BST 263 Final Project Outline - ReAdMission Impossible

Healthcare costs and expenditures are increasing and programs, such as the HRRP from CMS, have been created to reduce the burden on both patients and the healthcare system by minimizing the number of avoidable readmissions. In this project, we suggest a model to predict hospital readmission among those hospitalized for cardiac-related issues. These findings will not only guide healthcare providers in prioritizing patient follow-ups after discharge, but they may have implications in the decision-making of initial hospital stays (e.g. A patient stays 2 days longer because they have a very high probability of readmission).

The primary knowledge gap that we are aiming to fill is identifying major predictors of readmission among individuals hospitalized for cardiac issues. Understanding what patient characteristics and aspects of medical history are most important in predicting readmission can provide more information to physicians when determining whether or not to discharge a patient.

Our project will explore data collected from the Hero DMC Heart Institute, Unit of Dayanand Medical College and Hospital located in Ludhiana, Punjab, India. This data is provided under creative commons License and is readily accessible. The data was collected over 2 years, from 1 April 2017 to 31 March 2019, and it includes information on 14,845 admissions in the hospital's cardiology unit. Of these admissions, 1,921 of the 12,238 patients were admitted to the hospital multiple times, with several patients having upwards of 10 readmissions.

We will create a variable to indicate readmission to the hospital after a patient's initial visit. Therefore, only data on the initial patient admission will be used to predict the first readmission and any subsequent readmissions will not be considered. We will use demographic data (e.g. patient age, sex, locality) and patient history (e.g. drug use, chronic illnesses, blood profiles) as our primary predictors.

The overarching objective of this project is to develop a model to predict hospital readmission based on key characteristics from the patient's first visit. This would provide clinicians with a tool to look into specific admission features and/or patient characteristics that may result in a patient's readmission. Having such knowledge may also result in more careful monitoring, proper follow-up care, and more.

Our goal is to compare several statistical learning methods and their test performance and determine which model will most accurately predict whether or not someone will be readmitted at a later time (within a 2-year time frame, given the structure of the data). We also want to use these methods to look at the variation between what variables are considered the most important features. We will employ methods such as random forest, BART, boosting, KNN, and PCA.

We plan to split our dataset into a training set (80%) and a test set (20%) so that we have a fixed test set to test our model(s) on. We will perform cross-validation throughout the 80% training set to determine any tuning parameters needed in our model(s). Additionally, to evaluate and compare model performance, we will look at evaluation metrics such as out-of-bag error, MSE, and test error of our model(s). We will also look at the accuracy of predictions since we have that data available.