



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Applying data mining technique to analyze the causes of backlogs in University

Rajiv Senapati, Dr. G.Sambasiva Rao
Department of Information Technology
Gandhi Institute of Engineering and Technology
Gunupur, Odisha, INDIA

Abstract: Data is the heart of business. It happens to be the key concern in knowledge discovery. Every session huge amount of data being generated from universities. So data becoming a key challenge now a days. This paper addresses the application of data mining in student's data of an Engineering College. Association rule mining is one of the most preferred techniques used to obtain meaningful associations and valuable information which is not previously known. In this work student result of G.I.E.T, an engineering college of eastern India has been considered. As a result rules have been produced that will indicate the association between courses and the cause of failure.

Key words: association rule mining, education, apriori algorithm

I. Introduction

We have huge amount of data around us but the utility is very less. However if we arrange the data properly and if we extract the association among those data then we can produce some useful information. That is the reason why data mining plays a major role in this information era. Data mining is used to extract the hidden patterns from data stores. Due to this reason data mining is used as a tool which supports decision makers.

In this paper, data mining is applied on student result database of an Engineering college in eastern India. Here apriori algorithm is used in a data set of 25 students by considering their results in 24 subjects of 4 different semesters (each semester having 6 courses). A minimum support rate 7 is considered.

II. Method

STEP 1: In first steps the algorithm scans the entire student's result table. Where, 25 student's result is considered for analysis. This is represented in Table 1.

REGD NO	1ST SEM						2ND SEM						3RD SEM						4TH SEM						TOTAL BACK
	BEE	TD	CP	MATH1	CHEM1	CE	BE	MECH	DS	PHY1	MATH2	BCE	OOP	NT	MATH3	PSD	OB	AEC	DMS	EEC	SP	DAA	DE	DEC	
779	F	F				F		F	F	F				F	F									F	10
777		F						F	F	F	F			F	F							F			9
780			F						F		F		F	F	F	F		F	F		F	F			12
778		F																							1
776									F																1
429			F				F	F			F						F		F					F	7
426	F	F	F	F	F			F			F	F			F		F		F					F	12
437											F				F				F						3
439		F													F										2
436	F	F			F		F			F	F				F				F					F	9
435	F										F						F		F						4
431							F				F														2
427															F				F						2
425	F		F	F			F								F										4
434	F	F	F	F			F	F	F	F	F	F			F	F	F	F	F	F	F	F		F	19
441							F				F			F											3
438					F																				1
443											F				F										2
440		F	F				F	F			F				F			F	F					F	9
430																									0
433			F												F		F								3
428											F														1
442		F			F			F			F			F	F			F							7
432			F	F				F		F	F						F		F						7
781		F	F	F	F		F	F	F	F	F			F			F					F			12

Table 1: Student Result

STEP 2: In this step the algorithm count the frequency of each subject, which is the support count for them. That means the algorithm simply scans all the subjects in order to count the number of occurrences of each subject, which will produce candidate 1-subjectset C1.

The Candidate 1-subjectset C1 is represented in Table 2.

Subject	Support Count	Subject	Support Count	Subject	Support Count	Subject	Support Count
BEE	6	BE	8	OOPS	1	DMS	10
TD	10	MECH	9	NT	6	EEC	1
CP	9	DS	6	MATH3	13	SP	2
MATH1	5	PHY1	6	PSD	2	DAA	5
CHAM1	6	MATH2	17	OB	7	DE	0
CE	1	BEC	2	AEC	4	DEC	7

Table 2: Candidate 1 Subject Set C1

Here the minimum support count set to 7. Now the algorithm determine the set of frequent 1 subject set L1 , which is a candidate 1-subject set with support count 7.

Now L1 can be represented as:

Subject	BE	TD	CP	MECH	MATH2	MATH3	OB	DMS	DEC
Support Count	8	10	9	9	17	13	7	10	7

Table 3: L1

STEP 3: To discover set of 2 subject set the algorithm perform $L1 \times L1$ to generate candidate 2 subject set C2, which is represented in Table 4.

Subject	Support Count	Subject	Support Count	Subject	Support Count	Subject	Support Count
TDXCP	4	TDXBE	4	TDXMECH	7	TDXMATH2	7
TDXMATH3	8	TDXOB	3	TDXDMS	4	TDXDEC	5
CPXBE	5	CPXMECH	6	CPXMATH2	8	CPXMATH3	5
CPXOB	6	CPXDMS	6	CPXDEC	6	BEXMECH	4
BEXMATH2	8	BEXMATH3	3	BEXOB	3	BEXDMS	4
BEXDEC	4	MECHXMATH2	8	MECHXMATH3	6	MECHXOB	5
MECHXDMS	5	MECHXDEC	5	MATH2XMATH3	9	MATH2XOB	6
MATH2XDMS	9	MATH2XDEC	6	MATH3XOB	3	MATH3XDMS	7
MATH3XDEC	6	OBXDMS	5	OBXDEC	3	DMSXDEC	6

Table 4: Candidate 2 Subject Set C2

Now L2 can be generated from C2 by considering minimum support count i.e.7.L2 is represented in Table 5.

Subject	Support Count	Subject	Support Count	Subject	Support Count
TDXMATH3	8	BEXMATH2	8	TDXMECH	7
MATH2XDMS	9	CPXMATH2	8	MECHXMATH2	8
MATH2XMATH3	9	MATH3XDMS	7	TDXMATH2	7

Table 5: L2

STEP 4: To discover set of 3 subject set the algorithm perform $L2 \times L2$ to generate candidate 3 subject set C3, which is represented in Table 6.

Subject	Support Count	Subject	Support Count	Subject	Support Count
TDXMECHXMATH2	6	TDXMATH2XCP	4	TDXMATH3XDMS	4
TDXMECHXMATH3	6	TDXMATH2XBE	4	CPXMATH2XBE	6
TDXMATH2XMATH3	6	TDX MATH2XDMS	4	CPXMATH2XMATH3	4

Table 6: Candidate 3 Subject Set C3

The algorithm will terminate here because no 3 subject set is having 7 or more than 7 support count. Now L1 and L2 subject sets are used to find out association rules.

III. Result

In this part association rules are formed from the frequent subject sets which are having more than the considered support count. The rules are given with data in table 7.

Rules	Confidence
MATH2 $\rightarrow$ MATH3	52.94%
MATH2 $\rightarrow$ DMS	59.94%
TD $\rightarrow$ MATH3	80%
BE $\rightarrow$ MATH2	100%
CP $\rightarrow$ MATH2	88%
MATH3 $\rightarrow$ DMS	53.84%
MECH $\rightarrow$ MATH2	88%
TD $\rightarrow$ MECH	70%
TD $\rightarrow$ MATH2	70%

IV. Conclusion

In this paper a small study has been conducted by considering 25 students result performance in 24 subjects for 4 numbers of semesters. If we consider more number of data sets then we may get some more associations and some more interesting rules.

Some of the findings of this study are:

1. The students who are unsuccessful in numerical courses become unsuccessful again in numerical courses which are present in the higher semester.
2. It has been seen that the students who fail in basic electronics subject are not performing well in MATH2.
3. It has been observed that the students who are not performing well in MATH2, out of them around 53% students are not performing well in MATH3.
4. One interesting knowledge has come into picture i.e. the students who are not performing well in MATH subjects are also not performing well in computer programming.
5. It has been observed that the students those who are not performing well in thermodynamics and mechanics are not performing well in mathematics.

V. Reference

- [1] Agrawal,R.,Imielinski, T.,& Swami,A.(1993). Mining Association rule between sets of items in large data bases.
- [2] Gursakal, N.(2001).Sosyal Bilimlerde Arastirma Yontemleri, VIPAS,Bursa,189
- [3] Alatas,B., & Akin,E.(2004). Madenciliginde Yeni Yaklasimlar, YA/em'2004.