### **Instructions for the class project**

(More fine instruction will be provided when we learn the corresponding approach in the class).

### **Data Download**

### Download the respective datasets from Canvas.

* Dataset 1 → UG students
  + Users 1 to 12 performing 11 activities each (might have missing files)
* Dataset 2 → Grad students
  + Users 13 to 23 performing 15 activities each (might have missing files)
  + Users 24, 26. and 27 performing 12 activities each
  + User 25 performing 13 activities
* Each activity has 4 sensor data file associated with them - accelerometer, gyroscope, magnetometer, and pressure.
* Samling frequency of accelerometer and gyroscope are 100, its 20 for magnetometer, and 7 for pressure sensor.

### **Data Preparation**

* Ensure you have CSV files for accelerometer and gyroscope data. These files should have timestamped records with readings for the x, y, and z axes.
* The magnetometer and pressure data is not as important for feature extraction.
* Read the data files using Python and clean the data from missing files, etc. Use required pandas library to load and clean the required files.

### **Data Analysis**

* Create subplots using the matplotlib.pyplot library for all four types of sensors.
* For all sensor data, your x-axis should be the time elapsed in milliseconds.
* For accelerometer/gyroscope/magnetometer data, plot the x, y, and z values
* For the pressure data, simply plot the pressure values.
* The result should be four graphs for each of the sensor data files per experiment number.
* These graphs can be used to analyze the activities done at certain points in time.

### **Setup (Windowing time series data)**

* Set a value for a window\_size as an integer between 100-500 by intervals of 100.
* Initialize a loop that creates two windowed dataframes. One for the accelerometer data and the other for the gyroscope data.
  + This loop starts from 0 and ends once it reaches the length of the accelerometer data, moving at intervals determined by your window size.

### **Feature Extraction**

* Perform different calculations to act as your data’s features.
* Inside the initialized loop, perform calculations to obtain values that you will use as your dataset features.
  + For example: acc\_x\_mean = window\_acc\_data["x"].mean()
  + acc\_x\_mean will be a feature representing the mean of the accelerometer x values in the current accelerometer window.
* Create a dataframe using your extracted features.
* You may also try using some feature extraction python libraries.

### **Saving the DataFrame**

* To view your dataframe in the notebook, use the .head(10) to view the first 10 rows.
* To check the dimensions of the dataframe, use .shape
* To save your preprocessed data frame as a csv file, use the to\_csv() function.

**Applying Machine Learning Models**

* Apply various classification algorithms to the feature file (like Logistic regression, Naive Bayes, SVM, Decision Trees, Ensemble methods, Neural networks, etc)
* Perform cross-validation and hyparparameter tuning techniques
* Calculate standard metrics such as accuracy, precision, recall, F1 score, AUC, RoC, etc.

**Up to 10 Bonus points will be awarded for students who try**

more novel algorithms like

* AdaBoost, Gradient Boost, XGBoost,
* LSTM, CNN, DNN, etc and try different activation and loss functions

User Dataset 1 for training and dataset 2 for testing and vice versa