Predicting U.S. Hospital Ratings
By: Jiayu Li, Ben Oppenheimer, Xueting Pu, Bowen Sun
36-600: Overview of Statistical Learning and Modeling

Introduction
There are many factors, including medical costs and safety ratings, that are considered when hospitals are rated. We develop a model which predicts the overall ratings of American hospitals. We train our model using hospital data. Our data focus on ratings of the cost and value of certain procedures. The goal of our study is to find the best model to predict U.S. hospital ratings and to produce an accurate classification of high-rated and low-rated hospitals.

Data
Our dataset contains information about 1,739 American hospitals with 20 predictor variables and one response variable. The response variable is Rating which has been discretized to “low” (1-3 stars) and “high” (4-5 stars). The predictor variables are ratings about the facility and costs associated with particular procedures including heart attacks, heart failures, pneumonia, and hip-knee replacements.

Methods
- We train our model on 70% of the data and test it on 30% of the data.
- The predictor variables exhibit multicollinearity, but as our project goal is prediction and not inference, we do not remove variables with high variance inflation factors.
- We apply several statistical learning models to the data: logistic regression, decision tree, random forest, and linear-kernel SVM models.
- We build binary classifiers using logistic regression, decision tree, random forest, linear-kernel SVM.
- The highest AUC, 0.948, is from SVM. Logistic regression has an AUC of 0.946 which is almost as high as the AUC from the SVM model.
- Optimal class predictions for SVM are generated by maximizing Youden’s J statistic. Using this method ensures a low misclassification rate of 12%.

Conclusions
We choose the SVM model to predict U.S. hospital ratings, because it has the lowest AUC of 0.948. We find that our SVM model can predict hospital ratings with a low misclassification rate of 12%.

Reference: Data from: https://corgis-edu.github.io/corgis/csv/hospitals/