Deliverable 2.3
Data analytics and interpretation models

Lead Partner:	UNIPR
Authors:	Niccolò Mora, UniPR
Contributors:	
Date:	06-02-2017
Revision:	V1.0
Dissemination Level:	Public

ENSAFE

Project Acronym:	ENSAFE	
Project full title:	Elderly-oriented, Network-based Services Aimed at independent liFE	
AAL project number:	AAL	2014-1-112
With support of:		

Content

1 Introduction

The aim of the ENSAFE project is to detect relevant changes in subject's routines and behaviour (which, for instance, may be early indicators of incipient medical conditions), as well as to provide pro-active feedback to different end-users, including the subjects themselves or professional/informal caregivers. In order to do so, the ENSAFE project collects data from multiple sources, including smartphone, wearable, medical devices and home environmental sensors. Depending on the type of sensor, various analysis techniques are implemented to produce a full set of relevant information, including, typically:

- Routine/habits detection, e.g. frequent point of interest from the phone's GPS data, or patterns of interaction with home sensors.
- Trend analysis & anomaly detection, to spot, e.g., increasing toilet visits during night or anomalous usage patterns.
- Similarities between users' behaviours by means of suitably defined metrics.

In the following sections, the concepts at the core of the analytics services will be presented and discussed.

2 Home sensors domain

2.1.1 Sensor kit description

Data collected from smart home sensors offer a comprehensive set of information which can be exploited in many different ways: from the monitoring of very specific actions (such as user's sleeping patterns) up to building models of a *typical resident's day.*

Three major types of sensors are usually exploited to gather information about home residents' routines:

- *motion* sensors (mainly Passive InfraRed, PIR), which carry information about activities being carried out within their sensing range.
- *interaction* sensors, which require the user to specifically interact with them. Examples are: *(i)* magnetic switches for checking on doors, windows, cabinets, *(ii)* presence sensors, such as pressure pads for detecting bed/chair occupancy, *(iii)* energy meters for keeping track of an appliance's use, *(iv)* accelerometer-based units for detecting when a person moves them, and many others.
- *environmental* sensors, such as temperature, humidity, light, which can provide further context to the data.

Each one carries different information and, in general, the first type is most commonly encountered in literature, being the less obtrusive. However, the information they are able to convey is much coarser than that of the second type of sensors. In fact, interaction sensors, are much more specific and related to a given activity, but, on the other side, a larger amount of such sensors must be deployed to address different activity monitoring analyses.

The amount of data being generated by the home obviously depends on the number of sensors, as well as their type (motion sensors are, indeed, likely to be triggered much more than a *interaction sensor*, e.g. a bed sensor), and generally up to some thousand events per day can be generated, in a typical scenario. However, having such a volume of data does not guarantee to be able to extract and monitor each user's behaviour precisely. In fact, while some target behaviour may be quite specific and easy to monitor with a few sensors (e.g. toilet use), many other routines can vary so much in terms of duration, sequence of activated sensors and time of the day that modelling each possible scenario would require an intractably large amount of data.

The ENSAFE home sensor kit features the following sensing devices:

- Bed presence sensor
- Chair presence sensor
- Toilet sensor
- PIR motion sensor
- Magnetic contact (e.g. for medicine cabinet)
- Smart plug for electrical appliance monitoring

2.1.2 Cloud-based processing pipeline

Data gathered by the smart home kit are wirelessly forwarded to a home gateway, which acts both as a controller for the sensor network, as well as a bridge towards cloud services. All cloud functionalities are implemented using Microsoft Azure services.

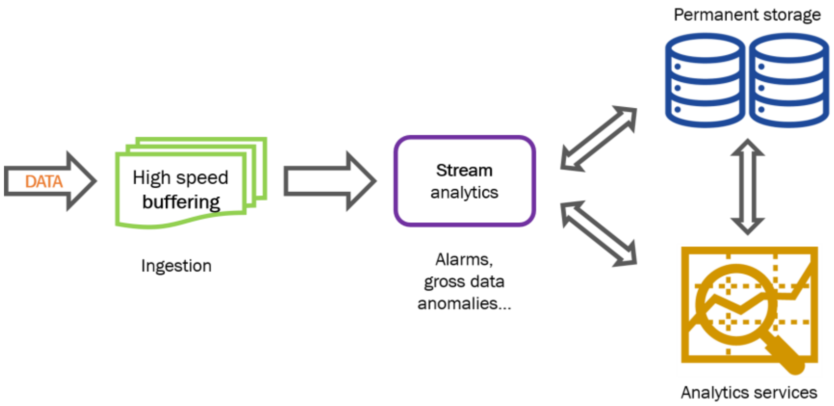

Figure 1: schematic of the cloud-based architecture

As shown in Figure 1, data sent to the cloud are immediately ingested by a high-bandwidth event buffer (IoT Hub). A stream analytics service can then perform fast analyses on demand (e.g. detection of alarms), or commit the data to permanent storage, where the analytics services will draw their inputs. The permanent storage is implemented with Azure SQL DB technology, whereas the analytics services are deployed as Azure ML Studio web services, which can be called via standard REST APIs. Such services can be either called on demand or via scheduled jobs.

2.1.3 Analytics examples

In this section, some examples of the smart home analytics are provided. Such analyses are used as building blocks for the ENSAFE system services, since they provide quantitative information to be further refined before interfacing to the users.

Three main category of analysis were developed, best detailed in the following subsections:

- *Regression analyses* with *outlier detection*. The aim of such investigations is to spot longitudinal trends, accounting for differences between week days and week-ends. These models also allow to detect data that do not fit well the general observed trend, and the system is able to flag them as anomalous, possibly triggering further investigations.
- *Activity curves*. Activity curves model the average user-sensor interaction throughout the day. Such curves may be used to detect changes in patterns of use, as well as a basis for measuring routines similarity between users
- *Sensor traces*. The most recent data for each sensor (the preceding 48 hours, cleaned and filtered) are available for display and further analyses.

(i) Regression analyses & outlier detection

Figure 2 shows an example of a regression plot focused on the detection of trends and anomalies in the sleeping behaviour along 3 months of data. After a first step of outlier removal, based on interquartile range criterions, a regression model is computed to best explain the data given three factors:

- *intercept*: the average effect

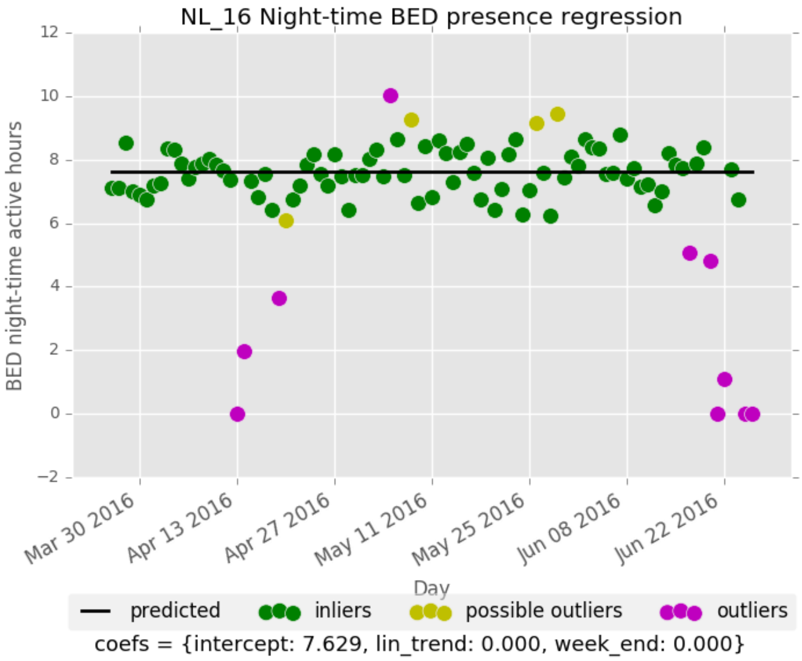

Figure 2 Regression and outlier detection for night-hours bed presence

- *lin_trend*: the effect of a linearly increasing trend on the hours slept
- *week_end*: the effect of considering differences between week days and week ends

The best model that gives the optimal tradeoff between data fitting and complexity is chosen using the Bayesian Information Criterion (BIC). As can be seen in the legend of the plot in Figure 2, the best model does not detect any significant linear trend effect, nor any difference between weekdays and weekends. Given the fitted model, we can also highlight possible outliers in the time series, as reported in the scatter plot (yellow dots for possible outliers and magenta for gross outliers).

Figure 3, instead, shows a case in which the regression model detects a linearly decreasing trend in chair usage throughout the day (≈-1 h in 2 months), compatible with the user being out of home more often. Possible outliers are detected as well, shown as yellow dots in the plot.

(ii) Activity Curves

Activity curves represent the expected interaction between users and specific sensors throughout the day. In order to compute such curves, the day is partitioned into 30 minutes time slots, and for each one of them the expected probability/percentage of time the user interacts with the sensor is predicted, along with confidence bounds. An example of such analysis is reported in Figure 4, where activity curves for bed, chair and toilet sensor are computed (shaded areas represent the 95% confidence intervals).

Such curves may be useful to check for user's behavioural consistency: it is possible to compare two different observation windows (e.g. the most recent 3 weeks vs. 3 reference weeks) and detect any statistically significant change. This analysis of significance may be further exploited by higher-level models to detect larger behavioural changes. An example of such analysis is shown in Figure 5, where the probability of visiting the toilet at least once in each 30 min. bins is compared between two

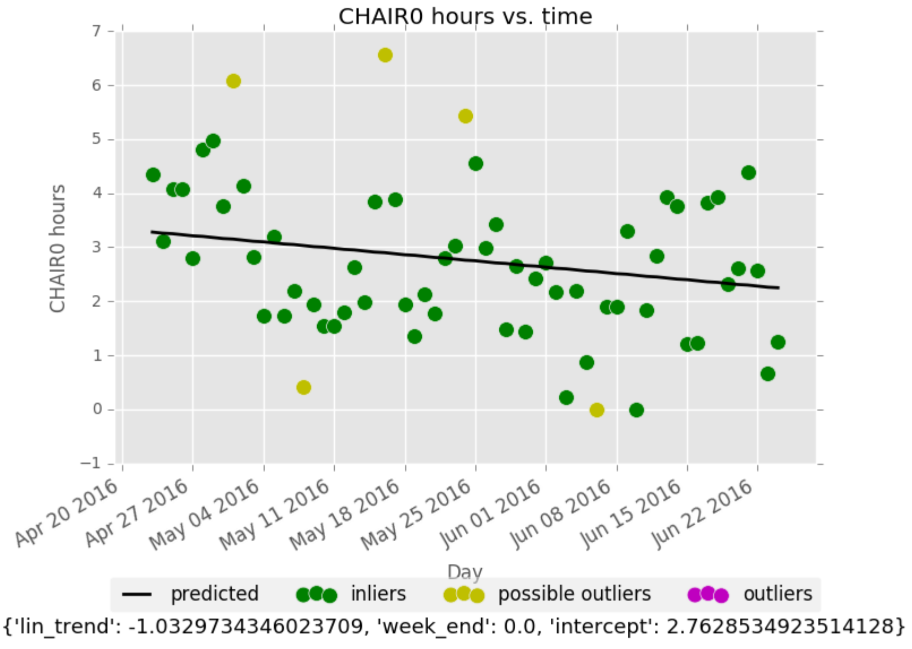

Figure 3 Regression and outlier detection for chair presence

different periods. Hypothesis testing is used to detect relevant differences between each time slot of the two periods.

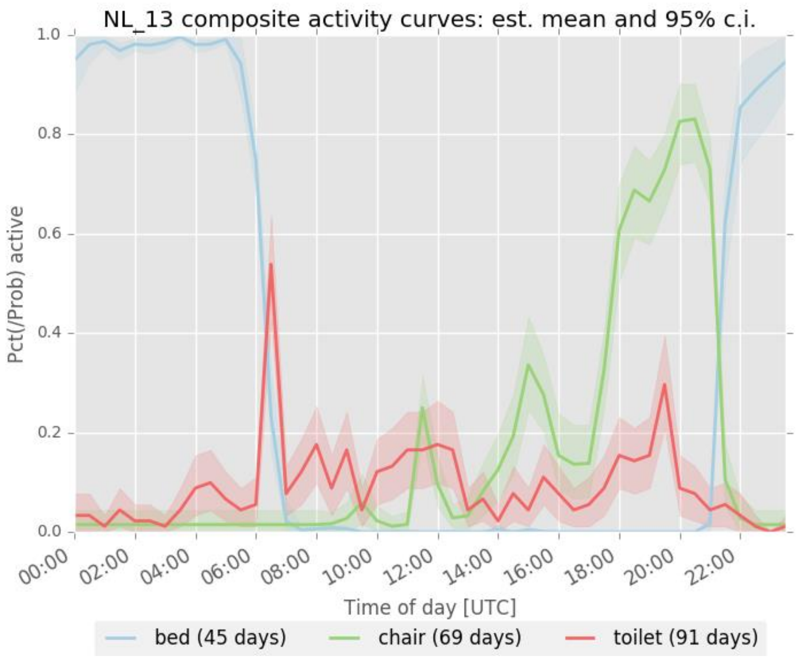

Figure 4 Activity curves of different home sensors

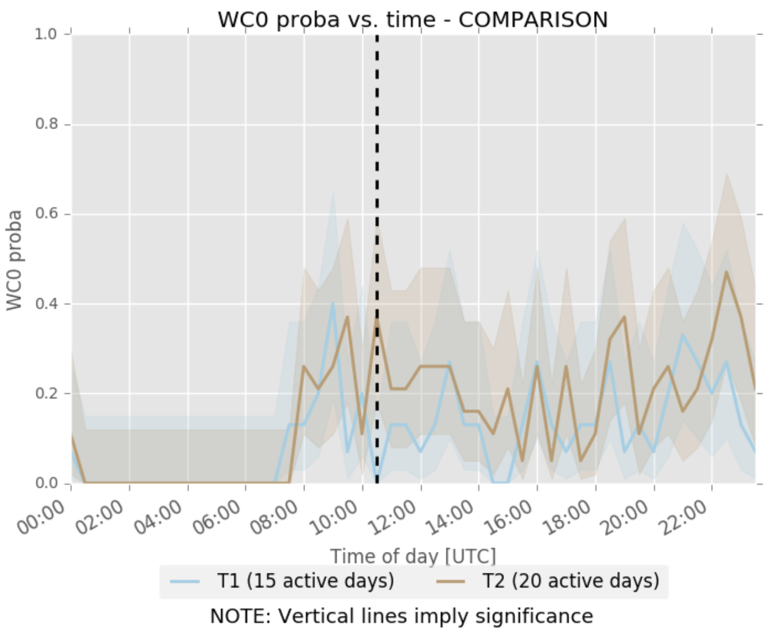

Figure 5 Activity curves comparison between different periods

(iii) Sensor traces

Sensor traces, i.e. minutes active within 15 minutes windows are also provided as output to enable further processing or user-friendly display of most recent history. An example is given in Figure 6.

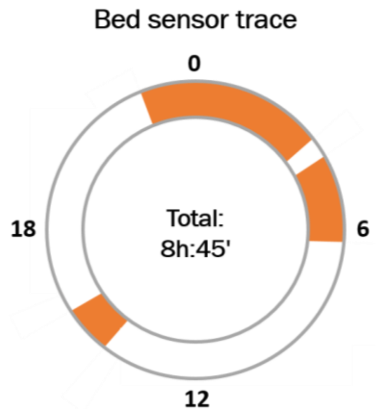

Figure 6 Example of bed sensor trace. Active periods are in solid orange colour.

3 Conclusion

This document presented the data-analytics services of the ENSAFE platform. The system gathers data from many different sources, including

- Smartphone
- Wearable devices
- Medical devices
- Home environmental sensors

Such data are then used to provide quantitative information about:

- Trend analysis & anomaly detection
- Routine/habits detection
- Similarities between users' behaviours
- Recap of last days

The services are implemented using cloud-computing frameworks. For example, Microsoft Azure is heavily leveraged to collect IoT data, create stream analytics tasks, run statistical analyses and publish results via web services.