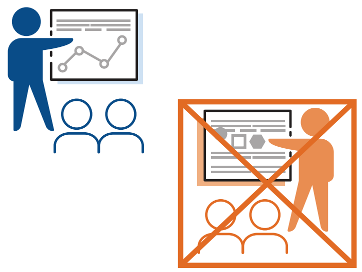
Predicting student enrollment: Can machine learning facilitate effective school district planning?

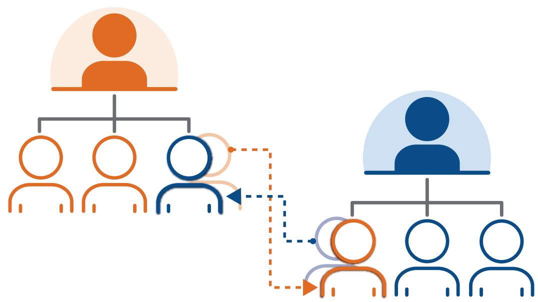
Fluctuations in school enrollment are a challenge for the School District of Philadelphia (SDP) and other school districts with substantial student mobility, school choice options, or both. Inaccurate enrollment predictions complicate planning efforts and attempts to allocate teachers to schools and classrooms. SDP uses data from February of each school year to forecast enrollments for each school and grade in the following school year and allocates teachers based on this forecast. The following fall, SDP reallocates teachers based on actual student enrollment, which can result in eliminating classrooms, reassigning students to different teachers, or reassigning teachers to a different grade level or school.

Possible results of reallocation

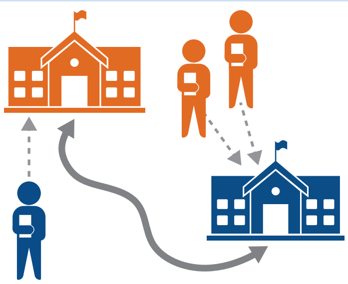
Some classrooms may be eliminated



Students may be assigned to different teachers



Teachers may be re-assigned to a different grade level or school



What if districts had a better way to predict student enrollment to minimize these disruptions?

A recent REL report **looked at whether three different machine learning methods can outperform a simple regression model in making enrollment predictions.** Regression models are widely used by large districts and can be implemented without advanced statistical software. Machine learning techniques generally require more sophisticated software, and analytic capabilities.

We partnered with SDP to examine the accuracy of the four methods using data from previous student cohorts.[1,2] The analysis used data from the 2016–2017 and 2017–2018 school years to develop the models and then tested their accuracy in predicting student enrollment using data from the 2018–2019 school year. Data provided by SDP included predictive data that most large school districts collect, such as student-level attendance, suspensions, demographics, and test scores. The study measured accuracy based on (1) the difference in the number of students predicted to enroll versus the actual number of students who enrolled and (2) the percentage of students in a cohort who would need to be reallocated to different classrooms after the start of the school year.

With the types of data currently available to the district, there was no advantage to using sophisticated machine learning algorithms over a simple regression model to predict fall enrollment.

All four methods performed with similar accuracy. All four predictions differed from the actual fall cohort sizes (typically around 60 students) by about 6 students, on average.

Each method produced results that would lead to a need to reallocate 20 to 30 percent of students to different teachers or classrooms in October.

The methods performed similarly across schools with larger proportions of Black students, economically disadvantaged students, and English learners.

What are the key takeaways for school districts with high rates of mobility or school choice options?

In districts with high rates of mobility, using conventional administrative data from the preceding February may not predict fall enrollments with the accuracy needed to prevent substantial disruptions. These districts might do better by gathering additional data later in the spring and early summer to improve predictions, regardless of which method they use.

Districts such as SDP may implement any of the four methods and produce similar predictions. However, regression models are easier and more cost-effective for districts to implement.

Endnotes

[1] The three machine learning algorithms in the study were least absolute shrinkage and selection operator (LASSO), elastic net, and random forest. The simple regression model used in the study was an ordinary least squares (OLS) model.

[2] Tanner, S., Terrell, J., Vislosky, E., Gellar, J., & Gill, B. (2021). *Predicting early fall student attrition in the School District of Philadelphia* (REL 2021–124). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4648.

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006, with REL Mid-Atlantic, administered by Mathematica. The content of the infographic does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.