Clustering Algorithms Applied in Educational Data Mining

Ashish Dutt, Saeed Aghabozrgi, Maizatul Akmal Binti Ismail, and Hamidreza Mahroeian

Abstract—**Fifty years ago there were just a handful of universities across the globe that could provide for specialized educational courses. Today Universities are generating not only graduates but also massive amounts of data from their systems. So the question that arises is how can a higher educational institution harness the power of this didactic data for its strategic use? This review paper will serve to answer this question. To build an Information system that can learn from the data is a difficult task but it has been achieved successfully by using various data mining approaches like clustering, classification, prediction algorithms etc. However the use of these algorithms with educational dataset is quite low. This review paper focuses to consolidate the different types of clustering algorithms as applied in Educational Data Mining context.**

Index Terms—**Clustering, educational data mining (EDM), learning styles, learning management systems (LMS).**

I. Introduction

According to the international consortium on educational data mining, EDM is defined as "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 they learn in" [1].

EDM focuses on analyzing data generated in an educational setup by the various intra-connected or disparate systems to develop model for improving learning experience and institutional effectiveness. Data mining also sometimes referred to as knowledge discovery in databases (KDD) is a known field of study in life sciences and commerce but the application of data mining to educational context is limited [2].

Various methods have been proposed, applied and tested in data mining field and it"s argued by some researchers that these generic methods or algorithms are not suitable to be applied to this emerging field of study.

It"s proposed that educational data mining methods must be different from the standard data mining methods because of multi-level hierarchy and non-independence in educational data [1]. Institutions are increasing being held accountable for student success [3] Since EDM emerged as a sub-discipline in DM there have been notable researches in student retention and attrition rates that have been conducted [4]. [5] applied predictive modeling technique to enhance student retention efforts. In a similar fashion, there have been various software"s like Weka, Rapid Mineretc.that have been developed to use a combination of DM algorithms or a specific algorithm to aid researcher"s or stakeholders to find answers to specific problems but the problem with such tools are that they need to be learned so as to use them. This means that for a novice computer user especially in the administration department of a college or a university, the usage of such tools is not that easy. Just like commercial e-commerce based websites are using recommender systems that collect user browsing data and recommend similar products there have been efforts to apply the same in the educational context but they have not been successful as they are highly domain dependent [6].

The objective and purpose of this research paper is to review, different clustering algorithms as applied to EDM context. Numerous studies have been conducted in this context, but with disparate associations. This research paper is to bridge this gap and present a comprehensive review of all types of clustering methodologies as applied to EDM till date.

This paper is organized as follows. Section II is a background of related works pertaining to Educational Data Mining (EDM); Section III discusses the various clustering algorithms/techniques applied to educational dataset. Section IV discusses on the application of clustering algorithms to learning styles of studentand learning management systems. Section V provides further discussion and finally Section VI shows the conclusion and future works.

II. Educational Data Mining

"EDM converts raw data coming from educational systems into useful information that could potentially have a greater impact on educational research and practice" [7]. Traditionally researchers have applied data mining methods like clustering, classification, association rule mining, text mining to educational context as outlined; [8], conducted a survey that provides a comprehensive resource of papers published between 1995 and 2005 on Educational Data Mining (EDM). Reference[9]Has suggested the application of data mining techniques to study on-line courses. [10] Had suggested association rules and clustering to support collaborative filtering for the development of more sensitive and effective e-learning systems. Reference [11] hasused a case study that uses prediction methods in scientific study to game the interactive learning environment by exploiting the properties of the system rather than learning the system.

Manuscript received July 2, 2014; revised September 4, 2014.

Ashish Dutt, Saeed Aghabozrgi, and Maizatul Akmal Binti Ismailis are with the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia (e-mail: {ashish_dutt, saeed}@siswa.um.edu.my, maizatul@um.edu.my).

Hamidreza Mahroeianis with University of Otago, New Zealand (e-mail: Hamidreza.mahroeian@postgrad.otago.ac.nz).

DOI: 10.7763/IJIEE.2015.V5.513

Reference [12] has provided tools that can be used to support educational data mining. [13] Had shown how educational data mining prediction methods can be used to develop student models. It must be noted that student modeling is an emerging research discipline in educational data mining [1]. While another group of researchers [14] have devised a toolkit that operates within the course management systems and is able to provide extracted mined information to non-expert users. Data mining techniques have been used to create dynamic learning exercises based on student"s progress through a course on English language instruction [15]. While most of the e-learning systems used by educational institutions are used to post or access course materials, they do not provide the educators the necessary tools that could thoroughly track and evaluate all the activities performed by their learners so as to evaluate the effectiveness of the course and learning process. [16].

III. Clustering Techniques

The theory of looking at didactic amounts of data whether it"s in digital or physical form and stored in diverse repositories be it book keeping records or databases of an educational institution is now termed as Big data [17]. According, to Manyika *et al.* [18] a data set whose computational size exceeds the processing limit of software can be categorized as big data. Several studies have been conducted in the past that have provided detailed insights into the application of traditional data mining algorithms like clustering, prediction, association to tame the sheer voluminous power of big data [9]. Traditional Data Mining algorithms have been applied to various kinds of educational systems as shown in Table I. Broadly, the educational system can be classified as two types, brick and mortar based traditional classroom's and the digital virtual classroom's better known as known as LMS Systems [19], web-based adaptive hypermedia systems [20] and intelligent tutoring systems (ITS) [21]. The application of various clustering algorithm has been applied in many a cases to educational data set in diverse studies. The following table consolidates the research work done on the application of clustering algorithms to educational dataset.

IV. Using Clusteringin EDM

In a learning environment the learning styles of student is a decisive factor. In many cases there has been a mismatch between personal learning styles and the learning demands of different disciplines. Reference [22], has utilized a two-step cluster analysis approach which examined the brain signals centroids that used electroencephalography (EEG) technology to measure the learning style of participants such that they were successfully able to classify it into 4 unique clusters. Students typically annotate texts while reading book by highlighting the context of interest or by underlining it or by writing comments in the side margins. This activity is called annotation. Researchers have [19] applied statistical clustering method like K-means clustering and Hierarchical clustering to student annotations. And they proved that by using these clustering methods, the creation of students with similar learning style cluster is improved and is faster. Comprehension reading is a very widely used classroom activity in schools and colleges. This helps in building a lifelong reading habit and learning process. This ability of the student behavioral learning patterns has been computationally mapped by applying the Forgy method for k-means clustering and combined with Bloom's taxonomy to determine positive and negative cognitive skills set in reference to reading comprehension skills. [20]. Yet in another study, [21] combined Web Based Instruction (WBI) programs with the cognitive learning style of the learner to study their effects on student learning patterns. K-means clustering algorithm was used to result in cluster of students that shared similar learning patterns that further leads to identification of the related cognitive style for each group.

Learning Management System (LMS) have become an integral part of educational institutions for teaching and learning. A typical LMS logs most of the user activities like course attempted, modules read, practice exam attempted, exam score, student-student interaction via chat logs or discussion boards similarly student-teacher interaction via discussion boards is also logged in the LMS. Several studies have been conducted in this regard.

Reference [35] studied the usage statistics that an LMS provides and worked on its statistical data analysis and the results were applied in the University of Valencia (Spain). Although they were successful in the statistical analysis of LMS usage data using SPSS but to standardize their methodology the subsequent automation process is yet to be completed and has been left as a future work. Performance in exams, usage statistics, regression, number of visits, top search terms, number of downloads of e-learning resources is presented in [24]. Several DM approaches and techniques (clustering, classification and association analysis) have been proposed for joint use in the mining of student's assessment data in LMS [25]. Association rules, clustering, classification, sequential pattern analysis, dependency modeling, and prediction have been used to improve web-based learning environments to subsequently enhance the degree to which the educator can evaluate the learning process [26]. Analysis of user access log in Moodle to improve e-learning and to support the analysis of trends is presented in [27] Comparison of different DM algorithms are made to classify learners (predict final marks) based on Moodle usage data [28]. Prediction of student's performance (final grade) based on features extracted from logged data is presented in [29] and university academic student performance is presented in [30]. Prediction of online students marks (using an orthogonal search-based rule extraction algorithm) is presented in [31].

Other studies have been conducted to predict student's performance from log and test scores in web-based instruction (using multi-variable regression), [32]. While [33] have used classification, clustering, association rule mining and regression for the discovery of possible dependencies among learner's mean performance and course characteristics. Their results confirm that students behaviour in an online learning platform affect their performance.

TABLE I: Clustering Algorithm Used in EDM

	Problem/ Objective	Algorithm/Method	Dataset/Data source
[34]	Evaluating undergraduate student academic performance	Using a combination of DM methods like ANN (Artificial Neural Network), Farthest First method based on k-means clustering and Decision Tree as a classification approach	Student data of the Computer Science department at Faculty of Science and Defence technology, National Defence university of Malaysia (NUDM)
[20]	To predict the potentiality of students performance who can fail during an online curriculum in a Learning Management System (LMS)	Expectation Maximization, Hierarchical Clustering, Simple k-Means and X-Means as provided in WEKA software has been used.	Real life dataset provided by Juriscampus accessible at http://www.juriscampus.fr/
[36]	Shows the applications of various DM techniques on student academic data.	Apriori Algorithm is applied to academic records of students to obtain the best association rules which help in student profiling-K-means clustering is used to group students categorically.	Student academic record file, no mention of where its obtained from.
[37]	To identify the significant variables that affects and influences the performance of undergraduate students	C-Means clustering method	Academic dataset from the state University of Santander (IUS). Contains basic student data like faculty, academic program, gender, age, origin, student category and academic achievements data.
[38]	To develop student profiles of learner behaviour from learner″s activity in an online learning environment and also to create click-stream server data	Two clustering methods used, Hierarchical clustering (Ward″s clustering) and Non-Hierarchical Clustering method (k-means clustering)	Information not provided
[33]	To analyze the web log data files of a Leaning Management System (LMS)	Markov Clustering (MCL) algorithm for clustering the students″ activity and a SimpleKMeans algorithm for clustering the courses	The dataset was collected from the Technological Education institute (TEI) of Kavala that uses the Open e-Class e-learning platform (GUNet, 2009). The data are from the spring semester of 2009 from the Department of Information management and involve 1199 students and 39 different courses. The data are in ASCII form and are obtained from the Apache server log file.
[39]	To predict student″s behaviour in future	UCAM (Unique Clustering with Affinity Measure) algorithm is for clustering which works without giving initial seed and number of clusters.	No information provided by author
[40]	Deals with clustering of student access patterns or surfing behaviour in an e-learning environment.	Fuzzy sets and Transitive closure	No real life data set used. Paper is based on hypothetical data
[41]	High dimensional categorical dataset are difficult to cluster therefore a fully automated algorithm has been proposed.	Introduced a Two Phase Clustering (TPC) algorithm that works as follows; First start with an initial partition that contains a single cluster (i.e. the whole dataset) and then continuously try to split the cluster within the partition into two sub-clusters. Now check the homogeneity of the two sub-clusters. If the homogeneity of the two sub-clutters is high then discard the initial cluster and add the two sub-clusters to the partition.	Author has mentioned that real life dataset has been used but from where that has been omitted!
[42]	How to teach a basic computer skills course to students from rural or urban backgrounds.	K-means Clustering method has been used	The dataset was collected in the form of a survey that was designed and then distributed to the students of a course in Masters in Computer Applications. The authenticity of the survey answer is based on the premise that student will answer the survey questions honestly.
[43]	To compare the emotional intelligence of students.	K-means clustering is applied on questionnaire data	823 students randomly selected from the Chongqing Electronic Engineering Vocational College, Chongqing Engineering Institute, Chongqing Industry Polytechnic College, Chongqing community Vocational College, Chongqing Real Estate Vocational College, etc. 5 Vocational College. 804 valid questionnaires scales are collected, the effective rate is 97.7%./K-means clustering algorithm of Dynamic Analysis is used.
[44]	To map out the approaches to teaching profiles of teachers in higher education on the basis of their scores on the ATI abbreviated for Approaches to Teaching Inventory (ATI)	A hierarchical cluster analysis was performed on the questionnaire data	Questionnaires and Interviews conducted in University X with 30 academics and response rate of 20%

In another study, researchers have shown how educational institutions can benefit from the data collected by LMS. They have proposed an algorithm called “Course Classification Algorithm”[45] when applied in the LMS (Open e-Class platform) that the institution uses can be used to determine and generate course content quality and student online usage reports. These reports are then sent to the instructors for evaluation and motivation purpose. [46] have proposed the usage of k-means clustering and self-organizing map to cluster learning objects (learning objects are educational resources like eBook, question paper, answer index etc.) so as to facilitate faster accessibility of such resources by searching in a LMS. [47] have proposed Particle Swarm Optimization (PSO)-based clustering for improving the quality of learning by integrating Personalized Learning Environment (PLE) in conjunction with the conventional Learning Management System (LMS)

However, one of the major problems that researchers encounter in finding interesting patterns from educational data set is the relatively small size of the dataset [48].

In another study [49]and [50] have applied Expectation-Maximization (EM) clustering algorithm to discover student profiles from course evaluation data and for finding associations between subjects that was based on student performance.

Employability of its graduates has been a primary goal of higher educational institutions. Knowledge workers are resorting to key educational courses using Massive Open Online Courses (MOOCs) being provided online by institutions of repute like MIT, Stanford, Harvard to name a few. The year 2012 was witness to a rapid development and expansion of several MOOEPs (Massive Open Online Education Platform) like Canvas, ClassToGo, Coursera, edX, NPTEL, Udacity to name a few. [51] had conducted a study to explore the scope of interdisciplinary education through MOOCs. Employability education is an integral component of higher education and an important path by which companies obtain excellent employees. It has been a sustainable argument that in the present socio-economic development, the employability based educational content becomes a mandate [52].

V. Discussion

So far we see that subject specific research has been done but what about domain specific i.e. how do institutions employ or apply data mining methods to improve on institutional effectiveness? Zimmerman‟s educational model states that maintaining and monitoring student‟s academic record is an integral activity of an educational institution. [53] Had used classification algorithm and Prediction algorithm namely decision table and One R algorithm on students‟ academic record from a previous semester to predict their performance in the current semester. An educational institution maintains and stores various types of student data, it can range from student academic data to their personal record like parents income, parent‟s qualification etc. In a study conducted by [54] they have proved that students performance can be predicted by using a data set that consisted of students gender, its parental education, its financial background etc. [55] have used Bayesian networks to predict the student outcome based on attributes like attendance, performance in class tests, assignments etc. Researchers have applied data mining methods like dimensional modeling into educational institutions be it a data warehousing solutions as applied in the department of Informatics of Aristotle University of Thessaloniki [56], storyboarding and decision trees [57] while others like [58] have used regression analysis and classification (CS5.0 algorithm which is a type of decision tree) to predict the academic dismissal of students and to predict the GPA of graduated students in e-learning center.

VI. Conclusion and Future Work

The application of data mining methods in the educational sector is an interesting phenomenon. It sets to uncover the previously hidden data to meaningful information that could be used for both strategic as well as learning gains. In this review paper, we have detailed the various disparate entities that are widely spread across in the educational foray. However, collectively they have not been addressed and this paper serves to bridge this gap.

We would continue to pursue our research in clustering algorithms as applied to educational context and will also be working towards generating a unified clustering approach such that it could easily be applied to any educational institutional dataset without any much overhead.

Acknowledgment

This work is supported by University of Malaya High Impact Research Grant no vote UM.C/625/HIR/MOHE/SC/13/2 from Ministry of Education Malaysia.

References

[1] R. S. J. D. Baker and K. Yacef, “The state of educational data mining in 2009: A review and future visions,” *Journal of Educational Data Mining*, vol. 1, no. 1, 2009.

[2] J. Ranjan and K. Malik, “Effective educational process: A data-mining approach,” *Vine*, vol. 37, no. 4, pp. 502-515, 2007.

[3] B. J. P. Campbell, P. B. Deblois, and D. G. Oblinger, “Academic analytics: A new tool for a new era,” *Educause Review*, vol. 42, pp. 40-57, 2007.

[4] J. Luan, *Data Mining and Knowledge Management in Higher Education*, Toronto, Canada, 2002.

[5] S. Lin, “Data mining for student retention management,” *J. Comput. Sci. Coll.*, vol. 27, no. 4, pp. 92-99, 2012.

[6] O. C. Santos and J. G. Boticario, “Modeling recommendations for the educational domain,” *Procedia Comput. Sci.*, vol. 1, no. 2, pp. 2793-2800, Jan. 2010.

[7] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Transactions on Systems Man and Cybernetics Part C*, vol. 40, no. 6, pp. 601-618, 2010.

[8] C. Romero, S. Ventura, J. A. Delgado, and P. De Bra, “Personalized links recommendation based on data mining in adaptive educational hypermedia systems,” *Ectel 2007 Lcns 4753*, vol. 4753, pp. 292-306, 2007.

[9] O. Zaıane and J. Luo, “Web usage mining for a better web-based learning environment,” presented at Conf. Adv. Technol., 2001.

[10] O. Zaíne, “Building a recommender agent for e-learning systems,” in *Proc. International Conference on Computers in Education*, 2002, pp. 55-59.

[11] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner, “Off-task behavior in the cognitive tutor classroom: When students „game the system‟,” 2004.

[12] Merceron and Yacef, “A web-based tutoring tool with mining facilities to improve learning and teaching,” presented at the 11th International Conference on Artificial Intelligence in Education, 2003.

[13] J. Beck and B. Woolf, “High-level student modeling with machine learning,” *Intell. Tutoring Syst.*, pp. 584-593, 2000.

[14] E. García, C. Romero, S. Ventura, and C. de Castro, “A collaborative educational association rule mining tool,” *Internet High. Educ.*, vol. 14, no. 2, pp. 77-88, Mar. 2011.

[15] Y. Wang and H.-C. Liao, “Data mining for adaptive learning in a TESL-based e-learning system,” *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6480-6485, Jun. 2011.

[16] M. Zorrilla, E. Menasalvas, and D. Marin, “Web usage mining project for improving web-based learning sites,” in *Proc. 10th International Conference on Computer Aided Systems Theory*, 2005, pp. 205-210.

[17] S. Sagiroglu and D. Sinanc, “Big data: A review,” in *Proc. 2013 International Conference on Collaboration Technologies and Systems (CTS)*, May 2013, pp. 42-47.

[18] J. Manyika, M. Chui, B. Brown, and J. Bughin, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, May, 2011.

[19] M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, and H. Y. Seong, “Predicting NDUM student‟s academic performance using data mining techniques,” in *Proc. 2009 Second Int. Conf. Comput. Electr. Eng.*, 2009, pp. 357-361.

[20] A. Bovo, S. Sanchez, O. Heguy, and Y. Duthen, “Clustering moodle data as a tool for profiling students,” in *Proc. 2013 Second Int. Conf. E-Learning E-Technologies Educ.*, Sep. 2013, pp. 121-126.

[21] S. Parack, Z. Zahid, and F. Merchant, “Application of data mining in educational databases for predicting academic trends and patterns,” in *Proc. 2012 IEEE Int. Conf. Technol. Enhanc. Educ. (Ictee 2012)*, 2012, pp. 1-4.

[22] A. Salazar, J. Gosalbez, I. Bosch, R. Miralles, and L. Vergara, “A case study of knowledge discovery on academic achievement, student desertion and student retention,” in *Proc. ITRE 2004 2nd Int. Conf. Inf. Technol. Res. Educ.*, 2004, pp. 150-154.

[23] P. D. Antonenko, S. Toy, and D. S. Niederhauser, “Using cluster analysis for data mining in educational technology research,” *Etr&D-Educational Technol. Res. Dev.*, vol. 60, no. 3, pp. 383-398, Feb. 2012.

[24] S. Valsamidis, S. Kontogiannis, I. Kazanidis, T. Theodosiou, and A. Karakos, “A clustering methodology of web log data for learning management systems,” *Educ. Technol. Soc.*, vol. 15, no. 2, pp. 154-167, 2012.

[25] 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.

[26] 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.

[27] 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, pp. 191-195.

[28] 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.

[29] F. Jing and K. Shiying, “Application of data mining for emotional intelligence based on cluster analysis,” in *Proc. 2010 Int. Conf. Artif. Intell. Educ.*, Oct. 2010, pp. 512-515.

[30] 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.

[31] 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.

[32] K. Ying, M. Chang, A. F. Chiarella, and J.-S. Heh, “Clustering students based on their annotations of a digital text,” in *Proc. 2012 IEEE Fourth Int. Conf. Technol. Educ.*, Jul. 2012, pp. 20-25.

[33] T. Peckham and G. McCalla, “Mining student behavior patterns in reading comprehension tasks,” *Int. Educ. Data Min. Soc.*, pp. 87-94, 2012.

[34] S. Chen and X. Liu, “An integrated approach for modeling learning patterns of students in web-based instruction: A cognitive style perspective,” *ACM Trans. Comput. Interact.*, vol. 15, no. 1, 2008.

[35] P. Moreno-Clari, M. Arevalillo-Herraez, and V. Cerveron-Lleo, “Data analysis as a tool for optimizing learning management systems,” in *Proc. Ninth IEEE Int. Conf. Adv. Learn. Technol.*, Jul. 2009, pp. 242-246.

[36] H. Grob, F. Bensberg, and F. Kaderali, “Controlling open source intermediaries-a web log mining approach,” *IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews*, vol. 1, pp. 233-242, 2004,

[37] M. Pechenizkiy, T. Calders, E. Vasilyeva, and P. De Bra, “Mining the student assessment data: Lessons drawn from a small scale case study,” *EDM*, 2008.

[38] Y. Psaromiligkos, M. Orfanidou, C. Kytagias, and E. Zafiri, “Mining log data for the analysis of learners‟ behaviour in web-based learning management systems,” *Oper. Res.*, vol. 11, no. 2, pp. 187-200, Jan. 2009.

[39] F. Getúlio, R. De Janeiro, F. G. V Online, M. A. Amaral, U. Universidade, P. Ffalm, and B. R. Km, “Analysing users „access logs in moodle to improve e learning analisando logs de acessos dos usuários do moodle para melhorar e-learning cássia blondet baruque alexandre barcellos joão carlos da silva freitas carlos juliano longo,” in *Proc. 2007 Euro Am. Conf. Telemat. Inf. Syst.*, 2007, pp. 1-4.

[40] C. Romero, S. Ventura, and E. García, “Data mining in course management systems: Moodle case study and tutorial,” *Comput. Educ.*, vol. 51, no. 1, pp. 368-384, Aug. 2008.

[41] B. M. I, D. A. Kashy, G. Kortemeyer, and W. F. Punch, “T2a predicting student performance,” 2003.

[42] D. Ibrahim and Zaidah, Rusli, “Predicting students‟ academic performance: comparing artificial neural network, decision tree and linear regression,” in *Proc. the 21st Annual SAS Malaysia Forum*, 2007, pp. 1-6.

[43] T. Etchells, À. Nebot, A. Vellido, P. Lisboa, and F. Mugica, “Learning what is important: Feature selection and rule extraction in a virtual course,” *ESANN*, pp. 26-28, 2006.

[44] P. Golding and O. Donaldson, “Predicting academic performance,” in *Proc. Front. Educ. 36th Annu. Conf.*, 2006, pp. 21-26.

[45] S. Valsamidis, S. Kontogiannis, A. Karakos, and I. Kazanidis, “Homogeneity and enrichment: Two metrics for web applications assessment,” presented at 2010 14th Panhellenic Conf. Informatics, Sep. 2010.

[46] A. S. Sabitha and D. Mehrotra, “User centric retrieval of learning objects in LMS,” in *Proc. 2012 Third Int. Conf. Comput. Commun. Technol.*, Nov. 2012, pp. 14-19.

[47] K. Govindarajan, T. S. Somasundaram, and V. S. Kumar, “Particle swarm optimization (pso)-based clustering for improving the quality of learning using cloud computing,” in *Proc. 2013 IEEE 13th Int. Conf. Adv. Learn. Technol.*, Jul. 2013, pp. 495-497.

[48] N. Jyothi, K. Bhan, U. Mothukuri, S. Jain, and D. Jain, “A recommender system assisting instructor in building learning path for personalized learning system,” in *Proc. 2012 IEEE Fourth Int. Conf. on Technol. Educ. (T4E)*, 2012, pp. 228-230.

[49] E. Trandafili, A. Allkoçi, E. Kajo, and A. Xhuvani, “Discovery and evaluation of student‟s profiles with machine learning,” in *Proc. Fifth Balk. Conf. Informatics - BCI „12*, 2012, pp. 174.

[50] A. Bogarín, C. Romero, R. Cerezo, and M. Sánchez-Santillán, “Clustering for improving educational process mining,” in *Proc. Fourth Int. Conf. Learn. Anal. Knowl. - LAK „14*, 2014, pp. 11-15.

[51] V. Subbian, “Role of MOOCs in integrated STEM education: A learning perspective,” in *Proc. 2013 IEEE Integr. STEM Educ. Conf.*, Mar. 2013, pp. 1-4.

[52] X. Zhang, S. Zhang, and X. Zou, “A Study of the Employability-Based Talent Cultivating Pattern,” in *Proc. 2010 Int. Conf. E-bus. E-Government*, May 2010, pp. 892-895.

[53] S. A. Kumar and M. N. Vijayalakshmi, “Mining of student academic evaluation records in higher education,” in *Proc. 2012 Int. Conf. Recent Adv. Comput. Softw. Syst.*, Apr. 2012, pp. 67-70.

[54] M. Quadri and D. Kalyankar, “Drop out feature of student data for academic performance using decision tree techniques,” *Glob. J. Comput.*, vol. 10, no. 2, pp. 2-5, 2010.

[55] B. K. Baradwaj, “Mining educational data to analyze students‟ performance,” *International Journal of Adcanced Computer Science and Application*, vol. 2, no. 6, pp. 63-69, 2011.

[56] N. Dimokas, N. Mittas, A. Nanopoulos, and L. Angelis, “A prototype system for educational data warehousing and mining,” in *Proc. 2008 Panhellenic Conf. Informatics*, Aug. 2008, pp. 199-203.

[57] R. Knauf, Y. Sakurai, K. Takada, and S. Tsuruta, “A case study on using personalized data mining for university curricula,” in *Proc. 2012 IEEE Int. Conf. Syst. Man, Cybern.*, Oct. 2012, pp. 3051-3056.

[58] M. Nasiri, B. Minaei, and F. Vafaei, “Predicting GPA and academic dismissal in LMS using educational data mining: A case mining,” in *Proc. 2012 Third International Conference on e-Learning and e-Teaching (ICELET)*, 2012, pp. 53-58.

Ashish Dutt has been graduated in December 2012 with M.Sc. in Computing from Staffordshire University, Malaysia Campus. This paper was part of his PhD work. He has a rich work experience of over 10+ years in education sector. At present he is working for his doctoral studies in educational data clustering.