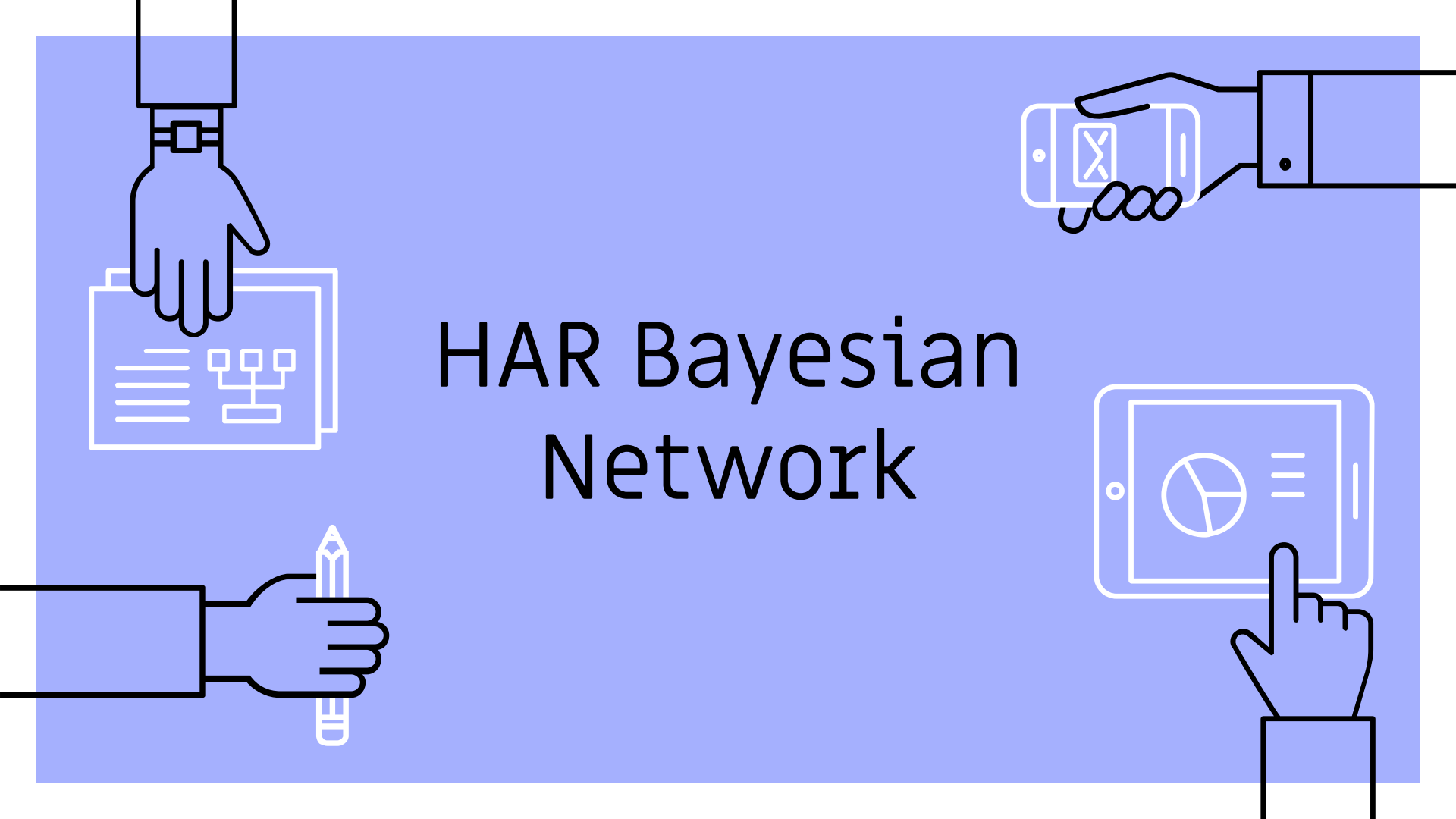
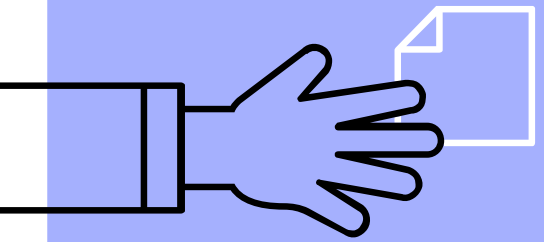
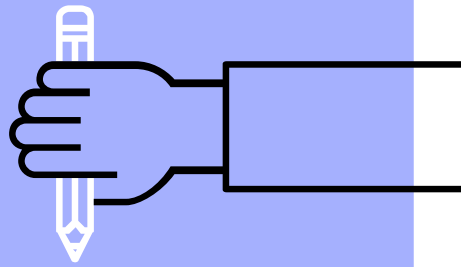
HAR Bayesian Network



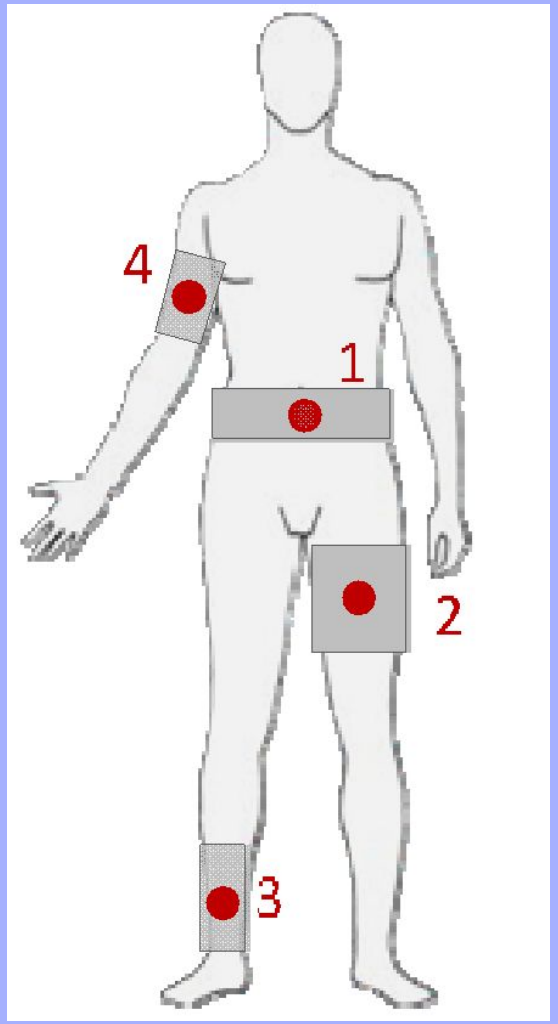
Data domain





Data source

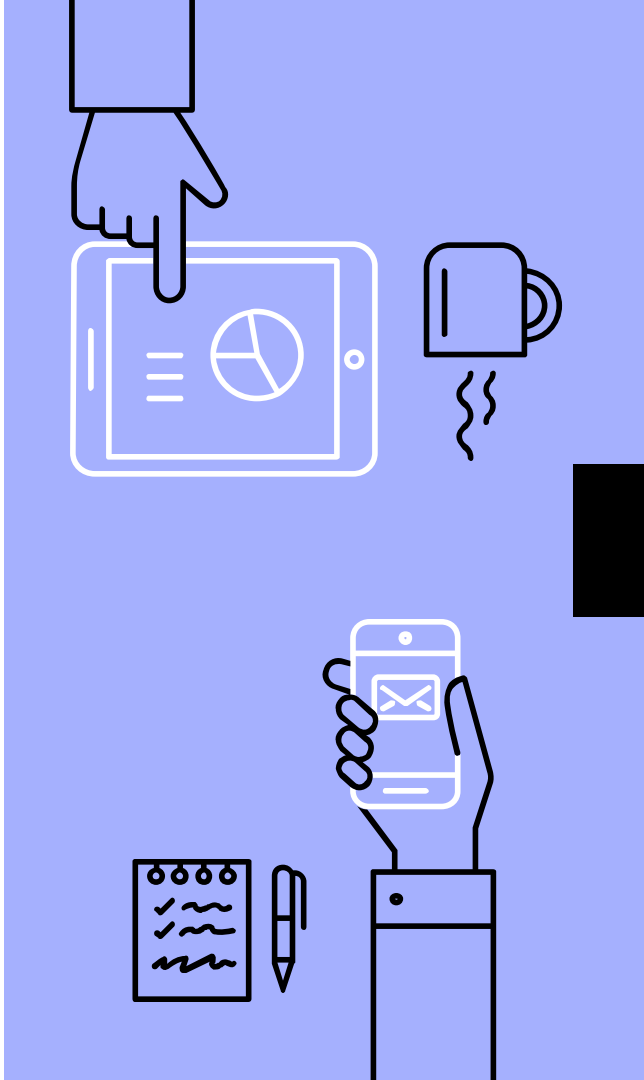
- User attributes like height, weight and body-mass-index
- Data raw snapshots provided by 4 accelerometer
- Sensors calibrated differently for each user
- Data not sensible at limb-rotations in space



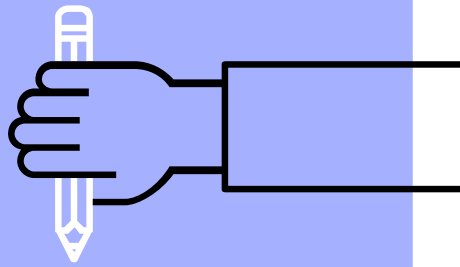
Sensors' statistic analysis

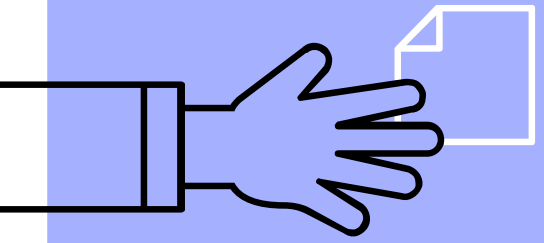
- Different data distributions
- Some data ranges are extremely big compared to standard deviation

Campo	Min	Max	Media	Moda	DevStd
x1	-306	509	-6.649327127	-1	11.61623803
y1	-271	533	88.29366732	95	23.89582898
z1	-603	411	-93.16461092	-98	39.40942342
x2	-494	473	-87.82750418	-492	169.4351938
y2	-517	295	-52.06504742	-516	205.1597632
z2	-617	122	-175.0552004	-616	192.8166147
x3	-499	507	17.42351464	38	52.63538753
y3	-506	517	104.5171675	108	54.15584251
z3	-613	410	-93.88172647	-102	45.38964613
x4	-702	-13	-167.6414483	-164	38.31134199
y4	-526	86	-92.62517131	-94	19.96861022
z4	-537	-43	88.29366732	-162	13.22102006



Data design





Features choices and cleaning

CHOOSEN

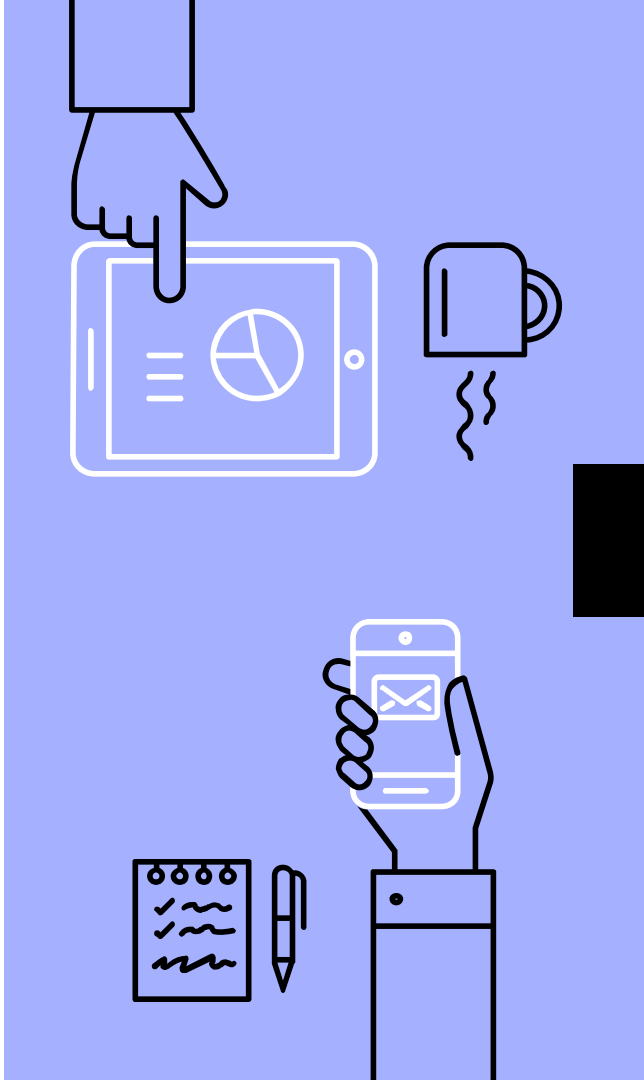
- xi
- yi
- zi
- class

IGNORED

- user
- gender
- age
- weight
- body_mass_index

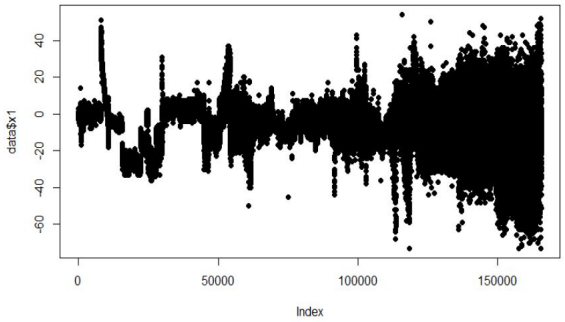
NOTES

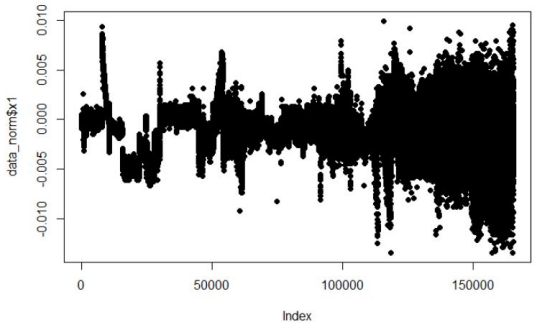
how_tall_in_meters is an important feature but it can be ignored because of the narrow range of the users' heights.



Normalization

Data have been normalized using a built in function provided by sklearn.preprocessing module. The range il from -1 to 1.

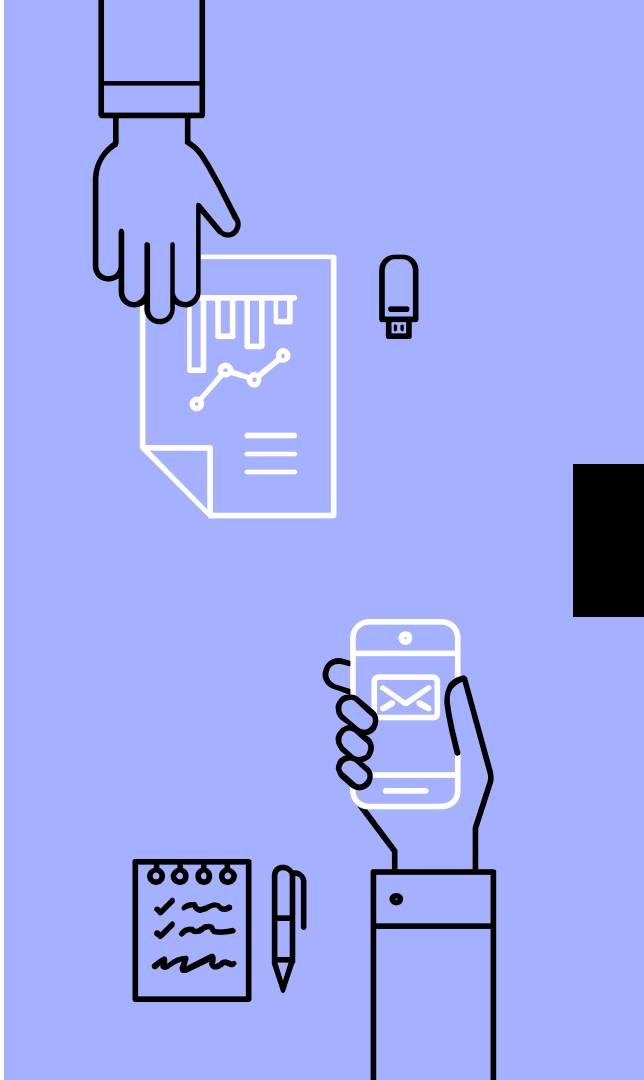




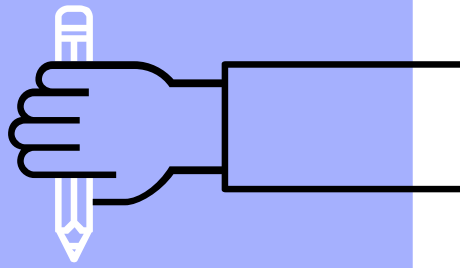
Discretization

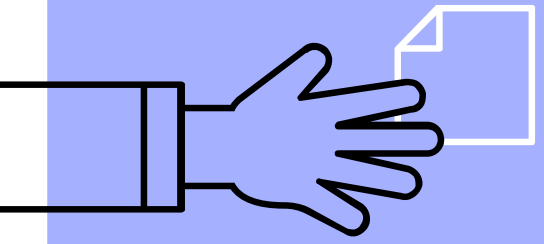
- Normalized data can assume all real values inside the specified range.
- The dimension of the data domain has been squeezed in order to reduce the complexity of the bayesian network and to drastically decrease the required computation time.

x1	y1	z1	x2	y2	z2	x3	y3	z3	x4	y4	z4	class
1	4	3	0	0	0	2	3	1	1	3	2	1
4	0	0	0	0	0	4	1	0	2	4	0	0
3	4	0	3	1	4	0	3	0	0	2	4	4
1	3	3	3	3	2	4	4	4	3	1	4	0
2	4	3	4	3	3	1	3	1	1	3	2	1
1	2	0	2	4	1	3	3	2	3	4	4	1
1	1	0	0	0	0	2	3	1	1	4	0	0
4	4	2	3	3	2	1	0	1	0	3	2	1
0	4	3	2	3	1	0	4	4	3	1	3	0



Models



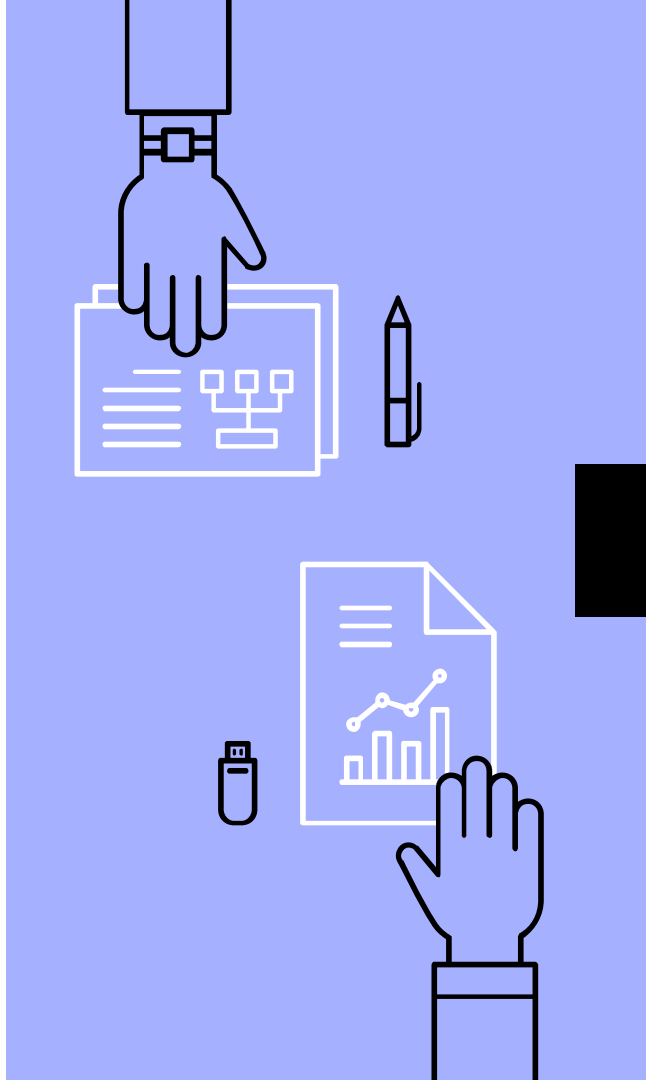


pgmpy

Correlation model (1/2)

Based on Pearson correlation coefficient $\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$

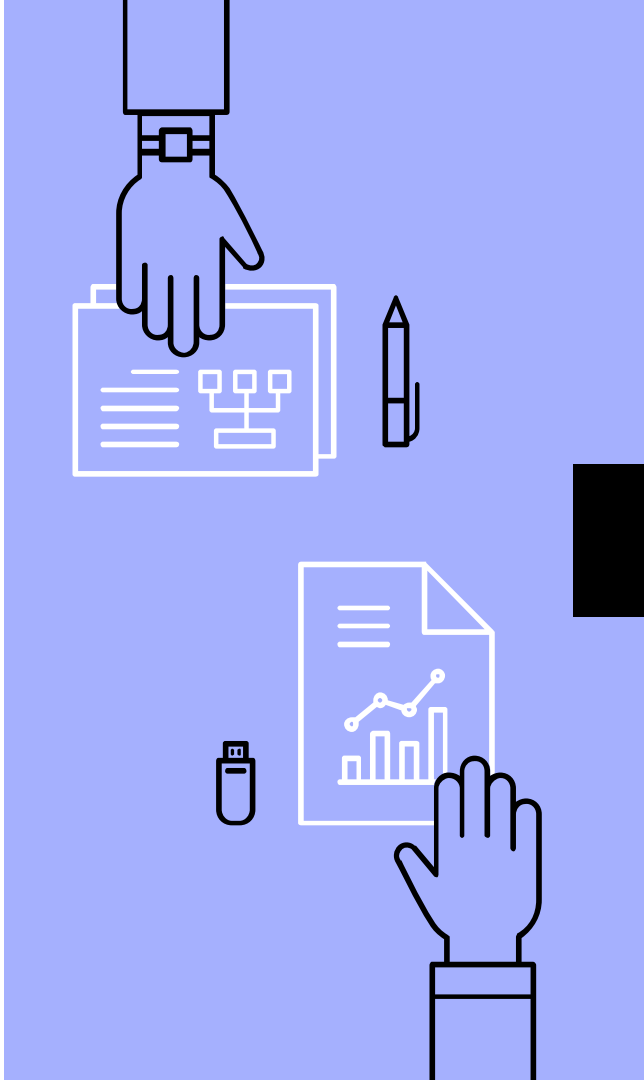
x1	y1	0.345808064	x2	x1	0.198233156	x3	x1	0.07121	x4	x1	0.099018401
x1	z1	0.030417499	x2	y1	0.043861069	x3	y1	-0.1816	x4	y1	-0.29045831
x1	x2	0.198233156	x2	z1	0.109970588	x3	z1	0.13028	x4	z1	0.50628732
x1	y2	0.202451431	x2	y2	0.977115735	x3	x2	0.00802	x4	x2	0.157007278
x1	z2	0.251138704	x2	z2	0.953083727	x3	y2	0.00636	x4	y2	0.081715165
x1	x3	0.071207709	x2	x3	0.008022688	x3	z2	0.00665	x4	z2	0.287828127
x1	y3	-0.136539934	x2	y3	-0.140455652	x3	y3	0.32843	x4	x3	0.166694382
x1	z3	0.00417011	x2	z3	0.107520027	x3	z3	0.27106	x4	y3	-0.111224183
x1	x4	0.099018401	x2	x4	0.157007278	x3	x4	0.16669	x4	z3	0.035357614
x1	y4	-0.142551934	x2	y4	-0.23983578	x3	y4	0.04272	x4	y4	-0.600982199
x1	z4	-0.025592835	x2	z4	0.164505939	x3	z4	-0.2023	x4	z4	-0.068008246
y1	x1	0.345808064	y2	x1	0.202451431	y3	x1	-0.1365	y4	x1	-0.142551934
y1	z1	-0.5159614	y2	y1	0.138219068	y3	y1	0.19162	y4	y1	0.228997237
y1	x2	0.043861069	y2	z1	0.017749163	y3	z1	-0.119	y4	z1	-0.405502292
y1	y2	0.138219068	y2	x2	0.977115735	y3	x2	-0.1405	y4	x2	-0.23983578
y1	z2	-0.0301789	y2	z2	0.918648041	y3	y2	-0.096	y4	y2	-0.15437806
y1	x3	-0.181573944	y2	x3	0.006358901	y3	z2	-0.2002	y4	z2	-0.389086436
y1	y3	0.191617913	y2	y3	-0.095987852	y3	x3	0.32843	y4	x3	0.042718189
y1	z3	0.109626835	y2	z3	0.120384491	y3	z3	0.67093	y4	y3	0.3239336
y1	x4	-0.29045831	y2	x4	0.081715165	y3	x4	-0.1112	y4	z3	0.076057385
y1	y4	0.228997237	y2	y4	-0.15437806	y3	y4	0.32393	y4	x4	-0.600982199
y1	z4	0.186791537	y2	z4	0.165709979	y3	z4	-0.0364	y4	z4	-0.117144404
z1	x1	0.030417499	z2	x1	0.251138704	z3	x1	0.00417	z4	x1	-0.025592835
z1	y1	-0.5159614	z2	y1	-0.0301789	z3	y1	0.10963	z4	y1	0.186791537
z1	x2	0.109970588	z2	z1	0.2172898	z3	z1	0.12427	z4	z1	-0.197813141
z1	y2	0.017749163	z2	x2	0.953083727	z3	x2	0.10752	z4	x2	0.164505939
z1	z2	0.2172898	z2	y2	0.918648041	z3	y2	0.12038	z4	y2	0.165709979
z1	x3	0.130282473	z2	x3	0.006650803	z3	z2	0.07906	z4	z2	0.160719903
z1	y3	-0.118961498	z2	y3	-0.200237814	z3	x3	0.27106	z4	x3	-0.202266103
z1	z3	0.124272757	z2	z3	0.07905866	z3	y3	0.67093	z4	y3	-0.036427754
z1	x4	0.50628732	z2	x4	0.287828127	z3	x4	0.03536	z4	z3	0.031079739
z1	y4	-0.405502292	z2	y4	-0.389086436	z3	y4	0.07606	z4	x4	-0.068008246
z1	z4	-0.197813141	z2	z4	0.160719903	z3	z4	0.03108	z4	y4	-0.117144404



Correlation model (2/2)

- Number of bins: 5
- CPTs estimation through the library's API: `MaximumLikelihoodEstimator`

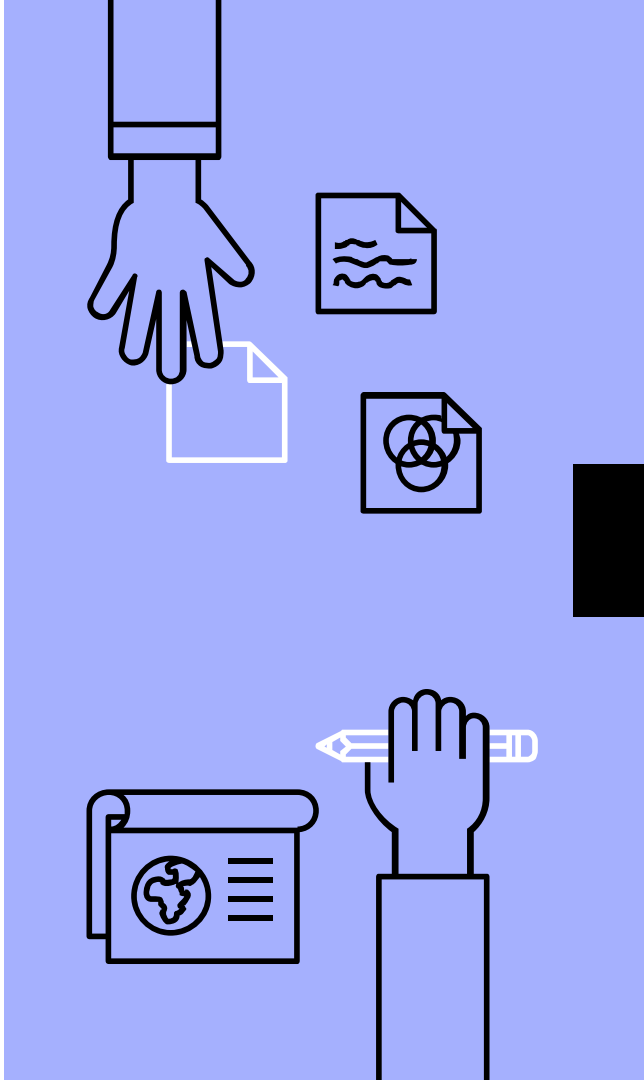
- Accuracy estimated: around **54%**
- Long computational times
- Model rejected

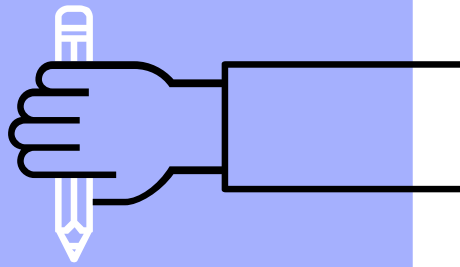


Generated model

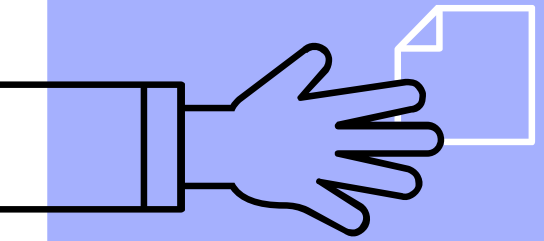
- Number of bins: 5
- Dependencies generation through the library's API: `build_skeleton`

- Method hasn't finished the computation after 8 hours. Then aborted
- Model rejected





Models

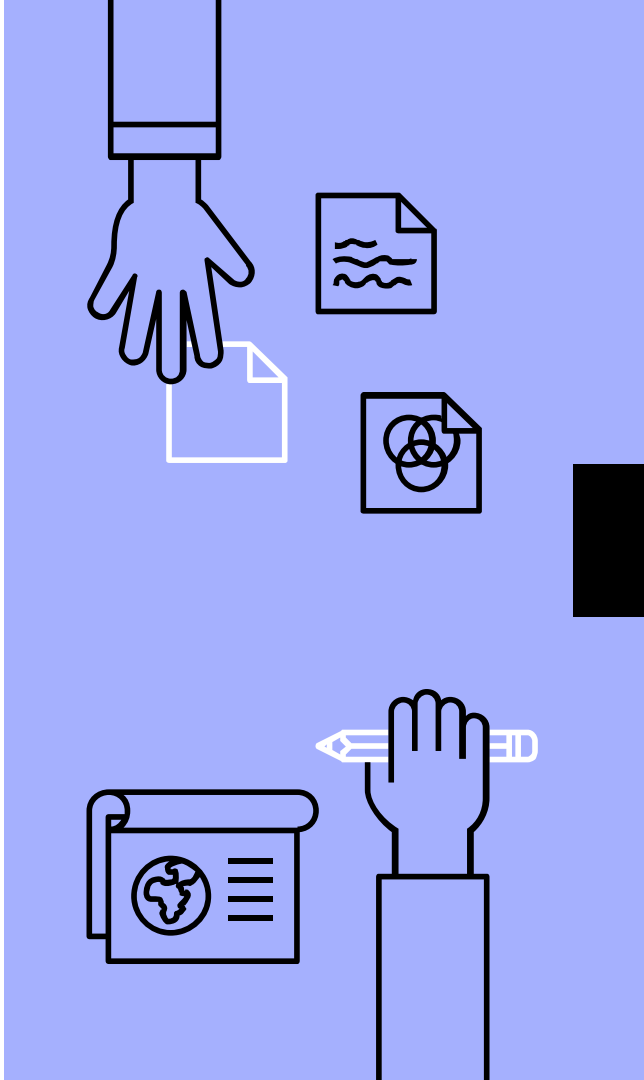


pomegranate

Generated model (1/3)

- Number of bins: 5
- Entire model estimation through the library's API: `from_samples`

- Accuracy estimated: around **97%**

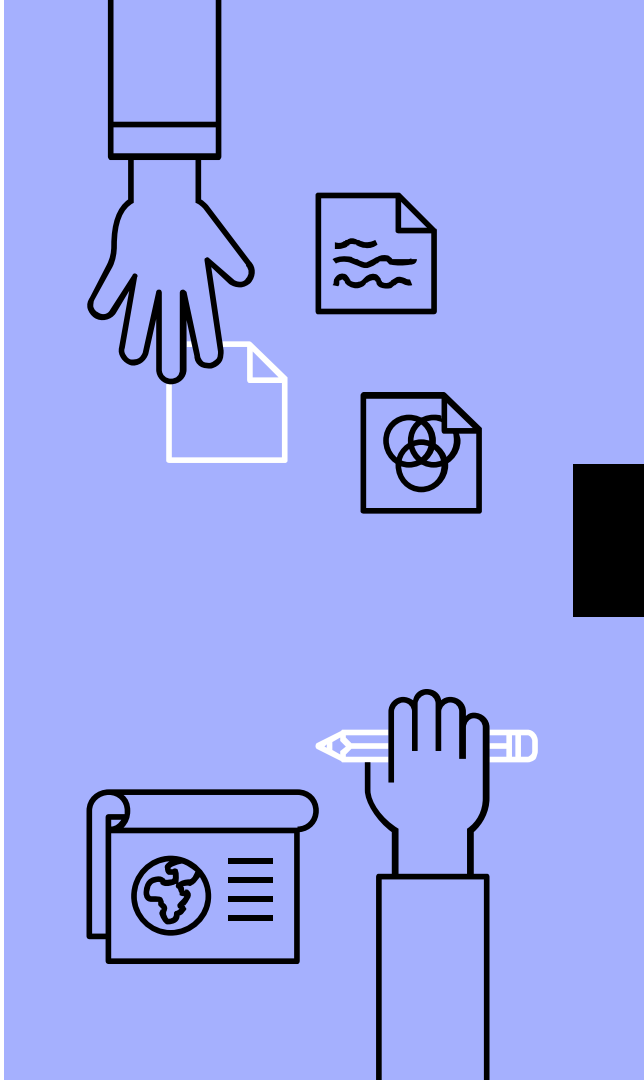


pomegranate

Generated model (2/3)

Confusion matrix

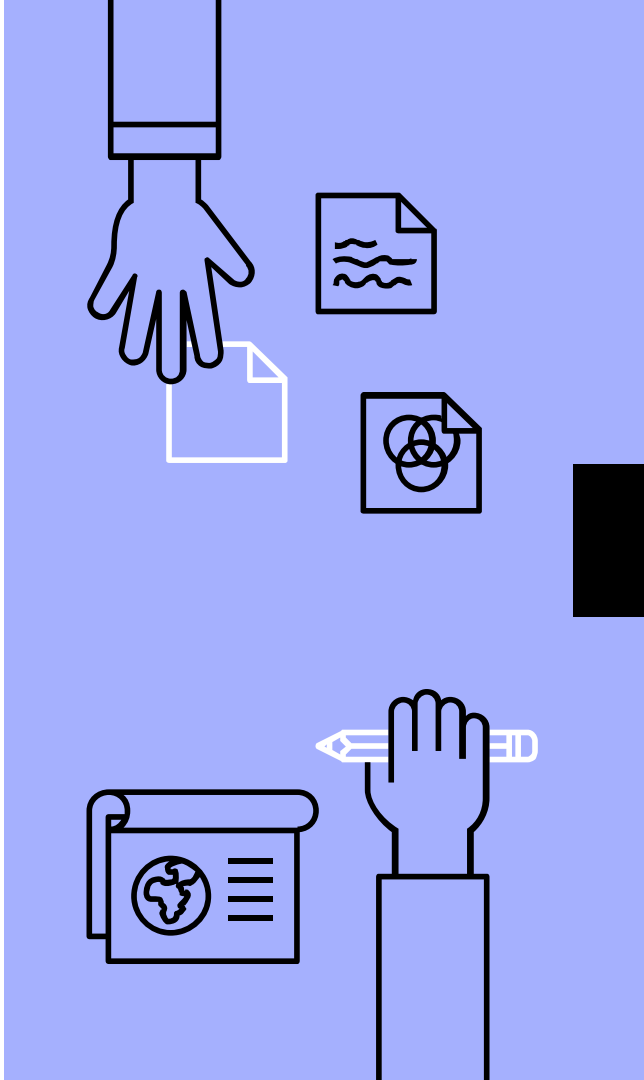
Pred\Exped	Walking	Standing	Standing Up	Sitting	SittingDown
Walking	4180	61	37	2	34
Standing	44	4753	19	0	7
StandingUp	30	20	1100	21	46
Sitting	4	0	9	4965	1
SittingDown	35	12	76	45	1047



Generated model (3/3)

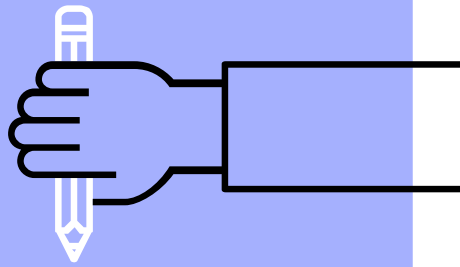
	Precision	Recall	F1 Score
Walking	97%	97%	97%
Standing	99%	98%	98%
Standingup	90%	87%	90%
Sitting	99%	99%	99%
Sittingdown	86%	92%	89%

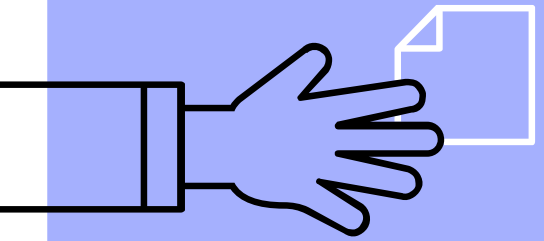
- Model accepted



pomegranate

Conclusions





- The classes are homogeneously distributed over the dataset
- *Learn structure by data* approach works well
- Python libraries are not optimized and complete
- The model has height/weight limits