1. As the project data set, we will work on the WISDM (Wireless Sensor Data Mining) dataset, which is available from the UCI machine learning repository of the same name.
2. We considered the WISDM (Wireless Sensor Data Mining) dataset as the dataset for this project
3. The dataset we will be working on is the WISDM (Wireless Sensor Data Mining) dataset, which is available from the UCI machine learning repository of the same name.

The dataset includes data collected from 51 subjects, each of whom were asked to perform 18 tasks for 3 minutes each. Each subject had a smart-watch placed on his/her dominant hand and a smartphone in their pocket.

The dataset contains the data collected from the accelerometer and gyroscope sensors of a smartphone and smartwatch

WISDM (Wireless Sensor Data Mining)

Collected data from the accelerometer and gyroscope sensors of a smartphone and smartwatch as 51 subjects performed 18 diverse activities of daily living.

Each activity was performed for 3 minutes, so that each subject contributed 54 minutes of data.

These activities include basic ambulation-related activities (e.g., walking, jogging, climbing stairs), hand-based activities of daily living (e.g., brushing teeth, folding clothes), and various eating activities (eating pasta, easting chips).

Non-hand-oriented activities, General Hand-oriented activities, eating Hand-oriented activities

These activities include basic Non-hand-oriented activities, hand-based daily life activities and various eating activities.

The data set contains the low-level time-series sensor data from the phone’s accelerometer, phone’s gyroscope, watches’ accelerometer, and watches’ gyroscope. All of the time-series data is tagged not only with the activity that was being performed, but with a subject identifier, which means that the data be used for building and evaluating biometrics models, as well activity recognition models.

Researchers in the WISDM Lab subsequently used a sliding window approach to transform the time-series data into labeled examples.

The data set is available from the UCI Machine Learning Repository as the “WISDM Smartphone and Smartwatch Activty and Biometrics Dataset.”

The raw time-series sensor data is recorded by the accelerometer and gyroscope on both the phone and watch at a rate of 20Hz.

* *What are the ‘modalities’ that apply to the data? (images, video, speech, text, tabular, categorical, numerical, time series, experimental measurements, etc.)*

The dataset we will be working on is the WISDM (Wireless Sensor Data Mining) dataset, which is available from the UCI machine learning repository of the same name.

The dataset includes time-series data collected from the accelerometer and gyroscope sensors of a smartwatch and smartphone, which were on the hand and inside the pocket of a subject, respectively.

To collect this dataset, 51 subjects were asked to perform 18 different activities in 3 minutes. These activities include basic Non-hand-oriented activities, hand-based daily life activities, and various eating activities.

* *What does an input look like? (show an example if possible, like an image, or a sound wave, or some features, or at least try to describe)*

Each line of the time-series sensor file is considered as our input.

The format of this line is identical across both types of sensors and it would be as follows:

*Subject-id, Activity Code, Timestamp, x, y, z;*

* *Subject-id;* is a symbolic numeric identifier that represents the subject and is a single value between 1600 and 1650. (Each number belongs to one of the 51 subjects)
* *Activity Code;* is a symbolic single letter which identifies one of the 18 activities represented in dataset. (A-S except N value)
* *Timestamp;* is a Linux time displayed by an integer.
* *x, y, z*; are sensor values for x, y, z axes respectively.

1600, A, 252207666810782, -0.36476135, 8.793503, 1.0550842;

*Subject-id*

*x value*

*Activity Code*

*Timestamp*

*z value*

*y value*

*Ex:*

* *For an example input, what does a desired output look like? (show an example if possible, or at least try to describe)*

Given that each activity was performed for 3 minutes, for each subject, there are 54 minutes of data in the dataset.

* *How many training and testing samples will there be?*

The data set contains a total of 15,630,426 lines/sensor readings.

Each smartphone and smartwatch has its own accelerometer and gyroscope sensors.

Thus, there are four sensors and the number of lines of data per each sensor is as follows:

* Smartphone accelerometer: 4,804,403
* Smartphone gyroscope: 3,608,635
* Smartwatch accelerometer: 3,777,046
* Smartwatch gyroscope: 3,440,342

In addition to the raw time-series data, the dataset includes labeled examples.

The transformation process for the generation of labeled examples has been carried out using a sliding window approach.

The raw time-series data for each sensor has been divided into 10-second non-overlapping segments, and then high-level features have been generated in each section.

(These time-series sensor data is recorded at a rate of 20Hz.)

The dataset includes data collected from 51 subjects, each of whom were asked to perform 18 tasks for 3 minutes each. Each subject had a smart-watch placed on his/her dominant hand and a smartphone in their pocket.

In addition to the raw time-series data, the dataset includes labeled examples.

The transformation process for the generation of labeled examples has been carried out using a sliding window approach.

The raw time-series data have been transformed into labeled examples using a sliding window approach.

The transformation process that was used to generate the labeled examples is described in Section 4.1,

The raw time-series data for each sensor has been divided into 10-second non-overlapping segments, and then high-level features have been generated in each section.