Using Procedural Evaluations To Predict Hospital Ratings

Group Members: *Ashritha Eswaran, Jakob Cervenak, Janvi Mirchandani, Prashanthi Kukkapalli, Shivam Mishra, Yash Dugar*

Introduction

- Our dataset contains information for ~1700 American hospitals, with 20 measurements that we seek to associate with the overall rating of the hospital.
- Originally, the hospitals were given an overall rating from 1 to 5 stars; here, that has been discretized to low (1-3 stars) and high (4-5 stars).
- Health care ratings are ratings or evaluations of health care. In the United States they have been an increasingly used tool to try to drive accountability among health care providers and in the context of classic supply/demand view of Health economics, to help health care consumers make better choices.
- In 1997, one research team said that the value of the ratings were limited because at that time public medical data was difficult to access, and a lack of quality data limited the usefulness of any ratings.
- Hospital ratings are essential in determining the standard of healthcare at a medical facility. In many cases, it is difficult to exactly determine a hospital's rating, in a manner that it can be compared with others. With our dataset, we will try to predict different hospital ratings, based on various factors that considers different aspects like safety, mortality, cost, quality etc.

Data

- **Predictors:** In our dataset, we have various features such as **type of facility** (i.e., Church, Government, Private or Proprietary), **rating in relative to national hospital** (such as mortality, safety, effectiveness etc.) Apart from this, **there are also three separate columns each (cost, quality and value)** for heart attack, heart failure, pneumonia and hip/knee conditions, which are numerical/ordinal.
- For our target variable, we have **hospital rating which has a value of low/high**.
- On the right *(Fig.1)*, we can see a quick snapshot of how the **various categorical variables (ordinal in nature) are distributed.**
- We can also see the **distribution of numerical features** *(in Fig.2)* for cost of the four conditions. mentioned above.)

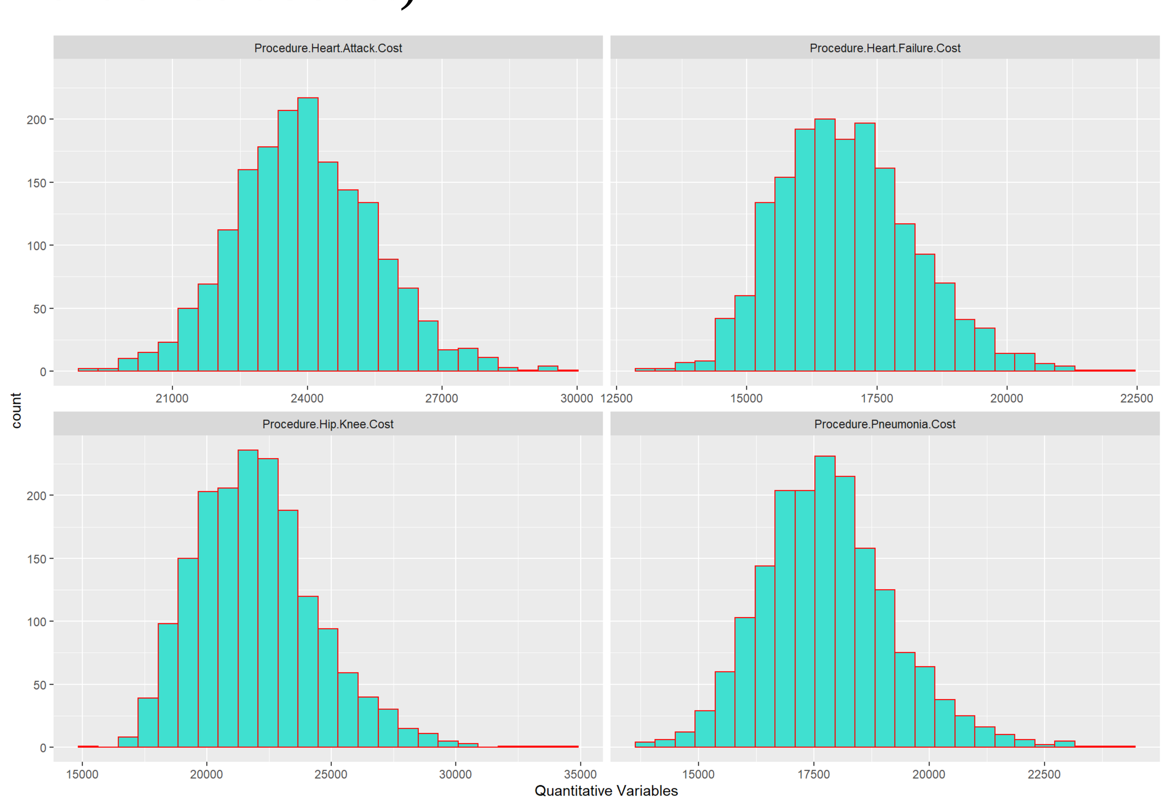

Fig.1 ***(Distribution of different procedures cost)***

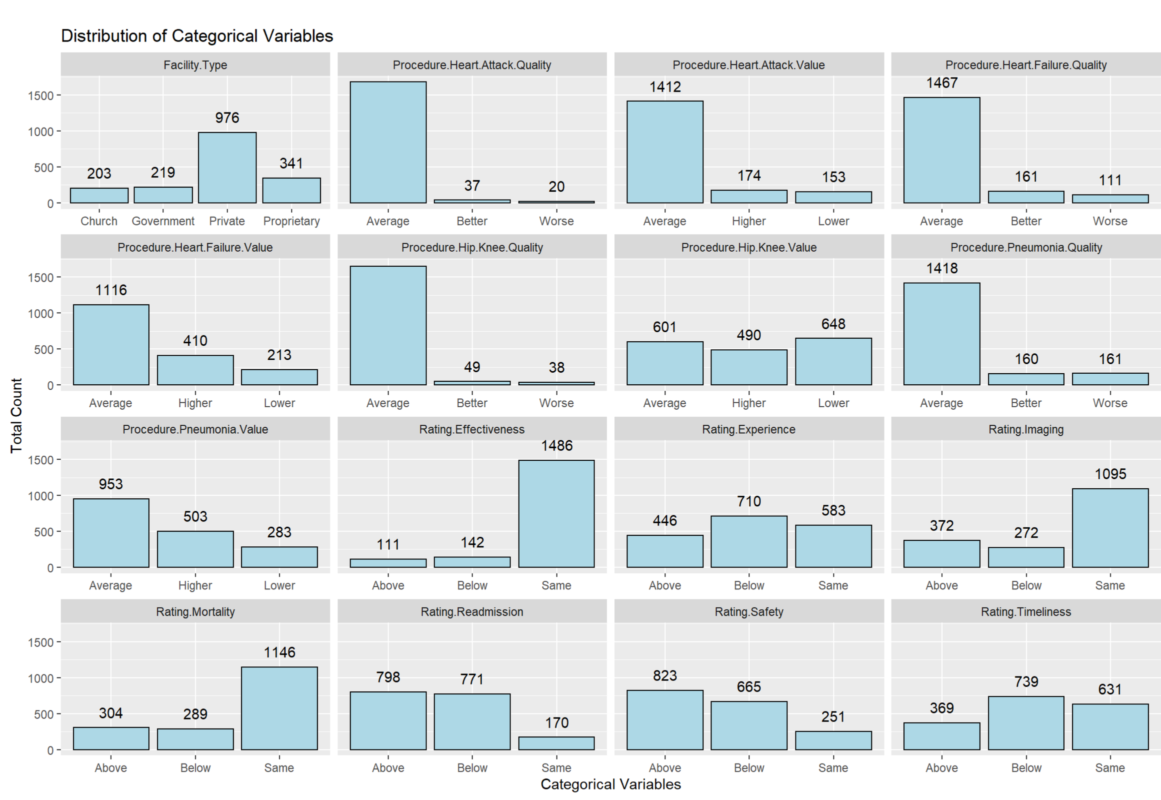

Fig.2 ***(Distribution of all categorical features)***

- **Response:** Our response variable is **Rating** which can take value of either low or high.

References

https://en.wikipedia.org/wiki/U.S._News_%26_World_Report_Best_Hospitals_Rankings
https://en.wikipedia.org/wiki/Health_care_ratings

Analysis

- We ran various models, to try and best predict our target variable, with the help of the different predictor variables present in our data. On the left *(Fig.3)*, we can see the **ROC curves for the different models.**
- **XGBoost and Random forest gave us the best results (with AUC of ~0.91)**. Looking at the feature importance, we find that **Rating.Safety and Rating.Readmission**, are the most important when determining the rating of a hospital *(Fig.4)*.

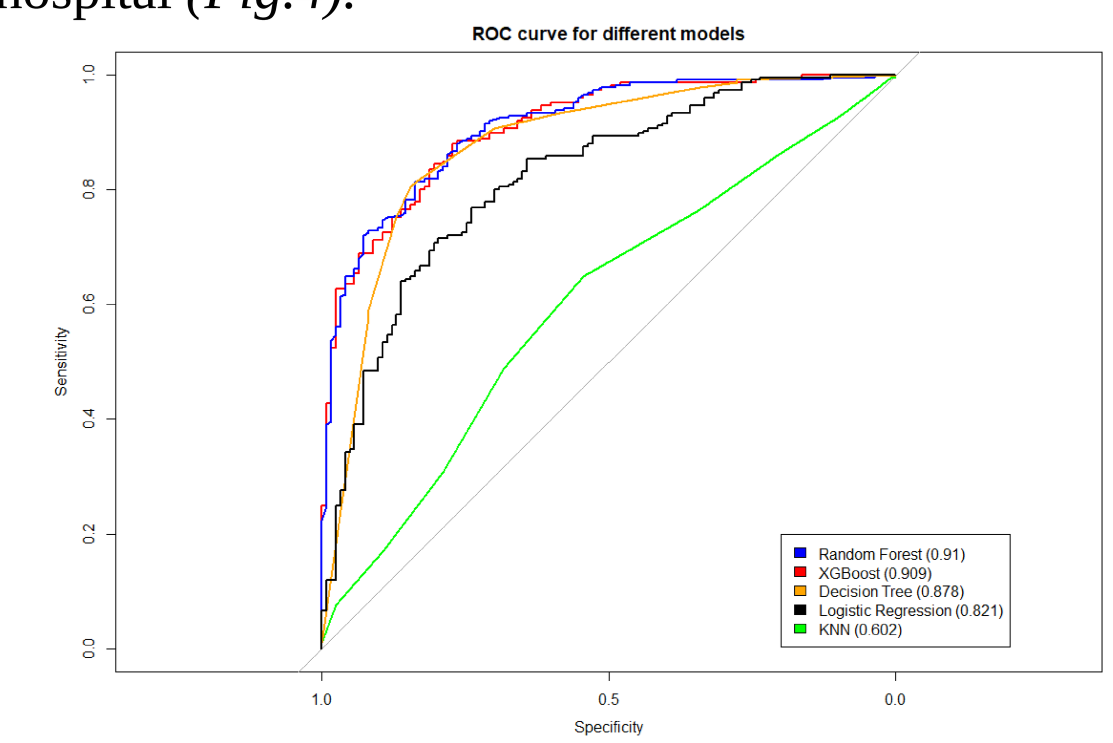

Fig.3 ***(ROC curve of different features)***

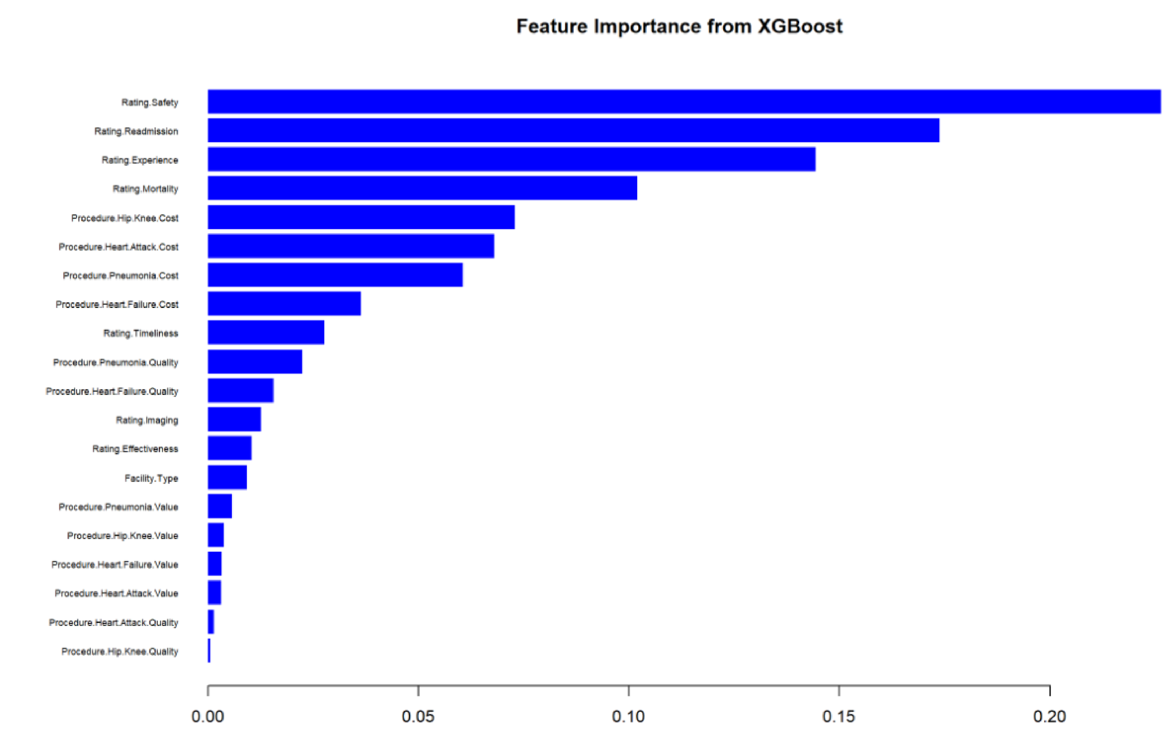

Fig.4 ***(Feature Importance using XGboost)***

- Below *(in Tab.1)*, we display the confusion matrix for the XGBoost and Random Forest models. We can see that **Random Forest have higher "True Negatives" but lower "True Positives", when compared to XGBoost.**

	Actual Value	
Predicted Value	High	Low
High	100	36
Low	23	189

	Actual Value	
Predicted Value	High	Low
High	99	32
Low	24	193

Tab.1 ***(Confusion Matrix for XGBoost above and Random Forest below)***

We determined optimal class thresholds using Youden's J statistic, which balances prediction performance across both classes.

a. We find the **optimal threshold for XGBoost to be ~0.75**, **while for Random Forest it is ~0.76,** which is used as the threshold for making class predictions for the respective models.

b. The associated **misclassification rate is ~0.17 for Extreme Gradient Boosting and ~0.16 for Random Forest.**

Results and Conclusion

- To the right *(Tab.2)*, we display area-under-curve (AUC) results for the five classification models we used.
- We were successfully able to create a predictive model for the rating of any hospital with the help of different predictor variables *(rating, facility type, cost + value + quality of different procedures).*
- Random Forest and XGBoost gives us similar results (with XGBoost having slightly low MCR, but similar AUC).
- We find that **Rating.Safety** (rating of a hospital's safety) is one of the important features when determining the rating of hospital, followed by **Rating.Readmission** (rating of a hospital's readmission) and **Rating.Experience** (rating of hospital's experience), respectively.
- These feature also makes sense, since the relative safety of a hospital should be of priority when comparing the rating of different hospitals, so it makes sense to have that as one of the important feature in classifying our data.
- Another interesting insight we saw from our results was that the cost of different procedures is on average very similar for both low- and high-rated hospitals (those more highly rated had costs approximately 1.5% lower, except for hip/knee replacement procedures).

Classification Model	AUC Value
Random Forest	0.91
XGBoost	0.91
Decision Tree	0.88
Logistic Regression	0.82
KNN	0.60

Tab.2 ***(AUC table for different models)***