ISSN 1512-1232

KNOWLEDGE DISCOVERY FROM EDUCATIONAL DATABASE USING APRIORI ALGORITHM

Abdulsalam Sulaiman Olaniyi[1], Hambali Moshood Abiola[2], Salau-IbrahimTaofeekatTosin[3], Saheed Yakub Kayode[4], Akinbowale Nathaniel Babatunde[5]

[1]Department of Computer Science Kwara State University Malete, Nigeria. sulaiman.abdulsalam@kwasu.edu.ng

[2]Department of Physical Sciences, Al-Hikmah University, Ilorin. yksaheed@alhikmah.edu.ng

[3]Department of Computer Science, Federal University, Wukari, Nigeria.hamberlite@gmail.com.

[4]Department of Physical Sciences Al-Hikmah University, Ilorin. ttsalau@alhikmah.edu.ng

[5]Department of Computer Science Kwara State University Malete, Nigeria.
akinbowale.babatunde@kwasu.edu.ng

*Corresponding Author

Abstract

Ability to predict student's performance has become very crucial in educational environments and plays important role in producing the best quality graduates. There are several statistical tools for analyzing students' performance for knowledge discovery from available data. This study presents data mining in educational sector that identifies students' failure pattern using Apriori algorithm. The results of 20 students in 25 courses taken in their 100 and 200 level of an educational institute in North Central Nigeria were considered as a case study. The patterns discovered were used to provide recommendations to academic planners so as to improve their level of decision making, restructuring of curriculum, and modifying the prerequisites of various courses. This study revealed some interesting patterns in failed courses as some failed courses have a relationship with other failed courses. A data mining software for mining student failed courses was developed, used to mine students result, and the analysis were described.

Keywords: *Association Rule Mining, Apriori Algorithm, Academic performance, Educational data mining, Curriculum, Educational database, Students' result repository.*

1.Introduction

Knowledge discovery in databases (KDD), often called data mining (DM) is a term used for exploration and analyzing of data in search of hidden and interesting patterns or knowledge. This involves applying various algorithms in order to discover and extract the said patterns.

Educational data mining (EDM) is fast becoming a fascinating research area which allows researchers to dig out useful, previously unknown patterns from the educational database for better understanding, improved educational performance and assessment of the student learning process (Dogan&Camurcu, 2008). It allows investigation of unique information from students' repository in academic institutions. Every higher institution must have a database where all relevant data to the academic activities are kept and managed. A students' result repository is part of an educational database. It is a large data bank which stores students' raw scores and grades in different courses they enrolled for during their academic year(s) in the institution. A student is expected to complete a number of courses prior to graduation. Some of the courses may have prerequisite(s) that a student must pass to ascertain the required body of knowledge taught in the prerequisite course. It is generally expected that students' who performed well in the prerequisites of a course or related course(s) would also perform well in the course. In contrast, students who do not perform well in the prerequisites of the course or related course(s) would not perform well in the course.

DM techniques have been employed to solve different problems in educational environment which includes students' classification based on their learning behaviors; prediction of students' final year grades; detection of irregular patterns; clustering according to e-learning system usage, etc. (Castro, Vellido, Nebot&Mugica, 2007; Abdulsalam, Babatunde, Hambali &Babatunde, 2015).

There are different data mining tools which can be used for the analysis of data, but the choice of data mining is mostly determined by the scope of the problem and the expected analysis result. In this study, we made uses of Association Rule Mining approach of DM which uses Apriori algorithm to analyze the relationship between students' academic failed courses and determine the prerequisite courses. This is done to discover hidden and important relationships that exist between students' failed courses in form of rules. The rules generated will be analyzed to bring about useful and constructive recommendations to the academic planners. This will help the academic planner to make a proper decision and aid in restructuring and modifying the curriculum which in turn improve student's performance and reduce failure rate among the students.

2. Related Works

There are lots of studies carried out in the area of DM in educational sector. Each of them is trying to enhance the educational system by discovering patterns among the great deal of data. In this section, we will discuss few of the existing works.

Romero and Ventura (2007) presented a survey of EDM from 1995-2005 and Baker and Yacef (2009) extended their survey covering the newest development until 2009.Khan (2005) worked on performance study of 400 students which comprises of 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a core objective of establishing the predictive value of various measures of cognition, personality and demographic variables for success at higher secondary level in science stream. He used cluster sampling technique to select from the entire population of interest which was divided into groups or clusters, and a random sample of these clusters were chosen for further analyses. It was established that girls with high socio-economic status had relatively better academic achievement in science stream and boys with low socio-economic status had relatively higher academic achievement in general.

Tissera, Athauda and Fernando (2006) present experimental work conducted in an ICT educational institute in Sri Lanka. They applied association rule mining on 20 courses in three fields of specialization, to find relationships between the subjects in undergraduate syllabus. This knowledge provides many insights into the syllabi of different educational programs in Sri Lanka Institute of Information Technology (SLIIT) and the results gained assisted in decision making that directly affects the quality of the educational programs. Al-Radaideh, Al-Shawakfa and Al-Najjar (2006) employed data mining techniques, particularly classification to build a model that help in improving the quality of the higher educational system by evaluating student data and detect some of the attributes that may affect the student performance in their courses. The extracted classification rules are based on the decision tree, ID3, C4.5 and Naive Bayes. The models were used to predict the students' final grade in a course under study. The outcome of their work indicated that Decision tree had better prediction than other models used.

Hijazi and Naqvi (2006) conducted a study on the students' performance by selecting 300 sample of students (225 males and 75 females) from a group of colleges affiliated to Punjab University of Pakistan. Their study investigate the factors that significantly related with students performance such as Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance. Galit et al. (2007) gave a case study of using students’ data to analyze their learning behavior in order to predict their results and warn the students at risk before their final exams.

Cortez and Silva (2008) predicted failure in the two core subjects (Mathematics and Portuguese) from two secondary school students of Alentejo region of Portugal by utilizing 29 predictive variables which including past school grades, demographics, social and other school related data. They used four DM algorithms such as Decision Tree (DT), Random Forest (RF), Neural Network (NN) and Supporting Vector Machine (SVM) to mine their data set of 788 students, who sat for 2006 examination. It was found that DT and NN algorithms had high predictive accuracy of 93% and 91% for two-class dataset (pass/fail) respectively. Also, it was reported that both DT and NN algorithms had the predictive accuracy of 72% for a four-class dataset.

Fadzilah and Mansour (2009) used DM techniques for understanding student enrolment data. They did comparative study of three predictive DM techniques namely Neural Network, Logistic regression and Decision tree. The results obtained can be used by the academic planners to formulate proper plan for University.

Hongjie(2010) works on student learning result based on DM. The research work was aimed at putting forward a rule-discovery approach suitable for the student learning result evaluation and applying it into practice so as to improve learning evaluation skills and better learning practicing.

The use of association based classification for relational data in web environment was presented in Bartik (2009). The intention of the study is to put forward modification of the fundamental association based classification technique that can be obliging in data gathering from Web pages. Damaševicius (2009) studied student results for informatics course. The research work try to improve students’ performance in the course by ranking course topics following their importance for final course marks based on the forte of the association rules and proposed which specific course topic should be enhanced to achieve higher student learning effectiveness and progress. Sumithra and Paul (2010) presented a distributed Apriori association rule mining and classical Apriori mining algorithms for grid-based knowledge discovery. Qiang, Shi-Xiong and Lei (2009) proposed association classification based method on compactness of rules. The proposed approach suffers from a difficulty of over fitting because the classification rules satisfied least support and lowest confidence are returned as strong association rules return to the classifier.

Oladipupo and Oyelade (2010) worked on association rule mining algorithm for enhancing the effectiveness of academic planners and level advisers in higher institutions of leaning. The analysis was done using undergraduate students’ result in the department of Computer Science and Management Information Science from a university in Nigeria.Their study was carry out using association rule data mining technique to identify student's failure patterns. They took a total number of 30 courses for 100 level and 200 level. Their study focused on constructive recommendation, curriculum structure and modification in order to improve students' academic performance and reduce failure rate.

3. Methodology

Data were collected from the University database using sampling method,20 students in 25 courses taken in their 100 and 200 level of Computer Science department of an educational institute in North Central of Nigeria were considered as a case study. The database is designed in phpMyAdmin environment, queried with MySQL connected to Java Programming Language. The model was evaluated and then trained using association rule mining technique. The proposed methodological framework used in this study is illustrated in figure 1.

¶

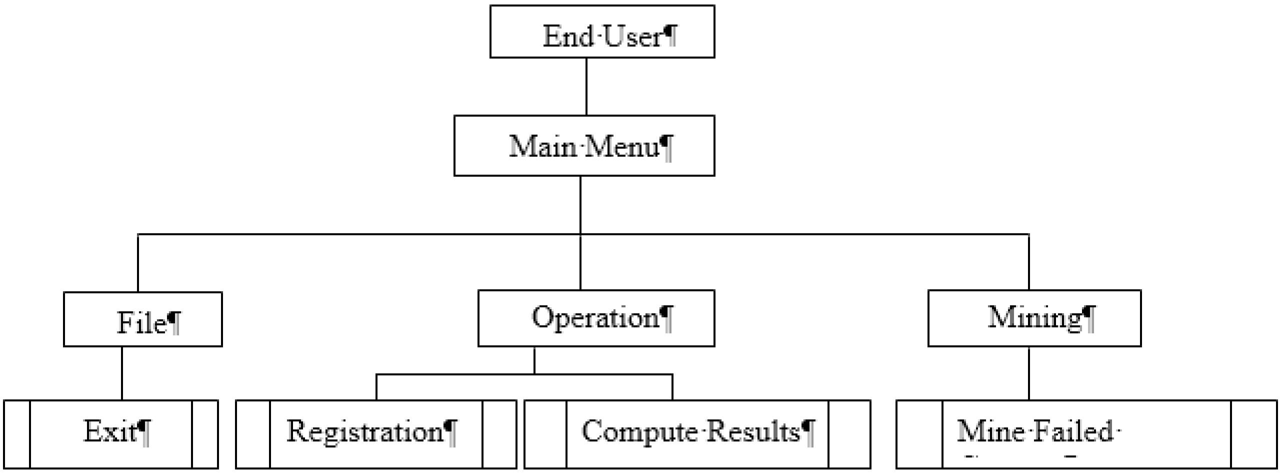


Figure 1: Proposed Methodological Framework

3.1 Data Input Request for Students Registration

The following student’s information required to be supplied into the system through Registration module:

i. **Matric Number**: This is a unique identification (transaction ID) number used to identify students to be considered in the study.
ii. **Level:** This is the level of the students in the university.
iii. **Semester**: This is the academic semester to be considered in the study.
iv. **Courses:** These are the courses registered by the students from which the failed courses will be selected, in order to carry out association mining so as to discover the relationship that exists between them using Apriori algorithm. Figure 2 shows the snapshot for student and course registration window.

3.2. **Data Input Request for Result Computation**

The necessary data required for result computation are:

i. **Matric Number**: This is a unique identification (transaction ID) number used to identify students to be considered in the study.
ii. **Level:** This is the level of the students in the university.
iii. **Semester:** This is the academic semester to be considered in the study.
iv. **Score**: This is the result of a student within a semester at a particular level. The grades of the result will be automatically generated as you input the scores. Figure 3 show a snapshot of a result computation window of the proposed system.

3.3.**Data Input for Generation of Rules**

The user or decision maker will be responsible for supplying the value of support and the confidence to the system in order to discover the relationship that exists between courses in the database.

i. **Support**: This will serve as a constraint that guides how the rules are generated from the failed courses. When the value of the support is low, more rules will be generated. When the support is high, few rules will be generated.

ii. **Confidence**: This is the degree to which the rule formed holds.

3.4. **Association Rule Analysis**

Association rule mining associates one or more attributes of a dataset with another attributes, in order to discover hidden and significant relationship between the attributes, producing an **If - Then** statement concerning attribute values in the form of rules (Lin &Pei-qi, 2001; Tan, Steinbach & Kumar, 2006; Abdulsalam, Adewole, Akintola&Hambali, 2014). An association rulebased on market basket analysis state that if we pick a customer at random and find out that he/she selected some item (that is, bought some products), we can be assured by indicating the quantity by a percentage that he/she also bought some other products (Abdulsalam et al., 2014).

According to Oladipupo&Oyelade (2010), an association rule is an implication expression of the form X => Y, where $X \subset I$, and $Y \subset I$, and X and Y are disjoint itemsets, i.e. $X \cap Y = \phi$. The strength of anassociation rule can be measured in terms of its support and confidence. The rule X => Y holds in the transaction set D with confidence *c* and support *s*, if c% of the transactions in D thatcontains X also contains Y, and s% of transactions in D contains $X \cup Y$. Both the antecedent and the consequent of therule could have more than one Item. The formal definitions ofthese two metrics are:

$$\text{Support, s } (X \Rightarrow Y) = \Sigma(X \cup Y)/N \qquad (1)$$

$$\text{Confidence, c } (X \Rightarrow Y) = \Sigma(X \cap Y)/\Sigma X \qquad (2)$$

Let $I = \{I_1, I_2 \ldots I_m\}$ be a set of literals called items and D be a set of transactions where each transaction T is a set of items such that $T \subseteq I$. Associated with each transaction is a unique identifier, called its TID. We say that a transaction T contains X, a set of some items in I, if $X \subseteq T$ (Han&Kamber 2006).Association rule mining process could be divided into two main phases to enhance the implementation of the algorithm. The phases are:

1. Frequent Item Generation: This is to find all the itemsetsthat satisfy the minimum support threshold. The itemsets arecalled frequent itemsets.

2. Rule Generation: This is to extract all the high confidence rules from the frequent itemsets found in the first step. Theserules are called strong rules.

According to Larissa (2003) association analysis is based on the rule that specifies in the form: If item A is part of an event then X % ofthe time (confidence factor) item B is part of the same event.

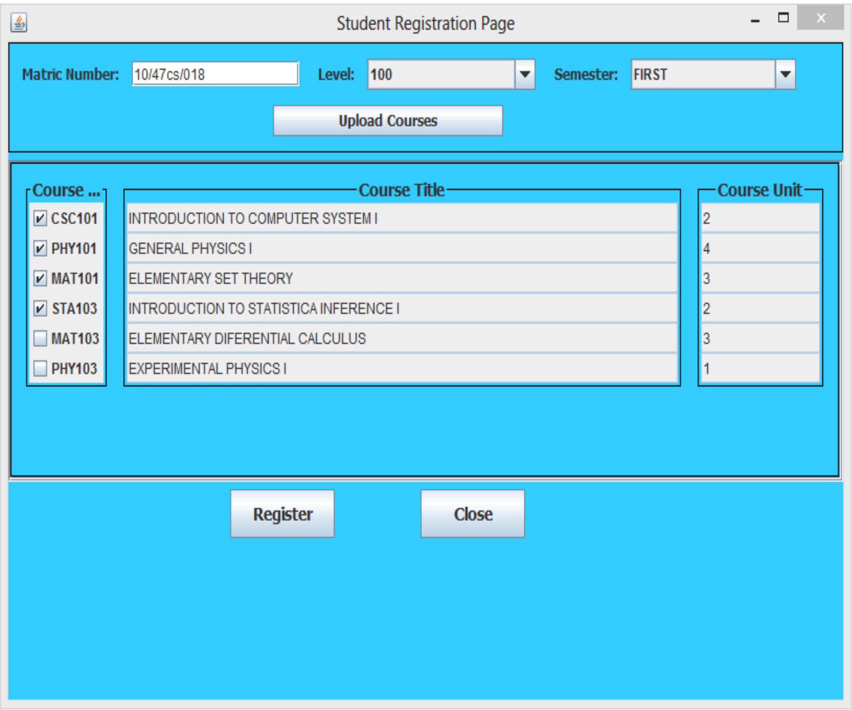
For instance:

- If a customer buys snacks, there may be 85%probability that the customer will also buy soft drinksor beer.
- If a person buys vacation airline tickets for an entirefamily, there may be 95% probability that he or shewill rent a full-size car at the vacation location.

4. Results and rules analysis

The design module of the proposed system includes:

i. **Student Registration Menu**: This is sub-menu of operation menu. This is where the courses taken by the students are registered. Figure 2 shows the student course registration window.

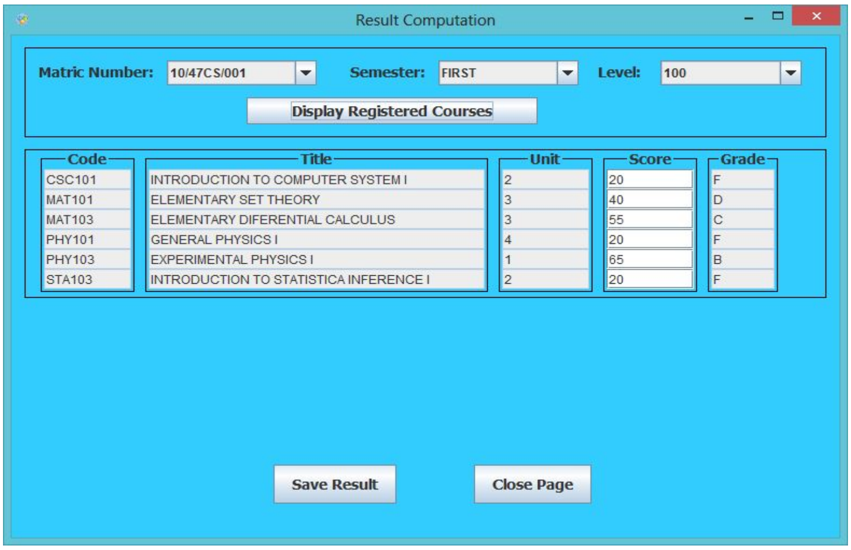


Matric Number: 10/47cs/018 Level: 100 Semester: FIRST

Course ...	Course Title	Course Unit
☑ CSC101	INTRODUCTION TO COMPUTER SYSTEM I	2
☑ PHY101	GENERAL PHYSICS I	4
☑ MAT101	ELEMENTARY SET THEORY	3
☑ STA103	INTRODUCTION TO STATISTICA INFERENCE I	2
☐ MAT103	ELEMENTARY DIFERENTIAL CALCULUS	3
☐ PHY103	EXPERIMENTAL PHYSICS I	1

Figure 2: Student Registration Window

ii. **Compute Result Menu**: This is another sub-menu of Operation menu, where all the results obtained by students in each courses are input into the system. Let us remember that the failed courses from the results are what the system will consider in discovering the patterns that exists between related courses. Figure 3 shows the result computation window.

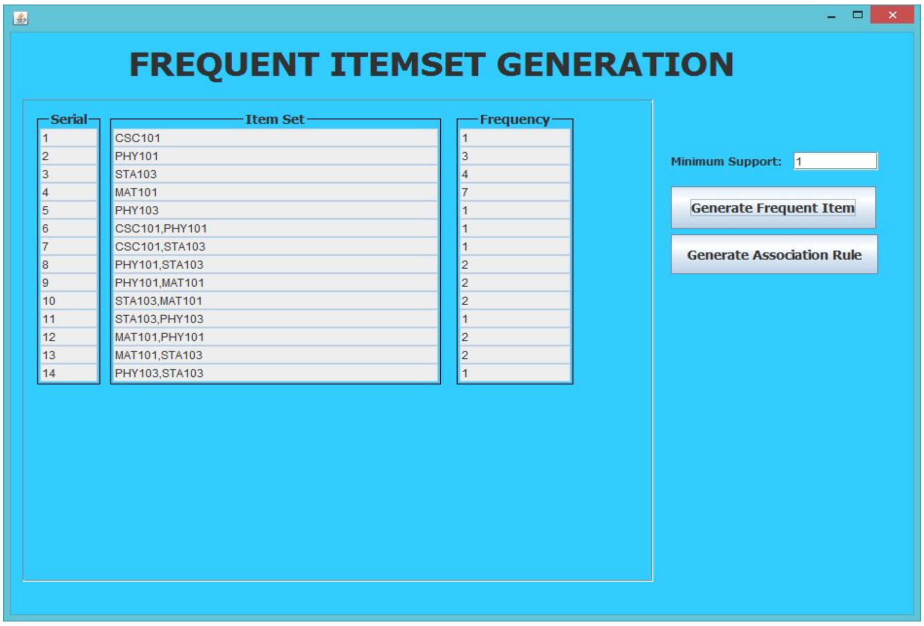


Matric Number: 10/47CS/001 Semester: FIRST Level: 100

Code	Title	Unit	Score	Grade
CSC101	INTRODUCTION TO COMPUTER SYSTEM I	2	20	F
MAT101	ELEMENTARY SET THEORY	3	40	D
MAT103	ELEMENTARY DIFERENTIAL CALCULUS	3	55	C
PHY101	GENERAL PHYSICS I	4	20	F
PHY103	EXPERIMENTAL PHYSICS I	1	65	B
STA103	INTRODUCTION TO STATISTICA INFERENCE I	2	20	F

Fig. 3: Result Computation Window

iii. **Mine Failed Courses Menu**: This is a sub-menu of Mining menu. This where the main objectives of the system will be performed, that is, generation of frequent itemsets and Association rule mining takes place. Figure 4 and 5show the frequent itemset generation and association rule mining respectively.

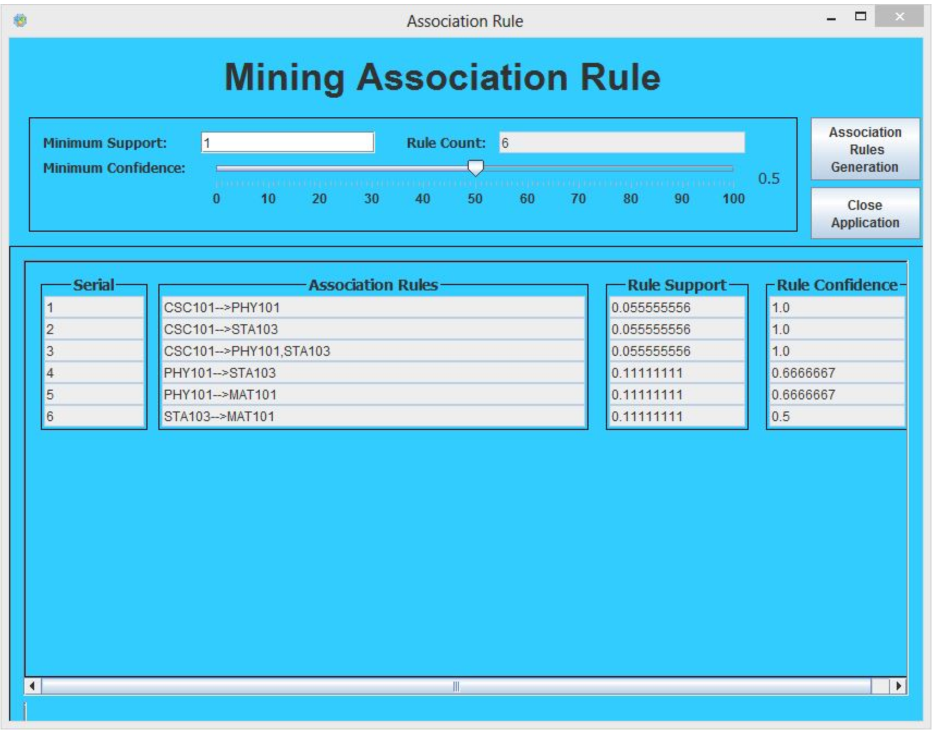


FREQUENT ITEMSET GENERATION

Minimum Support: 1

Serial	Item Set	Frequency
1	CSC101	1
2	PHY101	3
3	STA103	4
4	MAT101	7
5	PHY103	1
6	CSC101,PHY101	1
7	CSC101,STA103	1
8	PHY101,STA103	2
9	PHY101,MAT101	2
10	STA103,MAT101	2
11	STA103,PHY103	1
12	MAT101,PHY101	2
13	MAT101,STA103	2
14	PHY103,STA103	1

Figure 4: Frequent Itemset Generation Window



Association Rule

Mining Association Rule

Minimum Support: 1

Rule Count: 6

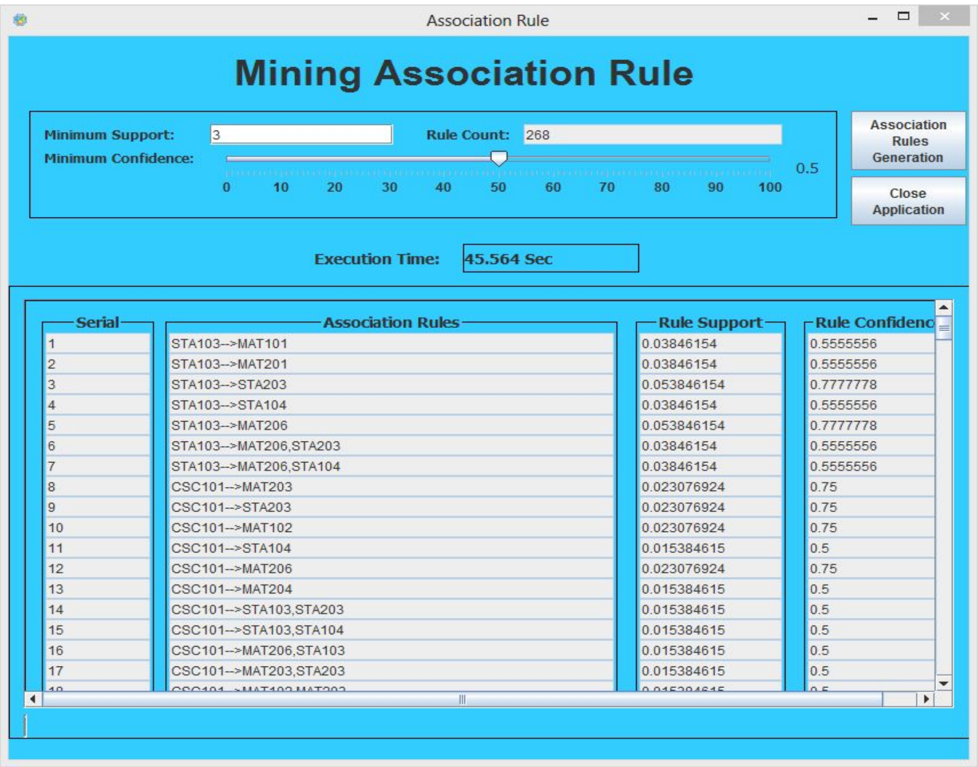
Minimum Confidence: 0.5

Serial	Association Rules	Rule Support	Rule Confidence
1	CSC101-->PHY101	0.055555556	1.0
2	CSC101-->STA103	0.055555556	1.0
3	CSC101-->PHY101,STA103	0.055555556	1.0
4	PHY101-->STA103	0.11111111	0.6666667
5	PHY101-->MAT101	0.11111111	0.6666667
6	STA103-->MAT101	0.11111111	0.5

Figure 5: Association Rule Mining Window

In this study, association rule mining analysis was performed based on students' failed courses. This identifies the hidden relationship between the failed courses, and suggests relevant causes of the failure to improve the low capacity students' performances. We observed that the lower the items minimum the larger the candidate generated. This adversely affects the complexity of the system.

Also, we observed that the execution time is also inversely proportional to minimum support, since it increases as minimum support decreases, which shows that an increase in system complexity and response time of the system as minimum support decreases. The figure 6 below shows the captured association rule mining process. For instance, in figure 6, if the item minimum support is 3 and the rule confidence is 0.5, 84 frequent item sets and 268 are rules generated.



Mining Association Rule

Minimum Support: 3

Rule Count: 268

Minimum Confidence: 0.5

Execution Time: 45.564 Sec

Serial	Association Rules	Rule Support	Rule Confidenc
1	STA103-->MAT101	0.03846154	0.5555556
2	STA103-->MAT201	0.03846154	0.5555556
3	STA103-->STA203	0.053846154	0.7777778
4	STA103-->STA104	0.03846154	0.5555556
5	STA103-->MAT206	0.053846154	0.7777778
6	STA103-->MAT206,STA203	0.03846154	0.5555556
7	STA103-->MAT206,STA104	0.03846154	0.5555556
8	CSC101-->MAT203	0.023076924	0.75
9	CSC101-->STA203	0.023076924	0.75
10	CSC101-->MAT102	0.023076924	0.75
11	CSC101-->STA104	0.015384615	0.5
12	CSC101-->MAT206	0.023076924	0.75
13	CSC101-->MAT204	0.015384615	0.5
14	CSC101-->STA103,STA203	0.015384615	0.5
15	CSC101-->STA103,STA104	0.015384615	0.5
16	CSC101-->MAT206,STA103	0.015384615	0.5
17	CSC101-->MAT203,STA203	0.015384615	0.5

Figure 6: Association Rule Mining Process

Table 1: Relationship between minimum confidence, minimum support and number of rules generated.

Minimum Item(s) Support = 3 Average Execution Time = 45sec			
Min. Conf.	No. of Rules	No. Frequent Itemsets	Exe. Time(seconds)
50%	268	84	45.564
60%	167	84	45.644
70%	21	84	45.620
80%	9	84	45.490
90%	6	84	45.472
100 %	6	84	45.349

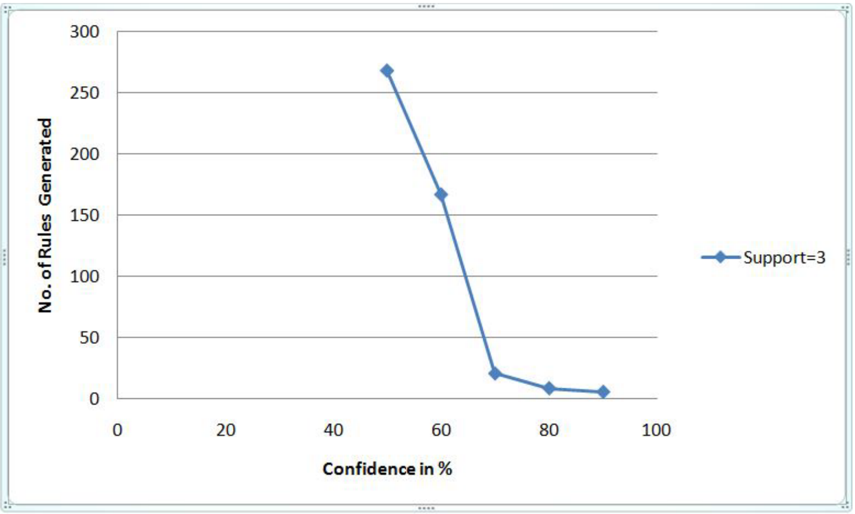


Figure: 7: Graphical representation of effect of minsup, minconf on Number of Rules

4.1 **Rule Analysis**

From figure 6, the rules with confidence 0.7 are strong rules, which implies that if a student fails the determinant (antecedent) course(s), such student will surely fail the dependent course(s). Such rules should not be overlooked when structuring the academic curriculum by academic planners. But if the rule support is high, it means that all the courses involved are failed together by most of the students considered. From rule number 1, we can see that STA103⇒MAT101 (s=0.03, c=0.5), this indicates that students who fails STA 103 will also fail MAT 101. This rule may not be very strong but it can be considered by academic planners because STA103 (STATISTICAL INFERENCE I) and MAT101 (LINEAR ALGEBRA I) are both mathematical courses. The two courses are related in such a way that a student who is not good in calculations will possibly fail all mathematical courses and related courses. From rule 2, STA103⇒MAT201 (s=0.03, c=0.5), the same decision that was considered in rule 1 can also be applied to rule 2. From rule 3, we can deduce that STA103⇒STA203(s=0.05, c=0.7), this indicates that a student that fails STA103 will also fail STA203. This rule is very strong because the confidence is high and the support is moderate. Since STA 103 is a 100 level course, while STA 203 is a 200 level course, Since the two courses are both "STATISTICAL COURSES", the academic planners can make the antecedent course (STA 103) a prerequisite to the dependent course(STA 203), that is, if a student has not passed STA103, such student will not be allowed to register for STA203 until he passes STA103. Also from rule 5 we can deduce that STA103⇒MAT206(s=0.05, c=0.7), this indicates that a student who fails STA103 will also fail MAT206. This rule is very strong and should be taken into consideration. The academic planners should structure the curriculum in such a way that a student who has not passed STA103 should not be allowed to register for MAT206. From rule 14, CSC101⇒STA103, STA203(s=0.01, c=0.5), this rule may not be considered by academic planners because it has an average confidence and the support is also low.

With all the observations above, if academic planners can make use of the extracted hidden patterns from students' failed courses using association rule mining approach, it will surely help them in restructuring the curriculum and help in monitoring students' ability and overall performance. This will also help academic planners in guiding students as to what course they should register for and not register for at a particular semester or session. This will definitely reduce the failure rate of students in different courses.

Table 2: An instance of the rules generated with support and confidence

Rule No.	Rule	Rule Support	Confidence
1	STA103⇒MAT101	0.03	0.5
2	STA103⇒MAT201	0.03	0.5
5	STA103⇒MAT206	0.05	0.7
14	CSC101⇒STA103, STA203	0.01	0.5

5. Conclusion

This study focused on how Association rule mining approach of data mining can be applied in an educational institution to discover hidden pattern or relationship that exist between students' failed courses. This research work has be able to explain the need for educational data analysis and shows the potential of association rule mining algorithm (Apriori algorithm) for improving the effectiveness of academic planners and level advisers in educational institutions in Nigeria, so that they can make better decisions that will improve students' academic performance.

A total number of 20 students were randomly selected in 25 courses registered for 100 and 200 level are considered as a case study. The final analysis revealed some interesting and hidden patterns of students' failed courses. It shows that some failed courses have a relationship that exist between them and some other courses. This relationship holds in such a way that if a student fails a particular course (determinant), the student will also fail the courses(dependent) that are related to that course. This could help academic planners in making academic decisions, restructuring and modification of curriculum which in turn improve the students' academic performance and reduce the rate at which students fail in different courses.

Most educational institutions are still using manual method of analyzing students' performance, other institutions use some software for their analysis. Nevertheless, the power of a data mining tool should not be underrated.

Besides developing a system proffered in this study for extracting interesting patterns or knowledge from an educational database, further research can be carried out, to know whether the accuracy of a particular data mining tool that uses a data mining technique is higher than another technique.

References

1. Abdulsalam Sulaiman O., BabatundeAkinbowale N., Hambali Moshood A. and Babatunde Ronke S. (2015). Comparative Analysis of Decision Tree Algorithms for Predicting Undergraduate Students’ Performance in Computer Programming. Journal of Advances in Scientific Research & Its Application (JASRA), 2, Pg. 79 – 92.
2. Abdsalam, S. O., Adewole, K. S., Akintola, A.G. and Hambali, M.A. (2014). Data Mining in Market Basket Transaction: An Association Rule Mining Approach. In International Journal of Applied Information Systems (IJAIS), 7(10): 15-20.
3. Al-Radaideh,Qasem A., Al-Shawakfa,Emad M., and Al-Najjar, Mustafa I. (2006).Mining Student Data Using Decision Trees. ACIT' 2006: The International Arab Conference on Information Technology.
4. Ayesha, S., Mustafa, T., Sattar, A. and Khan, I. (2010).Data Mining Model for Higher Education System. European Journal of Scientific Research, 43(1), pp. 24-29.
5. Baker, R. S. J. D., and Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1, pp. 3-17.
6. Bartik, V. (2009). Association based Classification for Relational Data and its Use in WebMining, CIDM ’09, IEEE Symposium on Computational Intelligence and DataMining, pp. 252-258.
7. Castro F., Vellido A., Nebot A., andMugicaF. (2007). Applying Data Mining Techniques to e-Learning Problems”. Evolution of Teaching and Learning Paradigms in Intelligent Environment ISBN: 10.1007/978-3-540-71974-8_8, 62, pp.183-221. Springer Berlin Heidelberg.
8. Cortez P. and Silva A. (2008). Using Data Mining to Predict Secondary Student Performance. In EUROSIS, A. Brito and J. Teixeira (Eds.), pp.5 -12.
9. Damaševicius R. (2009). Analysis of Academic Results for Informatics Course Improvement Using Association Rule Mining. Information Systems Development Towards a Service Provision Society, pp 357-363, published by Springer US.
10. Dogan B. and Camurcu A. Y. (2008). Association Rule Mining from an Intelligent Tutor. Journal of Educational Technology Systems, 36 (4), pp.433 – 447.
11. Fadzilah Siraj and Mansour Ali Abdoulha (2009). Uncovering hidden Information within University’s Student Enrollment Data using Data Mining. Third Asia International Conference on Modelling and Simulation.
12. Galit.et.al. (2007). Examining Online Learning Processes Based on Log Files Analysis: a Case Study. Research, Reflection and Innovations in Integrating ICT in Education.
13. Han, J. and Kamber, M. (2006). Data Mining: Concepts and Techniques, 2nd Edition, Morgan Kaufmann Publishers.

14. Hijazi, S. T. and Naqvi, R. S. M. M. (2006). Factors affecting student’s performance: A Case of Private Colleges. Bangladesh e-Journal of Sociology, 3(1).
15. Hongjie Sun (2010). Research on Student Learning Result System Based on Data Mining. International Journal of Computer Science and Network Security (IJCSNS), 10(4).
16. Khan Z. N. (2005). Scholastic achievement of higher secondary students in science stream. Journal of Social Sciences, 1(2), pp. 84-87.
17. Larissa, T. M. (2003). Introduction to Data Mining, London: Oxford University Press.
18. Lin, L. and Pei-qi, L. (2001). Study on an Improved Apriori Algorithm and its Application in Supermarket.
19. Oladipupo, O. O. and Oyelade, O. J. (2010). Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach, International Journal of Computer Science and Security, 4(2), pp.199-207.
20. QiangNiu, Shi-Xiong Xia, and Lei, Zhang (2009). Association Classification Based on Compactness of Rules, WKDD 2009, Second International Workshop on Knowledge Discovery and Data Mining, pp. 245-247.
21. Romero, C., and Ventura, S. (2007). Educational Data Mining: A Survey from 1995 to 2005. Expert Systems with Applications, 33, pp. 135-146.
22. Sumithra, R. and Paul, S. (2010). Using Distributed Apriori Association Rule and Classical Apriori Mining Algorithms for Grid Based Knowledge Discovery, International Conference on Computing Communication and Networking Technologies, pp. 1-5.
23. Tan, P. N., Steinbach M. and Kumar V. (2006). Introduction to Data Mining, Addison Wesley.
24. TisseraW.M.R., Athauda R.I.and FernandoH. C. (2006). Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining. IEEE International Conference on Information Acquisition.

Article received: 2017-01-24