

# README.md

Solenn Toupin

26/02/2022

## **Coursera project: Getting and Cleaning Data Course Project**

### **Step 1: Merges the training and the test sets to create one data set**

The datasets are imported from <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip> website.

- X\_train and X\_test are combined with rbind function to X\_merged dataset composed of 561 variables and 10299 observations.
- y\_train and y\_test are combined with rbind function to y\_merged dataset composed of 1 variable and 10299 observations.

### **Step 2: Appropriately labels the data set with descriptive variable names.**

The feature.txt is used to name the columns of the X\_merged dataset.

### **Step 3: Extracts only the measurements on the mean and standard deviation for each measurement**

- The 86 variables names containing “mean” and “std” are extracted to tidy\_dataset.

### **Step 4: Uses descriptive activity names to name the activities in the data set**

- Merge the activity\_label to y\_merged dataset
- The activity label column is then added to the tidy\_dataset

### **Step 5: From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject**

- Subject\_test and Subject\_train datasets are read and merged to Subject\_merged dataset with rbind() function
- Tidy\_dataset and Subject\_merged columns are combined with cbind() function
- The mean of each variable by subject and by activity is computed and stored in Tidy\_dataset\_2
- Tidy\_dataset\_2 is saved as .txt file