

WARRIOR STRONG



WAYNE STATE
UNIVERSITY

NORTHSTAR1

Lakshmi Poojitha Challagali | Paul Merica | Rachel Seifert | Sai Shiva Bhaskar Emani | Soundarya Alwa



Business Problem

Arbor Hospice wants to identify when a patient has 14 days, 7 days or less to live.

This prediction is helpful for any hospice

- To have additional time to maximize the resources available for an ailing patient
- To release more funding from the government(Medicare, Medicaid, etc.) which in turn helps improve patient care
- To allow the organization to receive reimbursement incentives from other funding organizations



Hospice Background

- North Star Care Community
 - Non-profit hospice care for both children and adults
 - Formed by two legacy hospices- Hospice of Michigan & Arbor Hospice
- Provides end-of-life and palliative care services to patients who are typically within 6 months or less from end-of-life
- Approximately 1,100 patients are cared for per day; constant churn of patients
- Data is collected during every visit; nearly 20 years of EHR (electronic health record) data was provided for analysis



Dataset Information

- Data was provided in the form of tables - of which below tables have useful features:
 - Visits (2373526, 94)
 - Admissions (33750, 38)
 - Hospice Item Set (33263, 28)
 - Statuses (109519, 29)
 - Clinical Measures (504300, 15)
 - PPS (336035, 3)
- Combination of numerical, date, string- categorical variables



Challenges/Data Issues for Modeling

- High level of missingness in the data
- Data collected at various time intervals leading to inconsistency
 - During the time of admission
 - At every visit
 - On recertification
- Patients may be admitted to hospice, leave and then be readmitted



Data Preparation

Visits

- Data has been prepared with a cutoff frequency of 7 days with each Monday as the starting point of the week from year 2015 to 2023
- The features from visits data has been created with respect to type of visits, Duration in minutes, Total Cost etc., happening in the last 0-7, 8-14, 15-30, 30-90, 90-365 days based on the cutoff date
- The target variable has been generated using the visits data where the patient had a death visit and defining 14 days before from the date of death as indication of patient health being deteriorated



Data Preparation

Admissions

- We built features such as number of days since the patient has been admitted to the hospice and building the dummy variables to the type of caregiver which has been provided to the patient and the referral type of the patient.

PPS

- Created the features based on the PPS score for the patient by calculating the Minimum, Maximum, Mean, Standard Deviation in the last 0-7, 8-14, 15-30, 30-90, 90-365 days based on the cutoff date



Data Preparation

Clinical Measures

- Created the features based on WeightValue, ExtremityValue, AnxietyActual, NauseaActual, Shortness of Breath in the last 0-7, 8-14, 15-30, 30-90, 90-365 while calculating the Mean, Standard Deviation

Statuses

- Created the features based on PRIMARY DIAGNOSIS GROUP DESCRIPTION by creating a dummy variable list for each individual disease type



Data Preparation

- We built an overall of 181 features based on the data provided by the Hospice using Visits, Admissions, PPS, Clinical Measures, Statuses and Hospice Item Set tables

```
Index(['PATIENT_ID', 'Visits_Up_To_Date', 'Days_Since_Admission',  
      'Age_in_days', 'Sum_Direct_Duration_Minutes_Since_Admission',  
      'Sum_Direct_Duration_Minutes_Over_Past_Week',  
      'Sum_Direct_Duration_Minutes_8_days_to_14_days',  
      'Sum_Direct_Duration_Minutes_15_days_to_30_days',  
      'Sum_Direct_Duration_Minutes_30_days_to_90_days',  
      'Sum_Direct_Duration_Minutes_90_days_to_a_year',  
      ...  
      'REFERRAL_TYPE_Generic_other', 'REFERRAL_TYPE_Hospice_Other',  
      'REFERRAL_TYPE_Hospital_Contracted',  
      'REFERRAL_TYPE_Hospital_non_contracted',  
      'REFERRAL_TYPE_NH_Contracted_Nonskilled_Bed',  
      'REFERRAL_TYPE_Other_Referrals', 'REFERRAL_TYPE_Patient_Task_Category',  
      'REFERRAL_TYPE_Physician_Practice_Clinic',  
      'REFERRAL_TYPE_Physician_Non_HOM',  
      'REFERRAL_TYPE_Referring_Agency_Misc'],
```

| | PATIENT_ID | Visits_Up_To_Date | Days_Since_Admission | Age_in_days | Sum_Direct_Duration_Minutes_Since_Admission |
|---|------------|-------------------|----------------------|-------------|---|
| 0 | 14716 | 3 | 1 | 34787 | 285.0 |
| 1 | 14930 | 1 | 1 | 21482 | 130.0 |
| 2 | 23315 | 2 | 1 | 32997 | 190.0 |
| 3 | 1517 | 8 | 4 | 23377 | 612.0 |
| 4 | 3343 | 6 | 4 | 18985 | 345.0 |

5 rows × 181 columns



Modeling Choice: XGBoost

- Good at dealing with data with smaller signals
- Can handle missing values
- Does not suffer issues from collinearity
- Was the most accurate model in our initial testing
- Offers hyperparameters to tune for better prediction

```
Percentage of 1's: 2.926760004230762  
Percentage of 0's: 97.07323999576923
```



Robustness and Hyper-Parameter Tuning

- Used RandomizedSearchCV to find the best hyperparameters based on cross-validation performance

Performed 5-fold Validation

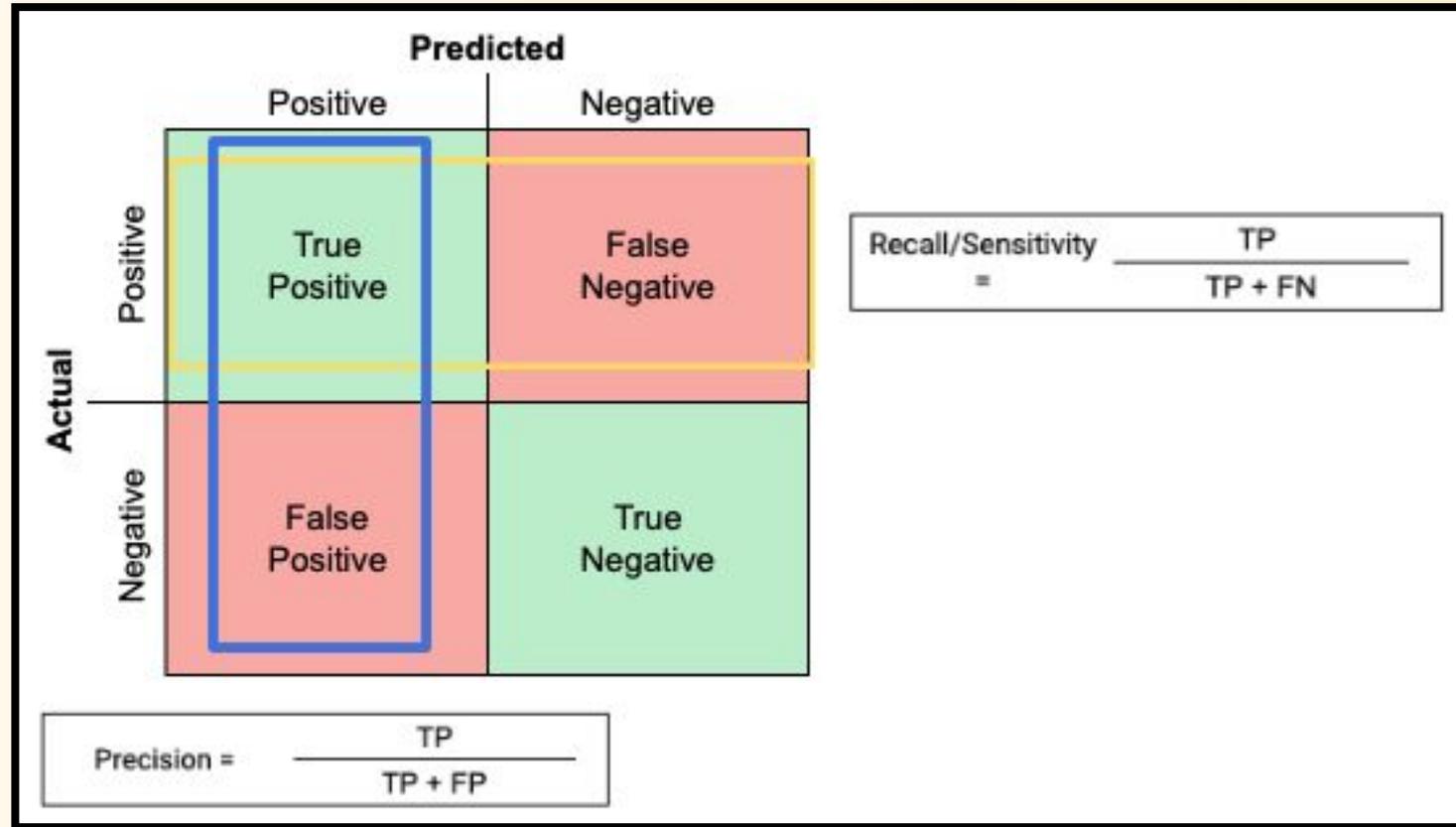
For accuracy model, used accuracy as the metric for best model

For aggressive model, used Sensitivity/Recall as the metric for the best model



Model Approach/Evaluation

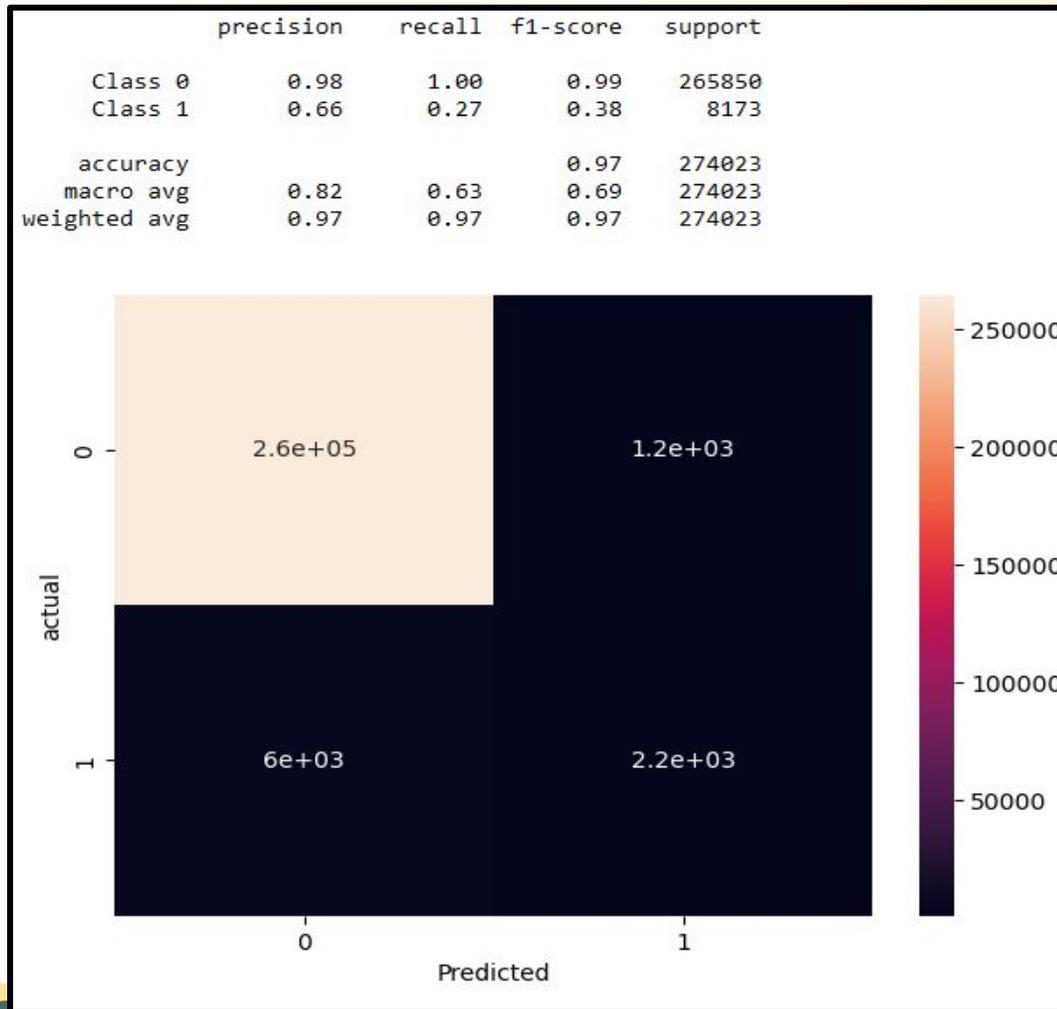
- Client wants to predict patients that will die within 14 days, not patients who won't.
 - With 97% of records being patients who didn't die in 14 days, you could get 97% accuracy guessing 0 every time.
- Thus, we developed two models one where accuracy was the goal, and one where Recall/Sensitivity of the positive class was prioritized as the client mentioned they care the least about false negatives.



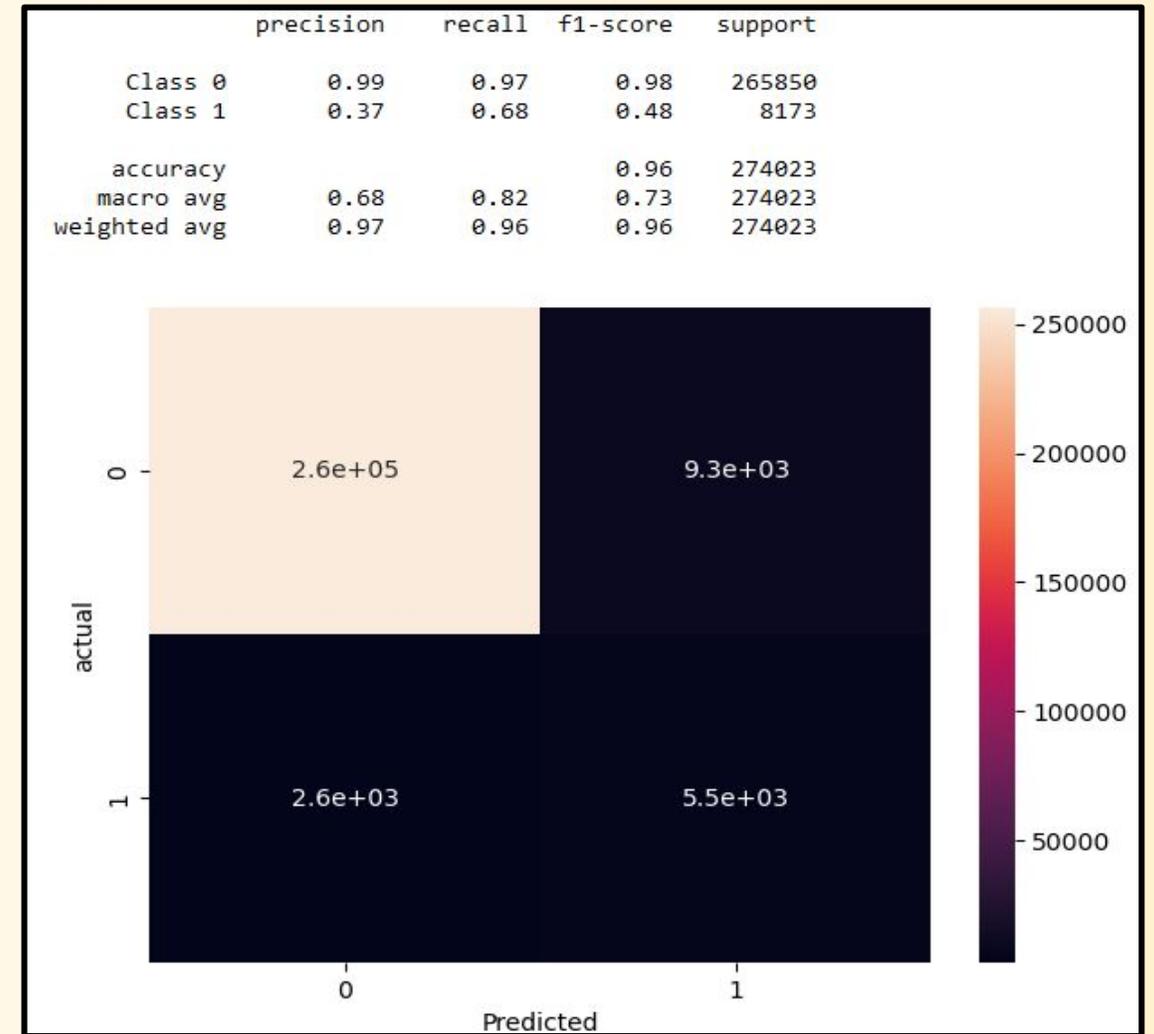
*Note True Positive (TP) = We predicted a patient would die in 14 days, that did die in 14 days. False Negative (FN) means we predicted a patient wouldn't die in 14 days and they did. False Positive means we predicted they would die in 14 day and they didn't. True Negative means we predicted they wouldn't die and they did not.



Unbalanced XGBoost (Accuracy Model)



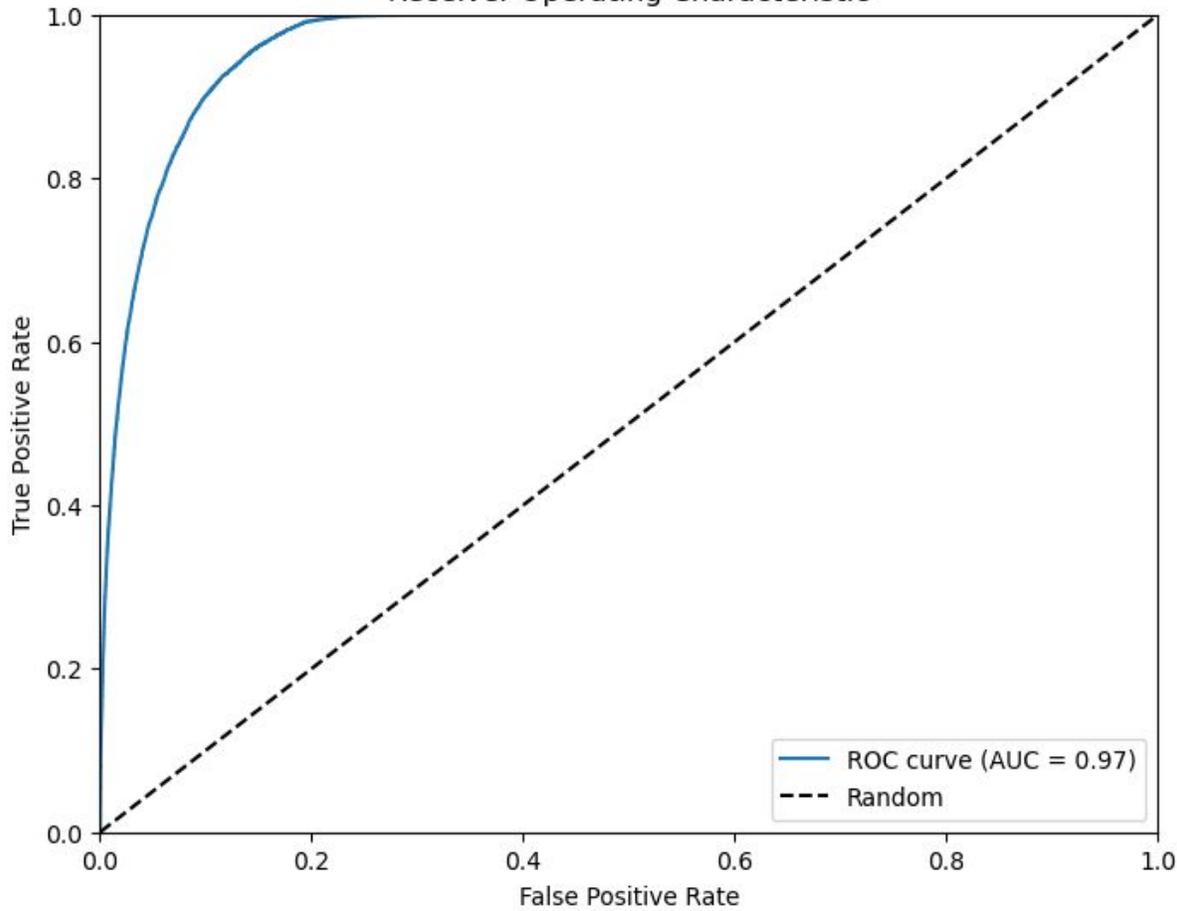
Balanced XGBoost (Aggressive Model)



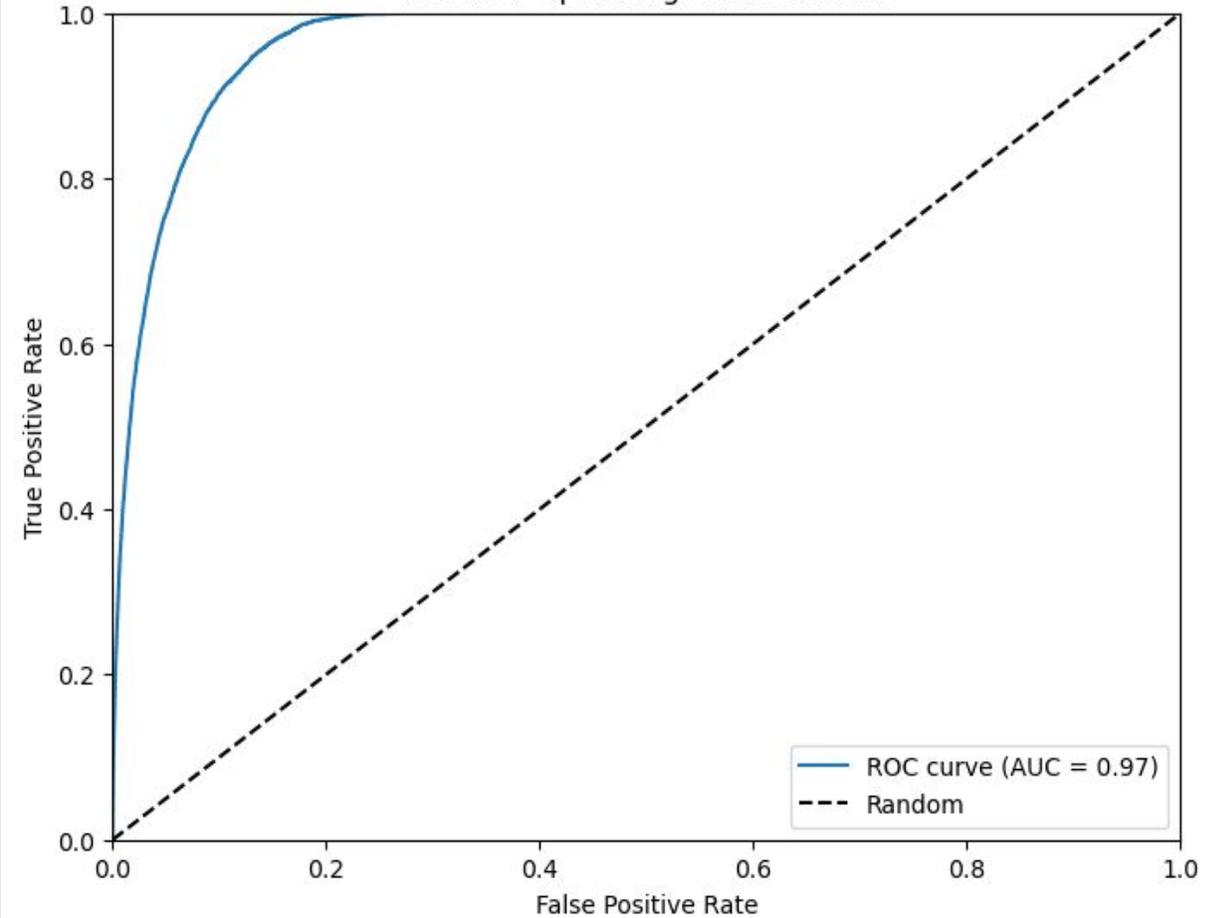
Unbalanced XGBoost (Accuracy Model)

Balanced XGBoost (Aggressive Model)

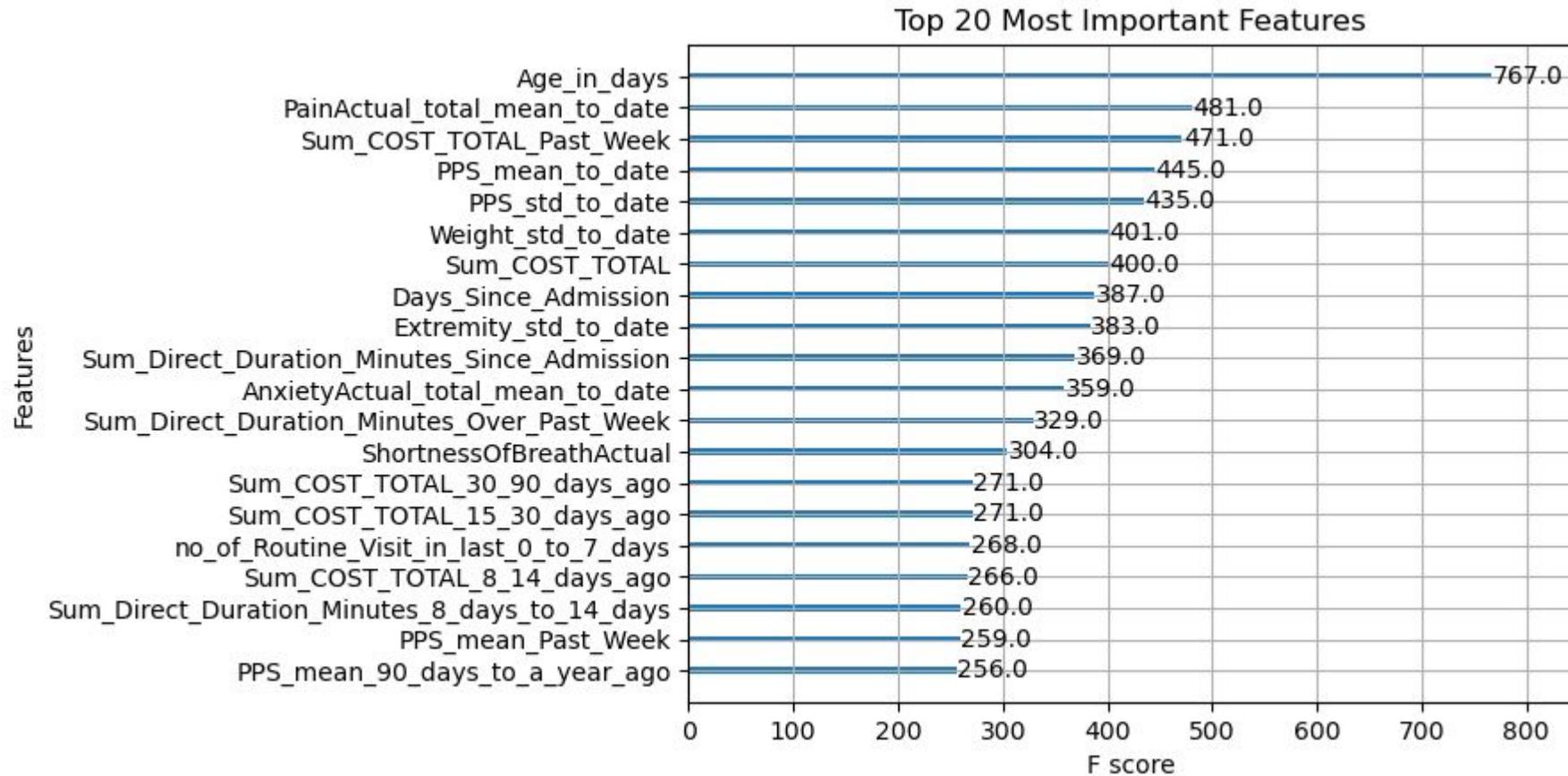
Receiver Operating Characteristic



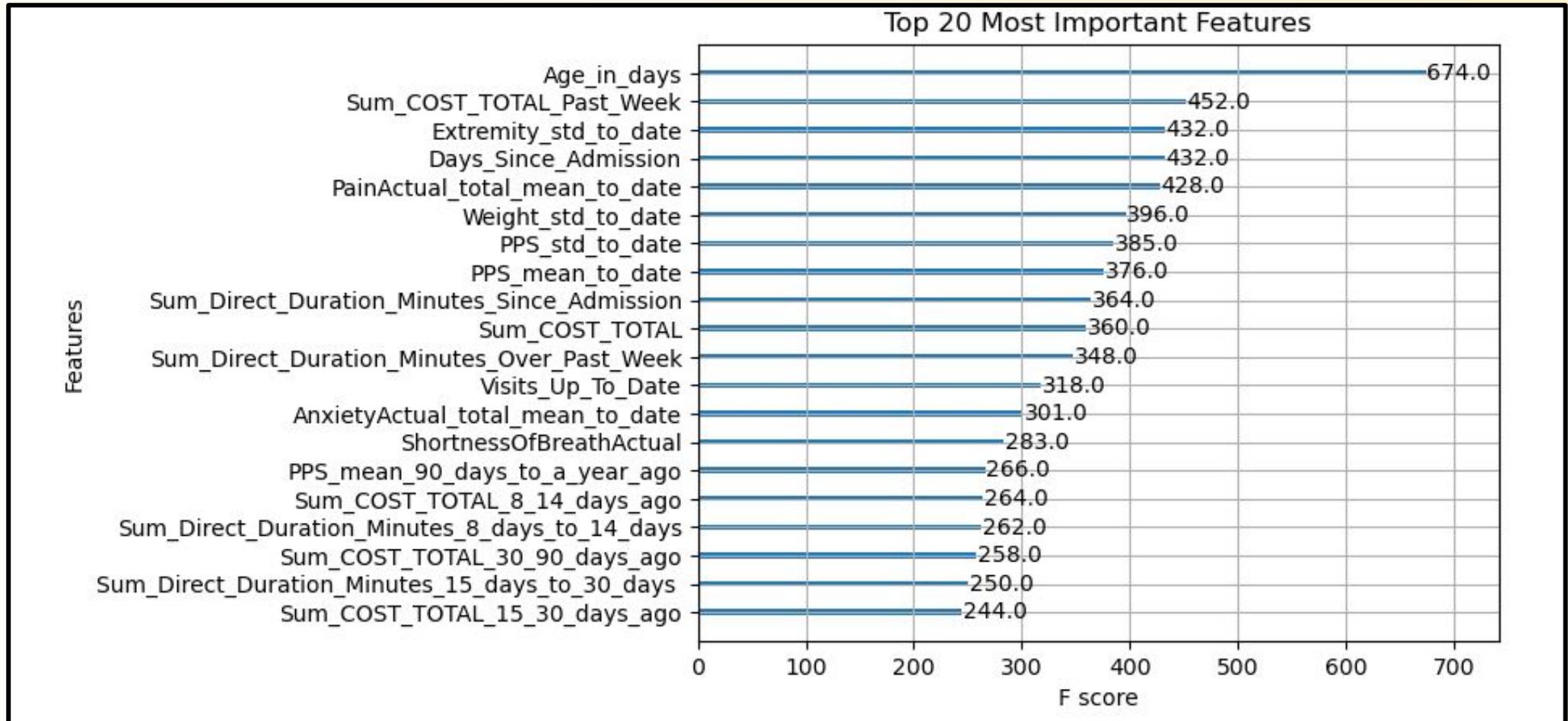
Receiver Operating Characteristic



Unbalanced XGBoost Top 20 features (Accuracy Model)



Balanced XGBoost Top 20 features (Aggressive Model)



Final Deliverables

- CSV's from both models with a probability of each patient dying in next 14 days (as shown on the right)
- Set of models predicting 7 & 14 days until death (Metrics provided in appendix for 7 day model)

| PATIENT_ID | Probability | TEAM |
|------------|-------------|---------------|
| 9644 | 92.39% | NW |
| 15564 | 87.54% | Arbor Hospice |
| 32017 | 83.81% | NW |
| 27342 | 82.08% | SE |
| 6360 | 81.02% | SE |
| 32001 | 77.13% | SW |
| 32039 | 74.59% | SE |
| 32031 | 72.81% | Arbor Hospice |
| 32040 | 72.30% | NW |
| 17382 | 70.28% | NE |
| 14272 | 65.56% | Arbor Hospice |
| 18381 | 64.37% | Arbor Hospice |
| 7782 | 62.59% | NE |
| 7782 | 62.59% | NE |
| 18476 | 61.70% | SE |



Areas for Improvement in the Future

Engineering more features into the model

Integrating a much more powerful
Hyperparameter tuning and Cross Validation

Making use of Hospice Item Set and Statuses
table in-depth



THANK YOU

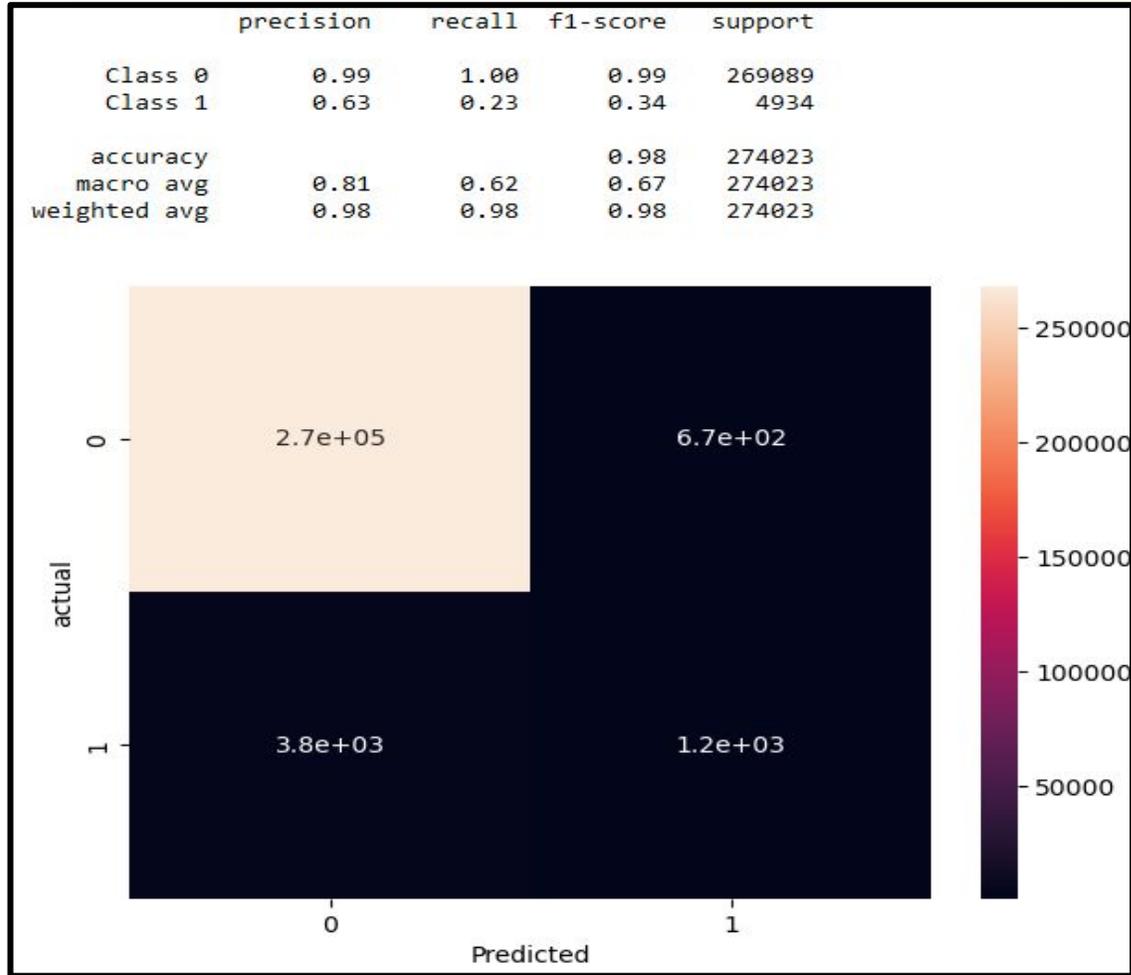
Team Northstar1



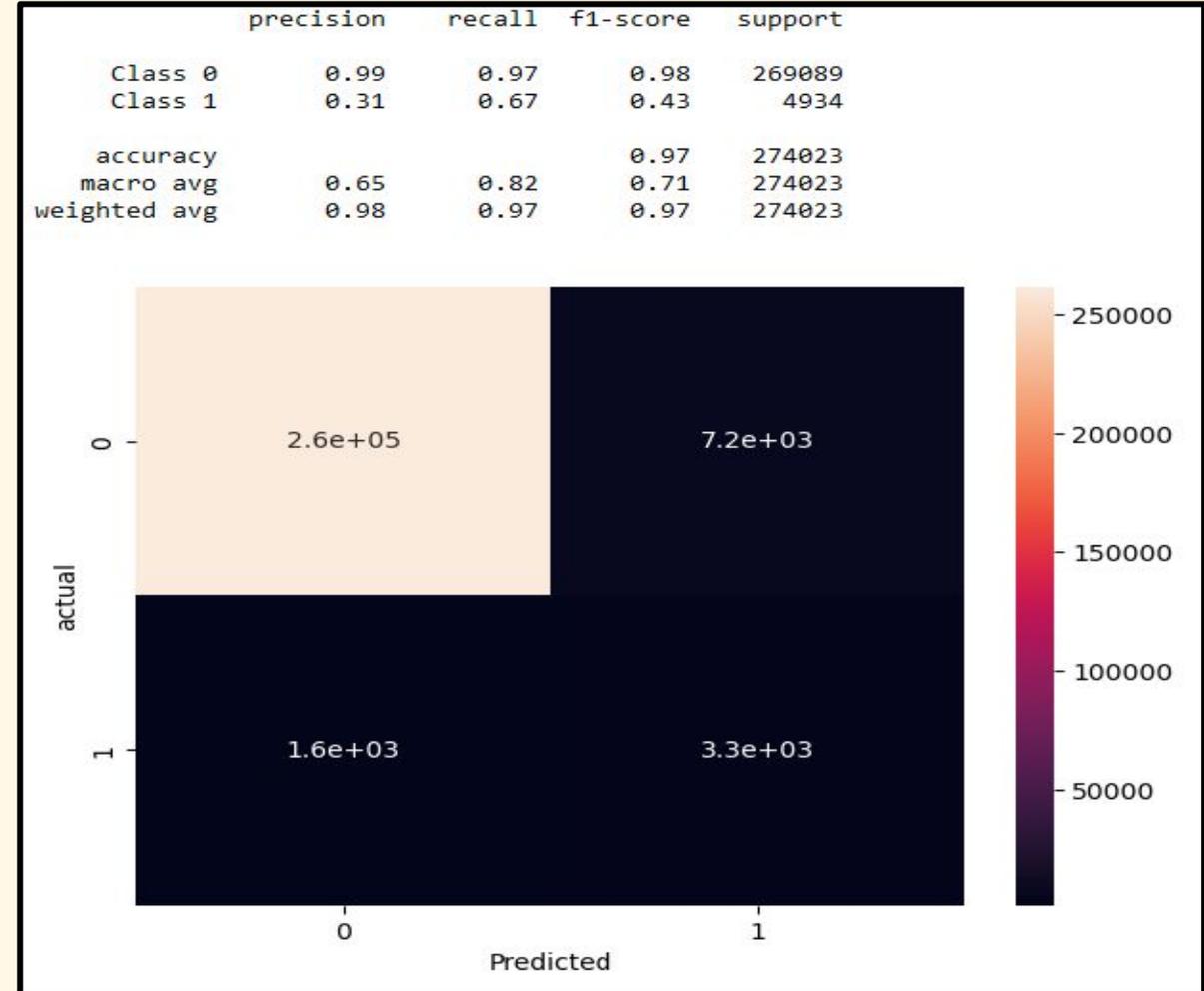
APPENDIX



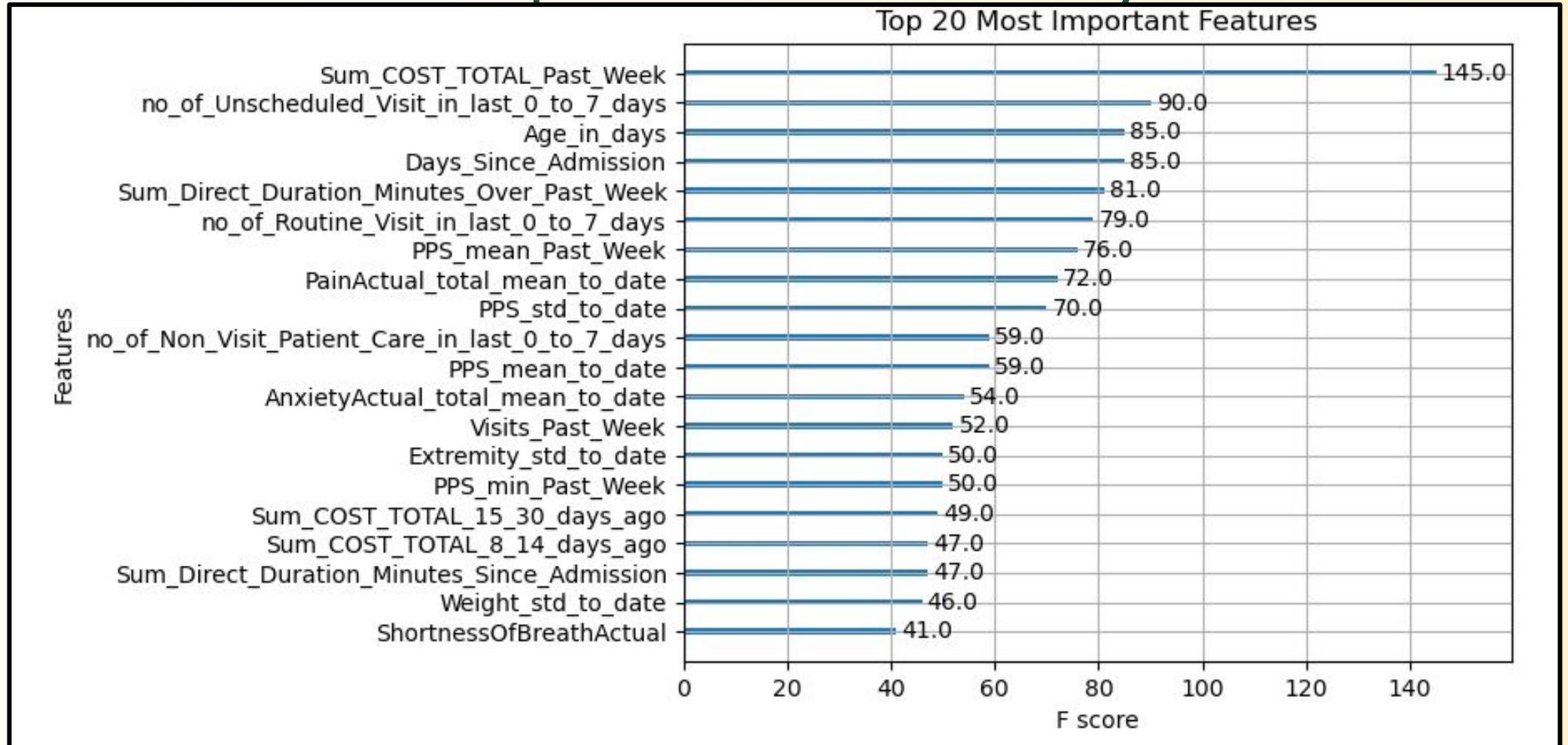
Unbalanced XGBoost (Accuracy Model 7 Days)



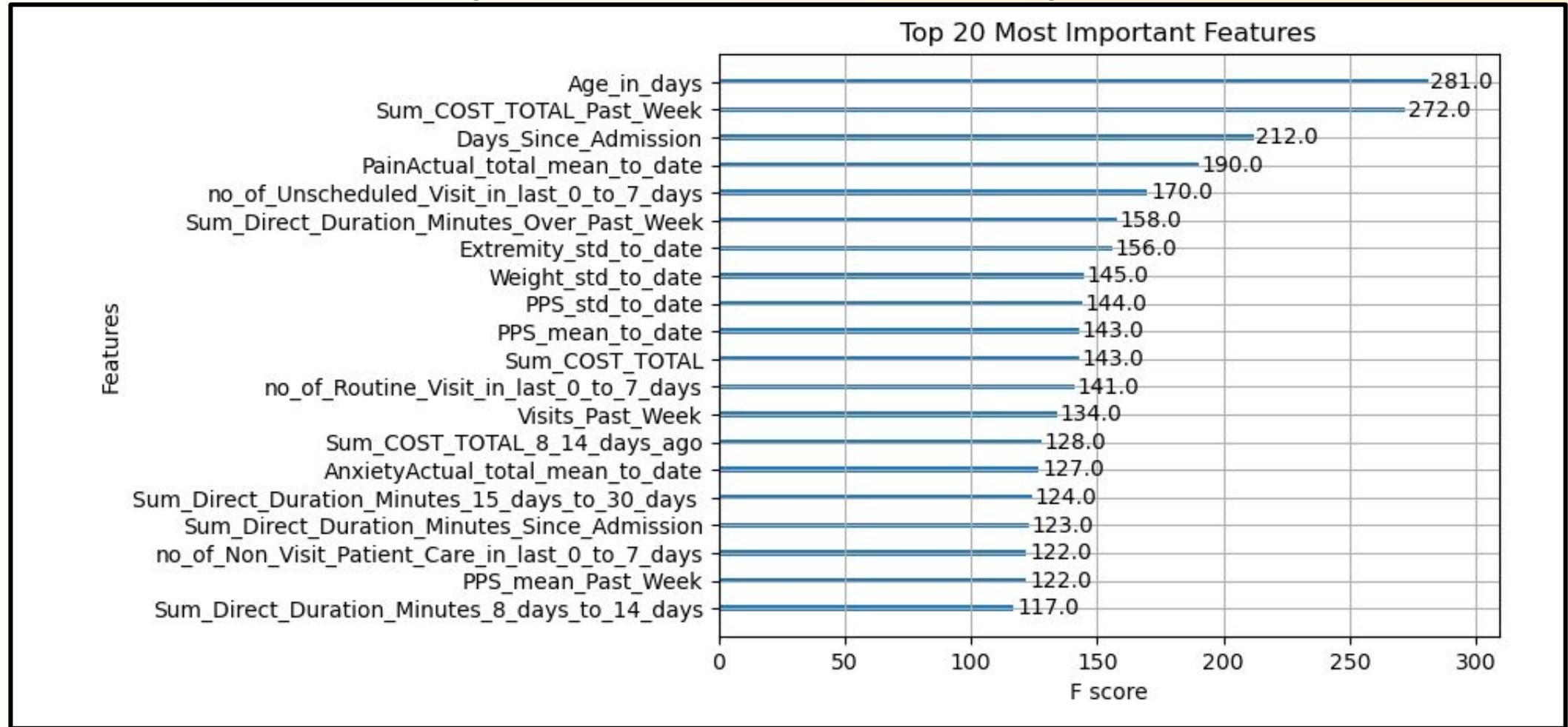
Balanced XGBoost (Aggressive Model 7 Days)



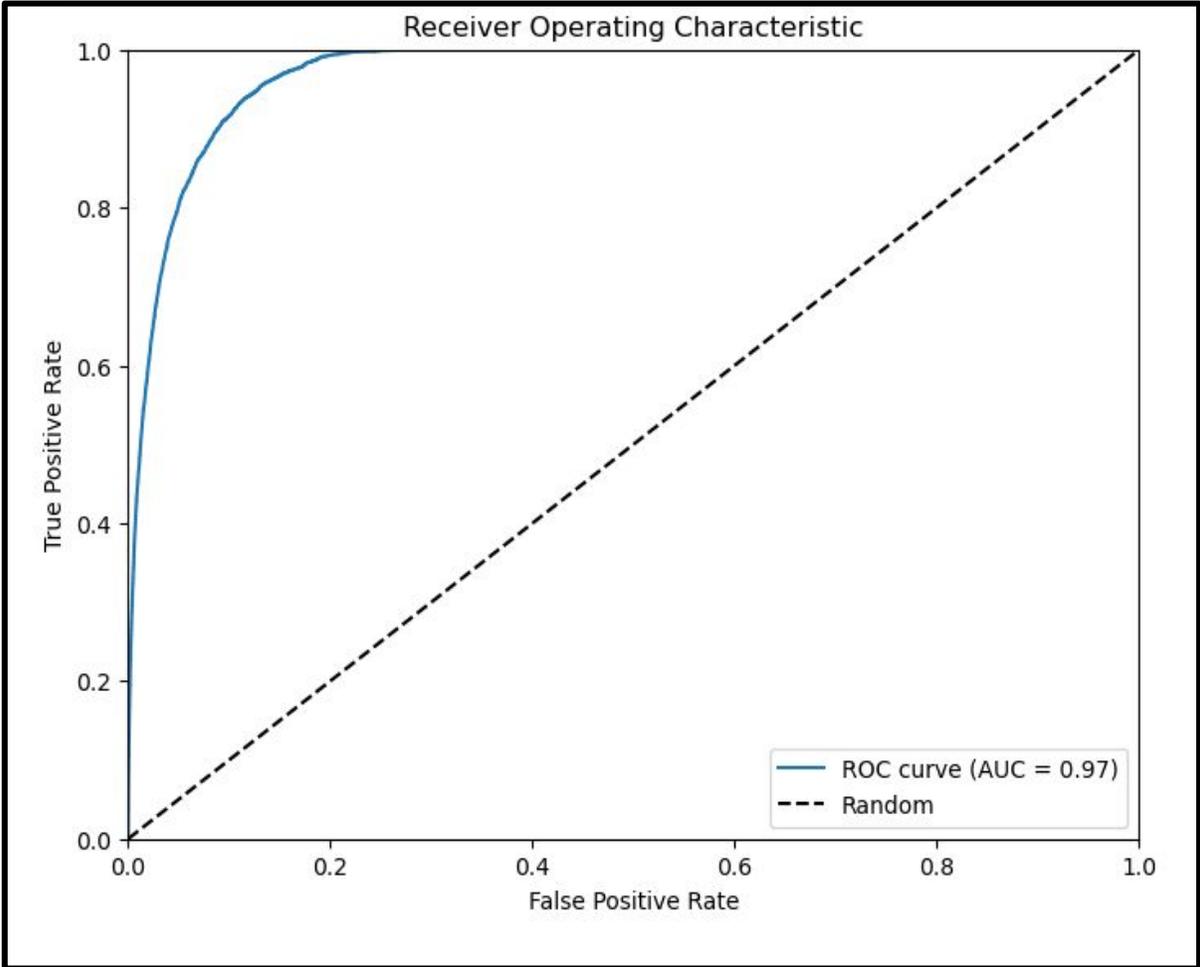
Unbalanced XGBoost Top 20 features – 7 Days



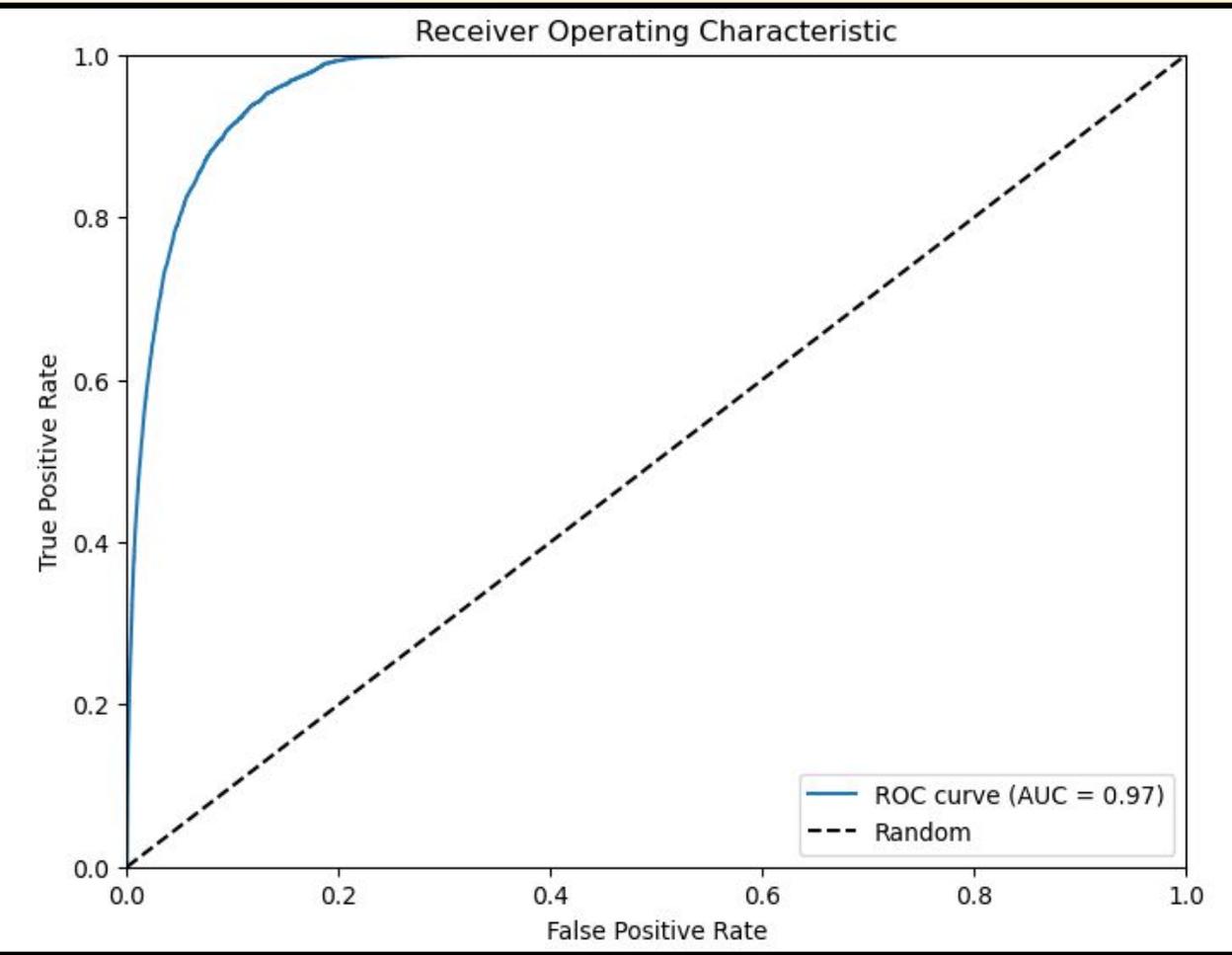
Balanced XGBoost Top 20 features – 7 Days



Unbalanced XGBoost (Accuracy Model 7 Days)



Balanced XGBoost (Aggressive Model 7 Days)



Logisitic Regression Coefficients 14 days & 7 days

```
Feature Coefficients:  
Age_in_days: 0.002133003696445677  
Extremity_std_to_date: 0.0004255757897767644  
Sum_COST_TOTAL_Past_Week: -2.7569775646919333e-05  
Days_Since_Admission: 0.0004923801437272406  
PainActual_total_mean_to_date: 0.0005026568253795939  
Weight_std_to_date: -0.0009440567951168735  
Sum_COST_TOTAL: -0.0025864001360141713  
PPS_mean_to_date: -0.002140810954600139  
Sum_Direct_Duration_Minutes_Since_Admission: -0.0004394171104683692  
AnxietyActual_total_mean_to_date: 0.008530550225116822
```

14 days

```
Feature Coefficients:  
Age_in_days: 0.0024488098699259786  
Extremity_std_to_date: -0.002366938194143127  
Sum_COST_TOTAL_Past_Week: -7.830746278984367e-06  
Days_Since_Admission: 0.0003883131244375402  
PainActual_total_mean_to_date: -0.0011692685152339078  
Weight_std_to_date: -0.001870034000611244  
Sum_COST_TOTAL: -0.002639416621393959  
PPS_mean_to_date: -0.0016434079218776772  
Sum_Direct_Duration_Minutes_Since_Admission: -0.0003665787738061906  
AnxietyActual_total_mean_to_date: 0.011868401836992998
```

7 days

