**Human Activity Recognition with Smartphones:**

**PROJECT PROPOSAL**

**1. What is the problem you want to solve?**

Data from the waist-mounted smartphone with embedded inertial sensors is collected from 30 study participants. The objective of this study is to classify these participants’ activities of daily living (ADL) into one of the 6 activities performed.

**2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?**

The Classification of ADL Activity have real world applications in health care and fitness monitoring. Physical activity has positive effects on all body functions and studies proved that the risk of cardiovascular diseases is up to 50% lower on physical active people.

In health care field, long term analysis of human activity could be helpful in early detection of diseases or even to encourage people to improve their activity level. It could also be useful for physiotherapy, helping to understand if the recommended exercises are being correctly performed or even to assist those with cognitive disorders.

**3. What data are you going to use for this? How will you acquire this data?**

Data was found from the “Human Activity Recognition with smartphones” dataset provided by Kaggle’s website. <https://www.kaggle.com/uciml/human-activity-recognition-with-smartphones>

**4. In brief, outline your approach to solving this problem (knowing that this might change later).**

We begin with Scrubbing or cleaning the data. This includes data imputation of missing or invalid data and fixing column names.

Exploring the data with exploratory data analysis will follow right after and allow further insight into what our dataset contains. Looking for any outliers or weird data. Understanding the relationship each explanatory variable has with the response variable resides here and we can do this with a correlation matrix. The creation or removing of features through the use of feature engineering is a possibility. The use of various graphs plays a significant role here as well because it will give us a visual representation of how the variables interact with one another. We will get to see whether some variables have a linear or non-linear relationship. Taking the time to examine and understand our dataset will then give us suggestions on what type of predictive model to use.

Interpreting the data is final. The Approaches used are Linear Discriminant Analysis (LDA); Multinomial Logistic Regression; Support Vector Machines (SVM) and kNN. The comparison is performed based on the confusion matrix of those classification results on the test data using the different training classifiers, presented by confusion matrices.SVM with linear kernel approach is the best fit classifier to our dataset. It gives the best performance for all the accuracies and error rates. SVM with the linear kernel is a flexible approach which is capable to reduce overfitting.

**5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

A python notebook containing source code, visualization and documentation; and a slide-deck of the project will be submitted for review.