The Profile of Student Attrition at Radford University

Stacy Austin
Radford University – Institutional Research
saustin4@radford.edu

ABSTRACT

As with most colleges, student retention at Radford University is of the utmost importance. This is something that is constantly monitored and has been studied by third party research groups in recent years, with the Fall to Fall retention rate hovering around 75% over the last 5.(Allen and Lovik) Additionally, with the recent implementation of President Brian O. Hemphill's 5 Year Strategic Plan and Strategic Planning Task Force, the Strategic Enrollment Growth Subgroup has committed to growing in-state freshman/out-of-state freshman/graduate student and new transfer headcount enrollment by 3%, veteran and active military student headcount enrollment by 10% and international student headcount enrollment by 50 students annually through the 2023-2024 academic year.(Strategic Planning Task Force) With the help of the joint efforts of the academic Department of Information Technology and the university's functional Division of Information Technology, this research will use new tools from complimentary analytics areas that are overlooked in the existing literature to examine the effects of specific student social interactions on the probability of success, thus contributing to the existing (Bradberry et al.) research providing a more predictive profile model of a successful or non-successful Radford University student. While in this study success is being defined by new freshman retention, this could also be determined by persistence, graduation or more granular indexes using other accolades achieved by students during their tenure.

INTRODUCTION

It is no wonder such a degree of importance is placed upon retention rates at Radford, and universities in general, as this is one of the first things parents look at when determining with their child where they will potentially study and/or spend the next 4-6 years of their life. With U.S. News Best Colleges rankings for example, out of their 7 major ranking indicators, the Graduation and Retention Rates indicator is tied with Undergraduate Academic Reputation for the highest weighted values at 22.5% each. (Morse et al.) These and other numbers guide most, if not every, facet of decision making at a university. This considered, third party research groups, administration and students alike have studied the student body in efforts to raise retention rates and have provided some insightful data on the subject, however with no real quantitative focus on social engagement, until recently. Reviewing student social interactions has clearly shown to be beneficial in past research(Kuh et al.), with regards to their success. Using new techniques from statistical computing and swipe access data obtained upon each students entry into a building of a particular focus, this study determines what correlation the amount of time spent at various locations, such as the student recreation centers, has to a student's probability of success. The results of the research will assist the following university stakeholders in making decisions on campus to either increase or decrease engagement in certain locations or at certain times aiding in the success of the student body. For example, if determined that a moderately engaged student has a higher probability of success, the Student Activities department could potentially schedule more or less events in certain locations where they know a higher concertation of students will or will not be based on their intentions. On the other hand, for instance, if it is determined that an excessively over-engaged student has a lesser probability of success, then the open and closed hours of certain locations could be adjusted at certain times as well. Finally, this information will be able to assist the universities Enrollment Management department and Office of Student Success and Retention in predicting overall likelihood of success, just to name a few.

REVIEW OF LITERATURE

While mining identification card swipe data is nothing new(Renaud et al.), nor is mining data for insight into attrition research(Hu), a combination of the two is a relatively new concept (Bradberry et al.). Likewise, the study of social effects on attrition has been researched since at least the 1970's (Mannan). While this research combines all three areas of focus, following is a summary of each found in previous research.

Swipe Data Mining From the Literature

(Bradberry et al.) uses knowledge from information systems to examine the benefits of card swipe data to aid in the prediction of retention. Finding a relation between attrition and the number of swipes or the time of swipes prove the benefits and gives reason to pursue further. With the authors intentions of adding additional dimensions to the model, the research in this study will apply its finding to strengthen the model used in the (Bradberry et al.)research.

(Du and Taylor) studies the use of mining that helped wait times at theme parks and Google hospitality locations as well as traffic patterns to reduce wait times at the recreation center on a college campus. This work, showed patterns of student workouts in terms of various factors (time, day, semester, etc.). Also learning a Decision Tree model to predict crowdedness at the recreation center for given time intervals. This research displayed the advantages of a Decision Tree model in this real-world application over Naive Bayes and Support Vector Machine classifiers.

(Renaud et al.) shows practical benefits of mining swipe data on campus. With a Pearson correlation analysis and patron, book, and swipe data the research team used data mining to determine library use patterns and successfully determined higher library use correlated to higher student grade point averages.

(Parry) reviews a study by Dr. Matthew Pittinsky at Arizona State using student ID cards swiped data. Dr. Pittinsky feels you can predict attrition by pinpointing changes in how a student uses a campus. "Say someone goes to Starbucks at 2 p.m. every day before a 2:15 p.m. class. Then stops." Additional information directly from the Pittinsky study is currently being pursued for further research.
(Jones) discusses the benefits of swipe card use in university library systems but no details on

methodologies of research. Similar to the (Renaud et al.) research, the author goes on to review software developed by Advanced Campus Services at Georgia State University to assist in a more user friendly review of the findings for administration.

Data Mining Attrition from the Literature

(Hu) demonstrates the effectiveness and efficiency of data mining in attrition analysis for retailing banks. The authors discuss the use of lift as a proper measure for attrition analysis and compare the lift of data mining model of decision tree, boosted naïve Bayesian network, selective Bayesian network, neural network and ensembles of those classifiers. The field test conducted by their clients proved that the data mining prediction model for attrition is very accurate and the target-oriented campaign is very effective.

(Miller et al.) offers the confirmation of the effectiveness of the logistic regression model using pre-matriculation. While this study only uses the pre-matriculation data, the findings may not as beneficial to this study as other sources.

(Shah) posits factors that are strong and significant predictors of the students' performances are academic integration, family background and social integration. Using Decision Trees, Bayesian Networks, Functions and Rules in Weka, the authors identify students who are not performing well and those who are at the risk of failing (below satisfactory) and eventually be dropped out (fail), Random Forest accuracy is the highest with accuracy of 95% and 85% respectively. Yet, in predicting students' academic performance Decision Trees were better.

(Delen) uses Cross Industry Standard Process for Data Mining, 8 years of institutional data, along with artificial neural networks, decision trees, and logistic regression to develop analytical predictive models of freshmen student attrition. Of the three model types, artificial neural networks performed the best, with an 81% overall prediction accuracy on the holdout sample. The variable importance analysis of the models revealed that the educational and financial variables are the most important among the predictors used in this study. However it is worth noting the social interaction variable was also considered important in the study.

(Simon) at University North Texas, Institutional Research used a stepwise Logistic Regression to study the correlation between student engagement and various statistical categories on campus. With regards to retention, students who engaged in Student Affairs activities were 1.21 times more likely to be retained.

Social Engagement on Attrition from the Literature

(Kuh et al.) determines the relationships between student behaviors and the institutional practices and conditions that foster student success. Merging student-level records from different types of colleges and universities the authors examine the links between student engagement and academic achievement or persistence. Using OLS and Logistic Regression the findings from this study point to two conclusions. First, student engagement in educationally purposeful activities is positively related to academic outcomes as represented by first-year student grades and by persistence between the first and second year of college. Second, engagement has a compensatory effect on first-year grades and persistence to the second year of college at the same institution.

(Robbins et al.) report on a large-scale study examining the effects of self-reported psychosocial factors on first year college outcomes. Using a sample of 14,464 students from 48 institutions, the authors constructed hierarchical regression models to measure the predictive validity of the Student Readiness Inventory, a measure of psychosocial factors. Controlling for institutional effects and traditional predictors, the authors tested the effects of motivational and skill, social, and self-management measures on academic performance and retention. Academic Discipline was incrementally predictive of academic performance (grade-point average) and retention. Social Activity and Emotional Control also helped predict academic performance and retention, whereas Commitment to College and Social Connection offered incremental prediction of retention.

(Mannan) explores the compensatory relationship between academic and social integration, and also assesses the differences of group specific academic and social integration in a small university in the Pacific as perceived by the students. Using Reliability analysis in SPSS, This study showed a strong negative relationship between academic and social integration, which indicates that less integration in the social domain of the university was compensated by higher academic integration leading to student persistence. Similarly, less academic integration might be compensated by higher social integration influencing students to continue study.

This literature being overall very insightful, particularly with regards to the determination of classifier and analysis types to use in the research, a common theme appears to be Bayesian Networks, Neural Networks, Decision Trees and Logistical Regressions. While (Hu), (Miller et al.), (Shah) and others have shown that data mining is highly beneficial in attrition prediction and (Du and Taylor), (Renaud et al.) and (Parry) have proven the usefulness of swipe data mining, each will be married in this paper to build upon the (Bradberry et al.) research, as well as quantitatively testing (Mannan)'s assertion to a strong negative relationship between academic and social integration and (Kuh et al.)'s assertion that student engagement in educationally purposeful activities is positively related to academic outcomes.

METHODS & INITIAL RESULTS

This population consist of first year freshmen at the university which had a total enrollment of 9418 students during fall 2017, when the sample was collected. The final dataset used was a combination of standard Institutional Research data (retention information) and card swipe data collected from the Division of Information Technology (including anonymous student id, door id, and timestamps). After filtering out for new freshman and rec center only swipes using Oracle SQL Developer, the dataset is made of 1698 (of the total 1848) new freshman and 218 (of the total 265) not retained students that swiped into the rec center while at the university. Finally, using timestamp information, the set matrix was transposed to aggregate the number of swipes by each student each week during the 14 week fall 2017 semester.

The analyzed sets columns are ID (anonymized student identification number), RETAIN (cell value = 1 if retained 0 if not), COUNT (total swipes per patron all semester) and WEEK1-14 (total swipe counts for each week of the semester). WEEK1 takes place between 25-AUG-17 and 31-AUG-17, and so on every 7 days, until 02-DEC-17. The first academic week started on 28-AUG-17, however student move in for all residence halls started on 25-AUG-17 so the data sets weeks do not directly correlate to the traditional academic weeks, being early by a few days.

The data was first modeled in R, a statistical computing programming language and environment, with student retention as a function of number of swipes per week. A statistically significant p-value, with an alpha level of 0.001, was found on weekly swipe count using a generalized linear model during weeks 1, 2 and 5. With a generalized linear model being preferred here for use with the retention variable, as it enables binary variables, the sum of binary variables, or polytomous variables (variables with more than two categories) to be modeled (dependent variable). Negative for weeks 1 and 2, while positive for week 5. This tells us that the total number of swipes a patron had at the rec center during weeks 1 and 2 is a statistically significant predictor of student *attrition;* and a statistically significant predictor of student retention during week 5, both with a strength of 93%. The results are shown in Table 1.

Additionally, the overall total number of rec center patron swipes were modeled, with student retention as a function of COUNT. A weak statistically significant p-value was found, with an alpha of level of 0.01, on total swipe count, again using a generalized linear model. This tells us that the total number of swipes a patron had at the rec center over the entire semester is a slight statistically significant predictor of student *retention* with a strength of 99%. The results are shown in Table 2. This is also illustrated in Fig.1 as retained patrons are shown to have not much of a higher average number of rec center swipes for the entire semester than non-retained patrons.

While this study concentrates mostly on students that actually used the facilities, it is interesting to note that out of the 150 new freshmen that did not, only 47 of them were not retained, as is displayed in Figure 2. However, when both patrons and non-patrons were modeled with student retention as a function of COUNT, a statistically significant p-value, with an alpha level of 0, was found. Telling us that the total number of swipes a student (both rec center patrons and non-patrons) had at the rec center during the semester is a statistically significant predictor of student retention with a strength of 95%, as illustrated in Table 3.

TABLES & FIGURES

Coefficient	Estimate	Std. Error	z-value	p-value	Sig.
Intercept	1.872636	0.102362	18.294	< 2e-16	***
WEEK1	- 0.129282	0.042078	- 3.072	0.00212	**
WEEK2	- 0.035070	0.064921	- 0.540	0.58906	**
WEEK3	0.132786	0.080202	1.656	0.09779	
WEEK4	0.007686	0.086080	0.089	0.92885	
WEEK5	0.228163	0.085476	2.669	0.00760	**
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table 1

Coefficient	Estimate	Std. Error	Z value	p-value	Sig.
Intercept	1.772502	0.096630	18.343	< 2e-16	***
COUNT	0.011518	0.005516	2.088	0.0368	*
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table 2

Coefficient	Estimate	Std. Error	Z value	p-value	Sig.
Intercept	1.333677	0.088594	15.054	< 2e-16	***
COUNT	0.033188	0.005375	6.174	6.66e-10	***
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table 3

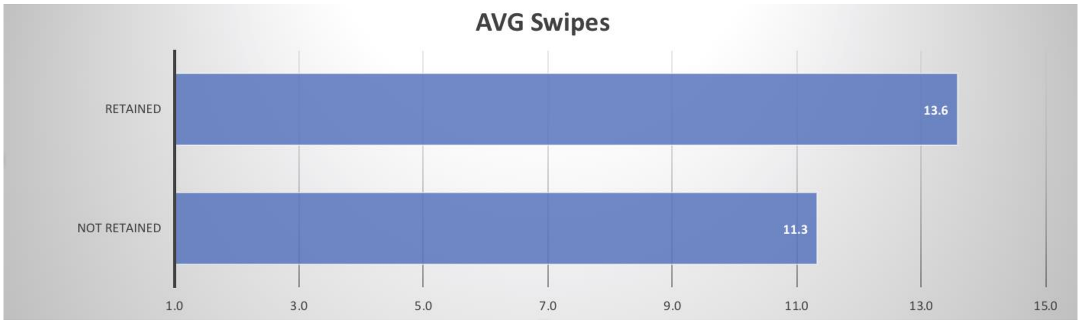

Fig. 1

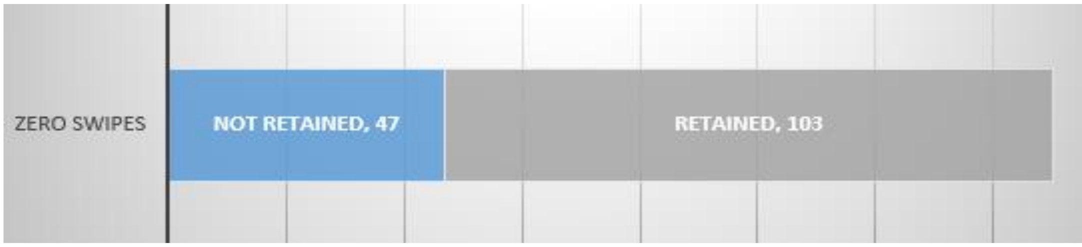

Fig. 2

CONCLUSION & NEXT STEPS

The initial findings bring favorable results and motivate additional analysis. Showing statistical significance for the activities in this one building illustrates the potential to use these same analysis in other buildings such as the dining halls and other social gathering places. Additionally, other points in which a student must swipe their id card can produce results as well, such as attendance at certain campus events like career fairs or sporting events.

The immediate next steps in this research would be to aggregate the dataset by day, day of week and hour. Specifically starting with weeks 1, 2 and 5 to gain further insight into their significance. The current dataset will also be modeled onto a neural network to compare results against the findings of the linear model. With all findings forwarded to be combined with and strengthen the predictability of the (Bradberry et al.) model. It will also be beneficial to cross reference the universities fall 2017 calendars with WEEK5 (22-SEP-17 - 28-SEP-17) to determine what events may have been taking place during that time that may correlate to rec center usage, further yielding student success. Again, aggregation by day and hour may provide further insight into this.

The weekly aggregated data proves beneficial because it allows us to dive deeper into rec center patron activity, giving us more specific data points to analyze. While, it is shown that student usage of the rec center is a predictor in retention, the weekly aggregation tells us when the center "should or should not" be used, with regards to predicting retention anyway.

It is possible, that as freshmen whom are spending an excessive amount of time in rec centers during the first 2 weeks of class, instead of gearing up for academics, are setting a precedent for their behavior for the remainder of the semester. On the other hand while a student's decision to use the facilities clearly predicts retention, once they are in there, it doesn't necessarily matter as much how often they visit as long as they do visit. Provided they have not visited excessively during the first 2 weeks of class of course.

WORKS CITED

Allen, Damien, and Eric Lovik. "New Freshmen Retention and Graduation Rates | Institutional Research | Radford University." *Electronic Fact Book*, 2017, https://ir.radford.edu/electronic-fact-book/chart.php?chart=RT01.

Bradberry, Caleb, et al. "Explaining and Predicting First Year Student Retention via Card Swipe Systems." *Twenty-Third Americas Conference on Information Systems*, 2017, https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1288&context=amcis2017.

Delen, Dursun. "PREDICTING STUDENT ATTRITION WITH DATA MINING METHODS." *J. COLLEGE STUDENT RETENTION*, vol. 13, no. 1, 2011, pp. 17–35, doi:10.2190/CS.13.1.b.

Du, Yunshu, and Matthew E. Taylor. "Work In-Progress: Mining the Student Data for Fitness." *Proceedings of the 12th International Workshop on Agents and Data Mining Interaction* , no. July 2015, 2016.

Hu, Xiaohua. "A Data Mining Approach for Retailing Bank Customer Attrition Analysis." *Applied Intelligence*, vol. 22, 2005, pp. 47–60, https://link.springer.com/content/pdf/10.1023%2FB%3AAPIN.0000047383.53680.b6.pdf.

Jones, Jennifer Link. *Using Library Swipe-Card Data to Inform Decision Making*. 2010, http://scholarworks.gsu.edu/univ_lib_facpres.

Kuh, George D., et al. "Unmasking the Effects of Student Engagement on First-Year College Grades and." *Source: The Journal of Higher Education*, vol. 79, no. 5, pp. 540–63, http://www.jstor.org/stable/25144692. Accessed 27 Feb. 2018.

Mannan, Abdul. "Student Attrition and Academic and Social Integration: Application of Tinto's Model at the University of Papua New Guinea." *Higher Education*, vol. 53, 2007, pp. 147–165, doi:10.1007/s10734-005-2496-y.

Miller, Thomas E., et al. "Using a Model That Predicts Individual Student Attrition to Intervene with Those Who Are Most at Risk." *Educational and Psychological Studies Faculty Publications*, vol. 28, 2009, http://scholarcommons.usf.edu/esf_facpub.

Morse, Robert, et al. "How U.S. News Calculated the 2018 Best Colleges Rankings | Best Colleges | US News." *The U.S. News Best Colleges*, 2017, https://www.usnews.com/education/best-colleges/articles/how-us-news-calculated-the-rankings.

Parry, Marc. *College Degrees, Designed by the Numbers*. 2012, https://immagic.com/eLibrary/ARCHIVES/GENERAL/CHRON_HE/C120718P.pdf.

Renaud, John, et al. "Mining Library and University Data to Understand Library Use Patterns." *The Electronic Library*, vol. 33, no. 3, 1108, pp. 355–72, https://doi.org/10.1108/EL-07-2013-0136.

Robbins, Steven B., et al. "Unraveling the Differential Effects of Motivational and Skills, Social, and Self-Management Measures from Traditional Predictors of College Outcomes." *Journal of Educational Psychology*, vol. 98, no. 3, 2006, pp. 598–616, doi:10.1037/0022-0663.98.3.598.

Shah, Najmus Saher. “PREDICTING FACTORS THAT AFFECT STUDENTS’ ACADEMIC PERFORMANCE BY USING DATA MINING TECHNIQUES.” *PAKISTAN BUSINESS REVIEW JANUARY*, 2012.

Simon, Jason F. *Gathering Direct Evidence of Student Engagement via ID Card Systems*. http://texas-air.org/conference/2013/presentations/F5.pdf. Accessed 1 Mar. 2018.

Strategic Planning Task Force. *2018-2023 Strategic Plan*. 2018, https://www.radford.edu/content/dam/departments/administrative/strategic-planning/UR-Strategic-Plan-121517.pdf.