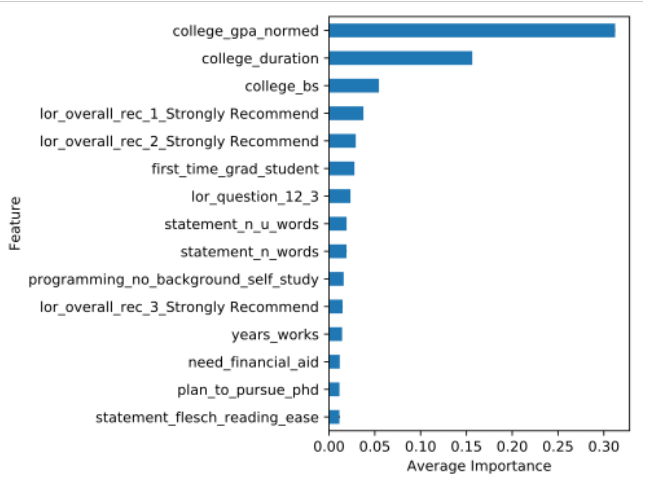

Figure 2. Feature importance in the top scoring model fit.

said, feature importance ranking provided in Figure 2 implicates the relative importance of the input parameters. However, correlation analysis is planned to be performed in the future phase of the project.

CONCLUSION AND FUTURE WORK

In this paper, we reported on the progress that we have made on building a robust model that can predict the admission status of applicants to the online Master's program at Georgia Tech. First, we were successfully able to process a massive amount of data from more than 9,000 candidates and reduce the dimensional complexity of the raw data. Through data processing, we identified 154 key variables to construct the matrix of the input variables for predictive modeling. We were also able to train the model and test model performance with reasonably high accuracy by using various machine learning algorithms. These results from our preliminary analyses suggest that data from graduate program applicants, including our OMSA applicant dataset, can provide a rich and promising basis so for applying machine learning techniques.

In the next steps of the project, we plan to expand the scope of our dataset and feature processing in several ways. Beyond whether an applicant was admitted to the OMSA program, we will use applicants' data to model the successful completion of the Analytics program. To do that, we will initially build models to predict students' grades in three core courses of the program. Then, based on those models' performance, we will construct models for predicting whether students drop out of the program, graduate on time, and land a job in analytics. Additionally, we plan to incorporate techniques such as Shapley Additive Explanation to ensure that our models do not discriminate based on demographic factors. Ultimately, we hope that our research will offer useful guidance for the OMSA program's admission process and help administrators make informed decisions contributing to program improvement.

REFERENCES

[1] April Bleske-Rechek and Kingsley Browne. 2014. Trends in GRE scores and graduate enrollments by gender and ethnicity. *Intelligence* 46 (2014), 25–34.

[2] Benjamin Halberstam and Fran Redstone. 2005. The predictive value of admissions materials on objective and subjective measures of graduate school performance in speech-language pathology. *Journal of Higher Education Policy and Management* 27, 2 (2005), 261–272.

[3] J Daniel House, James J Johnson, and William L Tolone. 1987. Predictive validity of the Graduate Record Examination for performance in selected graduate psychology courses. *Psychological Reports* 60, 1 (1987), 107–110.

[4] Chang-Ho C Ji. 1998. Predictive validity of the Graduate Record Examination in education. *Psychological Reports* 82, 3 (1998), 899–904.

[5] Daniel J Koys. 2005. The validity of the Graduate Management Admissions Test for non-US students. *Journal of Education for Business* 80, 4 (2005), 236–239.

[6] Nathan R Kuncel, Marcus Credé, and Lisa L Thomas. 2007. A meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education* 6, 1 (2007), 51–68.

[7] Nathan R Kuncel, Serena Wee, Lauren Serafin, and Sarah A Hezlett. 2010. The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement* 70, 2 (2010), 340–352.

[8] Casey W Miller. 2013. Admissions criteria and diversity in graduate school. *arXiv preprint arXiv:1302.3929* (2013).

[9] Bryan Nankervis. 2011. Gender Inequities in University Admission due to the Differential Validity of the SAT. *Journal of College Admission* 213 (2011), 24–30.