## Case Study on Human Activity Recognition using Random Forest

Human Activity Recognition or HAR for short is the problem of predicting what a person is doing based on a trace of their movement using sensors.

Movements are often normal indoor activities such as standing, sitting, jumping, and going up stairs. Sensors are often located on the subject such as a smartphone or vest and often record accelerometer data in three dimensions (x, y, z).

The idea is that once the subject’s activity is recognized and known, an intelligent computer system can then offer assistance.

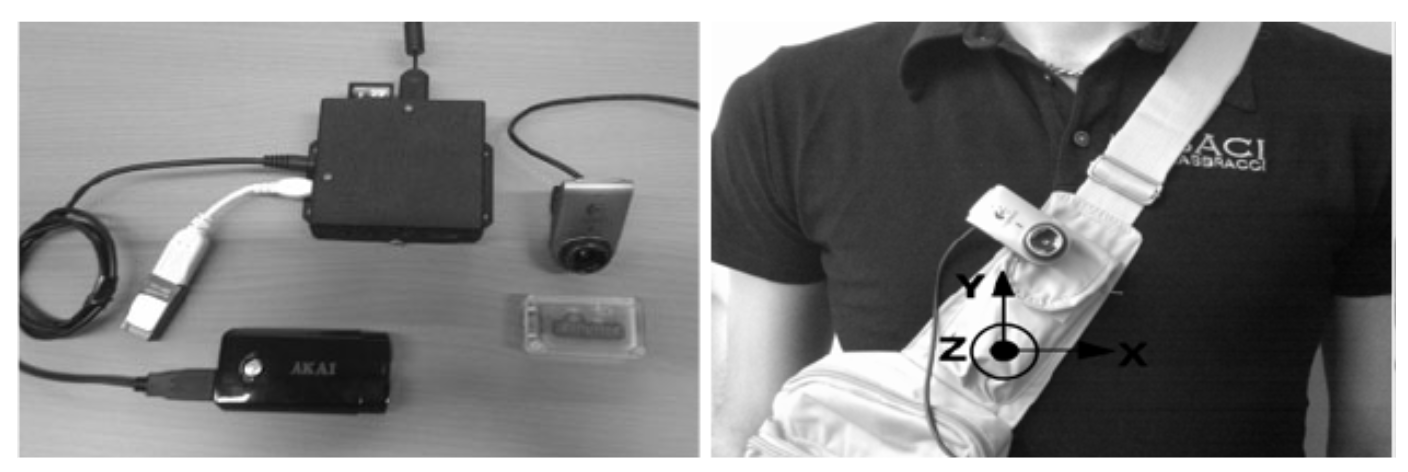
It is a challenging problem because there is no clear analytical way to relate the sensor data to specific actions in a general way. It is technically challenging because of the large volume of sensor data collected (e.g. tens or hundreds of observations per second) and the classical use of hand crafted features and heuristics from this data in developing predictive models

**Problem Description**

The dataset “Activity Recognition from Single Chest-Mounted Accelerometer Data Set” was collected and made available by Casale, Pujol et al. from the University of Barcelona in Spain. It is freely available from the UCI Machine Learning repository:

<https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer>

The dataset is comprised of un-calibrated accelerometer data from 15 different subjects, each performing 7 activities. Each subject wore a custom-developed chest-mounted accelerometer and data was collected at 52 Hz(52 observations per second).



**Data Description**

Un-calibrated Accelerometer Data are collected from 15 participants performing 7 activities. The dataset provides challenges for identification and authentication of people using motion patterns.

**Data Set Information:**

--- The dataset collects data from a wearable accelerometer mounted on the chest

--- Sampling frequency of the accelerometer: 52 Hz

--- Accelerometer Data are Un-calibrated

--- Number of Participants: 15

--- Number of Activities: 7

--- Data Format: CSV

**Attribute Information:**

--- Data are separated by participant

--- Each file contains the following information

---- Sequential number, x acceleration, y acceleration, z acceleration, label

--- Labels are codified by numbers

--- 1: Working at Computer

--- 2: Standing Up, Walking and Going updown stairs

--- 3: Standing

--- 4: Walking

--- 5: Going UpDown Stairs

--- 6: Walking and Talking with Someone

--- 7: Talking while Standing

## Read the data

## Import libraries and tools

## Load data set by creating a function

## Plot a subject

1. plot the x, y, z acceleration and activities for a single subject
2. create a plot for each column
3. plot activities for a single subject
4. **Plot Total Activity Durations**
5. return a list of dict, where each dict has one sequence per activity
6. calculate total duration in sec for each activity per subject and plot
7. calculate the lengths for each activity for each subject
8. plot durations
9. calculate\_durations(grouped, activities)
10. plot the x, y, z acceleration for each subject

# create a plot for each subject

# plot a histogram of x data

1. **Defining the variables**
2. **Train Data split**
3. evaluate the model by splitting into train and test sets
4. **Modelling**
5. calculate cross-validated AUC
6. use the model to make predictions with the test data
7. generate evaluation metrics-
8. Print out the confusion matrix
9. Print out the classification report, and check the f1 score
10. Find out the mean cross validation score/accuracy of the fitted model, use 5 cv steps
11. **Tree Visualisation**
12. Extract single tree
13. Export as dot file
14. Convert to png using system command (requires Graphviz)
15. Display in jupyter notebook