The Role of Educational Data Mining in the Quality of Educational Process

Dr. Hafida Mokhenfer
Senior Research Fellow, Class B
The Research Division of School and its Environment
National Institute for Research in Education, Algeria
Hafidha_19@hotmail.fr

Dr. Samiya Bedoui
Senior Researcher
The Research Division of Educational Governance
National Institute for Research in Education, Algeria
samiabedoui5@gmail.com

Abstract

Artificial Intelligence has changed many aspects of lives since it provides the technological services and tools that help to make decisions in everyday life. The present paper aims to highlight the role of education data mining in improving the quality of the educational process. Education data mining techniques can discover information that can assist educators in establishing pedagogical decisions when designing or modifying an environment or teaching approach. Also, education data mining can help predict students' failure or dropout. So present paper going to define educational data mining, why educational data mining is important, the goals of using educationanl data mining, and where educational data mining can be applied. Later, the paper develops some recent examples of studies that have had utilized educational data mining to enhance the educational process and predictive models that help educators in making educational decisions and building suitable prevention programs to improve the Quality of educational process.

Keywords: Education; Data Mining; Quality; educational process.

دور التنقيب في البيانات التعليمية في جودة العملية التعليمية

د. حفيظة مخنفر

أستاذ بحث قسم ب

قسم بحث المدرسة ومحيطها

المعهد الوطني للبحث في التربية، الجزائر

Hafidha_19@hotmail.fr

د. سامية بدوي

باحث دائم

قسم البحث في الحكامة التربوية

المعهد الوطني للبحث في التربية، الجزائر

samiabedoui5@gmail.com

المستخلص:

لقد غيّر الذكاء الاصطناعي العديد من جوانب الحياة لأنه يوفر الخدمات والأدوات التكنولوجية التي تساعد على اتخاذ القرارات في الحياة اليومية. تهدف هذه الورقة إلى تسليط الضوء على دور التنقيب في البيانات التعليمية في تحسين جودة العملية التعليمية. يمكن لتقنيات استخراج البيانات التعليمية اكتشاف المعلومات التي يمكن أن تساعد المعلمين في اتخاذ القرارات التربوية عند تصميم أو تعديل البيئة أو نهج التدريس. كما يمكن أن يساعد التنقيب في البيانات التعليمية في التنبؤ بفشل الطلاب أو تسربهم. لذا، ستحدد الورقة الحالية التنقيب عن البيانات التعليمية، وسبب أهمية التنقيب عن البيانات التعليمية، وأهداف استخدام التنقيب عن البيانات التعليمية، وأين يمكن تطبيق التنقيب عن البيانات التعليمية. ثم يعرض البحث بعض الأمثلة الحديثة للدراسات التي استخدمت التنقيب في البيانات التعليمية لتعزيز العملية التعليمية والنماذج التنبؤية التي تساعد المعلمين في اتخاذ القرارات التعليمية وبناء البرامج الوقائية المناسبة لتحسين جودة العملية التعليمية.

الكلمات المفتاحية: التربية، تنقيب البيانات، الجودة، العملية التعليمية.

The Role of Educational Data Mining in the Quality of Educational Process

Introduction:

Education and knowledge is the real source of wealth and the fact that education is one of the most important pillars of sustainable development in society, which requires continuous development and training to work for the development of skills and abilities in various fields, to be able to keep up with the demands of globalization and deal with the different technical and artistic aspects related to it at different levels.

According to several approaches, where Lee Harvey and Diana Green found that quality in education can be defined in the sense of excellence; where the primary goal of education institutions is to achieve excellence and outperform others (Janne Parri,2006,p.107). The definitions of quality can be summed up as the sum of the pedagogical and scientific qualities that enable the achievement of the required teaching and learning outcomes must be obtained with high-quality (James Jiang and all, 2002, p85). Nowadays, the use of technologies in education can help to achieve a high-quality of the education process.

The increasing use of information and communication technologies in the educational field entails the storage of large volumes of data in various formats such as archives, blogs, documents, images, videos, audios, scientific data, meta data or hyperlinks, and many new data formats (Fernández & Luján-Mora, 2017, pp. 81-85).

In addition, the increase of instrumental educational software the use of the Internet in education, and the establishment of state databases of student information has created large educational data. Traditional educational institutions have used for many year information systems that store lot of interesting information. Web-based educational systems have been rising exponentially and they led us to store a huge amount of potential data from multiple sources with different formats and with different granularity levels. but it is impossible to analyze it manually. So, tools to automatically analyze this kind of data are needed. Because of all this educational data can be explored and exploited to understand how students learn (Romero & Ventura, 2020). Today,

one of the biggest challenges that educational institutions face is the exponential growth of educational data and the transformation of this data into new insights that can benefit students, teachers, and administrators (Romero & Ventura, 2020) And reaching the highest levels of quality in educational outcomes.

The enormous educational data require the application of appropriate methods or techniques to process them and to extract knowledge. In the educational field, these techniques are classified into what is known as Educational Data Mining (Fernández & Luján-Mora, 2017).

Educational Data Mining provides the technological basis for new services and tools that help decision making in everyday life. Indeed, AI and machine learning tools can help to improve in several ways the learning process. A critical aspect in this context is the possibility of developing new predictive tools, which can be used to help students improve their academic careers (Del Bonifro, Gabbrielli, Lisanti, & Zingaro, 2020, p-p: 129–140). Based on the previous discussion, the paper aims to define Education Data Mining and its Role in improving the Quality of the educational processes.

Conception of Educational Data Mining

EDM has emerged from two independent conferences and communities. The first Educational Data Mining Conference was in Montreal, Canada in 2008 organized by the IEDM society (Romero & Ventura, 2020). The first book about EDM topics was published on 2006 and it was entitled Data Mining in E-Learning. Since then, an increasing number of books have been published, especially in the last years. from 2006 to 2014 the terms Data Mining in Education and Educational Data Mining were used in the titles (Romero & Ventura, 2020)

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings, and using those methods to better understand students, and the settings in which they learning. (Aleem & Gore, 2020, pp. 182-188) (EDM) is concerned with developing methods for exploring the unique types of data that come from educational environments (Bakhshinategh, Zaiane, ElAtia, & Ipperciel, 2018). It can be also defined as the application of data mining (DM) techniques to this specific type of dataset that come from educational environments to address important educational questions (Romero & Ventura,

2020). (EDM) employ the data mining techniques on the data collected through an educational activity (Aleem & Gore, 2020, pp. 182-188).

There are several major data mining techniques have been developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns; those techniques are used for knowledge discovery from data bases. (Kalyani, 2012)

Association is one of the best-known data mining techniques that discovered a pattern based on a relationship of a particular item on other items in the same transaction. For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together. (Kalyani, 2012)

Classification techniques provide an overview of the classification problem and introduce some of the basic algorithms. Classification is a classic data mining technique based on machine learning. it is used to classify each item in a set of data into one of predefined set of classes or groups. For example, teachers classify students' grades as A, B, C, D, or F. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistic (Kalyani, 2012).

Clustering techniques: is the process of organizing objects into groups whose members are similar in some way. A cluster is a collection of objects, which are similar between them and are dissimilar to the objects belonging to other clusters Prediction. The prediction as it names implied is one of a data mining technique that discovers relationship between independent variables and relationship between dependent and independent variables (Kalyani, 2012).

Sequential patterns analysis is one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data (Kalyani, 2012).

Clustering and classification are both classification methods. Clustering is an unsupervised classification and classification is a supervised classification. Classification and prediction are also related techniques. Classification predicts class labels, whereas prediction predicts continuous-valued functions. Clustering can group together a set of pages with similar contents, users with similar

navigation behavior or navigation sessions. Classification allows characterizing the properties of a group of user profiles, similar pages or learning sessions. And outlier detection can detect students with learning problems (Romero & Ventura, 2007).

EDM has a specific set of data mining methods that are exclusively applied to the educational data. There are six broad divisions: (1) Distillation of data for human judgment (2) Prediction methods, (3) Relationship mining methods, (4) Structure discovery methods (5) Discovery with models, (6) Miscellaneous other methods (Aleem & Gore, 2020)

Prediction methods used to address the issues like students' drop out of chool or poor performance of students. Initially, prediction methods build the model for the attribute whose value is to be predicted based on the values of the other attributes in the dataset. (Aleem & Gore, 2020, pp. 182-188).

Classification which includes decision tree, k-nearest neighbor, naïve Bayesian, rule based induction in association, and many others (Khanna, Narayan Singh , & Alam, 2016).

Density Estimation is a statistical technique employed for prediction. Density estimators predict the value of an attribute based on the probability density function of that attribute hidden in the given dataset.
(Aleem & Gore, 2020, pp. 182-188).

Latent Knowledge Estimation is also a statistical technique employed for prediction. Latent knowledge estimators assess the skills of students based on their responses to a problem solving exercise. The responses of a student are mapped with the required skills for a problem. The correctness or incorrectness pattern discovered for different skills estimates the knowledge of a student with respect to that skill. Prominent methods used for latent knowledge estimation are Naïve Bayes, Bayes net, Bayesian Knowledge Tracing (BKT), and performance factors assessment (Aleem & Gore, 2020, pp. 182-188).

The data in EDM can come from Traditional classroom system or on line tutoring systems used to evidence the significant information, be it student scores, answers to online quizzes, or events from an Intelligent Tutoring System. The main objective of orientation towards Educators is to get relevant denigration for training, assessing the organizations' course. content and its effectiveness,

perform classification to group learners on the basis of their requirements, identify learners behavior, identify the activities that adds on to learning, course personalization and restructuring, proficient content organization and constructing instructional plans, etc. The overall objective for Educators is to improve and enhance the learning among students. (Khanna, Narayan Singh , & Alam, 2016, pp. 1 - 7).

Educational data are gathered (Romero, Romero, & Ventura, 2014) from different sources such as the interaction between instructors, students and the behavior, input in quizzes, interactive exercises, forum messages, etc. administrative data, school and teacher information, demographic data, student affectivity , and so on. Educational environments can store a huge amount of potential data from multiple sources with different formats and with different granularity levels (from coarse to fine grain) or multiple levels of meaningful hierarchy (keystroke level, answer level, session level, student level, classroom level, and school level) that provide more or less data. Gathering and integrating all this raw data for mining are nontrivial tasks on their own and thus a preprocessing step is necessary.

The objectives of Educational Data Mining

Educational Data Mining is used to "Predict the Future and Change the Future" (Khanna, Narayan Singh , & Alam, 2016, pp. 1 - 7). The main objective of using EDM is to improving learning and teaching process. (Jones, Rubel, & LeClere, 2020, pp. 1227-1241). Moreover, provide opportunities for educator and educational technology systems to personalize students' educational experiences; according to their socioeconomic profile, personality characteristics, personal interests, and educational dispositions among other things. The EDM tools can provide instructors with information to interventions at the right time; they can also adjust the release of course content and construction of assessments based on a student's past school achievement and predicted levels of academic achievement (Jones, Rubel, & LeClere, 2020, pp. 1227-1241).

EDM help Learners to get an insight into their learning habits and give them the recommendations on how to further carry out improvement in their learning. It helps Educators to examine the learning among students, course developers in evaluating the course structure and how much effective it is in the educational setting. EDM understands the impact of various variables that effect teaching-

learning and in turn supports holistic decision making and therefore improves the organizational productivity (Khanna, Narayan Singh , & Alam, 2016). It explore the 'Big Data' from the field of education, solve the problems related to education, enhance the learner's knowledge and improve learning, help students and institutions in achieving success. (Aleem & Gore, 2020, pp. 182-188)

Application Areas Of Educational Data Mining

User Knowledge Modeling: "This aims at finding out the content a student know which may include specific, conceptual or technical understanding. The time student spend on practice, the number of times the student looked for hints and suggestions, the number of correct responses made, repeating the mistakes again and again frequently etc. can help in determining the Knowledge of the student" (Khanna, Narayan Singh , & Alam, 2016).

User Behavior Modeling: "Modeling the individual variances between students is of imperative importance in educational data mining research. The characteristics or behavior of students determine the echelon of learning and helps in predicting the students' knowledge and future performance. Most recently educational data mining methods have put importance on User behavior modeling as it facilitate researchers to study what factors are most important that can lead to greater edge in learning process among students" (Khanna, Narayan Singh , & Alam, 2016).

Domain modeling: "Domain in the broader area to describe the real world object or concepts. Domain modeling takes into consideration how a topic of domain under study is generalized which can in turn affect the learning among students. It depends on the depth a topic into key concepts at a particular stage of generalization. The more generalized the concept is; the better is the Learning". (Khanna, Narayan Singh , & Alam, 2016).

Trend Analysis: In education, trend analysis with the use of certain techniques extracts an essential pattern that further helps to identify the changes in student learning with time. It helps in study the gap in the existing learning methods which can be further enhanced to improve the learning process (Khanna, Narayan Singh , & Alam, 2016).

Samples of Studies That Used Education Data Mining:

The first study: Renato & Edna (2020), identify the factors that affect students who drop out of the University of Brasilia (UnB) using Data mining technique (Machine Learning) to provide a model for predicting which students will drop out of under graduate courses. The model used was a model belonging to the supervised classification (Renato & Edna, 2020, p-p:1637–1667).

The second study (Annalina, Lara, Simone, 2018), the purpose was to show the adequacy of a recent statistical technique, known as Bayesian Profile Regression as a useful EDM tool for detecting the main characteristics of students at risk of academic failure. The modelling approach, adopted in the analysis, goes beyond the most common methods which attempt to perform a direct regression of dropout against a number of variables, supposed to influence the academic failure. (Sarra, Lara, & Simone, 2018).

The third study: Kabathova & Drlik (2021). the main goal of the study is to emphasize the importance of the data understanding, data gathering phase, stress the limitations of the available datasets of educational data, compare the performance of several machine learning classifiers, and show that also a limited set of features, which are available for teachers in the e-learning course, can predict student's dropout with sufficient accuracy if the performance metrics are thoroughly considered. The data collected from four academic years were analyzed. The features selected in this study proved to be applicable in predicting course completers and non-completers. The prediction accuracy varied between 77 and 93% on unseen data from the next academic year. In addition to the frequently used performance metrics, the comparison of machine learning classifiers homogeneity was analyzed to overcome the impact of the limited size of the dataset on obtained high values of performance metrics. The results showed that several machine learning algorithms could be successfully applied to a scarce dataset of educational data. Simultaneously, classification performance metrics should be thoroughly considered before deciding to deploy the best performance classification model to predict potential dropout cases and design beneficial intervention mechanisms (Kabathova & Drlik, 2021).

Fourth study: Ribeiro & Canedo (2020)The purpose of this study is to identify the factors that affect students who drop out of the University of Brasilia (UnB) using Machine Learning to provide a model for predicting which students

will drop out of undergraduate courses; to reduce the dropout rate. The result of this work demonstrates that the courses with the most credits (workload), longer time to complete (5–6 year courses) and student's poorer academic performance (poor grades) influences student dropout rate. Also, social factors, such as quota holders or non-quota holders, also influence the dropout rate of undergraduate students at the University of Brasília (UnB) .

Conclution:

Educational Data Mining is a discipline that develops methods for exploring large-scale data from the educational sector and using those methods to understand better students' performance. ADM is the application of data mining (DM) techniques in educational environments to address important educational questions .it uses several major data mining techniques including association, classification, clustering, prediction, and sequential patterns; those techniques are used for knowledge discovery from databases. The main objective of using EDM is to improve the learning and teaching process. and provide opportunities for educators and educational technology systems to personalize students' educational experiences; according to their socioeconomic profile, personality characteristics, personal interests, and educational dispositions among other things. Educational Data Mining applied: User Knowledge Modeling, User Behavior Modeling, Domain Modeling, and Trend Analysis (Ribeiro & Canedo, 2020).

References:

Khanna, L., Narayan Singh , S., & Alam, M. (2016). Educational data mining and its role in determining factors affecting students academic performance: A systematic review. *1st India International Conference on Information Processing (IICIP)*, (pp. 1 - 7). Delhi, India. doi:doi: 10.1109/IICIP.2016.797

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated. *WIREs Data Mining Knowl Discov. e1355.* doi:https://doi.org/10.1002/widm.1355

Aleem, A., & Gore, M. (2020). Educational Data Mining Methods: A Survey. *IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)Gwalior,*, (pp. 182-188). India. doi:doi: 10.1109/CSNT48778.2020.9115734.

Fernández, D. B., & Luján-Mora, S. (2017). Comparison of applications for educational data mining in Engineering Education," 2017 IEEE World Engineering Education Conference (EDUNINE). *IEEE World Engineering Education Conference (EDUNINE)*, (pp. 81-85). Santos, Brazil. doi:doi: 10.1109/EDUNINE.2017.7918187.

Jones, K. M., Rubel, A., & LeClere, E. (2020). A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics. *Journal of the Association for Information Science and Technolo, 7*(10), 1227-1241. doi:https://doi.org/10.1002/asi.24327

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 1-9. https://doi.org/10.1186/s40561-022-00192-z

Renato Carauta Ribeiro & Edna Dias Canedo,Using Data Mining Techniques to Perform School Dropout Prediction: A Case Study, spring Nature Switzerland AG 2020, 17th International Conference on Information Technology–New Generations (ITNG 2020), Advances in Intelligent Systems and Computing v 1134, https://doi.org/10.1007/978-3-030-43020-7_28, p-p:1637–1667.

Annalina Sarra, Lara Fontanella, Simone Di Zio, Identifying Students at Risk of Academic Failure Within the Educational Data Mining Framework, https://doi.org/10.1007/s11205-018-1901-8, Springer Science+Business Media B.V, part of Springer Nature 2018, published online 19 april 2018.

C. Romero, S. Ventura, Educational data mining: A survey from 1995 to 2005, Volume 33, Issue 1, July 2007, Pages 135-146, 2006 Elsevier Ltd. All rights reserved, doi:10.1016/j.eswa.2006.04.005

Kalyani M Raval (B.Com, MSc IT), Data Mining Techniques, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012, Research Paper Available online at: www.ijarcsse.com

Kabathova, J.; Drlik, M. (2021). Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. Appl. Sci. 2021, 11, 3130. https://doi.org/10.3390/app11073130

Janne Parri,(2006), Quality in Higher Education,Vadyba/ Management, Index Copernicus International,Vol.11,Iss:02,Poland ,p 107.

James J.Jiang and all, (2002), Measuring Information System Service Quality: SERVQUAL from the Other Side, Journal of MIS Quarterly, Vol.26,No.2, p85.

Ribeiro, R.C., Canedo, E.D. (2020). 17th International Conference on Information Technology–New Generations (ITNG 2020). Advances in Intelligent Systems and Computing, vol 1134. Springer, Cham. https://doi.org/10.1007/978-3-030-43020-7_28