Abstract
Fitness is an important health marker that is assessed using an exercise test. However, there are different testing protocols, which can influence the determination of fitness (and patient safety during the test). Using data from FRIEND (a registry of fitness test data), I created different classification models to determine if an individual had a low absolute fitness level – therefore suggesting they perform a fitness test that starts at a low absolute workload. The features included measures typically assessed prior to an exercise test, such as age, height, weight, and disease status. The final model had a high recall indicating a strong ability to accurately identify those individuals needing to perform a lower workload fitness test.

Design
The American Heart Association recommends that fitness be considered a clinical vital sign that is assessed regularly in the same manner as other risk factors (i.e., blood pressure, cholesterol levels, and body weight). However, there is no standard recommendation for testing protocol to determine fitness. While every exercise test protocol involves a maximal effort, the starting workload of protocols differ and can influence the results. The creation of a classification model to determine the appropriate fitness test protocol can ensure accurate assessments of fitness and therefore improve patient risk stratification and patient care.

Data
The data for this analysis comes from the Fitness Registry and Importance of Exercise National Database (FRIEND). This is a database consisting of fitness test results from an exercise test as well as data from pre-test health screenings. There were ~6,500 tests that included all the features needed for this analysis.

Algorithms
Target: The target was classification of a low absolute fitness level, defined as a maximal oxygen consumption (VO_{2max}) <20ml/kg/min (approximately what is required to do 1.5 stages of a common exercise testing protocol).

Feature Engineering: Created some categorical features (BMI category, hypertension status).

Models and Evaluation: Created kNN, logistic, decision tree, extra trees, random forest, naïve Bayes, and XG Boost classification models. Data were split into 60% training, 20% validation, and 20% testing. Models were trained and the primary metric of interest was recall since the primary concern is correctly classifying individuals who have low fitness. Due to imbalanced groups (~10% of individuals were classified as having "low" fitness), I also explored different sampling methods (oversampling, SMOTE procedure, under-sampling), different decision thresholds, and different class weights.

Tools
Data cleaning and exploratory data analysis occurred using Numpy and Pandas. Classification model creation/testing was done with different Python libraries (SKlearn, Mlxtend, XGBoost). Data and model visualizations were created using Matplotlib and Seaborn.

Communication
Slides and visuals were presented and posted on my personal GitHub page.