IOSR Journal Of Humanities And Social Science (IOSR-JHSS)
Volume 27, Issue 6, Series 2 (June. 2022) 38-45
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Factors Influencing Diploma and Certificate Student's Progression in Kenyan Universities

Stanley Munga Ngigi
Tutorial Fellow, Zetech University, Nairobi, Kenya.

Abstract
The most important priority of a private academic university is financial stability, which is determined by 100% student progression to the next level of study. Poor student advancement will result in the university's demise, as student progression is the primary source of revenue for a private university. The institution can plan adequately for the next semester and determine the number of workers required without undue stress as a result of the students' progress. The study found that economic factors were most significant. Availability of financial resources should therefore be given a priority by stakeholders to ensure an improved progression of the students.
Keywords: Progression, Economic, Demographics, Diploma, Certificate.

Date of Submission: 01-06-2022 Date of Acceptance: 13-06-2022

I. Introduction

Diploma and Certificate students' progression is a major concern to guardians, sponsors, and university administrators, the majority of the students enroll in the universities for their diploma or certificate courses intending to progress until to the degree level. However, in many instances, this has not been the case. Majority of these students after completing their diploma or certificate courses they do not progress from certificate to diploma and diploma to degree (Gibbs, 2004). Students encounter a variety of challenges, including a lack of government funding by the Higher education loans board (HELB), lack of finances to further their academics until completion, missing marks from the respective department, geographical factors such as distance making it difficult for students to afford transport cost, job vows and health matters are challenges are some of the challenges or difficulties that students face among many others (Hayden, 2012). The most important mission of universities as academic institutions is to ensure that there is a one hundred percent progression of all the students enrolled for diploma and certificate courses because this will help the university maintain its population and generate income since the students are the major source of income.

The advancement of diploma and certificate students is a significant source of worry in educational policymaking sectors (Demetriou & SchmitzSciborski, 2011; Tinto, 2006). Around 30% of students enrolled in diploma and certificate programs around the world do not continue to pursue their degrees. In Kenyan universities, a similar problem of low diploma student development exists. Therefore, the study examined the factors influencing diploma and certificate student's progression in university

II. Literature Review

Motivation

In studies, motivation is important because they rely on self-direction and self-learning. Motivation, or a lack thereof, might result in a student's slow growth. (2016, Bawa). (Eric, 2010) student achievement is closely linked to accountability and motivation. He explains that interest in classes is strongly related to student aptitude and attitude toward learning. (K. L. Smart & J. J. Cappell,2006) Time required to complete a module, a lack of real-world examples, problems accessing a resource, and support systems are all shown to be key sources of motivating limitations.

Economics

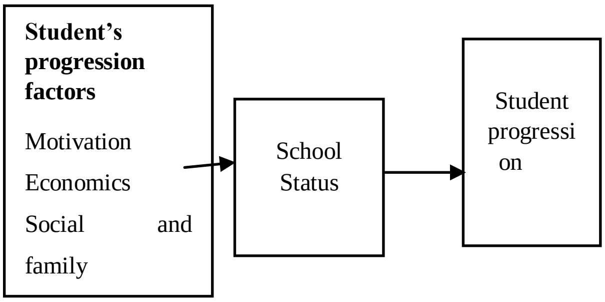
Face to face owing to the need for physical class attendance, which necessitates the use of time and travel expenses, among other things. According to research performed in Pakistan, the most common reason for dropout was a failure to pay fees on time (Darakhshan Muslim, Syed Muhhammad, Syeda Aneeqa, 2017).Learner dropout is heavily influenced by the financial difficulties of some families and the scarcity of financial aid in eLearning courses. Most governments do not subsidize diploma and certificate programs, and thus do not provide financial aid to students at lower levels.

Social and family

Home is the location where the foundation of learning and education takes place. To obtain good academic outcomes that contribute to kids' advancement, parents, children, and other family members must promote a learning environment in their homes. Parents are responsible for aiding pupils who are having difficulties in certain classes, for example. This help could be provided in the form of private tutoring. They equip kids with technology and other learning materials to help them enhance their academic performance at home. Parents have an important role in their children's development and growth (Kudari, 2016). Children frequently talk to their parents about any problems they are having in school, whether academic or otherwise. Parents offer their children security, encouragement, and aid in resolving their problems.

Demographics

Learner dropout is influenced by a variety of personal factors, including age and gender. (Darakhshan Muslim, Syed Muhhammad, Syeda Aneeqa, 2017) According to the author, some familial considerations, such as marriage, have an impact on female students and may cause them to drop out of classes. The student's age may play a role in dropout since younger students are more familiar with technology than older students, allowing them to use the online platform more simply.



Dependent variables independent variable

III. Research Methodology

The word "research design" refers to a plan that directs each study and assists in data collection, analysis, and interpretation of results. It can also be utilized by researchers as a template for deciding on methods and instruments to utilize in data gathering and analysis to answer the research questions (Cooper & Schindle, 2014). The Fayaz model for data mining was used in this investigation, as illustrated in Figure 3.

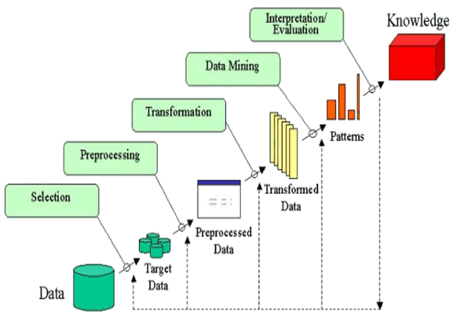


Figure 1: A model design approach for the study
Source: Fayyad et al. (1996)

3.1 Selection of Data

The entire student population was used in the study; it was necessary to conduct a census on the student population in the private universities in Kenya. The study worked with a representative dataset by doing data reduction after acquiring data from the university data source and saving it in an excel spreadsheet. As part of the numerosity reduction process, the data was replaced or estimated using alternative, smaller data representation methods such as means, clustering, sampling, and histograms. The data was then cleaned up by the researcher. The practice of discovering and eliminating faulty or erroneous records from a database is known as data cleansing. The binning technique was used to complete this operation. Binning methods were used to sort data value by checking its "neighborhood," or the values in its immediate vicinity. The sorted values were divided into several "buckets" or bins. The data was then divided into equal frequency bins before being smoothed using means. According to the research, several Private Universities in Kenya are divided into three faculties: ICT, hospitality, education, and business. The data from the faculties were used in the study using a convenience sampling methodology. This accounted for a third of the faculties and, as a result, 33% of the population. Mugenda and Mugenda (2013) claim that 10% of the population is indicative of the entire population. According to the data in the database, there will be occasions where diploma and certificate students earned university placement but did not enroll in their studies. There were instances where students entered in the second year, indicating that some of the students had previously completed diploma courses and qualified to participate in second-year university degree programs. As a result, the study undertook data cleaning to protect the students by filling in the missing first-year records with first-year enrolment averages. Due to repeated registration numbers, the database also featured a mix of registration numbers, with student information appearing more than once in the database. Dimension reduction was used to deal with this as well. Data reduction was carried out to improve the efficiency with which the data was handled for analysis. After that, the data was divided into training and test groups. Pesaran, Pick, & Timmermann (2011), as well as Andic &Ogunc (2012), have successfully used this strategy of sample splitting (2015).

3.2 Data Pre-processing and Transformation

Identifying the data to be used in training the model, testing the model, and assessing the output error with training and validation data are all part of this process. The type of data that enters into each data set, as established by the research sampling technique, is critical. The population that was identified was represented by the sample taken. The data was translated into three independent comma-delimited (CSV) format files, each providing information on student enrolling, deferment, and dropout. The student ids were used to link all three of these files. Following the preparation of data input files, data cleaning were carried out as follows:

i. Conversion of data types (character to numeric, character to factor)
ii. Aggregating data down to the student id level
iii. Identifying and deleting outliers as needed
iv. Removing or cleaning up missing values
v. Combining multiple datasets into a single one

The following factors datasets were used to determine the model efficacy through preprocessing. They include:
Motivation, Economic factor, social and family factor, and demographics.

After reading the CSV files, as the first step of the data preparation process, all the character type data were transformed into numeric data.

```
Motivation<- read.csv (file. Choose (), header= TRUE)
Economic<-read.csv (file. Choose (), header=TRUE)
Social and family<-read.csv (file. Choose (), header=TRUE)
Demographics<-read.csv (file. Choose (), header=TRUE)
Myfile<-as.data. Frame (Motivation, Economic, Social and Family, Demographics)
Attach (Myfile)
Myfile Motivation, Economic, Social and Family, Demographics
```

3.3 Data mining

The study will feed motivation, economic, social and family, and demographics to the model for processing, the model will be trained on the type of input data and the projected output of the training session. Data will be saved in an excel data sheet, which was converted to a CSV file (comma delimited) and read into the WEKA analysis tool.

When using the WEKA analyzing tool, the data was changed to an attribute relation file format (ARFF) file for ease of analysis. After reading the data into the WEKA Analyzing tool, the next step was to scale the data interval to ensure efficiency in working with the data. The data was then separated into training data and data sets to determine the efficiency in data collected, the efficacy on the model was determined in percentages of 100% preferred [70:30]. According to, seventy percent (70%) of the source data is utilized to train the model, whereas thirty percent (30%) was used to test the model (Minewiskan, 2018). The training data was then into the ANN model via the model neurons specified, and the process was repeated in several trials until the artificial neural network model converges on the correct efficacy.

3.4 Model Evaluation

Model evaluation is a process of ensuring that the system meets the user's requirement by observing the actual model output and the evaluated output model. It involved the use of several trials to find the accuracy of the model and validated to ascertain that it meets the user's requirement on the progression of the students in the university.

The model was evaluated by calculating the measure of performance through the use of root mean squared error and the mean absolute error. These measures were calculated against the actual recorded progression rate values in comparison with the progression rate values forecasted by the model over the same period. The model was run in several trials and the average accuracy of the model was determined. A confusion matrix was obtained to show the various performance measures and the accuracy of the model. The output of the results was displayed at the Knowledge gap. Several performance measures were conducted in this study.

3.5 Knowledge representation

Knowledge representation is a technique for visually representing data to the user. This included data that has been mined for information. Different strategies were used to generate the output of the model to ensure that the efficacy of the model is met.

3.6 Data Collection

The data was collected from the private university database system which involved several private universities in Kenya ranging from 2007 to 2019 in all semesters. It involved the raw data which was cleaned to get the right trend of student progression involving all attributes defined by this study.

The study used a stratified sampling method since it involved creating a section or group of population. In this study, it was put into consideration a certificate group and a diploma group which was used to determine the progression rate of one level to another. When using the stratified sampling method, it is easy to quantify because there is a clear number of users in each group.

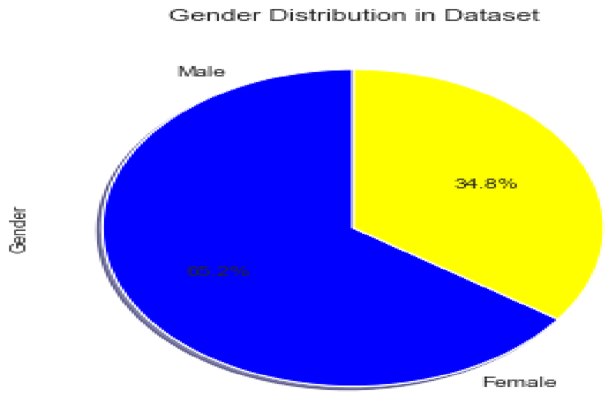
This method of sampling was adopted because it assisted in targeting a certain group of students from certificate to diploma and diploma to degree since in our Kenyan economy people who are dominating in the job market are degree holders thus bringing into the attention of making sure that there is a 100% students progression to reduce the gap of employment and job satisfaction to the community who need the service of the graduates. To get the sample it involved several private universities in Kenya from several faculties and to get the respondents the study used already collected and stored data in the database. Database administrators played a very great role in providing the raw data that assisted in sampling. This study used data that had already been stored in the database for all the sampled private universities in Kenya.

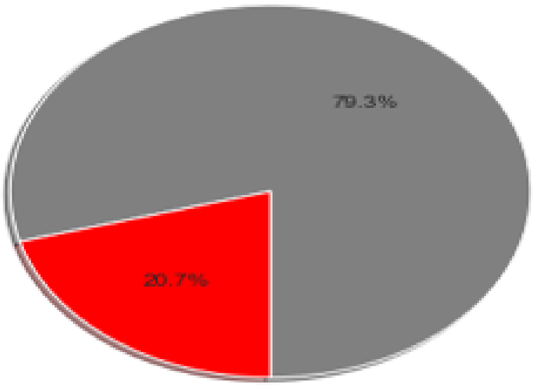
IV. Research Findings

4.1 Demographic Information on students Dataset

In the analysis, we focused mainly on the variables of the study during the period. Pre-processing was a crucial part to be done at the very beginning of any data science project. It included dealing with null values, detecting outliers, removing irrelevant columns through analysis, and cleaning the data in general. The data set used contained (4000) rows with four (4) columns. The four (4) attributes consisted of cost, motivation, social & family, demographics.

Participants in the study





The data was also normally distributed and hence appropriate for further analysis. This was evidenced by the histograms which showed balanced bins.

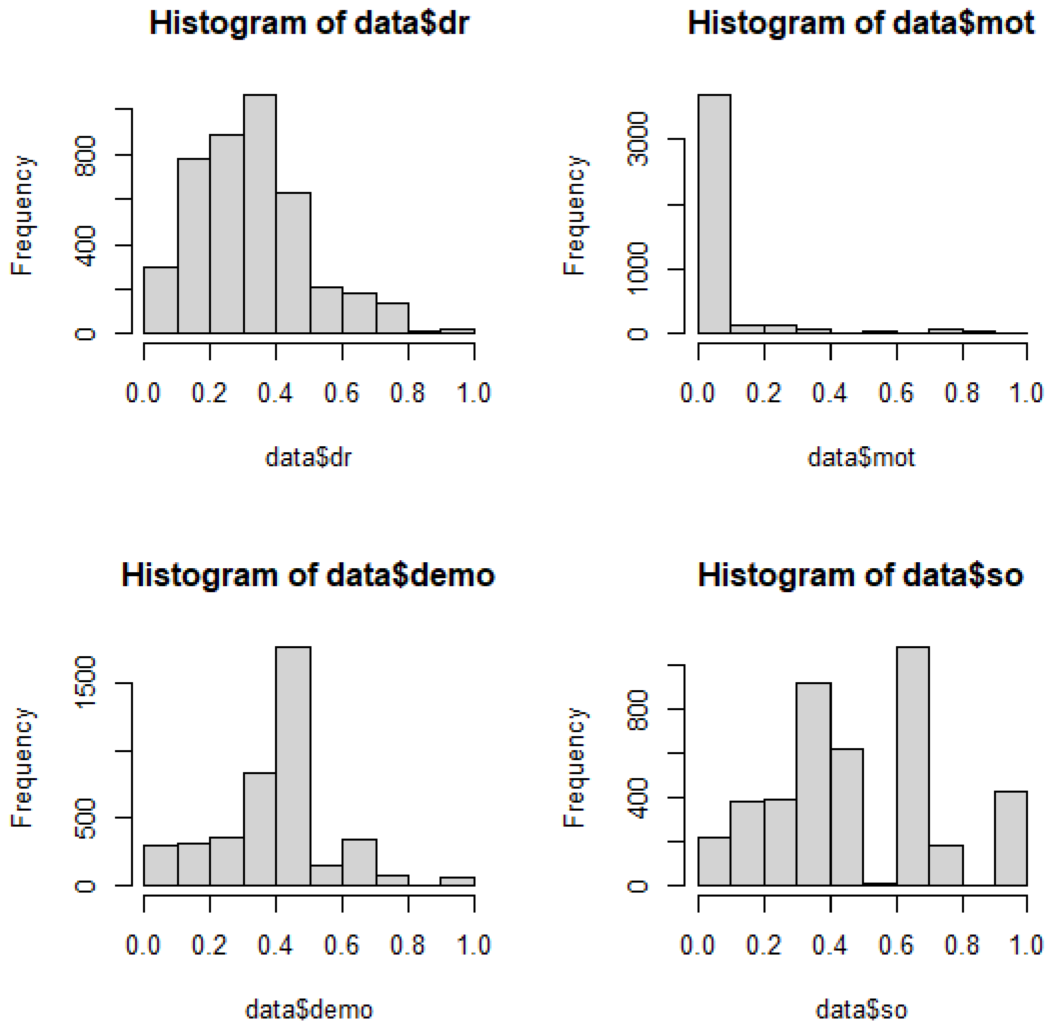


Figure 8:Histograms Results

Figure 4.1 Depiction of the Distribution of the Data

Where dr= Cost, mot=motivation, demo= demographics, so=social and family.

4.2 Factors influencing progression rate of students-Objective one

The researcher was interested in investigating the factors that influence the progression rate of students. The use of feature selection using Cramer's V Correlation was applied to the dataset to explore how the variables correlate with each other.

The use of Cramer's V Correlation to identify how the features or variables correlate with each other or if the variables increase at the same time they are said to correlate otherwise inversely if one variable increases while the other decreases, they anti-correlate. Cramer's V Correlation is similar to the Pearson Correlation coefficient. While the Pearson correlation is used to test the strength of linear relationships, Cramer's V correlation is used to calculate correlation in tables with more than 2 x 2 columns and rows. Cramer's V correlation varies between 0 and 1. A value close to 0 means that there is very little association between the variables. A Cramer's V correlation value of close to 1 indicates a very strong association.

Cramer's V	**Relationship**
0.25 or higher	Very strong relationship
0.15 to 0.25	Strong relationship
0.11 to 0.15	Moderate relationship
0.06 to 0.10	Weak relationship
0.01 to 0.05	No or negligible relationship

Table 1: Cramer's V Correlation

A co-efficient close to 1 means that there's a very strong positive correlation between the two variables. In our case, the blue shows very strong correlations. The diagonal line is the correlation of the variables to themselves thus, they will be 1. The correlation heat map of various variables is shown in Figure 4.5 below.
After checking the correlation matrix, it was observed that there are several attributes with high correlation.

Attributes	LEFT
Demographic factors	0.55
Social and family	0.41
Cost	0.36
Motivation	0.24

Table 2:Attributes Relationship

From the Correlation Matrix above, several attributes were more significant than others were. The revised framework is as in Figure 4.5 below. The key factors affecting the progression rate of students are motivation, cost, social and family, and demographic factors.

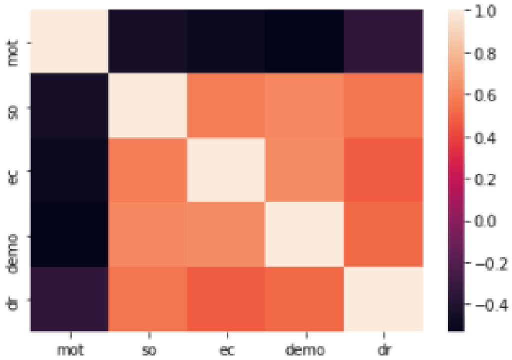


Figure 2:Revised Framework

4.3Findings based on the objectives of the study

The objective was to investigate and identify factors that determine certificate and diploma university students' progression rates. Four factors were discovered to be relevant in determining student progression rates. These were the students' demographic, financial, motivational, and social aspects. Table 3 shows the outcomes for objective one.

Table 3: Objective one results

Variable	Beta coefficient	Significance	Remark
Intercept	−3.8461		
Demographic	0.5973	0.000	Significant
Social and family	0.938	0.000	Significant
Cost	-1.8019	0.000	Significant
Motivation	0.8928	0.001	Significant

The study used the relevant data to obtain regression coefficients for diploma and certificate student progression rates in Kenyan private universities. The artificial neural network was configured to obtain a combination of demographic, social, and family, cost, and motivation data from the students. The regression analysis showed that demographic achieved alpha of 0.5973 percent which was significant at the 5 percent level while social and family achieved 0.938%, cost achieved 1.8% and motivation achieved 0.89%.

V. Conclusions and Recommendations

5.1 Conclusions

This study made important contributions to the knowledge base on certificate and diploma students' progression in the Kenyan private universities, which are currently facing a crisis of student numbers. The study concluded that diploma and Certificate student's progression is a major concern to guardians, sponsors, and university administrators, majority of the students enroll to the universities for their diploma or certificate courses to progress until to the degree level. Therefore, to identify significant predictors that can be used to predict the progression of these students to the next level of the study among the factors studied were motivation, economics, social and family, and demographic factors. The study found that economic factors were most significant. Availability of financial resources should therefore be given a priority by stakeholders to ensure an improved progression of the students.

5.2 Recommendations

This study brought interesting findings that can be believed to bring a positive change if implemented. First learning of students touches the heart of everyone, be it, the government, the parents, stakeholders, and even managers of the universities. Therefore, interesting comments can be borrowed from this study.

The identification of student predictor characteristics that are crucial in predictability accuracy is an important follow-up to this work. This is useful for three reasons. First, the government can intervene directly if they know why private university certificate and diploma scholars fail to proceed to the next study year and at the stipulated period frame, in conjunction with the university administrators as well as the Ministry of education. Future research could be performed to improve learning behavior and improve attributions. Academic management boards can provide data on learning behavior to help with input attributions for the predictive model, allowing for faster progression.

The government should provide more resources to help increase the number of students enrolling in the university and reduce the rate of deferment of students due to lack of financial aid. University management boards should pay keen attention to the matters to help discover leading causes of deferment and drop out of students. The outcomes of this research showed deferment of students' certificate and diploma students is on increase. A separate study can be done to help identify the main problem that leads to the increasing deferment of students.

The voices of students are critical components in educating educators about the complicated topic of dropout. This study demonstrates that early progress monitoring, academic support, and a safe and welcoming learning environment are all necessary for transformation to occur. If educators pay attention to the words of dropouts, they will be better able to prevent pupils from leaving. Students' voices are a valuable resource for educators looking for solutions to the dropout epidemic since their responses may result in stronger supports for those on the verge of dropping out. The findings suggest that progress should be monitored frequently that early communication between all stakeholders is established, that academic support is increased, and that safe and inviting learning settings be created. These findings suggest that other at-risk students will have a bright future.

Learner dropout is influenced by a variety of personal factors, including age and gender. some familial considerations, such as marriage, have an impact on female students and may cause them to drop out of classes and are therefore not able to proceed with their studies. The student's age may play a role in dropout since younger students are more familiar with technology than older students, allowing them to use the online platform more simply. Modern education is becoming more and more digital. Stakeholders should therefore invest more in the training of the disadvantaged group of students instead of assuming that they will learn on their own. In many instances, some groups of students are not able to make presentations via digital platforms such as KENET and zoom. This makes them fail in their assessments. The management should therefore consider revising their assessment methods.

References

[1] Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of educational data mining*, *1*(1), 3-17.

[2] Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). Springer, New York, NY.

[3] Demetriou, C., & Schmitz-Sciborski, A. (2011, November). Integration, motivation, strengths, and optimism: Retention theories past, present and future. In *Proceedings of the 7th National Symposium on student retention* (Vol. 201).

[4] Dissanayake, H. U. (2016). Predicting Student Retention: A Comparative Study of Predictive Models for Predicting Student Retention at St. Cloud State University.

[5] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.

[6] Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996, August). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *KDD* (Vol. 96, pp. 82-88).

[7] Ian, H. W., & Eibe, F. (2005). Data Mining: Practical machine learning tools and techniques.

[8] Jung, H. W., Varkoi, T., & McBride, T. (2014, November). Constructing process measurement scales using the ISO/IEC 330xx family of standards. In *International Conference on Software Process Improvement and Capability Determination* (pp. 1-11). Springer, Cham.

[9] Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003, September). Preventing student dropout in distance learning using machine learning techniques. In *International conference on knowledge-based and intelligent information and engineering systems* (pp. 267-274). Springer, Berlin, Heidelberg.

[10] Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, *103*, 1-15.

[11] Nghe, N. T., Janecek, P., & Haddawy, P. (2007, October). A comparative analysis of techniques for predicting academic performance. In *2007 37th annual frontiers in education conference-global engineering: knowledge without borders, opportunities without passports* (pp. T2G-7). IEEE.

[12] Niazi, AK, Kamran, M., Tariq, Z., Malik, M., Ilyas, K., & Zaman, MT Statistics Industrial Perceptive Sales Analysis using Data mining.

[13] Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27.

[14] Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368-384.

[15] Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30.

[16] Siemens, G., & Baker, R. S. D. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254).

[17] Sisimwo, J. (2016). *Electronic resources and its application in collection development practices in academic libraries: the case of United States International University* (Doctoral dissertation, University of Nairobi).

[18] Tinto, V. (2006). Research and practice of student retention: What next?. *Journal of college student retention: Research, Theory & Practice*, *8*(1), 1-19.

[19] Sisimwo, J. (2016). *Electronic resources and its application in collection development practices in academic libraries: the case of United States International University* (Doctoral dissertation, University of Nairobi), 1(1),5-20.

[20] Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., ... & He, C. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell*, *49*(1), 18-29.

[21] Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368-384.

Stanley Munga Ngigi. "Factors Influencing Diploma and Certificate Student's Progression in Kenyan Universities." *IOSR Journal of Humanities and Social Science (IOSR-JHSS)*, 27(06), 2022, pp. 38-45.