Academic Analytics

Si Chen
Murray State University, USA

INTRODUCTION

In recent years, as business analytics is used by more and more companies to drive their decision making, a similar phenomenon has been observed in higher education. In particular, an increasing number of colleges and universities have adopted data-driven decision making practices throughout their administrative and academic units. Such development has given rise to an emerging field known as Academic Analytics. The term "academic analytics" was first formally introduced by Goldstein et al., 2005 in which the authors describe it as the application of the principles and tools of business intelligence to academia. In this survey we further define academic analytics as the broad applications of both business intelligence, which traditionally focuses on query and reporting, as well as predictive modeling and data mining techniques for decision making in higher education.

This survey focuses on papers that deal with the methodology and application of academic analytics over the last decade. It also discusses two commercially developed academic analytics platforms.

BACKGROUND

Colleges and universities generate and manage a wide array of data from different sources. Analytics helps administrators and faculties extract useful information and knowledge from this data. Goldstein et al., 2005 surveys the applications of academics analytics at more than 380 higher education institutions and evaluates the impacts that analytics has had on institutional decision making, business processes and individuals. The authors conclude that academic analytics has great potential to improve an institution's ability to meet the needs of the students and to answer calls for greater accountability.

MAIN FOCUS

Academic analytics can be used in different areas in higher education. Antons et al., 2006 documents a successful application of data-mining techniques in enrollment management through a partnership between the admissions office, a business administration master's-degree program, and the institutional research office. Blanc et al., 2009 analyzes a large sample of 33,000 university alumni records and clusters them into six groups which are evaluated using discriminant analysis for different giving patterns. Nandeshwar et al., 2009 develops a decision tree model to explain admissions data at West Virginia University and finds that financial aids is the leading factor to attract students to enroll. As the demand for institutional accountability continues to rise, analytics for the purpose of improving student success has emerged as the main focus of academic analytics.

Student Success

Student success is often measured by student retention or student graduation rate. It has always been a major concern in the higher education community. Indeed, our universities' ability to produce high quality graduates plays a key role in the future of the nation as we confront the challenges of globalization. Additionally, it affects the amount of state funding received by public universities and the eligibility of for-profit schools

DOI: 10.4018/978-1-4666-5202-6.ch004

and certificate programs at public schools for federal student loans.

Improve Student Success with Academic Analytics

Campbell et al., 2007 defines the five steps of using analytics to improve student success as capture, report, predict, act and refine. In the capture step, data is obtained from multiple sources and may include both time-invariant variables (e.g., demographics, academic history) and time-variant variables (e.g., logs from course management systems such as Blackboard). In the report step, descriptive statistics and correlations from the data can be generated and presented by "dashboards". In the capture step, various data mining techniques can be used to develop predictive models. In the act step, appropriate actions such as interventions for at-risk students are carried out based on the results from the predictive models. In the refine step the usability of the models is tested with updated data and necessary changes are implemented.

As the greatest attrition tends to occur during the first two years of college, studies on student retention tend to focus on freshman and sophomore students. Salazar et al, 2004 uses C-means clustering algorithm to group students into different clusters. A set of decision rules is subsequently generated and used to gain insights into the factors influencing student academic achievement success and failure, student retention and student desertion. Bailey, 2006 uses the Integrated Postsecondary Education Data System (IPEDS) data to develop models that calculate predicted graduation rates for two- and four-year institutions. Superby et al., 2006 analyzes data from questionnaires completed by 533 first-year students using decision trees, random forests, neural networks and linear discriminant analysis. The authors develop predictive models that can be used to classify freshman students into different groups based on their likelihood to graduate. Sujitparapitaya, 2006 presents a case study where multinomial logistic regression, C5.0 rule induction and neural network are used to identify the critical predictors influencing the attrition of first-time freshmen. Herzog, 2006 compares the prediction accuracy of decision tree, neural networks and logistic regression on predicting freshmen retention. Their study suggests that the selection of a particular method may be determined by the level of complexity of the data used and the outcome predicted. Miller et al., 2008 applies logistic regression to predict the risk of individual student attrition based on pre-matriculation (to stay or to depart) characteristics. Delen, 2010 applies several data mining techniques (both individuals and as ensembles) to develop analytical models. These models are used to predict and to explain the reasons behind freshmen student attrition. Yu et al, 2010 explores the issue of student retention using classification trees, multivariate adaptive regression splines (MARS) and neural networks. They examine three sets of potential predictors, including demographic, pre-college or external academic performance indicators and online class hours as a percentage of total hours during the sophomore year.

Real-Time Analytics

Analytics can also be used to monitor students' progress towards obtaining a degree and provide real-time suggestions or interventions when at-risk students are identified. This type of application has seen an increase in recent years with the prevalence of course management systems. Minaei-Bidgoli et al., 2003 uses classification to predict students final grades based on their Web-use features. Students predicted to be at risk are provided with tutoring service. Baepler et al., 2010 describes several applications where data captured from course management systems is used to gain insights into student performances. Zhang et al., 2010 uses data mining and natural language processing technologies to monitor students' academic progress and provide suggestions when potential problems are identified.

Related Content