DMSL Course Project Proposal

Ju H Lee

Project topic: Predicting US hospital closures

Data source:

US hospital closure data: https://www.shepscenter.unc.edu/programs-projects/rural-health/rural-hospital-closures/

US hospital data: https://www.kaggle.com/datasets/thedevastator/hospitals-in-the-united-states-a-comprehensive-d

Description:

Hospital closures in the US have been increasingly common and often affect rural areas with lower income populations. By combining data sets on quality of patient care, ownership, and patient experience and the hospital closure data to highlight the hospitals that have closed, a logistic regression model can be used to assess the probability of future hospital closure. There are many data points that are available on various hospital performance metrics such as utilization, patient satisfaction, hospital financials, and surrounding area demographics. Analyzing these factors to choose the factors that are most highly correlated with hospital closures and visualizing the hospitals at risk for future closure is a valuable resource for hospital administrators, local residents, hospital employees, and local governments. This analysis can reveal geographic areas that are at risk for insufficient hospitals and care providers for the population.

The analysis will start with data cleaning and combining the two data sets to allow for supervised learning analysis. Variable/feature selection and exploratory data analysis will help identify the most relevant features to include in a logistic regression model. Heatmaps are an example of visualization that can present the results in a condensed format.