Capstone Project Ideas

Predicting Peer-to-Peer Loan Payback Amount

Peer-to-peer lenders risk a financial loss if a loan defaults and is not paid back in full. One way to try to minimize the loss is to try try to predict whether a loan will be paid back or not (classification). While predicting if a loan will default or not is valuable, it doesn't take everything into account. For example, there could be a situation where a high-interest loan defaults after being nearly paid off, that is actually pays back the lender more than a low-interest rate loan that is paid in full. Therefore, this project will take a different approach and predict the amount that will be paid back.

My client is an investor that wants to maximize their return on their lending. When an investor obtains a list of available loans they can invest in, they could run the model and generate a prediction of how much will be paid back from their investment, and choose the loans that will maximize their return.

The data for this project comes from http://lendingclub.com. The data is available for download in csv format.

Predicitng Medical Appointment No-Shows

Appointment no-shows are a significant problem in the medical industry. By calculating the probability of a pationt no-show, medical facilities can optimize their bookings to reduce the costs of under and over-booking.

Can patient information such as age and health problems, as well as time of day and day of week, be used to predict the probability of them making their appointments, and provide information that can be used to create a smarter booking system that can reduce the problems associated with more or fewer patients showing up than expected?

The dataset can be found on Kaggle at the following link: https://www.kaggle.com/joniarroba/noshowappointments. The data consists of ~110,000 medical appointments each with 15 variables, including appointment day, day scheduled, age, neighborhood, show or no-show, and patient medical conditions such as hypertension, diabetes, alcoholism, and handicap.

Predicting Income

There are many uses for determining income level in marketing. This project would explore using demographic variables to predict income class.

The data can be found at https://archive.ics.uci.edu/ml/datasets/census+income. The data consists of ~48,000 obeservations of 15 independent variables including age, gender, education, marital status, occupation, hours worked per week, and native country, as well as the dependent variable, which is whether the income level is above $50,000 or not.