I.J. Modern Education and Computer Science, **2019, 12, 1-13**
Published Online December 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2019.12.01

Visual Association Analytics Approach to Predictive Modelling of Students' Academic Performance

Udoinyang G. Inyang
Department of Computer Science, University of Uyo, Nigeria
Email: udoinyanginyang@uniuyo.edu.ng,udoiinyang@yahoo.com

Imo J. Eyoh
Department of Computer Science, University of Uyo, Nigeria
Email: imoeyo@gmail.com

Samuel A. Robinson
Department of Computer Science, University of Uyo, Nigeria
Email: Samulerobinson@uniuyo.edu.ng

Edward N. Udo
Department of Computer Science, University of Uyo, Nigeria
Email: edwardudo@uniuyo.edu.ng

Received: 19 August 2019; Accepted: 19 October 2019; Published: 08 December 2019

Abstract—Persistent and quality graduation rates of students are increasingly important indicators of progressive and effective educational institutions. Timely analysis of students' data to guide instructors in the provision of academic interventions to students who are at risk of performing poorly in their courses or dropout is vital for academic achievement. In addition there is need for performance attributes relationship mining for the generation of comprehensible patterns. However, there is dearth in pieces of knowledge relating to predicting students' performance from patterns. This therefore paper adopts hierarchical cluster analysis (HCA) to analyze students' performance dataset for the discovery of optimal number of fail courses clusters and partitioning of the courses into groups, and association rule mining for the extraction of interesting course-status association. Agglomerative HCA with Ward's linkage method produced the best clustering structure (five clusters) with a coefficient of 92% and silhouette width 0.57. Apriori algorithm with support (0.5%), confidence (80%) and lift (1) thresholds were used in the extraction of rules with student's status as consequent. Out of the twenty one courses offered by students in the first year, seven courses frequently occur together as failed courses, and their impact on the respective students' performance status were assessed in the rules. It is conjectured that early intervention by the instructors and management of educational activities on these seven courses will increase the students' learning outcomes leading to increased graduation rate at minimum course duration, which is the overarching objective of higher educational institutions. As further work, the integration of other machine learning and nature inspired tools for the adaptive learning and optimization of rules respectively would be performed.

Index Terms—Association Rule Mining, Predictive analytics, students' performance, hierarchal clustering, at-risk students

I. Introduction

The increasing dependency on technology and methods that are driven by information technology by academic institutions has accounted for the abundance of huge educational data repositories. Moreso, educators and educational administrators have intensified their efforts towards collecting and storing data providing information on the functionality of their educational systems. These repositories have the capacity of storing a large amount of student-related data and information. Students' data and information are key requirements in educational learning systems, especially in the planning, monitoring, and assessment of educational management systems (EMS). EMS is a platform for data acquisition and collection, validation and processing, analysis and communication of information relating to administrators, students, teachers, staff and infrastructure in the educational environment [1]. EMS is a rapidly progressing field of data mining which concentrates on the search and discovery of interestingly new patterns, techniques, tools, and models for intelligent exploratory analysis and visualization of large educational dataset. EMS is aimed at the extraction of novel and interpretable structures that will enhance comprehensibility of students, their processes and environments [2,3]. Among the important modules of EMS are students' management (SM), human resource, infrastructure management, school management, and graduands' management module. The SM module captures and stores students' data with a unique student enrolment number as the primary key, demographic data, academic status amongst others. Associated with EMS, is learning analytics (LA) and educational data mining (EDM), concerned with the exploratory analysis of the educational dataset and utilizing the outcome directly on the students, teachers and other components in the learning process. EDM and LA try to interpret how students cope and interact with educational resources at their disposal, their learning behavioural patterns, likely final academic outcome, and most importantly, their likelihood of completing their program within the minimum stipulated timeframe. EDM depends more on methods, tools, and techniques while LA focuses on the description of data, knowledge and resultant patterns. Machine learning techniques, statistics, visual analytics, link analysis, opinion, and sentiment analysis are some of the widely used tools of LA while classification, clustering, Bayesian modeling, relationship mining, discovery with models and predictive/prescriptive modeling are often associated with EDM [4]. Predictive analytics employs a range of statistical approaches ranging from machine learning and predictive modelling to data mining to competently analyze the historical and operational data and information to enable predictions about the unidentified future event. Its application cut across several domains including academic performance prediction.

Academic achievement is a key factor considered by recruiting organizations and motivates the monitoring of students' performance during their academic pursuits. Students have to work hard for outstanding grades, in order to rise up to the potentials of recruiting organizations and meet the expectations of parents/guardians, educators and administrators. Persistent and quality graduation rates of students are increasingly important indicators of progressive and effective educational institutions. Any educational system characterized by high rates (frequency) of drop-outs (students who leave an institution precipitately without completing the desired programme of study), transfer-outs (students who started in one course of study or one institution and, thereafter move to another course or educational institution to enable him/her graduate) stop-outs (students who voluntarily withdraw and leave for a period of time, and then re-enroll in order to complete their programme), and spill-over (students who spend extra year(s) in due to poor performance) is said to fail [5,6]. The early identification of students' weaknesses during their academic career will guide in the effective provision of necessary pedagogical interventions, suggesting behavioural changes to enhance students' learning processes and also ensure students' on-time and satisfactory graduation [7-9]. However, educational systems in most developing countries) lack facilities for automatic predictions of fail or pass percentages of students and cannot account for the number of drop-outs, stop-outs, transfer-out or spill-over students but rather concentrate more on successful students. They have no information about what patterns lead to these at-risk students and cannot identify students who are likely to struggle in their academics at an early stage of their academic pursuits. In consideration of these challenges, this paper employs EDM and LA methodologies (cluster analysis and association rule mining) to model students' learning processes, for informed decisions and timely pedagogical interventions. Association rule mining [1]. [10] attempts to extract relevant and interesting relationships among items in a database. This paper aims at identifying relationships among courses offered by the students, and the effects of such correlations to learning and academic performance vis-a-vis status. It will also employ cluster analysis to reveal the optimal number of course clusters and their association with student's status at the end of the minimum duration of the programme.

The rest of the paper is organized as follows. Section II presents literature review with emphasis on cluster analysis and performance-course association rule mining. In section III, the methodological framework is conceptualized for the implementation of the system. Course association rule mining procedure and results are described in Section IV. Discussion of results, and conclusions and further work is presented in section V and VI respectively.

II. Literature Review

Learning analytics and EDM can discover and extract trends in data, and also act as a medium for promoting educational activities by identifying and avoiding failure (or poor performance) trends and patterns while exploiting and utilizing success patterns. In Ref. [11], EDM and learning analytics promise to make sustainable impact on learning and teaching to transform slow learners into effective and better learners [12]. Reference [13] points out that learning analytics involves two major operations namely predicting student learning successes and providing proactive feedbacks. Reference [14] proposed a multivariate based method of predicting students' results in learning courses associated with web learning while reference [15] reported that, to make sense of large amounts of educational data, intelligent systems must be developed to automatically process the data and provide reports to stakeholders. In reference [16] a LA dashboard to enhance students' learning performance was developed. The system in reference [16] works by tracking and mining massive online student data and visualizing results so they can be comprehended at a glance. Experimental evaluation indicates that although the LA model did not have a significant impact on student achievement, there was an overall student satisfaction with the dashboard which impacts on students' understanding level.

Reference [13] proposed video annotation as a tool to aid student learning. In their study, self-reflection was translated into grades and association between these grades and the variables used in the study was carried out in the context of their research. Their findings indicate that students make their self-reflection much early in the video timeline. Reference [17] investigated student academic performance using features like family expenditures and students' personal profile gotten from students who are on scholarship in different universities in Pakistan. Analysis of findings show that associating family expenditure with personal information features significantly affect the performance of the students. Reference [18] utilized genetic programming in the modelling of students' centered performance prediction method. Their proposed student prediction performance system leverage on tools from learning analytics, educational datamining and human computer interaction. The data for analysis was obtained from the Virtual Mathematics Teams in a geometry based environment solving session. Experimental analysis indicates that genetic programming-based model is interpretable with an optimized prediction rate compared to traditional prediction algorithms. The increasing drop-out rates in school, students who earn weak classes of degree, and those who exceed the specified duration of programme motivated the work reported in [2]. Their study explored k-means and self-organizing map (SOM) in mining pieces of knowledge relating to the optimal cluster numbers in students' dataset and the correlation between the selected input features; demographic, pre-admission and first year performance. Cluster analysis and association rule mining are among the commonest EDM methods and are adopted in this work.

A. Cluster Analysis

Cluster analysis (CA) is a statistical approach of splitting observations in a dataset into sub-groups based on similarity or dissimilarity of their features. The sub-groups are created so that each observation is similar to other members of the sub-group such that the within sub-group variance is minimized (low intra-cluster dissimilarity) and variance between the points in the other sub-groups are maximized (high inter-cluster dissimilarity) [19]. Partitional and hierarchical clustering are the two categories of clustering; where partitional clustering algorithm divides a dataset such that every record belongs to a distinct group and hierarchical clustering analysis (HCA) produces a of a tree like structure of clusters, called dendrogram [20] The root of the tree comprises a big singleton cluster having all observations as members while the leaves map to each observation.

HCA methods are classified into two types; divisive (top-down) and agglomerative (bottom-up). Agglomerative clustering (AHC) starts by placing each observation ($x_1,\ldots,x_n$) in a distinct cluster ($c_1, \ldots, c_n$), and iteratively merging the nearest clusters to form successively larger clusters until all the observations are grouped into a single cluster. Divisive algorithms begin by treating all the records in the dataset as a single cluster and iteratively splitting them into smaller sub-clusters. The output is a hierarchical tree or dendrogram. AHC is implemented in two stages; the first focuses on the choice of a distance measure of similarity (or dissimilarity) between the records units (which includes Euclidean, Manhattan/city block, cosine, squared Euclidean and so on); and the second concentrates on choosing a suitable linkage technique for forming cluster — the distance defining the degree of (dis)similarity between any pairs of clusters. The linkage criterion that have been proposed includes; single, average, complete, ward's, centroid, median [21]. The Single linkage (also known as nearest neighbor) returns the smallest value of all pairwise distances between observations in a cluster and those in other clusters as distance between two clusters while complete (maximum, furthest neighbour) linkage adopts the largest value. For any two clusters of failed courses, C_i and C_j, the distance between the two clusters is given in (1) and (2) for single and complete linkages respectively [22,23].

$$f_s(C_i, C_j) = \min_{a\in C_i, b\in C_j} \|a - b\| \tag{1}$$

$$f_c(C_i, C_j) = \max_{a\in C_i, b\in C_j} \|a - b\| \tag{2}$$

where a and b are data points in C_i and C_j correspondingly, and $\|a - b\|$ is the chosen distance measure. Mean (average) linkage computes the mean distance between all pairs of objects in a cluster and the objects in another cluster as the distance separating any two clusters as in (3).

$$f_a(C_i, C_j) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \|a_i - b_j\| \tag{3}$$

where a_i and b_j are data points in C_i and C_j respectively, and $\|a - b\|$ measures Euclidean distance and N and M are number of instances in C_i and C_j respectively. Centroid linkage determines the distance separating two clusters by returning the difference between centroids of the two clusters as defined in (4).

$$f_m(C_l, C_k) = \left\| \left(\frac{1}{N}\sum_{i=1}^{N} a_i\right) - \left(\frac{1}{M}\sum_{j=1}^{M} b_j\right) \right\| \tag{4}$$

Ward's method is a unique method since it relies on a sum-of-squares, generating clusters that enhances within-group compactness when merging pairs of clusters. It is widely applied because it is capable of searching for clusters in a multivariate Euclidean space [24]. AHC is more frequently used than divisive because it is suited to both consistent and inconsistently structured clusters and the knowledge of the number of clusters a priori is not a requirement [25].

B. *Course Performance Association Mining*

Association-rule mining (ARM) is a widely used relationship mining and EDM tools. It creates and uncovers interesting and regular patterns (associations or correlations) among data items in a dataset [26]. It aims at detecting pattern(s) in a dataset that causes another pattern by examining the frequency of the joint occurrence of such patterns in a particular dataset. These relations or patterns are represented as *If-Then* rules characterized by antecedent and consequent parts. A strong and interesting association rule must satisfy some quality measures like support, confidence and lift. The support of a rule measures the amount of the data items in the dataset that contains items depicted in both the antecedent and consequent of the rules (frequency of the rule in the dataset) while confidence measures the proportion of the data items containing antecedent and also containing the consequent. Lift of a rule gives the ratio of the rule confidence and expected rule confidence rule. The expected confidence of a rule is computed by multiplying the support of the rule antecedent and the rule consequent and dividing by the support of the rule antecedent. These measures are formally defined in (5) and (6) as in [27].

$$Support, s(X \rightarrow Y) = \frac{\sigma(X \rightarrow Y)}{N} \tag{5}$$

$$Confidence, c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \tag{6}$$

where $\sigma(X \rightarrow Y)$ represents the frequency of transactions containing items X and Y, N represents the sum number of transactions, $\sigma(X \cup Y)$ depicts the count of transactions containing X and Y, $\sigma(X)$ is the support count of transaction containing X. However, an item with a high support might occur by chance while those with high confidence can also be misleading because confidence does excludes the correlation of the antecedent and action parts of association rules during rule extraction [28]. Lift is an interestedness factor which is the ratio of the confidence to support of an item in a rule and given in (7).

$$Lift = \frac{c(X \rightarrow Y)}{s(Y)} = \frac{confidence(X \rightarrow Y)}{support\ (Y)} \tag{7}$$

A lift greater than unity depicts a positive correlation while a negative value reveals a negative correlation between items in the antecedent with item in the consequent component of rule. A lift equals to unity denotes an independent correlation. In this work, association rules are interesting if they fulfill support, confidence and lift thresholds.

Frequent pattern growth (FP-Growth) and apriori algorithms are the main algorithms for discovering association rules. FP-Growth builds FP-tree and generates frequent itemsets directly from it while Apriori algorithm builds and extracts candidate itemsets and determine whether they are frequent or infrequent [29]. Apriori algorithm, is the most preferred algorithm for mining association rules [30-32] and can be summarized in two phases, frequent item generation—searches for all the generated frequent itemsets satisfying the support threshold value (frequent itemsets) and rule generation — extracting all the high performing confidence of rules from frequent itemsets. Detailed description of a priori algorithm is given in [26, 32,33].

III. Methodological Workflow

The major tasks for this predictive analytics, as shown in Fig. 1, comprise dataset preparation, pre-processing, predictive modeling and visualization. The predictive analytics is a core task where useful but hidden knowledge and patterns are discovered via cluster analysis and association rule mining techniques. Visualization task is implemented at every stage for enhanced insight and knowledge of the data, processes and resulting models.

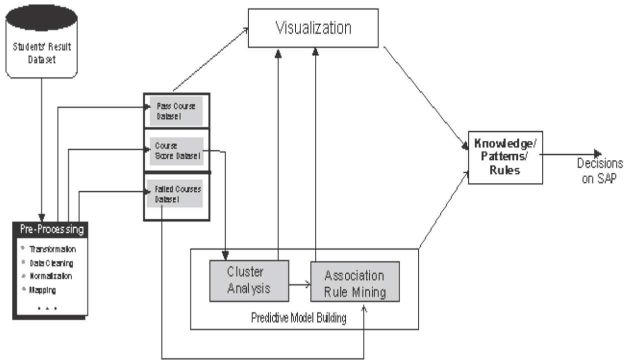


Fig.1. Methodology of the Predictive Analytics of Students Academic Performance

A. *Dataset preparation and pre-processing*

The predictive analytics is targeted at discovering patterns of failed courses resulting in a given status at the expiration of the minimum period of study by students. This was demonstrated on a dataset 846 records. This dataset represents nine sets of Bachelor of Science degree graduands with academic and demographic attributes. The academic records (result dataset) consist of results for each academic session organized in semesters.

A semester result contains raw scores, grades earned by students in the courses they enrolled and the grade point average (GPA) [3] derived from the earned scores, and the respective status of students. The final performance of the student, which is a function of performances in all preceding semesters, is summarized in the final semester with the cumulative grade point average (CGPA), and status of the student depicting the class of degree (CoD) – first class, second class upper ('21'), second class lower ("22", third class("3") and pass(P), depending on their CGPA. Any student, who exceeded the specified minimum period of programme on account of poor academic performance, is said to *'spill'*; such students are known as 'spillover" or "stay-over" while drop-outs, stop-outs and transfer-outs students are classified as "voluntary withdrawal". First-year results and the status of the students at the completion of the minimum duration of the course are of interest in this predictive analytics task.

Dataset pre-preprocessing involved the splitting of the students result dataset into three distinct groups, the first group, Course Score Dataset (CSD) — has scores earned by students in each course, while courses passed by each student at the completion of the first year of their programme make up the second group and labeled Pass Course Set (PSD). The Failed Course Set (FCS) is the third group which consist of row listing of courses failed by students at the end of their first year of study and their respective status at the end of the specified period of study. Each row of the FCS represents a student record and rows without any failed course were deleted from the dataset. The FCS had a total of 506 records with varying number of columns. The maximum number of columns is twenty two (22) representing the count of courses used in the analysis and status of each student. Student status is the rightmost entry of every row while blank cells in-between courses in a row were deleted. The CSD was represented as a 21-by-4 matrix. The four columns comprise the students' status, voluntary withdrawal (VW), second class lower (lower), second class upper (upper) and spillover students (SPO)) while the rows represent courses and the elements are average of students' scores group by status. The datasets were represented in the comma separated values (CSV), a suitable format for R environment.

B. *Students' Performance Predictive Analytics*

Predictive analytics is a sub field of data mining which concentrates on the extraction of information from data, for trends, relationships and behavioural patterns prediction. Predicting students' performances which is mostly applied to assist the trainers and learners in improving their instruction and learning activities, has become more difficult due to the increasing large volume and varying nature of data in educational data repositories [34]. Predictive analytics adopts data mining (DM) techniques for either classification or prediction tasks or both. However, the decisive goal of DM is prediction; therefore, predictive analytics is the commonest and widely applied DM task[35]. This paper employs descriptive and prescriptive techniques of DM for the identification and discovery of students' performance patterns for the predication of their performance status at the expiration of the minimum period of study. Cluster analysis was used for identification and grouping of courses into distinct groups based on (dis)similarity of students' performances, while association rule mining was used to extract sets of interesting association rules or fail course patterns that are useful for the prediction of students' status.

The HCA is implemented in RStudio – an Integrated Development Environment (IDE) for R programming language. R is open source and freely distributed programming environment with multiple integrated set of software tools for data manipulation, computations, and visualization [36]. It has rich built-in and user-defined iterative functions as well as facilities for input and output handling. R has overwhelming strengths in non-deterministic computations, resource management, computational efficiency, and scoping, and widely accepted for data science and analytics [37]. This paper implements HCA in the following steps; data pre-processing, objects and cluster similarity analysis, assessment of distance linkages, determining cut-off point into clusters and evaluating the goodness of fit of the clustering solution.

In the data pre-processing phase, the CSD was transformed into a 21-by-4 matrix containing measurements of aggregated raw scores (students' performances) based on course, and students' status. The matrix was normalized and then scaled to zero (0) mean and variance of unity as in [21]. The standardization of variables before the commencement of CA is a major requirement for some distance metrics that are sensitive to varying magnitude or scales of input feature space. It prevents outweighing attributes values from having a large influence over those with smaller values or units thereby equalizing the dimensions and variability of all features [38]. The performance of the two types of HCA algorithms (agglomerative and divisive clustering) was compared on the CSD with four linkage measures in order to discover a suitable HCA method, capable of identifying stronger clustering structures within the students' performance dataset. The divisive clustering *(diana function)* yielded a coefficient of 83.14% while agglomerative coefficient (AC) was highest with the ward's (minimum variance) linkage measure of 0.902. The performances of other linkages with agglomerative clustering are as follows; average (77.12%), single (0.5892) and complete (0.8582). The Euclidean distance measure was used in all linkage methods, since the CSD has continuous numerical values [39]. A higher AC gives an indication of a better clustering quality and fit of the dendrogram.

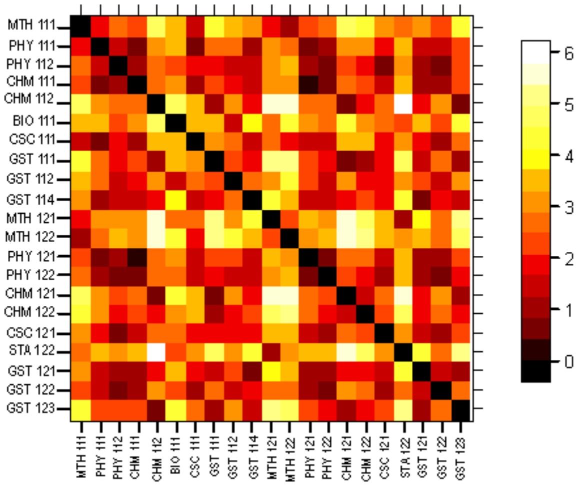


Fig. 2. Visual of intra-course performance similarity

This work adopts AHC based on ward's linkage in the partitioning of the CSD. The similarity information between every pair of objects is contained in distance matrix obtained via the "*dist*" and utilizing the Euclidean distance measure to compute the distances between the rows of CSD matrix presented visually in Fig. 2, with six colour levels ranging from black (0) to white (6) (the minimum value of dissimilarity at 0 (identical) and maximum is at 6) denoting the magnitude of similarity.

The performance distance matrix is in the range [0.368,5.7847] for minimum and maximum values representing the distances between the pairs {PHY 121,CHM 111} and {STA 122, CHM 112} respectively. This shows that the extent of similarity of the performance of students in PHY 121 and CHM 111 is the closest while STA 122 and CHM 112 performances have the widest distance separating them. High dissimilarity in the failure patterns of students is observed when mathematics based courses (MTH 111, MTH 121, MTH 122, STA 121) are compared with chemistry based courses and some university wide courses (CHM 122, CHM 112, CHM 121, GST 123 and GST 111). This implies a very high failure rate in one group than the other group while other courses depict moderate similarity with the earlier mentioned courses.

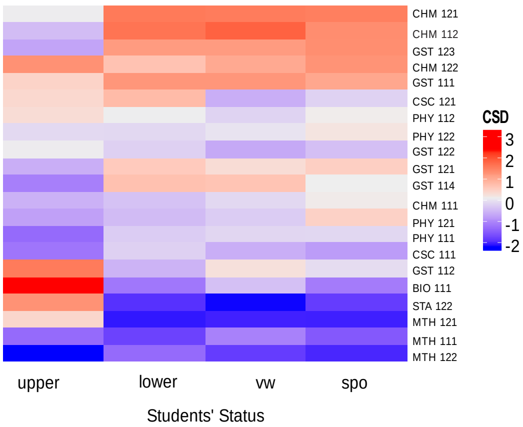
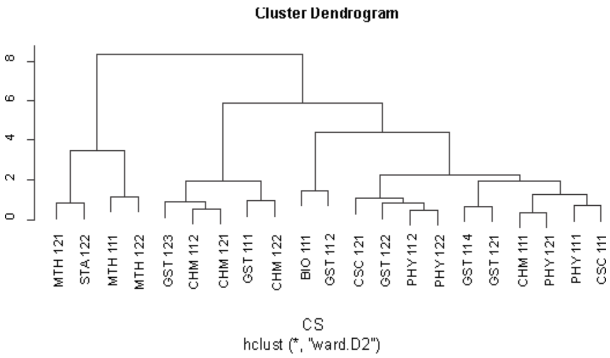


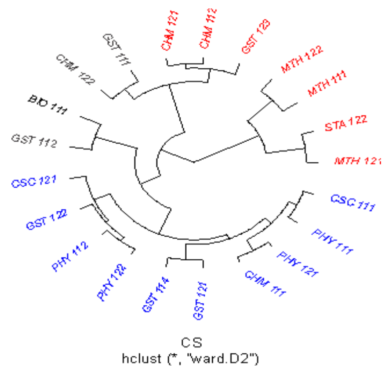
Fig. 3. Heatmap of course and status similarity

The relationship between performances in each course and students status is represented in a heatmap (Fig. 3). It shows the relative intensity of poor performances in each course on student status. Poor performances in MTH 122, MTH 111, MTH 121 and STA 121 are strongly associated with all the students' status except second class upper (upper) while failure grade in CHM 112, GST 123, CHM 122 and GST 111 are less common with students from SPO, VW, lower and upper students (high dissimilarity). Poor performances in other courses are moderately associated with upper students except MTH 122 which is closely followed by MTH 111 and PHY 111. It implies that instructional interventions are required for courses with high intensity on students' performance status.

The *'hclust'* function performs AHC transforming the course performance similarities into dissimilarities for the 21 objects (courses) being clustered. It begins by allocating each course to a distinct cluster and iteratively joining any two most similar clusters until the cluster becomes one. Figure 4 shows the dendrogram (hierarchical cluster tree) for the clustering solution. It reveals that the pair of courses with the lowest height ((dis)similarity) are {STA 122, MTH 121} and {PHY121, CHM 111}, this marks the beginning of the clustering operation. The attributes of CSD are partitioned into twelve (12) clusters with GST 123, CSC121 and GST 122 as singleton clusters at the initial stage, and thereafter merging pairs of clusters having the least between-cluster distance. This implies that GST 123, CSC121 and GST 122 are outliers and do not depict a consistent pattern of failure as exhibited by the other courses, therefore failure in these courses may largely depend on the students, though other factors might contribute as well.



a) Horizontal dendrogram



b) radial dendrogram

Fig. 4. Dendrogram of the AHC clustering of CSD (a) Horizontal view b) radial view

A widely used cluster validity index, the cophenetic correlation coefficient (CCC), was used to assess the efficiency of the dendrogram. CCC computes the correlation between the actual dissimilarity matrix and computed dissimilarity matrix [40]. It is represented by a cophenetic matrix (CM) generated after the HCA algorithm re-computes the (dis)similarities [21]. The CCC value of 0.82% returned by the cophenetic function reveals an excellent degree of accuracy by which the dendrogram preserves the pair wise distances in CSD [41]. The CSD lacks natural boundaries therefore does not have information on the inherent number of clusters. The criteria adopted in the discovery of the optimal number of clusters present in the CSD include; silhouette criterion [3.8], elbow method [42,43] and bayesian information criterion (BIC) [44]. Elbow method considers the extent of variability as a function of the natural number of clusters. The elbow is depicted in a plot of the extent (percentage) of variability explained by clusters against the count of the clusters [43]. BIC is a criterion based on likelihood corrected by the model complexity — the number of parameters in the model [44]. Silhouette plots display the closeness measure of each point in one cluster to points in the neighboring clusters. Silhouette function (*sil*) of R's cluster package was used to calculate the mean silhouette width of clusters. The experiment was iterated for 10 clusters ($1 \leq k \leq 10$) for each criterion [45]. The BIC criterion showed Fig. 5 produced three (3) as the optimum number of clusters and five (5) clusters as the second-best number of clusters. The result from elbow method (Fig. 6) is in tandem with that of the silhouette criterion (Fig. 7) which produces 5 clusters (k=5) as the best number of clusters.

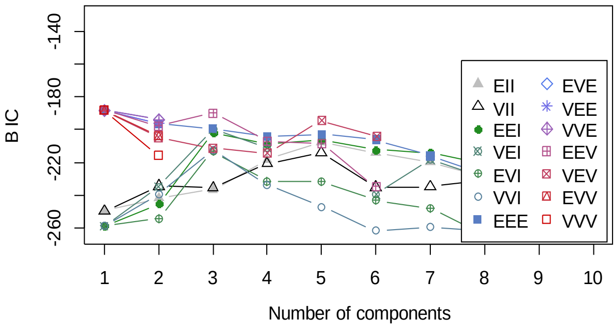


Fig. 5. BIC plot for Optimal Number of Cluster

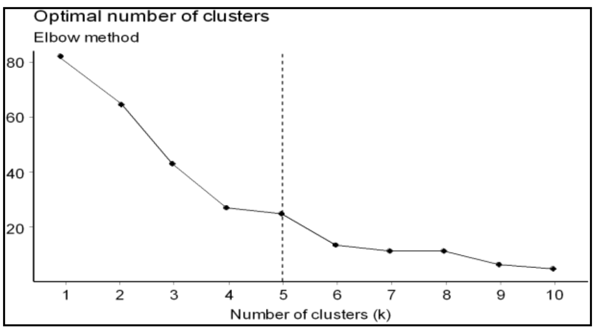


Fig. 6. Elbow plot for optimal number of clusters

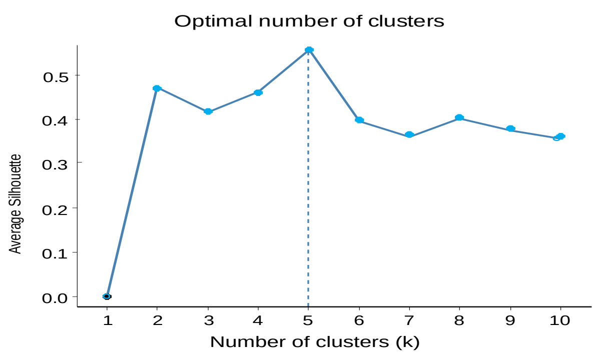


Fig. 7. Silhouette Plot for 10 clusters

The resultant five (5) clusters is considered appropriate, since two approaches point in that direction, and BIC also depicts some information relating to this number. This silhouette width index explains the compactness based on the pairwise distances between all courses in their own cluster, and separation based on pairwise distances between all courses in the cluster and all courses in the nearest other clusters [46]. The silhouette width plot (Fig. 8) validated the cluster compactness and separation with an average silhouette width of 0.57 depicting a satisfactory structure and cluster topology.

The summary of the cluster size shows that clusters 1, 4 and 5 have 6 members with 28.57%, —2 members each accounting for 9.52% each, cluster 2 has 10 members (47.6%) while 23.8% belongs to cluster 3. Extraction and visualization of the actual cluster membership is presented in Fig. 9.

Fig. 8. Silhouette width criterion plot of cluster validation

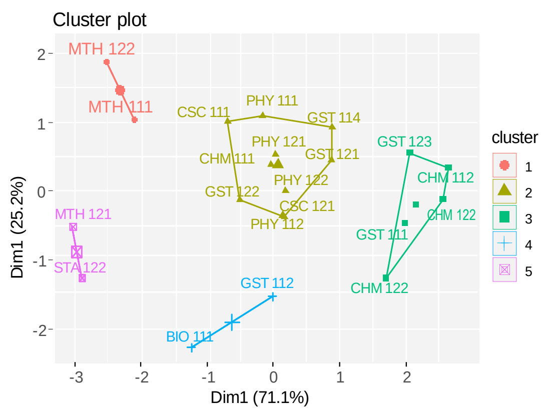


Fig. 9. Course cluster membership plot

Fig. 9, shows that MTH 111 and MTH 122 belongs to cluster 1 and exhibit the same pattern of failure while STA 122 and MTH 121 are members of cluster 5. BIO 111 and GST 112 are in the same cluster while cluster 2 has the highest membership of courses {CSC 111, CSC 121, PHY 111, PHY 112, PHY 121, PHY 12, CHM 111, GST 122, GST 114, GST 121}. In cluster 3, the membership includes GST 123, GST 111 and the courses taught by chemistry department except CHM 111.

IV. Course Association Rule Mining

The association of the discovered clusters patterns with the students' status after the minimum duration of the programme was mined using *apriori* algorithm on the Fail Course Set (FCS). The main aim is to discover courses that are frequently failed together and to assess the sensitivity of the discovered clustered course patterns on the ruleset, in relation to each status of student. An interesting and useful application of production rules is the accurate prediction of the consequent of the rule from the premise or condition part. However, not all the generated rules may be interesting or strong enough to satisfy this important requirement. Hence, reference [47] list the necessary qualities of strong rules. Secondly, premise of a production rule must specify the minimum requirements that implies the action part of the rule. In addition, confidence of a predictive rule must be relatively stable with respect to the period determined by the area of application. In consideration of the desirable properties, and ensuring the creation and extraction of useful, stronger and interesting rules, minimum support threshold and confidence levels were fixed at 0.5%, and 80% respectively and lift of greater than 1. The maximum length (maxlen) of the antecedent was also set at 10, to correspond to the minimum number of courses offered in a semester. The first step was to discover the frequent failed courses in FCS, which occur with a frequency greater than the user-specified threshold [48]. Figure 10 shows the relative frequency(RF) of failed courses for all the student's record while Fig. 11 - 13 are representations of RF of failed courses for spillover students, students who withdrew voluntarily from studies and students who graduated with second class lower, in that order.

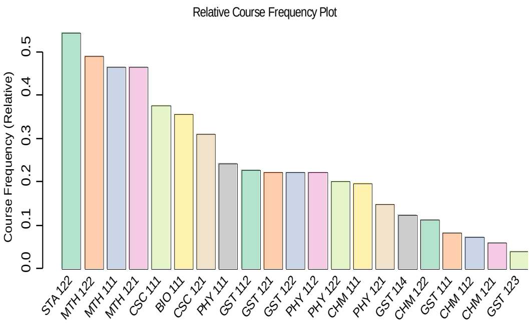


Fig. 10. Relative frequency of failed courses for all students

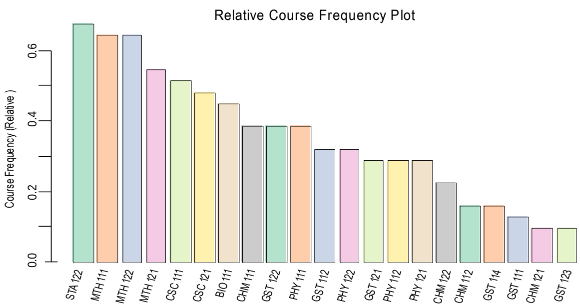


Fig. 11. RF of failed courses for voluntary withdrawal students

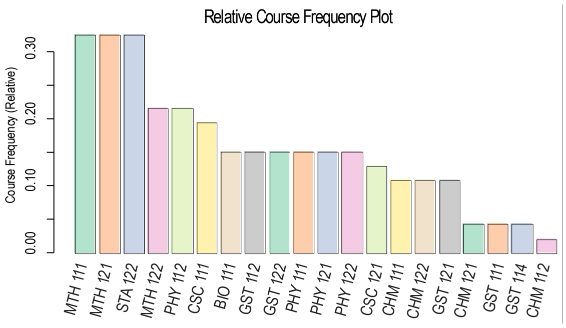


Fig. 12. Relative Frequency of Failed Courses for Second Class Lower Students

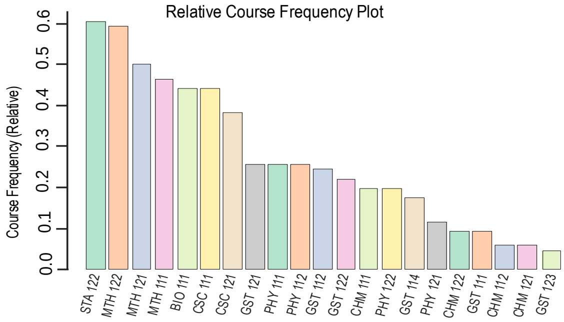


Fig. 13. Relative Frequency of Failed Courses for Spillover Students

V. Discussion

Analysis of results show that VW students have the highest average RF followed by SPO students while students with second class lower exhibited the lowest average RF. A set of seven(7) courses {STA 122, MTH 111, MTH 121, MTH 122, CSC 111, CSC 121, BIO 111} top the list of failed courses (Frequent Failed Course Set) in SPO, VW and all students category, though with varying RFs. The list is the same for second class lower students except for CSC 121 replaced by PHY 112 as the 5th in the RF rank. These frequent failed courses account for significantly for SPO and VW status. It is therefore necessary to put measures in place to improve the teaching of these courses for improved performances by students. STA 111, MTH 111 and MTH 122 have high RF index in all categories of students while GST 123, CHM 112, CHM 122, GST 111 have insignificant RFs and may not frequently occur in the list of failed courses but may significantly differentiate the status of any students. For example, GST 123 has a high pass rate and is not among the list of failed courses in the second class lower students while showing a relatively higher RFs for VW students as compared to SPO students. It is therefore obvious that any student who fails GST 123 has a higher likelihood of withdrawing from studies than spending extra year(s) in the programme. The pattern of RFs confirms the clustering results obtained in Section 3, cluster 1, 4 and 5 courses are in the top ten failed courses and therefore account for SPO and VW status. For example STA 122 and MTH 11 each has RF of about 60% for SPO category, which implies that 60% of spillover students fail STA 122 and MTH 122 while less than 5% of SPO students fail GST 123. Courses in cluster 3 have high pass rate therefore may be a means of discriminating weak students. These information will help administrators not only in identifying students at-risk of withdrawing voluntarily, or spending extra year(s) but also in providing remedial measures like extra class, tutorial classes, improving students-teacher ratio and provision of adequate teaching aids for these courses, to enhance students performances.

A total of 359,958 rules were extracted with a maximum rule length of 6, however, pruning of redundant rules [32] resulted in 1034 rules with a maximum length of 5. Rules of length 2 are 30, 604 rules have a length of three (3), 387 rules comprises 4 courses while 13 rules have length 5. A summary of the descriptive statistics of rule quality measures is given in Table 1, while Figure 14 depicts a two-key scatter plot of the resultant rules. As shown in Table 1, all the rules have a lift greater than 1 and support values between 0.005 and 0.29. The mean confidence is almost 1 (0.93). These values confirm that the rules are strong and suitable for the prediction of students' status. The visualized ruleset (Figure 14) shows that majority of rules are of order 3 with support between 0.005 - 0.1 and a confidence of 0.0 - 1 unlike order 4 rules that clusters around the support of 0.15. Order 2 and 5 rules consist of rules with higher support but relatively low confidence.

Table 1. Distribution of interestingness measures for the extracted Ruleset

	support	confidence	lift	Count
Minimum	0.005	0.8	1.470	1.000
1st Quartile	0.0051	0.8571	2.115	1.000
Median	0.0103	1.0000	2.681	4.000
Mean	0.0135	0.9379	3.272	6.114
3rd Quartile	0.0155	1.0000	16.083	8.000
Maximum	0.2901	1.0000	56.125	56.000

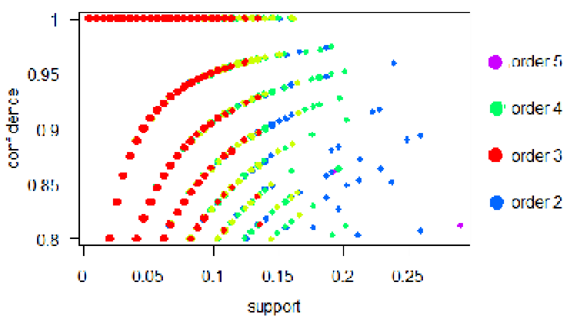


Fig. 14. Visualization of ruleset for Students status prediction

The rules were categorized into three based on the consequent. Out of the 1034 rules, five rules have Second Class Lower as consequent, 38 rules are for the prediction of SPO students while the remaining rules are modeled for VW. The ten topmost rules representatives for SPO students and VW students are visualized in Figures 15-17. It is seen that GST 111 and CHM 111 are central in rules that associates with SPO students. It implies that courses which have these courses as pre-requisites have high likelihood of failure by SPO students, therefore pedagogical interventions should be put in place to improve the performances of students in these courses. CHM 111 is not a core course but GST 111 is a general course (Use of English). These two courses (CHM 111 and GST 111) have low RFs (20% and 15% respectively) meaning that more than 80% of students pass these courses. This show that this category of students are weak students and require special attention in teaching and infrastructure. The top ten rules for second class status prediction hover around CHM 111 and GST 121 while VW students rule depends on GST 123 and CHM 112 as failed courses. It is seen that cluster 3 courses have high RFs in PSD and are actually used to discriminate performance status of students. Each rule is visualized independently in the parallel coordinate plots depicted in Figures 17-19 and their interpretation provided in Table 2. It is observed that the top most rules in each category have lift values greater than 2 which confirms a high degree of interestingness.

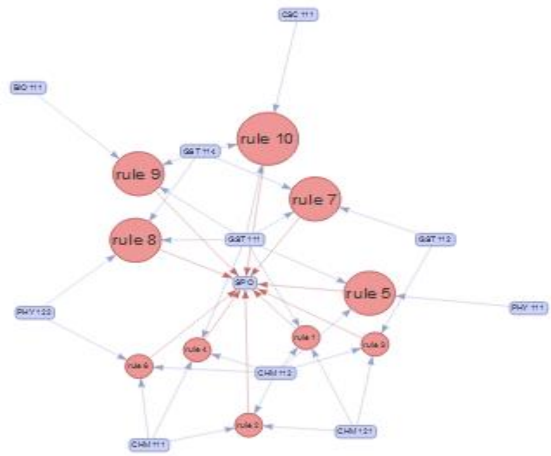


Fig. 15. Graph-based visualization of ten top rules as vertices for SPO Students

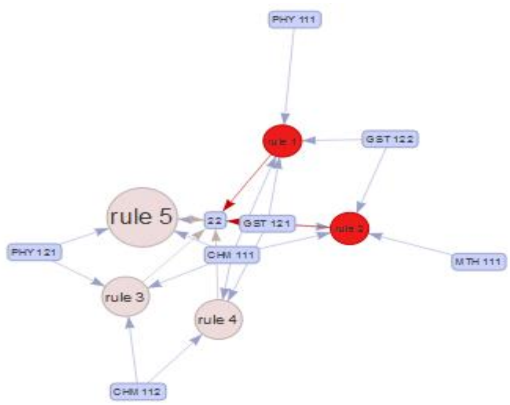


Fig. 16. Graph-based visualization of rules as vertices for second class lower students

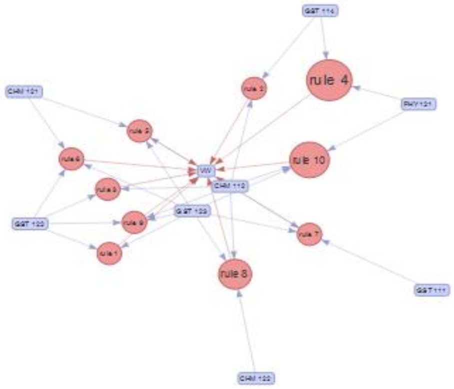


Fig. 17. Graph-based visualization of 5 rules as vertices for VW students

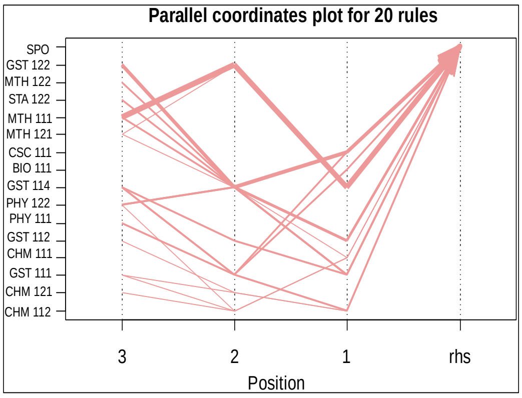


Fig. 18. Coordinate plots for SPO rules

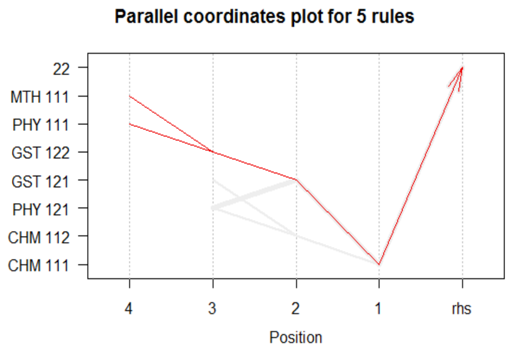


Fig. 19. Coordinate plots for 22 rules

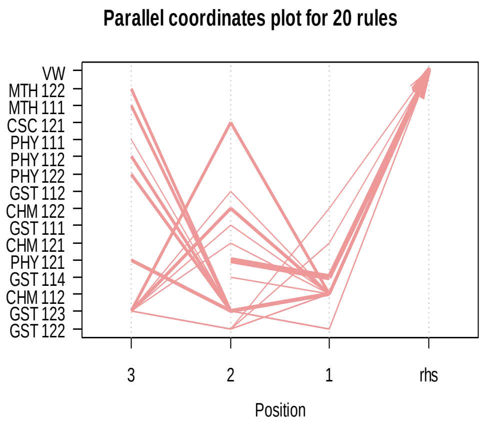


Fig. 20. Coordinate plots for vw rules

VI. Conclusion

This paper proposes a methodology driven by HCA and ARM for the mining of students' data with the intention to identifying performance patterns of students who are at risk of failure, and to provide timely academic interventions. CA was adopted to identify and group the students' courses into distinct groups based on similarity of students' performances, while ARM was used to extract interesting course sets that are useful for the prediction of students' status. In particular, failed courses and their possible associations/effect with performance status was analyzed. CA identified five distinct clusters with corresponding courses in each of students' cluster based on their impact on students' status. ARM identified and extracted frequently failed courses and their relationship to student status. Analysis of findings indicates predictive patterns about frequently occurred courses that might put students at risk of failure, spill over, voluntary withdrawal or drop out. A set of seven

Table 2. Listing of Topmost rules for each status

	LHS	RHS	support	Confi-dence	lift
[1]	{CHM 111, CHM 112, PHY 121}	{22}	0.010	0.667	2.738
[2]	{CHM 111, CHM 112, GST 121}	{22}	0.010	0.667	2.738
[3]	{CHM 111, GST 121,PHY 121}	{22}	0.021	0.667	2.738
[4]	{CHM 111, GST 121,GST 122,PHY 111}	{22}	0.005	1.000	4.106
[5]	{CHM 111, GST 121, GST 122,MTH 111}	{22}	0.005	1.000	4.106
[4]	{CHM 111, GST 121,GST 122,PHY 111}	{22}	0.005	1.000	4.106
[5]	{CHM 111, GST 121,GST 122,MTH 111}	{22}	0.005	1.000	4.106
[1]	{GST 122, GST 123}	{VW}	0.005	1.000	7.423
[2]	{CHM 112, GST 114}	{VW}	0.005	1.000	7.423
[3]	{CHM 112, GST 122}	{VW}	0.005	1.000	7.423
[4]	{GST 114, PHY 121}	{VW}	0.016	1.000	7.423
[5]	{GST 122, PHY 121}	{VW}	0.036	0.875	6.495
[6]	{CHM 112, CHM 121, GST 123}	{VW}	0.005	1.000	7.423
[7]	{CHM 121, GST 122, GST 123}	{VW}	0.005	1.000	7.423
[8]	{CHM 112,GST 111,GST 123}	{VW}	0.005	1.000	7.423
[9]	{CHM 112,CHM 122,GST 123}	{VW}	0.010	1.000	7.423
[10]	{CHM 112, GST 122,GST 123}	{VW}	0.005	1.000	7.423
[1]	{CHM 112,CHM 121,GST 111}	{SPO}	0.005	1.000	2.413
[2]	{CHM 111,CHM 112,CHM 121}	{SPO}	0.005	1.000	2.413
[3]	{CHM 112,CHM 121,GST 112}	{SPO}	0.005	1.000	2.413
[4]	{CHM 111,CHM 112,GST 111}	{SPO}	0.005	1.000	2.413
[5]	{CHM 112,GST 111,PHY 111}	{SPO}	0.010	1.000	2.413
[6]	{CHM 111,CHM 112,PHY 122}	{SPO}	0.005	1.000	2.413
[7]	{GST 111,GST 112,GST 114}	{SPO}	0.010	1.000	2.413
[8]	{GST 111,GST 114,PHY 122}	{SPO}	0.010	1.000	2.413
[9]	{BIO 111,GST 111, GST 114}	{SPO}	0.010	1.000	2.413

courses were found to frequently occur together in the failed course set and are identified to be critical to students' academic performance. The results of CA were in tandem with ARM both in their RF patterns and in the extracted rules. Courses in cluster 5 play pivotal in the discriminating students' performance status, while the failure patterns in clusters 1,4 and 5 depict high failure rates implying major contributors to poor students status.

The findings from this work, will help instructors in the provision of appropriate pedagogical measures for teaching of courses in each cluster, identify sources of poor performances and also identify students at risk of failure as well as providing adequate interventions to these clusters of students. Particularly, instructors should be more connected with their students by providing individualized or cluster-based supports and advice. They should identify and proffer timely academic interventions to at risk students such as organizing extra classes, attendance monitoring, one-on-one interactions, and reassessment of instructional activities. This will help to boost student learning and consequent improvement in students' performance and overall students' quality and satisfaction. As further work, the integration demographic attributes of students and other performance based attributes in the dataset for a holistic predictive modelling and utilizing other nature inspired tools for the adaptive learning and optimization of rules respectively would be necessary for utilization of the extracted association rules.

References

[1] Hussein, S., Dahan, N. A., Ba-Alwib, F. M. and Ribata, N. "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA Indonesian Journal of Electrical Engineering and Computer Science" 9(2), (2018), 447~459,. DOI: 10.11591/ijeecs.v9.i2.pp447-459

[2] Ray, S., and M. Saeed. "Applications of educational data mining and learning analytics tools in handling big data in higher education". In *Applications of Big Data Analytics*, 135-160, 2018. Springer, Cham.

[3] Inyang, U, G, Umoh, U. A., Nnaemeka, C and S. Robinson. "Unsupervised Characterization and Visualization of Students' Academic Performance Features". 12(2), 103-105, 2019. doi.org/10.5539/cis.v12n2p103

[4] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, no. 6 (2010): 601-618.

[5] Sklyar, Eduard. "Exploring First-Time Community College Transfer Students' Perception of Their Experience as They Transition to a Large Public Four-Year Institution." PhD diss., Northeastern University, 2017

[6] O'Keeffe, Patrick. "A sense of belonging: Improving student retention." *College Student Journal* 47, no. 4 (2013): 605-613.

[7] K. E. Arnold, and M. D. Pistilli. Course signals at Purdue: using learning analytics to increase student success. *In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. 267–270, 2012. doi:10.1145/2330601.2330666.

[8] U. G. Inyang and E. E Joshua. "Fuzzy Clustering of Students' Data Repository for At-Risks Students Identification and Monitoring". *Computer and Information Science*, 2013. *6(4)*, 37-50.

[9] J., Xu, Moon, K. H., and M. Van Der Schaar, "A machine learning approach for tracking and predicting student performance in degree programs" *IEEE Journal of Selected Topics in Signal Processing*, 2017. *11*(5), 742-753

[10] R. Agrawal, T. Imielinski and A. Swami "Mining association rules between sets of items in large databases". in: *Proceedings of the ACM SIGMOD Conference on Management of Data*, (1993). 207-216.

[11] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic. "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success". *The Internet and Higher Education*, 2016. *28*, 68-84.

[12] O. Zaïane "Web usage mining for a better web-based learning environment", In proceeedings of t*he 4th International Conference on Advanced Technology for Education (CATE'01)*, 27–28 June 2001, Banff, Canada.

[13] D. Gašević, N. Mirriahi, and S. Dawson. "Analytics of the effects of video use and instruction to support reflective learning". In *Proceedings of the fourth international conference on learning analytics and Knowledge*. 2014. 123-132

[14] N. Zacharis "A multivariate approach to predicting student outcomes in web-enabled blended learning courses, Internet and Higher Education", 2015, 27, 44–53.

[15] J. Ruipérez-Valiente, P. Muñoz-Merino, D. Leony, and Kloos Delgado. "ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform". *Computers in Human Behavior*, 2015. 47, 139–148.

[16] Y. Park, and L. Jo. "Development of the Learning Analytics Dashboard to Support Students' Learning Performance" *Journal of Universal Computer Science*, 2015. 21(1), 110-133

[17] A. Daud, N. Aljohani, R. Abbasi, M. Lytras, F. Abbas, and J. Alowibdi. "Predicting Student Performance using Advanced Learning Analytics, International World Wide Web Conference Committee (IW3C2)", 2017, 415-421.

[18] X. Wanli, G. Rui, P. Eva, and G. Sean. "Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory" *Computers in Human Behaviour*, 2015. 47. 168–181.

[19] J. Hair, R. Anderson, Tatham; and C. Black. *Análise multivariada de dados*, Bookman, 2005, Porto Alegre, Brazil

[20] R. W. Sembiring, J. M. Zain, and A. Embong. "A comparative agglomerative hierarchical clustering method to cluster implemented course". Journal of Computing, 2,(12), December 2010, ISSN 2151-9617 *Arxiv Preprint Arxiv:1101.4270*.

[21] Singh, E. Hjorleifsson, and G. Stefansson. "Robustness of fish assemblages derived from three hierarchical agglomerative clustering algorithms performed on Icelandic ground fish survey data" *Journal of Marine Science*, 2011, 68(1), 189 –200. doi:10.1093/icesjms/fsq144

[22] O. Yim, and K. T. Ramdeen. "Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data". *The quantitative methods for psychology*, *11*(1), 2015, 8-21.

[23] Z. Li, and M. deRijke. "The impact of linkage methods in hierarchical clustering for active learning to rank". In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, 941-944). ACM.
[24] F. Murtagh, and P. Legendre. "Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?" *Journal of classification*, *31*(3), 2014, 274-295.
[25] J. Vesanto, and E. Alhoniemi. "Clustering of the self-organizing map. *IEEE Transactions on neural networks*, *11*(3), 2000, 586-600.
[26] N. Ye. "*Data mining: theories, algorithms, and examples"*. 2014. CRC press.
[27] P. D. McNicholas, T. B. Murphy, and O'Regan, M. Standardizing the lift of an association rule. *Computational Statistics and Data Analysis*. *52*(10), 2008, 4712-4721.
[28] N. Hussein, A. Alashqur and B. Sowan. "Using the interestingness measure lift to generate association rules". *Journal of Advanced Computer Science & Technology*, *4*(1), (2015, 156.
[29] F. Verhein. "Frequent pattern growth (FP-growth) algorithm". *School of Information Studies, The University of Sydney, Australia*, 2008, 1-16.
[30] J. Han, M. Kamber and J. Pei. *"Data mining: Concepts and techniques"* (3rd ed.). 2012, San Francisco: Morgan Kaufmann Inc
[31] M. Y. Avcilar, E. Yakut. "Association Rules in Data Mining: An Application on a Clothing and Accessory Specialty Store". *Canadian Social Science.* 10(3), 2014.75-83. DOI: 10.3968/4541
[32] M. Dimitrijevic and Z. Bosnjak. "Pruning statistically insignificant association rules in the presence of high-confidence rules in web usage data". *Procedia Computer Science*, 35, 2014. 271-280.
[33] Mandave, Pratibha, Megha Mane, and Sharada Patil. "Data mining using Association rule based on APRIORI algorithm and improved approach with illustration." *International Journal of Latest Trends in Engineering and Technology (IJLTET), ISSN* (2013).
[34] A. M. Shahiri, and W. A. Husain. "Review on predicting student's performance using data mining techniques". *Procedia Computer Science*, 72, 2015. 414-422
[35] Meng, Xue-Hui, Yi-Xiang Huang, Dong-Ping Rao, Qiu Zhang, and Qing Liu. "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors." *The Kaohsiung journal of medical sciences* 29, no. 2 (2013): 93-99.
[36] W. Venables; and D. Smith. "The R Core Team, An introduction to R" (2017). https://cran.r project.org/doc/manuals/r-release/R-intro.pdf. Accessed on 6th June, 2019
[37] R. Ihaka, and R. Gentleman. "R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*. 5(3), 299-314,996:DOI: 10.1080/10618600.1996.10474713
[38] I. Mohamad and D. Usman. "Standardization and Its Effects on K-Means Clustering Algorithm. Research" *Journal of Applied Sciences, Engineering and Technology*. 6(17), 2013. 3299-3303
[39] M. Pathak. "Hierarchical Clustering in R". (2018) https://www.datacamp.com/community/tutorials/hierarchical-clustering-R. Accessed June 28, 2019.
[40] A. Timofeeva. "Evaluating the robustness of goodness-of-fit measures for hierarchical clustering". *In Journal of Physics: Conference Series*, January 2019, 1145(1), 012049. IOP Publishing.
[41] P. Carvalho, C. Munita, and A. Lapolli1. "Validity Studies among Hierarchical Methods of Cluster Analysis Using Cophenetic Correlation Coefficient" *International Nuclear Atlantic Conference - INAC 2017 Belo Horizonte*, MG, Brazil, October 22-27, 2017
[42] R. Gove. "Using the elbow method to determine the optimal number of clusters for k-means clustering". URL: https://blocks. org/rpgove/0060ff3b656618e9136b, 17-19. (2017)
[43] P. Bholowalia, and A. Kumar. "EBK-means: A clustering technique based on elbow method and k-means in WSN". *International Journal of Computer Applications*, 105(9), (2014). 17-24
[44] Wolzinger, Renah, and Henry O'Lawrence. "Student Characteristics and Enrollment in a CTE Pathway Predict Transfer Readiness." *Pedagogical Research* 3, no. 2 (2018): n2..
[45] M. Charrad, N. Ghazzali, V. Boiteau, and Niknafs, A. "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set". *Journal of Statistical Software* 61(6), (2014). 1-36
[46] T. Van Craenendonck and H. Blockeel. "Using internal validity measures to compare clustering algorithms". In *AutoML workshop at ICML 2015*, 1-8.
[47] J. Deogun and L. Jiang. "Prediction mining–an approach to mining association rules for prediction". *In International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, 98-108, (August 2005). Springer, Berlin, Heidelberg.
[48] J. Thakkar and M. Parikh. "An Efficient Approach for Accurate Frequent Pattern Mining Practicing Threshold Values". *International journal of Engineering and Technology.* 4(4), (2018) 2394-4099

Authors' Profiles

Udoinyang G. Inyang received the B.Sc. degree in Computer Science from University of Uyo, Nigeria in 1996. He received the M. Sc and PhD. degrees in Computer Science from The Federal University of Technology Akure, in 2005 and 2012, respectively. He is a Senior lecturer in the department of Computer Science University of Uyo, Nigeria and has many publications in reputable international journals. His research interests include Soft Computing, Speech Technology, Data Mining and Knowledge Engineering. He is a member of IEEE, Computer Professionals Registration Council of Nigeria (CPN) and Nigerian Computer Society among other professional bodies.

Eyoh, Imo Jeremiah received the Ph.D. in Computer Science from the University of Nottingham, United Kingdom in 2018. She received the B.Sc and M.Sc degrees in Computer Science from the University of Uyo and Ibadan, all in Nigeria in 1999 and 2005 respectively. She is currently working in the Department of Computer Science, University of Uyo, as a Senior lecturer. Her research areas include artificial intelligence, optimization and machine learning with particular interest in uncertainty modelling using type-2 fuzzy logic systems. She has published many papers in reputable national and international journals and conferences. She is a member of IEEE, Computer Professionals Registration Council of Nigeria (CPN) and Organisation for Women in Science for the Developing World (OWSD).

Mr. Samuel A. Robinson is a lecturer in the Department of Computer Science. He holds a Bachelor of Science degree in Computer Science from University of Uyo, Nigeria in 2009, a Master degree in Computer Science from University of Benin, Nigeria in 2016. His research interest includes; Data Communication Network, Computational Intelligence Systems and Machine Learning. He has published widely in reputable local and international journals. He received the award of the best president of Nigerian Association of Computer Science Students UniUyo Chapter in 2008.

Dr. Edward Udo is a lecturer in the department of Computer Science, University of Uyo, Uyo. He holds a Bachelor of Science degree in Computer Science from University of Uyo, Nigeria, a Masters degree in Computer Science from University of Port Harcourt, Nigeria and a Ph.D degree from University of Benin, Nigeria. He is a professionally certified Network Administrator by CISCO and a recognized CISCO Certified Academy Instructor. He is also certified as a Management Trainer by Centre for Management Development (CMD). He is a member of Nigeria Computer Society (NCS), a member of Computer Professionals Registration Council of Nigeria (CPN) and a member of IEEE (Computer section). He earned FGN/TETFUND award in 2008 and 2012 for M.Sc and Ph.D research respectively. His research interest includes among others, object oriented software engineering, software metrics and measurement, networking and systems security.

How to cite this paper: Udoinyang G. Inyang, Imo J. Eyoh, Samuel A. Robinson, Edward N. Udo, " Visual Association Analytics Approach to Predictive Modelling of Students' Academic Performance ", International Journal of Modern Education and Computer Science(IJMECS), Vol.11, No.12, pp. 1-13, 2019.DOI: 10.5815/ijmecs.2019.12.01

