



Assignment

“First insights in the data”

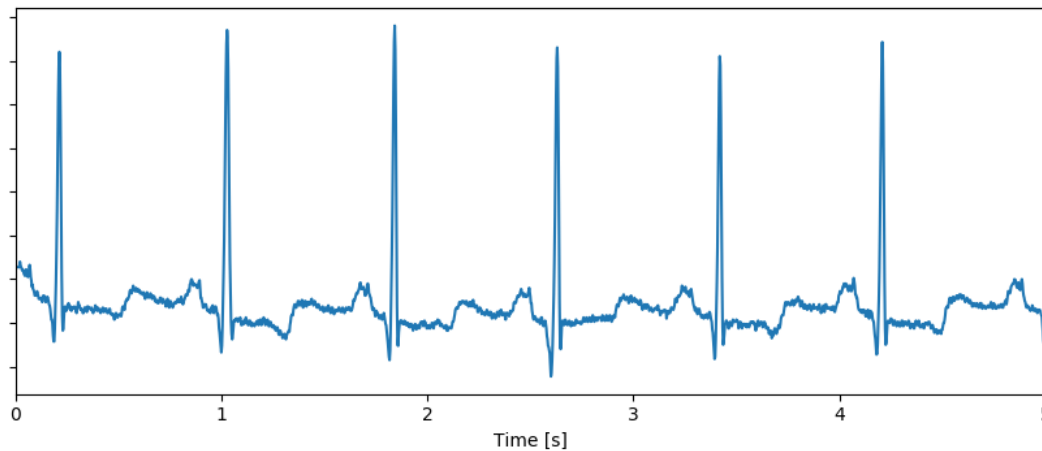
for the Full-Time-Position *Data Scientist*

at monikit

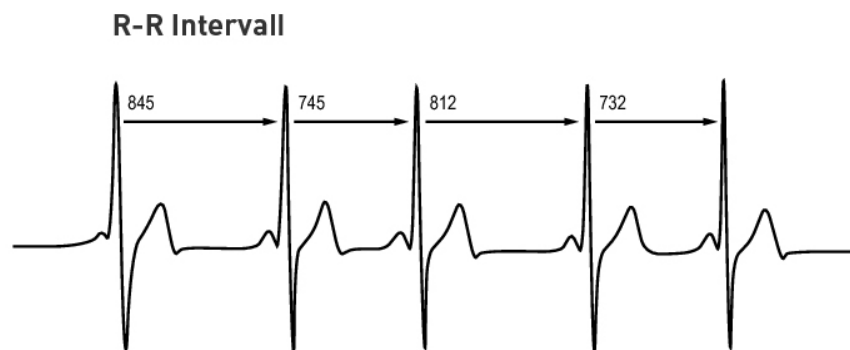
1. Background

ECG

This is what a typical ecg signal looks like. The prominent peaks are called 'R-Peak' and can be seen as a heartbeat.



In our project we analyse the time between two R-Peaks, the so-called RR intervals like shown in this picture:



These RR intervals vary all the time, in healthy people as well as in epileptic patients. Our main goal is to find specific patterns in the variation of the RR intervals, which only occur if a patient has an epileptic seizure.

Epileptic seizures

There are many different types of epileptic seizures but for this task we will only deal with simple and complex partial seizures. These are generally seen as 'small' seizures.

2. Description of the data

We will provide you with a file called `data.zip`. This file contains a scientific paper and three comma-separated text files:

- `data.csv` contains two columns: the timestamp of each RR interval and the length of this interval. Both given in milliseconds. This data is from one continuous record of about 160 hours.
- `data_header.csv` contains general information about the record which was the source for the data
- `seizures.csv` contains information about the seizures occurred in the given record.

Important columns are:

- classification which can be 'SP' for simple partial, 'CP' for complex partial and 'UC' for unclassified
- `eeg_onset_secs` which gives the beginning of the seizure given in seconds from start of the record

All other columns can be ignored. Refer to the seizures only by IDs given by their row number (1:22) and ignore the `seizure_id` column.

3. Your Tasks

1. Load the data and plot it. What do you generally notice when you look at the data? Describe it shortly.
2. Implement the following three features:
 - a. Mean
 - b. Standard deviation
 - c. Cardiosympathetic Index (CSI) as described in the given paper
3. Calculate the features on the given data with a window size of 30 seconds and overlap between windows of 5 seconds
4. Make plots of the features for all seizures where the data is good enough
5. Do you think these features are helpful for our main goal, the detection of epileptic seizures? What would you do next?
6. Prepare a 15-20 Minute presentation with all your outcomes and present it to us during the interview. Choose whatever Presentation-Tool you want (PowerPoint, Keynote, Pen and Paper,...).