Capstone Project Proposal

summary

Dialysis is a treatment that does some of the things done by healthy kidneys. It's a treatment that takes over your kidney functions if those organs stop doing their job. There are several dialysis facilities registered with Medicare in the country where patients often visit. Besides the importance of the quality of care they provided, collecting additional health data of their patients could be helpful in improving their approach towards the treatment. Most of the dialysis centers have separate health data related to their patients. Some of the data include mortality rate (deaths), hospitalizations, blood transfusions, incidents of hypercalcemia (too much calcium in the blood), percentage of waste removed during hemodialysis in adults and children, percentage of waste removed in adults during peritoneal dialysis, percentage of AV fistulas, percentage of catheters in use over 90 days and others. The Medicare's "DFC Star Program developed by Medicare assigns 1 to 5 stars to these dialysis facilities by comparing the health of the patients in their clinics to the patients in other dialysis facilities across the country. However, no reports were found that shows a relationship pattern or prediction of mortality rate based on the other health data. This project will try to accomplish that. What would the pattern of mortality rate look like if a prediction is made based on the health data listed above?

Objective of this project

The goal of this capstone project is to predict a pattern of mortality rate based on the other health data provided.

My hypothesis:

Some of the health data such as readmission rate, hospitalization rate, standard infection rate are expected to show positive relationship to mortality rate. Different results can happen when we to try to predict the mortality rate from the other health data. The higher or lower values of each field will define the mortality rate pattern in positive or negative way respectively. For example, patients with arteriovenous_fistulae has a lower risk of infection than patients with catheters, and so does their mortality rate.

Data

The data for this project was publised by Centers for Medicare & Medicaid Services and was downloaded from the DATA.MEDICARE.GOV. The dataset is comprised of data on anemia management, phosphorus levels, transfusion rate, dialysis adequacy, vascular access, mineral and bone disorder, hospitalization rate, readmission rate,infection ratio, scale rate of the facility and others.The data was collected from 2012 to 2015 and can be downloaded from:

https://data.medicare.gov/Dialysis-Facility-Compare/Dialysis-Facility-Compare-Listing-by-Facility/23ew-n7w9/data

The details of the data fields with their term definitions can be found at the following link by clicking the “get supporting documents tab":

https://data.medicare.gov/data/dialysis-facility-compare

Method/Approach

I will treat this project as a unsupervised learning problem. The following libraries will be used for data loading, wrangling, cleaning, feature selection, matrix manipulation, data visualization, classifier models, model evaluation etc.

- pandas
- numpy
- matplotlib
- scikit-learn

Deliverables

- Code along with document
- Power point slides followed by presentation