02-604 Project Proposal: An Interpretable Tool for Predicting In-patient Length of Stay and Integrated Bed Assignment

Problem & Motivation

Predicting patient length-of-stay (LOS) is a key challenge in healthcare analytics. Accurate prediction of LOS can greatly aid key logistical decisions for the hospital such as in bed assignment, for example. While the industry standard is to estimate LOS by the average or median of past hospital admissions, without consideration of patient characteristics. There are numerous statistical methods in the literature to build more sophisticated models for this prediction task. While some of these papers use neural-networks, there is evidence that simple models can work comparably well, while having the advantage of interpretability (e.g., here).

This in mind, there are 3 goals for this project:

- Develop an interpretable model (or models) for accurate LOS prediction and compare it to the interpretable (and non-interpretable) models currently in the literature
- Use the LOS prediction to assist in the bed assignment problem
- Formalize this workflow in a Python tool where users can use data to train their model, assign beds, and visualize results

Approach, Data, & Timeline

For the task of LOS prediction, there are a variety of potential interpretable methods that I can try. From the current literature, it seems that decision trees could work well but I would certainly consider other approaches such as building a GLM. There seems to be test data online (e.g. here). Because predicting the LOS will likely be an iterative process where various types of methods and features are taken into account to build an effective model, I expect that it will take several weeks.

For the task of bed assignment from LOS prediction, I will plan to use a classic approach to optimization under uncertainty—stochastic programming. Using SP does require some sort of estimate of the uncertainty of the LOS model prediction, so depending on what can be determined in the initial step, the optimization paradigm may change. A key challenge to optimizing will be to accurately model the logistics of a hospital. This will require some research into how hospitals typically assign beds (perhaps available in the research literature) in order to formulate the optimization problem correctly. I will plan to work on this after the LOS model has been constructed and, between literature review and modeling, it will take a few additional weeks.

Finally, putting both of these things components into a useable Python tool should be a relatively quick process since most of the code would have already been written.

I see this project as quite feasible since it is already a well-studied problem and really my goals is to find some improvements in the margins of interpretability, performance, and practicality.