PREDICTING ACADEMIC OUTCOMES WITH DAACS: DIAGNOSTIC ASSESSMENT AND ACHIEVEMENT OF COLLEGE SKILLS

Daniel Moscoe

Abstract

Enhanced advising has been shown to improve the academic performance of at-risk college students (Bailey, Bashford, Boatman, Squires, Weiss, Doyle, Valentine, LaSota, Polanin, Spinney, Wilson, Yeide, and Young, 2016). Because institutions have scarce resources for student academic support, and usually cannot offer enhanced advising or other special academic support to every student. As such they may seek an "early alert system" that can identify students least likely to persist in their studies or attain other academic benchmarks. This study compares Super Learner ensemble models trained on demographic and DAACS assessment data to predictive models based on random forests alone. Super Learners do not attain greater accuracy than random forest models in predicting student outcomes. Predictions based on Super Learners can form the basis of an early alert system that directs academic interventions to the students who need them the most. The early alert system is more effective when it incorporates DAACS assessment data than when it relies on demographic data alone.

Introduction

Students entering college for the first time face an array of challenges unlike any they may have encountered previously (Credé and Niehorster, 2011). Academically, they are expected to master increasingly sophisticated content with greater independence. Socially, they must negotiate their belonging within new peer groups and an institution that may feel impersonal (Daugherty and Lane, 1999). In addition, they may be academically underprepared for college-level work, or they may bring with them attitudes and beliefs about themselves or about school that are inimical to learning (Mokher, Barnett, Leeds, and Harris, 2019). The result is that student performance in college varies widely, with about a third of entering undergraduates never earning a degree, although attrition rates vary widely across schools (Leonhardt & Chinoy, 2019).

The response from institutions to variability in student academic preparation, as well as to the high numbers of students who drop out, has generally been to provide remedial coursework to students who perform poorly on placement exams. Poor performance is common: The Community College Research Center reports that "about half of all entering college students take at least one remedial course, and among those who take any, the average is 2.6 remedial courses" (Community College Research Center, 2015, p. 1). However, the effectiveness of remedial coursework has recently come under increased scrutiny. Bailey and Cho (2010) report that "less than one quarter of community college students who enroll in developmental education complete a degree or certificate within eight years of enrollment in college" (p. 1). Several studies indicate that "students who participated in remediation did no better on several outcome measures than similar students who enrolled directly in college-level courses" (Bailey and Cho, 2010, p. 2).

A different model for student support is described by the What Works Clearinghouse (WWC) in their *Strategies for Postsecondary Students in Developmental Education- A Practice Guide for*

College and University Administrators, Advisors, and Faculty. This model includes strategies for delivering streamlined yet effective academic remediation, for strengthening students' sense of belonging at their institution, and for improving students' ability to regulate their own learning (Bailey et al., 2016).

One recommended strategy is that schools "require or incentivize regular participation in enhanced advising" (Bailey et al., 2016). Enhanced advising provides intensive, holistic support to students, and includes sustained counseling on "a range of issues, including course selection, registration, financial aid, other financial issues, tutoring, work-based learning efforts, juggling school and work, career aspirations, and personal issues" (Scrivener, Weiss, and Teres, 2009, p. ES-3).

At many institutions, capacity for enhanced advising is constrained by high student-to-advisor ratios. Six hundred students per advisor is not uncommon, and this ratio can rise as high as 1,000 students per advisor (Scrivener et al., 2009).

Because institutions have limited advising resources, and because enhanced advising can significantly reduce the likelihood that a student will fall behind, accurately matching at-risk students with advisors is critical. To match at-risk student with advisors, institutions may implement an "early alert system," the purpose of which is to "identify and support students at risk of attrition in order to improve student success, retention and persistence" (Lynch-Holmes, Troy, and Ramos (n.d.), p. 2).

This project seeks to improve and extend predictive models in DAACS to provide an early alert system that identifies students least likely to persist and succeed in their studies. DAACS is a suite of assessments and intervention strategies that implements the WWC recommendations for developmental education for undergraduates (Bryer, Akhmedjanova, Andrade, and Lui, 2021). DAACS incorporates predictive models based on random forests for assessing college readiness, both in terms of academic skills (are students prepared for college-level coursework?) and in terms of progress toward degree (are students likely to drop out or fail to meet certain benchmarks of progress?). This project seeks to improve DAACS random forest predictive models by combining linear, tree, and Bayesian component models together in an ensemble model called a Super Learner. With these Super Learner models, I then build an early alert system to identify students in greatest need of academic support. This early alert system can help colleges that implement DAACS direct enhanced advising and other support services to students who need them the most.

Literature Review

The literature on predictive models for academic progress in college describes two broad categories: models that predict grades (either overall GPA or grades in coursework in a specific subject area), and models that predict attrition. Many entering college students (including nearly all those entering 2-year institutions) complete assessments to predict their performance in specific coursework. These assessment scores are then used to route them to remedial or credit-bearing classes (Scott-Clayton, 2012).

Although research supports the use of multiple measures of college readiness (Scott-Clayton, 2012), no institutions in the literature I reviewed were described as considering anything other than placement exam scores when making placement decisions or otherwise predicting the academic success of incoming students. A second measure of college readiness, high school GPA, has attracted attention as a statistically significant predictor of college performance among researchers (for example, Belfield and Crosta, 2012). One study found that "When we control for high school GPA, the correlation [between performance on placement tests and college GPA] disappears" (Belfield & Crosta, 2012, p. 2). However, "Placement test scores are positively associated with college credit accumulation even after controlling for high school GPA" (Belfield & Crosta, 2012, p. 2).

Scott-Clayton (2012) reports that two college placement exams dominate the market: ACCUPLACER, developed by the College Board, and COMPASS, developed by ACT. She found that "observed correlation coefficients [between exam performance and course performance] (available only for ACCUPLACER) are generally higher for math than for reading/writing and are generally higher for a B-or-higher success criterion than for a C-or-higher success criterion" (p. 7). Placement accuracy rates range between 60 and 80 percent for both ACCUPLACER and COMPASS. (Placement accuracy "is calculated as the sum of 'observed true positives'—students who are placed at the college level and actually succeed there—and 'predicted true negatives'—students who are not predicted to succeed at the college level and are 'correctly' placed into remediation") (p. 7).

Other variables that have been found by various researchers to predict undergraduate academic performance include a student's academic preparation in high school, gender, race, (Chingos, 2018), psychological factors including growth mindset (Mesmin, 2018), self-regulated learning (Bryer et al., 2021), and the quality of the institution (Leonhardt & Chinoy, 2019).

Several studies construct predictive models for undergraduate attrition, but in the literature I reviewed there was no mention of an institution employing such a model for incoming undergraduates.

Daugherty and Lane (1999) examined attrition behavior of college students at an all-male military institution over a period of four years. "A linear combination of academic ability, family legacy status, specific stress perceptions, and self-perceived social alienation was found to predict attrition status" with accuracy of about 70% (p. 355). They found that several measures of the student's perceived belonging at the school, including family legacy status, influenced their attrition behavior. This finding is consistent with that of Scrivener et al. (2009), who note that "the more integrated, engaged, and generally satisfied student will be more likely to succeed in school" (p. 6).

A second study echoed the claim that non-academic psychological factors can play an important role in student retention. Mesmin (2018) found that "approaches that incorporate psychological factors—such as encouraging growth mindsets, linking classroom work to real-world aspirations, and using online modules that help activate students' motivation and sense of belonging—can improve student success in higher education."

DAACS as it is currently implemented demonstrates the main recommendations from the literature. It includes measures of academic attainment and self-regulated learning. Its predictions of undergraduate retention, incorporating demographic and assessment data, are around 70% accurate. These DAACS models are currently based on random forests.

There is good reason to think that a random forest model may be nearly optimal for this data. Fernández-Delgado, Cernadas, Barro, and Amorim (2014) evaluated 179 classifiers on 121 data sets from the UCI machine learning database. They found that "The classifiers most likely to be the best are the random forest (RF) versions" (p. 3133). In the absence of theoretical knowledge about the underlying structure of the data, a random forest model is often a fruitful starting point. These researchers found similarly high performance across a range of data sets using support vector machines with Gaussian and polynomial kernels.

Combining a random forest model together with other types of classifiers in a Super Learner may improve model performance. "The Super Learner is a prediction method designed to find the optimal combination of a collection of prediction algorithms. The Super Learner algorithm finds the combination of algorithms minimizing the cross-validated risk" (Polley & van der Laan, 2010, abstract). A Super Learner is also "asymptotically optimal," since in the case where a single component model outperforms any linear combination of component models, the Super Learner simply collapses to that specific model.

Research Question

Can a Super Learner ensemble model predict student academic outcomes with accuracy greater than that attained by random forest models alone?

Data and Variables

This study uses demographic data and DAACS assessment results to predict the following academic outcomes:

- Term 1 success
  - At EC, *True* if the student accrues at least 3 credits during term 1, and *False* otherwise.
  - At WGU, *True* if the student accrues at least 12 credits during term 1, and *False* otherwise.
- Term 2 success
  - Defined analogously to term 1 success.
- Retention
  - At both institutions, *True* if the student attempts at least 1 credit in term 2, and *False* otherwise.
- Term 1 credit ratio
  - At both institutions, *True* if credits attempted in term 1 is equal to credits earned in term 1, and *False* otherwise.
- Term 2 credit ratio
  - Defined analogously to term 1 credit ratio.

The variable “total positive outcomes” is the sum of the five response variables.

Data from incoming students at Western Governors University (WGU) and Excelsior College (EC) are used to train predictive models. This data was collected during an initial study of DAACS. Students in the control group at WGU did not take any of the DAACS assessments, and students in the treatment group completed all the assessments as part of their orientation process. At EC, students in the treatment group were asked to complete the DAACS assessments, but only 58% of these students completed at least one assessment.

The goal of this project is to predict academic outcomes using demographic variables and DAACS assessment scores. Therefore, I use only the observations that contain scores for all the DAACS assessments. Missing demographic data is imputed using multiple imputation with chained equations as implemented in the R package Mice.

DAACS assessments consist of groups of subtests in math, reading, and self-regulated learning (SRL). The writing assessment asks students to construct a single essay reflecting on the results of their SRL assessment. Total math and reading scores are included in this study, along with SRL component scores for grit, self-regulated learning strategies, metacognition, anxiety, mastery orientation, mindset, and self-efficacy. In the tables that follow, SRL scores are summarized by total score.

Variables used for predictive modeling are described in the tables on the following pages.

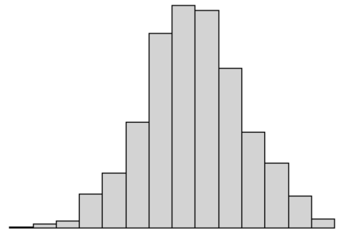
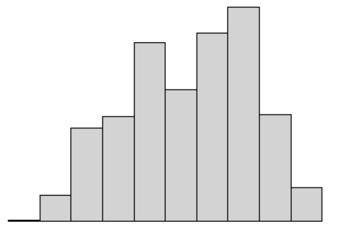
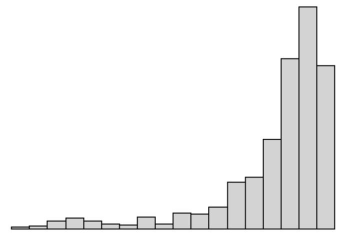
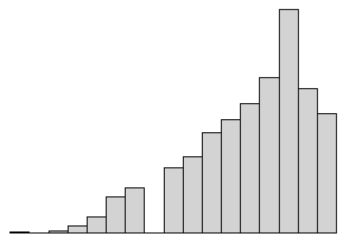
Demographic and response variables, Excelsior College
(*n* = 2532)

		n	%
Demographic: Gender	Female	941	37
	Male	1591	63
Ethnicity	Asian	95	4
	Black or African American	409	16
	Hispanic	331	13
	White	1517	60
	Unknown	180	7
First generation	No	2075	82
	Yes	457	18
Active military	No	1346	53
	Yes	1186	47
Veteran	No	2125	84
	Yes	407	16
Employment status	Not employed	371	15
	Employed	2038	80
	NA	123	5
English language native	No	160	6
	Yes	2348	93
	NA	24	1
Program division	Health Sciences	163	6
	Liberal Arts	771	30
	Nursing	439	17
	Public Service	193	8
	Technology	555	22
Income	Less than $25k	177	7
	Less than $35k	189	7
	Less than $45k	194	8
	Less than $55k	200	8
	Less than $70k	235	9
	Less than $85k	215	8
	Less than $100k	131	5
	Less than $120k	119	5
	Greater than or equal to $120k	136	5
	NA	936	37
Response: Retained	True	1524	60
	False	1008	40
Term 1 success	True	2030	80
	False	502	20
Term 2 success	True	1353	53
	False	1179	47
Term 1 credit ratio	True	2017	80
	False	515	20
Term 2 credit ratio	True	1292	51
	False	1240	49
Total positive outcomes	5	1096	43
	4	144	6
	3	155	6
	2	700	28
	1	295	12
	0	142	6

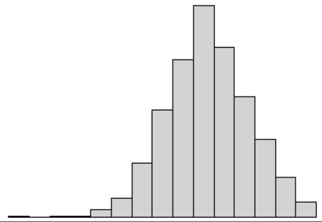
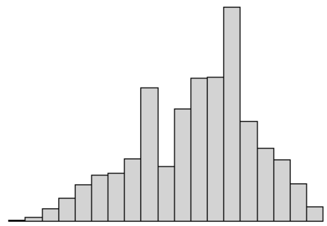
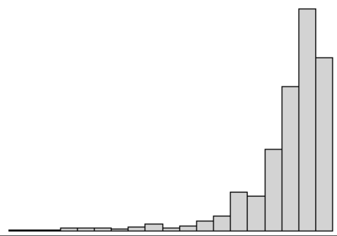
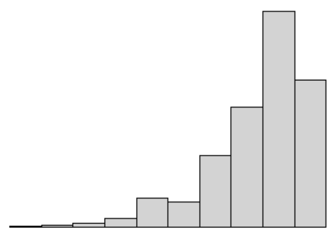
Demographic and response variables, Western Governors University
(n = 6260)

		n	%
Demographic: Gender	Female	3644	58
	Male	2611	42
	NA	5	<1
Ethnicity	Asian	179	3
	Black	717	11
	Hispanic	698	11
	White	4706	75
	NA	142	2
First generation	No	3743	60
	Yes	2517	40
Military	No	5582	89
	Yes	678	11
Employment status	Not employed	807	13
	Part time	797	13
	Full time	4326	69
	NA	330	5
Citizenship status	Noncitizen	59	1
	Nonresident alien	30	<1
	U. S. Citizen	6104	98
	NA	67	1
Income	Less than $16k	493	8
	Less than $25k	649	10
	Less than $35k	878	14
	Less than $45k	824	13
	Less than $65k	1125	18
	Greater than or equal to $65k	1889	30
Response: Retained	True	4720	75
	False	1540	25
Term 1 success	True	4344	69
	False	1916	31
Term 2 success	True	2432	39
	False	3828	61
Term 1 credit ratio	True	3622	58
	False	2638	42
Term 2 credit ratio	True	1860	30
	False	4400	70
Total positive outcomes	5	1523	24
	4	745	12
	3	1324	21
	2	603	10
	1	1205	19
	0	860	14

Total DAACS assessment scores, Excelsior College

Variable	*Q1*	*Median*	*Q3*	*Histogram*	*Notes*
Total SRL score	2.51	2.78	3.07		Total SRL scores are normally distributed.
Total math score	0.44	0.61	0.72		Total math scores are highly variable.
Total reading score	0.83	0.89	0.94		Total reading scores are skewed left.
Total writing score	0.67	0.83	0.89		Total writing scores are skewed left.

Total DAACS assessment scores, Western Governors University

Variable	*Q1*	*Median*	*Q3*	*Histogram*	*Notes*
Total SRL score	2.68	2.95	3.21		Total SRL scores are normally distributed.
Total math score	0.50	0.63	0.75		Total math scores are highly variable.
Total reading score	0.83	0.94	0.94		Total reading scores are skewed left.
Total writing score	0.72	0.83	0.89		Total writing scores are skewed left.

Statistical Methods

Can a Super Learner ensemble method predict student academic outcomes with accuracy greater than that attained by random forest models alone?

The main research question of this paper compares the accuracy of Super Learner ensemble models to that of random forests when predicting student academic outcomes. For each response variable, a Super Learner and a random forest model were constructed at each institution using either demographic data only ("base models") or demographic data combined with DAACS assessment data ("DAACS models"). The five response variables modeled at two institutions using two sets of variables altogether resulted in 20 pairs of Super Learner and random forest models.

The component models of each Super Learner model are:

- *k*-nearest neighbors (SL.kernelKnn), (*k* = 10, 15, 20, 25)
- Generalized linear model, (SL.glmnet), (α = 0.25, 0.50, 0.75, 1.00)
- Random forest (SL.randomForest)
- Classification trees with bagging (SL.ipredbagg)
- Mean (SL.mean)

KNN and GLM models were included with a variety of hyperparameter values. This allows for model selection to be performed by the SuperLearner package itself. The particular KNN and GLM models associated with the best performance are included in each ensemble, and component models with poor performance are discarded (that is, assigned a coefficient of 0 in the resulting ensemble).

The accuracy of each Super Learner was computed for a holdout set, and this was compared to the accuracy of the corresponding random forest model. Accuracies were also compared to the "naïve accuracy," equal to the fraction of observations in the majority class.

Are predictions of student academic outcomes based on Super Learners statistically significant?

Because there is no clear systematic difference between naïve accuracy, random forest accuracy, and Super Learner accuracy, I performed a χ^2 test for independence for predicted positive outcomes. I predicted the total positive outcomes for each student by adding together the predictions of each of the 5 relevant Super Learners. The null hypothesis of the test is that the distribution of predicted positive outcomes cannot be distinguished from predictions from a null model. The alternative hypothesis is that predictions of academic outcomes by Super Learners are significantly different from predictions from a null model.

How can predictions of student academic outcomes be used as an early alert system?

An early alert system should identify students who would, without additional support, attain the fewest positive outcomes. To assess whether models in this study identify such students, I sorted students at each institution into two groups depending on whether their predicted total positive outcomes lay above or below the median. I then performed a *t*-test for difference of means of actual total positive outcomes among these groups. The null hypothesis is that mean positive outcomes among these groups are equal, and predicted positive outcomes are not useful in sorting students according to future academic attainment. The alternative hypothesis is that the means are not equal, and this study's models do identify students in greater need of academic support.

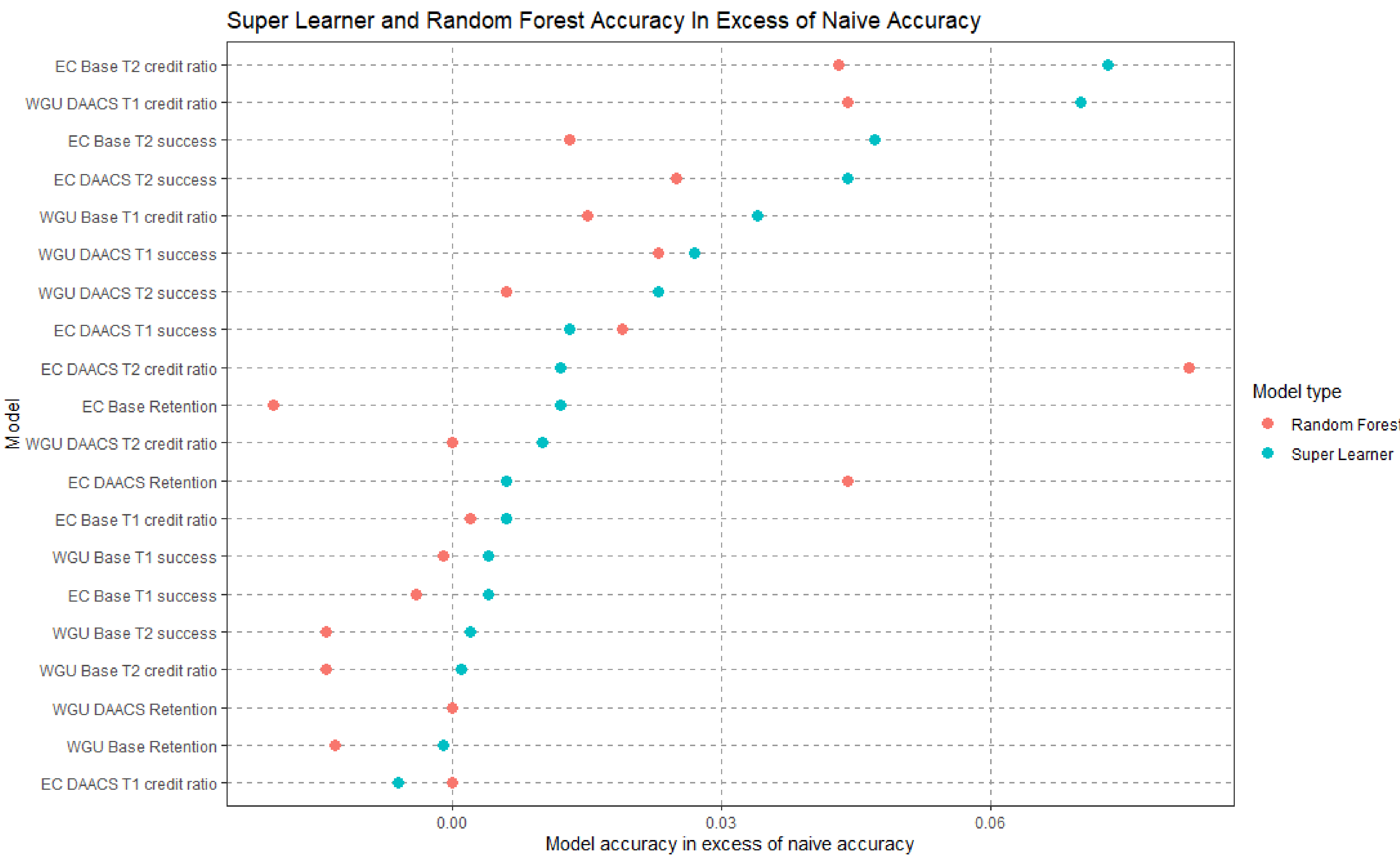
What variables are important in predicting student academic outcomes?

Variable importance was measured as mean decrease in accuracy in DAACS random forest models.

Results

Can a Super Learner ensemble method predict student academic outcomes with accuracy greater than that attained by random forest models alone? This table compares the accuracies attained by naïve, Super Learner, and random forest models. The final column shows the difference of Super Learner accuracy and random forest accuracy.

School	*Explanatory variables*	*Response variable*	*AUC*	*Naïve accuracy*	*SL accuracy*	*RF accuracy*	*Difference in favor of SL*
EC	Base	T1 success	0.636	0.797	0.801	0.793	0.008
EC	Base	T2 success	0.579	0.525	0.572	0.538	0.034
EC	Base	Retention	0.590	0.592	0.604	0.572	0.032
EC	Base	T1 credit ratio	0.634	0.813	0.819	0.815	0.004
EC	Base	T2 credit ratio	0.595	0.505	0.578	0.548	0.03
EC	DAACS	T1 success	0.558	0.777	0.790	0.796	-0.006
EC	DAACS	T2 success	0.583	0.567	0.611	0.592	0.019
EC	DAACS	Retention	0.563	0.650	0.656	0.694	-0.038
EC	DAACS	T1 credit ratio	0.511	0.828	0.822	0.828	-0.006
EC	DAACS	T2 credit ratio	0.552	0.580	0.592	0.662	-0.07
WGU	Base	T1 success	0.614	0.696	0.700	0.695	0.005
WGU	Base	T2 success	0.581	0.617	0.619	0.603	0.016
WGU	Base	Retention	0.570	0.761	0.760	0.748	0.012
WGU	Base	T1 credit ratio	0.622	0.569	0.603	0.584	0.019
WGU	Base	T2 credit ratio	0.591	0.700	0.701	0.686	0.015
WGU	DAACS	T1 success	0.697	0.702	0.729	0.725	0.004
WGU	DAACS	T2 success	0.662	0.622	0.645	0.628	0.017
WGU	DAACS	Retention	0.594	0.754	0.754	0.754	0
WGU	DAACS	T1 credit ratio	0.680	0.596	0.666	0.640	0.026
WGU	DAACS	T2 credit ratio	0.675	0.704	0.714	0.704	0.01


Super Learner and Random Forest Accuracy In Excess of Naive Accuracy
EC Base T2 credit ratio
WGU DAACS T1 credit ratio
EC Base T2 success
EC DAACS T2 success
WGU Base T1 credit ratio
WGU DAACS T1 success
WGU DAACS T2 success
EC DAACS T1 success
EC DAACS T2 credit ratio
EC Base Retention
WGU DAACS T2 credit ratio
EC DAACS Retention
EC Base T1 credit ratio
WGU Base T1 success
EC Base T1 success
WGU Base T2 success
WGU Base T2 credit ratio
WGU DAACS Retention
WGU Base Retention
EC DAACS T1 credit ratio
Model
0.00
0.03
0.06
Model accuracy in excess of naive accuracy
Model type
Random Forest
Super Learner

Are predictions of student academic outcomes based on Super Learners statistically significant? The tables below compare predicted positive outcomes compared to actual positive outcomes. For each table, a χ^2 test of independence is performed. Tables of expected values are not shown.

Predicted positive outcomes at WGU, DAACS models

Actual positive outcomes	*0*	*1*	*2*	*3*	*4*	*5*
0	1	43	30	79	14	3
1	0	44	41	115	24	2
2	0	16	9	79	29	3
3	0	29	35	150	40	12
4	0	7	9	93	37	7
5	1	5	18	167	84	26

$\chi^2 = 1271.83$, df = 25, $p < 0.001$

Predicted positive outcomes at EC, DAACS models

Actual positive outcomes	*0*	*1*	*2*	*3*	*4*	*5*
0	0	0	0	1	3	3
1	0	0	0	1	7	14
2	0	0	0	4	10	21
3	0	0	0	0	0	8
4	0	0	0	0	2	5
5	0	0	0	3	13	62

$\chi^2 = 184.62$, df = 25, $p < 0.001$

Predicted positive outcomes at WGU, base models

Actual positive outcomes	*0*	*1*	*2*	*3*	*4*	*5*
0	0	16	11	125	1	0
1	0	26	19	204	4	0
2	1	7	5	113	6	0
3	0	16	8	242	3	0
4	0	6	5	141	5	0
5	0	9	8	259	11	1

$\chi^2 = 3266.90$, df = 25, $p < 0.001$

Predicted positive outcomes at EC, base models

Actual positive outcomes	*0*	*1*	*2*	*3*	*4*	*5*
0	0	0	5	11	6	12
1	0	1	6	16	10	17
2	0	0	15	36	29	62
3	0	0	1	9	12	14
4	0	0	1	6	12	15
5	0	0	13	36	47	115

$\chi^2 = 697.67$, df = 25, $p < 0.001$

How can predictions of student academic outcomes be used as an early alert system? When students are sorted into groups based on predicted total positive outcomes, there is a statistically significant difference in the number of positive outcomes they actually attain. This difference is increased by the inclusion of DAACS assessment data in predictive models.

Excelsior College	*Predicted fewer positive outcomes*	*Predicted more positive outcomes*	*Difference*
Mean positive outcomes (base)	2.97 (n = 253)	3.47 (n = 254)	0.50 (p = 0.001)
Mean positive outcomes (DAACS)	3.09 (n = 78)	3.71 (n = 79)	0.62 (p = 0.027)

Western Governors University	*Predicted fewer positive outcomes*	*Predicted more positive outcomes*	*Difference*
Mean positive outcomes (base)	2.39 (n = 626)	3.02 (n = 626)	0.63 (p < 0.001)
Mean positive outcomes (DAACS)	2.23 (n = 626)	3.23 (n = 626)	1.00 (p < 0.001)

What variables are important in predicting student academic outcomes? The table below shows the top three variables by importance for each model. Because variable importance for Super Learners is not easily interpreted, these measures were obtained from random forest models.

School	*Explanatory variables*	*Response variable*	*Important variable*	*Mean decrease in accuracy*	*Important variable*	*Mean decrease in accuracy*	*Important variable*	*Mean decrease in accuracy*
EC	DAACS	T1 success	Initial transfer credits	22.77	Program division	12.8	SRL metacognition	8.46
EC	DAACS	T2 success	Initial transfer credits	12.51	Income	4.62	SRL grit	4.3
EC	DAACS	Retained	Initial transfer credits	13.28	Total reading score	4.8	Income	4.61
EC	DAACS	T1 credit ratio	SRL metacognition	11.73	SRL mastery orientation	10.49	Age	7.07
EC	DAACS	T2 credit ratio	Initial transfer credits	10.53	Veteran	5.8	Total reading score	3.24
WGU	DAACS	T1 success	Total reading score	28.15	Total math score	16.97	Ethnicity	12.82
WGU	DAACS	T2 success	Total reading score	18.13	Total math score	14.54	Total writing score	5.03
WGU	DAACS	Retained	Total reading score	13.85	SRL strategies	10.35	SRL anxiety	8.86
WGU	DAACS	T1 credit ratio	Total reading score	26.04	Total math score	16.62	Ethnicity	15.88
WGU	DAACS	T2 credit ratio	Total math score	16.44	Total reading score	12.27	SRL grit	8.12

Discussion

Can a Super Learner ensemble method predict student academic outcomes with accuracy greater than that attained by random forest models alone?

This study does not provide any evidence that Super Learner models are more accurate than random forest models when predicting student academic outcomes. There are many aspects of this study that could be varied to possibly yield a different result. Choosing different Super Learner component models; predicting different academic outcomes; selecting different explanatory variables; or examining data from a wider range of institutions could all potentially provide evidence of the superiority of Super Learners.

The possibility of a different result is qualified by the fact that the nearly equal performance of Super Learners and random forests in this study extended across multiple institutions, multiple response variables, and multiple sets of explanatory variables. The findings of Fernández-Delgado et al. further qualify this possibility. (Recall that Fernández-Delgado et al. found that when it comes to classification models, random forests are nearly optimal in a variety of contexts.) Despite this, there may be reasons other than improved accuracy to explore Super Learners for a classification problem: the ability to compare a variety of models simultaneously in a Super Learner can lend insight into underlying structure in the data. In a counterfactual scenario, Super Learners in this study might have given large weights to KNN models with $k = 10$. Even if accuracy did not exceed that of a random forest, the utility of the KNN model suggests underlying clustering in the data. Examining these clusters could provide theoretical insight into student performance.

Are predictions of student academic outcomes based on Super Learners statistically significant?

The results of χ^2 tests of independence on predictions for total positive outcomes showed that these predictions were statistically significantly different from those obtained by a null model. This significance extended across both institutions and both sets of explanatory variables. While improvements in accuracy for each individual Super Learner over naïve accuracy averaged 0.019, when Super Learners are used together to predict total positive outcomes, they show a significant difference from null models. Combining models for each outcome to predict a student’s total positive outcomes can form the basis of an early alert system.

How can predictions of student academic outcomes be used as an early alert system?

Predicted total positive outcomes for a student can be used to construct an early alert system. At both EC and WGU, using both sets of explanatory variables, students for whom predicted positive outcomes were in the bottom 50% did in fact attain fewer positive outcomes than students for whom predicted positive outcomes were in the top 50%. Using DAACS assessment data improved the model’s ability to differentiate between higher and lower performing students. At EC, the mean difference in total positive outcomes for groups constructed using base models was 0.50 ($p = 0.001$). This difference increased to 0.62 when DAACS assessment data were incorporated in the model. At WGU, the mean difference in total positive outcomes for groups constructed using base models was 0.63 ($p < 0.001$). This difference increased to 1.00 when DAACS assessment data were incorporated.

If an institution such as EC had capacity to offer enhanced advising and other academic and social support to only half its students, then the model shown here could help them direct this support toward students needing it most: namely, those who are on average less likely to attain positive academic outcomes. Incorporating DAACS assessment data into predictive models would allow them to target at-risk students even more effectively.

What variables are important in predicting student academic outcomes?

Variables of importance showed interesting differences across institutions. At EC, the most important variable in 4 out of 5 DAACS models was the student's number of transfer credits. DAACS assessment results also played an important role, representing six of the 15 most important variables. DAACS variables were even more important at WGU, where all but two of the most important variables were DAACS results. The DAACS results that were most important at EC were mostly related to self-regulated learning. At WGU, the reading and math assessments were most important, followed by some self-regulated learning components. At both institutions, demographic variables appeared only rarely among lists of the most important variables. At WGU, Ethnicity appears twice. At EC, Income appears twice.

These results suggest that models for predicting academic outcomes can vary greatly across institutions. This variation is a further recommendation for Super Learners, which automatically select component models and can be adapted to perform variable selection as well. (Though aspects of random forests perform similar functions.) Results also suggest that, while DAACS assessments provide important information about student success, the value or meaning of that information can vary from school to school. A general model for predicting academic outcomes that functions similarly across institutions remains elusive and would be an interesting area for further research.

General comments

While the models developed in this study do support the development of an early alert system for students at risk of negative academic outcomes, there are several caveats. First, there remains significant variation in positive academic outcomes within each group of students created by the early alert system. For example, some students in every group attained all five positive outcomes. Some students in each group attained none. For students with predicted outcomes near the middle of the distribution, predictive models were less accurate. There may be good reasons not captured by these predictive models for directing any particular student toward or away from additional support resources. In the event of a conflict between model outcome and stakeholders' preferences for support, the model outcome should be considered one piece of evidence among many.

Second, this study does not consider any costs that might accrue to students receiving additional support. If additional support such as enhanced advising is at worst useless, then there is no harm in offering it to some students who would have been successful without it. However, if additional support entails significant student costs—such as a fee, or a loss of opportunity to participate in other campus programming—then the early alert system described here should be implemented with caution.

In dividing students based on predicted positive outcomes, I've assumed that support resources are available for half of all students. The reality is that these resources, especially enhanced advising, may be far scarcer. Institutions may modify this early alert system to flag any range of prediction percentiles they choose. Choosing a smaller group of students will result in a greater mean difference between the targeted and untargeted groups.

Conclusion

When trained on the data in this study, Super Learners do not demonstrate greater accuracy than random forest models in predicting academic outcomes for undergraduates. Data from DAACS assessments combined with demographic data can form the basis of a model that predicts academic outcomes. This model can be used to direct scarce academic support resources to students who need them the most. However, decisions about directing students to additional intervention should consider model results together with exogenous factors. Models for predicting academic outcomes for undergraduates varied significantly across institutions. This suggests that a general model for predicting academic outcomes may require data other than that considered here.

References

Bailey, T., Bashford, J., Boatman, A., Squires, J., Weiss, M., Doyle, W., Valentine, J. C., LaSota, R., Polanin, J. R., Spinney, E., Wilson, W., Yeide, M., & Young, S. H. (2016). *Strategies for postsecondary students in developmental education—a practice guide for college and university administrators, advisors, and faculty*. Institute of Education Sciences, What Works Clearinghouse (ERIC Document Reproduction Service No. ED570881).

Bailey, T. and Cho, S. (2010). *Issue Brief: Developmental Education in Community Colleges*. Community College Research Center, Columbia University.

Belfield, C. R. and Crosta, P. M. (2012). *Predicting success in college: the importance of placement tests and high school transcripts* (CCRC Working Paper No. 42). Community College Research Center, Columbia University.

Bryer, J., Akhmedjanova, D., Andrade, H., and Lui, A. (2021). The use of predictive modeling for assessing college readiness. In *Enhancing effective instruction and learning using assessment data* (pp. 83-107). Information Age Publishing.

Chingos, M. M. (2018). *What matters most for college completion? Academic preparation is a key of success*. American Enterprise Institute. https://www.thirdway.org/report/what-matters-most-for-college-completion-academic-preparation-is-a-key-of-success

Community College Research Center, Columbia University. (2015, July). *Improving the Accuracy of Remedial Placement*. https://ccrc.tc.columbia.edu/media/k2/attachments/improving-accuracy-remedial-placement.pdf

Credé, M., & Niehorster, S. (2012). Adjustment to college as measured by the student adaptation to college questionnaire: A quantitative review of its structure and relationships with correlates and consequences. *Educational Psychology Review*, 24(1), 133-165.

Daugherty, T. K., & Lane, E. J. (1999). A longitudinal study of academic and social predictors of college attrition. *Social Behavior & Personality: an international journal*, 27(4).

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. *The journal of machine learning research*, 15(1), 3133-3181.

Lynch-Holmes, K. B., Troy, A. B., & Ramos, I. (n.d.). Early alert & intervention: Top practices for retention. Retrieved from https://issuu.com/connectedu/docs/early_alert_white_paper

Leonhardt, D. and Chinoy, S. (2019, May 23). The college drop-out crisis. *The New York Times*. Retrieved from http://www.nytimes.com

Mesmin, D. (2018). *Leveraging psychological factors: a necessary component to improving student outcomes*. American Enterprise Institute. https://www.thirdway.org/report/leveraging-psychological-factors-a-necessary-component-to-improving-student-outcomes

Mokher, C. G., Barnett, E., Leeds, D. M., & Harris, J. C. (2019). Re-envisioning college readiness reforms: Florida's statewide initiative and promising practices in other states. *Change: The Magazine of Higher Learning*, 51(2), 14-23.

Polley, E. C. and van der Laan, M. J. (2010). *Super Learner in prediction*. U. C. Berkeley Division of Biostatistics Working Paper Series, Paper 266. http://biostats.bepress.com/ucbbiostat/paper266

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* (CCRC Working Paper No. 41). Community College Research Center, Columbia University. (ERIC Document Reproduction Service No. ED529866).

Scrivener, S., Weiss, M. J., & Teres, J. J. (2009). More guidance, better results? Three-year effects of an enhanced student services program at two community colleges. New York, NY: MDRC.