

A Survey on the Result Based Analysis of Student Performance using Data Mining Techniques

K. Govindasamy¹, T.Velmurugan²

¹Research Scholar, VELS University, Chennai, India.

²Associate Professor, PG and Research Department of Computer Science, D. G. Vaishnav College, Chennai, India
E-Mail: mphilgovind@gmail.com, velmurugan_dgvc@yahoo.co.in

Abstract—Extraction of information available in various data base repositories is a tedious task. A composed works of Data Mining (DM) method is accessible for different category of applications for the same work. Many researchers involve analyzing the student's performance using some relevant DM techniques. This attracting little field is named as Learning Data Mining (LDM). The organizations of the syllabus also increase a very big contact about the growth of the student's information and their performance. Among the different data mining methods, classification plays a vital role in learning data mining. The primary intention of this research work is to cross the data mining methods which are apply for the improvement of the student's performance and also identify the most excellent appropriate structure of syllabus for the new environment. This study investigates about the use of ID3 and C4.5 classification algorithms for the improvement of student performance evaluation system. A comparative analysis of various works is carried out in this survey to identify the best classification algorithms for LDM.

I. INTRODUCTION

Data Mining (DM) is used to extract the hidden information from the huge volume of data storage. Data Mining plays a key role in a particular field especially in a termed learning analytics or educational data mining to examine for a quality assortment. It seeks to find ways to make helpful use of the enlarging amount of data about learners to understand the process of learning and the social motivational factors encircle learning [7]. The data collected from different applications involve the proper method of extracting knowledge from a large repositories for enhanced decision - making. Knowledge discovery in databases (KDD), often called data mining, aims at a discovery of useful information from a large collections of data [5].

These techniques supply a route to a multiple level of ranking, quest which gives a new perception of how people can become proficient in these educational sectors. Data mining has various techniques such as Classification, Clustering, Prediction, Association rules, Decisions Tress, Neural Networks and many others. Among these, classification algorithm such as Decision tree algorithms ID3 and C4.5 plays an essential role in EDM [7]. With these potential techniques, one can explore an essential hidden pattern for finding future direction and helps in decision - making. Data mining helps the educators by managing effective educational units, redesigning the curriculum and improving the hearing methods. There are several arrangement algorithms in DM, in which the decision tree algorithms are mostly used because of its easy implementation. Self-efficacy has been related to persistence, tenacity, and achievement in educational settings [8]. A meta-

analysis of research in educational settings found that self-efficacy was related both to academic performance and persistence. The contribution of self-efficacy to the educational achievement is based both on the increased use of particular cognitive activities, strategies and on the positive contact of efficacy beliefs on the broader, more general classes of meta cognitive skills and coping abilities.

Computer based experiment have been used since 1960s to experiment an information and problem solving skills, [22]. A computer based evaluation systems have enabled learners and trainers promoting to authors, schedule, deliver and report on investigation, quizzes, experiment and exams. An effective method of student assessment technique is necessary in assessing the student knowledge. Due to an augment in students number, ever-escalating work commitments for academic staff, and the advancement of internet technology, the use of computer assisted evaluation has been an attractive proposition for several senior learning institutions. The rest of the paper is structured as follows. Section II discuss about the data mining methods which are used for educational purpose. The applications of educational DM using DM methods are illustrated in section III. Section IV reviews about the use of first year student adjustment in curriculum development. Finally section V concludes the investigation work.

II. DATA MINING TECHNIQUES FOR STUDENT PERFORMANCE

Han and Kamber [12] describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process. The primary objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here [5]. DM Techniques have to do with the discovery of hidden association that being in educational databases. As we know wide range of data is stored in educational databases, so in order to get required data and to find the hidden relationship, for that different data mining techniques are developed and used [7].

They extend a user-friendly software tool which is based on neural network classifiers for guessing the student's presentation in the course of "Mathematics" of the final year of

Lyceum. Based on their numerical experiments, they conclude that the MSP-trained neural networks display more consistent activities and illustrate better classification results than the other classifiers [18]. This paper reviews an application of data mining in learning system to explore the final year students and presented their result investigation in WEKA device. To improve the academic performances they have increased a system which can schemes the complete of students from their last performances using the techniques of data mining under Classification. They enclose investigation the in order set containing information about students, such as sexual category, marks achieve in the panel test of classes X and XII, marks and grade in opening examinations and consequences in first year of the last group of students. By applying the ID3 and C4.5 classification algorithms on this information, they guess the universal and behavior performance of admitted students in future test.

III. APPLICATIONS OF CLASSIFICATION ALGORITHMS

Educational data mining is explicates as an area of scientific query which occur mainly in and around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to a better understandings in students and the settings which they learn [7]. Mining Educational Data to Analyze Students' Performance is discussed by Brijesh Kumar Baradwaj and Saurabh Pal in [5]. In this research, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, particularly the decision tree method is used. Predicting Academic Success from Student Enrolment Data using Decision Tree Technique done by C. Anuradha and T. Velmurugan [7]. They presented the classification techniques to predict the performance of the students based on the data. Among the various techniques researcher preferred Decision tree method is used to create model and prediction of student's performance also found 92.5% accuracy for the FAIL class.

Learning data mining is nothing but explaining part of scientific qualm which comes to mind at first and around the progress of methods for making breakthrough within the singular kinds of data that come from a learning settings, and using those methods to a better understanding in students and the settings which they learn [15]. Institutions across the globe are migrating toward the use of Computer Based Testing (CBT) to experiment students' knowledge. The advantages of using computer knowledge for learning measurement in a global sense have been recognized and these include lower administrative cost, time saving and less demand upon teachers among others. Some experiment takers reported that, it is more difficult to navigate back about grades, attitudes about convenience, control and validity. Some examiners have a general anxiety about the computer itself, while others are more concerned about their level of computer experience (John et al., 2002). Their result reveals that classifier accuracy shows the true positive rate of the model for the FAIL class is 0.84 for ID3 and C4.5 decision trees that means model is successfully identifying the students who are likely to fail. We have developed a mining tool in order to help the teacher to carry out all this process. This application has

been integrated into the well known AHA! [9] (Adaptive Hypermedia Architecture) system. This application is a Java Applet, just like other AHA! authoring tools. In order to use it, the author has to identify himself, once the user has logged, it is shows the main window of the tool. This window is divided in the main areas and the descriptive study used paper-based surveys and interviews for all data collections are exclusively displayed. To obtain information about the students' perceptions of online assessment, a Web site has developed and implemented. The course instructor was responsible for the instructional design, content creation, and all activities for the course, but one researcher designed and developed the online assessment Web site. The Web site was database driven and developed using Active Server Pages (ASP), a Microsoft Access Database, and Cascading Style Sheets (CSS)[13].

IV. CLASSIFICATION ALGORITHMS FOR CURRICULUM DEVELOPMENT

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at a hefty. This approach frequently employs decision tree or neural network-based classification algorithms 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 groups. In educational data mining, clustering has been used to group students according to their behavior. According to clustering, clusters distinguish the student's performance according to their behavior and activates [1].

Qiuxiang Shi et al. have initiated a special coaching method and analyze learning elements by the application of DM. The researcher creatively realizes the various teaching ideas and network curriculum integration through DM application in the development of network instructional platform of "Recent Learning Knowledge". Learner performance in university courses is of huge concern to the superior learning managements where a number of factors may involve the performance. This paper is an attempt to use the data mining processes, particularly classification, to help in enhancing the quality of the superior learning system by evaluating learner data to study the main attributes that may affect the learner performance in courses. CRISP framework for data mining is used for mining learner related academic data. The classification rule of a generation process is based on the decision tree as a classification method where the generated rules are studied and evaluated. A system that facilitates the use of the generated rules is built which allows learners to predict the final grade in a course under study. The foundation on a choice of data mining methods (DMM) and the submission of piece of equipment learning progression, rules are consequential that enable the classification of learners in their predicted classes. The exploitation of the prototyped solution, integrates measuring, round and reporting measures in the new system to optimize guess accuracy. An academic curriculum is a (legal) record defining a particular learning program that puts few types of constraints on how students are made use of the courses. Curricula authoring is a tough job involving various factors and different kinds of knowledge. The output of their experiment helps management staffs and academic planners to improve overall performance of students

and consequently reduces unsuccessful rate in most academic institutions work carried out by Abeer and Ibrahim in [1]. In their paper, Classification task is used to evaluate the student performance. A comparative analysis of various student performance used by different researchers is given in the table 1, which gives an insight about the each and every paper taken for analysis in this survey work.

Table 1: Comparative Analysis

Paper Ref. No.	Proposed Method	Results		
		Accuracy	Specificity	Sensitivity
1	The decision tree (ID3) method	The decision tree method is used on student's database to predict the student's performance on the basis of student's database.		
2	ID3 classifier has the potential to significantly improve the conventional classification methods for use in performance	95%	97%	99%
3	Progressive mastery enhanced perceived self-efficacy, efficient analytic thinking, challenging goal setting, self-reaction, and organizational performance. Relative decline undermined these self-regulatory factors and produced a growing deterioration of organizational performance.	-	89%	96%
4	New method of predicting preparation for future learning based on quantitative analysis of the Moment-by-Moment Learning Graph the patterns of which were previously shown to be associated with PFL.	96%	98%	99%
5	Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases	96%	98%	99%
6	Apriori algorithm, ID3 algorithm and C4.5 algorithm	ID3 and C4.5 are used to identify the various categories of students' performances		
7	Bayesian Classifiers algorithm, kNN algorithm, Rule Learners Classification Algorithm	The students based on the attributes selected reveals that the prediction rates are not uniform among the algorithms. The range of prediction varies from 61-75 %. Moreover, the classifiers perform differently for the five classes.		
8	cfsSubsetEval algorithm, GainRatio AttributeEval algorithm	Naïve Bayes classifiers have been applied on the selected features. From the results, it is concluded that Correlation Based Feature Subset evaluator performs well with the Naïve Bayes Classifier as compared with Gain-Ratio Attribute Evaluator.		
9	We want to experiment with huge student's usage information of different AHA! courses and other sequence mining algorithms as SPADE, FreeSpan, CloSpan, PSP	95.5%	97.5%	99%
10	Knowledge discovery with genetic programming for providing feedback to courseware authors.	91%	85%	95%
11	Data mining in course management systems: Moodle case study and tutorial.	-	90%	95%
12	Data mining: concepts and techniques.	Concepts of DM are elaborated in a broad method and discussed		
13	Using online assessment requires close cooperation of academic and technical units. First, preparing questions for online settings requires extra effort.	89%	-	94%
14	This study has raised a number of questions about how mode may have affected the performance of some children.	-	89%	93%
15	The chosen approach relies on the notion of competence, thus introducing an abstract perspective, in which courses do not directly depend on one another but rather have knowledge	-	89%	93%

	prerequisites and effects on the learner model.			
16	This use of model checking is inspired. LTL formulas are used to describe and verify the properties of a composition of Web Services.	98%	100%	97.5%
17	A rule schema allows user expectations representation and permits to the user to supervise association rule mining, meanwhile operators guide the post-processing task by pruning and filtering discovered rules.	88%	90%	95%
18	Experience worldwide has shown it to be generally popular with candidates and efficient for marking and delivery	90%	94%	99%
19	Data mining is applied in network course in order to solve different teaching. Then, introduces different teaching and data mining, analyzes network instructional platform of Modern Education Technology	79%	90%	80%
20	Learners in this category will exist in almost every class, yet at present a systematic way of identifying and supporting them does not exist.	-	83%	97%
21	An examination of computer anxiety related to achievement on paper-and-pencil and computer-based aircraft maintenance knowledge testing of united states air force technical training students	-	-	90%
22	Classroom Evaluation affects students in many different ways.	-	87%	93%
23	We are assuming a rather narrow view of machine learning—that is, supervised and unsupervised inductive learning from examples.	-	88%	95%

V. CONCLUSION

This survey work addresses the significance of LDM for finding secret association in learning information. A number of articles reviewed in this work to identify the students' performance by considering various benchmarks. Based on the observation, many DM algorithms are suitable for LDM. From the various researchers' viewpoints, it is identified that to enhance the nature of higher education, the syllabus setup is the key point. For that, the investigators recommend a new methods and patterns to improve the student's performance through LDM techniques. Moreover classification algorithms ID3 and C4.5 are used to identify the various categories of students' performances via their behavior and results in various types of examinations. For the chosen data concept by the researchers, the behavior and performance of both the algorithms differ. Among the choice of classification algorithms, the performance of C4.5 performs well in processing the student's data.

References

[1] Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method", World Journal of Computer Application and Technology, Vol. 2, Issue 2, 2014, pp. 43-47.

[2] Ajay Kumar Pal and Saurabh Pal, "Analysis and Mining of Educational Data for Predicting the Performance of Students", International Journal of Electronics Communication and Computer Engineering Volume 4, Issue 5, pp. 1560-1565.

[3] Albert Bandura and Forest J. Jourden, "Self-Regulatory Mechanisms Governing the Impact of Social Comparison on Complex Decision Making", Journal of Personality and Social Psychology, Vol. 60, issue. 6, 1991, pp. 941-951.

[4] Arnon Hershkovitz, Ryan S.J.d. Baker, Sujith M. Gowda and Albert T. Corbett, "Predicting Future Learning Better

Using Quantitative Analysis of Moment-by-Moment Learning", Educational Data Mining 2013.

[5] Brijesh Kumar Baradwaj and Saurabh Pal, "Mining Educational Data to Analyze Students Performance", International Journal of Advanced Computer Science and Applications, Vol. 2, issue. 6, 2011, pp. 63-69.

[6] C. Anuradha and T. Velmurugan, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance", Indian Journal of Science and Technology, Vol 8, issue15, July 2015, pp. 1-12.

[7] C. Anuradha and T. Velmurugan, "A Data Mining based Survey on Student Performance Evaluation System", IEEE International Conference on Computational Intelligence and Computing Research, 2014, pp. 452-455.

[8] C. Anuradha and T. Velmurugan, "Feature Selection Techniques To Analyse Student Academic Performance Using Naïve Bayes Classifier", The 3rd International Conference on Small & Medium Business, January 19 - 21, 2016, pp. 345-350.

[9] C. Romero Morales A. R. Porras Pérez S. Ventura Soto C. Hervás Martínez and A. Zafra Gómez, "Using sequential pattern mining for links recommendation in adaptive hypermedia educational systems", Current Developments in Technology-Assisted Education ,2006, pp. 1016-1020.

[10] Cristobal Romero, Sebastian Ventura and Paul De Bra, "Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors", Kluwer Academic Publishers, 2004, pp. 1-48.

[11] Cristobal Romero, Sebastian Ventura, Enrique Garcia, "Data mining in course management systems: Moodle case study and tutorial ", Computers & Education Vol. 51, 2008, pp. 368-384.

[12] Han J, Kamber M. Data Mining: Concepts and Techniques, New Delhi: Morgan Kaufmann Publishers; 2006. ISBN: 978-81-312-0535-8.

- [13] M. Yasar Özden, Ismail Ertürk, and Refik Sanli, "Students' Perceptions of Online Assessment: A Case Study", *Journal of Distance Education Revue De L-Education À Distance Spring*, 2004 Vol. 19, issue 2, pp. 77-92.
- [14] Martin Johnson and Sylvia Green, "On-line assessment: the impact of mode on student performance", 2004
- [15] Matteo Baldoni, Cristina Baroglio, Ingo Brunkhorst, Nicola Henze, Elisa Marengo and Viviana Patti, "Constraint Modeling for Curriculum Planning and Validation", *Interactive Learning Environments*, Vol.19.issue 1, 2011, pp. 1-36
- [16] Matteo Baldoni, Cristina Baroglio, and Elisa Marengo, "Curricula Modeling and Checking", *AI* IA, Artificial Intelligence and Human-Oriented Computing*. Springer Berlin Heidelberg, 2007, pp. 471-482.
- [17] P. Ajith, B. Tejaswi, M.S.S.Sai, "Rule Mining Framework for Students Performance Evaluation", *International Journal of Soft Computing and Engineering*, Vol. 2, Issue 6, 2013, pp. 203-206.
- [18] Peter Cantillon, Bill Irish, David Sales, "Learning in practice Using computers for assessment in medicine", *BMJ*, Vol. 329, 11 September 2004, pp. 606-609.
- [19] Qiuxiang Shi, Jianying Li and Liying Wang., "Application of Data Mining in Network Instructional platform of Modern Educational Technology for Different Teaching", *International Journal of Database Theory and Application* Vol.7, No.3, 2014, pp.143-158.
- [20] Rashmi Rekha Borah, "Slow Learners: Role of Teachers and Guardians in Honing their Hidden Skills", *International Journal of Educational Planning & Administration*, Vol. 3, issue. 2, 2013, pp. 139-144.
- [21] Richard B and McVay. B. A., "An examination of computer anxiety related to achievement on paper-and-pencil and computer-based aircraft maintenance knowledge testing of united states air force technical training students.", *Diss. University of North Texas*, 2002., pp. 1-65.
- [22] Terence J. Crooks, "The Impact of Classroom Evaluation Practices on Students", *Review of Educational Research* Winter, Vol. 58, issue. 4, 1988, pp. 438-481.
- [23] William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus, "Knowledge Discovery in Databases: An Overview", *AI Magazine*, Vol.13 issue. 3 1992, pp.57-70.