EDUCATIONAL DATA MINING

[1]Dr.M.RATHAMANI MCA.,MPhil.,Ph.D., and [2]K. SAVUMIA

[1]Assistant Professor, Department Of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, India.

Email:rm32233@gmail.com

[2]Department Of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, India.

Email:savumiakarthikeyan@gmail.com

ABSTRACT

Data mining methods and techniques are applied to educationally related data as part of the burgeoning field known as educational data mining (EDM). Education data mining is the name of an emerging multidisciplinary research field that applies data mining (DM) to the field of education. Its focus is on creating tools for examining the distinctive kinds of data that emerge from educational settings. The discipline's main goal is to analyse educational data and create models that will enhance learning opportunities and institutional effectiveness. a survey of the literature on issues related to personal recommender systems in education, student retention and attrition, and the use of data mining to examine course management system data.

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

INTRODUCTION

Data mining is the process of obtaining significant and practical information from massive collections of data. The goal of educational data mining (EDM) is to create, investigate, and use automated methods to find patterns in massive collections of educational data that would otherwise be difficult or impossible to evaluate owing to the sheer volume of data they exist within. EDM is a new field of study that aims to analyse the special types of data that emerge in educational settings to address problems with educational research. EDM is actually the practise of applying data mining (DM) techniques to a certain kind of dataset that originates from educational environments to tackle crucial educational issues.

In the newly growing discipline of educational data mining (EDM), challenges relating to education are solved in novel ways by using data mining techniques. Higher education institutions are under pressure to offer current data on institutional effectiveness. Institutions are being held more and more liable for the achievement of their students. Finding novel approaches to use analytical and data mining techniques to educationally related data is one way to deal with this strain.

Although data mining (DM) has been used in many different fields and industries, it hasn't been widely used in educational settings. Rich educational data sets from course management systems like Blackboard, Angel, WebCT, and Moodle have been discovered by researchers to be amenable to data mining. It presents the most recent research on educational data mining (EDM). An emerging field of study called educational data mining focuses on
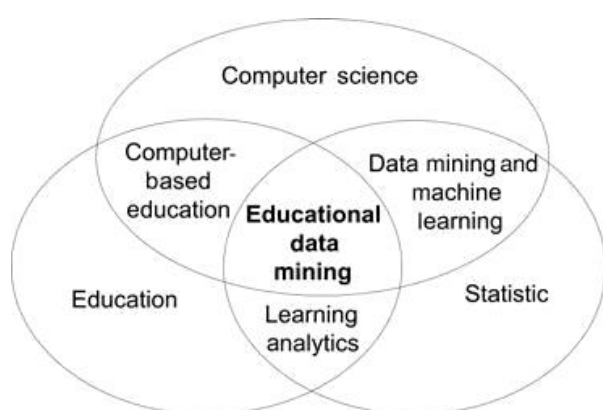

Figure 1: Areas related to educational data mining.

In order to introduce EDM to researchers, teachers, and advanced students without a

BACKGROUND STUDY OF DATA M INING

Data mining is also a step in the overall knowledge discovery process used by businesses to extract new knowledge from data and use it to support decision-making. Data mining and knowledge discovery can be seen as decision-making and organisational effectiveness tools. The data analytics community has established a standard procedure for data mining activities as a result of the complexity of

using data mining tools and techniques on educationally relevant data. Researchers Focus areas for EDM include everything from using data mining to increase institutional effectiveness to using it to enhance student learning processes. While there are many different themes in educational data mining, this essay will only discuss how data mining is applied to enhance student achievement and procedures directly associated with student learning.

solid experience in the area, this paper offers an updated review of the field's present state of knowledge. The structure of the essay is as follows. The history of EDM is first discussed. The major categories of educational environments are then displayed together with their statistics. The main objectives and the specific knowledge discovery process in EDM are described in the following sections. The most common EDM techniques are then discussed. Then, several examples of tasks or applications in educational settings are shown, along with some instances of particular DM tools.

data mining.

Machine learning, artificial intelligence, computer science, and statistics are the foundations of data mining. Data mining methods and techniques come in many forms, including clustering, classification, and association rule mining. A huge data set can be quantitatively analysed using any of these methods to uncover hidden patterns and significance. Although it is an exploratory technique, data mining can also be utilised for confirmatory research. Data mining is

distinct from other methods of searching and analysing since it is mostly exploratory, whereas other analyses are frequently problem-driven and confirmatory.

Data mining has been used in many other sectors, including government, the military, retail, and banking, but it hasn't gotten much attention in educational settings. The subject of educational data mining involves analysing and using data mining to address issues in education. By using data mining in this way, researchers and practitioners can find novel approaches to sift through vast amounts of educational data to find patterns and trends.

BACKGROUND STUDY OF EDUCA TIONAL DATA MINING

Beginning with research on adaptive and intelligent educational hypermedia, intelligent tutoring systems (ITS), artificial intelligence in education (AIED), user modelling (UM), and technology-enhanced learning (TEL), EDM has become a stand-alone study field in recent years (AIEH). Its roots can be found in a series of seminars (see Table 1) that were brought together into linked conferences and started in 2000. The inaugural "Educational Data Mining" workshop took place in 2005, and the International Working Group on Educational Data Mining established the biennial International Conference on Educational Data Mining in 2008.

The inaugural conference, EDM2008, took place in Montreal, Canada. Subsequent conferences followed in Cordoba, Spain, Pittsburgh, USA, Eindhoven, The Netherlands, Chania, Greece, and Memphis, United States. EDM typically co-locates with a few other conferences that are closely related (see Table 2). With the exception of the LAK conference (International Conference on Learning Analytics and Knowledge), all of them are older than EDM. The inaugural LAK conference took place in Canada's Banff in 2011 and the second one in Canada's Vancouver in 2012.

There are different ways that educational data mining is defined. Campbell and Oblinger (2007) defined academic analytics as the use of statistical There are various definitions for educational data mining. Academic analytics, as described by Campbell and Oblinger (2007), is the application of statistical methods and data mining to enable professors and advisors to be more proactive in detecting at-risk students and taking appropriate action. In this way, data mining findings might be applied to boost student retention. Academic analytics is concerned with activities that take place within departments, units, or colleges and universities. This form of analysis has a global viewpoint because it does not concentrate on the specifics of each particular course. It is possible to classify academic analytics as a branch of educational data mining.

Organizational data mining can provide inspiration for educational data mining. The goal of organisational data mining (ODM) is to help businesses maintain their competitive advantage (Nemati & Barko, 2004). The primary distinction between ODM and DM is that ODM uses organisational theory as a discipline of reference (Nemati & Barko, 2004). Companies that effectively turn their data into knowledge and information should reap tremendous advantages like improved decision-making, increased competitiveness, and significant financial

rewards (Nemati & Barko, 2004). Hence, organisational theory is also used in the

EDM sector. This connection is significant because EDM research can explore phenomena at a range of analytical levels, including social, organisational, unitary, and individual levels.

Data mining uses statistical, machine learning, and artificial intelligence approaches, hence the research done within EDM generally focuses on quantitative analysis. Several of the studies discussed in this literature review are case studies of data mining projects that were carried out at particular institutions using data from just those institutions. In order to supplement case studies in EDM, qualitative methods like document analysis and interviews are also used.

The most common study paradigm is quantitative, and outcomes typically take the form of connections, clusters, or classifications. The problem with some of the case studies now in use is that the conclusions may not necessarily apply to other institutions. This indicates that the findings are strongly linked to a particular institution at a particular moment. EDM research should look towards how to make data mining results more generalizable.

APPLICATIONS OF DATA M INING

The next step is a survey of related academic research in educational data mining. It emphasises the use of data mining in activities that directly affect student learning and student achievement. Course management system (CMS) data can be mined in a variety of ways to uncover novel patterns of student behaviour, according to studies on

educational data mining. Findings can help instructors and staff members

enhance instruction and other educational procedures, which enhances institutional performance.

Student Retention and Attrition

Data mining can be used to identify at-risk students and assist schools in becoming much more proactive in identifying and assisting those individuals, according to research (Luan, 2002). used data mining to forecast the types of students who would leave their courses. return to school later after a break. To determine which children are most likely to return to school, he used educational data and the data mining approach known as classification and regression trees (C&RT). In this case study, Luan used both quantitative and qualitative research methods to identify the elements that contribute to student achievement. This study is significant since it showed how data mining technologies can be effectively used to support attempts to retain students. As previously mentioned, the case study method for EDM frequently yields conclusions that cannot be broadly generalised. Yet, the method by which academics use data mining is generalizable and applicable in different situations. Simply put, the outcomes of the data mining models might not be generalizable.

The effectiveness of a writing centre was evaluated using data mining in an effort to examine student accomplishment and progress to the following grade (Yeats, Reddy, Wheeler, Senior, & Murray, 2010). In an effort to increase institutional performance, their work proved the capacity

to evaluate a specific educational assistance procedure, namely the writing centre. They combined quantitative work and case study analysis in their research strategy. The mixed-methods approach to data mining helped us learn a lot more about the practical applications of the technique.

The findings of their study, which showed that students who use writing centres typically perform better in their classes, were not particularly surprising. Yeats et al(2010) .'s study used a unique method for examining student achievement because it linked participation in the writing centre with academic performance. Although the study did not make a connection between these three ideas—attendance at the writing centre, student grades, and retention—it might in the future.

Using data mining approaches, academic achievement and student success can be predicted. To classify students into three categories as early as possible in the academic year, one research team used data mining (Vandamme, Meskens, & Superby, 2007). Low-risk, medium-risk, and high-risk kids made up the three categories. The authors employed a number of data mining methods, including decision trees, random forests, and neural networks. The likelihood of a kid in the high risk group failing or leaving school was very high. These studies are crucial because they provide teachers and staff with a proactive technique to spot the students who are at risk for dropping out since "once a student decides to leave, it is hard to convince them to stay."

Personal Learning Environments and Recommender Systems

Personal recommendation systems (PRS) and personalised learning environments (PLEs) are also intimately related to educational data mining. In order for the system to instantly adjust to students' learning needs, personalised learning environments put a strong emphasis on offering a variety of tools, services, and artefacts (Mödri tscher, 2010). The majority of research on recommender systems is quantitative and is used extensively in e-commerce. For instance, Amazon.com uses recommender systems to personalise each user's browsing experience. Suggestions show comparable goods that customers might buy. In order to help its users find the kinds of movies they will likely enjoy, Netflix also uses recommender algorithms. Because the recommendations should be in line with educational objectives, recommender systems must be modified for usage in educational situations. Because existing recommender systems are heavily domain based, it is not possible to directly apply them to educational data. The use of recommender systems in a learning environment presents two key issues. The system must first make an effort to comprehend or identify the demands of learners. Second, faculty members ought to have some control over the recommendations made to their students. These issues are often not addressed by existing recommender systems in the educational space, which presents further study opportunities for the EDM research community.

On the basis of a student's online browsing habits and increased student accomplishment, suggestions for additional learning activities were produced. A data mining methodology was developed to provide new, personalised content recommendations exclusively for course management systems by annotating browsing events with contextual information. Based on surfing history and prior academic performance, the findings demonstrated that data mining may give

highly tailored content. As a result of being able to read through the content at their own leisure, this also boosted student learning. The contextual browsing approach is also significantly more efficient than association rule mining models, according to the researchers.

CONCLUSION

EDM is a field that offers many intriguing potential for researchers and practitioners. In order to increase institutional performance and student learning, this sector helps higher education institutions in practical and efficient methods. Data mining is a crucial technique

The research being done in EDM is diverse, ranging from using app-lying data mining to assess student retention and attrition to developing new techniques for providing individualised learning recommendations to each individual student. There are numerous options to investigate EDM, ranging from an examination of the organisational unit to an analysis of the individual course levels. Some of the study is of a strategic character, while other research is more technical.

for assisting businesses in improving decision-making and examining fresh patterns and connections among a lot of data. Computer scientists, learning scientists, psychometricians, and scholars from various domains come together in the interdisciplinary community of EDM. EDM utilises emerging technologies from analysis of data gathered from teaching and learning, testing of learning theories, and decision-making in educational practise using statistics, machine learning, and data mining.

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).

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. (2007). Academic analytics. Washington, DC: Edu cause. Chacon, F., Spicer, D., & Valbu ena, A. (2012). Analytics in Support of Student Retention and Success. Louisville, CO: Educause Cente r for Applied Research