<The Effects of Perceived Quality and Satisfaction on Hospital Rating>

Introduction

There is complexity in evaluating the efficacy of a health care system due to a combination of scientific, technological, economic, and epidemiological imperatives [1]. As such, the common proxy for quality of hospital care is the patients' ratings of their hospital [2]. The perceived quality and satisfaction of patients is of particular importance as it can be a salient indicator used to draw detailed contextual descriptions of hospital's different components to identify patterns that may have an impact on other outcomes of a health care system. In addition, the findings of these comparative analyses can recommend policies that are aimed at system-level improvements. I will evaluate the official hospital compare datasets provided by Center for Medicare and Medicaid (CMS) for hospital quality comparison [3]. Classification methods will be used to predict the overall hospital rating and identify the characteristics of hospitals. For predictive modelling, three classification algorithms will be explored: (1) decision tree, (2) Naïve Bayes, and (3) logistic regression. These methods will be implemented using a combination of Weka and R.

Literature Review

The quality of health care is hard to quantify as it is contingent on several factors. Based on the literature review, there is a lack of exploration on the hospital performance on measurements of patients' ratings and performance with respect to one characteristic of hospitals (e.g., mortality national comparison) related to performance with respect to another characteristic (e.g., readmission national comparison). In addition, the quality metrics for characterizing a hospital of high quality has yet to be discussed. For example, do patients who receive care in hospitals with key characteristics (not-for-profit hospital, below national average timeliness, etc) report better experiences than patients in hospitals without these characteristics? DeLancey et al. in 2017 published their results on the associations between hospital characteristics, measure reporting, and the CMS Overall Hospital Quality Star Ratings [4]. Based on this article, hospitals with star ratings between 4 and 5 were considered to have high quality while hospitals with 1 to 3 stars were considered to have low quality. The principal investigators have concluded that smaller hospitals more frequently achieved a high star rating compared with larger hospitals.

National policies to improve health care quality have largely focused on clinical provider outcomes and payment reform; however, the study conducted by Tsai et al. in 2015 revealed the association between hospital leadership and quality [5]. Also, the paper outlines the increase in quality of care with hospitals practicing more effective management. In particular, the hospital ownership type determined the management style which validates the relevance and inclusion of the Hospital Ownership attribute in the analysis. In addition, Bloom and his team explored whether different management style affected hospital performance [6]. They published that higher competition results in higher management quality and improved hospital performance.

With regards to the meaningful use of electronic health records (EHRs), it is a challenge to validate the relevance as the external rules, regulations and pressures have influenced the credibility of such a metric. The recent passage of the American Recovery and Reinvestment Act (ARRA) of 2009, which includes the Health Information Technology for Economic and Clinical Health (HITECH) Act, made available over $20 billion dollars for health care practitioners who become "meaningful users" of health IT. Thus, ARRA introduces the single largest financial incentive ever to facilitate electronic health record (EHR) implementation. However, recent studies reveal strong evidence that supports the use of EHRs, as such, the attribute regarding meaningful use of EHRs will be retained [7].

Dataset

The dataset used in this analysis is sourced from the Hospital ratings dataset from Kaggle (https://www.kaggle.com/center-for-medicare-and-medicaid/hospital-ratings). This is derived from the official dataset used on Medicare.gov (https://www.medicare.gov/) for hospital quality comparison.

The dataset consisted of 4812 observations and 28 attributes. The nominal variables include Hospital Name, Address, City, State, County Name, Hospital Type, Hospital Ownership, Emergency Services, Meets criteria for meaningful use of EHRs, and the following variables had themselves and the respective footnotes: Hospital Overall Rating, Mortality National Comparison, Safety of Care National Comparison, Readmission National Comparison, Patient Experience National Comparison, Effectiveness of Care National Comparison, Timeliness of Care National Comparison, and Efficient Use of Medical Imaging National Comparison.

Given that the analysis aims to discover the attributes that significantly impact the quality of hospital care, the Hospital Overall Rating was used as the class attribute. Weka filter was used to remove observations without a class attribute. The observations without a class attribute was due to the fact that there were too few measures or measure groups reported to calculate a star rating or measure group score.

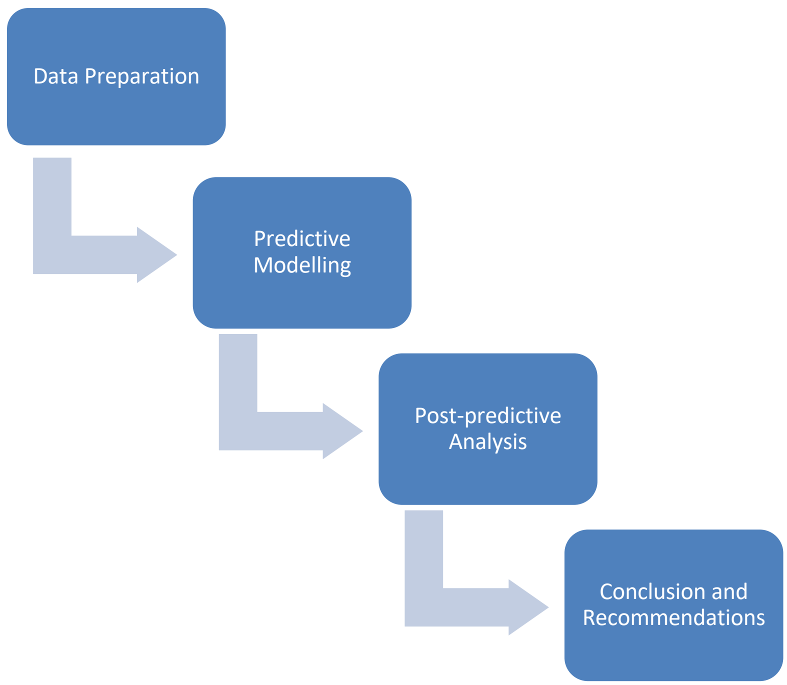
In addition, attribute footnotes were also removed given their redundancy and irrelevance to the analysis. Other attributes that were not pertinent to the scope of the analysis were removed (i.e., Provider ID, Hospital Name, Address, City, State, ZIP Code, County Name, and Phone Number). Once the observations with missing class, footnote, and irrelevant attributes were removed, the dataset consisted of 3399 observations and 12 attributes. Lastly, Weka filter was used to detect outliers and found no outliers (i.e., hospital overall rating was normal).

The following codes were used for attribute names:

Original Attribute Name	Abbreviated Attribute Name
Hospital Type	h_type
Hospital Ownership	h_ownership

Emergency Services	h_es
Meets criteria for meaningful use of EHRs	h_ehr
Hospital Overall Rating	h_rating
Mortality National Comparison	h_mortality
Safety of Care National Comparison	h_soc
Readmission National Comparison	h_ra
Patient Experience National Comparison	h_pex
Effectiveness of Care National Comparison	h_eoc
Timeliness of Care National Comparison	h_toc
Efficient Use of Medical Imaging National Comparison	h_imaging

Approach



Step 1: <Data Preparation>

Data was cleaned and processed using Weka filters and R script. Further cleaning may be necessary through the analysis.

https://github.com/fleejy/ckme136-capstone/

Step 2: <Predictive Modelling>

For predictive modelling, three classification algorithms will be explored: (1) decision tree, (2) Naïve Bayes, and (3) logistic regression.

1. Decision Tree

    a. The first classification algorithm to use
        i. Over-sample true instances
        ii. Down-sample false instances
            1. J48 classifier from Weka
    b. Model can handle categorical features correctly and the machine learning model processes categorical features correctly as categoricals.
2. Naïve Bayes
    a. The second classification algorithm to use
        i. Input values must be nominal (numerical inputs are supported by assuming a distribution)
        ii. This model is suitable because it can handle nominal attributes.
        iii. Given the use of simple implementation of Bayes Theorem (hence naïve) where the prior probability for each class is calculated from the training data and assumed to be independent of each other (conditionally independent), this assumption allows for faster and easier calculation of the probabilities.
3. Logistic Regression
    a. The third classification algorithm to use
        i. Converts binary classification problems into linear regression one
        ii. In this case, the model predicts the probability of hospital overall rating for a hospital by fitting data to a logit function with factors affecting hospitals' ratings.
4. Ensemble method
    a. This method will help improve machine learning results by combining the aforementioned models. This approach allows the production of better predictive performance compared to a single model.
    b. Ensemble method is a meta-algorithm that combine the three models into one predictive model in order to decrease variance (bagging), bias (boosting), or improve predictions (stacking).

Step 3: <Post-predictive Analysis>

1. Clustering analysis vs Association rule?
    a. Clustering requires quantitative whereas, association requires categorical attributes; therefore, association will be tested.

Step 4: <Conclusion and Recommendations>

Once the analysis is complete, the conclusion will summarize the results and recommendations will be given where detailed contextual descriptions of hospital's different components may be drawn to identify patterns that may have an impact on other outcomes of a health care system. In addition, recommendations will be made based on the findings of the comparative analyses that are aimed at system-level improvements.

References

[1] Mechanic, D., & Rochefort, D. A. (1996). Comparative medical systems. Annual Review of Sociology, 22(1), 242.

[2] Young, G. J., Meterko, M., & Desai, K. R. (2000). Patient satisfaction with hospital care: effects of demographic and institutional characteristics. Medical care, 325-334.

[3] Data.Medicare.gov. (2018, July 25). Hospital Compare datasets. Retrieved from https://data.medicare.gov/data/hospital-compare

[4] DeLancey, J. O., Softcheck, J., Chung, J. W., Barnard, C., Dahlke, A. R., & Bilimoria, K. Y. (2017). Associations Between Hospital Characteristics, Measure Reporting, and the Centers for Medicare & Medicaid Services Overall Hospital Quality Star Ratings. Jama, 317(19), 2015-2017.

[5] Tsai, T. C., Jha, A. K., Gawande, A. A., Huckman, R. S., Bloom, N., & Sadun, R. (2015). Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health Affairs*, *34*(8), 1304-1311.

[6] Bloom, N., Propper, C., Seiler, S., & Van Reenen, J. (2015). The impact of competition on management quality: evidence from public hospitals. *The Review of Economic Studies*, *82*(2), 457-489.

[7] Jones, S. S., Rudin, R. S., Perry, T., & Shekelle, P. G. (2014). Health information technology: an updated systematic review with a focus on meaningful use. *Annals of internal medicine*, *160*(1), 48-54.