

Background

Virginia Tech University is a land-grant university situated in Blacksburg attracting students from across the country from a diverse set of backgrounds. The Provost's office is tasked with oversight of all educational activities and student affairs. Ken Smith, the Vice Provost charged our team with exploring and identifying trends around the university's ability to attract students with particular emphasis on vulnerable populations.

Variables

The project leveraged and combined several data sources to identify a number of meaningful variables that break out across the following categories:

- **Individual Background** Demographic and family characteristics
- **Academic Performance** Scholarly performance, either for applicants during their high school years or for students within their degree program
- **Competitive Factors** University competitive factors, such as rank or graduation rate

Results

LASSO highlighted the variables as having significant impact on enrollment decisions. Of the original 36 variables, 18 were selected. A correlation plot was drawn against the variables to identify any potential variables with collinearity, and the resulting plot is shown in Figure 1. The plot indicates a correlation score of 0.5 between in-state tuition, residency-college state match, and high school college-goingness culture.

Logistic regression was run on these 18 variables to model enrollment decisions and the resulting model's coefficients are

Conclusion

In conclusion, we were able to successfully model the factors that influence enrollment decisions. Ultimately, 14 out of 18 variables were identified as important attributes for applicants' decision making process. Many of these variables were binary (such as gender or first-generation status), while some had multiple levels. For example, applicants were offered acceptance to one of the 7 VT colleges, a notably significant factor for enrollment decisions.

Moreover, 4 out of 14 significant variables showed a negative odds ratio, indicating that these variables decreased enrollment likelihood. These included gender, underrepresented minority status, VT college offered, and high school GPA. We believe that the Provost's office might want to explore potential strategies to attract applicants who show vulnerability in these areas.

Methods

The study utilized logistic regression with LASSO-selected features to model enrollment decision and successful student progression behaviors. LASSO (least absolute shrinkage and selection operators) is a regression analysis method that seeks to reduce the dimensionality of the data. A second technique, binomial logistic regression, was applied to the reduced datasets' variables to better model the relationships between our responses and the independent variables and capture the accuracy of our model.

diagrammed at left. The regression estimates indicate the number of times the probability shifts with that variable, while the bars show the standard deviation.

The model's prediction showed an accuracy rate of 82%. The ROC diagrammed at left shows the performance of the model, well above the baseline rate of 50%.

Next Steps

The 4 significant variables with negative impact on enrollment decisions are gender, underrepresented minority, college offered, and high school GPA. Future research into enrollment decision trends could be spent exploring patterns between VT colleges and identifying successful strategies for attracting applicants.

Potential strategies to attract applicants within these riskier populations might include:

- Incorporate and weight applicant interviews and essays alongside high school GPA as academic indicators.
- Conduct one-on-one welcome meetings for international student applicants.¹
- Continue to build a supportive and safe college environment for LGBTQ and minority applicants.

Data Sources

We utilized applicant data from a wide range of sources for this project:

- National Student Clearinghouse
- Virginia Department of Education
- College Board
- US News

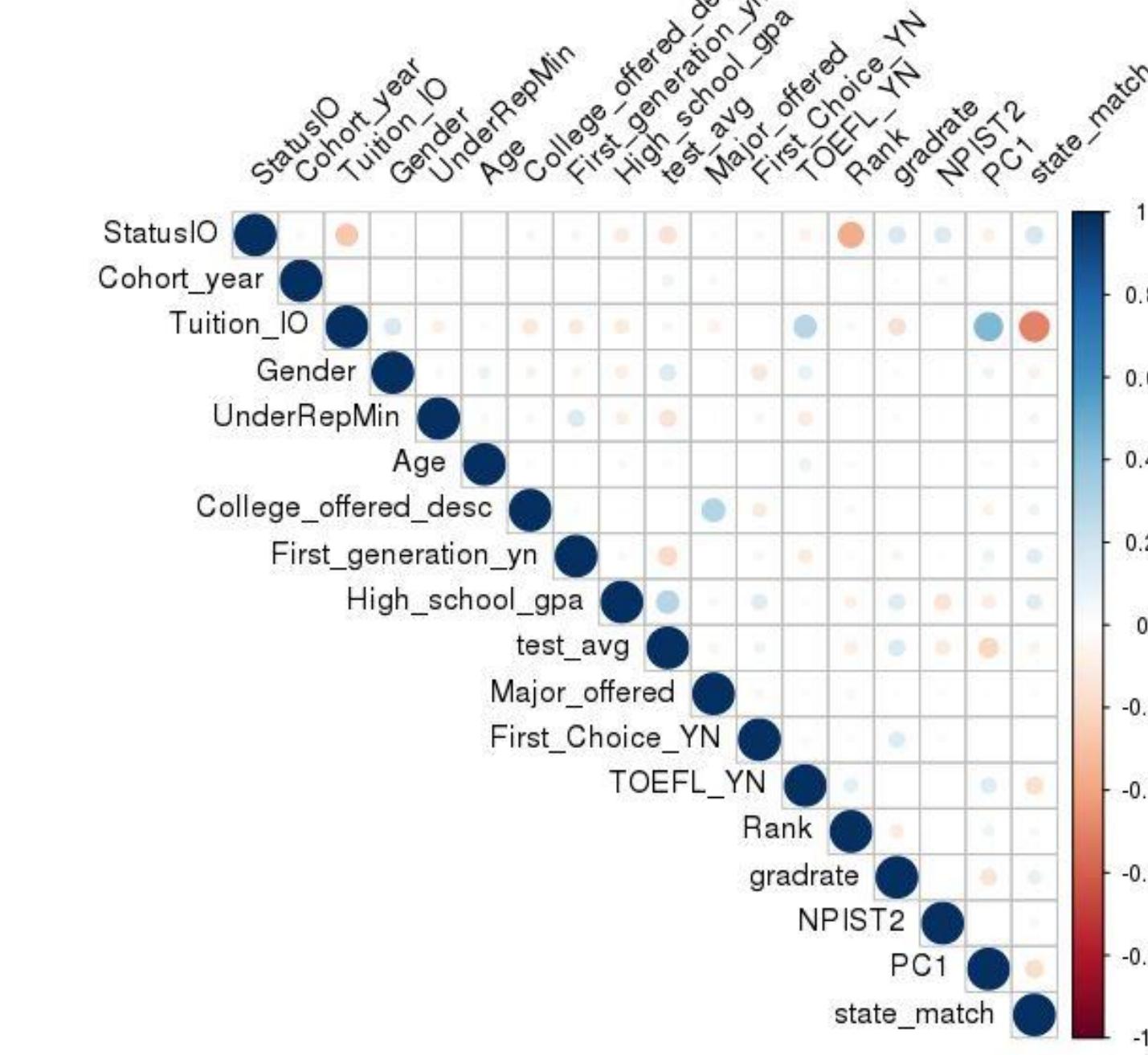


Figure 1. Correlation plot for all the 18 variables to show potential collinearity between variables.

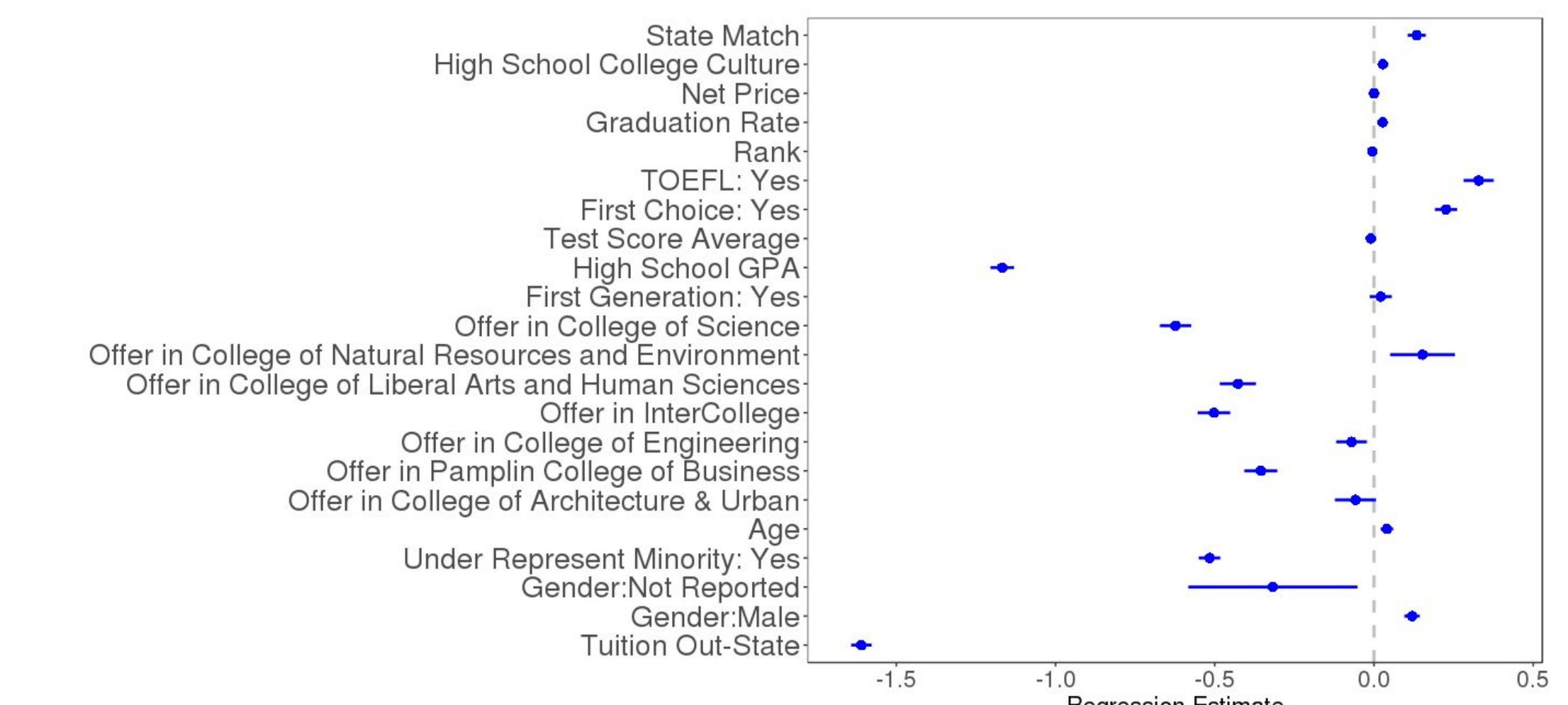


Figure 2. Coefficient plot of significant variables indicated by logistic regression: points show magnitude of log-odds ratio for each variable while bars indicate size of standard deviation.

References

1. Papandrea, Dawn. "International admissions: getting to know you: 5 ways virtual interviews are improving the process of admitting (or not) international student applicants." *University Business*, June 2016, p. 32+. *ducators Reference Complete*, http://link.galegroup.com.ezproxy.lib.vt.edu/apps/doc/A455058055/PROF?u=viva_vpi&sid=PROF&xid=fe36e197. Accessed 12 June 2018.
2. Calisir, Fethi, et al. "Predicting academic performance of master's students in engineering management" *College Student Journal*, vol. 50, no. 4, December 2016, p. 501+. *Educators Reference Complete*, http://link.galegroup.com.ezproxy.lib.vt.edu/apps/doc/A474548660/PROF?u=viva_vpi&sid=PROF&xid=b47b9880. Accessed 19 June 2018.
3. Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B (Methodological)* 58(1):267-288

