Student Success Predictive Model Report – South Dakota State University

September 29, 2020

Introduction

Overview

This document provides information about your institution's custom Student Success Predictive Model (SSPM). It describes how the model was built; details the top success indicators or "predictors" used in the model and provides metrics characterizing the predictive power of the model.

The SSPM uses your school's student records to predict the likelihood that any chosen student **will persist to the next fall (or graduate before then)**. This is done by first "training" a statistical model using the records of historical students in order to determine—and assign values to—the items derived from those records that are "predictors" of persistence outcomes.

The model outputs a success score between zero and one estimating the probability that a selected student will persist to the next fall. That is, each student's success score corresponds to the model's estimate of their likelihood of persisting to the next fall. Since it is not possible to build a perfectly prescient model, it is important to state that a score of one does not guarantee a student's persistence. Nor does a score of zero guarantee their failure. A success score of 0.7 for instance, may be interpreted as our expectation that, on average, seven of ten students with this score will persist to the next fall.

Methodology

The SSPM uses the latest advances in data science to estimate persistence likelihood for each student, from incoming freshmen to nearly-graduating seniors. A customized set of predictors are constructed from student records, and then combined and weighted using an automated training process. EAB's Data Science team customizes this process for each partner, and uses a variety of optimization tools to ensure the best possible performance given the data available.

As described above, the model is trained from recent historical student records; in particular, students satisfying the following criteria were used:

- Matriculated between 2010-08-30 and 2018-08-20.

- Had at least one registered term.

- Were seeking a degree.

Technical details: The model is a combination of several penalized logistic regression models applied to different subgroups of students. The predictors include simple lookups of student

records (e.g., high school GPA), as well as composite attributes derived from them whose details are proprietary.

Executive summary

EAB has built a customized Student Success Predictive Model (SSPM) for your institution that predicts the persistence likelihood of your students. Your SSPM incorporates the latest breakthroughs in statistics and data science, placing your institution at the cutting edge of student-insight technology. It is a powerful tool for promoting your students because it gives you invaluable insight into their likelihood of academic success. This document provides an overview of the SSPM, describes how it was built and extensively customized and optimized for you, and details benchmarks of its predictive performance.

Your Institution's Model

The SSPM includes a wide variety of success indicators called "predictors" in order to ensure maximal predictive power. We use your institution's historical data to determine the best set of predictors that most accurately reflects the underlying patterns of your students. The items below were found to be good predictors for your institution. The predictors in these lists are not equally important and may not be the same for all subgroups; the statistical model learns how to identify and assign values to the best predictors for each subgroup of students. For instance, we might expect high school GPA to be highly relevant for freshmen, but minimally important for seniors.

Your Predictors

The lists below describes the predictors for each subgroup of students in the model. We are sharing these predictors to help you understand how the model works and the types of variables that are predictive of success at different points in a student's academic career. Knowing these variables can help build understanding of the model and may provide insight into where to start conversations with different groups of students. However, multivariate machine learning models are very complicated, and sometimes unintuitive, so the individual variables should be interpreted cautiously. The SSPM is designed to maximize predictive accuracy, not to maximize our understand the impact of any individual input variable. This is not the same as a controlled study on the influence of these variables, and the inclusion of any variable on this list does not imply a causal effect. Many variables are highly correlated with one another, and therefore "High Impact Predictors" may change from one model iteration to another even if the training data and model structure are similar. Therefore, EAB does not recommend using this list to drive specific actions around individual variables except when used in conjunction with external studies on causality and, of course, the human intelligence of subject matter experts.

For each sub-model, the predictors are organized into two sections: "High Impact Predictors" and "Other Predictors." *"High Impact Predictors" are the predictors that are responsible for more than 5% of the variance in scores across all of the students in a credit bin. This may mean that the variable has a moderate impact on the scores of many students, or a high impact on the scores of just a few students.* Some variables have a significant impact on the students they affect, but affect only a low number of students and therefore do not count as "High Impact Predictors." Just because a variable is or is not a "High Impact Predictors" for the

population does not mean that it is or is not an important factor for an individual student. The "Other Predictors" are variables that the model identified as statistically significant predictors, but are responsible for less than 5% of the variance in scores across the population. They may be mildly predictive for many students or highly predictive for a very low number of students. Although the "Other Predictors" are individually weak predictors, collectively they are responsible for a significant portion of model performance.

In the following section, we enumerate the "High Impact Predictors" for each sub-model. To see the full list of "High Impact" and "Other" Predictors for each sub-model, please refer to Appendix near the end of the report.

High Impact Predictors by Credit Range Sub-Model

- **Pre-Enrollment Students High Impact Predictors**

  High School Percentile

  High School GPA

  Admit Code

- **Day 1 Students High Impact Predictors**

  High School Percentile

  High School GPA

  Admit Code

  In State Resident Indicator

- **Students with Between 1-60 Accumulated Credits High Impact Predictors**

  Cumulative GPA

  Admit Code

  Number of Completed Terms

  First Term Transfer Credits

  Ratio of Earned to Attempted Credits

- **Students with Between 61-120 Accumulated Credits High Impact Predictors**

  Admit Code

  Ratio of Earned to Attempted Credits

  High School Percentile

  Credits Attempted Current Term

  Cumulative GPA

Proportion of Transfer Credits

Trend in Term GPA

- **Students with More Than 120 Accumulated Credits High Impact Predictors**

Admit Code

High School Percentile

SAT/ACT Verbal Score Percentile

High School GPA

Ratio of Credits Attempted Current Term to Prior Term

Ratio of Earned to Attempted Credits

Model Performance

Your SSPM is well-calibrated and its performance has been thoroughly characterized using a "test set" of your historical students that was set aside from the training set NA Blind Campaign model that randomly targets students. This section describes the most insightful performance benchmarks and compares your SSPM to these other notional models.

The primary metric EAB uses to benchmark model performance is high-risk student identification rate. It is based on the most common use case for the model: that you are designing a campaign targeting high-risk students but only have the capacity to advise a limited subset of your total student population. In this case, your goal is to efficiently use your constrained resources to reach as many of your school's actual high-risk students as possible.

Lift

We may divide the percentage of actual high-risk students identified by a given model by the percentage found by a Blind Campaign to create a new metric called "lift". For instance, a lift of two would mean that a campaign based on your SSPM identified twice as many high-risk students as a Blind Campaign, while a lift value less than one would indicate that your model identified fewer actual high-risk students than simply choosing from your student population at random. Considering a campaign that includes 25% of your total student population, lift is 5.68, 2.16, 1.51, and 1.00 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

High-Risk Student Identification Rate for Murky Middle and Top Performing Students

Your Student Success Predictive Model's performance varies across different subgroups of students. This section provides plots and tables evaluating model performance in terms of high-risk student identification rate for two student subgroups: Murky Middle and Top Performing. The same plots provided for the overall student population in the main body are shown in this section for two student subgroups.
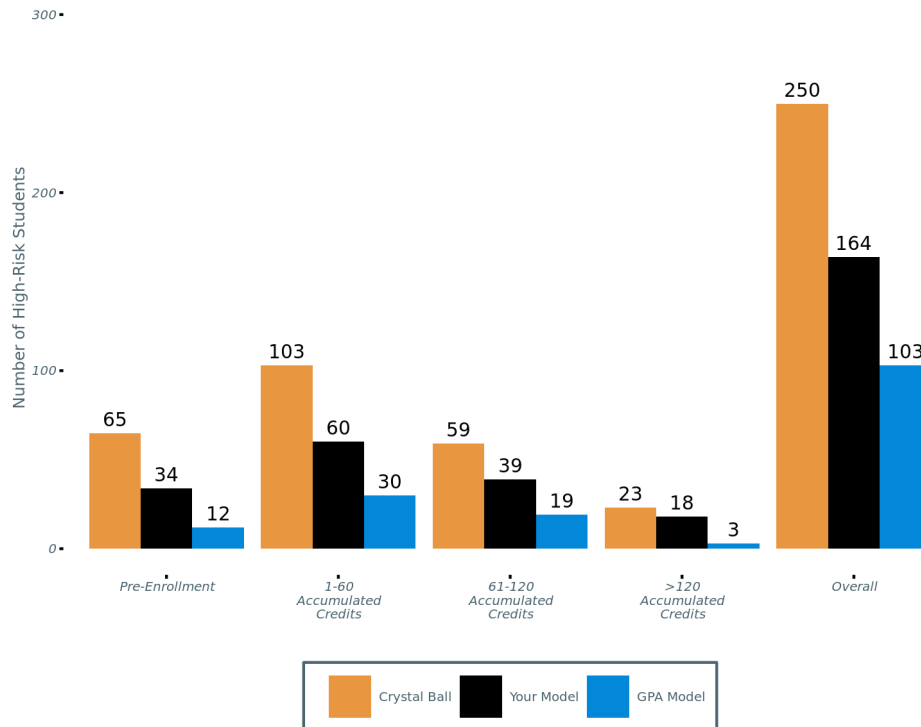
Murky Middle

Murky Middle students are defined as those students whose cumulative GPAs are between 2.0 and 3.0.
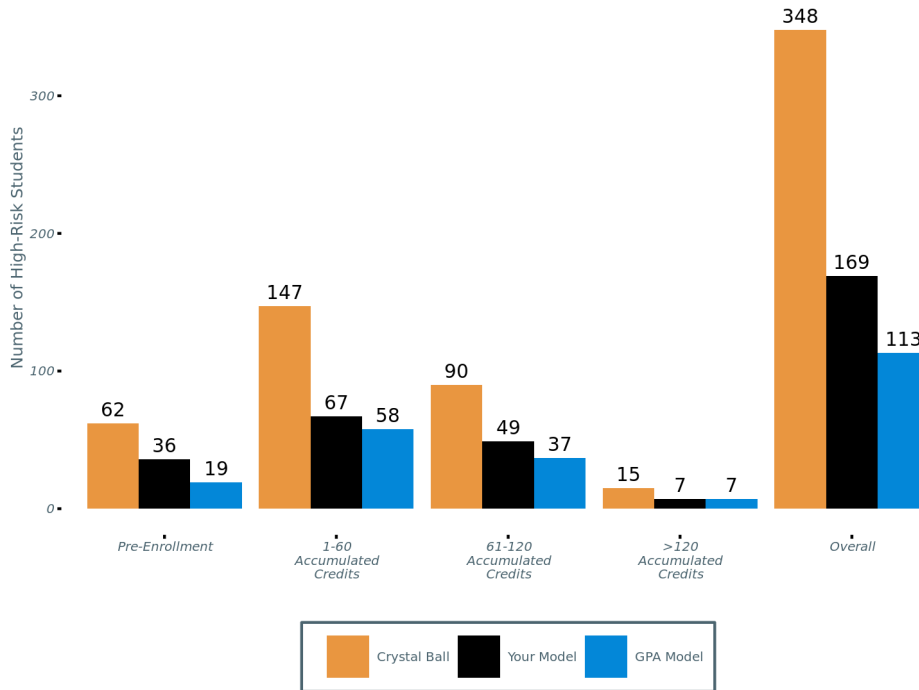
Top performing students

Top performing students are defined as those students whose cumulative GPAs are greater than 3.

Number of High-Risk Students

| Pre-Enrollment | 1-60 Accumulated Credits | 61-120 Accumulated Credits | >120 Accumulated Credits | Overall |

Crystal Ball: 65, 103, 59, 23, 250
Your Model: 34, 60, 39, 18, 164
GPA Model: 12, 30, 19, 3, 103

Legend: Crystal Ball | Your Model | GPA Model

Evaluating AUC

We commonly use AUC to measure and tune the performance of your Student Success Predictive Model across your institution's entire student population and different subgroups. AUC stands for Area Under the Curve and is a measure used extensively in data science, which ranges from 0.5 (pure chance) to 1.0 (Crystal Ball). We evaluate your SSPM's AUC in comparison to the notional GPA Model; your SSPM's larger AUC indicates that it identifies high-risk students more accurately than the GPA Model. This is the type of rule-of-thumb based approach that academic advisors intuitively know is useful.

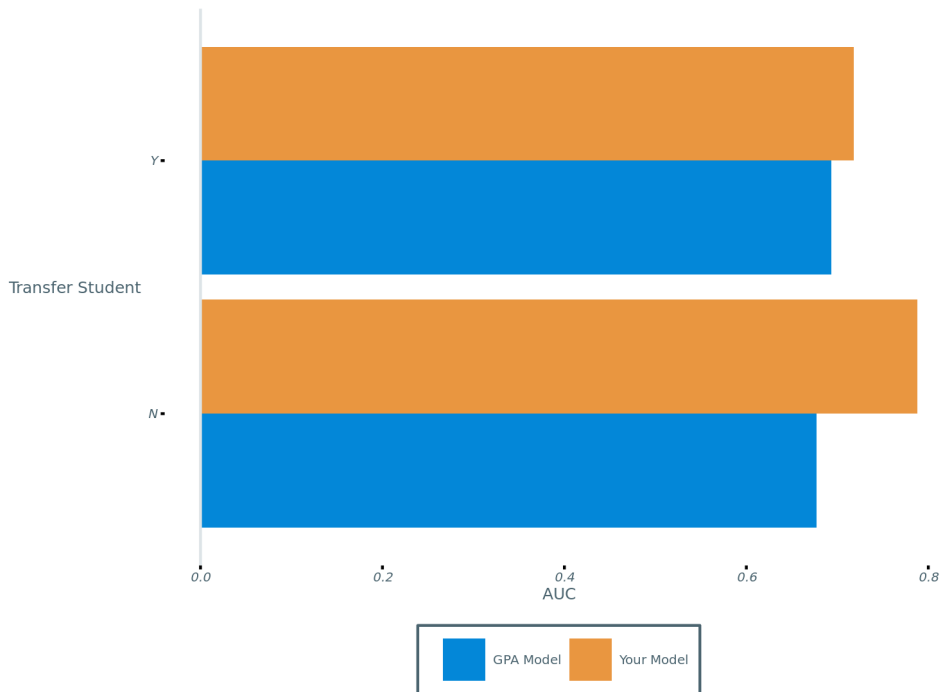The table below shows AUC values for your SSPM and the GPA Model.

Model	AUC
GPA Model	0.68
Your Model	0.78

As part of validating your SSPM, we examine subgroups of students to ensure that it consistently performs. The figures below show the AUC values for students with different levels of accumulated credits and for Transfer/Non-Transfer students.

AUC for Students with Different Numbers of Accumulated Credits

AUC vs Accumulated Credits (Pre-Enrollment, 0, 60, 120, 121+) for GPA Model and Your Model.
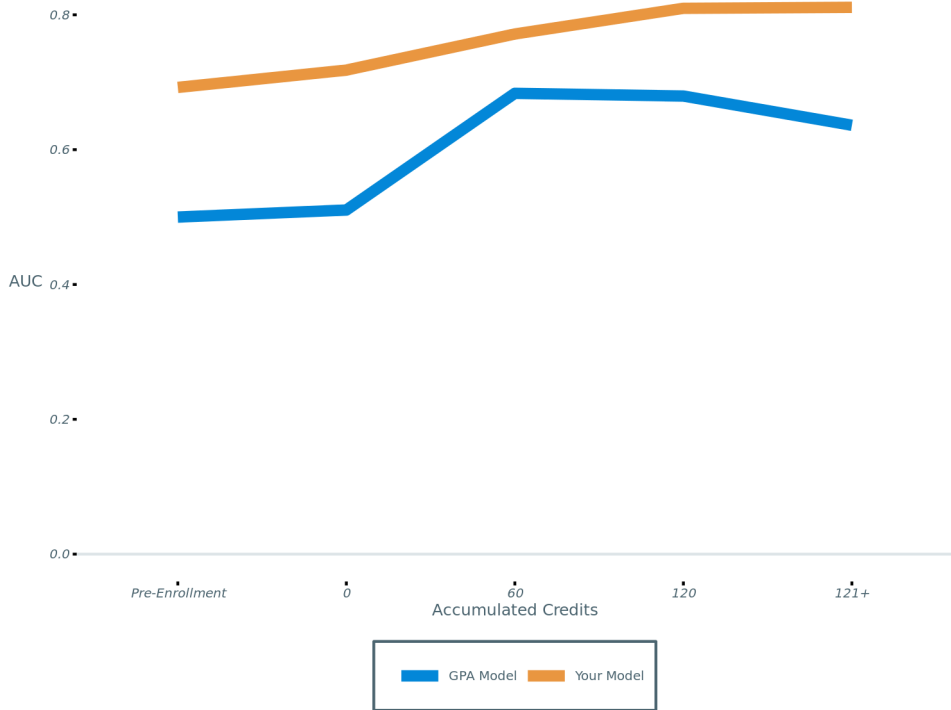
Accuracy for Transfer/Non-Transfer Students

AUC for Transfer Student (Y/N) for GPA Model and Your Model.

Conclusion

The performance of your institution's Student Success Predictive Model has been extensively optimized and evaluated; the model will provide your school and its advisors with invaluable and otherwise unobtainable insight into your students' likelihood of academic success. The model incorporates the latest breakthroughs in statistics and data science and places your institution at the cutting edge of student-insight technology. Your advisors may use it with confidence to both assess individual students and design effective and efficient targeted campaigns.

Appendix – Additional Predictor Information

Predictor Descriptions

The list below provides detailed descriptions of all the predictors used in your model. We discussed the most important among these in the "Your Predictors" section of the report. This list is ordered alphabetically.

- A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.: A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major. This percentile score ranks students in comparison to the performance of their peers' in the same major; e.g., a sociology student with a score of 80 has a higher cumulative GPA than 80% of all students declared in the sociology major. Students declared in multiple majors are assigned a percentile value that corresponds to the mean average of their scores for each major.

- Admit Code: A student's admission type (i.e., first time freshman, first time transfer, conditional admit, etc.)

- Age at First Term: A student's age upon starting their first term at your institution.

- Average Credits Attempted per Term: The average number of credits a student has attempted per term.

- Average Success Outcome of Students Declared in Same Major: This score indicates the average success outcome of all students enrolled in a given student's chosen major. E.g., if the model's success outcome is whether a student eventually graduates, and 90% of chemistry students do, then the score will be 90% for all chemistry students. Students declared in multiple majors, however, are assigned the mean average score across all of their majors.

- Credits Attempted Current Term: The number of credits a student is attempting in the current regular term. (The number of credits a student attempted in the most recent regular term is used in the case that a regular term is not currently in session.)

- Cumulative GPA: A student's cumulative GPA.

- Estimated Skills: A student's estimated academic skills. More specifically, we identify underlying patterns in the grades students earn in different courses – e.g., some students may have a history of excelling in math-related courses but not writing-related courses – and call the discrete factors behind these patterns "skills".

- First Term Transfer Credits: The number of credits a student transferred from other institutions upon matriculation.

- Gender: A student's gender.

- High School GPA: A student's high school GPA.

- High School Percentile: A student's high school rank in terms of percentile.

- High School Size: The size of an individual's high school student body.

- In State Resident Indicator: A "Yes" or "No" indicator of whether a student is a resident of your institution's home state.

- International Indicator: "Yes" or "No" indicator of whether an individual is an international student.

- Number of Completed Terms: The number of terms a student has completed at your institution.

- Proportion of Transfer Credits: The proportion of a student's credits that were earned at another institution.

- Race/Ethnicity: A student's race/ethnicity.

- Ratio of Credits Attempted Current Term to Prior Term: The number of credits a student attempted in the current regular term as to compared to the number of credits they attempted in the prior regular term. (The most recent regular term and the one prior to it are used in the ratio in the case that a regular term is not currently in session.)

- Ratio of Earned to Attempted Credits: The overall number of credits a student has earned divided by the number of credits they have attempted.

- Recent Change in GPA: The difference in a student's GPA from the prior two complete terms

- SAT/ACT Math Score Percentile: A student's highest percentile achieved in either the SAT or ACT math test. We calculate a student's math percentile as the highest percentile they earned in either the SAT or ACT math tests. A percentile score ranks students in comparison to their peers' performance; e.g., a percentile score of 80 indicates that a student outperformed 80% of his peers in either the SAT or ACT math tests.

- SAT/ACT Verbal Score Percentile: A student's highest percentile achieved in either the SAT or ACT verbal test. We calculate a student's verbal percentile as the highest percentile they earned in either the SAT or ACT verbal tests. A percentile score ranks

students in comparison to their peers' performance; e.g., a percentile score of 80 indicates that a student outperformed 80% of his peers in either the SAT or ACT verbal tests.

- Transfer Indicator: "Yes" or "No" indicator of whether the student transferred from another institution.

- Trend in Term GPA: A measure of the change over time in a student's term GPAs.

All Predictors

The list below enumerates all predictors used in each submodel, including "Other Predictors" that were not important enough to be included in the "Your Predictors" section of the report.

- **Pre-Enrollment Students High Impact Predictors**

   High School Percentile

   High School GPA

   Admit Code

- **Pre-Enrollment Students Other Predictors**

   Race/Ethnicity

   Transfer Indicator

   Gender

   In State Resident Indicator

   International Indicator

   High School Size

- **Day 1 Students High Impact Predictors**

   High School Percentile

   High School GPA

   Admit Code

   In State Resident Indicator

- **Day 1 Students Other Predictors**

   SAT/ACT Math Score Percentile

   Average Success Outcome of Students Declared in Same Major

   Age at First Term

   Race/Ethnicity

Ratio of Earned to Attempted Credits

Gender

Cumulative GPA

A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.

Estimated Skills

Ratio of Credits Attempted Current Term to Prior Term

High School Size

Average Credits Attempted per Term

SAT/ACT Verbal Score Percentile

Credits Attempted Current Term

International Indicator

Number of Completed Terms

- **Students with Between 1-60 Accumulated Credits High Impact Predictors**

Cumulative GPA

Admit Code

Number of Completed Terms

First Term Transfer Credits

Ratio of Earned to Attempted Credits

- **Students with Between 1-60 Accumulated Credits Other Predictors**

High School Percentile

In State Resident Indicator

Proportion of Transfer Credits

Credits Attempted Current Term

High School GPA

A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.

Gender

Transfer Indicator

Average Credits Attempted per Term

Ratio of Credits Attempted Current Term to Prior Term

Average Success Outcome of Students Declared in Same Major

SAT/ACT Math Score Percentile

Trend in Term GPA

International Indicator

Recent Change in GPA

SAT/ACT Verbal Score Percentile

High School Size

Age at First Term

Estimated Skills

Race/Ethnicity

- **Students with Between 61-120 Accumulated Credits High Impact Predictors**

Admit Code

Ratio of Earned to Attempted Credits

High School Percentile

Credits Attempted Current Term

Cumulative GPA

Proportion of Transfer Credits

Trend in Term GPA

- **Students with Between 61-120 Accumulated Credits Other Predictors**

Ratio of Credits Attempted Current Term to Prior Term

Recent Change in GPA

Age at First Term

Transfer Indicator

SAT/ACT Verbal Score Percentile

Gender

Estimated Skills

Average Credits Attempted per Term

Race/Ethnicity

International Indicator

First Term Transfer Credits

High School Size

A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.

High School GPA

In State Resident Indicator

SAT/ACT Math Score Percentile

Number of Completed Terms

Average Success Outcome of Students Declared in Same Major

- **Students with More Than 120 Accumulated Credits High Impact Predictors**

  Admit Code

  High School Percentile

  SAT/ACT Verbal Score Percentile

  High School GPA

  Ratio of Credits Attempted Current Term to Prior Term

  Ratio of Earned to Attempted Credits

- **Students with More Than 120 Accumulated Credits Other Predictors**

  High School Size

  In State Resident Indicator

  Recent Change in GPA

  Credits Attempted Current Term

  Average Credits Attempted per Term

  Transfer Indicator

  Age at First Term

  Average Success Outcome of Students Declared in Same Major

  Trend in Term GPA

  Race/Ethnicity

A student's cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.

First Term Transfer Credits

Number of Completed Terms

Estimated Skills

Proportion of Transfer Credits

International Indicator

Cumulative GPA

SAT/ACT Math Score Percentile

Gender