Predicting Hospital Readmission Rates

By Matthew Emanuel

Value

Personalized healthcare

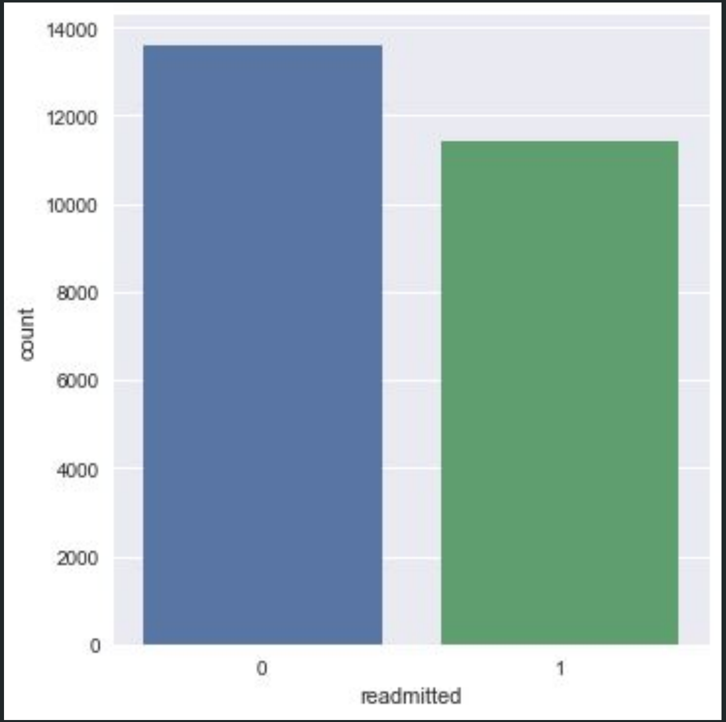
Clinical insights

Cost benefit

Data

25,000 patients

Target variable



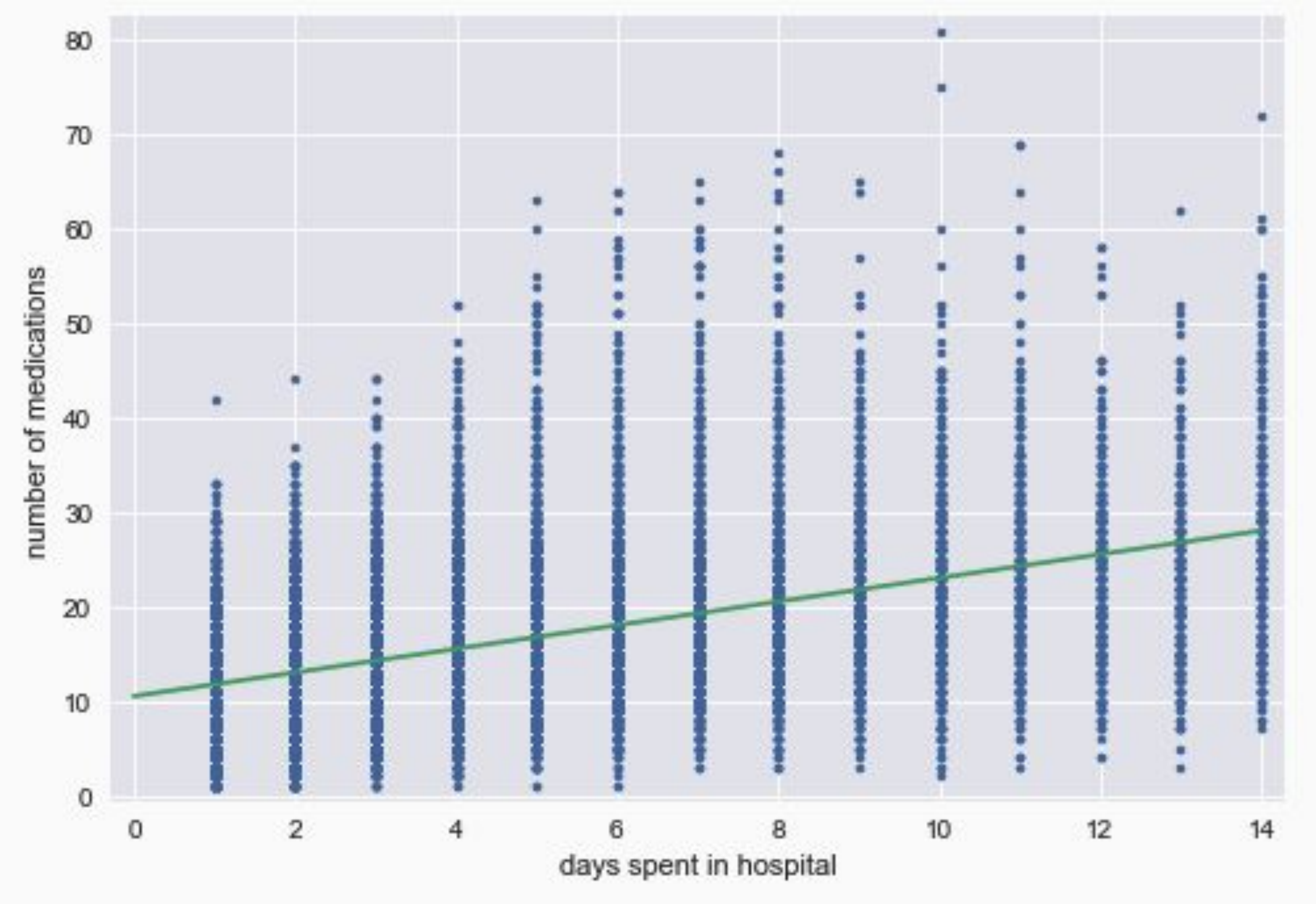
Features

Hospital stay information

Hospital history information

Medications in use

Diagnoses



number of medications
days spent in hospital
0
10
20
30
40
50
60
70
80
0
2
4
6
8
10
12
14

Preprocessing

Dummy

race_African_American	race_Caucasian	race_Other	gender_Female	gender_Male
0	1	0	0	1
0	1	0	1	0

Scale

days_in_hospital	num_lab_procedures	num_procedures	num_medications
1.000000	0.320	0.0	0.1250
0.076923	0.232	0.0	0.1375

Split

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

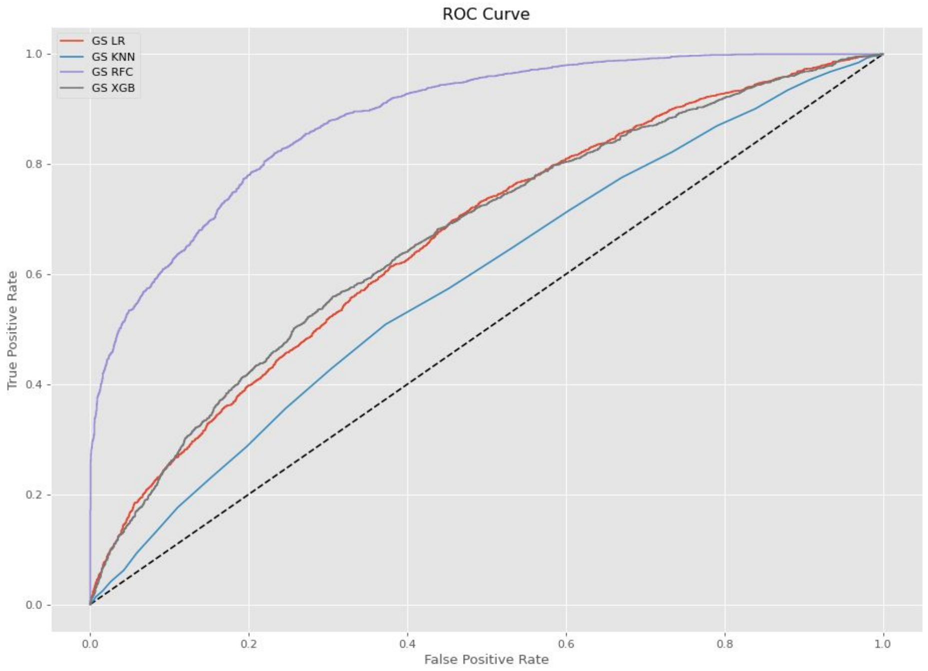
Modeling

Logistic Regression

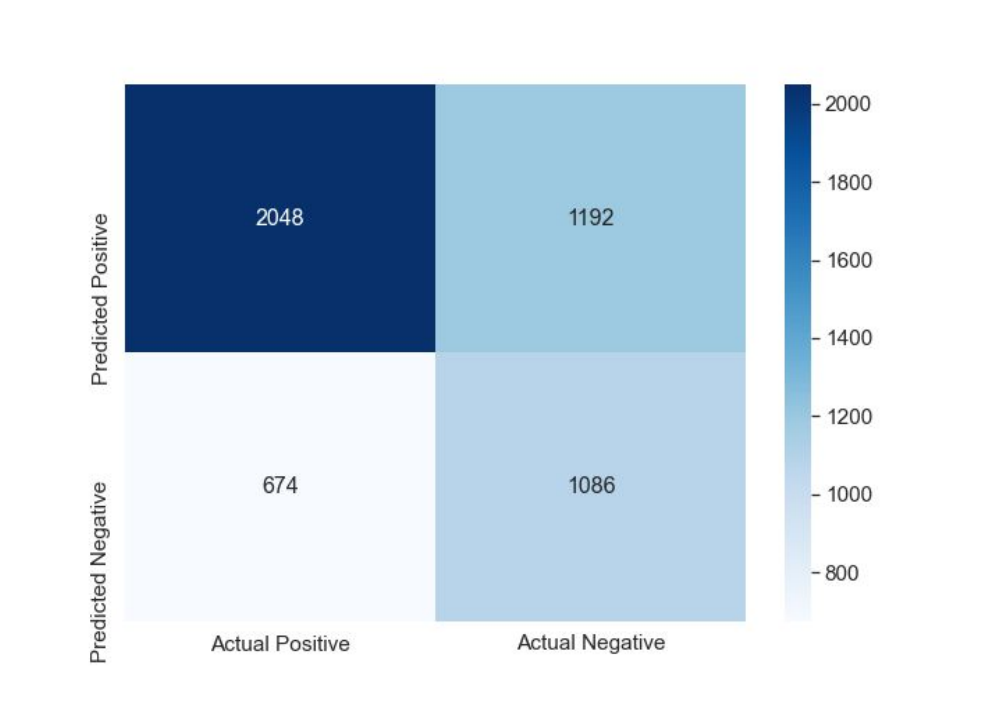
K Nearest Neighbors

Random Forest Classification

Gradient Boosting



Random Forest Classification Model



Model Scores

Labels	Default_accuracy_scores	Fit_times	GS_accuracy_scores	AUC_scores
RFC	0.6118	4.106606	0.7734	0.883675
XGB	0.6202	0.454879	0.6268	0.668189
LR	0.6152	0.558409	0.6146	0.665155
KNN	0.5400	0.557119	0.5740	0.586294

Importance

	Importance	Feature
0	0.149487	number_inpatient
1	0.090562	num_medications
2	0.083157	num_lab_procedures
3	0.057175	number_diagnoses
4	0.054365	days_in_hospital
5	0.045655	number_emergency
6	0.042379	number_outpatient
7	0.036866	num_procedures
8	0.015015	Dx1_Heart_Failure
9	0.013838	paycode_payer_code_MC
10	0.012966	diabetesMed_Yes

Number of inpatient visits

Number of medications

Number of lab procedures

Limitations

Model Fit time

Type I Error Rate

Available data

Further Exploration

Bayesian optimization

Explore other data:

- living situation, annual income
- BMI, A1c, blood pressure