Current State of Art of Academic Data Mining and Future Vision

Insha Majeed

Department of Computer Science and Engineering,
School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi-110062, India
wani.insha06@gmail.com

Sameena Naaz*

Department of Computer Science and Engineering,
School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi-110062, India
snaaz@jamiahamdard.ac.in

Abstract - Academic Mining or Educational Data Mining is multidisciplinary research area of Mathematics, Data Mining/ Machine Learning and Statistics to enable student, teacher, civil society and educational administrators in better understanding and decision making. Academic mining is still evolving and is closely associated with learning analytics which enables us to discover how students understand & choose and how teachers deliver right content in view of changing academic standards. This is derived from large repositories of data available in academic institutions, though data is non standardised and heterogeneous in nature. In this paper effort has been laid to understand academic mining, its root, the present scenario and future vision has been discussed. The work of major contributors in this field has been reviewed and compared in order to make academic mining or Education Data Mining understandable to researchers.

Keywords: Academic Mining or Education Data Mining; Data Mining; Predictive Data Mining; Predictive Analytics.

1. Introduction

Academic as an adjective means associated with academia or relating to school, education, college, academy, or other educational institution, specifically one for advanced higher education. It implies relating to areas to contemplate that are fundamentally professional or applied, as the humanities or pure mathematics, and not relating to anything real or practical. Academic is hypothetical or theoretical and it's not expected to produce an immediate or practical result. As a noun academic means an educator, a student or a teacher who works at a college or university i.e. a person who is scholarly in background, techniques, attitudes, strategies, methods, etc.

A branch of knowledge or information that is learned, taught, instructed, examined and researched as part of higher advanced education is known as **Academic discipline** or an **Academic field**. It covers people, individuals, ventures, groups, challenges, projects, studies, communities, enquiry and research fields that are emphatically related with a given scholarly educational branch of knowledge or ecology department e.g. in the field of science – its branches like chemistry, physics, mathematics, and computer science are commonly referred to as scientific disciplines. The people that are associated with these academic scientific disciplines are commonly called as specialists or professional experts.

According to Zongyi deng - "An academic discipline is a branch of learning or field associated with a scholarly department inside a college or university detailed for the progression and betterment of research and scholarship and the expert training of scientists, scholastics, specialists, researchers and academics. [Bashar, (n.d)]

Data mining is a multidisciplinary subfield of computer science. It is the computing process of discovering and extracting patterns in large data sets including techniques at the intersection of machine learning, artificial intelligence and database systems. It is a process that searches for the precious nuggets from the raw data [Han *et al.* (2006)]. The general objective of data mining (DM) process is to extract information from a data set and convert it into a comprehensible structure for additional use. The analysis step of the KDD (Knowledge discovery in databases process) is Data mining. The term is often applied to any type of vast scale information or data processing like collection, analysis, warehousing, extraction and statistics. Also, the term is frequently applied to any utilization and application of computer DSS (Decision support system) which includes machine learning, artificial intelligence and business knowledge or intelligence. The term is misnomer, on the grounds that the objective is extraction of knowledge and patterns from extensive large and huge amount of data, &not simply the mining (extraction) of data.

[Baradwaj and Pal (2011)] described data mining as a technique that allows the user to analyze, categorize, and summarize the data which are recognized during the process of mining. The automatic or semi-automatic analysis of large amounts of data to extract formerly unknown, interesting patterns like groups of data records (clustering or cluster analysis), dependencies (association rule mining) and records which are unusual (anomaly detection) are actually the real tasks of data mining. These interesting patterns would then be able to be viewed as a sort of input data, and might be utilized as a part of further analysis, e.g. predictive analytics, machine learning, etc.

Predictive data mining [Predictive Data Mining (n.d.)]: The data mining that is done on the data to predict or forecast trends or which is improved for the purpose of using business knowledge or intelligence is known as Predictive Data Mining(PDM). The predictive analytics is supported by predictive data mining. This predictive data mining can enable business pioneers to make better decisions and can increase the value of the endeavors of analytics team. For the utilization of data to forecast the results we use the predictive data analytics. The algorithm-based tools may be used by PDM to go through a database of customer to look at past exchanges and transactions with a specific end goal to support and help theories about possible future volumes of exchanges and transactions. In other words, the data may help to predict what will happen later on in the business, enabling business pioneers to plan and design accordingly.

Predictive analytics [Predictive Analytics (n.d.)] : portrays a scope of scientific, analytical and statistical techniques utilized for creating models that may be used to predict and anticipate future events, behaviors or practices. There are distinctive types of predictive models that vary based on the behavior, event, conduct or occasion that is being predicted. Almost all the predictive models produce a score; a higher score shows that a given behavior, occasion or event is probably going to happen. The data mining techniques along with predictive analytics and predictive models depends on advanced or time series regression models and multivariate analysis techniques. These techniques enable organizations to choose and decide trends, relationships and predict future events, occasions or behaviors.

Educational data mining (EDM) concerns with developing methods for discovering knowledge from data that come from educational environment [El-Halees (2009)]. Educational Data Mining (EDM) is an application of data mining techniques and aims to evaluate the educational data issues by availing existing techniques in data mining [Barnes *et al.* (2009)]. EDM is said 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 [Romero and Ventura (2010)].

The Educational data mining which is defined by the website www.educationaldatamining.org , known as the Educational Data Mining Community website, is as follows : "Education Data Mining is an egressing rising discipline, concerned about creating strategies for exploring the types of data which are unique that originate from educational settings, and utilizing those methods or techniques to better comprehend students, and the settings which they learn in." [International Educational Data Mining Society (n.d)].

It has been recommended that educational data mining methods are frequently unique in relation to standard data mining methods, because of the need to absolutely account for (and the chances to exploit) the multi-level chain of command and non-independence in educational data [Baker in press]. Hence, it is progressively normal to see the utilization of models drawn from the psychometrics literature in EDM publications [Barnes (2005), Desmarais and Pu (2005), Pavlik *et al.* (2008)].

In 2009, an academic journal was set up for educational data mining (EDM) known as the journal of Educational Data Mining. This journal was set up for sharing, circulating and dissipating research results. Also, later the **International Educational Data Mining Society** was set up by the educational data mining scientists in 2011, to integrate EDM researchers and keep on developing and growing the education field. [Nithya *et al.* (2016)]. 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 [Huebner (2013)]. Using data mining, different users or candidates observe educational information through different approaches according to their own perception, aim etc. [Hanna (2004)]. The teaching-learning methods can be improved by discovering knowledge form EDM algorithms [Merceron and Yacef (2005)], and thus, there has been an exponential growth in the number of publications that evolved from time-to-time in the development of EDM [Romero and Ventura (2010)].

Education Data Mining also obtains a lot from the techniques like data mining and machine learning. In reality, the term ―Educational Data Mining (EDM) is a minor misnomer in which ―data mining is generally affiliated with substantial databases and great part of the study is focused on developing rapid and proficient algorithms for finding significance in the data. Although there are definitely databases with thousands or even massive records, but that is just as common to deal with such amount of data. EDM is most likely to face the problems of petite data, rather than the general data mining problem of hefty data. Universal machine learning

methods, especially unsupervised or semi-supervised learning has a more direct influence on EDM. It is, however, imperative to note that EDM does contribute to some usability characteristics with general data mining. Most considerably, well-built and generalized EDM methods should have few parameters and require small or no user intrusion.

In the educational Systems, the application of data mining can be coordinated to support the particular need of each one of the members in the process and the education system itself. Students are required to include the proposal for teaching material, additional activities and undertaking that would support and enhance his/her learning procedure. Teachers would have the feedback, potential outcomes to classify students into group's base on their requirement for direction, guidance and monitoring, to find effective actions, mistakes and locate the compelling activities [Bhushan and Deshmukh (2014)].

Currently the **need of data mining in Education System**, for presenting information in an integrated and consistent format has increased, along with the growing needs of the user, thus requiring an optimal integration of information technology (IT) for supporting the process of presenting the appropriate information. Basically, information acquired from every operational data is handled and processed in accordance with the stakeholder requirements. Involvement of data in the operational implementation of the company today has an important role to support any activities. Therefore, the warehouse is required by the companies for integrating existing data. The system that can support the company's operations is known as database system. The quantity of data stored in the educational databases is increasing rapidly. In order to take an advantage of the large amount of the data, it is necessary to find relationships between variables by using the data mining models. The information from an academic system can be rapidly assessed for the probability factors of student failing in the academic system, and the alerts can be sent to the parent and academic staff to warn them about low performance score of the student. And based on the integration of data mining and data warehouse, it has got a significance focus for decision making process, especially in the education environment.

2. Literature review

The past several decades witnessed an active and vital growth in both the areas of software and databases pertaining to student's information reflecting how they learn [Baker (2010), Peterson *et al.* 2010)]., which has turned out to be a gold mine in educational research area [Mostow and Beck (2006)].

The utilization of World Wide Web (www) in educational area has yield a new medium known as e-learning in which huge amount of teaching i.e. measure of educating and learning behaviors can be produced and analyzed [Castro *et al.* (2007)].There has been multifold studies carried by various analysts in the zone of Educational data mining and among these the most pronounced, reasonable, relevant and significant is of C. Romero and S. Ventura [Romero and Ventura (2007)], who concentrated on use of data mining in the areas like: Intelligent and versatile education systems which are web-based; well-known learning content management systems; Web-based courses and Traditional Educational Systems. In these frameworks, the data can be gathered from heterogeneous sources with the thought of discovering knowledge. Different data mining strategies like clustering, classification 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 [Romero and Ventura (2007)].

In the field of education, there has been a great contribution in the form of workshops, journals, international congress and few books [Romero and Ventura (2007)] that demonstrates the noticeable quality of data mining. The researchers set forth their view and highlighted some of the areas that should be taken into account for the development, improvement, growth and advancement in Educational data mining. The content of mining of tools, generalization of tools & techniques, that can be applied to any educational system, integration data mining tasks into a single application (E-learning), data mining techniques are specific to EDM [Romero and Ventura (2007)].

[Romero and Ventura (2010)] classified educational tasks and the various techniques & strategies   that are associated with Education Data Mining. The literature is an enhanced, updated and considerably much more elaborated than the past one Romero & Ventura [Romero and Ventura (2007)] and has visualized the recurrence of each research area published till 2009 and consequently, revealed most of the research lines that have touched the heights: Analysis & Visualization; predicting performance, providing feedback, student modelling, recommendation systems, detecting behaviors, grouping students. Moreover, the aggregate references being utilized till date (2009) are related with the type of data like: Web-based education/e-learning (54 references), Traditional Education (36 references) and intelligent tutors (29 references) etc [Romero and Ventura (2010)]: It demonstrated that the quantity of references with the highest score resulted in E-learning/Web-based Education. The former review classified data mining techniques utilized by use different authors.

In [Minaei-Bidgoli *et al.* (2003): an approach was proposed by author's, based on the characteristics to categories students in order to predict their final marks grade. In an education web-based system, these characteristics were mined from logged information data. On an online data set, the author's designed, implemented, and evaluated a progression of pattern classifiers and thus analyzed their performance. To confine the students', four classifiers were applied. Thus, for the performance of classification, the sequence of multiple classifiers prompt to a significant progress.

In [Sheikh *et al.* (2003)]: for analyzing educational data, the researchers interpreted that how association rules are useful in EDM (Educational data mining). Also, the cosine and added value were explained and clarified by researchers, that are well suited and appropriate to educational data and thus the educator or instructor can interpret their outcome result easily. From Learning Management System (LMS), the case study with data was provided by them [Sheikh *et al.* (2003)].

In [El-Halees (2009)]: researchers explained that in higher education, how data mining is helpful predominantly to improve the students' performance. For this purpose, they utilized the database course and furthermore composed all data which is available including their usage of Model learning facility. They used classification techniques by using association rule, clustering analysis i.e. clustered the students into the group using education mining clustering (EM-clustering), decision tree and utilizing outlier analysis that discovered the outlier in the data. All this knowledge of classification techniques was used to improve the performance of students'.

Another major effort in which researchers supervised the data mining application to traditional educational system mainly intelligent web based educational system, web-based courses, learning content management system [Romero and Ventura (2007)].

To a great extent, from the analysis of logs of student-computer interaction the EDM has rose up and egressed. This emerging principle of EDM is possibly most clearly publicized by earlier EDM workshops – e.g. At AIED2005, the workshop on Usage Analysis in Learning Systems [CHOQUET *et al.* (2015)].

[Romero and Ventura (2007)] listed the methods as web mining methods. These methods are quite prominent both in mining of web information data and in mining presently different types of educational data. A second perspective on EDM is specified by [Baker (2010)], which analyzes and classifies various tasks in EDM as follows:

- Association rule mining
- Causal data mining
- Classification
- Clustering
- Correlation mining
- Density estimation
- Discovery with models
- Distillation of data for human judgment
- Prediction
- Regression
- Relationship mining
- Sequential pattern mining

There are number of top high-quality research studies supervised by pronounced authors in the area of educational data mining which has been valuable in the area of improving students' success and student-learning outcomes. Among them, a research team [Garcia *et al.* (2011)] built up a basic and simple toolkit for data mining so that non-expert users can fetch information for their courses easily. This research exploration is very significant and noteworthy because of the fact, that this effortless, toolkit does not require profound knowledge of various data mining tools, visualization & statistics & machine learning algorithms [Romero and Ventura (2010)]

The other research contributions towards Educational Data Mining incorporated that of [Ramesh *et al.* (2011)]., they inspected the accuracy of 5 data mining techniques, with the intend of finding most optimal technique that could be utilized on a given dataset, and to predict the probability of student's placement based on final semester.

[Tair and El-Halees (2012)] gathered data of fifteen years [1993-2007] from College of Science and Technology and applied educational mining techniques with the establishment of discovering useful and valuable information from institutional domain. The data that is related to education viz. Education data is considered to be a key component in acceleration, progression and advancement of any country [Agarwal *et al.* (2012)]. Hence, if data mining techniques are applied at educational course, it can help in enhancing student's performance, course selection, and retention rate of an organization [Goyal and Vohra (2012)] and, consequently decrease the dropout probability [Yadav *et al.* (2012)].

Moreover, in 2013, various researchers aimed for improving the nature and quality of education and the performance of students' by applying different data mining techniques at various levels of progressive systems within the educational system (Graduation, Post-Graduation) [Priya and Kumar (2013), Bhise *et al.* (2013), Kumar and Chadha (2012)].

[Ahmed and Elaraby (2014)] proposed an examination, keeping in mind the end goal to distinguish weak students' in order to decease the failing ratio and taking relevant action at the exact specific time. The students' can utilize certain helping tools that will encourage the students to take best decisions in view of the courses to select [Osofisan *et al.* (2014)].

Table 1. Prediction Accuracy Results of various classification techniques

Attributes	Techniques	Results (Accuracy)	Authors
Internal assessments	• K-Nearest Neighbor • Decision Tree • Neural Networks(NN)	82% 76% 81%	[Bigdoli *et al.* ,(2003)] [Romero *et al.* ,(2008)] [Wang and Mirovic, (2002)]
Psychometric factors	• SVM (Support Vector Machine) • K-Nearest Neighbor • Decision Tree • Neural Networks	83% 69% 65% 69%	[Sembiring *et al.*,(2011)] [Gray *et al.* ,(2014)] [Gray *et al.* ,(2014)] [Gray *et al.* ,(2014)]
Internal assessments, CGPA, Extra-curriculum activities	• SVM (Support Vector Machine) • Naïve Bayes • Decision Tree • K-Nearest Neighbor	80% 73% 66% 83%	[Mayilvaganan and Kapalnadevi, (2014)]
External assessments	• Decision Tree • Neural Networks	85% 97%	[Bunkar *et al.* ,(2012)] [Arsad *et al.* ,(2013)]
CGPA	• Naïve Bayes • Neural Networks • Decision Tree	75% 75% 91%	[Jishan *et al.* ,(2015)]
High school background, CGPA, Social Network Interaction, Student Demographic, Scholarship	• Neural Networks • Decision Tree • Naïve Bayes	71% 73% 76%	[Osmanbegovic and Suljic, (2008)]
High school background, Student Demographic	• Decision Tree • Neural Networks	65% 72%	[Ramesh *et al.* ,(2013)]
Internal assessments, External assessments	• Neural Networks	98%	[Anupama and Vijayalakshmi, (2012)]

[Shahiri *et al.* (2015)]: has reviewed in their paper the past studies on predicting the performance of students' with different explanatory techniques. The vast majority of the specialists have utilized total review point normal (CGPA) and inner evaluation as information sets. The primary target of this research paper is to give an outline on the data mining strategies that have been utilized to predict the students' performance. The authors [Shahiri *et al.* (2015)] reviewed attributes, analyzed by the various techniques of classification – Support Vector Machine(SVM), K-Nearest Neighbor, Naïve Bayes, Neural Networks(NN) & Decision Tree, to compare that which one is the best method of prediction for analyzing the performance of students'. The prediction accuracy results summarized by [Shahiri *et al.* (2015)] are in Table 1.

Among all the methods, the classification method is every now and again utilized as a part of educational data mining area [Shahiri *et al.* (2015)].The classification techniques reviewed by [Shahiri *et al.* (2015)] are: Neural Networks (NN), Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbor and Naive Bayes. Among these techniques, it has been shown that the highest prediction accuracy is of Neural Networks (NN) which is (98%) trailed by Decision tree which is (91%). Next it has been seen that the K-Nearest Neighbor and SVM gave exact the same accuracy of (83%)[Shahiri *et al.* (2015)] also reviewed, the technique that has bring down forecast exactness i.e. the technique which has got the lowest accuracy among all is Naive Bayes which is (76%) only. Also, the authors have concluded that the meta-analysis of predicting the performance of students' has persuaded them to complete further research to be applied and practiced in our environment .Thus, in a systematic way, our environment will push the educational system to observe and monitor the performance of students' .[Shahiri *et al.* (2015)].

[Nithya *et al.* (2016)]: has described in their paper that educational Data Mining alludes to methods, apparatuses, and look into intended for naturally removing significance from huge storehouses of information created by or identified with individuals' learning exercises in educational settings. Also, the authors have surveyed that it is a rising control, worried about creating techniques for investigating the special sorts of information that originate from educational settings, and utilizing those strategies to better comprehend understudies, and the settings which they learn in.

The various data mining applications, in the sector of education like Grouping Students, Enrolment Management, Visualization and Analysis of Student Data, Planning and scheduling, Organization of Syllabus, Predicting Students Profiling, Predicting Students Performance, User Modeling, and Detecting Cheating in Online Examination has been described in the paper [Nithya *et al.* (2016)].

Apart from this, Applications of Algorithms in Education Mining are also included like Apriori Algorithm, SVM (Support Vector Machine), The Expectation Maximization Algorithm, C.4.5 and Page Ranker. These algorithms have demonstrated an amazing change in methodologies like course diagram arrangement, teacher student understanding and high yield output and turn out proportion [Nithya *et al.* (2016)].

Recently in 2017, [Dutt *et al.* (2017)]: have described in their paper that by and by, educational establishments arrange and store gigantic volumes of information, for example, understudy enrolment and participation records, and in addition their examination comes about. According to them, mining such information yields fortifying data that serves its handlers well. Fast development in instructive information focuses to the way that refining enormous measures of information requires a more complex arrangement of calculations. This issue prompted the development of the area of instructive information mining i.e., Educational Data Mining (EDM). Conventional data mining algorithms can't be specifically connected to educational issues, as they may have a particular goal and capacity. This suggests a preprocessing calculation must be authorized first and final then some particular data mining techniques can be connected to the issues. One such preprocessing calculation in EDM is Clustering. [Dutt *et al.* (2017)] has used the Clustering technique and according to the authors, many investigations on EDM have concentrated on the utilization of different information mining calculations to instructive traits.

The paper [Dutt *et al.* (2017)] gives over three decades long (1983-2016) precise and systematic literature review on clustering algorithm and its relevancy, pertinence and ease of use with regards to EDM. Future bits of knowledge are sketched out in light of the writing looked into, and roads for additionally explore are distinguished. Since [Dutt *et al.* (2017)] has displayed more than three decade's methodical review on gathering calculation and its appropriateness and convenience with respect to EDM.. This paper has moreover outlined and sketched out a few future bits of knowledge on educational data clustering based on the current literary works assessed [Dutt *et al.* (2017)].

3. Conclusion

Academic mining is far from conclusion and is still very much evolving, however with right kind of tools and research this area is not only going to benefit students and teachers but also parents, civil society etc. However, there is still no standardization neither in academic standards nor in data generated in academic institutions besides academia is in state of contradiction globally thus adding challenges to researchers, but this area has to be established in order to have standardization in academia. Major reforms that can be made in this field in the future are more consistent and versatile datasets and use of hybrid data mining techniques. These techniques can be implemented at the school education level also which can help students choose better fields for them as per their interests.

References

[1] 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.

[2] 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), 140-144

[3] Arsad, P. M., Buniyamin, N. & Manan, J.-l. A.(2013, November). A neural network students' performance prediction model (NNSPPM). In Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on (pp. 1-5). IEEE.

[4] Baker, R. S. J. D. (2010). Data mining for education. International encyclopedia of education, 7(3), 112-118.

[5] Baradwaj, B., & Pal, S. (2011). Mining Educational Data to Analyse Students Performance. International Journal of Advanced Computer Science And Application, 2(6), 63-69.

[6] Bashar. Subject and discipiline. Slideshare.net. Retrieved 18 March 2018, from https://www.slideshare.net/abubashars/subject-and-discipiline

[7] Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (2009). Educational Data mining 2009: Cordoba,Spain.: 2nd International Conference on Educational Data mining, Proceedings.

[8] Barnes, T. (2005, July). The Q-matrix method: Mining student response data for knowledge. In American Association for Artificial Intelligence 2005 Educational Data Mining Workshop (pp. 1-8).

[9] 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) ISSN, 2279-0837,6(6), 18-21

[10] Bhushan, S., & Deshmukh, V. (2014). Application of Data Mining Technique for Prediction of Academic Performance of Student A Literature survey. International Journal on Recent and Innovation Trends in Computing And Communication (IJRITCC), 2(12), 3962-3965.

[11] Bunkar, K., Singh, U. K., Pandya, B., & Bunkar, R. (2012, September). Data mining: Prediction for performance improvement of graduate students using classification. In Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on (pp. 1-5). IEEE.

[12] 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.

[13] CHOQUET, C., LUENGO, V., & YACEF, K. (2015). Proceedings of "Usage Analysis in Learning Systems" workshop. Amsterdam, The Netherlands: conjuction with Artificial Intelligence in Education AIED 2005.

[14] Desmarais, M. C., & Pu, X. (2005). A Bayesian student model without hidden nodes and its comparison with item response theory. International Journal of Artificial Intelligence in Education, 15(4), 291-323.

[15] Dutt, A., Ismail, M., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. IEEE Access, 5, 15991-16005. http://dx.doi.org/10.1109/access.2017.2654247.

[16] El-Halees, A. (2009). Mining students data to analyze e-Learning behavior: A Case Study.

[17] García, E., Romero, C., Ventura, S., & De Castro, C. (2011). A collaborative educational association rule mining tool. The Internet and Higher Education, 14(2), 77-88.

[18] Goyal, M., & Vohra, R. (2012). Applications of Data Mining in Higher Education. IJCSI International Journal of Computer Science Issues, 9(2), 113-120.

[19] Gray, G., McGuinness, C., & Owende, P. (2014, February). An application of classification models to predict learner progression in tertiary education. In Advance Computing Conference (IACC), 2014 IEEE International (pp. 549-554). IEEE.

[20] Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques .2nd edition.

[21] Hanna, M. (2004). Data mining in the e-learning domain. Campus-wide information systems, 21(1), 29-34.

[22] Huebner, R. A. (2013). A Survey of Educational Data-Mining Research. Research in higher education journal, 19.

[23] International Educational Data Mining Society www.educationaldatamining.org

[24] Jishan, S. T., Rashu, R. I., Haque, N., & Rahman, R. M. (2015). Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. Decision Analytics, 2(1), 1-25.

[25] Kumar, V., & Chadha, A. (2012). Mining Association Rules in Student's Assessment Data. International Journal Of Computer Science Issues, 9(5), 211-216. Retrieved from http://www.ijcsi.org/

[26] Mayilvaganan, M., & Kalpanadevi, D. (2014, December). Comparison of classification techniques for predicting the performance of students academic environment. In Communication and Network Technologies (ICCNT), 2014 International Conference on (pp. 113-118). IEEE.

[27] Merceron, A., & Yacef, K. (2005, May). Educational Data Mining: a Case Study. In AIED (pp. 467-474).

[28] Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003, November). Predicting student performance: an application of data mining methods with an educational web-based system. In Frontiers in education, 2003. FIE 2003 33rd annual (Vol. 1, pp. T2A-13). IEEE.

[29] Mostow, J., & Beck, J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. Natural Language Engineering, 12(2), 195-208.

[30] Nithya, P., Umamaheswari, B., & Umadevi, A. (2016). A survey on educational data mining in field of education. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 5(1).

[31] Osmanbegović, E., & Suljić, M. (2012). Data mining approach for predicting student performance. Economic Review, 10(1), 3-12.

[32] 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.

[33] Pavlik Jr, P. I., Cen, H., Wu, L., & Koedinger, K. R. (2008). Using Item-Type Performance Covariance to Improve the Skill Model of an Existing Tutor. In Proceedings of the 1st International Conference on Educational Data Mining, 77-86.

[34] Peterson, P., Baker, E., & McGaw, B. (2010). International encyclopedia of education. Elsevier Ltd.

[35] Priya, K., & Kumar, A. (2013). Improving the Student's Performance Using Educational Data Mining. International Journal of Advanced Networking And Applications, 4(4), 1680-1685.

[36] 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).

[37] Ramesh, V. A. M. A. N. A. N., Parkavi, P., & Ramar, K. (2013). Predicting student performance: a statistical and data mining approach. International journal of computer applications, 63(8).

[38] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.

[39] Romero, C., Ventura, S., Espejo, P. G., & Hervás, C. (2008, June). Data mining algorithms to classify students. In Educational Data Mining 2008.

[40] 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.

[41] Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., & Wani, E. (2011, April). Prediction of student academic performance by an application of data mining techniques. In International Conference on Management and Artificial Intelligence IPEDR (Vol. 6, No. 1, pp. 110-114).

[42] Shahiri, A., Husain, W., & Rashid, N. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. Procedia Computer Science, 72, 414-422. http://dx.doi.org/10.1016/j.procs.2015.12.157.

[43] Sheikh, L. M., Tanveer, B., & Hamdani, M. A. (2004, December). Interesting measures for mining association rules. In Multitopic Conference, 2004. Proceedings of INMIC 2004. 8th International (pp. 641-644). IEEE.

[44] Tair, M. M. A., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. International Journal of Information and Communication Technology Research, 2(2), 140-146.

[45] What is Predictive Analytics? - Definition from Techopedia. Techopedia.com. Retrieved 26 October 2017, from https://www.techopedia.com/definition/180/predictive-analytics.

[46] What is Predictive Data Mining? - Definition from Techopedia. Techopedia.com. Retrieved 26 October 2017, from https://www.techopedia.com/definition/30597/predictive-data-mining.

[47] Yadav, S., Bharadwaj, B., & Pal, S. (2012). Mining Education Data to Predict Student's Retention: A comparative study. (IJCSIS) International Journal of Computer Science And Information Security, 10(2), 113-117.