

Application of Fuzzy Association Rule Mining for Analysing Students Academic Performance

Olufunke O. Oladipupo¹, Olanrewaju. J. Oyelade² and Dada. O. Aborisade³.

¹Department of Computer Science and Information Sciences
Covenant University, Ota, Ogun State, Nigeria

²Department of Computer Science and Information Sciences
Covenant University, Ota, Ogun State, Nigeria

³Department of Computer Science, Federal University of Agric, Abeokuta

Abstract

This study examines the relationship between students' pre-admission academic profile and academic performance. Data sample of students in the Department of Computer Science in one of Nigeria private Universities was used. The pre-admission academic profile considered includes 'O' level grades, University Matriculation Examination (UME) scores, and Post-UME scores. The academic performance is defined using students' Grade Point Average (GPA) at the end of a particular session. Fuzzy Association Rule Mining (FARM) was used to identify the hidden relationships that exist between students' pre-admission profile and academic performance. This study hopes to determine the academic profile of students who are most admitted in the session. It determines students' performance ratings as against their pre-admission academic profile. This can serve as a predictor for admission committee to enhance the quality of the new in-take and guide for the academic advisers.

Keywords: Data mining, Fuzzy Association Rule Mining, Student's Pre-admission academic profile, Academic Performance, Grade Point Average.

1. Introduction

In recent years a passionate concern has been expressed by the faculty and staff about the continuing decline in student academic performance in the universities. At the end of first semester from 2009/2010, 2008/2009, 2007/2008 admission process we have 30.5% ; 24.7% ; 20.6% of students below 2.5 respectively. This gives a clear view of continuing decline in student academic performance in the universities. This failure rate reduces as they move higher in levels because some of them must have been withdrawn or ask to repeat. It is also observed from experience that if a student at the end of 100 level has a large number of carry over courses with the grade

point average (GPA; a commonly used indicator of academic performance) less than 2.0, and is allowed to progress to 200 level, he will likely end up changing to another department, repeating or withdrawing, because of some fundamental problems. The question now is: what is actually responsible for the continuing decline in students' academic performance and how can we check the failure rate?

Many reasons have been deduced and several factors have been identified. These factors include socio-economic, demography, academic factors and stressors such as time management, financial problems, sleep deprivation, social activities and for some student having children can also pose treat to their academic performance [1],[2]. A number of study shows that female students' levels of academic performance were higher than their male counterparts irrespective of race [1],[3],[4]. Also some other factors such as test competence and academic competence, strategic studying, text anxiety can also play important factors in evaluating academic performance [5].

In responding to this question several studies have been carried out. In [3] students' cognitive data from high school and their non-cognitive self beliefs were examined to determine their impact on students' academic success. Neural network and other three modelling methodologies: logistic regression, discriminate analysis and structural equation modelling were used for the prediction. The relationship between academic integration and self efficacy with regard to institution types and student's major in two different programmes was determined in [1]. MANOVA was used to analyse the effect and the result was reported. A

research on academic profile of students failing in the first two years of medical school was carried out to identify the characteristics of medical students who fail at 100 and 200 levels. Age, sex, 'O' level grades, University Matriculation Exam (UME) scores, Pre-Degree Science (PDS) scores, 100 level cumulative grade points average (CGPA), 200 level Physiology scores and comprehensive examination results of student in 1999/2000 session were recorded to carry out the analysis. For the result analysis t-tests was carried out [6]. Also in [7] the correlation of admissions criteria with academic performance in Dental students' study was carried out to compare the admissions criteria as predictors of dental school performance in underachieving and normally tracking dental students. The analysis was carried out and evaluated with descriptive statistics, correlation and regression analysis. As a submission by Ritcher, academic performance is predicted by educational qualification and social skill. Experience and maturity are not significantly related to academic performance and pro-social behaviour. However, in some courses such as medicine the most widely acceptable identified factor is the quality of the pre-admission academic profile [8] of the admitted student. For instance a student who has a very low credit in mathematics might find it hard-hitting to break through in studying computer science or any engineering courses because of his weakness in mathematics. Such student ends up changing to another department. In all these reviews one thing is still found missing: to relate the present approved National University Commission (NUC) admission academic qualification criteria (Pre-admission academic profile) which include 'O' level results, UME scores and Post UME scores with the students' academic performance.

The aim of the present study is to identify the hidden relationships that exist between the students' pre-admission academic profile and their academic performance. This will obviously reveal the characteristics of students who are most admitted in a session and students that performed better. Also, the result will be helpful in determine the academic profile of students who are likely to repeat or may be advised to withdraw at the end of the first years and to determine the characteristic of students who are likely to have a high rating academic performance. At large this can serve as a predictor for admission committee to enhance the quality of the new in-take and guide for the academic advisers. To achieve this over the year, Data mining has proved to be effective in discovery of previously unknown, potentially useful and hidden knowledge in educational databases [9],[10],[11],[12],[13]. Association rule mining [14] is one of the best studied models for data mining. The discovery of association rules from,

databases in recent years has become an important and highly active research topic in the data mining field [15].

Association rule mining searched for interesting relationship among items in a given dataset. However, for a data mining analysis with quantitative dataset traditional association rule mining is limited because it is necessary to transform each quantitative attribute into discrete intervals. Over the year Fuzzy logic has been demonstrated to be effective in interpretation of these discrete intervals [16]. Fuzzy association rule mining efficiency has been prove in its application in different field including educational domain [17],[18],[19],[20]. Therefore, in this paper Fuzzy association rule mining techniques is used as the instrument for analysis because of the quantitative nature of the analysed data.

2. Methods

This study utilized a cross-sectional survey design and was conducted by administering a questionnaire to students 100 level of a private university in Nigeria. A non-probabilistic convenience sampling procedure was used. Participation in the study was mandatory to cover the entire 100 level student in analysis.

The survey instrument consisted of a single page with 8 items to obtain students pre-admission profile. Student reported their programme of study, age, O'level result in English, Mathematics and any other three subject of interest, Jamb score, Post Jamb interview score and their current Grade Point Average (GPA) at the time they completed the questionnaire. Cumulative GPA was the primary indicator of academic performance and was measure on a scale ranging from 0 to 5 in 2 significant figures. The O'level result was calibrated into 4 groups; distinction, credit, pass and fail. Joint Admission Matriculation Board (JAMB) score was calibrated into 4 groups; Low, Average, High and Very High. Post JAMB interview result was grouped into 3 groups; Poor, Average and Good. The grade point average was calibrated into 4 groups; First class, Second class upper division, Second class lower division and Third class. Table 1 shows the description of the analysed data.

For the analysis of the data Fuzzy Association Rule Mining apriori-like algorithm was adopted because of the quantitative nature of the analysed data [21]. Using the fuzzy set concept, the discovery rules are more understandable to human and also, fuzzy set soften the effect of sharp boundary problem [22]. During the mining attributes were considered as linguistic variables which include two O'level subjects results (English and Maths), the JAMB score, Post JAMB score and GPA.

GPA is the output linguistic variable and others serve as input linguistic variables. For each linguistic variable fuzzy set was defined. This comprises of linguistic values. This is shown in Table 2. In order to normalize the data value, fuzzy membership expressions are defined for each linguistic value as stated in equation 1-5.

Table1: Descriptions of Analyzed Dataset

Data Attribute	Ranges
O'Level Result	
<i>Distinction</i>	<i>A1-B3</i>
<i>Credit</i>	<i>C4-C6</i>
<i>Pass</i>	<i>D7-E8</i>
<i>Fail</i>	<i>F9</i>
JAMB Score (J)	
<i>Low</i>	$J < 180$
<i>Average</i>	$180 \leq J \leq 200$
<i>High</i>	$200 \leq J \leq 250$
<i>Very High</i>	$250 \leq J \leq 300$
Post JAMB Interview (P)	
<i>Poor</i>	$P < 50$
<i>Average</i>	$50 \leq P \leq 60$
<i>Good</i>	$60 \leq P \leq 100$
Grade Point Average (G)	
<i>1st Class</i>	$4.5 \leq G \leq 5$
<i>2¹</i>	$3.5 \leq G < 4.5$
<i>2²</i>	$2.5 \leq G < 3.5$
<i>3rd Class</i>	$1.5 \leq G < 2.5$
<i>Fail</i>	< 1.5

Table 2: Fuzzification of Linguistic variables

Linguistic Variables	Fuzzy set
English Lanuage	{EngDistinction, EngCredit, EngPass}
Mathematics	{MathsDistinction, MathsCredit, MathsPass}
JAMB Score	{Low, Average, High, VeryHigh}
Post JAMB Interview	{Poor, Average, Good}
Grade Point Average	{GPA1, GPA21, GPA22, GPA3}

For O'level English value, fuzzy membership expression will be as:

$$\begin{array}{lll}
 \text{EngDistinction} & A1 & \rightarrow 1.0 \\
 & B2 & \rightarrow 0.8 \\
 & B3 & \rightarrow 0.5 \\
 \\
 \text{EngCredit} & C4 & \rightarrow 1.0 \\
 & C5 & \rightarrow 0.8 \\
 & C6 & \rightarrow 0.5 \\
 \\
 \text{EngPass} & D7 & \rightarrow 1.0 \\
 & E8 & \rightarrow 0.5
 \end{array} \quad (1)$$

For O'level Maths value, fuzzy membership expression will be as:

$$\begin{array}{lll}
 \text{MathsDistinction} & A1 & \rightarrow 1.0 \\
 & B2 & \rightarrow 0.8 \\
 & B3 & \rightarrow 0.5 \\
 \\
 \text{MathsCredit} & C4 & \rightarrow 1.0 \\
 & C5 & \rightarrow 0.8 \\
 & C6 & \rightarrow 0.5 \\
 \\
 \text{MathsPass} & D7 & \rightarrow 1.0 \\
 & E8 & \rightarrow 0.5
 \end{array} \quad (2)$$

For JAMB score value (let J) fuzzy membership expressions using triangular membership function (trimf) will be as:

$$\mu_{\text{poor}}(j) = \begin{cases} 1 & j < 180 \\ \frac{200-j}{20} & 180 \leq j \leq 200 \end{cases}$$

$$\mu_{\text{average}}(j) = \begin{cases} \frac{j-180}{20} & 180 \leq j \leq 200 \\ 1 & j = 200 \\ \frac{250-j}{50} & 200 \leq j \leq 250 \end{cases}$$

$$\mu_{high}(j) = \begin{cases} \frac{j-200}{50} & 200 \leq j \leq 250 \\ 1 & j = 250 \\ \frac{400-j}{50} & 250 \leq j \leq 300 \end{cases}$$

$$\mu_{veryhigh}(j) = \begin{cases} \frac{j-250}{50} & 250 \leq j \leq 300 \\ 1 & 300 \leq j \leq 400 \end{cases}$$

(3)

For Post JAMB interview score value (let p) fuzzy membership expressions using triangular membership function (trimf) will be as:

$$\mu_{poor}(p) = \begin{cases} 1 & p < 50 \\ \frac{60-p}{10} & 50 \leq p \leq 60 \end{cases}$$

$$\mu_{average}(p) = \begin{cases} \frac{p-50}{10} & 50 \leq p \leq 60 \\ 1 & p = 60 \\ \frac{70-p}{10} & 60 \leq p \leq 70 \end{cases}$$

$$\mu_{good}(p) = \begin{cases} \frac{p-60}{10} & 60 \leq p \leq 70 \\ 1 & 70 \leq p \leq 100 \end{cases}$$

(4)

For Grade Point Average (GPA) grade (let G) fuzzy membership expressions using linear membership function will be as:

$$\mu_{GPA1}(g) = (2 \times g) - 9$$

$$\mu_{GPA21}(g) = \left(\frac{g}{0.99}\right) - 3.54$$

$$\mu_{GPA22}(g) = \left(\frac{g}{0.99}\right) - 2.53$$

$$\mu_{GPA3}(g) = \left(\frac{g}{0.99}\right) - 1.52$$

(5)

3. Results and Discussion

Eighty students were administered the questionnaire and 69 students completed the questionnaire for an initial response rate of 86.25%. All the 69 completed samples were included in the analysis. A correlation was run on the Jamb score and the Post Jamb score, the correlation result is 0.374. This indicates that there is a reasonable correlation between the two attribute but not very strong. It shows that to a greater extent a very good score in Jamb might not necessary indicate a good score in Post Jamb examination. This is an irony. This can be view and trace to the effect of corruption in the examination processes.

The data sample records were fuzzified to normalize the data to a range of [0,1]. This is done to avoid the sharp boundary problem in quantitative data set (22). Fig. 1 shows the snapshot for the fuzzification system. Total number of 432 rules was generated with 108 antecedent rules, this is determined based on the number of attributes and their subspaces. For each of the rules the percentage rule item support and the rule confidence values were determined as results from fuzzy association rule mining process. This is represented on Fig. 2.

The percentage item support (Fig. 2, column 2) reveals the probability nature of the type of candidates given admission in the session under consideration. It was observed that "EngCredit, MathsDis, JambScoreHigh, PostJambScoreAverage" has the higher support percentage of 99, The implication is that a higher numbers of candidate with Credit in English language, Distinction in Mathematic, Jamb score between 200-250 and Post Jamb score within the range of 50-60 were considered mostly for the admission that session.

Also, the rule with EngPass, MathPass, JambScoreVeryHigh, and PostJambScorePoor has 13% supports. This implies that candidates with Pass English, Pass in Mathematic and Jamb Score within the range of 250-300 and Post Jamb less than 50 were less considered for admission. This category of students can be suspected for cheating during the Jamb Exam. And if giving admission definitely they might not cope and end up being withdrawn from the Department.

The rule confidence percentage indicates the degree at which each rule antecedent implies the rule consequent. The rules antecedents show different combinations of pre-admission profile and the rule consequent is the academic performance measure by Grade Point Average

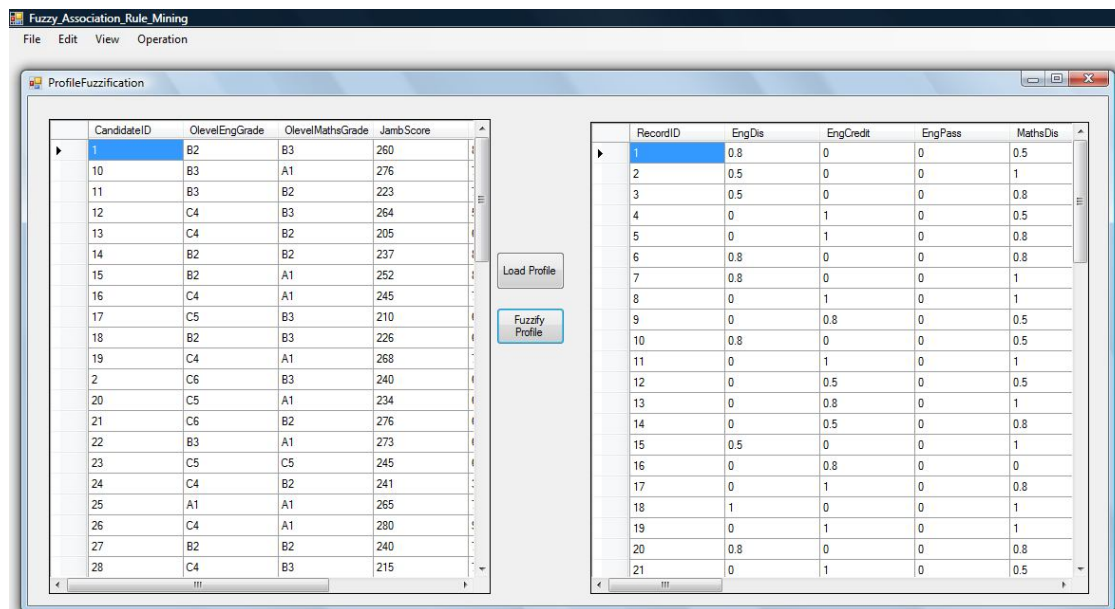


Fig. 1: The snapshot for fuzzification process

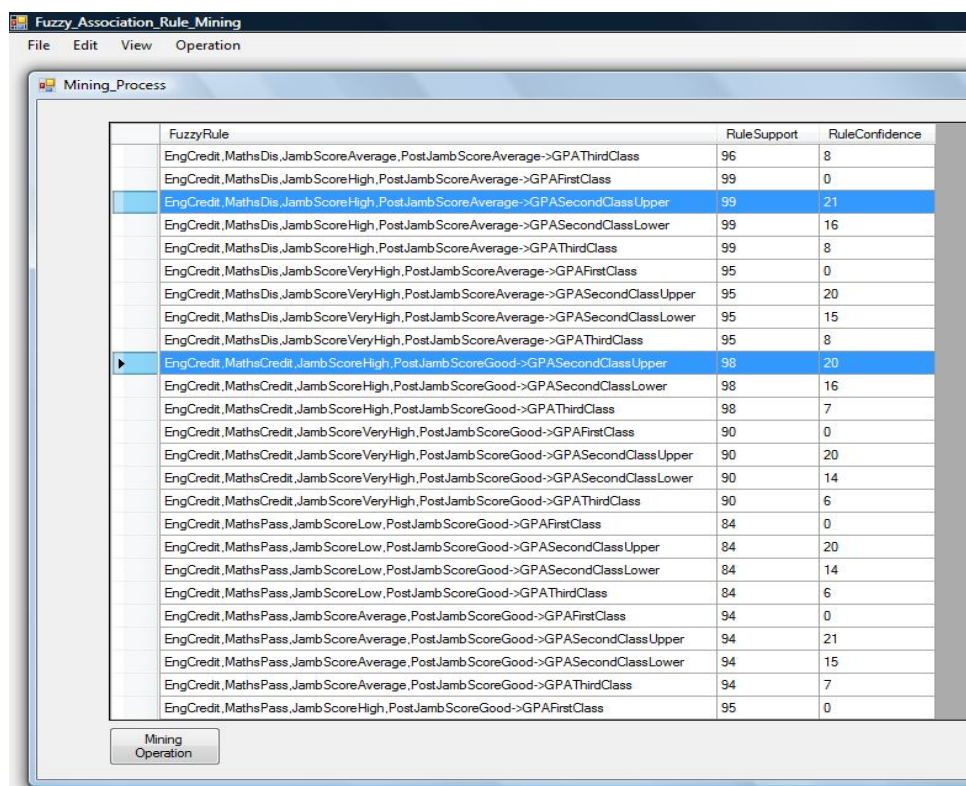


Fig. 2: Snapshot for mining process

(GPA). Therefore, the rule confidence reveals the extent at which each pre-admission profile combination yields a particular grade. Based on the data set used, it was observed that all the combinations implied First Class GPA with zero percentage. This means that none of the students is in First class as of the time of data collection. The academic implication of this is that the student pre-admission profile is not enough to determining the student academic performance in the higher institution. If a student with high profile did not work hard he/she might not be able to sustain the high academic performance. For instance, Table 3 shows the support and confidence of four set of candidates with highest percentage support. That is, the categories of candidate that were majorly admitted in the considered session. From Table 3, for the four groups, none of the students was in first class upon their pre-admission profile. They

are more in second class upper and second class lower. Notwithstanding, few of them still ended up with third class. The academic implication is that if they work harder some of them can still move up to first class and leave third class to second class. To a greater extent, one can actually say if better candidates are admitted it can give better result.

This fuzzy mining result, reveals the characteristic of student who are likely to repeat or likely to have high academic rating. It can also serve as a predictive model to the admission office in an institution to know the performance of their intake right from their year one. With this they can determine a corrective measure for subsequent admission processes. Also, this will serve as guidance to the level adviser for proper monitoring and effective advice for the students to enhance their academic performance.

Table 3: The rules support and confident table

Group	Rule	Sup	Conf.
A	EngCredit,MathsDis,JambScoreHigh,PostJambScoreAverage->GPAFirstClass	99	0
	EngCredit,MathsDis,JambScoreHigh,PostJambScoreAverage->GPASecondClassUpper	99	21
	EngCredit,MathsDis,JambScoreHigh,PostJambScoreAverage->GPASecondClassUpper	99	16
	EngCredit,MathsDis,JambScoreHigh,PostJambScoreAverage->GPASecondClassUpper	99	8
B	EngCredit,MathsCredit,JambScoreHigh,PostJambScoreGood->GPAFirstClass	98	0
	EngCredit,MathsCredit,JambScoreHigh,PostJambScoreGood->GPASecondClassUpper	98	21
	EngCredit,MathsCredit,JambScoreHigh,PostJambScoreGood->GPASecondClassLower	98	16
	EngCredit,MathsCredit,JambScoreHigh,PostJambScoreGood->GPAThirdClass	98	7
C	EngCredit,MathsDis,JambScoreAverage,PostJambScoreGood->GPAFirstClass	98	0
	EngCredit,MathsDis,JambScoreAverage,PostJambScoreGood->GPASecondClassUpper	98	21
	EngCredit,MathsDis,JambScoreAverage,PostJambScoreGood->GPASecondClassLower	98	16
	EngCredit,MathsDis,JambScoreAverage,PostJambScoreGood->GPAThirdClass	98	8
D	EngCredit,MathsDis,JambScoreHigh,PostJambScoreGood->GPAFirstClass	98	0
	EngCredit,MathsDis,JambScoreHigh,PostJambScoreGood->GPASecondClassUpper	98	21
	EngCredit,MathsCredit,JambScoreHigh,PostJambScoreGood->GPASecondClassLower	98	16
	EngCredit,MathsCredit,JambScoreHigh,PostJambScoreGood->GPAThirdClass	98	7

4. Conclusion

Result of this study underline the importance of new student pre-admission profile in evaluating academic success. Focusing efforts to utilize Fuzzy Association Rule Mining technique in analyzing student profile would be helpful for admission office in determining the characteristics and profile of candidate to be considered for admission. Also, it would intimate the level advisers the basis upon which they could monitor each student academic performance appropriately.

References.

- [1] F. Weng, F. Cheong and C. Cheong, "IT education in Taiwan relationship between Self-efficacy and Academic Integration among Students" Pacific Asia Conference on Information Systems (PACIS), Association for Information Systems, 2009, 22
- [2] L. P. Womble, "Impact of Stress Factors on College Students Academic Performance" Undergraduate Journal of Psychology. 2003;16
- [3] J.J.J. Lin, P.K. Imbrie and K.J. Reid "Student Retention Modelling: An Evaluation of Different Methods and their Impact on Prediction Results" Proceedings of the Research in Engineering Education Symposium 2009, Palm Cove, QOD, pp. 1-6.
- [4] M.M. Campbell "Motivational Systems Theory and the Academic Performance of College Students" Journal of College Teaching & Learning, , vol 4(7), 2007, pp. 11-24.
- [5] S.S. Sansgiry and K. Sail "Factors that affect Academic Performance among Pharmacy Students" American Journal of Pharmaceutical Education ; 70 (5) Article 104, 2006,pp. 1-9
- [6] A.O. Afolabi, V.O. Mobayoje, V.A. Togun and A.S. Oyadeyi "The Academic Profile of Students Failing in the First Two Years of Medical School" Middle-East Journal of Scientific Research 2 (2): 2007, pp. 43-47, ISSN 1990-9233 © IDOSI Publications, 2007
- [7] D.A. Curtis, S.L. Lind, O. Plesh and F.C. Finzen "Correlation of Admissions Criteria with Academic Performance in Dental Students" American Dental Education Association: *Critical Issues in Dental Education* ,Dent Educ 71(10), 2007, 1314-1321.
- [8] K.L. McCall, E.J. MacLaughlin, D.S. Fike and D. Ruix "Preadmission Predictors of PharmD Graduates' Performance on the NAPLEX" American Journal of Pharmaceutical Education 2007; 71 (1) Article 05, 1-7
- [9] C. Kuok, A. Fu, and M. Wong " Mining fuzzy association rules in databases". SIGMOD Record 17(1): 1999, pp 41-6.
- [10] L.M. ElFangary "Mining of Egyptian Missions Data for Shaping New Paradigms" International Journal of Engineering and Technology Vol.1(1), 2009, pp.14-22.
- [11] C. Vialardi, J. Bravo, L. Shafti and, A. Ortigosa "Recommendation in Higher Education Using Data Mining Techniques" Educational Data Mining, 2009, pp. 190-199
- [12] S. Herzog "Estimating Student Retention and Degree-Completion Time: Decision Trees and Neural Networks Vis-à-Vis Regression" NEW DIRECTIONS FOR INSTITUTIONAL RESEARCH, no. 131, Fall 2006 © Wiley Periodicals, Inc. Published online in Wiley InterScience.
- [13] A.B. Adeyemo and G. Kuye "Mining Students' Academic Performance using Decision Tree Algorithms" Journal of Information Technology Impact Vol. 6, No. 3, 2006, pp. 161-170,
- [14] R Agrawal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases", Proc. ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'93), Washington, DC, July,1993, pp.45-51.
- [15] Y. He, Y. Tang, Y. Zhang and R. Sunderraman "Adaptive Fuzzy Association Rule mining for effective decision support in biomedical applications" *Int. J. Data Mining and Bioinformatics*, Vol. 1, No. 1, 3-18. Copyright © 2006 Inderscience Enterprises .

- [16] M. Delgado , D. Sanchez, . M.J Martin-Bautista and MA. Vila “Mining association rules with improved semantics in medical database” *Artificial Intelligence in Medicine* 21,2001, pp.241-245.
- [17] J. Luo and S.M. Bridges “Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection”, *International Journal of Intelligent Systems*, Vol. 15, No. 8, 2000, pp.687–704.
- [18] S.S.C. Wong and S. Pal “Mining fuzzy association rules for web access case adaptation” ,*Workshop on Soft Computing in Case-Based Reasoning, International Conference on Case-Based Reasoning (ICCBR’01)*.2001
- [19] W.H. Au and K.C.C. Chan ‘Mining fuzzy association rules in a bank-account database’, *IEEE Transactions on Fuzzy Systems*, Vol. 11, No. 2, 2003, pp.238–248.
- [20] O.O. Oladipupo and O.J. Oyelade “Knowledge Discovery from Students’ Result Repository: Association Rule Mining Approach” *International Journal of Computer Science and Security* Vol.4 (2). CSC press, CSC Journals. 2010, pp.199-207.
- [21] A. Gyenesei., “A fuzzy approach for mining quantitative association rules”. *Acta Cybernetica*, 15:2001, pp.305–320.
- [22] O.O. Oladipupo, O.C Uwadia and C.K. Ayo) “On sharp boundary problem in rule-based expert systems in medical domain”. *International journal of Healthcare Information Systems and Informatics*, 5(3) , 2010, pp.14-26.

Oyelade, O. J. is a senior faculty member in the department of Computer and Information Sciences, Covenant University, Ota, Nigeria. His research interests are in Bioinformatics, Clustering, Fuzzy logic and Algorithms. He is a member of International Society for Computational Biology (ISCB), Africa Society for Bioinformatics and Computational Biology (ASBCB), Nigeria Society of Bioinformatics and Computational Biology (NISBCB), the Nigerian Computer Society (NCS), and Computer Professional Registration Council of Nigeria (CPN).

Aborisade Dada Olaniyi holds Bsc, Msc in Computer Science. He is a lecturer in the Department of Computer Science, College of Natural Sciences, Federal University of Agriculture, Abeokuta (FUNAAB), Nigeria. He has ten years of teaching with four years of Teaching/Research in University Environment. His research interests are Human Computer Interaction (HCI), Information Security and Fuzzy logic. He is a member of Microsoft IT Academy, Nigeria Computer Society (NCS) and Microsoft Certified Professional (MCP)

Olufunke. O. Oladipupo recieved her Bachelor degree in Computer Science, 1999, M.Sc degree in Computer Science, 2004 and Ph. D in Computer Science. **Dr. Oladipupo, O. O.** is a senior faculty member in the Department of Computer and Information Sciences, Covenant University, Ota, Nigeria. Her research interests include Artificial Intelligent, Data Mining, and Soft Computing Technique. She is a member of Nigerian Computer Society (NCS), and Computer Professional Registration Council of Nigeria (CPN).

Olanrewaju Jelili Oyelade recieved his Bachelor degree in Computer Science with Mathematics (Combined Hons) and M.Sc degree in Computer Science from Obafemi Awolowo University, Ile-Ife, Nigeria. He obtained his Ph. D in Covenant University, Ota, Nigeria. **Dr.**