Various Data Mining Techniques used to Study Student's Academic Performance

Ms. Priti S. Patel[1], Dr. S.G.Desai[2]

[1]Assistant Professor, Department of Computer Science, SRICEAS, Surat & Research Scholar, Uka Tarsadia Univ., Bardoli,
priti_patel22@yahoo.co.in
[2]Professor, SAL Institute of Technology and Engineering Research, Ahmadabad,
subhash1948@yahoo.com

Abstract:

To extract decision from huge amount of data is tedious work. And for this now days various data mining tools are available in market which are focused of data mining techniques to find pattern recognition and able to support decision making process using these tools. In this paper we are discussing some research papers on the students progression based on their performance.

Keywords: data mining tools, data mining techniques, decision making

1. Introduction:

Data mining is a process which finds useful patterns from large amount of data. The process of extracting previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions[2].Data are generally stored in various formats like text, images, audio, video, animated scripts etc in the repository. Han and Kamber define data mining as the process of discovering hidden images, patterns and knowledge within large amount of data and making predictions for outcomes or behaviors [1]. Data Mining is a multidisciplinary field, encompassing areas like information technology, machine learning, statistics, pattern recognition, data retrieval, neural networks, information based systems, artificial intelligence and data visualization. [3]

2. Educational Data Mining: (EDM)

EDM is emerging discipline, with a suite of computational and psychological methods and research approaches to understanding how students learn and setting which they learn in it. Educational data mining (EDM) is an emerging discipline which focuses applying data mining tools and techniques to educationally related data. [15]

An educational data mining is a broader term that focuses on nearly any type of data in educational institutional, while academic analytics is specific to data related institutional effectiveness and student retention issues. The scope of educational data mining includes areas that directly impact students. Other areas within EDM include analysis of educational processes including admissions, alumni relations and course selections. [15]

In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: Classification, categorization, estimation, and visualization. Classification identifies associations and clusters, and separates subjects under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay." Estimation includes predictive functions or likelihood and deals with continuous outcome variables, such as GPA and salary level. Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. [16] In recent years, the analysis and evaluation of students

performance and retaining the standard of education is a very important problem in all the educational institutions [3].

3. Literature Survey: Research papers

Ramaswamy N has written research paper which focused on predicts student's characteristics or academic performances in various educational institutions. This paper focus on students performance as a slow learner or fast lerner. For that they applied various data mining techniques and compare the accuracy based on students attributes[6]. For assessing the goodness of a predictor, an extensive study on the student data set was conducted by applying five individual classifiers J48 (J48), Bayesian Net (BN), Neural Net (NN), Decision Tree (DT), and Naïve Bayes (NB).

The other research paper which is written by Mrs. M.S. Mythili, Dr. A.R.Mohamed Shanavas to use data mining methodologies to study and analyze the school students' performance based on classification techniques which is useful to gauge students' performance and deals with the accuracy, confusion matrices and the execution time taken by the various classification data mining algorithms. The decision tree classifier C4.5 (J48), Random Forest, Neural Network (Multilayer Perception) and Lazy based classifier (IB1) Rule based classifier (Decision Table) were enforced in weka. [3]

Socio Economic Status has always affect and influenced the Academic Achievement of students. Academic Achievement varies for children from different social backgrounds. To fucus on the social economy and category wise students achievements. The High SES students have better exposure and environment. [19].In comparison to Low SES category Average SES category students show higher academic achievement. The other paper used to apply selected data mining algorithms for classification on the university sample data reveal that the prediction rates are not remarkable (vary between 52-67 %). Several different algorithms are applied for building the classification model, each of them using different classification techniques. The WEKA Explorer application is used, each classifier is applied for two testing options – cross validation (using 10 folds and applying the algorithm 10 times – each time 9 of the folds are used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing).[20]

Using K-means clustering which used for pattern recognized classification and clustered students according to their class performance, sessionals and attendance record. .Using K-Means Clustering clustered the students based on their Class Performance, sessionals and Attendance in class. Centroids are calculated from the educational data set taking K-clusters. This study is helpful to notify the students with less attendance and slow performance in sessionals but also enhances the decision-making approach to monitor the performance of students. Also on increasing the value of K, the accuracy becomes better with huge dataset and Kmeans can find the better grouping of the data. The results obtained help to cluster those students who need special attention. [23]

Here in one case study authors try to extract useful knowledge from graduate student's data collected from the college of Science and Technology – Khanyounis. This paper discovered association rules and sorted the rules using lift metric. Then other used two classification methods which are Rule Induction and Naïve Bayesian classifier to predict the Grade of the graduate student. Finally, outlier detection to detect all outliers in the data, two outlier methods are used which are Distance-based Approach and Density-Based Approach. Each one of these tasks can be used to improve the performance of graduate student. [22]

The comparison of data mining algorithms for clustering published by Shiv Pratap Singh Kushwah, Keshav Rawat, Pradeep Gupta .These algorithms are among the most influential data mining algorithms in the research community. A Knn algorithm is more sophisticated approach, k-nearest neighbor (kNN) classification, finds a group of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood. KNN classification is an easy to understand and easy to implement classification technique. Despite its simplicity, it can perform well in many situations. [17]

Apriority is a seminal algorithm for finding frequent itemsets using candidate generation. This paper is applied partition based apriori algorithm and divide problem in smaller partition. The dataset is divided into n non-overlapping partitions such that each partition fits into main memory and each partition is mined separately. [17]
Apart from one other research paper, classification of the data collected from students of polytechnic institute has been discussed. This paper used Weka tool and also compares results of classification with respect to different performance parameters. This study shows more attention should be given to improve basic fundamentals in mathematics and language based on some academic attributes of students. Comparing results of DT and NB algorithms is observed that Decision tree (J48) gives better result than Naïve Bayesian algorithm in terms of accuracy in classifying the data. [13]

A research paper surveys the most relevant studies carried out in EDM using Apriority algorithm. The improved Apriority algorithm proposed in this research uses bottom up approach along with standard deviation functional model to mine frequent educational data pattern.[5]This paper presents an Improved Apriority algorithm based on Bottom up approach and Support matrix to identify frequent item set. This algorithm works in 2 phases, Support Matrix Generation and Bottom Up approach to mine frequent items set based upon calculated minimum support. [5] The Major advantage of this algorithm is that it avoids comparison of currently chosen transaction with other transaction to mine the frequent item set if the total Support value or count of the other transactions on which comparison needs to be done is lesser than the chosen transaction.

4. Conclusion and future work:

Majority researchers works on classification, clustering and association rules and for this research WEKA data mining tool was mostly preferred for experiments. Above research papers are included academic performance of student, analysis of result to find out poor academic record of students and find out solution to improve it. Now a day's so many tools are available on market which are free source. So, in future we will try to implement our work on various tools with various techniques which will help us to find out more quality of attributes which affect the lower performance of student.

5. References:

[1] S. Saranya, N.Tamilselvi , P.Usha, M.Yasodha, V.Padmapriya, **,** **"**Data Mining Techniques in EDM for Predicting the Pupil's Outcome", *International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 6, June 2013,*ISSN (Print) : 2319-5940
[2] Kalyani M Raval , " Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012 ISSN: 2277 128X

[3] Mrs. M.S. Mythili, Dr. A.R.Mohamed Shanavas ," An Analysis of students' performance using classification algorithms",*IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. III (Jan. 2014), PP 63-69 www.iosrjournals.org.*

[4] Brijesh Kumar Baradwaj , Saurabh Pal , "Mining Educational Data to Analyze Students" Performance" , *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011*

[5] Jayshree Jha and Leena Ragha , "Educational Data Mining using Improved Apriori Algorithm", International Journal of Information and Computation Technology. ,ISSN 0974-2239 Volume 3, Number 5 (2013), pp. 411-418,© International Research Publications House, http:,,www. irphouse.com ,ijict.htm

[6] Ramaswami M, "Validating Predictive Performance of Classifier Models for Multiclass Problem in Educational Data Mining" , IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014 ,ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784 ,www.IJCSI.org

[7] Azwa Abdul Aziz, Nur Hafieza Ismail, Fadhilah Ahmad, " Mining Students' Academic Performance" , Journal of Theoretical and Applied Information Technology 31st July 2013. Vol. 53 No.3 © 2005 - 2013 Jatit & Lls. Issn: 1992-8645

[8] B R Prakash, Dr.M. Hanumanthapp, Vasantha Kavitha , "Big Data in Educational Data Mining and Learning Analytics" , International Journal of Innovative Research in Computer and Communication Engineering, *(An ISO 3297: 2007 Certified Organization),* Vol. 2, Issue 12, December 2014

[9] Er. *Rimmy Chuchra* "Use of Data Mining Techniques for the Evaluation of Student Performance:A Case Study" „International Journal of Computer Science and Management Research Vol 1 Issue 3 October 2012 ,ISSN 2278-733X, www.ijcsmr.org

[10] Samrat Singh, 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 (IJCSN) Volume 1, Issue 4, August 2012 www.ijcsn.org ISSN 2277-5420 ,Page | 121

[11] Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás, "Data Mining Algorithms to Classify Students".

[12] K.Shanmuga Priya & A.V.Senthil Kumar , "Improving the Student's Performance Using Educational Data Mining", Int. J. Advanced Networking and Applications, Volume: 04 Issue: 04 Pages:1680-1685 (2013) ISSN : 0975-0290

[13] Vaibhav P. Vasani, Rajendra D. Gawali, "Classification and performance evaluation using data mining algorithms", International Journal of Innovative Research in Science, Engineering and Technology ,*(An ISO 3297: 2007 Certified Organization),* Vol. 3, Issue 3, March 2014

[14] Sonali Agarwal, G. N. Pandey, and M. D. Tiwari, "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012.

[15] Richard A. Huebner, "A survey of educational data- mining research". Research in Higher Education Journal.

[16] Jing Luan,"Data mining application in higher education".

[17] Shiv Pratap Singh Kushwah, Keshav Rawat, Pradeep Gupta , " Analysis and Comparison of Efficient Techniques of Clustering Algorithms in Data Mining" , International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-3, August 2012
[18] Han and Kamber ,Data Mining concepts and techniques.

[19] Ritu Chandra and Prof. (Dr.) Shaikh Azimuddin , " Influence of Socio Economic Status On Academic Achievement Of Secondary School Students Of Lucknow City" , International Journal of Scientific & Engineering Research, Volume 4, Issue 12, December-2013 ISSN 2229-5518

[20]Dorina Kabakchieva," Predicting Student Performance by Using Data Mining Methods for Classification", BULGARIAN ACADEMY OF SCIENCES
CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **13**, No 1, Print ISSN: 1311-9702; Online ISSN: 1314-4081.

[21]Edin Osmanbegović , Mirza Suljić , "DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE ", Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012

[22]Mohammed M. Abu Tair, Alaa M. El-Halees , "Mining Educational Data to Improve Students' Performance: A Case Study ", International Journal of Information and Communication Technology Research , Volume 2 No. 2, February 2012 ISSN 2223-4985

[23]Pratiyush Guleria, Manu Sood , "Mining Educational Data Using K-Means Clustering ", International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 3, Issue 8 October 2014