



Knowledge Discovery in Academia: A Survey on Related Literature

Mudasir Ashraf

Research Scholar, Department of Computer Science
University of Kashmir, J&K, India

Dr. Muheet Ahmed

Scientist D, Department of Computer Science
University of Kashmir, J&K, India

Dr. Majid Zaman

Scientist D, Directorate of IT&SS
University of Kashmir, J&K, India

S. Jahangeer Sidiq

Research Scholar, Department of Computer Science
University of Kashmir, J&K, India

Abstract: Educational Data Mining (EDM) is an emerging, promising and interdisciplinary research area, which has an immense impact on student's academic performance by analyzing educational data for improving teaching, learning experiences and, of course improving the institutional effectiveness. One of the most key facts in educational data is the significant growth of data and these statistics are rising swiftly without any benefit to the institution. In this short study, different studies conducted on the theme "EDM and interdisciplinary" were consulted and examined. Most of the authors have conducted studies on different systems which mostly include e-learning systems, student performance recommender systems (SPRS), conventional educational systems, adaptive and intelligent web based educational systems.

The study of the past research has revealed that immense progress has been done in various domains of education and in different phases as well, but there still exists reasonable gap which needs to be examined in details as a future research study.

Keywords: Educational Data Mining; Neural Networks; Classification; Association rules; Clustering; Prediction; k-means.

I. INTRODUCTION

Data Mining (DM) is a practice that searches for the precious nuggets from the raw data [1]. K. B. Brijesh and P. Saurabh (2011) described data mining as a technique that allows the user to analyse, categorise, and summarize the data which are recognised throughout the entire process of mining [2].

Data Mining can additionally be explicated as the process of extracting comprehensible, credible and novel information from substantial amount of data [3]. From the review of the past studies, it has been observed that data mining has attained central importance in the cognate and interdisciplinary fields of business, scientific, education and other government and private sectors to make a proximate examination of all elements of data like census data, airline records, and supermarket data-set that produces market research reports, and valuable information [1].

Educational Data Mining (EDM) is an application of data mining techniques that aims to evaluate the educational data issues by availing subsisting techniques in data mining [4]. EDM is supposed to be approaching its maturity height as a number of journals, conferences and specific tools for data mining in EDM are growing at a reasonable pace. So, it can be said that EDM is no longer an emerging field. But, at the same time has not yet reached to its mellowness plane [5].

The past several decades witnessed a brisk growth in both the areas of software and databases pertaining to student's information reflecting how they learn [6], which has proved to be a gold mine in educational research area [7]. The use of World Wide Web (www) in educational domain has produced a new context known as e-learning in which tremendous amount of teaching and learning behaviours can be generated and analysed [8]. There has been numerous surveys carried out by different researchers in the area of Educational data mining

and among these the most prominent, valid and pertinent is of C. Romero and S. Ventura (2005), who fixated on application of data mining in the areas: Web-based courses; Traditional Educational Systems; well-known learning content management systems; and adaptive & intelligent web-based education system. In all of these systems, the data can be collected from heterogeneous sources with the notion of discovering knowledge. Various data mining techniques like: classification, clustering and outlier detection; statistics and visualization; association rule mining and pattern mining and text mining can be applied after pre-processing the data in each case [9].

There has been a great contribution in the form of journals, international congress, workshops and few books (Romero & Ventura, 2006) that demonstrates the eminence of data mining in the field of education. The researchers (Romero & Ventura, 2006) put forward their view and highlighted some of the areas that should be taken into consideration for the development, improvement and enlargement in Educational data mining: Ease of mining of tools, Generalization of tools & techniques that can be applied to any educational system, integration of data mining tasks into a single application (E-learning), data mining techniques specific to EDM [9].

Romero and Ventura (2010), classified educational tasks and the various techniques associated with EDM. The literature is an improved, updated and much more elaborated than the previous one (Romero & Ventura, 2005) and has visualised the frequency of each research area published till 2009 and consequently, revealed most of the research lines that have touched the heights: Analysis & Visualization; providing feedback, recommendation systems, predicting performance, student modelling, detecting behaviours, grouping students. Besides, the total references being used till date (2009) are associated with the type of data like: Traditional Education (36

references), Web-predicated inculcation/e-learning (54 references), intelligent tutors (29 references) etc [5]. It illustrates that the number of references with the highest score resulted in E-learning/Web-predicated Education. The former review [9] classified data mining techniques used according to various authors.

Educational data has certain special characteristics and difficulties which need to be resolved through various mining techniques and there are number of specific techniques which are applied while as others are not [5]. Using data mining, different users or candidates observe educational information through different approaches according to their own perception, aim etc [10]. The teaching-learning methods can be improved by discovering appropriate facts and knowledge form EDM algorithms [11] and thus, there has been an exponential augmentation in the number of publications that evolved from time-to-time in the advancement of EDM [5].

There are various high-quality and proficient research studies conducted by prominent authors in the field of educational data mining which has been valuable and in the area of improving student-learning outcomes and student success. Among them, a research team (Gracia, Romera, Ventura, & de Castro, 2011) developed a simple toolkit for data mining so that non-expert users can fetch information for their courses with ease. This research is very significant because of the fact, that this effortless, toolkit does not require deep knowledge of various data mining tools, visualization & statistics & machine learning algorithms [5].

The other research contributions in the direction of EDM includes that of Parkavi, Ramesh and Yasoda (2011), examined the accuracy of 5 data mining techniques, with the intend of finding most optimal technique that could be employed on a specified dataset, and to envisage the probability of student's placement based on final semester [12]. Mohammad M. Abu Tair and AlaaM.El-Hakees (2012), collected data of fifteen years [1993-2007] from College of Science and Technology, and applied educational mining techniques with the foundation of discovering useful information from institutional domain [13]. The education data is considered to be a vital element in escalation and advancement of any nation [14]. Therefore, if data mining techniques are applied at educational route, it can lend a hand in improving student's performance, course selection, and retention rate of an organization [15] and, consequently reduce the dropout likelihood [16]. Furthermore, in 2013, various researchers aimed at enhancing the quality of education and student's performance by applying various data mining techniques at different levels of hierarchies within the educational system (Graduation, Post-Graduation) [17], [18], [19]. Recently, Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby, 2014, proposed a study, in order to identify weak students so as to decrease the failing ratio and taking suitable action at the precise time [20]. The students can employ certain assisting tools that will facilitate the students to take best decisions based on the courses to enrol [21].

The objective of this research paper is to endow with a survey on educational data mining, and present the research gaps and further breaks in contemporary literature. The study also endeavours to highlight few categorical and generic applications of educational data mining, and subsequently

provides general view on the research studies conducted in the field and the areas to follow in line of investigation. This paper aims at providing comprehensive background of recent as well as past related literature, which provides an idea of which data mining techniques are useful for specific data in different research scenarios and how mining can be constructive in diverse educational systems.

II. EDM AND VARIOUS SCIENTIFIC PROCEDURES EMPLOYED

Several techniques have been useful in the area of EDM such as: classification, regression, clustering and association rule mining. The highlighted techniques and their deployment has been conversed in the following subsection that categorises among various data mining tasks and the most frequently used data mining techniques/methods that have emerged from the review of prior revision are decision trees and neural networks.

A. Statistical Analysis

Statistics is a science of mathematical background, which is concerned in the areas akin-to compilation, analysis, descriptive and data presentation [22]. In order to get the basic description and statistics of educational data, it is comparatively effortless to apply software well-known as SPSS (Statistical Package for the Social Sciences). Educational data in software generates descriptive analysis reports or summaries the learning behaviour of the student while taking the various data characteristics into consideration [23]. The instructors find records in databases excessively cumbersome to examine [5]. So, statistical analysis of educational databases can notify the most important factors such as: frequency of a particular pattern, the most pages surfed, where the students start and terminate [24]. It is useful to get statements or details assessing, how much period a student has worked in that course of time, what is the percentage of problems being resolved. In addition, it helps to predict performance levels and score attained [25].

The main objective of analysis is to emphasize on useful information and the most widely used techniques, providing assistance for administration or stakeholders to obtain a general view of student's learning for decision making. Through various figures or facts, one can come to know what the success percentage is, what parameters need to be considered and what the correlation of different attributes is. Based on these parameters, we can come to a decision whether this area is significant or vice versa.

B. Visualization

To better understand and analyse data, information, visualization can be relatively useful and be of great assistance [26]. The visualization of education data, which is gigantic in size existing in the form of database, warehouse etc can be comprehended and recognized in the graphical representation with much more simplicity and in an interesting manner. Several studies have been dedicated towards educational data visualization such as: Annual patterns, months, daily, weekly and hourly behaviour of users on online discussions [27]. Also, statistical visualization/representation of assignment, exam score, other curriculum activities etc has been represented. [28].

Statistic and Visualization are the two techniques that are associated to each other. Once a user applies a statistical technique, there is always high probability of being followed by visualization practice.

It represents the data in an interesting graphical manner, making it easier for a client to analyze, comprehend and visualize data.

C. Student Forecasting

Mostly the variables that are predicted in educational data are marks or score and results etc. These values can be both categorical/numerical. Based on this result whether discrete/continuous, two techniques are supposed to be applied; if the response is continuous, then one of the most frequently-used and finest options is regression-analysis. Else, classification can be used for discrete values. Regression is a statistical technique which is most frequently used for continuous/numeric prediction. While as classification is the method of searching a model for whose class tag is unfamiliar and model can be used for the purpose of explaining and differentiating data classes[1]. The model being derived is based on the training set whose class labels for the data objects are known.

Several techniques and models have been applied in educational data such as: rule based systems, neural networks, regression and co-relation, Bayesian networks etc. In order to forecast the decisive grades of students, diverse kinds of neural network models have been applied such as feed-forward and back-propagation neural networks [29]. We can also forecast the performance of learners from aggregate marks attained in a test, using counter propagation and back propagation [30]. Furthermore, by exploiting multilayer perception topology, the likelihood of a student being admitted in a university can be predicted as well [31].

There are lots of regression techniques used for the forecasting of students marks in an open university including model trees, linear regression, locally weighed linear regression, neural networks and support vector machines [32], Multi-variable regression model to prophesy the student's performance in web-predicated systems [33], Stepwise linear regression to forecast the performance in academic pursuit [4], regression analysis and decision tree for examining and predicting outcomes in distance courses using linear regression [34].

Similarly, Bayesian networks can be used to predict the candidate's performance [35] and rule based systems can be applied in e-learning for a student performance through fuzzy association rules [36] and co-relation analysis have been employed to predict performance in on-line classes [37].

The aim of the prediction method is to determine the unknown values of a variable prior to its outcome. Predicting student performance can be valuable in circumstances like: enrolment, identifying intellectual and feeble students, students worthy of fellowships, course selection and grades etc. In such cases, the most frequent and common types of techniques used are: Neural Network and Decision Trees that have demonstrated reasonable results. Both these methods have few shortcomings in the form of speed and classification. In EDM, the neural

network has the finest capability of classification and prediction, but suffers from speed, while as decision tree being swift suffers from the drawback of classification [21].

D. Clustering students

In order to cluster/group the learners, broad collection of classification algorithms have been applied in past such as: neural networks, discriminant analysis, random forest and decision trees have also been employed in different universities for classifying students into varied risk categories of high, medium and low respectively [38]. This provides an opportunity to administration so as to identify the students at risk, and assists institution in responding to those students with constructive policy making. Furthermore, classification techniques can be applied to improve student's performance by extracting useful information from semester marks [17]. There are also several clustering algorithms for instance agglomerative clustering, K-means and model based clustering to discover bunches of students with parallel expertise [39]. With the aim of analysing learning behaviour of student, parameters in the vein of exam assessment like: mid, final and quiz tests are evaluated using k-means algorithm [18].

The objective of clustering different students is to group them according to certain common characteristics and features, and to implement a model that can classify colossal records at large. The stakeholders, administrators or decision makers can analyse and evaluate those clusters/groups of students for producing effective teaching-learning system. There are typically two types of data mining techniques used for grouping; supervised and unsupervised learning. In former technique, the training record set is provided with the class label while in later technique, the training set for each tuple is unknown [1].

E. User models

The user model procedure is yet another measure adopted in the area of educational data mining. The objective of user modelling is to build cognitive models of students/users, their ability and knowledge [5]. In order to mechanize the learning behaviour and user characteristics (motivation, performance, learning etc), various data mining techniques are applied for the development of student models [40]. Moreover, several data mining algorithms have been exerted for this particular task such as: Navie Bayes, Bayes Net, Support Vector Machines, Logistic Regression & Decision Trees have been used in intelligent tutoring systems [41]. The most generic data mining technique that has been exploited for this purpose is mainly Bayesian network. Also, techniques like supervised and unsupervised have been utilized to reduce the experimental cost in constructing user models [42]. To measure the online learning performance, clustering and classification has been used [43] and decision tree algorithm (id3) has been used to predict/forecast the dropout students in academia [44].

Modelling in particular has increased our prediction capability in terms of potential performance and student acquaintance. Models have forced researchers to analyse the various aspects that lead students to make learning decisions in specific areas. They have also increased our prediction accuracy (48%) by

consolidating models of guessing and slipping into student prediction performance [45].

III. TOOLS AND PERTINENCE OF EDUCATIONAL DATA MINING

The recent advancements in the area of statistical courses and in the capacity of computer science have made life more facile and competitive for researchers in the field of educational data mining. Data mining tools allow researchers to contemplate techniques and algorithms exploited by other researchers. There are several educational data mining tools that have emerged in past and recent years for instance: GISMO, LOCO, MINEL, O3R, TADAED, LISTEN Mining tool, EPrules etc. [46]. In addition to these tools, other open platform and free tools are also available such as: Weka, R, Rapid Miner, SNAPP and KEEL. These tools assist in preparation of data which is supported by algorithms that puts into practice the methods for carrying out statistical validation of models and visualizing data. An analytical tool known as Waikato Environment for Knowledge Analysis (Weka), being an open source and written in java is a wide-ranging mining tool in which number of tools have been incorporated like Rapid-Miner. This specific tool being admired in educational data mining is flexible for delivering out validation. Another popular open source tool for conducting statistical and advanced analytics is R, assists researchers in addressing R packaging in precise quarter [47]. There are few special purpose and commercial tools available such as developmental tools for model prediction [48], tools to exhibit student's behaviour and performance patterns for different fields [49]. Business/ industrial related tools such as IBM Cognos, SAS etc. can also be used which are better at examining, reporting, monitoring and highly developed statistical analytics. Another conspicuous Apache Hadoop, is an open source high speed data processing tool that processes the nodes carrying data in parallel, but foremost weakness in this specific tool is the complexity in setting the environment.

There are quite a lot of areas where data mining has been used: business, education, medical, meteorology and in scores of fields as well. However, this study would focus on educational data based on related literature and consequently, try to find out methods of teaching and learning processes which can lend a hand in improving the institutional achievements.

Research community has revealed that data mining can be used to recognize the students at menace and aid educational society in discovering and act in response to these students [50]. The researcher (Luan) applied two mining techniques-classification and regression, in order to predict types of students to drop out. To classify students according to success and academic performance, the research team branched students into three groups: low, high, and medium-risk students [38]. Students with high-risk were those students which were potentially weak, high likelihood failure and required high attention of staff and faculty so as to assist the high risk students in every possible way. Besides, there are quite a few influential factors or parameters such as residency, ethnicity etc. that determine as forward planners of student retention after applying various mining techniques in the form of neural networks, multivariate adaptive regression splines, and classification trees [51]. In an additional revision, researchers examined demographic settings of students based

on their performance [52]. In contrary to this study, another research group noticed that demographic features are insignificant predictors of student achievements [53]. These results explain that different research discoveries related to performance and retention of students under diverse demography, do not have dramatic impact. Thus, this field needs more analysis and centre of attention in order to come up with significant factors.

In a related study of student retention efforts, the investigators in the field found that certain techniques of machine learning's can be endowed with valuable prediction of students [54]. The study also generated a model that predicted short time precision of students and consequently tested what kind of students would benefit from such model on university grounds. One more team of investigators proposed a system for recovering and supporting retention based on data mining methods [55]. This specific system assists the institution to react, classify and retain risk prone students.

Another major application in educational environment is recommendation method. The specific system aims at making recommendations not only in educational domain, but is also widely used in business environment. Recommendations can be in different form such as jobs to be done, links to be surfed, courses to be adapted, learning tasks etc. and these recommendations can support the learning behaviour of students proficiently and successfully.

In an attempt to advance students end results, researchers applied recommender systems while predicting student performance using mining [56]. Based on student's web surfing activities and achievements, further recommendations for learning drills were proposed. To construct fresh individual recommendations explicitly for course management system, a data mining model was put together that inferred browsing proceedings with background aspects [57]. There are number of purposes that a recommender system can serve such as: recommendations for web-learning users for studying material [58], recommendations for individual learning's in e-learning systems [57], recommendations for pertinent debates [59], recommendations related to test scores [60], recommendations for course-authors to look up settled in courses [61] and recommendations for selecting most favourable optional courses [62]. Besides, in industrial domain, recommender system can provide each user with custom browsing knowledge and exhibit each user/customer with related items for consumption that a user/customer might procure. For instance, flipkart, snapdeal, olx, amazon etc are the various organizations which exercise recommender systems to assist its users to find items they will possibly feel affection for.

Other imperative applications in EDM are technical breakthroughs about teaching-learning behaviour, the principles and method of instruction made available by the learning program, and developing student models which enable beneficiaries to make privileged presumptions about students, for instance to comprehend about students personal rationalization, playing and making mistakes within the system [63]. Open learning course management system such as Modular Object Oriented Developed Learning Environment (MOODLE) is a valuable application of data mining for online tutors [64]. In this study conducted by (Cristobal Romero, Sebastian Ventura, Enrique Garcia) revealed that in order to

obtain interesting patterns in a faster and proficient way, we can also apply various mining techniques mutually, even though the researcher has shown these techniques independently.

IV. DISCUSSION

After going through the related literature of selective academia, researcher came to the conclusion that performance is based on number of aspects:

1. Previously outstanding academic record of a candidate/student gives sufficient credibility that a student is expected to perform well and is largely associated with dropout ratio. Therefore, it can be concluded that overall success and retention of a student is directly proportional to the good academic performance or inversely proportional to low risk coupled with a student having profound expertise in his/her subject and has been experimentally shown by many researchers [65]. Since past performance of a student is an indication of hard work, intelligence and dedication towards the educational domain, efforts should be made to develop a model for cognitive and other psychological attributes of a candidate. The deficiency in aptitude generation of a student that previous research studies have highlighted should be taken in cognizance and it broadens the scope for the state authorities to widen the restoration work in academia.
2. From the review of research studies conducted in quondam in the area of knowledge discovery in academia have brought in lime light some important aspects that can help in predicting the performance of a student. This mainly includes choosing a particular course for next phase of student's career which primarily is based on grade obtained in previous facet, and potency of a student as well. The attributes such as potential of a student and complexity of course can be tested by recommender systems that show a fare performance under real life conditions [66]. Therefore, before proceeding to next phase and to have inexorable success, attributes of varied candidates need to be properly scrutinized.
3. Scientists in the field of Data Mining need to put their greater efforts and strive to develop other techniques that could consequently check and validate outcome accuracies of various methodologies like Neural Networks, Decision Tree, and Regression etc adopted under diverse environments. In this regard, it becomes imperative to implement hybrid techniques and compare results obtained through different methodologies. At the same time it becomes significant to compare the training and testing time so that it gives enough credibility to accept the technique being highly reliable.
4. The authorities need to develop strategic intent and must familiarize the application of EDM in institutions and make the stake holders aware as to how productive EDM can be.

V. EXAMINATION OF SECONDARY DATA PERTAINING TO EDM

After going through the literature and secondary data available in the field of EDM, researcher came to the conclusion that there are various research lines and valid facts which can assist in exploring more collaborative and unified studies. Fig.1 (Data Types verses references) displays the number of references against the type of data. From the figure it is clearly visible that Web-based Education/E-learning is having the highest number of references. This signifies that most researchers have shown interest in this field. But, at the same time, there are other data types which need to be taken into cognizance, that otherwise receive lesser attention or lack of interest from researchers.

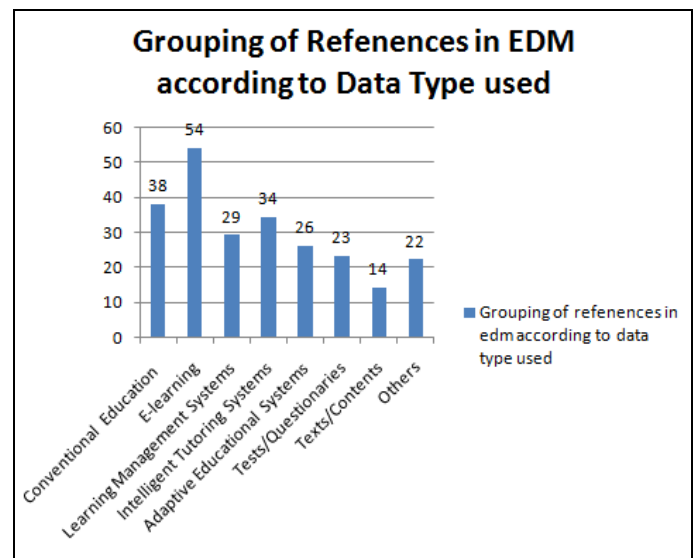


Figure 1

Furthermore, Fig.2 shows the pace of "Conferences and Workshops" and "Journal Papers and Book Chapters" over various years. This is quite evident from the figure that workshops, conferences always dominated over book chapters, Journal Papers, in terms of research work and papers. Subsequently, there has been a reduction in the number of papers being published in both cases over 2009 onwards.

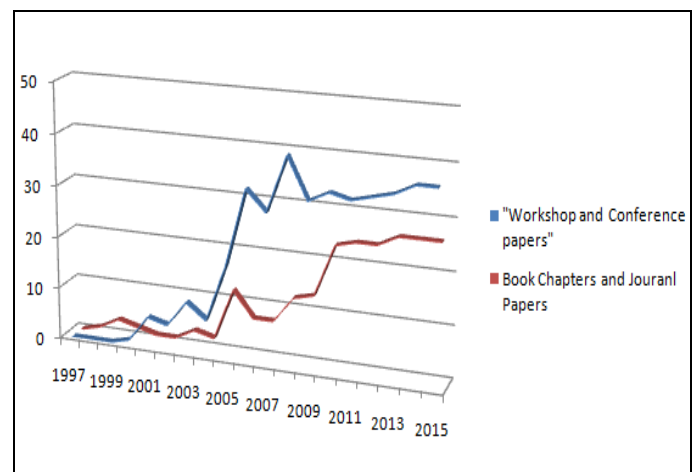


Figure 2

Making further advances in this direction, papers published in various categories of EDM was explored for determining the

trend existent in the specific field. The Fig. 3 displays the various categories of EDM that have been least studied. Therefore, these categories of EDM can be analysed and explored to improve the performance of both students and institutions as whole. These are unexplored fields in EDM, which needs to be given significance. In addition, it can open new gateways and methods for research, which can assist an organization in predicting better future of their students.

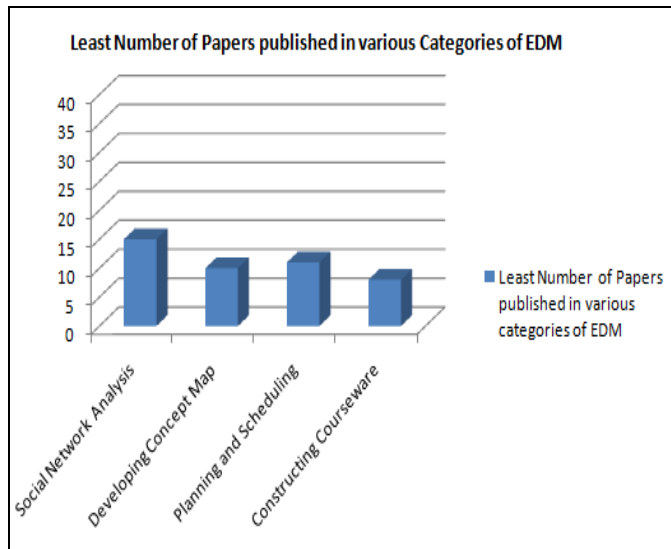


Figure 3

VI. CONCLUSION

In this review work, the core findings have been put forward which are associated with various studies related to Educational Data Mining and Academics. An attempt was made to conceptualize the study of Knowledge discovery in academia using various data mining approaches and educational data was classified according to various data mining methods associated with each area. It revealed how useful, constructive and explosive data mining can be in educational domain particularly to improve student's performance. This study also reviewed some of the contemporary tendencies of prediction in data mining as prognosticating is concerned to be one of the modern themes in educational data mining. Besides, number of thriving research studies have conducted in the domain of EDM, and considerable efforts are indispensable in the up gradation of traditional techniques and existing tools which no doubt have reduced the complication than before a decade. But there are still several issues like user friendly mining tools, specific pre-processing tools and education mining methods which needs to be given due consideration for the benefit of a common user.

VII. REFERENCES

- [1] Han, J., & Kamber, M. (2006). Data mining: Concepts and Techniques.
- [2] Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417.
- [3] Piatetsky-Shapiro, G. (1996). Advances in knowledge discovery and data mining (Vol. 21). U. M. Fayyad, P. Smyth, & R. Uthurusamy (Eds.). Menlo Park: AAAI press.
- [4] Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (2009). Educational Data Mining 2009: 2nd International Conference on Educational Data Mining. Proceedings. Cordoba, Spain.
- [5] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.
- [6] Baker, R. S. J. D. (2010). Data mining for education. International encyclopedia of education, 7, 112-118.
- [7] Mostow, J., & Beck, J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. Natural Language Engineering, 12(02), 195-208.
- [8] Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. In Evolution of teaching and learning paradigms in intelligent environment (pp. 183-221). Springer Berlin Heidelberg.
- [9] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.
- [10] Hanna, M. (2004). Data mining in the e-learning domain. Campus-wide information systems, 21(1), 29-34.
- [11] Merceron, A., & Yacef, K. (2005, May). Educational Data Mining: a Case Study. In AIED (pp. 467-474).
- [12] Ramesh, V., Parkavi, P., & Yasodha, P. (2011). Performance analysis of data mining techniques for placement chance prediction. International Journal of Scientific & Engineering Research, 2(8), 1.
- [13] Tair, M. M. A., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. International Journal of Information, 2(2).
- [14] Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2012). Data mining in education: data classification and decision tree approach. International Journal of e-Education, e-Business, e-Management and e-Learning, 2(2), 140.
- [15] Goyal, M., & Vohra, R. (2012). Applications of data mining in higher education. International journal of computer science, 9(2), 113.
- [16] Yadav, S. K., Bharadwaj, B., & Pal, S. (2012). Mining Education data to predict student's retention: a comparative study. arXiv preprint arXiv:1203.2987.
- [17] Priya, K. S., & Kumar, A. S. (2013). Improving the Student's Performance Using Educational Data Mining. International Journal of Advanced Networking and Applications, 4(4), 1806
- [18] Bhise, R. B., Thorat, S. S., & Supekar, A. K. (2013). Importance of data mining in higher education system. IOSR Journal Of Humanities And Social Science (IOSR-JHSS), 6(6), 18.
- [19] Kumar, V., & Chadha, A. (2012). Mining association rules in student's assessment data. International Journal of Computer Science Issues, 9(5), 211-216.
- [20] Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer Application and Technology, 2(2), 43-47.
- [21] Osofisan, A. O., Adeyemo, O. O., & Oluwasusi, S. T. (2014). Empirical Study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database. African Journal of Computing & ICT, 7, 191-193.
- [22] Freedman, D., Pisani, R., & Purves, R. (1978). Statistics, 2007.
- [23] Wu, A. K., & Leung, C. H. (2002, December). Evaluating learning behavior of web-based training (WBT) using web log. In Computers in Education, 2002. Proceedings. International Conference on (pp. 736-737). IEEE.
- [24] Ingram, A. L. (1999). Using web server logs in evaluating instructional web sites. Journal of Educational Technology Systems, 28(2), 137-157.
- [25] Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assessment

- system. *Technology Instruction Cognition and Learning*, 3(1/2), 63.
- [26] Mazza, R. (2009). *Introduction to information visualization*. Springer Science & Business Media.
- [27] Burr, L., & Spennemann, D. H. (2004). Patterns of user behaviour in university online forums. *International Journal of Instructional Technology and Distance Learning*, 1(10), 11-28.
- [28] Shen, R., Yang, F., & Han, P. (2002). Data analysis center based on e-learning platform. In *The Internet Challenge: Technology and Applications*(pp. 19-28). Springer Netherlands.
- [29] Gedeon, T. D., & Turner, S. (1993, October). Explaining student grades predicted by a neural network. In *Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of 1993 International Joint Conference on* (Vol. 1, pp. 609-612). IEEE.
- [30] Fausett, L. V., & Elwasif, W. (1994). Predicting performance from test scores using backpropagation and counterpropagation. In *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on* (Vol. 5, pp. 3398-3402). IEEE.
- [31] Oladokun, V. O., Adebajo, A. T., & Charles-Owaba, O. E. (2008). Predicting students' academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology*, 9(1), 72-79.
- [32] Kotsiantis, S. B., & Pintelas, P. E. (2005, July). Predicting students marks in hellenic open university. In *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)* (pp. 664-668). IEEE.
- [33] Yu, C.H., Jannasch-pennell, A., Digangi, S., Wasson, B.(1999). Using On-line interactive statistics for evaluating Web-based instruction, In *Journal of Educational Media International*, 35, 157-161.
- [34] Wang, A. Y., & Newlin, M. H. (2002). Predictors of web-student performance: The role of self-efficacy and reasons for taking an on-line class. *Computers in human behavior*, 18(2), 151-163.
- [35] Hien, N. T. N., & Haddawy, P. (2007, October). A decision support system for evaluating international student applications. In *2007 37th Annual Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports* (pp. F2A-1). IEEE.
- [36] Nebot, A., Castro, F., Vellido, A., & Mugica, F. (2006, January). Identification of fuzzy models to predict students performance in an e-learning environment. In *The Fifth IASTED international conference on web-based education, WBE* (pp. 74-79).
- [37] Romero, C., Gutiérrez, S., Freire, M., & Ventura, S. (2008, June). Mining and visualizing visited trails in web-based educational systems. In *Educational Data Mining 2008*.
- [38] Superby, J. F., Vandamme, J. P., & Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Workshop on Educational Data Mining* (Vol. 32, p. 234).
- [39] Ayers, E., Nugent, R., & Dean, N. (2009). A Comparison of Student Skill Knowledge Estimates. *International Working Group on Educational Data Mining*.
- [40] Frias-Martinez, E., Chen, S. Y., & Liu, X. (2006). Survey of data mining approaches to user modeling for adaptive hypermedia. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 36(6), 734-749.
- [41] Rus, V., Lintean, M., & Azevedo, R. (2009). Automatic Detection of Student Mental Models during Prior Knowledge Activation in MetaTutor. *International Working Group on Educational Data Mining*.
- [42] Saini, P., Sona, D., Veermachaneni, S., & Ronchetti, M. (2004). Making e-learning better through machine learning. In *International Conference on" Methods and Technologies for Learning* (pp. 38-45).
- [43] Hershkovitz, A., & Nachmias, R. (2008, June). Developing a log-based motivation measuring tool. In *Educational Data Mining 2008*.
- [44] Rai Sweta, Saini IIPriyanka, Jain IIIAjit Kumar (2014): Model for Prediction of Dropout Student Using ID3 Decision Tree Algorithm, *International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014) 2014, IJARCST Vol. 2 Issue 1 Ver 2 Jan-March 2014*.
- [45] Baker, R. S., Corbett, A. T., & Aleven, V. (2008, June). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *International Conference on Intelligent Tutoring Systems* (pp. 406-415). Springer Berlin Heidelberg.
- [46] Romero, C., and Ventura S.(2013). *Data Mining in Education. WIREs Data Mining and Know.Dis.,Vol.3,pp.12-27*.
- [47] Muenchen, R. A. (2012). The popularity of data analysis software. <http://r4stats.com/articles/popularity>
- [48] Rodrigo, M., Mercedes, T., d Baker, R. S., McLaren, B. M., Jayme, A., & Dy, T. T. (2012). Development of a Workbench to Address the Educational Data Mining Bottleneck. *International Educational Data Mining Society*.
- [49] Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A Data Repository for the EDM Community. In C. B. Romero, S. N. Ventura, M. Pechenizkiy, & R. S. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 43-55). CRC Press.
- [50] Luan, J. (2002). *Data Mining and Knowledge Management in Higher Education-Potential Applications*.
- [51] Yu, C. H., DiGangi, S., Jannasch-Pennell, A., & Kaprolet, C. (2010). A data mining approach for identifying predictors of student retention from sophomore to junior year. *Journal of Data Science*, 8(2), 307-325.
- [52] Yorke, M., Barnett, G., Evanson, P., Haines, C., Jenkins, D., Knight, P., & Woolf, H. (2005). Mining institutional datasets to support policy making and implementation. *Journal of Higher Education Policy and Management*, 27(2), 285-298.
- [53] Thomas, E. H., & Galambos, N. (2002). What Satisfies Students? Mining Student-Opinion Data with Regression and Decision-Tree Analysis. *AIR 2002 Forum Paper*.
- [54] Lin, S. H. (2012). Data mining for student retention management. *Journal of Computing Sciences in Colleges*, 27(4), 92-99.
- [55] Chacon, F., Spicer, D., & Valbuena, A. (2012). Analytics in support of student retention and success. *Research Bulletin*, 3, 1-9.
- [56] Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., & Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. *Procedia Computer Science*, 1(2), 2811-2819.
- [57] Wang, F. H. (2008). Content Recommendation Based on Education-Contextualized Browsing Events for Web-based Personalized Learning. *Educational Technology & Society*, 11(4), 94-112.
- [58] Markellou, P., Mousourouli, I., Spiros, S., & Tsakalidis, A. (2005). Using semantic web mining technologies for personalized e-learning experiences. *Proceedings of the web-based education*, 461-826.
- [59] Abel, F., Bittencourt, I. I., Henze, N., Krause, D., & Vassileva, J. (2008, July). A rule-based recommender system for online discussion forums. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 12-21). Springer Berlin Heidelberg.
- [60] Chu, H. C., Hwang, G. J., Tseng, J. C., & Hwang, G. H. (2006). A computerized approach to diagnosing student learning problems in health education. *Asian Journal of Health and Information Sciences*, 1(1), 43-60.

- [61] García, E., Romero, C., Ventura, S., & De Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modeling and User-Adapted Interaction*, 19(1-2), 99-132.
- [62] Wen-Shung Tai, D., Wu, H. J., & Li, P. H. (2008). Effective e-learning recommendation system based on self-organizing maps and association mining. *The Electronic Library*, 26(3), 329-344.
- [63] Shih, B., Koedinger, K. R., & Scheines, R. (2011). A response time model for bottom-out hints as worked examples. *Handbook of educational data mining*, 201-212.
- [64] Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384.
- [65] Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.
- [66] Vialardi, C., Chue, J., Peche, J. P., Alvarado, G., Vinatea, B., Estrella, J., & Ortigosa, Á. (2011). A data mining approach to guide students through the enrollment process based on academic performance. *User modeling and user-adapted interaction*, 21(1-2), 217-248.