

Predicting Diabetes Re-hospitalizations

By Maxim Belov



Diabetes in the US



- 34.2 million Americans (about 10%) have diabetes.
- \$237 billion were spent in 2017 on direct care.
- \$102 billion of that is related to direct hospital admissions.
- Patients with diabetes are more likely to be re-admitted following a hospitalization.
- What can we do to allow early identification and intervention for high-risk patients?

Agenda



Data Overview



Exploratory Data
Analysis

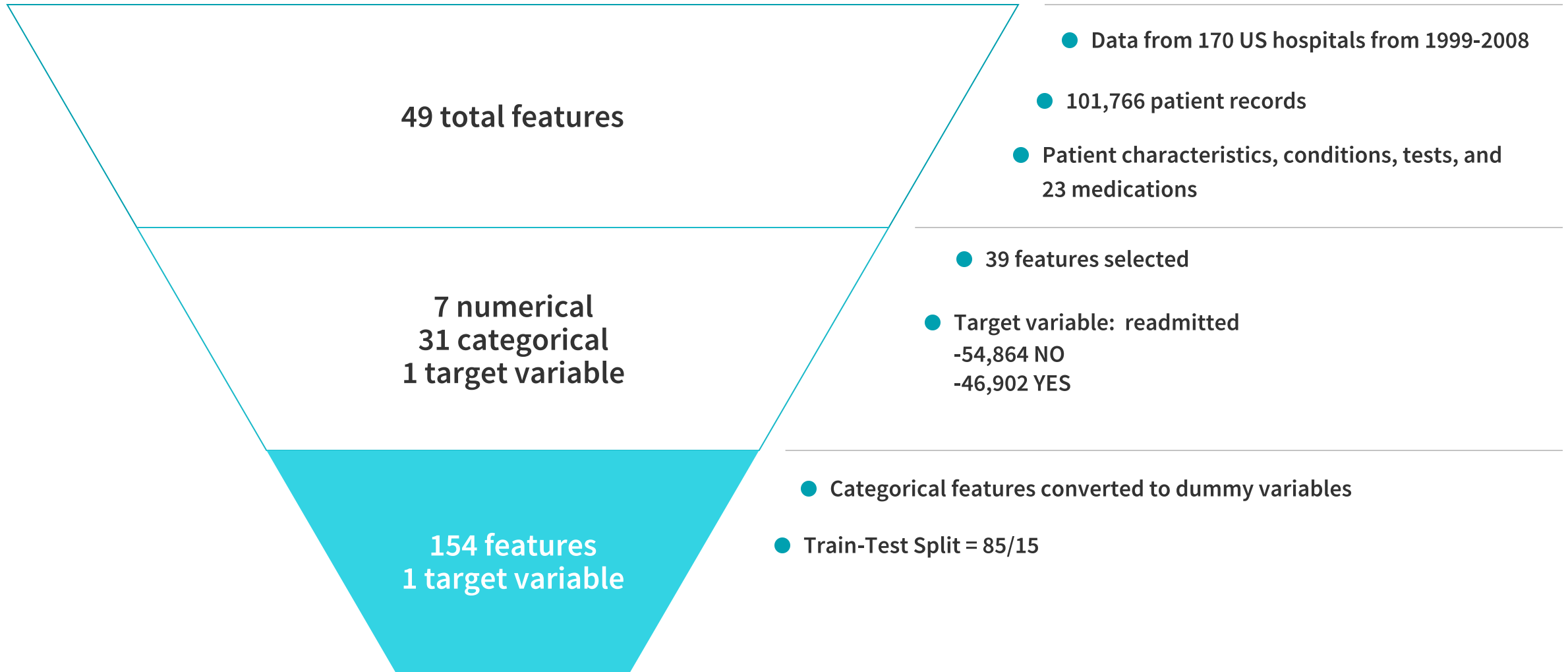


Supervised Machine
Learning Models

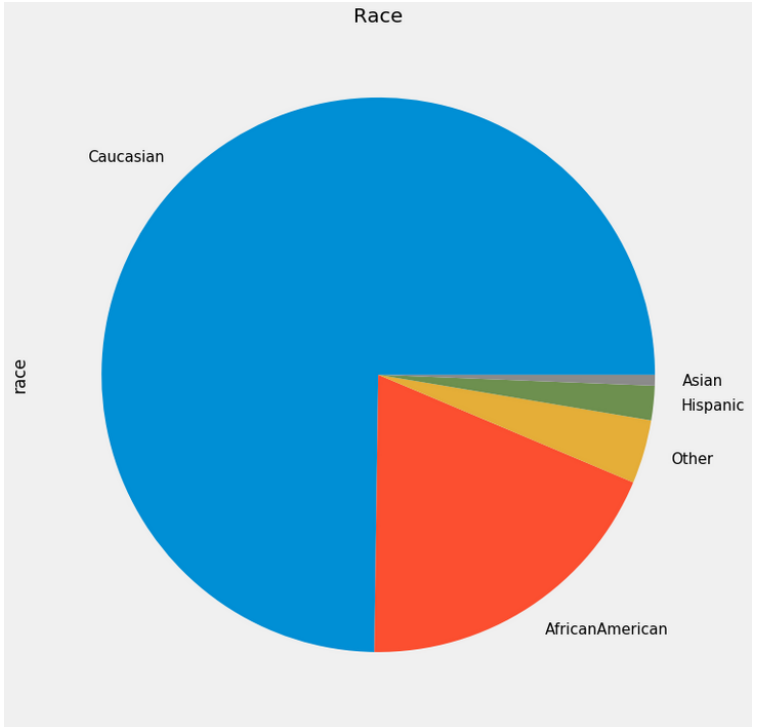
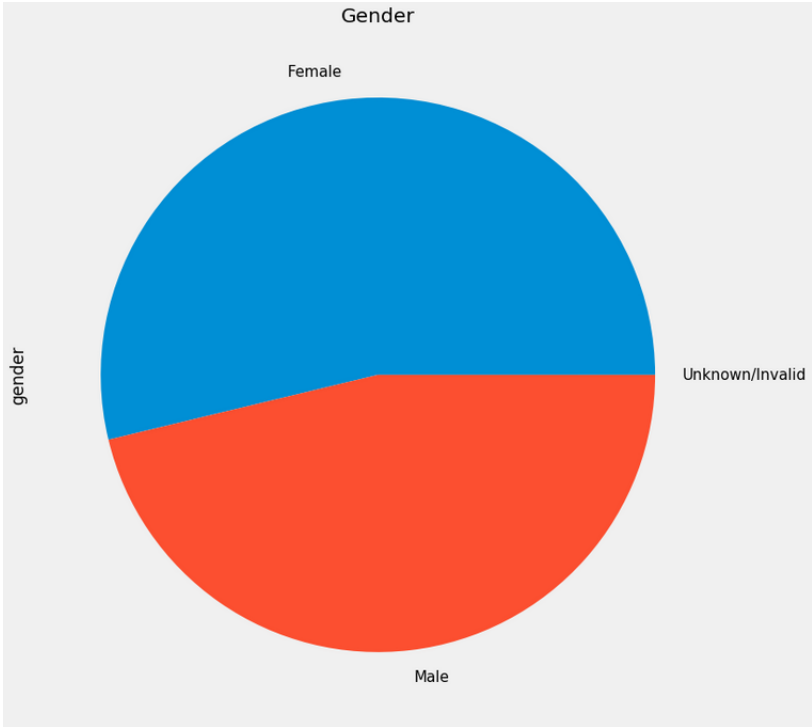
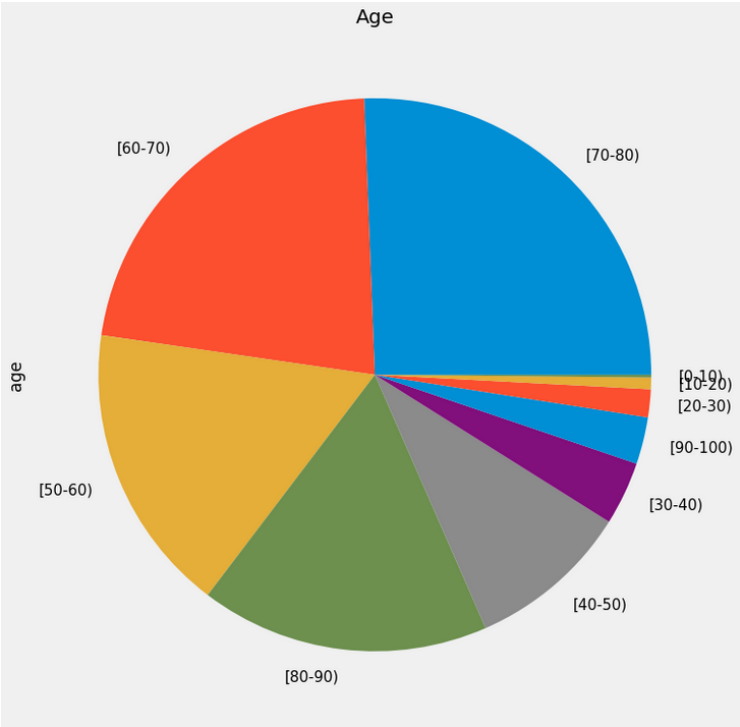


Results

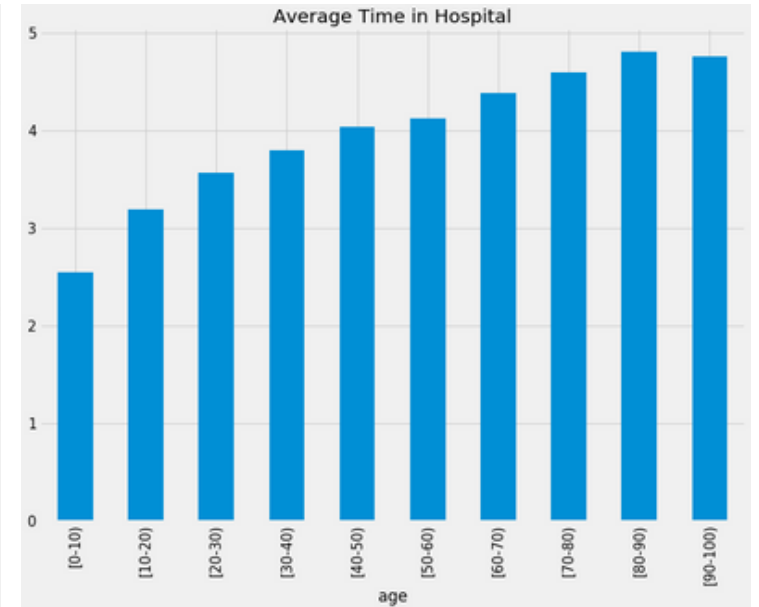
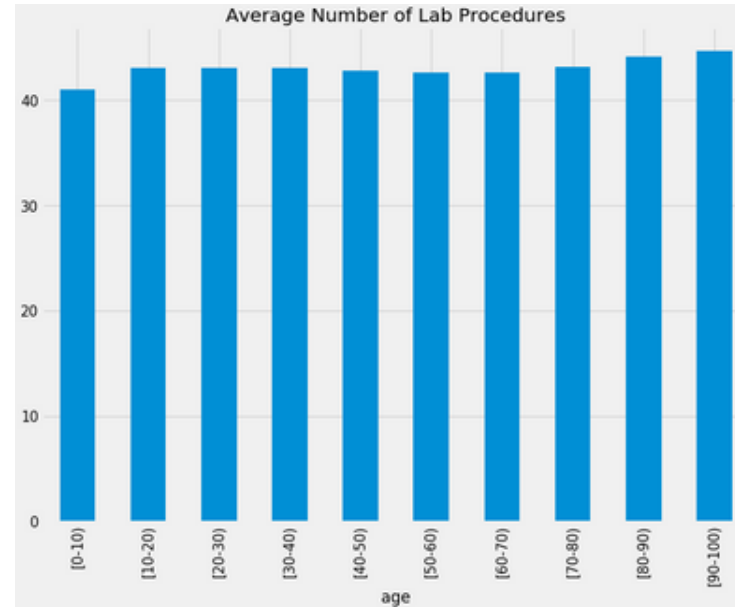
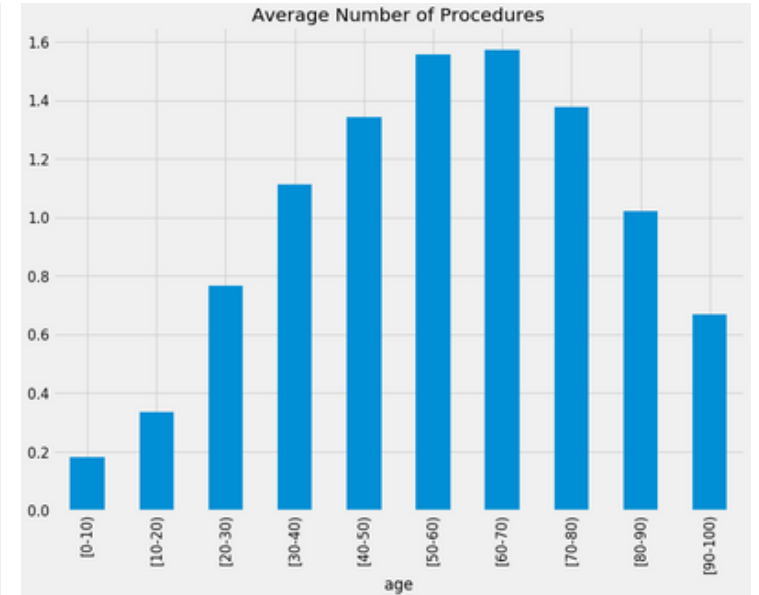
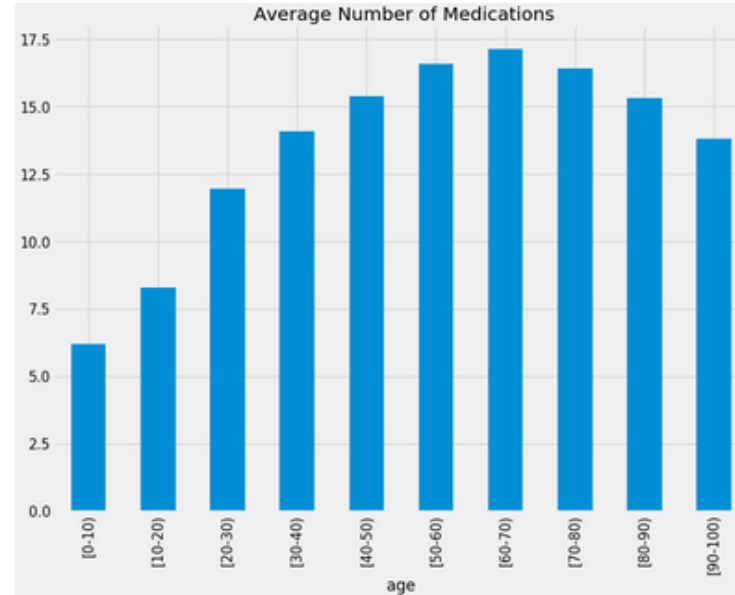
Data Overview



Exploratory Data Analysis

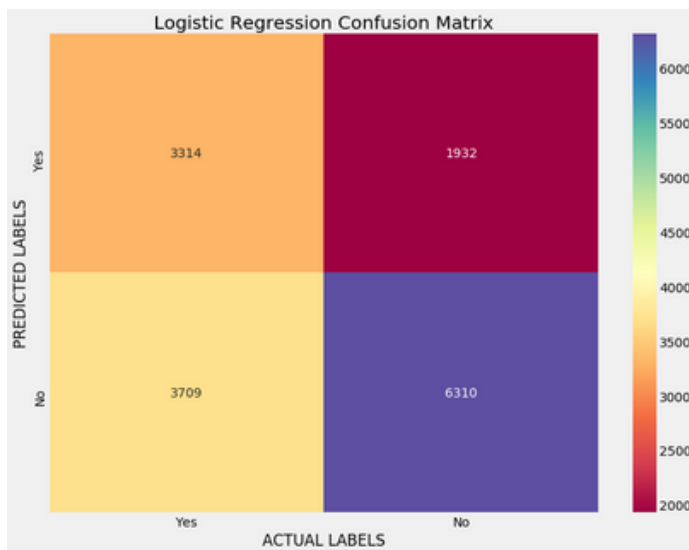


Exploratory Data Analysis



*Train-Test Split = 85/15

Base ML Models



Logistic Regression

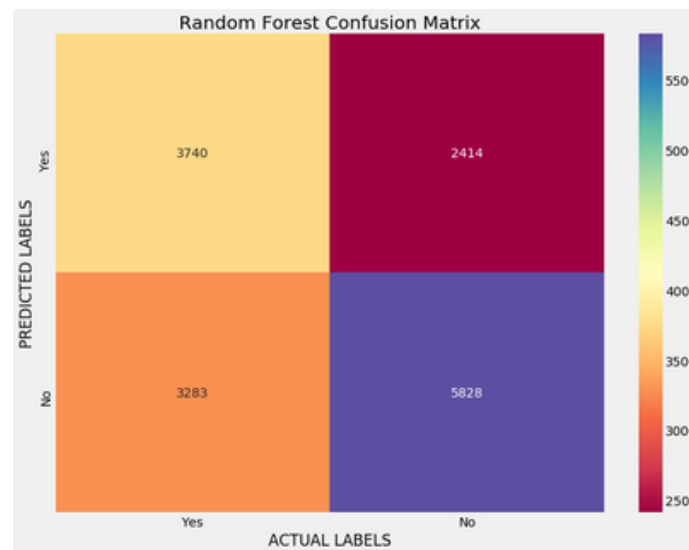
Accuracy: 0.63

Precision: 0.63

Recall: 0.62

F1: 0.62

AUC: 0.69



Random Forest Classifier

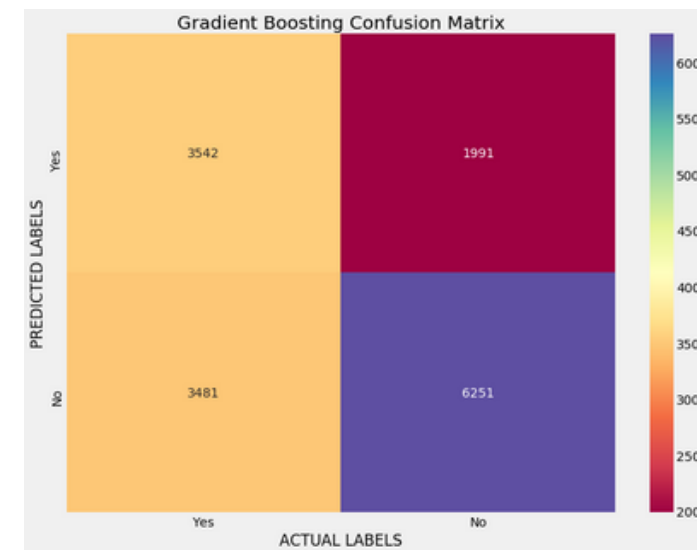
Accuracy: 0.64

Precision: 0.63

Recall: 0.62

F1: 0.62

AUC: 0.68



Gradient Boosting Classifier

Accuracy: 0.65

Precision: 0.64

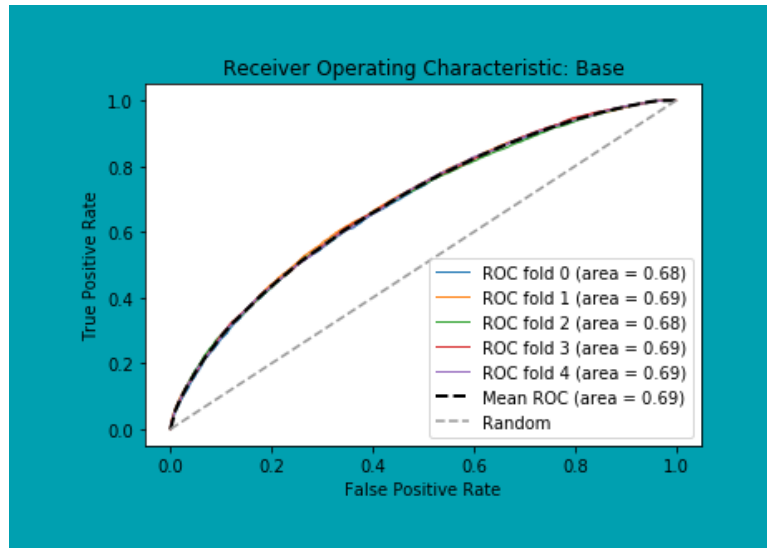
Recall: 0.63

F1: 0.63

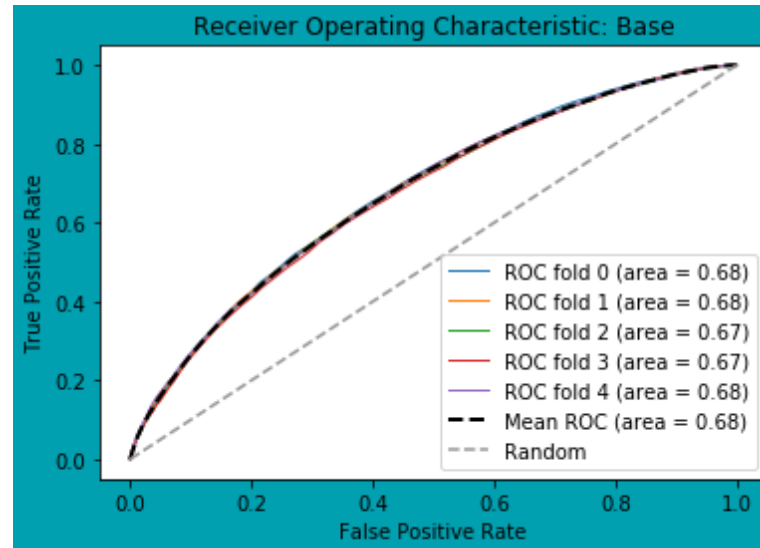
AUC: 0.69

ROC Curves - Base

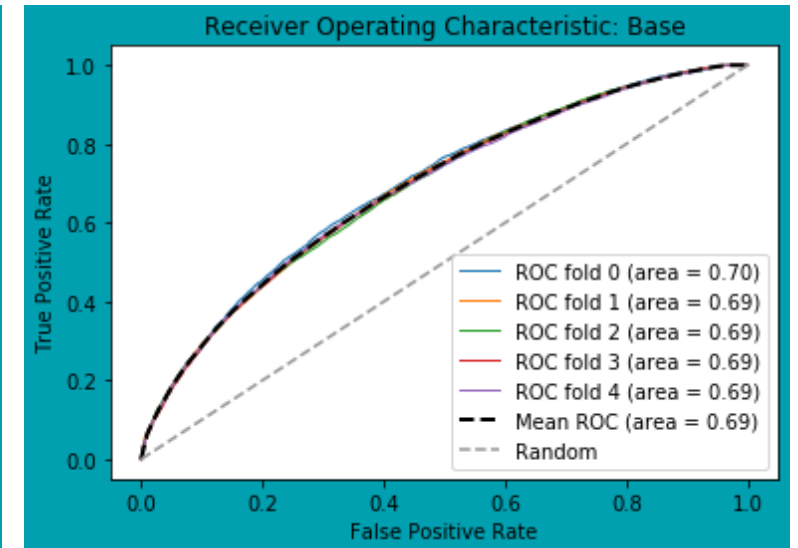
Logistic Regression



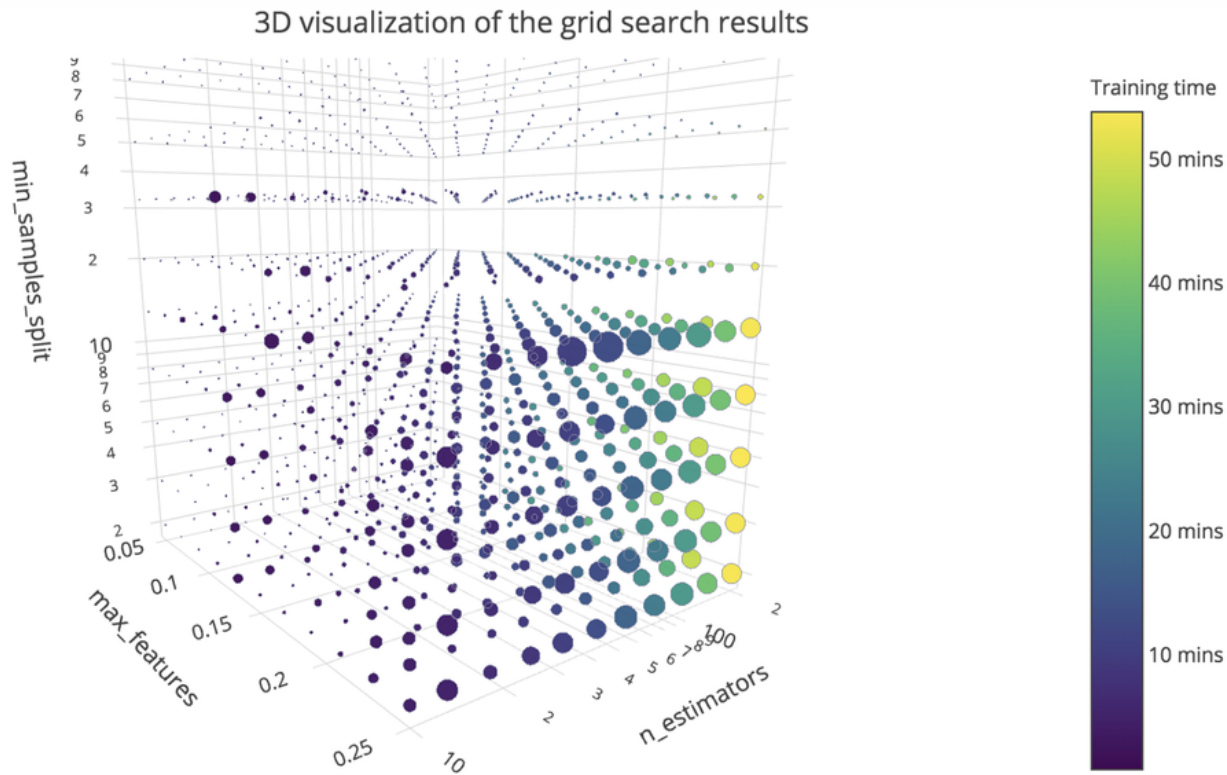
Random Forest



Gradient Boosting



Optimization - Grid Search



- **Logistic Regression**

penalty='l2', C=0.01

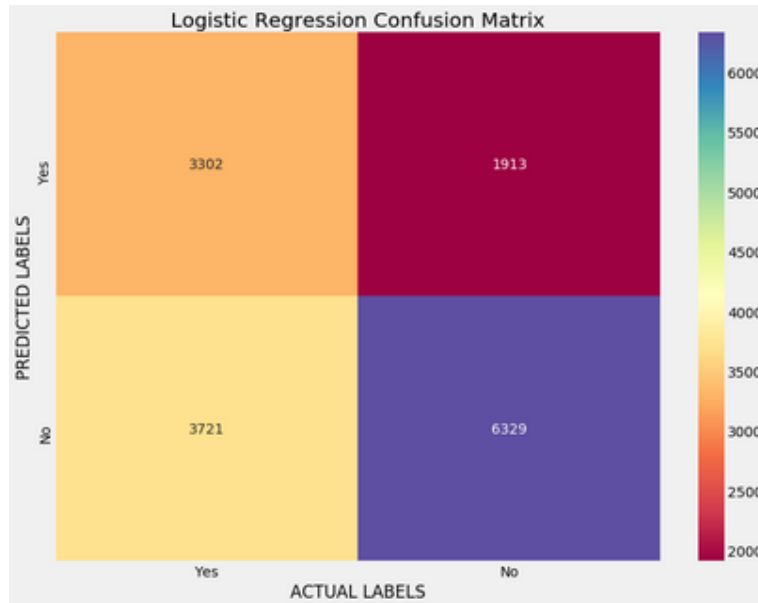
- **Random Forest Classifier**

max_depth=None, max_features='sqrt',
min_samples_leaf=2, min_samples_split=2,
n_estimators=80

- **Gradient Boosting Classifier**

learning_rate=0.50, n_estimators=120

Optimized ML Models



Logistic Regression

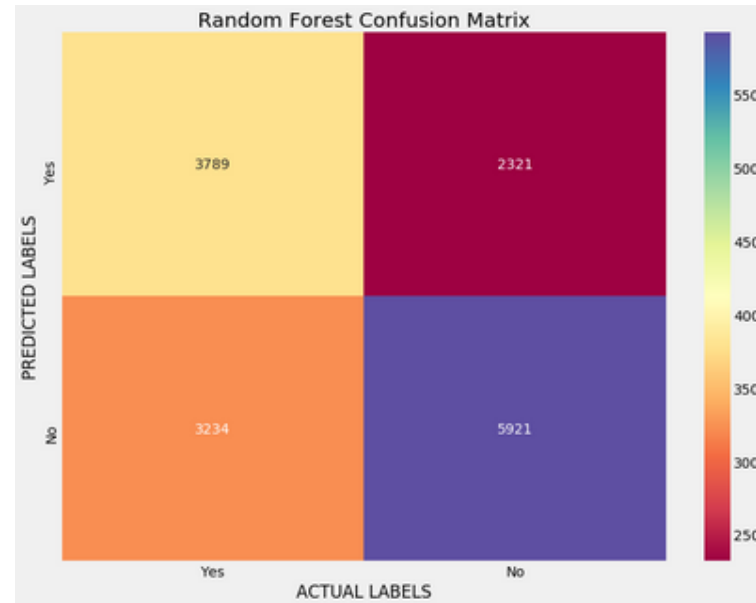
Accuracy: 0.63

Precision: 0.63

Recall: 0.62

F1: 0.62

AUC: 0.69



Random Forest

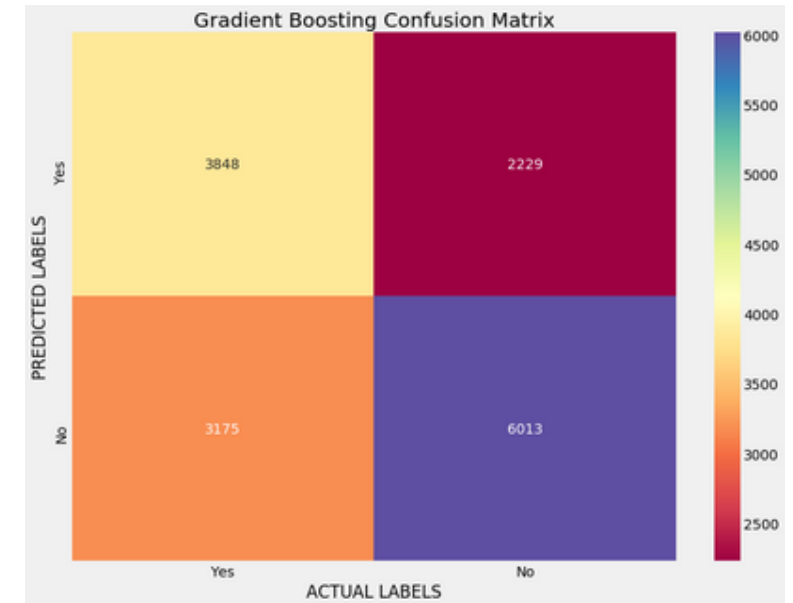
Accuracy: 0.64

Precision: 0.63

Recall: 0.63

F1: 0.63

AUC: 0.69



Gradient Boosting

Accuracy: 0.65

Precision: 0.64

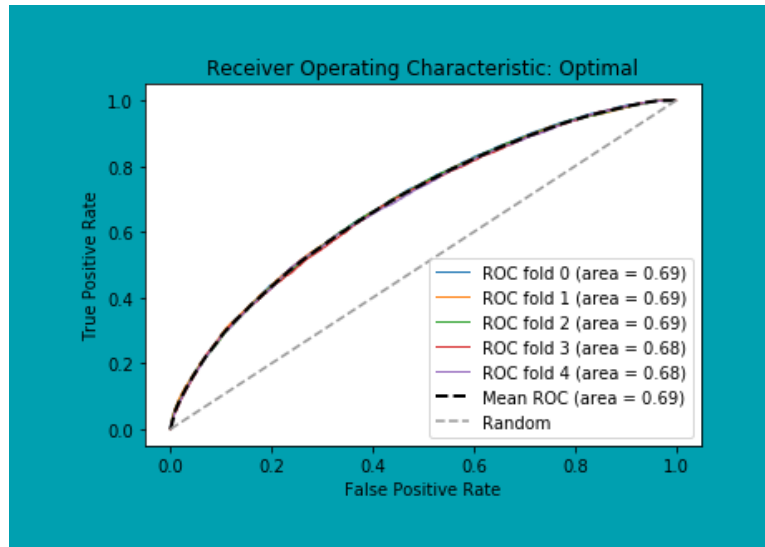
Recall: 0.64

F1: 0.64

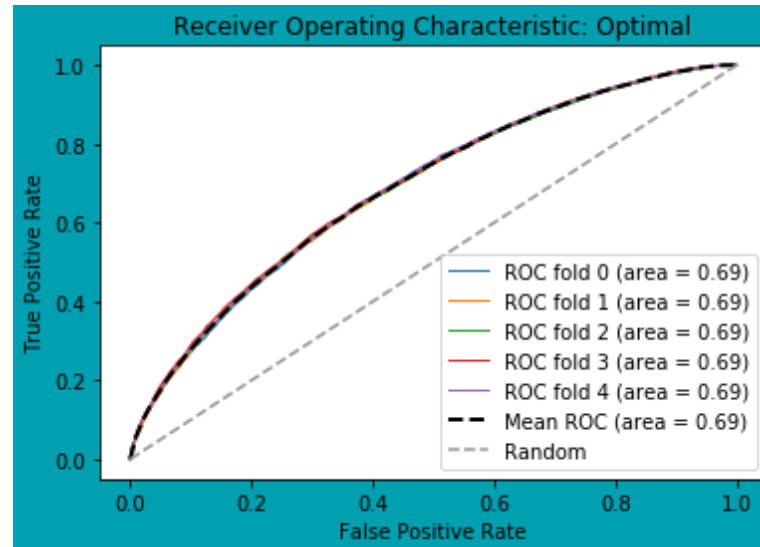
AUC: 0.70

ROC Curves - Optimized

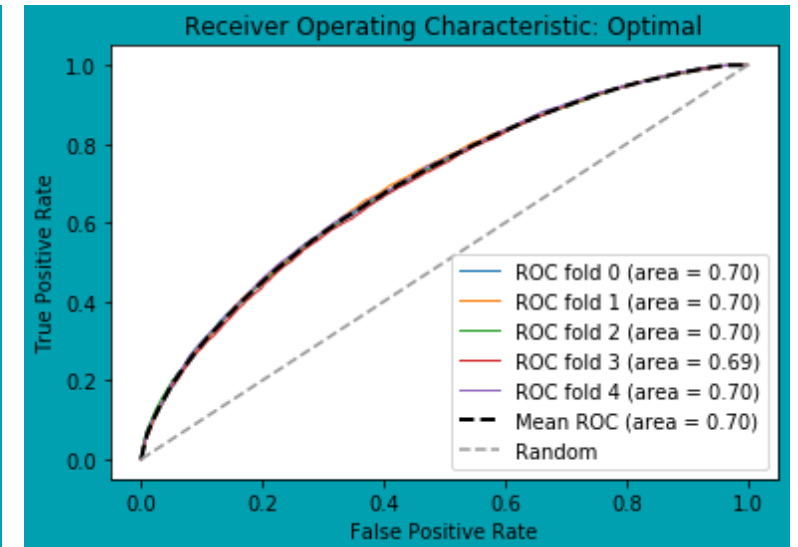
Logistic Regression



Random Forest



Gradient Boosting



Conclusion

- **Performance**

- All 3 models performed pretty well out-of-the box.
- Improvements after optimizing hyper-parameters in 2 of the models.

- **Next steps**

- Different approach to pre-processing.
- Only test select features.
- Redefine target classes - limit re-admissions to <30.
- Try different models.

Thank you!