

CKME 136 Data Analytics: Capstone Course  
Final Report  
Chang School of Continuing Education, Ryerson University

Title: The Effects of Perceived Quality and Satisfaction on Hospital Rating

Jae Yong (Francisco) Lee  
#500 883 333  
Supervisor: Dr. Karim El Mokhtari, [elmkarim@ryerson.ca](mailto:elmkarim@ryerson.ca)

## Table of Contents

<b><i>Introduction.....</i></b>	<b><i>3</i></b>
<b><i>Literature Review.....</i></b>	<b><i>3</i></b>
<b><i>Dataset .....</i></b>	<b><i>4</i></b>
<b><i>Approach .....</i></b>	<b><i>5</i></b>
Step 1: Data Preparation.....	5
Step 2: Predictive Modelling .....	7
1. Decision Tree.....	7
2. Naive Bayes .....	8
3. Logistic Regression.....	9
Step 3: Post-predictive Analysis .....	11
<b><i>Results and Recommendations .....</i></b>	<b><i>11</i></b>
<b><i>Conclusions.....</i></b>	<b><i>13</i></b>
<b><i>References .....</i></b>	<b><i>14</i></b>
<b><i>Appendix.....</i></b>	<b><i>15</i></b>
Figure 1.1: Summary Statistics of Relevant Numerical Variable .....	15
Figure 1.2: Kruskal-Wallis Rank Sum Test.....	15
Figure 1.3: Summary of key metrics, default dataset .....	15
Figure 1.4: Summary of key metrics, balanced dataset .....	15
Figure 1.5: Key Metrics with regards to the Size of Tree:.....	16
Figure 1.6: Final decision tree .....	16
Figure 1.7: Within Cluster Sum of Squared Errors .....	17

## **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.

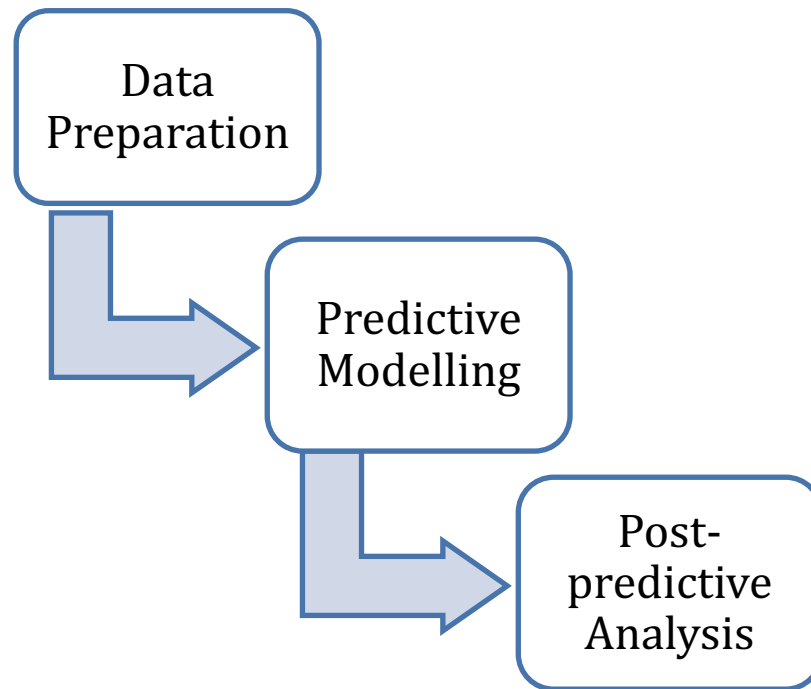
With regards to the meaningful use of electronic health records (EHRs), it is a challenge to validate the relevance as the external rules, regulations and pressures have influenced the credibility of such a metric. The recent passage of the American Recovery and Reinvestment Act (ARRA) of 2009, which 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. Recent studies reveal strong evidence that supports the superficial validation and implementation of the meaningful use of EHRs criteria, as such, the attribute regarding meaningful use of EHRs will be eliminated [7].

## **Dataset**

The dataset used in this analysis is sourced from the Hospital ratings dataset from Kaggle [8]. 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. The criteria to include and/or exclude certain attributes will be further discussed in the Approach section, in particular with reference to redundant attributes.

## Approach



### Step 1: Data Preparation

The detailed approach accompanied by the respective programming can be found on the GitHub page (<https://github.com/fleejy/ckmel36-capstone/>).

The dataset used in this analysis consisted of 4812 observations and 28 attributes. Based on literature review, the *Hospital overall rating* variable which act as the key prediction variable was converted to a nominal variable. Other nominal variables include: Hospital Name, Address, City, State, County Name, Hospital Type, Hospital Ownership, Emergency Services, Meets criteria for meaningful use of EHRs, Hospital overall rating footnote, Mortality national comparison, Mortality national comparison footnote, Safety of care national comparison, Safety of care national comparison footnote, Readmission national comparison, Readmission national comparison footnote, Patient experience national comparison, Patient experience national comparison footnote, Effectiveness of care national comparison, Effectiveness of care national comparison footnote, Timeliness of care national comparison, Timeliness of care national comparison footnote, Efficient use of medical imaging national comparison, and Efficient use of medical imaging national comparison footnote.

The numeric variables consist of Provider ID, ZIP Code, and Phone Number. They were of no significance to the purpose of this analysis; therefore, the overall dataset is qualitative such that summary statistics was not appropriate. Consequently, Principal Component Analysis (PCA)

could not be conducted for dimensionality reduction; however, Multiple Factor Analysis (MFA) which is dedicated to datasets where variables are structured into groups was considered [9]. However, once the missing data and outlier treatment was conducted the dimensionality of the dataset was reduced to manageable size.

From the Exploratory Data Analysis (EDA), the primary objective was to characterize hospital overall quality rating and to find a typology of the hospital overall quality rating. Attributes that did not contribute to the modelling purpose were eliminated (i.e., Provider ID, Hospital Name, Address, City, State, ZIP Code, County Name, Phone Number, and all attributes' footnotes) [Figure 1.3].

In addition, outlier detection was not suitable for a categorical dataset; therefore, chi-squared test was conducted for all attributes. Chi-squared test was chosen because it is able to determine if the attributes are dependent or not, given that both response and predictors are categorical variables (i.e., hospital overall rating and all other attributes) [Figure 1.2]. For all attributes except Emergency Services attribute, the null hypotheses were rejected because the p-values of the observed H statistics were smaller than 0.05. This indicates that the attributes differ in their overall hospital rating, suggesting a significant effect in the overall rating. As such, these attributes were included in the analysis. On the other hand, Emergency Services' null hypothesis was retained because the p-value of the observed H statistic was larger than 0.05 which suggest that the attribute was not significant in its effect in relation to the overall hospital rating. As such, the Emergency Services attribute was removed.

With regards to missing data, the Hospital Type required careful analysis with respect to the national comparison attributes. It is worth noting that all instances of the Critical Access Hospitals and Children's Hospital had missing data for the national comparison attributes.

This may be explained through the nature of these hospital types. Critical Access Hospital (CAH) is a designation given to eligible rural hospitals by the Centers for Medicare and Medicaid Services (CMS). Furthermore, any rural hospitals that are CAH designated receive certain benefits, such as cost-based reimbursement for Medicare services which help to reduce their financial vulnerability while improving access to healthcare by keeping essential services in rural communities. As such, it is expected any tracking of information was not available; therefore, any hospital types of CAH was removed. For Children's Hospital, the confidentiality clause of minor prevents the collection of any metrics required for this analysis. As such, Children's Hospital instances were further removed.

The complexity Children's Hospital and CAH are both instance specific such that they are deemed unsuitable for the purpose of this analysis. Furthermore, the uniqueness of both Children's hospitals and CAH require dedicated studies. Given these characteristic and elimination of the two classes, the Hospital Type attribute was additionally removed.

After the data was cleaned, there was a total of 2200 observations and 9 attributes and the following abbreviations were used for attribute names:

Original Attribute Name	Abbreviated Attribute Name
Hospital Ownership	h_ownership
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

The dataset showed significant imbalanced distribution with 1588 False for not HHQ (i.e., LHQ) and 612 True for HHQ. I used *ClassBalancer* in Weka filter to equalize the number of False to make the ratio of the two classes equal. *ClassBalancer* reweighs the instances in the data so that each class has the same total weight, while maintaining the total sum of weights across all instances. Therefore, the final number of samples was 2200 (i.e., 1100 for both HHQ and not HHQ while maintaining 612 True for HHQ and 1588 False for HHQ). All the attributes were statistically significant, and with no missing values in the dataset, the dataset was considered clean [Figure 1.3].

## Step 2: Predictive Modelling

### 1. Decision Tree

The first classification algorithm used for predictive modelling was the decision tree model. With the prepared dataset, initial assessment of the models displayed high accuracy rates around 90% [Figure 1.3]. The primary metric of recall was significantly lower compared to the overall accuracy rate. From the initial assessment, I concluded that the model was biased toward the false instances as a consequence of the imbalance in distribution of the class attribute. The imbalance resulted in high true negative rate at the cost of low true positive rate.

Two approaches were explored to resolve the imbalance issue. The first approach was to over-sample true instances, and the second approach was to down-sample false instances. The over-sampling option was discarded because there was large enough dataset to down-sample from and oversampling generally makes overfitting more likely. In order to overcome the imbalance, *ClassBalancer* in Weka filter was used to equalize the number of False to make the ratio of the two classes equal. Proceeding with the class-balancing approach, the instances were reweighed such that each class had the same total weight, while maintaining the total sum of weights across all instances.

With the balanced dataset, J48 Classifier from Weka was used to construct decision tree model. Within different parameters present in Weka, *minNumObj* parameter was used to obtain models with different tree sizes. Summary of key metrics on [Figure 1.5 and Figure 1.6] shows that class-balancing significantly enhanced the recall of the model. The model with the second smallest tree size among six models with various tree sizes was chosen, because the model displayed highest recall, maintained acceptable rate of accuracy, and was most comprehensible. The recall of the final model was 83.20% and the accuracy was 82.64%.

The final decision tree model consisted of nineteen leaf nodes with three attributes as decision rules [Figure 1.6]. The three attributes that played vital role in the decision-making process were Safety of Care, Readmission, and Patient Experience. Hospitals with Safety of Care above national average, hospitals with Readmission below national average, and hospitals with Patient Experience above national average were more likely to have HHQ.

## 2. Naive Bayes

The second classification algorithm used for predictive modelling was the Naive Bayes model, which assumes that the input values are nominal (numerical inputs are supported by assuming a distribution). Naive Bayes uses a simple implementation of Bayes Theorem (hence naive) where the prior probability for each class is calculated from the training data and assumed to be independent of each other (conditionally independent), which is an unrealistic assumption because we expect the variables to interact and be dependent. However, such an assumption allows for faster and easier calculation of the probabilities.

With the initial dataset, the assessment of the models displayed high accuracy rates around 87%; however, the recall and accuracy varied depending on the type of test option which will be explained in a following section. This accuracy is promising because Naive Bayes model is can handle categorical inputs by default, as such the accuracy is representative of the model's strength.

For this data set, Naive Bayes is suitable. This is due to the strength of Naive Bayes in handling nominal attributes. If the attributes numeric, the model would require treatments such as (1) discretizing the numeric variables or (2) using probability density functions. Discretizing the continuous variable would be a simple solution; however, it is subjective which causes loss of information, but it is used as a quick way to prepare data before applying Naive Bayes classification.

By default, a Gaussian distribution is assumed for each numerical attribute in Weka. In order to handle numerical attributes, the numerical attributes are automatically converted to nominal attributes with the *useSupervisedDiscretization* parameter set as TRUE. The particular implementation of Naive Bayes algorithm would be changed to use a kernel estimator with the *useKernelEstimator* argument which can be used to better match the actual distribution of the



attributes in datasets. It is important to note that *useKernelEstimator* and *useSupervisedDiscretization* parameters are mutually exclusive.

After conducting various test options on the class-balanced dataset, the test option which produced the best recall rate and accuracy was 70% split.

Test Option	Recall (T)	Accuracy
70% split	89.6%	85.34%
Use training set	87.6%	84.59%
10-fold CV	87.6%	84.47%
3-fold CV	87.4%	84.33%

*Summary of Test Options*

Going back to data preparation, in order to treat for imbalanced class distribution, I utilized class-balancing which resulted in limited number of samples; therefore, I ruled out using the training set as the test option due to its small sample size. More importantly, training set was not used because it produced generalization error (i.e., using only one data set, the model could achieve high accuracy by simply learning this particular set, but not the general concept). In accordance with the former reasons, the percentage splits were ruled out due to generalization error [10]. In addition, the use of training/test sets and cross-validation are conceptually equivalent. Cross-validation simply takes a more rigorous approach by averaging over the entire data set. Consequently, the  $n$ -fold cross-validation (CV) was chosen [11].

Between the 10-fold and 3-fold CV, the 10-fold CV was chosen over 3-fold because a larger  $n$ -fold has less bias towards overestimating the true expected error (i.e., training folds closer to the total dataset), but higher variance and running time (i.e., approaches the limit case of leave-one-out CV) [12]. In addition, higher  $n$  gives you more samples to estimate a more accurate confidence interval on your estimate [13].

Based on the classification using Weka, I have found that the 10-fold cross validation was most suitable with a recall rate of 87.6% and accuracy of 84.47%.

### **3. Logistic Regression**

Besides Decision Tree and Naive Bayes, logistic regression was chosen as the third classification algorithms to build the predictive model. Logistic regression converts binary classification problems into linear regression ones. That is, by means of proper transformation, it is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables. In our case, it predicts the probability of hospital overall quality for a hospital by fitting data to a logit function with factors affecting hospitals' ratings.

I have found that for logistic regression in particular, there was of little benefit to creating a balanced sample. Compared to balancing the class, using the original datasets containing all the information available improves classifier performance. Thus, I fitted the baseline model in Weka with 70% training set from the original dataset, with outliers removed and attributes with little contribution to the model construction dropped.

From the model summary, the estimated coefficients of *Patient Experience National Comparison*, *Mortality National Comparison*, and *Readmission National Comparison* are statistically significant variables suggesting strong associations with the probability of high hospital quality. The logistic regression coefficients give the change in the log odds of the outcome for a one unit increase in the predictor variable. For example, the positive coefficient for *Readmission* predictor suggests that all other variables being equal, the hospitals who have *Readmission* predictor are more likely to be a high-quality hospital.

When selecting the best performing algorithm, the goal is to select the algorithm which maximizes the true positives and true negatives because the dataset has imbalanced class distribution. If the accuracy is the only metric of evaluation, logistic regression seems to be the best model to make the predictions. However, accuracy works best if true positives and true negatives have similar cost. In our case, the cost of true positives and true negatives are very different, so it is better to look at both precision and recall since we are emphasizing on hospitals which are more likely to have high hospital quality rating.

The conclusions I drew from logistic regression is consistent with the analysis I did from the decision tree—hospitals with Mortality rate above the national average, Satisfaction of Care above the national average, and Readmission rate above the national average are more likely to have high hospital quality rating.

However, even though decision tree provides higher recall than logistic regression, logistic regression outperforms other classifiers in all other metrics (in particular the ROC area). As discussed in the previous section, Naive Bayes is not suitable for our dataset. Thus, logistic regression was selected as the best performing algorithm in this case.

The comparison is shown as follows:

	Accuracy	Precision (True)	Recall (True)	ROC Area
Decision Tree	87.49%	84.4%	91.8%	0.911
Logistic Regression (on full dataset with classification threshold = 0.5)	86.89%	85.6%	88.5%	0.943
Logistic Regression (on further eliminated dataset)	87.85%	85.3%	91.5%	0.944

### Step 3: Post-predictive Analysis

For the post-predictive analysis, clustering analysis was selected based on the purpose of the analysis which is to find groups of hospitals with high quality rating. The association rule was not examined, because the analysis of discovering relationships between attributes was not suitable.

To successfully identify the characteristics of high-quality hospitals, I applied SimpleKMeans Clusterer to the filtered dataset with different numbers of clusters.

I examined six different models with number of clusters ranging from two to seven. As shown in [Figure 1.7], the sum of squared errors decreased the most when the number of clusters was increased from two to three. It was decided that the best number of clusters is three because adding another cluster beyond this point did not give significantly better modelling of the data.

On average, cluster 1 showed higher Mortality rate than the national average, cluster 2 showed higher Satisfaction of Care than the national average, and cluster 3 showed higher Readmission rate than the national average.

#### Recommendations:

- Cluster 1: One possible explanation behind this cluster is that there are patients experiencing problems with referrals, so they tend to have longer wait times until they reach the specialist that usually reside in a tertiary care center. I recommend further technical analysis on this cohort such that I can identify methods to expedite the referral process and surveying patients on their experience with that process.
- Cluster 2: I suspect patients of this hospital are satisfied because their overall experience with the patient relations was very high. I advise implementing loyalty plans to ensure the continued satisfactory patient relations for frequent patients.
- Cluster 3: The reason behind hospitals with high Readmission rate could be current patient management delivering communication functionality and quality. I recommend auditing the current patient plan to ensure its success.

### Results and Recommendations

Having the ability to predict and act on insights for hospital quality rating is a priority for any effective hospital. This can be achieved through data analytics and optimizing workflows based on those findings. Through my analyses, I have found several key points that may be of significant contribution to further developing such an ability to understand our customers leading to robust business execution.

After initial analysis, the data was prepared for investigation. The nineteen non-relevant attributes were removed from 28 total attributes resulting in 9 attributes and 1 class attribute (hospital rating). In addition, the dataset showed significant imbalanced distribution with 1588 False for not HHQ (i.e., LHQ) and 612 True for HHQ. I used *ClassBalancer* in Weka filter to equalize the number of False to make the ratio of the two classes equal. Therefore, the final number of samples was 2200 (i.e., 1100 for both HHQ and not HHQ while maintaining 612 True for HHQ and 1588 False for HHQ). All the attributes were statistically significant, and with no missing values in the dataset, the dataset was considered clean.

For predictive modelling, three types of classifiers were tested: (1) Decision Tree, (2) Naive Bayes, and (3) Logistic Regression. The explanation on criterion for ruling out the classifier is detailed in the classifier's respective section. The Naive Bayes classifier was ruled out due to its assumptions which were not suitable for our analysis. The Decision Tree was ruled out because the accuracy of the model was high with the recall being disproportionately low. The Logistic Regression classifier was chosen as the most suitable classifier for our analysis, because the model displayed highest recall, maintained acceptable rate of accuracy, and was most comprehensible.

For the logistic regression classifier, I used recall (i.e., recall on true instances) and accuracy (i.e., proportion of correct classifications among all classifications) which assigns equal cost to false positives and false negatives. When the data set is imbalanced, the accuracy is a poor measure; however, the data set was balanced. Thus, the problem was eliminated. The recall of the final model was 91.5% and the accuracy was 87.85%.

For the post-predictive analysis, clustering analysis was selected based on the purpose of the analysis which is to find groups of hospitals with high quality rating. The association rule was not examined, because the analysis of discovering relationships between attributes was not suitable. Using the association rule would result in suboptimal performance of the model (i.e., association rule requires categorical attributes that have more than the grouping currently present in the dataset of interest). Using k-means clustering algorithm, I examined six different models with number of clusters ranging from two to seven. When the number of clusters was increased from two to three, the maximum decrease in sum of squared errors was observed. The best number of clusters is three, because adding another cluster did not significantly contribute to a better modelling of the data.

From the logistic regression, I have identified three key attributes: (1) Mortality rate, (2) Satisfaction of Care, and (3) Readmission rate. These attributes should be carefully monitored as they are the key attributes which lead to the high-quality outcome of hospitals. Therefore, I recommend monitoring hospital departments on these attributes to identify their likelihood of high-quality outcome. This insight is of uttermost importance to the efficacy of the hospital, because it is much less expensive to prevent illnesses than it is to treat. Preventive medicine is more cost-effective as the healthcare system can control risk factors; therefore, having insight to predict high quality outcome is critical to the delivery of hospital efficacy.

Based on the post-predictive analysis, hospitals and policy makers must pay careful attention to hospitals who portray a higher Mortality rate than the national average, a higher Satisfaction of Care than the national average, and a higher Readmission rate than the national average are more likely to be high quality hospitals.

- Prediction Model showed that these attributes lead to high quality ratings:
  - Mortality rate
  - Satisfaction of Care
  - Readmission rate
- Post-predictive Model demonstrated that high quality hospitals had these attributes:
  - Higher Mortality rate than the national average
  - Higher Satisfaction of Care than the national average
  - Higher Readmission rate than the national average

The findings from the prediction model and post-predictive analysis differ in that prediction model suggests which key attributes healthcare system must monitor to track potential hospitals that have high quality outcome and the post-predictive analysis suggests characteristics of hospitals that have high quality; therefore, providing insight into which healthcare system strategies must be employed to maintain quality healthcare system.

## **Conclusions**

There is complexity in evaluating the efficacy of a health care system due to a combination of multiple factors. As such, the patients' ratings of their hospital is the common proxy for quality of hospital care. The perceived quality and satisfaction of patients has the potential as an 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.

Unfortunately, we do not have any prediction model to gain insight on why certain hospitals underperform nor a metric to track and predict hospital quality. Given such metric's saliency, I have utilized predictive modelling and post-prediction analyses to gain insight into how we may be able to predict which hospital have a high-quality outcome. Based on logistic regression classifier, the recall on true instances, hereinafter simply referred to as recall, was 91.5% and the accuracy was 87.85% (accuracy defined as the proportion of correct classifications among all classifications). Based on k-means clustering, I have identified three clusters of hospital type which indicate characteristics of hospital that have high quality outcome. Based on these analyses, our healthcare system and policy makers must monitor the hospitals that possess/portray these characteristics: (1) high Mortality rate, (2) high Satisfaction of Care, and (3) high Readmission rate. This insight is critical to the success of the healthcare system, because determining which preventive measures and treatments are most efficient can bring about substantial aggregate improvements in health at an acceptable cost.

With ever increasing constraint on resources, there is a realistic way of achieving better health results: conduct careful analysis to identify evidence-based opportunities for more efficient delivery of health care, mainly through the monitoring of hospitals with high-quality outcome, and then restructure the system to create incentives that encourage the appropriate delivery of efficient interventions.

For subsequent analysis, the dataset can be tailored for the Canadian healthcare system. There may be opportunity to utilize the datasets available through the Canadian Institute for Health Information (CIHI). In addition, future analysis can look at other performance metrics. In particular, the sensitivity and F1 score which is based on the needs of this project. In terms of visualization, an overlay of the ROC for each instances of all classifiers would be helpful. For classifier implementation, the specific algorithms could be explored to optimize the performance.

## 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.
- [8] Kaggle (2018, September 25). Hospital Compare datasets. Retrieved from <https://www.kaggle.com/center-for-medicare-and-medicaid/hospital-ratings>
- [9] Escofier, B., & Pages, J. (1994). Multiple factor analysis (AFMULT package). *Computational statistics & data analysis*, 18(1), 121-140.
- [10] Nadeau, C., & Bengio, Y. (2000). Inference for the generalization error. In *Advances in neural information processing systems* (pp. 307-313).
- [11] Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, 5(Sep), 1089-1105.
- [12] Kearns, M., & Ron, D. (1999). Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. *Neural computation*, 11(6), 1427-1453.
- [13] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, 111-147.

## Appendix

**Figure 1.1: Summary Statistics of Relevant Numerical Variable**

Attributes	Max	Min	Mean	Standard deviation
Hospital Overall Rating (h_rating)	5.000	1.000	2.994	0.866

**Figure 1.2: Kruskal-Wallis Rank Sum Test**

Attributes (h_rating by)	Kruskal-Wallis chi-squared	Degrees of Freedom	P-value
h_mortality	136.15	2	< 2.2e-16
h_soc	453.12	2	< 2.2e-16
h_ra	673.62	2	< 2.2e-16
h_pex	719.14	2	< 2.2e-16
h_eoc	79.619	2	< 2.2e-16
h_toc	173.25	2	< 2.2e-16
h_imaging	11.624	2	0.002992
h_ownership	69.485	8	6.222e-12
h_es	0.053771	1	0.8166

**Figure 1.3: Summary of key metrics, default dataset**

Size of Tree	Recall(T)	Accuracy
16	66.80%	84.86%
16	65.20%	84.45%
88	71.40%	87.50%

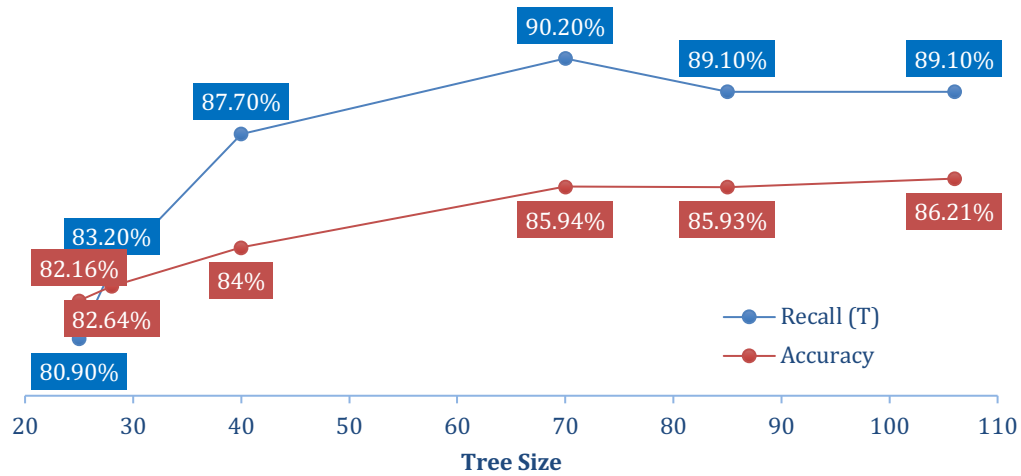
\*minNumObj = 50, 20, 2, respectively

**Figure 1.4: Summary of key metrics, balanced dataset**

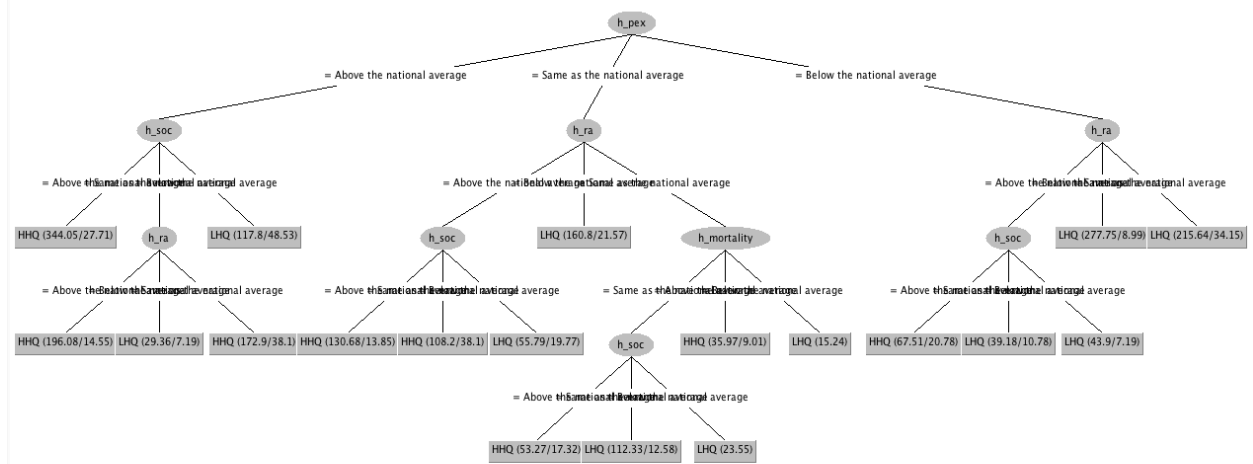
Size of Tree	Recall(T)	Accuracy
25	80.90%	82.16%
28	83.20%	82.64%
40	87.70%	83.92%
70	90.20%	85.94%
85	89.10%	85.93%
106	89.10%	86.21%

\*minNumObj = 40, 30, 20, 10, 5, 2, respectively

***Figure 1.5: Key Metrics with regards to the Size of Tree:***



***Figure 1.6: Final decision tree***





***Figure 1.7: Within Cluster Sum of Squared Errors***

