**Human Activity Recognition Using Smartphones Data Set**

***Introduction:***

The raw data that was taken up for analysis was produced through experiments carried out by a group of 30 volunteers, in the age group of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Signals from sensors on the smartphone (accelerometer and gyroscope) were captured, which provided the raw data. This was further processed, and the processed data was used for this analysis.

***Objective:***

The goal of the initial exercise was to create a ‘tidy data set’, as follows:

1. Merge training and test data files
2. Extract columns that pertained to mean and standard deviation measurements, and drop the remaining columns
3. Include the following columns:
   1. ‘Subject-ID’, indicating the identifier of the volunteer who performed the experiments
   2. ‘Activity-ID’, indicating the type of activity performed
   3. ‘Activity Description’, corresponding to the respective Activity-ID.
4. After the tidy data is created as above, summarize the data such that each mean and standard deviation measurement is averaged and grouped by the Subject ID and Activity ID,

***Data Processing Steps:***

1. Relevant data files were downloaded from the source destination and saved to local disk
2. The test data file, “X\_test.txt” was opened from R, into a data set
3. The subject data file, “subject\_test.txt” was opened into an R data set, and appended as a column, to the test data set in step 2
4. The activity data file, “y\_test.txt” was opened into an R data set and appended as a column to the test data in step 2
5. Steps 2-5 were performed on the training data as well, and the following files were used to prepare the training data set: “X\_train.txt”, “subject\_train.txt”, “y\_train.txt”.
6. The training and test data sets created above were merged and proper column names were added
7. The unwanted columns were then stripped out, leaving only the columns pertaining to mean and standard deviation (indicated by column names ending in “mean()” and “std()”.
8. The data was then grouped by ‘Subject ID’ and ‘Activity ID’, and the measurements reflected in each data column were averaged over each occurrence of Subject ID and Activity ID.
9. Another column ‘Activity Name’ was added to the data set to provide additional description to the Activity Code.
10. The data was sorted by Subject ID, and by Activity ID, for better readability.
11. The dataset was finally written to a file on the disk, and is provided as part of this submission.