Exploring Student Matriculation and Admissions Melt in Graduate Business Programs

Liye Sun
Department of Management, Purdue University Krannert School of Management, 403 West State Street, West Lafayette, IN 47907, sun783@purdue.edu

Lorena V. Bustamante
Department of Management, Purdue University Krannert School of Management, 403 West State Street, West Lafayette, IN 47907, bustamal@purdue.edu

Matthew A. Lanham
Department of Management, Purdue University Krannert School of Management, 403 West State Street, West Lafayette, IN 47907, lanhamm@purdue.edu

Kimberly J. Lanham
Indiana University Kokomo, 2300 South Washington Street, Kokomo, IN 46902, klanham173@gmail.com

Abstract

With increasing demand of analytics talents in companies, both the demand and supply of master programs in Business Analytics and Data Science has increased rapidly in recent years. In this study, we aim to build a predictive model that can accurately predict whom would matriculate in an analytical graduate program, if provided an admission offer. For this, we analyzed the application and matriculation rates over the past two years to develop a predictive model of whom is most likely to attend if provided an offer. From a planning perspective, being able to estimate the matriculation rate could provide valuable decision-support for program preparation.

Keywords: Admissions, Predictive Analytics, Melt, Business Planning

1. Introduction

Matriculation and melt are a serious problem within higher education that deserves much more attention from key stakeholders. If too many students are admitted to a university or program, then students will in turn suffer the consequences. For example, there will be a large student to faculty ratio, leaving little time for student questions and one-on-one discussion opportunities. Alternatively, if too little students are admitted, the university and individual programs will suffer due to lack of necessary revenues. When higher education administrators admit a certain number of students, they do so to ensure that tuition dollars provide at least enough revenue to account for overhead (e.g. utilities, space, faculty salaries, etc.). Furthermore, meeting target enrollments provides a university the resources and flexibility to offer opportunities to their students that would not exist otherwise. Finally, staff and faculty salaries, as well as resources provided to them, often depend on matriculation revenues.

Nevertheless, just as in the business world, there is fierce competition among universities. As a result, there is always a certain percentage of students that apply, are admitted, and accept admittance, yet do not enroll. Nevertheless, in order to remain profitable, administrators must be able to estimate student enrollment as accurately as possible. Thus, being able to estimate this melt is a useful metric for administrators. More accurate admissions decisions correlate to better performance measures and admissions rates for the university and thereby reduce the risk of admissions melt. If directors and administrators can more accurately estimate and thereby predict the number of students that will matriculate, they can therefore plan curriculum and provide appropriate resources more accordingly.

This study reviews student application and matriculation data at a tier-one public research university in the Midwest. In 2016, the school was tied for the 18th-best public university in the United States, tied for 56th among national universities, and ranked 103rd on the international scale. U.S. News & World Report also rated the university 12th in the most innovative category (2016). This study focuses on a fairly new business graduate program housed within the school of management. The university's graduate management programs encompass about 7.4% of the total university graduate student enrollment. The department of management attracts a global student base with 42% of the 2018 class coming from foreign countries and 29% coming from Asia (Undisclosed, School of Management, 2017). Top student placements for 2016 included Amazon, Procter & Gamble, Bank of America and Ford (Forbes, 2017).

The graduate program analyzed in this study is an exciting new addition on campus (currently within its second year). The program is a full-time graduate degree that focuses on balancing analytical methodologies and frameworks with information management tool and software to develop its students into future industry leaders. Admitted students usually come from backgrounds in technology, engineering, science or mathematics (Undisclosed, School of Management, 2017). Because the program is new, admissions melt and enrollment numbers are particularly of great concern for its directors and administrators. For the program to continue, certain enrollment numbers and revenues must be reached. Moreover, the results of this study will facilitate more accurate enrollment planning and will be highly conducive to a more selective admissions process.

The primary objective of the study is to estimate the number of students who will actually enroll after they have applied and have been accepted to the university's graduate program. Applying both logistic regression and classification tree predictive models, the study offers an empirical-based statistical solution that provides decision-support to educational administrators by giving them insights about enrollment expectations, as well as potential drivers that could be used to market to future students to increase the chance that they will matriculate to a university. Using predictive modeling in higher education for enrollment decisions provides many opportunities for staff and students. Building a predictive model using student attributes such as demographics, test scores, previous professional and educational experiences may provide insight into which components coincide with enrollment. The subsequent data analysis from predictive enrollment models can provide a probability for how likely each student is to enroll in the program.

2. Literature Review

The need for accurate enrollment and admissions decisions and planning is evident in such cases where schools are forced to close due to lack of revenues and resources. In an article discussing the closure of St. Joseph, a 28-year-old private Catholic college in Indiana, Colias (2017) noted the corresponding negative effects on students, employees, and the local community. For instance, around 900 students had to find a new institution to complete their degrees and 200 employees were left with an uncertain future upon the school's closure.

The use of predictive analytics to analyze student application and enrollment data is gaining ground as an innovative approach in higher education planning and decision making. In a comprehensive analysis, Ekowo and Palmer reviewed the role of predictive analytics in higher education. The authors noted, "*Schools use this information to help forecast the size of incoming and returning classes. They also use it to narrow their recruitment and marketing efforts to target those students most likely to apply, enroll, and succeed at the institution*" (2016). Similarly, Ramos and Jansen stated, "*Using characteristics of students who have enrolled in the past, predictive models can help institutions determine the chances that a student will enroll*" (2013). Furthermore, Wagner-Clews explained that universities "*can find characteristics that influence enrollment and also weigh the amount of their influence on enrollment, then apply that model against each student in your pool to see how much they fit the profile of a student who did enroll*" (2015).

Howard Community College used analytics to predict future enrollment at the school. The university partnered with ASR Analytics in Maryland to use collected data to help predict and anticipate enrollment trends several years into the future (ASR Analytics, 2014). The study reviewed the school's past enrollment data, unemployment data from the Bureau of Labor and state data to estimate anticipated high school graduates (ASR, 2014). Using past enrollment data can be a primary indicator of future enrollment. Furthermore, gathering high school graduate information is an interesting component to analyze in an enrollment management study. It can be useful to track the number of students who were admitted and enrolled from different regions of a state in order to make accurate estimations on an ongoing basis. In this instance, the predictive model provided enrollment forecasts for seven years to help program directors make decisions on student resources and other needs.

Similarly, Sampath, Flagel, & Figueroa (2009) used logistic regression to predict the likelihood of enrollment of incoming freshmen at George Manson University based on demographic information, quantitative data such as GMAT scores and undergraduate GPA, and work experience. The authors noted that using this model to estimate enrollment probability has the following advantages: The estimated data lies anywhere between $-\infty$ to $+\infty$, the model performs even when the enrollment probabilities are non-normal, and the model has a linear format and the parameter estimates can be directly related to enrollment.

Moreover, an enrollment study conducted at Wichita State University used a probability methodology to predict future enrollment using limited resources. The school's chief data officer assigned prospective students a probability score from 0–100 as a means to predict future enrollment (Felton, 2015). Felton explained, "*The scores are based on factors such as gender, race, ethnicity, standardized test scores, grades from high school, and whether parents went to college*" (2015). Although quantitative measures such as standardized test scores and past grades are likely good predictors of future enrollment, factors such as gender, race, and ethnicity are more subjective and thus do not seem reliable for an enrollment prediction study. Furthermore, basing enrollment decisions on gender, race, and ethnicity can be problematic in that they can be considered discriminatory. Ekowo and Palmer maintained, "*Predictive tools can produce discriminatory results because they include demographic data that can mirror past discrimination included in historical data*" (2016). Further, they noted, "*Enrollment managers must balance several priorities, such as say they are constantly trying to balance competing priorities such as increasing the quality of each incoming class, enrolling students with an ability to pay, and increasing the diversity of the student body*" (2016). Thus, universities must make clear indications of why certain variables are used in predictions in order to show transparency and remain unbiased in future enrollment decisions.

Many university administrators agree that admissions melt is a serious issue that must be combatted with a multitude of techniques and methods. Many schools are employing aggressive tactics to target melt. This includes faculty phone calls to students to enforce positive direct interaction, as well as the use of technology, text automation, and feedback to help keep students prepared, engaged, and on track (Belkin, 2017). Moreover, in an article on the Education Advisory Board (EAB), the author noted that most campuses are using front-end and back-end solutions to combat admissions melt. The research indicated the importance of front-end approaches as a proactive method of communication with students before they show signs of melt. It also maintained that back-end approaches have their place as well, though they are more reactive (EAB, 2016).

These methods allow direct contact between admissions staff and at-risk students, which can be an effective means of enrolling and retaining students. Predictive modeling can be used to strengthen these efforts and identify those students most likely to enroll in a university or program to tailor communications to these students for targeted enrollment efforts (Ramos & Jansen, 2013). Once specific students are targeted, more resources and direct contact can be initiated with these students in order to ensure enrollment and/or retention. Direct outreach to targeted applicants becomes crucial in combatting admissions melt. Predictive modeling can "*target applicants based on a model score so* *that you can maximize the enrollment yield potential of the entire applicant pool*" (2013). Ramos and Jansen also identified the importance of maintaining contact with all applicants, however, prioritizing communications based on statistical modeling. The authors noted, "*Instead of choosing who you will communicate proactively with, you can now reach out to all students, but in order of highest to lowest probability to enroll based on the predictive model*" (2013).

As discussed previously, there are several benefits in using a predictive model in early admissions and recruitment efforts. Predictive models such as this can be utilized as a tool to estimate applicant behavior for admissions and program managers. This data can then aid managers in making appropriate decisions for their departments in regard to resource and budgetary matters. Universities can identify potential prospects from a large set of applicants by summarizing and analyzing the student data. Managers can then identify which students are most likely to enroll based on these factors. As, Ramos and Jansen noted, "*Predictive modeling can help you qualify names, and instead of guessing, you have a statistical model to guide you and add more students to your funnel who are already more likely to enroll*" (2013).

Nevertheless, predictive modeling also has its weaknesses. Utilizing data-based methods for estimation and decision making can have drawbacks such as inaccurate or often insufficient data, which can result in lower predictive accuracy. Moreover, though a model may be successful at any specific moment in time, there are always factors that could influence student behavior that were not able to be measured or known. For example, many schools ask students what other schools they are considering in their application, but they often do not know how much scholarship or financial assistance those programs are offering the student, which are likely to affect where the student matriculates. These factors can be impossible to predict or anticipate. Even asking the applicant, one should be skeptical in the data provided as the applicant might try to game the school by stating school A provides more financial benefits than was actually offered. Thus, constant revision of such models is often necessary to reflect an accurate portrayal of the current landscape, as well as considerations in how to obtain model features that affect the student's enrollment decision is required to generate accurate enough predictions that can serve as a basis for sound decision-making.

3. Data

To predict whether an admitted student will indeed enroll in the specified program, 44 independent variables were included in the model. Demographic attributes (age, gender, ethnicity, citizenship), standardized testing scores (GMAT overall, TOEFL), academic measures (GPA, field of degree, prestige of education institution), employment (total months of work experience, type of employer, location of employer), and subjective textual metrics from the candidate's interview (e.g., professional appearance, overall preparation, maturity/self-awareness, desire to attend) were included in the model.

The data was collected during the application periods corresponding to the 2016 – 2017 and the 2017 – 2018 academic years. Most of the data used in this study was provided by the students themselves at the time of application to the graduate program. Moreover, the variables pertaining the interview were collected directly from the lone interviewer's notes on the applicant's file. Each of the attributes collected during the interview were measured using a 1-5 Likert-scale.

4. Methodology

This study uses both logistic regression and a classification tree to analyze and predict whether an admitted student would indeed enroll in the specified program. These methods are appropriate because our response variable a categorical variable. The outcome variable, "Status" (labeled Y in the data analysis), has three different possibilities: Confirmed, Declined, or Deferred to Next Year, as shown in Figure 1.

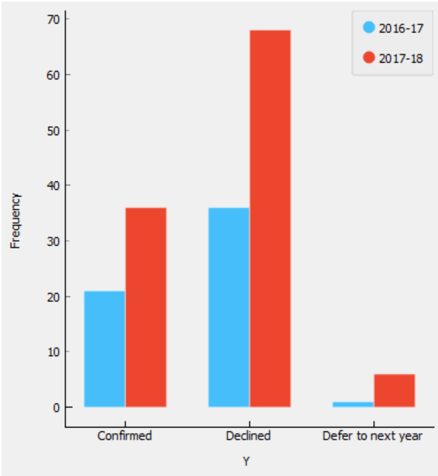

Figure 1: Bar graph of accepted students for 2016-17 and 2017-18 year

It is important to note that for our analysis we excluded those applicants who chose to defer their admission to future years and focused only on those records where students accepted or declined the admission offer. This led to having 161 rows in the dataset, where 57 confirmed and 104 declined their admission letter. Of the confirmed students, 24 students were female, and 33 were male, marking a fairly equal gender distribution in cohorts.

In relation to the standardized testing scores, adjustments were made for some students. First, some of the students had taken the GMAT, while other had taken the GRE. The school had a converter that could transform the GRE score into a predicted GMAT score. Thus, for this analysis the variable GMAT comes from those that took the GMAT or is a translated GMAT from those that took the GRE. Furthermore, several variables measured items regarding English speaking skills. Many international students were required to take the TOEFL exam. However, a few took an accepted alternative exam known as the IELTS test.

Furthermore, some variables required clean-up and processing. The "Major" variable had many different values (e.g., engineering disciplines, business fields, etc.) indicating the student's most recent major. For this, four dummy variables were created pertaining to major type: engineering, computer science/computer engineering, statistics, and business. Similarly, the student's school was also coded as four dummy variables based on the rankings for each school: World university ranking, BRICS ranking, US top 100, and Asia ranked[1]. This variable was believed to be a potential important variable at predicting melt, suggesting that students coming from more prestigious programs would be more likely to melt and attend other schools, compared to those coming from less prestigious schools. Finally, the applicant's latest employer was also coded as seven dummy variables that measured the following characteristics about the firm: was the company considered a consulting firm, firm primary based overseas, multinational firm, US-based firm, analytics-focused firm, financial-services-focused firm, or was the last employer a university or not. Program administrators believed that those coming from trendy analytics-focused companies (e.g. Google), US-based, multinational companies, consulting, or analytics-focused were more likely to melt and attend more prestigious schools more so than other students. Conversely, those coming from overseas firms, financial-services firms, or from a university were less likely to melt.

[1] World university ranking means that the school the student came from had a world rank or not (1=Yes, 0=No). BRICS ranking means that a "BRICS & Emerging Economics Ranking" existed for that school or not. US top 100 means that they came from a school that was ranked in the top 100 schools in the United States. Lastly, Asia ranked means that they came from a school that had a ranking among the top Asian universities.

Finally, in the case of missing values (e.g. GPA), we employed a model-based imputation approach, which fills in the missing values using its own predictive model of variables that do not have missing values.

Once data cleaning was completed, and numerical features were standardized, the authors entailed building the models. Predictive analytics bases the likelihood of student enrollment on common characteristics with students who have enrolled or not enrolled in the program in the past. Logistic regression and classification decision trees were applied in this study because they allowed for more interpretability than more sophisticated methods, which was important for administrator decision-making. Moreover, since our dataset was relatively small (n=161) we felt more flexible methods would lead to overfitting.

Logistic regression was used for analyzing data in which the outcome is measured with a categorical variable. A ridge regression technique was used to alleviate multicollinearity amongst the predictor variables in the model. Classification trees were used to predict an outcome based on several input variables and provided a nice visual for admissions staff to identify the most pertinent variables.

Finally, the last step in our methodology, we evaluated the performance of both classification models applied using a confusion matrix table and the classification accuracy rate to evaluate the performance of the two classification models applied.

5. Results and Analysis

5.1 Descriptive Statistics

Overall, 57 students confirmed their acceptance while 104 declined their admission letter. Table 1 shows that 63.2% (36/57) melted for the academic year of 2016-17, while 65.4% (68/104) melted in the academic year of 2017-18. Based on this data, an administrator might expect that 35.4% (57/161) of students that were admitted into the program will indeed matriculate.

Table 1: Table of confirmed students versus declined for 2016-17 and 2017-18 years

	Confirmed	**Declined**	**Total Admission Offers**
2016-2017	21	36	57
2017-2018	36	68	104
Total Admission Offers	57	104	161

Examining the differences between confirmed and declined admission letters, several insights can be drawn. First, in regard to age and gender, there is little difference between those students who confirm, compared to those who decline. In terms of number of months of work experience and TOEFL score, applicants who decline their admission offer have higher scores than those who confirm. Conversely, the mean GPA and GMAT score for those applicants who declined are lower than those who confirmed but the difference was not statistically significant (refer to Table 2).

Table 2: Comparison of confirmed vs. declined students

	Confirmed	**Declined**
% of Males	36.3%	63.7%
% of Females	34.3%	65.7%
Mean Age	25.02	25.6
Mean Number of Months of Previous Work Experience	27.58	35.72
Mean GPA	3.41	3.37
Mean GMAT	679.12	676.44
Mean TOEFL	106.20	107.62

5.2 Prediction Results

The goal of the classification tree is to classify the "Status" of the student into one of two possible outcomes (Confirmed vs. Declined) based on his/her own characteristics (e.g., TOEFL score, type of employer, etc.). The classification decision tree obtained is shown below in Figure 2; a top-down read provides insight regarding the importance of the variables in explaining the student's status. As shown in Figure 2, experience in months was the first independent variable used, suggesting this is the most important variable at predicting a confirm or decline decision.

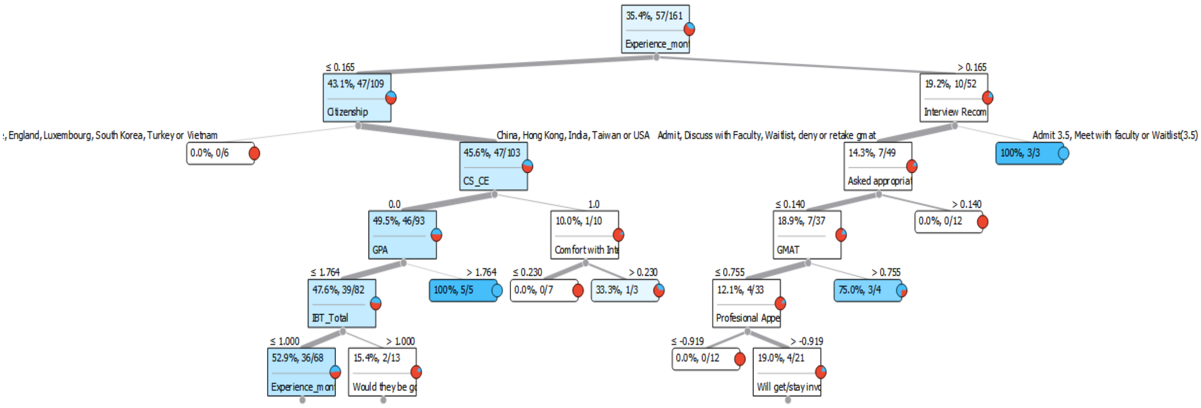

Figure 2: Melt classification decision tree

5.3. Model Evaluation

A confusion matrix was used to show the number of times the model predicted correctly and incorrectly. As shown in Table 3, the logistic regression model accurately predicted that 20 students will confirm enrollment and accurately detected that 73 students will decline their enrollment. On the other hand, as shown in Table 4, the decision tree model accurately predicted that 26 students will confirm enrollment and accurately detected that 66 students will decline their enrollment (Table 4). Based on these results, it can be concluded that the logistic regression model is more accurate in predicting decline of enrollment, compared to the decision tree model. Conversely, the decision tree model is more accurate in predicting confirmation of enrollment, compared to the logistic model.

Table 3: Confusion matrix for the logistic regression model

			Predicted	
		Confirmed	*Declined*	*Total*
Actual	*Confirmed*	20	37	57
	Declined	31	73	104
	Total	51	110	161

Table 4: Confusion matrix for the decision tree model

			Predicted	
		Confirmed	*Declined*	*Total*
Actual	*Confirmed*	26	31	57
	Declined	38	66	104
	Total	64	97	161

The models were compared based on their classification accuracy, which is the proportion of the time the model correctly identified confirmed students and declined students (in other words, the proportion of correctly identified applicants). Additionally, the precision of the models was measured, meaning the proportion of true positives (e.g., proportion of predicted confirmed students who confirmed). As shown in Table 5, the classification decision tree is accurate 57.1% of the time, while the logistic regression is correct 57.8% of the time.

Table 5: Model comparison of accuracy and precision

	Classification Accuracy	**Precision**
Decision Tree	0.571	0.583
Logistic Regression	0.578	0.568

Finally, the models were evaluated based on their accuracy to predict the overall attrition ("melt") rate. The logistic model predicts a melt rate of 68.3% (110/161), as shown in Table 3. Conversely, as shown in Table 4, the decision tree model predicts a melt rate of 60.2% (97/161). Since the actual melt percentage is 64.5% (refer to Table 1), we can conclude that both these models are fairly accurate in predicting overall melt rates, yet the logistic model is slightly better in predicting overall melt percentages.

6. Conclusion

Using predictive modeling in higher education for enrollment decisions provides many opportunities for staff and students. Each year, directors must review existing policies and, if needed, adapt to changing circumstances. Taylor and Miroiu noted the importance of determining long term goals and objectives, and then adapting and allocating resources to best carry out these goals (2002). Thus, building predictive models to determine the probability that specific students will enroll is pivotal in higher education planning and management, allowing more time and resources to be spent on these students, as opposed to little attention to all applicants in the admissions pool.

Our initial results suggest that the logistic regression model (57.8%) yields a slightly higher accuracy rate, compared to the decision tree (57.1%). When we delve further into these results, we identified that the logistic regression model is more accurate in predicting decline of enrollment, compared to the decision tree model. Conversely, the decision tree model is more accurate in predicting confirmation of enrollment, compared to the logistic model. Finally, when comparing the overall admissions melt rate predicted by the two models, the logistic model predicts a rate (68.3%) that is closer to the actual melt percentage of 64.5% (refer to Table 1).

We are continuing to develop our predictive modeling solution to achieve greater predictive accuracy, so administrators are comfortable using the results to support admissions decisions and melt estimation. Therefore, we need to continue to strengthen our findings. In building a model, it is important to first gather as much historical data as possible to gain an accurate representation to make appropriate enrollment decisions. As noted on an eBook discussing optimizing enrollment with predictive modeling, "*It is recommended that you use at least three years of historical data to build your models*" (Rapid Insight, n.d.). Because the program of study in this paper in just beginning its third year, we intend to extend our model by including an additional year of data. The 2018-2019 program's cohort began in June 2018; therefore, we will test the accuracy of our models with this new cohort. Per Rapid Insight, the model can be used to "*score your admitted applicant pool to get each applicant's probability of enrolling, which you can then use to make decisions and to understand what your incoming class will look like*" (n.d.). Moreover, to strengthen our results, we will also evaluate additional variables (e.g., scholarship offers) and more flexible models (e.g., random forest, support vector machines) with the hope of increasing predictive accuracy while not overfitting. Finally, we would like to extend our modeling approach to other graduate programs within the business school so that all stakeholders can be clear regarding the planning of future student intakes.

References

[1] ASR Analytics. (2014). ASR demonstrates student enrollment forecasting with Howard Community College. Retrieved from https://www.asranalytics.com/about-asr/news/asr-demonstrates-student-enrollment-forecasting-howard-commu

[2] Belkin, D. (2017). College admissions officers won't just chill about 'summer melt'. Retrieved from http://www.cetusnews.com/life/College-Admissions-Officers-Won%E2%80%99t-Just-Chill-About-%E2%80%98Summer-Melt%E2%80%99-.SJQ4Sc8sv-.html

[3] Colias, M. (2017). Rensselaer stunned after announcement of St. Joseph's closure. Retrieved from http://www.chicagotribune.com/suburbs/post-tribune/news/ct-ptb-rensselaer-st-joseph-closure-st-0212-20170210-story.html

[4] EAB. (2016). Understanding—and stopping—summer melt. Solutions to tackle summer melt aren't one-size-fits-all. Retrieved from https://www.eab.com/daily-briefing/2016/08/15/understanding-and-stopping-summer%20melt

[5] Ekowo, M. & Palmer, I. (2016, October). The promise and peril of predictive analytics in higher education: A landscape analysis. Retrieved February 25, 2018, from https://na-production.s3.amazonaws.com/documents/Promise-and-Peril_4.pdf

[6] Felton, E. (2015). Colleges shift to using 'big data' — including from social media — in admissions decisions. Hechinger Report. Retrieved from https://hechingerreport.org/colleges-shift-to-using-big-data-including-from-social-media-in-admissions-decisions/

[7] Forbes (2016). Undisclosed School of management.

[8] Ramos, C. & Jansen, B. (2013, April). How predictive modeling benefits enrollment managers. Retrieved from http://blogem.ruffalonl.com/2013/04/09/predictive-modelingbenefits-enrollment-managers

[9] Rapid Insight. (n.d.). Optimizing enrollment management with predictive modeling. Retrieved from http://www.rapidinsightinc.com/wordpress/wp-content/uploads/em-ebook.pdf

[10] Sampath, V., Flagel, A., & Figueroa, C. (2009). A logistic regression model to predict freshmen enrollments. Retrieved from http://analytics.ncsu.edu/sesug/2009/SD016.Sampath.pdf

[11] Taylor, J. & Miroiu, A. (2002). Policy-making, strategic planning, and management of higher education. Papers on higher education. Retrieved from https://files.eric.ed.gov/fulltext/ED475531.pdf

[12] Undisclosed. (2017). School of management.

[13] U.S. News and World Report. (2016). Best colleges rankings. Retrieved from https://www.usnews.com/best-colleges/

[14] Wagner-Clews, M. (2015). Predictive modeling benefits. Retrieved from https://www.linkedin.com/pulse/predictive-modeling-benefits-margie-wagner-clews

[15] World University Rankings. (2018). Retrieved from https://www.timeshighereducation.com/world-university-rankings