

**BOOK REVIEW**  
**EDUCATIONAL DATA MINING: APPLICATIONS AND TRENDS**  
Edited by Alejandro Peña-Ayala

Res. Assist. Aylin OZTURK  
Department of Distance Education  
Open Education Faculty  
Anadolu University, Eskisehir, TURKEY

<b>ISBN</b>	978-3-319-02737-1 (Hardcover) 978-3-319-02738-8 (e-Book)
<b>Publication Date</b>	2014
<b>Publication Formats</b>	Hardcover and e-Book (PDF)
<b>Publisher</b>	Springer International Publishing



#### INTRODUCTION

Educational Data Mining (EDM) is a developing field based on data mining techniques. EDM emerged as a combination of areas such as machine learning, statistics, computer science, education, cognitive science, and psychometry. EDM focuses on learner characteristics, behaviors, academic achievements, the process of learning, educational functionalities, domain knowledge content, assessments, and applications. Educational data mining is defined by Baker (2010) as “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 settings which they learn in”. EDM is concerned with improving the learning process and environment.

In this book, entitled “Educational Data Mining: Applications and Trends” and edited by Peña-Ayala, computer-based learning environments such as e-Learning, LMS and MOOCs are mainly considered. This book includes four parts and sixteen chapters that are prepared with contributions of experts in different fields. In the first part of this book, EDM is explained with a conceptual view. In the second part, applications of student modeling are examined. In the third part, studies in the field of assessment are discussed. In the last part, new trends in EDM field such as text mining and social network analysis are mentioned.

This book is important in terms of explaining the theoretical base and sample applications of EDM and leading new studies. In this book review, topics discussed in each section are summarized and the importance of the book for researchers has attempted to highlight.,

### **Part 1: Profile**

With a conceptual view of the EDM, as described in this section, it is indicated that valuable information about learning process can be obtained by analyzing the raw data produced in the educational systems. In e-learning systems by using DM techniques, new and different paths can be found in the solution of learning problems, and also learner behaviors and performance can be predicted.

The first section of the book consists of three chapters. In the first chapter, bibliographic review of the studies conducted in EDM field is made. It is indicated that prediction and clustering techniques are usually used in this field. In the second chapter, preprocessing of educational data is examined. Main tasks and issues in the preprocessing of educational data are explained with Moodle data. It is emphasized that preprocessing is one of the most important phases for EDM. In this chapter, steps of preprocessing are explained, but in addition, it is stated that there is no formula for this process. The third chapter illustrates that EDM can support government policies for enhancing education. For this reason, an application that includes EDM for enhancing learning quality performed in Mexico is explained.

### **Part 2: Student Modeling**

Student Modeling is one of the most important applications, especially for Computer-Based Educational Systems. With Student Modeling, learning environments can be designed according to learners' needs and expectations. This part consists five chapters which are focused on developing systems for identifying learners and estimate their performance automatically with machine learning, especially in intelligent tutoring systems and adaptive learning environments. Learner attributes such as learner characteristics, behaviors, learning styles, educational experiences, personality, academic achievements, learning system usage data, and assessment results are used for the development of these systems. It is aimed to assist tutors and to increase the quality of education, learners' satisfaction and achievement with these systems. It is emphasized that estimating performance is especially important for determining at-risk students in the first years and for retention of those students. It is stated that student models can be used in all learning environments. However, using them especially in e-learning systems in which learners and instructors are not in the same place can be more effective. In the fourth chapter, knowledge discovery processes and development of the Student Knowledge Discovery Software which is created with the determination of learner features has been explained. Student performance is modeled using data mining techniques. In the fifth chapter, automatic learner modeling is proposed towards detecting personality of the players with data mining techniques in educational games. It is emphasized that determining learner behaviors are important in order to develop user adaptable systems. In the sixth chapter, it has tried to predict learners' performance using the Multi-Channel Decision Fusion. In the seventh chapter, the model which is developed in the study entitled "Predicting Student Performance from Combined Data Sources" is tested in Open University (OU) modules. In this chapter, it is emphasized that virtual learning environment activities are beneficial data sources for predicting learner outcomes and it has been demonstrated that different modules require different methods. In the eighth chapter, the accuracy of the learners' answers by following their eye movements has been tried to estimate by using random forest. It is indicated that estimates can be made, especially in e-learning systems, intelligent tutoring systems, and online quizzes by using learners' eye movements and mouse clicks.

### **Part 3: Assessment**

This section which consists of four chapters is focused on techniques that can be used in the evaluation process. In the assessment process, learners' domain knowledge acquisition, skills development, and achieved outcomes are taken into account. In addition to these, it is underlined that reflection, inquiring, and sentiments are significant for computer-based educational systems. In the ninth chapter, a coherence analyzer is designed to be used in an Intelligent Tutoring System. With this kind of work, learners can evaluate their drafts early and simplify the reviewing process of the instructor. The tenth chapter offers an approach to create test automatically. In this chapter, a model has been developed by using EDM techniques to estimate the difficulty of the items in the computer-based tests. In the eleventh chapter, an instrument has been developed to support instructors' understanding of learner activities and visualize them. In the study, new techniques for augmenting substantial pedagogical software have been developed. In the twelfth chapter, a new approach to find the most dependent test items in learner response data has been proposed. It is highlighted that students' response data can be used to identify what is learned by learners and it can also be used to discover the relationship between the test items. These data can be used effectively in the development of the test.

### **Part 4: Trends**

This part which consists of five chapters includes new applications in the field of EDM. In particular, text mining and social network analysis (SNA) applications are described. In the thirteenth chapter, evaluation of learners' productions with text mining is described. The ReaderBench software is used for this evaluation. The authors stated that text mining techniques based on advanced natural language processing and machine learning algorithms. Text mining software allow users for cohesion-based assessment, reading strategies identification and textual complexity evaluation and to make a positive contribution to evaluation process of instructors. In the fourteenth chapter, learners' comments are evaluated by using statistical methods and text mining techniques. This kind of studies can be used particularly for revealing views of learners in online learning environments and MOOCs which have a high number of learners. In the fifteenth chapter, an application for the use of SNA and data mining in the field of education was conducted. Social network analysis becomes necessary in conjunction with the proliferation of the use of social networks in educational environments. Learner interaction in e-learning environments can be understood by using SNA methods. SNA is the methodical study that examines relationships occurs in the connected actors from a social point of view. In the sixteenth chapter, it is indicated that at-risk learners can be identified by analyzing users' interaction data collected from LMS and MOOC platforms and thus learning processes can be supported and developed. In this chapter it is stated that EDM and Learning Analytics (LA) is not very far from each other; however, at the same time, it is emphasized that they are different from each other. In EDM; prediction, classification, clustering, and association methods are used for the purpose of answering questions about educational practice. On the other hand, LA focuses on data collection, measurement, and reporting by analyzing. This book which consists of four parts will be a valuable source of interest to researchers, practitioners, academicians, and learners who aim to have information on the EDM field, to learn current practices related to the EDM and to see the trends in this area.

## BIODATA and CONTACT ADDRESSES of the AUTHOR



**Aylin OZTURK** is currently working as a Research Assistant in the Department of Distance Education, Open Education Faculty at Anadolu University, Turkey. She graduated from Primary Education Mathematics Teaching Department and Computer Education and Instructional Technology Department of Eskisehir Osmangazi University. OZTURK had her master's degree in Distance Education Department at Anadolu University and currently she is a PhD candidate at the same university in Department of Distance Education. Her research interests are open and distance education, online learning, educational data mining, personal learning environments, adaptive learning, information and communication technologies, learning analytics, and differentiated instructional design.

Research Assistant Aylin OZTURK  
Open Education Faculty,  
Anadolu University, 26470 Eskisehir, TURKEY  
Phone: +90 (222) 335 0580 ext:2772  
E-mail: [aylin\\_ozturk@anadolu.edu.tr](mailto:aylin_ozturk@anadolu.edu.tr)

## REFERENCES

- Pena-Ayala, A. (Ed.) (2014). *Educational Data Mining: Applications and Trends* (Vol. 524). Springer.
- Baker, R. S. J. D. (2010). Data mining for education. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education* (3<sup>rd</sup> edition), 7, 112-118. Oxford, UK: Elsevier.