Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 1, January 2014, pg.428 – 433

RESEARCH ARTICLE

Mining Association Rules to Improve Academic Performance

Rakesh Kumar Arora[1], Dr. Dharmendra Badal[2]

[1] Dept. of Computer Science, Krishna Engineering College, Ghaziabad, UP, India

[2] Dept. of Mathematical Sciences & Computer Applications, Bundelkhand University, Jhansi, UP, India

[1] rka1211@gmail.com

***Abstract*—The main objective of higher education institutions is to provide quality education to its students. Institutions hope to improve the quality of education by identifying the set of students that needs special focus to clear the exams such that appropriate steps can be taken to improve the overall success ratio of the students. This will result in excellent placements and hereby increasing the quality intake of students in subsequent years. A system to analyze the performance of students using association analysis algorithm is being described in this paper. This paper will assist the academic planners in identification of students that need more attention such that the extra efforts can be employed on these set of students to improve the results.**

***Keywords*—Data Mining; Business Intelligence; WEKA; Data Visualization; Association Analysis**

I. INTRODUCTION

A large number of self financing private institutes and universities have opened over the last decade with the objective of providing quality education to students in various fields of engineering and other professions. Numbers of foreign universities have also got approval from the ministry of Human Resource and Development to compete with the Indian Universities. Global competition and rapid opening of these educational institutions may result in admission crisis in number of institutes in near future until and unless quality of education and good placements is being provided by these institutes. To have good results early identification of the set of students that requires extra attention is required such that the institute can arrange for the extra classes or can run some introductory courses at the beginning of the programme. [1]

This paper uses Educational Data Mining Technique (EDM) in identification of students that need special attention in order to improve the overall results of the institute. Data mining, the extraction of hidden predictive information from large databases is a powerful technology

with great potential to help head of departments in the institutes in distribution of subjects. It discovers information within the data that queries and reports can't effectively reveal. After gathering data submitted by students at the time of admission regarding the percentage of marks obtained in graduation and from the faculties regarding the percentage of marks obtained in post-graduation, data mining technique need to be applied to identify set of students in subsequent years that need more focus in order to improve their result in post graduation[6][7].

With the help of data mining techniques, such as clustering, decision tree or association analysis it is possible to discover the key characteristics from the details of students and possibly use those characteristics for future prediction. This paper presents association analysis algorithm as a simple and efficient tool to analyze the students details collected over the years such that effective measures can be taken in subsequent semesters.

II. METHODOLOGY

Association analysis is useful for discovering interesting relationships hidden in large data set. The uncovered relationships are represented in the form of association rules.[2] A common strategy adopted by many association rule mining algorithms is to divide the problem in two subtasks[2]

1. Frequent Itemset Generation - The main objective is to find all the itemsets that satisfy the minsupport threshold. These itemsets are called frequent itemsets.
2. Rule Generation – Its objective is to extract all the high confidence rules from frequent itemsets found in previous step.

III. APRIORI ALGORITHM

Apriori is a seminal algorithm proposed by R. Agarwal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemset properties. The following lines state the steps in generating frequent itemset in Apriori algorithm.[3]

Let C_k be a candidate itemset of size k and L_k as a frequent itemset of size k. The main steps of iteration are:
1. Find frequent set L_{k-1}
2. Join step: C_k is generated by joining L_{k-1} with itself(cartesian product $L_{k-1} \times Lk\text{-}1$)
3. Prune step (apriori property): Any $(k-1)$ size itemsetthat is not frequent cannot be a subset of a frequent k size itemset, hence should be removed
4. Frequent set Lk has been achieved [3]

Fig 1: Apriori Algorithm

The analysis using association analysis is being done with the help of WEKA tool. WEKA, formally called Waikato Environment for Knowledge Learning supports many different standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection. WEKA is an open source application that is freely available under the GNU general public license agreement. Originally written in C the WEKA application has been completely rewritten in Java and is compatible with almost every computing platform. It is user friendly with a graphical interface that allows for quick

set up and operation. WEKA operates on the predication that the user data is available as a flat file or relation, this means that each data object is described by a fixed number of attributes that usually are of a specific type, normal alpha-numeric or numeric values. The WEKA application allows novice users a tool to identify hidden information from database and file systems with simple to use options and visual interfaces. [4]

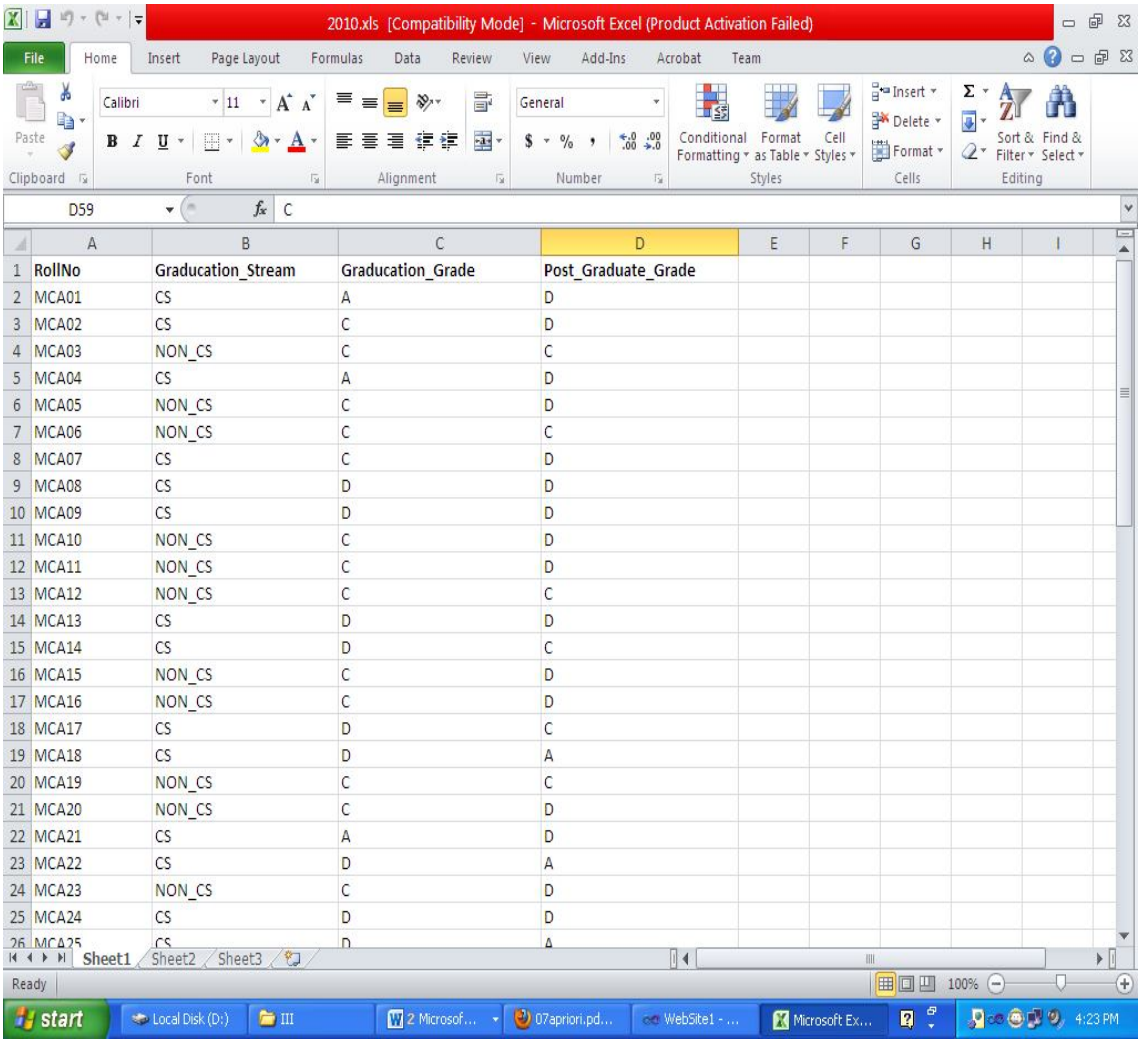
IV. RESULTS

The model was applied on students that have taken admission in MCA department in year 2010 in reputed Engineering College of Ghaziabad. The analysis is being performed on the basis of percentage of marks obtained by these students in graduation and post graduation (MCA). The minimum requirement for a candidate to take admission in MCA is that a candidate should be graduate and should have mathematics as subject either at XII level or at graduate level. It indicates that students that are not from computer science background can also take admission in MCA course. The numbers of students involved in analysis are 61 .The performance of students is compared by associating the following grades with various percentage intervals as represented in Table 1.

TABLE I
PERFORMANCE INDEX

Percentage	Grades
75 and above	A
70 – 74	B
65 – 69	C
60 - 64	D
55 – 59	E
50 – 54	F
Below 50	F

The details for students of MCA Batch (2010-2013) are represented in Fig. 2. The batch considered for analysis consists of 32 students from computer science background and 29 students from non computer science background.



	A	B	C	D
1	RollNo	Graduation_Stream	Graduation_Grade	Post_Graduate_Grade
2	MCA01	CS	A	D
3	MCA02	CS	C	D
4	MCA03	NON_CS	C	C
5	MCA04	CS	A	D
6	MCA05	NON_CS	C	D
7	MCA06	NON_CS	C	C
8	MCA07	CS	C	D
9	MCA08	CS	D	D
10	MCA09	CS	D	D
11	MCA10	NON_CS	C	D
12	MCA11	NON_CS	C	D
13	MCA12	NON_CS	C	C
14	MCA13	CS	D	D
15	MCA14	CS	D	C
16	MCA15	NON_CS	C	D
17	MCA16	NON_CS	C	D
18	MCA17	CS	D	C
19	MCA18	CS	D	A
20	MCA19	NON_CS	C	C
21	MCA20	NON_CS	C	D
22	MCA21	CS	A	D
23	MCA22	CS	D	A
24	MCA23	NON_CS	C	D
25	MCA24	CS	D	D
26	MCA25	CS	D	A

Fig 2: Sample Database of students

The data file normally used by WEKA is in ARFF (Attribute-Relation File Format) file format, which consist of special tags to indicate different things in the data file. Figure 2 shows the sample view of dataset and Figure 3 shows the ARFF format of desired dataset. To convert an Excel format into ARFF format an Excel to ARFF convertor is being used.

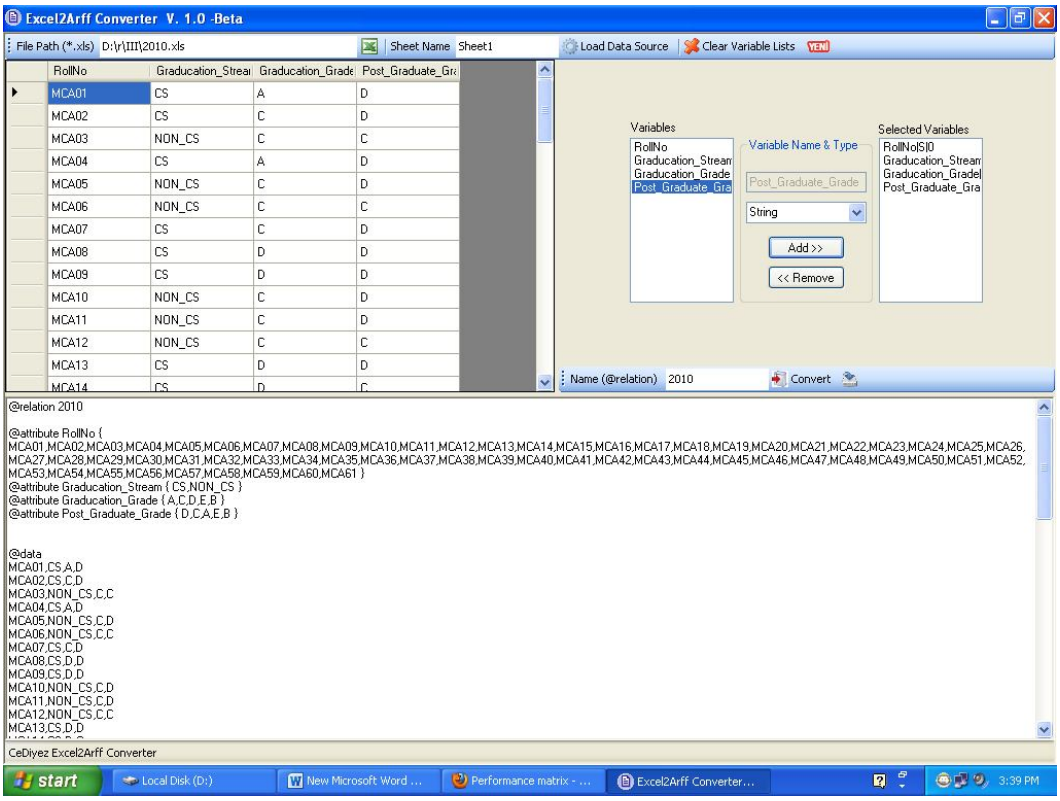


Fig 3: ARFF Format of the sample database of students

The results generated using WEKA is represented using Fig 4.

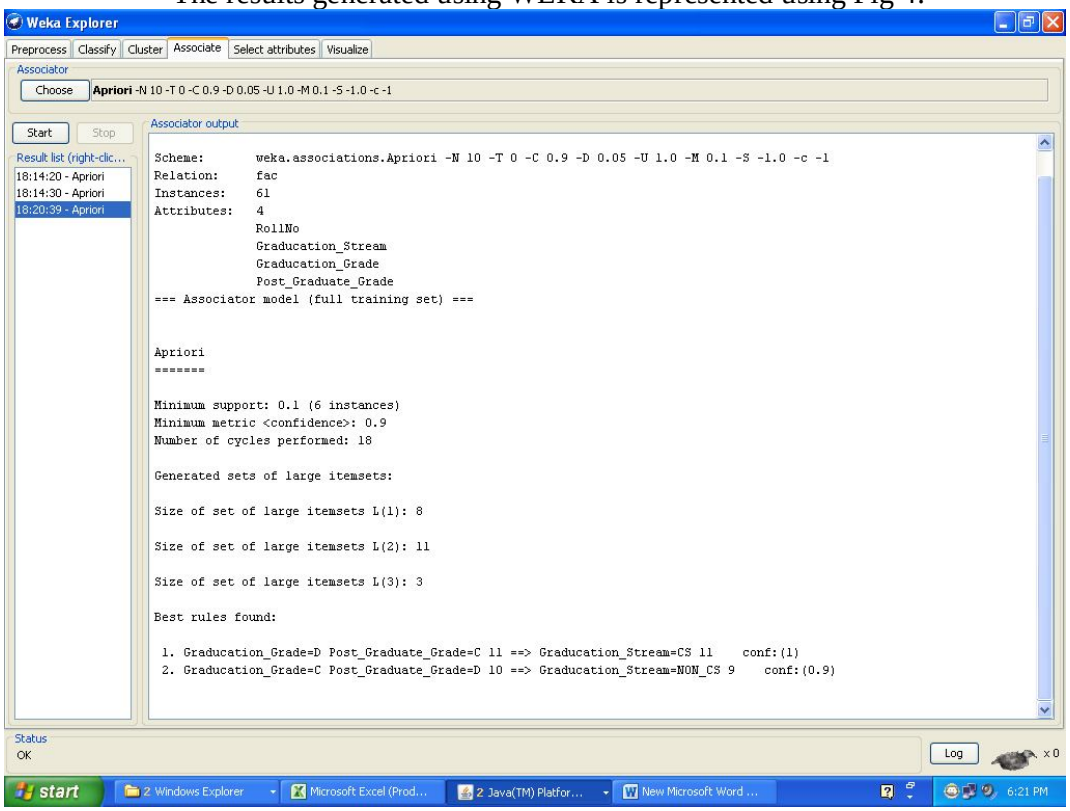


Fig 4: Output generated using WEKA

The above output clearly indicates that the performance of students having computer science background has significantly improved in MCA. The grades of 11 students from computer science background have improved from D to C while in the case of students with non computer science background the condition is reverse. The grades of 9 students have been degraded from C to D. It clearly indicates that the students from non computer science background are facing difficulty in competing with students having computer science background. In order to improve performance of students from non computer science background , some introductory courses need to be conducted at initial level or some extra lectures can be taken by experts.

V. CONCLUSION

The image of the institute is greatly affected by the results of the students. In this chapter, a simple methodology based on association analysis algorithm is being used to determine set of weak students in current semester by comparing the performance of students of previous semesters on the basis of marks obtained at graduate and post graduate level. This methodology will assist the academic planners to take the steps at the initial level in the subsequent years such that the overall result of the institute can improve. This will result in excellent placements and hereby increasing the quality intake of students. Hence this model will play important role for academic planners to determine the set of students that requires extra effort from institute side.

REFERENCES

[1] Arora K. Rakesh, Badal Dharmendra,“ Subject Distribution using Data Mining ”, International Journal of Research in Engineering and Technology, Volume 2, Issue 12, Dec 2013
[2] http://www-users.cs.umn.edu/~kumar/dmbook/ch6.pdf
[3] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth. “The KDD process for extracting useful knowledge from volumes of data”, CACM 39 (11), pp. 27-34, 1996.
[4] www.gtbit.org/downloads/dwdmsem6/dwdmsem6lman.pdf
[5] N.V.Anand Kumar Research Scholar, Department of Computer Science and engineering Anna university, Chennai G.V.Uma Assistant professor, Department of Computer Science and Engineering Anna university, Chennai “Improving Academic Performance of Student By Appling Data Mining Techniques
[6] Arora K Rakesh, Badal Dharmendra, “Location wise student admission analysis”, International Journal of Computer Science, Information Technology and Security, Dec 2012.
[7] Arora K. Rakesh, Badal Dharmendra,“ Evaluating Student’s Performance Using k-Means Clustering”, IJCST Vol. 4, Issue 2, April - June 2013

