HEALTHCARE CAPSTONE PROJECT

- RATING PREDICTION MODEL AND PROVIDER ANALYSIS

By-
SAJAL SOMANI

CMS STAR RATING -

- CMS collects, analyses the data and produces research reports. It then works to eliminate the instances of fraud and abuse within the healthcare system. The CMS rates hospitals in the US on a scale of 1-5. Objective of the analysis is to:
  - Identify potential root causes for the issues faced by Evanston Hospital.
  - Recommend ways for Evanston Hospital to improve their current Star Rating of 3 to 4 at-least.
- Analysis is divided into 4 parts:
  - Data Understanding – Groups and Measures.
  - Identifying important measures affecting star ratings.
  - Predictive Modelling of Star Ratings.
  - Recommending ways for Evanston Hospital to improve their rating.

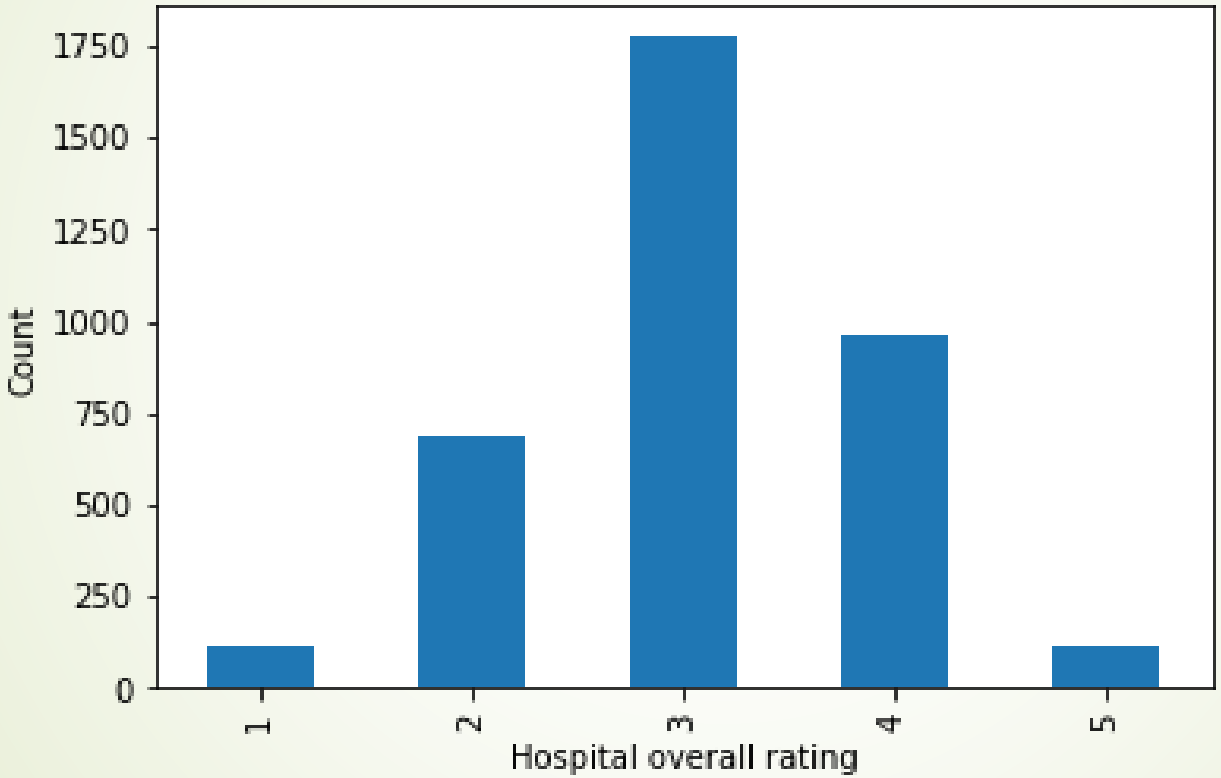
Data Understanding – Groups and Measures

- CMS included 62 measures classified under 7 groups having a certain weightage. Files considered are:
  - Complication – Hospital
  - HCAHPS
  - Healthcare Associated Infections
  - Hospital General Information
  - Outpatient Imaging Efficiency
  - Readmissions and Deaths
  - Timely and Effective Care

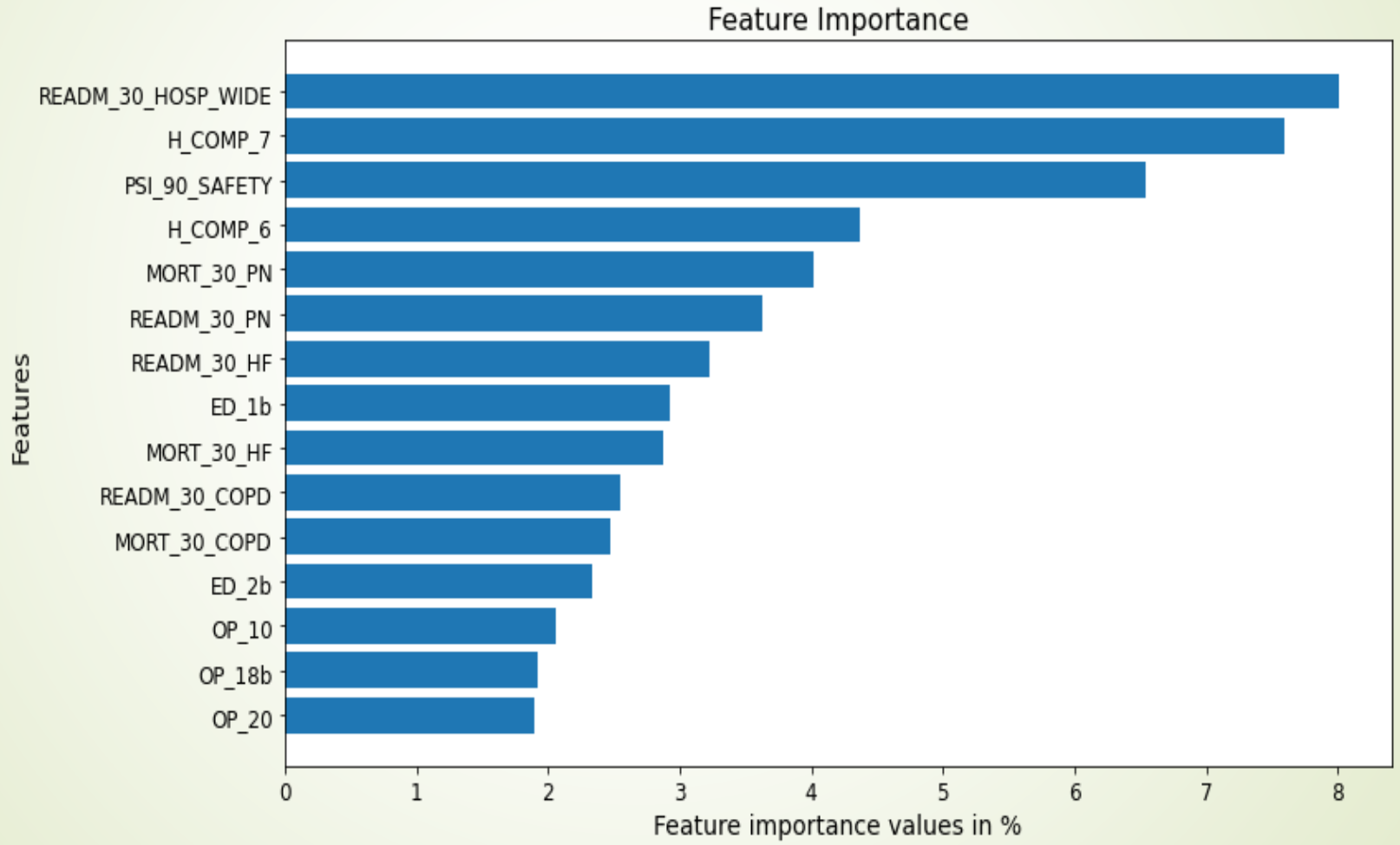
Steps followed for Data Cleaning-

- Reading the respective CSV files for Measure indicators.
- Keeping Measure ID, Provider ID, Score and Hospital Rating.
- Remove the duplicate rows and Null rows.
- Pivot the tables individually to get the unique rows for unique provider ID.
- Performing outer merge on all measure tables.
- Filter Hospital General Information and keep Provider ID and Rating column.
- Finally merging all tables of measure indicator.
- Removing rows with Hospital Rating as 0.
- Final Shape of the table comes to be 3648 x 64.

Provider Rating Distribution –



Relative Feature Importance -



Linear Regression Model -

- Steps performed-
  - Splitting the Data
  - Feature Scaling
  - Using RFE and checking VIF and p-value value.
  - Residual Analysis
  - Fitting in Test Data
- The overall accuracy after doing Linear Regression was found out to be 59.5%.
- Equation for best fitted line -
  - Rating = 3.090873 - 0.357644 x PSI_90_SAFETY - 0.054019 x OP_10 + 0.050121 x OP_13 + 0.058894 x OP_8 -0.053359 x MORT_30_AMI - 0.119120 x MORT_30_COPD - 0.074737 x MORT_30_HF - 0.172764 x MORT_30_PN - 0.043535 x MORT_30_STK + 0.048406 x READM_30_CABG + 0.132755 x READM_30_COPD + 0.071689 x READM_30_HF + 0.030086 x READM_30_HIP_KNEE - 0.354792 x READM_30_HOSP_WIDE - 0.174265 x ED_1b + 0.118215 x IMM_2 + 0.065893 x IMM_3 + 0.054150 x STK_1 + 0.082043 x VTE_1 + 0.112600 x H_COMP_7 - 0.072443 x HAI_1

Logistic Regression -

- Steps followed:
  - Splitting the data.
  - Feature Scaling
  - Applying Logistic Regression.
  - Predicting the accuracy on Test Data.
- After performing Logistic Regression we got an accuracy of 70%.

Random Forest -

- Steps included-
  - Standard Scaling.
  - Feature Importance.
  - Hyper-parameter Tuning.
  - Fitting the Final Model.
  - Evaluation Matrix.
  - PCA and Clustering.

Model:

Sno	Type of Model	Accuracy Score
1.	Linear Regression	59.54 %
2.	Logistic Regression	70 %
3.	Random Forest	72 %
4.	XGBoost	81.4%

From above we can see that the best model was found using XG Boost.
It gave an maximum accuracy of around 81.4% amongst both the other models.

	Models	Train f1 score	Test f1 scores	Train Accuracy	Test Accuracy
0	Logistic Regression	0.717	0.695	0.721	0.700
1	Random Forest	1.000	0.700	1.000	0.720
2	XGBoost Classifier	1.000	0.810	1.000	0.814

From the above values for important feature we can see that if Evanston Hospital improve on few measures, it can improve it's rating from 3 to 4 atleast if:-

1. READM_30_HOSP_WIDE from 15.2 to 15.04.
2. PSI_90_SAFETY from 1.45 to 0.62.
3. H_COMP_6 fromHF4 to 85.73.
4. MORT_30_PN from 13 to 14.95.
5. READM_30_PN from 16.6 to 15.85.
6. READM_30_HF from 21.2 to 19.93.
7. MORT_30_HF from 9.5 to 11.02.
8. MORT_30_COPD from 5.7 to 7.15.
9. ED_2b from 76 to 77.4.
10. ED_1b from 245 to 223.6.

Areas where Evanston Hospital needs improvement are-

- Safety
- Readmission
- Mortality
- Timeliness of Care

Conclusion and Recommendations -

- Key Measures of improvement are:
  - 1. Readmission: Hospital-wide readmission
  - 2. Timeliness of Care: Median Time from ED arrival to ED departure and Admit decision time to ED departure time for admitted patients.
  - 3. Mortality: Heart failure and chronic obstructive Pulmonary Disease.
  - 4. Safety: PSI
  - 5. Patient Experience: Discharge Information.

THANK YOU