

# Introduction Into Data Science

---

Project: Activity Recognition

**Mirco Pyrtek, Ralf Bleymehl**

## Content

1. Motivation.....	3
2. Overview of the Data.....	4
3. Implementation of the model.....	6
4. Conclusion .....	7

# 1. Motivation

The increasing number of comfort enhancing devices and services impact the health and fitness of the society. Due to speech-controlled homes and orders executed with one click, people don't need to put any effort in activities, which used to be connected to at least a small amount of walking or other body related work.

Therefore, we looked at some data, which determines the activity the subject was doing at the point of collecting. Vital signs, as well as movement data was used to predict one out of twelve different activities like jogging, waist-bends and sitting.

With the above-mentioned problem in mind, we propose a model that gathers vital signs and movement patterns and predicts the executed activity. Activities are mapped to some score and the total number of points represents the fitness per day. This kind of service might support several business ideas, e.g. insurance packages, smart home alerts and general medical support.

In the insurance sector, it is possible to skip prevention courses and replace them with voluntary activity measuring. The calculated score represents the customers general health level and may be used as a classification in an insurance plan.

Furthermore, a smart home could offer an alert whenever it detects a very inactive way of life, i.e. it constantly measures the activities, calculates the scores and give hints like „You were sitting all day, you should go for a walk!“ or „You were very active today, good job!“.

As a last point, medical doctors might use this data to quickly check on the fitness level of their patients.

The services mentioned above should all be offered on a voluntary basis.

In the following sections, we provide an overview of the data, how the model looks like and which models were used.

## 2. Overview of the Data

Given datasets were collected as part of a master thesis<sup>1,2</sup>. For the training of this model, the datasets that were used were collected by ten probands. Their vital signs and movements were measured during twelve different physical activities using three sensor devices on the chest, the right wrist and the left ankle. The datasets consist of 24 columns each. Table 1 shows the collected data and the meaning of the columns.

Column	Data	Unit
1	acceleration from the chest sensor (X axis)	$\frac{m}{s^2}$
2	acceleration from the chest sensor (Y axis)	$\frac{m}{s^2}$
3	acceleration from the chest sensor (Z axis)	$\frac{m}{s^2}$
4	electrocardiogram signal (lead 1)	<i>mV</i>
5	electrocardiogram signal (lead 2)	<i>mV</i>
6	acceleration from the left-ankle sensor (X axis)	$\frac{m}{s^2}$
7	acceleration from the left-ankle sensor (Y axis)	$\frac{m}{s^2}$
8	acceleration from the left-ankle sensor (Z axis)	$\frac{m}{s^2}$
9	gyro from the left-ankle sensor (X axis)	$\frac{deg}{s}$
10	gyro from the left-ankle sensor (Y axis)	$\frac{deg}{s}$
11	gyro from the left-ankle sensor (Z axis)	$\frac{deg}{s}$
12	magnetometer from the left-ankle sensor (X axis)	<i>local</i>
13	magnetometer from the left-ankle sensor (Y axis)	<i>local</i>
14	magnetometer from the left-ankle sensor (Z axis)	<i>local</i>
15	acceleration from the right-lower-arm sensor (X axis)	$\frac{m}{s^2}$
16	acceleration from the right-lower-arm sensor (Y axis)	$\frac{m}{s^2}$

<sup>1</sup> Banos, O., Garcia, R., Holgado, J. A., Damas, M., Pomares, H., Rojas, I., Saez, A., Villalonga, C. mHealthDroid: a novel framework for agile development of mobile health applications. Proceedings of the 6th International Work-conference on Ambient Assisted Living an Active Ageing (IWAAL 2014), Belfast, Northern Ireland, December 2-5, (2014).

<sup>2</sup> Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado, J. A., Lee, S., Pomares, H., Rojas, I. Design, implementation and validation of a novel open framework for agile development of mobile health applications. BioMedical Engineering OnLine, vol. 14, no. S2:S6, pp. 1-20 (2015).

17	acceleration from the right-lower-arm sensor (Z axis)	$\frac{m}{s^2}$
18	gyro from the right-lower-arm sensor (X axis)	$\frac{deg}{s}$
19	gyro from the right-lower-arm sensor (Y axis)	$\frac{deg}{s}$
20	gyro from the right-lower-arm sensor (Z axis)	$\frac{deg}{s}$
21	magnetometer from the right-lower-arm sensor (X axis)	<i>local</i>
22	magnetometer from the right-lower-arm sensor (Y axis)	<i>local</i>
23	magnetometer from the right-lower-arm sensor (Z axis)	<i>local</i>
24	Label (0 for the null class)	-

Table 1 - Meaning of columns

The collected data predicts a total of twelve different, physical activities. These activities are shown in table 2.

Label	Activity	Duration/Repetitions
L1	Standing still	1 min
L2	Sitting and relaxing	1 min
L3	Lying down	1 min
L4	Walking	1 min
L5	Climbing stairs	1 min
L6	Waist bends forward	20 repetitions
L7	Frontal elevation of arms	20 repetitions
L8	Knees bending (crouching)	20 repetitions
L9	Cycling	1 min
L10	Jogging	1 min
L11	Running	1 min
L12	Jump front and back	20 repetitions

Table 2 - Activity set

### 3. Implementation of the Model

We sequentialized the data and counted the amount of activities. Figure 1 shows the distribution of activities of one participant that have not been classified in the given dataset, but were predicted by the proposed model and visualized in a pie chart.

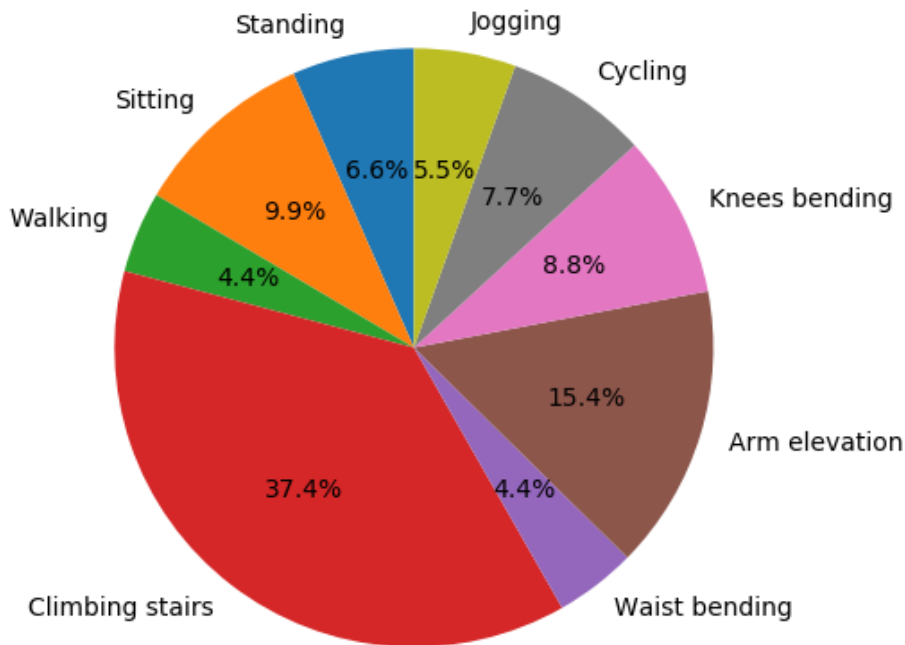


Figure 1 - Activities of a single participant

The predicted data field is column 24, the label of the activity. The model will use all the gathered vital signs to predict this value.

The data, since it is delivered in several files, is concatenated and sequentialized. In order to train the model, we splitted the provided data set to 50% training data. The rest of the data was splitted into testing and validation equally, i.e. 25% testing data and 25% validation data. The model itself was tested with three different methods: K nearest neighbors, support vector machine and gradient boosting. K nearest neighbors resulted in a validation accuracy of 93.27%, the support vector machine scored with 97.51% validation accuracy and gradient boosting delivered a validation accuracy of 98.13%. Therefore, the best performing model, the gradient

boosting approach, was selected for further computations, resulting in an 98.44% test accuracy.

As mentioned earlier, the model supports the possibility to assign scores to a user, i.e. the users activities receive a score and thus, it is possible to calculate if the user is active or not active. Using this information, several parties are able to react to it. Information on how to use the model and a more detailed explanation can be found in the readme.md file.

## **4. Conclusion**

The presented activity score model detects the currently executed activity and assigns a score to it. As an extension of the model, controlled experiments as well as field studies could determine additional insights, e.g. “what is fit?” or “what is lazy behaviour?”.

Furthermore, it is possible to add new data to the model. The activity score could benefit from information about the user. Personal information like age, gender, body – mass – index could improve the activity score and enrich the outcome.

Integration of the model into daily life is the main goal of this approach. There derivation of different scenarios is possible. In the healthcare sector, it offers information about upcoming health risks from a lack of physical activity. This enables medical personal to prescribe preventive measures. The Smart Home sector could also benefit from the activity score model. Possible scenarios are for example a cozy environment after having a hard day. Activities are captured over the day and the smart home reacts to the user by offering a cozy environment with appropriate light and heating atmosphere or by setting up cold water if the user returns from a sports session. Another important aspect in the smart home sector would be a motivation that enables the user to do some sports after sitting for the duration of the day, e.g. by playing motivation music. As a last sector, insurance companies would also benefit from the model. They could create new business models dynamically around the activity behavior of the customers. Customers would also benefit from the model, since they could have an opportunity of lower prices when behaving healthy. Lower

prices could also lead to a change in behavior of the users, since they're motivated by them.