ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

Data Mining Techniques: Educational System

S.Haseena, Research Scholar, St Peter's University Avadi, Chennai. haseenaakm@gmail.com

Dr.R.Latha, Professor and Head, Department of CSA, St.Peter's University, Avadi, Chennai.

***Abstract*—Data Mining Techniques (DMT) is provided wisdom support for educational based multi dimensional data engineering research area that handles the development of methods to explore data appearing in educational fields. Computational approaches used by DMT are to examine student's data in order to study educational questions. Weka, open source data mining software is used to explore the student's academic progress. The objective of this research is to introduce educational data mining by describing step by step process using technique, RepTree, BFTree of Decision Tree, RandomForest, Bayes and NaiveBayes of BayesNetwork, RBFNetwork functions and JRip rule. As a result, it provides intrinsic knowledge of teaching and learning process for effective education system.**

***Keywords*— data mining, Decision Tree, Bayesin Networks.**

I. INTRODUCTION

Predicting results of students at an early stage of the degree program help education institution not only to concentrate more on the bright students but also to apply more efforts in developing programs for the weaker ones in order to improve their progress while attempting to avoid student failures. Weka (Waikato Environment for Knowledge Analysis) is selected as a data mining tool Weka, is a free software available under the GNU General Public License and it supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. Weka provides various algorithms grouped in different classifying methods. The aim is to compare these algorithms in predicting students' performance.

II. DATA MINING IN HIGHER EDUCATION SYSTEM

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes.

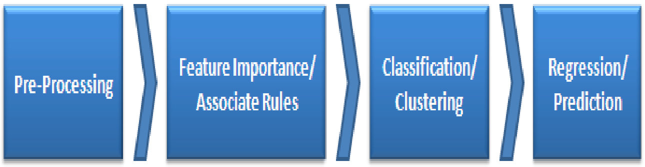
Higher Education faces many challenges, such as predicting the paths of students to graduation. Many institutions would like to know which students will need assistance in order to graduate and the kind of assistance required. Data Mining help an institution to take action before a student's results, or to predict the number of students to enhance understanding of learning process focus on identifying, extracting and evaluating variables related to the learning process.

Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well mannered. Mining in educational environment is called Educational Data mining, concern with developing new methods to discover knowledge from educational database in order to analyze student's trends and behaviors towards education. Educational data mining is used to identify and enhance educational process which can improve their decision making process. Lack of deep and enough knowledge in higher educational system may prevents system management to achieve quality objectives, data mining methodology can help bridging this knowledge gaps in higher education system.

i. Data mining Process

The data exploration and presentation process consisted of following steps

(a): Data mining process


Pre-Processing
Feature Importance/ Associate Rules
Classification/ Clustering
Regression/ Prediction

III. ATTRIBUTES SELECTION

In the entire data mining process, the data cleaning process is utilized in order to eliminate irrelevant items. The discovery of patterns will be only useful if the data represented in files offer a real representation of the enrolment process and the actions or decisions taken by the past student. After filtering process of the data, our research is o discover patterns that will be used to predict a loyal or not loyal student previously to his enrollment or not enrollment to a particular institute. By taking information from other students with similar information, in this sense, we can know the role of each attribute and the implicit relations among them.

Student data were collected from the institute was categorized in five groups Very good, Good, Satisfactory, Below Satisfactory, Fail.

IV. Results

The following tables are the classification accuracy of each classifier applied based on data.
Decision trees are a collection of nodes, branches, and leaves. Each node represents an attribute; this node then split into branches and leaves until the data are classified to meet a stopping condition.

Table i: J48 Tree

Actual Class	Original Data				Re-Sampled Data			
	Satisfactory	Below Satisfactory	Good	Fail	Satisfactory	Below Satisfactory	Good	Fail
Satisfactory	97	16	2	0	194	22	1	1
Below Satisfactory	65	16	0	0	26	156	3	0
Good	23	3	1	0	7	7	30	2
Fail	5	3	0	0	1	1	0	11
%Hit	84%	20%	4%	0%	89%	84%	65%	85 %

Table 4: classifier J48 decision tree, there was no prediction of the number of students failing but 84% students have been predicted as satisfactory performers. Even the class of "Very Good" was not considered in confusion matrices of every classifier, the reason being that those students who participated in the survey, their first year CGPA did not fall in this group. Therefore, confusion matrices for all classifiers only represented four classes.

Table ii: RandomForest

Actual Class	Original Data				Re-Sampled Data			
	Satisfactory	Below Satisfactory	Good	Fail	Satisfactory	Below Satisfactory	Good	Fail
Satisfactory	76	36	3	0	203	13	1	1
Below Satisfactory	46	33	2	0	8	176	1	0
Good	18	8	1	0	3	6	37	0
Fail	3	4	1	0	0	1	1	11
%Hit	66%	41%	4%	0%	93%	95 %	80%	85%

RandomForest's finds students results as 80 % and 85% respectively on resampled data. It showed that resampling of data have a significant impact on predicting capability of a classifier. The predictions for "Satisfactory" and "Below Satisfactory" classes have also improved considerably.

Table iii : Bayesian Network

Actual Class	Original Data				Re-Sampled Data			
	Satisfactory	Below Satisfactory	Good	Fail	Satisfactory	Below Satisfactory	Good	Fail
Satisfactory	80	32	2	1	152	61	2	3
Below Satisfactory	42	37	2	0	72	106	6	1
Good	18	8	1	0	28	9	9	0
Fail	6	2	0	0	4	3	2	4
%Hit	70%	46%	7%	0%	70%	57%	20%	31 %

Bayesian network is based on decision theory. It is a branch of probability and statistics which investigates how to minimize risk and loss when making decisions based on uncertain information. It is a graphical model that encodes relationships among variables that it models

V. CONCLUSION

On comparing all classifiers, RandomForest is most effective in student's performances with accuracy. The least effective is the BayesNet and NaïveBayes.

In the above performance of all the classifiers on the data of students, it is proved that Decision tree classifiers are better, to find students performance.

In this paper, the classification method is used on student database to predict the student's records on the basis of previous year database like Attendance, Assessments, Internal and External marks. As there are many methods that are used for data classification, the decision tree method is used here, to predict the performance at the end of the semester. Data mining can predict with a reasonable certainty, that a student might achieve and also helps students and teachers to improvise the area of the student. This study will also work to identify those students which needed special attention to reduce fail ratio and taking appropriate action for the next semester examination.

VI. FUTURE WORK

In this study we use data mining process in a student's database to predict students result. The information generated after the implementation of data mining technique may be helpful for a institution as well as for students. For future work we redefine our techniques in order to get more valuable and accurate outputs useful for institutions to improve the students learning process. Some different software's may utilize for accurate outputs.

REFERENCES

[1] Afzal, H.,Imran,A., Khan,M.A. and Kashif,H. (2010) A study of university students' motivation and its relationship with their academic performance, International Journal of Business and Management,Vol. 5, 4.

[2] Affendey, L.S., Paris,I.H.M., Mustapha,N., Sulaiman, Md.N., and Muda,Z.(2010) Ranking of Influencing Factors in Predicting Students' Academic Performance.Information Technology Journal Vol. 9, 4. [1] Shaeela Ayesha,(2010),data mining model for higher education system, J. of scientific research, vol -43, pp24-29.

[3] Sunita B. Aher,(2011), data mining in education system using WEKA, International J. of Computer Applications, pp20-25.

[4] Alaa el- Halees,(2009), Mining Students Data to Analyze e-learning behavior. A case study.

[5] [4] Connolly T., C. Begg et al,(1999) Database System: A practical approach to design, Implementation and management(3 rd edition), Harlow; Addison-Wesley,687.

[6] Erdogan and Timor (2005) A data mining application in a student database. Journal of Aeronautic and Space Technologies July 2005 Volume 2 Number 2 (53-57)

[7] Han,J. and Kamber, M., (2006) "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.

[8] ZhaoHui. Maclennan.J, (2005). Data Mining with SQL Server 2005 Wihely Publishing, Inc.

[9] Brijesh Kumar Baradwaj , Mining Educational Data to Analyze Students Performance ,international journal of advanced computer science and application, Vol. 2, No. 6, 2011

[10] Business Research Institute Conference, las Vegas Vandamme, J.P. and Superby,J.F(2007) "Predicting Academic Performance by Data Mining Methods."Education Economics, Vol.15 No(4), pp.405–419.

[11] Cardoso AR, Verner D (2007). School Dropout and push-out factors in Brazil: The role of early parenthood, child labor and poverty.IZA discussion on Paper. No 2515 Bon: Institute for study of labor (IZA)

[12] Davies, B. "Education for sexism: a theoretical analysis of the sex/gender bias in education." Educational Philosophy and Theory 21, no. 1 (1989): 1-19.

[13] Kokkodis,M. and Akabay,M. CS 235 Project Report: UCSD Data Minning ContestMannan, MD.A.(2007) "Student Attrition And Academic And Social Integration: Application of Tinto's Model at the University of Papua New Guinea.", Higher Education, Vol 53, pp. 147-165.

[14] Umesh KUmar Pandey et al, Data Mining : A prediction of performer or underperformer using classification, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2) , 2011, 686-690

[15] Jonassen, D. H., Computers in the Classroom, Englewood Cliffs, NJ:Merrill, Keefe, J. W. (1987), in "Learning Style".\

[16] Peters, J., Jarvis, P. et al., Adult Education, San Francisco, CA, Ed Rogers, A., Teaching Adults, Buckingham: Open University Press

[17] Jemni, M., & Nasraoui, O. (2009). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. Educational Technology & Society, 12(4), 30–42.

[18] Liang, G., Weining, K. & Junzhou, L. (2006). Courseware recommendation in e-learning system. Advances in Web Based Learning – ICWL2006, Springer Berlin/Heidelberg, 10-24.

[19] Khairil Imran Ghauth and, Nor Aniza Abdullah, (2009) An Empirical Evaluation Of Learner Performance In E-Learning Recommender Systems And An Adaptive Hypermedia System, pp 141-152.