Education Advisory Board | Student Success Collaborative

Predicting Student Risk Through Advanced Analytics

Methodology

The Student Success Collaborative data platform predicts individual students' likelihood to graduate in a wide range of majors on each campus. Predictions are generated by comparing key academic factors for each student against historical patterns of success and failure. To develop these algorithms, data on historical student cohorts is collected from member information systems and used to generate an initial series of candidate academic metrics. Advanced multivariate statistical and machine learning techniques perform variable selection and conduct hypothesis testing. Results are cross-validated for accuracy and compared with known research on student success outcomes.

We start with a student's courses and grades…

…search historical records for academically similar students…

…and predict their likelihood to graduate in a given major

Student Data Inputs

- Courses attempted
- Course grades earned
- Overall credits attempts
- Overall credits completed
- GPA trend over time
- Difficulty of major
- Pre-College Academic Data
- Demographic Data

Predictive model incorporates articulated transfer courses

Predictive Outputs

Risk in current major

Risk in alternate majors

High
Medium
Low
Difficulty of upcoming courses

Model Accuracy

Regressions testing the fit of our predictions against actual graduation outcomes generate an R-squared of 0.94, confirming a high degree of model accuracy (max $R^2 = 1.0$)

Why Do We Focus on Transcript Data?

Academic data from transcripts offers several key advantages over other data types:

1. **Strong Predictive Power** Demonstrated to be more predictive than high school, demographic, or financial variables
2. **Comprehensive Coverage** Data is obtainable all students with no extra effort on the part of staff or instructors
3. **Robust Historical Records** Most institutions can provide at least a decade of electronic transcript records
4. **Highly Actionable Insight** Students can improve predictions by making targeted changes to their academic performance

What Other Data Could We Add to Further Improve Model Precision?

While transcripts will continue to be central to our algorithms, we are continuously working with Collaborative members to refine and improve our models through the evaluation and incorporation of additional types of student information.

Types of Data for Further Evaluation

- Financial Data
- LMS Gradebook Data
- Non-articulated Transfer Data
- Student Engagement

2445 M Street NW ▪ Washington DC 20037 ▪ P 202-266-6400 ▪ F 202-266-5700