

Stat159 Project 3: Add a title

Liang Hao, Bret Hart, Andrew Shibata, Gary Nguyen

December 3, 2016

1 Introduction

For this project, we are assuming the role of an NGO which desires to provide at-risk, low-income, minority students a portfolio of colleges which would best serve their needs: a high-quality, low-cost education. This will ultimately be done using the shiny tool: students will enter in various criteria (their GPA, location, desired field of study, etc) and it will output a list of colleges ranked according to a score which we arbitrarily, but intelligently create as an ad-hoc indicator of quality of school for these types of students. This college quality score will be our response variable, and will be compiled in multiple ways: one, by searching through the columns of data and doing basic research on what makes an education good, such as the average salary of each institutions graduates after college, graduation rate for at-risk students, or median debt of graduates; in addition, we can choose salient response variables by using various methods to cluster the data and examine what variables seem indicative or dependent upon a good education. We will then combine these responses into one college quality score for the school: if there turn out to be five really profound responses, each response would be 0.2 of the final quality score. In addition, we will be using the last 5 years of college data for our model, but each earlier year will be given less weight, in a polynomial manner, to still let the earlier data inform our model, but to not give it equivalent importance to the more recent years. To choose our predictor set, which will partially be entered in by students and partially inferred to predict how a student may do in respect to the score, we can use many different feature selection algorithms, such as lasso regression or pruning random forests or support vector machines for predictor selection. The interesting specificity of our model is that it is geared not toward creating a score for general school quality and a prediction for the average student, but specifically for helping at-risk students who need assistance in deciding which schools are worth the application in their specific circumstance.