Contemporary Scientific and Technological Aspects towards an Entrepreneurial Approach

THE 16ᵀᴴ EDITION OF THE INTERNATIONAL CONFERENCE
EUROPEAN INTEGRATION
REALITIES AND PERSPECTIVES

Data Mining in Educational Data – Useful Tool for Sustainable Learning Development

Corina Simionescu[1], Mirela Danubianu[2], Corneliu Octavian Turcu[3]

Abstract. Transforming data to knowledge can be realized through data mining methods, and that is a continuous challenge to discover helpful information in huge data sets. Data mining (DM) techniques can be used to extract knowledge and identify specific patterns that can be harnessed in optimal solutions to answer identified issues. By collecting huge data sets and analyzing them, data mining can produce major changes that enhance the quality of educational activities, management, and teachers' continuous training. The educational potential of data mining can be argued through the changes it produces in teaching and learning. Analyzing the context that subjects perceive as encouraging learning, the predictability of learning, and the understanding of real learning behavior, data mining techniques can be used to inform and develop teaching methodologies than enhance the learning results.

Keywords: educational data mining; learning management systems; online education

1. Introduction

The general development that the IT field is constantly experiencing, and in particular e-learning, as the area of application of information technology in education, have given rise to the field of Educational Data Mining, which can be described as a relatively new multidisciplinary field - combining techniques from statistics, artificial intelligence, machine learning, neuronal networks, database systems, data visualization. It aims to discover how people learn effectively and to identify aspects that can improve educational processes by exploiting educational data sets through the use of data mining techniques. Fig. 1 shows the relationships between data mining and its related fields.

[1] Stefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, Romania, Corresponding author
[2] Stefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, Romania, E-mail: mdanub@eed.usv.ro
[3] Stefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, Romania, E-mail cturcu@usm.ro

Figure 1. Relationship between Data Mining and Related Fields

The data used in EDM systems are extracted from various sources. The largest volumes come from e-learning platforms and educational software. (Al-Razgan; Al-Khalifa; Al-Khalifa, 2014)

Knowledge extracted through data mining techniques can be used to identify patterns that lead to improved decision-making processes and find optimal solutions to solve the identified issues.

In this paper we aim to make a brief overview of educational data mining.

The paper is structured as it follows: the second section refers to the phases of the EDM process, the third section addresses the current state of DM in education and research directions, the fourth section contains a brief analysis of the most used techniques in education followed by conclusions.

2. Educational Data Mining Process

The EDM process has four main phases:

a. defining the problem. In this phase, the objectives are formulated, as well as the main research directions.

b. data extraction and preparation phase. It can take up to 80% of all processing time. As the obtained knowledge essentially depends on the mined data, extraction step is very important. Data extraction is an iterative process; it doesn't stop when a certain solution is implemented. This can be a new input for a new data extraction process. (Zorić, A.B., 2020) Data quality is a major challenge in data mining. At this stage, the source data must be identified, cleaned and formatted in a pre-specified format.

c. the modeling and evaluation phase in which the parameters are set to optimal values and different modeling techniques are selected and applied;

d. the implementation phase in which the results of data exploitation are organized and presented through graphs and reports.
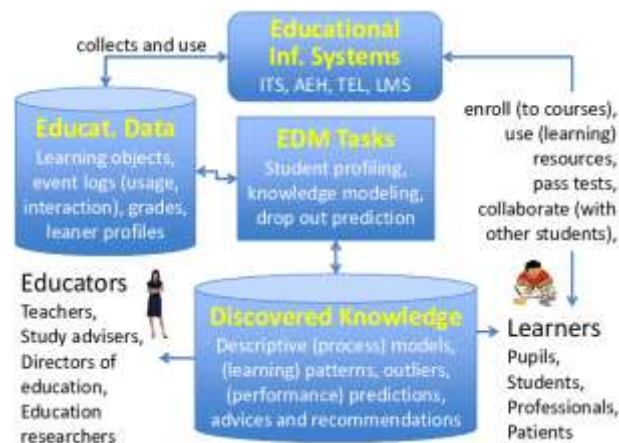
Figure 2. Educational data mining in a nutshell (Calders, Toon & Pechenizkiy, Mykola, 2012)

3. The Current State of DM in Education and Research Directions

EDM research has evolved from workshops (2005) to international conferences (since 2008 in Montreal, Canada) to the establishment of EDM communities. The Journal of Educational Data Mining began publication in 2009, and in 2011 the International Educational Data Mining Society was founded to support collaboration and scientific development in this new discipline, by organizing the EDM conference series and the Journal of Educational Data Mining.

Although EDM is a relatively new field, there are already many valuable works from which we can extract some guidelines:

- current data extraction techniques and tools are complex and can't be used by teachers who should know, first of all, to select the most appropriate tool from those available

- EDM tools should be designed to have an intuitive interface, good visualization so that they are easy to use by non-data mining experts;

- data extraction algorithms should be pre-configured so that their use can easily generate patterns

- integration in e-learning environments and standardization of input data;

There are authors who have suggested the use of XML as a data specification. Other authors have used PMML (Predictive Modeling Markup Language) for statistical and data extraction models. (https://educationaldatamining.org). To select the correct algorithms, researchers should initially design the data and make it compatible with the output data. There is a need for tools to integrate knowledge in the field of education into algorithms for data exploitation.

In the near future, researchers will conduct studies on EDM and Industry 4.0 and one of its application areas, the Internet of Things-IoT. EDM has the potential to substantially help researchers discover patterns in interaction data in Learning Management Systems (LMS) and in designing more effective learning environments.

It also gives designers the opportunity to develop personalized learning. Currently, EDM focuses more on the "student model", which is a key component of learning systems, compared to learning analysis.

The most important study area of EDM is Intelligent Tutoring Systems (ITS) where there is no human tutor. Research on the integration of LMS and ITS to combine EDM and LA forces has increased recently. Promising topics for the future of EDM include updating, optimizing and improving algorithms based on machine learning and expert systems in artificial intelligence applications in education (Muhittin & Halil, 2020).

4. Data Mining Techniques Used in Education

The educational system stores huge data that can come from both traditional and online learning environments: data about teachers, data about students, data about graduates, courses, educational programs, etc.

We present, below, a summary of the most used data mining techniques in the educational field, and their application.

Table 1. Data Mining Techniques Used in Educational Field (Villanueva, Moreno & Salinas, 2018)

Technique \ Domain	Dropping out or Retention	VLO or VLE Analysis	Performance and student's	Generation of Educationa	Learning pattern Identificati	Students patterns Identificati	Students related Prediction
Correlation Analysis		1				1	
Decision Trees	5	3	8	2	2	6	2
Regression Trees			1				
Markov Chains				1			
Classification	4	2	4		3	1	3
Clustering		7	3	5	3	9	2
Differential Sequence						1	
Sequential Patterns		4	1	1	7	3	2
Bayesian Networks		2	1		1	1	6
Neural Networks	1	2	2		1		5
Association rules		8	1	7	14	9	1
Linear regression					1		1

5. Conclusions

The daily storage of large amounts of data in education, through the massive use of IT technologies, is a target for researchers around the world.

The exploitation of educational data (EDM) helps to create and develop interesting, interpretable, useful data extraction methods to obtain new information, which can lead to a better understanding of the student and the optimal learning environment in order to obtain the best results. (Algarni & Abdulmohsen, 2016).

6. Acknowledgement

References

Abdulmohsen, Algarni (2016). Data Mining in Education. *International Journal of Advanced Computer Science and Applications*, 7.

Al-Razgan M.; Al-Khalifa A.S. & Al-Khalifa H.S. (2014) Educational Data Mining: A Systematic Review of the Published Literature 2006-2013. *In: Herawan T.; Deris M.; Abawajy J. (eds) Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013). Lecture Notes in Electrical Engineering, vol 285. Springer, Singapore*. https://doi.org/10.1007/978-981-4585-18-7_80, Online ISBN 978-981-4585-18-7.

Calders, Toon & Pechenizkiy, Mykola (2012). Introduction to the special section on educational data mining. *Sigkdd Explorations*, 13. 3-6.10.1145/2207243.2207245. https://educationaldatamining.org.

Muhittin Ş. & Halil Y. (2020) Educational Data Mining and Learning Analytics: Past, Present and Future. *Bartin University Journal of Faculty of Education*, 9(1), pp. 121-131.

Villanueva, A., Moreno, L. G. & Salinas, M. J. (2018). Data mining techniques applied in educational environments: *Literature review, Digital Education Review* - Number 33. http://greav.ub.edu/der.

Zorić, A. B. (2020). Benefits of Educational Data Mining. *Journal of International Business Research and Marketing*, 6(1), pp. 12-16. https://www.researchgate.net/publication/224160756_Educational_Data_Mining_A_Review_of_the_State_of_the_Art.