

# **PREDICTING ACADEMIC OUTCOMES WITH DAACS**

---

Daniel Moscoe

CUNY SPS

12/6/2021

# Overview

- How can we support students who need it most, *before* they fail a course?
- Predictive models:
  - Demographic / assessment data
  - SuperLearner / Random Forest

# Overview

- **SL  $\approx$  RF**
- **Models are useful for categorizing students as higher or lower risk.**
- **Including assessment data improves models.**

# Context

- About 1/3 of entering undergraduates never earn a degree.
- Widespread models for remediation show poor results.
- Negative academic outcomes are costly.
- DAACS implements best practices for student support from assessment to advising.

# Response variables

- Measured 5 binary academic outcomes:
  - “Success” in term 1, term 2 (binary)
  - “Retained” at the end of term 1 (binary)
  - “Credit Ratio” at the end of term 1, term 2 (binary)
  - Total positive outcomes

# Explanatory variables

- Student data from two institutions, Western Governors' University ( $n = 6,260$ ) and Excelsior College ( $n = 2,532$ )
- Demographic: Gender, ethnicity, first-gen, military, income, etc.
- DAACS: SRL, math, reading, writing.

# Methods

- **SL and RF model for each of:**
  - 2 institutions
  - 5 outcomes
  - 2 exp var subsets
- **Total = 20 pairs of SL, RF models.**

# Methods

- **SL component models:**
  - *k*NN ( $k = 10, 15, 20, 25$ )
  - GLM ( $\alpha = 0.25, 0.50, 0.75, 1$ )
  - RF
  - Bagged trees (SL.ipredbagg)
  - Mean



# Are SL preds statistically significant?

- Chi-square test of independence
- Null hyp: predicted count of positive outcomes is randomly distributed across students.
- Result: For both institutions, for both groups of exp vars, reject null.

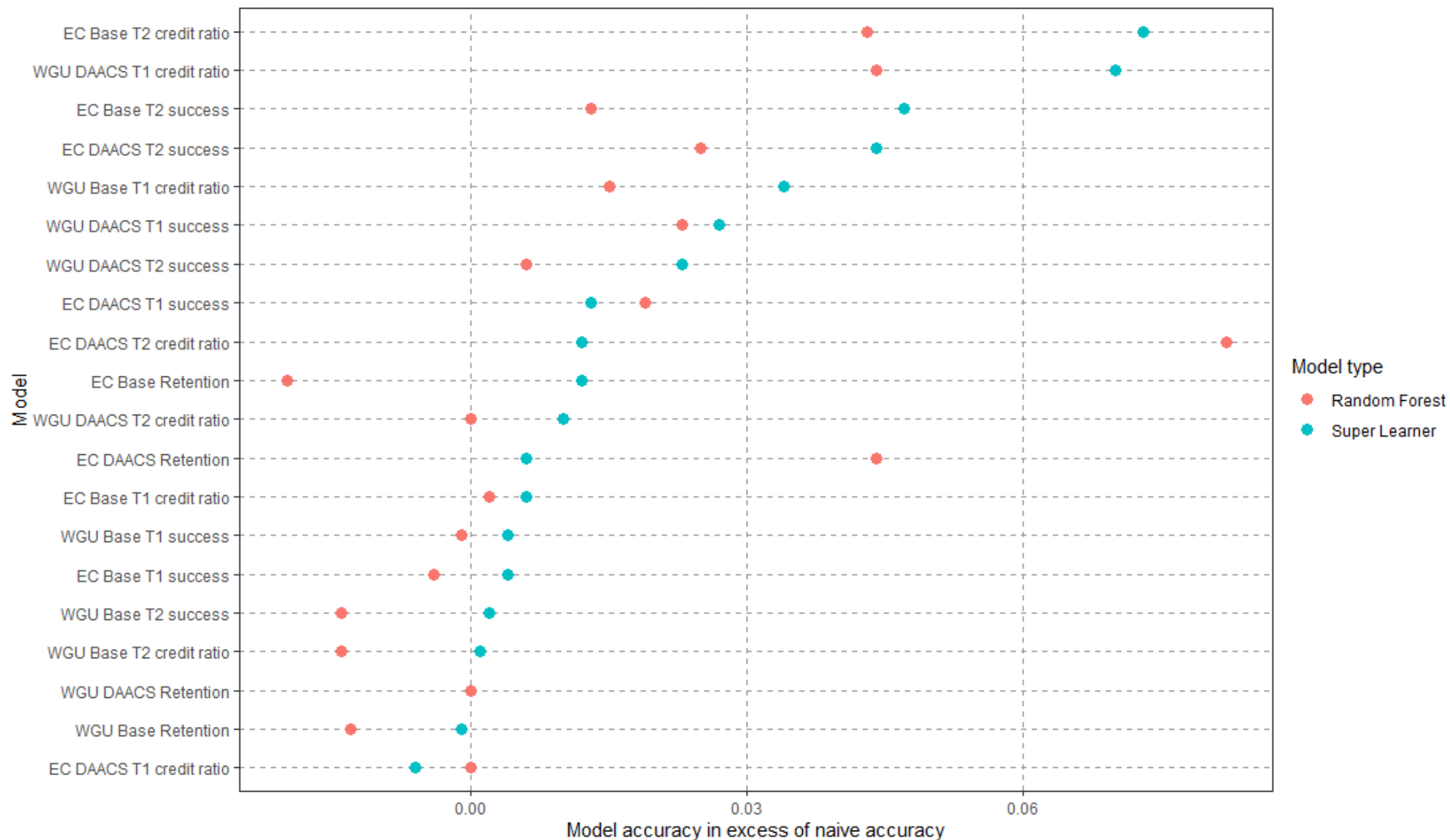
# How can we target students for support?

- For each institution, for each group of EVs, sort students as higher or lower risk.
- *t*-test for difference of means.
- Null hyp: mean positive outcomes are equal.
- Result: Reject null for all inst/EV combinations.

# Targeting students for support

	Predicted fewer positive outcomes	Predicted more positive outcomes	Difference
EC mean positive outcomes (demo)	2.97 (253)	3.47 (254)	0.50 ( $p = 0.001$ )
EC mean positive outcomes (DAACS)	3.09 (78)	3.71 (79)	0.62 ( $p = 0.027$ )
WGU mean positive outcomes (demo)	2.39 (626)	3.02 (626)	0.63 ( $p < 0.001$ )
WGU mean positive outcomes (DAACS)	2.23 (626)	3.23 (626)	1.00 ( $p < 0.001$ )

### Super Learner and Random Forest Accuracy In Excess of Naive Accuracy



# Final points

- Assumes additional support is never costly/harmful to students.
- Summing positive outcomes “made the signal louder.”
- Predictive models vary across institutions.