Analysis of student performance through data mining techniques. Study case: Learning management system at UET

Jugerta GURABARDHI
European University of Tirana, Tirana, Albania
jgarabardhi@uet.edu.al

Teuta XHINDI
European University of Tirana, Tirana, Albania
teuta.xhindi@uet.edu.al

Abstract

Industry 4.0, like in any other field, also in education, has made it possible for the activities of teaching to continue to develop and the efforts made are for distance learning or e-learning. This is related to the integration of systems that synchronize with computers, mobile phones and technology that can manage the learning system electronically.

Nowadays, modeling user preferences is one of the most important tasks challenging in e-learning systems. This research aims to use Data Mining (DM) for it analyzed the data collected from the learning management system (LMS) used in e-learning systems. The main goal is to predict the individual learning style by using the Moodle LMS platform and analyze the data through Data Mining techniques. With a large volume of data, such as the time spent on the page, as well as the actions taken by students on the platform, it is intended to adapt models to their current preferences.

In this context, the research focuses on the use of Data Mining to improve the quality of education and identify models in educational environments. For her to accomplish this, the study uses well-known data mining techniques and uses an environment analysis called RapidMiner. The study describes how RapidMiner can be used to extract information from the raw data of students in the management system to the students. This paper uses student data captured in the UET LMS management system online teaching and analyzes different algorithms to choose the most suitable ones for the given model. In particular, the analysis of 10 million data records was carried out usage collected from the Learning Management System in 450 online university courses from the period March-June.

Keywords: *Student Performance, Data Mining, Learning Management System, Analysis, Forecasting, RapidMiner, LMS.*

Introduction

We witness the importance of modern education in people's lives. The growing need for an educated workforce in society requires education to be a leader in economic reform. Educational organizations store their data in electronic form and the amount of information does grow every day. To analyze this data, a process is needed which includes inspection, data cleaning, transformation, and modeling. This process is called Educational Data Mining (EDM) in the educational field and converts raw student data from systems educational into useful information for the educational process.

The methodology of online learning in higher education requires the analysis of a large volume of its data generated by planned activities in different subjects. This data is stored in dedicated spaces and includes various interactions such as log files associated with communication between the system and students. To deal with this huge volume of data, data mining techniques provide a way to analyze and extract knowledge from them. This knowledge has a significant impact on the process of decision making.

According to the number of students and activities in the analyzed subjects, we encounter a large volume of data that cannot be analyzed by traditional methods, be they automated or manual analysis. This is why data extraction techniques, such as data mining, including the logic model for knowledge extraction, provide a clear visualization of the results of user application. This knowledge makes an important contribution to the process of decision making. In many bibliographic sources, the classifications of extraction techniques data, such as classification and clustering, allow the creation of data sets, especially

when the analyzed group is large. In this paper, we aim to describe the most widely used, accessible and powerful available to researchers or practitioners of Educational Data Mining (EDM) or Learning Analytics (LA). The discussion in this paper will follow a course like that of a research question or analysis. In the context of extracting educational data, as well as in other areas of data science, the main challenge is the transformation of the data unstructured and integrated into comprehensible variables. Often, data comes in the form of formats that are not ready for parsing. To make it possible, not only should are transformed into a more understandable format but must also be developed into variables meaningful (see section 3.3 in Baker, 2015 or Veeramachaneni et al., 2015, for discussion of details of this process).

In addition, the data must often be cleaned to eliminate unusual cases and values. After cleaning the data, transforming it into a better format fit and feature engineering, the next question that presents itself to EDM or LA researchers are that of analysis - what tests can be performed, what models can be built, what relationships can be mapped and explored, and how can we validate our results?

Methodology

This paper emphasizes the necessity of improving the e-learning method of education to effectively align it with rapid technological advances, especially after pandemic interruptions in the teaching process.

The main objective is to explore the application of EDM in e-learning systems from a perspective 'improving', considering the need to improve e-learning.

Question questions:

- Can EDM be applied to e-learning to improve the design of e-learning systems?
- How to predict student engagement and performance in e-learning systems?
- What are the possible areas in e-learning where educational institutions should focus?

The methodology used in this study is analytical, using data available in e-learning platform (UET LMS) to draw conclusions and create models that help to improve the student learning experience. Hence, this paper highlights the increased necessity for applying EDM techniques to improve e-learning systems in the future.

Literature review

E-Learning: Learning Management System (LMS)

Recently, the field of E-Learning has served as an opportunity not only to reflect on the role of technologies in the teaching process, but also to review the way we conceive the learning process itself. E-Learning brings an important difference in comparison with traditional learning environments offering the possibility to track the actions of users while exploring Electronic Learning Environments (ELEs). These data are unique and expressed in numerical form. Therefore, data-driven approaches should experiment to be analyzed.

Specifically, the Learning Management System (LMS) was developed to improve independent learning of students. LMS is a software used for presentation of materials teaching, management of learning activities and assessment of results based on the web. LMS includes administration, content delivery, assessment, tracking and monitoring, cooperation and communication. LMS users can be educators, students, and administrators. The development of the Learning Management System aims to help students in the process of learning. There has been a wide spread of LMS applications in the business world, but still are not fully utilized. A way that can be used to evaluate the use of this application is the activity of users, such as students and lecturers who use it, and the results they get. Through the LMS, the lecturer can distribute the syllabus of his courses, manage files, create exam questions, quizzes, assignments, monitor student activity, deliver assessment, communicate with students through forums and email, and to follow the progress of students. Through the LMS, students can access the program curriculum, materials, submit assignments, view assessment results, solve problems, quizzes, manage files, communicate with other students and lecturers through forums and sending e-mail. While through the LMS administrator, you can register the course, define the faculty and courses, register and manage the internet pages for data storage.

In the literature, there is a growing interest in this topic in different communities' research such as Data Mining, User Modeling and Intelligent Tutoring Systems or E-Learning.

Data Mining

Data Mining is an analytical process used for exploring large amounts of data, usually related to business or market, in search of consistent patterns and systematic relationships between variables. After the discovery of these models,

they are validated by applying the models to new data. This practice has attracted the attention of many researchers and scientists due to the availability of large amounts of data in various forms such as records, texts, files, sounds, images and others.

There is a concern about missing values, noisy data, data e rarity, static data, dynamic data, attractiveness, heterogeneity, importance, size of its data, algorithm efficiency and complexity. The data we have is often large and noisy, meaning they are imprecise and have a complex structure. In this case, purely statistical techniques will not work, so data mining is one solution (Aruna & Butey, 2014).

At a time when processed data is increasingly available and their complexity increases, the acquisition of data and the extraction of knowledge from them are becoming more and more important and useful. This is especially true due to the increase in applications of Internet-connected systems. But applying the mining methods of data, hidden relationships that are not immediately apparent can be discovered and models that help in making decisions. Thus, data mining becomes a vital tool to handle large amounts of data and extract value from it through in-depth analysis and finding useful patterns.

Data mining uses many techniques to extract useful information where the main purpose of these techniques is to discover patterns using algorithms and different methods. The application of DM is wide and includes various fields such as its visualization data, statistics, Machine Learning, database systems and finding information. They can be divided into two main parts: descriptive mining and mining predictors, each with different functions and techniques. Data mining techniques fall into three main groups: statistical techniques, machine learning techniques, and techniques of artificial intelligence. Each of these techniques has its own algorithms to create patterns to get the best solution (Mustafa Abdalrassual Jassim & Abdulwahid, 2021).

In the field of education

In the traditional educational model, lecturers have a key role in the teaching process. They have the duty to share their knowledge and experience with the students, which they are expected to have a certain base of knowledge and skills.

Before the internet age, there were several models of distance education that included programs television, manual or recorded audio/video. In these models, the lecturers were available to resolve issues by phone or mail. Although these models enabled learning from home and had a flexible schedule, lack of interactivity hindered the teaching process.

Today, the education system has changed drastically with the advent of the Internet. Many institutions now offer online courses, taking advantage of the

benefits of the internet. These courses do not are limited to a specific geographic location or regular schedule, which made possible the growth of the number of potential students. This has led to the creation of universities that offer exclusive online education, while traditional universities have expanded their offerings with online courses and hybrid classes.

E-learning promotes a more personalized learning process, where students play an active role. E-learning courses can be offered through the Systems of Learning Management Systems (LMS) such as Moodle, Sakai and ILIAS, or Learning Platforms such as Knewton and Dream Box. A characteristic of these courses is the large amount of data that can be collected. In addition to the student's history and performance data, any action I performed (reading files, participating in forums, sending messages or visiting links recommended, for example) leaves a digital footprint. From a general perspective, it can be argued that EDM focuses more on techniques and methodologies. Therefore, the contributions of this work are:

a) analyze the origins and characteristics of these fields of research.
b) to provide an overview of the accompanying literature.
c) to examine how both fields of knowledge have evolved in recent years and to discuss their possible convergence, and
d) to present some of the new challenges and trends, including those related to Big Data and MOOCs.

E-Learning Data Analysis (EDM) methods are used for analyzing data on learning and behavior of students in educational environments. The application of EDM methods contains a series of steps that include planning the study design, extraction of data from the educational environment, pre-processing of data, modeling data and interpretation of models.

The importance of DATA MINING in education according to the literature

Educational Data Mining (EDM) is an emerging research field that includes the concept of 'Data Mining in Education' (Grigorova et al., 2017). It revolves around four main parties interested: students, teachers, administrators and researchers. EDM involves the use of wide range of Data Mining techniques to analyze large volumes of collected data from educational institutions (Shahiri et al., 2015). As defined by the International Society EDM (The International EDM Society) and (Baker et al. 2010), EDM is a discipline in development focused on developing methods to explore unique educational data and for gain a better understanding of students and their learning environments. His purpose is to extract valuable

hidden information from large educational data sets in all levels, from schools to universities (Bhardwaj and Pal, 2012). EDM plays a role crucial in uncovering accurate information about student behavior and evaluating the effectiveness of the learning process (Sana et al., 2019).

Researchers use EDM in e-learning environments to evaluate the process of teaching and learning and to suggest improvements. Common applications of EDM in e-learning include the analysis and visualization of educational data, the examination of student behavior, predicting student performance, providing feedback related to learning, providing feedback to teachers, curriculum design and planning of school activities (Romero and Ventura, 2010). The goal of education institutions is to improve existing education systems by empowering administrators with better decision-making skills (Silva and Fonseca, 2017).

Currently, EDM research efforts in e-learning mainly focus on four main areas: understanding student learning behaviors, assessing or predicting i student performance/grades, assessment of engagement and satisfaction levels of students and analyzing the reasons for the student dropout rate. There was a growing interest among researchers in the field of e-learning, as evidenced by the growth of papers published on e-learning systems (Fig. 1, based on data from the ScienceDirect database). Educational institutions have increasingly adopted e-learning mechanisms in recent years, which has contributed to the interest in EDM. The figure illustrates an increase of continuous in published papers on e-learning systems in institutions education from 1998 to 2015. However, between 2015 and 2018, there was a decline, potentially due to the shifting focus of researchers towards Artificial Intelligence and Blockchain. However, research on e-learning systems regained momentum after 2018, especially with the renewed attention on EDM techniques following the global impact of the Covid-19 pandemic, which caused a significant transformation in the education sector worldwide.

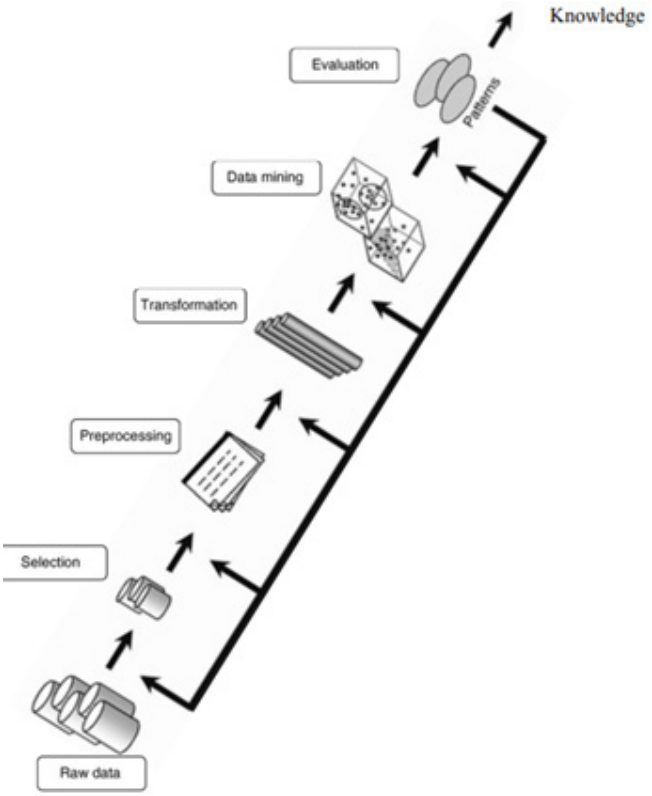
FIGURE 1. Increase of published works on the E-Learning system in educational institutions.

(Retrived from International Journal of Educational Development 101, 2023)

Definitions and techniques of Data Mining

Data mining, also known as knowledge discovery in the database, involves extracting or "mining" knowledge from a large amount of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships useful for the decision-making process. While data mining and discovery knowledge based on data are often treated as synonyms, data mining is a part of the process of discovering knowledge. The sequence of steps identified in extracting knowledge from its data is shown in Figure 8.

FIGURE 2. Steps of extracting knowledge from data



Different algorithms and techniques such as Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Connection Rules, Decision Trees, The Genetic Algorithm, the Close Unit Method, etc., are used to discover knowledge from the basics of data. These techniques and methods in data mining should be briefly mentioned and better understood.

Discussions and Conclusions

Discutions

Several data mining techniques such as Attribute Weighting have been applied in this study (Weighting by information Gain, Relief, Hi-Squared, Uncertainty), Grouping (K-Means), Classification (Tree Inductions), and Association Mining (Apriori, FPGrowth, Create Association Rules).

The use of clustering methods was aimed at dividing the data into groups of data that naturally form a group. Student actions were grouped together to investigate student behavior patterns in the LMS System. The K-Means algorithm was used to define groups, starting without first knowing the groups in the data. Using prediction techniques, the aim was to develop a model that could predict the expected variables (predicted variables) from the predictor variables (predictors). The Inductive Decision Tree algorithm was chosen as the classification method. The expected variable was the categorical variable of the final assessment (Final Mark). The aim was to determine the variables that have a significant impact on the final evaluation. By implementing ARM (Association Rule Mining), the aim was to discover relationships between variables. Algorithms FPGrowth, Creation of Association Rules (Create Association Rule) and APRIORI as data mining techniques of association rules. The rules of findings can be explained in the form that if a set of certain variables is present, another variable has a high probability of having a specific value.

Conclusions

In this work I have presented research on the implementation of data mining in course management systems, along with a case study about the LMS system. I have described how various data mining techniques can be used to improve courses and student learning. All these techniques can be applied independently in the same system or together in a hybrid system. Although I have described the most popular and general techniques of data mining, there are also other specific data mining techniques that are also used in e-learning, such as unusual value analysis and social network analysis. (They can be as an indication for the work next time).

Also in this research, some concrete models of the mine were proposed data for LMS data based on several techniques (mentioned above). This mining work of educational data allowed the identification and localization of information on

the processes of e-Learning that need improvement, as well as those that perform very well and can be used as good examples. Educational data mining that was investigated in this research allows for better analysis and understanding of learning and teaching processes applying data mining techniques. Experimental results have shown that the presented data mining model was able to obtain comprehensible answers, actionable and logical from the learning management system data describing student learning behavior patterns.

This work was focused on the overall performance of the management system courses at the European University of Tirana and the data mining process of LMS UET. LMS data mining allowed the identification of the most effective ways of the process teaching that can be used to improve the educational process. To test further the effectiveness of the proposed model and to expand the generality of this research, should conduct more extensive experiments using even larger amounts of data course management system.

In the future, it would be very useful to have tools to mine it data oriented specifically for e-learning environments. Today, tools of the mine data drives are usually designed more for power and flexibility than for simplicity. Most of the current data mining tools are too complex to be used by educators and their features exceed the scale of what an educator can to want to do. Therefore, these tools should have a more intuitive and friendly interface to the user, with parameter less data mining algorithms to simplify configuration and execution, and with good visualization services to make the results of understandable for every eLearning leader and designer. Also, it is necessary for the data mining tool to be integrated into e-learning environments as another tool for authors. Thus, all data mining processes can be performed in a single application and the results and feedback gained can be directly applied to the e-learning environment.

Bibliography

Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F. (2014, June 24). A Survey of Data Mining Techniques for Social Media Analysis. Journal of Data Mining & Digital Humanities, 2014.

Alcalá, R., Gacto, M. J., & Alcalá-Fdez, J. (2017). Evolutionary data mining and applications: A revision on the most cited papers from the last 10 years (2007–2017). WIREs Data Mining and Knowledge Discovery, 8(2).

Ali, R. H. (2022). Educational Data Mining for Predicting Academic Student Performance Using Active Classification. Iraqi Journal of Science, 3954–3965.

Amutha, P., & Priya, D. (2018). A survey on educational data mining techniques in predicting student's academic performance. International Journal of Engineering & Technology, 7(3.3), 634.

Altalhi, A. H., Luna, J. M., Vallejo, M. A., & Ventura, S. (2017). Evaluation and comparison of open-source software suits data mining and knowledge discovery. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7(3), 1204.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. Technology, Knowledge and Learning, 19(1–2), 205–220.

Caprotti, O. (2017, July 5). Shapes of Educational Data in an Online Calculus Course. Journal of Learning Analytics, 4(2).

Dabhade, P., Agarwal, R., Alameen, K., Fathima, A., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. Materials Today: Proceedings, 47, 5260–5267.

Dhika, H., Destiawati, F., Surajiyo, & Jaya, M. (2021). Data Mining Approach for Learning management system. IOP Conference Series: Materials Science and Engineering, 1088(1), 012013.

Dhingra, K., & Sardana, K. D. (2017). Educational data mining: a review to its future vision. International Journal of Technology Transfer and Commercialisation, 15(3), 309.

Ghosh, A., & Freitas, A. (2003, December). Guest editorial data mining and knowledge discovery with evolutionary algorithms. IEEE Transactions on Evolutionary Computation, 7(6), 517–518.

Gupta, P., Sharma, A., & Jindal, R. (2016). Scalable machine-learning algorithms for big data analytics: a comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 6(6), 194–214.

Xiahou, X., & Harada, Y. (2022). K-Medoids Clustering Techniques in Predicting Customers Churn: A Case Study in the E-Commerce Industry. International Journal of Computer Trends and Technology, 70(2), 22–28.