Analytical Approach for Predicting Dropouts in Higher Education

Garima Jaiswal, Indira Gandhi Delhi Technical University for Women, New Delhi, India
Arun Sharma, Indira Gandhi Delhi Technical University for Women, New Delhi, India
Sumit Kumar Yadav, Indira Gandhi Delhi Technical University for Women, New Delhi, India

ABSTRACT

In the world of technology, tools and gadgets, a huge amount of data is produced every second in applications ranging from medical science, education, business, agriculture, economics, retail and telecom. Higher education institutes play an important role in the overall development of any nation. For the successful operation of these institutions, continuous monitoring for improving the quality of education and students is required, which is the subject of this article. A huge amount of data that education systems produce increases every year and it is difficult by traditional techniques to manage, predict and analyze this data. This challenge can be addressed through mining large amount of data. It enables the institutions to use their present reporting trends to unmask hidden patterns and identify data relationships. Through this, institutions easily predict which students are likely to dropout, and their performance. Present paper conducts a detailed and exhaustive study on techniques and approaches implemented in education mining for predicting dropouts.

KEYWORDS
Classification, Data Mining, Dropout, Education Data Mining, Prediction

1. INTRODUCTION

In the internet age or the dot com age, a huge amount of data transformation is observed. There is an exponential rise in data generation. According to an IBM Marketing Cloud study, the numbers of internet users are multiplying at a fast rate. In 2014, the number of internet users was 2.4 billion. In 2016, it was raised to 3.4 billion. In 2017, 300 million new users joined – this aggregated to 3.8 billion internet users (as of April 2017). The figures conveyed that the number of internet users increased by 42% in past three years (Schultz, 2017). Do we ever wonder, what is happening every day on the internet? Each day data is produced by the users of social media, Twitter is flooded with 656 million tweets per day, each day around 4 million hours of data is uploaded on YouTube, Instagram is populated with 67,305,600 posts each day. In 2015, there were 1.44 billion Facebook users and in the beginning of 2016, the number grew to 1.65 billion. At present, the figure has crossed over to 2 billion users (Schultz, 2017).

This huge amount of data generated every minute can be used to produce meaningful information which can be used to serve the nation in a better way. Due to large volume of data is generated and stored in databases, traditional approaches and database tools are no more adequate for analyzing such a huge, voluminous amount of data. The biggest problem for the educational institutes is the storage of huge volume of data which is generated and how to utilize this data for improving the

DOI: 10.4018/IJICTE.2019070107

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
intake and retention of students, improving academic programs, facilities, services and management (Abaidullah, Ahmed, & Ali 2015; Delavari, Phon-Ammuaisuk, & Beikzadeh, 2008; Goyal & Vohra, 2012). Higher education institutes implement various “conventional and unconventional" strategies based on “qualitative and quantitative” approaches, which keep them away from achieving their quality targets (Abaidullah et al., 2015; Delavari et al., 2008). The approaches used by the educational institutions are mainly based on their consistent formats and reports in the form of student feedback. These methods have the shortcoming to unfold the hidden information like student performance, admission intake, predicting student’s weak areas (Abaidullah et al., 2015).

The hidden information from the large dataset can be best unfolded by a data analysis methodology known as data mining techniques (Luan, 2002; Han, Pei, & Kamber, 2011). Data mining is now widely used in higher education, due to its latent qualities or abilities that may be developed and lead to future success or usefulness to educational institutes, a rising field known as educational data mining has evolved (Kumar & Chadha, 2011; Romero & Ventura, 2007). The education data mining community website www.educationaldatamining.org defines data mining as follows: “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 which they learn in” (Baker & Yacef, 2009). Educational data mining is about making predictions for the educational entities like students, faculties, staff and management with the objective to cater quality education among the students (Baradwaj & Pal, 2012).

Predictive analysis is about predicting the future considering current and past data. Predictive analysis is used to uncover the hidden patterns and capture the relationships of data (Baker & Yacef, 2009). Correct prediction, can be used to increase student’s grade, performance, retention rate, intake in a particular course, allocation and utilization of staff and resources in a more productive manner, decrease student dropout rate, identify the weak domain of students, increase institutes ranking (Delavari et al., 2008; Goyal & Vohra, 2012). If done at the right time, will not only help the management of the institute but in fact can help in the overall development of the student. Right guidance, right direction, right mentorship are necessary aspects for the academic as well as career growth of a student. In the present education system, grading of students is done on the basis of written exams, quizzes, assignments or tutorials. The result of these exams is to simply find the pass, failed or topper students. There is no way in the present system to identify the reasons why the student has failed, his/her weak areas, domain which he/she needs to focus, his/her intense areas, which results in the student’s dropout.

At present, student detainment is utmost important topics in higher education. It is observed that graduation rates have dropped which must be addressed immediately. According to the statistics given by Times of India in August 2016, it was unbelievable to know that 2000 students dropped out of IITs, IIMs in 2 years. According to them the data collected from the individual institutes for 2014 to 2016, it was astonishing to know that IIT-Delhi recorded the highest number of dropouts with 699 students. The second position was chased by IIT-Kharagpur with 544 students and third position was hacked by IIT-Bombay with 143 students (Chhapia, 2016). Student’s dropout of India’s best academic institutes is illustrated in Figure 1.

Diverse changes are experienced in the sphere of education; the revolution is tackled through analytics. The analytics process acts as an instrument for the education institutes to enhance the operational efficiency and gain competitive advantages (Sivarajah, Kamal, Irani, & Weerakkody, 2017). To understand why the student dropout rates are flourishing and retention rate is declining, it is critical to understand how the students select their courses and then why they dropout?
From the Cold War to Internet Cathedral
www.igi-global.com/article/cold-war-internet-cathedral/2264?camid=4v1a