January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

Prediction of Technical Education Student Performance using ARM

Dr. B. Raja Srinivasa Reddy, Department of CSE, Sri Vasavi Institute of Engineering and Technology, Nandamuru, brs_121@yahoo.com

Bala Brahmeswara Kadaru, Department of CSE, GEC, Gudlavalleru,balukadaru2@gmail.com.

Dr. Ch. Suresh Babu, Department of CSE, GEC, Gudlavalleru,sureshdani2004@gmail.com

Dr.N. Siva Chintaiah, Department of CSE, GEC, Gudlavalleru,shiva.gecmca@gmail.com

Article Info
Volume 82
Page Number: 8657 - 8664
Publication Issue:
January-February 2020

Abstract:
From the last few years, statistical techniques are utilized to analyze the Performance of Students in Academics by considering some parameters. In Present days because of state and central governments schemes, technical education is having admissions from more rural area students. In the same manner urban development also having influences in technical education admissions. This paper focuses implementing association rule mining to identify powerful rules from the existing data, which is used to discover the importance to the student performance related to the instructive environment where they will study. We have recognized the association among dissimilar attributes of educational background i.e., college locality, college type, diverse societal groups, dissimilar courses etc., and thereby dig up powerful association rules. For the administrators of technical education, from the existing data the unidentified rules are extracted and analyzed to take better decisions for growth of the institutes. These rules are also useful for a right perceptive of fire instructive location aids in course structure and other required up gradations to get better students' educational performance. This paper focuses on association rule mining to identify powerful rules from the existing data of higher education institutes which will be used to know the success patterns of students of different colleges based on societal groups. Additionally we have analyzed processed the available data to find the pattern of support for these rules from time to time.

Article History
Article Received: *5 April 2019*
Revised: *18 Jun 2019*
Accepted: *24 October 2019*
Publication: *08 February 2020*

Keywords: *Student Academic Performance, Data Mining, Technical Education, Association Rule Mining, Powerful rules*

I. Introduction

From the last 10 years, some statistical methods have been utilized to study the students' performance in technical education from based on different parameters. In the state of Andhra Pradesh, the government is offering fee reimbursement scheme to the students those who are having their family income is below one lakh rupees. Up to 2018-19 academic year, the students who got 10000 below Eamcet rank are eligible for full fee reimbursement, for example if a student got allotment in Computer Science and Engineering program through web counselling process in a College which is having tuition fee as 90,000 rupees, the state government will reimburse the full tuition fee to the concerned college, for remaining eligible students the colleges will be reimbursed 35,000 rupees and remaining will be collected from the students.

For Scheduled Cast and Scheduled Tribes central government of India will reimburse full tuition fee to the colleges irrespective of their Eamcet rank. From 2019-20 academic year onwards, the state government of Andhra Pradesh is ready to pay full tuition fee to all eligible students irrespective of their social group. This scheme is giving chance to economically poor students to study technical education in private engineering colleges also without paying tuition fee and getting better opportunities for further studies and placements in

TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

multinational companies. In urban areas, people are getting awareness of technical education to survive in this digital world. This paper focuses on association rule mining of data mining technique to identify powerful rules from the existing data of higher education institutes which will be used to study the success patterns of students of different colleges based on their societal groups.

In the constant changes of world, the requirement for the skilled work labour to meet needs is very high [2]. In present days, the main requirement is to develop the Public/Private Universities and other Colleges for adopting most resourceful, efficient and precise educational policies. Data Mining is provided with list of techniques i.e. association, classification and clustering used to identify hidden patterns from the institutional data, and also possible to analyze the existing data for getting added information all the stakeholders of the educational institute.

In this work, we selected Apriori Association mining algorithm to extract strong rules from the given data. We analyzed 8 years of B.Tech students' results of JNTUK Affiliated Colleges, from 2010-11 academic year to 2018-19 academic year. Pre-processing is applied for the data to go well with the requirements of mining technique.

II. Problem Statement

In this work, we would like to learn the students' academic results in diverse societal group categories in village and city areas in Autonomous and Non-Autonomous private engineering colleges, different programs by using association rule mining. This is to find out the unknown associations that exist among dissimilar under graduate categories in different colleges. We collected the data from JNTUK Affiliated Colleges, from 2010-11 academic year to 2018-19 academic year.

As we know that, a popular trust is that the quality of study in city/town colleges is faraway than from remote located colleges mostly in villages. In present situation, majority of rural engineering colleges getting funds through MODROS scheme for infrastructure establishment and DST projects for improving research labs, AICTE and UGC are recommending to subscribe national and international reputed journals i.e. print or online, and MHRD is providing NPTEL learning with free of cost through which the students are able to listen the lectures on a subject of IIT professors. Some of the colleges are placing this NPTEL courses as self-learning and mandatory and credit courses.

In present days, as per statistics rural backdrop colleges also achieving 80 percent placements. Results as well as placements are based on the facilities of the college, human resources of the college, interaction with industry, research oriented teaching, and effective course structure. Based on the analysis of existing data we can show the proofs for the above statements with the support of Data Mining technique in consideration with different parameters like college locality, college type, diverse societal groups, and dissimilar courses etc. By applying this Apriori technique we know the performance of engineering students of different groups and different environments. We can discover the trend of a particular group of students in a period of time.

III. Method

The dataset which we are using in this method consists of eight academic years' students' data of different private engineering colleges which are affiliated to JNTUK. The available data is related to examinations, so additional parameters of student which related to social status, type of the college, area of the college etc. The general activities are divided as below listed steps:

- Dataset preparation
- Applying pre-processing methods
- Applying selected Data Mining Technique
- Analysis of results

TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

Dataset Preparation

We gathered student exam results dataset from nearly 670 affiliated engineering colleges of JNTUK University from 2010-11 academic year to 2018-19 academic year.

The dataset consists of end semester exam results for B.Tech programs of these colleges. We collected nearly five lakh student records as the dataset, additionally student personal information which consists of their social status, college type and location of the college is also collected from the respective engineering colleges.

By combining the above mentioned data sets the main database record is having 48 parameters such as diverse societal groups, the sub categories, rural and town areas and Government sector and Private engineering colleges, list of courses of each student.

Data cleaning is required to apply for the collected data because of different sources are involved in dataset preparation like handling missing data, noisy data, identification or removal of outliers, and resolve the inconsistencies.

After completion of pre-processing technique, we get the cleaned data in the form tables which is passed as the input to the selected data mining technique

Data Pre-processing

The data which is collected is very large and consists of needless particulars [1]. So, the dataset is once again processed and the below listed attributes have been derived. The detailed list of attributes is shown in table 3.1.

TABLE 3.1
Structure of Basic Data

SNO	ATTR. NAME	TYPE	DESCRIPTION
1	D	CHAR	District Name
2	T	CHAR	Town Name
3	CC	Integer	College/Institute Code
4	C1	CHAR	'R'-Rural,'U'-Urban
5	C2	CHAR	'A'-Autonomous , 'NA'-Non-Autonomous
6	Grp	CHAR	CE,EEE,ME,ECE, CSE,IT
7	CN	CHAR	College/Institute Name
8	OC-M-D	Integer	OpenCategory-Male-Distinction class pass
9	OC-F-D	Integer	OpenCategory-Female-Distinction class pass
10	BC-M-D	Integer	BackwardCast-Male-Distinction class pass
11	BC-F-D	Integer	BackwardCast-Female-Distinction class pass
12	OC-M-I	Integer	OpenCategory-Male-First class pass
13	OC-F-I	Integer	OpenCategory-Female-First class pass
14	BC-M-I	Integer	BackwardCast-Male-First class pass
15	BC-F-I	Integer	BackwardCast-Female-First class pass
16	OC-M-S	Integer	OpenCategory-Male-Second class pass
17	OC-F-S	Integer	OpenCategory-Female-Second class pass
18	BC-M-S	Integer	BackwardCast-Male-Second class pass
19	BC-F-S	Integer	BackwardCast-Female-Second class pass
20	OC-M-P	Integer	OpenCategory-Male-Third class pass
21	OC-F-	Integer	OpenCategory-

January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

	P		Female-Third class pass
22	BC-M-P	Integer	BackwardCast-Male-Third class pass
23	BC-F-P	Integer	BackwardCast-Female-Third class pass

The above table consists of students' data which is not allowed to process directly. So once again we will process the table to separate the students' data based their societal status, gender, course and college. Resultant data is represented by considering the attributes listed in table 3.2.

TABLE 3.2
Data Structure for the segregated data

SNO	ATTR.NAME	TYPE	DESCRIPTION
1	C1	CHAR	'R'-Rural, 'U'-Urban
2	C2	CHAR	'A'-Autonomous 'N'-Non-Autonomous
3	Group	CHAR	CE,EEE,ME,ECE,CSE,IT
4	C3	CHAR	'OC' for Open Category, 'BC' for Backward category
5	PD	CHAR	'O' for Distinction class Pass , 'A' for First Class Pass, 'B' for Second Class Pass
6	cou	Integer	Colleges Count

Data Processing

To extract powerful association rules we are using ARM technique with Apriori algorithm.

Step 1: We removed unwanted attributes by applying AOI. We considered 2014 year data for this analysis purpose.

TABLE 3.3
Analysis of Year 2014 dataset

C1	C2	Grp	C3	PD	Cou
R	A	CE	OC	O	8
R	A	CE	OC	A	13
R	A	CE	OC	B	1
U	A	CE	OC	O	1
U	A	CE	OC	A	8
U	A	CE	OC	B	1
R	NA	CE	OC	O	31
R	NA	CE	OC	A	33
R	NA	CE	OC	B	11
U	NA	CE	OC	O	24
U	NA	CE	OC	A	31
U	NA	CE	OC	B	7
R	A	CE	BC	O	14
R	A	CE	BC	A	4
R	A	CE	BC	B	3
U	A	CE	BC	O	11
U	A	CE	BC	A	3
U	A	CE	BC	B	19
R	NA	CE	OC	O	36
R	NA	CE	OC	A	17
R	NA	CE	OC	B	23
U	NA	CE	OC	O	27
U	NA	CE	OC	A	14
U	NA	CE	OC	B	6
R	A	EEE	OC	O	6
R	A	EEE	OC	A	12
R	A	EEE	OC	B	2
U	A	EEE	OC	O	5
U	A	EEE	OC	A	10
U	A	EEE	OC	B	0
R	NA	EEE	OC	O	12
R	NA	EEE	OC	A	21
R	NA	EEE	OC	B	3
U	NA	EEE	OC	O	14
U	NA	EEE	OC	A	16
U	NA	EEE	OC	B	3
R	A	EEE	BC	O	9
R	A	EEE	BC	A	16
R	A	EEE	BC	B	2
U	A	EEE	BC	O	6
U	A	EEE	BC	A	10
U	A	EEE	BC	B	0
R	NA	EEE	BC	O	14
R	NA	EEE	BC	A	23
R	NA	EEE	BC	B	5

January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

U	NA	EEE	BC	O	11
U	NA	EEE	BC	A	18
U	NA	EEE	BC	B	5

Step 2: In this step, algorithm scans entire data from table 3.3. To form C1, we will use joining step on C3 & PD attributes with min-sup. C1 item set is displayed as shown below

C1

C3	PD	COU
Backward-Cast	O	92
Open Category	O	102
Backward-Cast	A	156
Open Category	A	145
Backward-Cast	B	51
Open Category	B	29

Step 3: We will find out L1 using C1 with min-sup, and we will remove the attributes which are not having min-sup.

Cl

C3	PD	Cou
BC	O	92
OC	O	102
BC	A	156
OC	A	145
BC	B	51
OC	B	29

>=min sup(25) --------------→

LI

C3	PD	Cou
BC	O	92
OC	O	102
BC	A	156
OC	A	145
BC	B	51
OC	B	29

Step 4: To generate C2 by using L1 join L1 and also perform pruning process with min-support. Finally, we will get C2.

C2

Grp	C3	PD	cou
CE	BC	O	40
EEE	BC	O	51
CE	OC	O	37
EEE	OC	O	64
CE	BC	A	67
EEE	BC	A	88
CE	OC	A	59
EEE	OC	A	85
CE	BC	B	12
EEE	BC	B	38
CE	OC	B	8
EEE	OC	B	20

Step 5: We will generate L2 item-set by using C2 with min-sup.

C2

Grp	C3	PD	cou
CE	BC	O	40
EEE	BC	O	51
CE	OC	O	37
EEE	OC	O	64
CE	BC	A	67
EEE	BC	A	88
CE	OC	A	59
EEE	OC	A	85
CE	BC	B	12
EEE	BC	B	38
CE	OC	B	8
EEE	OC	B	20

>=min sup(25) --------------→

L2

Grp	C3	PD	Cou
CE	BC	O	40
EEE	BC	O	51
CE	OC	O	37
EEE	OC	O	64
CE	BC	A	67
EEE	BC	A	88
CE	OC	A	59
EEE	OC	A	85
EEE	BC	B	38

Step 6: In this step we are generating C3 by using L2 join L2 and also perform pruning process with min-support. Finally, we will get C3.

C3

C2	Grp	C3	PD	Cou
A	CE	BC	O	15
NA	CE	BC	O	25
A	EEE	BC	O	9
NA	EEE	BC	O	42
A	CE	OC	O	11
NA	CE	OC	O	26
A	EEE	OC	O	9
NA	EEE	OC	O	55

January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

A	CE	BC	A	26
NA	CE	BC	A	41
A	EEE	BC	A	25
NA	EEE	BC	A	63
A	CE	OC	A	22
NA	CE	OC	A	37
A	EEE	OC	A	21
NA	EEE	OC	A	64
A	CE	OC	B	7
NA	CE	OC	B	31

Step 7: In this 7th step, we will generate L3 item-set by using C3 with min-sup.

C3

C2	Grp	C3	PD	Cou
A	CE	BC	O	15
NA	CE	BC	O	25
A	EEE	BC	O	9
NA	EEE	BC	O	42
A	CE	OC	O	11
NA	CE	OC	O	26
A	EEE	OC	O	9
NA	EEE	OC	O	55
A	CE	BC	A	26
NA	CE	BC	A	41
A	EEE	BC	A	25
NA	EEE	BC	A	63

>= min sup(25) ---------→

L3

C2	Grp	C3	PD	Cou
NA	CE	BC	O	25
NA	EEE	BC	O	42
NA	CE	OC	O	26
NA	EEE	OC	O	55
A	CE	BC	A	26
NA	CE	BC	A	41
A	EEE	BC	A	25
NA	EEE	BC	A	63
NA	CE	OC	A	37
NA	EEE	OC	A	64
NA	CE	OC	B	31

A	CE	OC	A	22
NA	CE	OC	A	37
A	EEE	OC	A	21
NA	EEE	OC	A	64
A	CE	OC	B	7
NA	CE	OC	B	31

Step 8: To generate C4 by using L3 join L3 and also perform pruning process with min-support. Finally, we will get C4.

C4

C1	C2	Grp	C3	PD	cou
R	NA	CE	BC	O	14
U	NA	CE	BC	O	11
R	NA	EEE	BC	O	19
U	NA	EEE	BC	O	23
R	NA	CE	OC	O	12
U	NA	CE	OC	O	14
R	NA	EEE	OC	O	31
U	NA	EEE	OC	O	24
R	NA	CE	BC	A	16
U	NA	CE	BC	A	10
R	NA	EEE	BC	A	23
U	NA	EEE	BC	A	18
R	NA	CE	OC	A	14
U	NA	CE	OC	A	11
R	NA	EEE	OC	A	36
U	NA	EEE	OC	A	27
R	NA	CE	BC	B	21
U	NA	CE	BC	B	16
R	NA	EEE	BC	B	33
U	NA	EEE	BC	B	31
R	NA	CE	OC	B	17
U	NA	CE	OC	B	14
R	NA	EEE	OC	B	25
U	NA	EEE	OC	B	35

TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

Step 9: We will generate L4 item-set by using C4 with min-sup.

C4

C1	C2	Grp	C3	PD	cou
R	NA	CE	BC	O	14
U	NA	CE	BC	O	11
R	NA	EEE	BC	O	19
U	NA	EEE	BC	O	23
R	NA	CE	OC	O	12
U	NA	CE	OC	O	14
R	NA	EEE	OC	O	31
U	NA	EEE	OC	O	24
R	NA	CE	BC	A	16
U	NA	CE	BC	A	10
R	NA	EEE	BC	A	23
U	NA	EEE	BC	A	18
R	NA	CE	OC	A	14
U	NA	CE	OC	A	11
R	NA	EEE	OC	A	36
U	NA	EEE	OC	A	27
R	NA	CE	BC	B	21
U	NA	CE	BC	B	16
R	NA	EEE	BC	B	33
U	NA	EEE	BC	B	31
R	NA	CE	OC	B	17
U	NA	CE	OC	B	14
R	NA	EEE	OC	B	25
U	NA	EEE	OC	B	35

↓ >=Min sup (25)

L4 freq item set

C1	C2	Grp	C3	PD	Cou
R	NA	CE	OC	O	31
R	NA	CE	BC	A	36
U	NA	CE	BC	A	27
R	NA	CE	OC	A	33
U	NA	CE	OC	A	31

Step 10: We will generate L4 item-set with min-sup. Finally, we will get strong association rules.

List of Strong Association Rules

C1	C2	Grp	C3	PD	Cou
R	NA	CE	OC	O	31
R	NA	CE	BC	A	36
U	NA	CE	BC	A	27
R	NA	CE	OC	A	33
U	NA	CE	OC	A	31

IV. RESULT AND ANALYSIS

We analyzed 8 years of B.Tech students' results of JNTUK Affiliated Colleges, from 2010-11 academic year to 2018-19 academic year. By using ARM technique to generate strong association rules[3]. The rules are listed as below in table 4.1.

TABLE 4.1
Strong rules for the year 2010-2018

S NO	YEAR	C1^C2^Group^c3)->Pass Div	Cou	Support	Confidence
1	2010	(U^NA^CE^BC)->C	28	0.3181	0.5833
2	2011	(U^NA^CE^BC)->C	27	0.3033	0.5510
3	2011	(U^NA^CE^OC)->C	25	0.2808	0.5813
4	2012	(R^NA^CE^OC)->A	31	0.3364	0.5714
5	2012	(R^NA^CE^BC)->C	36	0.2523	0.428
6	2012	(U^NA^CE^BC)->C	27	0.3084	0.5156
7	2012	(R^NA^CE^OC)->C	33	0.3084	0.515625
8	2012	(U^NA^CE^OC)->	31	0.2897	0.4843

We are representing various support & confidence of strong rules by using charts for the selected years.

Rules representation using Charts from 2010 to 2018

(U^NA^CE^BC)-A

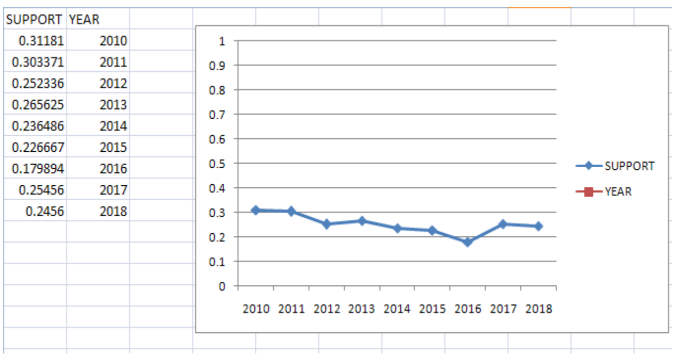

Figure 4.1: Pattern Rule 1: (U^NA^CE^BC)-A

January-February 2020
ISSN: 0193-4120 Page No. 8657 - 8664

(U^NA^CE^OC)-A

SUPPORT	YEAR
0.2808	2010
0.2897	2011
0.25	2012
0.23648	2013
0.21333	2014
0.15345	2015
0.179894	2016
0.256756	2017
0.258	2018

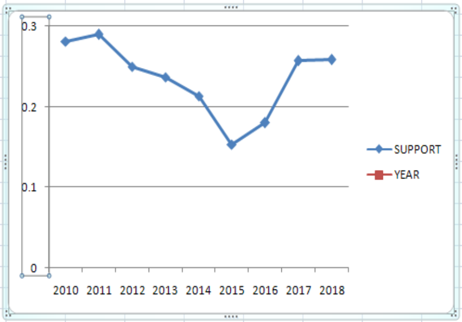

Figure 4.2: Pattern Rule2:(U^NA^CE^OC)-A

(R^NA^CE^OC)->O

SUPPORT	YEAR
0.2897	2010
0.34375	2011
0.3785	2012
0.3758	2013
0.258	2014
0.15345	2015
0.35	2016
0.256756	2017
0.27	2018

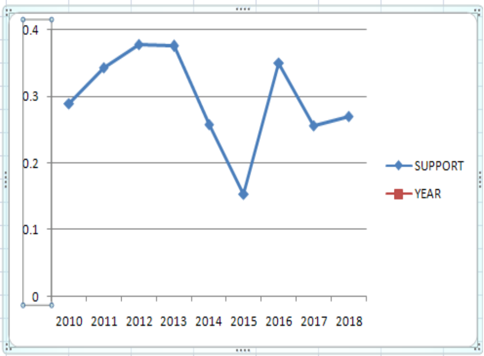

Figure 4.3: Pattern Rule3:(R^NA^CE^OC)->O

(R^NA^CE^BC)->B

SUPPORT	YEAR
0.336	2010
0.42	2011
0.47	2012
0.4451	2013
0.4462	2014
0.51	2015
0.523	2016
0.53	2017
0.56	2018

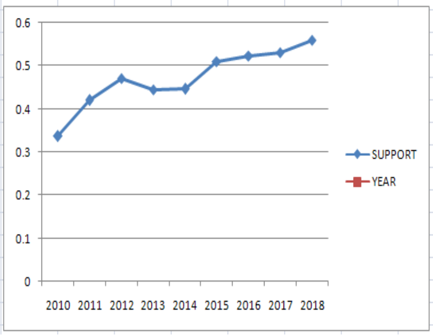

Figure 4.4: Pattern Rule4:(R^NA^CE^BC)->B

V. CONCULSION

Based on the observed results from this study, we can say that technical education institutes infrastructure, facilities and human resources are influencing the students' performance for their placements and getting chance to study higher education. This work was helped to find the relation among diverse attributes of educational institutes, namely, college locality, college type, societal groups, and list of courses etc. and they are used to extract strong rules by applying ARM technique[4]. Based on the results we conclude the performances of the students from village background institutes are achieving constant growth in rate of success but in slow manner. Interestingly, the performance of city environment colleges is gradually reducing.

VI. References

1. Yutang Liu, Qin Zhang, "Research on Association Rules Mining Algorithm Based on Large Data", "Revista de la Faculta de Ingenieria U.C.V.", vol.32,pp.229-236,2017.
2. Panadya Jalpa. P,MorenaRustom D. ",A Survey on Association Rule Mining Algorithms Used Different Application Areas", "International Journal of Advanced Research in Computer Scinece",ISSN:0976-5697,vol.8,pp. 1430-1436 ,2017.
3. Gurpreet Singh, Sonia Jassi, "A Review Paper: A Comparative Analysis on Association Rule Mining Algorithms", "International Journal of Recent Technology and Engineering(IJRTE), ISSN:2277-3878,vol.6, pp.1-3 2017.
4. Ashwini Rajendra Kulkarni, Dr. Shivaji, D. Mundhe, "Data Mining Technique: An Implementation of Association Rule Mining in HealthCare", "IARJSET", ISSN: 2394-1588, vol.4, pp. , 2017.