Application of EDM to Understand the Online Students’ Behavioral Pattern

Luis Naito Mendes Bezerra, UNIP - Universidade Paulista, São Paulo, Brazil
Márcia Terra da Silva, UNIP - Universidade Paulista, São Paulo, Brazil

ABSTRACT
In distance learning, the professor cannot see that the students are having trouble with a subject, and can fail to perceive the problem in time to intervene. However, in learning management systems (LMS’s) a large volume of data regarding online access, participation and progress can be registered and collected allowing analysis based on students’ behavioral patterns. As traditional methods have a limited capacity to extract knowledge from big volumes of data, educational data mining (EDM) arises as a tool to help teachers interpreting the behavior of students. The objective of the present article is to describe the application of educational data mining technics aiming to obtain relevant knowledge of students’ behavioral patterns in an LMS for an online course, with 1,113 students enrolled. This paper applies two algorithms on educational context, decision tree and clustering, unveiling unknown relevant aspects to professors and managers, such as the most important examinations that contribute to students’ approval as well as the most significant attributes to their success.

KEYWORDS
Clustering, Data Analyzes, Decision Tree, Distance Learning, Higher Education, LMS, MOOC, Students’ Behavioral

INTRODUCTION
Distance Learning for undergraduate courses is continuously increasing all around the world. According to the last survey conducted by the INEP (2013), Brazil has 1.1 million students in this modality and a growth rate of 24% from 2010 to 2013, i.e. an average growth rate of 6% per year. It is greater than the numbers of US, where the growth rates in applications, for at least one course in this modality, represented 3.9%, overpassing the previous year rate of 3.7%. The segment was responsible for 5.8 million applications only in 2014 (Allen & Seaman, 2015; Baker, Isotani, & Carvalho, 2011).

In the online education context, a new modality should also be considered – MOOCs (Massive Open Online Courses). They are completely online, application fee free, not demanding any previous requirements and, comprising an expressive number of students from several countries (Cooper & Sahami, 2013; Hyman, 2012; Liyanagunawardena, Adams, & Williams, 2013; López-Meneses, Vázquez-Cano, & Román, 2015; Martinez Abad, Rodriguez

DOI: 10.4018/JITR.2019070109

Copyright © 2019, IGi Global. Copying or distributing in print or electronic forms without written permission of IGi Global is prohibited.
Conde, García-Peñalvo, 2014). As an example, Think Again: How to Reason and Argue offered by University of Duke in partnership with Coursera in 2012, reached 226,652 students’ applications. However, it is not usual a course with more than 100,000 students as a MOOC typically has an average of 25,000 students (Jordan, 2015).

Students normally use either Learning Management Systems (LMS) that are commercial or open-source systems, similar to the virtual environments employed by MOOCs providers, as Coursera or Udacity. Consequently, an enormous amount of data regarding the platform usage and students’ progress can be registered and collected allowing analysis about their behavior in the virtual environment. These environments include modules recording each event automatically in the LMS (Long & Siemens, 2014; Pardo & Kloos, 2011; You, 2016).

It is worth noting that during the traditional face-to-face courses, teachers have the possibility to interact with students in order to follow up their understanding regarding the subject and also their performance. So, it is possible making adjustments on the subject based on the students’ behavior. That model cannot be applied to distance learning once there is not a face-to-face component between teacher and student. That model cannot be applied to distance learning once there is not a face-to-face component between teacher and student. Instead of it, there is a significant volume of data generated by the LMS students that can be collected and stored but cannot be manually examined by basic computational methods such as spreadsheets. These data cannot be interpreted by traditional technics due to the big volume of registers, attributes, data absence, qualitative and not quantitative data, and others. Then, it is necessary the application of more sophisticate computational solutions (Castro & Ferrari, 2016).

In this context, the usage of adequate computational systems to analyze the collected data allows understanding regarding students’ behavioral patterns in the LMS (Castro & Ferrari, 2016; Goldschmidt, Bezerra, & Passos, 2015; Muñoz-Merino, Ruipérez-Valiente, Alario-Hoyos, Pérez-Sanagustín, & Delgado Kloos, 2015).

Educational data mining (EDM) arises in this context as a tool that enables the transformation of raw data, generated by education environments, to useful information for education researchers (Castro & Ferrari, 2016; Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Cristóbal Romero & Ventura, 2010).

It is also important to consider that it is possible to find in the literature research papers that approach the analysis of educational data from the application of statistical methods, such as, for example, the works (Credé, Roch, & Kieszczynka, 2010; Goos & Salomon, 2017; Park, Nam, & Cha, 2012). However, this article uses another approach, based on concepts and techniques of educational data mining.

Hence, the objective of the present article is to describe the application of educational data mining methods transforming data into new and relevant knowledge about students’ behavioral patterns on an online course.

It is important to emphasize that this knowledge allows teachers and managers to plan and act beforehand, mainly to enhance the learning process maintaining the students’ permanence on the course.

In this manner, this article contributes improving the data processing and decision-making processes to teachers and managers, allowing them to act beforehand, mainly to enhance the learning process maintaining the students’ permanence on the course.

Besides the introduction, this article is divided into five more sections. The initial section presents the theoretical background regarding data mining and educational data mining as well as a literature review. Subsequently, we present the research methodology, a study case, conclusions, and finally the references.