2445 M Street NW | Washington DC 20037
P 202.266.5600 | F 202.266.5700 | eab.com

EXHIBIT B

Student Success Collaborative

Student Success Predictive Model Report – University of Houston

Executive summary

EAB has built a customized Student Success Predictive Model (SSPM) for your institution that predicts the graduation 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.

Performance Summary

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.

Methodology

The SSPM uses the latest advances in data science to estimate graduation 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 2007-08-20 and 2012-08-27.
- Had at least one registered term.
- Were seeking a degree.

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.

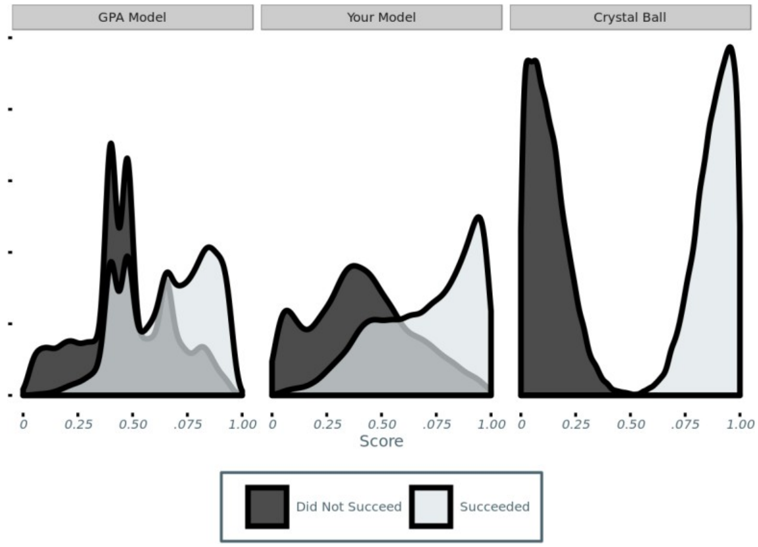
In the following section, we enumerate the "High Impact Predictors" for each sub-model. To see the full list of "High Impact" and "Other" Predictors for each sub-model, please refer to Appendix III near the end of the report.

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.

2445 M Street NW | Washington DC 20037
P 202.266.5600 | F 202.266.5700 | eab.com

2445 M Street NW | Washington DC 20037
P 202.266.5600 | F 202.266.5700 | eab.com



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.

2445 M Street NW | Washington DC 20037
P 202.266.5600 | F 202.266.5700 | eab.com

EXHIBIT C
Navigate FOIA Data Request

Population: University of Houston students in EAB Navigate; enrolled Fall 2020; undergraduates classified as Freshman, Sophmore, Junior, or Senior

Risk Level

Risk Level	Total Students
Low	26,901
Moderate	9,250
High	1,967
Unknown	125

Risk Level by Gender

Risk Level	Male	Female
Low	12,429	14,472
Moderate	5,101	4,149
High	1,297	670
Unknown	60	65

Risk Level	Asian	Black/African American	Not Specified	White	Hispanic/Latino	Native Hawaiian/Other Pacific Islander	American Indian/Alaska Native
Low	6,841	2,813	692	5,869	9,784	32	49
Moderate	2,228	1,077	138	1,931	3,553	13	22
High	370	255	20	395	859	0	8
Unknown	43	10	<5	33	30	0	0

Note: Excludes 1,174 students who do not have a race/ethnicity group stored in Navigate

Risk Level by Race/Ethnicity and Gender

	Asian		Black/African American		Not Specified		White		Hispanic/Latino		Native Hawaiian/Other Pacific Islander		American Indian/Alaska Native	
Risk Level	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Low	3,371	3,470	1,134	1,679	344	348	2,936	2,933	4,258	5,526	12	20	27	22
Moderate	1,297	931	552	525	78	60	1,116	815	1,883	1,670	10	<5	12	10
High	253	117	164	91	11	9	276	119	555	304	0	0	<5	<5
Unknown	20	23	<5	8	<5	<5	20	13	12	18	0	0	0	0

Note: Excludes 1,174 students who do not have a race/ethnicity group stored in Navigate

Risk Level by GPA Group

Risk Level	GPA 0-1.999	GPA 2.0-2.999	GPA 3.0-4.0	GPA Unknown
Low	14	6,186	18,286	2,415
Moderate	125	1,914	833	6,378
High	795	243	57	872
Unknown	<5	6	66	50

Risk Level by Major

Major	Low	Moderate	High	Unknown
Accounting, BBA	627	11	<5	40
Adult Admission Opt, NDO UN	0	0	0	7
Advertising, BA	97	7	0	0
African American Studies, BA	8	<5	<5	0
American Sign Language Interpreting,BA	27	7	<5	0
Anthropology, BA	51	17	7	0
Anthropology, BS	66	24	10	0
Applied Music, BM	87	28	<5	0

Architecture, BARCH	505	160	31	0
Art History, BA	32	7	<5	0
Art, BA	167	23	<5	0
Art, NDO UN	0	0	0	<5
Biochemical & Biophysical Sciences, BA	<5	9	<5	0
Biochemical and Biophysical Sciences, BS	379	149	23	0
Biochemical and Biophysical Sciences, NDO UN	0	<5	0	0
Biology, BA	6	15	6	0
Biology, BS	1238	511	110	<5
Biology, NDO UN	0	0	0	<5
Biomedical Engineering, BSBE	191	91	16	0
Biotechnology, BS	304	123	38	0
Chemical Engineering, BSCHE	346	121	33	0
Chemistry, BA	9	<5	0	0
Chemistry, BS	130	79	18	0
Chemistry, NDO UN	0	0	0	<5
Chinese Studies, BA	23	<5	<5	0
Civil Engineering, BSCE	249	65	15	0
Communication Disorders, BA	26	<5	0	0
Communication Disorders, BS	40	<5	0	0
Communication Studies, BA	54	65	15	0
Communication Unspecified, DEG UN	<5	0	0	0
Computer Engineering Technology, BS	263	180	34	0
Computer Engineering, BSCPE	157	77	28	<5
Computer Information Systems, BS	839	240	60	<5
Computer Science, BS	895	347	73	0
Computer Science-Systems, BS	<5	0	<5	0
Construction Management, BS	562	163	34	<5
Dance, BA	14	<5	0	0
Dance, BFA	12	7	<5	0
Digital Media, BS	295	149	54	0

Earth Science, BA	7	<5	<5	0
Economics, BA	95	69	27	0
Economics, BS	369	126	23	0
Electrical Engineering, BSEE	309	99	26	0
Electrical Power Engineering Technology, BS	96	38	<5	0
Engineering Unspec, DEG UN	0	0	<5	0
Engineering, NDO UN	0	5	14	0
English, BA	337	83	22	0
Entrepreneurship, BBA	23	0	0	0
Environmental Design, BS	66	19	5	0
Exercise Science, BS	1103	309	57	0
Exploratory Studies, DEG UN	600	1360	142	0
Exploratory Studies, NDO UN	0	0	0	<5
Finance, BBA	766	<5	0	<5
Fitness & Sports, BS	99	51	<5	0
French, BA	7	<5	0	0
Geology, BS	58	11	6	0
Geophysics, BS	20	5	0	0
Graphic Design, BFA	149	80	16	0
Health Communication, BA	46	22	5	0
Health, BS	780	161	17	0
Health, DEG UN	83	15	11	0
History, BA	248	90	32	0
Honors Biomedical Sciences, BS	54	8	0	0
Hotel & Restaurant Management, BS	687	116	21	0
Human Development & Family Studies, BA	28	35	12	0
Human Development & Family Studies, BS	217	54	11	0
Human Development & Family Studies, DEG UN	28	<5	<5	0
Human Nutrition & Foods, BS	529	172	39	0
Human Nutrition & Foods, NDO UN	0	0	0	<5
Human Resource Development, BS	294	56	15	0

Industrial Design, BS	88	20	5	0
Industrial Engineering, BSIE	91	18	<5	<5
Information Systems Technology, NDO UN	0	0	<5	0
Integrated Communication, BA	42	<5	<5	0
Interior Architecture, BS	44	24	<5	0
Interpersonal Communication, BA	27	<5	0	0
Italian Studies, BA	<5	0	0	0
Journalism, BA	222	56	9	0
Journalism, DEG UN	<5	0	<5	0
Kinesiology, BS	<5	<5	0	0
Kinesiology-Exercise Science, NDO UN	0	0	0	<5
Liberal Arts and Social Sciences-Unspec, DEG UN	17	11	25	0
Liberal Arts and Social Sciences-Unspec, NDO UN	0	0	0	<5
Liberal Studies, BA	93	35	10	0
Liberal Studies, BS	62	29	18	0
Management Information Systems, BBA	570	<5	0	0
Management, BBA	262	<5	<5	0
Marketing, BBA	425	<5	0	0
Mathematical Biology, BS	37	<5	<5	0
Mathematics, BA	13	9	<5	0
Mathematics, BS	267	71	9	0
Mechanical Engineer Tech, BS	367	133	51	0
Mechanical Engr, BSME	741	250	54	0
Media Policy/Media Studies, BA	13	<5	0	0
Media Prod, BA	376	100	22	0
Media Prod, DEG UN	<5	0	0	0
Music, BA	9	<5	<5	0
Music, BM	210	44	5	0
Natural Sciences and Mathematics Unspecified, DEG UN	9	<5	<5	0
Natural Sciences and Mathematics Unspecified, NDO UN	0	0	0	<5
Nursing, BSN	48	5	<5	0

Organizational/Corporate Comm, BA	50	<5	<5	0
Painting, BFA	53	9	0	0
Petroleum Engineering, BSPETE	79	23	8	0
Philosophy, BA	58	39	10	0
Photography/Digital Media, BFA	60	18	<5	0
Physics, BA	<5	0	<5	0
Physics, BS	61	30	16	0
Political Science, BA	318	139	47	0
Political Science, BS	267	93	21	0
Political Science, NDO UN	0	0	0	<5
Pre-American Sign Language, DEG UN	23	19	<5	0
Pre-Business Administration, DEG UN	1704	846	163	0
Pre-Communication Disorder, DEG UN	148	78	<5	0
Pre-Nursing, DEG UN	<5	42	0	0
Pre-Pharmacy, DEG UN	<5	0	0	0
Pre-Psychology, DEG UN	1212	799	168	0
Pre-Sociology, DEG UN	51	39	27	0
Psychology, BA	176	<5	<5	0
Psychology, BS	557	20	<5	0
Public Relations, BA	129	<5	0	0
Religious Studies, BA	14	<5	<5	0
Retailing & Consumer Science, BS	198	33	14	0
Sculpture, BFA	21	<5	0	0
Sociology, BA	36	7	<5	0
Sociology, BS	64	<5	0	0
Spanish, BA	53	14	5	0
Spanish, NDO UN	0	0	0	<5
Sports Administration, BS	201	54	10	0
Strategic Communication, BA	255	81	14	0
Supply Chain & Logistics Technology, BS	475	82	24	<5
Supply Chain Management, BBA	550	<5	<5	<5

Teaching and Learning, BS	848	145	15	0
Teaching and Learning, DEG UN	70	6	<5	0
Technology Leadership and Innovation Management, BS	167	71	36	0
Technology-Unspec, DEG UN	<5	0	<5	0
Theatre, BFA	105	30	<5	0
Visiting Student, NDO UN	0	0	0	53
Women's, Gender, Sexuality, BA	13	7	<5	0
World Culture & Literature, BA	34	6	<5	0

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

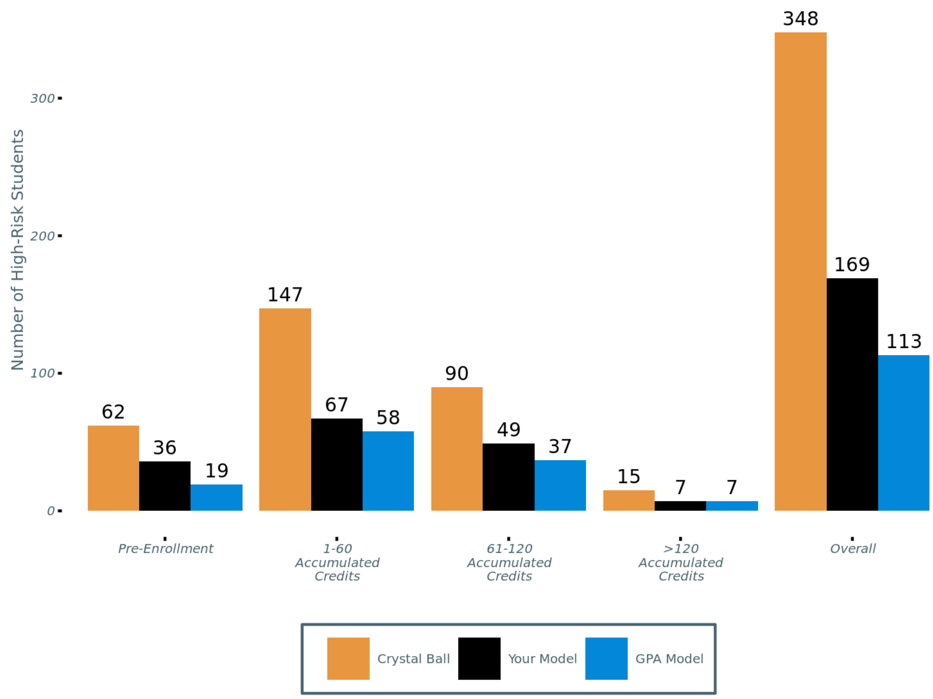
We may divide the percentage of actual high-risk students identified by a given model by the percentage found by a Blind Campaign to create a new metric called "lift". For instance, a lift of two would mean that a campaign based on your SSPM identified twice as many high-risk students as a Blind Campaign, 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.

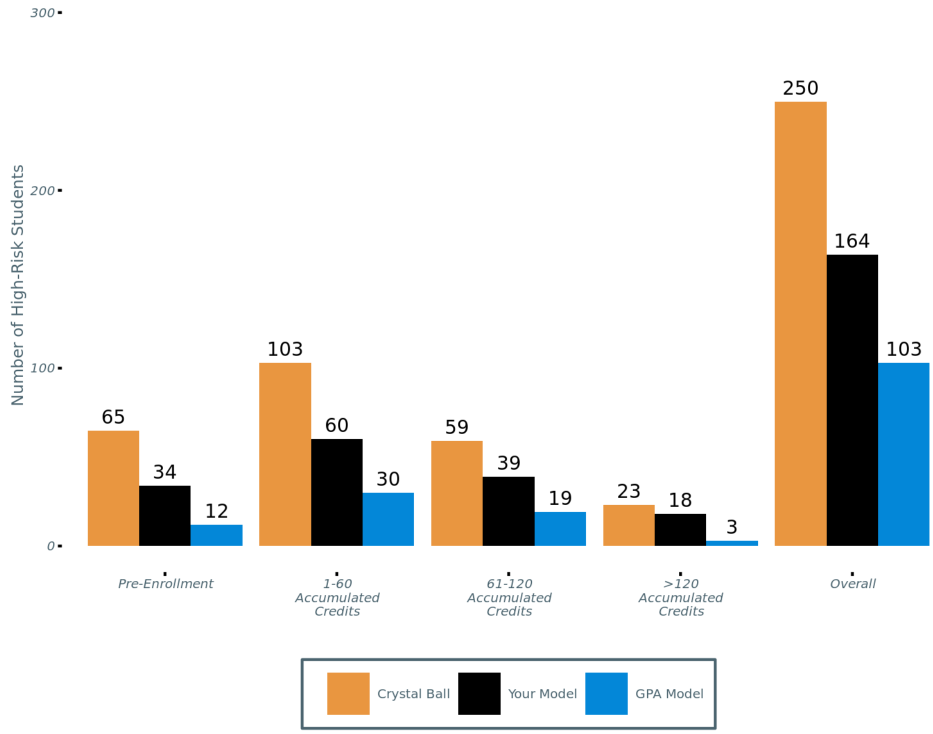
Murky Middle

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.



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.

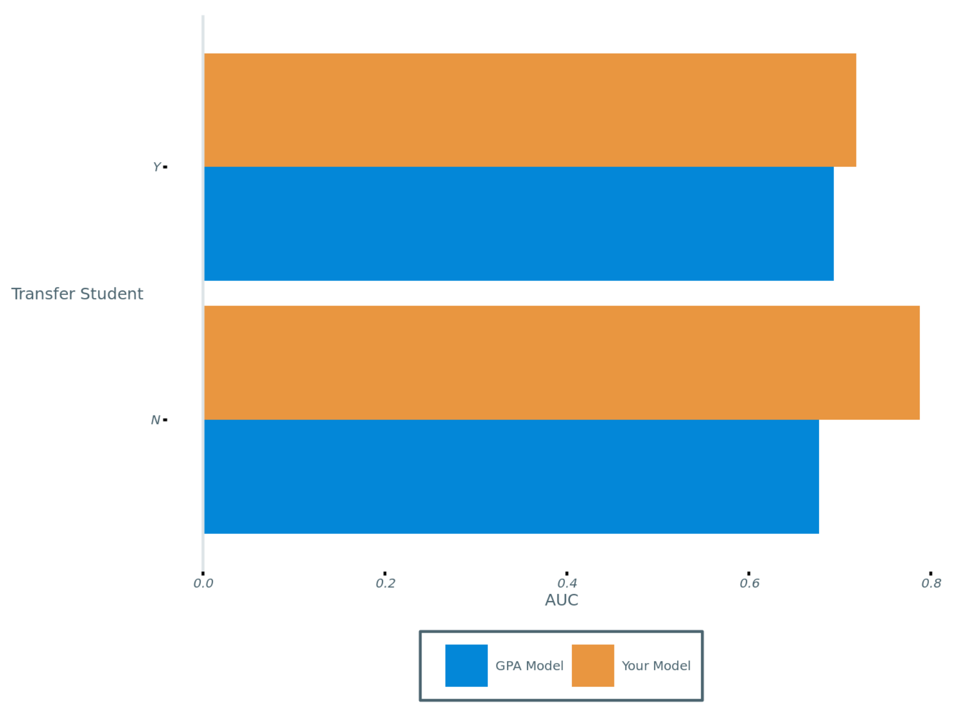
The table below shows AUC values for your SSPM and the GPA Model.

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

Accuracy for Transfer/Non-Transfer Students



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

Student Success Predictive Model Report – Texas A & M University-College Station

March 18, 2020

Executive summary

EAB has built a customized Student Success Predictive Model (SSPM) for your institution that predicts the graduation 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.

Performance Summary

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.

The table below summarizes your SSPM's performance and compares it to the following notional models:

- A fictitious, perfectly prescient model (Crystal Ball).
- A model based exclusively on students' cumulative GPAs (GPA Model).
- A model that randomly targets students (Blind Campaign).

The columns assume different percentages of your total student population that you are able to cover in the campaign, while the rows provide the percentage of your school's actual high-risk students that will be identified in the campaign based on each model.

The bottom row highlights the substantial relative percentage gains achieved in going from the simple GPA Model to your advanced Student Success Predictive Model and demonstrates that your model is much better at distinguishing between students who are on track to graduate and those that need intervention in order to succeed.

Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	9%	17%	43%	85%
Your Model	8%	16%	38%	71%
GPA Model	8%	15%	36%	67%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**0%**	**7%**	**6%**	**6%**

Your SSPM is high-performing; it can be used confidently to both assess individual students and efficiently design effective, targeted intervention campaigns.

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 graduate within four years from matriculation date (from any major). 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 graduation outcomes.

The model outputs a success score between zero and one estimating the probability that a selected student will graduate within four years. That is, each student's success score corresponds to the model's estimate of their likelihood of graduating within four years. 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 graduation. 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 graduate within four years.

Methodology

The SSPM uses the latest advances in data science to estimate graduation 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 2010-08-28 and 2012-08-27.
- 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.

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.

Separate models were trained for students in the following subgroups, as it was found to improve performance:

- Transfer Indicator

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.

- Pre-Enrollment|transfer_student=N High Impact Predictors

  Admit Code

  High School Size

  First Generation Indicator

  In State Resident Indicator

- Pre-Enrollment|transfer_student=N Other Predictors

  Veteran Indicator

  High School GPA

  Race/Ethnicity

  High School Percentile

  International Indicator

  Gender

- Pre-Enrollment|transfer_student=Y High Impact Predictors

  Admit Code

- Pre-Enrollment|transfer_student=Y Other Predictors

  Admit Code

  Race/Ethnicity

  First Generation Indicator

  Gender

  High School GPA

  High School Percentile

  High School Size

  In State Resident Indicator

  International Indicator

  Veteran Indicator

- |binned_credits=0 High Impact Predictors

  Race/Ethnicity

Gender

Admit Code

High School Percentile

- |binned_credits=0 Other Predictors

  SAT/ACT Math Score Percentile

  Credits Attempted Current Term

  First Generation Indicator

  Age at First Term

  SAT/ACT Verbal Score Percentile

  High School Size

  High School GPA

  Average Credits Attempted per Term

  International Indicator

  Average Success Outcome of Students Declared in Same Major

  …and 1 more features.

- |binned_credits=120|transfer_student=N High Impact Predictors

  Ratio of Earned to Attempted Credits

  Admit Code

  Gender

  Estimated Skills

  First Term Transfer Credits

  Average Credits Attempted per Term

- |binned_credits=120|transfer_student=N Other Predictors

  Average Success Outcome of Students Declared in Same Major

  Age at First Term

  Cumulative GPA

  Race/Ethnicity

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

  Proportion of Transfer Credits

  Credits Attempted Current Term

Number of Completed Terms

Trend in Term GPA

Recent Change in GPA

…and 12 more features.

- |binned_credits=120|transfer_student=Y High Impact Predictors

  First Term Transfer Credits

  Number of Completed Terms

  Ratio of Earned to Attempted Credits

  Proportion of Transfer Credits

  Admit Code

- |binned_credits=120|transfer_student=Y Other Predictors

  Lifetime Attempted Credits

  Average Credits Attempted per Term

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

  Lifetime Transfer Credits

  Average Success Outcome of Students Declared in Same Major

  Cumulative GPA

  International Indicator

  Race/Ethnicity

  Age at First Term

  …and 14 more features.

- |binned_credits=60|transfer_student=N High Impact Predictors

  Race/Ethnicity

  Cumulative GPA

  Admit Code

  Gender

  Number of Completed Terms

  Ratio of Earned to Attempted Credits

  First Term Transfer Credits

  Estimated Skills

- |binned_credits=60|transfer_student=N Other Predictors

  Credits Attempted Current Term

  High School Percentile

  First Generation Indicator

  Proportion of Transfer Credits

  Average Credits Attempted per Term

  Lifetime Accumulated Credits

  Lifetime Transfer Credits

  Average Success Outcome of Students Declared in Same Major

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

  SAT/ACT Math Score Percentile

  …and 10 more features.

- |binned_credits=60|transfer_student=Y High Impact Predictors

  Number of Completed Terms

  First Term Transfer Credits

  Admit Code

  Gender

- |binned_credits=60|transfer_student=Y Other Predictors

  In State Resident Indicator

  Ratio of Earned to Attempted Credits

  Age at First Term

  Race/Ethnicity

  Ratio of Credits Attempted Current Term to Prior Term

  First Generation Indicator

  Average Credits Attempted per Term

  Cumulative GPA

  Lifetime Attempted Credits

  …and 15 more features.

- |binned_credits=Inf|transfer_student=N High Impact Predictors

  Admit Code

Credits Attempted Current Term

Number of Completed Terms

Cumulative GPA

Lifetime Attempted Credits

Race/Ethnicity

Proportion of Transfer Credits

- |binned_credits=Inf|transfer_student=N Other Predictors

Lifetime Transfer Credits

High School Percentile

First Term Transfer Credits

Age at First Term

Average Success Outcome of Students Declared in Same Major

High School Size

Estimated Skills

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

First Generation Indicator

Lifetime Accumulated Credits

…and 10 more features.

- |binned_credits=Inf|transfer_student=Y High Impact Predictors

High School GPA

Ratio of Earned to Attempted Credits

Admit Code

Average Credits Attempted per Term

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

Estimated Skills

International Indicator

- |binned_credits=Inf|transfer_student=Y Other Predictors

Proportion of Transfer Credits

SAT/ACT Math Score Percentile

Ratio of Credits Attempted Current Term to Prior Term

Average Success Outcome of Students Declared in Same Major

In State Resident Indicator

High School Size

Lifetime Transfer Credits

Veteran Indicator

Cumulative GPA

Recent Change in GPA

…and 11 more features.

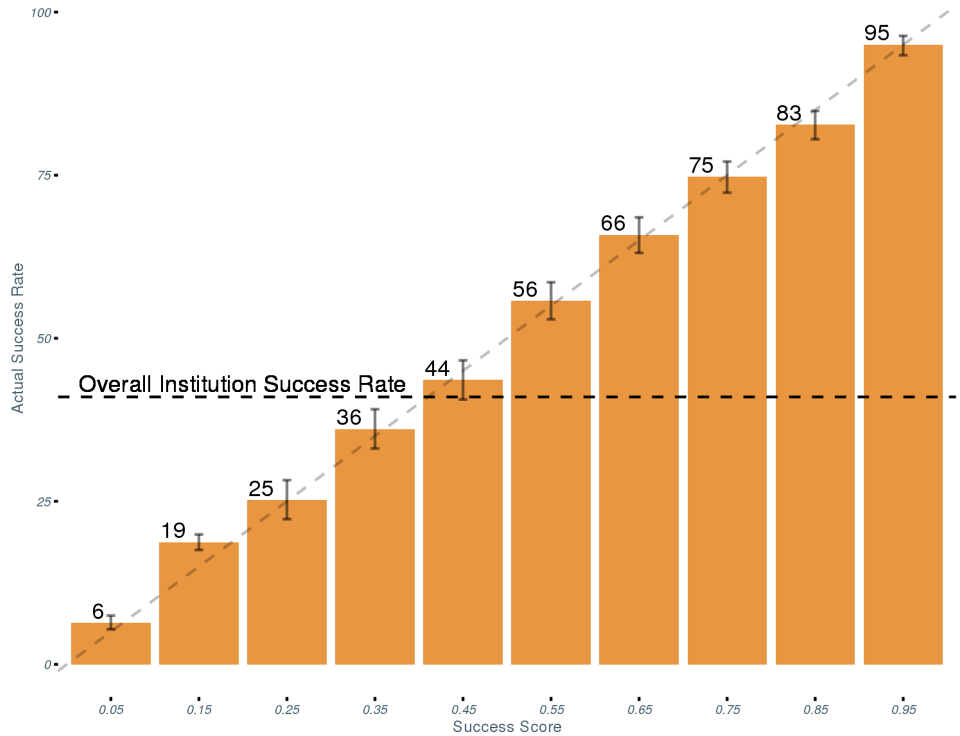
Model Performance

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 and compared to other notional models. These notional models include a fictitious, perfectly prescient Crystal Ball; a GPA Model based exclusively on students’ cumulative GPAs; and a Blind Campaign model that randomly targets students. This section describes the most insightful performance benchmarks and compares your SSPM to these other notional models.

Calibration

Calibration offers an intuitive way to evaluate a model by capturing how close its estimated probability scores are to reality.

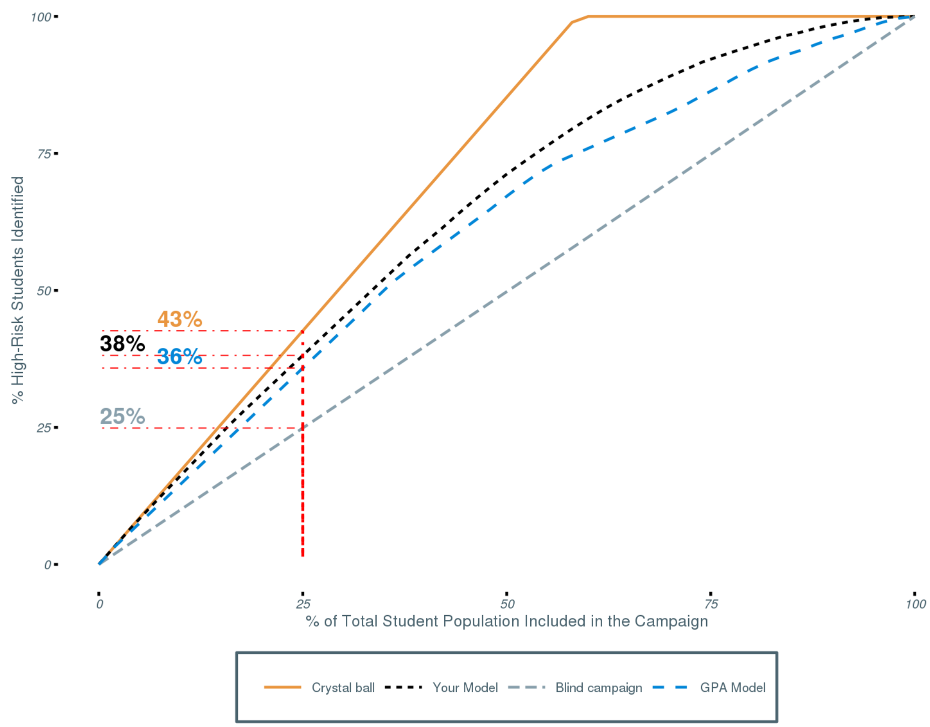
Students are divided into different bins along the horizontal axis according to their success score, while the vertical height of each bin indicates the actual graduation rate of the historical students it contains. The horizontal line shows the overall percentage of students that graduated within four years from matriculation date.



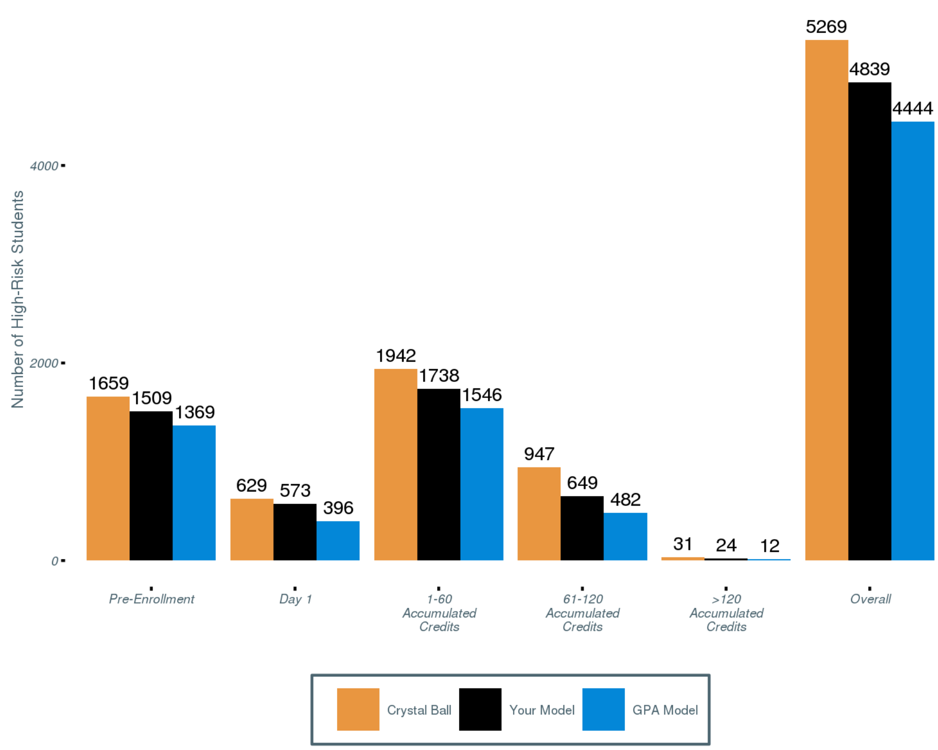
High-Risk Student Identification Rate

The SSPM enables you to rank students by order of risk (i.e., success scores from low to high) so that you can most efficiently target as many high-risk students for intervention as your institution or office/department can effectively handle. Let's assume, for instance, that you are designing an intervention campaign targeting high-risk students and have the capacity to advise N students. Let's assume you use different predictive models to generate lists of N targeted high-risk students, and step forward in time to compare their performance by evaluating the percentage of those N students that did not graduate within four years. This performance comparison is summarized in the high-risk student identification rate chart below, which shows the percentage of actual high-risk historical students (i.e., students that did not graduate within four years.) identified by the model vs. the percentage of the total student population targeted in the campaign. For example, if you design a campaign that includes 25% of your total student population, then the percentage of your school's high-risk students identified by the campaign will be 43, 38, 36, and 25 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

High-risk student identification rate provides a powerful and transparent performance benchmark of model performance; the large performance enhancement gained in going from the simple GPA Model to your advanced SSPM is clearly visible in the chart.



High-risk student identification rates can also be converted to actual numbers of students and compared across different accumulated credit subgroups, as shown in the figure below for campaigns targeting 25% of the total student population.

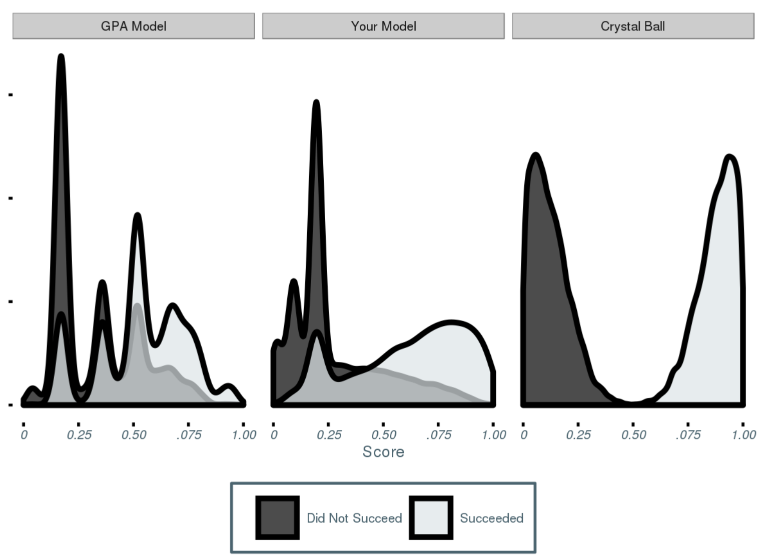


Lift

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 1.71, 1.54, 1.49, and 1.00 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

Separation

Displaying the distributions of success scores for students in the historical test set who did and did not graduate within four years from matriculation date also provides an intuitive sense of a model’s performance. We see in the charts below that successful students (light gray) typically have higher scores than unsuccessful ones (dark gray) for both your SSPM and the GPA Model, as you would expect, but that your SSPM is much better at separating these two student populations from each other. That is, the graphic demonstrates that your SSPM ascribes high success scores to successful students and low success scores to unsuccessful students more accurately than the GPA Model. A perfect prediction would result in complete separation between the students (shown in the Crystal Ball chart on the right).



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 I: Evaluating AUC

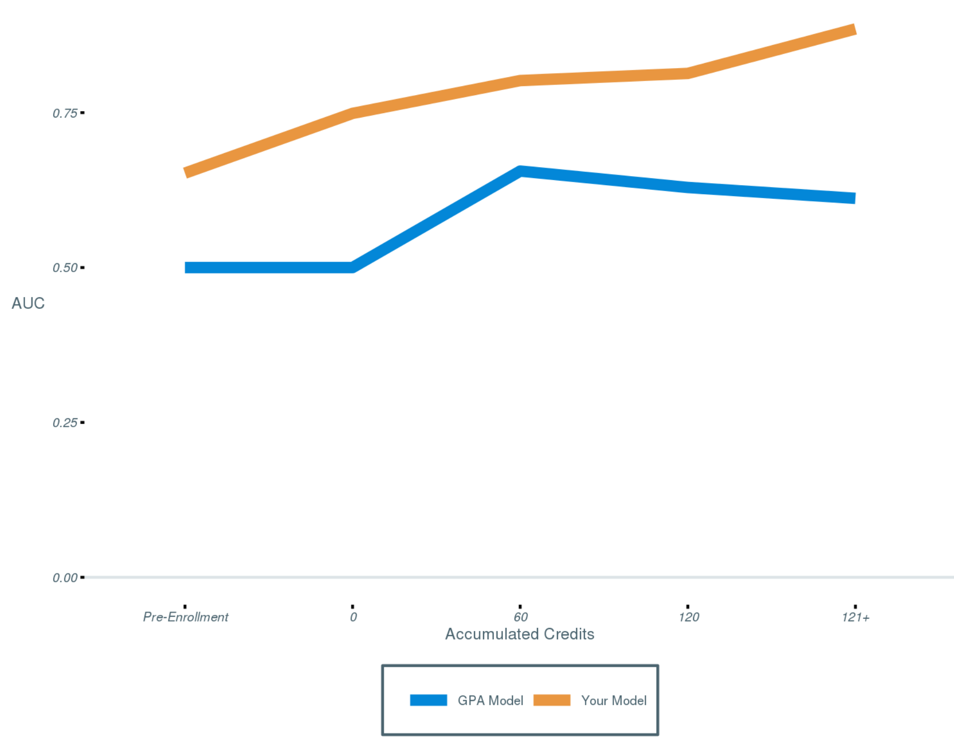
We commonly use AUC to measure and tune the performance of your Student Success Predictive Model across your institution's entire student population and different subgroups. AUC stands for Area Under the Curve and is a measure used extensively in data science, which ranges from 0.5 (pure chance) to 1.0 (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.

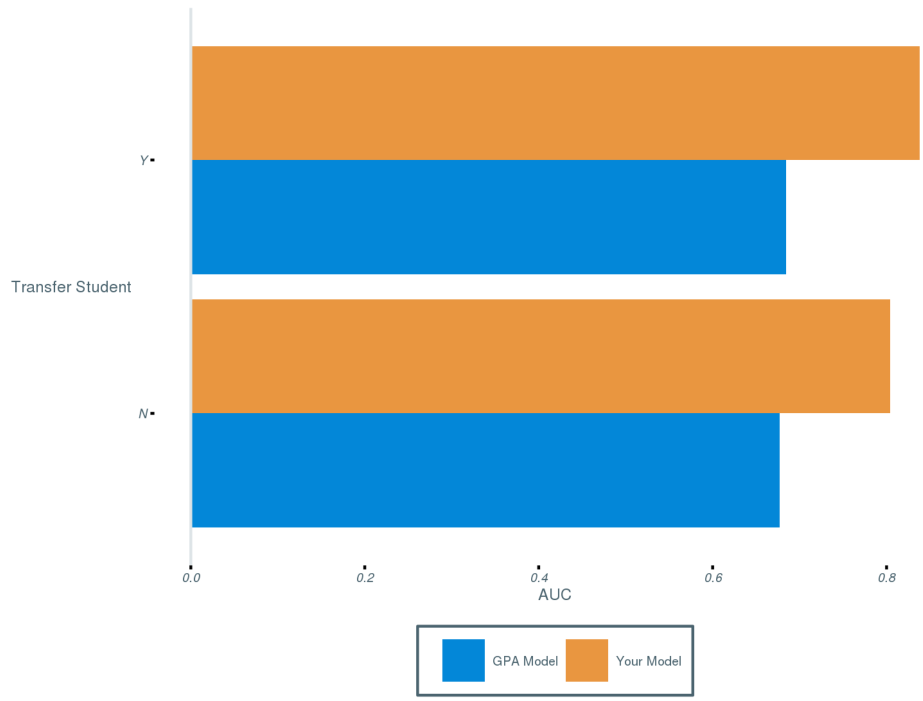
Model	AUC
GPA Model	0.70
Your Model	0.82

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



Accuracy for Transfer/Non-Transfer Students

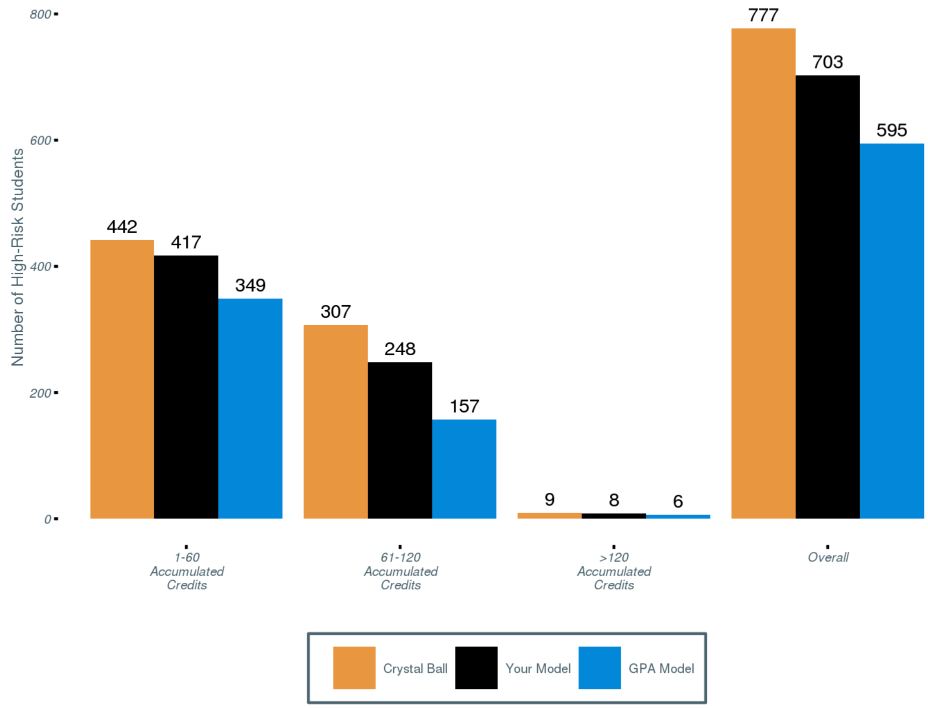


Appendix II: High-Risk Student Identification Rate for Murky Middle and Top Performing Students

Your Student Success Predictive Model's performance varies across different subgroups of students. This appendix 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 appendix for two student subgroups.

Murky Middle

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

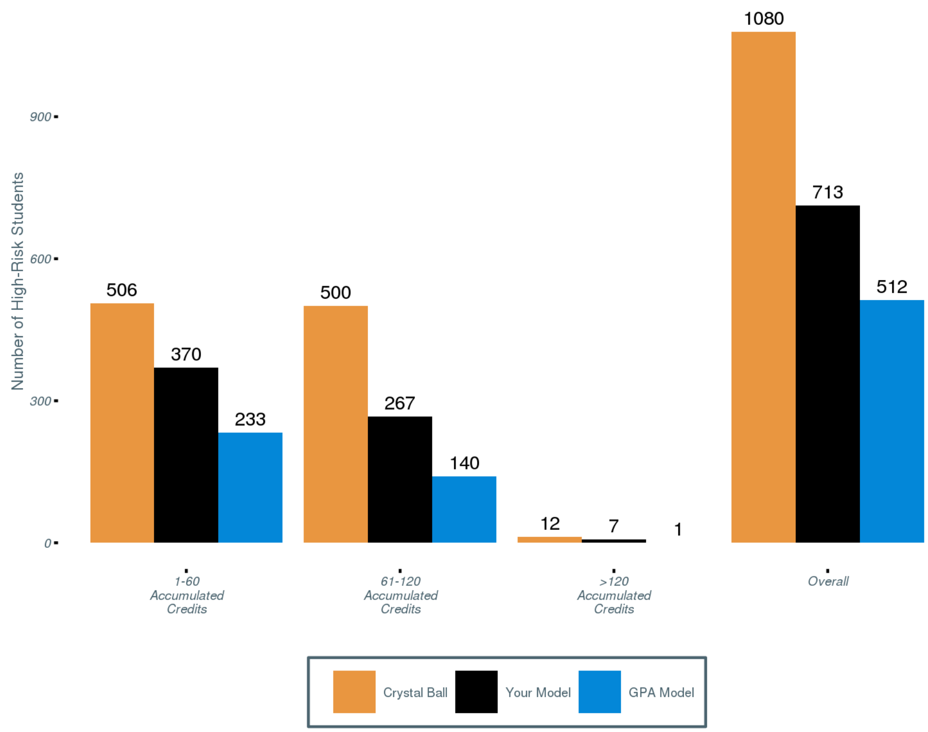


Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	9%	18%	45%	90%
Your Model	9%	17%	41%	71%
GPA Model	8%	15%	34%	62%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**12%**	**13%**	**21%**	**15%**

Top performing students

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



Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	15%	31%	77%	100%
Your Model	14%	25%	51%	78%
GPA Model	8%	14%	36%	62%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**75%**	**79%**	**42%**	**26%**

Appendix III – 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 Generation Indicator: “Yes” or “No” indicator of whether any of an individual’s parents have ever earned a bachelor’s degree.
- 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.
- Lifetime Accumulated Credits: The total number of credits a student has accumulated across all institutions.
- Lifetime Attempted Credits: The total number of credits a student has attempted at your institution.
- Lifetime Transfer Credits: The total number of credits a student has transferred from other institutions.
- 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.
- Trend in Term GPA: A measure of the change over time in a student’s term GPAs.
- Veteran Indicator: “Yes” or “No” indicator of whether a student is a veteran of the United States Armed Forces.

EAB

Student Success Collaborative

Student Success Predictive Model Report – uwm_10132

August 26, 2020

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.

Performance Summary

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.

The table below summarizes your SSPM's performance and compares it to the following notional models:

- A fictitious, perfectly prescient model (Crystal Ball).
- A model based exclusively on students' cumulative GPAs (GPA Model).
- A model that randomly targets students (Blind Campaign).

The columns assume different percentages of your total student population that you are able to cover in the campaign, while the rows provide the percentage of your school's actual high-risk students that will be identified in the campaign based on each model.

The bottom row highlights the substantial relative percentage gains achieved in going from the simple GPA Model to your advanced Student Success Predictive Model and demonstrates that your model is much better at distinguishing between students who are on track to graduate and those that need intervention in order to succeed.

Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	20%	40%	98%	100%
Your Model	14%	25%	47%	75%
GPA Model	9%	16%	35%	63%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**56%**	**56%**	**34%**	**19%**

Your SSPM is high-performing; it can be used confidently to both assess individual students and efficiently design effective, targeted intervention campaigns.

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 term of the regularly scheduled academic year (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 term. That is, each student's success score corresponds to the model's estimate of their likelihood of persisting to the next term. 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 term.

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 2006-08-26 and 2018-09-04.
- 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.

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 III near the end of the report.

- Pre-Enrollment Students High Impact Predictors

  Admit Code

  High School Percentile

  Race/Ethnicity

  Transfer Indicator

  High School GPA

  …and 5 low impact predictors.
- Day 1 Students High Impact Predictors

  Admit Code

  Credits Attempted Current Term

  Race/Ethnicity

  High School Percentile

  Average Credits Attempted per Term

  Average Success Outcome of Students Declared in Same Major

  High School GPA

  …and 17 low impact predictors.
- Students with Between 1-60 Accumulated Credits High Impact Predictors

  Credits Attempted Current Term

  Number of Completed Terms

  Cumulative GPA

  First Term Transfer Credits

  Average Success Outcome of Students Declared in Same Major

  …and 22 low impact predictors.
- Students with Between 61-120 Accumulated Credits High Impact Predictors

  Credits Attempted Current Term

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

  Average Credits Attempted per Term

  Admit Code

Cumulative GPA

Ratio of Earned to Attempted Credits

…and 21 low impact predictors.

- Students with More Than 120 Accumulated Credits High Impact Predictors

Cumulative GPA

Credits Attempted Current Term

Average Success Outcome of Students Declared in Same Major

Average Credits Attempted per Term

Proportion of Transfer Credits

Ratio of Earned to Attempted Credits

Age at First Term

…and 19 low impact predictors.

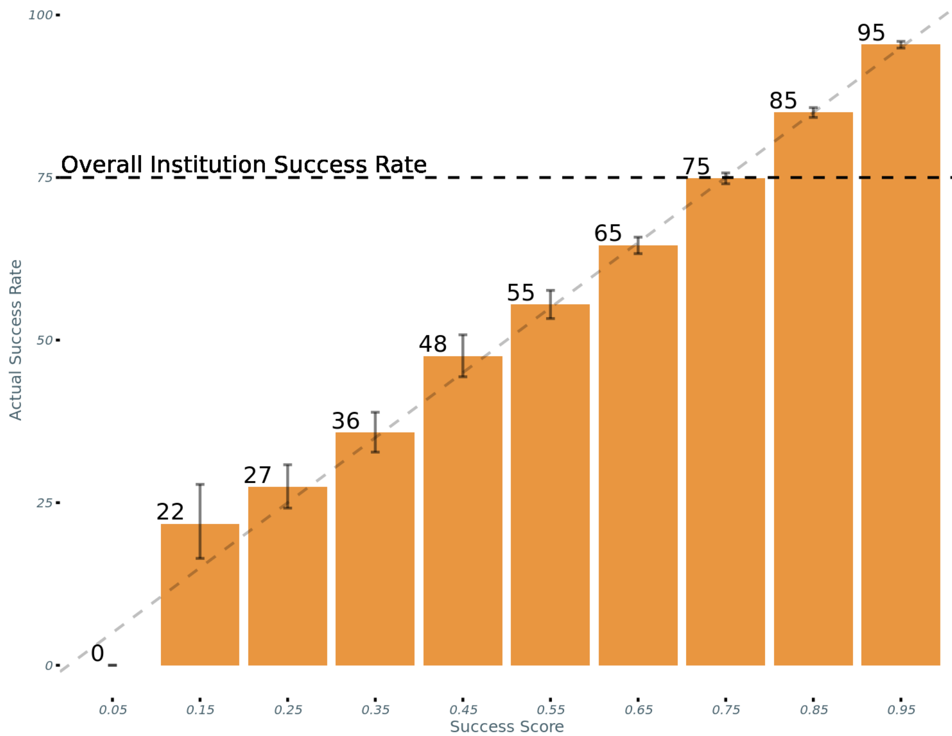
Model Performance

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.

Calibration

Calibration offers an intuitive way to evaluate a model by capturing how close its estimated probability scores are to reality.

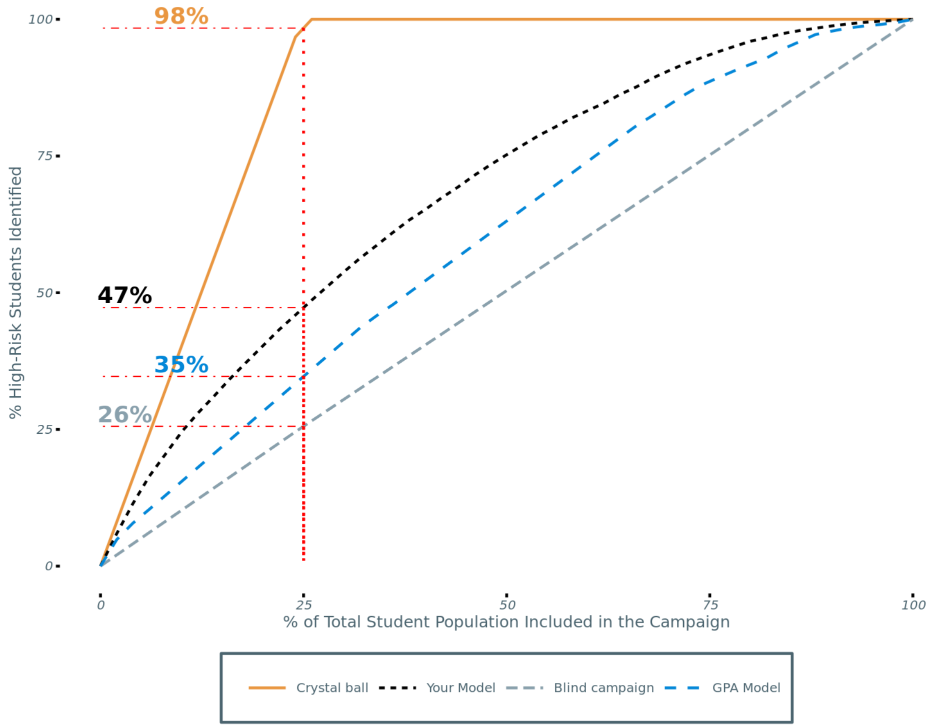
Students are divided into different bins along the horizontal axis according to their success score, while the vertical height of each bin indicates the actual persistence rate of the historical students it contains. The horizontal line shows the overall percentage of students that persisted to the next term.



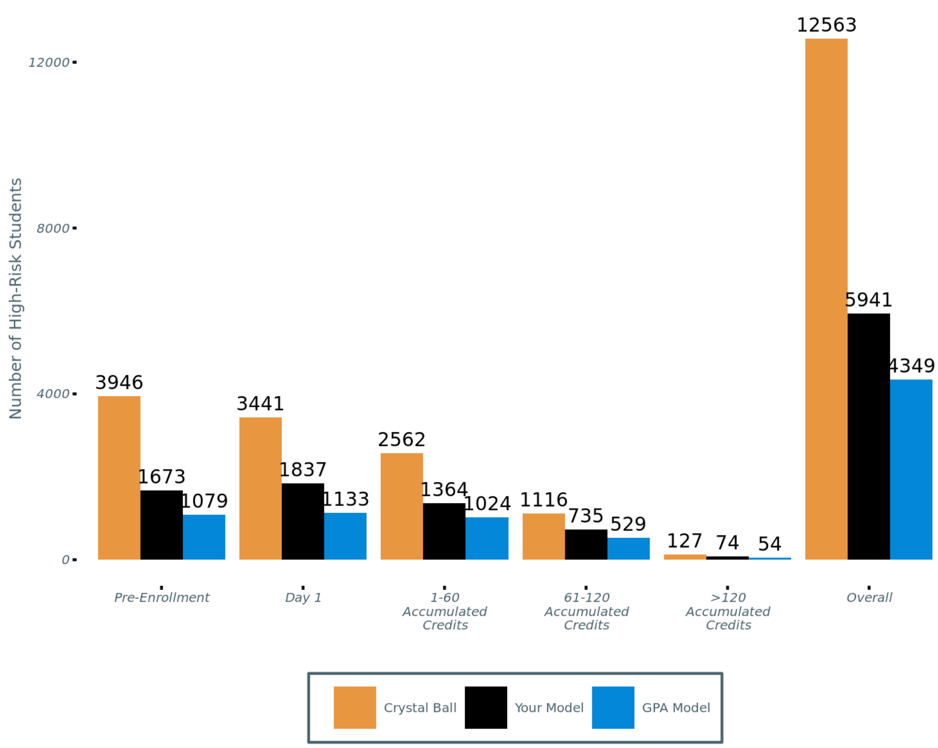
High-Risk Student Identification Rate

The SSPM enables you to rank students by order of risk (i.e., success scores from low to high) so that you can most efficiently target as many high-risk students for intervention as your institution or office/department can effectively handle. Let's assume, for instance, that you are designing an intervention campaign targeting high-risk students and have the capacity to advise N students. Let's assume you use different predictive models to generate lists of N targeted high-risk students, and step forward in time to compare their performance by evaluating the percentage of those N students that did not persist to the next term. This performance comparison is summarized in the high-risk student identification rate chart below, which shows the percentage of actual high-risk historical students (i.e., students that did not persist to the next term.) identified by the model vs. the percentage of the total student population targeted in the campaign. For example, if you design a campaign that includes 25% of your total student population, then the percentage of your school's high-risk students identified by the campaign will be 98, 47, 35, and 26 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

High-risk student identification rate provides a powerful and transparent performance benchmark of model performance; the large performance enhancement gained in going from the simple GPA Model to your advanced SSPM is clearly visible in the chart.



High-risk student identification rates can also be converted to actual numbers of students and compared across different accumulated credit subgroups, as shown in the figure below for campaigns targeting 25% of the total student population.

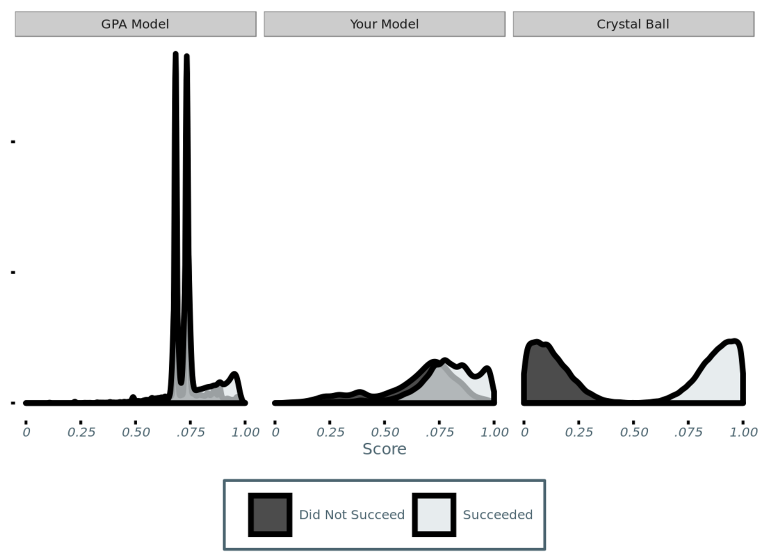


Lift

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 4.01, 1.89, 1.5, and 1.00 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

Separation

Displaying the distributions of success scores for students in the historical test set who did and did not persist to the next term also provides an intuitive sense of a model’s performance. We see in the charts below that successful students (light gray) typically have higher scores than unsuccessful ones (dark gray) for both your SSPM and the GPA Model, as you would expect, but that your SSPM is much better at separating these two student populations from each other. That is, the graphic demonstrates that your SSPM ascribes high success scores to successful students and low success scores to unsuccessful students more accurately than the GPA Model. A perfect prediction would result in complete separation between the students (shown in the Crystal Ball chart on the right).



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 I: Evaluating AUC

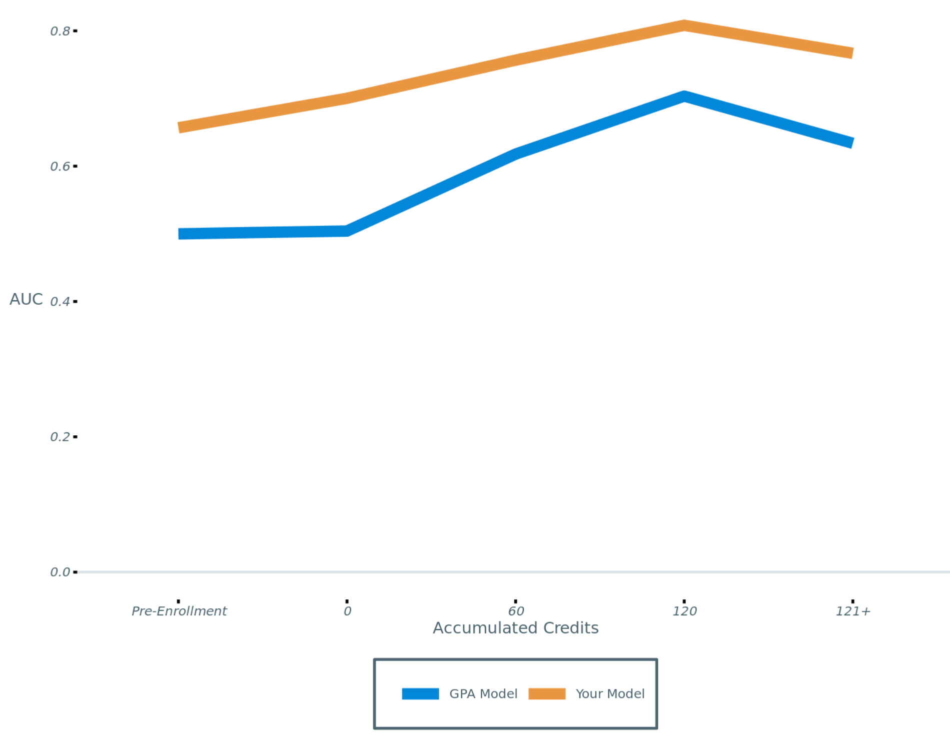
We commonly use AUC to measure and tune the performance of your Student Success Predictive Model across your institution's entire student population and different subgroups. AUC stands for Area Under the Curve and is a measure used extensively in data science, which ranges from 0.5 (pure chance) to 1.0 (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.

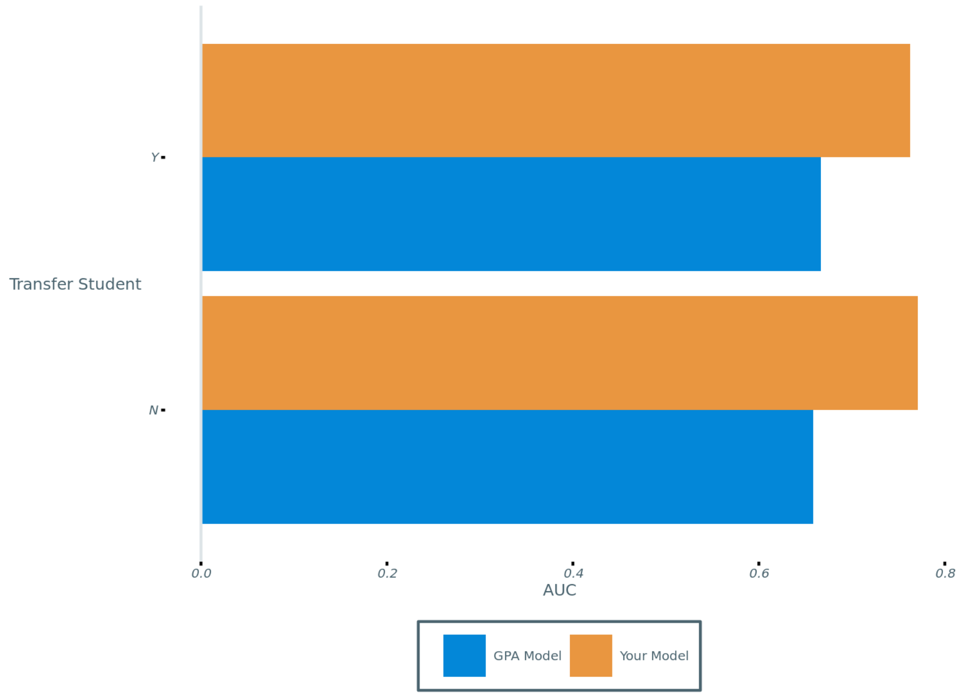
Model	AUC
GPA Model	0.67
Your Model	0.77

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



Accuracy for Transfer/Non-Transfer Students

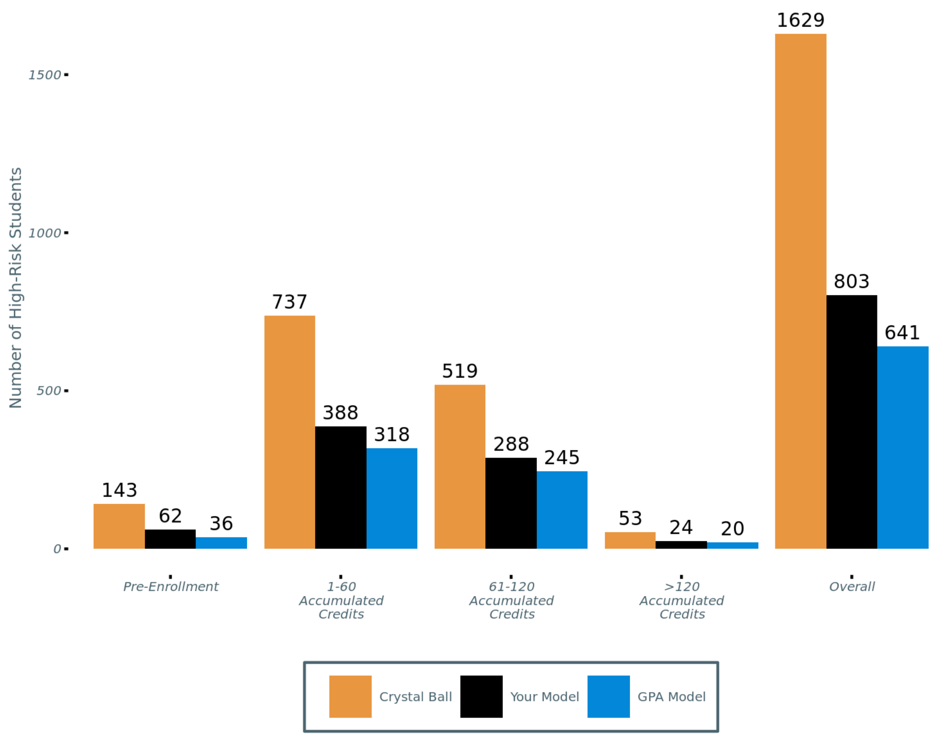


Appendix II: High-Risk Student Identification Rate for Murky Middle and Top Performing Students

Your Student Success Predictive Model's performance varies across different subgroups of students. This appendix 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 appendix for two student subgroups.

Murky Middle

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

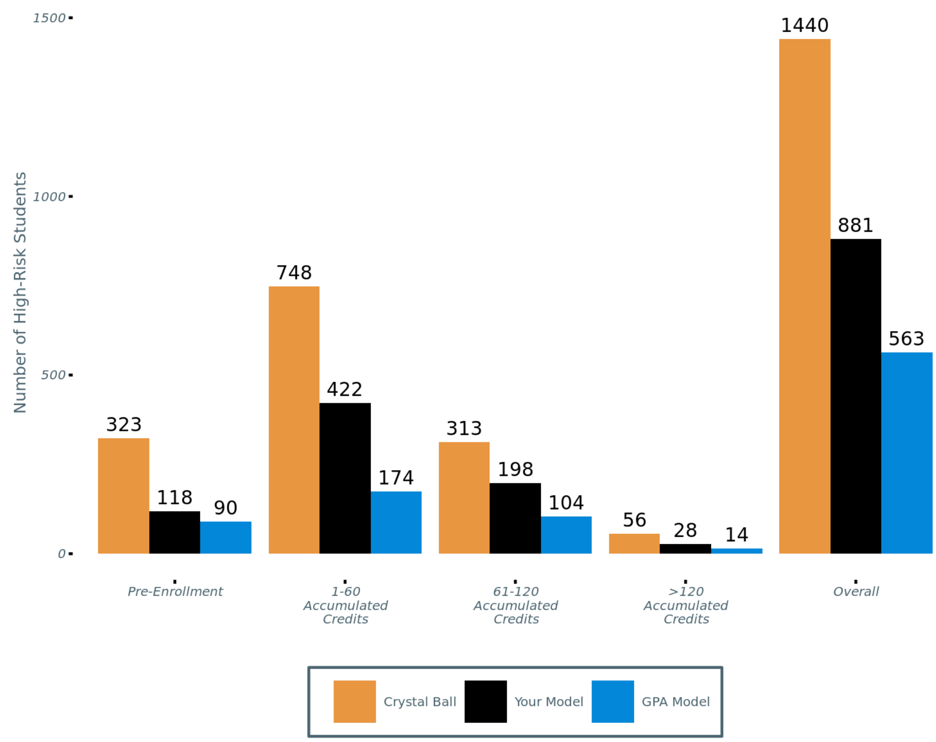


Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	22%	44%	100%	100%
Your Model	13%	25%	49%	77%
GPA Model	11%	19%	39%	68%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**18%**	**32%**	**26%**	**13%**

Top performing students

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



Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	40%	81%	100%	100%
Your Model	19%	34%	61%	86%
GPA Model	9%	17%	39%	77%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**111%**	**100%**	**56%**	**12%**

Appendix III – Additional Predictor Information

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

Admit Code

High School Percentile

Race/Ethnicity

Transfer Indicator

High School GPA

- Pre-Enrollment Students Other Predictors

    International Indicator

    First Generation Indicator

    Gender

    In State Resident Indicator

    High School Size

- Day 1 Students High Impact Predictors

    Admit Code

    Credits Attempted Current Term

    Race/Ethnicity

    High School Percentile

    Average Credits Attempted per Term

    Average Success Outcome of Students Declared in Same Major

    High School GPA

- Day 1 Students Other Predictors

    First Generation Indicator

    International Indicator

    Ratio of Credits Attempted Current Term to Prior Term

    SAT/ACT Verbal Score Percentile

    Gender

    SAT/ACT Math Score Percentile

    High School Size

    In State Resident Indicator

    Age at First Term

    Cumulative GPA

    Ratio of Earned to Attempted Credits

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

    Number of Completed Terms

    Proportion of Transfer Credits

    Trend in Term GPA

Estimated Skills

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

  Credits Attempted Current Term

  Number of Completed Terms

  Cumulative GPA

  First Term Transfer Credits

  Average Success Outcome of Students Declared in Same Major

- Students with Between 1-60 Accumulated Credits Other Predictors

  Average Credits Attempted per Term

  Proportion of Transfer Credits

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

  Admit Code

  High School Percentile

  Estimated Skills

  Trend in Term GPA

  Race/Ethnicity

  International Indicator

  Transfer Indicator

  Gender

  Ratio of Earned to Attempted Credits

  Recent Change in GPA

  In State Resident Indicator

  First Generation Indicator

  SAT/ACT Verbal Score Percentile

  High School GPA

  High School Size

  Age at First Term

  SAT/ACT Math Score Percentile

  Ratio of Credits Attempted Current Term to Prior Term

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

  Credits Attempted Current Term

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

Average Credits Attempted per Term

Admit Code

Cumulative GPA

Ratio of Earned to Attempted Credits

- Students with Between 61-120 Accumulated Credits Other Predictors

Average Success Outcome of Students Declared in Same Major

Trend in Term GPA

Number of Completed Terms

Age at First Term

Recent Change in GPA

High School Percentile

SAT/ACT Verbal Score Percentile

Ratio of Credits Attempted Current Term to Prior Term

Gender

Proportion of Transfer Credits

First Generation Indicator

Race/Ethnicity

Estimated Skills

International Indicator

Transfer Indicator

First Term Transfer Credits

High School Size

SAT/ACT Math Score Percentile

High School GPA

In State Resident Indicator

- Students with More Than 120 Accumulated Credits High Impact Predictors

Cumulative GPA

Credits Attempted Current Term

Average Success Outcome of Students Declared in Same Major

Average Credits Attempted per Term

Proportion of Transfer Credits

Ratio of Earned to Attempted Credits

Age at First Term

- Students with More Than 120 Accumulated Credits Other Predictors

Trend in Term GPA

Race/Ethnicity

SAT/ACT Verbal Score Percentile

Admit Code

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

Gender

Ratio of Credits Attempted Current Term to Prior Term

High School Percentile

High School GPA

In State Resident Indicator

SAT/ACT Math Score Percentile

International Indicator

Estimated Skills

First Term Transfer Credits

Number of Completed Terms

Recent Change in GPA

First Generation Indicator

High School Size

Transfer Indicator

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 Generation Indicator: “Yes” or “No” indicator of whether any of an individual’s parents have ever earned a bachelor’s degree.
- 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.

Student Success Predictive Model Report – Kansas State University

November 8, 2019

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.

Performance Summary

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.

The table below summarizes your SSPM's performance and compares it to the following notional models:

- A fictitious, perfectly prescient model (Crystal Ball).
- A model based exclusively on students' cumulative GPAs (GPA Model).
- A model that randomly targets students (Blind Campaign).

The columns assume different percentages of your total student population that you are able to cover in the campaign, while the rows provide the percentage of your school's actual high-risk students that will be identified in the campaign based on each model.

The bottom row highlights the substantial relative percentage gains achieved in going from the simple GPA Model to your advanced Student Success Predictive Model and demonstrates that your model is much better at distinguishing between students who are on track to graduate and those that need intervention in order to succeed.

Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	39%	78%	100%	100%
Your Model	27%	43%	68%	88%
GPA Model	25%	34%	52%	74%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**8%**	**26%**	**31%**	**19%**

Your SSPM is high-performing; it can be used confidently to both assess individual students and efficiently design effective, targeted intervention campaigns.

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 term of the regularly scheduled academic year (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 term. That is, each student's success score corresponds to the model's estimate of their likelihood of persisting to the next term. 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 term.

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 member, 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 2008-08-25 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.

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.

Separate models were trained for students in the following subgroups, as it was found to improve performance:

- Transfer Indicator

Your Predictors

The lists below rank the top ten predictors for each subgroup of students included in the model. Please note that transfer credits are incorporated in credit bin determinations.

- |Pre-Enrollment|transfer_student=N

Admit Code

First Generation Indicator

Median Income by Admission Zip Code

High School GPA

Legacy Indicator

International Indicator

High School Percentile

High School Size

Veteran Indicator

In State Resident Indicator

- |Pre-Enrollment|transfer_student=Y

Legacy Indicator

Median Income by Admission Zip Code

Admit Code

High School GPA

High School Percentile

High School Size

First Generation Indicator

Veteran Indicator

- |binned_credits=0|transfer_student=N

Average Credits Attempted per Term

High School GPA

First Generation Indicator

Admit Code

Credits Attempted Current Term

Median Income by Admission Zip Code

International Indicator

Age at First Term

Gender

Average Success Outcome of Students Declared in Same Major

- |binned_credits=120|transfer_student=N

Average Success Outcome of Students Declared in Same Major

Average Credits Attempted per Term

First Term Transfer Credits

Gender

Admit Code

Cumulative GPA

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

Number of Completed Terms

International Indicator

Ratio of Earned to Attempted Credits

- |binned_credits=120|transfer_student=Y

Average Credits Attempted per Term

Cumulative GPA

Average Success Outcome of Students Declared in Same Major

First Term Transfer Credits

Proportion of Transfer Credits

Trend in Term GPA

Ratio of Credits Attempted Current Term to Prior Term

High School Percentile

Number of Completed Terms

High School GPA

- |binned_credits=60|transfer_student=N

Average Credits Attempted per Term

Cumulative GPA

First Term Transfer Credits

Admit Code

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

Proportion of Transfer Credits

Ratio of Earned to Attempted Credits

High School GPA

First Generation Indicator

Gender

- |binned_credits=60|transfer_student=Y

Average Credits Attempted per Term

Cumulative GPA

Proportion of Transfer Credits

First Term Transfer Credits

Number of Completed Terms

Admit Code

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

High School GPA

Legacy Indicator

Ratio of Credits Attempted Current Term to Prior Term

- |binned_credits=Inf|transfer_student=N

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

    Average Success Outcome of Students Declared in Same Major

    Credits Attempted Current Term

    Admit Code

    Cumulative GPA

    Trend in Term GPA

    High School GPA

    International Indicator

    Ratio of Earned to Attempted Credits

    SAT/ACT Verbal Score Percentile
- |binned_credits=Inf|transfer_student=Y

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

    Average Success Outcome of Students Declared in Same Major

    Proportion of Transfer Credits

    Average Credits Attempted per Term

    Credits Attempted Current Term

    Cumulative GPA

    First Term Transfer Credits

    Trend in Term GPA

    Admit Code

    Legacy Indicator

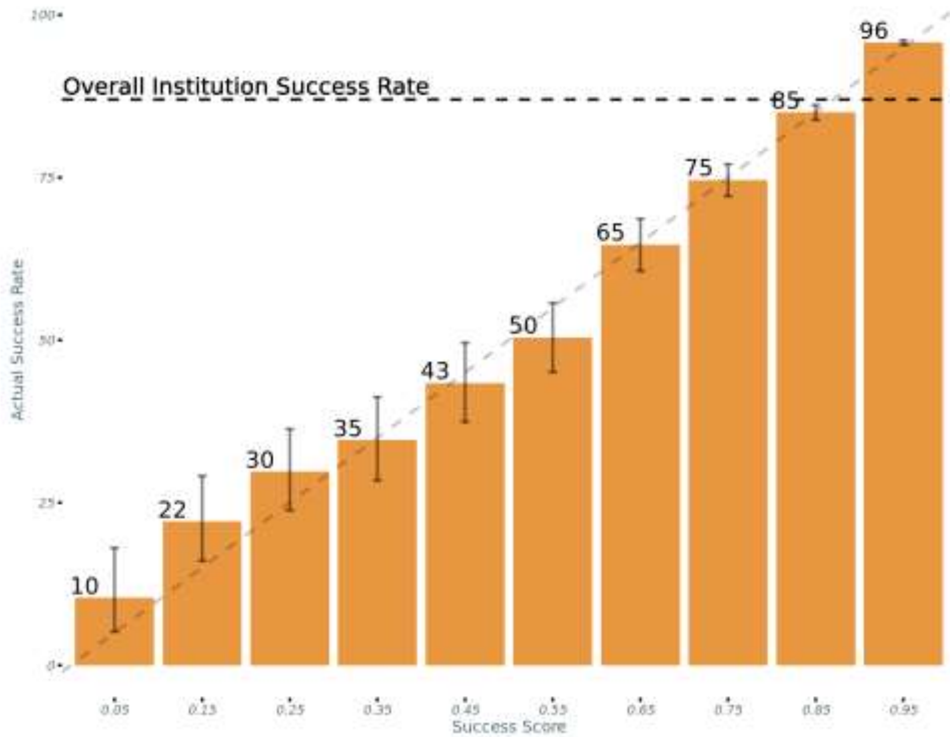
Model Performance

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.

Calibration

Calibration offers an intuitive way to evaluate a model by capturing how close its estimated probability scores are to reality.

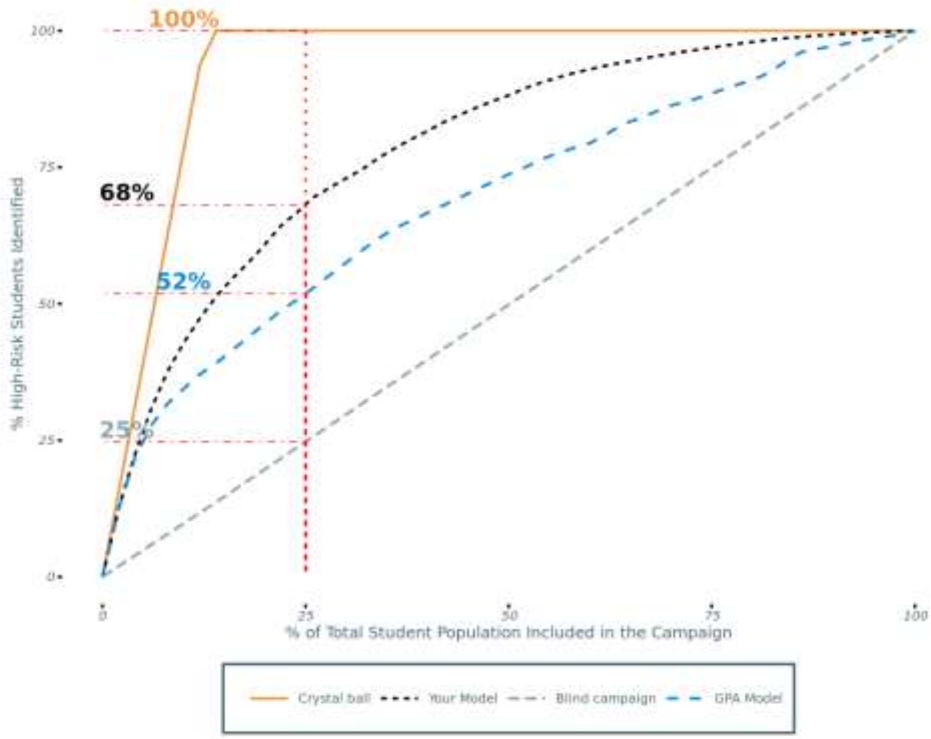
Students are divided into different bins along the horizontal axis according to their success score, while the vertical height of each bin indicates the actual persistence rate of the historical students it contains. The horizontal line shows the overall percentage of students that persisted to the next term.



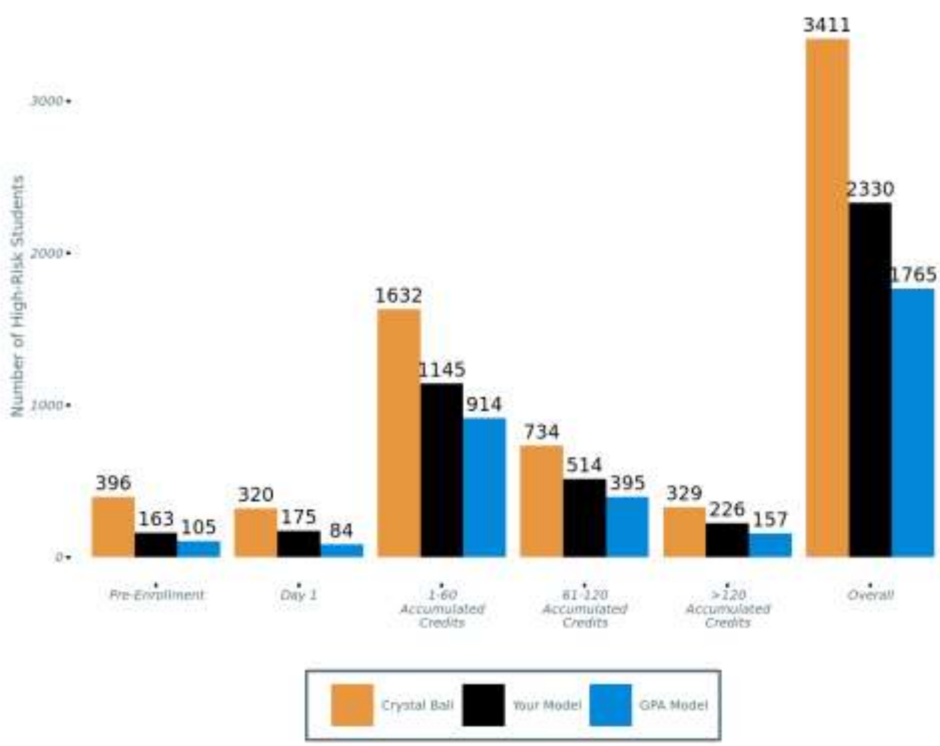
High-Risk Student Identification Rate

The SSPM enables you to rank students by order of risk (i.e., success scores from low to high) so that you can most efficiently target as many high-risk students for intervention as your institution or office/department can effectively handle. Let's assume, for instance, that you are designing an intervention campaign targeting high-risk students and have the capacity to advise N students. Let's assume you use different predictive models to generate lists of N targeted high-risk students, and step forward in time to compare their performance by evaluating the percentage of those N students that did not persist to the next term. This performance comparison is summarized in the high-risk student identification rate chart below, which shows the percentage of actual high-risk historical students (i.e., students that did not persist to the next term.) identified by the model vs. the percentage of the total student population targeted in the campaign. For example, if you design a campaign that includes 25% of your total student population, then the percentage of your school's high-risk students identified by the campaign will be 100, 68, 52, and 25 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

High-risk student identification rate provides a powerful and transparent performance benchmark of model performance; the large performance enhancement gained in going from the simple GPA Model to your advanced SSPM is clearly visible in the chart.



High-risk student identification rates can also be converted to actual numbers of students and compared across different accumulated credit subgroups, as shown in the figure below for campaigns targeting 25% of the total student population.

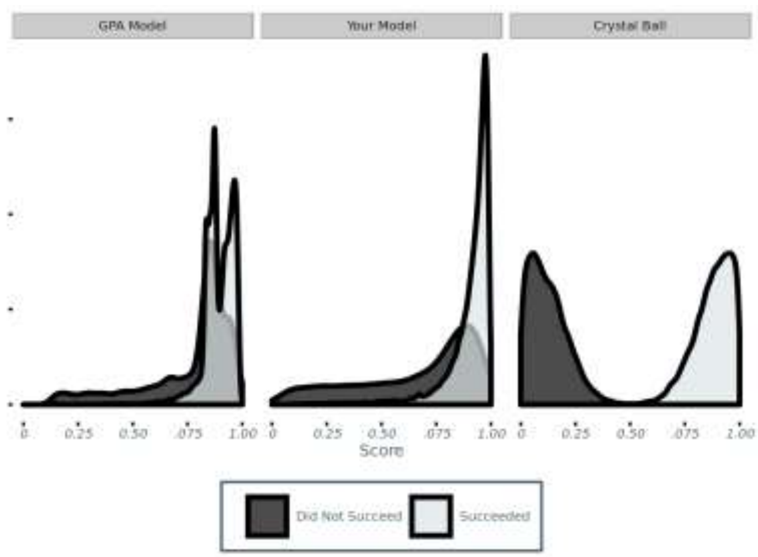


Lift

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 6.86, 2.73, 2.08, and 1.00 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

Separation

Displaying the distributions of success scores for students in the historical test set who did and did not persist to the next term also provides an intuitive sense of a model's performance. We see in the charts below that successful students (light gray) typically have higher scores than unsuccessful ones (dark gray) for both your SSPM and the GPA Model, as you would expect, but that your SSPM is much better at separating these two student populations from each other. That is, the graphic demonstrates that your SSPM ascribes high success scores to successful students and low success scores to unsuccessful students more accurately than the GPA Model. A perfect prediction would result in complete separation between the students (shown in the Crystal Ball chart on the right).



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 I: 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.

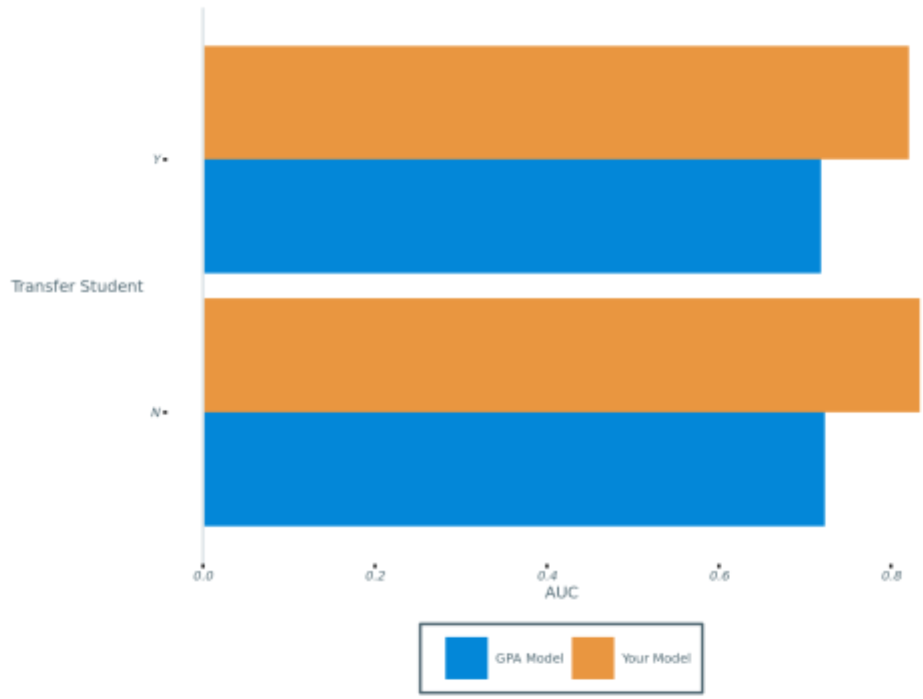
The table below shows AUC values for your SSPM and the GPA Model.

Model	AUC
GPA Model	0.72
Your Model	0.83

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

Accuracy for Transfer/Non-Transfer Students

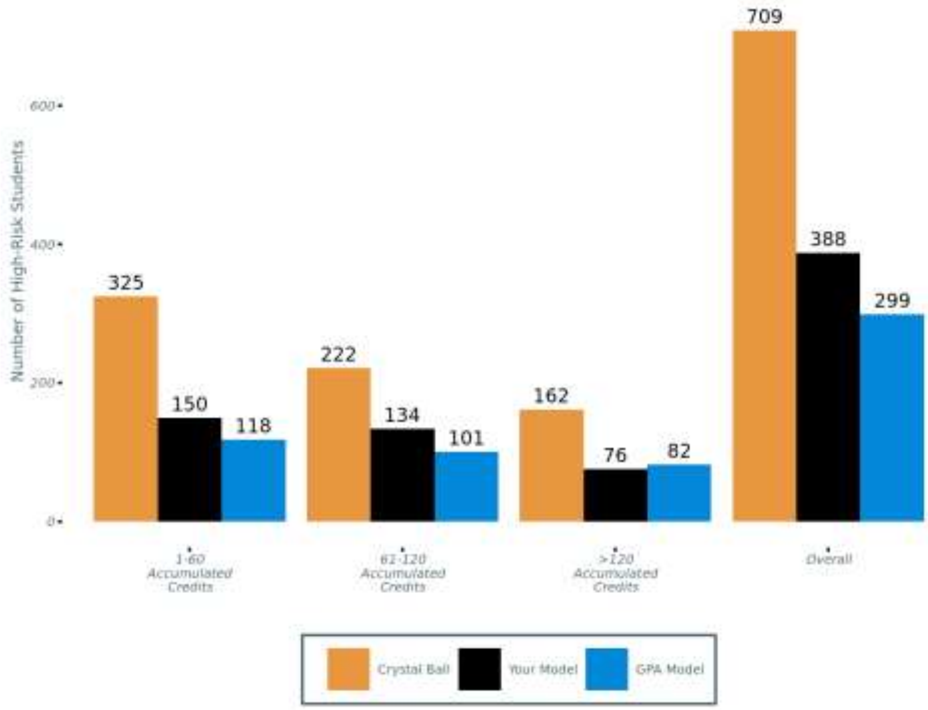


Appendix II: High-Risk Student Identification Rate for Murky Middle and Top Performing Students

Your Student Success Predictive Model's performance varies across different subgroups of students. This appendix 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 appendix for two student subgroups.

Murky Middle

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

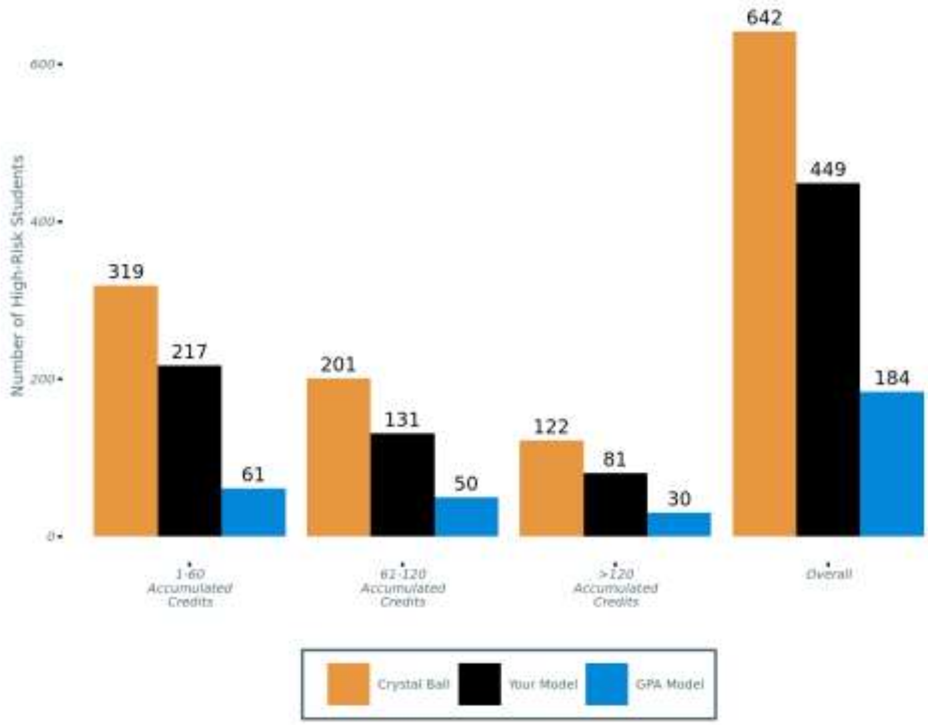


Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	43%	87%	100%	100%
Your Model	16%	27%	54%	80%
GPA Model	10%	21%	42%	70%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**60%**	**29%**	**29%**	**14%**

Top performing students

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



Summary of high-risk student identification rates vs. model.

Model	5%	10%	25%	50%
Crystal ball	77%	100%	100%	100%
Your Model	32%	47%	70%	87%
GPA Model	7%	14%	28%	53%
Blind campaign	5%	10%	25%	50%
Relative Percentage Gain	**357%**	**236%**	**150%**	**64%**

Appendix III – 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.
- First Generation Indicator: “Yes” or “No” indicator of whether any of an individual’s parents have ever earned a bachelor’s degree.
- 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.
- Legacy Indicator: “Yes” or “No” indicator of whether an individual is a legacy student.
- Median Income by Admission Zip Code: The median household income in the zip code of a student’s home at the time of their admission.
- 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.
- 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.
- Trend in Term GPA: A measure of the change over time in a student's term GPAs.
- Veteran Indicator: "Yes" or "No" indicator of whether a student is a veteran of the United States Armed Forces.

EAB | Student Success Collaborative™

Risk Model Report -- Texas Tech University

November 2, 2015

Introduction

This document provides information about the Student Success Collaborative (SSC) Risk Model, as customized and evaluated for your institution. It includes descriptions about the data used for the model, as well as metrics about the predictive power of the model, and details about the "skills" that are a subcomponent of the model.

The risk model's primary task is to predict the likelihood that a currently-enrolled student will eventually graduate. It does so by extracting information ("predictors") from the transcript and other data about each current student, and pushing the result through a complex statistical model that was previously calibrated on historical student data from your institution.

The result is an estimate, a score between 0.0 and 1.0. A score of 1.0 does not mean that the student is guaranteed to graduate, nor does a score of 0.0 mean that there is no hope. The model is not a crystal ball. A good way to interpret the score is as a probability -- a score of 0.5 means that if you had ten very similar students, the model would estimate that about five of them would graduate.

It is worth reiterating that the model uses the broadest possible definition of success -- graduation, in any major, at any point in time. Other definitions of success, such as 6-year graduation, or graduation in (at least one of) current major(s), or persistence, could be used in the future. Note that the fact that the student has a particular major or majors is used by the predictive model, but the student is still considered successful even if they change majors.

It's also worth noting the relationship between risk score, risk color (Red, Yellow, Green), and accuracy. Risk score is a probability, and ranges between 0 and 1. If you draw a line at 0.5, and say that a student above the line is predicted to graduate, and they do, then that is an accurate prediction. But for the purposes of identifying students at risk, the 0.5 theshold may not be most useful. Therefore, the SSC application uses risk color, which draws threshold at other values. By default, less than 0.33 is Red, between 0.33 and 0.67 is Yellow, and above 0.67 is Green, but other boundaries may make more sense for your institution, your students, and your advisors. See the "Thresholds" section, below.

Model Description

The Student Success Collaborative Risk Model uses best practices in data science to estimate graduation likelihood for each student, from incoming Freshmen to nearly-graduating Seniors. A customized set of "predictors", or facts derived from the students' academic and other records, are combined and weighted using an automated learning process. EAB's Data Science team optimizes this process for each member, and uses a variety of validation tools to ensure the best possible performance given the data available.

Technical details: The model is a combination of several penalized logistic regression models that each focus on different subgroups of students. Predictive features include simple lookups of facts about a student, as well as processed or derived facts whose details are proprietary.

Data

To build the model, we used transcript and other records from historical students provided by your institution. The following students were used:

- Started between 2006-08-21 and 2009-05-08.
- Had at least one registered term.
- Were seeking a baccalaureate degree.
- Transfer students were included.

Within the student records, some information was excluded. Courses without grades or with grades that were not mapped to a standard set of grades were excluded from the analysis. Unless you specified otherwise, grades from transfer courses were excluded during model creation.

Process

The SSC Risk Model is fit initially when your institution joins, and annually thereafter, as additional graduation data becomes available. Occasionally the model may be updated off-cycle, to address data or other changes.

We use industry-standard processes for ensuring a high-quality statistical model, including separating historical data into separate "training" and "test" splits. The model learns from the training data, and all results reported below were from the test set. The test set is a random 20% of the total data. We expect that the students in the test set will be similar to the current set of students at your institution, so that the metrics we use below should be similar to what you will see in practice.

Your Institution's Model

The EAB Data Science team explores a variety of ways of describing students, based on the data your institution provided, to build the best model for your institution. The data points below are those that were found to be predictive and useful. The data in this list are not equally important; the statistical model learns how to identify the most predictive types of data for each sub-group of student and makes the best use of it. For instance, some facts about a student are highly relevant for Freshman, but minimally important for Seniors.

- Attempted Credits Trend
- Completed Terms
- Credits Since Last Major Change
- Cumulative GPA
- Current Major Frequency

- D-F-W Counts
- D-F-W Trend
- Estimated Skills
- First Generation Student Indicator
- GPA Trend
- Grade Variance
- International Student Indicator
- Lifetime Attempted Credits
- Major(s)-Skills Alignment
- Number of Current Majors
- Number of Former Majors
- Readmitted Student Indicator
- Standardized SAT/ACT Exam Scores
- Student Ethnicity
- Student Gender
- Term GPA
- Transfer Credit Proportion
- Transfer Student Indicator
- Veteran Indicator

Skills

An important component of the SSC risk model is the process of identifying underlying patterns in the grades that students earn in courses. Although students receive higher or lower grades for many reasons, there is some structure to the pattern. A student who earned a high grade in one mathematics-intensive class might be expected to earn a high grade in another mathematics-intensive class; a student who received a low grade in one class requiring writing research reports might be expected to receive a low grade in another class with similar requirements. We call these underlying factors "Skills," and use sophisticated mathematical techniques to identify the skills and estimate them for current students. Skills are neither destiny nor fixed over time, but they provide additional information not available elsewhere.

We use the alignment between estimated skills and the major or majors that a student has declared as part of the risk model. Research has shown that this alignment, between a student's skills and the major that they choose, has a substantial correlation with their eventual success. In addition to using this alignment internally by the model, the alignment score is displayed as the Risk Score Analysis in the SSC platform.

The number of skills in your model is chosen by the EAB Data Science team during model training, and typically ranges between two and seven. The number that best helps the overall risk model predict outcomes for students was used, even if other numbers of skills may be more interpretable.

Skills are based on grades in courses. Here are, for each skill, the courses that are most strongly associated with primarily that skill. SSC Consultants will work with your institution to identify interpretable names and descriptions for the commonalities among these courses. These namesand descriptions will be displayed in the SSC application. Note that there may be unexpected courses in these lists -- this may happen if there are relatively few students in a course, or if the grading pattern was historically unusual.

skill_01

MATH2360, MATH3350, ME2322, CHEM3305, ZOOL2403, PHYS2401, IE3301, MATH3342, MATH2350, ZOOL2404, CHEM1308, HIST3310, PSS1411, PHYS1403, POLS2302, ART1309, HIST2301, MCOM3320, ENGL2305, HIST2300, NS1325, ANTH1301, ASTR1400, ESS1301, POLS1301, THA2304, GEOG2351, PHYS1408, HDFS2303, ENGL1302, BIOL1402, MCOM3300, ENGL2307, PFP3301, MATH1352, ADV3310, SOC1301, ECO2301, GEOL1303

skill_02

IS1100, PFW1117, PFW1114, MUHL3310, CHEM1105, GEOL1101, EDIT2318, ENGL2351, COMS3358, CFAS2300, ADRS2310, THA2304, CHEM1107, ENGL2311, ENGL1301, CHEM1108, COMS2300, HDFS3301, SOC1301, ENGL1302, HDFS2303, MUHL1308, ENGL2305, ESS1301, PSY1300, MCOM1300, ENGL2307, ASTR1400, SPAN1507, ATMO1100, ART1309, PSS1411, SPAN2302, GEOG2351, HIST2301, SPAN2301, MATH1320, HIST2300

skill_03

CHEM1301, MATH1351, MATH1321, MATH1550, MATH1330, CHEM1307, MATH2300, MATH1352, MATH1320, SPAN2301, SPAN2302, MATH1300, ECO2305, PHYS1408, ECO2301, SPAN1507, CHEM1305, ECO2302, ATMO1300, CHEM1107, PHYS1403, BIOL1402, GEOL1303, ADV3310, PFP3301, POLS1301, CHEM1105, MCOM3300

Thresholds

As noted above, different score thresholds may make more or less sense for different institutions, depending on your population, how you use the SSC platform, and how you think about and talk about risk. Here are four options that may be considered, but other thresholds are possible. Note that score is likelihood x 100.

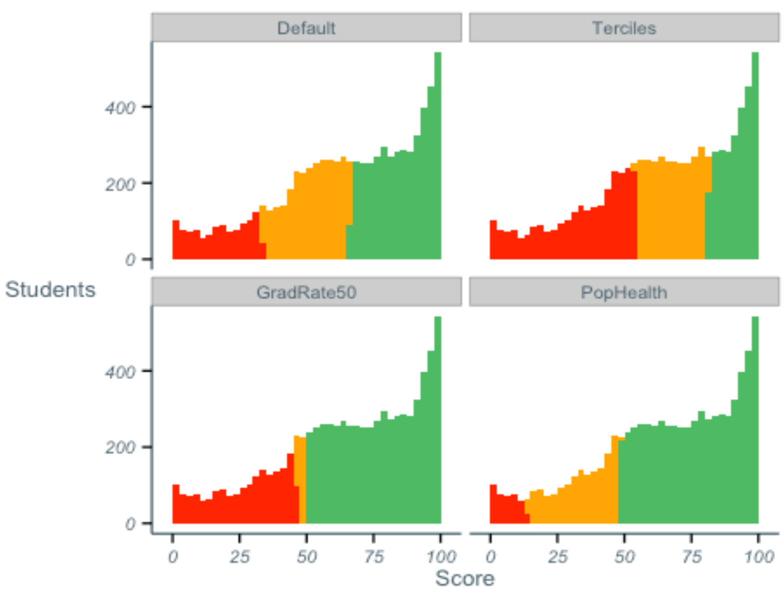
1. *Default*: 33 and 67, dividing the scores into thirds by likelihood.
2. *Terciles*: Divide historical students into three equally sized groups, based on their score ranks, and use the resulting thresholds: 55, 81, for your institution. Use this option to see about the same number of Red, Yellow, and Green students in the platform.
3. *Grad Rate and 50%*: Use thresholds based on your overall institution graduation rate, and 50%: 46, 50, for your institution. Provides intuitive distinctions for institutions with graduation rates substantially higher or lower than 50%.

4. *Population Health*: Use thresholds determined as the 5th and 25th percentiles of historical students, as is frequently done in healthcare settings: 13, 48, for your institution. Valuable for institutions focused on students at highest risk.

Model Quality

We use a variety of metrics and comparisons to understand and tune the performance of your model overall, compared to a baseline model, and for various subpopulations. This section describes the results of your model when evaluated on the test set.

We use several commonly-used metrics. *Kappa* is a corrected version of accuracy that takes chance likelihood into account -- it is typically called Cohen's Kappa in the academic literature. If your institution had a 75% graduation rate, and you predicted everyone would graduate, you would be 75% accurate, but Kappa would be 0. If you had a crystal ball and could perfectly predict the future, Kappa would be 1. A common rule of thumb used by researchers is that a Kappa above 0.4 is considered fair to good; for the social sciences, including educational outcomes research, a Kappa much above 0.6 would be extremely rare. The predictiveness of SSC risk models vary between institutions and depends on your student population and the data that is available.



AUC is a related measure used extensively in data science, which ranges from 0.5 (chance) to 1.0 (crystal ball). AUC is better able to capture a model's ability to rank students, but does not account for calibration, which is described below.

Your model is compared to a *baseline* model that only uses accumulated credits and cumulative GPA as predictors. This is the type of rule-of-thumb based approach that academic advisors intuitively know is useful. Additional accuracy over the baseline model indicates that SSC's risk model provides more information and guidance for the task of identifying students at risk of not graduating.

We also compare model accuracy for students with varying numbers of accumulated credits and in several subgroups. It is important that a predictive model be able to provide value across the years that a student is at your institution, and broadly across the student population.

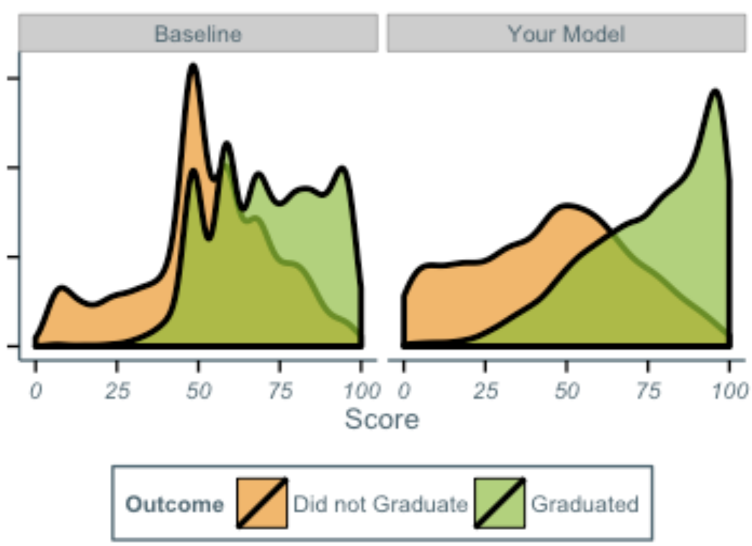
Overall Performance

Model	Kappa	AUC
Baseline	0.359	0.751
Your Model	0.453	0.825

A useful analysis is to rank the test set of students by predicted graduation likelihood, then calculate the average *predicted* and *actual* graduation rates for each group. An *accurate* model will have widely different predicted graduation rates for each group (read vertically in the table below). A *well-calibrated* model will have similar predicted and actual graduation rates for each group (read horizontally in the table below).

Model	Risk Score	Pred. Grad Rate	Actual Grad Rate
Baseline	Bottom Quartile	0.39	0.36
Baseline	Second Quartile	0.58	0.64
Baseline	Third Quartile	0.72	0.73
Baseline	Top Quartile	0.90	0.87
Your Model	Bottom Quartile	0.28	0.27
Your Model	Second Quartile	0.58	0.58
Your Model	Third Quartile	0.78	0.79
Your Model	Top Quartile	0.94	0.94

A graphical view of predictive performance is to plot the distribution of prediction scores, for historical test-set students who did and did not graduate, for different models. A more powerful model will pull more students away from the center of the graph, towards the left (confident will not graduate) and right (confident will graduate) edges. This graph uses density plots, which can be thought of as smoothed histograms.



Lift Over Baseline

Another useful metric is Lift. Lift is frequently used in sales and advertising analytics, and is closely tied to a common use-case for SSC -- identifying students at highest risk of failing to eventually graduate. Consider a scenario where you want to reach out to 25% of your students who are at highest risk of failing to graduate. If you reach out randomly, the proportion of students who will eventually not graduate will be equal to your campus graduation rate. If you use our baseline model, which incorporates only GPA and accumulated credits to identify students, you will reach a higher proportion. The metric we provide is how much more lift, or how many more students your model identifies in this scenario (cutoff at 25%), compared to the baseline model: *(your model's negative predictive value) / (baseline negative predictive value) - 1*

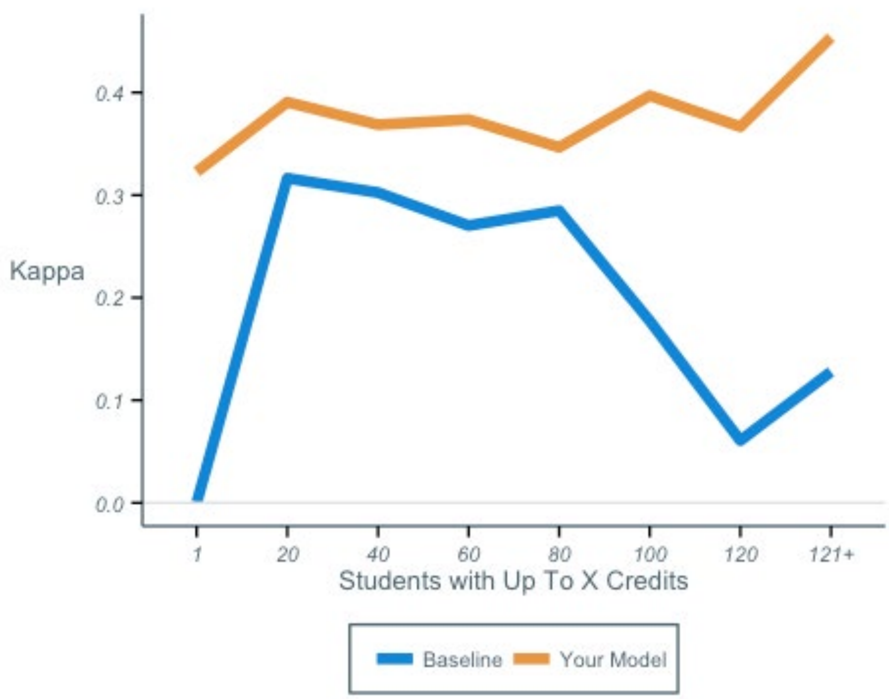
For your model, Lift Over Baseline is:

$$\frac{0.73}{0.649} - 1 = 12.4\%$$

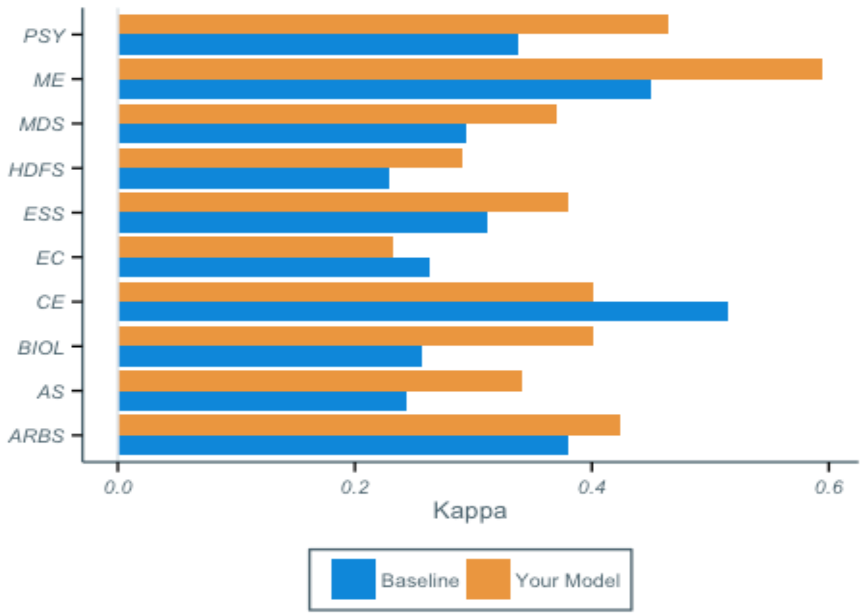
Subgroups

During validation, we examine subgroups of students to ensure that your model is consistently valuable. Note that you may see a few instances where your model may not exceed the baseline model's accuracy. This is expected, and depends on the number and properties of students in each subgroup.

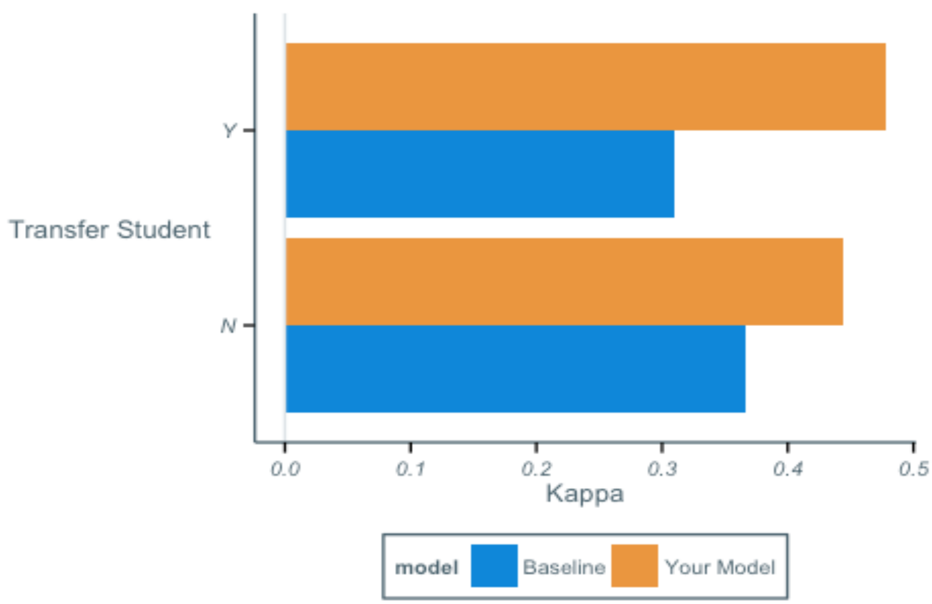
Accuracy for Students with Accumulating Credits



Accuracy for Students Who Declared a Top 10 Most Enrolled Major



Accuracy for Transfer/Non-Transfer Students



Conclusion

Based on these results and our extensive validation process, we are confident that this model will help your advisors productively identify students to target for proactive outreach, and otherwise prioritize their interactions with current students. We look forward to hearing your feedback, suggestions, and success stories.