Capstone 2 Proposal: Predicting Hospitalization Features

Project Description

Healthcare records have become increasingly more digitized. This is an open an opportunity to analyze and obtain patient data at an unprecedented detail and scale. While the potential to gain greater insights, cost reduction and efficiencies in the healthcare space exists, great challenges remain. Some of these include data availability and complexity of services. Healthcare is particularly complex due to overlapping systems, diversity of providers, services and health issues.

This project would use hospitalization data from Brazil to make predictions about key features of a hospitalization request. There are three potential prediction targets. Implementation of these will likely depend on computational resources available.

1. Predict how many days will pass between request for hospitalization and discharge.
2. Predict whether the hospitalization request will be rejected or accepted.
3. Predict procedure performed based on the we know about the patient in the hospitalization request.

Ability to accurately predict these three features of hospitalization can yield significant benefits. For example, knowing how many days a patient can be expected to be in the hospital, will help hospital managers manage their capacity (especially in areas where beds are scarce). Length of stay and likely procedures can inform service charges and help all parties involved navigate the healthcare charge system better so likely costs are known in the front end.

Moreover, predicting healthcare expenditures can be tricky for insurers, providers and particularly consumers. One of the main factors that have been cited as a cause of rising healthcare expenditures is the inability of consumers to know in advance the cost of the healthcare services they consume.

Predicting whether the hospital request will be rejected or accepted can help reduce administrative burden of thoroughly reviewing each hospitalization request.

Dataset

The data that will be used is from the Authorization for Hospital Admission. This dataset is part of Brazil's SIHSUS Hospital Information System. This system manages the coordination and payment by Brazil's public healthcare system (covers around 34% of

Brazil's population). The data is publically available as .dbc files. In this application, I will be using data from 2015 – 2018. This represents 3.5 years' of information.

A record in the AIH database is created when a hospital or healthcare unit generates a request for hospitalization. Providers submit demographic and health information about the patient. This request is approved, reduced, rejected, or rejected due to an error. While the patient is in the hospital, the record is updated to also contain information about procedures performed and discharge.

Methodology

For this problem, I will use a deep learning approach (neural network approach). The structure of the neural network will be informed by exploratory data analysis and the prediction task.

Project Deliverables

The deliverables for this project are: (1) code notebook with a complete implementation in the python language, and (2) report with accompanying a slide deck explaining the problem at hand, methods and results.