Physical Activity Classified by Tri-Axial Accelerometer

Dominique Barnes

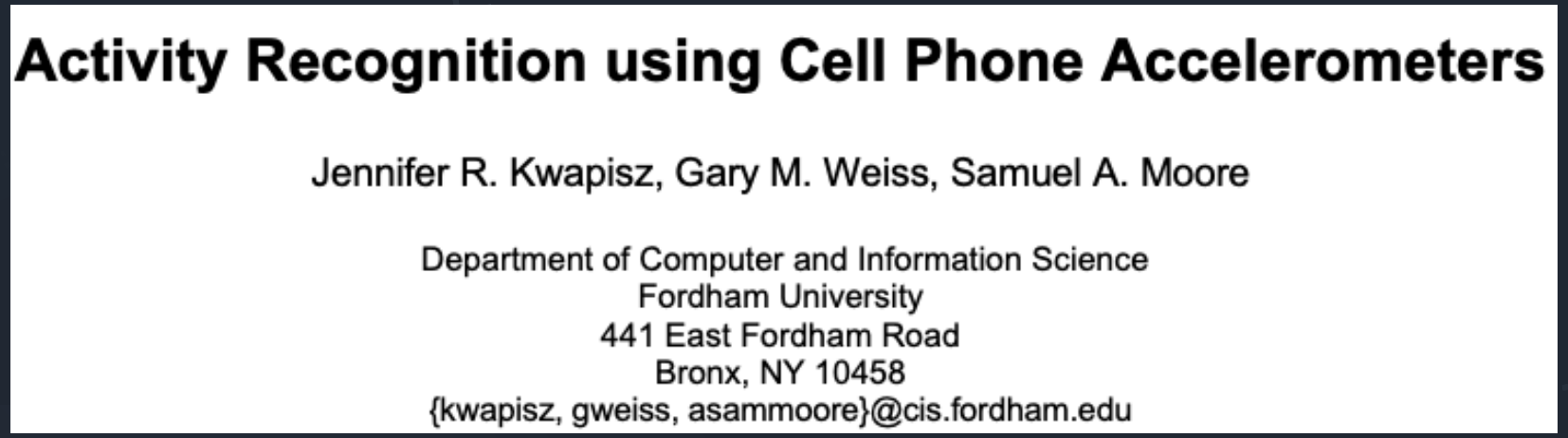
Data Science Institute

December 7th, 2023

https://github.com/dbarnes16/DATA1030-Final.git

Introduction
Dataset

- Kaggle dataset: Activity-Detection-IMU-Sensor
- Wireless Sensor Data Mining (WISDM) Lab
  - Fordham University
- 29 volunteer subjects



Activity Recognition using Cell Phone Accelerometers

Jennifer R. Kwapisz, Gary M. Weiss, Samuel A. Moore

Department of Computer and Information Science
Fordham University
441 East Fordham Road
Bronx, NY 10458
{kwapisz, gweiss, asammoore}@cis.fordham.edu

Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, *12*(2), 74-82.

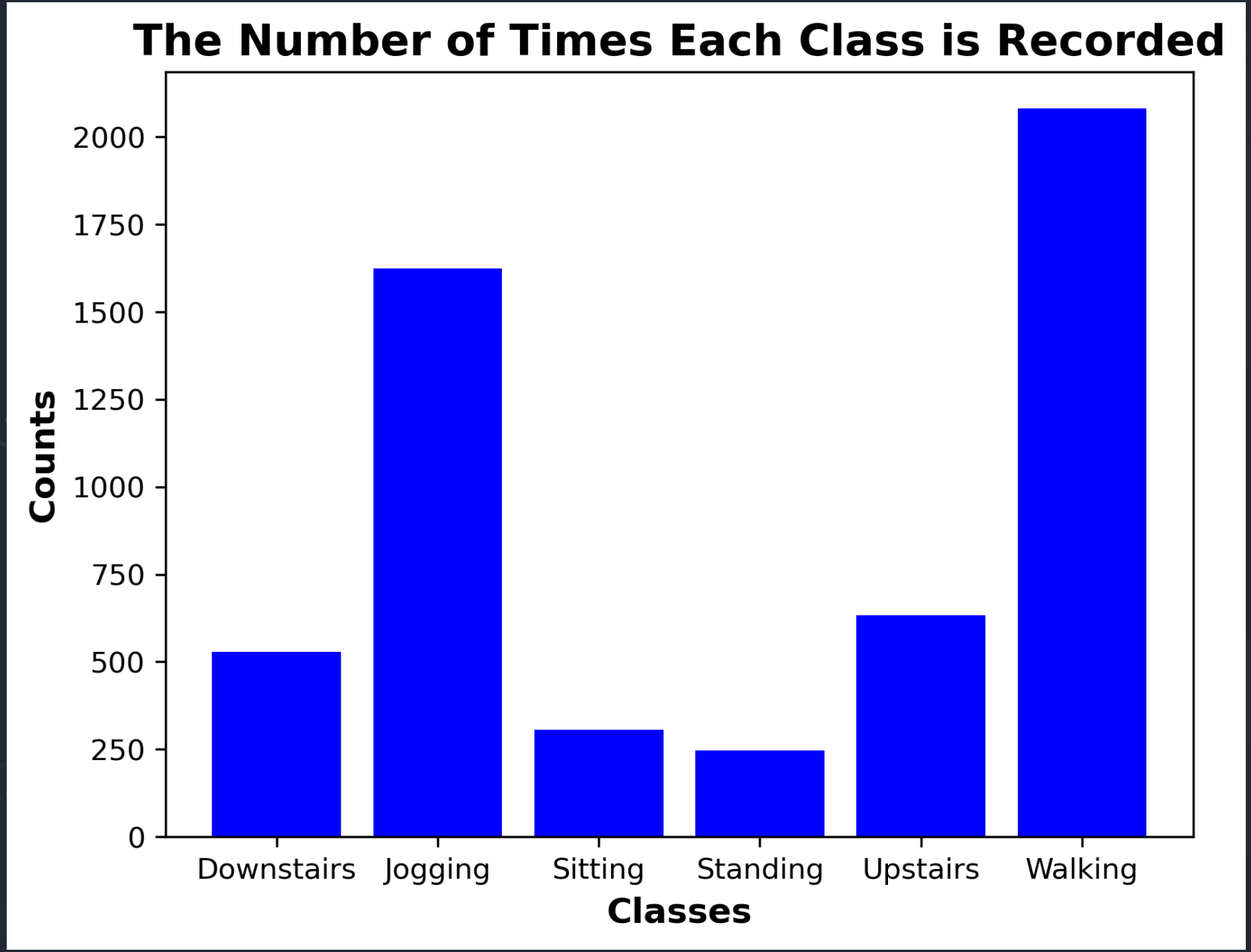
Recap

Goal

- **Classify** the user's motion by the phone-based accelerometers

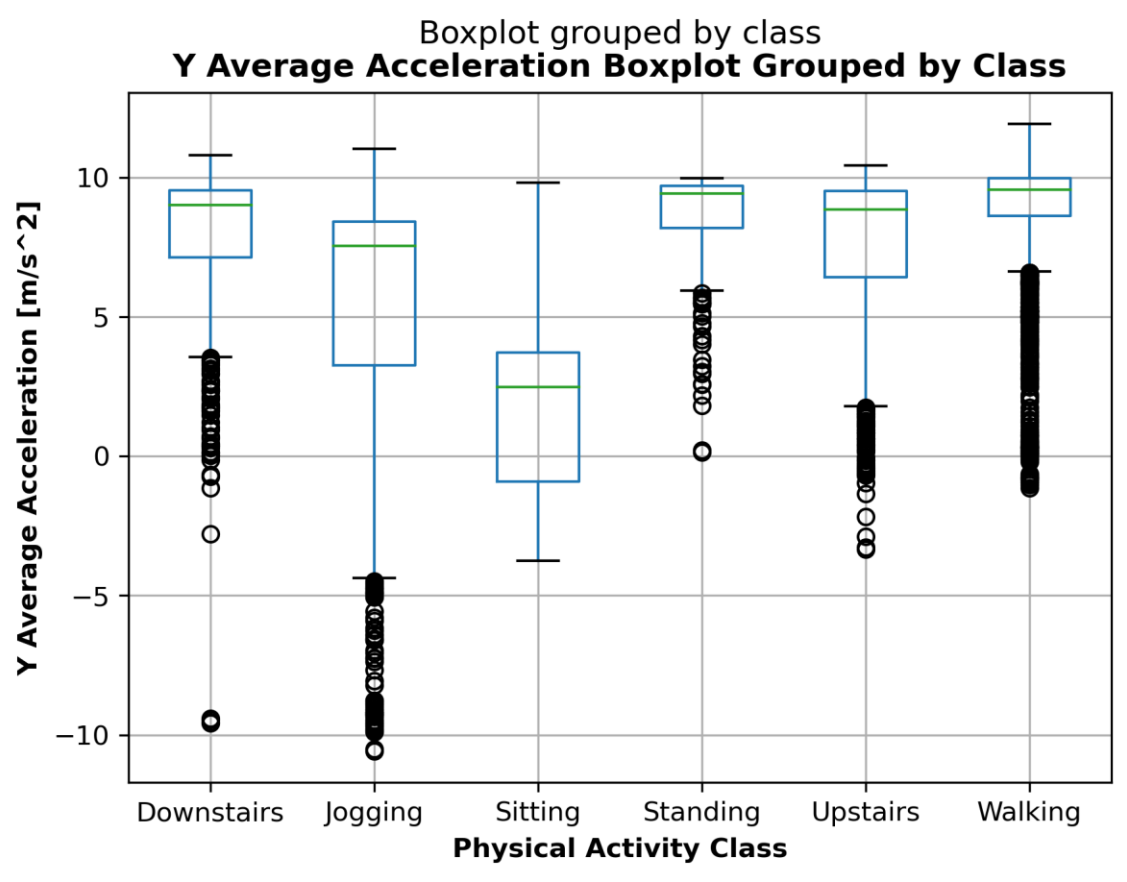
Exploratory Data Analysis

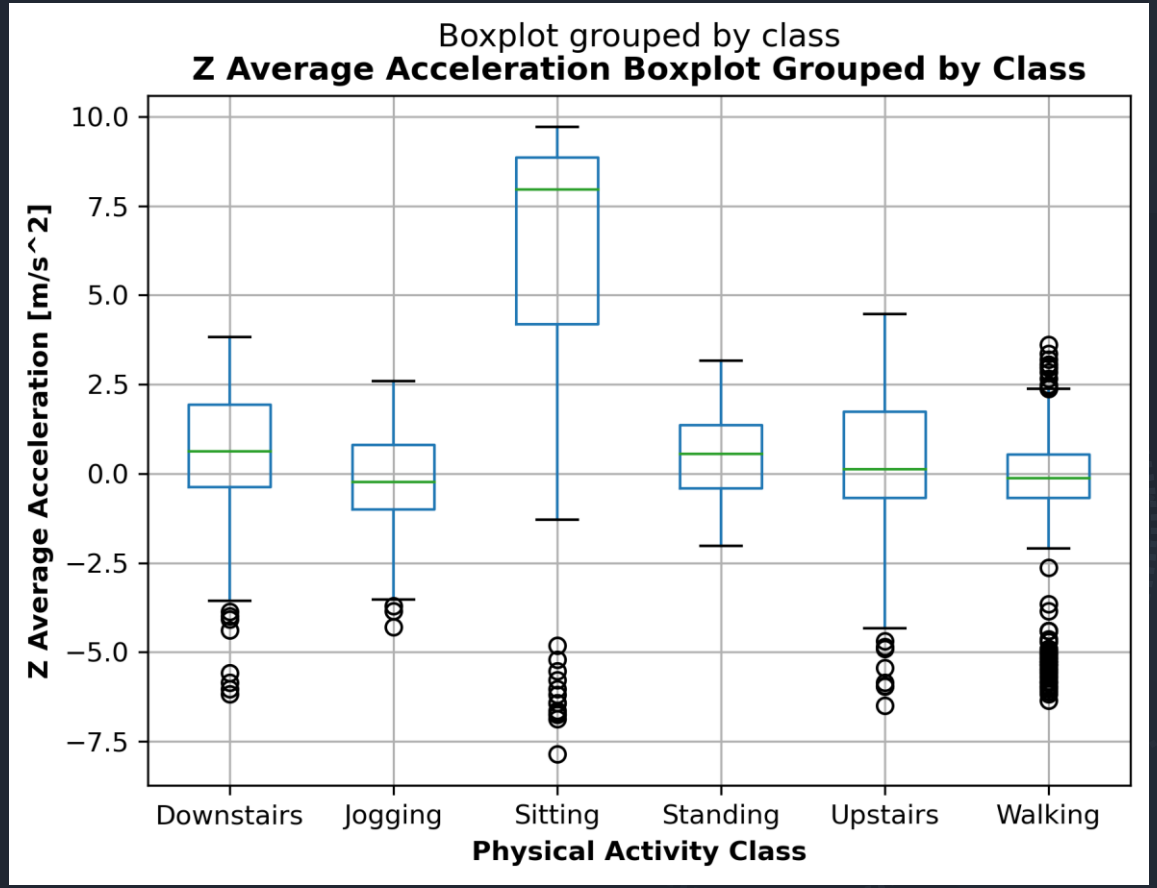
Data Imbalance



Exploratory Data Analysis

Y and Z Average Acceleration





Cross-Validation

Splitting
&
Preprocessing

Is the data IID?

No. It is group-structured

What will the model predict?

Physical activity class on unseen **users**

Group Kfold & Standard Scaler

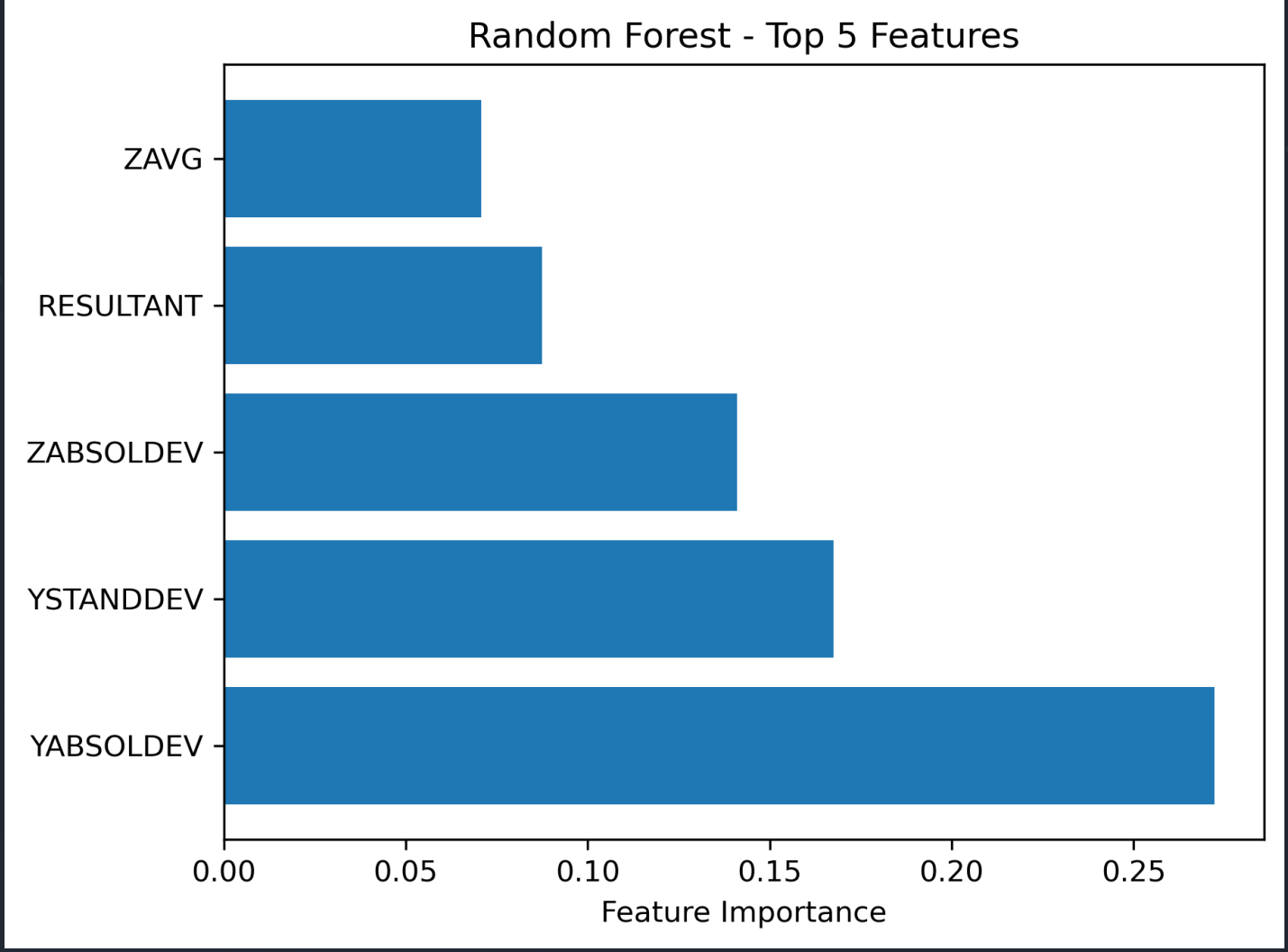
Cross-Validation

Machine Learning Algorithms & Parameters

XGBoost	• N_Estimators: [50,100,500] • Max_Depth: [1,3,10,30]
Logistic Regression L2	• C: [0.001,0.1,1,10,100] • Max_Iter: [1000,5000]
Random Forest	• Max_Depth: [1, 3, 10] • Max_Features: [0.5,0.75,1.0]
SVC	• C: [0.1, 0, 1] • Gamma: [0.001, 0.1, 1, 1000]

Results

Global Feature Importance

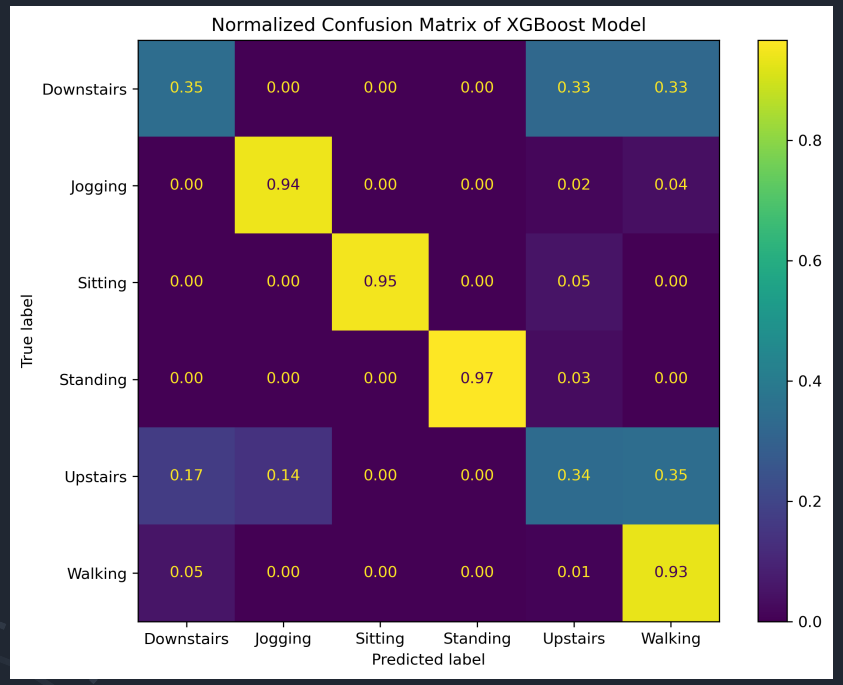


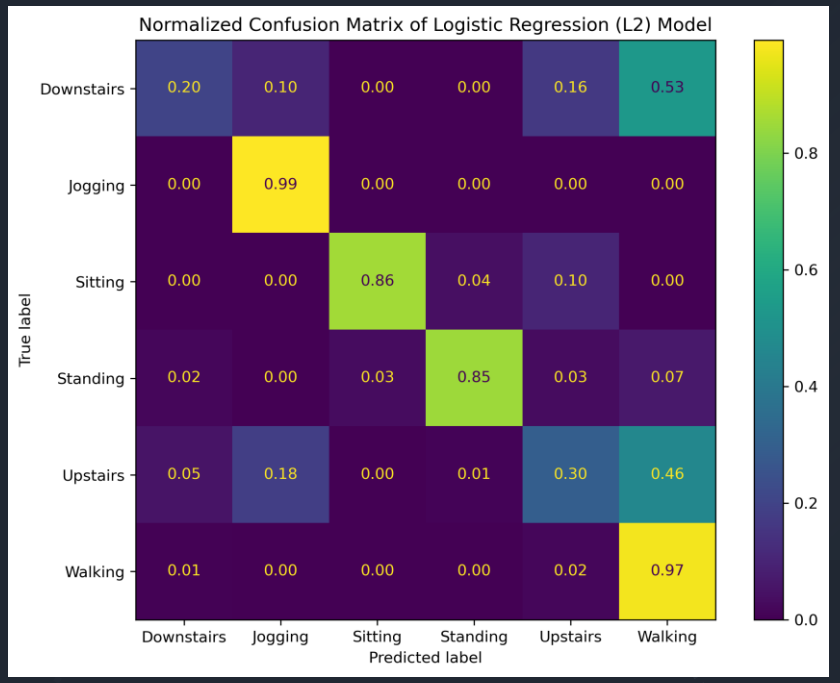
XGBoost &
Logistic Regression (L2)

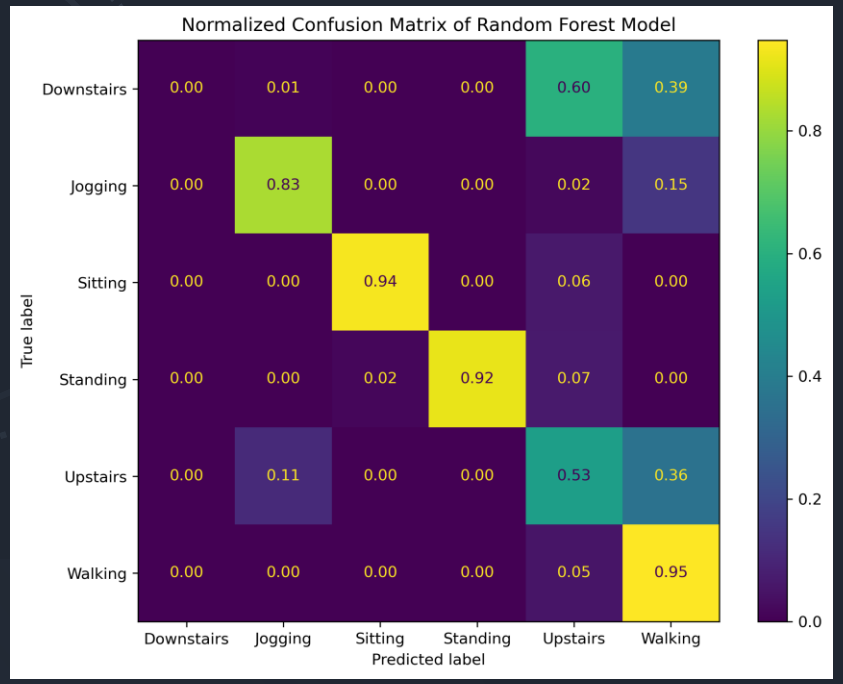
Results

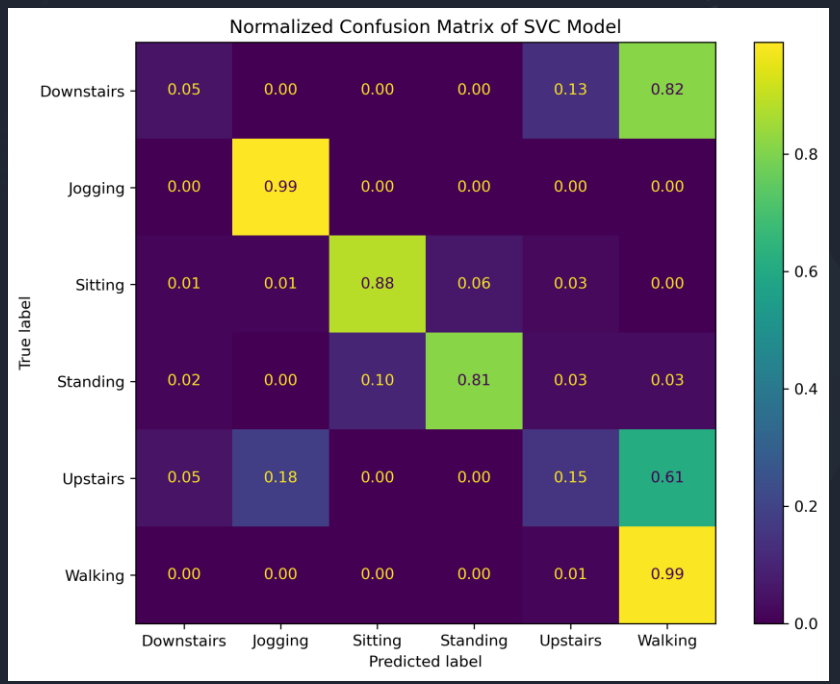
Normalized Confusion Matrices

Random Forest & SVC



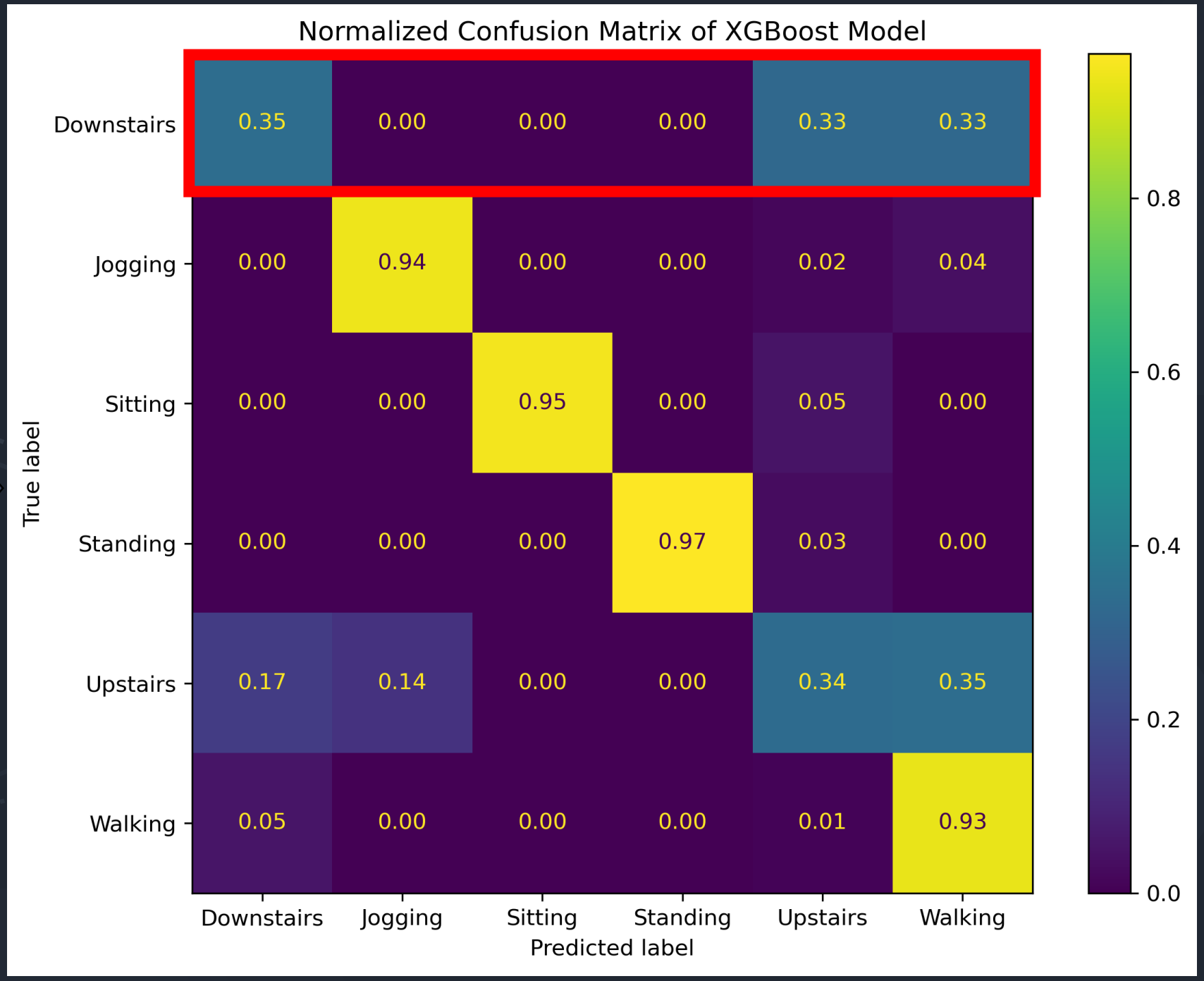






Results

XGBoost Normalized Confusion Matrices



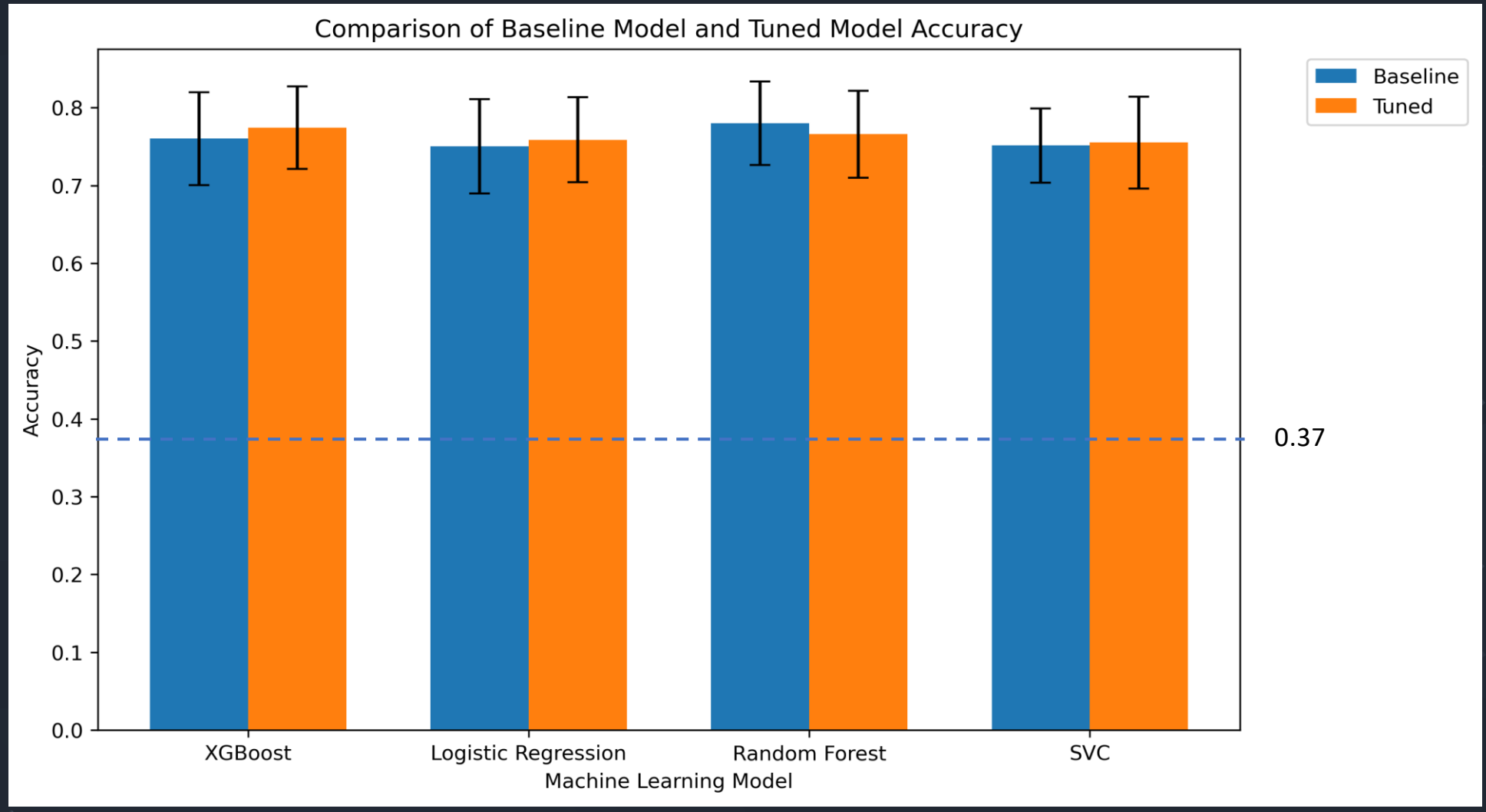
Results

Average Test Scores

	XGBoost Mean +/-StDev	Logistic Regression L2 Mean +/-StDev	Random Forest Mean +/-StDev	SVC Mean +/-StDev
Accuracy	**0.774 +/-0.053**	0.759 +/-0.054	0.762 +/-0.054	0.752 +/-0.057
Precision	0.755 +/-0.050	0.725 +/-0.067	**0.770 +/-0.046**	0.712 +/-0.062
Recall	**0.774 +/-0.053**	0.759 +/-0.054	0.762 +/-0.054	0.752 +/-0.057
F1 Beta Score	**0.751 +/-0.53**	0.718 +/-0.060	0.749 +/-0.052	0.703 +/-0.057

Results

Average Baseline Scores Comparison



Outlook

- Tune more hyperparameters
- Explore other machine-learning models
- Increase the dataset sample size and create a balanced set
- Incorporate muscle activity data

Questions?

Dominique_Barnes@Brown.edu

https://github.com/dbarnes16/DATA1030-Final.git