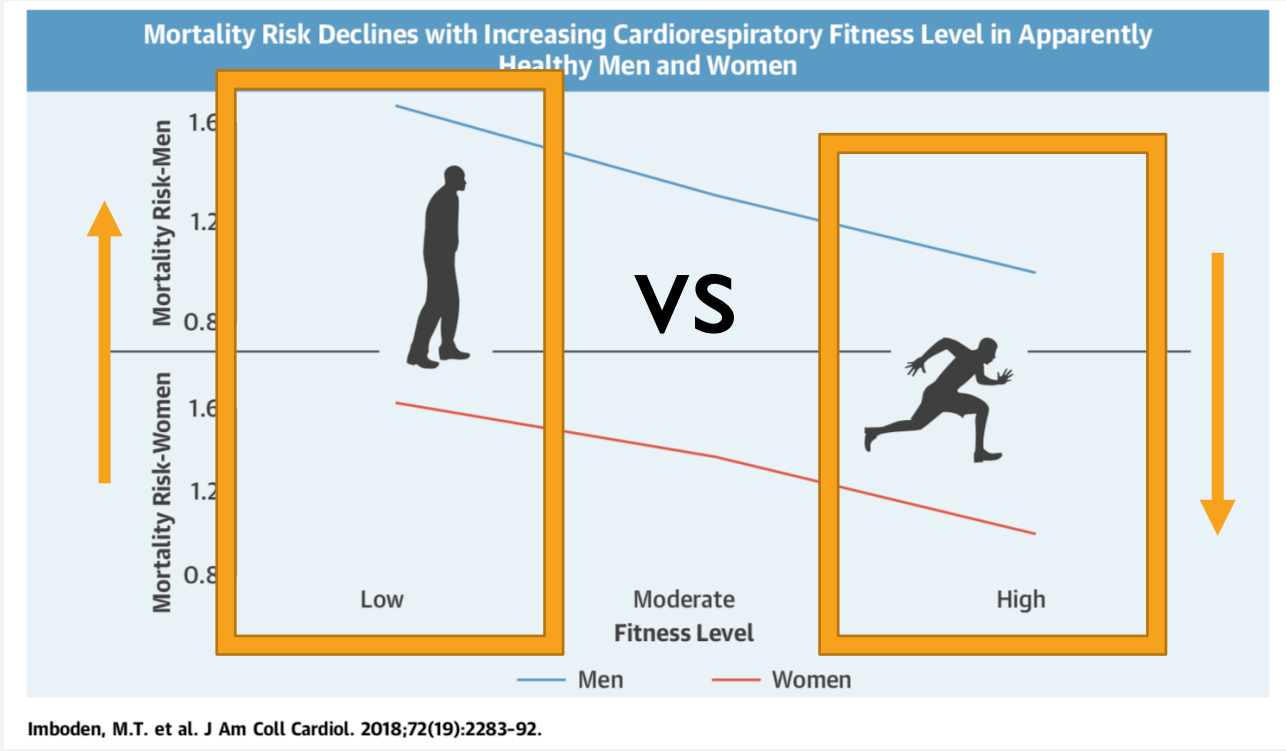
CREATING A CLASSIFICATION MODEL TO DETERMINE FITNESS TESTING PROTOCOL

Jim Peterman, PhD

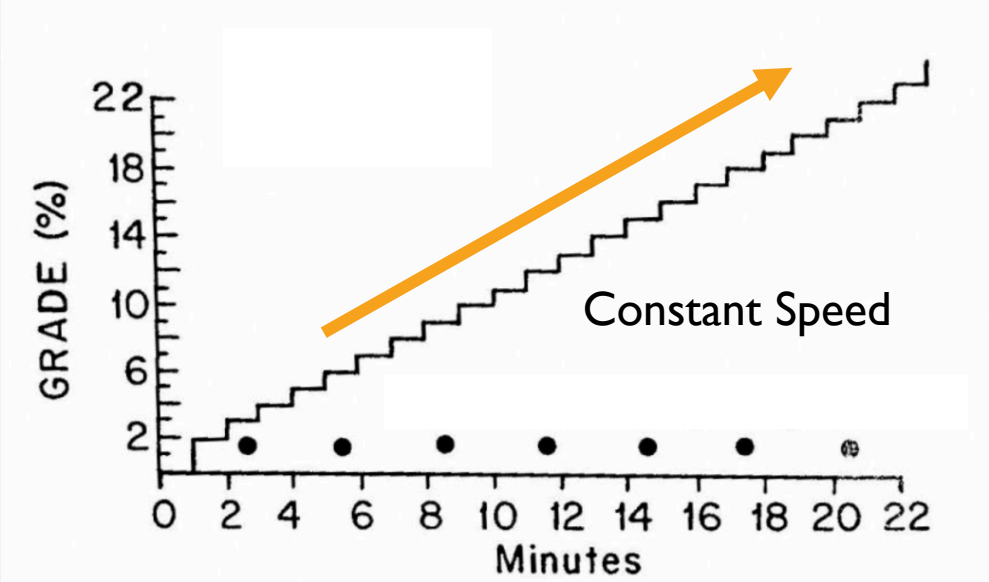
GitHub: JimPeterman

BACKGROUND

- Cardiorespiratory fitness ("fitness") is associated with health outcomes.
- The gold-standard for assessing fitness is with an exercise test in which the intensity gradually increases to maximal.

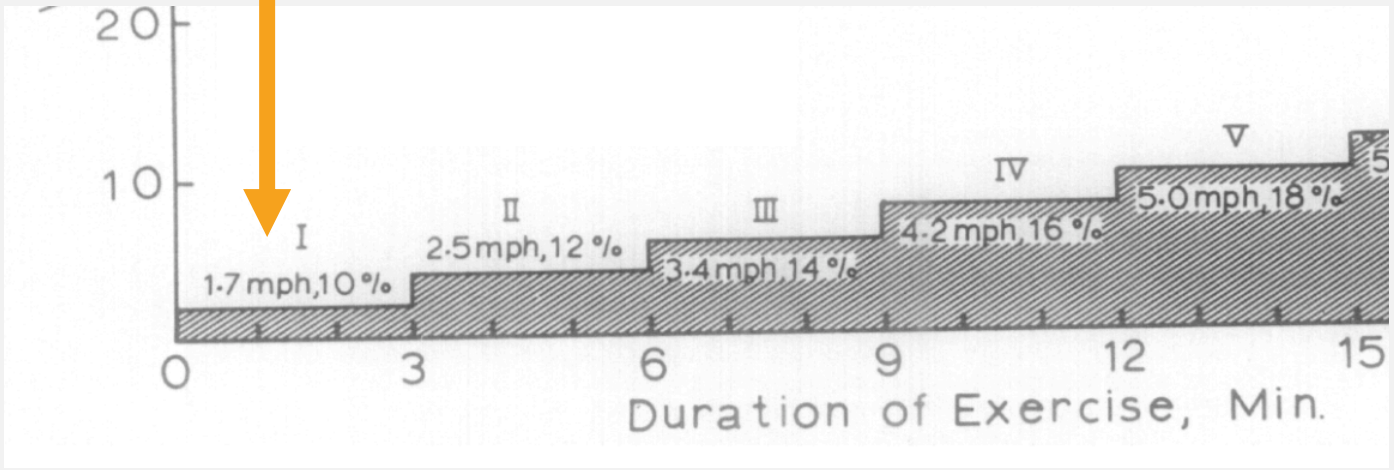


Imboden, M.T. et al. J Am Coll Cardiol. 2018;72(19):2283-92.



BACKGROUND

- The most common protocol begins at an absolute workload that is too high for individuals with a low absolute fitness level.
- Selecting the proper test protocol ensures accurate assessments of fitness and can improve patient risk stratification and patient care.



BACKGROUND

Objective

- Build a classification model using features traditionally assessed before an exercise test to determine whether someone likely has a low absolute fitness level.

Goal

- Develop a model that can determine an appropriate fitness test protocol based on pre-test values.

METHODS: DATA

- Data source: Fitness Registry and Importance of Exercise International Database (FRIEND).
  - Collection of pre-test and exercise test data.
  - Sample consisted of 6,252 tests.

Features

- *Numeric*: age, height, weight, BMI, percent body fat, resting heart rate, resting blood pressure.
- *Categorical*: status for sex, CVD, COPD, asthma, meeting physical activity guidelines, beta-blocker medication.
- *Created:* BMI category, hypertension status.

METHODS: MODELS

Exploratory Data Analysis

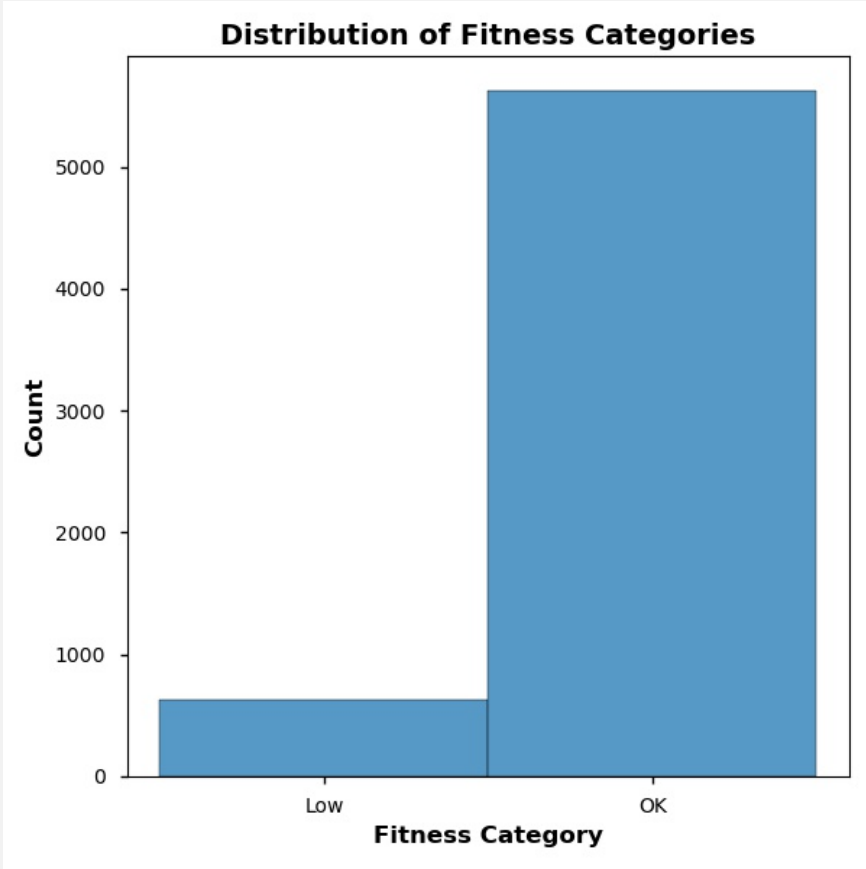
- Imbalanced groups

Classification Models

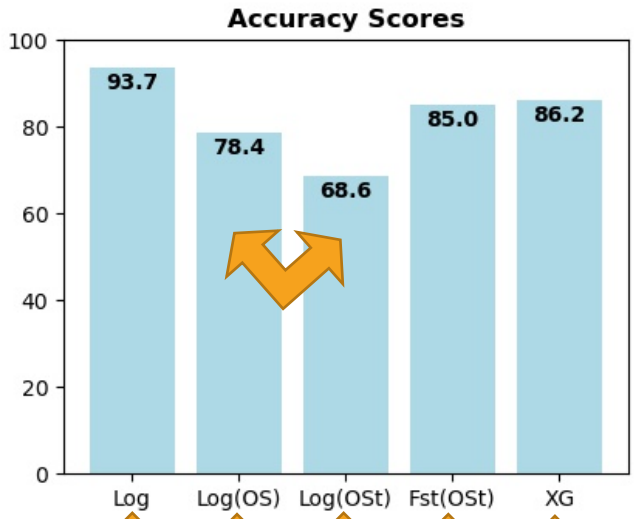
- kNN, logistic, decision tree, extra trees, random forest, naïve Bayes, XG Boost
- Models trained using cross validation and then tested on holdout data (60/20/20 split).
- Tested different sampling methods, decision thresholds, class weights, and ensemble methods.

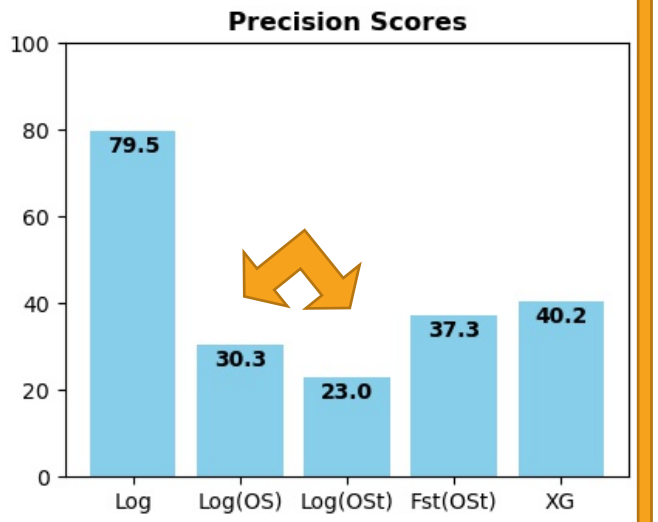
Metrics

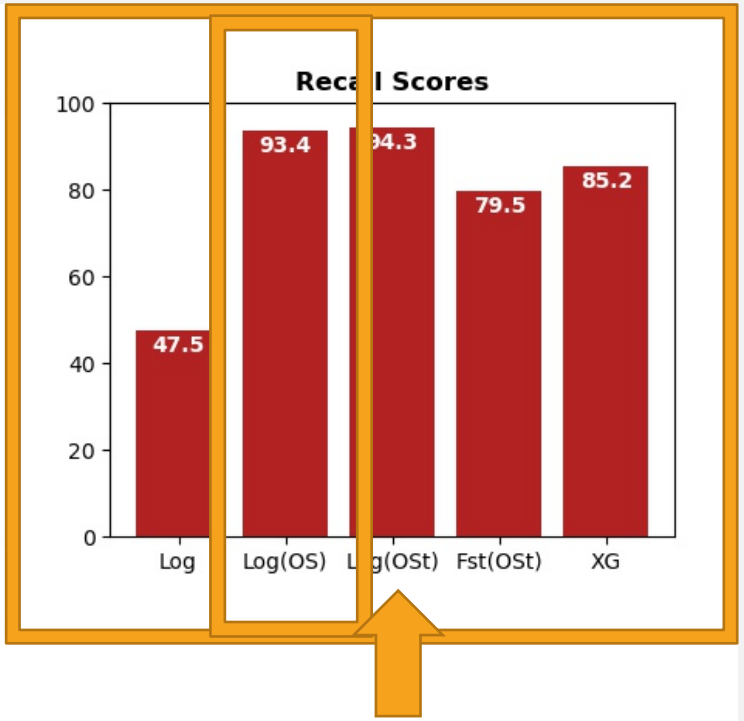
- Primary metric of interest was recall (less concerned about misclassifying someone as low fit).

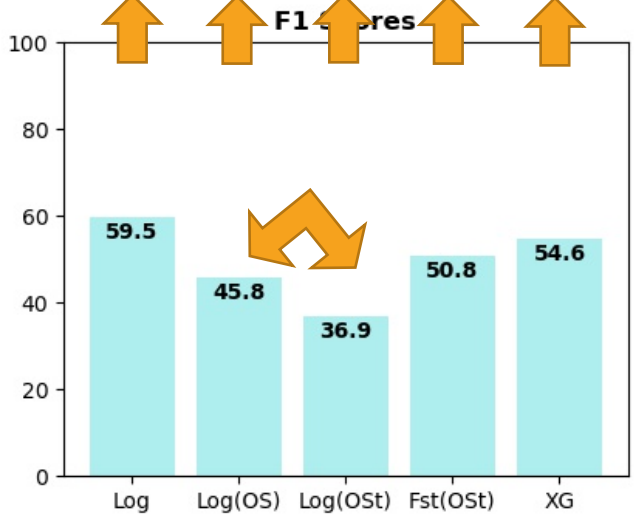


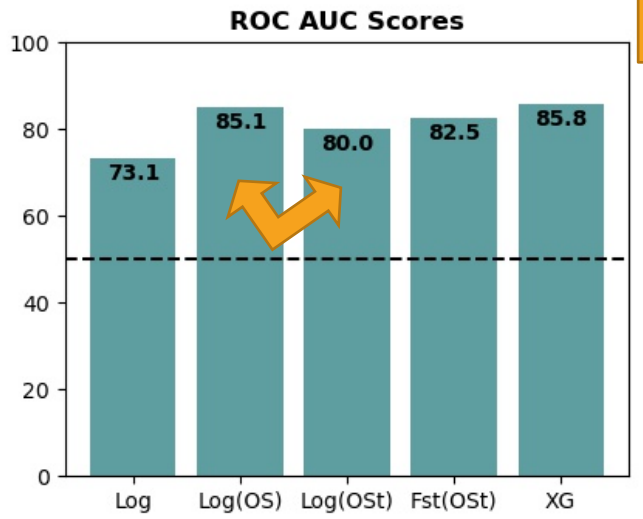
PERFORMANCE METRICS FROM A SAMPLE OF MODELS





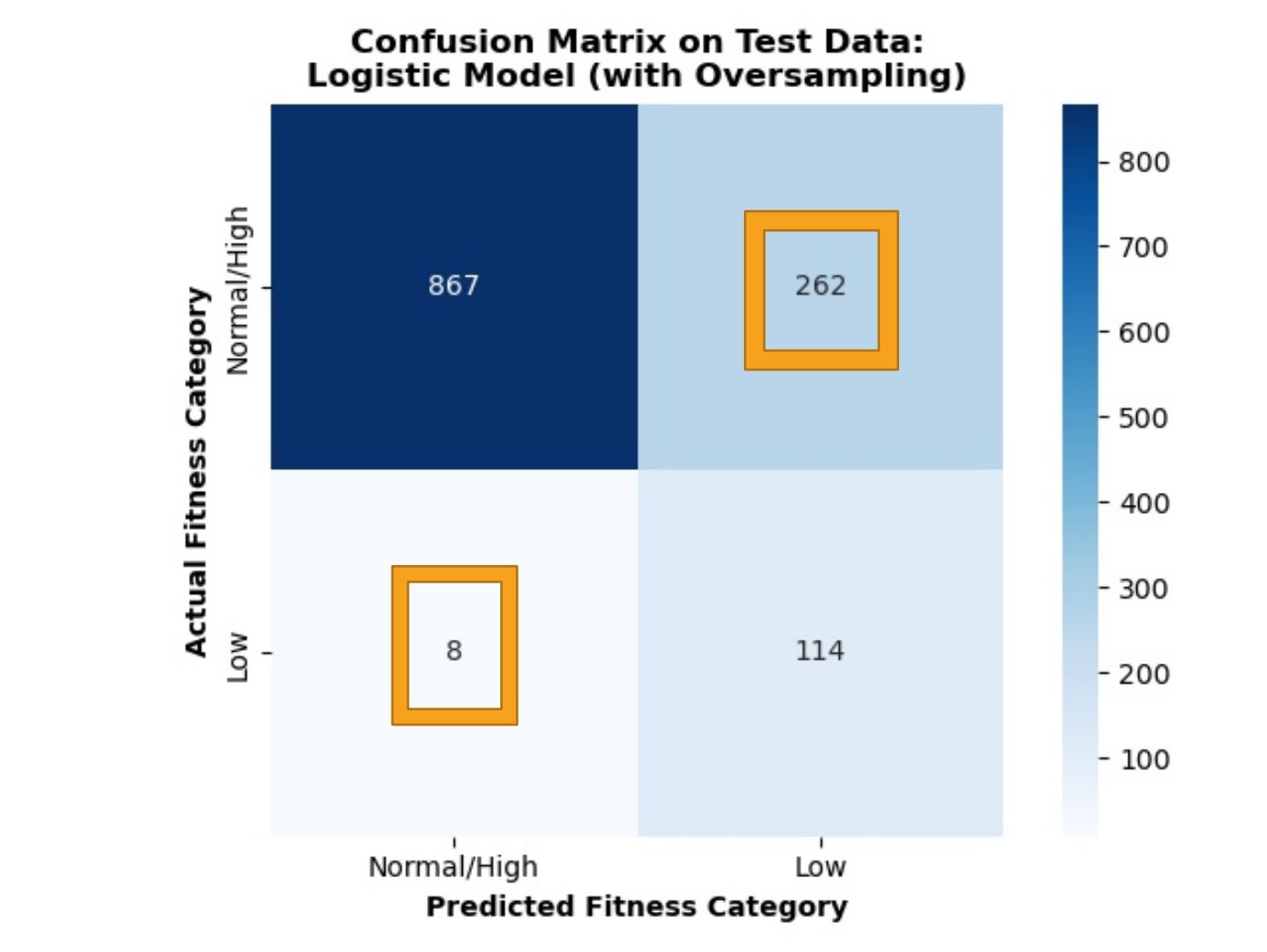






Log = Logistic Regression model
Fst = Random Forest model
XG = XG Boost model
(OS) = Oversampling procedure
(OSt) = Oversampling *and* lower decision threshold

CONFUSION MATRIX FOR PREFERRED MODEL

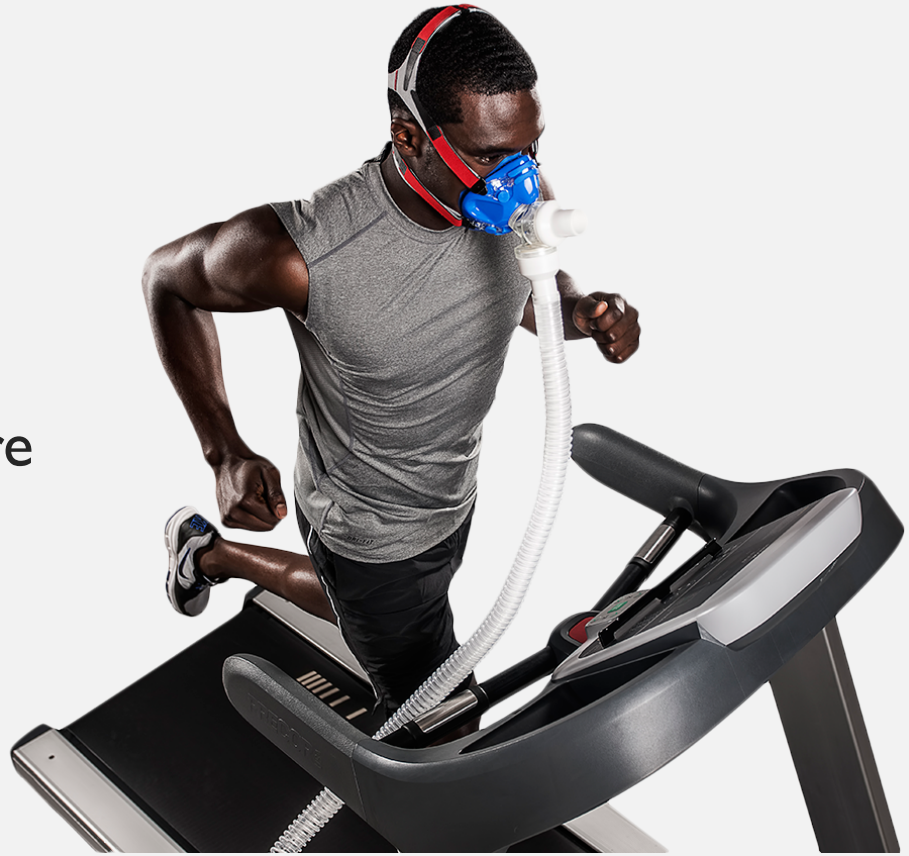


Misclassified as Normal/High: 7%

Misclassified as Low: 23%

CONCLUSION

- The best performing classification model was the logistic model that used oversampling.
- Low absolute fitness could be identified with this model using resting, pre-test measures.
  - Suggests an appropriate fitness test could be determined for different individuals and therefore improve patient care.
  - Also suggests a model could be used to predict fitness and potentially eliminate the need for an actual exercise test.



FUTURE STEPS

- Explore the benefits of adding other pre-test measurements.
  - Body fat percentage improved initial predictions but is not commonly assessed.
- Optimize the hyperparameters of the XG Boost model.
- Study the impact of misclassification on fitness assessments.

FEATURE IMPORTANCE FOR PREFERRED MODEL