

Overview of algorithms in Educational Data Mining for Higher Education: An Application Perspective

Dipesh Walte¹, Hari Reddy², Vivek Ugale³, Amol Unwane⁴
^{1,2,3,4} Pimpri Chinchwad College of Engineering
 Pune, India

Abstract— Even with the ever increasing Data Generated in the Higher Educational Sector, most of the colleges still use classic statistical and regression techniques for Decision Making purpose. Educational Data Mining (EDM) proposes various approaches to procure meaningful information and eventually knowledge which would greatly benefit the Higher Educational Institutes to Improve the function of their various activities. We intend to summarise the various applications of these techniques.

Index Terms—Educational Data Mining, Applications of EDM, Classification, Association, Clustering.

I. INTRODUCTION

After initial works in Educational Data Mining in 1995[2] a lot of work has been done in the data mining techniques applied to the educational sector. Although Simple Regression and Statistical and Visualisation Techniques can give a good understanding of data and are used heavily in almost every Educational Organization Researchers have now applied various techniques including Classification, Association, Clustering are used to give a variety of results that find a variety of applications. It is essential for anybody using EDM to know the possible applications and outcomes of the various data mining techniques used. This paper thus summarises the works done in this domain, the various algorithms used, and the applications of these in improving the overall functioning of the higher education sector.

A. Classification

It is a supervised learning technique in which a tuple is assigned a previously known class using the model.

- *Decision Trees*

A decision tree is widely used mining technique much because of the fact that the output is very easy and intuitive to interpret. It is flow-chart-like tree structure where all internal nodes have two or more child nodes. The leaf nodes denote the decision made or the class label, the arcs the condition that we have applied and the internal nodes denote the attributes. Decision trees are amongst the favourite among researchers applying the data mining techniques on the educational domain, which is evident from the table 1.0

- *Bayesian*

Bayes classifier is based on Bayes rule of conditional probability. Bayes rule is a technique to estimate the

likelihood of a property given the set of data as evidence or input.

Bayes rule or Bayes theorem is-

$$P(h_i | x_i) = P(x_i | h_i) P(h_i) / P(x_i | h_1) + P(x_i | h_2) P(h_2)$$

B. Association Rule Mining

The central task of association rule mining is to find sets of binary variables that co-occur together frequently in a database. Association Rule Mining is used in applications where we want to find the relation between the attributes.

C. Clustering

Clustering method is an unsupervised learning technique which groups similar data.. This technique is especially applicable to find out groups that behave similarly and whose class is not predetermined, thus avoiding prejudgement of a dataset.

II. RELATED RESEARCH

Carlos, Cristobal and Sebastian[1] analysed the data of 670 middle-school students by applying whitebox classification methods and predicted the drop out and failure rate with accuracy of over 90%.

In the very first papers of application of Data Mining techniques to Higher Education Luan, Jingdiscuss various potential applications including marketing, alumni fund raising, to survival analysis, persistence and many others. [3]

Pandey and Pal[4] try to find out the adept teacher dealing with Students by taking a psychometric test that converts qualitative variables into quantitative and further applying association rules

Kabakchieva[5] applied selected data mining algorithms for classification on the Bulgarian university sample data revealed that the prediction rates are not remarkable and the classifiers perform differently for the five classes.

Ajay Pal and Sourabh Pal from Sai Nath University and VBS Purvanchal Universities respectively have proposed a model based on classification approach so as to predict the student placement. The study was done with an aim to extract the relations between the academic achievements and the student's recruitment.They concluded that the Naïve Bayes

algorithm works as accurately as 86.15% with a lowest average error at 0.28 compared to others. [6]

Umesh Kumar Pandey, Brijesh Kumar Bhardwaj, and Saurabh pal in their paper, Data Mining as a Torch Bearer in Education Sector have portrayed the roadmap of research done in EDM owing to various education sector segments. The research was mostly done in order to get to inferences which may be used to improve the quality of education. [7]

Mohammed M. Abu Tair and Alaa M. El-Halees aimed at using EDM to improve the student's performance. In the case that they undertook, a database of student information of about 15 years was acquired and data mining techniques to discover association, clustering and outlier detection was applied on it. The paper showed how a student performance may be improved with the use of data mining. Rule Induction and naïve Bayesian classifier methods for classification were adopted. K means and outlier detection methods including distance and density based approaches were also used for forming clusters. [8]

Ying Zhang and Tony Clark from the Thames Valley University studied data mining and have described the MCMS (Mining Course Management Systems), which proposes to build a knowledge management system based on data mining for improving the student retention.[9]

Sourabh Pal, in his paper, Mining Educational Data to Reduce Dropout rates in engineering students, explains how EDM is vital in estimating the students likely to drop out due to unfortunate conditions which may be foreseen with an educated prediction. Machine learning algorithms such as the ID3 learning algorithm can learn effective predictive models from the student dropout data accumulated from the previous years. The empirical results show that we can produce short but accurate prediction list for the student dropout purpose by applying the predictive models to the records of incoming new students. [10]

Senol Zafer Erdogan from Maltepe University, Istanbul and Mehpare Timor from Istanbul University have proposed a data mining application in a student database. Cluster analysis and K means algorithm has been used to study the relationship between students' university results and success patterns they have shown. [11]

Oladipupo O.O and Oyelade O.J from Covenant University, Nigeria have proposed Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach. They have used association rule mining techniques to identify student failure patterns. Such patterns may be further analyzed and constructive recommendations may be made in support of the decision making process. This study has bridge the gap in educational data analysis and shows the potential of the association rule mining algorithm for enhancing the effectiveness of academic planners and level advisers in higher institutions of leaning. [12]

Samrat Singh and Vikesh Kumar from the Neelkant Institute of Technology, Meerut, India, have proposed the Classification of Student's data Using Data Mining Techniques for Training & Placement Department in Technical Education. The aim of the study was to identify the final grade of the student for the placement purpose by the use of decision tree method for classification, owing to the past results of the records

undertaken for the analysis. The information generated after the analysis of data mining techniques on student's data base is helpful for executives for training & placement department of engineering colleges. This work classifies the categories of student's performance in their academic qualifications. [13]

III. SUMMARISATION OF RESULTS

Table Head	Application of Various Techniques and Algorithms		
	Technique	Algorithm	Application
1	Classification	Decision Tree	Predicting Failure [1]
			Predicting Dropout[1][3][13]
		Bayesian	Alumini Fund Raising[3]
			Classify the students as Bad Average Very Good and Excellent[5]
2	Association	Aproiri	Prediction of Placements[6]
			Student Retention[9]
3	Clustering	K-Means	Classify the students as Bad Average Very Good and Excellent[5]
			Academics & Recruitment [6]
4	Outlier Detection		Finding Adept Teacher Dealing with Students[4]
			Failure pattern extraction[11]
			Persistent and Non-Persistent student Comprehension[3]
			Grouping Similar Students [11]
			Rare Events Analysis and understanding[8]
			Understanding irregular events[8]

IV. CONCLUSION

We would like to summarise the applications in a tabular format, presenting the various researches done in an application perspective. The research varies as the domain knowledge of the various institutes and the researchers carrying out the research vary. Also the factors and the algorithms taken into consideration are heterogeneous. Decision Trees are widely used for failure and dropout prediction whereas other techniques have found a variety of other uses.

REFERENCES

- 1) C. Marquez-Vera, Autonomous University of Zacatecas, Mexico
- 2) C. Romero and S. Ventura, University of Cordoba, Spain - Predicting School Failure Using Data Mining.
- 3) C. Romero and S. Ventura, University of Cordoba, Spain - Educational Data Mining A survey from 95-05
- 4) Luan, Jing - Data Mining and Knowledge management in Higher Education - Potential Applications
- 5) Umesh Pandey, PSRIET, Sourabh Pal, VBS Purvanchal University, India - Mining Data to Find Adept Teachers in Dealing with Students

- 5) Dorina Kabakchieva, Sofia University, Bulgaria - Predicting Student Performance by Using Data Mining Methods for Classification
- 6) Ajay Pal, Sai Nath University, Ranchi, Jharkhand, Saurabh Pal VBS Purvanchal University, Jaunpur, India - Classification Model of Prediction for Placement of Students
- 7) Technical Journal of LBSIMDS - Data Mining as a Torch Bearer in Education Sector - Umesh Kumar Pandey, Brijesh Kumar Bhardwaj, Saurabh pal
- 8) International Journal of Information and Communication Technology Research - Mining Educational Data to Improve Students' Performance: A Case Study - Mohammed M. Abu Tair, Alaa M. El-Halees
- 9) Use Data Mining To Improve Student Retention In Higher Education : A case Study – Ying Zhang, Samia Oussena, Tony Clark, Hyeonsook Kim
- 10) Sourabh Pal – Mining Educational Data to reduce dropout rates of engineering students – I.J. Information Engineering and Electronic Business.
- 11) Senol Zafer Erdogan and Mehpare Timor – A data mining application in a student database – Journal of Aeronautics and Space Technologies
- 12) Oladipupo O.O and Oyelade O.J. - Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach
- 13) Samrat Singh and Dr. Vikesh Kumar - Classification of Student's data Using Data Mining Techniques for Training & Placement Department in Technical Education – International Journal of Computer Science and Network

IJERT