HAR

(Human Activity Recognition)

(C01 – C09)

Dataset Description

- Number of attributes:
  - 3 (X, Y, Z) - Accelerometer
  - 3 (X, Y, Z) – Gyroscope

- Missing attribute values:
  - None

- Sampling Rate :
  - 100 Hz

Sample_Train.csv

Acc_X	Acc_Y	Acc_Z	Gyro_X	Gyro_Y	Gyro_Z
0.58507	0.958694	-0.91408	0.508786	0.003394	0.024296
0.554663	1.122004	-0.84205	0.543408	0.002409	0.034015
0.527227	1.208196	-0.75273	0.531366	-0.00668	0.071343
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…

3000 Rows ++

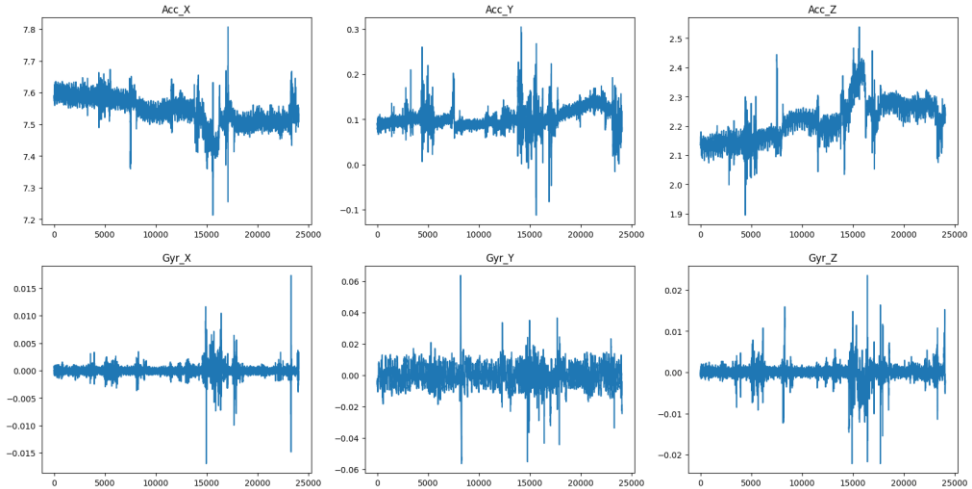
Sample_Test.csv

500 Rows

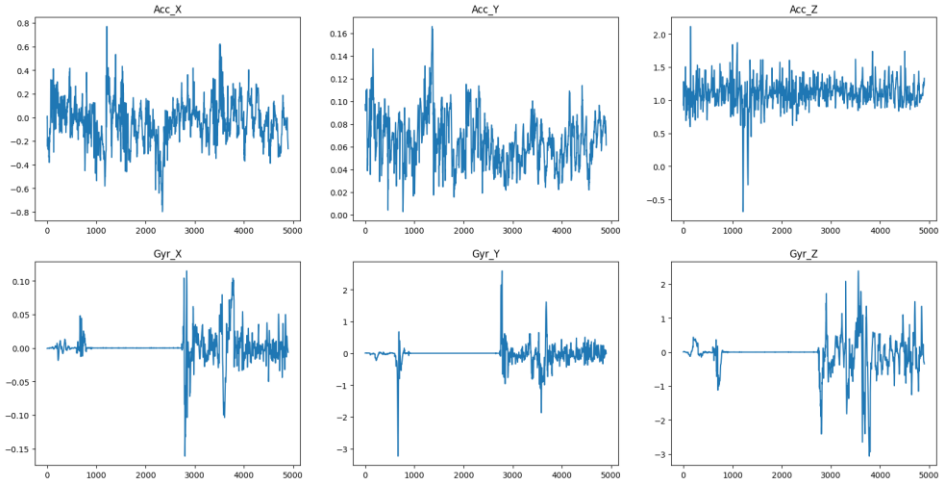
Acc_X	Acc_Y	Acc_Z	Gyro_X	Gyro_Y	Gyro_Z
0.58507	0.958694	-0.91408	0.508786	0.003394	0.024296
0.554663	1.122004	-0.84205	0.543408	0.002409	0.034015
0.527227	1.208196	-0.75273	0.531366	-0.00668	0.071343
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…

EDA

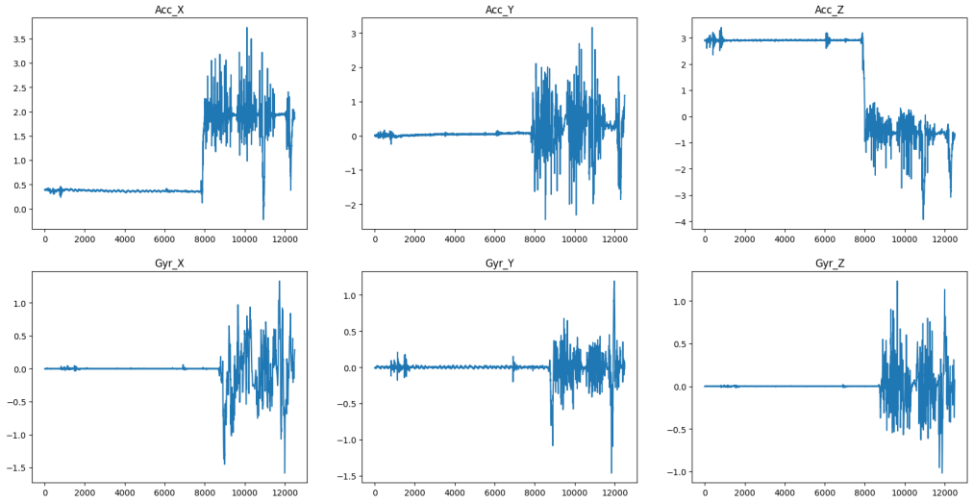
C01



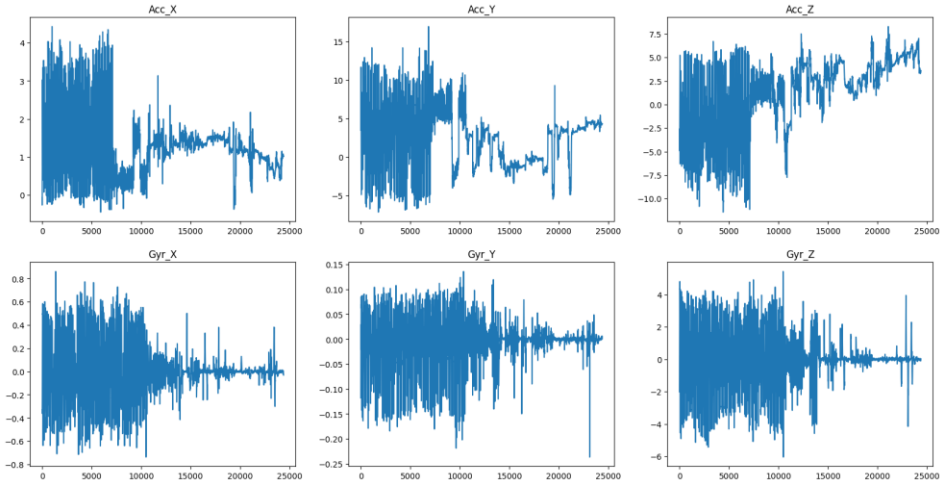
C02



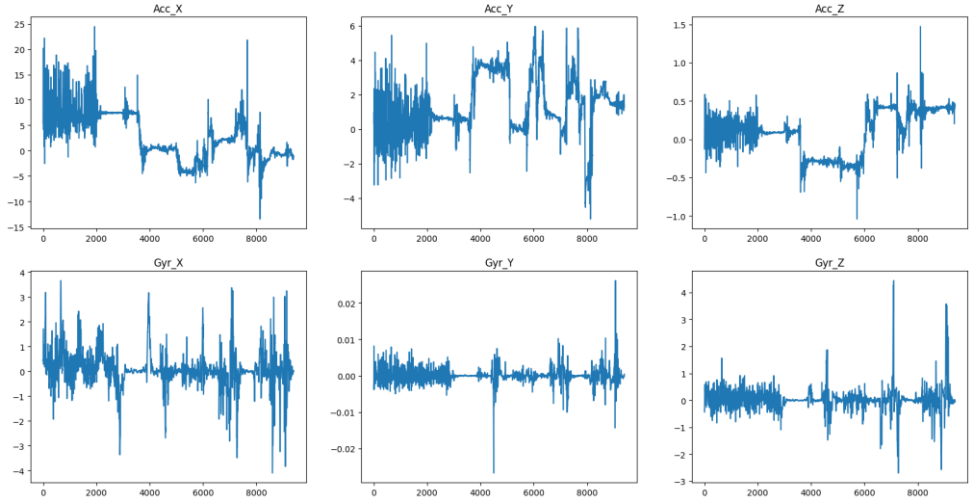
C03



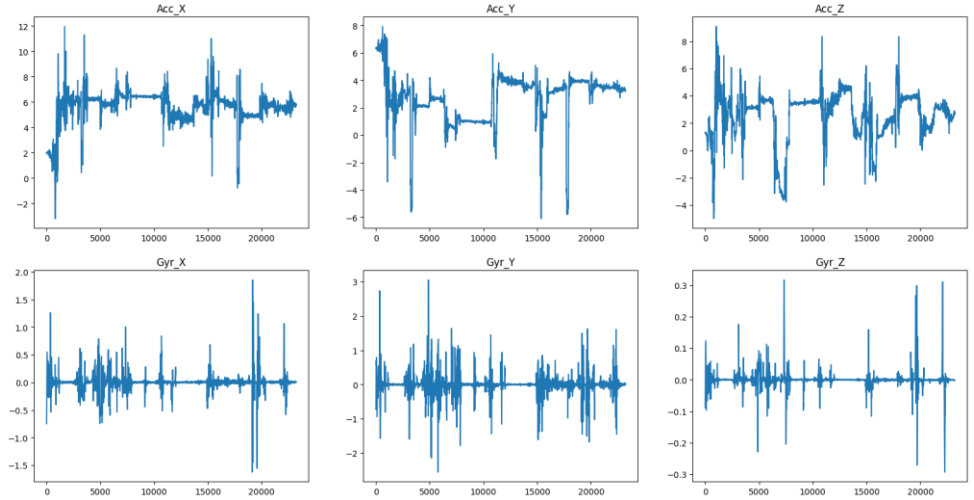
C04



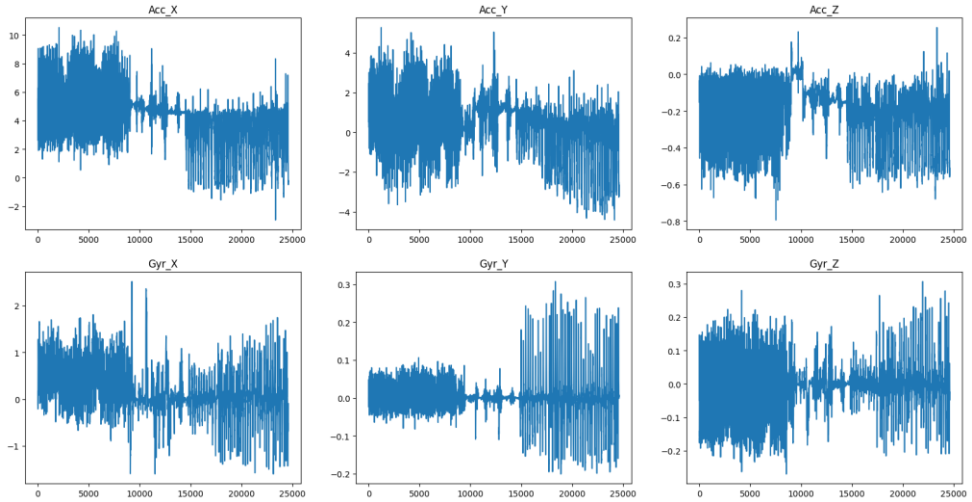
C05



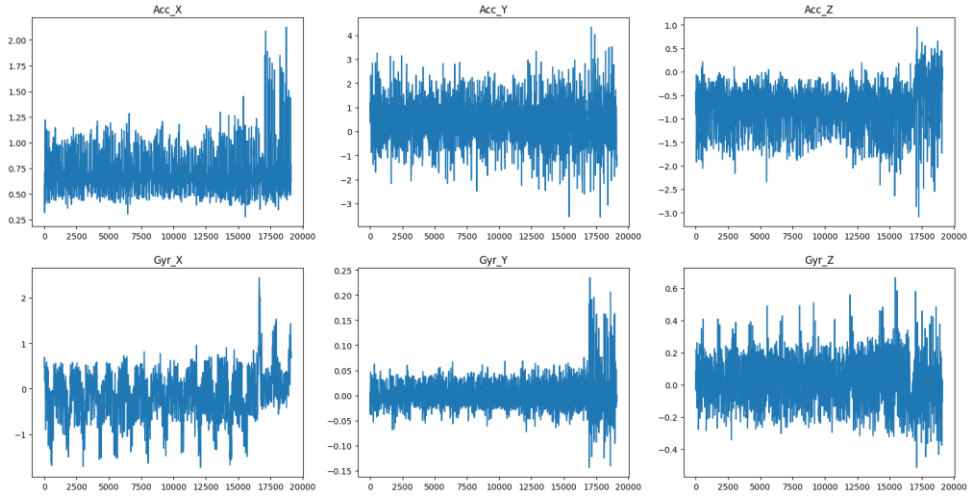
C06



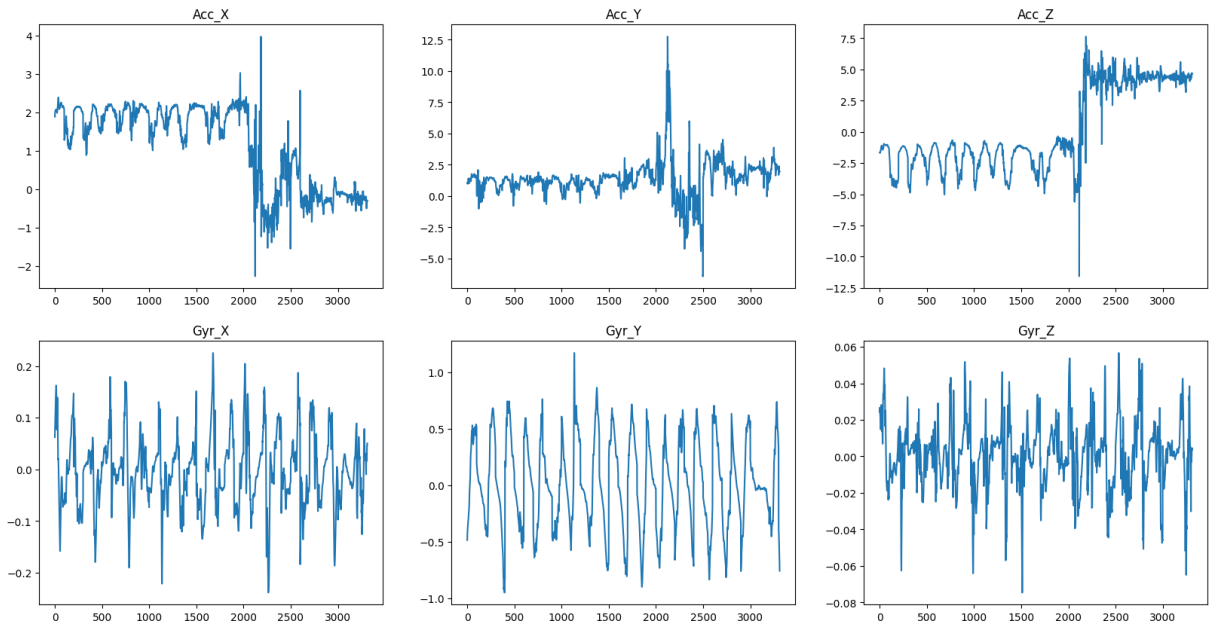
C07

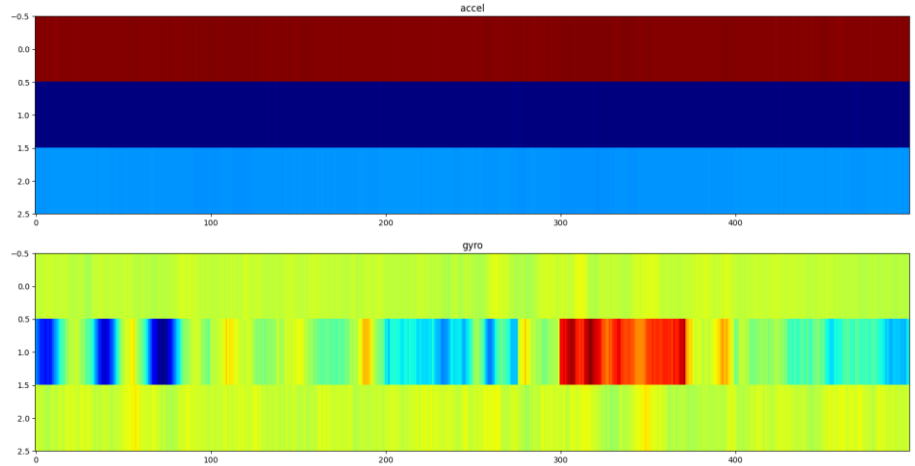


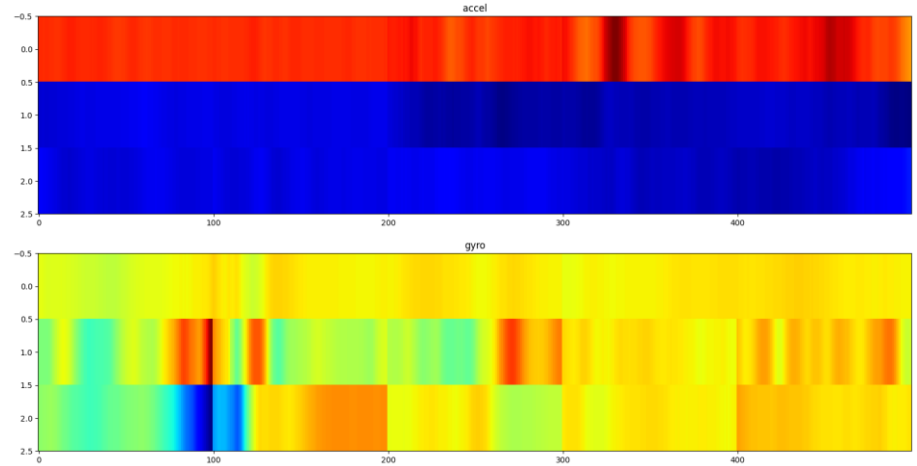
C08

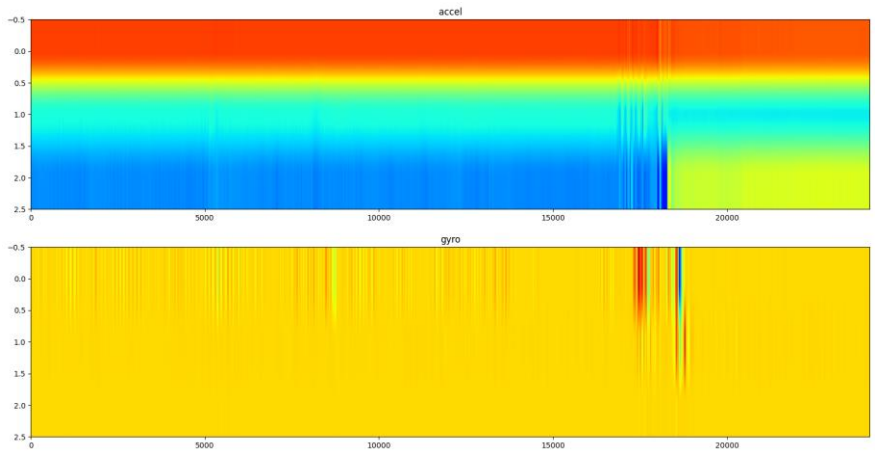


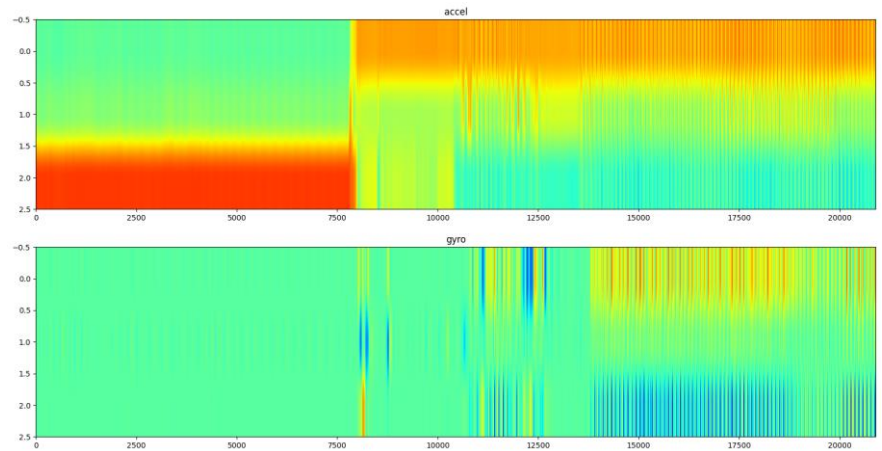
C09

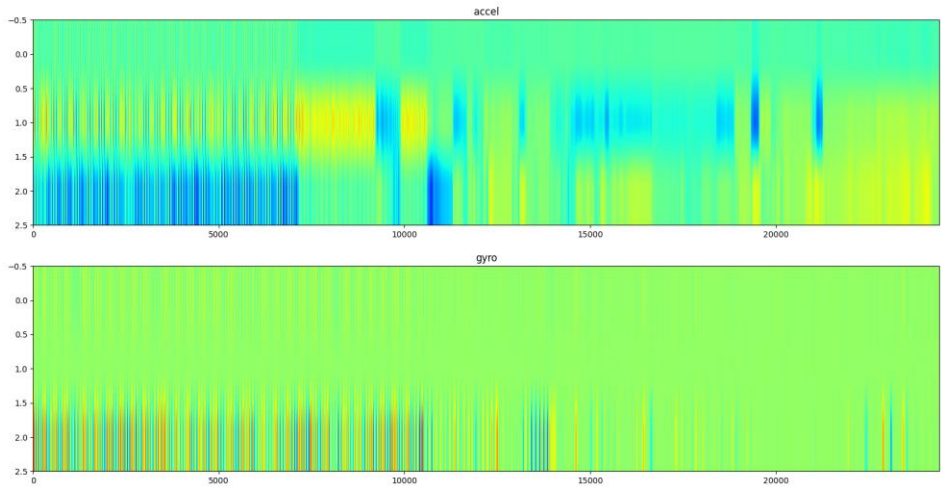


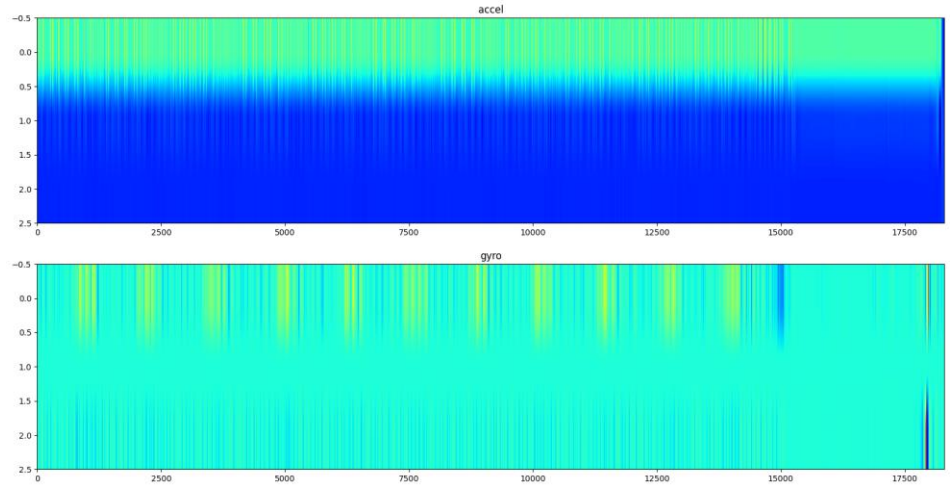

accel
gyro

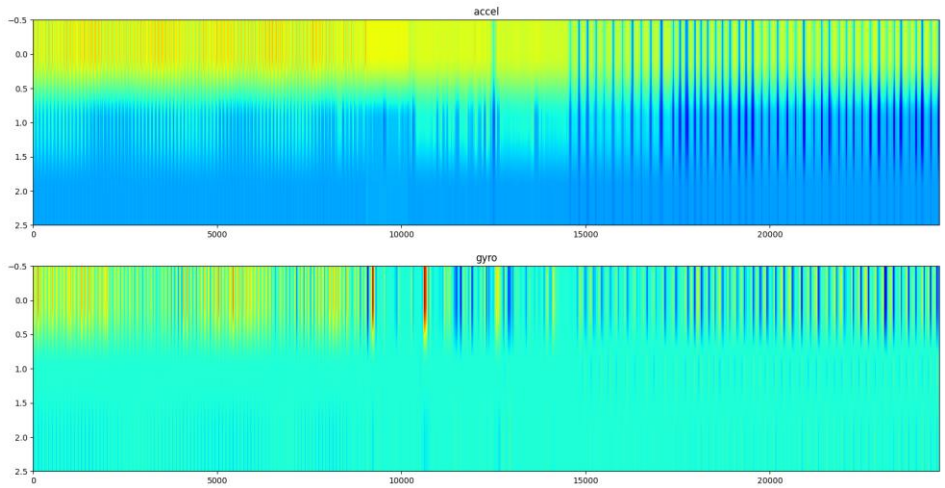

accel
gyro

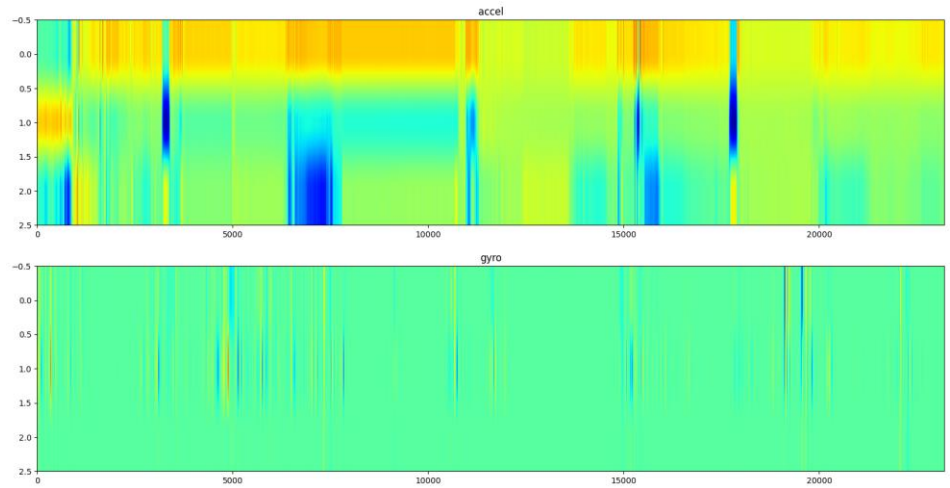

accel
gyro

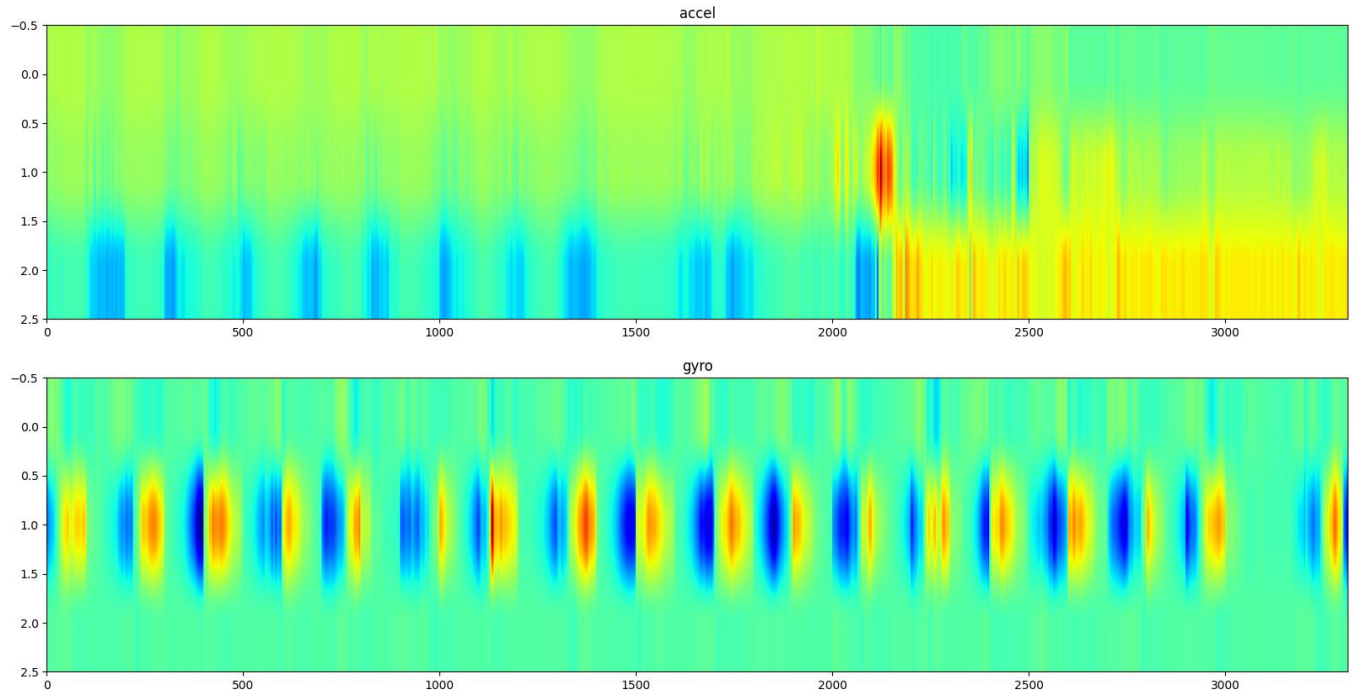

accel
gyro


accel
gyro


accel
gyro


accel
gyro


accel
gyro


accel
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
3000
gyro
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
3000

Pipeline #1
Lightgbm+caret experiment

Add Features

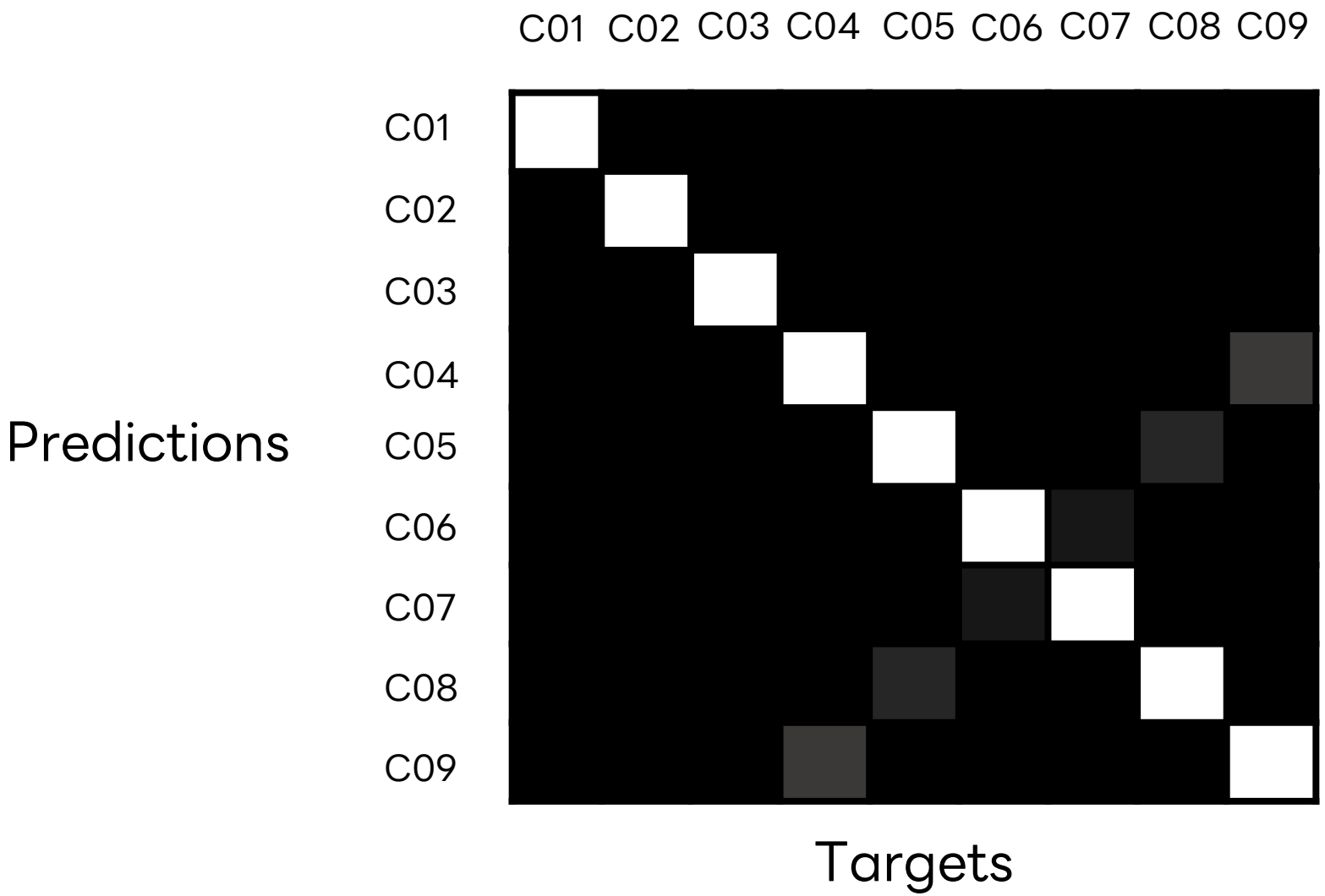
- Mean
- Variance
- Lagged+Downsample
- Multiple predictions per cell

Lagged Features

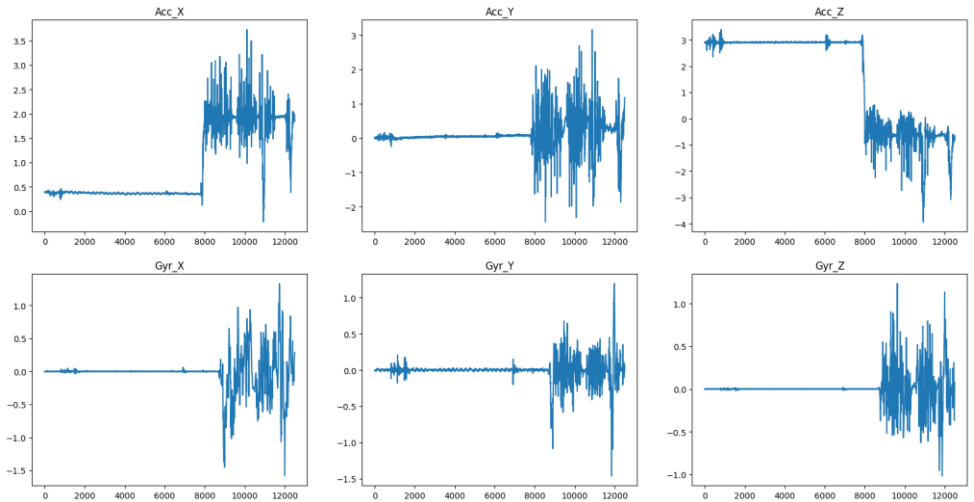
# index	# 0	# Acc_X1	# Acc_X2	# Acc_X3	# Acc_X4	# Acc_X5	# Acc_X6
	M 647203 (>99%) Distinct: 9 (<1%)	Missing: 1 (<1%) Distin... 69204 (11%)	Missing: 2 (<1%) Distin... 69204 (11%)	Missing: 3 (<1%) Disti... 69204 (11%)	Missing: 4 (<1%) Disti... 69204 (11%)	Missing: 5 (<1%) Disti... 69204 (11%)	Missing: 6 (<1%) Distin... 69204 (11%)
	Min 0.0 Max 8.0	Min -13.4... Max 38.8...	Min -13.46... Max 38.8...	Min -13.4... Max 38....	Min -13.4... Max 38....	Min -13.4... Max 38.8...	Min -13.46... Max 38.8...
0	0.0	*Missing value*	*Missing value*	*Missing value*	*Missing value*	*Missing value*	*Missing value*
1	0.0	7.909174341745075	*Missing value*	*Missing value*	*Missing value*	*Missing value*	*Missing value*
2	0.0	7.921172913243259	7.909174341745075	*Missing value*	*Missing value*	*Missing value*	*Missing value*
3	0.0	7.93517194271138	7.921172913243259	7.909174341745075	*Missing value*	*Missing value*	*Missing value*
4	0.0	7.959169085707749	7.93517194271138	7.921172913243259	7.909174341745075	*Missing value*	*Missing value*
5	0.0	7.969168034500284	7.959169085707749	7.93517194271138	7.921172913243259	7.909174341745075	*Missing value*
6	0.0	7.965167953824697	7.969168034500284	7.959169085707749	7.93517194271138	7.921172913243259	7.909174341745075
7	0.0	7.961168708413398	7.965167953824697	7.969168034500284	7.959169085707749	7.93517194271138	7.921172913243259
8	0.0	7.943170433533979	7.961168708413398	7.965167953824697	7.969168034500284	7.959169085707749	7.93517194271138


submission.csv
Complete · 402829-ธราธร · 20h ago · Light gbm + 10 lag + mean +var
0.93018
0.92402

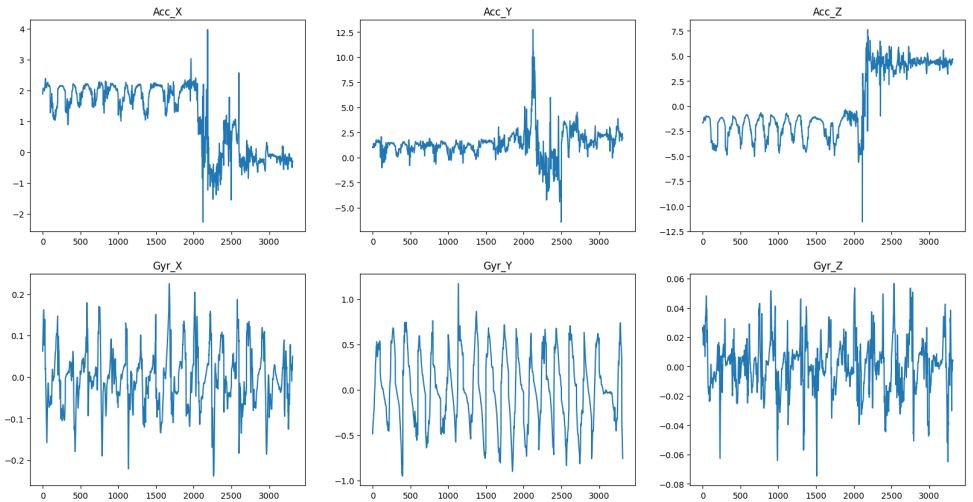
Confusion Matrix



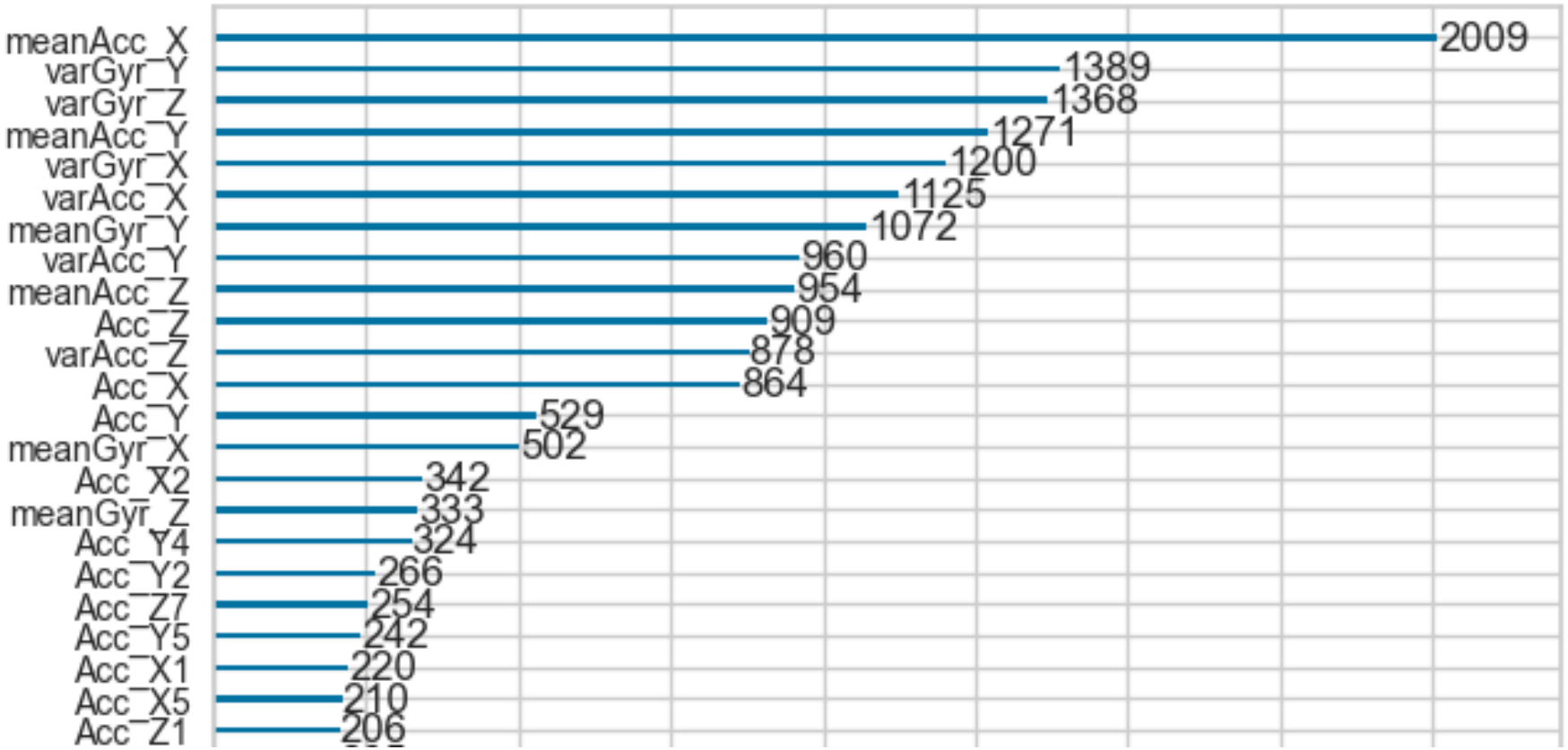
C03



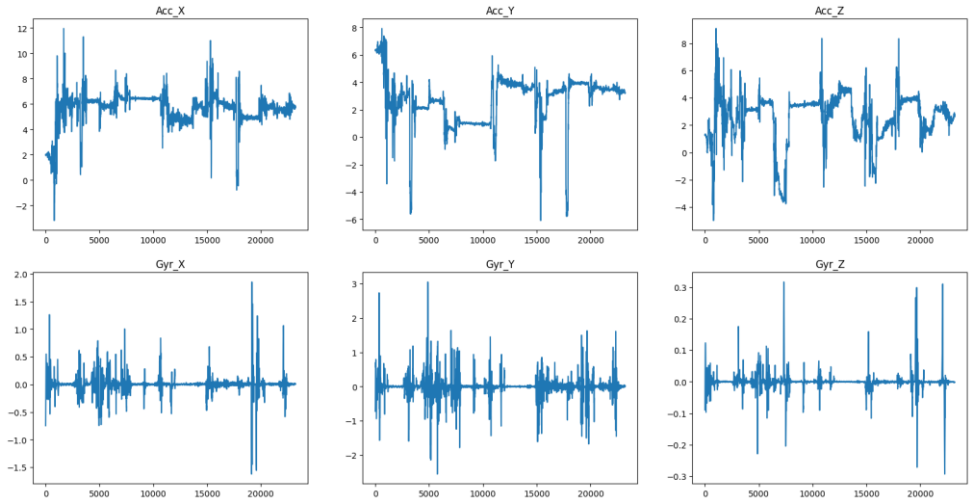
C09



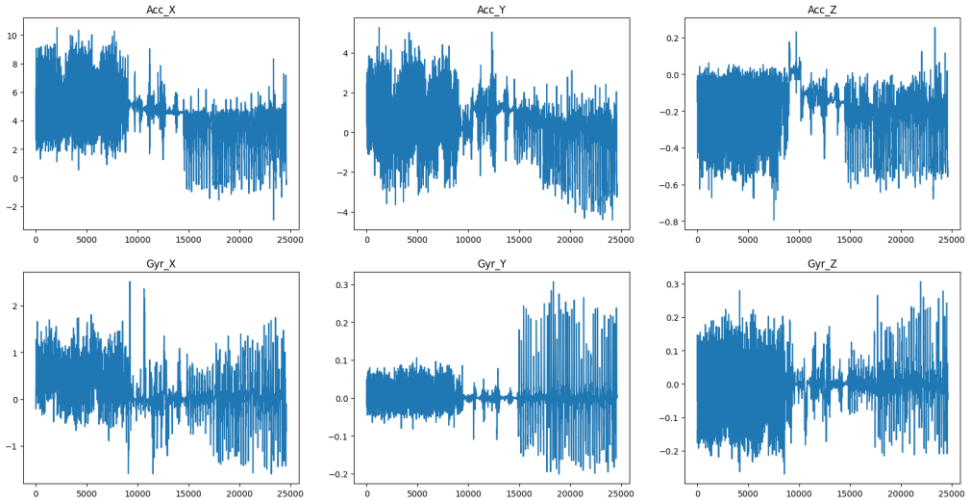
Feature Importance



C06



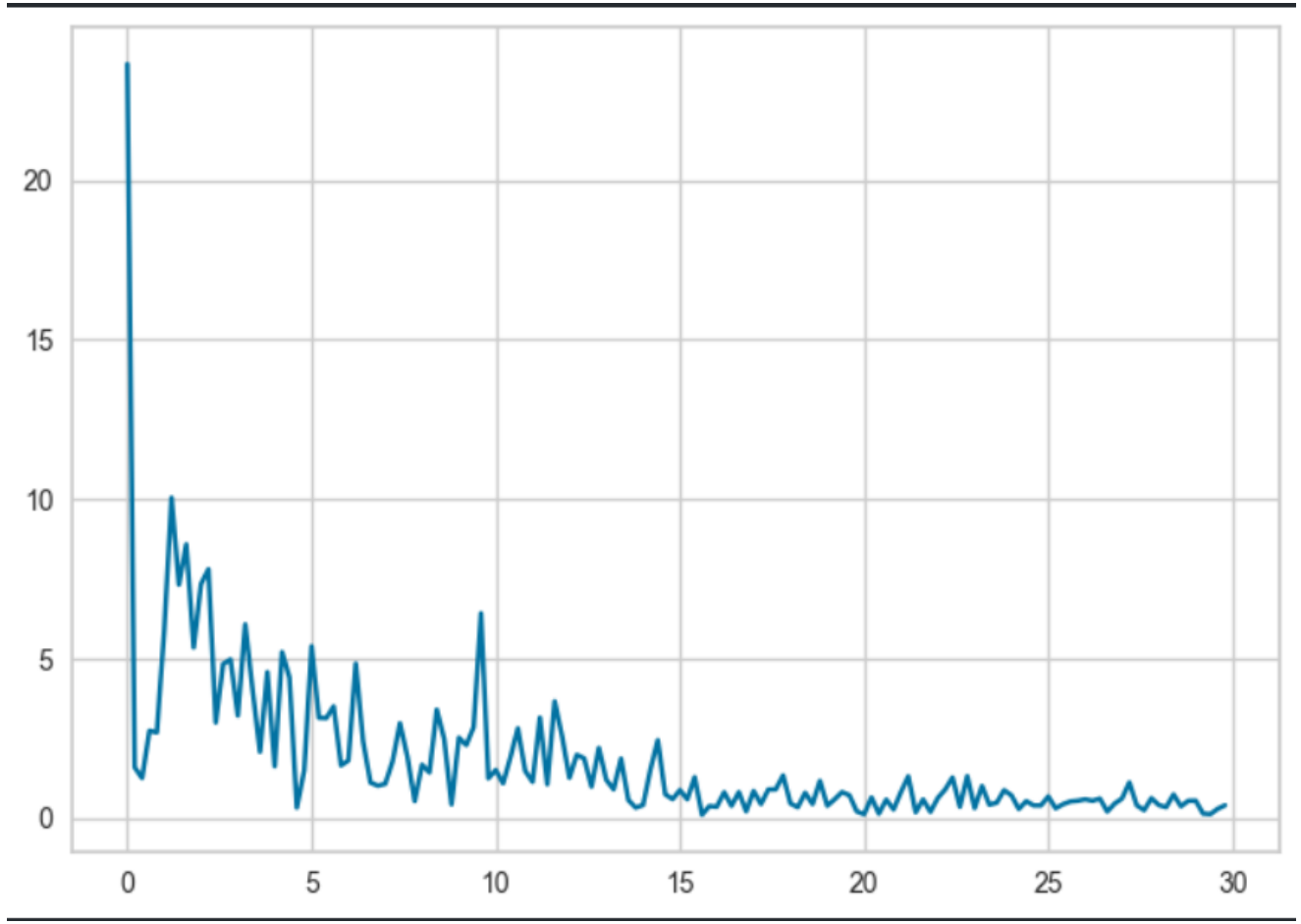
C07



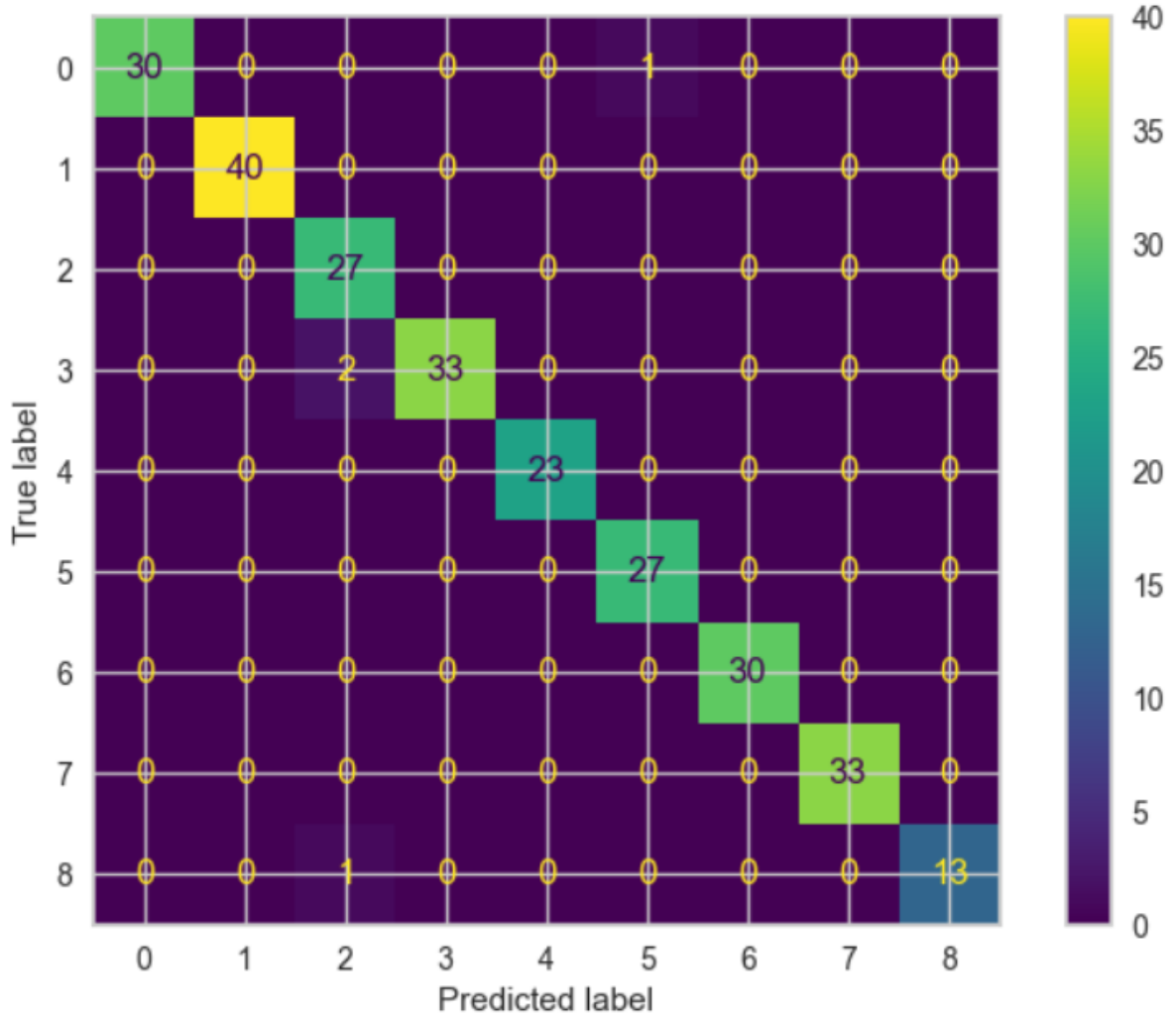
Pipeline#2
LightGBM / Catboost

Add Features

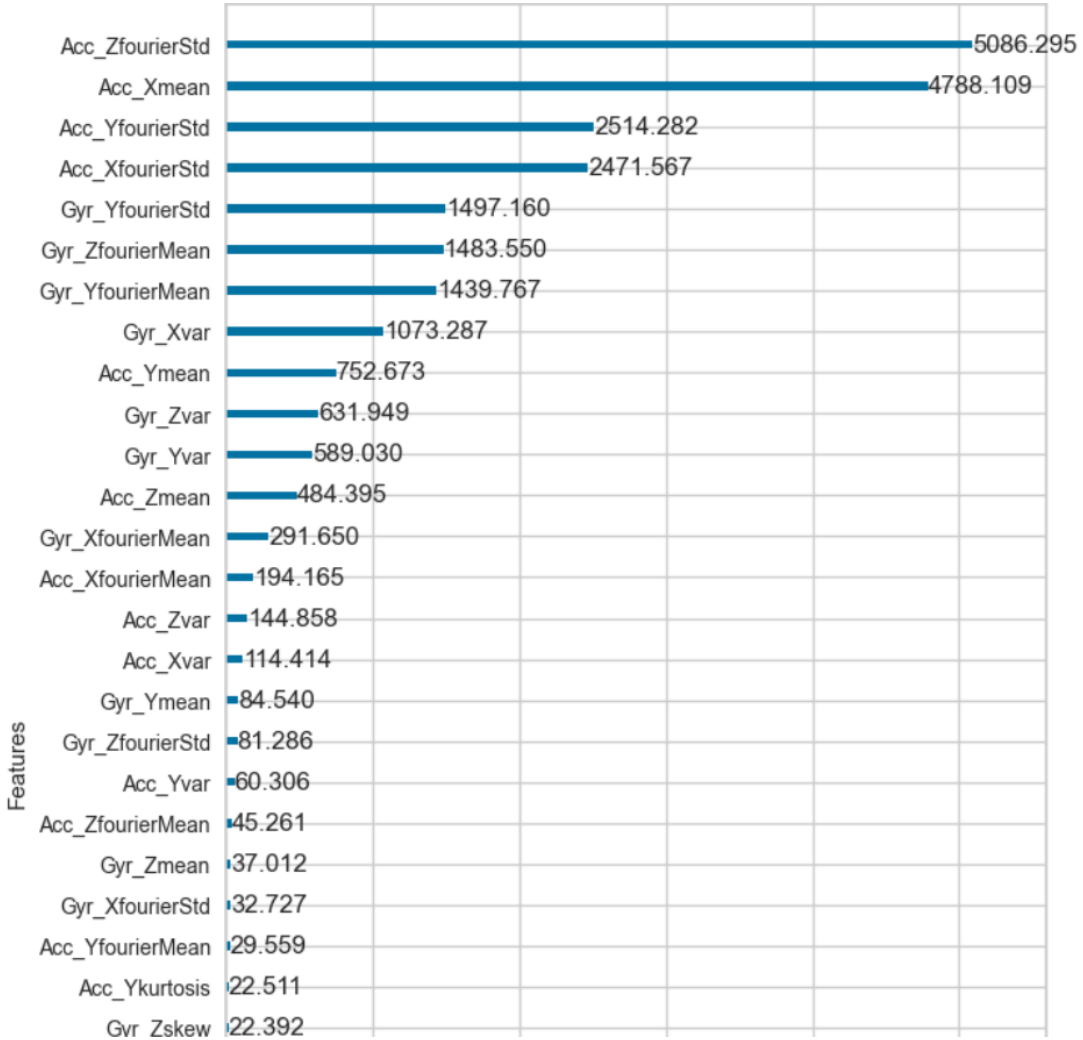
- Fourier Transform
- Kurtosis/Skew
- Stratified 5-Fold

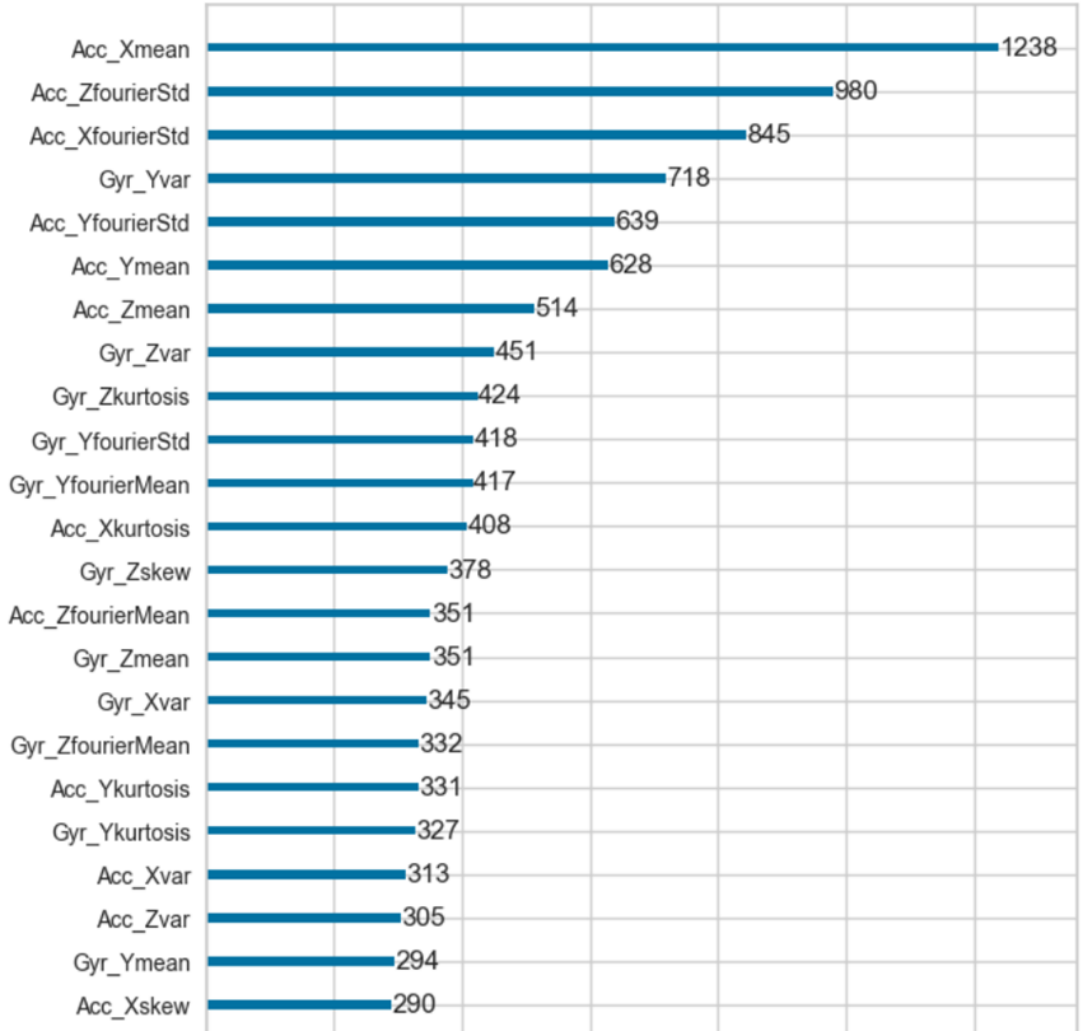


Confusion Matrix



Feature Importance(gain/split)



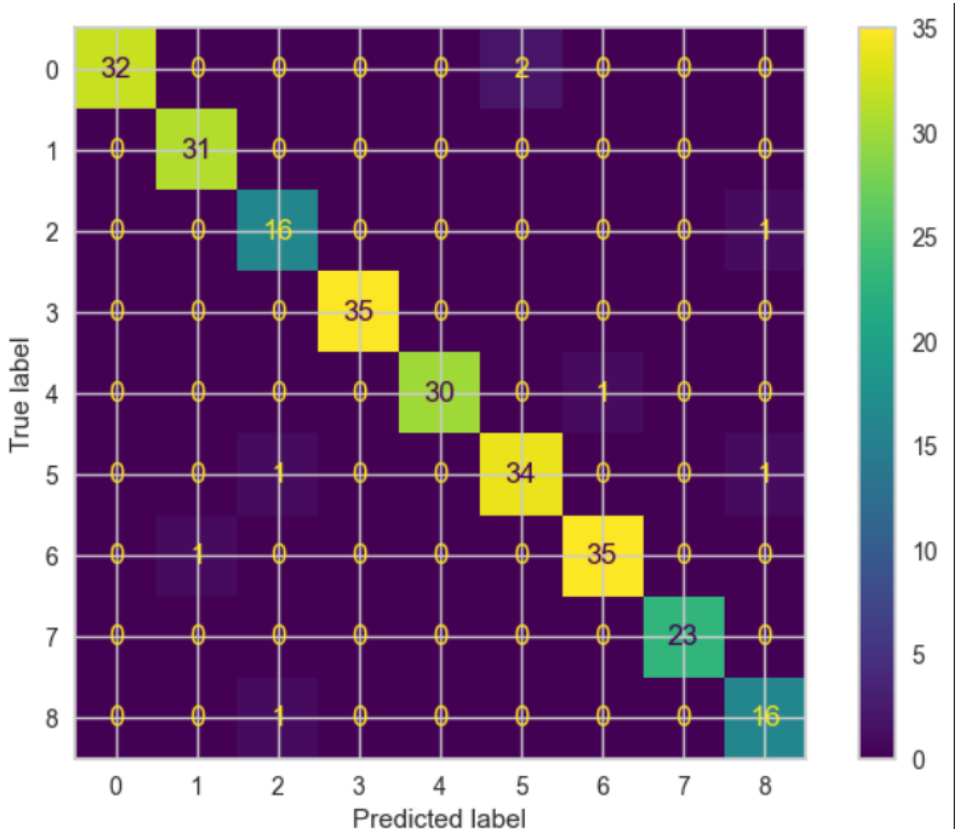


Results: inference time 1+0.8 second

Submission and Description	Private Score	Public Score	Selected
submission_lgbm_5fold.csv Complete · 400256-ธนพัฒน์ · 5h ago	0.97125	0.99178	☑
submission_catboost_5fold.csv Complete · 400256-ธนพัฒน์ · 5h ago · Before	0.98562	0.98767	☑

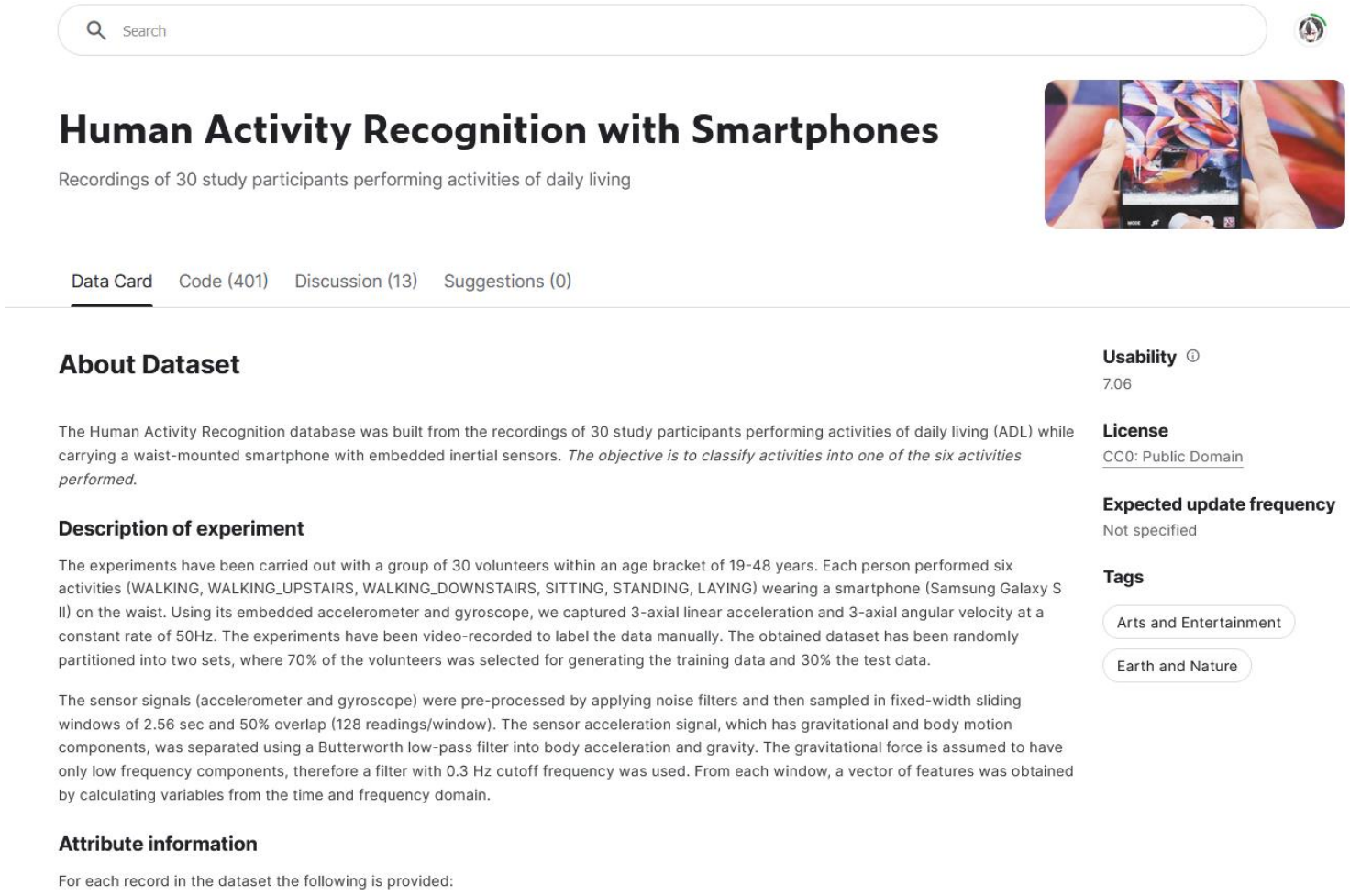
Same confusion matrix?

✓	**submissionCleanedData.csv** Complete · 402829-ssтss · 11h ago · amogus	**0.86447**	**0.88706**
✓	**submissionFixed.csv** Complete · 402829-ssтss · 17h ago · fixed rolling window	**0.93634**	**0.94866**
✓	**submission.csv** Complete · 402829-ssтss · 18h ago · Light gbm + 10 lag + mean	**0.93018**	**0.92402**



Future plans

- Data oversampling
- Training on extra datasets



Future plans

- More visualization (Sankey-diagram)
- More Feature engineering(entropy, PSD, energy, jerk, Freq. binning)
- Optuna implementation
- Experiment with filters
- Experiment with ViTs

Q & A