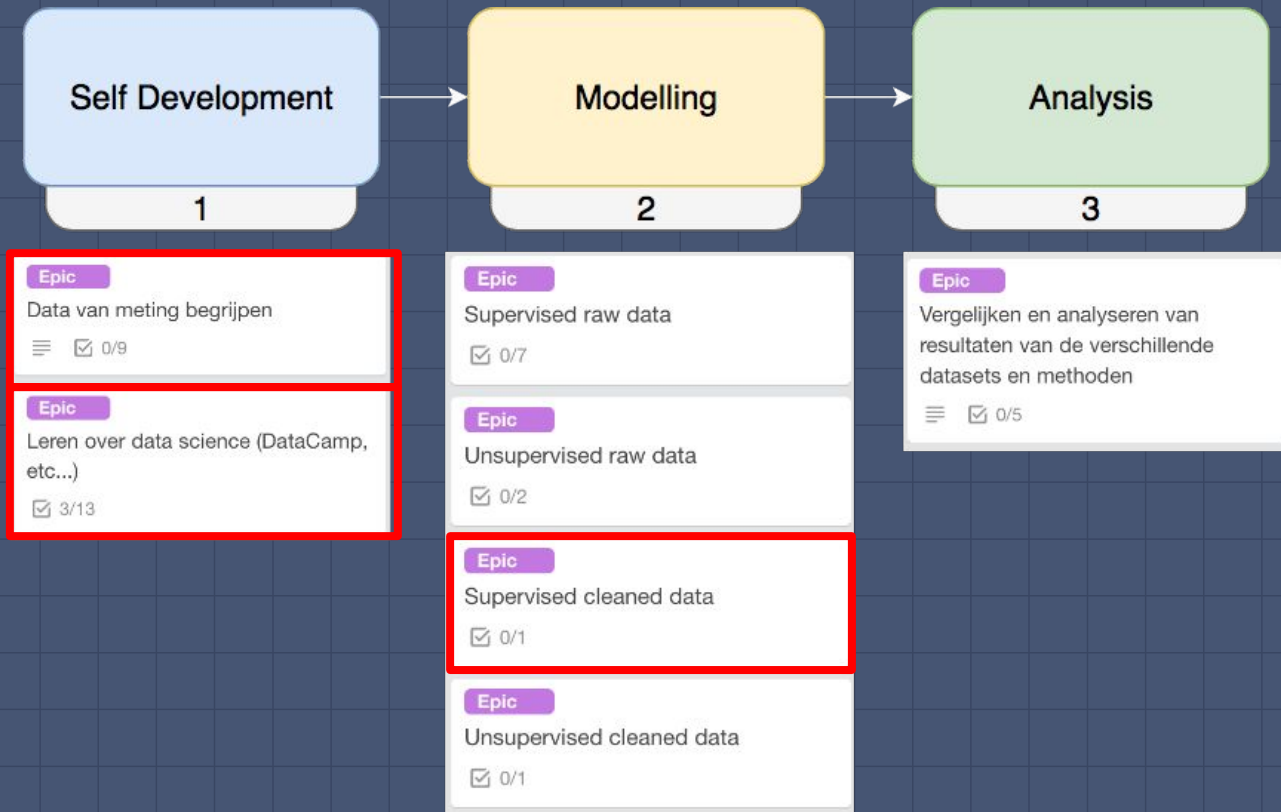


Ortho Eyes

Tony Andrioli

- Kasper van der Hoofd
- Vincent van den Oord
- Rogier Zitman
- Luke de Keijzer

Approach



Sprint planning

Sprint goal: The creation of a classifier

The two components:

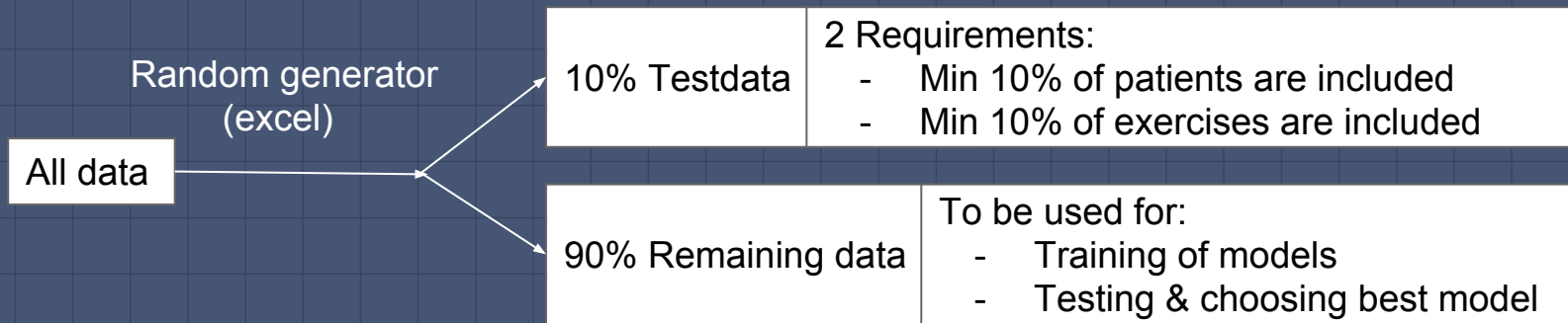
1. Preparing data
2. Building the classifier

– DataCamp & Coursera

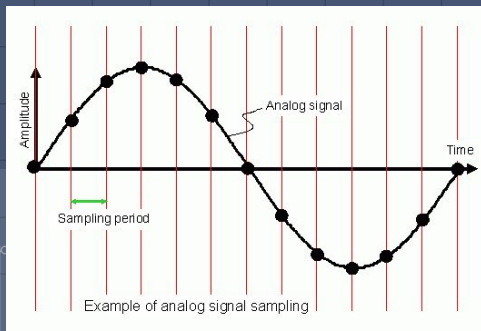
A decorative background graphic at the bottom of the slide. It features a white line graph with circular markers at various points, showing an overall upward trend with some fluctuations. Below the line graph is a series of vertical bars of varying heights, creating a bar chart effect. The entire graphic is rendered in a light blue/white color against the dark blue grid background.

1. Preparing data

1. Setting test-data aside (10% - 90%)



2. Dealing with changing amount of 'frames' in each measurement (time series)



Our approach:

Splitting the time series up in single samples

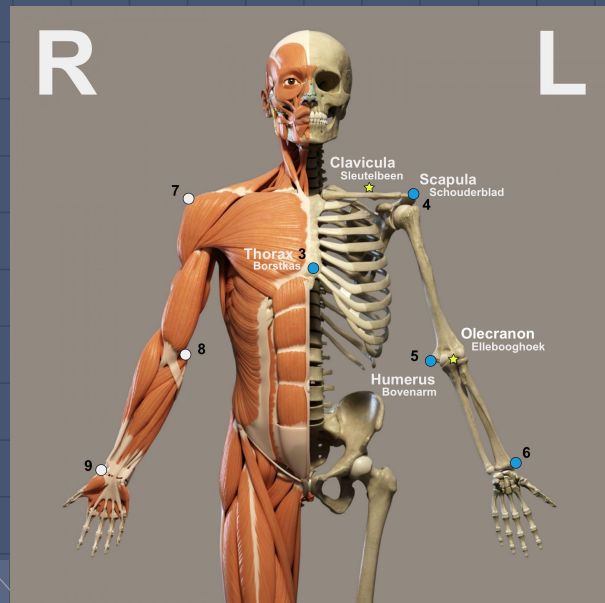
Result:

Csv 32 x 173621

2. Building the classifier (next week)

Sprint goal: The creation of a classifier

1. Building a simple (but scalable) classifier
 - Simple to validate the code
2. Adding more parameters to classifier (if possible)
 - Finishing DataCamp



What have we achieved this sprint?

- Finished Coursera week 3
- Built a classifier
 - Only two features, as a test
 - Added differences in the left and right arm movement, as features
 - Using more features



Data structure

- 4 categories (= patient groups)
- 119 patients, unevenly divided in the 4 categories
- 21 different types exercises, not all exercises are present in all categories
- 1396 individual exercises.
- 191292 data samples, 26 features/parameters each

1st attempt: **make a classifier to see if cat4 data samples can be distinguished from the other category samples.** (The end goal is to classify patients, not individual data samples)

Result

resultaten op testset	pos	neg
Predicted pos	650	119
Predicted neg	1396	32151

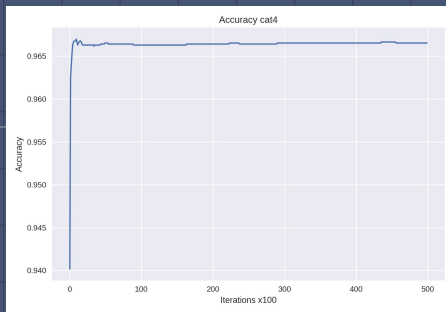
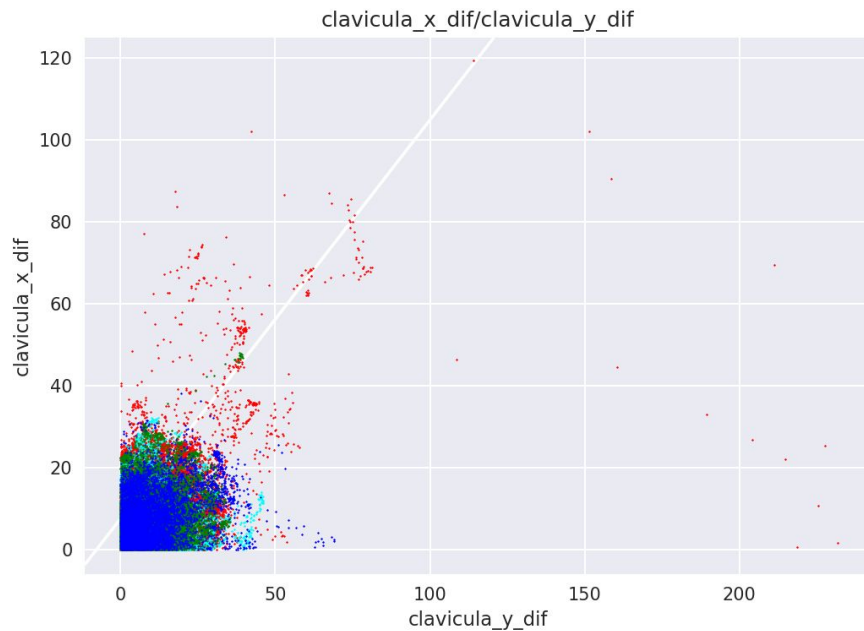
recall 0.322

precision 0.845

accuracy 0.957

Model is trained on:

- 27 features
- 20% of the training dataset



Only 2 of the 27 features are plotted here!

Result

Last sheet was on sample basis.

What if we summarize all samples per patient?

All cat4 patients have more than 10% of samples classified as Cat4.

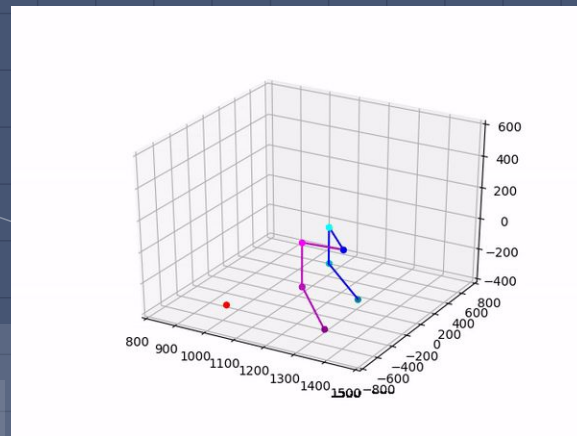
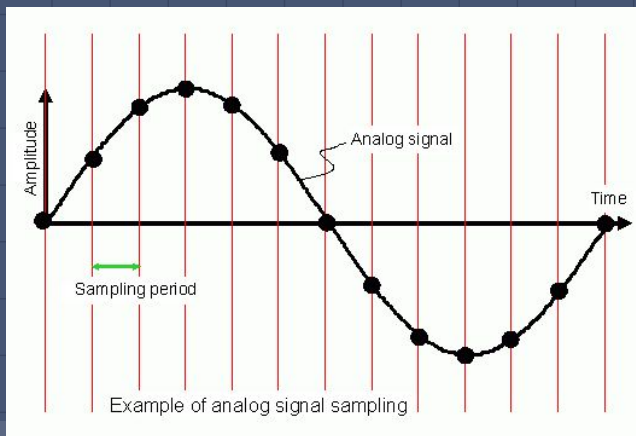
(Other patents van less than 10% classified as cat4)

pat	Rest	Cat4	percentage
4006.0	356.0	569.0	61.513.514
4023.0	423.0	653.0	60.687.732
4014.0	625.0	424.0	40.419.447
4015.0	530.0	219.0	29.238.985
4012.0	724.0	297.0	29.089.128
4011.0	539.0	218.0	28.797.886
4020.0	1389.0	412.0	22.876.180
4024.0	908.0	236.0	20.629.371
4017.0	1402.0	199.0	12.429.731
3031.0	1957.0	170.0	7.992.478
3008.0	3250.0	130.0	3.846.154

Cat 4 patients

What have we planned?

- Meeting with de Groot (LUMC) to make a measurement ourselves
- Classifying with more meaningful features / results
- Tackling the issue of samples frequency



The problems we face or expect to face

- ▣ Memory error on full dataset/parameters
- ▣ Slow on bigger datasets
- ▣ Deciding what is meaningful information
- ▣ Receiving the raw data





Any questions or suggestions?