PGDDA – Healthcare Capstone project

Sasidhar K, Rajesh U, Sanjay M, Shashank K

Background

- CMS (Center for Medicare and Medicaid) is a federal agency within the United States Department of Health and Human Services that administers the Medicare program. CMS is also responsible for maintaining the consumer-oriented website called Hospital Compare that provides information on how well hospitals provide recommended care to their patients.
- This information can help consumers make informed decisions about where to go for health care. Hospital Compare allows consumers to select multiple hospitals and directly compare performance measure information related to heart attack, heart failure, pneumonia, surgery and other conditions.
- CMS rates hospitals in the US on a scale of 1-5 with the objective to make it easier for patients and consumers to compare the quality of hospitals.
- The ratings directly influence the choice of the hospital made by consumers and may have a significant impact on the revenue earned by hospitals. Thus, it is extremely important for hospitals to understand the methodology used by CMS for calculating the ratings so that they can work on improving the factors that influence them.

Problem Statement

- This project is focused on developing an approach to calculate hospital ratings and using it to identify areas of improvement for certain hospitals. The most important part is to identify 64 measures of quality for rating the hospitals and categorize them into 7 broad categories to arrive at a group score for each of the seven categories which can be used to evaluate the final star rating for any hospital under the purview of CMS.
- The three important parts of the problem are:
  - Data Understanding
    - Identify all 64 measures and categorize them into 7 broad categories.
  - Data Modelling
    - Create a supervised and unsupervised Machine learning models to arrive at the final rating
  - Recommendations for hospitals
    - Provide recommendations to hospitals on improvement opportunities to improve the overall rating

Approach

- The data (Hospital_Revised_FlatFiles_20161110) downloaded from hospital compare website has 57 csv format files. These files correspond to various quality related data for each of the 7 broad categories identified below. Some of them are structural and duplicate measures which need not be considering for arriving at the rating.

The methodology followed to calculate the Star Rating is listed below:

- Step 1: Selection and standardization of measures for inclusion in the Star Rating
- Step 2: Assignment of measures to groups
- Step 3: Perform Factor analysis on model group scores
- Step 4: Calculation of hospital summary scores as a weighted average of group scores
- Step 5: Application of clustering algorithm to categorize summary scores into star ratings

Step- 1: Measure Selection Criteria

- CMS used the following criteria to exclude measures from the Star Rating calculation
- Measures suspended, retired, or delayed from public reporting on *Hospital Compare*
- Measures with no more than 100 hospitals reporting performance publicly
- Structural measures are excluded
- Measures for which it is unclear whether a higher or lower score is better (non-directional)
- Measures not required for Inpatient Quality Reporting (IQR) Program or Outpatient Quality Reporting (OQR) Program
- Overlapping measures (for example, measures that are identical to another measure, or measures with substantial overlap in cohort and/or outcome).

Standardization of Measures

- The measure scores for all measures will be standardized and brought into one common scale.

Step 2: Assignment of measures to groups

Group	Number of Measures (N=64)
Mortality	7
Safety of Care	8
Readmission	8
Patient Experience	11
Effectiveness of care	18
Timeliness of care	7
Efficient use of Medical Imaging	5

Filename - Quality measure mapping	
File name (Measure Id)	**Measure description**
Healthcare Associated Infections (HAI_1_SIR)	SafetyOfCare (8)
Healthcare Associated Infections (HAI_2_SIR)	SafetyOfCare
Healthcare Associated Infections (HAI_3_SIR)	SafetyOfCare
Healthcare Associated Infections (HAI_4_SIR)	SafetyOfCare
Healthcare Associated Infections (HAI_5_SIR)	SafetyOfCare
Healthcare Associated Infections (HAI_6_SIR)	SafetyOfCare
Complications – Hospital (COMP_HIP_KNEE)	SafetyOfCare
Complications – Hospital (PSI_90_Safety)	SafetyOfCare
HCAHPS – Hospital (H_CLEAN_STAR_RATING)	PatientExperience(11)
HCAHPS – Hospital (H_COMP_1_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_COMP_2_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_COMP_3_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_COMP_4_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_COMP_5_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_COMP_6_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_COMP_7_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_HSP_RATING_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_QUIET_STAR_RATING)	PatientExperience
HCAHPS – Hospital (H_RECMND_STAR_RATING)	PatientExperience
Readmissions and Deaths – Hospital (READM_30_HF)	Readmission(8)
Readmissions and Deaths – Hospital (READM_30_CABG)	Readmission
Readmissions and Deaths – Hospital (READM_30_COPD)	Readmission
Readmissions and Deaths – Hospital (READM_30_HIP_KNEE)	Readmission
Readmissions and Deaths – Hospital (READM_30_STK)	Readmission

Quality measure - Filename mapping	
File name (Measure Id)	**Measure description**
Readmissions and Deaths – Hospital (READM_30_PN)	Readmission
Readmissions and Deaths – Hospital (READM_30_HOSP_WIDE)	Readmission
Readmissions and Deaths – Hospital (READM_30_AMI)	Readmission
Readmissions and Deaths – Hospital (MORT_30_HF)	Mortality(7)
Readmissions and Deaths – Hospital (MORT_30_CABG)	Mortality
Readmissions and Deaths – Hospital (MORT_30_COPD)	Mortality
Readmissions and Deaths – Hospital (MORT_30_PN)	Mortality
Readmissions and Deaths – Hospital (MORT_30_STK)	Mortality
Readmissions and Deaths – Hospital (MORT_30_AMI)	Mortality
Complications - Hospital (PSI_4_SURG_COMP)	Mortality
Timely and Effective Care – Hospital (OP_5)	Timeliness of Care(7)
Timely and Effective Care – Hospital (OP_21)	Timeliness of Care
Timely and Effective Care – Hospital (OP_3b)	Timeliness of Care
Timely and Effective Care – Hospital (ED_1b)	Timeliness of Care
Timely and Effective Care – Hospital (ED_2b)	Timeliness of Care
Timely and Effective Care – Hospital (OP_18b)	Timeliness of Care
Timely and Effective Care – Hospital (OP_20)	Timeliness of Care
Outpatient Imaging Efficiency - Hospital (OP_8)	Efficient Use of Medical Imaging(5)
Outpatient Imaging Efficiency - Hospital (OP_10)	Efficient Use of Medical Imaging
Outpatient Imaging Efficiency - Hospital (OP_14)	Efficient Use of Medical Imaging
Outpatient Imaging Efficiency – Hospital (OP_13)	Efficient Use of Medical Imaging
Outpatient Imaging Efficiency – Hospital (OP_11)	Efficient Use of Medical Imaging

Quality measure - Filename mapping	
File name (Measure Id)	**Measure description**
Timely and Effective Care – Hospital (STK_1)	Effectiveness of Care
Timely and Effective Care – Hospital (STK_6)	Effectiveness of Care
Timely and Effective Care – Hospital (STK_8)	Effectiveness of Care
Timely and Effective Care – Hospital (VTE_1)	Effectiveness of Care
Timely and Effective Care – Hospital (VTE_2)	Effectiveness of Care
Timely and Effective Care – Hospital (VTE_3)	Effectiveness of Care
Timely and Effective Care – Hospital (IMM_2)	Effectiveness of Care
Timely and Effective Care – Hospital (IMM_3_OP_27_FAC_ADHPCT)	Effectiveness of Care
Timely and Effective Care – Hospital (OP_4)	Effectiveness of Care
Timely and Effective Care – Hospital (OP_29)	Effectiveness of Care
Timely and Effective Care – Hospital (OP_30)	Effectiveness of Care
Timely and Effective Care – Hospital (PC_01)	Effectiveness of Care
Timely and Effective Care – Hospital (OP_22)	Effectiveness of Care
Timely and Effective Care – Hospital (VTE_6)	Effectiveness of Care
Timely and Effective Care – Hospital (CAC_3)	Effectiveness of Care
Timely and Effective Care – Hospital (OP_23)	Effectiveness of Care
Timely and Effective Care – Hospital (HF_2)	Effectiveness of Care

Step 3: Factor analysis

- Perform factor analysis on model group scores on the combined dataset.
- Using factor analysis, we will calculate loadings for each measure in the group which signifies the national performance on the measure and the measure's relationship with other measures in the group and the group's latent variable.

Step 4: Calculation of hospital summary scores

- Groups will be assigned weights as per the table below. The group weights has been arrived at by CMS by taking feedback from TEP and other stakeholders.

Group	Star ratings weight
Mortality	22%
Effectiveness of care	4%
Safety of care	22%
Readmission	22%
Patient experience	22%
Timeliness of care	4%
Efficient use of Medical Imaging	4%

Winsorization

- Winsorization is performed on the group summary scores to set extreme outliers to a specified percentile of data.
- The lower limit will be set to 0.5 percentile and higher limit will be set to 99.5 percentile.

Step 5: Application of clustering algorithm

- K-means clustering algorithm will be used in order to bucket the group summary scores into 5 different buckets
- As part of grouping the measures, the minimum threshold for each group is 3 to attain a reliability of 0.75.

Data Preparation and cleaning

- The data shall be cleansed for missing values. Some files have “NA” and some have “Not available” as column data. The missing value imputation techniques have to be applied uniformly across all datasets.
- The data shall be treated for outliers using quantile() function in R
- The file Readmissions and Deaths – Hospital.csv has measures listed as column values. The MeasureId column has to be transposed.
- The data shall be standardized for all measures using scale() function in R
- Finally, the data across all datasets shall be merged together into a single dataset containing all 64 measures
- Further, factor columns can be converted to continuous variables by creating dummy variables.
- The final training dataset should ideally contain all 64 parameters required for arriving at the CMS 5 star rating.
- Discard the measures which is having rating from less than 100 hospitals.
- EDA is performed on all continuous and categorical variables

Supervised Data Modelling

- The target variable (Hospital overall rating) for supervised modelling is obtained from Hospital general Information.csv file.
- A regression Random forest model will be used to predict the 5 star rating for a given hospital.
- Variable importance plot is also a useful tool and can be plotted using **varImpPlot** function. Top 5 variables will be selected and plotted based on Model Accuracy and Gini value.

Recommendations for Hospital

- The hospital with Provider ID = 140010 (EVANSTON HOSPITAL) has a rating of 3. So we will check the group summary scores for this hospital and select the group which has the least score
- From the group we will arrive at the measures that have low scores which can be used to recommend for improvement.