

A stylized, dark grey smartphone graphic is positioned on the left side of the slide. It features a large screen area, a small notch at the top, and a circular home button at the bottom. The phone is tilted slightly to the right.

Human Activity Recognition with Smartphones

A solid dark grey rectangular box is located below the main title, containing the text 'KELOMPOK 2'.

KELOMPOK 2

Meet Our Team

KELOMPOK 2



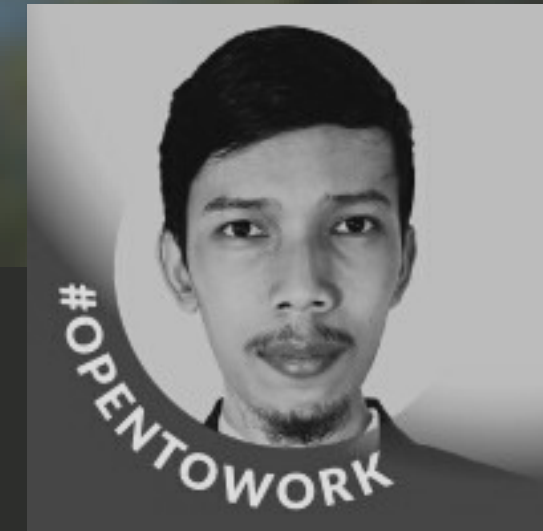
RIZKI RACHMADI

Project Leader



**MUHAMMAD DWI
HARYANTO**

Programmer



**MUHAMMAD ADIB
MAJDI**

Team support

Materi yang akan kami diskusikan

TOPIC OUTLINE



Business Understanding
Analytic Approach
DATA REQUIREMENT
DATA COLLECTION
DATA UNDERSTANDING
DATA PREPARATION
MODELING
EVALUATION



Business Understanding



DESKRIPSI

Dataset Pengenalan Aktivitas Manusia diambil dari rekaman 30 peserta penelitian yang melakukan aktivitas kehidupan sehari-hari sambil membawa smartphone yang dipasang di pinggang dengan sensor inersia (akselerometer dan giroskop) yang terdapat di smartphone tersebut.

TUJUAN

Tujuannya adalah untuk mengklasifikasikan kegiatan menjadi salah satu dari enam kegiatan yang dilakukan (WALKING, WALKINGUPSTAIRS, WALKINGDOWNSTAIRS, SITTING, STANDING, LAYING).

SUMBER

<https://www.kaggle.com/uciml/human-activity-recognition-with-smartphones>

Business Understanding

	tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z	tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z	tBodyAcc-mad()-X	tBodyAcc-mad()-Y	tBodyAcc-mad()-Z
0	0.288585	-0.020294	-0.132905	-0.995279	-0.983111	-0.913526	-0.995112	-0.983185	-0.983185
1	0.278419	-0.016411	-0.123520	-0.998245	-0.975300	-0.960322	-0.998807	-0.974914	-0.974914
2	0.279653	-0.019467	-0.113462	-0.995380	-0.967187	-0.978944	-0.996520	-0.963668	-0.963668
3	0.279174	-0.026201	-0.123283	-0.996091	-0.983403	-0.990675	-0.997099	-0.982750	-0.982750
4	0.276629	-0.016570	-0.115362	-0.998139	-0.980817	-0.990482	-0.998321	-0.979672	-0.979672

5 rows x 563 columns

```
1 df_test.columns

Index(['tBodyAcc-mean()-X', 'tBodyAcc-mean()-Y', 'tBodyAcc-mean()-Z',
      'tBodyAcc-std()-X', 'tBodyAcc-std()-Y', 'tBodyAcc-std()-Z',
      'tBodyAcc-mad()-X', 'tBodyAcc-mad()-Y', 'tBodyAcc-mad()-Z',
      'tBodyAcc-max()-X',
      ...,
      'fBodyBodyGyroJerkMag-kurtosis()', 'angle(tBodyAccMean,gravity)',
      'angle(tBodyAccJerkMean,gravityMean)',
      'angle(tBodyGyroMean,gravityMean)',
      'angle(tBodyGyroJerkMean,gravityMean)', 'angle(X,gravityMean)',
      'angle(Y,gravityMean)', 'angle(Z,gravityMean)', 'subject', 'Activity'],
      dtype='object', length=563)
```

Attribute

- Akselerasi triaxial dari accelerometer.
- Kecepatan Triaxial Angular dari gyroscope.
- Fitur vektor dengan menghitung variabel dari domain waktu dan frekuensi.
- Label setiap kegiatan (activity).
- An identifier of the subject who carried out the experiment.

Rumusan Masalah

- Aktivitas apa yang paling sering dilakukan oleh manusia?
- Bagaimana cara memisahkan antara aktivitas statis dan dinamis?
- Bagaimana cara mengolah data hasil sensor pada smartphone untuk dapat melakukan klasifikasi aktivitas manusia?

Analytic Approach

MLP

(Multilayer Perceptron)

RFC

(Random Forests
Classifier)

SVM

(Support Vector Machine)

NAIVE BAYES
CLASSIFIER

ENSEMBLE

DATA REQUIREMENT

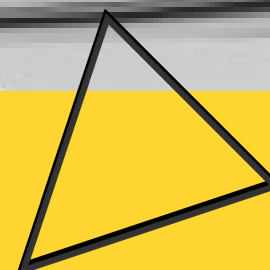
- Label aktivitas manusia
- Nilai perubahan guncangan pada sensor accelerometer
- Nilai sudut antara X-axis dan arah gravitasi pada sensor gyrometer



A stylized yellow smartphone graphic is positioned on the left side of the slide. It features a yellow border and a dark screen area. At the top, there is a small notch-like detail. At the bottom, there is a circular cutout representing a camera or speaker. The phone is tilted slightly to the right.

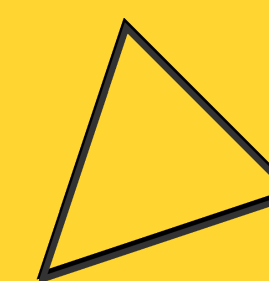
DATA COLLECTION

- Dataset berjudul "Human Activity Recognition with Smartphones":
<https://www.kaggle.com/uciml/human-activity-recognition-with-smartphones>
- Deskripsi dataset:
https://rpubs.com/Jb_2823/55939



DATA UNDERSTANDING

- Nilai 0 pada dataset bukanlah missing value, melainkan nilai dari pembacaan sensor.
- Data label Activity terdapat 6 jenis, yaitu (WALKING, WALKINGUPSTAIRS, WALKINGDOWNSTAIRS, SITTING, STANDING, LAYING).
- tBodyAccMag-mean() -> Membaca goncangan tubuh.
- angle(X,gravityMean) - > Membaca sudut antara X-axis dan gravityMean



DATA PREPARATION

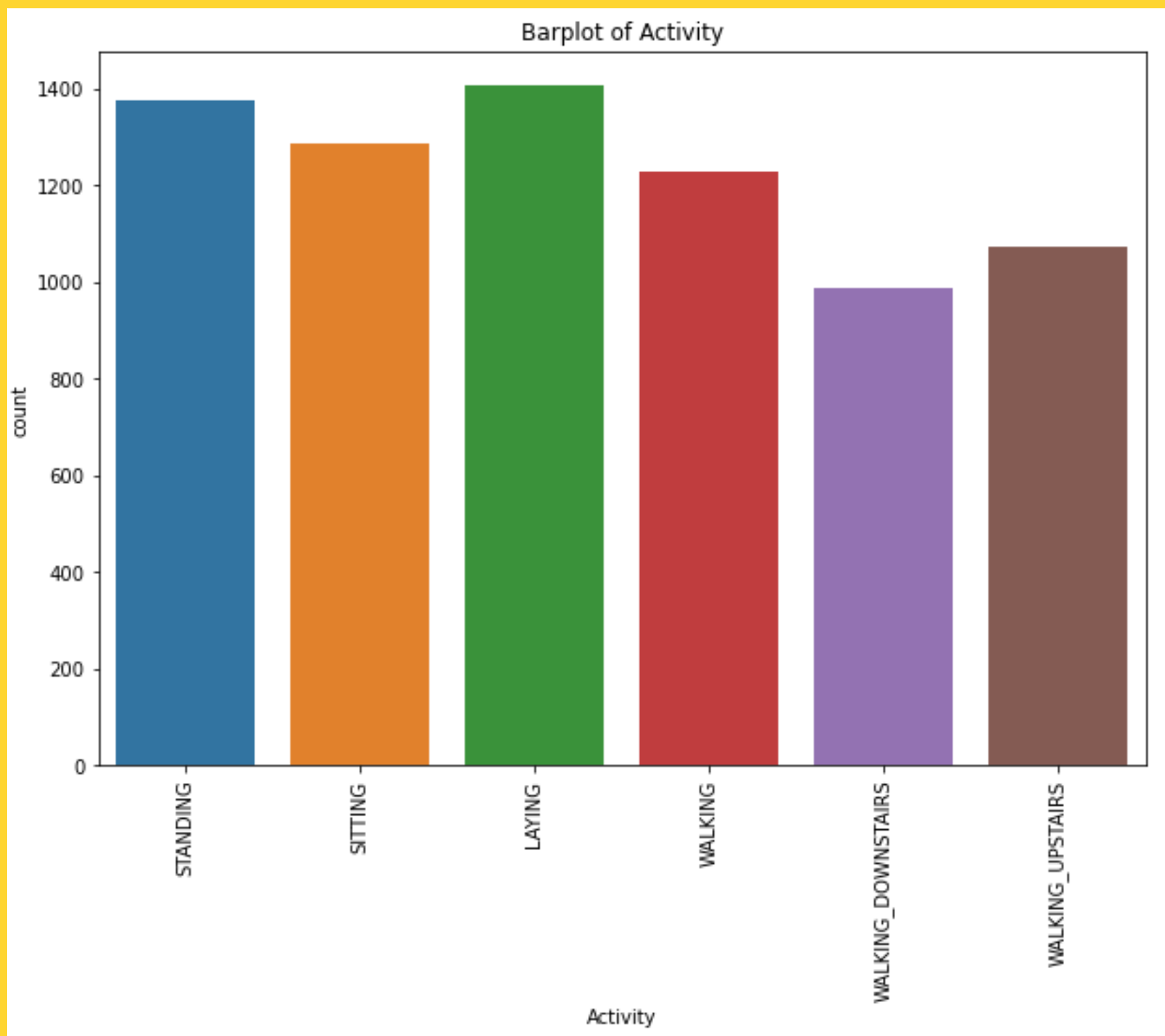
Pada dataset yang digunakan, dilakukan pengecekan duplikasi data dan pengecekan missing value.

```
1 #test untuk mengetahui apakah terdapat data duplicate pada data test dan train
2 print('Number of duplicates in train : ',sum(train.duplicated()))
3 print('Number of duplicates in test : ', sum(test.duplicated()))
```

```
Number of duplicates in train : 0
Number of duplicates in test : 0
```

```
1 #cek untuk mengetahui apakah terdapat missing value pada data train dan tes tersebut:
2 print('Total number of missing values in train : ', train.isna().values.sum())
3 print('Total number of missing values in train : ', test.isna().values.sum())
```

```
Total number of missing values in train : 0
Total number of missing values in train : 0
```



```
1 #untuk melakukan pengecekan terhadap balance data dari setiap aktivitas:
2 plt.figure(figsize=(10,7))
3 plt.title('Barplot of Activity')
4 sns.countplot(train.Activity)
5 plt.xticks(rotation=90)
```

```
(array([0, 1, 2, 3, 4, 5]), <a list of 6 Text major ticklabel objects>)
```

Aktivitas apa yang paling sering dilakukan oleh manusia?

Pada dataset yang digunakan, dilakukan pengecekan terhadap balance data / keseimbangan data. Pengecekan ini sangat penting karena klasifikasi memerlukan jumlah data yang seimbang tiap kelasnya. Berdasarkan pengecekan ini juga menjawab pertanyaan tersebut, yaitu LAYING adalah aktivitas yang paling sering dilakukan.

Exploratory Data Analysis

Berdasarkan sifat umum aktivitas, secara garis besar kami dapat menempatkannya dalam dua kategori.

- DUDUK, BERDIRI, BERLAPIS dapat dianggap sebagai aktivitas statis tanpa adanya gerakan.
- WALKING, WALKING_DOWNSTAIRS, WALKING_UPSTAIRS dapat dianggap sebagai aktivitas dinamis dengan melibatkan sejumlah besar gerakan.

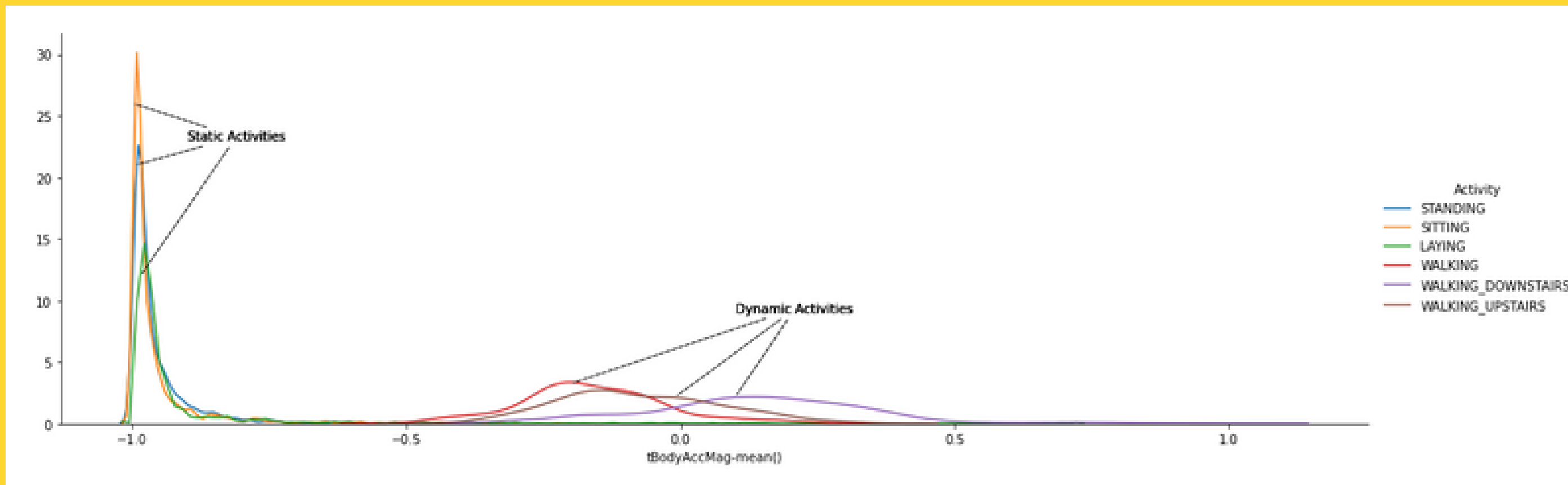
Mari pertimbangkan fitur tBodyAccMag-mean () untuk membedakan di antara dua rangkaian aktivitas .



Exploratory Data Analysis

Jika kita mencoba membangun model klasifikasi sederhana untuk mengklasifikasikan aktivitas menggunakan satu variabel dalam satu waktu, maka fungsi kepadatan probabilitas (PDF) sangat membantu untuk menilai pentingnya variabel kontinu `tBodyAccMag-mean()`.

Dengan menggunakan plot kepadatan, kami dapat dengan mudah menemukan kondisi untuk memisahkan aktivitas statis dari aktivitas dinamis.



Exploratory Data Analysis

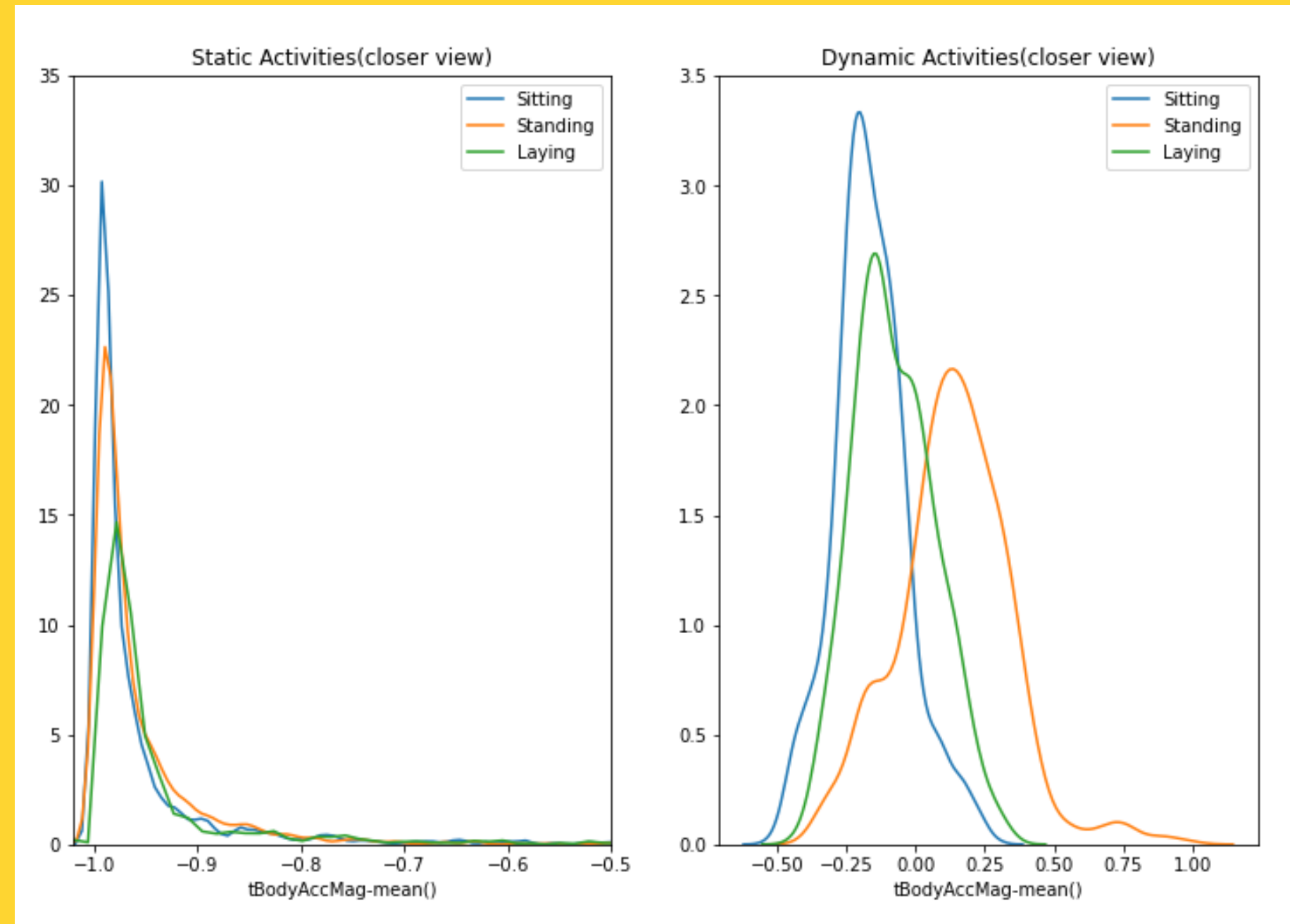
if(tBodyAccMag-mean())<=-0.5): kurang dari/= 0.5

Activity = "static"

else:

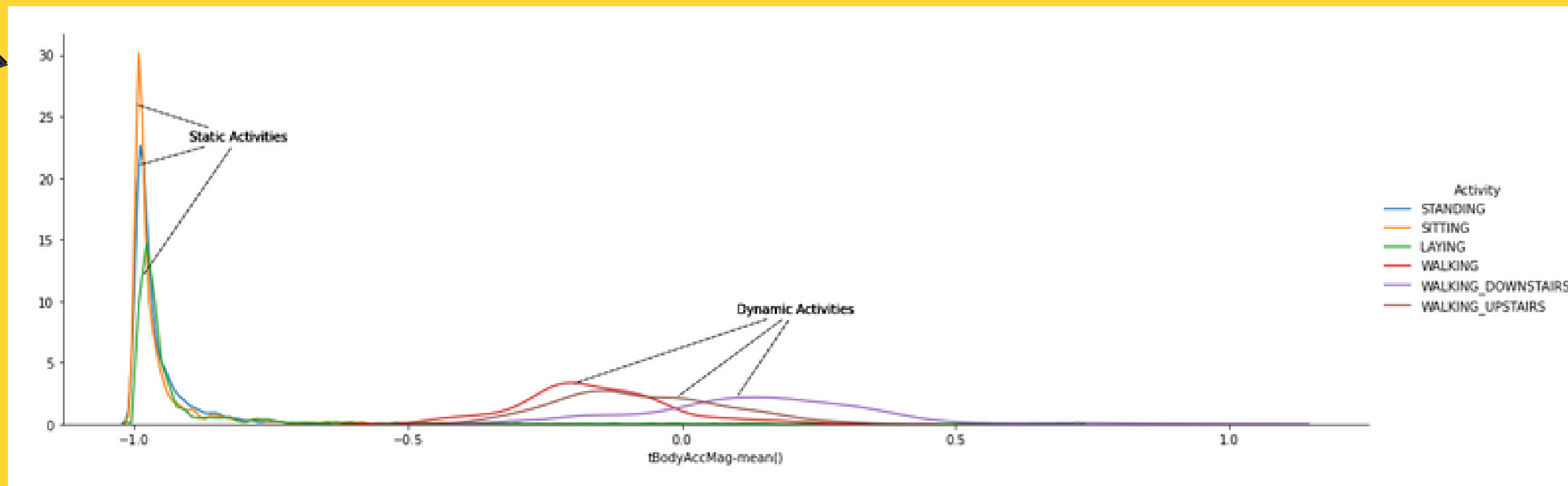
Activity = "dynamic"

Dapat kita lihat pada hasil plotting grafik tersebut.



Exploratory Data Analysis

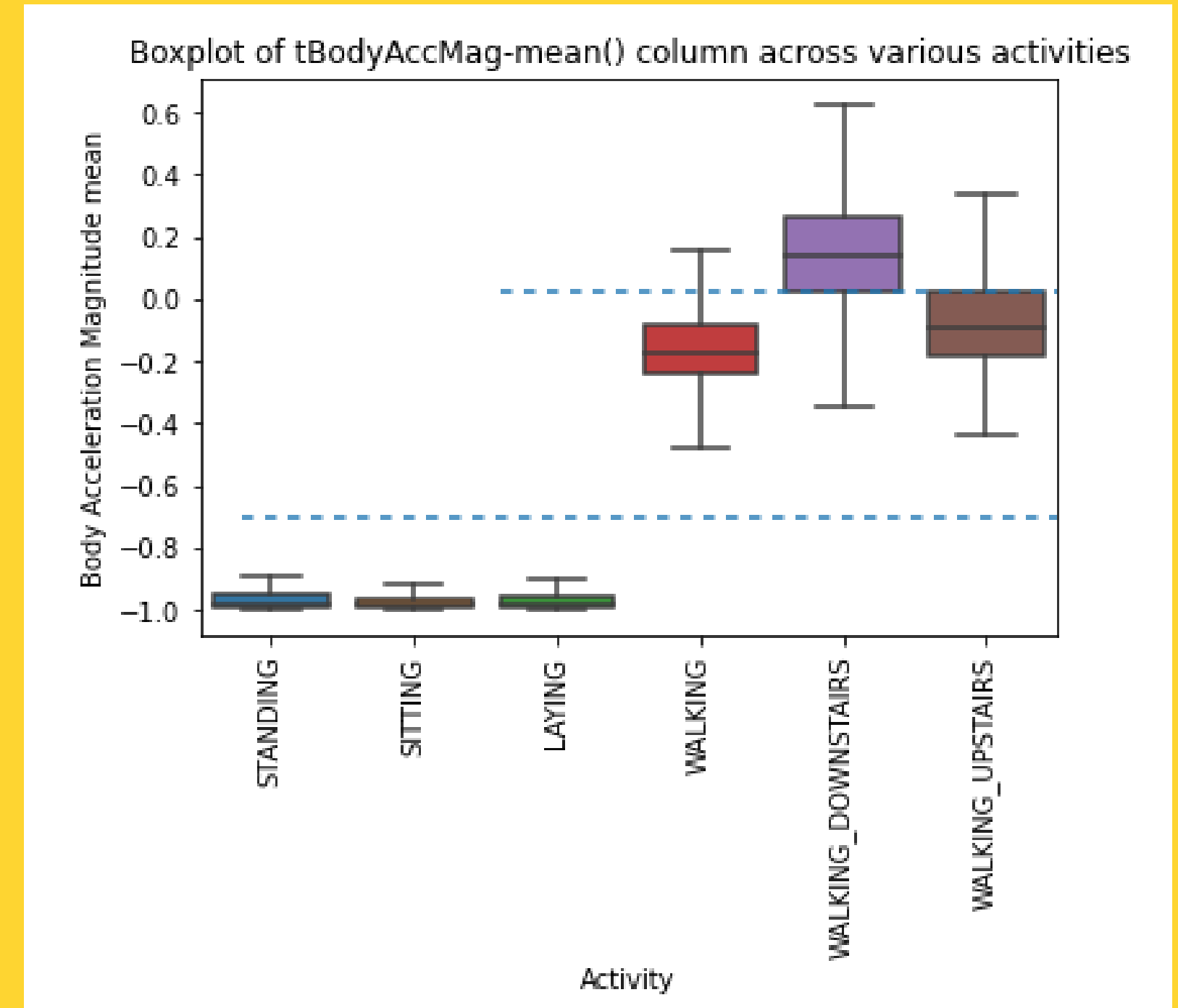
Dengan menggunakan plot kepadatan, kami dapat dengan mudah menemukan kondisi untuk memisahkan aktivitas statis dari aktivitas dinamis.



Exploratory Data Analysis

Menggambarkan box plot mean Body Acceleration Magnitude (tBodyAccMag-mean()) di keenam kategori tersebut

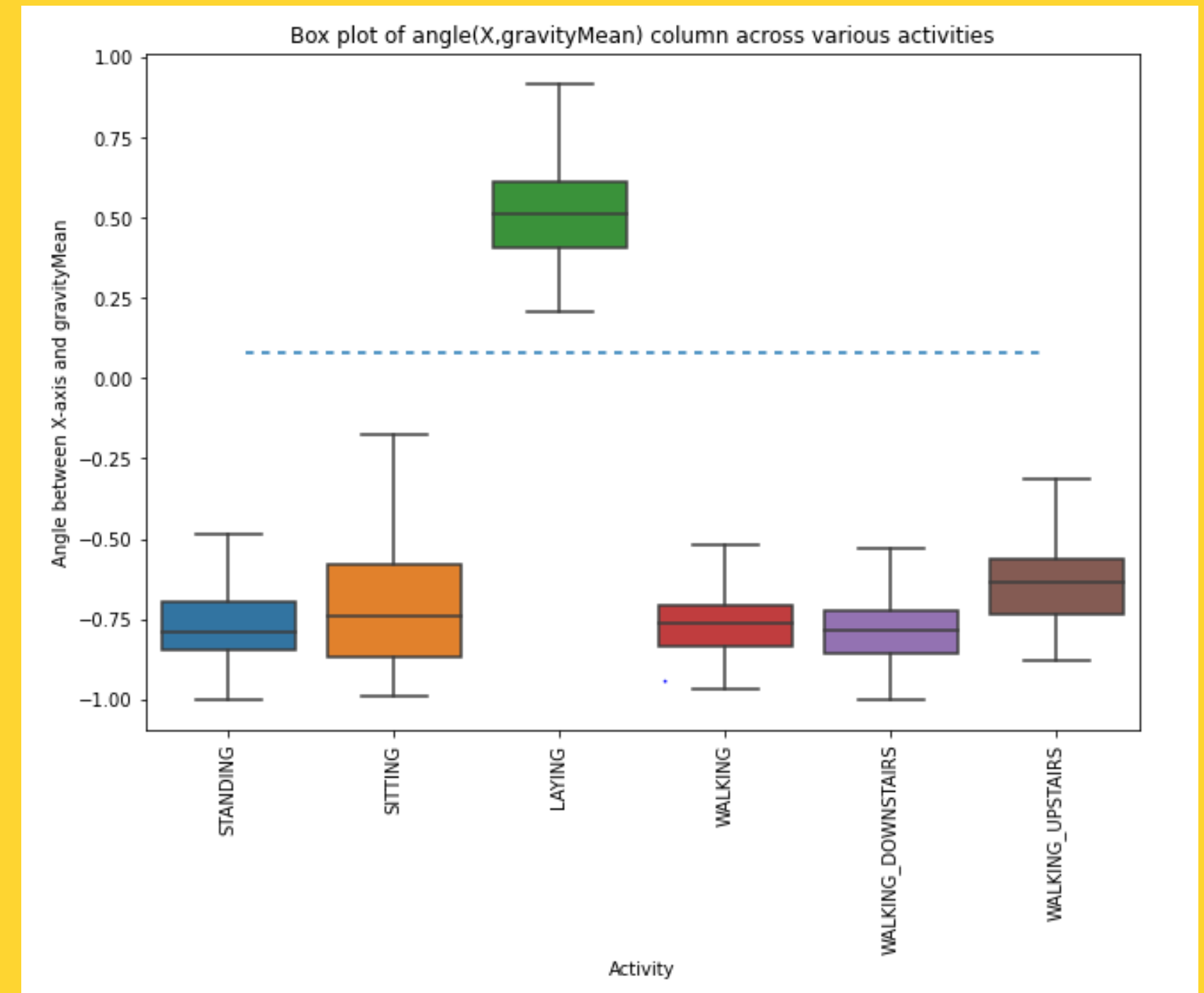
Berdasarkan dari hasil plotting box tersebut kita juga dapat memperhatikan perbedaan klasifikasi dari aktivitas statis dan dinamis



Exploratory Data Analysis

Menganalisis Sudut antara sumbu X dan fitur gravityMean.

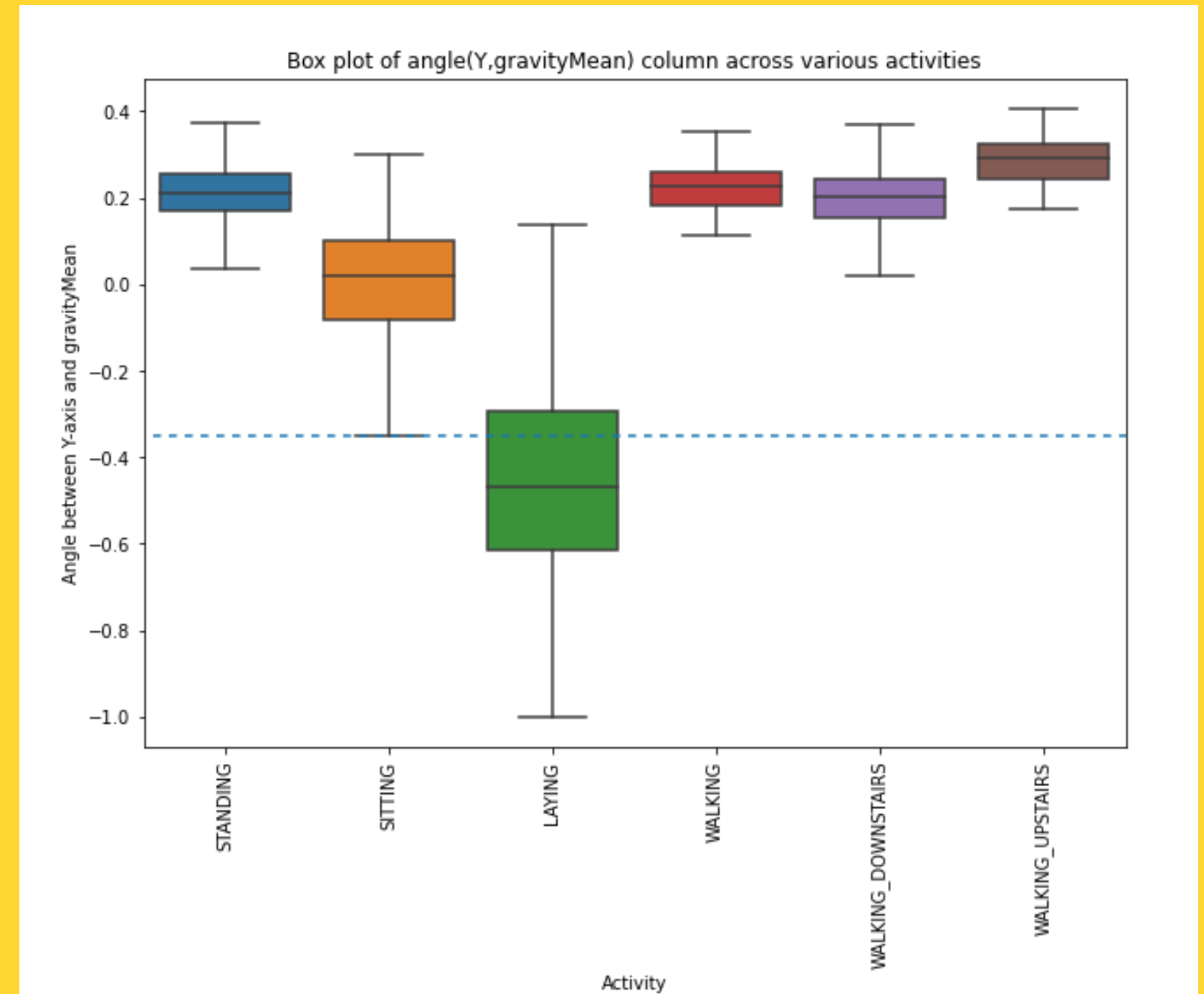
Dari diagram kotak kita dapat mengamati bahwa sudut (X, GravityMean) memisahkan LAYING dari aktivitas yang lain.



Exploratory Data Analysis

Menganalisis Sudut antara sumbu Y dan fitur GravityMean

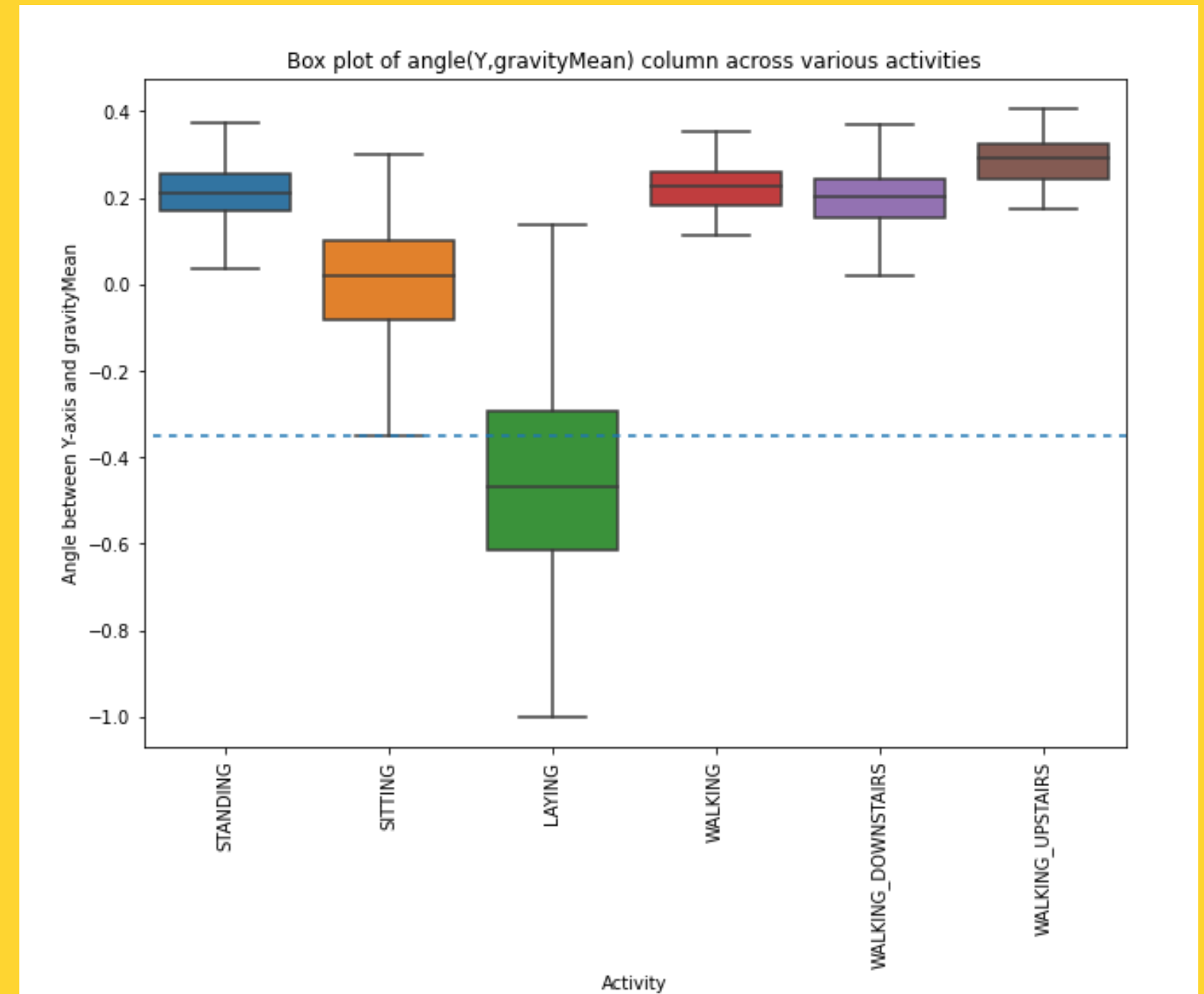
Demikian pula, dengan menggunakan Sudut antara sumbu Y dan gravitasiMean. Kita dapat memisahkan LAYING dari aktivitas yang lain, tetapi menyebabkan beberapa kesalahan klasifikasi



Exploratory Data Analysis

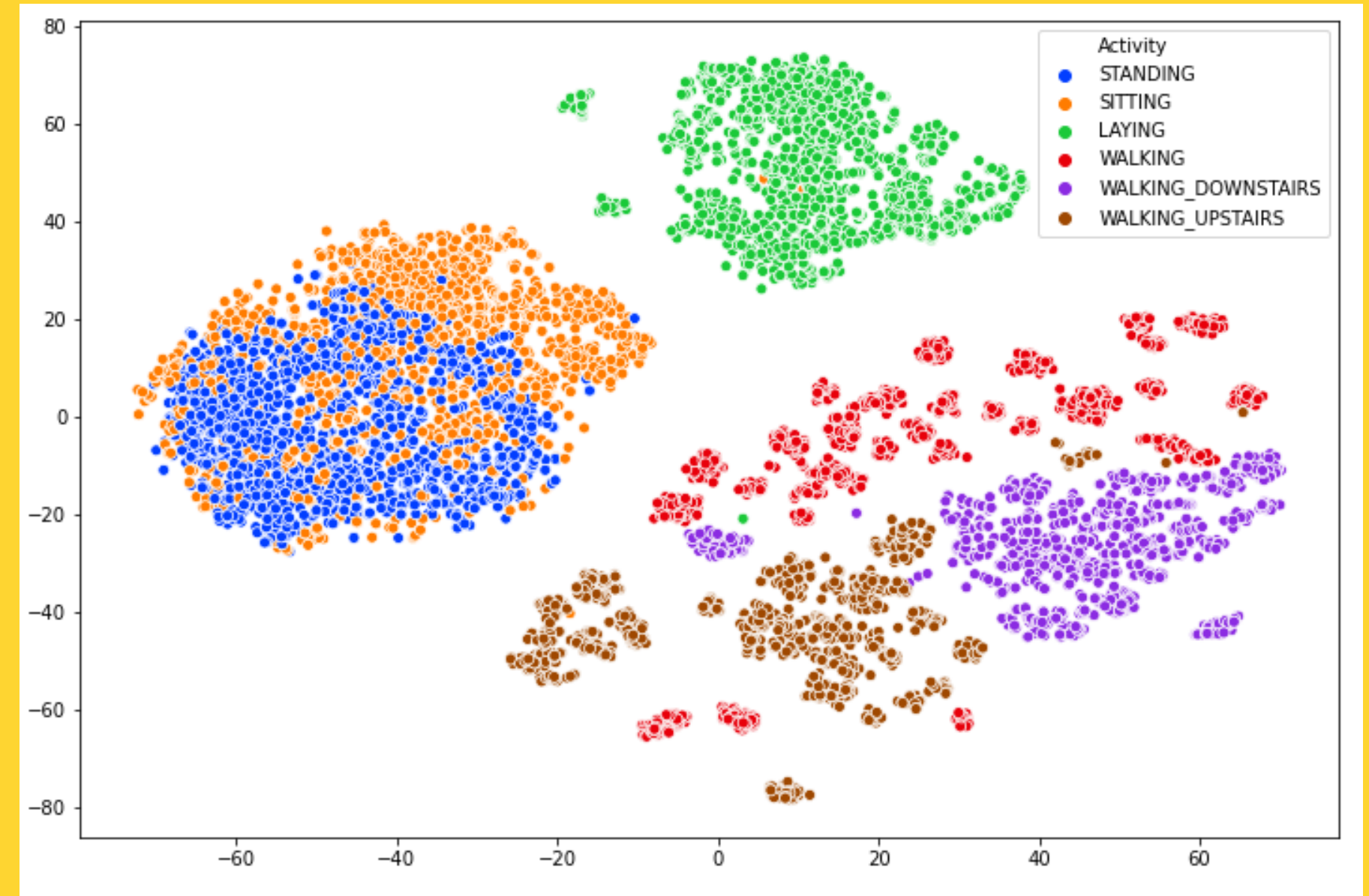
Menganalisis Sudut antara sumbu Y dan fitur GravityMean

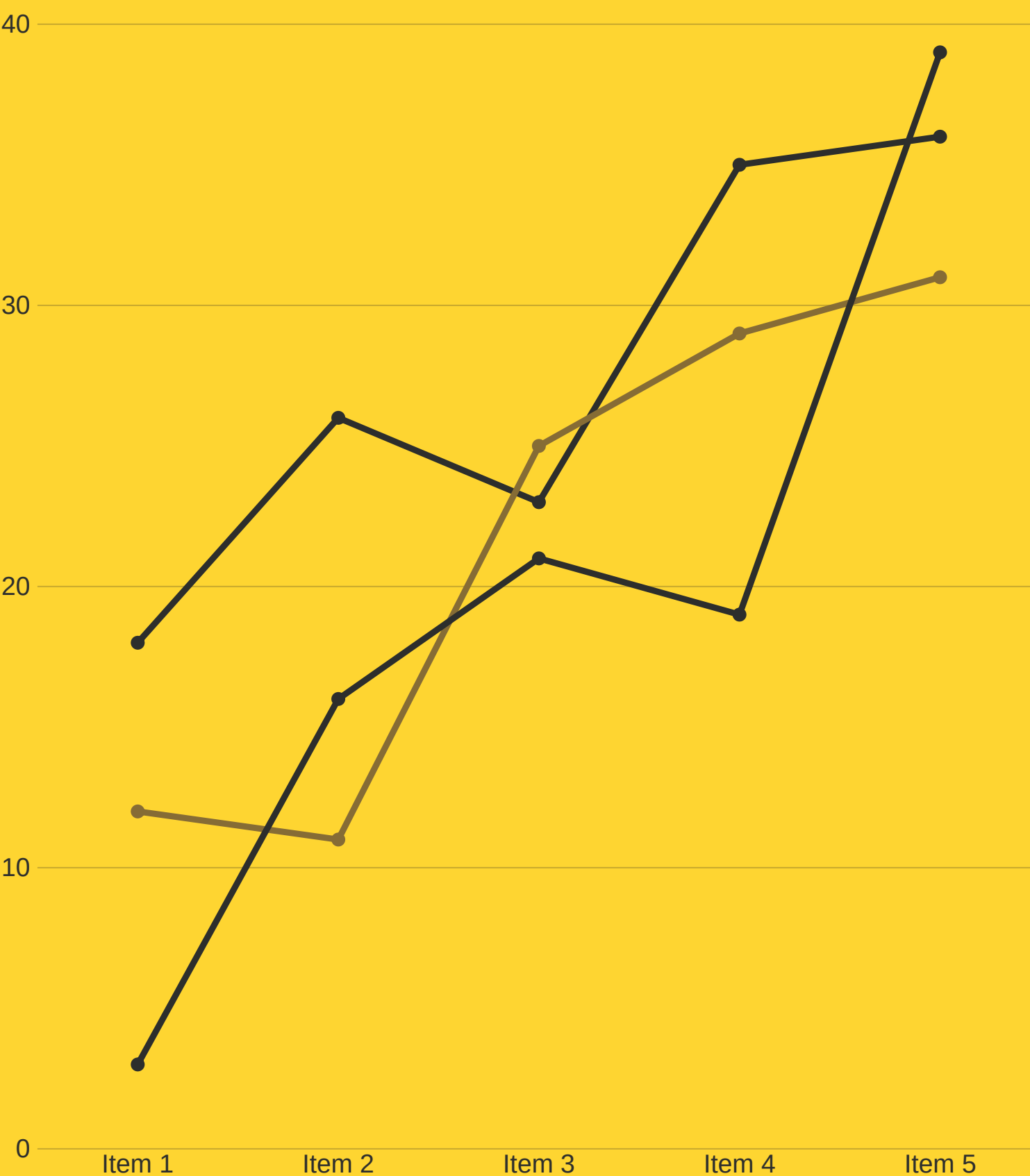
Dengan Sudut antara sumbu Y dan gravitasiMean. Kami dapat memisahkan LAYING dari aktivitas yang lain, tetapi menyebabkan beberapa kesalahan klasifikasi



Exploratory Data Analysis

Menggunakan t-SNE, kami memvisualisasikan dan memisahkan keenam aktivitas ke dalam ruang 2D.





MODELING

MLP

(Multilayer Perceptron)

RFC

(Random Forests
Classifier)

SVM

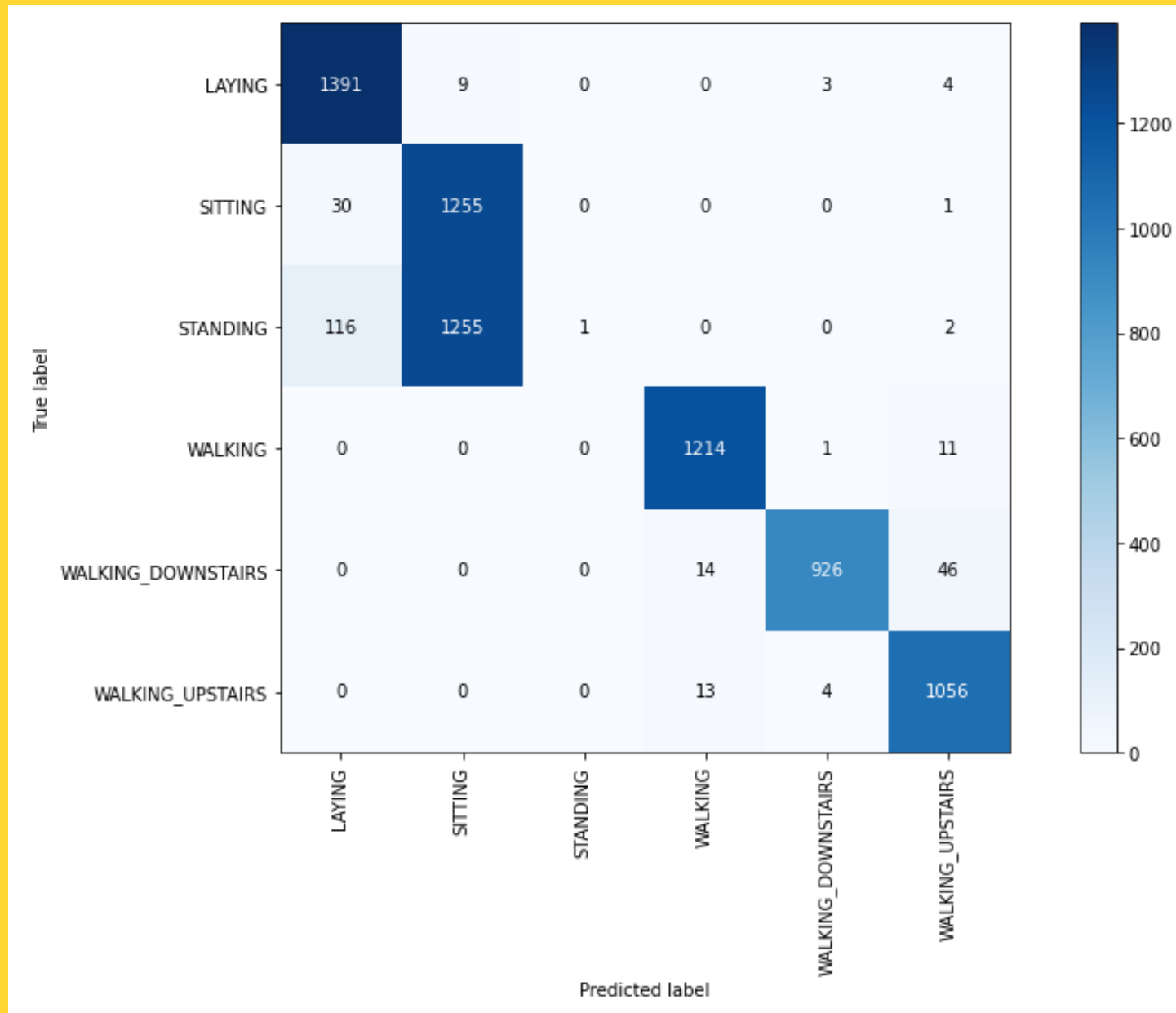
(Support Vector Machine)

ENSEMBLE

NAIVE BAYES
CLASSIFIER

MODELING

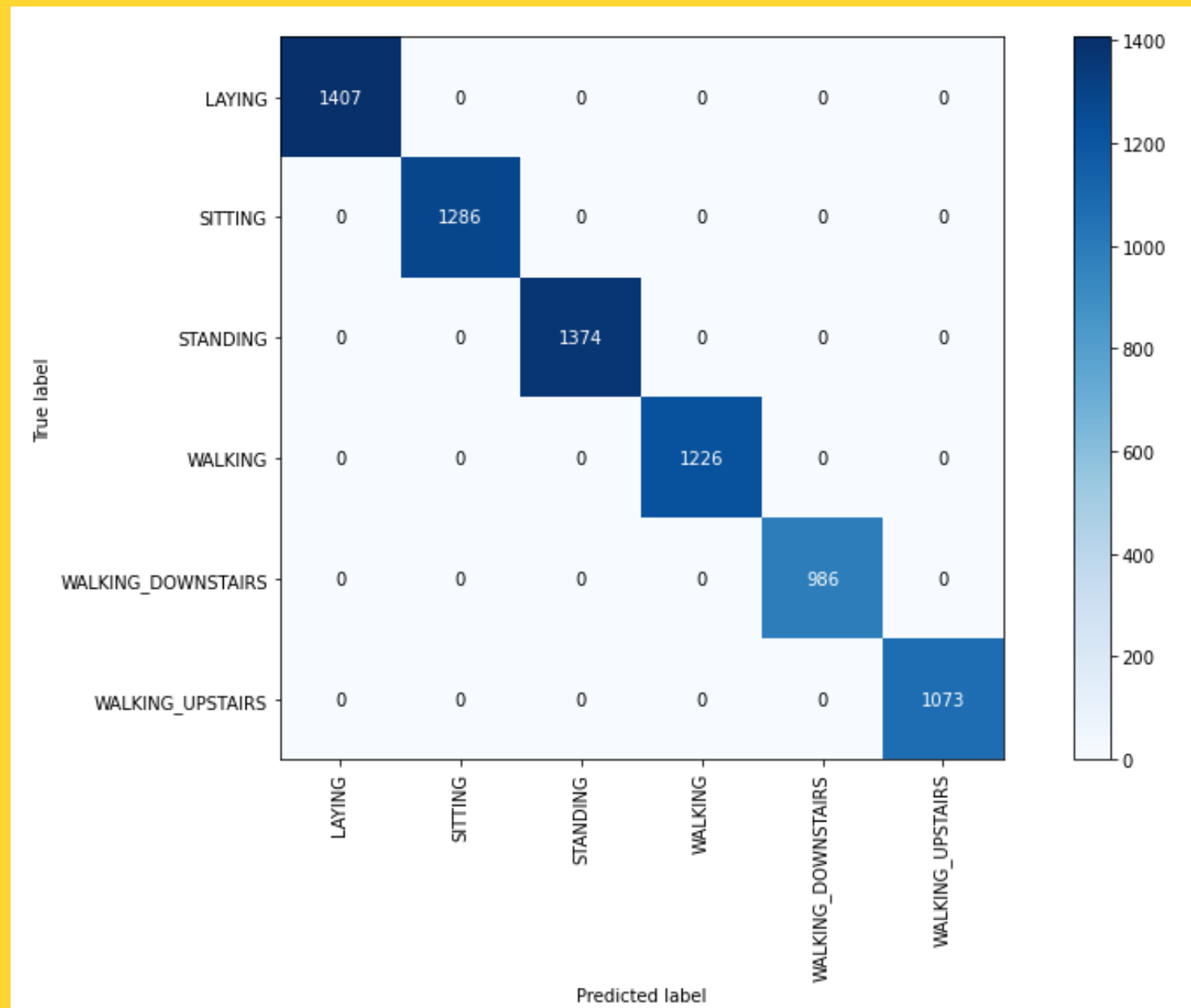
MLP
(Multilayer Perceptron)



Accuracy Score on test data : 0.7947497279651795

MODELING

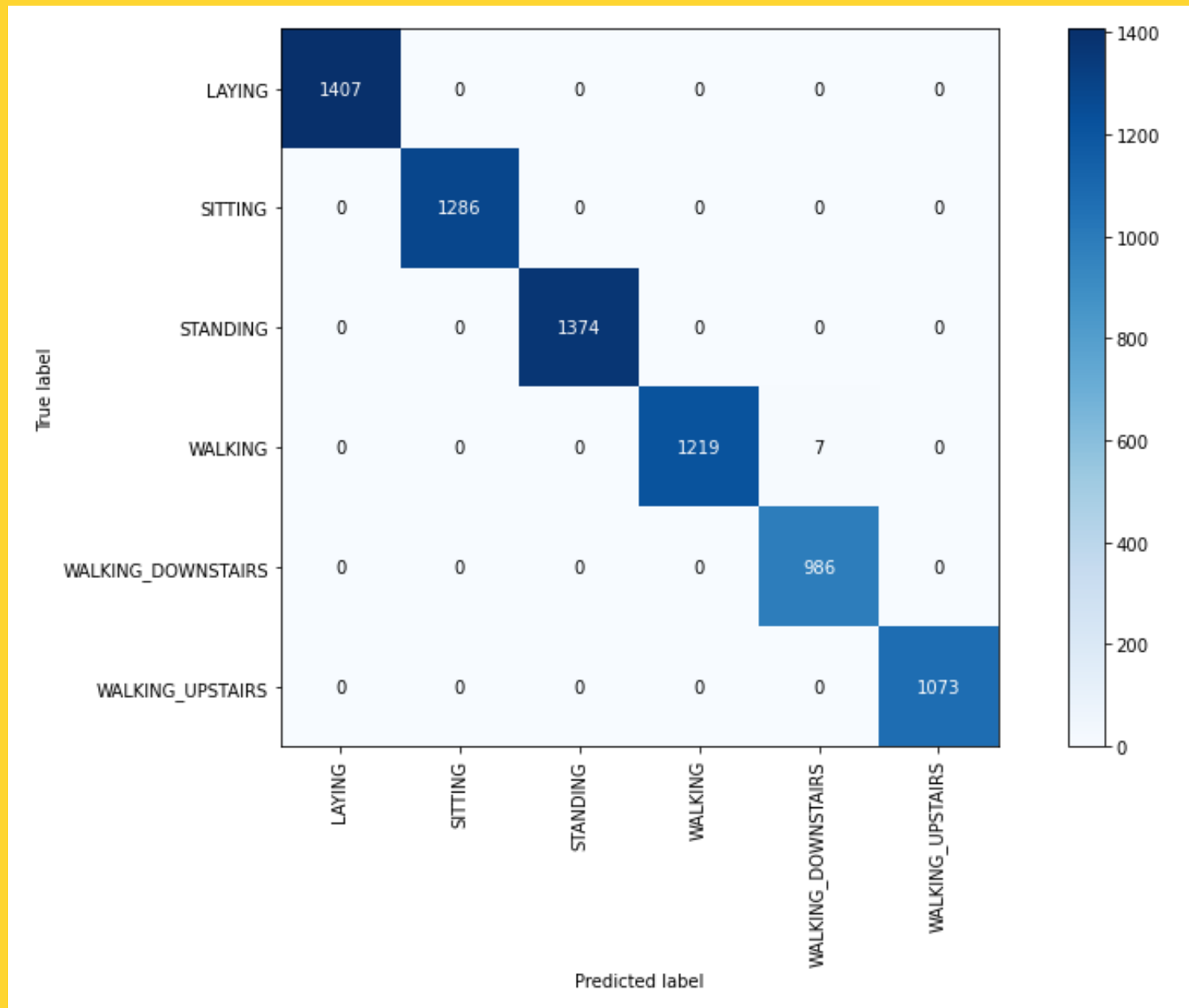
RFC
(Random Forests
Classifier)



Accuracy Score on test data : 1.0

MODELING

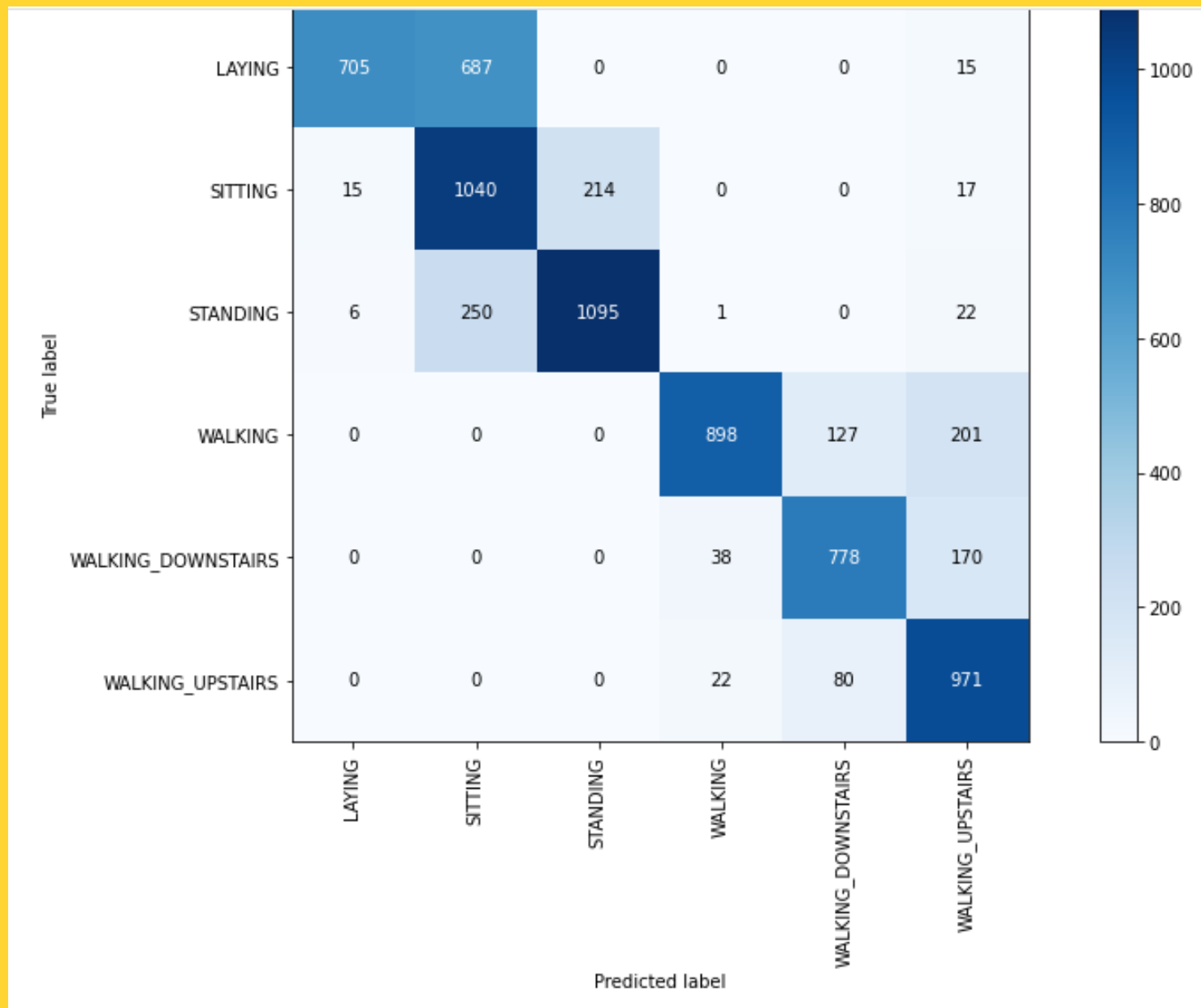
SVM
(Support Vector Machine)



Accuracy Score on test data : 0.9990478781284005

MODELING

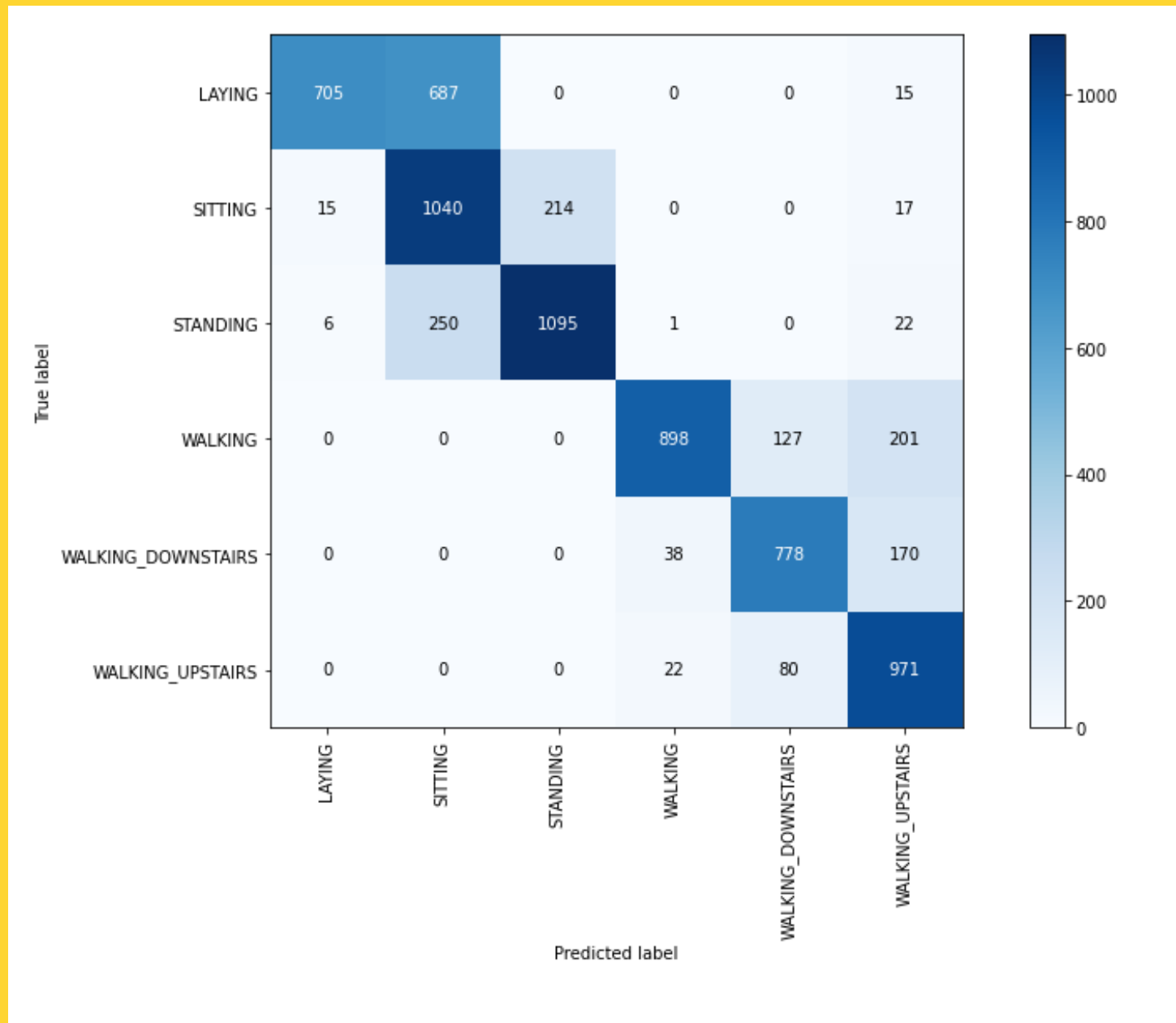
NAIVE BAYES CLASSIFIER



Accuracy Score on test data : 0.7463275299238302

MODELING

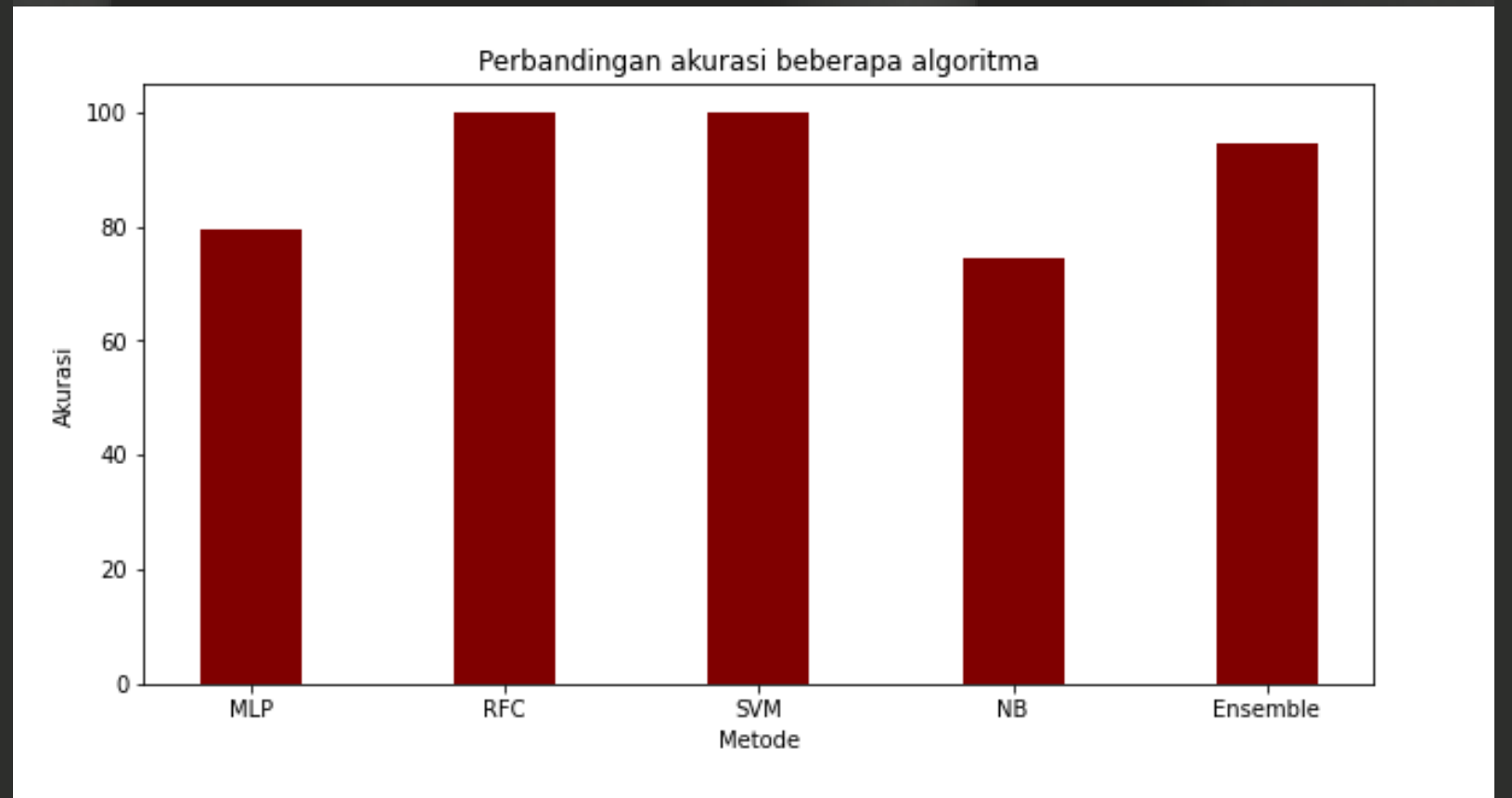
ENSEMBLE



0.9440968443960827

EVALUATION

- #1 RFC (100%)
- #2 SVM (99.90%)
- #3 ENSEMBLE (94.40%)
- #4 MLP (79.47%)
- #5 NB (74.63%)





KESIMPULAN



- Aktivitas yang paling sering dilakukan oleh manusia adalah LAYING
- Dengan grafik pada attribute `tBodyAccMag-mean()`, dapat dibedakan menjadi dua kelompok. Untuk `tBodyAccMag-mean()`, ≤ -0.5 dikategorikan menjadi statis dan ≥ 0.5 dikategorikan menjadi dinamis.
- Dalam database yang digunakan, tidak ada nilai missing, membuat cleaning menjadi mudah, namun pembacaan sensor sering berakibat pada nilai minus dikarenakan pembacaan yang berubah-ubah, nilai ini berdampak besar pada model yang bergantung pada nilai negatif besar, contoh : MLP dengan menggunakan relu, dimana relu sangat rentan pada nilai -, yang akan mengganggu proses update weight.



KESIMPULAN



Dengan menggunakan accelero, seseorang dapat mengukur percepatan linier dari suatu object, sedangkan dengan gyro, seseorang dapat mengukur kecepatan sudut dari suatu object, dengan menggabungkan kedua aspek tersebut, maka klasifikasi gerak dapat dilakukan dengan menerapkan sampling dalam waktu tertentu. Dari 5 model yang diuji, RFC yang paling tinggi, disusul dengan SVM dan Ensemble, 3 algoritma yang tidak memerlukan tuning yang rentan. MLP menghasilkan akurasi rendah karena beberapa alasan, dimulai dari activator relu yang rentan pada nilai minus, solver SGD yang sering terjebak di global optima, dan juga nilai momentum/learning rate yang kurang ditune. NB justru lebih rendah karena nilai data tidak mengikuti persebaran gaussian, maupun multinomial.



TERIMA KASIH