U of M Summer 2022
Data Analytics and Visualization Capstone Project
Group 2:
Lindsey Teeters
Kate Heise
Lalit Toshniwal
Nick Petcoff

In Hospital Mortality: Prediction of Heart Failure

Executive Summary

During the 2010's the CDC did a study for heart failure patients within the United States. The results of that study show that in the 2010's on average there were more than $6 million patients a year that had heart failure as a cause for being in hospital care. From that an average of 400,000 Americans a year passed due to the leading cause being heart failure. Along with lives being lost the average cost to treat heart failure cost America more than $30 million a year.

We wanted to look at the predictability of survival for patients who come into the hospital. Can we build a tool that will help predict the survival of the patient through machine learning?

We sought out to answer this question through several avenues of analysis. The first being the exploratory database analysis. The database we used was quite large, with 1177 rows and 51 columns. Our ERD review allowed us to review the shape of the data and the results were telling, as it showed that we would not need to use every column of data to provide us with a dataset that could aid in survival predictability.

Through the ERD process the refined dataset was further reviewed through Tableau visualizations. One being focused on heart failure and its relationship with comorbidities. The second dashboard showed the relationship between patient statistics, such as BMI, heart rate, EF, and others.

The refined dataset was also put through a machine learning module using SMOTEEN, with an Extra Tree Classifier. This was done as our dataset was imbalanced due to most of the patients surviving to hospital discharge. The classifier was needed to aid in balancing the data to get a more accurate prediction, without it, the prediction would be heavily leaning towards survivability.

The project was then organized in a website and deployed to public access via Heroku. The website itself provides a one-stop review for the ERD, database, Tableau dashboard, and machine learning analysis.

One additional feature on the website is a filterable prediction chart. Once filled out this prediction chart will show predictability of if the patient will survive to discharge.

The labs and other items that can be filtered showed the greatest correlation to our target variable of outcome, or if the patient died or lived within hospital care.

With our analysis it seems that the older a patient is, the more their stats deviate from the average for the correlated lab work, and if that patient has greater comorbidities, the greater the patient has to not survive while in care. With that being said the population of America is rapidly aging. It will be important for the healthcare system to get tools out to the medical staff around the country to aid in saving time, lives, and funding.