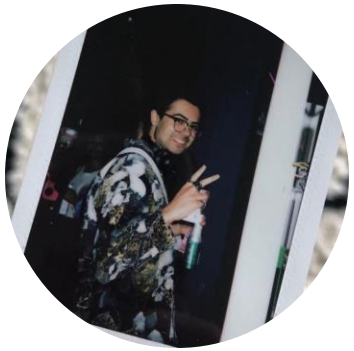
ActivPal External Presentation #5

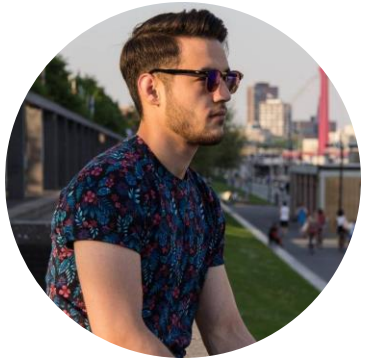
Adnan Akbas
Ali Safdari
Mark Boon
Matthew Turkenburg
Colin Werkhoven

Project group



Colin

Mark

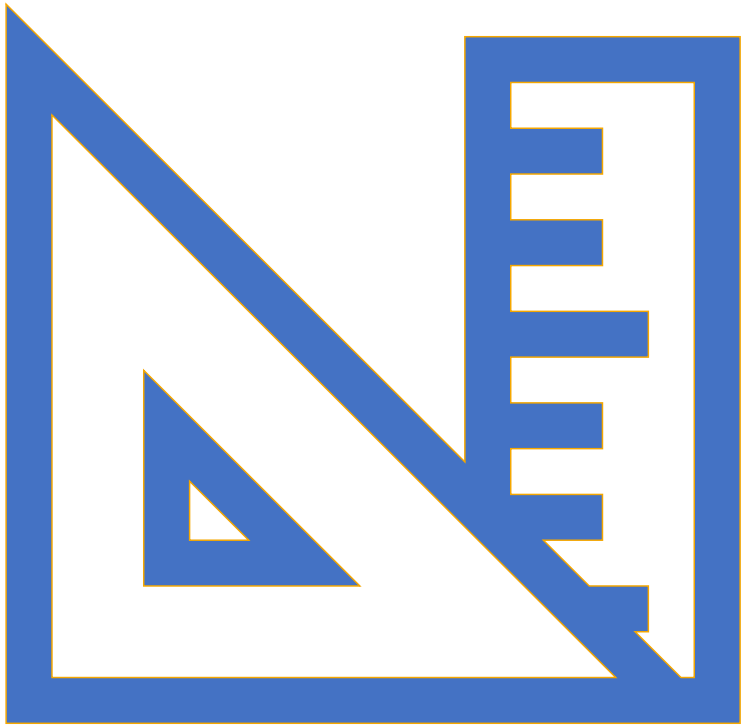


Ali

Adnan

Matthew

Table of content



Problem domain

- SQUASH-questionnaire was not accurate enough
- Find a more accurate way to estimate if people did enough exercising

Assignment

Research questions:

1. How can Machine Learning be used to predict the intensity of activities performed in a lab situation by a person, who is being monitored with Vyntus One device and wearing an ActivPal accelerometer?
2. How can Machine Learning be used to predict the intensity of activities performed by a person wearing only the ActivPal accelerometer, based on the data gathered from the Vyntus One device and the ActivPal accelerometer in the lab situation?
3. How can Machine Learning be used to determine whether people did their 150 minutes of moderate activity in the ActivPal accelerometer data of an entire week?

Data

- 31 respondents
- Accelerometer data
- Vyntus data
- Test data

Our approach

- Scrum sessions
- Literature research
  - What are x,y,z axis
  - MET predictions/models
  - Activity predictions/models
- Reviews
- Application

MET Regression Method

1. Research different Regression models

2. Configure the chosen models in almost the same way

3. Evaluated both model results to pick a model for each activity

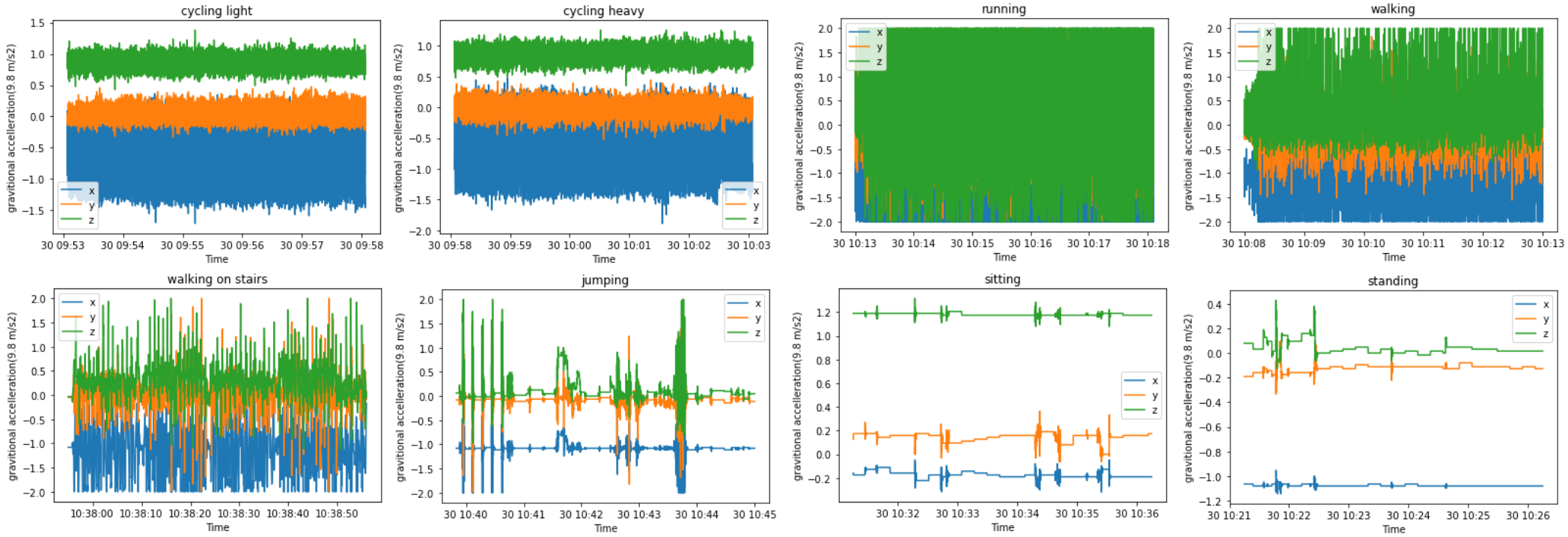
MET Regression Results

Activity	Random Forest	Random Forest Test		XGBoost	XGBoost Test	
	R Squared Score Cross Validation	R^2	MSE	R Squared Score Cross Validation	R^2	MSE
Walking	0.40 (+/- 0.15)	0.40	0.65	0.32 (+/- 0.12)	0.29	0.70
Running	0.60 (+/- 0.20)	0.58	1.57	0.63 (+/- 0.15)	0.45	1.80
Cycling Light	0.41 (+/- 0.28)	-0.70	1.78	0.41 (+/- 0.23)	-1.03	1.95
Cycling Heavy	-0.39 (+/- 1.18)	0.02	1.67	0.12 (+/- 0.59)	-0.08	1.23
Standing	-0.54 (+/- 0.55)	-0.02	0.38	-0.85 (+/- 0.78)	0.36	0.30
Sitting	-0.23 (+/- 0.41)	-0.05	0.37	-0.17 (+/- 0.42)	-0.24	0.40
Note: Model comparison between Random Forest and XGBoost						

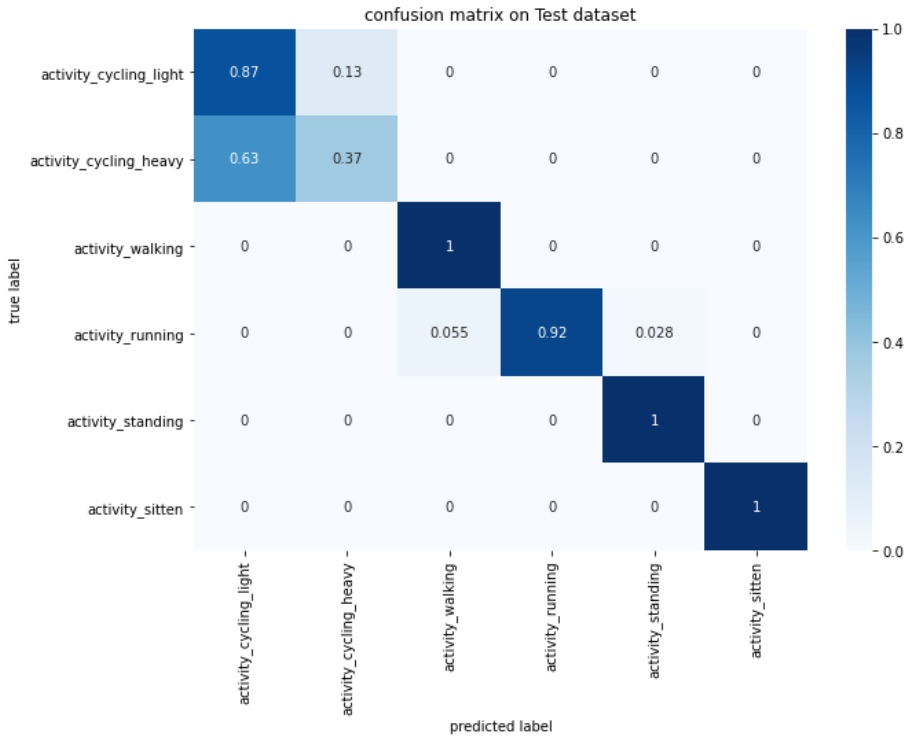
Activity	Chosen Model
Walking	Random Forest
Running	Random Forest
Cycling Light	Random Forest
Cycling Heavy	XGBoost

Activity recognition

- Resampled data to certain segment size
- Features:
  - Mean of X, Y and Z
  - Standard deviation of X, Y and Z
- Created model for each segment size
  - Segment range 1 tot 14 seconds with steps 0.1
  - Looked at accuracy from cross validation on the model
  - Looked for the best number of trees



Activity recognition results



Validation

Segment size	Number of trees	Accuracy	Precision(micro)	Recall(Miicro)
12.1	93	0.96	0.96	0.96
8.9	171	0.94	0.94	0.94
7.0	203	0.95	0.95	0.95

N-fold cross-validation

Segment size	Number of trees	Accuracy	Precision(micro)	Recall(Miicro)
12.1	93	0.82 ± 0.05	0.84 ± 0.04	0.82 ± 0.05
8.9	171	0.82 ± 0.04	0.84 ± 0.04	0.82 ± 0.04
7.0	203	0.83 ± 0.04	0.84 ± 0.04	0.83 ± 0.04

Test

Segment size	Number of trees	Accuracy	Precision(micro)	Recall(Miicro)
12.1	93	0.85	0.85	0.85
8.9	171	0.86	0.86	0.86
7.0	203	0.84	0.84	0.84

Application

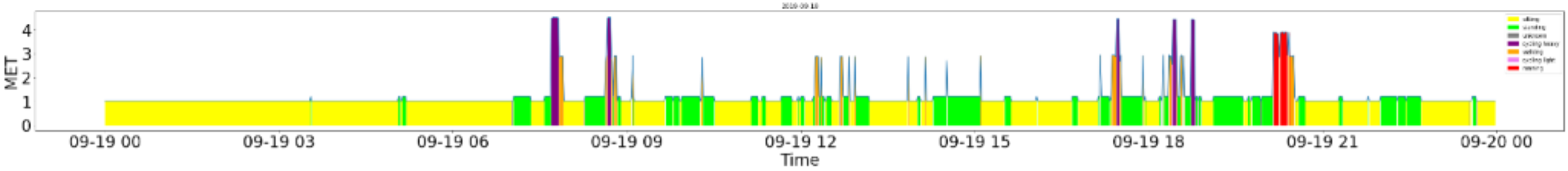
- Application combines classification and regression models
- Firstly, the activities are being identified
- Secondly, the MET-values per activity is predicted
- Thirdly, the rows with MET-values lower than 3 (threshold for moderate intense PA) are being filtered out
- Resulting in a dataset containing rows with at least 3 MET
- Each row represents a minute
- Amount of minutes performing at least moderate intense PA is equal to the amount of rows in the dataset

Application Results

- Activpal data of 7 days from 28 respondents analysed
- Results:
  - Mean(rounded): 469 minutes
  - Std(rounded): 225 minutes
  - 27 of 28 respondents has met their weekly moderate activity recommendations
- The trustworthy of these results depends on accuracy of the MET regression models

Application Results

Activity	Colour
Sitting	Yellow
Standing	Lime
Walking	Orange
Running	Red
Cycling light	Violet
Cycling heavy	Purple



Conclusion

- The activity classification model
  - Performs generally well
  - Has a bias for standing
  - Has difficulties with cycling light and heavy

- The MET regression models
  - Not good enough
  - Models are overfitting
  - Models handle unseen data poorly
  - Mainly because there is too little data

- Therefore, the models can not accurately predict whether a person has performed at least 150 minutes of moderate intense PA

Advice

- More respondenten
- More features
- Furture research into MET prediction

Questions?