



A Survey of Data Mining Approaches in Performance Analysis and Evaluation

Nikitaben Shelke, Shriniwas Gadage

Computer Engineering, Savitribai Phule Pune University,
Maharashtra, India

Abstract— *Education Domain is very big and spread across complete geographical locations. It is widespread in other domain as well. Considering the influence of education, there is enormous amount of data is available. This data is related to teaching, learning, and assessment of students across all educational institutes. In recent years, EDM is considered the most efficient system to overcome the statistical problems of performance evaluation and prediction due to advanced data mining techniques.*

This paper aims to survey the research done in the field of education data mining over past decade. This paper describes various applied data mining techniques for student performance analysis, prediction and evaluation.

Keywords— *Data Mining, EDM, Performance Evaluation, Analysis, Prediction*

I. INTRODUCTION

In recent years, there has been enormous growth of data related to students' academics. This data can be used to improve the quality by analysing the patterns of performance during the complete tenure of student in an institute. Analysis and prediction of student performance play a major role in improvement of quality of education in various aspects. Educational database contain the useful information for predicting a students' performance, rank factor & details. Student's academic performance in educational environment is based upon the psychological and environmental factor is predicted by different educational data mining techniques.

In general, EDM considers all data collected during any educational process i.e. e-learning data along with the data collected from traditional learning system. Online Learning Management System is being used to create virtual learning groups from different geographical locations that help in rapid growth of collaboration in learning. This approach is being very useful in the educational data mining in recent times. In this case, discovering the learning patterns of students by analysing the system usage can play a most important role.

The student performance can be assessed with the help of few applications such as tools which gives learning recommendations to students based on analysis, approaches providing the evaluation and tools involving the feedback to students based on the evaluation performed.

The techniques studied in this paper are classification, clustering, decision tree, sequential pattern mining, text mining, and many other. Also the summary of the tools useful for educational data mining has been given along with their functions, associated data mining techniques and the supporting platforms.

II. LITERATURE REVIEW

Md.Hedayetul Islam Shovon [1] have done the analysis of the student's learning behavior by using the k-means clustering. He has taken the external and internal assessment of university students into consideration from student's database. This model tries to indicate students who are the weak in learning before the final assessment.

Oyelade, O. J [2] presented a method of using k-means clustering algorithm along with deterministic model as an efficient tool to supervise the progress of students' performance in different courses provided by the institutions. They have used Euclidean distance measure of similarity[2]. The results shown in this paper are the numerical in nature showing the student's performance evaluation.

D. Kabakchieva [3] have the paper which is focused on the techniques and methods of educational data mining for knowledge discovery from the data collected by universities. The main goal of their work is to recognize the patterns in data to predict the student's performance. They have done the analysis using classification algorithm such as Bayes classifiers and a Nearest Neighbour classifier as well as rule learners, a decision tree classifier [3]. They have implemented these two classifiers in WEKA; a open source tool used for data mining algorithms.

Yaday, Pal[4] have shown the data mining techniques for predicting the results of students. The current results have been analyzed and then comparative analysis of the results has been done. This shows that the prediction has helped the average students to perform better in the final examinations.

Edin, Mirza [5] have investigated the different factors which impacts the results of students. They have done the comparison of Bayesian classifier, neural networks and decision trees. The results of their work is shown by distribution graphs and different class labels. This analysis was conducted after the training and testing of the algorithms, making it possible to conclude on possible predictors of students' success [5].

Merceron, Yacef [6] have presented the results of evaluation of students by using the clustering algorithms based on the mistakes done by the students. The data of students have been gathered from the web-based tool used by School of Information Technologies at the University of Sydney [6]. They have used k-means clustering and hierarchical clustering for the experiment of the evaluation of students. Their results have shown the advantages of hierarchical clustering over k-means clustering where the results of hierarchical clustering have neutral results than the k-means clustering methods.

Tair, El-Halees[7], have presented the case study of the work done to improve the performance of students and help the weak students. They have collected students' data of 15 years from College of Science and Technology – Khanyounis. The techniques used for this research are classification, clustering, association and outlier detection [7]. They have used Rapid Miner. They have considered 18 attributes for the analysis of performance of students.

El-Halees [8], gave a case study that for educational data mining to understand the learning behavior of students. The data considered for this case study includes academic and course records, logs from the e-learning system as well as the personal records. After applying association rules and classification rules with decision tree, there has been EM clustering [8] to group students. This discovered knowledge have reveled the benefits of data mining techniques to help students improve their overall performance.

Baradwaj and Pal [9], applied the decision tree classification algorithm as data mining technique to evaluate student' performance. The goal of their study is to extract knowledge that describes students' performance in end semester examination. They used student' previous database containing Previous Semester Marks, Class Test Grade, Seminar Performance [9], Assignment Marks, General Proficiency, Attendance, and Lab Work. This study helps to identify the students need special attention and teachers can also give the guidance based on this analysis.

Al-Radaideh [10], have applied classification technique for knowledge extraction and extracted classification rules are based on the decision tree as a classification method. The analysed and evaluated results allows predicting final grade of students. This prediction helps in finding the parameters affecting good performance of students. By overcoming these parameters, the higher education institutes can improve the quality of educational system.

Banumathi, Pethalakshmi[11] have proposed several Data Mining approaches and techniques(clustering, classification and association analysis) for combined use in the mining of student's assessment data in Learning Management System.

Liu, Zhao [12] have suggested non-linear correlation techniques for EDM. For the improvement of e-learning environments they have used techniques viz., Association rules, clustering, classification, sequential pattern analysis, dependency modelling, and prediction [12]. This will enhance the extent of learning behaviour which will help the teacher to evaluate in better manner.

Ayesha [13] has proposed the system in their paper which have considered the students performance data in previous years such as mid-exam and final exam marks, class quiz and assignment marks and have applied K-means algorithm to predict the failing ratio beforehand. This work had helped the teachers to take appropriate measures and help students to improve their grade.

Lahane, Kharat[14] have applied the clustering on Moodle Log data for the improvement of e-learning environment. Tie, Zhuang[15] have also used the Moodle log data and have studied the comparison of various data mining algorithms to classify students to predict the final grades.

Prediction of student's performance (final grade) based on features extracted from logged data is presented in Jing, Shiyong [16] and university academic student performance is presented in Stes, Petegem [17]. Prediction of students' online examination marks (using an orthogonal search-based rule extraction algorithm) is presented in Rashid, Taib [18].

III. OPEN SOURCE TOOLS

There has been rapid growth in the research of educational data mining and the use of open source tools have also become popular because of the efficiency and ease of use of those tools. The following table summarizes the open source tools available for Data mining in EDM. The table shows their functions, compatible platforms, techniques and area of applications.

Table 1: useful open source tools for data mining

Name of Tool	Function	Techniques	Area of Application	Technology	Platform
KNIME (Konstanz Information Miner)	Data analysis and visualization- KNIME allows users to visually create data flows (or pipelines), selectively execute some or all analysis steps and later inspects the results, models, and interactive views [20].	ETL, Clustering, Text mining, Image mining,	CRM, BI Financial Data Analysis, Enterprise Reporting [20].	Written in Java and based on Eclipse	32 bit – Linux and Windows 64 bit- LINUX, WINDOWS and Mac OS X
RAPIDMINE R (Formerly)	RapidMiner functionality can be extended with	(Extract, transform, load (ETL), Data	Business and Industrial	Written in the Java	LINUX, WINDOWS

Known As Yale)	additional plug-ins. The Rapid Miner Extensions marketplace provides a platform for developers to create data analysis algorithms and publish them to the community.	pre-processing and visualization, Predictive analytics and statistical modelling	Applications, Education, Training, Rapid Prototyping,	Programming language	and Mac OS X
WEKA (Waikato Environment for Knowledge Analysis)	Collection of visualization tools and algorithms for data analysis and predictive modelling, graphical user interfaces for easy access to this functionality.	Clustering, Classification, Regression, and Feature selection	Educational Research (EDM)	Java	LINUX, WINDOWS and Mac OS X
CARROT	A native C# / .NET API is available. Other non-Java platforms, such as PHP or Ruby, can call Carrot clustering through its REST interface. Offers ready-to-use components for fetching search results from various sources including Google API, Bing API, e-Tools Meta Search, SOLR, and more [21].	Clustering	Text mining and Search Results cluster Analysis	Java	LINUX, WINDOWS and Mac OS X
R-Programming	The R language is widely used among data miners for its ease of use and extensibility. Used for linear and nonlinear modelling, classical statistical tests, time series analysis [22]	Classification, Clustering,	Development of Statistical Software and Data Analysis	Primarily in developed C language & Fortran. A lot of its modules are written in R language	LINUX, WINDOWS and Mac OS X
ORANGE	Designed to simplify the assembly of data analysis workflows and crafting of data mining approaches from a combination of existing components	Classification, association, clustering, regression	Bioinformatics and Text Mining	Scripts in Python, which build upon C++ implementations [23]	LINUX, WINDOWS and Mac OS X
iDA (iData Analyzer)	An Excel-Based Data Mining Tool, iDA consists of a pre-processor, three data mining tools, and a report generator	Clustering	Business Intelligence and CRM, ERP	-	Windows, Linux, Solaris

IV. CONCLUSIONS

There are various tools and techniques available for data mining in this rapid growth of educational and e-learning domain. These tools have different compatibilities with different types of applications. In this paper, the survey of data mining techniques used for the performance analysis and evaluation is done. Also, the summary of open source tools for data mining is presented. Even though, an elaborate explanation is required, we hope this survey gives an insight about the techniques for performance analysis, prediction and evaluation.

REFERENCES

- [1] Md. Hedayetul Islam Shovon, "Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 2(7), July 2012.
- [2] Oyelade, O. J, "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance", (*IJCSIS*) *International Journal International Journal of Computer Science and Information Security*, Vol. 7, No. 1, 2010

- [3] D. Kabakchieva, “Analyzing University Data for Determining Student Profiles and Predicting Performance”, *Cybernetics and Information Technologies*, Vol.1(3), March 2013.
- [4] Yadav, Pal, “ Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification”, *World of Computer Science and Information Technology Journal (WCSIT)* , 2012
- [5] Edin Osmanbegović , Mirza Suljić “Data Mining Approach For Predicting Student Performance”, *Economic Review – Journal of Economics and Business* ”, May 2012
- [6] Agathe Merceron, Kalina Yacef., “Clustering Students to help Evaluate Learning”. *Jean-Pierre Courtiat, Costas Davarakis, Thierry Villemur. Technology Enhanced Learning, 2004, Toulouse, France, 2004*
- [7] Tair, El-Halees, “ Mining Educational Data to Improve Students’ Performance: A Case Study ”, *International Journal of Information and Communication Technology Research*, February 2012
- [8] El-Halees, A. “Mining Students Data to Analyze Learning Behavior: A Case Study”, *The 2008 international Arab Conference of Information Technology (ACIT2008) – Conference Proceedings, University of Sfax, Tunisia*, Dec 15- 18.
- [9] Baradwaj, B. and Pal, S. “Mining Educational Data to Analyze Student s’ Performance”, *International Journal of Advanced Computer Science and Applications*, 2011
- [10] Al-Radaideh, Q., Al-Shawakfa, E. and Al-Najjar, M. “Mining Student Data Using Decision Trees”, *The 2006 International Arab Conference on Information Technology (ACIT'2006) – Conference Proceedings*.
- [11] A. Banumathi and A. Pethalakshmi, “A novel approach for upgrading indian education by using data mining techniques,” in *Proc. 2012 IEEE Int. Conf. Technol. Enhanc. Educ. (ICTEE 2012)*, 2012, pp. 7-11.
- [12] B. Liu and J.-H. Zhao, “Non-linear correlation techniques in educational data mining,” in *Proc. 2009 Sixth Int. Conf. Fuzzy Syst. Knowl. Discov.*, 2009, pp. 270-274.
- [13] Ayesha, S. , Mustafa, T. , Sattar, A. and Khan, I. (2010) “Data Mining Model for Higher Education System”, *European Journal of Scientific Research*, 2010
- [14] S. V Lahane, M. U. Kharat, and P. S. Halgaonkar, “Divisive approach of clustering for educational data,” in *Proc. 2012 Fifth Int. Conf. Emerg. Trends Eng. Technol. (ICETET 2012)*, 2012
- [15] Z. W. Tie, R. Jin, and H. Zhuang, “The research on teaching method of basics course of computer based on cluster analysis,” in *Proc. 2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010)*, 2010, pp. 2001-2004.
- [16] F. Jing and K. Shiyong, “Application of data mining for emotional intelligence based on cluster analysis,” in *Proc. 2010 Int. Conf. Artif. Intell. Educ.*, Oct. 2010,
- [17] A. Stes and P. Van Petegem, “Profiling approaches to teaching in higher education: A cluster-analytic study,” *Stud. in High. Educ.*, vol. 39, issue 4, pp. 1-15, 2014.
- [18] N. A. Rashid, M. N. Taib, S. Lias, and N. Sulaiman, “Classification of learning style based on Kolb’s learning style inventory and EEG using cluster analysis approach,” in *Proc. 2010 2nd Int. Congr. On Eng. Educ. (ICEED)*, pp. 64-68, 2010
- [19] Romero, C. and Ventura, S. “Educational data mining: A Survey from 1995 to 2005”, *Expert Systems with Applications* (33), 2007
- [20] <http://en.wikipedia.org/wiki/KNIME>
- [21] [http:// project.carrot2.org](http://project.carrot2.org)
- [22] www.r-project.org/
- [23] Demsar, Curk, Erjavec “Orange: Data Mining Toolbox in Python”, *Journal of Machine Learning Research* 14 (2013)