



Improving Teaching and Learning Outcomes: An Outlook on Data Analytics in Education

Vandna Dahiya,
Maharshi Dayanand University, Haryana
vandanadahiya2010@gmail.com

Abstract

Educational data mining (EDM) is the term used for analyzing the data originating from the educational contexts with the prime aim of improving the quality and effectiveness of education for the overall development of students. Educational data is evaluated using various computational methods to better understand the students and their learning environment. The rising use of technology in educational systems has led to the storage of large amounts of student data, which makes it significant to use EDM to improve teaching and learning processes. Education helps in building social skills and enhances the problem-solving skills for an individual.

EDM is useful in many different areas including identifying at-risk students, priority learning needs for different groups of students, increasing graduation rates, effectively assessing institutional performance, maximizing campus resources, and optimizing subject curriculum renewal. The aim of this study is to present the importance of data mining in education and implementing automated systems in education to enhance the learning of students.

ISSN : 2348-5612 © URR



Keywords: Data mining; Data management, Educational data mining; Teaching and learning

I. Introduction

One of the primary goals of any educational system is to equip the students with the knowledge and skills needed to transition into successful human beings. Education helps the individuals in building various abilities for their overall growth. With the growing number of schools, colleges and coaching institutes, education has achieved a new dimension. How effectively global educational systems meet the goals of education is a major determinant of both economic and social progress. Some countries provide free education for all citizens from grade one through the university years. Therefore, a large number of students enter universities every year. It has become difficult to provide high-quality teaching and guidance to such a large number of students. As a result, many students fail to complete their degrees within the required periods. Educational data mining (EDM) can present a clear picture of specific hindrances to student learning. For example, students can fail in advanced subjects because they did not learn the basic information from the prerequisite subjects. Using data mining (DM) methods to examine student information can help classify possible reasons for student disappointments. Data mining provides many methods for information investigation. A huge quantity of information currently residing in student records overdoes the human ability to examine and extract the most useful information without help from automatic analysis methods. Knowledge discovery is the procedure of nontrivial extraction of implied, unidentified, and theoretically useful information from a large database. DM is a powerful artificial intelligence tool, which can ascertain useful information by exploring data from many viewpoints or extents, categorize that information, and review various relations acknowledged in the records. Subsequently, this information helps make or improve decisions. In DM, algorithms can be used either independently or together to achieve the desired results. Some algorithms can explore data; others extract a specific outcome based on that data. A large amount of data in databases today exceeds the human ability to analyze and extract the most useful information without help from automated analysis techniques. The accurate discovery of patterns through DM is influenced by several factors, such as sample size, data integrity, and support from domain knowledge, all of which affect the degree of certainty needed to identify the patterns. Typically, DM uncovers a number of patterns in a database; however, only some of them are interesting. Useful knowledge constitutes the patterns of interest to the user. It is important for users to consider the degree of confidence in a given pattern when evaluating its



validity. The knowledge discovery process is interactive and examines many decisions made by the user. Loops can occur between any two steps in the process, which are needed for further iteration. First, it is important to develop an understanding of the application domain, including relevant prior knowledge, and identify the end users goal. Second, choose a target data set and focus on the subset of variables or data samples targeted for examination. Third, clean and preprocess the data by reducing noise, designing strategies for dealing with missing data, and accounting for time-sequence information and known changes. Fourth (the data reduction and projection phase), find useful features to represent the data such as dimensionality reduction or transformation methods. Finally, use the discovered knowledge by taking action and documenting or reporting the knowledge.

The increasing use of skill in educational methods has made a huge quantity of information available. EDM provides a significant amount of relevant information and offers a clearer picture of learners and their learning processes. Such methods are then used to enhance knowledge about educational phenomena, students, and the settings in which they learn. Developing computational approaches that combine data and theory will help improve the quality of teaching and learning. From a practical point of view, EDM allows users to extract knowledge from student data. This knowledge can be used in different ways such as to validate and evaluate an educational system, improve the quality of teaching and learning processes, and lay the groundwork for a more effective learning process. Data mining and analytics do not address all the issues and decisions that are faced by any educational institute but they can be an integrated part of various administrative decisions. The purpose of this paper is to present the evolving use of data analytics in the field of education to improve the existing academic models.

II. Data Mining in Education

According to the community of educational data mining, “ Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the setting which they learn in.” It is an interdisciplinary area of research to explore the data of educational contexts. Mining with education data can help various stakeholders such as students, parents, faculty, administrators, funders and policy-makers at different levels. Similar ideas have been applied successfully, especially in business data, in different datasets, such as e-commerce systems, to increase sales profits. However, while EDM has made comparatively less progress in this direction than other fields, this situation is changing due to increased interest in the use of DM in the educational environment. Many tasks or problems in educational environments have been managed or resolved through EDM. Castro considered EDM tasks into four different parts: applications that deal with the assessment of students learning presentation, course alteration and learning commendations to modify scholars learning based on separate students activities, developing a way to appraise resources in online courses, tactics that use response from students and teachers in e-learning courses, and detection models for revealing student learning activities.

The main goal of EDM is to mine valuable information from educational data. The mined knowledge can enhance the procedure of training and learn in the scholastic method. There are three important factors in EDM- objectives, data and techniques.

Objectives: Applying methods of data mining to any specific data is led by the objectives. The main objective of using EDM is to improve teaching and learning processes. Research objectives, such as gaining a deeper understanding of the teaching and learning phenomena, occasionally influence the objectives. The application of DM to any problem is driven by the objectives of the research and the type of data at hand. Therefore, smearing data mining successfully for educational information needs an exact assumption. The assumption can be for either the DM systems or pre-processing of the data.

Data: Applying EDM methods to educational data can make extracting specific knowledge either quite simple or more complicated such as in applying relational mining. EDM is especially apprehensive with emerging techniques to discover the unique types of information in scholastic settings. Such methods are used to enhance knowledge about educational phenomena, students, and the settings in which they learn. Developing computational approaches that combine data and theory will help improve the quality of teaching and learning processes. The increasing use of technology in educational systems has made a



large amount of data available. Educational data mining provides a significant amount of relevant information. Therefore, the main source of data used in EDM can be categorized as follows:

- *Offline education*, also known as traditional education, is where knowledge transfers to learners based on face-to-face contact. Data can be collected by traditional methods such as observation and questionnaires. It studies the cognitive skills of students and determines how they learn. Therefore, the statistical technique and psychometrics can be applied to the data.
- *E-learning and learning management systems (LMS)* provide students with materials, instruction, communication, and reporting tools that allow them to learn by themselves. Data mining techniques can be applied to the data stored by the systems in the databases.
- *Intelligent tutoring systems (ITS)* and adaptive educational hypermedia systems (AEHS) try to customize the data provided to students based on student profiles. As a result, applying data mining techniques is important for building user profiles. The data generated by that system could then assist in further research.

Based on the three categories established by Romero, we can group EDM research according to the type of data used: traditional education, web-based education (e-learning), learning management systems, intelligent tutoring systems, adaptive educational systems, tests questionnaires, texts contents, and others.

Techniques: The type of technique we are using drives the application of data mining to any problem. Various techniques such as classification, clustering, regression etc. have different goals to mine. Specific tools are available for each type of technique such as Weka, DBMiner, R, Intelligent Miner, Rattle etc. Techniques that can be used in EDM for data mining are presented as follows:

Classification: It is a DM technique where data is categorized into predefined categories by a trained classifier. So, it is called as supervised learning. Various algorithms for classification are decision tree induction, genetic algorithms, naïve Bayesian classifiers, and neural networks, SVM etc. For example, students can be classified in the categories of average, below average, bright etc.

Clustering: In this technique also, data is classified into different categories. But these categories are not predefined. It is an unsupervised learning where data itself form various groups based on the similarity characteristics. For example, based on the style of learning, students can be categorized into groups of Visual, Aural and Kinesthetic.

Regression: Here, some kind of mathematical formula is used to predict that data will fall into some known data formats like linear or non-linear etc.

Prediction: It also uses some mathematical formulas to predict the value of a single aspect of the data called as predicted variable based on the values of other aspects of data called as predictor variables.

Association Rules: This DM technique uncovers relations among various attributes of the data. It defines rules of association with an if-then kind of model. For example, if the parents of a student are highly educated, the student would pass his exam.

III. Parameters predicting the Performance of Students

There are various characteristics of the students, which are used as parameters by various data mining techniques. These are used to prepare the datasets. Some of them can be described as below:

- Financial status of the student.
- Punctuality of the student to attend the class.
- Qualifications of parents.
- Company of friends.
- Internet facility for learning.
- Teacher's way of delivering lectures.
- Mother is working or not.
- Gender of the student.



- Effect of previous grades.
- Teacher-Student relationship.
- Effect of seminars/workshops.
- Practical/Lab work.
- If hostler, facilities at the hostel.
- If day scholar, facilities at home.
- Extracurricular activities.
- Parent's cohabitation.
- Size of the family.
- Travelling time to school.
- Extra classes/tuitions.
- Current health status.
- Category of the student (R, UR)
- Medium of teaching.
- Other habits (Drinking, smoking, etc.)
- Annual Income of the family.
- Reading habit. (Morning, night)
- Use of social media.
- Use of mobile.

IV. Application Areas of Educational Data Mining

Many studies have been developed in the area of EDM. A framework for examining learner's behaviors in online education videos was recommended by Alexandro & Georgios. The proposed framework consisted of capturing learner performance data, designing a data model for storing the activity data, and creating modules to monitor and visualize learner viewing behavior using captured data. Researchers relied on most of the students to watch videos in the few days prior to exams or an assignment due date. Moreover, pausing and resuming was mainly observed in videos associated with an assignment. One lamentation was that the author did not study what affected learner viewing behavior or why some learners refrained from viewing online videos altogether. In other research, Saurabh Pal built a model using data mining methodologies to predict which students would likely drop out during their first year in a university program. That study used the Nave Bayes classification algorithm to build the prediction model based on the current data. The result of the system was promising for identifying students who needed special attention to reducing the dropout rate.

Various applications of educational data mining can be described as follows:

Analysis and Visualization of Data

Pattern analysis is performed to mine hourly, daily and monthly activities of students. Important information is highlighted to support the decision-making. Various statistics for a student are visualized. For example, to track the Internet usage activities like time spent on Internet, the number of web pages browsed, browsing time for each page etc. are statistically analyzed. This gives the periodical trend of student activities with different usage summary.

Student Performance Prediction

Values of variables that define the performance of a student such as marks, score etc. are predicted using various variables of known values. Different data mining techniques are applied such as prediction and regression to predict the performance of student's success in various courses, their marks and levels of knowledge. Classification techniques like the decision tree, neural networks are used to combine individual attributes and predict the performance. Linear and multiple regression techniques are applied to predict the time spent for learning a page and success of various courses.

Enrolment Management

Practices like marketing of school or college, financial aids, retention programs, policies for admission are planned to increase the enrolment in an institution. Enrolment management is planned systematically to meet the goals of an institution.

Grouping Students



Classification and clustering techniques are applied to categorize the students into various groups and personalized learning can be provided. Groups are created based on their individual learning style like visual, aural etc. and effective training can be provided accordingly.

Predicting Students Profiling

Techniques like neural network can be used to identify the students based on different demographic, geographic and psychographic traits using the information providing them.

Planning and scheduling

Various data mining techniques like classification, clustering etc. can be used to estimate and visualize the planning for various future courses, resource allocating, counseling, curriculum development etc. Planning can enhance the traditional process of education like course-enrollment preferences, recruitments and admissions.

User Modeling

Knowledge, behavior and experience can be applied to user modeling with the help of educational data mining. Knowledge of the user, experience, motivations and learning can be incorporated in modeling with EDM.

Syllabus Organization

Syllabus of an institute is influenced by many factors like affiliation, competitions, collaboration with other organizations, availability of teachers etc. These may not be the best factors to frame the syllabus for all the students. Mining various characteristics of students can help in assisting better organization of syllabus than the existing one.

V. Conclusion

The increased use of technology in education is generating a large amount of data every day, which has become a target for many researchers around the world. The field of educational data mining is growing quickly and has the advantage of containing new algorithms and techniques developed in different data mining areas and machine learning. EDM can be used in various diverse parts including classifying at-risk students, classifying significances for the learning requirements of altered groups of students, increasing graduation rates, efficiently measuring the performance of the institution and improve the overall quality of education. Many tools are available for mining data of educational contexts but standardization of tools for EDM is required. This paper defines techniques and the common tasks that are employed in EDM. To design a specific framework and investigation of various parameters, which affects the process of teaching and learning will be the future work of research to enhance the overall development of the student.

Motivation

Every student is not same. A particular learning process may be the best suitable for some of the students but may not suitable for the rest of them. For example, some students learn by visuals whereas some may prefer aural. The study here aims to implement an automated system for students, which can classify them into different categories so that the tailoring learning environment can be provided to them. Data mining in education field can enhance the overall learning environment of the students, which is necessary for the quality education.

References

- [1] www.educationaldatamining.org. ^[1]_{SEP}
- [2] Ahmed E., Ibrahim S. E., "Data Mining: A prediction for Student's Performance Using Classification Method." World Journal of Computer Application and Technology, Volume: 2, Issue: 2, 2014
- [3] Bhise R.B., Thorat S.S., Supekar A.K., "Importance 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
- [4] E. Osmanbegović and M. Suljić, "Data Mining Approach for Predicting Students Performance", Economic Review, Volume: 10, Issue: 1, 2012
- [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] Syed Tanveer Jishan, Raisul Islam Rashu, Naheena Haque and Rashedur M Rahman. "Improving Accuracy



- of Students' final Grade Prediction Model Using Optimal Equal Width Binning and Synthetic Minority Over-Sampling Technique", Springer Open Journal, 2015
- [7] Shakil, Kashish Ara, Shuchi Sethi, and Mansaf Alam. "An Effective Framework for Managing University Data Using a Cloud Based Environment." 2nd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2015
- [8] A. 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 ^[1]_[5EP]
- [9] Raju Ranjan¹, JayanthiRanjan, "Modelling Key parameters in Higher Education using Logistic Regression: An Indian case based Data Analysis" 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), 2015
- [10]K sreenath, G Jeyakumar, "Evolutionary Algorithm Based rule(s) Generation for Personalized Coursework Construction in Educational Data Mining," IEEE International conference on computational intelligence and computing research, 2016, doi:10.1109/ICCIC.2016.7919631
- [11] Prakash Kumar, Nisha Shamra, SK Jha, "Educational Data Mining and Big Data Framework for E-Learning Environment," 5th International conference on reliability, infocom and optimization (Trends and Future Directions), 2016, doi: 10.1109/ICRITO.2016.7784961