**Human Activity Recognition Using Smartphones Data Set**

**CODE BOOK**

1. Data set download link:

<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip> (90MB)

1. Project Information web link:

<http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

1. Files in the data set with description: After you uncompress it a folder UCI HAR Dataset is created and this has the following files and subfolders:

|  |  |  |
| --- | --- | --- |
| activity\_labels.txt | file | Six activities along with IDs are defined where the first  Column is the activity ID and the second column is the activity  1 WALKING  2 WALKING\_UPSTAIRS  3 WALKING\_DOWNSTAIRS  4 SITTING  5 STANDING  6 LAYING |
| features\_info.txt | file | Contains an excellent description of the various features listed in the file features.txt |
| features.txt | file | 561 features are listed here that are created after processing the raw sensor data in the directory test and train. Each row in this file contains the feature ID and the feature label (partially listed below):  1 tBodyAcc-mean()-X  2 tBodyAcc-mean()-Y  3 tBodyAcc-mean()-Z  4 tBodyAcc-std()-X  5 tBodyAcc-std()-Y  6 tBodyAcc-std()-Z  ….  …. |
| README.txt | file | Contains information pertaining to all the files |
| test | folder | Contains the following files and folders:  Inertial Signals – Folder that contains 9 files each file has 128 columns of numerals and 2947 rows  subject\_test.txt – contains 2947 rows of integers in the range [1-30]  X\_test.txt – contains 2947 rows and 128 columns  y\_test.txt – contains one column of numbers in the range [1-6] each corresponding to the activity ID. |
| train | folder | Contains the similar files as in test by the number of observations is 7352 |

1. From all the above data items a tidy data set was created (tidydataoutput.txt). It contains 180 observations and 88 columns. The first three columns are observation number, activity, and subject no. A sample is shown below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| obsno | activity | subjectno | tbodyaccmeanx | tbodyaccmeany | tbodyaccmeanz |
| 1 | LAYING | 1 | 0.221598244 | -0.040513953 | -0.113203554 |
| 2 | SITTING | 1 | 0.261237565 | -0.001308288 | -0.104544182 |
| 3 | STANDING | 1 | 0.278917629 | -0.01613759 | -0.110601818 |
| 4 | WALKING | 1 | 0.277330759 | -0.017383819 | -0.111148104 |
| 5 | WALKING\_DOWNSTAIRS | 1 | 0.28918832 | -0.009918505 | -0.107566191 |
| 6 | WALKING\_UPSTAIRS | 1 | 0.25546169 | -0.023953149 | -0.097302002 |
| 7 | LAYING | 2 | 0.281373404 | -0.01815874 | -0.10724561 |
| 8 | SITTING | 2 | 0.277087352 | -0.015687994 | -0.109218272 |

Each observation is a aggregate that is performed for each activity for each subject on partial subset of the 561 features. These features were selected based on partial matching, that is looking for feature labels that contains the word “mean” or “std” (standard deviation).