KWANTUM

Analytics

Business Problem

The institutional research office at a public university needs help to increase the student retention rate. Knowing the students' attribute such as family income, ACT score, etc., the office want us to 1. Build a model to identify the students who are likely to drop; 2. Find the important drivers that causing them to drop and suggest possible actions to prevent the students from dropping.

Challenge

Student data has a skew structure which means if we predict naively that no one drops, we can have an 86% accuracy since only 14% of students actually dropped. Therefore regular prediction models will fail. For the second questions, the survey data has many attributes that are highly correlated thus finding the significant variable is almost impossible which means a large amount of feature engineering is needed.

Our solution Kwantum Analytic built an ensemble model which combines 5 supervised learning algorithms for prediction. Instead of the accuracy, the model recalls as many students who dropped as possible while keeping the false alarm relatively low. Important features are identified using ANOVA (analysis of variance) and mean decrease accuracy. Additional features representing student's financial situation, high school performance, etc., were created to boost the prediction results.

Impact

Our model was able to identify 30-50% (depending on the data) dropped students one year before they actually drop. The office can use it to send alerts to students and their advisor. The model also provide a table of variables ranked in statistical significance and impact. With the information, the office is working on designing activities such as freshman seminars and a tour of major in the summer before freshman year. We are also working with the office to publish a paper on the research project. The tuitions of dropped students make huge impact on the school's financial situation thus our research was very valuable. It is easy to see that data science has great application in institutional research yet not many researches/projects have been done. We see great potential in this area, not limited to just student retention analysis, and are looking forward to working more on projects from different school.