Using Business Intelligence in College Admissions: A Strategic Approach

W. O. Dale Amburgey, Saint Joseph’s University, USA
John C. Yi, Saint Joseph’s University, USA

ABSTRACT

Higher education often lags behind industry in the adoption of new or emerging technologies. As competition increases among colleges and universities for a diminishing supply of prospective students, the need to adopt the principles of business intelligence becomes increasingly more important. Data from first-year enrolling students for the 2006-2008 fall terms at a private, master’s-level institution in the northeastern United States was analyzed for the purpose of developing predictive models. A decision tree analysis, a neural network analysis, and a multiple regression analysis were conducted to predict each student’s grade point average (GPA) at the end of the first year of academic study. Numerous geodemographic variables were analyzed to develop the models to predict the target variable. The overall performance of the models developed in the analysis was evaluated by using the average square error (ASE). The three models had similar ASE values, which indicated that any of the models could be used for the intended purpose. Suggestions for future analysis include expansion of the scope of the study to include more student-centric variables and to evaluate GPA at other student levels.

Keywords: Decision Trees, Enrollment Management, Enrollment Strategy, Neural Networks, Predictive Models

INTRODUCTION

Higher education has long been rich in data but slow in converting that data into useful information. In institutions ranging from large public research universities to small liberal arts colleges, vast amounts of data are collected by every internal entity. Some of the largest amounts of data are captured within universities’ enrollment management divisions. Admissions offices are inundated with geodemographic data on prospective students. Financial aid offices constantly collect data points relating to the personal or family financial situations of prospective and current students. Retention offices collect data to help identify students that may be at risk of dropping out. Enrollment management divisions are among the largest data collectors in higher education; however, they tend to lag behind the corporate world in conversion of data into usable information.

With the voluminous amounts of data collected within enrollment management divisions,

DOI: 10.4018/jbir.2011010101
only within the past decade has there been a concerted effort to use that data to develop predictive models. Consulting groups have added enrollment management services to capitalize on the popular cultural shift to use of historical data to develop predictive analytics.

One of the most common uses of predictive analytics in enrollment management is for forecasting future first-year student enrollments. Many institutions, especially private colleges and universities, are tuition dependent, with most of their net revenues generated by student tuition. Being able to accurately forecast the number of entering students each year enables them to better plan and strategize improved benefits and services for all members of the college or university community.

PURPOSE OF THE STUDY

The purpose of this study is to develop a predictive model to assist undergraduate admissions officers in determining the likelihood of academic success for entering first-year students.

Incorporating into the admissions process a predictive model to identify the potential for success can be very advantageous. University admissions offices are seeing an increasing percentage of the applicant pool fall into a marginal category. Marginal applicants are loosely defined as those who are not definite admits or definite denials. These students’ academic credentials are not as sound as those of the upper-tier applicants but significantly better than those of unsuccessful applicants. Using a predictive model to determine applicants’ potential to have a strong grade point average (GPA) at the end of the first year should help alleviate most of the conjecture currently applied to making admissions decisions about marginal applicants.

As pertinent data is collected during the initial inquiry stage, these predictive models may be used to shape recruitment strategies and to target a specific message to the many audiences in the inquiry pool. For example, marketing messages relating to tutoring services or student success programs may be directed to applicants identified as having a low likelihood of earning a high end-of-first-year GPA.

Admissions counselors may also use predictive models to better counsel prospective students during their college search. Admissions representatives can counsel prospective students who display characteristics known to indicate academic distress about the possibility of future success. These discussions can help prospective students determine whether the rigor of the institution’s academic environment is suitable to their skills and abilities.

STUDY DESIGN AND METHODOLOGY

The primary methodology of the study consists of analysis of historical student data to determine the best-fit model to predict applicants’ end-of-first-year GPA. Three types of analytical models will be developed, and comparison testing will be conducted to determine the model displaying the lowest error.

Data stewards of the institution representing the Office of the Registrar, the Office of Financial Assistance, and the Director of Enrollment Analysis conferred to develop standards of acceptable use of the historical data. All agreed that the potential results of the study were significant enough to justify use of the data and that the study had to strive to protect the anonymity of the student data.

DATA COLLECTION AND ANALYSIS

Data was collected for entering first-year students at a midsized private university in the northeastern United States. The data was collected for first-year students beginning studies in the fall academic terms from 2006 through 2008. The data was collected from numerous sources within the enrollment management division and in other divisions of the university. The Common Application, the university application, the College Board, and the Free Application
International Journal of Business Intelligence Research, 2(1), 1-15, January-March 2011

for Federal Student Aid (FAFSA) were some of the sources for collection of applicant data. To assure the individual students’ anonymity, unique identification numbers were generated to replace any identifiable student identification numbers in the data set. Three types of predictive models—a decision tree analysis, a neural network analysis, and a multiple regression analysis—were developed for comparison of predictability as determined by the average square error (ASE).

SIGNIFICANCE AND LIMITATIONS

Several limitations were encountered during the study. An initial limitation is that data was collected only from entering first-year students at the participating institution. Data from applicants who had not completed the application process or who had been admitted but had chosen to attend another institution was not considered.

Another limitation is that all data is assumed to be truthful and accurate. The data collection took place by manual data entry and electronic uploads. It is assumed that there were no significant data entry errors and that all electronic upload processes were functioning correctly.

Since data from the FAFSA was used, the possibility of a large number of missing variables is present because entering students are not required to complete the form. At the participating institution, it is assumed that approximately 40% of entering first-year students will complete the FAFSA.

The FAFSA data also represents the most recent financial aid data available for entering students. An entering student may have submitted several iterations of the FAFSA, and the potential for numerous changes within the data variables is possible.

LITERATURE REVIEW

Use of data to aid decision making has long been accepted in the corporate world but has been slower to gain momentum in higher education. Some higher education professionals anecdotally state that it is not unusual for the lag time for adopting corporate best practices in an academic setting can be as long as a decade. This statement, if accurate, describes a process that could be a mixed blessing for those in the academic world. Practices and procedures tend to undergo several revisions until they become best practices, and higher education may benefit from this delay by not having to be concerned about establishing a practice until the revision cycle is complete. If anything, higher education is not an area where constant change is embraced.

A shift in the mindset of enrollment management professionals must take place if the true potential of using business intelligence to aid in achieving goals and objectives is to be realized. It is imperative to gain an overview of enrollment management to better understand the historical philosophies of the practice and the motivations behind the implementation of those philosophies.

Overview of Enrollment Management

Approximately four decades ago, the need to develop more structured marketing communications and the benefits of developing relationships between prospective students and the institution were identified. During the early 1970s, the term enrollment management was first used in conjunction with student recruitment and retention. During the 1980s, institutions began developing and adhering to detailed enrollment plans. These plans coupled the use of integrated marketing communication plans with the strategic use of financial aid packages (Merante, 2009, p. 5).

As funding for higher education fluctuated in the late 1980s and certain demographics began to shift, the enrollment plans became more aggressive to overcome obstacles in the marketplace. The enrollment plans during this period became more vigorous, and the use of direct mail and targeted marketing campaigns
became established practices at many institutions (Fiske, 2008; Merante, 2009, pp. 5-6).

By the middle of the 1990s, enrollment management continued to develop and incorporated a more strategic nature than had been used previously. During this period, the multitude of departments that represented functions of enrollment management was collectively grouped to form a unique division. Also during this period, the marketplace became better informed and began to display the desire for institutional data. Prospective students and their families began asking questions about placement rates, graduation rates, and descriptive academic statistics as had never been seen previously (Fiske, 2008; Merante, 2009, p. 6).

The new millennium has brought developments in enrollment management that could not have been foreseen a decade earlier. The college view book was at one time the most visible and significant piece of marketing that an institution’s admissions office could produce. Now, view books have become much smaller as the focus has switched toward use of websites and social media to promote the institution to prospective students. The pace of technological innovation is ensuring that the future of recruitment and retention activities has the possibility of being personalized and delivered in a format that recipients choose and at any time they want (Fiske, 2008; Merante, 2009, p. 6).

Enrollment management is a term that may seem understandable in theory but proves difficult to define in practice. Sutton (2007) defines enrollment management as “simply a name that has been given to the evaluation of data and implementation of planning that will result in a more efficient academic institution” (p. 1). This definition seems rather broad, but it does incorporate all the activities that may fall under the enrollment management umbrella. Stewart (2004) states that “as a process, enrollment management helps institutions: develop a keener awareness of their purpose and character in relation to the student marketplace, improve ties to prospective client groups, and attract students into and through the institution” (p. 21). This description of enrollment management begins to encompass the “whole life” nature of the process, beginning with the initial inquiry and concluding with the transformation from a graduated student to an active alumnus.

In some institutions, enrollment management is used interchangeably with the undergraduate admissions office. Often, this is perpetuated by the massive infusion of new revenue to the institution that results from the enrollment of each new undergraduate entering class. This is especially relevant at private institutions, where tuition revenue is especially vital to funding institutional operations (Antons & Maltz, 2006, p. 69).

The two most critical components of the fiscal nature of enrollment management are the yield rate of admitted students and the tuition discount rate. The yield rate is calculated by estimating the percentage of admitted students who will matriculate at the institution. Fluctuations in the yield percentage can have significant effects on net tuition revenue. The discount rate “is the projected financial aid to be allocated to students as a percentage of tuition” (Antons & Maltz, 2006, p. 69). The ability to offer financial assistance to students is an important component of prospective student recruitment and current student retention. A discount rate that is too high can result in the institution’s overspending its financial aid budget and placing considerable strain on its other institutional resources. One that is too low can lead to its under-achieving its enrollment goals and drastically reducing its net tuition revenue (Antons & Maltz, 2006).

The composition of enrollment management divisions lacks complete consistency but it does have a few root components. As mentioned earlier, some see enrollment management as being synonymous with undergraduate admissions. This is usually because undergraduate admissions are the largest and most visible component of the enrollment management division. The financial aid office is another important component of the enrollment management division that has a presence throughout the entire student life cycle. This office is instrumental in the student recruitment process by making the institution affordable to entering students.
It also plays a crucial role in student retention by assisting current students with the financial obligations they have undertaken.

Although many may not see the connection, the future of university admissions and the strategic use of financial aid are built on the foundations of business intelligence. As technological advances permit more prospective student data to be captured for analysis, the number of data-mining professionals housed institutionally and of external consulting groups offering predictive analytical services is poised for tremendous growth.

Developing predictive models to assist in the marketing, recruitment, and admission of prospective students holds the potential to leverage an institution's history in making the best strategic decisions not only for the organization, but for the prospective student as well. The scarcity of fiscal resources, in concert with the increasing number of educational alternatives, will necessitate the use of business intelligence to compete in the higher education marketplace.

Use of Predictive Analytics to Determine Educational Outcomes

Some have embraced the use of business intelligence principles within higher education. The primary goal of data mining should be to analyze data and convert it to useful information on which informed decisions can be made. The vast amount of data collected at colleges and universities provides a ripe environment to yield the benefits of business intelligence.

One of the most important uses of data within enrollment management, and especially within a university admissions office, is its use in predicting outcomes. The ability to analyze variables to determine the individual’s potential for success should be an important component of decision making and strategic planning in an admissions office.

There are numerous examples where predictive analytics have been used to forecast educational outcomes. The art and science of making college admissions decisions can greatly benefit from the use of business intelligence and predictive analytics. Some variables used in the admissions decision process are standardized while a great number may be unique to an individual high school.

Tam and Sukhatme (2004) evaluated the potential for a student’s college academic success on a newly constructed high school percentile rank variable (p. 12). The researchers used data collected from the enrolling freshmen cohort at the University of Illinois at Chicago (UIC) for fall term 1994. The academic progress of the cohort was tracked over a six-year period with the definition of success being graduation from the institution within that time window (p. 13). Historically, the institution arrived at freshman admissions decisions by calculating a selection index determined by the ACT score and the percentile class rank. Since the ACT is a standardized test, all students had an equal chance of achieving a high test score. However, earning a higher-percentile class rank proved more difficult because it would be more difficult for a student to achieve a higher class rank in a more academically challenging high school. This motivated the researchers to measure the academic quality of the participants’ high schools. However, they were faced with a multitude of variables that could be considered when defining high school quality (Tam & Sukhatme, 2004, pp. 12-14).

The researchers decided to use a high school’s average ACT score for all students as the means of defining academic quality. They made this choice primarily because of the availability of data for analysis. In evaluating the admissions decisions made over the six-year period, Tam and Sukhatme found that much better admissions decisions could have been made by using the high school academic quality indicator. The results led to the generation of a new variable called the modified student high school percentile rank to be incorporated into the UIC admissions decision criteria. Including the new criterion in the selection process proved to yield much better admissions decisions as measured by student success (Tam & Sukhatme, 2004, pp. 12-14).
Predictive analytics are beneficial to more areas than just undergraduate admissions. Making decisions at the graduate level sometimes proves to be as difficult as or even more difficult than at the undergraduate level. Naik and Ragothaman (2004) evaluated ways to improve prediction of MBA (Master of Business Administration) students’ performance by using predictive analytics to assist with admissions decision making.

In their research, Naik and Ragothaman compared the outcomes of using neural networks, logit, and probit models to make MBA admissions decisions. They incorporated 10 explanatory variables into their research, encompassing areas such as campus location, undergraduate major, GMAT (Graduate Management Admission Test) scores, and undergraduate institution. The applicants were divided into successful and marginal groups based on many of the explanatory variables defined (Naik & Ragothaman, 2004, pp. 143-144).

The purpose of Naik and Ragothaman’s research was to determine whether neural networks could be used in the MBA admissions decision-making process with the same level of effectiveness as the more traditional statistical models such as probit and logit models. The outcomes proved that the neural network performed as well as the probit and logit models in predicting MBA student performance. Their final recommendation was that the incorporation of neural networks into the MBA decision-making process could assist MBA directors and business school deans in making better admissions decisions on which applicants to accept in their respective programs (Naik & Ragothaman, 2004, pp. 146-147).

Often, those charged with making admissions decisions at any level do not think of incorporating any method of business intelligence into the process. The innate instinct is to try to evaluate the applicant based on the academic and professional credentials contained in his or her record. However, the potential to make even better decisions using statistical data while relying less on the proverbial “gut instinct” lies with employing business intelligence principles.

Another interesting study, conducted by Johnson (2008), examined the relationship of the type of high school a student attended to enrollment, persistence, and graduation rates at a public university. The study attempted to build on previous studies relating to college choice and persistence by incorporating the effects of high school–related and individual student variables on academic success (p. 777).

The research questions guiding the study centered on the characteristics of high schools that had a higher probability of graduating students who would enroll in and succeed at the university, along with the effect of student attrition rates, depending on the type of high school attended. The research was conducted at a doctoral/research university with an enrollment of approximately 12,000 students. It focused on the traits of in-state students at the institution, who composed almost 80% of the 2,000 entering first-year students enrolling directly from high schools. Data was collected from students over the course of the fall 2001-2005 entering cohorts. Some of the high school attributes examined related to academic quality, poverty levels, and ethnic composition. Data collection used internal, institution-specific data, along with external sources, such as a U.S. Department of Education database (Johnson, 2008, p. 779).

The data analysis included development of a matriculation model, a persistence model, and a graduation model to estimate the desired effects. The matriculation model was used to estimate the effects on the odds of a student’s enrolling in the institution. The persistence model evaluated the odds of an enrolled student’s being retained after the first year. Finally, the graduation model evaluated the likelihood of an enrolled student’s graduation from the institution within five years (Johnson, 2008, p. 793).

The results of Johnson’s (2008) research with the three independent models indicated that the type of high school did matter when evaluating the dependent variables of enrollment, persistence, and graduation. Understanding how the individual-related and high school–related variables can be evaluated to provide the odds
of the dependent variables can be critical to enrollment managers and admissions officials. The models used in this study can have strong implications on planning recruitment and enrollment strategies because they can effect decisions ranging from marketing messages to enrollment decisions.

Allen and Robbins (2008) conducted a large-scale research project to develop a predictive model that would assess the probability of a first-year student’s persisting in his or her chosen major area of study. They used three hypotheses to guide the purpose of their study: The first was that “major persistence is predicted by interest-major fit” (p. 65). The second stated “that students with higher first-year GPA are more likely to persist in their entering major” (p. 65). The final hypothesis was “that indicators of pre-collegiate academic preparation (high school GPA and ACT Composite score) are related to major persistence, but are mediated by high school GPA” (p. 65). For the purpose of their research, major persistence was determined to have occurred if the student remained in his or her initial major in the third year.

The initial sample population included almost 88,000 first-time entering students from 25 four-year colleges and universities. The outcome of concern for the researchers was the third-year major. To remain in the sample, a student had to be enrolled in a major during his or her third year of study, and his or her initial major program of study selection had to be known. Applying this parameter to the sample reduced the sample size to just over 48,000. This sample size allowed for the sample to be split into an Estimation group and a Validation group. The analysis led to the development of a regression equation generated from the Estimation group that was tested against the Validation group (Allen & Robbins, 2008, p. 65).

The research supported the claims that interest-major fit and the academic performance at the end of the first year were both instrumental in predicting whether or not a student remained in his or her major of choice. Understanding this relationship can be extremely beneficial to academic advisors at colleges and universities as they assist and counsel students early in their college career. The predictive model that was developed could aid students in finding a major closely aligned with their interest and abilities, while providing the framework to improve the quality of their college experience as a whole (Allen & Robbins, 2008).

**Use of Predictive Analytics to Determine Potential Enrollment**

Business intelligence has tremendous potential to help admissions and enrollment management professionals make decisions in areas other than the applicants’ potential success. The voluminous amounts of data that filter through an enrollment management division may be analyzed to determine the potential enrollments of prospective students. The benefits of using business intelligence go well beyond predicting the number of enrolling students. Decisions about recruitment strategies and budget expenditures can be affected by the same predictive models that forecast enrollments.

Chen (2008) examined ways predictive modeling could be used to determine enrollments at a university in the Midwest. What makes Chen’s discussion interesting is that the research was truly a longitudinal study that analyzed enrollment data from 1962 to 2004. The research evaluated the predictive nature of 15 variables and incorporated individual student indicators, along with economic indicators. These variables were used to develop two predictive models for comparison in predictability (p. 1).

An autoregressive integrated moving average (ARIMA) model and a linear regression model were developed to predict institutional enrollments. When comparing the two models, the ARIMA model yielded an $r^2$-squared value of 0.96 and a mean absolute percentage error (MAPE) of 2.11%. The linear regression model had an $r^2$-squared value of 0.97, just slightly better than the ARIMA model. The MAPE of the linear regression model was 1.62%, which was almost half a percentage point less than the ARIMA model. The high $r^2$-squared values indicate that well over 95% of the variability
in the data set is accounted for by the variables contained within the respective model. Also, the state high school graduate variable and the one-year lagged institution enrollment variables emerged as primary enrollment functions of the institution in both models (Chen, 2008).

The research of Goenner and Pauls (2006) involved the development of predictive analytics to aid in the marketing efforts an institution made in response to enrollment inquiries. The researchers hoped to prove that analysis of the inquiring student’s behaviors could be used to positively affect the institution’s marketing efforts. The interesting approach to the research was that the analysis focused on the initial inquiry and the inquirers’ eventual enrollment decision rather than on those who had applied or been accepted to the institution (p. 936).

Because the manner by which a prospective student inquires often results in limited amounts of data, the researchers also built in geodemographic data to aid in developing the model. A sample of almost 16,000 inquiring students was used to develop the predictive model, with the dependent variable being a calculated binary variable associated with enrolling. Given the presence of the binary variable, a logistic regression model was used for the analysis (Goenner & Pauls, 2006, pp. 946-947).

The results of the analysis yielded a model that predicted enrollment behavior almost 90% of the time. The model had a sensitivity, or the ability to predict the number of students who eventually enrolled, of 36%. Conversely, the model had a specificity, or the ability to predict the number of inquirers who would not enroll, of 97%. Using the model allowed the institution to better segment marketing campaigns to prospective students. The predictive model had been used to score ZIP codes so that marketing campaigns could be directed to a more concentrated area, where the potential for enrollment truly lay (Goenner & Pauls, 2006, p. 953).

Chang (2006) also evaluated the use of business intelligence to predict admissions yield through a case study at a large state university. One of the questions that Chang’s research attempted to answer was whether applicants enrolled randomly or in identifiable patterns. Also, the presence of enrollment trends within specific groups and the accuracy of enrollment forecasts were questions that guided the study (p. 54).

The data set of the research contained nearly 3,000 admitted first-year, transfer, and nondegree students at a large state university. Data was analyzed, and an enrollment status indicator variable was added for purposes of modeling. The study contained C&RT (classification and regression trees), neural networks, and logistic regressions as the predictive models for analysis. The data set was divided into a Training group, to be used for model development, and a Testing group, to be used for model validation. All the models used generated predictions at the individual student level (Chang, 2006, pp. 62-63).

The research concluded that use of a data-mining approach to predictive modeling yielded better results than use of traditional statistical forecasting tools such as logistic regression. The data-mining models provided outcomes that could be considered extremely actionable and preferable because of the potential to be connected to a live database. Modeling on live data offers a tremendous opportunity to make enrollment decisions in the moment and to amend strategies and goals in a more immediate manner (Chang, 2006, p. 68).

Summary

Enrollment management is an extremely dynamic component of higher education where the advantages of business intelligence could aid in the process of decision making on a multitude of issues. Though the structure of an enrollment management division may be unique to an institution, the vast amounts of data collected remain consistent. Also consistent is the general lack of motivation to mine the data to determine potential use of the information that may be present. As population demographics continue to shift in the future, the necessity to apply business intelligence practices to higher education will become critically important.
Research has shown that there are instances where predictive analytics have been used to successfully address issues within higher education. The most common themes present in those instances are the ability to predict applicants’ future academic success or to forecast the size and composition of an entering class. The potential benefits that lie with having accurate predictive models can only aid enrollment managers as they are forced to become more competitive in recruitment of prospective students and increasing retention of current students.

Demographic shifts occurring in some regions of the United States may force institutions to adopt the use of predictive analytics in enrollment and marketing strategies. As the number of high school graduates trends downward in some areas, higher education institutions must look to predictive analytics to maximize the yield from a potentially smaller pool of inquirers and applicants. Conversely, as demographic trends track upward in other areas, predictive analytics become an integral component in enrollment planning. A growing inquiry and applicant pool has implications for institutional selectivity and increases the potential for over enrollment. Applying business intelligence principles to these situations will provide institutional leadership with the road map needed to achieve goals while maximizing net tuition revenue.

DATA ANALYSIS

In the present study, the data for analysis was retrieved from a private, midsized university in the northeastern United States. The data represented first-year students enrolling for the institution’s fall academic term from 2006 through 2008. The data set included 3,576 student records, and it contained 40 variables that represented an array of academic and demographic data. The dependent variable for this analysis was the interval data value End-of-First-Year GPA. This variable represented the GPA at the end of the spring semester and was the focus of prediction for this study. All variables used in analysis are found in Table 1.

Data was collected in cooperation with the various data stewards at the participating institution, and necessary precautions to maintain student anonymity were enacted. This included generating a dummy identification variable to

| Table 1. Variables used for decision tree, neural network, and multiple regression analysis |
|---------------------------------------------|------------------|------------------|
| Variable                        | Role            | Level            |
| End-of-First-Year GPA            | Target          | Interval         |
| Citizenship Type                | Input           | Nominal          |
| Gender Description              | Input           | Nominal          |
| High School Quality Index       | Input           | Interval         |
| Institutional Characteristic Description | Input           | Nominal          |
| Major Description               | Input           | Nominal          |
| Religion Description            | Input           | Nominal          |
| SAT Mathematics Score           | Input           | Interval         |
| SAT Verbal Score                | Input           | Interval         |
| SAT Writing Score               | Input           | Interval         |
| SAT Total Score                 | Input           | Interval         |
| Secondary School GPA            | Input           | Interval         |
| State_Province                  | Input           | Interval         |

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
prevent personally identifiable features within the data set from being used for identification purposes. Data was also analyzed to identify any erroneous variables. The data is considered to be truthful and representative of the demographics of the study sample.

The data analysis represented two methods pertaining to data mining and a traditional statistical analysis method. The decision tree and neural network analyses are more closely associated with data mining and predictive analytics, while the multiple regression analysis is more rooted in the realm of traditional statistical analysis. Comparing the results of the three analyses allows for the selection of a model that will most accurately predict the level of success a student may experience during his or her first year of college academic study. The resulting model may then be used in a variety of ways to make an enrollment management decision, especially in the undergraduate admissions office.

Decision Tree Analysis

Decision tree analysis is a data-mining tool used to reduce a large number of data records into smaller and smaller collections by applying a series of decision rules. Decision trees are based on use of a series of “if-then” statements that sequentially divides the data into smaller, more homogenous groups. The eventual outcome is to arrive at a decision tree that will identify numerous target groups in relation to the numerous variables input into the model. Decision trees also have an interactive nature that permits the ability to prune the tree in numerous ways (Berry & Linoff, 2004, Matignon, 2007).

In this study, the data set was partitioned into a Training group and a Validation group for the purpose of conducting a decision tree analysis. The End-of-First-Year GPA was used as the target variable, and the remaining variables were defined (see Table 1) as input variables. The basis for measuring the overall performance for the model was the ASE of the validation population.

The decision tree analysis yielded an ASE of 0.331 for the Training group and of 0.379 for the Validation group. The assessment plot indicated that the ASE was the smallest with a decision tree containing 12 leaves. However, using the ASE of the Validation group, the decision tree could be trimmed to three leaves and increase the ASE by only 0.009. This is represented in Figure 1, where a noticeable flattening of the ASE values for the Validation group occurs after Leaf 3.

The decision tree analysis output also generated a ranking of the importance of the variables. This value was based on the number of occasions when a variable was used in splits within the model. The Secondary School GPA was the variable of most importance, with a value of 1.00. The High School Quality Index, with a value of 0.425, was the only other variable with an importance value greater than 0.40.

Choosing to use the 3-leaf tree versus the 12-leaf tree proves to be an interesting outcome. Ideally, a decision tree with fewer leaves would be preferred. However, an enrollment manager may want to see how the remaining variables contribute to model performance. One of the advantages of decision tree analysis is the ability to allow the researcher to make determinations about the inclusion and exclusion of the leaf variables.

Multiple Regression Analysis

Regression is one of the statistical procedures more commonly used to predict an outcome. Multiple linear regression analysis is used to forecast a target variable that contains interval-level data. Multiple regression models may contain numerous input variables that span both categorical and continuous data types. A multiple regression model is also used to explain the variability of the predicted value in relation to the input variables contained within the model. Ideally, the best regression model is one that provides the least error and has the fewest parameters (Matignon, 2007).
A stepwise multiple regression analysis was conducted to determine a predictive model for the target variable End-of-First-Year GPA. The data was portioned into a Training group and a Validation group for purposes of analysis. The Training group was used to develop the model, and the Validation group was applied to measure performance. The ASE was the basis of model performance for the multiple regression analysis.

The stepwise multiple regression model yielded an ASE of 0.347 for the Training group and of 0.382 for the Validation group. The selected model occurred during step 8 and used the following effects: Gender Description; High School Quality Index; Institutional Characteristic Description; Major Description; SAT Mathematics Score; SAT Writing Score; Secondary School GPA; and State_Province. Figure 2 both provides a visual display of the ASE leveling off at step 8 and shows the variability between the ASE values for the Training and Validation groups. The selected model had an $r$-squared value of 0.4647, which indicates that almost 47% of the variance in End-of-First-Year GPA could be attributed to the variables in the model.

**Neural Network Analysis**

The neural network is a popular tool in business intelligence because of past successes in use of numerous data-mining and decision-making analysis applications. Neural networks are composed of basic structures that attempt to imitate the human brain. Three basic structures compose a neural network: the input layer, which represents the variables to be input; the hidden layer, which may contain one or many levels; and the output level, which represents the predicted value of target variables (Berry & Linoff, 2004; Matignon, 2007).

In this study, a neural network analysis was completed to create a predictive model to estimate the target variable End-of-Year GPA. As in the prior completed analyses, the data set was split into a Training group and a Validation group. The performance of the model was determined by the ASE of the Validation group.

The neural network analysis yielded an ASE of 0.359 for the Training group and of 0.373 for the Validation group. The neural network was a large model, as indicated by the model degrees of freedom value of 247.
Figure 3 contains a visual representation of the ASE yielded from the neural network analysis. A large divergence in the ASE values for the Training and Validation groups occurs at training iteration 3. Primarily, this type of occurrence takes results from a large number of weights in the fitted neural network model. The effect of this phenomenon may be lessened by reducing the number of inputs in the neural network model. This reduction of inputs would lead to a smaller number of model weights and perhaps have a positive impact on model performance.
Comparison of Model Performance

The purpose of this study was to create a predictive model that may be used by undergraduate admissions officers to help them arrive at admissions decisions for applicants, especially whose academic credentials may be marginal. Further, the predictive nature of the model may also be used to assist in marketing and recruitment strategies.

Three models were created to forecast the GPA that an applicant could be expected to attain at the end of the first year of academic study at the institution. The decision tree analysis and neural network analysis are closely associated with business intelligence and have not been incorporated to a significant degree within higher education analysis. The multiple regression analysis can be considered more traditional because it is firmly rooted in traditional statistical analysis and has been more widely used in higher education analysis.

Numerous methods may be used to determine the performance of a predictive model. The usefulness of the models developed in this study was determined by the calculated average square error for each model, as determined by the Validation group. Comparing the three models developed in this analysis showed that the neural network model yielded the lowest ASE value (0.373); the decision tree model yielded the next-lowest ASE value (0.379); and the multiple regression model yielded the highest ASE value (0.382).

Based on the calculated ASE values, the neural network would be the model of choice. However, because the ASE values of all three models were similar, an argument could be made that any of the models could be used to forecast an applicant’s End-of-First-Year GPA. Each model contained a variety of data variables, and some may be more appealing to admissions officers than others. Situations like this are prime examples of the intersection of the science and art of data analysis.

SUMMARY AND CONCLUSIONS

The decision-making environment of the corporate world uses business intelligence principles on a daily basis. Mining the vast quantities of quantitative and qualitative data gathered has led to the creation of business intelligence units within organizations to manage such analyses. Higher education historically trails the corporate world in embracing principles that could improve performance and lead to more certain decision outcomes. Using predictive analytics in a grander scheme could allow educational institutions to operate with higher levels of efficiency and effectiveness.

Implications for Enrollment Management Leadership

The changing demographics of the educational marketplace are almost forcing the need for educational institutions to embrace business intelligence principles. Incorporating the models developed in this study could be of great assistance to any enrollment management division in many ways.

First, understanding the potential academic success of an applicant can help eliminate much of the guesswork in the evaluation of applicants with marginal credentials. On many campuses, such decisions are currently made by admissions committees representing a variety of areas. These decisions often include anecdotal analysis comparing the marginal applicant to similar past applicants. This methodology is far from scientific and relies heavily on individuals’ recalling specific details and circumstances of past applicants. A more consistent, equitable approach would be to use a predictive model to generate a forecasted level of academic success and combine that outcome with other mitigating circumstances, such as the type of high school curriculum the applicant had pursued.

Predictive models may also be incorporated earlier in the admissions funnel than the...
applicant stage. Understanding how inquirers may perform academically could lead to a more targeted marketing campaign earlier in the admissions process. Inquiries forecast to be not so academically successful may be sent promotional materials about academic, mentoring, and student success resources to raise their awareness of ways these services can aid in the transition from secondary academics to a college curriculum. Further, students who have excelled in their secondary school studies and are forecast to excel in a college curriculum may receive targeted communications regarding honor societies and academic research opportunities.

These predictive models would provide enrollment managers a tool that could be applied consistently across all inquirers or applicants and reduce dependence on some of the historical methods of making decisions. Using an institution’s historical data for decision making would appear to be more reliable than using arbitrary values of standardized test scores and secondary school GPAs.

Suggestions for Future Study

This study’s methodology focused on enrolling first-year students at the participating institution over a three-year period. However, enrolling students compose a small facet of the overall admissions funnel. Opportunities exist for future study, using the same parameters contained in the current study and expanding the number of years’ worth of data used for analysis. The current study used three years’ worth of data, but future analysis might use as many years’ worth of data as have been collected. Conducting a true longitudinal study may yield a specific type of model that may be more accurate.

Another aspect to be considered for future study would be incorporation of more student-centric variables during the first year of study. While the current study used few descriptive variables of the first-year student environment, incorporating variables to the analysis such as the number of credit hours a student attempts, the number of hours a student studies each week, and whether or not a student holds some sort of employment could make the model more robust.

Finally, it is recommended that the scope of analysis be expanded to increase the predictive nature of the study from the end of the students’ first year of study to include other level classifications. Understanding the relationship between the first-year GPA and future GPAs can allow academic advisors the opportunity to better serve their students and possibly be alerted to potential academic deficiencies much earlier. Early detection of any academic weakness is critical in determining a course of action to modify the behavior that may lead to the academic deficiency.

Most important, higher education institutions must embrace business intelligence techniques to remain competitive in the academic marketplace. The projected shrinkage in the number of high school graduates in certain regions means that colleges and universities need to use every tool within their grasp to achieve institutional goals. Business intelligence principles contain the pathway to organizational success only if higher education chooses to embrace them.

REFERENCES


W.O. Dale Amburgey is Director of Enrollment Analysis at Saint Joseph’s University. His research interests are in enrollment forecasting models, employee motivation, financial aid leveraging, and enrollment operations analysis. Prior to his current position, Dr. Amburgey served as Associate Director of Data Management and Analysis at Drexel University and as Director of Operations for Undergraduate Admissions at Embry-Riddle Aeronautical University. Dr. Amburgey earned degrees from the University of Central Florida, University of Louisville, Saint Joseph’s University, and Transylvania University.

John C. Yi is an Assistant Professor of Decision and System Sciences in the Haub School of Business, Saint Joseph’s University. His research interests are in modeling complex business problems using integration of quantitative methods including operations research, data mining, and knowledge management. Prior to his academic position, Dr. Yi was managing analytical projects involving marketing resource optimization, new and in-line product forecasting, strategic decision planning, sales operations optimization, product portfolio optimization, and manufacturing process improvement at Sanofi-Aventis (2 years), Johnson & Johnson (8 years), and GEC (2 years). Dr. Yi earned degrees from University of Pennsylvania, University of California at Berkeley, and Rutgers University.