How One Higher Education Organization Is Helping The Military Population By Using Analytics

The Challenge

A global Higher Education organization wanted a way to identify at-risk-students and discover predictive factors in the success of a student. This way, they could create effective policies for helping more students graduate, make better decisions, and prioritize advisory contact.

Like many mature organizations, the university had a lot of data on student demographics, financial aid, coursework, and more. Before we partnered with the university, they were using a solution that produced insufficient results, integrated few sources of data, didn’t produce a data model, and stored data inefficiently.

The university wanted the ability to perform periodic analyses to remain relevant in their predictive factors, and to identify the most at-risk-students. They needed a solution that was in-house, flexible and scalable.

The Solution

First, we started by narrowing our solution focus — to undergraduate students seeking a Bachelor’s degree — so we could quickly deliver measurable value in a short amount of time. Once the solution has been proven, we could scale the framework to include graduate students, specialty degree students, and more.

To develop the solution for the undergraduate students we:

1. Partnered with the university to understand their data and business objectives — the most important part of the project. Failure to describe target variables could result in inaccurate results, findings, and modeling. For the university, their target variable included first semester completion and graduation.

2. Designed, installed and configured a data platform built with scalability and flexibility in mind. To process and store the high volume and variety of data, we installed a Cloudera Hadoop Cluster. This allows the university to scale their solution by storing data, and having it accessed, by multiple machines.
Enabled data integration and analysis by using Apache Spark, Apache Sqoop, and Impala technologies.

Designed and developed integration strategies once the data was accessed and the data framework was in place. We collected data from the student ERP system, financial aid system, and their student care system, and more.

Used R language, a predictive modeling language, to analyze 110 variables and over 60 features. From the analysis we dropped, cleaned, and transformed data to accept it into the model. We selected 21 features and determined a logistic regression algorithm was a best fit to provide prediction scores. This resulted in a 12-point improvement over the null hypothesis.

Delivery

All business logic that was used in the model was vetted and made transparent through detailed documentation by partnering with the university. They are using the model to test their hypotheses on student success as well as other ideas gathered during the interviews.

Executives are using the visual dashboard to filter the student population and discover a success rate based on the values being filtered. For example, the university places an emphasis on retraining the military population, so we created a “military dashboard” that allows executives to explore success factors among the military students.

The Results

The university was able to find key predictors that allowed them to identify at-risk-students. From their finding, they developed a better advisory contact system, were able to make better decisions, target at-risk-students appropriately, and improve the graduation rate for their students.

For example, the university found that onsite females are more likely to leave before completing their degrees, meaning they are at a higher risk than their male counterparts. They created an appropriate contact system to help onsite females stay on track.

With our solution, the university now has a scalable and flexible solution they can use to perform periodic studies. This allows the university to stay relevant and create the most effective policies as their student population evolves.