A SURVEY ON EDUCATIONAL DATA MINING

M. Shirgaovkar [1] , V.I.Pujari [2] , Dr. R.B.Patil [3]

[1] Student M.Sc. I.T., I.C.S. College, Khed, Ratnagri.

[2] Asst. Prof., Department of I.T., I.C.S. College, Khed, Ratnagiri.

[3] H.O.D, Department of I.T., S.K. Somaiya College of Arts, Science and Commerce, Mumbai,

ABSTRACT

Educational Data Mining (EDM) is an developing field to explore the data from numerous educational contexts. It delivers essential knowledge about informing the education, which is used to improve the excellence of teaching and learning. Effective planning can deliver personalized education. This paper discuss a survey on numerous components of educational data mining along with its objectives.

KEYWORDS: Data Mining, Educational Data Mining, Knowledge Discovery

INTRODUCTION review

Data mining, in overall, can be described as the methodical processing of large datasets and finding hidden facts and patterns. The determination of data mining is to appreciate the data trends and develop new and effective visions. This helps with upcoming pacing and delivers a vision. If we look at the current past; the teaching sector has seen an extreme vertical as well as flat growth. This kind of huge growth carries along a promotion in educational data. This is where the educational data mining develops an important feature of the analysis and further growth in the field of education. Hence, applications of technical tools are used to understand the designs and tendencies of the educational data. Educational data mining is authoritative today; otherwise, it would not be likely to analyze the huge amount of educational data. With educational data mining, the first and the leading task is to establish and classify the educational data received from numerous sources. The determination of introducing mechanization in the education is to more professionally understand the evidence-of-learning. It is an ever-evolving process that repeatedly proceeds towards the creation and development of core values of any educational organization. These values are typically to cultivate the aptitude and capabilities of the students. A part of these educational task is to address students' matters with the proactive knowledge to meet learning and management goals. This task can divide into two types: student's based task- operative support of primary investors to fulfill learning objectives and decision-making task- for the constant participation of the crossbreed group of investors to fulfill management-oriented objectives.

OBJECTIVES OF EDUCATIONAL DATA MINING

The purposes of educational data mining can be categorized into two categories- the educational objectives and administrative objectives. The third category is the commercial or the market objectives, which is recently becoming very popular. The complete look at the objectives of the educational data mining is as bellow:

Educational Objectives

They are also known as the academic objectives and are of great importance when it comes to conception and designing of educational content. They help in doing time-to-time development and improvement required in the study material. Another critical educational use is to bring improvement in academic performance of the learners. So, talking on formal terms the educations objectives can be divided into the following sub-categories:

Individual/ Person-oriented- these objectives have a direct contribution to teaching and learning.
The person- oriented purposes of educational data mining that occur around us are -
· Student learning
· Risk
· Modeling
· Behavior
· Cognitive learning
· Predicting right enrollment decision
· Performance analysis
There are many more generic and specific factors, both in the conventional and digital environment for faculty modeling, which helps in job performance analysis and satisfaction analysis.
Department/Institution -oriented- this is precisely related to definite department or organization with respect to-time, structure, and demand. EDM helps organizations and numerous educational departments in
· Redesigning new courses according to industry requirements,
· Understand real-time field problems
· Effective research analysis
· Enhance learning experience and process.
Domain-oriented- as the name recommends this is connected to a specific branch/institution. Particular branches of education can apply EDM to specific courses and use it for improvising and detailing like designing methods/tools, techniques, and Knowledge Discovery based Decision Support System (KDDS) for a specific application.

Administrative or Management Objectives
Educational data mining can be functional to optimization of the school, college or any other learning organization. It can be of prodigious use when it comes to the maintenance of educational organization, which is a main administrative oriented objective and includes the direct contribution of higher authorities and senior administration. Other objectives are to
• Automatic the decision process of management by examining diverse educational data from traditional as well as intelligent tools.
• Investigate with the latest intelligent tools and application of operative techniques used in EDM and
• Understand numerous real-time challenges faced in EDM.
Commercial Objectives
Commercial objectives are mainly important in case of private education, such as the formation of a niche and capturing the market in terms of employments. The data can be efficiently used in marketing for student registration and capturing the scholars for private institutions and the founding of a network for innovative research and practices.

COMPONENTS OF EDUCATIONAL DATAMINING
Educational data mining traces and touches numerous aspects of the education industry. The main components of EDM are - shareholders of education, numerous data mining tools and techniques, educational data, educational atmosphere and task and how they meet the educational objectives.
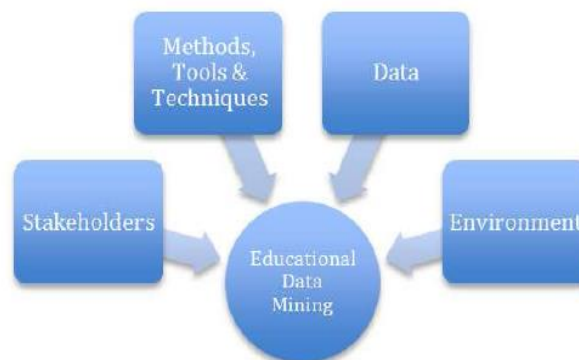

Figure 1: Components of Educational Data Mining

Shareholders

Keeping in mind all facets of education, i.e. primary to higher education, shareholders of the education can be majorly considered into the following:

· **Learners/Students:** The most important and wedged component is the beginners. As students are involved straight in the process of learning, they fall in the primary group of shareholders. EDM can help them with the personalized education based on numerous endorsements and can increase the interestingness of education for students towards learning. Diverse learning tasks can be expressed in the diverse group of students based on their needs.

· **Abilities:** Instructors, Teachers are promoted as they can determine which student need extra support. The estimate of student performance becomes informal. Another influence is it helps in the cataloguing of learners into groups. It can also deliver an insight into the designs in which students can learn- Regular and irregular. Teachers can examine the data and govern the most commonly made errors. Outside just academics, the analysis of the student's learning and performance can also be done to detect if they require any extra provision during the process of education. Teachers are also the primary shareholders.

· **Parents:** Parents are the part of the subordinate group. They are likely for helping their kids to get them enroll in the most appropriate courses for them.

• **Course Researchers and Educational Developers:** They are the persons who design and adjust the course. They are responsible for the development of education. Developers fall into the group of secondary shareholders.

• **Administrators:** They can also be called as the crossbreed users. EDM is useful for effective application of resources; it can help in decisive what are the proposals that can capture more apprentices into numerous programs and courses . They are blamable for various administrative choices such as infrastructure development and paying the expert faculty.

Data for EDM

Analysis of the huge amount of educational data is involved in decision-making and future planning. The educational data is a mix of structured as well as unstructured data collected from simple as well as complicated sources. When collected from various mechanized sources, this will generate big data. This data could be in the form of responses given by large numbers of students to various questions or be a huge pool of student's texts received from collectives like online papers and other general descriptive data. Such information from a totally digital process can actually be separated into two types of information. These two classifications are structured data and unstructured data.

Structured Data: This data is already structured and leave a minor possibility of being too vast and too vague. As a result, the data is self- explanatory and more controlled as compared to unstructured data. At the same time, this makes the data free from human interference and brings in crystal-clear evidence away from biases and judgments. Several sources of collecting structured data that could be called as capable of including embedded and determinative
assessment are as follows-

• Intelligent tutors
• Simulations
• Semantic mapping tools
• Learning management systems

Unstructured Data: This kind of data does not come from one exact source and there is no predefined data model. This may include education information that comes from the web. This contains information such as a learner's IP address or her/his username and may relate to numerous texts from sources like Internet opportunities, video clips, or audio files.

Here are some of the possible sources from where unstructured data can be collected from-

• Learning games
• Social interaction analyses
• Affect meters

• Body sensors

As can be seen, that instructive data are in abundance, and this makes educational data mining an important exercise that can improve the learning development across all verticals of the education industry. All this composed, structured and unstructured data when analyzed will be of great help when it comes to meeting the purposes and to determine exact goals of education. The data could be both mixed and classified too. The data can be produced from online and offline sources-

• **Offline Data:** As the name proposes, offline data are generated through real-time conditions and settings. Setups like traditional classroom tests, communication based contemporary classes, teacher-student interactions, student-to-student interactions, real-time data derived from different courses and numerous departments of any institute like schools, colleges, and universities. Other factors are levels of contribution from the students, students' attendance, behavior and attitude related scores.

• **Online Data:** Unlike offline data, online data is not dependent upon any kind of geographical location. The information are derived from weblogs, E-mails, spreadsheets, transcripts telephonic conversations, medical records, legal information, and publication databases etc.

Data Mining Methods

Some of the most popular and effective approaches of data mining can be classified as follows:

• **Classification:** Classification is training and testing method, which classifies the collected data into some preset groups. It is a useful method for predicting student presentations, risk analysis, student monitoring systems, and detection of errors etc.

• **Clustering:** Like classification, this method puts similar data together into clusters, but not under preset categories. This technique is helpful in distinguishing the preferences of different learners. Examination of students' comprehensive character and methods suitable for cooperative learning are done with the help of clustering.

• **Statistics:** Statistics are useful for course management system and contributions in the determination of thrilling deviations from the mean. It records, statistical functions like mean, mode and helps in managing of a student response system.

• **Prediction:** Extremely useful in passing the future education business trends. It is a technique that is applied to predict success rate, dropout rate and designing methods of retention.

• **Association Rule Mining**: This is an important data mining method in finding numerous relations among the attributes of data, such as admission, migration, parents-faculty-students relation etc. numerous patterns for reasons of student's failure can be found out. The other effective data mining approaches used in the EDM business are neural networks, regression, SVM etc.

Useful Tools of Educational Data Mining

• **WEKA (Waikato Environment for Knowledge Analysis):**The Weka workbench consists of several tools, algorithms and graphics methods that lead to the analysis and predictions. Most of the algorithms are inbuilt in this tool.

• **KEEL (Knowledge Extraction Based on Evolutionary Learning):**KEEL as an application is a set of machine learning software that is designed for providing a resolution to numerous data mining problems. It has a collection of software techniques that are involved in data manipulation and analysis before and after the process as well. It applies soft-computing approaches in mining information about learning and knowledge.

R (Revolution): This is a numerical calculating software/ language that is extensively used by data miners to perform statistics for learning development solutions. R is an enormously versatile tool that is not only systematically designed but is also easy to use. So, smearing stats and formulas in R are suitable.

• **KNIME (Konstanz Information Miner):**This platform is a widely used open source for data analytics, reporting, and integration. Traditionally used for medicinal research, this commercial analysis tool is now widely used for Educational Data Mining.

• **ORANGE:** Orange is a component-based data mining software suite that is suitable for explorative data analysis, visualization, and predictions. It operates perfectly for various exploration techniques and also aids in scoring and filtering data as a part of the post-processing operation.

CONCLUSIONS

Educational data mining has an effect on many parts of the education organizations and sure to be enormously beneficial in the visualization of realities, predicting student performance, prediction of student performance, grouping and classification of students, predicting students profiling, planning, and development. Despite being an enormously useful exercise, there are challenges in Educational Data Mining. The first and the leading is that educational data derived from numerous sources are evolutionary in nature and is always increasing. Secondly, the storage and the maintenance of the data becomes challenging as well. Also, the concern of the checking of the operational data sources designates the learner's thoughts, aims and its result in a specific learning institution. Another concern is the order and procedure of this dynamic educational data and its knowledge. The situations like appropriating time, context and its sequence must be maintained .Lastly, utilization and exploitation of human, as well as machine resources, must be to the fullest.

REFERENCES

1. S. Harikumar, "A study on educational data mining," International Journal of Computer Trends and Technology, Volume: 8, Issue: 2, pp: 90– 95, 2014.

2. C. Romero and S. Ventura, "Data mining in education," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Volume: 3, Issue: 1, pp: 12–27, 2012.

3. Jain, A. K., Murty, M. N., & Flynn, P. J., "Data clustering: A review," ACM Computing Surveys, Volume: 31, Issue: 3, pp: 264–323, 1999 _

4. Merceron, A., &Yacef, K., "Mining student data captured from a web-based tutoring tool: Initial exploration and results," Journal of Interactive Learning Research, Volume: 15, Issue: 4, pp: 319–346, 2004 _

5. Bekele, R., Menzel, W. "A bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students". Journal of Information Science, 2013

6. SonaliAgarwal, G. N. Pandey, and M. D. Tiwari, Data Mining in Education: Data Classification and Decision Tree Approach, 2012

7. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," Procedia Computer Science, Volume: 72, pp: 414–422, 2015

M. Berland, R. S. Baker, and P. Blikstein, "Educational data mining and learning Analytics: Applications to Constructionist research," Technology, Knowledge and Learning, Volume: 19, Issue: 1, pp: 205–220, 2014. _

9. Cristobal Romero Sebast¿¿an Ventura, "Educational Data Mining: A Review of the State of the Art", IEEE Transactions on system, man and cybernetics-Part C: Application and Reviews, Volume: 40, Issue: 6, pp: 601- 618, 2010 _

10. Bhise R.B., Thorat S.S., Supekar A.K., "mportance of Data Mining in Higher Education System," In IOSR Journal Of Humanities And Social Science (IOSR-JHSS). Volume 6, Issue 6, pp:18-21, 2013 11. Dahiya V, "Improving Teaching and Learning Outcomes: An Outlook on Data Analytics in Education,"Universal Research Reports, Volume: 4, Issue: 7, pp:-146-151, 2017

12. Surjeet Kumar Yadav and Saurabh Pal, "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification" Volume: 2. Issue: 2 2012

13. Devi, S. V. S. G. "A survey on distributed data mining and its trends." International Journal of Research in Engineering & Technology (IMPACT: IJRET) 2.3 (2014): 107-120.

14. Samrat Singh and Dr.Vikesh Kumar, "Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques," Volume: 3 Issue: 2, 2013

*15. Mohammed M.AbuTair and AlaaM.El-Halees, "Mining Educational Data to Improve Students'
Performance: A case Study", International journal of information and Communication Technology
Research, ISSN: 2223-4985, Volume: 2, Issue: 2, 2012*

*16. M.S. Mythili, Dr. A.R.MohamedShanavas, "An Analysis of students ″performance using classification
algorithms", IOSR Journal of Computer Engineering, Volume: 16, Issue: 1, 2014*

*17. Dr. MohdMaqsood Ali, "Role of data mining in education sector," International Journal of Computer
Science and Mobile Computing Volume: 2, Issue: 4, 2013*