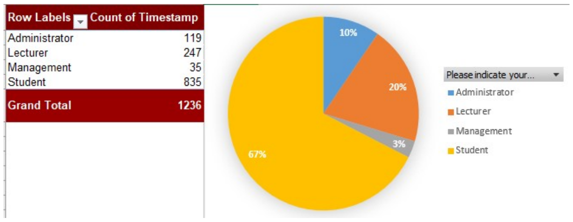

Figure 3.1. Demographic characteristics of participants

It is important to note that the study paid special attention to some eleven (11) Universities that are in a consortium that is participating in a pipeline project seeking support for online learning and requisite infrastructure. These are; Kenyatta University (KU), Daystar University, University of Kabianga (UOK), University of Nairobi (UON), Jomo Kenyatta University of Science and Technology (JKUAT), South Eastern Kenya University (SEKU), Zetech University, Kibabii University, Pwani University, Dedan Kimathi (DeKUT) and Garissa University. Of these Daystar and Zetech are private Universities in Kenya. That is the reason the number of respondents from these Kenyan Universities was relatively higher compared to the rest.

3.1 Overview of the Questionnaire Results

The results of the continental survey includes findings concerning different aspects of teaching and learning such as quality, access and affordability of University education (Aristovnik et al., 2020).

3.2 Quality issues

Universities around the world, Africa included, cancelled their face-to-face classes and shifted their teaching and learning processes, and some research activities, to online media. For some Universities the online mode of teaching was not new, unlike others who were encountering the mode for the first time. The transition was sudden with inadequate lead time for adequate planning, given that the quality of teaching and learning in these new circumstances needs proper attention (Zhu & Liu, 2020). On the other hand, students from marginalized and remote areas had problems with poor Internet connectivity or even lack of electricity. Further, poverty was also a factor, leading to a negative attitude towards the online mode (Shenoy et al., 2020). Nevertheless, 49.7% the students were satisfied with the quality of curriculum implementation (Figure 3.2).

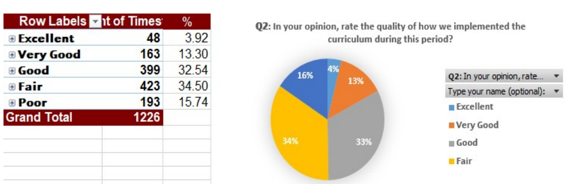

Figure 3.2. Rating of quality of curriculum implementation

The effectiveness of online learning depends on the designed and prepared learning materials, the lecturer's engagement in the online environment, and lecturer–student or student–student interactions (Rashid & Yadav, 2020; Zhu & Liu, 2020). Further, while studying online from home, students must have an opportunity to ask questions and expect timely answers. Therefore, in the context of quality of education, students were asked about the quality of the content provided online. The students agreed that the quality of e-content was important (Figure 3.3).

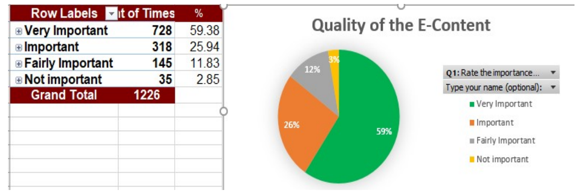

Figure 3.3. Extent to which students' expectations were met

About 44% of the student respondents indicated that their expectations were met (Figure 3.3). This is a pointer to the need for concerted efforts to ensure that the identified issues by students, for improved performance, are addressed.

Virtual labs and studios constituted an important issue of interest. Results from a survey conducted in June 2021, targeting ODeL directors and ICT directors of the eleven (11) universities participating in the pipeline project, revealed that the Kenya Education Network (KENET) is the main internet service provider. Surprisingly, only 18.8% of the institutions had virtual (simulation-based) labs. Internet problems, lack of training and difficulty in following instructions emerged as the major issues in using virtual and remote labs. Remote teaching classrooms and studios were almost non-existent (6.7%) in the participating institutions with Wacom and Digitizer tablets the most common multimedia authoring tools. A majority of these institutions (73.3%) do not have infrastructures for differently-abled learners.

3.3 Access and Affordability issues

Analysis of data on students' admission and registration revealed a decreasing trend in Admission and Registration of freshmen and women in the period 2018-2020 (see Figure 3.4 below). However, despite their registration, COVID-19 greatly affected curriculum delivery due to disrupted student progression. Therefore, access is declining. The gains which have been made over the years are threatened to be eroded by the COVID-19 pandemic.

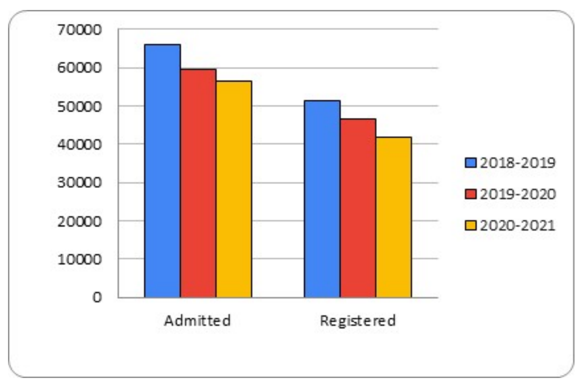

Figure 3.4. Trends on students' admission and registration

Analysis of an open-ended question on what needs to be done differently revealed that over half of the respondents, based on commonly used words from this question, indicated that they have issues with internet access (Figure 3.5).

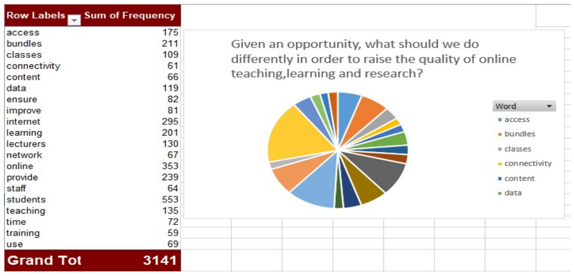

Figure 3.5. The future outlook

Words mentioned that essentially mean the same issue of Internet Access include access, bundles, classes attendance, connectivity, content and data as shown in Figure 3.5 above.

4. DISCUSSION

While the world, Africa included, was facing up to the outbreak of COVID-19 pandemic, higher education institutions were crucially affected at their core: the students. For them, the period was undoubtedly unprecedented and very stressful as onsite classes were moved online, semesters were postponed and examinations adjusted, among other measures. Accordingly, there is an urgent need for in-depth studies on how the pandemic crisis has impacted students' lives around the world. The paper by Aristovnik et al., 2020 was the first large-scale global survey among students from different study perspectives since the COVID-19 outbreak. The current study provides premier insights about the impacts of COVID-19 on teaching and learning in Africa, and Kenya in particular. There is need to improve the quality, access and affordability of online teaching and learning in African Universities. This is in agreement with the study by Aristovnik et. al., 2020, conducted between 5 May and 15 June 2020. An attempt was made to illustrate what student life looked like during the COVID-19 pandemic from academic, social, emotional, financial, and other perspectives. In this respect, a number of insights into student life during the lockdown period were observed. First, the students' academic work and academic life aspects were studied. Due to the physical closure of higher education institutions, the majority of teaching and learning processes went online, that is, 86.7% of all respondents claimed that their onsite classes had been cancelled and substituted with online lectures in the form of real-time video conferences, sending presentations to students, video recordings, and written communication in terms of forums and chats. The students were the most satisfied with real-time video conferences, video recordings, and written communication, with Oceania and Europe emerging as global frontrunners while developing countries, from Asia and Africa, significantly lagged behind (Aristovnik et al., 2020).

5. CONCLUSION

Quality of curriculum implementation is such that 55% success was achieved in online teaching and learning. There is need for the issues that were expressed by 45% of respondents to be addressed.

Most important issues in e-learning have been identified to be Internet access including devices and electrical power; Quality of e-content; E-content development and Capacity Building .

Implementation challenges faced by universities in Kenya are characterized by institutional, individual and national issues. Central to this is the need for Infrastructural strengthening. In addition, there is a need for policy formulation to address the digital learning in Kenya.

The following are the key recommendations from this study:

(a) Need for adequate support for eLearning infrastructure development for continued delivery of the curricula.

(b) Providing subsidized data bundles and devices to students to maintain/improve access to Internet for learning purposes.

(c) Support to mount virtual labs and studios.

(d) Developing policies, guidelines and processes on eLearning for effective implementation of the teaching programs using technology.

(e) Refocused planning and training by Universities on eLearning for both staff and students.

6. ACKNOWLEDGMENTS

This work was done under the auspices of Education for Sustainable Development in Africa (ESDA) phase 2 project suported by the African Development Bank (AfDB). We thank these institutions for their pivotal role in this study. A number of Universities and colleagues also played a part and we are indebted to them. We thank the eleven participating Universities in the consortium and their representatives in the project. These representatives coordinated the data collection exercise and we single them out, with much gratitude, within their respective Universities. For Daystar University we had Dr. Caroline Ayuya, Mr. Samson Mburu and Dr. Sylvia Tuikong; Dedan Kimathi University of Technology: Prof. G K Muthakia, Dr. John M. Wandeto, Mr. Muchiri Michael Njoki and Dr. Moses Ollengo, Garissa University: Dr. Aden Ali Abdi and Mr. Felix Wasike, Jomo Kenyatta University of Agriculture and Technology: Prof. Urbanus N. Mutwiwa, Prof Rebecca W. Waihenya, Prof. Stephen Kimani and Dr. Kipchumba Cherono, Kenyatta University: Dr John Muriuki and Anthony Mbaabu, Kibabii University: Dr. John Wanjala Makokha, Dr. Anselemo Peters Ikoha, Dr. Alice Nambiro Wechuli, Prof. Samwel Mungai Mbugua, Dr. Ben Nyongesa Wekesa, Pwani University: Dr. Leornard Kiti Alii, Dr. Fred Namasaka and Dr. Rose Kigathi, South Eastern Kenya University: Dr. Andrew Omala, Kim Musyoka Musila and Dr. David Musembi, University of Kabianga: Dr. Patrick Bii and Mr. Geoffrey Sowek, University of Nairobi: Prof. Christopher Gakuu and Mr. Paul Kariuki and Zetech University: Dr. Alice Njuguna, Dr. Catherine Njoki and Cyrus Baraza.

7. REFERENCES

[1] Aristovnik, A., Keržič, D., Ravšelj, D., Tomaževič, N., & Umek, L. 2020. Impacts of the COVID-19 Pandemic on Life of Higher Education Students: A Global Perspective. Sustainability 2020, 12(20), 8438; https://doi.org/10.3390/su12208438.

[2] Marelli, S., Castelnuovo, A., Somma, A., Castronovo, V., Mombelli, S., Bottoni, D., Leitner, C., Fossati, A., & Ferini-Strambi, L. 2021. Impact of COVID-19 lockdown on sleep quality in University students and administration staff. Journal of Neurology, 268(1), 8–15. https://doi.org/10.1007/s00415-020-10056-6

[3] Popa, D., Repanovici, A., Lupu, D., Norel, M., & Coman, C. 2020. Using Mixed Methods to Understand Teaching and Learning in COVID 19 Times. Sustainability 2020, 12(20), 8726. https://doi.org/10.3390/su12208726

[4] Rashid, S., and Yadav, S. S. 2020. "Impact of Covid-19 Pandemic on Higher Education and Research". *Indian Journal of Human Development.* 14(2): 340–343. https://doi.org/10.1177/0973703020946700

[5] Selvi, D. V. D., & Veilatchi, K. (2021). "Economic Impact of COVID-19." *BSSS Journal of Commerce.* XIII(1), 15

[6] Shenoy, Veena, Sheetal Mahendra and Navita Vijay. APRIL, 2020. "COVID 19 – Lockdown: Technology Adaption, Teaching, Learning, Students Engagement and Faculty Experience". *Mukt Shabd Journal.* IX(IV). ISSN NO : 2347-3150

[7] Zhu, X., and Liu, J. 2020. "Education in and After Covid-19: Immediate Responses and Long-Term Visions". *Postdigital Science and Education*, 2(3), 695–699. https://doi.org/10.1007/s42438-020-00126-3