SPACS: Students' Performance Analysis and Counseling System using Fuzzy logic and Association Rule Mining

Ritu Banswal
PG Student
Department of Computer Science and Engineering
Lovely Professional University, Punjab

Vishu Madaan
Assistant Professor
Department of Computer Science and Engineering
Lovely Professional University, Punjab

ABSTRACT

Education plays an important role in one's life if the right education and right environment is not provided then it somewhat distract the person from his/her path. In this regard, fuzzy expert system known as SPACS (Students' Performance Analysis and Counselling System) is designed which helps to analyze the students' academic performance by identifying all the critical factors and counsels him/her against those affecting parameters. In this paper, the study highlights on the critical factors that degrades the academic performance of the student based on the some interesting relationships between those factors. The performance of the overall system depends on the accuracy of the fuzzy knowledge base. Hence, association rule mining technique helps to generate the accurate results and also decreases the complexity of the system.

General Terms

Data mining, Expert System, quality education, academic performance, and fuzzy rule based expert system.

Keywords

Association rule mining, fuzzy expert system, fuzzy logic, apriori algorithm, itemsets, fuzzy set.

1. INTRODUCTION

As we know education plays an important role in one's life for the overall growth in the society. If the right education and right environment is not provided then it somewhat distracts the person from his/her path. There would be arising some factors like university environment, family issues, teaching factor, personal factor etc that affects the overall performance of the student. For solving this current problem, we designed an automated system which helps to analyze and identify all the critical factors that affects the performance of the student and then counsels him/her against those affecting parameters. The integration of data mining and expert system results best approach our study we designed an expert system known as SPACS which works on the data sample of students collecting within the LPU, Punjab and makes use of fuzzy logic and association rule mining technique [1].

The expert system helps to take effective decision by using the expertise knowledge but it somehow not able to provide precise solution for the real problems. So the concept of fuzzy logic is used with expert system to enhance its capabilities and becomes fuzzy expert system. But there is some limitation while working with fuzzy expert system that it would not be able to find all the hidden rules that occur frequently and also the performance of the system depends on the accuracy of the fuzzy rules. So for this association rule mining approach is used to find the hidden patterns that having relationships among the datasets and generates the strong association rules thus generates the accurate fuzzy rules for the fuzzy expert system. On the basis of that the system would be able to analyze the performance of the student inconsistently increasing, inconsistently decreasing, consistent increasing, consistent decreasing or remains consistent.

The idea comes from the previous researches on FARM technique [3] but in our present work, we do the reverse of FARM process and applied it on the application to analyze the student performance with the help of fuzzy expert system.

In 2014, Prateek Agarwal *et.al* proposed fuzzy rule based expert system which evaluates the overall performance of the students using fuzzy logic [2]. They have considered the five major factors that affect the performance. These are teaching factor, university system, university environment, personal factor, and family issues. On the basis of that a fuzzy expert system is designed that realizes the purpose of education but this system not fully efficient to generate the results. Its capabilities are not much enhanced using fuzzy approach.

In 2012, Olufunke O. Oladipupo *et.al* examines the students' performance ratings as against their pre-admission academic profile and relationship between the academic performances [3]. The study focuses to utilize fuzzy association rule mining technique in analyzing student profile that helps to take the admission according to his characteristics and the profile of the candidate. Fuzzy association rule mining (FARM) was used to analyze and identify the hidden relationships between the student's admission profile and academic performance. This study helps to determine the academic profile of those students that are most admitted in the session and thus intimate the advisors to monitor the performance of student. The proposed approach uses the fuzzy logic for uncertain data set and then applies association rule mining to obtain and analyzing the results. The academic performance is evaluated using the Grade Point Average (GPA) for each particular session.

In 2012 Hamid Eslami Nosratabadi *et.al* proposed the fuzzy expert system that classifies the customers of the bank using classification rules and association rules with the help of Apriori algorithm and CRISP-DM process, specifies the Credit Degree of banks' customers [7][8]. It combines the extracted rules of association mining and the knowledge of experts. The CRISP-DM (Cross-Industry Standard Process for Data Mining) model method is followed by this standard to study the 436 records of customers collected for the Saman Iranian Bank. It considers the four financial factors related to the customers, Current ratio, Debit ratio, Net Benefit Ratio, Claims collection period are inputs of database and the credit degree of customers as the output. The association rules are

used to classify the data using Apriori algorithm [1]. Then they concluded fuzzy expert system has been created based on these selected rules for evaluating the credit degree of Bank's customers.

In 2014 Charanjit Bambrah *et.al* proposes the Association rule mining to analyze the performance of the students and to enhance the quality of education in the educational institutions [4]. Apriori algorithm is used to extracts the set of rules, analyzes the given data to classify the student based on their performance in academics. Students are classified based on assignment, internal assessment, tests, attendance etc., for analyzing the performance of the student whether good, poor, or excellent. Then a report is generated to improve the result of student. But in this classical Association rule mining is not efficient in terms of processing time, accuracy, redundant rule generation, etc.

In 2014, Hua Wang *et.al* studied the application of improved apriori algorithm in the course management system [5]. It mines the relationships between the courses using improved and classical apriori algorithm to analyze tha data and compares the performance.

In 2013, K. Rajeswari Professor *et.al* designs a decision support system for medical diagnosis of patients and assists the doctors to predict risk of disease [6]. Medical data is used for mining frequent patterns using association rule algorithm, further genetic and fuzzy logic will optimize those generated rules to take effective decisions. First the data is given to the apriori algorithm to generate the interesting patterns and association between the items that is hidden in the data. Then the results are optimised with the help of genetic algorithm and further given to the fuzzy logic module. From this module a knowledge discovery is designed to predict the risk of disease.

2. PROBLEM

Association rule mining helps to find the hidden association rules which would definitely affecting the student's performance and on the basis of this fuzzy logic would be applied to develop a fuzzy expert system which helps to analyze and counsels the student's academic performance. The main objective of this study is to decrease the complexity of the fuzzy expert system. Here, the system works with the certain data for mining and then deals with the uncertainty.

2.1 Association Rule Mining

Association rule mining [1] is most widely used data mining technique in all real life applications like in business analysis, identifying the customer behaviour, medical diagnosis etc. This mining technique is used to find the hidden and interesting itemsets that occurs frequently and have some relationships among the large datasets. The strong association rules are generated being a complex task if they satisfy the minimum support threshold and minimum confidence threshold. The association rule is defined as a state when X occurs then Y also occurs having confidence level i.e.

$X \rightarrow Y$,

where $X \cap Y \neq \emptyset$.

Rule Evaluation:

Support: It defines the percentage (s %) of transactions in D contains the union of both X and Y itemsets.

$$\text{Support}(X \rightarrow Y) = P(X \cup Y)$$

Confidence: It defines the percentage (c %) of transactions in

D having X that also contain Y itemset.

$$\text{Confidence } (X \rightarrow Y) = P(X \mid Y)$$

$$= \text{Support}(X \cup Y)/\text{support}(X)$$

In this paper we use classical apriori algorithm to discover hidden association rules that occurs frequently in the data. Apriori algorithm is the most significant and efficient algorithm to mine the association rules. It uses the prior knowledge and follows the bottom up approach in which k itemsets used to generate k+1 itemsets.

The algorithm involves two concepts

- Find all frequent item sets that have minimum support.

- Based on frequent itemsets, generating the strong association rules.

The pseudo-code of apriori algorithm as shown in the figure 2.1 and it works under prescribed procedure.

Works with two procedures

- *Join*: The candidate sets are generated by joining the frequent item sets in step-wise.

- *Prune:* Discard the frequent items if subset is not frequent.

$\text{Apriori}(T, \epsilon)$
$\quad L_1 \leftarrow \{\text{large } 1 - \text{itemsets}\}$
$\quad k \leftarrow 2$
$\quad \textbf{while } L_{k-1} \neq \emptyset$
$\quad\quad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\}$
$\quad\quad \textbf{for } \text{transactions } t \in T$
$\quad\quad\quad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$
$\quad\quad\quad \textbf{for } \text{candidates } c \in C_t$
$\quad\quad\quad\quad count[c] \leftarrow count[c] + 1$
$\quad\quad L_k \leftarrow \{c \mid c \in C_k \wedge count[c] \geq \epsilon\}$
$\quad\quad k \leftarrow k + 1$
$\quad \textbf{return } \bigcup_k L_k$

Figure 2.1 Pseudo-code of apriori algorithm [1]

Working of Apriori algorithm

1. Scan the database for identifying frequent items.

2. Candidate generation is done by joining process which helps to find the frequent itemsets at different levels.

3. Pruning is done at each level if frequent itemsets is not found and also does not satisfy the minimum support threshold.

4. Terminate the algorithm if there are no frequent itemsets left.

5. Finally on the basis of resulted frequent itemsets, strong association rules are generated which holds the condition that it satisfy the both minimum support and confidence threshold value.

2.2 Fuzzy Expert System

Fuzzy expert system is an expert system that defines the membership functions and rules using fuzzy logic. The rules are defined in the form of IF-Then statements.

"If value of A is low and value of B is high then C= Medium",
where A and B works as input, C as an output.

In this low, high and medium are the fuzzy sets defined on A, B & C resp. The set of rules defined in fuzzy expert system is stored in knowledge base according to the particular domain. These rules are in human readable form and they are easy to understand and interpret the results. The fuzzy expert system is somewhat similar to traditional expert system but the difference lays in that it uses fuzzy set theory while inference the rules. Fuzzy Logic is an expert system technique that deals with uncertain data. The uncertain data defines in terms of 0 and 1.

Two Concepts involved in Fuzzy Logic:
Linguistic Variables: are the variables whose values may either words or sentences in a natural and synthetic language.

Fuzzy If-Then Rule: in which the antecedent and consequent are the basic propositions that contain linguistic variables.

Fuzzy logic includes the concept of fuzzy sets which is useful for representing linguistic variables. A fuzzy set Z in X is defined by the function $\mu_z(x)$.

$$Z = \{x, \mu_z(x) \mid x \in X\}$$

where $\mu_z(x)$ is called the membership function of x in Z. This function maps each element of X to a membership value between 0 and 1. A membership function is a curve which shows the mapping of an input space to a membership value between 0 and 1.

Components of Fuzzy Expert System

The development of fuzzy rule based expert system consists of fuzzification, inference, knowledge base, defuzzification processes as shown in figure 2.2.
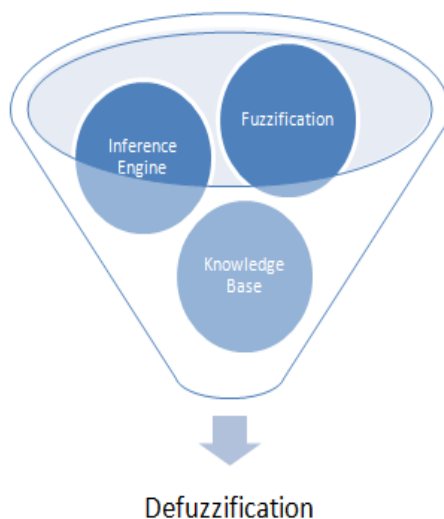

Figure 2.2 Fuzzy rule based expert system

- **Fuzzification** is a process of transforming the crisp values into degree of membership functions for linguistic variables of fuzzy sets.

- **Fuzzy Knowledge Base:** The Fuzzy system takes effective decisions with the help of this fuzzy knowledge base. The knowledge base consists of facts and rules about the particular domain. The rules are defined in the form of fuzzy if-then rules and the knowledge is being collected from domain experts.

- **Fuzzy Inference Engine:** The Fuzzy expert system provides a user interface system with the help of inference engine. The facts are defined by the user and then inference engine provides expertise solution by applying some logic based on fuzzy knowledge base.

- **Defuzzification** is an inverse transformation process that converts the fuzzy sets into crisp set. The output obtained is a fuzzy set so to obtain a single output value, defuzzification is needed. Different methods like centroid, max membership, bisector, SOM are used to defuzzify the fuzzy set.

The advantage of the integration of fuzzy logic with expert system is that it enhances the decision making capability of the expert system.

3. METHODOLOGY
The methodology first uses the association rule mining and then fuzzy logic approach. The process is shown in figure3.1.
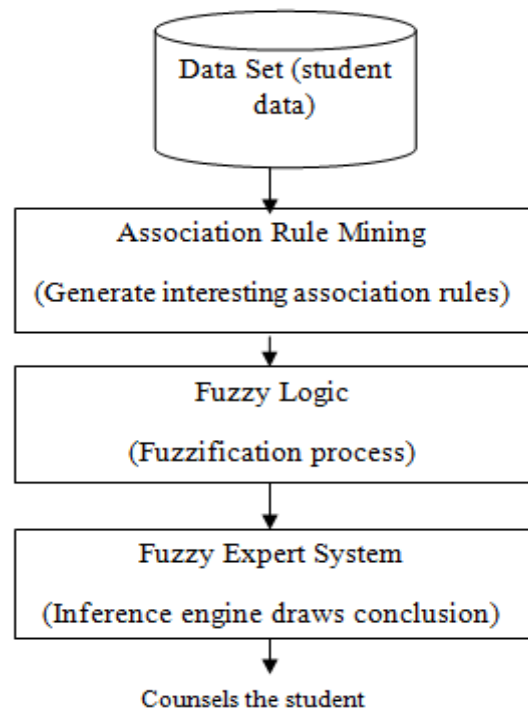

Figure3.1 Process of Research Methodology

Step 1:- Clarify the objective
Here, the main objective of our proposed work is to decrease the complexity and reduce the computation time of the existing work [2].The study focuses on the application that is going to analyze the academic performance of the students which would automatically tells whether the performance of the student will increase, decrease or remains consistent. According to these factors it would also counsels the student for his/her personal growth.

Step 2:- Data collection
The present study works on the existing student data that is collected within Lovely Professional University (LPU) Punjab by one to one interaction with students and highly experienced authorities of the university. The reason is to identify the critical factors that may affect the students' academic performance [2].

Step 3:- Data Preparation

From the data collection step, we have collected the 569 records of the students for analysis. As data is not appropriate for mining, it goes under the data pre-processing step. Here, considered the 36 factors related to students with 569 records as transactional data. Each factor represented as items for mining the association rules. 36 factors that degrades the student performance includes confidence level, hesitation, university environment, teaching quality, teaching experience, etc

Step 4:- Association rule mining

Based on the dependencies between the itemsets, association rules have been generated and selected as the useful and proper rules. Classical apriori algorithm is used for mining and is implemented in MATLAB programming tool.

In our research, the input is taken as $m \times n$ matrix of binary transactions where m is the no. of transactions and n is the no. of attributes in the dataset. The algorithm works under the min_support of 0.1 and with 0.7 of min_confidence. These values are user defined and we have selected the average range for the minimum support and minimum confidence threshold. Under these considerations the algorithm mines the frequent item sets that occur frequently in the dataset and from these patterns association rules are generated. The rules are as shown in the figure 3.2.

```
1   Rule   (Support, Confidence)
2   factor 27 -> factor 32   (16.7254%, 71.9697%)
3   factor 25,factor 33 -> factor 23   (16.0211%, 71.6535%)
4   factor 33,factor 36 -> factor 32   (14.9648%, 72.0339%)
5   factor 25,factor 32 -> factor 23   (14.6127%, 72.1739%)
6   factor 2,factor 5 -> factor 33   (13.9085%, 73.1481%)
7   factor 7 -> factor 33   (13.2042%, 70.7547%)
8   factor 13 -> factor 33   (12.1479%, 71.134%)
9   factor 21 -> factor 32   (12.1479%, 75%)
10  factor 24 -> factor 26   (12.1479%, 72.6316%)
11  factor 27,factor 32 -> factor 33   (12.1479%, 72.6316%)
12  factor 27,factor 33 -> factor 32   (12.1479%, 75%)
13  factor 12 -> factor 32   (11.7958%, 71.2766%)
14  factor 12 -> factor 23   (11.6197%, 70.2128%)
15  factor 5,factor 19 -> factor 33   (11.6197%, 72.5275%)
16  factor 21 -> factor 26   (11.4437%, 70.6522%)
17  factor 1 -> factor 32   (11.2676%, 71.9101%)
18  factor 23,factor 27 -> factor 32   (11.0915%, 73.2558%)
19  factor 5,factor 23,factor 32 -> factor 33   (10.5634%, 70.5882%)
20  factor 11 -> factor 23   (10.3873%, 72.8395%)
21  factor 26,factor 27 -> factor 32   (10.3873%, 73.75%)
22  factor 11 -> factor 26   (10.2113%, 71.6049%)
23  factor 2,factor 19 -> factor 33   (10.2113%, 72.5%)
24  factor 5,factor 25 -> factor 33   (10.2113%, 71.6049%)
25  factor 26,factor 27 -> factor 33   (10.0352%, 71.25%)
26  factor 2,factor 23,factor 32 -> factor 33   (10.0352%, 70.3704%)
27  factor 25,factor 32,factor 33 -> factor 23   (10.0352%, 77.027%)
```

Figure 3.2 Association rules

Step5: Fuzzy Expert System

This step involves the fuzzification process to transform the crisp values into degree of membership functions for linguistic variables of fuzzy sets.

The factors that occur frequently in the mining process are selected as input variables for fuzzy sets. In our study, we have extracted the 17 factors as input like communication phobia, confidence level, and hesitation, bad company, financial issues etc from the 36 factors. Now these factors are categorized into five major categories as shown in table3.1. These major categories are taken as inputs for the fuzzy inference system and performance factor as output.

Table 3.1 Categorization of input variables

Input variables (Major factors)	Sub-factors
Personal factor	*communication phobia, confidence level, hesitation, communication phobia, confidence level, hesitation, bad company*
Family factor	*financial issues*
University factor	*distance among blocks, university atmosphere, overcrowded strength of students, campus disturbances*
Academic factor	*placement policy, examination pattern, academic schemes*
Teaching factor	*Teaching quality, teaching experience, teaching style.*

Once the input and output variables are defined then fuzzy sets will be specified using the concept of fuzzy logic. The variables are fuzzified with the help of defined membership functions. The calculation of output remains transparent to the user and is based on fuzzy logic. In our study, Mamdani's fuzzy inference engine is used.

Fuzzy sets

In this system, for each input factor three fuzzy sets are defined i.e. low, medium, high and for output five fuzzy sets are defined i.e. inconsistent increase, inconsistent decrease, consistent increase, consistent increase, and consistent. Triangular shaped built-in membership function is used to plot the fuzzy sets is described as "trimf" function. The values of the fuzzy sets for each input and output variables are as shown in the table 3.2.

Table 3.2 Fuzzy linguistic variables and their membership values

Factors	Membership range for fuzzy sets				
	Low		Medium		High
Input Variable	0-30		20-60		50-100
	Consistent Decrease	Inconsistent Decrease	Consistent	Inconsistent Increase	Consistent Increase
Output Variable	0-30	20-50	40-60	50-80	70-100

Fuzzy rules

The knowledge base of fuzzy expert system stores knowledge in the form of fuzzy rules and the system draws conclusion with the help of these rules. The performance of the overall system depends on the accuracy of the fuzzy rules.

In this system, 204 fuzzy rules are generated based on the results of association rules. The results are shown in figure3.3.
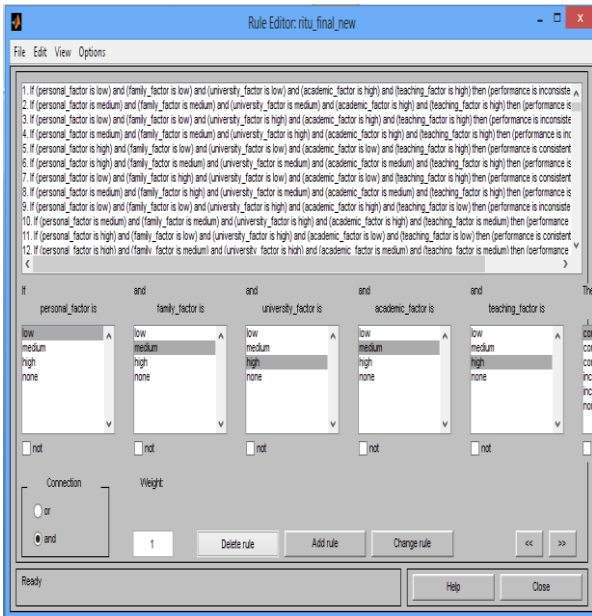
Figure3.3 Rule Editor

Rules are viewed under the rule viewer that helps to edit the values of each input as to check the possible output. The results are shown in figure 3.4.
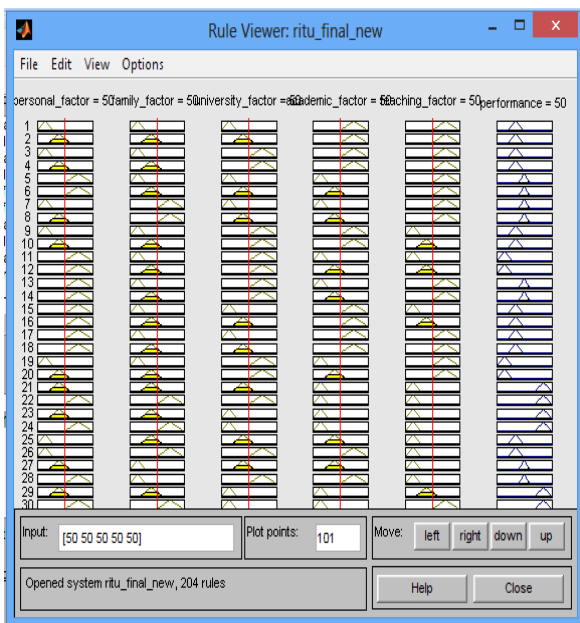

Figure 3.4 Rule viewer

Defuzzification

In this step defuzzification of the fuzzy output values are done. Different methods like centriod, max membership, bisector, SOM are used to test the robustness of the rules and knowledge engineering. The test shows that the system performs better for all methods but not on SOM method because the ranking does not match with the other results as shown in table 4.3.

Table 4.3 testing results

Sno	Personal Factor	Family factor	University factor	Academic factor	Teaching factor	Defuzzification using different techniques									
						centroid		Lom		Mom		Som		bisector	
						values	ranks	Values	Rank	values	rank	values	Rank	values	rank
S1	75	0	30	45	65	50	4	50	4	50	4	50	5	50	4
S2	90	15	45	25	75	58.6	5	56	5	50	5	44	4	57	5
S3	79	7	65	48	8	15	1	23	1	15	1	7	2	15	1
S4	60	18	90	30	20	15	2	24	2	15	2	6	1	15	2
S5	18	39	30	83	91	35	3	44	3	35	3	26	3	35	3
S6	0	45	50	90	60	50	4	50	4	50	4	50	5	50	4

4. EXPERIMENTAL RESULTS

The experimental results shows that this system generates the 204 rules for analyzing the performance of the student either inconsistent increase, inconsistent decrease, consistent, consistent increase, or consistent decrease when compared with previous work[2] which generates the 243 rules for the fuzzy expert system using only fuzzy logic approach.

This shows that the current system with new approach using association rule mining and fuzzy logic decreases the complexity of fuzzy expert system and also draws the accurate conclusions by the fuzzy inference engine.

5. CONCLUSION AND FUTURE SCOPE

The system is developed that helps to analyze the students' academic performance against the critical factors considered while collecting the information. It works under the two approaches one is association rule mining technique and the other is fuzzy logic. First, the association rule mining approach is applied to generate the strong association rules from the frequent itemsets that occurs frequently in the dataset. Then fuzzy logic is used to define the fuzzy rules for the inference engine based on the results of the mining process .Hence the study concludes that by using association rule mining technique, the complexity of the fuzzy rule based expert system decreases. The accurate results are generated by the system and thus reduce the computation time for the analysis of the students' performance. The testing shows that it works better for all defuzzification methods except SOM method.

The scope of the study can be extended by including more generalised association rules that are extracted based on the higher confidence value specified by the user. Besides using the classical apriori algorithm, we can extend the system to use the improved mining algorithm. The system can also be improved by considering the more interesting and non-academic factors that may affect the performance of the student.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. "Mining association rules between sets of items in large databases." In ACM SIGMOD Record, vol. 22, no. 2, pp. 207-216. ACM, 1993.

[2] Kaur, Parwinder, Prateek Agrawal, Sanjay Kumar Singh, and Leena Jain. "Fuzzy rule based students' performance analysis expert system." In Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on, pp. 100-105. IEEE, 2014.

[3] Oladipupo, Olufunke O., Olanrewaju J. Oyelade, and Dada O. Aborisade. "Application of Fuzzy Association Rule Mining for Analysing Students Academic Performance." International Journal of Computer Science Issues (IJCSI), Vol. 9, Issue 6 No. 3, 2012.

[4] Bambrah, Charanjit, Minakshi Bhandari, Nirali Maniar, and Vandana Munde. "Mining Association Rules in Student Assessment Data." In International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol. 3, Issue 3, 2014.

[5] Wang, Hua, Ping Liu, and Hongyang Li. "Application of improved association rule algorithm in the courses management." In Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on, pp. 804-807. IEEE, 2014.

[6] K.Rajeswari, Mahadev Shindalkar, Nikhil Thorawade, Pranay Bhandari. "DSS Using Apriori Algorithm, Genetic Algorithm and Fuzzy Logic" International Journal of Engineering Research and Applications (IJERA), Vol. 3, Issue 4, 2013, pp.132-136

[7] Nosratabadi, Hamid Eslami, Ahmad Nadali, and Sanaz Pourdarab. "Credit Assessment of Bank Customers by a Fuzzy Expert System Based on Rules Extracted from Association Rules." International Journal of Machine Learning and Computing, Vol. 2, No. 5, (2012)

[8] Nosratabadi, Hamid Eslami, Ahmad Nadali, and Sanaz Pourdarab." "Evaluating Discovered Rules from Association Rules Mining Based on Interestingness Measures Using Fuzzy Expert System" International Journal of Machine Learning and Computing (IJMLC), Vol.2,No.5,IEEE2011