

Chapter 94

Assessing Data Mining Approaches for Analyzing Actuarial Student Success Rate

Alan Olinsky
Bryant University, USA

Phyllis Schumacher
Bryant University, USA

John Quinn
Bryant University, USA

ABSTRACT

One way to enhance the likelihood that more students will graduate within the specific major that they begin with is to attract the type of students who have typically (historically) done well in that field of study. This chapter details a study that utilizes data mining techniques to analyze the characteristics of students who enroll as actuarial students and then either drop out of the major or graduate as actuarial students. Several predictive models including logistic regression, neural networks and decision trees are obtained. The models are then compared and the best fitting model is determined. The regression model turns out to be the best predictor. Since this is a very well understood method, it can easily be explained. The decision tree, although its underpinnings are somewhat difficult to explain, gives a clear and well understood output. Not only is the resulting model a good one for predicting success in the major, it also allows us the ability to better counsel students.

INTRODUCTION

One of the significant issues confronting higher education is the area of enrollment management with specific attention paid to student retention. In an analogous way, on the departmental level,

university faculty and administration members are often interested in determining what percentage of students who begin in a certain major will eventually graduate in the same major. This information is important for individual departments in determining how to allocate limited resources in making decisions as to the appropriate number of classes and sections to offer and the number

DOI: 10.4018/978-1-4666-2455-9.ch094

of faculty lines needed to staff the department. It is also important for admissions departments in deciding which students to target for specific programs and majors and for enrollment management in addressing student retention.

One way to enhance the likelihood that more students will graduate within the specific major that they begin with is to attract the type of students who have typically (historically) done well in that field of study. This chapter will detail a study that utilizes data mining techniques to analyze the characteristics of students who enroll as actuarial mathematics students and then either drop out of the major or graduate as actuarial students.

BACKGROUND

In a previous paper (Schumacher et al., in press), data mining techniques were applied in a study that investigated the likelihood that incoming college freshmen majoring in Actuarial Mathematics (AM) will graduate in this major. The study applied data mining to an earlier study which predicted success using only traditional logistic regression. The original study contained data spanning seven years of incoming university freshmen who started as AM majors in the years 1995-2001 (Smith and Schumacher, 2006).

There have been other recent studies utilizing the various techniques of data mining applied to issues within higher education. For example, in one comprehensive paper (Davis et al., 2008), predictive models were generated for three important educational concerns: student retention, student enrollment and donor giving. In another study (Herzog, 2006), used logistic regression, decision trees and neural nets to predict student retention and degree completion time for new and transfer students. Similarly, student retention was analyzed through six-year graduation predictive models which were developed with the use of various data mining techniques (Campbell, 2008).

Furthermore, data mining methods, including neural nets and random forests were applied to an investigation (Vandamme et al., 2007) of academic success among first year college students. A data mining approach to predicting the disposition of admitted students as enrollees or nonenrollees was completed in another investigation (Antons and Maltz, 2006). There are additional papers involving applications of data mining within a university setting which are cited in the previous study (Schumacher et al., in press).

Since the goal of the original study was to predict whether or not a student graduated in the major, this became the target variable. The input variables included gender, math and verbal SAT scores, percentile rank in high school class and percentile rank on a department mathematics placement test. These variables were chosen from among data available from the admissions department collected from incoming students because they were known to be relevant to forecasting the student's grade point average in their concentration (Smith and Schumacher, 2005). The variable high school rank in class did have more missing values than high school GPA, which was available and could have been used. However, rank in class had previously been shown to be a better predictor of student success in the 2005 study and so was used in the 2006 and the 2010 studies as well.

The data mining procedures used in the previous study included logistic regression, decision trees and neural networks. The main conclusion drawn from this study was that these three approaches provided similar results with respect to predictive ability but that the decision trees had the additional advantage of treating cases with missing data without imputation. It was further concluded that if one technique was to be specified, then the pruned tree appeared to be the best choice in that it provided a simple set of rules to follow and avoided over-fitting of the data, that is, the phenomena of obtaining an exceptionally well fitting model that is not applicable to new data.

14 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/assessing-data-mining-approaches-analyzing/73524

Related Content

Analyzing Social Emotions in Social Network Using Graph Based Co-Ranking Algorithm

Kani Priya, Krishnaveni R., Krishnamurthy M. and Bairavel S. (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 329-341).

www.irma-international.org/chapter/analyzing-social-emotions-in-social-network-using-graph-based-co-ranking-algorithm/308495

Automated Integration of Heterogeneous Data Warehouse Schemas

Marko Banek, Boris Vrdoljak, A Min Tjoand Zoran Skocir (2008). *International Journal of Data Warehousing and Mining* (pp. 1-21).

www.irma-international.org/article/automated-integration-heterogeneous-data-warehouse/1815

A Hybrid Decomposition and Deep Learning Model for Photovoltaic Power Forecasting Under Variable Meteorological Conditions

Liusong Huang, Adam Amril bin Jaharadak, Nor Izzati Ahmad and Jie Wang (2025). *International Journal of Data Warehousing and Mining* (pp. 1-22).

www.irma-international.org/article/a-hybrid-decomposition-and-deep-learning-model-for-photovoltaic-power-forecasting-under-variable-meteorological-conditions/388673

Parallel Computing for Mining Association Rules in Distributed P2P Networks

Huiwei Guan (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 107-124).

www.irma-international.org/chapter/parallel-computing-mining-association-rules/73436

Visual Data Mining for Collaborative Filtering: A State-of-the-Art Survey

Marenglen Biba, Narasimha Rao Vajjhala and Lediona Nishani (2017). *Collaborative Filtering Using Data Mining and Analysis* (pp. 217-235).

www.irma-international.org/chapter/visual-data-mining-for-collaborative-filtering/159505