**Human Activity Recognition with Smartphones:**

**PROJECT PROPOSAL**

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

The Human Activity Recognition database is built from the recordings of 30 study participants performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. ***The objective is to classify activities into one of the six 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 idea is that once the subject’s activity is recognized and known, an intelligent computer system can then offer assistance. These days, in addition to Smartphones, we are also using Smart-Watches like Fitbit or Apple-Watch, which help us to track our health. They monitor our each activity throughout the day check how many calories we have burnt. How many hours have we slept.

Also by automatically monitoring human activities, home-based rehabilitation can be provided for people who are suffering for any specific diseases.

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

The 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).**

Scrubbing or cleaning the data is the first step. 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 the suggestions on what type of predictive model to use.

Interpreting the data is final. The Approaches used are: Linear Discriminant Analyses (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 linear kernel being 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.**

The deliverables would be my python notebook, which includes the source code, visualizations, and some documentation.