Supporting Student Success: First Year Retention Modeling

Rapid Insight®
data driven decisions™

Why Student Success Matters

A college's retention rate is a widely available metric that prospective students can use to evaluate or compare colleges

A more successful student body stabilizes a school's yearly revenue

More students are borrowing more money to attend college, and default rates on student loans are rising, with adverse impacts to the institution

Many states are implementing performance-based funding formulas which rely heavily on graduation and student success rates

Students attending college are seeking success. College graduates have a higher employment rate, earn more, and are more satisfied at work than non-graduates.

Successful, graduating students make for more successful, generous alumni

Achieving Student Success

There are many people on campus who can reach out to struggling students to help them to be successful, including instructors, advisors, deans, and academic support personnel.

Retention models enable you to use your limited time and resources to decide which students are most in need of intervention.

So much data, so little insight

Using historical data about students who have been successful and students who have not, predictive modeling can help to prioritize which students are most at-risk.

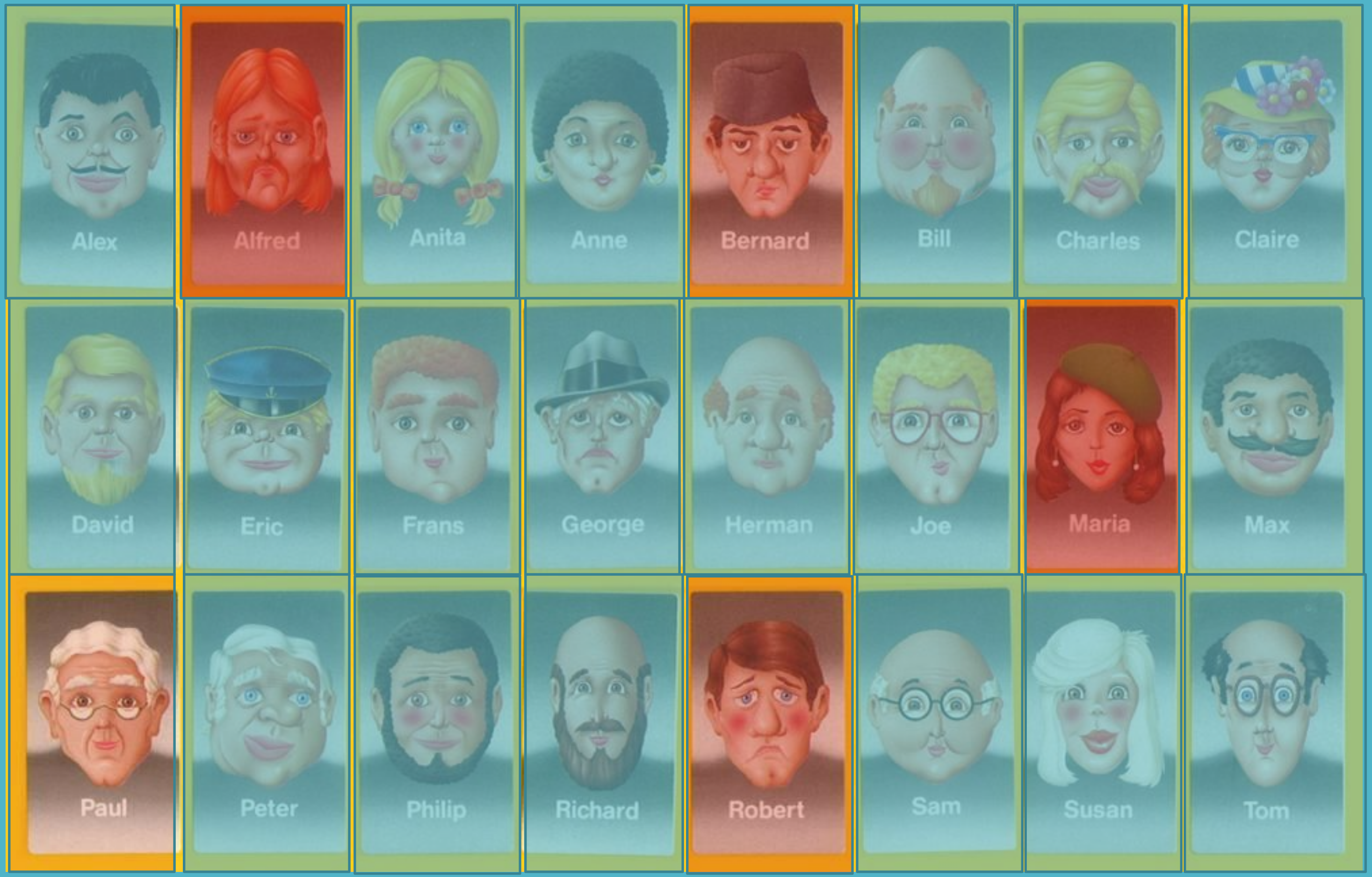
No student outreach is a waste of time, but knowing which students are most at-risk allows you to distribute your resources efficiently by focusing on the students who need them the most.

What is Predictive Modeling?

Predictive modeling is the process of applying statistical techniques to your historical data to predict future events or behaviors.

A predictive model is a mathematical equation that weights several predictor variables (x's) to predict a given outcome (y). The weights (coefficients) are determined by a modeling algorithm, either by manually programming or loading your data into modeling software. Once your formula is assembled, you'll be able to apply the equation to incoming students to predict their future behavior.

How Predictive Modeling Can Help with Retention?



At-Risk Student List

Name	Attrition Probability
Alfred	87%
Bernard	68%
Maria	79%
Paul	45%
Robert	53%

A model will identify students who are considered at-risk in order to intervene as early as possible.

Types of First Year Student Success Models

You can predict:

- Likelihood of attrition
- GPA

For:

- End of first term
- End of first year

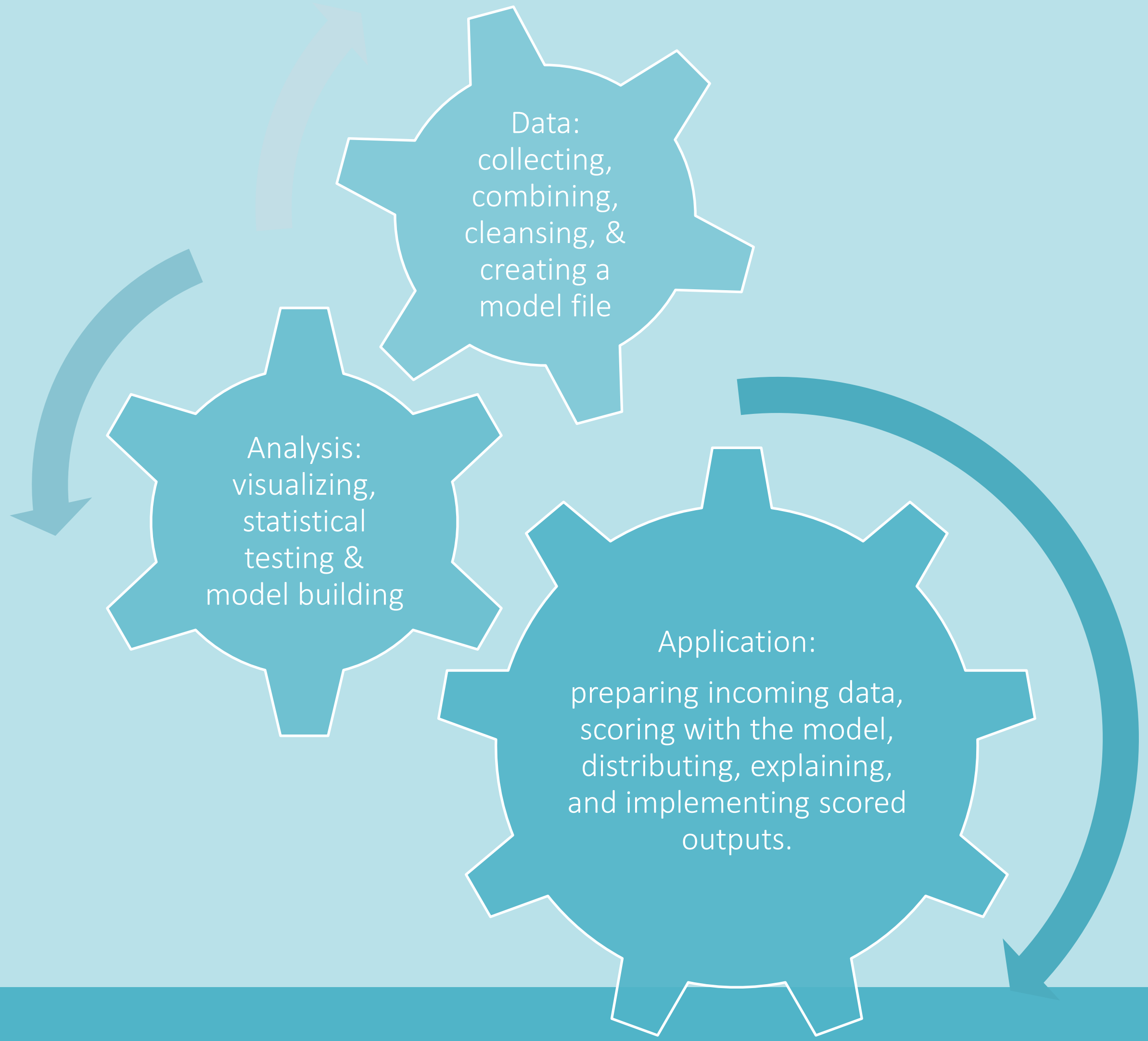
Which model is best for you?

To decide which model is right for you, think about which students you'd like to focus on and how you'd like to use the results.

For example, if you'd like to focus on retaining students who are not likely to return for a sophomore year, you might choose the "fall to fall" attrition model. If you want to focus on students who are academically underperforming, building a first semester GPA model and comparing the predictions to a student's actual end-of-semester GPA might be a better fit.

Predictive modeling breaks down into three pieces:

1. the data
2. the modeling and analysis
3. the scoring and application



The Modeling Process

What data should I use?

Most of the data used for freshmen modeling comes from the student application itself. This data includes (but is not limited to):

- Whether the student inquired before applying
- Gender
- State (and/or in-state flag)
- County
- ZIP (used to calculate distance from campus)
- International student flag
- High school code (used to calculate # of applicants and # enrolled from each person's HS)
- HS type
- Standardized test scores (SAT Math/Verbal, ACT)

- Application date
- Age or birthdate
- Application type (online app, paper app, etc.)
- Program (or major) applied for
- Legacy student flag
- Visited campus flag
- Applied for FAFSA flag
- Expected financial contribution
- Financial aid offered
- Scholarship amount offered

Pre-Modeling Discussions

Before you begin, make sure that you and your team have a clear understanding of:

1) What you want to predict

2) How you will use the results

3) At what point you'd like to use the predictions (for example, after the first semester or before a student gets to campus)

Why does it matter?

These three ideas will inform the rest of the modeling process. For example, if you'd like to know which students are at-risk of attrition before their sophomore year (1) before they step foot on campus (3) to intervene right away (2), that model might use different information than one that predicts attrition at the end of the first semester. In the latter model, you'll have much more college-level data available for each student, whereas the former model would be mostly based on high school performance.

Data preparation

You might be surprised to learn that 80% of the process can be described as data preparation.

Data preparation is a general term which encompasses many actions. This includes merging together data from multiple files, tables, and sources across multiple years, creating your Y variable, ensuring that codes remain consistent, that missing values are handled properly, and ensuring that your historical dataset is a good representation of your current student population.

This is a big step, and often it takes an analyst a few tries to get it right.

The Y-Variable

The y-variable is what you'd like to predict (in this case, attrition or GPA), but it's important to be a bit more discerning…

Defining "retained" for first-year students as a student who persists to sophomore year is a great start, but you'll still need to clarify further. For example, is a student who returns but switches from full time to part time still considered "retained"? These clarifications usually circle back to how you plan to use the model. If you'll be using your model to forecast the number of freshmen who enroll as sophomores for revenue purposes, you might answer the earlier question differently than if you'll be using the model to ensure student success by flagging at-risk students.

The model file contains the historical data you'll use to make your predictions, and is the file you'll focus on while building a model.

Guidelines for building your modeling file:

- Three years of historical data is a good starting point
- Contains both students who were retained and students who were not
- Historical data chosen is an accurate representation of your current student profile

The Model File

How to build a model

The "how" of building a predictive model will vary widely depending on the tool you're utilizing to build it. These tools can range from manually programming exploratory analyses and performing statistical tests to almost fully automated predictive modeling.

Building Your Model

Once your modeling dataset has been assembled and prepared, it's time to statistically analyze it and build a predictive model.

1. The solution that is best for you will depend on your technical background and level of comfort with statistics. Find variables that are statistically related to your y-variable.
2. Create the model using this predictive subset of variables.
3. Validate your model on a holdout sample and holdout years.

Applying the Model

After building the model, you can apply it to score your students. Keep in mind:

- The student data that you apply the model to should be prepared exactly the same as your modeling data, including any data cleanup operations.
- You will get one score (either predicted GPA or probability of attrition, depending on the model you chose) for each freshman.

The raw output

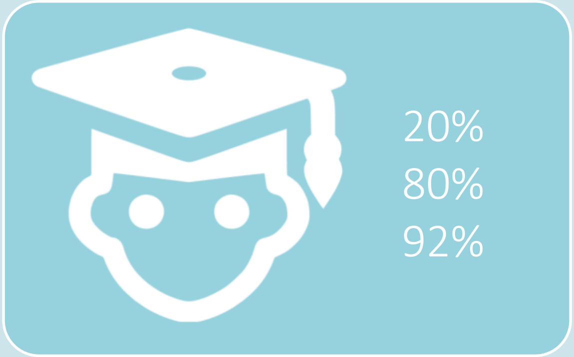
When modeling attrition likelihood, you might get scores like .09, .67, or .38, which would mean that a student is 9%, 67%, or 38% likely to attrit, respectively.

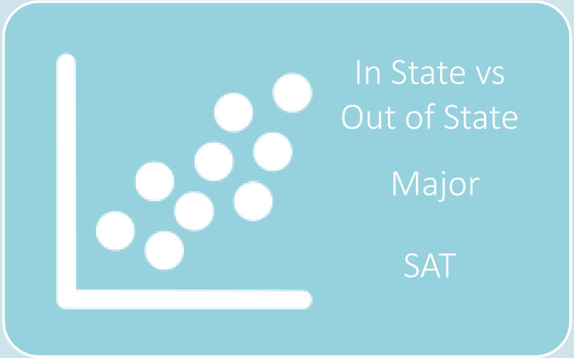
When predicting GPA, you might see scores like 3.82, 2.03, or 3.29, which would mean that student's predicted GPA would be equal to the outputted number.

Scored Students

Name	Attrition Probability
Alfred	87%
Bernard	68%
Maria	79%
Paul	45%
Robert	53%

Insights Gained





Individual students

Using the raw scores, you have an individualized probability of attrition for each student you choose to score with the model.

Overall predictions

By summing the probabilities, you get total predicted enrollment. If you sort these sums by gender, program, etc., you get the projected enrollment for each characteristic.

Additional findings

The analysis will help to understand which factors are closely tied to student retention.

Implementing Results: UNC Greensboro

The University of North Carolina at Greensboro enrolls about 2,600+ new students per year and knew that they didn't have the resources for their advisors to call all of these students, so they turned to predictive modeling as a way to prioritize at-risk students. They've also based the type of intervention a student receives (call from academic advisor vs. call from faculty member) on a how likely a student is to attrit.

By organizing and prioritizing their retention efforts, UNCG has realized a 7% growth in retention rates over the past two years.

UNCG

Growth in retention rate of 7%

Implementing Results: Dickinson College

Dickinson College is a small liberal arts college located in Carlisle, PA. Their retention rate was hovering around a fairly steady 90%, and the task of increasing this rate proved difficult.

They decided to use predictive modeling to identify students who were not meeting their expected potential by predicting first term GPA and comparing it to actual GPA to see which students were underperforming. They also involved multiple offices across the system, and their end result was a team well-versed in each component of the retention initiative, a proactive approach, and an increase in retention rate beyond 90%.

Increased retention rate beyond **90%**

Implementing Results: Ball State University

Ball State University lost 700 students over 5 years and saw retention dip below 75%.

Led by Student Affairs, the university added in retention specialists and began to look at new and creative ways they could analyze student data to help address the problem. By bringing their predictive modeling retention efforts in house, they are now incorporating over 250 variables into their models with up to date data.

Through all of their campus-wide initiatives, Ball State has improved their retention rate by over 6 points.

They are also experiencing the largest five-year increase in on-time graduation rates of any public institution of higher education in the state of Indiana.

Increased retention rate by 6 points

ROI for Students/Schools

Using a data-driven approach can achieve significant increases in the number of students retained, as well as a hefty return on investment.

Paul Smith’s College in upstate NY has an overall enrollment of about 1,000 students. Over a five year period, they observed an ROI in excess of $6,000,000 through their data-driven retention efforts.

Paul Smith’s
COLLEGE

Calculating ROI

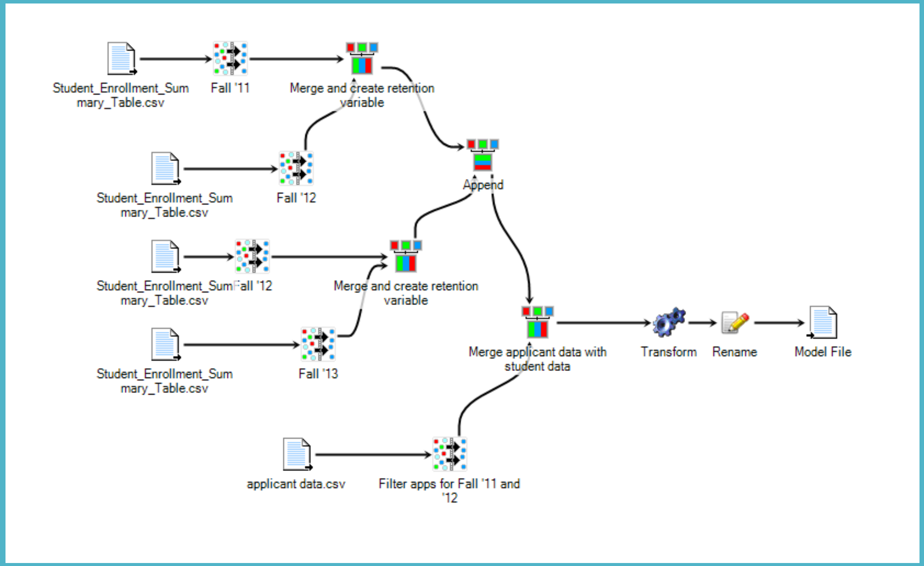
In order to demonstrate that predictive modeling is providing value, it is possible to calculate the ROI of a given model.

The ROI is equal to the amount of tuition saved by retaining a student who might not otherwise have been retained.

So, if your modeling efforts increase your retention numbers by 10 students who each pay $40,000 in tuition, your ROI for the model is equal to $400,000.

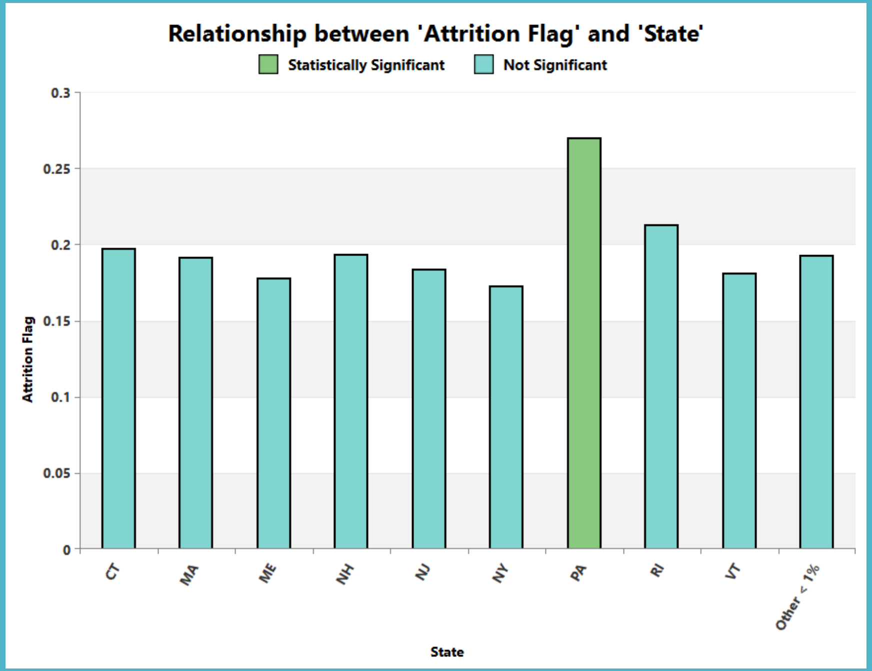
Rapid Insight: Tools for Student Success

Veera



Data extraction, blending, transformation, and reporting without any coding or programming

Analytics



Automated data analysis and predictive modeling without a PhD

Additional Resources

For additional information related to data, analytics, and student success:

Case Study: Ball State University – Boosting Retention with Predictive Analytics

Free Online Predictive Modeling Education Videos

Request Demo of Rapid Insight Software

Rapid Insight® data driven decisions™