Methods

Because the data I collected does not provide enough information, and I would need NLP to proceed my goals which I am half the way. I used data from lab 4.2. It is a dataset about data of motivation trackers. First, the device type is collected as dummy variables for future use.

The attribute activity is categorical so it could be used as y dataset for training, also for labeling. The x data would be the rest data that is cleaned. Then the datasets are separated to test and train for prediction. SVM with Linear, Polynomial, BRF and Sigmoid kernels are applied for predicting.

After predicting, the confusion matrix shows how both test and train are distributed compared to the prediction results.

Class distribution

From looking at the Linear SVM result, most of the users are running 5 METs, but it returned low F1 scores, which means the precision is not as accurate.

The rest of the models shows most of the users are lying with higher recall scores. The polynomial works for this dataset the best, by looking at the scores.

Baseline model for comparison (train to test)

Linear

	precision	recall	f1-score	support
Lying	0.33	0.31	0.32	1096
Running 7 METs	0.07	0.00	0.00	783
Running 5 METs	0.31	0.49	0.38	809
Running 3 METs	0.39	0.48	0.43	868
Sitting	0.27	0.09	0.13	703
Self Pace walk	0.19	0.35	0.24	752
accuracy			0.29	5011
macro avg	0.26	0.29	0.25	5011
weighted avg	0.27	0.29	0.26	5011

	precision	recall	f1-score	support
Lying	0.32	0.29	0.30	283
Running 7 METs	0.00	0.00	0.00	167
Running 5 METs	0.29	0.50	0.37	193
Running 3 METs	0.42	0.47	0.44	246
Sitting	0.22	0.05	0.09	186
Self Pace walk	0.18	0.35	0.24	178
accuracy			0.29	1253
macro avg	0.24	0.28	0.24	1253
weighted avg	0.26	0.29	0.26	1253

Polynomial

	precision	recall	f1-score	support
Lying	0.22	1.00	0.36	1096
Running 7 METs	0.31	0.01	0.02	783
Running 5 METs	0.00	0.00	0.00	809
Running 3 METs	0.00	0.00	0.00	868
Sitting	0.00	0.00	0.00	703
Self Pace walk	0.00	0.00	0.00	752
accuracy			0.22	5011
macro avg	0.09	0.17	0.06	5011
weighted avg	0.10	0.22	0.08	5011

	precision	recall	f1-score	support
Lying	0.22	0.99	0.37	283
Running 7 METs	0.00	0.00	0.00	167
Running 5 METs	0.00	0.00	0.00	193
Running 3 METs	0.00	0.00	0.00	246
Sitting	0.00	0.00	0.00	186
Self Pace walk	0.00	0.00	0.00	178
accuracy			0.22	1253
macro avg	0.04	0.16	0.06	1253
weighted avg	0.05	0.22	0.08	1253

BRF

	precision	recall	f1-score	support
Lying	0.22	0.97	0.36	1096
Running 7 METs	0.55	0.01	0.02	783
Running 5 METs	0.67	0.00	0.00	809
Running 3 METs	0.30	0.08	0.12	868
Sitting	0.47	0.01	0.02	703
Self Pace walk	0.17	0.00	0.01	752
accuracy			0.23	5011
macro avg	0.40	0.18	0.09	5011
weighted avg	0.39	0.23	0.11	5011

	precision	recall	f1-score	support
Lying	0.23	0.94	0.36	283
Running 7 METs	0.00	0.00	0.00	167
Running 5 METs	0.00	0.00	0.00	193
Running 3 METs	0.30	0.07	0.11	246
Sitting	0.60	0.02	0.03	186
Self Pace walk	0.00	0.00	0.00	178
accuracy			0.23	1253
macro avg	0.19	0.17	0.08	1253
weighted avg	0.20	0.23	0.11	1253

Sigmoid

	precision	recall	f1-score	support
Lying	0.21	0.54	0.30	1096
Running 7 METs	0.15	0.01	0.01	783
Running 5 METs	0.18	0.38	0.25	809
Running 3 METs	0.19	0.08	0.11	868
Sitting	0.18	0.02	0.04	703
Self Pace walk	0.05	0.00	0.00	752
accuracy			0.20	5011
macro avg	0.16	0.17	0.12	5011
weighted avg	0.17	0.20	0.13	5011

	precision	recall	f1-score	support
Lying	0.19	0.47	0.27	283
Running 7 METs	0.14	0.01	0.01	167
Running 5 METs	0.15	0.35	0.21	193
Running 3 METs	0.17	0.06	0.09	246
Sitting	0.18	0.03	0.05	186
Self Pace walk	0.14	0.01	0.01	178
accuracy			0.18	1253
macro avg	0.16	0.15	0.11	1253
weighted avg	0.17	0.18	0.12	1253

Final results

The polynomial SVM has the highest f1 score. By looking at that, the most frequent activity is lying for the users.

Conclusions:

Users tend to lying in the most of cases. The secondary frequent activity is run 3 METs.