Using Machine Learning to Understand Physics Graduate School Admissions

Nicholas T. Young[1,2], Marcos D. Caballero[1,3,4]

[1]Department of Physics and Astronomy, Michigan State University
[2]Department of Computational Mathematics, Science, and Engineering, Michigan State University
[3]CREATE for STEM Institute, Michigan State University
[4]Department of Physics and Center for Computing in Science Education, University of Oslo

Highlights

- Studied graduate admissions by analyzing the applications
- GPA & physics GRE score are most predictive of being admitted to this program.
- Can predict with 75% accuracy who is admitted using just GPA & physics GRE score
- Metrics heavy admissions practices can limit number of gender & racial minority students in graduate physics.

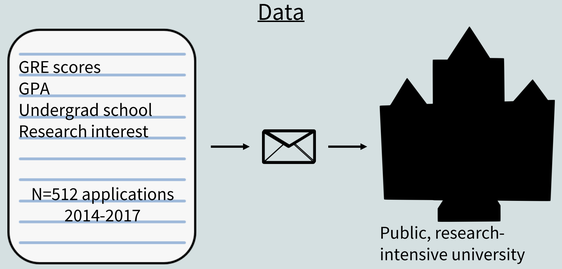
Background

- Studies of the faculty[1] and the process[2], but not the applications

- Mix of data types & binary outcome makes problem ideal for machine learning
- **Goal: Use machine learning to identify factors most predictive of applicant being admitted to a physics graduate program.**

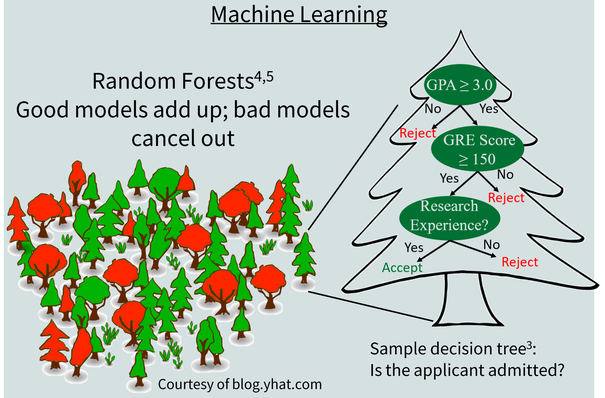
Methodology

Data



Machine Learning

Random Forests[4,5]
Good models add up; bad models cancel out



- Important variables determined by changes to model when removing that variable[6]
- Meaningful variables determined by minimum needed to make "good enough" model[7]

Results

Undergraduate GPA & physics GRE scores are the main predictors of being admitted to a physics graduate program

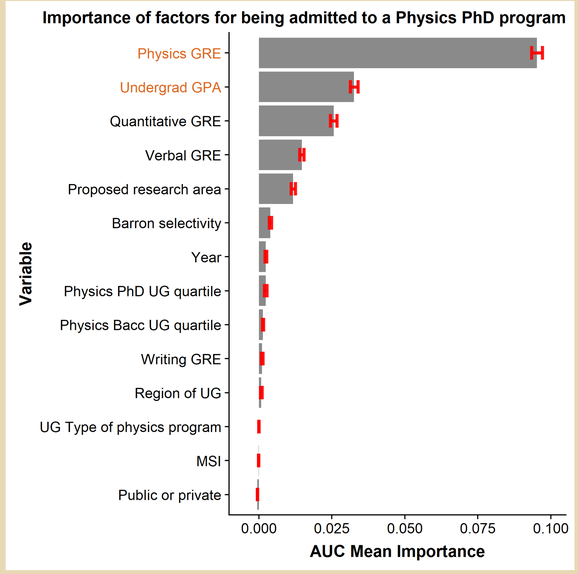


Figure 1: Importances of the factors used in our admissions model. Metrics tend to be more important than characteristics of the applicant's undergraduate institution.

Is applicant admitted to the physics graduate program?		Actual Decision: Not Admitted	Actual Decision: Admitted
Model Prediction	Not Admitted	**40.3%**	14.9%
	Admitted	9.1%	**35.7%**

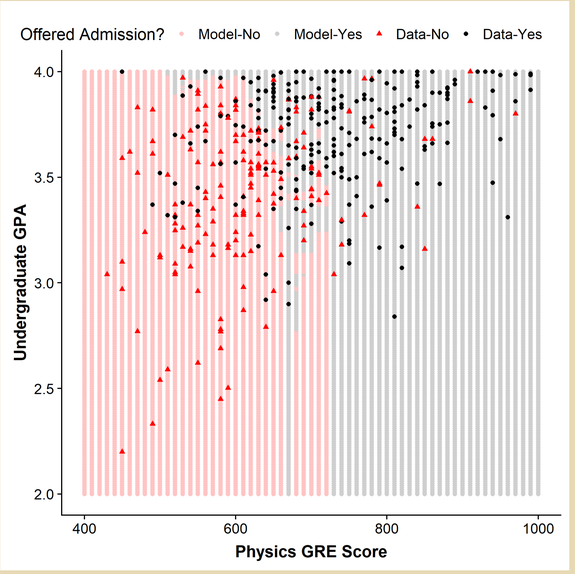


Figure 2: Admitted status of actual applicants (data) & for hypothetical students (Model) with given score & GPA. Higher GPA seems to compensate for lower GRE score.

How does your program compare?

We want to study your program! Please scan the QR code if you are interested.

Discussion & Future Work

- GRE scores biased against underrepresented minorities[8]
- Racial minority students may not apply to graduate programs if they don't believe they will be admitted[9].
- New admissions process considering non-cognitive factors and fit at this university. Will metrics still be valued as highly?
- How representative of graduate admission practices is this program?

Acknowledgements & Citations

We would like to thank the graduate program directors at the studied institution who graciously provided admissions data and insight into their admission process. We would also like to the thank the various members of PERL who have provided feedback on this project.

[1]G. Potvin, D. Chari, and T. Hodapp, Investigating approaches to diversity in a national survey of physics doctoral degree programs: The graduate admissions landscape, Physical Review Physics Education Research 13, 020142 (2017).
[2]J. R. Posselt, Inside graduate admissions (Harvard University Press, 2016).
[3]Rokach, Lior, and Oded Maimon. Data mining with decision trees: theory and applications. World scientific, 2014.
[4]Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.
[5]Strobl, Carolin, et al. "Bias in random forest variable importance measures: Illustrations, sources and a solution." BMC Bioinformatics 8.1 (2007): 25.
[6]Janitza et al. "An AUC-based permutation variable importance measure for random forests." BMC Bioinformatics 14:119 (2013)
[7]Diaz-Uriarte, Ramon and Sara Alvarez de Andres. "Gene selection and classification of microarray data using random forest." BMC informatics 7:3 (2006).
[8]C. Miller and K. Stassun, A test that fails, Nature 510, 303 (2014).
[9]G. L. Cochran, T. Hodapp, and E. E. A. Brown, Identifying barriers to ethnic/racial minority students' participation in graduate physics, in Physics Education Research Conference Proceedings, PER Conference (Cincinnati, OH, 2018) pp. 92–95