

Chapter 7

Using Machine Learning Techniques in Student Dropout Prediction

Rajeev Bukralia

Dakota State University, USA

Amit V. Deokar

Dakota State University, USA

Surendra Sarnikar

Dakota State University, USA

Mark Hawkes

Dakota State University, USA

ABSTRACT

This chapter outlines a case of identifying students at-risk of dropping out of online courses by using institutional research data. The case delineates this step-by-step process that includes identification of appropriate constructs and variables, data collection, data pre-processing, data analysis, and model evaluation to develop a predictive model for student dropout in online courses at a small, public, Midwest university in the United States. Included is a comparative data analysis of various machine learning techniques, such as Artificial Neural Networks (ANN), Decision Trees, and Support Vector Machines (SVM), with statistical Logistic Regression (LR) analysis. The chapter provides steps for data analysis and predictive modeling using the open source, downloadable data mining software, WEKA. The chapter concludes with a discussion on the challenges and suggestions for building a predictive model in the context of institutional research.

DOI: 10.4018/978-1-60960-857-6.ch007

BACKGROUND

Student retention is a key issue in higher education. To be competitive in today's knowledge driven global economy, it is crucial that students pursue higher education and complete their degree in a timely manner. Obtaining a college degree is instrumental in securing employment opportunities. Due to the tremendous growth in e-learning in recent years, there is a pressing need to address the issue of student retention and attrition in online programs. Online enrollments have been growing at rates far in excess of the total higher education student population (Allen & Seaman, 2007). Dropout rates for distance classes have been consistently higher than those of traditional classes. Some studies have shown that dropout rates for distance education students are estimated to be 10-20% higher than the dropout rates of face-to-face students (Carr, 2000; Diaz, 2002). Lynch (2001) found student dropout rates in online courses were as high as 35 to 50%, as compared to 14% in traditional classes. Online courses have catered to distance learners, and they are becoming increasingly popular with on-campus students due to their greater flexibility. Successful completion of courses is essential for the successful completion of programs.

In this chapter, we present an example case of a small Midwestern university with a Carnegie classification of baccalaureate diverse fields. The university has an enrollment of 4722 students (duplicated head count, as of Fall 2010). Retention and the graduation rate have been priorities for the institution, along with enrollment. The graduation rate is approximately 25%, with freshman retention at 57%. Freshman retention has hovered around 57% for the last five years, with a slight increase in Fall 2010. The university has dedicated resources and implemented new ideas to improve the retention rate. One idea was to strengthen the Office of Institutional Research by hiring a full-time institutional research director in 2009. In the fall of 2010 the university initiated

the implementation of a comprehensive strategy for enrollment management. The overarching goal for the enrollment management strategy was to improve enrollment, as well as to achieve improved retention and graduation rates. The strategy included an analysis of all academic and student support procedures and policies, including promoting the university to potential students and parents, student recruitment, advising, preregistration, registration, analyzing the university environment to improve social integration of students, closely monitoring student progress, greater involvement of the institutional research office to facilitate data-oriented decision making, and restructuring some academic and student support units for improved enrollment management.

Outlined in the university strategic plan is the goal to improve student retention and boost enrollment. Some ideas, including a greater emphasis on recruiting better academically prepared students and improving the preregistration process, helped the university increase enrollment by nearly 650 additional students in Fall 2010 from the previous year. The increase in enrollment invigorated the university's enrollment management strategy, and the university is continuing to move forward with the comprehensive enrollment strategy. An important objective of the enrollment strategy is to identify students who are more likely to drop out—students who are at-risk of dropping out. At-risk students are those who may be at greater risk to drop out of a course or program. It is important to identify at-risk students at both the course and program level because dropping out of a program affects retention and graduation rates, and prior to dropping out most students exhibit a pattern of course dropout or failure.

CASE DESCRIPTION

The institution has witnessed growth in online course enrollment in recent years. This increased enrollment occurred partly because of an increased

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/using-machine-learning-techniques-student/60843

Related Content

Text Mining by Pseudo-Natural Language Understanding

Ruqian Lu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1942-1946).

www.irma-international.org/chapter/text-mining-pseudo-natural-language/11085

Stages of Knowledge Discovery in E-Commerce Sites

Christophe Giraud-Carrier (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1830-1834).

www.irma-international.org/chapter/stages-knowledge-discovery-commerce-sites/11067

Distributed Data Aggregation Technology for Real-Time DDoS Attacks Detection

Yu Chen and Wei-Shinn Ku (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 701-708).

www.irma-international.org/chapter/distributed-data-aggregation-technology-real/10897

Search Engines and their Impact on Data Warehouses

Hadrian Peter (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1727-1734).

www.irma-international.org/chapter/search-engines-their-impact-data/11051

Data Streams

João Gama and Pedro Pereira Rodrigues (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 561-565).

www.irma-international.org/chapter/data-streams/10876