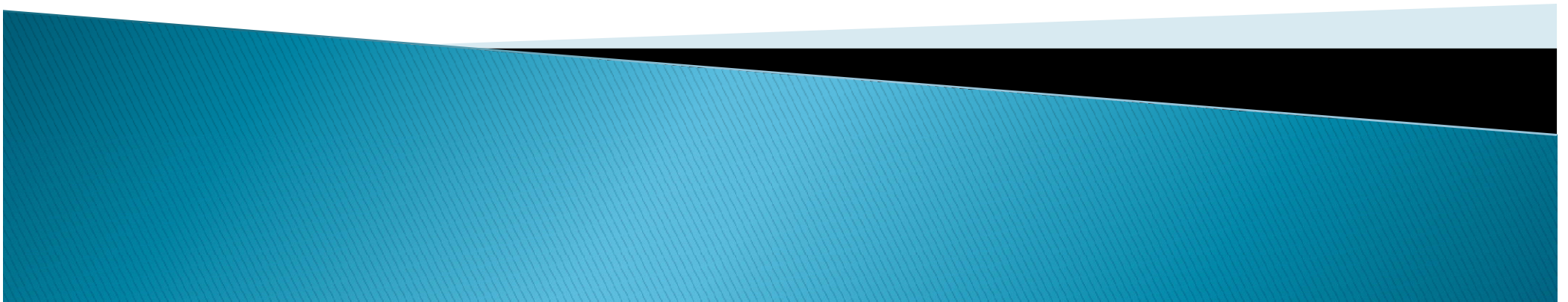# Healthcare Capstone Project
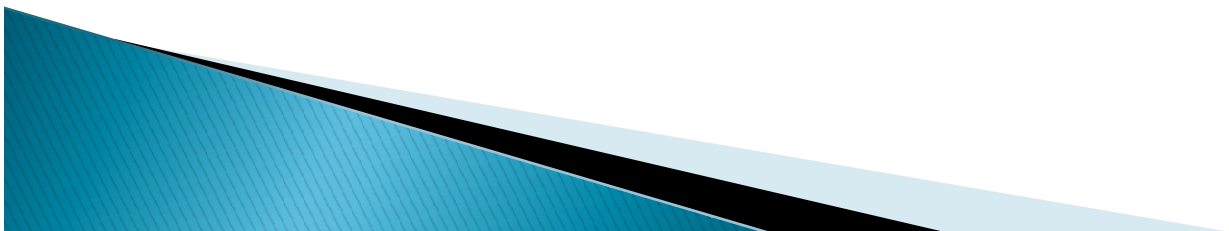
Hospital Rating Prediction Model

By Ruchika Desure

Business Problem
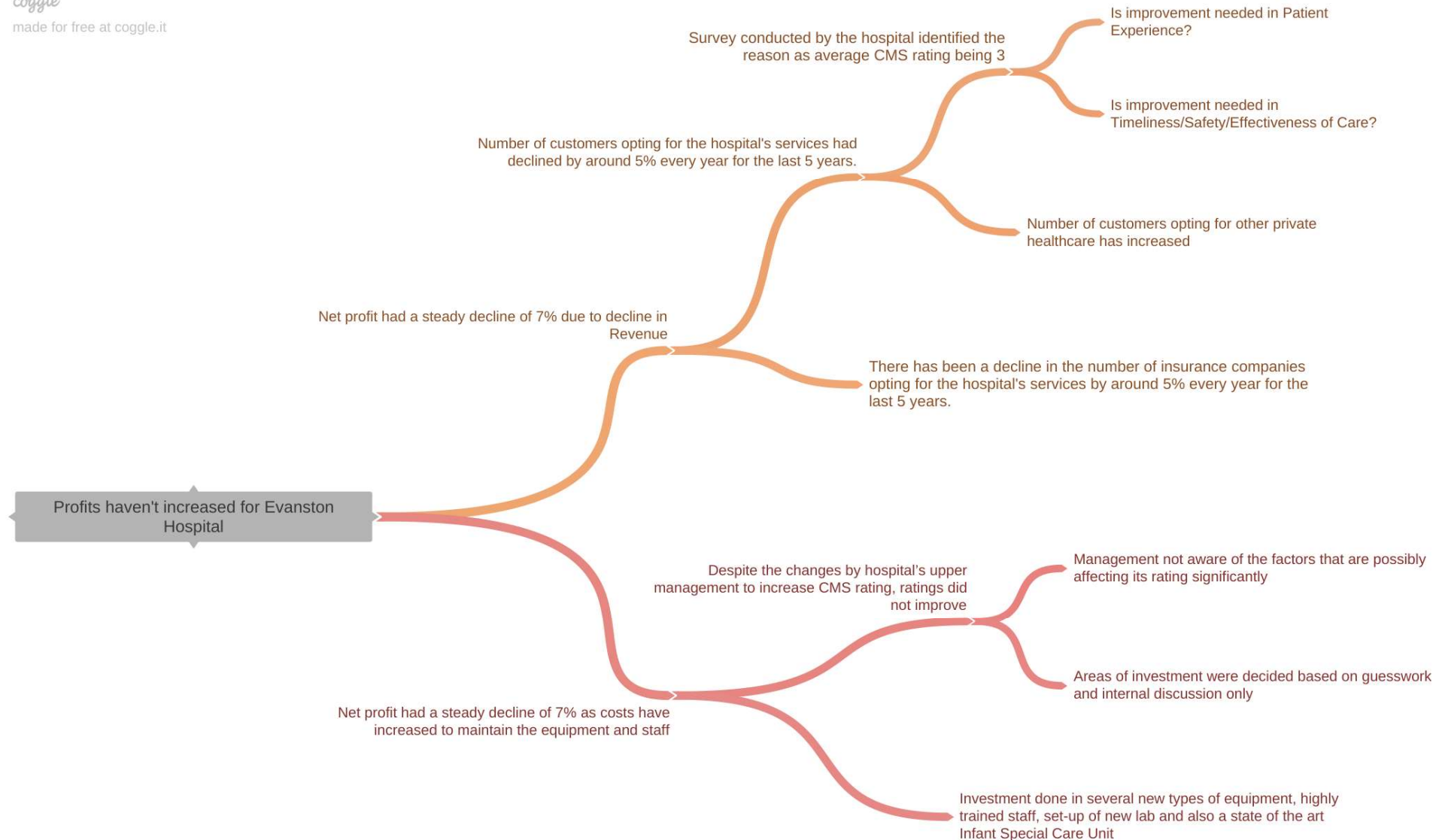
▸ The Centres for Medicare & Medicaid Services (CMS), rates the US hospitals on a scale of 1-5. Their objective is to make it easier for patients and consumers to compare the quality of services offered by various hospitals.

▸ Evanston Hospital is a comprehensive acute-care facility in Illinois, US. Despite spending a considerable amount of resources on improving its services, the hospital's CMS rating has remained at 3 for the past 5 years. This has led to a steady decline in revenue for the hospital.

▸ Main task is to identify possible root causes for the hospital getting such an average rating and recommend measures to mitigate this problem.

▸ A rating prediction model can be built and used to determine those factors.

Root Cause Analysis

Survey conducted by the hospital identified the reason as average CMS rating being 3

Is improvement needed in Patient Experience?

Is improvement needed in Timeliness/Safety/Effectiveness of Care?

Number of customers opting for the hospital's services had declined by around 5% every year for the last 5 years.

Number of customers opting for other private healthcare has increased

Net profit had a steady decline of 7% due to decline in Revenue

There has been a decline in the number of insurance companies opting for the hospital's services by around 5% every year for the last 5 years.

Profits haven't increased for Evanston Hospital

Despite the changes by hospital's upper management to increase CMS rating, ratings did not improve

Management not aware of the factors that are possibly affecting its rating significantly

Areas of investment were decided based on guesswork and internal discussion only

Net profit had a steady decline of 7% as costs have increased to maintain the equipment and staff

Investment done in several new types of equipment, highly trained staff, set-up of new lab and also a state of the art Infant Special Care Unit

3

Understanding the Data

▸ After identifying the necessary datasets, data was cleaned, important columns were kept and 62 measures were considered.

▸ Data from multiple datasets were combined into a single data frame. Cleaned data had 3648 rows of data.

Data Standardization

▸ Hospital quality measures include different types of scoring information. Some measures are in time, some are in terms of percentage, etc. Quality measures also have two directions - either "lower is better" or "higher is better". Therefore, each measure should be converted into a common scale with a common direction of measurement.

▸ The measure scores are standardize by calculating the Z-score for each measure. Also, the direction would be reversed if necessary while evaluating the values.

$$Z\text{-score} = (x - \text{mean}) / \text{standard deviation}$$

Winsorization

To avoid extreme outlier performance, data was winsorized at a Z-score of -3 and +3. z-scores below -3 were set to -3 and z-scores above 3 were set to 3.

Model Building

▶ **Stratified KFold Cross Validation** – For forming the train-test data sets, this validation method works perfectly well for imbalanced data having multiple classes. Each fold in stratified cross-validation will have a representation of data of all classes in the same ratio as in the whole dataset.

▶ The performance of various classifiers with their default parameter values were first seen on the available data.

▶ Started with Logistic Regression first as this is a classification problem. The one-vs-rest (OvR) scheme has been used as it is a multiclass case.

▶ Below are the models considered:

Logistic Regression (with OVR)

Random Forest

Decision Tree

Linear Support Vector Classification

Logistic Regression model has the highest accuracy as well as F1-score among these and therefore it was further hyperparameter tuned to improve model performance.

Model Evaluation

	Model	Accuracy	Precision (Weighted Avg)	Recall (Weighted Avg)	F1-Score (Weighted Avg)
0	Logistic Regression (with OVR)	0.7933	0.8092	0.7933	0.7806
1	Random Forest	0.7892	0.7995	0.7892	0.7787
2	Decision Tree	0.6689	0.6718	0.6689	0.6691
3	Linear Support Vector Classification	0.7695	0.7834	0.7695	0.7582
4	Logistic Regression Hyperparameter Tuning	0.8658	0.8737	0.8658	0.8662

- The Accuracy, Precision, Recall and F1-Score has increased after tuning the hyperparameters of the Logistic Regression model (with OVR).

- Accuracy is 0.87

- Precision (Weighted Avg) is 0.87

- Recall (Weighted Avg) is 0.87

- F1-Score (Weighted Avg) is 0.87

- The above scores are decent. Hence, finalizing this model to predict the Hospital overall rating.

Provider Analysis

Feature Importance of each Measure
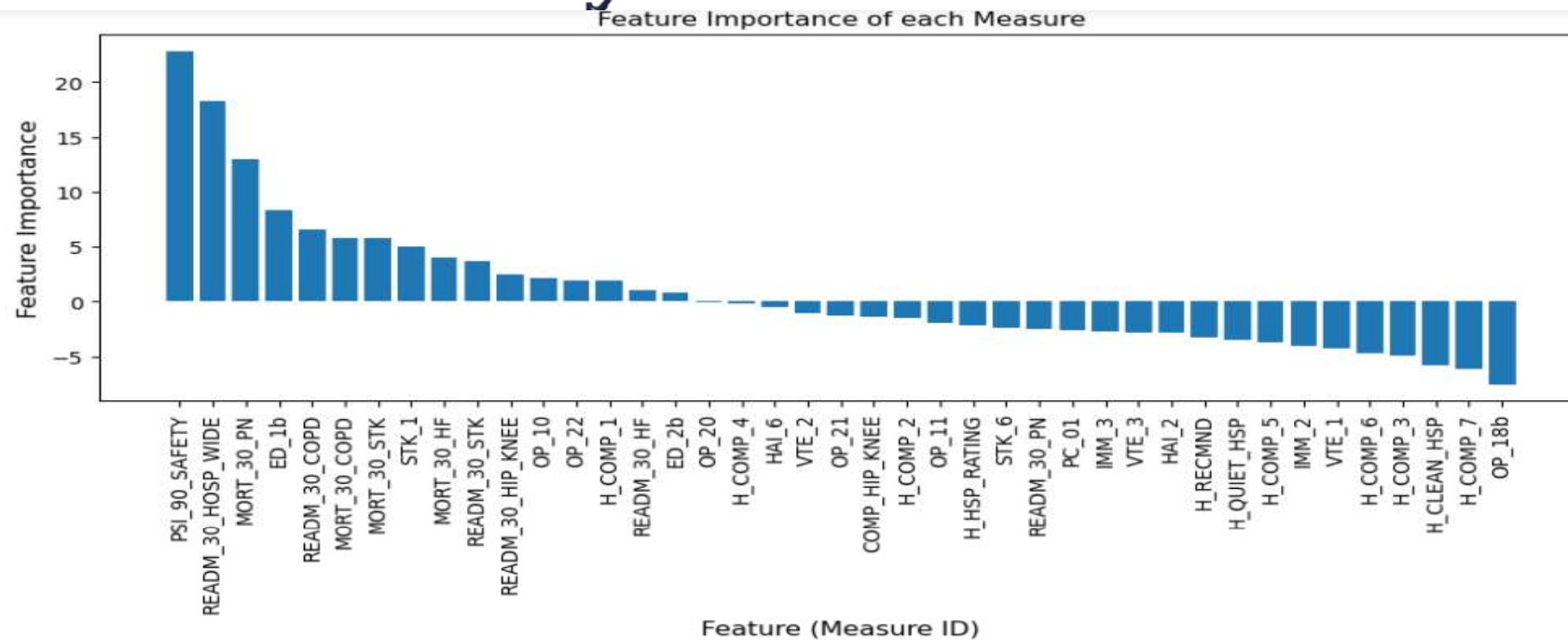
- For Evanston Hospital, the final model correctly predicted the Hospital overall rating i.e. 3 from the current measure scores.
- A dataframe was created to store the median value of each measure ID of all the hospitals having current rating as 4
- The Evanston hospital current measure scores where compared with these median values.
- Then the some features with positive feature importance and some with negative feature importance were considered to be improved.

The following Measure IDs were updated for Evanston Hospital:

- PSI_90_SAFETY -> 3 to -0.234236
- MORT_30_PN -> -1.600837 to -0.36813
- MORT_30_COPD -> -2.176896 to -0.165093
- MORT_30_STK -> -1.680969 to -0.0665
- MORT_30_HF -> -1.840738 to -0.183862
- OP_18b -> 0.543667 to 0.353265
- H_COMP_7 -> -0.675235 to 0.740328
- H_COMP_3 -> -0.284223 to 0.792852
- H_COMP_6 -> -0.373408 to 1.085093
- H_COMP_5 -> -1.240653 to -0.094493

▸ Positive Feature Importance–
PSI_90_SAFETY, MORT_30_PN,
MORT_30_COPD, MORT_30_STK,
MORT_30_HF

▸ Negative Feature Importance –
OP_18b, H_COMP_7, H_COMP_3,
H_COMP_6, H_COMP_5

With the improved scores, the final model predicted the Hospital overall
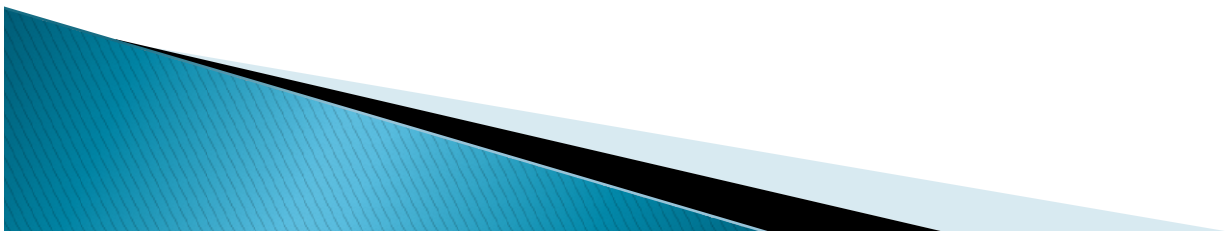rating as 4 for Evanston Hospital!
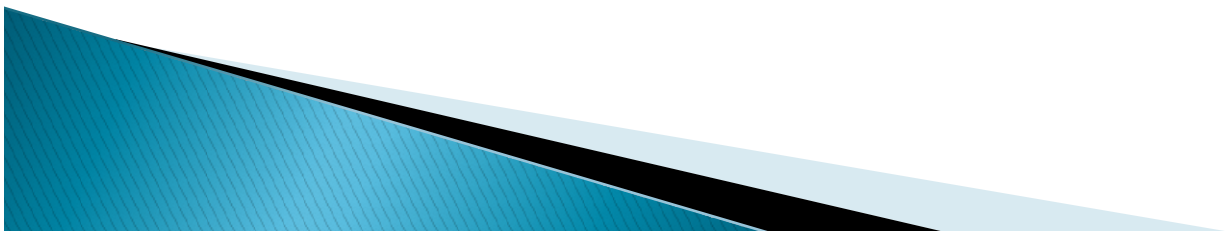
Recommendations

- Evanston Hospital should improve the below few measures to increase it's rating from 3 to at least 4.

  - PSI_90_SAFETY -> Serious complications measure PSI_90 is composed of 11 other measures. This needs to be low but is currently very high for Evanston Hospital. Hospital needs to level up inpatient hospital care and see to it that patients do not get serious complications.

  - MORT_30_PN -> Death rate for pneumonia patients, this needs to be as low as possible. Pneumonia patients need to be handled with extra care.

  - MORT_30_COPD -> Death rate for chronic obstructive pulmonary disease (COPD) patients. COPD patients need to be handled with extra care.

  - MORT_30_STK -> Death rate for stroke patients. This should be low. STK patients need to be handled with extra care.

  - MORT_30_HF -> Death rate for heart failure patients. This should be low. Patients with the risk of heart failure need to be prioritized.

  - OP_18b -> Average (median) time patients spent in the emergency department before leaving from the visit. This need to be low. Emergency department needs to increase its services.

- H_COMP_7 -> Care Transition. This needs to be high. The hospital needs to make sure patients understood their care when they leave the hospital.

- H_COMP_3 -> Staff responsiveness. This needs to be high. The hospital needs to make sure patients receive help as soon as they want it. Hospitals should train the staff accordingly.

- H_COMP_6 -> Discharge Information. This needs to be high. The hospital needs to make sure patients receive information about what to do during their recovery at home. Hospitals should train and keep some resources specifically to do this task.

- H_COMP_5 -> Communication about medicine. This needs to be high. The hospital needs to make sure staff explain about the medicines to the patients before giving it to them. Hospitals should inform this to the appropriate staff resources.

▸ Thus, following are the groups that the Evanston Hospital's management need to focus on:
  - Safety of Care
  - Mortality
  - Patient Experience
  - Timeliness of Care

Business Impact

- Advantages of The proposed 'Rating Prediction' model are –
  - It can help to compare the hospital data and identify the features that need improvement.
  - It can also be used to see the feature areas which the hospitals are performing good and maintain the same.
  - The Feature Importance gives us the importance of each feature (in this case the Measure IDs) in the prediction of rating for the hospitals.

- The average rating was impacting hospitals' revenue there was a steady decline in the profits. This was majorly because the management was unaware of the factors that were affecting its rating significantly.

- The proposed 'Rating Prediction' model would be helpful for the Evanston hospital to improve their services accordingly for the next year.

Thank You!