Evaluation of Student Academic Performance using Improved Prism

S. Umamaheswari, PhD Associate Professor, School of IT and Science, Dr.G.R.Damodaran College of Science, Tamilnadu

ABSTRACT

Data mining techniques (DMT) are extensively used in educational field to find new hidden patterns from student's data. Today, one of the main challenges that educational institutions face is the explosive growth of educational data of students and to use this data to improve the quality of decision-making. The study's curiosity is in predicting the paths of students, thus identifying the student achievement. This paper takes consideration with the various factors on student performance in education in order to analyse the student achievement is high over schooling or in graduation.

Keywords

DMT, KDD, EDM

1. INTRODUCTION

Data Mining (DM), or Knowledge Discovery in Databases (KDD), is an approach to discover useful information from large amount of data [3]. The advantage of data mining has landed in the plentiful fields mainly the educational research will commonly known as Educational Data Mining (EDM). There are increasing research interest in using educational data. This new emerging discipline, concerned with developing methods that extract knowledge from data comes from the educational content. DMT has a potential in performance monitoring of High school and graduation education offering historical perspectives of students' performances [1]. Evaluating student performance is a complex issue. The current data collected resides in the database of Dr.G.R.Dmodaran College of Science educational institution. The student data is personal and academic. This study is based on the Student's database of B.Sc., (Computer Science) and B.Sc., (Information Technology) for the past 3 years. Pre-processing techniques of Educational Data Mining were applied on selected dataset. After pre-processing the data, apply data mining techniques to discover classification. Data Mining is one of the approaches, which can provide an efficient assistance in revealing complex relationships behind the grades.

2. RELATED WORK

A methodology developed by derivation of performance prediction indicator, it mainly focusing on performance monitoring of students continuous assessment and examination scores in order to predict their final achievement status upon graduation. Based on various data mining techniques and the application of machine learning processes, rules are derived that enable the classification of students in their predicted classes [1]. Educational data mining improves graduate students' performance, and overcome the problem of low grades of graduate students. In their case study, the data include fifteen years period (1993-2007). They applied data mining techniques to discover knowledge. Particularly they K.S. Divyaa Research Scholar, Master of Philosophy, School of IT and Science, Dr.G.R.Damodaran College of Science, Tamilnadu

discovered association rules and used two classification methods which are Rule Induction and Naive Bayesian classifier to predict the Grade of the graduate student. Also clustered the students into groups using K-Means clustering algorithm[4]. The use of data mining techniques is to improve the efficiency of higher education institution. Data mining techniques such as clustering, decision tree and association are applied to higher education processes, it would help to improve the student performance and in selection of course[5]. The three elements needed to make prediction on Students' Academic Performances which are parameters, methods and tools. Also author proposes a framework for predicting the performance of first year bachelor students in computer science course. Naive Bayes Classifier is used to extract patterns using the Data Mining Weka tool. The framework can be used as a basis for the system implementation and prediction of Students' Academic Performance in Higher Learning Institutions[6]. Predictive data mining model for students' performance is to identify the difference between high learners and slow learners student. As a result, they had 300 student records, which were used for by Bayes classification prediction model construction.[7]

3. ACADEMIC PERFORMANCE EVALUATION

Academic achievement and academic attainment is the outcome of education and it is known as academic performance. It organizes the extent to which a student, teacher or institution has achieved their educational goals. The performance is mainly measured by examinations or continuous assessments. This paper deals with that student data as well as mark details. The data will be selected, pre-processed and classified.

The data is collected in two different phases. Initially, the data collected at SSLC, HSC with school name and UG level Mark/Percentage data. Data is collected from the Department of School of Information Technology and science of the Dr.G.R.Damodaran College of Science. The Data collection is done from the student database of B.Sc., (Computer Science) and B.Sc., (Information Technology) for the past 3year's i.e.2010-2012, 2011-2013, 2012-2015. Secondarily the main parameters are considered.

In Data processing the data set used in this work contains graduate students information collected from the college. The graduate student consists of 350 records and 15 attributes. Table 1 presents the attributes and their description that exists in the data set as taken from the source database. The selected attributes and description are the selected for the analysis process.

Table 1:	The	Student	Data	Set	Description
----------	-----	---------	------	-----	-------------

Attribute	Description				
Stud_Rollno	Student ID/ Roll Number				
Stud_Name	Name of the Student				
Gender	The gender of the student				
Dob	The date of birth of the student				
Enrol_year	The year of enrolment in the college				
Gradu_year	The year of graduation from the college				
Home_Loca	Location of the student home				
Tel_no	The telephone number of the student				
HSC_Perc	Percentage in the Higher Secondary Education				
HSC_School	School in which the Student have studied				
HSC_Loca	Location of the Higher Secondary Education				
UG_Perc	Percentage in the Under Graduation				
UG_Loca	Location of the UG College				
\$1,\$2,\$3\$6	Semester wise mark List				
UG_Major	Major of the Degree				

The data set contains some missing values in various attributes from 350 records; the records with missing values are ignored from the data set since it doesn't consider a large amount of data. The number of records is reduced.

After applying pre-processing and preparation methods, analyse the data graphically and figure out the grade of students using MAT Lab. The Data Set is shown in figure 1.

		-		-	-		-					-	
1	S.No	Stud_Rollno	Stud_Name	Gender	HSC_Perc	HSC_Loc	HOME_LOCA	S1	52	53	54	\$5	\$6
2	1	11B/T01	AARTHI V	F	88	U	U	87	85	83	81	77	76
з	2	11B/T04	ABIRAMI B	F	66	U	U	70	57	62	60	61	67
5	4	11B/T06	ARASU RAMANAN A	M	63	U	U	61	62	63	64	54	59
6	5	11B/T08	ARJUN E G	М	60	R	R	70	69	62	53	55	55
7	6	11B/T09	ARUN MURUGAPPAN	М	60	R	R	66	62	62	57	59	66
8	7	118/T10	ARUNKUMAR R	М	91	R	R	79	69	69	77	70	68
9	8	118/711	DARYL JOSEPH C	M	62	U	U	49	45	58	51	55	61
10	9	11B/T13	GOPINATH T K	М	65	U	U	52	42	43	50	38	60
11	10	11BIT14	GOWTHAM N R	М	74	U	U	61	58	60	61	57	65
12	11	11BIT16	HARSHINI A	F	70	R	R	71	67	63	73	68	67
13	12	118/T17	INDHUMATHI R	F	89	R	R	79	67	71	79	72	72
14	13	118/T18	JAIANAND S A	М	69	U	U	61	54	62	59	53	56
15	14	118/719	JENIFER ROSELIN S	F	73	U	U	83	82	90	84	86	86
16	15	11BIT20	KEERTHANA M	F	79	U	U	77	84	80	78	73	71
17	16	11B/T22	KINGSLY P	М	75	U	U	72	72	75	62	63	67
18	17	11BIT25	MANOJ KUMAR S	М	82	U	R	85	84	84	77	78	75
19	18	11BIT26	MEENA R	F	58	U	u	67	67	69	71	64	70
20	19	118/727	MITHILA D	F	83	U	u	83	74	82	79	73	69
21	20	11BIT28	NANDHA KUMAR S	M	59	U	U	67	69	55	62	55	66
22	21	11BIT29	NAVIN ARAVINTH N	М	82	U	U	59	60	61	56	59	63
23	22	11B/T31	PAPPU SHANKAR M	М	66	U	R	70	64	66	61	48	47

Figure 1 : Sample Data

4. PROPOSED WORK

Proposed system introduces improved prism algorithm to cluster more efficiently rather than existing algorithms. Initially the parameters are considered as location such as urban and rural, HSC, UG and percentage. The Work consider the school and college mark of particular student's dataset. For all class estimate the low, medium and high mark results by using improved prism clustering algorithm. The attributes are such as urban student, rural student and rural school, urban school also urban college. For the HSC dataset, evaluate urban student with urban school, urban student with rural school, and rural student with urban school and rural school with rural student. Then the conditions are checked and satisfied results are cluster into one group. Compute the low, medium and high marks on this data then improved prism algorithm cluster the three groups more efficiently.



Figure 2 : Phase 1

Similarly, for UG dataset, the attributes are considered as urban student, urban college and rural student. As depicted in phase 1,the conditions are checked and the corresponding results are grouped. As depicted in phase 2, the improved prism algorithm will group the low, medium and high marks for UG degree dataset.



Figure 3 : Phase II

The Improved Prim improves the attribute constraints efficiently. This scenario cluster the marks more effectively than the existing system. The Improved Prism increases the performance of proposed scenario prominently.

4.1 Pseudo Code

For each class C (Marks high, marks low and marks medium)

Initialize E to the instance set

While E contains instances in class C

Create a rule R with an empty left-hand side that predicts class C

```
Do
```

{

```
StuUrban == HSCUrban
```

StuRural == HSCUrban

```
StuUrban == HSCRural
```

```
StuRural == HSCRural
```

}

While

Compute high, medium, low percentage for all HSC

Do

{

StuUrban == UGUrban

StuRural == UGUrban

}

While

Compute high, medium, low percentage for all UG

Until R is perfect (or there are no more attributes to use) do

For each attribute A not mentioned in R, and each value v,

Consider adding the condition A = v to the left-hand side of R

Select A and v to maximize the accuracy p/t

(break ties by choosing the condition with the largest p)

Add A = v to R

Remove the instances covered by R from E

For each HSC

Find the male candidates high, low and medium percentage for urban

Find the male candidates high, low and medium percentage for rural

Compute the female candidates high, low and medium percentage for urban

Compute the female candidates high, low and medium percentage for rural

Obtain results

end

For each UG

Find the male candidates high, low and medium percentage for urban

Find the male candidates high, low and medium percentage for rural

Compute the female candidates high, low and medium percentage for urban

Compute the female candidates high, low and medium percentage for rural

Obtain results

end

Based on the classified rules the instances are covered and separated. Hence the important attributes are classified and it is used to identify the particular student's performance efficiently.



Figure 4 : Work Flow

The Student data will be having the Name, Roll no., Home Location, HSC Location, UG Location, UG Percentage and Gender. Using Improved Prism the B.Sc., Computer Science and Information Technology mark will be categorized and compared. Finally the report will be displayed. The above mentioned work is depicted in Figure 4.

5. EXPERIMENTAL RESULTS

The calculations are performed in MATLAB. The categories will be Urban-Urban-Urban, Urban-Rural-Urban, Rural-Urban-Urban and Rural-Rural-Urban. The low, medium and high percentage is performed. The Performance will be increased or decreased in their schooling or in graduation. whether the performance is increasing or decreasing for the urban or rural background student is experiemented.

RSG CS 2010-2013 Total number of students 40 Number of students with lowest percentage 23 Number of students with medium percentage 23 Number of students with highest percentage 30

Figure 5 : Performance Evaluation

The Figure 5 depicts the performance of the student in each class. The student data is of B.Sc., Computer Science and Information Technology (2010-2013, 2011-2014, 2012-2015) Batches. For all class estimate the low, medium and high mark results by using improved prism algorithm. The attributes are such as urban student, rural student and rural school, urban school also urban college. For the HSC dataset evaluate urban student with urban school, urban student with rural school, and rural student with urban school and rural school with rural student. Then the conditions are checked and satisfied results are cluster into one group. Now compute the low, medium and high marks on this data then improved prism algorithm cluster the three groups more efficiently. Similarly, for UG dataset, the attributes are considered as urban student, urban college and rural student. The conditions are checked and the corresponding results are grouped. The improved prism algorithm group the low, medium and high marks for UG degree dataset successfully.

				Increased	Decreased	No Of students	
Urban - Urban - Urban				59	1	60	
	Urban - Ru	ıral - Urban		16	1	17	
	Rural - Url	oan - Urban		18	2	20	
	Rural - Ru	ral - Urban		57	1	58	

Figure 6 : Performance Analysis Table

In Figure 6 Performance is analyzed for the class B.Sc., Computer Science 2010-2013 Batch. Likewise for each Class the report will be an output in Improved Prism Algorithm.

6. CONCLUSION

This study conducted analyses to examine the issue of differences in students' academic achievement from rural/urban area. Improved Prism Algorithm is used for predicting the paths of the students. Improved Prism improves the attribute constraints efficiently. Based on the results presented in the foregoing, the findings emerged. To analyze the academic achievement of urban and rural area students in order to identify superior performance is over schooling or in graduation. The analyzing is performed for class wise data. This shows that the students are performing well in their graduation. By the conclusion, Dr.G.R.Damodaran college of science is performing well in the student achievement.

Our future work includes investigating more parameters like extra curriculam activities and analyze the each semester mark in order to see the overall performance is over schooling or in graduation. Considering such parameters would result in better accuracy.

7. REFERENCES

[1] Ogor, Emmanuel N. "Student academic performance monitoring and evaluation using data mining techniques." *Electronics, Robotics and Automotive Mechanics Conference, 2007.CERMA 2007.IEEE, 2007.*

- [2] Baker, R.S.J.d.: Data Mining for Education. In: McGaw, B., Peterson, P., Baker, E. (eds.) To appear in *International Encyclopedia of Education*, 3rd edn. Elsevier, Oxford (2010)
- [3] Sachin,R.B., & Vijay, M. S, "A Survey and Future Vision of Data Mining in Educational Field", *Paper* presented at the AdvancedComputing & Communication Technologies(ACCT), Second International Conference on 7-8 Jan. 2012.
- [4] Tair, Mohammed M. Abu, and Alaa M. El-Halees. "Mining educational data to improve students' performance: a case study." *International Journal of Information* 2.2 (2012).
- [5] Goyal, Monika, and RajanVohra. "Applications of data mining in higher education." *International journal of computer science* 9.2 (2012): 113.
- [6] Abdul Aziz, Azwa, NurHafieza Ismail, and Fadhilah Ahmad."MININGSTUDENTS'ACADEMICPERFORM ANCE." Journal of Theoretical and Applied Information Technology 53.3 (2013): 485-485.
- Bhardwaj, Brijesh Kumar, and Saurabh Pal. "Data Mining: A prediction for performance improvement usingclassification." *arXivpreprintarXiv:1201.3418*(2012).
- [8] M.Kebritchi, and A.Hirumi, Examining the pedagogical foundations of modern educational computer games. *Computers and Education*, 5 (4): 1729-1743, 2008
- [9] A. McFarlane, N. Roche, and P. Triggs, Mobile Learning:Research Findings. Becta, July 2007. http://partners.becta.org.uk/uploaddir/downloads/page_d ocuments/research/mobile_learning_july07.pdf (accessed February4,2008),2007.
- [10] MoMath, Mobile Learning for Mathematics: Nokia projectinSouthAfrica.SymbianTweet,http://www.symbia ntweet.com/mobile-learning-formathematics-in-southafrica (access October 19, 2010), 2010.
- [11] D. Perry, Handheld computers (PDAs) in schools. Becta,http://partners.becta.org.uk/page_documents/resear ch/handhe lds.pdf (accessed September 27, 2010), 2003.
- [12] J-H. Valk, A. T. Rashid, and L. Elder, Using Mobile Phonesto Improve Educational Outcomes: An Analysis of Evidencefrom Asia. *The International Review of ResearchinOpenand Distance Learning*, 11 (1), 2010.