A New Predictive Model for Hospital Readmissions

Samantha Lee, Ruojie Lu, & Clayton Neill
Advisors: Ian Duncan & Jiyoun Myung, PSTAT
University of California, Santa Barbara

Abstract

With the enactment of the Hospital Readmissions Reduction Program under the ACA, hospitals need ways to identify patients who have a high risk of readmission. A popular risk-scoring method is the LACE index, which has limitations. We developed a logistic regression model for Medicare patients using the components of LACE together with other available data . We also utilized third party software, DataRobot, to develop machine learning models (gradient boosted trees were best-performing). Our resulting models identified patients with a high risk of readmission, allowing hospitals to allocate their resources efficiently to reduce readmission risk.

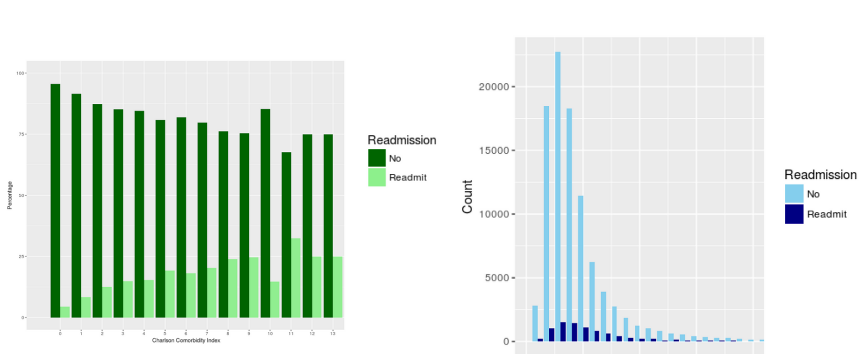
Introduction

Hospital readmission occurs when a patient, who has been discharged from a hospital, is admitted again to the same or different hospital within thirty days. Readmissions are costly and disruptive for both patients and hospitals. For the patient, a readmission increases their possibility of hospital-acquired infections and complications. For the hospital, readmissions lead to higher costs and are logistically inefficient.

Data Description

LACE

Length of Stay	Acuity	Charlson Comorbidity Index	Emergency room visits in previous 6 months



Hierarchical Condition Categories (HCC)

A risk score that identifies high risk individuals based on their demographic data and diagnosis data, and is calibrated for persons over sixty-five years old

Diagnosis Related Group Type

Specifies whether the procedure was medical or surgical based on the diagnosis provided

Note: two datasets for separate model building
General data for general hospital population
Medicare data for patients patients who are readmitted with CMS penalized diagnoses

Methods

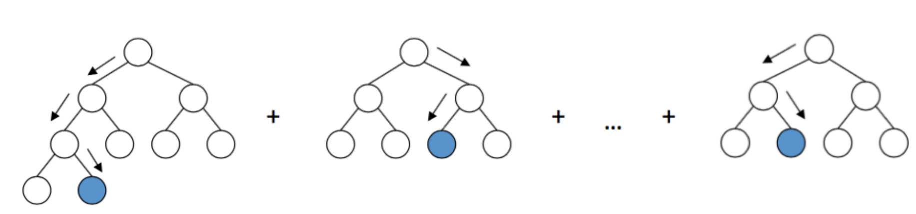
Logistic Regression

A statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a binary variable. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest

$$Ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Gradient Boosted Trees

The general idea is to compute a sequence of simple trees, where each successive tree is built for the prediction residuals of the preceding tree. The additive weight expansions of trees can produce an excellent fit of the predicted values to the observed values.



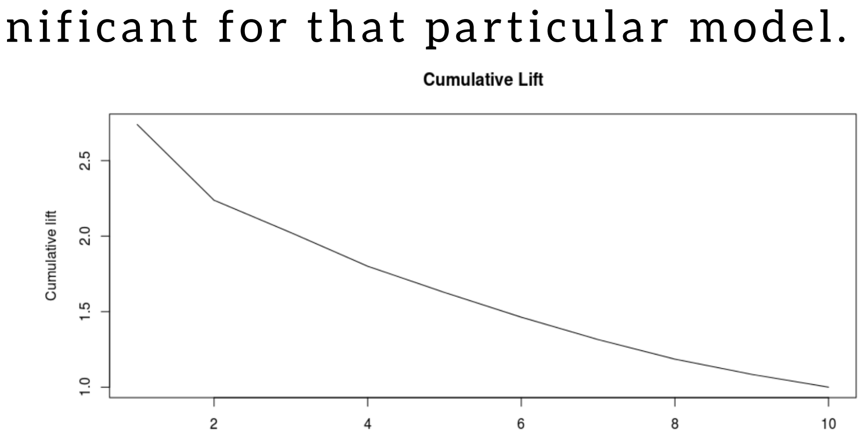
DataRobot

An automated machine learning platform that makes it fast and easy to build and deploy accurate predictive models.

Results

Generalized Linear Model

The following table displays the odds ratios of the significant covariates at a 0.05 significance level along with their 95% confidence intervals. A value of "N/A" indicates that the variable was not significant for that particular model.



Variable	Odds Ratio (General)	95% Confidence Interval (General)	Odds Ratio (Medicare)	95% Confidence Interval (Medicare)
Intercept	0.424	(0.377,0.478)	0.516	(0.475,0.562)
Age	0.997	(0.995,0.999)	1.020	(1.018,1.022)
Sex Male (vs. Female)	1.203	(1.168,1.239)	1.093	(1.043,1.146)
Non-emergent admission	0.744	(0.721,0.767)	0.915	(0.869,0.963)
HCC Score	1.302	(1.277,1.328)	N/A	N/A
DRG Surgical (vs. Medical)	0.544	(0.527,0.563)	0.043	(0.039,0.048)
ER visits in previous 6 months	1.439	(1.417,1.462)	N/A	N/A
Length of Stay (days)	1.213	(1.199,1.227)	N/A	N/A
CCI	1.113	(1.102,1.124)	1.157	(1.141,1.173)
Education	0.943	(0.928,0.958)	0.697	(0.633,0.767)
Income	N/A	N/A	0.714	(0.661,0.771)
Poverty	N/A	N/A	0.665	(0.595,0.744)

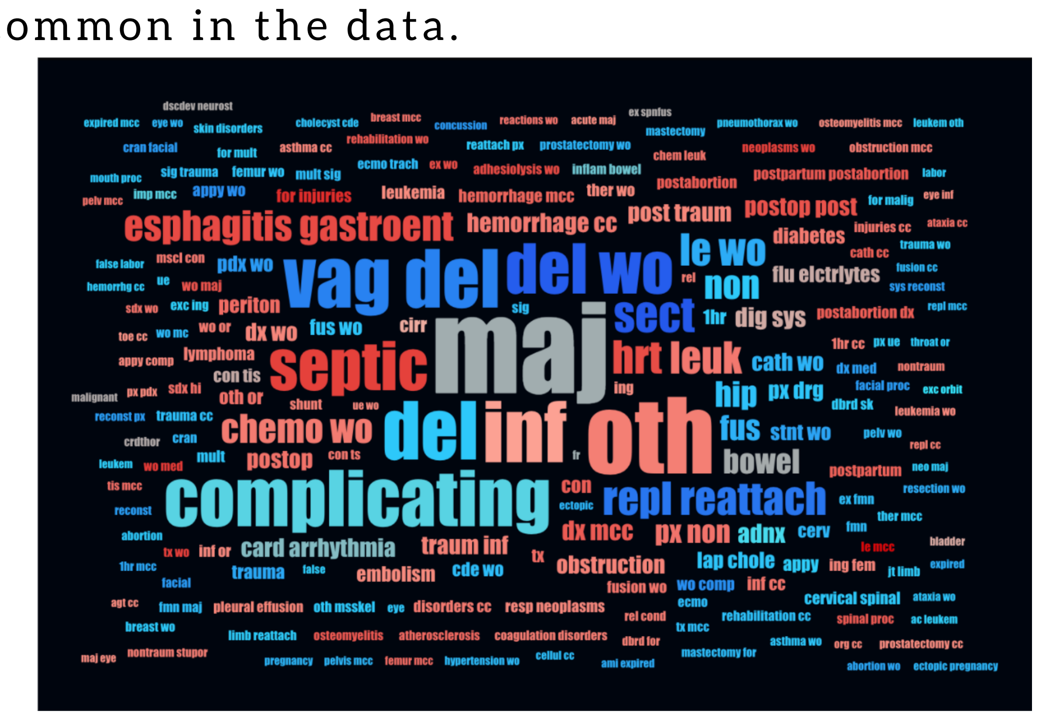
Machine Learning Models

The performances of these models were compared on the basis of F1 score, balanced accuracy, and Top 10%/20% measures. This table shows the performance metrics of these models along with the results from the GLM. The cutoff value for the DataRobot models was chosen to maximize the F1 score on the training data.

	LGBM (General -DataRobot)	Logistic Regression (General -R)	LACE (General)	GBM (Medicare -DataRobot)	Logistic Regression (Medicare -R)
Cutoff Value	0.438	0.474	N/A	0.578	0.596
Error	0.325	0.349	0.337	0.158	0.322
Sensitivity	0.739	0.700	0.620	0.460	0.730
Specificity	0.670	0.647	0.667	0.864	0.675
PPV	0.170	0.153	0.145	0.161	0.113
F_1 Score	0.276	0.251	0.236	0.239	0.196
Accuracy	0.675	0.651	0.663	0.842	0.678
Balanced Accuracy	0.568	0.556	0.548	0.563	0.546
AUC	0.771	0.730	N/A	0.797	0.771
Top 10%	32.4%	27.4%	25.5%	31.0%	26.8%
Top 20%	50.8%	44.7%	40.6%	54.8%	46.8%

Word Cloud

This word cloud generated on the DRG Descriptions displays the red words as correlated with readmission and the blue and grey words as not. Larger words are more common in the data.



Conclusions

The LACE index is simple and has moderate predictive power to identify risky patients. The best performing models were the light gradient boosted tree for the General data and the gradient boosted tree for the Medicare data; each outperformed the logistic regression models for their respective datasets by a considerable margin. Until the model has been validated on patient data outside of the area it was created, we believe our model should be used as an assessment of hospital quality and performance.

Works Cited

1. Van Walraven, C., et al. "Derivation and Validation of an Index to Predict Early Death or Unplanned Readmission after Discharge from Hospital to the Community." CMAJ : Canadian Medical Association Journal 182.6 (2010): 551–557. PMC. Web. 15 Mar. 2018.
2. Bekkar, M., Djemaa, H. K., & Alitouche, T. A. "Evaluation Measures for Models Assessment over Imbalanced Data Sets." Journal of Information Engineering and Applications 3.10 (2013): 29.
3. Huynh, N., Fallah, H., & Robbins-Kelley, D. "Improving the Prediction of Re-admissions Amongst Medicare Patients in a California Hospital." ARCH: Actuarial Research Clearing House, 2017.1 (2017).