Introduction

The primary focus of this analysis was to determine which variables are most predictive of four-year graduation for each college. In other words, on a college by college basis, which variables could classify students into successful vs. unsuccessful four-year graduates the most efficiently. Many of the variables under study are severely impacted by survivorship bias. If this bias was not accounted for, a few variables would appear to be far more predictive than they really are, and therefore, every college would appear to be subject to just those few variables. To account for this, every single analysis, statistic, and student in the following analysis has been adjusted in the following way. Only students who either graduated within four years, or who were still taking classes spring semester of 2013, or who took a class summer of 2013 were included in any of these analyses. There were 4,840 of such students out of 6,166 students in the entire cohort (78.5%). This analysis, then, compares students on their four-year graduation outcomes based on the differences in behavior while they were here, not comparing to students who were not here anymore due to dropping out or transferring.

The order of the colleges represents the size of remaining cohort, after adjusting for survivorship bias, from largest to smallest. To do the analysis I used Chi-square Automatic Interaction Detection (CHAID) to determine the most significant variables that could be used to split the data into the purest groups, defined as groups that are most uniform in four-year outcome. I will present the fitted classification tree and then do a brief analysis of each node (split) in the tree to quantify the differences in outcomes. It will become apparent that, among the smaller colleges, there will only be a couple of variables included because there were not enough student observations to continue splitting the data.

First Year Engineering
4yr Grad Rate vs. Co-Op Indicator
First Year Engineering Students
With Spring 2010 CUM GPA (3.0, 3.5) and Attended Summer School

Co-Op Indicator

<table>
<thead>
<tr>
<th>Co-Op Indicator</th>
<th>4yr Grad Rate</th>
<th>College Avg.</th>
<th>Subset Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>48.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16.1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4yr Grad Rate vs. Co-Op Indicator
First Year Engineering Students
With Spring 2010 CUM GPA (3.5, 4.0)

Co-Op Indicator

<table>
<thead>
<tr>
<th>Co-Op Indicator</th>
<th>4yr Grad Rate</th>
<th>College Avg.</th>
<th>Subset Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>77.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18.8%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
College of Liberal Arts
Here I made a second classification tree because, despite having many observations, the split achieved above was not great. This is partly due to how high the four-year graduation rate is overall.
**4yr Grad Rate vs. Spring 2010 CUM GPA Range**

**College of Liberal Arts Students**

`GPA Range`  
- [0, 2.0]  
- (2.0, 2.5]  
- (2.5, 3.0]  
- (3.0, 3.5]  
- (3.5, 4.0]  

- `4yr Grad Rate`  
- `College Avg.`  
- `Subset Avg.`

- 15.2%  
- 45.3%  
- 65.3%  
- 76.8%  
- 88.0%

**4yr Grad Rate vs. Transfer Credit Earned Indicator**

**College of Liberal Arts Students**

- `Transfer Credit Earned Indicator`  
- `None`  
- At least 1  

- `4yr Grad Rate`  
- `College Avg.`  
- `Subset Avg.`

- 63.9%  
- 77.9%
College of Science
College of Technology

4yr Grad Rate vs. Summer School Indicator
College of Science Students
With Fall 2009 and Spring 2010 CUM GPA [3.5, 4.0]

Summer School Indicator

- No: 83.3%
- Yes: 69.6%

College of Technology
4yr Grad Rate vs. Spring 2010 CUM GPA
College of Technology Students

4yr Grad Rate vs. Summer School Indicator
College of Technology Students
Undergraduate Studies Program

The following analysis used all of the variables that the other by-college analyses had, plus one more. The new variable, labelled “Terms_in_USP” in the tree below, provides a count of the number of semesters a student was in the Undergraduate Studies Program. Again, the analysis was survivorship bias corrected, otherwise low values of Terms_in_USP would look bad for dropouts, and good for non-dropouts, which would provide no useful information.

This plot is the 4YR grad rate vs. number of terms as an explorer before I combined the insignificantly different levels seen in the tree above. Additionally, 2 students had 6 terms in the program. Neither of them graduated in 4yrs. SAS did not want to plot those, so they’re not seen. Having few observations of each category resulted in insufficiency between some levels despite apparent differences.
The green line and an above plot show poor outcomes for students with more than 3 terms in the program. We see here, however, that students with a good GPA of 3.0+ have similar outcomes to the program as a whole. Furthermore, in the mosaic plot below we can see students with more terms in the program are significantly overrepresented in the low GPA category.
This tree had issues with having too few observations in some of the “leaves” or terminal nodes (the bottom squares), most notably Node 8 and 9. I decided to include this analysis for two reasons, one being statistically significant findings, and second, the interesting features of Nodes 6, 7 which are briefly discussed further below.
This is a particularly interesting finding. Despite the small number of observations (59 in the “no” group and 24 in the “yes”), a significant increase in four year graduation rate occurred when students with low GPA attended summer school at Purdue.
College of Agriculture

4yr Grad Rate vs. Spring 2010 CUM GPA
College of Agriculture Students

- Node 3 (n = 20)
- Node 4 (n = 68)
- Node 6 (n = 50)
- Node 7 (n = 41)
- Node 9 (n = 96)
- Node 10 (n = 56)
College of Pre-Pharmacy

4yr Grad Rate vs. Fall 2009 GPA Range
College of Pre-Pharmacy Students

- 36.1% for GPA range [0, 3.5]
- 14.6% for GPA range (3.5, 4.0)