Predicting 30-day hospital readmissions

Background

Performance of hospitals/healthcare providers is measured using several factors, often referred to as the “Measures that Matter”. While there are several measures that are reported (to adhere to national standards and to receive financial incentives for superior performance from the government), CMS (Center for Medicare and Medicaid Services) uses 7 outcome measures that determine the quality of care delivery. These 7 measures also help CMS rate hospitals (hospitals with lower star ratings get penalized for their poor performance and the ones that do well receive heavy financial benefits from the government). Hospital readmissions is one of the 7 measures that receives a 22% weightage when hospitals are to be evaluated for care delivery.

Data set

The data has been retrieved from the Cerner database, now available in the Machine Learning Repository, it represents clinical data from 1999-2008. The database has over 100,000 diabetes inpatient encounters (diabetes was a diagnosis on these encounters). Information about laboratory tests and the medications that were administered are also available. Further, we have information about the admission type, discharge disposition, and demographic information such as race, age, and gender.

Goals

The main goal of this project is to predict if a patient would be readmitted to the hospital within 30 days of initial discharge. An attempt will be made to make associations/analyze correlations (if any) between the admission type, discharge disposition, demographics and the length of stay. This analysis will be used towards stratifying the risk of patient readmission.

Approach

1. Understand the features of the database
2. Data wrangling
3. Explorative analysis
4. Predictive analysis, machine learning
5. Construct models around the data

Output

1. Paper
2. Data analysis with visualizations
3. Jupyter notebooks