

Project Proposal Title:

Predicting Hospital Readmission Rates for Diabetes

Members:

Steven Worthington (stevenworthington@g.harvard.edu)

Background and Motivation:

Management of high blood sugar (hyperglycemia) in hospitalized diabetes patients has a major bearing on their health outcomes, in terms of both morbidity and mortality. However, while intensive care units (ICU) employ rigorous standards for hyperglycemia management the same cannot be said for most non-ICU inpatient admissions. An analytically sound national assessment of diabetes care during hospitalization could serve as a baseline for change, which might lead to improvements in patient safety.

Data:

The data we will use is from the UC Irvine Machine Learning Repository (<https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008>) and represent ten years (1999-2008) of clinical care at 130 U.S. hospitals. There are just over 100k observations, which represent records for hospitalized patients diagnosed with diabetes. The target variable represents days to inpatient readmission, with values "<30 days", ">30 days", and "No" for no record of readmission. There are 46 features, which cover patient demographics, health outcomes, and medications, as well as hospital-level attributes. The data are relatively clean and organized, and most features have minimal missing values.

Scope:

The focus of the project is on predicting early (<30 days) hospital readmission as a function of patient and hospital-level features. A major goal is to assess relative feature importance. As such, any model capable of classification is a potential candidate model, though we will focus on bagging and boosting ensemble models and generalized linear models. The project will be an excellent bootcamp for learning and applying these methods to real World data.