ISSN:0975-9646

S. Lakshmi Prabha et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4440-4443

Application of Educational Data mining techniques in e-Learning- A Case Study

S. Lakshmi Prabha[1] Dr.A.R.Mohamed Shanavas[2]

[1] *Ph.D Research Scholar, Bharathidasan University & Associate professor,Department of Computer Science, Seethalakshmi Ramaswami College, Tiruchirappalli, Tamilnadu, India,*

[2] *Associate professor,Department of Computer Science, Jamal Mohamed College, Tiruchirappalli, Tamilnadu, India,*

***Abstract*-An emerging interdisciplinary research field known as educational data mining (EDM) is concerned with developing methods and applying data mining techniques for exploring the unique types of data that come from educational systems. Its goal is to better understand how students learn and identify the settings in which they learn to improve educational outcomes. This paper explains how the data mining tasks classification, prediction, and clustering can be applied to data taken from an e-learning system. The performance of sixth grade school students are taken for the analysis and present the results achieved with WEKA tool.**

Keywords
EDM, Classification, Clustering, Association, WEKA

I. INTRODUCTION

Data Mining is an effective tool to extract meaningful and interesting patterns from the current and historical data stored in data warehouses or data repositories which may be analyzed to predict future trends. Seeking knowledge from massive data is one of the most desired attributes of Data Mining. Data could be large in two senses: in terms of size & in terms of dimensionality. With the help of data mining tools and emerging research trends in this field, the data miner may extract knowledge from the large data marts / data warehouses very efficiently and quickly which may be used for the betterment of the organisations and the society.

Data mining had touched many fields including bioinformatics, ecommerce, fraud detection and now in the field of education as well. The data mining in the field of educational research is known as Educational Data Mining (EDM) EDM often tries to simulate a student models which may be used for the improvement of students by predicting the future trends [11]. This paper explains the methods of EDM through an example data taken from an e-learning platform. This paper is organised as follows: section 2 describes about EDM and important methods classification, prediction and clustering. Section 3 lists the related works in this area. Section 4 explains how WEKA tool is used for classifying and clustering the students' data taken from MATHSTUTOR [12] and section 5 concludes the work.

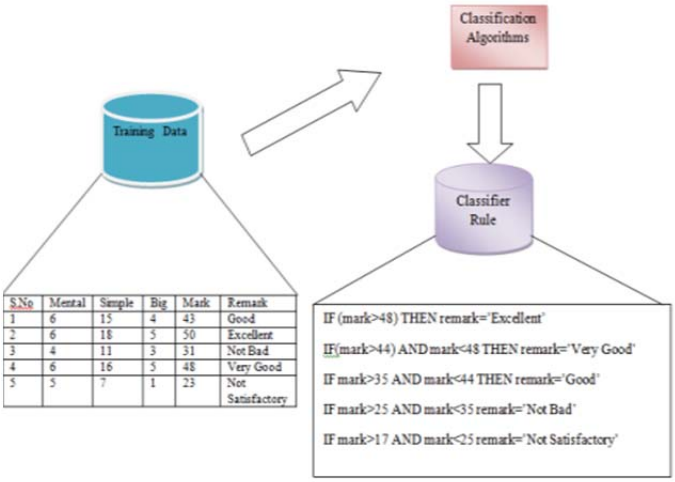
II EDUCATIONAL DATA MINING

Education is increasingly occurring online or in educational software, resulting in an explosion of data that can be used to improve educational effectiveness and support basic research on learning. Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [1]. There are varieties of popular data mining tasks within the educational data mining e.g. classification, clustering, outlier detection, association rule, prediction etc. We can use the EDM methods in educational system as: predicting drop-out student, relationship between the student examination results & their success, predicting student's academic performance, discovery of strongly related subjects in the school syllabi, knowledge discovery on academic achievement, classification of students' performance in all courses according to learning style.

2.1 Classification

Classification is a data mining task that maps the data into predefined groups & classes. It is also called as supervised learning .It consists of two steps:

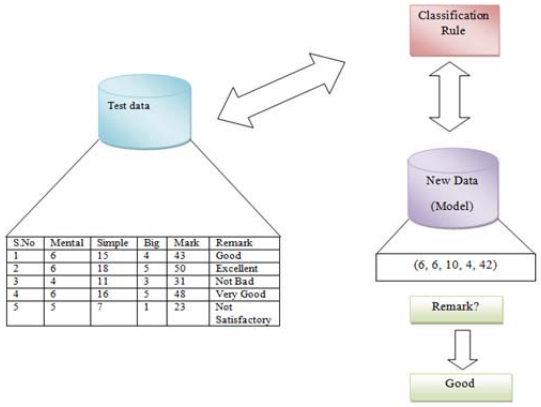
1**. Model construction:** It consists of set of predetermined classes. Each tuple /sample is assumed to belong to a predefined class. The set of tuple used for model construction is training set. The model is represented as classification rules, decision trees, or mathematical formulae. This model is shown in figure 1.



S.No	Mental	Simple	Big	Mark	Remark
1	6	15	4	43	Good
2	6	18	5	50	Excellent
3	4	11	3	31	Not Bad
4	6	16	5	48	Very Good
5	5	7	1	23	Not Satisfactory

Figure 1: model construction

2. **Model usage:** This model is used for classifying future or unknown objects. The known label of test sample is compared with the classified result from the model. Accuracy rate is the percentage of test set samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur. This model is shown in figure 2. The decision tree is used to represent logical rules of student final remark.

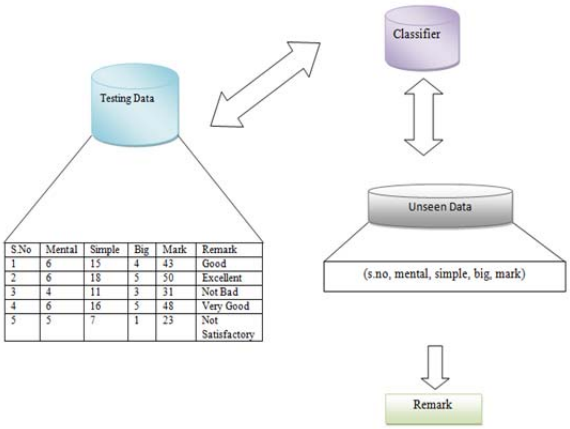


S.No	Mental	Simple	Big	Mark	Remark
1	6	15	4	43	Good
2	6	18	5	50	Excellent
3	4	11	3	31	Not Bad
4	6	16	5	48	Very Good
5	5	7	1	23	Not Satisfactory

Figure 2: Model Usage (Classification)

2.2 Prediction

It is used to model continuous-valued functions, i.e., predicts unknown or missing values. In this model we deduce single aspect of data from some combination of other aspect of data. In educational data mining prediction can be used to detect student behavior, predicting or understanding student educational outcomes. This model is shown in figure 3.



S.No	Mental	Simple	Big	Mark	Remark
1	6	15	4	43	Good
2	6	18	5	50	Excellent
3	4	11	3	31	Not Bad
4	6	16	5	48	Very Good
5	5	7	1	23	Not Satisfactory

Figure 3: Prediction model

2.3 Clustering

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group [7] Clustering can be considered the most important unsupervised learning technique. Clustering & its classification is shown in figure 4.

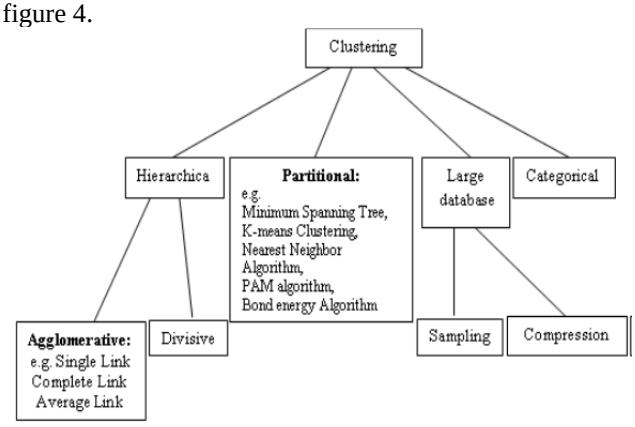


Figure 4: Classification of clustering algorithm

In educational data mining, clustering has been used to group the students according to their behaviour e.g. clustering can be used to distinguish active student from non-active student according to their performance in activities.

III. Related Work

As educational data mining has become a significant research area, a number of studies have been conducted on educational data mining in order to discover the effect of using it on students' performance [2]-[9]. The researchers will give an overview of a few representative works in this section.

The paper [2] the authors applied classification algorithms Decision Tree, Neural Network and K-Nearest Neighbour and identified that Neural Network predicts with higher accuracy.

In paper [3] the authors identified the students with low AGPA using neuro-fuzzy inference systems and the experimented results showed 97% accuracy.

In this paper [4], students' performance is evaluated using association rule mining algorithm.

Paper [5] presents a classification model based on decision tree approach to predict students' academic performance.

Paper [6] used k-means clustering to analyze students' learning behaviour. Class quizzes, final exams and assignments are considered for the study. This study helped to identify the dropout ratio and improve the performance.

The authors [7] analyzed the performance of final year Under Graduate IT course students using data mining techniques in WEKA.

In paper [8] the authors used Bayes classification to predict the students; results on the basis of previous year database.

The author in paper [9] used association rule minig and the analysis revealed that the students' university performance is dependent on unit test mark, assignment, attendance, and graduation percentage.

IV. WEKA As EDM Tool

For the purposes of this study, we select WEKA (Waikato Environment for Knowledge Analysis) software that was developed at the University of Waikato in New Zealand**.** WEKA is open source software issued under the GNU General Public License. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is portable & platform independent because it is fully implemented in the Java programming language and thus runs on almost any modern computing platform and is now used in many different application areas, in particular for education & research.

Data set used

This paper considers the MathsTutor [12] as an e-learning environment implemented for Tamilnadu State Board Mathematics syllabus used in India for the school students of grade 6th to 8th in the area of mensuration. The students learn the problem through examples, understand through doing exercises and test his knowledge through test. Each of these categories has three types of problems (mental, simple, big) according to number of steps required for

solving and knowledge requirement. The number of problems a student solves is 60(example-20+exercise-20+test-20). Sixty students of 6th grade participated in this project. So total of 3600 entries are stored along with the present status of student (like completed the learning material/completed how many problems in each category etc.) keystrokes, steps, results, feedback, number of hints used, time spent in solving a problem. This makes the data big and EDM tools are required for analysis.

Here we consider the log of sixty sixth grade students from MATHSTUTOR [12] who attended the test module of lesson1. The Students are solving three kinds of problems namely Mental Problem, Simple Problem and Big Problem. According to their performance in these tests the mark is evaluated and remark attribute is assigned with Excellent/Very good/ Good/ Not bad/ Not satisfactory. Figure 6, shows the graph for remark attribute, 15 students got the remark Good, 17 students got the remark Excellent, 3 students got the remark Not Bad, 23 students got the remark Very Good& 2 students got the remark Not Satisfactory.

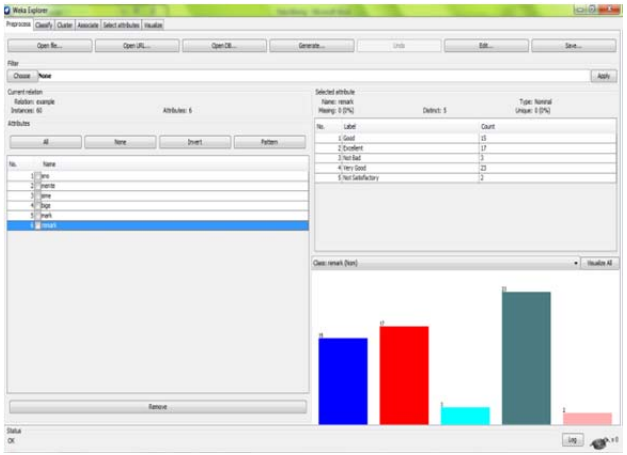


Figure 6: Weka 3.6.9 with Explorer window open with Example dataset

In WEKA, J48 classifier predicts the majority of class in training data. It predicts the mean for numeric value & mode for nominal class. As shown in figure 7, given 60 instances 58 are classified correctly and 2 incorrectly. The accuracy is 96.667% .The mean absolute error and Root mean squared error are 0.01 and 0.066 respectively and they are very minimum. In the detailed accuracy report by class specifies TP rate (True positive rate) is 1 for all classes except for 'Not satisfactory' class. Precision and F-measure are 0 for 'Not satisfatory', 0.6 and 0.75 for 'Not Bad' and 1 for other 3 classes. The confusion matrix shows how instances classified as:

a- Good- 15
b- Excellent – 27
c- Not bad -3
d- Very Good -23
e- Not Satisfactory-2

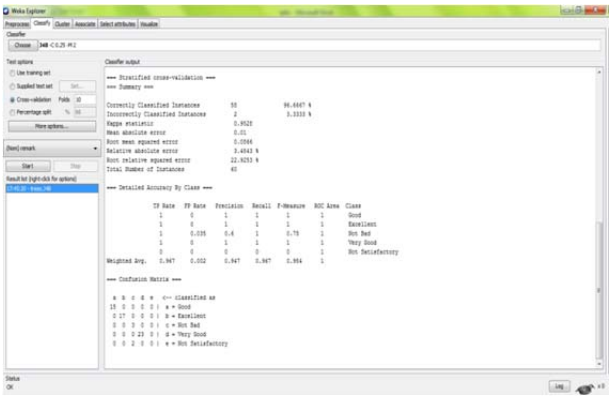


Figure 7: Result of Classifier

The following Figure 8. Shows the visualizer for the result of J48 classifier algorithm.

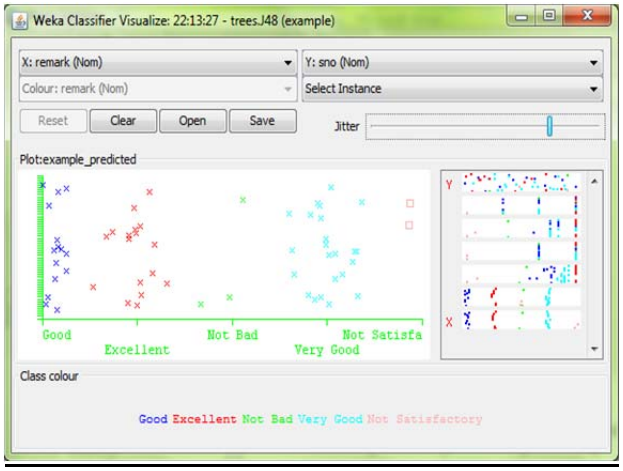


Figure 8: Result of Classifier Visualizer

Now consider the clustering algorithm EM. This algorithm is used to create the cluster with a minimum size & density. This algorithm also handles the outlier problem. The result of training set using this algorithm is given in figure 9. Five clusters are derived according to the remark field and figure 10 shows visualize for the same.

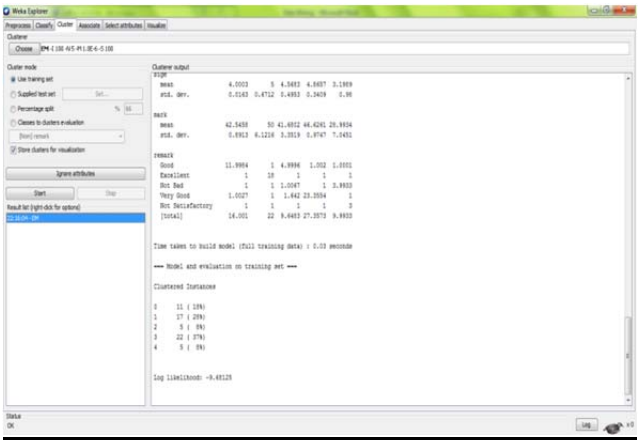


Figure 9: Result of Cluster

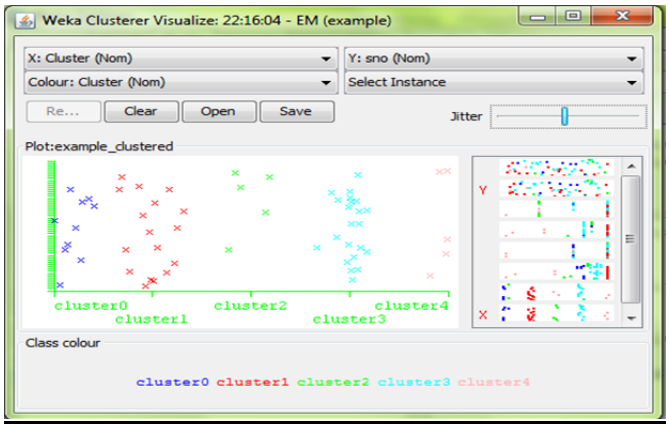


Figure 10: Result of Cluster Visualize

V. Conclusion & Future Work

This paper, studied how EDM methods can be applied to educational data, particularly to improve students' performance. We used students' data from the database of sixth grade students of a school who worked in Mathstutor[12]. The training set contains 60 records each for a student. This paper provides only limited number of screen shots applied on student data. By classifying the knowledge level of students with test marks and grouping them will make easier for the teacher to concentrate the areas for week students.

We applied data mining techniques to discover knowledge. We discovered classification using J48 algorithm. Also we clustered the student into group using EM clustering algorithm. Each one of this knowledge can be used to improve the performance of a student. It is tested with testing data and accuracy is above 90%, which assures that these algorithms are well suited for classifying large set of student data.

For future work more granular data about student activity log having hints used and time taken for solving the problems are to be analyzed.. And data mining algorithms could be embedded into e- learning system so that one using the system can be benefited from the data mining techniques.

References

[1] C. Romero, S. Ventura, Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, Vol. 40*, n. 6, pp. 601-618, 2010. http://dx.doi.org/10.1109/TSMCC.2010.2053532

[2] Dorina Kabakchieva," Student Performance Prediction by Using Data Mining Classification Algorithms", International Journal of Computer Science and Management Research Vol 1 Issue 4 November 2012.

[3] A. B. E. D. Ahmed, I. S. Elaraby, Data Mining: A prediction for Student's Performance Using Classification Method. *World Journal of Computer Application and Technology, Vol. 2, n.* 2, pp.43-47, 2014.

[4] S. Borkar, S. Rajeswari, Predicting Students Academic Performance Using Education Data Mining. *IJCSMC International Journal of Computer Science and Mobile Computing*, Vol. 2, Issue. 7, July 2013, pg.273 – 279

[5] Sonia Joseph, Laya Devadas, Student's Performance Prediction Using Weighted Modified ID3 Algorithm, International Journal of Scientific Research Engineering & Technology, Volume 4, Issue 5, May 2015,pg.57-575.

[6] Md. Hedayetul Islam Shovon , Mahfuza Haque, Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 7, July 2012,pg 353-355.

[7] Sunita B Aher and Lobo L.M.R.J.. Data Mining in Educational System using WEKA. *IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT)* (3):20-25, 2011.

[8] Umesh Kumar Pandey S. Pal, "Data Mining : A prediction of performer or underperformer using classification," International Journal of Science and e and Information Technologies, Vol. 2 (2) , 2011, 686- 690.

[9] Suchita Borkar and K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining," International Journal of Computer Science and Mobile Computing. IJCSMC, Vol. 2, Issue. 7, July 2013, pg.273 – 279

[10] A.Peña-Ayala, Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications, Vol. 41*, n. 4, pp. 1432-1462, 2014. http://dx.doi.org/10.1016/j.eswa.2013.08.042

[11] Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3-17.

[12] Lakshmi Prabha S. And AR. Mohamed Shanavas, Implementation of E-Learning Package for Mensuration-A Branch of Mathematics, IEEE, ISBN: 978-1-4799-2876-7,Pg: 219 – 221, DOI: 10.1109/WCCCT.2014.37