ISSN–2278-5655

AMIERJ

Aarhat Multidisciplinary International Education Research Journal

Volume–IX, Issues– VI Nov-Dec 2020

AN ANALYTICAL VIEW ON EDUCATIONAL DATA MINING

Kunal Manohar Bhamare

Class T.Y. Bsc.IT

UG Student, Department of Information Technology, B.K Birla College of Commerce & Science (Autonomous), Kalyan, Maharashtra, India

Abstract:

Educational Data Mining (EDM) is an emerging field to explore the data from various educational contexts. It provides inherent knowledge about imparting the education, which is used to enhance the quality of teaching and learning. Effective planning can provide personalized education. This paper presents a survey on various components of educational data mining along with its objectives

Keywords: *Data Mining, Educational Data Mining, Knowledge Discovery, Component.*

INTRODUCTION:

One of the primary goals of any educational system is to equip students with the knowledge and skills needed to transition into successful careers within a specified period. How effectively global educational systems meet this goal is a major determinant of both economic and social progress. Data mining, in general, can be described as the systematic processing of large datasets and finding hidden facts and patterns. The purpose of data mining is to understand the data trends and develop new and effective insights. This helps with future pacing and provides a vision. If we look at the recent past; the education sector has seen an extreme vertical as wellas horizontal growth. This kind of enormous growth brings along an elevation in educational data. This is where the educational data mining becomes an important aspect of the analysis and further development in the field of education. Hence, applications of technological tools are used to understand the patterns and trends of the educational data. Educational data mining is imperative today; otherwise, it would not be possible to analyze the gigantic amount of educational data. With educational data mining, the first and the foremost task is to organize and categorize the educational data received from various sources. The purpose of introducing automation in the education is to more efficiently understand the evidence-of-learning. It is an ever-evolving process that continually proceeds towards the formation and development of core values of any educational institute. These values are mostly to nurture the talent and capabilities of the students. A part of these educational task is to address students' matters with the proactive expertise to meet learning and management objectives. This task can divide into two types: *student's based task*-effective support of primary stakeholders to fulfill learning objectives and *decision-making task*- for the constant involvement of the hybrid group of stakeholders to fulfill management-oriented objectives.

OBJECTIVES OF THE STUDY:

1. To predicting learner’s behaviors by improving student models.
2. To discovering or improving knowledge domain structure models.
3. To identifymost effective pedagogical support for student learning that can be achieved through learning systems.
4. To identify core influential components of learning to enable the designing of better learning systems.

LITERATURE REVIEW:

Many investigations have been carried out to demonstrate the importance of the "Data Mining" techniques in education, demonstrating that this is a new concept for the purpose of extracting valid and accurate information about the behavior and effectiveness in the learning process [10][11].

In the field of education techniques "Data Mining" has also been used to analyze the curriculum and subject of the current research topics, as well as to analyze the students’ performance [12]. There have been several investigations made under this proposed study object. For example, Bhardwaj used the Naïve Bayes algorithm to predict student performance based on 13 variables [13]. The results were used to build a model that is used to predefine the students who are at risk of failure and thus activate a guidance and counselling program. Varghese, Tommy and Jacob [14] in their research used the "K means" algorithm to cluster 8000 students based on five variables (input average in the University average scores of the tests / exams, average scores of papers, seminars notes and notes the work by frequency). The results showed a strong relationship between attendance and student performance. Gulati and Sharma [15] claim that knowledge through analysis by "Data Mining" can improve the education system in orientation, student performance and organizations management.

Many investigations have been carried out to demonstrate the importance of the "Data Mining" techniques in education, demonstrating that this is a new concept for the purpose of extracting valid and accurate information about the behaviour and effectiveness in the learning process [10][11].

In the field of education techniques "Data Mining" has also been used to analyze the curriculum and subject of the current research topics, as well as to analyze the students’ performance [12]. There have been several investigations made under this proposed study object. For example, Bhardwaj used the Naïve Bayes algorithm to predict student performance based on 13 variables [13]. The results were used to build a model that is used to predefine the students who are at risk of failure and thus activate a guidance and counselling program. Varghese, Tommy and Jacob [14] in their research used the "K means" algorithm to cluster 8000 students based on five variables (input average in the University average scores of the tests / exams, average scores of papers, seminars notes and notes the work by frequency). The results showed a strong relationship between attendance and student performance. Gulati and Sharma [15] claim that knowledge through analysis by "Data Mining" can improvethe education system in orientation, student performance and organizations

Management. Many investigations have been carried out to demonstrate the importance of the "Data Mining" techniques in education, demonstrating that this is a new concept for the purpose of extracting valid and accurate information about the behavior and effectiveness in the learning process [10][11].

In the field of education techniques "Data Mining" has also been used to analyse the curriculum and subject of the current research topics, as well as to analyze the students' performance [12]. There have been several investigations made under this proposed study object. For example, Bhardwaj used the Naïve Bayes algorithm to predict student performance based on 13 variables [13]. The results were used to build model that is used to predefine the students who are at risk of failure and thus activate a guidance and counselling program. Varghese, Tommy and Jacob [14] in their research used the "K means" algorithm to cluster 8000 students based on five variables (input average in the University average scores of the tests / exams, average scores of papers, seminars notes and notes the work by frequency). The results showed a strong relationship between attendance and student performance. Gulati and Sharma [15] claim that knowledge through analysis by "Data Mining" can improve the education system in orientation, student performance and organizations management. Many investigations have been carried out to demonstrate the importance of the "Data Mining" techniques in education, demonstrating that this is a new concept for the purpose of extracting valid and accurate information about the behavior and effectiveness in the learning process [10][11].

In the field of education techniques "Data Mining" has also been used to analyze the curriculum and subject of the current research topics, as well as to analyze the students' performance [12]. There have been several investigations made under this proposed study object. For example, Bhardwaj used the Naïve Bayes algorithm to predict student performance based on 13 variables [13]. The results were used to build model that is used to predefine the students who are at risk of failure and thus activate a guidance and counselling program. Varghese, Tommy and Jacob [14] in their research used the "K means" algorithm to cluster 8000 students based on five variables (input average in the University average scores of the tests / exams, average scores of papers, seminars notes and notes the work by frequency). The results showed a strong relationship between attendance and student performance. Gulati and Sharma [15] claim that knowledge through analysis by "Data Mining" can improve the education system in orientation, student performance and organizations management. Many investigations have been carried out to demonstrate the importance of the "Data Mining" techniques in education, demonstrating that this is a new concept for the purpose of extracting valid and accurate information about the behavior and effectiveness in the learning process . In the field of education techniques "Data Mining" has also been used to analyze the curriculum and subject of the current research topics, as well as to analyze the students' performance. There have been several investigations made under this proposed study object. For example, Bhardwaj used the Naïve Bayes algorithm to predict student performance based on 13 variables. The results were used to build a model that is used to predefine the students who are at risk of failure and thus activate a guidance and counselling program. Varghese, Tommy and Jacob in their research used the "K means" algorithm to cluster 8000 students based on five variables (input average in the University average scores of the tests / exams, average scores of papers, seminars notes and notes the work by frequency). The results showed a strong relationship between attendance and student performance. Gulati and Sharma claim that knowledge through analysis by "Data Mining" can improve the education system in orientation, student performance and organizations management.

Limitations of this research

This survey work studied around 50 EDM research papers from various journals/conferences of repute in the context of DM techniques/methods, Tools, citation nos, Dataset used, educational outcomes, useful commercial open sources/ open access tools with their features, data set and links. Since it is not possible to cover all the research papers, from all corners and explores eacLimitations of this research This survey work studied around 50 EDM research papers from various journals/conferences of repute in the context of DM techniques/methods, Tools, citation no's, Dataset used, educational outcomes, useful commercial / open sources/ open access tools with their features, data set and links. Since it is not possible to cover all the research papers, from all corners and explores eac

RESEARCH METHODOLOGY:

The research is completely based on secondary data which is collected through the published sources, magazines, journals, websites and books.

LIMITATIONS:

This survey work studied around 50 EDM research papers from various journals/conferences of repute in the context of DM techniques/methods, Tools, citation no's, Dataset used, educational Outcomes, useful commercial / open sources/ open access tools with their features, data set and Links. Since it is not possible to cover all the research papers, from all corners and explores each and every mentioned tools with their functional points, popular tools, techniques and most cited Research papers were discussed which may be considered as representatives of this research area. The features discussed in this work are comprehensive rather than inclusive.

COMPONENTS OF EDUCATIONAL DATAMINING:

Educational data mining touches and affects numerous aspects of the education industry. The major components of EDM are - stakeholders of education, various data mining tools and techniques, educational data, educational environment and task and how they meet the educationalobjectives.

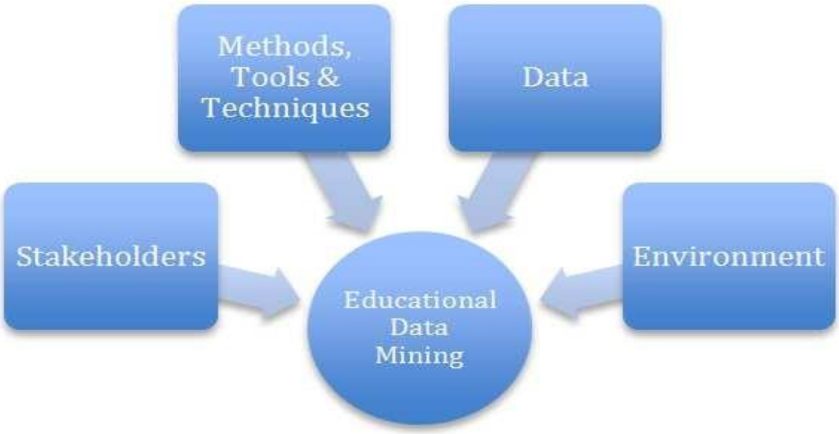

Figure 1: Components of Educational Data Mining

Stakeholders:

Keeping in mind all aspects of education, i.e. primary to higher education, stakeholders of the education can be majorly categorized into the following:

- **Learners/Students:** The most important and impacted component is the learners. As students are involved directly in the process of learning, they fall in the primary group of stakeholders. EDM can help them with the personalized education based on various recommendations and can increase the interestingness of education for students towards learning. Different learning tasks can be formulated in the different group of students based on their needs.
- **Faculties:** Educators, Teachers, and Instructors are benefitted as they can determine which student require extra support. The prediction of student performance becomes easy. Another impact is it helps in the classification of learners into groups. It can also provide an insight into the patterns in which students can learn- Regular and irregular. Teachers can analyze the data and determine the most commonly made errors. Beyond just academics, the analysis of the student's learning and behavior can also be done to detect if they require any extra support during the process of learning. Teachers are also the primary stakeholders.
- **Parents:** Parents are the part of the secondary group. They are liable for helping their kids to get them enroll in the most suitable courses for them.
- **Course Researchers and Educational Developers:** They are the people who design and modify the course. They are responsible for the growth of education. Developers fall into the group of secondary stakeholders.
- **Administrators:** They can also be called as the hybrid users. EDM is useful for effective utilization of resources; it can help in determining what are the offers that can capture more pupils into various programs and courses. They are responsible for various administrative decisions such as infrastructure development and employing the expertfaculty.

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 essays and other generic descriptive data Such information from a completely digital process can actually be divided into two types of information. These two categories are structured data and unstructured data.

Structured Data: This data is already organized and leave a lesser possibility of being too vast and too vague. As a result, the data is self- explanatory and more regulated as compared to unstructured data. At the same time, this makes the data free from human intervention and brings in crystal-clear evidence away from prejudices and judgments. Several sources of collecting structured data that could be termed as capable of incorporating embedded and formative assessment are as follows-

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

Unstructured Data: This kind of data does not come from one specific source and there is no predefined data model. This may include learning information that comes from the web. This includes information such as a learner’s IP address or her/his username and may relate to various texts from sources like Internet forums, video clips, or audio files. Here are some of the potential sources from where unstructured data can be collected from-

- Learning games
- Social interaction analyses
- Affect meters
- Body sensors

As can be seen, that educational data are in abundance, and this makes educational data mining an important exercise that can enhance the learning development across all verticals of the education industry. All this collected, structured and unstructured data when analyzed will be of great help when it comes to meeting the objectives and to determine specific goals of education. The data could be both assorted and classified too. The data can be generated from **online and offline sources-**

- **Offline Data:** As the name suggests, offline data are generated through real-time situations and settings. Setups like traditional classroom tests, interaction based contemporary classes, teacher-student interactions, student-to-student interactions, real-time data derived from different courses and various departments of any institute like schools, colleges, and universities. Other factors are levels of participation 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 data are derived from weblogs, E-mails, spreadsheets, transcripts telephonic conversations, medical records, legal information, and publication databasesetc.

Data Mining Methods:

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

- **Classification:** Classification is training and testing technique, which categorizes the collected data into some preset groups. It is a useful technique for predicting student performances, risk analysis, student monitoring systems, and detection of errorsetc.
- **Clustering:** Like classification, this technique puts similar data together into clusters, but not under preset categories. This technique is helpful in differentiating the preferences of different learners. Analysis of students’ comprehensive character and techniques suitable for collaborative learning are done with the help of clustering.
- **Statistics:** Statistics are useful for course management system and assists in the determination of extreme 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 industry trends. It is a technique that is utilized to predict success rate, dropout rate and designing methods of retention.
- **Association Rule Mining**: This is an important data mining technique in finding various relations among the attributes of data, such as admission, migration, parents-faculty-students relation etc. various patterns for reasons of student’s failure can be found out.

The other effective data mining methods used in the EDM industry are neural networks, regression, SVM etc.

Useful Tools of Educational Data Mining

- **WEKA (Waikato Environment for Knowledge Analysis):** The Weak 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 methods in extracting information about learning and knowledge.
- **R (Revolution):** This is a statistical computing software/ language that is widely used by data miners to perform statistics for learning development solutions. R is an extremely versatile tool that is not only scientifically designed but is also easy to use. So, applying stats and formulas in R are convenient.
- **KNIME (Konstanz Information Miner):** This platform is a widely used open source for data analytics, reporting, and integration. Traditionally used for pharmaceutical research, this business 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.

CONCLUSION:

Educational data mining has an impact on many parts of the education industry and bound to be extremely beneficial in the visualization of facts, predicting student performance, prediction of student performance, grouping and categorization of students, predicting students profiling, planning, and scheduling. Despite being an extremely useful exercise, there are challenges in Educational Data Mining. The first and the foremost is that educational data derived from various sources are evolutionary in nature and is always increasing. Secondly, the storage and the maintenance of the data becomes challenging as well. Data mining is a powerful analytical tool to enhance decision making and analyzing new patterns and relationships for organizations. And EDM contains techniques including data mining, statistics, machine learning. DM need to analyze data coming from teaching and learning, tests learning theories, and policy decision-making etc.

REFERENCES:

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

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.

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

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

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

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.

Cristobal Romero Sebastian 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.