Project Name: Student Attrition

Group Members: Ryan Kallicharran and Tereza Shterenberg

Data: June 21, 2016

Presentation Summarry:

Slide intro [Ryan and Teraza]

Hello everyone, my name is Tereza and this is Ryan and together we are working on the Student Attrition project.

For our capstone project, we will be conducting a research project on currently enrolled students that have a high risk of attrition. What is student attrition? Student attrition is the reduction in numbers of students attending courses/program/school as time goes by. Since we are doing a research project, we will be talking mostly about the research we have done to arrive at our hypothesis. When it came to the research portion of our project we split up the work evenly, we found six useful research papers that guided us to our hypothesis.

Student attrition has been an important problem for students as well as educational establishments. By conducting our research we will be adding to the numerous past and ongoing projects, which focused on this topic and by doing so improved the academic counseling methods and helped students achieve their goals.

Slide #1 [Ryan]

Why They Leave: Understanding Student Attrition from Engineering Majors by D. Raj Raman & Brandi N. Geisinger. This paper mostly focused on engineering majors and why they decided to leave the program. This paper aimed to explain why students may leave the majors in detail. One category is classroom and academic climate, which include inadequate teaching and advising: lack of faculty guidance (encouragement, support, and attaching), competitive or hostile environment and Inadequate teaching style. Individualistic culture: lack of sense of engagement or belonging and sense of isolation. While this may be true it is difficult to quantify this category because it more of a psychological view point of the student. Not to say that as a computer scientist that this cannot be used but in our project we just wanted to just work with quantitative entities. Alright, another category is self-efficacy and self-confidence this includes high school preparation, inadequate mathematics, science, physics, and chemistry preparation. Also inadequate overall high school GPA, inadequate high school class rank, ACT/SAT scores. The GPA and ACT/SAT scores were considered but this would not reflect the older students well since they are coming back to school after such a long time. Other categories included were interest, goals, race and gender. Again this not quantitative entities. Lastly, the category that stood out the most was grades and conceptual understanding. Here we only care about the grades

(GPA) since there is no ambiguity it is easier to work with. Low course grades drive students away (regardless of conceptual understanding)

Slide #2 [Ryan]

Understanding Student Attrition in the Six Greater Toronto Area (GTA) Colleges by Tet S. Lopez-Rabson (Seneca College) et al. Study conducted in Canada across six colleges that seeks to better understand the factors motivating college departure and identify post-attrition pathways that college dropouts undertake. This paper had a great deal of information but most of it could not be used since we did not care so much about the post-attrition. However, the information we found was very insightful about student age. The age group below 25 year olds had general common reason for leaving. It shows them having a high rate of attrition due to financial, academics (changes/issues) and interest in program while personal reasons were low. Personal reasons being external factors such as family related or factors unrelated to college. The ages between 26 to 36 year olds had a high dropout rate in personal reasons while the others were low. The 37 year olds student will be ignored because it has small percentage of old students and most of are just going for fun rather than obtaining a degree.

Slide #3 [Ryan]

Student Attrition: Consequences, Contributing Factors, and Remedies by Ascend Learning, LLC. This paper mostly focuses on nursing schools and talks about the interactions between program and student. Also mentioning several strategies, policies and processes that help improve student attrition. But the most profound aspect of the paper is about admission tests. They have been shown that it can improve the student success rate in the program therefore lowering student attrition.

Slide#4 [Tereza]

The following study we investigated is The "Big Picture": Key Causes of Student Attrition & Key Components of a Comprehensive Student Retention Plan, by Joe Cuseo who is associated with Marymount College. This study showcases the root causes of attrition which include:

Academic roots – which is essentially student withdrawal related to either inadequate preparation to meet the academic demands of college coursework or disinterest with the content of courses or their method of delivery to the student

Motivational roots – student attrition related to either low level of commitment to college in general or perceived irrelevance of the college experience

Psychosocial roots – social factors or emotional issues

Financial roots – student attrition related to inability to afford the total cost of college or perception that the cost of college overweighs its benefits

The author of this study also proposes a 12-principal retention programs and methods of its delivery

List 12 principals.

Slide #5 [Tereza]

Another study that we have examined is The Institutional Costs of Student Attrition by Nate Johnson, who is associated with the Postsecondary Analytics LLC

It focuses largely on the total cost that academic establishments bear for the students who drop out. It claims that 33% of students entering a 4-year college leave one without a degree. Therefore reducing student attrition is crucial to enhancing overall productivity for both students and academic entities. The data tested by this study are the Beginning postsecondary data (BPS) and the Integrated Postsecondary Education Data System (IPEDS). Both data sets consider national statistics and they collect data on a variety of topics, including student demographic characteristics, school and work experiences, persistence, transfer, and degree attainment.

Slide #6 [Tereza]

The last study we found useful is College Student Attrition and Retention, by Leonard Ramist, College Board Report No.81-1. This study claimed that between 10 and 35 % of students' dropout without a degree and those students were surveyed by the reasons of them dropping out. There were various reasons all of which include but are not limited to Academic matters, financial difficulties, motivational problems personal considerations, dissatisfaction with college, military service, full-time jobs, need for practical non-academic experience, and lack of initial plan to obtain a degree. These supported all of the previous study claims, unrelated to this particular one. Which brings us to our hypothesis.

Slide #7 [Tereza]

What he would like to test is whether ***Student's age, GPA, and standardize test score predict students attrition.*** If we are able to do so will be able to identify those students in high risk of attrition and intervene with resolution in a timely manner.

Slide #8 [Tereza]

The data set we have is sensitive and contains private information, pertinent to students of Hunter College.

It has about 1 million data points each carrying above 200 attributes. Some characteristics include GPA, Age, SAT score, Admission scores, Gender, Race Part-time or full time employment

If our work bring fruitful results, we will be able to proceed even further and address more factors of attrition.

Slide #9 [Ryan]

The method is pretty simple. We will clean the data we have for usage and this is where we expect most of our difficulties because we need to convert from DFA to CSV then work form there. This sound easy but the problem is there are one million data point and when the conversion occurs something may have change in the data which could be difficult to find. We are still considering how to make sure the data stays how want it. Next, we will select student that have not graduated within 6 years, basically assuming that they are drop outs. We will then use two machine learning algorithm to see if they have similar result. This will confirm our hypothesis if result are relatively similar.

If our hypothesis is true, we will would be able to go a step further and predict which students are at risk by doing the following. Select the student currently enrolled meaning student have not graduate and under 6 years duration in college. Compute the risk factor for each component. Standardize test score will have a static risk factor (numeric value). GPA will have to be dynamic risk factor (numeric value) depending on how many credit the student have earned. This is important because student usually have high risk in there sophomore year as our research suggested. Then we will sum the each component which give us the risk factor for the student.

Slide #10 [Tereza]

We will be using a data mining software called Weka. It is written in Java and it is open source, it contains a large number of various machine learning algorithms and is also an excellent visualization tool.

We also will utilize Stata, which is a data analysis and statistical software tool that will help us sort and modify the data we have.

Python is a very useful tool also and in particular we will apply Bokeh and Pandas libraries for data visualization.

We will also use R and Mathematica for quantitative results and we hope to get our hands on a server in order to be able to process this large data file we will be working with.

Slide #11 [Ryan]

We will try to prove our hypothesis: Student's age, GPA, and standardize test score predict students' attrition. If this is true we can take it a step further to compute each student risk factor. Ultimately, trying to lower student attrition by intervene before they decide to leave the program/college.