Predict 101 for Enrollment Management

The Promise of Prediction and Advanced Analytics

August 2018

element451
intelligent admissions

a product of SPARK451

A Promise Not Yet Realized

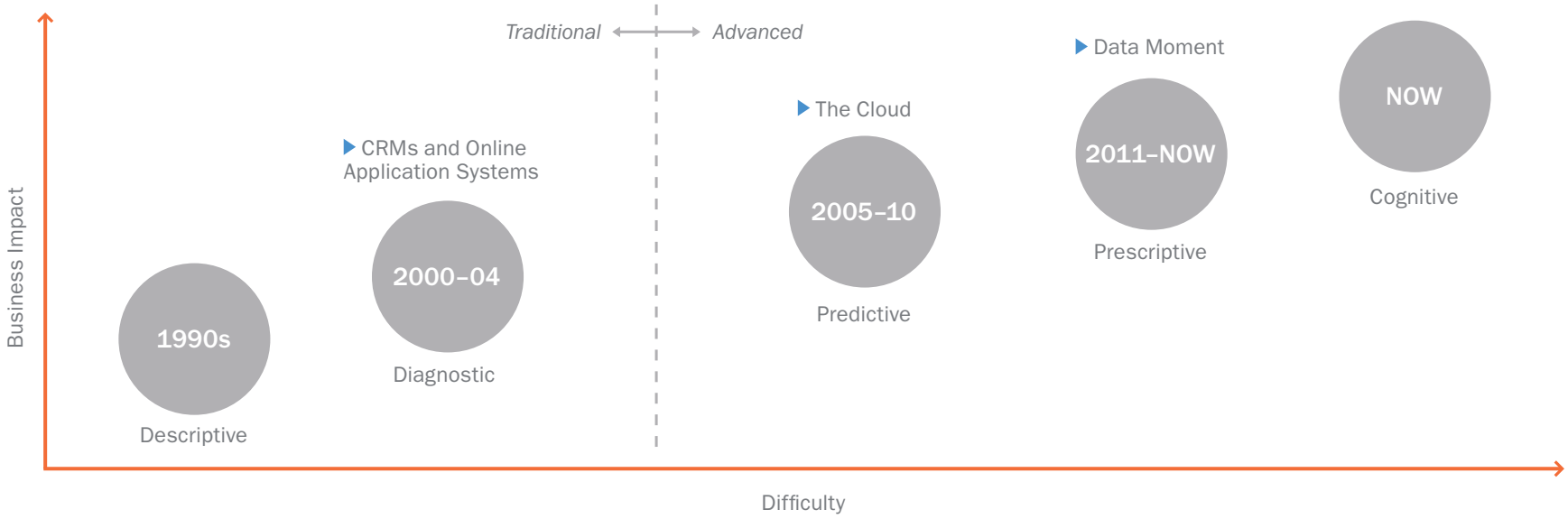
Cautionary news reports and enthusiastic companies alike remind us daily that predictive models can learn from data, identify patterns, and make decisions with minimal human intervention. The promise is big. Data and advanced analytics will unlock the mystery of what students want. Operations will be more efficient. And ultimately, recruitment will increase. The technology behind this promise is presented as machine learning (ML) or data science (DS), and more recently as artificial intelligence (AI).

Within enrollment management and student recruitment, the promise has not fully materialized. CRMs like Hobsons and online applications systems like ApplyYourself helped advance higher-education marketing. But analytics in higher ed hasn't kept up with strides in big data, machine learning, and cloud-based solutions. Instead, many admissions professionals still rely on intuition for their search and yield campaigns. They also use best-case recommendations and repeat tactics that have worked in the past.

> "Data and advanced analytics will unlock the mystery of what customers want. Operations will be more efficient. And ultimately, **profits will increase**."

A Brief History of the Evolution of Analytics

Let's take a look at how analytics has matured. The chart shows that as time progresses, the advances in business impact and analytics complexity also moves forward. And for a good portion of the evolution, higher ed admissions and recruitment kept pace.



We start with descriptive analytics and exploratory data analysis, which fully emerged in the 1990s. In this decade, "the recruitment funnel" and "funnel analysis" emerge as the techniques to mine the number of prospects, applicants, admits, and enrollments in a recruitment cycle. For admissions teams, the analytical capability during these years centers around the ability to forecast at the top or bottom of the funnel; "yield" becomes a common term. We begin to see examples of massive infrastructure to safeguard data. Big, private-sector players can support data centers with global footprints, whereas universities use local and modest IT infrastructure.

There is no doubt that the early 2000s were significant for all higher education professionals. Admissions teams begin using CRMs and online applications systems. Essentially, we go online! With these technologies, organizations capture lots of transactional data from students that permit ratio analysis, such as "open rates," as well as survival and retention analysis. By this time, the higher education sector is keeping up with other industries in its ability to analyze and understand student churn and audience segmentation.

But a split occurs between 2005–10. Aided by cloud-based software that reduces computational costs, along with mobile devices, private sector companies move into predictive analytics. Those that have successfully stored customer data over time are now able to estimate the likelihood of an action. Will Customer A buy this product? What is the probability that Customer B will churn?

Companies also begin to understand the payoff in interpreting behavioral data. This spurs businesses that collect and secure user behavior across digital platforms.

The divide picks up even more with the massive expansion of unstructured data from social media, enabled by wide adoption of smartphones and other mobile devices. Organizations begin to detect hidden patterns in text, behavior, and customers' attitudes. They realize that real-time analysis of multiple data sources is critical to be competitive.

There are some early movers within student recruitment that learn the "likeliness to apply" models that predict the probability of a prospective student submitting an application. But overall, the cost, lack of technical expertise, and data-shy culture impede higher ed from adopting predictive analytics.

That brings us to today. Organizations can predict anything and everything. Moreover, AI is blurring the line between human and machines. Autonomous vehicles, automated appliances, chatbots, and more are becoming commonplace.

Universities, on the other hand, lag somewhat behind, but are increasingly seeing data as a strategic asset, and analytics its natural complement. More admissions professionals are turning to machine learning and AI to improve recruitment practices. New admissions and recruitment platforms have machine learning technology embedded in them. Slowly and steadily, we're transitioning from an intuition-driven practice to a data-driven one. But the reality remains: there's a 10-year gap between the private sector and higher ed when it comes to analytics.

"More admissions professionals are turning to machine learning and AI to **improve recruitment practices**."

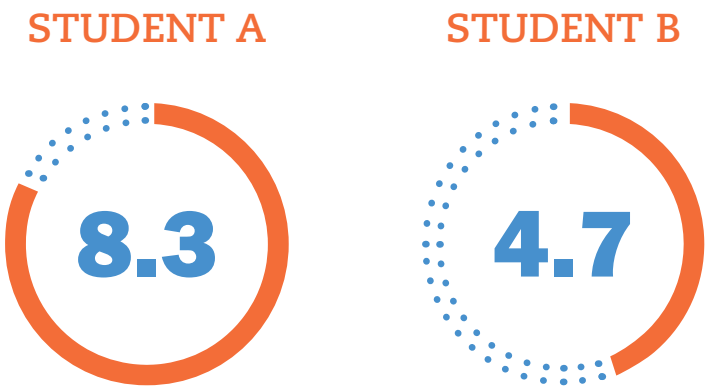
PART 1

Understanding How Predictive Modeling Works

To understand how machine learning helps colleges and universities predict what prospects and admits will do, we'll first look at a scenario based on one of our clients, a small liberal arts school in the Northeast.

An admissions committee offered an applicant entrance to the university, along with a robust financial aid package. Typically, what happens next is waiting and hoping for a deposit rather than an "offer declined" notice. With machine learning, we convert that waiting period into a game of probability, where admissions staff receive a number associated with the admitted student representing the likelihood he will accept the admissions offer and submit a deposit.

In this example, Student A is given a score of 8.3 (out of 10) and Student B, a 4.7. Intuitively, we can say that Student A has a higher probability to deposit than Student B. Specifically, we can say that the chance of depositing from Student A is very high, almost twice as high compared to Student B. In practice, and depending on the context of a school, an admissions officer could choose to move resources to the student with the lower score. For example, a one-to-one conversation with a faculty member or an on-campus event that attempts to persuade Student B to commit to the school.



Making the Model

With that scenario in mind, let's look at how a predictive model is built to arrive at a score. Note that in this paper, we are focusing on a specific type of technique called supervised learning, which trains a model on known input and output data so that it can predict future outputs. Another related technique called unsupervised learning finds hidden patterns or natural structures in input data.

To build a model, we need to:

1. Access historical data (e.g. demographic and behavioral information such as inquires, event attendance, email clicks)
2. Split the data into two groups: training and test
3. Train a model using the training data set
4. Validate and evaluate the performance of the model using the test data set
5. Tune the model—the traditional rinse and repeat

The results from a machine learning algorithm could be in the form of classification or regression predictive models. Classification techniques predict discrete response—for example, whether a student starts an application or not. Regression techniques predict continuous responses, such as changes in the scholarship package or financial aid in a new student class.

Techniques and Approaches

Classification Technique

predicts discrete response, such as when an application is submitted

Regression Technique

predicts continuous responses, such as changes in the scholarship package or financial aid in a new student class

Supervised Learning

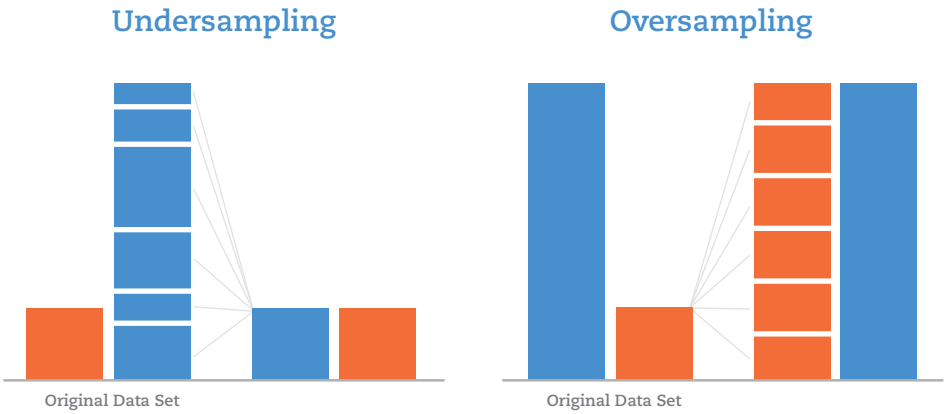
trains a model on known input and output data, so that it can predict future outputs

Unsupervised Learning

finds hidden patterns or natural structures in input data

About Historical Data

In our case, we start by selecting data and, in an ideal situation, three years of recruitment cycles. The data includes variables such as ZIP code, gender, age, family income, intended program of study, GPA, and academic load. Because our objective is to model the likelihood that someone will deposit, an essential requirement for choosing the right data is that we aim for a 50/50 representation of students who deposited and those who didn't. In machine learning, we call this positive vs. negative data. This representation of "positive" vs. "negative" deposits is important because the model will learn which condition takes place with the students who have instances of "positive" deposits.



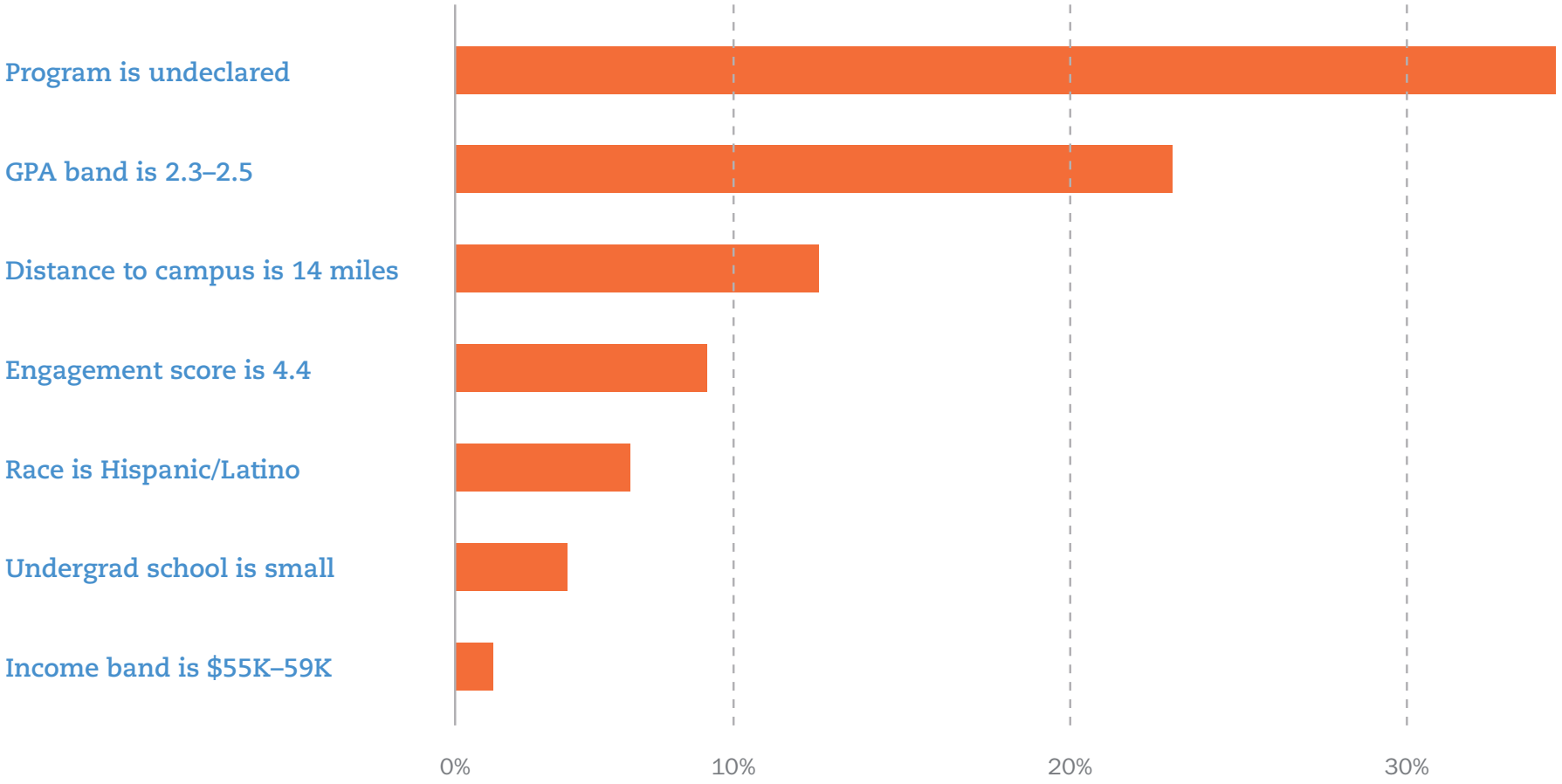
Training the Model: Factors of Importance

When our machine analyzed the data, the model it created generated a list of seven factors of importance. A factor of importance is a data point that has significance in determining the outcome of what we're trying to predict. Our machine learned that the most predictive factors for students who deposited were having an undeclared intended major and a GPA of 2.3–2.5. The model estimates having an undeclared major contributes 34 percent to the likeliness of depositing or not. A GPA of 2.3–2.5 is 22 percent important. Other factors of importance include distance to campus, race, and income band.

The prediction model estimates that these seven variables will help guide the direction of who is more likely to submit a deposit. Note that this does not mean that the same seven factors determine how every admit will decide to deposit or not. Instead, it indicates that the chances of depositing are lower if a student plans to study business, versus a student who hasn't declared a major. We can also say that a student whose family income band is $55K–59K is more likely to deposit, versus students in other income brackets.

There are two important aspects concerning how the model came to be. First, it did not use all the data we secured (see more in the data split section). Second, it deployed a level of advanced statistics, such as entropy and information gain, to divide the training data into smaller chunks until it found the attributes that reduced data impurity. We will spare the technical details in this paper, but search YouTube for an excellent selection of tutorials in classification algorithms.

Factors of Importance



How Good is the Model?

Naturally, our attention should now turn to the reliability of the model. How good or representative are the seven factors? Can we trust the predictions coming from the machine learning model? The answer is YES!

Splitting Data

Let me explain with an example. As a professor, I am used to assessments and evaluation via exams, homework, and good ol' "leading questions." I ask students questions that should be answered in a particular way. This is how I test for learning.

In machine learning, we want to expose our algorithms to the same level of assessment. We know the question and the answer. If the algorithm gets it right, it gets the point. If wrong, its score is penalized, just like a student who receives an A, versus one who receives a B. To accomplish this with the model, machine learning deploys a rather simple technique of splitting the data set. The historical data set is divided in two: (1) a large chunk for training the model; and (2) a smaller chuck for evaluation. The large chunk is usually 70–80 percent of the data set, whereas the evaluation is generally at 5–30 percent.

Say we have a total of 1,000 records in our historical database. Using a 70/30 split strategy, our machine learning will have produced an algorithm by looking at 700 records and which factors are more relevant in the prediction of the objective. Then comes the evaluation. We reserved 300 records with actual information. Based on the deposit predictive model example, we would have 300 students that either paid deposits or didn't. As such, our model predicts binary actions like "deposit" vs. "not a deposit" that can be compared against actual information in four possible results.

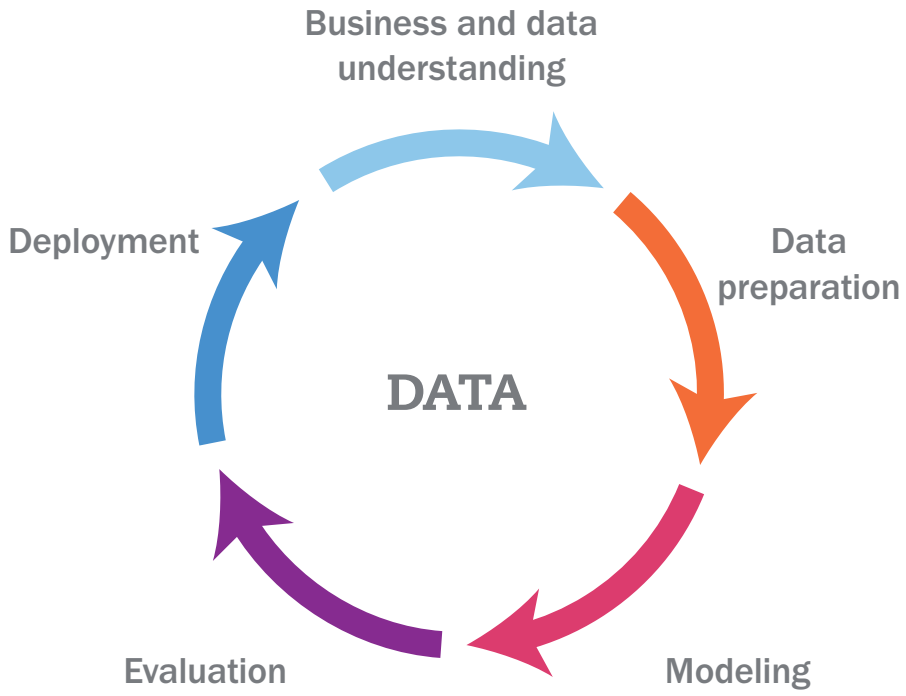
To compute the final model accuracy, we compare the number of accurate predictions over the total number of predictions. We can observe a total number of four predictions. From these four, the total number of accurate predictions is two. Accurate predictions receive a score of 1; inaccurate predictions get a 0. In our example, we have 300 predictions to be conducted. If our algorithm accurately predicts 185 of them, then our performance is 185/300 = 61.7 percent.

Scoring New Data

Prediction	Actual	Scoring
Deposit	Deposit	1
Deposit	Not a Deposit	0
Not a Deposit	Deposit	0
Not a Deposit	Not a Deposit	1

Rinse and Repeat: Tuning

The last component in the modeling process is to tune the algorithm. In machine learning, a data expert will have access to many statistical parameters that could be tweaked to increase model performance. This is where data scientists and people with advanced mathematical training make a difference in modeling. We will not be covering details in model tuning, but leaving it to our Ph.D.s of the world to make our models more predictive and more reliable.



Operationalizing the Model

To recap: We know that we need to split historical data to train an algorithm and use a subset of the data to evaluate the model's performance. What we get is a list of factors of importance such as a GPA range of 2.3–2.5, along with a score indicating the model's accuracy, e.g. 61.7 percent.

Now, let us focus on interpreting the model. Say we recognized that the GPA of 2.3–2.5 is below the student class average of 3.1. Let's also hypothesize that in this recruitment cycle, we were invested in a diversity recruitment strategy. If we were to follow the model's recommendation, do we only end up with low-income Latino students, with GPAs of 2.3–2.5, who live within a radius of 14 miles from campus and have undeclared majors? The answer is NO.

Other students will continue to enroll as they have in the past. The difference is that predictive modeling and the scores it generates should be a guide to how recruitment resources are distributed. There isn't a one-size-fits-all approach. Take these examples:

"Let Them Be"

With this approach, schools move resources to students with lower scores to hedge their bets. They prioritize efforts towards students with a lower likelihood to deposit, leaving the high probability students with a score higher than 50 to chance.

"Win Them Over"

Here, schools focus their energy on students with higher probabilities of enrolling. They spend less of their budget on recruiting students with low probability scores.

Aligning the Model to Enrollment Strategy

Properly interpreting a model's results requires calibrating the good factors of importance. This allows us to put the model in context of a school's enrollment operation and its strategic priorities. In this example, race and school size is seen as positive because it matches the enrollment agenda of the university. On the other hand, having an undeclared major traditionally delays graduation rates, making this factor less desirable. Note how factors such as "Distance to campus is 14 miles" and "Engagement score is 4.4" are neutral, thus not included.

Categorizing Factors of Importance

Positive Factors	Negative Factors
Race is Hispanic/Latino	Program is undeclared
Undergrad school is small	GPA band is 2.3–2.5
	Income band is $55K–59K

Scoring New Data

We move now from historical data to active students in our pipeline. We've built our model and determined that it's accurate. Now, we'll use it to guide ongoing recruitment.

Let's recall that a model tells us what factors are important in making predictions, and how much the factor contributes to the prediction. We can think of the factors of importance and their significance as a data rubric, then use the rubric to evaluate new data. If a condition is met, we assign the contribution score. If not, the contribution is zero.

You can see two scoring examples in the table. Student A was admitted with a physics major, a GPA of 2.45, living 12 minutes away from campus, an affinity for small schools, and a family income of $79K. Four of these variables do not match the predictive model's factors. As such, those scores are zero. The final score is calculated with the weight of the matching variables. Student A received a total score of 37 percent, whereas Student B gets 79 percent.

Scoring New Data

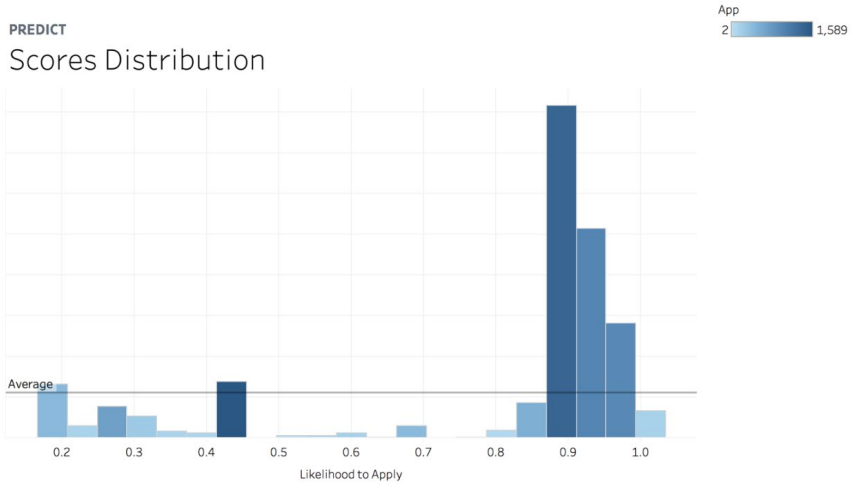
MODEL		STUDENT A		STUDENT B	
	Weight		Score		Score
Program is undeclared	35%	Major: Physics	0%	Major: undeclared	35%
GPA band is 2.3–2.5	22%	GPA: 2.45	22%	GPA: 2.45	22%
Distance to campus is 14 miles	11%	Distance: 12mi	11%	Distance: 9mi	11%
Engagement score is 4.4	6%	Engagement: 0	0%	Engagement: 0	0%
Race is Hispanic/Latino	4%	Race: N/A	0%	Race: Hispanic	4%
Undergrad school is small	3%	School: Small	3%	School: Small	3%
Income band is $55K-59K	1%	Income: $78K	0%	Income: $118K	0%
			36%		75%

Highlighted variables have a zero contribution toward the predictive model's scoring

Scoring New Data

Another relevant technique for scoring new records is to visualize scores in groups. For this, we borrow some of the tools in descriptive analytics. A standard approach in exploratory data analysis is using histograms to quantify the number of occurrences of a subgroup. Because our scores range from 0–100 percent, we can create 20 subgroups (bins) with 5 percent increments each. Subsequently, we count the occurrence at each subgroup and plot the frequency of scores in each bin.

The resulting histogram showcases a distribution of scores concentrated between 80–100 percent, with the highest bin at 85 percent. We can also observe a smaller concentration on the left of the histogram with scores between 0–40 percent.



Ethical Considerations

"Interpretability" is a commonly used term in machine learning. There isn't a standard definition for it. We think of it as being able to easily see how a model arrived at its recommendation. As you can imagine from the examples we've reviewed, when models get more complex, we get further away from understanding how they arrived at a prediction.

Not only is interpretability—and its cousin "transparency"—important for the performance of a model, but we also need to consider the ethical implications of using predictive models in a field like education that impacts students' lives.

Interpretability is not about understanding every single detail of the model for all of the data points. It's more about being able to audit the decision process and ensure it is not discriminatory or violates any regulation. With the rise of data and privacy protection regulation like GDPR, interpretability and context will become even more essential. A single incorrect prediction can seriously impact a student. Additionally, being able to explain how a model works in a straightforward way makes it more likely that members of an organization new to machine learning will see its potential and adopt it.

"With the rise of data and privacy protection regulation like GDPR, interpretability and context will become even more essential. A single **incorrect prediction** can seriously **impact a student**."

PART 2

What We Can Predict: Use Cases

In Part 1, we looked at how machine learning allows us to detect hidden patterns in our student data that ultimately can be used to predict actions, such as starting an application or submitting a tuition deposit. With the right historical data, we can train an algorithm and evaluate its performance. To do all this responsibly, it's essential to use interpretable models, properly understand the context of a model's recommendations, and always have the goals of the recruitment cycle we represent in mind when we formulate enrollment tactics.

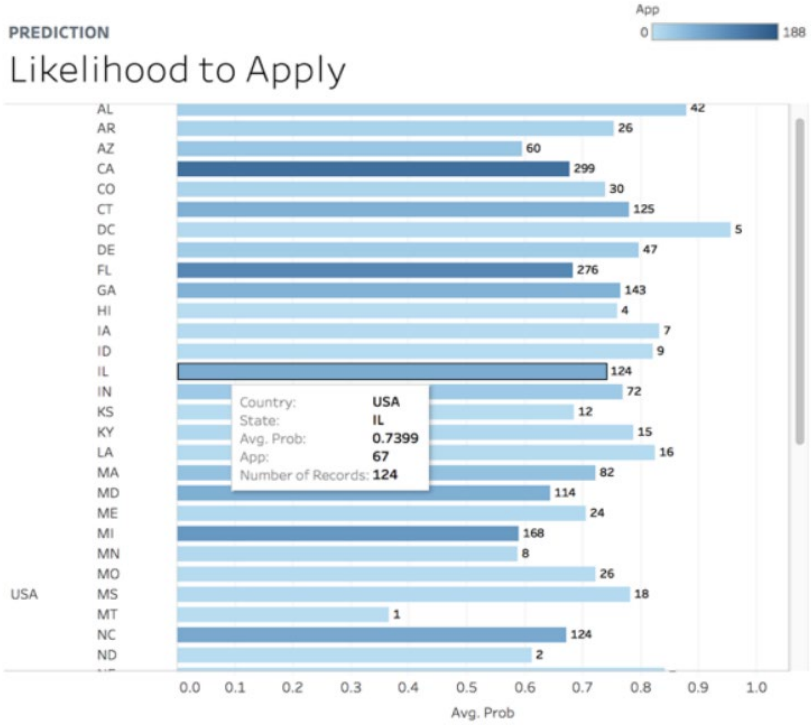
Let's now look at three examples that showcase good modeling, interpretation, and operationalization into the marketing enrollment mix. These cases are based on real models and scenarios we worked on with clients.

Case 1: Alev, a director of admissions, wants to reduce marketing spending by targeting high impact territories.

Admissions Director Alev is contending with multiple states, EPS markets, and dozens of counties and ZIP codes to derive a territory management strategy. As usual, Alev's budgets are limited, and her recruitment team consists of three staff members. How can Alev's predictive modeling identify high-impact territories?

Using three years of historical data, we created a model that predicts the likelihood to start an application with 81 percent accuracy. To overlay Alev's recruitment goals, we hone in on the geographical variables of the model. The result is a territory prioritization tool that Alev and her team can use to select states for upcoming information sessions.

The graph shows the probability that a student will start an application in a given state. The length of the bars indicates the average likelihood; the color intensity represents the number of applications received in the past in each state. When the team needs to decide between traveling to Illinois or North Carolina, we can see both states generated 124 applications last year. However, the average probability of starting an application in IL is 67 percent, versus 0.49 in NC. As application generation is a goal, Alev and the team decided to include IL in their upcoming travel schedule and will not be visiting NC this year.



Case 2: Alev and the recruitment and admissions team want to determine who to visit during the upcoming travel season.

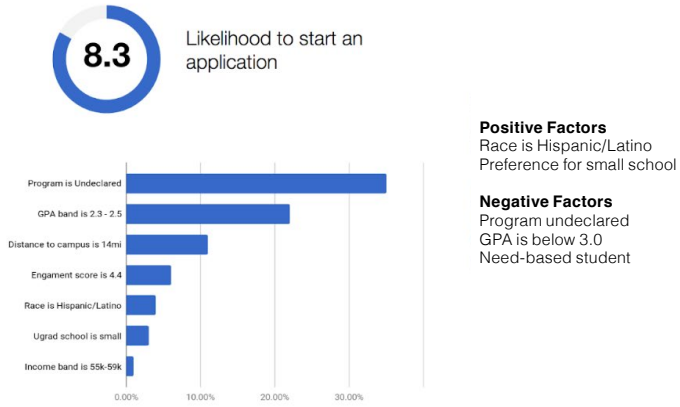
Using historical student data, the team now needs to shift to individual features in the predictive model. Specifically, how can the model suggest which students to visit?

The answer to this problem is found among the features of importance the model produces. Say we have a model with seven features of importance at 81 percent accuracy. The team focuses on the features of "Program is undeclared" and "Undergrad school is small." These features indicate that prospective students are likely to matriculate without declaring a major and have a desire to attend a small school. In a nutshell, what this predictive model suggests is that students with these characteristics are more likely to start an application.

Notice how the team avoids "GPA band is 2.3–2.5," "Distance to campus is 14 miles," "Race is Hispanic/Latino," and "Income band is $55K–59K." This is because these parameters do not fit the academic and financial aid requirements in the recruitment strategy.

At this point, the team generates a list of emails and names from all students who have indicated "undeclared major" and preference for "small schools." They send them personal invitations for upcoming information sessions near them. The team also prepares special content for the identified audience: brochures and a video centered on the experience of the "undeclared major" student at the school.

Predict451 - "likelihood" Visualization

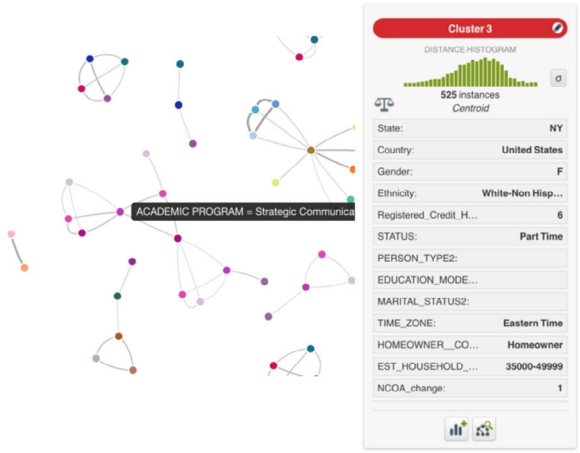


Case 3: A graduate school's marketing team wants to create a highly personalized academic program recruitment campaign.

In this case, we explore a new way of using machine learning algorithms: persona modeling. A marketing persona is a composite sketch of a key segment of an audience. Content marketers rely on them for producing content that is relevant and useful, and thereby, more effective than generic, one-size-fits-all content.

How can machine learning help create audience personas?

Using associations and cluster analysis, we reveal hidden relationships within smaller groups of students or segments (called "clusters") who convert at a higher rate—in other words, a more efficient group to yield across a recruitment funnel. An important distinction in a cluster is that these student groups do not represent the average student in a class profile. While the incoming class consists of 55 percent females, out-of-state, with GPAs of 3.2, the clusters we identify don't follow those exact averages.



The graph shows clusters and associations of students in one program. We find that 523 students are from New York, are "White-non Hispanic," aim to register "part time," are "homeowners," and have an estimated household income of "$35,000–49,999."

The marketing team translates these statistics into an audience description they can use to create personalized content: females with critical financial responsibilities at home for whom work-life balance is a crucial conversation because they are approaching studies on a part-time basis.

After a round of brainstorming using this data to develop a marketing campaign, the team comes up with the concept of "Female Alpha" for the strategic communications degree program. The resulting materials highlight the flexibility and affordability of the program.

Looking Ahead: Operational Efficiencies Are Great, but There's More

Machine learning is helping firms unlock shifts in market trends, identify changes in customer behaviors, and enhance their products and services. Our challenge and opportunity is to replicate this power within student recruitment, admissions, and enrollment management. Why? Because with predictive modeling, we can prioritize efforts and allocate resources more strategically during each recruitment cycle. For example, predictive modeling can help us identify and target optimal territories for student search. It also helps us personalize content for today's students who are accustomed to the sophisticated techniques of brands with bigger budgets most schools enjoy.

But there's more. The unrealized promise of machine learning and predictive technologies goes beyond unlocking student preferences and meeting enrollment targets. Machine learning may be a path to better matching our students with academic programs, faculty, and schools. This means increasing academic success and personal satisfaction for students. And successful, satisfied students mean more engaged students who become proud alumni and ambassadors for our institutions.

To talk about how the approaches and technology discussed in this paper can be applied at your institution, get in touch: connect@element451.com

About the Author

Dr. JeanCarlo (J.C.) Bonilla is the Chief Data Scientist and Head of Analytics of Element451, the leading admissions marketing and CRM platform company. He also oversees the analytics practice of Spark451, a higher education enrollment strategy, technology, and marketing firm, and parent company of Element451. J.C. has helped schools use data to optimize their enrollment activities for more than a decade.