Case Study:

Machine Learning to Support Success & Combat Inequity

ASA research

Know the facts. Act with insight.

Summary

ASA Research, LLC, has developed a model that predicts next-term retention at a large four-year public university approximately three to five times better than alternatives. The model leverages 80 variables from a single university's student-level administrative data and 33 variables from public data, captured over 12 years, represented by the 2010 through 2021 student cohorts. Like all machine learning models, ours mines historical data for mathematical relationships between predictor variables and the outcome of interest, then "learns" to generalize these relationships so that it can predict future outcomes when only the predictor variables are known.

While developing the model, our team worked closely with the university's undergraduate education division to prioritize transparency and equity. To ensure that the model combats existing inequities instead of reinforcing them, we included a range of voices when making key decisions, including students' perspectives, and we carefully tested our model for algorithmic bias.

Once machine learning algorithms predicted which fall cohort students were at highest risk of not persisting into the winter term, our team worked with advisors in undergraduate education to provide on-time interventions to students identified in the model, with an emphasis on equity and not perpetuating bias. We provided 385 students in the intervention group with early advising from the advising team.

Estimated Results[i]

Figure 1: Observed fall-to-winter retention results with predictive analytics, compared to estimated results without.

Retention Rate:	Student Body		Intervention Students		Difference
Observed (With Predictive Analytics)	**96%**	-	**90%**	=	**6pp**
Estimated (Traditional Alternative)	**96%**	-	**86%**	=	**10pp**

*Estimated **4 percentage-point** difference in differences with predictive analytics.*

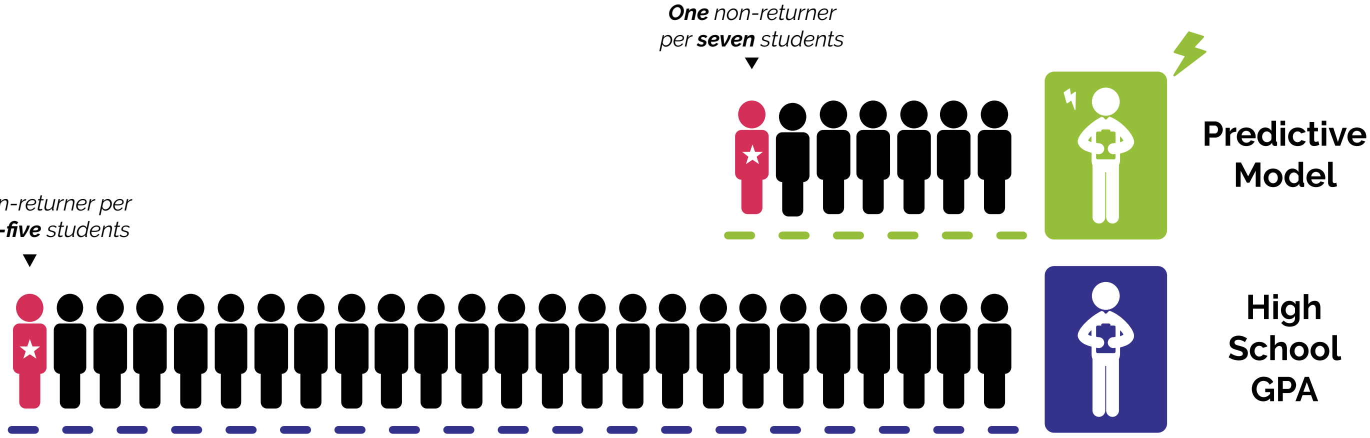
This estimated 4.6 percentage point difference is conservatively equated to a $300,000 annual cost savings of in-state tuition that otherwise would have been lost due to students dropping out of the institution. This is, of course, coupled with the intrinsic academic benefits of students being retained at the institution and the regional economic benefits of a trained labor force.

Model Development & Testing

Internally, the model is built on the XGBoost machine learning framework, which seeks to produce accurate and generalizable predictions by combining many simple models into an "ensemble" model designed to perform better than any one model could on its own. To evaluate the model's performance, we compared its predicted outcomes against actual outcomes within the 2021 student cohort, which was withheld from the model during development.[ii] Among the 2021 cohort, the model identified 35% of all non-returners. Focusing specifically on potentially vulnerable students – defined here as first-generation students and those belonging to traditionally underserved races and ethnicities – the model identifies 45%.

For comparison, a more traditional approach based directly on high school GPA would identify only 11% of non-returners, and a random lottery would identify only 10%. By identifying students using our model instead of identifying students with low high school GPAs, results suggest that advisors equipped with our model could proactively reach out to 3.2 times as large a share of non-returners, and 3.4 times as large a share of potentially vulnerable non-returners.

Figure 2: Illustration of model test performance with 2021 validation cohort, as compared to high school GPA alternative.



Equity & Ethics

ASA placed equity and fairness at the core of our machine learning process. Before beginning this project, we pledged to put the final model into service only if it met our standards for fairness and demonstrably resisted existing biases, rather than cementing them. Moreover, we pledged to do this in consultation with stakeholders like the undergraduate education division, data ethics scholars, and student leaders.

Through a participatory process, we determined that it would be acceptable for the model to predict non-retention at higher rates for potentially vulnerable groups than for their less-vulnerable counterparts. In a contrasting example, evaluators might forbid a model used for bank loan decisions from issuing approvals at a higher rate for white applicants than for applicants of color. However, since our model aims to allocate a supportive resource where it is needed most, and we know that need can relate to traits like race due to existing systemic inequities, it was deemed sensible and unproblematic for our model to allocate early advising at higher rates for some groups than for others, including along sensitive lines such as race and gender.[iii]

Further, ASA and stakeholder groups determined that it would be unacceptable for the model to serve potentially vulnerable groups with a substantially lower level of predictive accuracy than it did their less-vulnerable counterparts. Having arrived at an understanding of equity in the context of this model, we developed a set of concrete criteria against which the model would be evaluated to test whether it aligned or did not align with our understanding of equity. The model was put into service only after passing these tests.

Conclusion

Consulting with a range of stakeholders at a large university, ASA developed a machine learning model that performs up to five times as well as alternatives. We then co-created requirements for equity and fairness and tested the model against them. This initiative can serve as a blueprint for responsible use of machine learning to promote student success and advance equity. Our model is customizable for institutions that have the data and aligned senior leadership to implement AI solutions for predicting student success outcomes and providing more resources to those students who need it most.

i As this intervention has not yet been subject to formal evaluation, estimated retention without the intervention is drawn from simulations based on historical data. Performance comparisons should be taken only as illustrations based on best-estimate scenarios.

ii Assessments of performance and equity are based on the model's handling of 2021 data, which it had never encountered before. However, at the conclusion of the modeling process, the model was re-trained with 2021 data included so that it could make the best possible predictions for the incoming 2022 cohort. This is standard practice in machine learning.

iii Under the same rationale, we chose to supply the model with race and other sensitive demographics, rather than withhold them. This choice is especially sensible given the well-documented ability of complex models to infer sensitive attributes from other data points, such as ZIP code.