

# **AbleTo Data Science Take-Home Assessment Presentation**

**Matthew Melendez to Data Science Team**

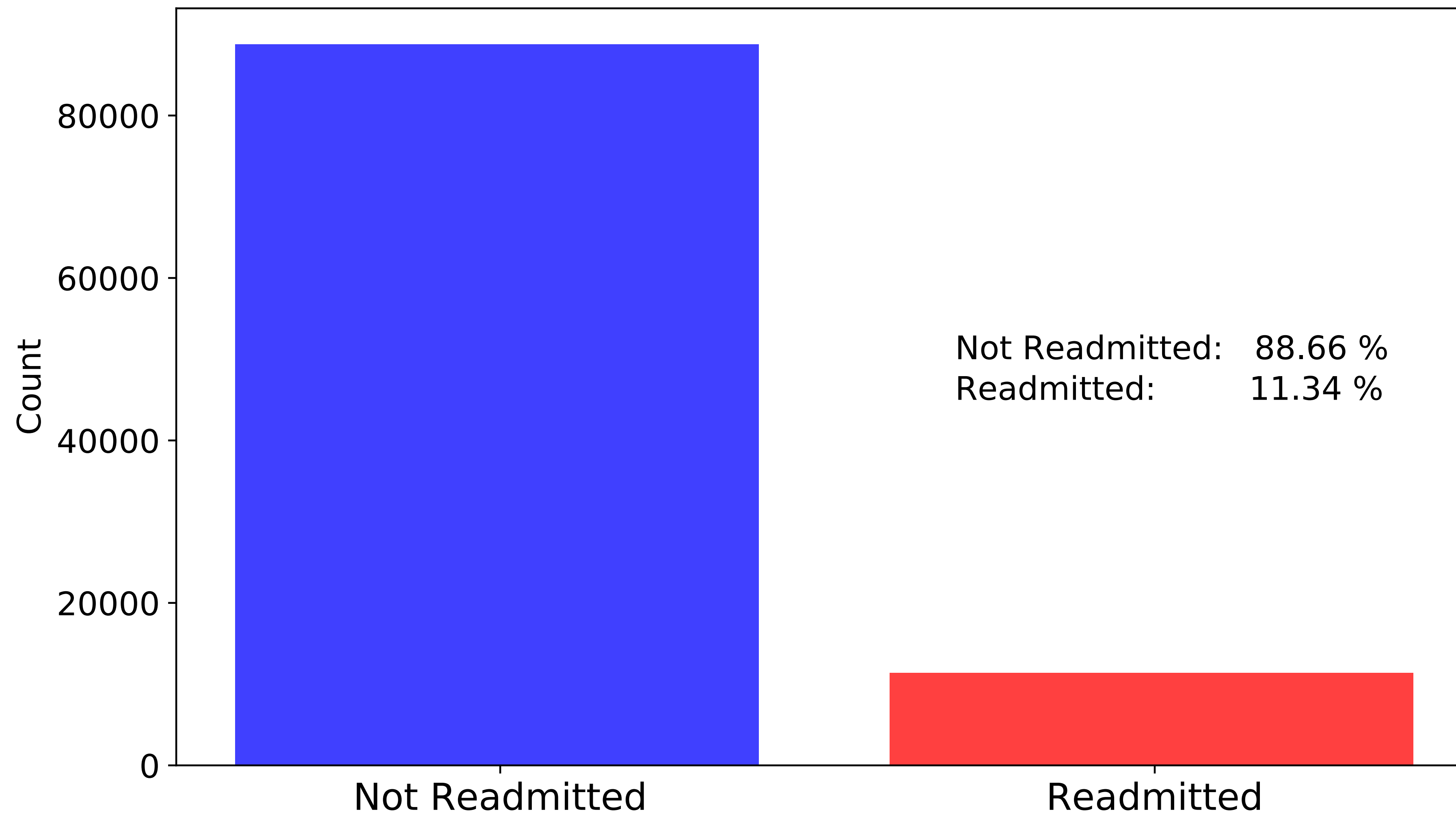
July 13, 2021

# Project Details

- Create a Machine Learning Algorithm to determine if patients were likely to be readmitted to the hospital within 30 days of being discharged
- Data available included:
  - Patient Information: Age, Weight Class, Race
  - Hospital Information: admission type, time in hospital, number of medications and procedures, top diagnoses, lab results, and medication change info
  - Whether or not the patient was readmitted within 30 days

# Data Summary

Patient Readmittance Imbalance



# Data Summary

- No repeat patients
- Imbalances within some of the features
  - weight only had ~3% filled in
  - payer\_code and medical\_specialty have ~1/2 missing values
- Some medications had only a few people taking them, and a smaller percentage of changes to said medication

```
data1.pioglitazone.value_counts()
```

No	94438
Steady	6976
Up	234
Down	118

Name: pioglitazone, dtype: int64

# Feature Engineering

- One-Hot-Encoded: race, gender, change of medication, medication
- Categorized: age, weight, payer code, medical specialty, tests, and medication information
- ICD  $\rightarrow$  CCS codes for diagnoses
- Removed 'expired' patients
- Didn't feel the need to combine any categories, any correlated features?
- Overall DataFrame dimensions: 100114 rows  $\times$  48 columns

# Machine Learning Algorithm

- XGBClassifier with GridSearchCV for hyper-parameter tuning with EarlyStopping
- Other feature engineering variations were also tested along with the above

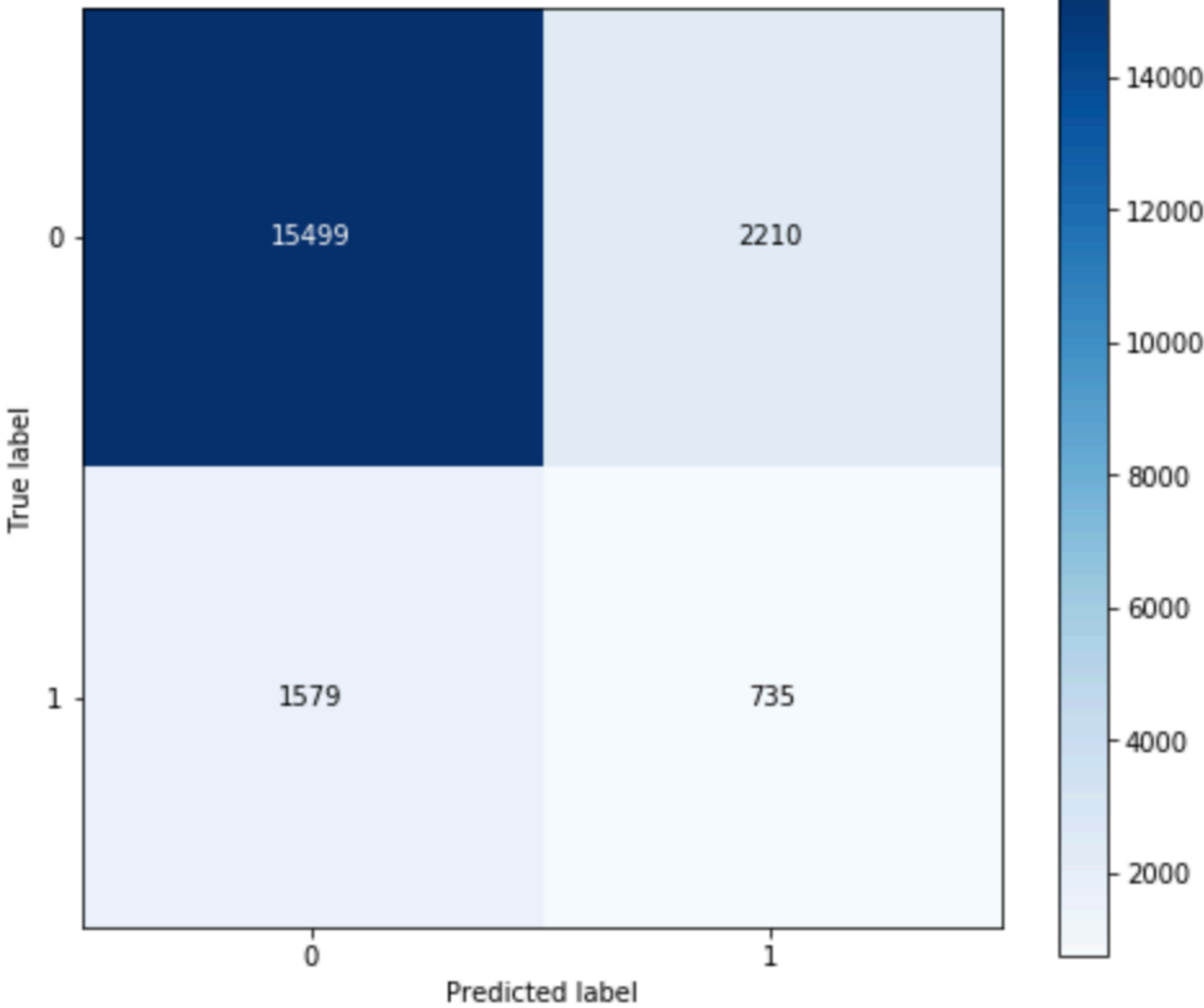
```
params = {'n_estimators': [400, 500, 600], 'max_depth': [3, 4, 5],  
          'learning_rate': [0.05, 0.1, 0.15], 'early_stopping_rounds': [10],  
          'scale_pos_weight': [1, 5, 10]}  
  
xgb_model = GridSearchCV(x_cfl, param_grid = params, verbose = True, n_jobs = -1,  
                        scoring = 'f1', cv = 3)
```

- Final Model:

```
clf_mod = XGBClassifier(scale_pos_weight = 5, learning_rate = 0.1, max_depth = 3,  
                        n_estimators = 600, early_stopping_rounds = 10, n_jobs = -1)
```

# Results

Confusion matrix



	precision	recall	f1-score	support
0	0.91	0.88	0.89	17709
1	0.25	0.32	0.28	2314
accuracy			0.81	20023
macro avg	0.58	0.60	0.59	20023
weighted avg	0.83	0.81	0.82	20023

Accuracy: 81.08%

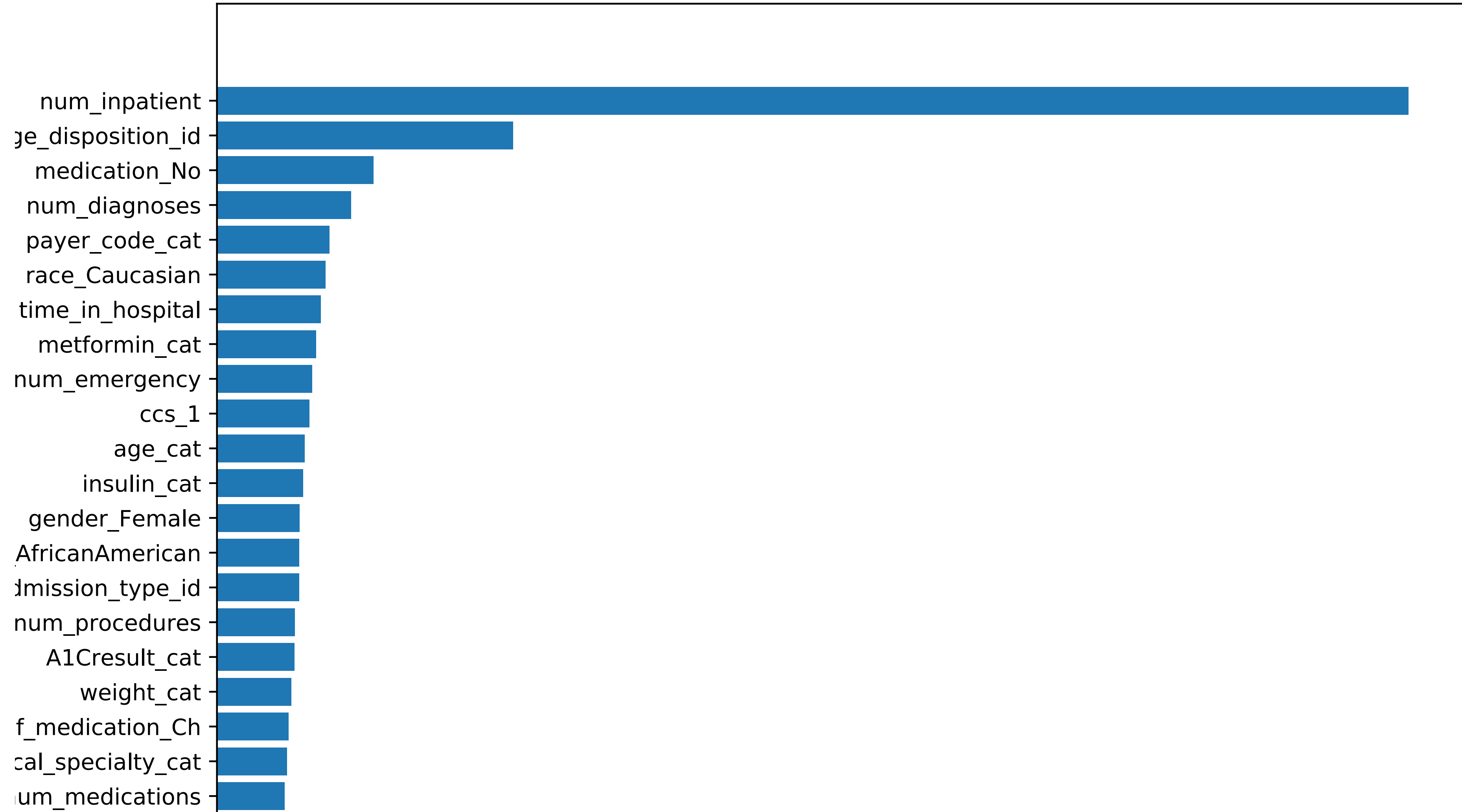
Precision: 57.86%

F1 Score: 58.53%

Recall: 59.64%

# Feature Importance

XGBClassifier Model Feature Importance





# Results

- Most important features to determine if a patient will be readmitted within 30 days:
  - Number of inpatient visits the previous year
  - Discharge ID
  - Number of Medications
  - Number of Diagnoses
  - Payer Code
  - ...

# Next Steps?

- Look into creating new features based on admission\_type:
  - If listed as Emergency, Urgent, or maybe Trauma then can give a 1 in new column as those might be likely to result in a higher chance of being readmitted
- Similar with discharge\_type:
  - If sent to another facility/ICU or 'expected to return for outpatient services' then can create a flag/new feature with 1
- Would want to talk to someone who knows the data more for this
- Do more hyper-parameter tuning
- Test out other algorithms (would need to change feature engineering)

# **AbleTo Data Science Take-Home Assessment Presentation**

**Matthew Melendez to Data Science Team**

July 13, 2021