

# CodeBook

Sankar

02/05/2020

## CodeBook Description of the steps performed

The ‘run\_Analysis.R’ file is coded in such a way that it gets the data of test and training data sets and converts it into a tidy dataset following the below steps.

### Step 1: Assigning the text files to Different variables

The text files *activity\_labels.txt*, *features.txt*, *subject\_train.txt*, *X\_train.txt*, *y\_train.txt*, *subject\_test.txt*, *X\_test.txt*, *y\_test.txt* are read into the r file into a table format.

\* **activities** <- ‘**activity\_labels.txt**’ List of different activities performed by the subjects and the codes representing those activities. Size = 6 rows and 2 columns

\* **features** <- ‘**features.txt**’ List of different measurments taken from the accelerometer and gyroscope 3-axial raw signals. Size = 561 rows and 2 columns

\* **subject\_train** <- ‘**subject\_train.txt**’ List of different subject members taken as training sets for the dataset. Size = 7352 rows and 1 column

\* **subject\_test** <- ‘**subject\_test.txt**’ List of different subject members taken as testing sets for the dataset. Size = 2947 rows and 1 column

\* **X\_train** <- ‘**X\_train.txt**’ Dataset containing the different observations for different measurements of the subjects in the training dataset. Size = 7352 rows and 561 columns

\* **X\_test** <- ‘**X\_test.txt**’ Dataset containing the different observations for different measurements of the subjects in the testing dataset. Size = 2947 rows and 561 columns

\* **y\_train** <- ‘**y\_train.txt**’ List containing the codes of different types of activities performed by the subjects in the training dataset. Size = 7352 rows and 1 column

\* **y\_test** <- ‘**y\_test.txt**’ List containing the codes of different types of activities performed by the subjects in the testing dataset. Size = 2947 rows and 1 column

## Merging Operation

This step is divided into four steps:

**Step 1:** Merging the X training and testing datasets into a single dataset. size = 10299 rows, 561 columns

**Step 2:** Merging the y training and testing datasets into a single dataset. size = 10299 rows, 1 column

**Step 3:** Merging the subject training and testing datasets into a single dataset. size = 10299 rows, 1 column

The first 3 steps are performed using rbind (i.e., row bind operation).

**Step 4:** Merging all the datasets formed from the above 3 steps into a single 'Merged\_Data' Dataset using cbind (i.e., column bind operation). size = 10299 rows, 563 columns

## Creation of Tidy Dataset

The tidy dataset is then created from the *Merged\_Data* dataset by extracting only the subject, activity code and measurements on the mean and standard deviation for each measurement and stored in a separate dataset called the **Tidy\_Data**.

Size = 10299 rows, 88 columns

## Renaming by descriptive activity names

The codes representing the activities performed by the subject in the tidy data set is then entirely changed into the names of the activities to describe the activity literally than depicting those activities by a code.

## Renaming the column Names of the Tidy Dataset

The Column Names of the *Tidy\_Data* Dataset are changed:

- 'subject' column renamed to *SubjectID*
- 'Activity.code' column renamed to *NameOfActivity*
- All characters starting with 't' renamed to *Time*
- All characters starting with 'f' renamed to *Frequency*
- All characters that contains 'Acc' renamed to *Acceleration*
- All characters that contains 'Mag' renamed to *Magnitude*
- All characters that contains 'Gyro' renamed to *Gyroscope*
- All characters that contains 'BodyBody' renamed to *Body*
- 'gravity' column renamed to *Gravity*
- 'angle' column renamed to *Angle*

## Creation of Final DataSet

This dataset is created from the *Tidy\_Data* dataset summarizing the tidy dataset by the averages of all the measurements or variables for each subject and for each activity.

Size = 180 rows, 88 columns

This dataset is then exported into a text file named **Final\_Dataset.txt**.