Analysis of College Rankings using Machine Learning

Casey Watkins, Prof. Bradley A. Hartlaub
COMP 401 Seminar, Kenyon College

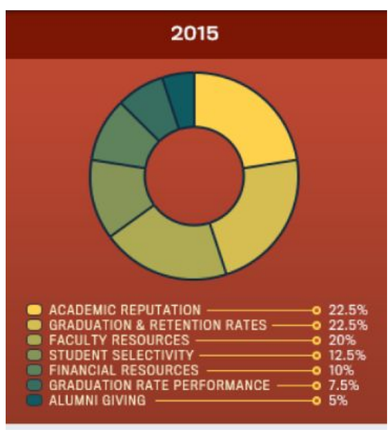
Motivation

- College rankings aim to provide insight into which schools are generally the ‘best’ and these lists are widely used by applicants
- However, these same rankings have faced controversy for their data policies and lack of transparency [1]
- Additionally, 70% of students believe these rankings are neither transparent nor reproducible [2]

US News’ Released Model

- This graphic [3] shows what *US News & World Report* include in their model

Figure 2: Through CFA, we find a Root Mean Square Error of Approximation (RMSEA) of .803 with n = 386, meaning this infographic’s model is insufficient



Model Analysis

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.92647    0.13191 370.913  < 2e-16 ***
fa1          0.48286    0.21133   2.285 0.023124 *
fa2          1.09829    0.09023  12.172  < 2e-16 ***
fa3         -1.36419    0.15119  -9.023  < 2e-16 ***
fa4          2.74349    0.34850   7.872 9.34e-14 ***
fa5          5.96171    0.41894  14.230  < 2e-16 ***
fa6         -3.10710    0.43776  -7.098 1.20e-11 ***
fa7         -3.42759    1.04346  -3.285 0.001160 **
fa8         -6.01218    0.80402  -7.478 1.15e-12 ***
fa9         -8.27379    2.25173  -3.674 0.000289 ***
fa10        51.64181   12.18561   4.238 3.13e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.175 on 261 degrees of freedom
Multiple R-squared:  0.9848,    Adjusted R-squared:  0.9842
F-statistic:  1694 on 10 and 261 DF,  p-value: < 2.2e-16

> |
```

Figure 1: This regression using derived factors can successfully predict *US News’* Overall Score for schools

- Confirmatory Factor Analysis demonstrates that the model to the left is insufficient to explain the rankings list provided
- Thus, we can use exploratory unsupervised learning to find that ten factors are required to explain the scores given
- This method uses many metrics and aggregates them into ‘factors’. Examples of metrics are financial aid rank or endowment

Designing Custom Rankings

- Using survey data and the exploratory factor analysis in Figure 1, we selected a myriad of different possible metrics to include in a new rankings list function
- From there, I built an app which allows users to select weights to give to each metric, customizing each list generated
- There are also dynamic rankings for metrics like geographic proximity or desired school size
- I deployed this app using shinyapps.io and RStudio for public access

App Access

Figure 3: A demo version of the app that’s still in development. Any feedback is welcome, but this should serve as a more transparent college search tool.

Amending these Rankings Lists

- The problem is the lack of transparency in these lists which disproportionately affects low-income and international applicants who rely on these rankings lists as opposed to tours and visits
- The goal was to create a more transparent list which adjusted to user inputs and desires, particularly for these marginalized groups in an anti-DEI era of admissions

INSTNM	CITY	SAT_AVG_ALL	ZIP	STABBR	REGION	LATITUDE	LONGITUDE	PRESTIGE_NOTEST	PRESTIGE_TEST	SIZE_SMALL	SIZE_MED
Princeton University	Princeton	1503	08544-0070	NJ	2	40.34873	-74.65936	89.43541	92.15095	93.53200	96.31973
Yale University	New Haven	1517	06520	CT	1	41.31116	-72.92669	87.55683	90.95911	88.34642	94.30974
Duke University	Durham	1516	27708	NC	5	36.00113	-78.93762	81.94079	86.72082	85.46222	93.75903
University of Pennsylvania	Philadelphia	1492	19104-6303	PA	2	39.95093	-75.19391	81.69719	86.22392	73.27100	94.21540
Harvard University	Cambridge	1520	02138	MA	1	42.37447	-71.11831	90.77777	93.86782	80.35077	92.44550
Vanderbilt University	Nashville	1514	37240	TN	5	36.14659	-86.80337	77.11798	83.94456	86.83522	95.19125
Northwestern University	Evanston	1508	60208	IL	3	42.05036	-87.67986	80.20060	85.85588	79.53262	93.72291
Massachusetts Institute of Technology	Cambridge	1545	02139-4307	MA	1	42.35924	-71.09323	88.08927	91.69518	90.85860	29.47434

Figure 4: An example of a ranking list with sortable subrankings from the research

References / Acknowledgements

- Thanks once again to Professor Hartlaub for his support, as well as Professor James Skon for offering this seminar and the Admissions Office.
- [1] Ali, S. (2022) The scandal facing college ranking lists, explained, The Hill.
- [2] Davis, M. (2016) Can College Rankings Be Believed?, She Ji: The Journal of Design, Economics, and Innovation, Volume 2, Issue 3, Pages 215-230
- [3] Boyington, B. (2013) Infographic: 30 editions of the U.S. News Best Colleges Rankings, US News & World Report.