International Journal of Recent Development in Computer Technology & Software Applications
Vol. 1, Issue 1- 2017

REVIEW ON DATA MINING IN EDUCATION SYSTEM

AMOL E DAHAKE*, PRIYANKA S MITKARI*, PRADIP SINGLE*

ABSTRACT

Data mining is the vast field for extracting the meaningful information from data warehouse. Development of every country education is the essential thing. To make the education easier the board term called data mining is used. Data mining in education gives various techniques, methods and tools for the better understanding. Education data mining fields describes research field concern with application of data mining, machine learning and statistics. EDM muffled the learning management system, student attraction student retention.

The EDM consist of the phase likewise discovering relationship in data discover relationship must be validated to avoid over fitting, prediction about future in event in learning environment, prediction are used to support decision making process. The various application of data mining is analysis and visualization of data recommendation for student, student modeling, grouping of student, social network analysis. This type of approach gives high confidence to student in their studies. It also highlights the opportunities for future. The various data mining techniques such as cluster technique, decision tree technique which can apply to education system. The various techniques such as CLUSTRING technique, APRIORI TECHNIQE decision tree, which can apply to education system.

KEYWORDS*:* Data Mining, Classification, Decision Tree, ID3, C4.5, CHAID, C5.

INTRODUCTION

The concept of data mining is the techniques of extracting previously unknown information with the large relevance from database, in order to use it in the decision-making process. Data mining is a process of extracting previous unknown, valid, potential useful and hidden patterns from large data sets. As the huge amount of data stored in Educational database is increasing rapidly, in order to find the hidden relation between variables using different data mining techniques.

- Data mining techniques are analytical tools that can be used to extract meaningful knowledge from these large data sets.
- To face these challenges different systems are used such as Enterprise Resource Planning, Online Transaction Process etc. Some definition of data mining has been used. Most of these definition has been concentrated on relation of data mining with education and some of them oriented to the

*Department of Information Technology, Anuradha Engineering College, Chikhli.
Correspondence E-mail Id: editor@eurekajournals.com

analytical processing of the educational data, while EDM is becomes speedier and have been converge with a few zones, for example, training, e-learning, information mining ,information investigation, savvy framework and so forth limit degree and meaning of EDM in some relate work. The Educational Data Mining people group site, characterizes instructive information mining as Educational Data Mining is a rising control, worried about creating strategies for investigating the interesting sorts of information that originate from instructive settings, and utilizing those techniques to better comprehend understudies, and the settings which they learn in.

- Today, one of the greatest difficulties that instructive establishments confront is the touchy development of instructive information and to utilize this information to enhance the nature of administrative choices and understudy's execution. The fundamental target of advanced education foundations is to give quality training to its understudies. One approach to accomplish most elevated amount of value in advanced education framework is by finding learning for expectation with respect to enrolment of understudies in a specific course, estrangement of customary classroom showing model, location of out of line implies utilized as a part of online examination, discovery of anomalous esteems in the outcome sheets of the understudies, forecast about understudies' execution.

- This paper plans to reason the utilization of Data mining systems to enhance the productivity of higher instructive establishments. In the event that information mining procedures, for example, bunching, choice tree and affiliation can be connected to advanced education forms, it can help enhance understudy's execution. With a specific end goal to get required advantages from such huge information and to discover concealed connections between factors utilizing diverse information mining procedures created and utilized. Grouping and choice tree are most broadly utilized procedures for future expectation. The primary objective of bunching is to segment understudies into homogeneous gatherings as per their attributes and capacities.

LITERATURE SURVEY

Learning that initially started in the class room was based on three models namely behavioural, cognitive and constructivist models.[1] The behavioural models relay on observable changes in the behaviour of the student to assess the learning outcome. The cognitive models are based on the active involvement of teacher in the learning which helps in guided learning. In the constructivist models, the students have to learn on their own from the knowledge available to them. Siemens (2004) [2] proposed a new model termed “Connectivism” which was characterized as the “amplification of learning, knowledge and understanding through the extension of personal network”. According to this model, learning is no longer an internal activity Connectives proposed learning in a network of nodes which improved the learning experience of students and reduced the need for the direct involvement of an instructor. Since then, traditional learning environments have gradually mutated into community based learning environments.

- This section presents some related studies on data mining in education, the section categorized these studies in according to the educational areas, providing an over view of each application and techniques used in these areas.
- A number of publications about EDM in higher education have grown exponentially over the past few years. Romero & Ventura [3] mentioned the compression, in higher educational institutions to provide up – to date information on institutional effectiveness. Institutions are also increasingly held accountable for student success It also increases the responsibility of educational institutions for the success of the student.

- One reaction to this weight is to discover better approaches for applying the techniques for investigation and information mining to instructively significant information. At a beginning time, Baker and Yacef [4] said that it is useful to have a more exhaustive scientific categorization of the diverse zones of concentrate inside EDM despite the fact that analysts have effectively settled a fundamental scientific classification.
- There are diverse ways that information mining can be utilized as a part of advanced education organizations. For example, Course administration frameworks, Student practices, choice emotionally supportive network in advanced education understudy maintenance and whittling down.

MAJOR TOPICS OF INTEREST

- COURSE MANAGEMENT SYSTEM
- DISTENCE EDUCATION
- STUDENT BEHAVIOUR
- STUDENT MODELING
- PERFORMANCE PREDICTION

DATA MINING METHODOLOGY

- Before applying the data mining techniques on the data set, there ought to be a philosophy that oversees our work. Figure 1 portrays the work technique utilized as a part of this paper, which depends on the structure proposed in.
- The approach begins from the issue definition, at that point preprocessing which are examined in the presentation and the informational index and preprocessing segments, at that point we go to the information mining strategies which are affiliation, order, grouping, and anomaly identification, trailed by the assessment of results and examples, at last the learning portrayal process.
- In this segment we depict the after effects of applying the information mining procedures to the information of our contextual analysis, for every one of the four information mining errands; Association, characterization, grouping and anomaly location, and how we can profit by the found learning.

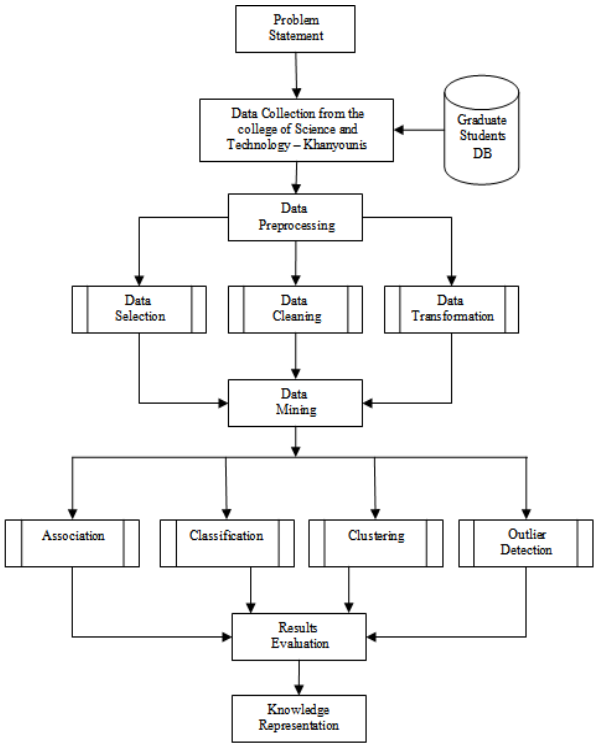

Figure 1.Methodology of Data mining

DATA MINING IN EDUCATION SYSTEM

Data mining can be connected to information originating from two sorts of instructive frameworks: conventional classroom and separation training. It is important to bargain independently with the utilization of data mining systems in each sort because of the way that they have diverse information sources and destinations.

TRADITIONAL CLASSROOMS

- Traditional classroom situations are the most broadly utilized instructive frameworks. It depends on up close and personal contact amongst instructors and understudies composed through teachers. There are a variety of subtypes: private and state funded training, rudimentary and essential instruction, grown-up training, higher, tertiary and scholarly training, specialized curriculum, and so on. They have been censured in light of the fact that they support aloof learning, overlook singular contrasts and needs of the students, and don't focus on critical thinking, basic reasoning, or other higher request thinking abilities.
- In traditional classrooms, instructors endeavor to improve directions by checking understudy's learning forms and breaking down their exhibitions by paper records and observation[10]. They can likewise utilize data about understudy participation, course data, educational programs objectives, and individualized arrangement information.
- Students may wish to know how best to choose courses in view of expectation of how well they will perform in the courses chose. Teachers may wish to realize what realizing encounters are most contributive to general learning results, why is one class beating the other, comparative gatherings of students,[10] and so on. There are a few works about the utilization of information mining in customary training.
- One of the primary articles about the utilization of information mining in instruction to comprehend the understudy enlistment was composed by Sanjeev and Zytkow (1995). They apply information disclosure as articulations ''Pattern P holds for information in Range R'' to college databases. The outcomes were exhibited to a senior college executive keeping in mind the end goal to settle on vital choices about the institutional arrangements.
- Another deal with the utilization of KDD to personality and comprehend whether educational modules modifications can influence understudies in a Brazilian college was finished by. They confirm the subjective effect of updates and assess it utilizing various methods, for example, synopsis, affiliation, arrangement. In a related work, the goal is to choose the frail understudies to go to medicinal classes (Maet al., 2000). They utilize a scoring capacity that depends on affiliation rules.

DISTANCE EDUCATION

- Distance instruction or separation learning comprises of systems and techniques giving access to instructive projects to understudies who are isolated by time and space from speakers. E-Learning frameworks do not have a nearer student– teacher relationship (balanced).
- There are diverse subtypes of separation training: paper-based correspondence instruction, tape instruction, PC helped training (sight and sound instruction, web instruction or online training), and so forth. As of now, the most utilized is online training enabling understudies to helpfully learn through the Internet. Online training is a type of separation instruction conveyed over the Internet (Johnson 2000).
- Today, there is a considerable measure of terms used to elude to electronic instruction, for example, e-learning, e-preparing, online

guideline, online learning, online preparing, electronic guideline, and so on. Also, there are distinctive kinds of online frameworks: synchronous and offbeat, shared and non-communitarian, shut corpus and open corpus, and so forth. These electronic training frameworks can typically record the understudy's gets to in web logs that give a crude hint of the students' route on the website.

- There are a few sorts of logs (Srivastava, 2000) Server log document. This constitutes the most broadly utilized information hotspot for performing information mining, containing only the uncovered subtle elements of timing, way, and information reaction. There is an assortment of arrangements, for example, regular log design (CLF), expanded log organizes (ELF) and so on. (Koutri, Avourisand Daskalaki, 2004). Ordinarily, there is a solitary log petition for all understudies customer log document. This comprises of an arrangement of log documents, one for every understudy, and contains data about the cooperation of the client with the framework. Can be actualized by a remote specialist (such Javascripts, Java Applets), changing the source code of a current program, or utilizing treats.[11]

METHODS OF DATA MINING IN EDUCATION

COURSE MANAGEMENT SYSTEMS

Countless inside EDM centre specifically on course administration frameworks and how they can be enhanced to help understudy learning results and understudy achievement. Around there, instructive information mining (EDM) has been utilized for using information mining methods and research approaches for seeing how understudies learn. Intelligent e-learning strategies and instruments have opened up chances to gather and investigate understudy information, to find out examples and patterns in these information, and to detail new revelations and test presumptions about how understudies learn. Innovation improved learning depends intensely on learning administration frameworks (LMS) or course administration frameworks (CMS). These LMS/CMS consequently record the keystrokes of individual clients as server logs. Mining these logs give examples to instructors to distinguish moderate students and can modify educating systems.

Course management system (CMS) can tender a great variety of procedure and workspaces to facilitate information sharing and communication among participants in a course. The Course management systems (CMS)also let educators distribute information to students, produce content material, prepare assignments and tests, engage in discussions, manage distance classes and enable collaborative learning with forums, chats, file storage areas, news services, etc. Some of the most commonly used one is Moodle (modular object oriented developmental learning environment), a free learning management system, enabling the creation of powerful, flexible and charming online courses and experiences.

STUDENT BEHAVIOURS

Beck and Woolf [13] referred to in their article, how instructive information mining expectation strategies can be utilized to create understudy models. They utilize an assortment of factors to foresee whether an understudy will make a right answer. This work has roused a lot of later instructive information mining work – understudy displaying is a key subject in current instructive information mining, and the worldview of testing EDM models' capacity to foresee future rightness.

Information mining was utilized to evaluate complex understudy practices as for a three – week programming task. Blikstein discovered

outcomes that demonstrated diverse sorts of understudy programming practices in an online course. These log documents contained distinctive kinds of occasions as every understudy finished them. The occasions included coding and non - coding exercises in the online course. This quantitative information mining research found distinctive programming procedures utilized by understudies, and created three programming conduct profiles: duplicate and-pasters, blended mode, and independent [14].

SYSTEM DESIGN

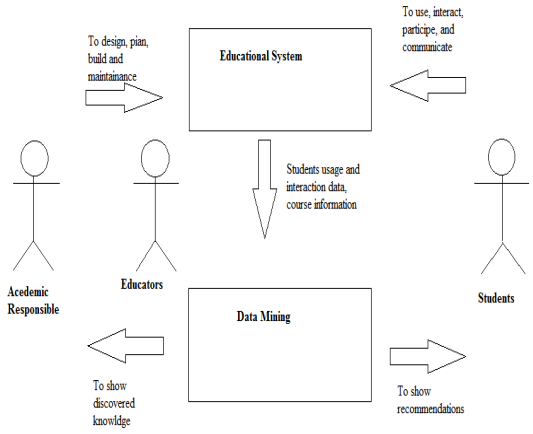

Figure 3.Lifecycle of Data Mining in Education

As we can see in Fig. 3, teachers and scholastics mindful are responsible for outlining, arranging, fabricating and keeping up the instructive frameworks. Understudies utilize and communicate with them. Beginning from all the accessible data about courses, understudies, use and cooperation, diverse information mining methods can be connected to find valuable information that enhances the e-learning process. [15]. The found information can be utilized by teachers as well as by possess clients understudies. Thus, the utilization of information mining in instructive frameworks can be arranged to various understudy with every specific perspective.

DATA MINING TECHNIQUES

The challenges faced in processing Data mining [14] technologies are overcome by using various techniques. The most popular techniques used in educational data mining are listed below.

DECISION TREE

Decision tree is a choice help apparatus that uses a tree-like diagram or model of decision and their conceivable results, including chance occasion results, asset expenses, and utility. It is one approach to show a method.

Decision trees are usually utilized as a part of tasks look into, particularly in decision examination, to help distinguish a methodology well on the way to achieve an objective. Another utilization of decision trees is as an enlightening means for figuring restrictive probabilities. Decision trees can be utilized to break down the confirmation criteria of an organization as appeared in fig.3.

Decision trees are easy to comprehend and decipher and in addition they give great outcomes even with little information. This approach may not be appropriate for information

incorporating unmitigated factors with various numbers of levels.

CLUSTERING ANALYSIS

Clustering involves grouping of records that are similar by identifying the distance between them in an n-dimensional space where n is the number of variables.

This approach is used in many fields like machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.[15] The objects are clustered or grouped based on the principle of maximizing the intra class similarity and minimizing the inter class similarity. Each cluster that is formed can be viewed as a class of objects, from which rules can be derived. Application of clustering in education can help institutes group individual student into class of similar behavior. Partition the students into clusters, so that students within a cluster (e.g. Average) are similar to each other while dissimilar to students in other clusters (e.g. Intelligent).

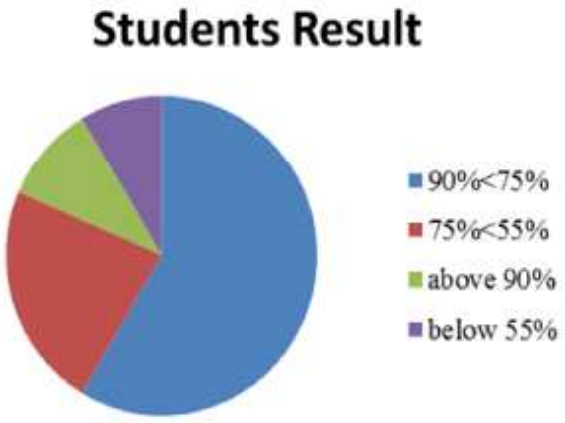

Figure 4.Clustering of student as per performance

There are many clustering techniques reported in the seminar. One has to take many decisions while choosing the appropriate technique for a particular problem. Some of them are:

- **Connectivity Model:** These are based on distance connectivity.
- **Centroid Models:** This technique represents each cluster by a single mean vector.
- **Distribution Models:** Clusters are modeled using statistic distributions, such as multivariate normal distributions used by the Expectation-maximization technique.
- **Density Models:** It defines clusters as connected dense regions in the data space
- **Subspace Models:** Clusters are modeled based upon both cluster members and relevant attributes.

OPEN SOURCE TOOLS FOR DATA MINING

Several Open source tools exist which help in taming Big Data [16] some of the top tools are listed below.

- **MONGODB** is a cross platform document oriented database management system. It uses JSON like documents instead of a table based architecture
- **HADOOP** is a framework that allows distributed processing of big datasets across clusters of networked computers using simple programming models.
- **MAPREDUCE** is a programming model and framework used by had oop. It enables processing huge amount of data in parallel on large clusters of compute nodes.
- **ORANGE** is a python based tool for processing and mining big data. It has an easy

to use interface with drag & drop functionalities with variety of add-ons.

- **WEKA** is a java based tool for processing large amount of data. It has a vast selection of techniques that can be used in mining data.

PROPRIETARY TOOLS

SAP HANA is a proprietary in-memory RDBMS capable of handling large amount of data. It uses Parallel in Memory relational query techniques, Columnar stores and Compression technology to overcome the challenges faced in handling Big Data.

APPLICATION OF DATA MINING IN EDUCATION FIELD

1. Analys is and visualization of data.
2. Predicting student performance
3. Enrolment management
4. Grouping students
5. Predicting students profiling
6. User modeling
7. Organization of syllabus.

CURRENT AND FUTURE SCOPE

The main goal of data mining inn education is to overcome the problem in online examination. While during the online examination some student do chitting in the exam hall,to avoid this the education data mining work on it.

CONCLUSION

From this paper we conclude that data analysis play an important role for any type of education. This paper represent the state of art in EDM, it try to provides a boundary definition to the term Education Data Mining, hence we found that EDM covers bigger area and many interdisciplinary such as education, e-learning, data mining, data analysis, intelligent system and so on, The paper covers most relevant work in the area of EDM in course management systems, student behaviors, decision support system, and Student Retention and Attrition.

This paper also provides a comparison study between some of research work in such areas. The paper concludes that EDM is growing fast and many new ideas and technologies can be merge in this discipline, in addition well-established research and application have provide a considerable contribution. Thus, we expect that EDM will become more useful, and fully operative and available even for external users.

REFRENCES

[1]. Peggy A. Ertmer and Timothy J. Newby, “Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective”, *Performance improvement quarterly*, Vol. 6, No. 4, pp. 50-72, 1993.

[2]. George Siemens, “Connectivism: A learning theory for the digital age”, *International Journal of Instructional Technology & Distance Learning*, Vol. 2, No. 1, 2005.

[3]. Romero, Crostobal, Sebastian Ventura,and Garcia “Data mining in course management system.

[4]. Baker, R., &Yacef, K. (2009). The State of Educational Data mining in 2009: A Review andFuture Visions. Journal of Educational Data Mining, 1.

[5]. Wikipedia, “Big data-Wikipedia, The Free Encyclopedia”, https://en.wikipedia.org/w/index.php? title=Big_data&oldid=669888993. Accessed 2015.

[6]. Han, J. and Kamber, M. (2006) Data Mining: Concepts and Techniques, 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.

[7]. Johnson, Arago, Shaik, & Palma-Rivas, 2000 Comparative study of learner satisfaction and learning outcomes in online and fact-to-face learning environment 11(1), 29-49.

[8]. Arockiam, L., S. Charles, and M. AmalaJayanthi. "An Impact of Emotional Happiness and Personality in Students' Learning Environment." Data Mining and Knowledge Engineering 7.2 (2015): 69-74.

[9]. Beck, J.E., Woolf, B.P.: High-Level Student Modeling with Machine Learning. In: Gauthier, G., et al. (eds.): Intelligent Tutoring Systems, ITS 2000. Lecture Notes in Computer Science, Vol. 1839. Springer, Berlin Heidelberg New York (2000) 584-593.

[10]. Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. Paper presented at the Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, Canada.

[11]. Crist´obal Romero and Sebasti´an Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics-Part c: Applications and Reviews, vol. 40, no. 6, 2010, pp. 601-618.

[12]. Jiawei Han and Micheline Kamber, Data Mining- Concepts and Techniques: Elsevier Publishers, 2006.

[13]. S. Parack, Z. Zahid, and F. Merchant, "Application of data mining in educational databases for predicting academic trends and patterns," in, 2012, pp. 1-4.

[14]. Cynthia Harvey, "50 Top Open Source Tools for Big Data", http://www.datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html. Accessed 2012.

