IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

# Survey of Educational Data Mining

**P. Padmapriya[1], Dr. Antony Selvadoss Thanamani[2]**

Ph.D (Research Scholar) Computer Science, NGM College, Pollachi[1]

Head of Department, Associate Professor of Computer Science, NGM College[2]

**Abstract:** Educational data mining (EDM) is an emerging discipline that focuses on applying data mining tools and techniques to educationally related data. Applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining. It is concerned with developing methods for exploring the unique types of data that come from educational environments. The discipline focuses on analyzing educational data to develop models for improving learning experiences and improving institutional effectiveness. A literature review on educational data mining topics such as student retention and attrition, personal recommender systems within education, and how data mining can be used to analyze course management system data.

**Keywords:** data mining, educational data mining, academic analytics, learning analytics.

## I. INTRODUCTION

The process of extracting important and useful information from large sets of data is called Data Mining. Educational data mining (EDM) is concerned with developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist.EDM has emerged as a research area in recent years aimed at analyzing the unique kinds of data that arise in educational settings to resolve educational research issues In fact, EDM, can be defined as the application of data mining (DM) techniques to this specific type of dataset that come from educational environments to address important educational questions.

The emerging field of educational data mining (EDM) examines the unique ways of applying data mining methods to solve educationally related problems. There is pressure in higher educational institutions to provide up-to-date information on institutional effectiveness. Institutions are also increasingly held accountable for student success. One response to this pressure is finding new ways to apply analytical and data mining methods to educationally related data.

Even though data mining (DM) has been applied in numerous industries and sectors, the application of DM to educational contexts is limited. Researchers have found that they can apply data mining to rich educational data sets that come from course management systems such as An gel, Blackboard, WebCT, and Moodle. The recent literature related to educational data mining (EDM) is presented. Educational data mining is an emerging discipline that focuses on applying data mining tools and techniques to educationally related data. Researchers within EDM focus on topics ranging from using data mining t o improve institutional effectiveness to applying data mining in improving student learning processes. There is a wide range of topics within educational data mining, so this paper will focus exclusively on ways that data mining is used to improve student success and processes directly related to student learning.

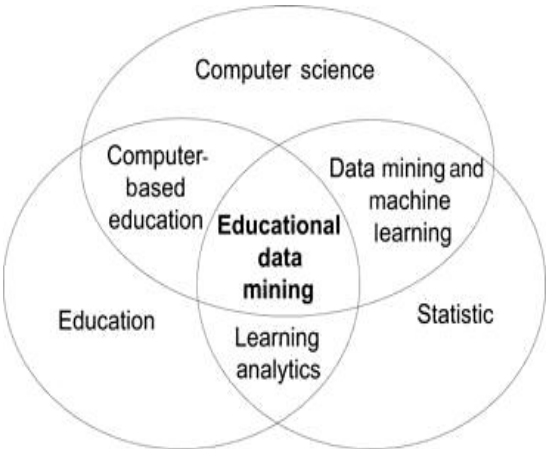



**Figure 1:** Areas related to educational data mining.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

This paper provides an updated overview of the current state of knowledge in EDM with the objective of introducing it to researchers, instructors and advanced students without a strong background in the field. The paper is organized as follows. First, the background of EDM is described. Then the main types of educational environments and their data are shown. The following sections describe the main goals and the specific knowledge discovery process in EDM. Next, the most popular methods used in EDM are presented. Subsequently, some examples of applications or tasks in educational environments and some examples of specific DM tools are listed.

## II. BACKGROUND STUDY OF DATA MINING

Data mining is also one step in an overall knowledge discovery process, where organizations want to discover new information from the data in order to aid in decision-making processes. Knowledge discovery and data mining can be thought of as tools for decision-making and organizational effectiveness. The complexity of data mining has led the data analytics community to establish a standard process for data mining activities.

Data mining has its roots in machine learning, artificial intelligence, computer science, and statistics. There are a variety of different data mining techniques and approaches, such as clustering, classification, and association rule mining. Each of these approaches can be used to quantitatively analyze large data sets to find hidden meaning and patterns. Data mining is an exploratory process, but can be used for confirmatory investigations. It is different from other searching and analysis techniques in that data mining is highly exploratory, where other analyses are typically problem-driven and confirmatory.

While data mining has been applied in a variety of industries, government, military, retail, and banking, data mining has not received much attention in educational contexts. Educational data mining is a field of study that analyzes and applies data mining to solve educationally-related problems. Applying data mining this way can help researchers and practitioners discover new ways to uncover patterns and trends within large amounts of educational data.

## III. BACKGROUND STUDY OF EDUCATIONAL DATA MINING

EDM has emerged as an independent research area in recent years, starting with research in intelligent tutoring systems (ITS), artificial intelligence in education (AIED), user modeling (UM), technology-enhanced learning (TEL), and adaptive and intelligent educational hypermedia (AIEH). Its origins lie in a series of workshops (see Table 1) organized into related conferences that began in 2000. The first workshop, referred to as 'Educational Data Mining', took place in 2005 and culminated in 2008 with the establish-ment of the annual International Conference on Educational Data Mining organized by the International Working Group on Educational Data Mining.

The first conference EDM2008 was held in Montreal, Canada; then EDM2009 in Cordoba, Spain; EDM2010 in Pittsburgh, USA; EDM2011 in Eindhoven, The Netherlands; EDM2012 in Chania, Greece; and the next EDM2013 will be held in Memphis EEUU. There are some other closely related conferences (see Table 2) in which EDM is colocated most years. All of them are older than EDM with the exception of the LAK conference (International Conference on Learning Analytics and Knowledge), which is younger. The first LAK conference, was in Banff, Canada, in 2011 and the second in Vancouver, Canada, in 2012.

There are different ways that educational data mining is defined. Campbell and Oblinger (2007) defined academic analytics as the use of statistical techniques and data mining in ways that will help faculty and advisor s become more proactive in identifying at-risk students and responding accordingly. In this w ay, the results of data mining can be used to improve student retention. Academic analytics focuses on processes that occur at the department, unit, or college and university level. This type of analysis does not focus on the details of each individual course, so it can be said that academic analytics has a macro perspective. Academic analytics can be considered a sub-field of educational data mining.

Educational data mining can draw upon ideas from organizational data mining. Organizational data mining (ODM) focuses on assisting organizations with sustaining competitive advantage (Nemati & Barko, 2004). The key difference between DM and ODM is that ODM relies on organizational theory as a reference discipline (Nemati & Barko, 2004). Organizations that transform their data into useful information and knowledge, and do so efficiently, should gain tremend us benefits such as enhanced decision-making, increased competitiveness, and potential financial gains (Nemati & Barko, 2004). Therefore, the EDM field draws upon organizational theory as well. This is an important relationship because the focus of research within EDM ca n examine phenomena at different levels of analysis, from societal, organizational, unit, or individual level.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

The type of research don e within EDM focuses primarily on quantitative analyses, which is necessary because data mining employs statistics, machine learning, and artificial intelligence techniques. Many of the studies presented in this literature review are case studies where data mining projects were done at a specific institution, with a single institution's data. Qualitative techniques such as interviews and document analysis are also used to support case studies in EDM.

The dominant research paradigm is quantitative, with results coming in th e form of predictions, clusters or classifications, or associations. The drawback with some of the existing case studies is that the results are not necessarily generalizable to other institutions. This means that the results are highly associated with a specific institution at a specific time. Research in EDM should examine ways for data mining results to be more generalizable.

## IV. APPLICATIONS OF DATA MINING

A review of related literature in educational data mining follows. It focuses on how data mining is used for improving student success and processes directly related to student learning. Educational data mining research examines different ways that course management systems (CMS) data can be mined to provide new patterns of student behavior. Results can assist faculty and staff with improving learning and supporting educational processes, which in turn improve institutional effectiveness.

### Student Retention and Attrition

Research has shown that data mining can be used to discover at-risk students and help institutions become much more proactive in identifying and responding to those students (Luan, 2002). Applied data mining as a way to predict what types of students would drop out of school, and then return to school later on. He applied classification and regression trees (C&RT) – a specific data mining technique – to educational data in order to predict which students are unlikely to return to school. In this case study, Luan applied both quantitative and qualitative research techniques t o uncover student success factors. This research is important because it demonstrated the successful application of data mining tools to assist in student retention efforts. As noted earlier, the case study method for EDM may often produce results that are not generalizable. However, the process by which researchers apply the data mining can be generalized and used in other contexts. It is simply the results of the data mining models that may not be generalized.

Data mining was used to assess the efficacy of a writing center in an effort to analyze student achievement and student progress to the next grade (Yeats, Reddy, Wheeler, Senior, & Murray, 2010). Their work demonstrated the ability to assess a specific educational support process, i.e., the writing center, in an effort to improve institutional effectiveness. Their research approach used a combination of quantitative work and case study analysis. The mixed-methods approach to data mining was helpful in understanding much more about the ways data mining can be used in an actual implementation. Their research results were not surprising in that it found students who attend writing centers tend to do better in their classes. The research by Yeats et al. (2010) took a different approach to analyzing student achievement in that it made the connection between writing center attendance and student grades. It did not make the link to student retention issues, but a future study could examine the relationship between these three concepts: writing center attendance, student grades, and retention.

Academic performance and student success can be predicted by using data mining techniques. One research team used data mining to classify students into three groups as early as they could in the academic year (Vandamme, Meskens, & Superby, 2007). The three groups included low-risk, medium risk, and high-risk students. The authors used several data mining techniques including neural networks, random forests, and decision trees. The student in the high risk group had a high probability of failing or dropping out of school. These types of studies are important in that they give faculty and staff a way to identify the at-risk students in a proactive way, because "once a student decides to leave, it is hard to convince them to stay".

### Personal Learning Environments and Recommender Systems

Personal learning environments (PLEs) and personal recommendation systems (PRS) also directly relate to educational data mining. Personalized learning environments focus on providing the various tools, services, and artifacts so that the system can adapt to students' learning needs on the fly (Mödri tscher, 2010). Much of the work done related to recommender systems is quantitative and is widely used in eCommerce. For example, Amazon .com uses recommender systems in order t o customize the browsing experience for each user. Recommendations display related products that a consumer might purchase. Netflix also employs recommender systems to help its subscribers find the types of movies t hat they will probably like.

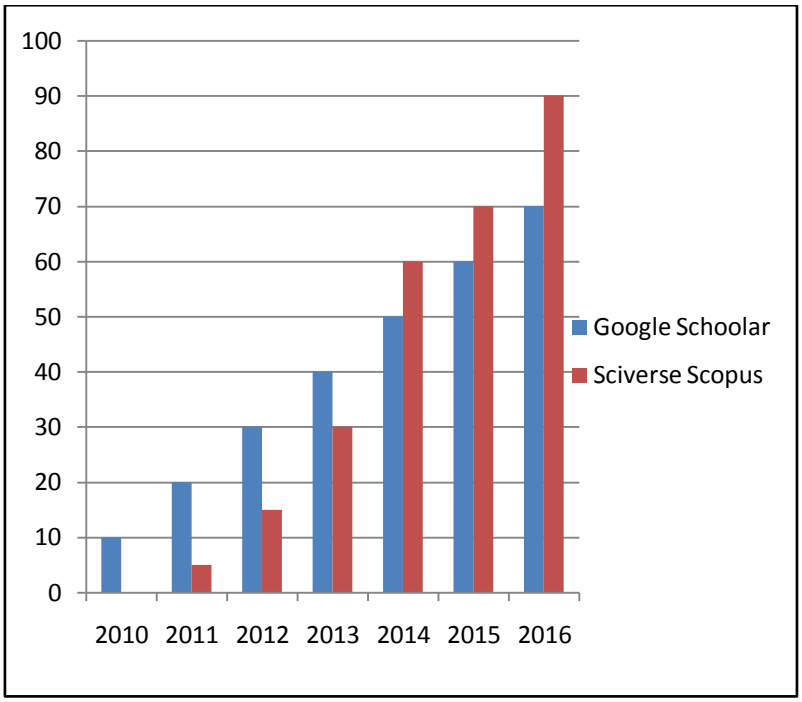
Recommender systems must be adapted when they are used in educational contexts because the recommendations should coincide with educational objectives. The reason is that it is not possible to apply existing recommender systems

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

directly to educational data because they are highly domain dependent. There are two significant challenges with respect to applying recommender systems in an educational context. First, the system must attempt to understand or determine the needs of learners. Second, there should be some way for faculty members to control recommendations for their learners. Existing recommender systems in the educational domain typically do not address these concerns, which open up additional research opportunities for the EDM research community.




**Graph 1:** Number of educational data mining references in Google Scholar and Sciverse Scopus by year.

Recommendations for further learning exercises were made based on a student's web browsing behavior and improved student achievement. A data mining model was established that annotated browsing events with contextual factors, to produce new individualize d content recommendations specifically for course management systems. The results showed that data mining can deliver highly personalized content, based on browsing history and history of student achievement. This also improved student learning because students could move through the material at their own pace. The researchers also discovered that the contextual browsing model is much more effective than using association rule mining models.

## V. CONCLUSION

Educational data mining (EDM) is an area full of exciting opportunities f or researchers and practitioners. This field assists higher educational institutions with efficient and effective ways to improve institutional effectiveness and student learning. Data mining is a significant tool for helping organizations enhance decision making and analyzing new patterns and relationships among a large amount of data.EDM brings together an interdisciplinary community of computer scientists, learning scientists, psycho metricians, and researchers from other fields. EDM applies techniques coming from statistics, machine learning, and data mining to analyze data collected during teaching and learning, tests learning theories, and informs decision-making in educational practice.

A broad sense of the types of research currently being conducted in EDM was presented, from app lying data mining for understanding student retention and attrition to finding new ways of making personalized learning recommendations to each individual student. Many opportunities exist to study EDM from an organizational unit of analysis to individual course-levels of analysis. Some work is strategic in nature and some of the research is extremely technical.

## REFERENCES

1. Baker, R., & Yacef, K. (2009). T he State of Educational Data mining in 2009: A Review and Future Visions. Journal o f Educational Data Mining, 1(1).

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

2. Koedinger K, Cunningham K, Skogsholm A, Leber B. An open repository and analysis tools for fine-grained, longitudinal learner data. In: First Interna-tional Conference on Educational Data Mining. Mon-treal, Canada; 2008, 157–166.
3. Mostow J, Beck J. Some useful tactics to modify, map and mine data from intelligent tutors. J Nat Lang Eng 2006, 12:195–208.
4. Bala M, Ojha DB. Study of applications of data mining techniques in education. International J Res Sci Tech-nol 2012, 1: 1–10.
5. Romero C, Ventura S, Pechenizky M, Baker R. Hand-book of Educational Data Mining. Data Mining and Knowledge Discovery Series. Boca Raton, FL: Chap-man and Hall/CRC Press; 2010.
6. Blikstein, P. (2011). Using learning analytics to assess students' behavior in ope n-ended programming tasks. Pape r presented at the Proceedings of the 1st Interna tional Conference on Learning Analytics and Knowledge, Banff, Alberta, Canada.
7. Calders, T., & Pechenizkiy, M. (2012). Introduction to the special section on ed ucational data mining. SIGKDD Explor . Newsl., 13(2), 3-6. doi: 10.1145/2207243.2207 245
8. Campbell, J., & Oblinger, D. (20 07). Academic analytics. Washington, DC: Edu cause. Chacon, F., Spicer, D., & Valbu ena, A. (2012). Analytics in Support of Student Retention and Success (Research Bulletin 3, 2012 ed.). Louisville, CO: Educause Cente r for Applied Research.
9. Thomas, E. H., & Galambos, N. (2004). What Satisfies Students?: Mining Stude nt-Opinion Data with Regression and Decision Tree Analysis. Research in Higher Educat ion, 45(3), 251-269
10. Wang, Y.-h., & Liao, H.-C. (2011). Data mining for adaptive learning in a TESL -based e-learning system. Expert S ystems with Applications, 38(6), 6480-6485. doi: 10.1016/j.eswa.2010.11.098

## BIOGRAPHIES

**Padmapriya.P** received her BSC IT (Information Technology) From Dr. MCET Pollachi, India. She Completed her Master Degree MCA from Karpagam Engineering College, Coimbatore. Currently she is a Research Scholar at Department of Computer Science, NGM College, Pollachi, India. She participated in a International Conference. Her area of interest includes Data mining, Big Data, Web Mining.

**Dr. Antony Selvadoss Thanamani** is presently working as Professor and Head, Research Department of Computer Science, NGM College, Pollachi, Coimbatore, India. He has published more than 100 papers in international/national journals and conferences. He has authored many books on recent trends in Information Technology. His areas of interest include E-Learning, Knowledge Management, Data Mining, Networking, Parallel and Distributed Computing. He has to his credit 25 years of teaching and research experience. He is a senior member of International Association of Computer Science and Information Technology, Singapore and Active member of Computer Science Society of India, Computer Science Teachers Association, New York.