Hospidata

For all of your healthcare-related, data needs.

Director of Analytics:
Paul Tluczek

Summary

The Problem’s History

The Problem’s Cost

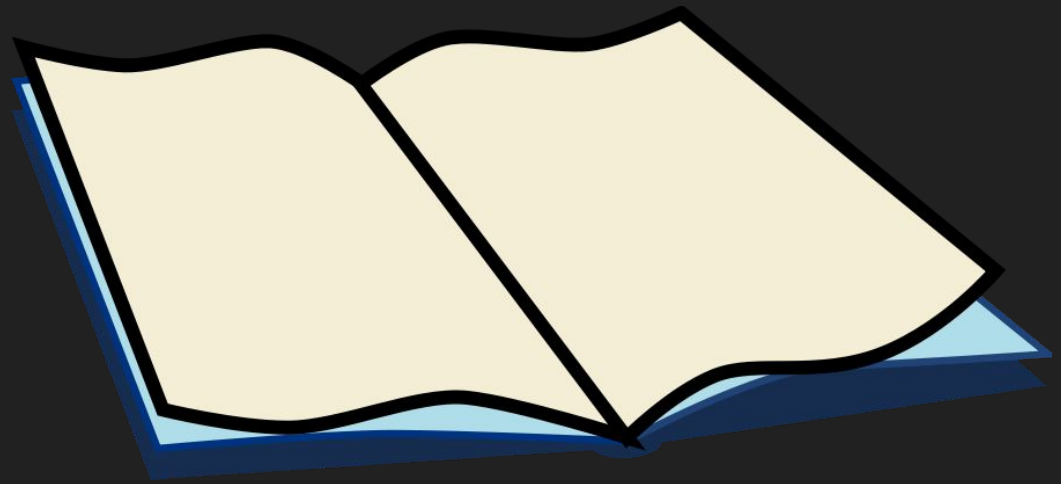
Initial Solution

Data

Data Science

Our Solution

Conclusion



Hospital Readmissions

Concerns readmission upon initial discharge from hospital.

20% Readmitted after 30 Days.

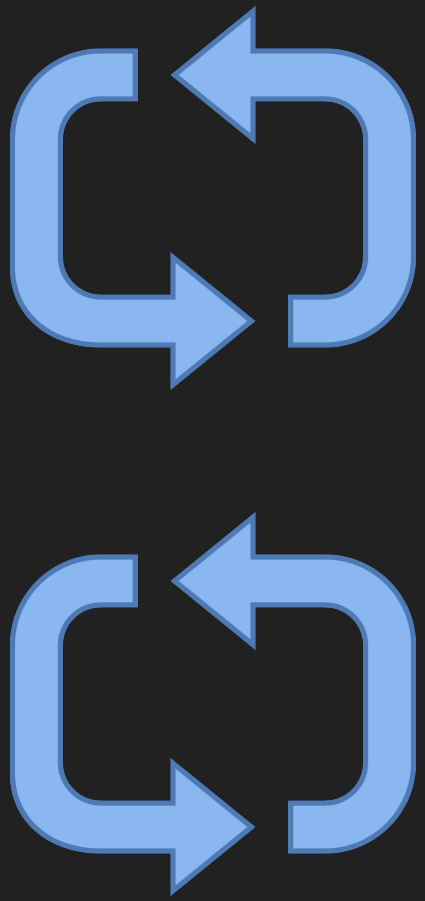
34 % Readmitted after 90 Days.

Costs 17.4 Billion Annually.

Potential Quality Problems

Lack of coordination in follow up care after discharge.

Misaligned Financial Incentives.



Readmission Penalties

2010 Affordable Care Act

- Establishes hospital readmission reduction program.
- Financial incentives to reduce readmissions

Hospitals with a 3-Year Rolling Readmission rate that exceed their risk adjusted target, penalized on a portion of their Medicare reimbursements

For 2012, As much as 1% of total reimbursements penalized

By 2014, As much as 3% of total reimbursements penalized

Tahoe Healthcare Systems

18% of revenues were Medicare reimbursements

Over 750,000 in fines for 2012

Under 2014 regulations, loss in reimbursements rises to $8000 per readmitted patient within 30 days

CareTracker program

Tahoe healthcare systems think they have come up with a solution

Involves personnel education on patients, during/post hospitalization monitoring, periodic home monitoring after discharge

Early data shows a reduced readmission rate by 40% compared to a control group

Cost of program equates to $1200 per patient

THE BIG QUESTION.

18% of total revenues are from Medicare reimbursement for the three HRRP conditions

The cost of CareTracker per patient is $1200

Only 40% success rate, 60% still ends up fined

Should CareTracker be deployed

If CareTracker is rolled out for no one…..
If CareTracker is rolled out for everyone….

	N	Cost - $
0	None	7984000.0
1	All	10048800.0

Our Initial Data Analysis

The following features were provided:

Sex

Age

ED Admit

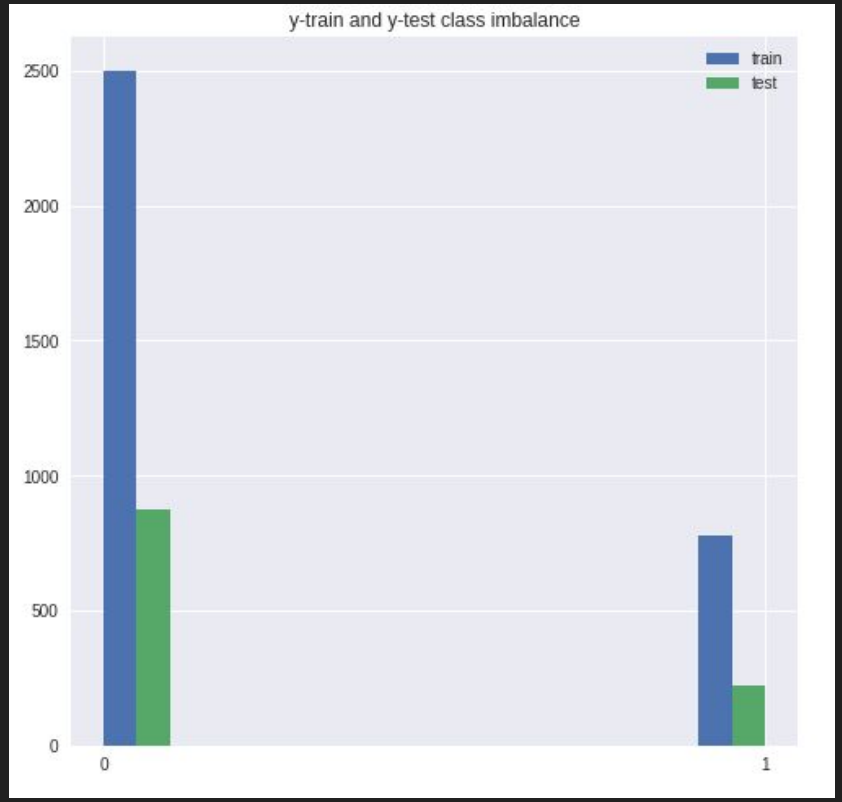
Flu Season

Severity Score

Comorbidity Score

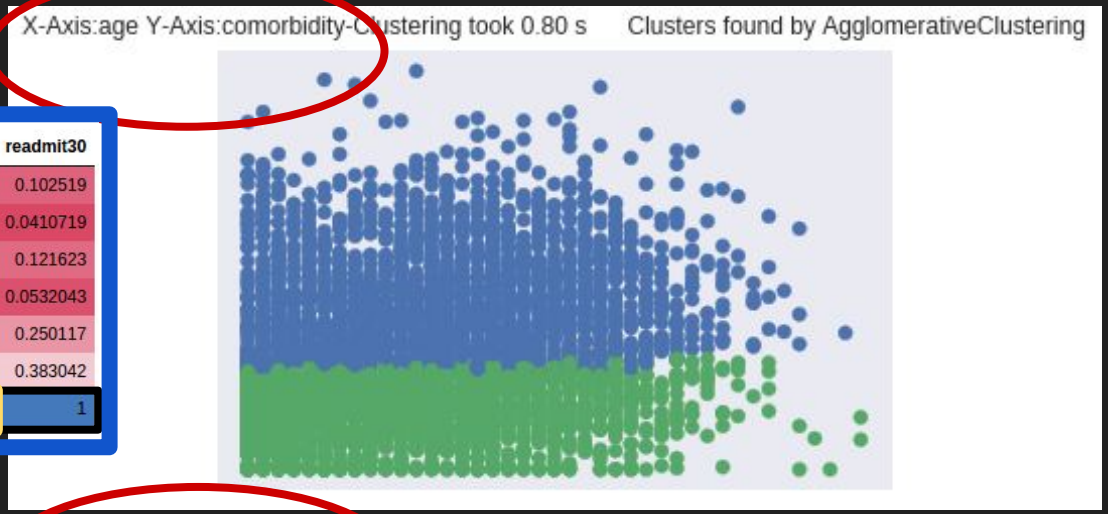
Readmitted in 30 Days

CLASS IMBALANCE



Patient Segments

	age	gender	flu_season	ed_admit	severity	comorbidity	readmit30
age	1	0.149386	-0.0116744	0.139991	0.280944	0.161004	0.102519
gender	0.149386	1	0.0156708	0.121463	0.0942044	0.00498498	0.0410719
flu_season	-0.0116744	0.0156708	1	-0.0115095	0.00356251	-0.0126297	0.121623
ed_admit	0.139991	0.121463	-0.0115095	1	0.333882	0.0291476	0.0532043
severity	0.280944	0.0942044	0.00356251	0.333882	1	0.196449	0.250117
comorbidity	0.161004	0.00498498	-0.0126297	0.0291476	0.196449	1	0.383042
readmit30	0.102519	0.0410719	0.121623	0.0532043	0.250117	0.383042	1



- The highest two correlations with what we are trying to predict are severity and comorbidity
- Age is a close third, however the relationship between age and comorbidity is very similar to the relationship between severity and comorbidity
- These clustering methods show us there exists a fine line somewhere in our dataset that was once invisible
- This line becomes a new feature in our dataset to help with predictions

Result Prediction Methods

- Random Forest
- Logistic Regression

Evaluation strategy for the prediction system

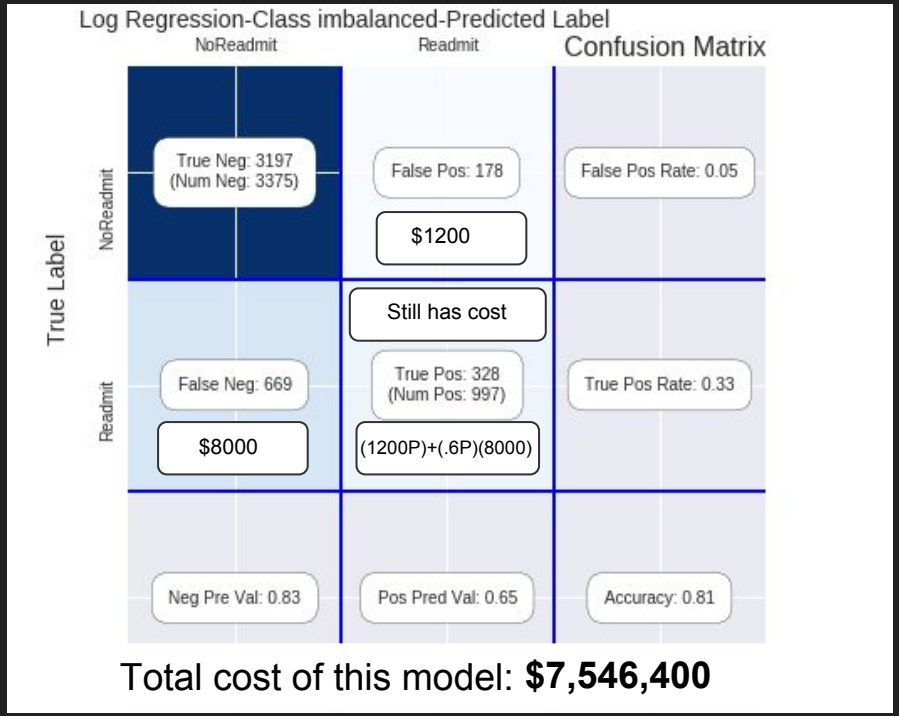
Confusion Matrix:

- A quick reference guide to actual and predicted values

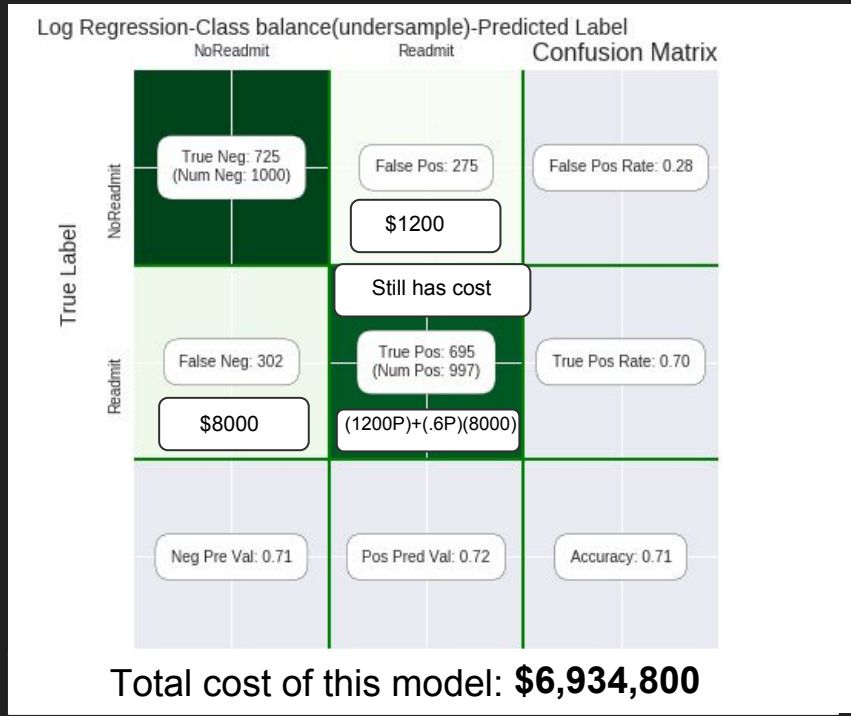
- A table that is often used to describe the performance of a classification model on a set of data for which the actual outcomes are known

Logistic Regression

(Essentially, this is defined by drawing a curved line that relates an individual patient's attributes to outcomes.)

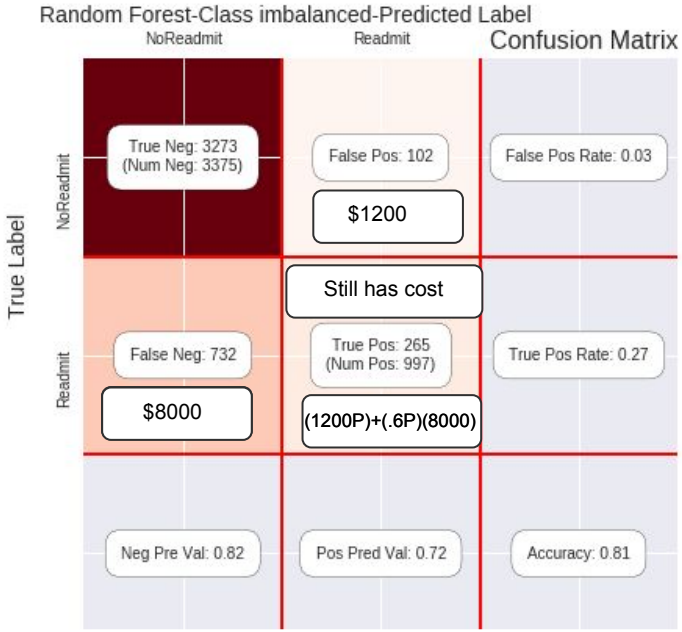


Total cost of this model: **$7,546,400**

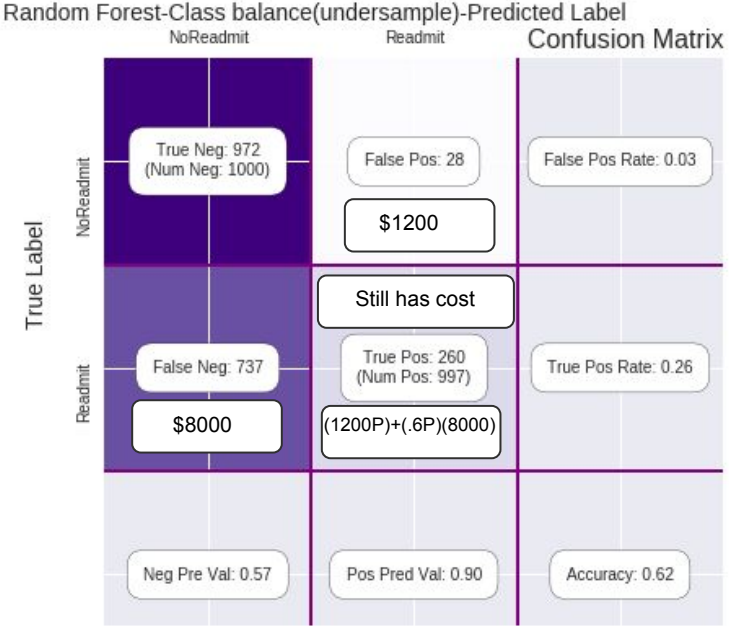


Total cost of this model: **$6,934,800**

Random Forest

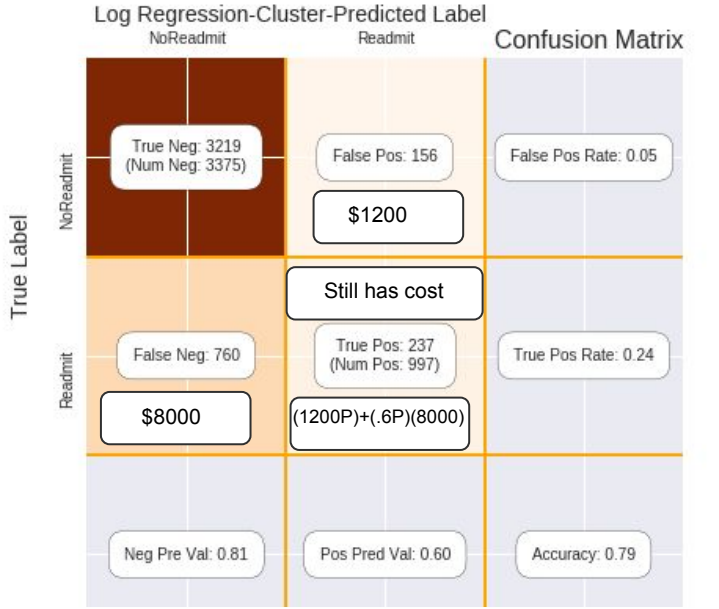


Total cost of this model: **$7,654,000**



Total cost of this model: **$7,613,200**

CareTracker and Machine Learning



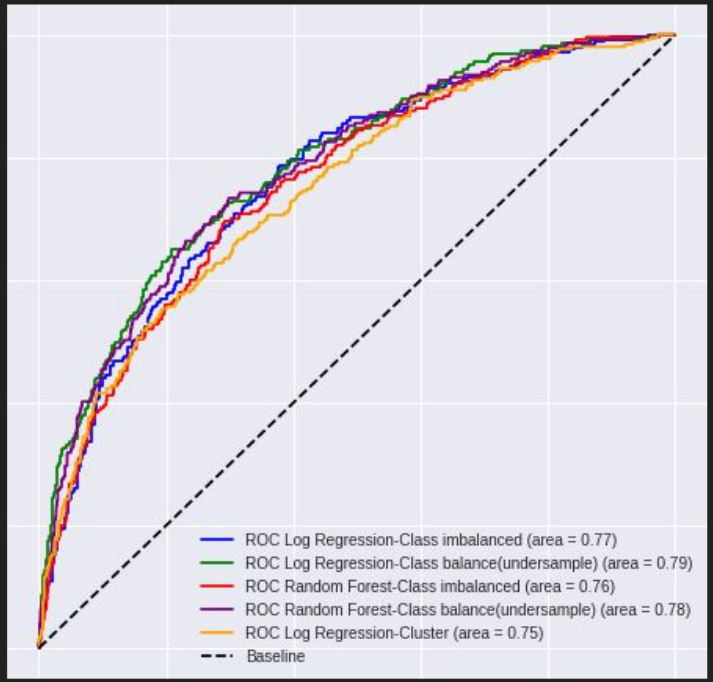
Total cost of this model: **$7,699,200**

	Model based Care	Cost - $
0	None	7984000.0
1	All	10048800.0
2	Predict Log Regression-Class imbalanced	7546400.0
3	Predict Log Regression-Class balance(undersample)	6934800.0
4	Predict Random Forest-Class imbalanced	7654000.0
5	Predict Random Forest-Class balance(undersample)	7613200.0
6	Predict Log Regression-Cluster	7699200.0

Savings Strategies

Strategy	Savings
NONE	0
ALL	- $2,064,800
LOGISTIC REGRESSION (IMBAL)	$437,600
LOGISTIC REGRESSION (BAL)	$1,049,200
RANDOM FOREST (IMBAL)	$330,000
RANDOM FOREST (BAL)	$370,800
CLUSTERED LOG REGRESSION	$284,800

RISK PATIENT CARE
VS
INCREASED
CARETRACKER COVERAGE



Conclusion

Using logistic regression will save money.

Roll out revised Beta CareTracker system to a control group at selected hospitals (A/B)

Production version of CareTracker will require more data, and **soon**.