Final Project: Proposal

#### Gustavo Felhberg, Jorge Marcano, Muhammad

### **1. List 3 questions that you intend to answer (1 point)**

* Which sensors provide the most useful information in order to classify Falls? Are there any sensors that can be removed?
* Is it possible to classify Falls vs Non-Falls without false positives using the given dataset?
* (UHHHH...not sure what else?)

### **2. List all the datasets you intend to use (1 point)**

So far the plan is to use the dataset provided by the person that proposed the project, it can be found here :

<https://researchdata.sfu.ca/islandora/object/islandora%3A9085>

It contains data for 10 subjects, each one having data for “Daily activities”, “Near Falls” and “Falls”, and each of those containing data for each of the trials performed by the subjects, the size of the dataset is approximately 600MB and has 64 feature columns.

### **3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)**

* Data Collection: **The data has been provided to us by the person that proposed the project**
* Data Exploration: Do you need to conduct EDA in order to understand the data? : **We want to do an statistical analysis on the data in order to understand it, including most likely some data visualization steps.**
* Data Cleaning: Do you need to clean data? How to clean them? : **The data, specifically the magnetic field columns (one for each part of the body where the sensors were placed) appear to be noisy, we'll need to find a way to deal with this as well as any missing data or outlying values.**
* Data Integration: Do you need to integrate data from multiple sources? : **We need to integrate the data from the different subjects, trials and actions (Falls,Near Falls and Daily Activities) but these all share the same structures, they are just in different files and directories.**
* Data Analysis: What analysis do you intend to do? (e.g., SQL, Statistics, Deep Learning) How to evaluate your analysis results? (e.g., evaluation metrics, confidence intervals, benchmark) : **We want to do an initial statistical analysis on the data in order to explore it and then we want to create machine learning models in order to predict Falls vs Non-Falls. Ideally (and following the project outline) we want to have no false positives while still maintaining an acceptable true possitive rate (>= 90%)**
* Data Product: What product do you want to build? (e.g., visualizations, an interactive web app, a jupyter notebook) : **As of right now the main objective of the project is to create a Machine Learning model to predict Falls. We'll most likely present a Jupyter Notebook with this along with the statistical analysis done.**

### **4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)**

If we do manage to create a Machine Learning model that can classify Falls vs Non-Falls, this could potentially be extended to predict falls before the fall impact happens, with these, falls can either be prevented, or if they do happen, actions can be taken in a timely manner.