



Machine Learning (Praktikum)

SI – I₂

Principal Component Analysis

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA

082111633002 Mochamad Taufiqul Hafizh
082111633017 Muammar Qois Al Qorni
082111633094 Ferry Triwantono

Ekstraksi dan Seleksi Fitur

```
import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

from sklearn.decomposition import PCA

from sklearn.preprocessing import MinMaxScaler

from sklearn.preprocessing import StandardScaler

from sklearn.datasets import load_breast_cancer
```

Load untuk memberi label dan data. Untuk mengambil data, perlu memanggil .data dan untuk mengambil label .target.

```
cancer=load_breast_cancer()

(cancer.keys())

Out[10]: dict_keys(['data', 'target', 'frame', 'target_names', 'DESCR',
'feature_names', 'filename', 'data_module'])
```

Menampilkan 5 baris pertama dengan metode head()

```
df=pd.DataFrame(cancer['data'],columns=cancer['feature_names'])

df.head(5)
```

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	...	worst smoothness	worst compactness	worst concavity	worst concave points	worst symmetry	worst fractal dimension
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	...	0.1622	0.6656	0.7119	0.2654	0.4601	0.11890
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	...	0.1238	0.1866	0.2416	0.1860	0.2750	0.08902
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	...	0.1444	0.4245	0.4504	0.2430	0.3613	0.08758
3	11.42	20.38	77.58	386.1	0.14250	0.28390	...	0.2098	0.8663	0.6869	0.2575	0.6638	0.17300
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	...	0.1374	0.2050	0.4000	0.1625	0.2364	0.07678

Meratakan nilai kumpulan data secara merata dengan menggunakan StandardScaler()

```
scaler=StandardScaler()

scaler.fit(df)

StandardScaler(copy=True, with_mean=True, with_std=True)

scaled_data=scaler.transform(df)

scaled_data

Out[18]:
```

```

[[ 1.09706398 -2.07333501  1.26993369 ...  2.29607613  2.75062224
   1.93701461]

 [ 1.82982061 -0.35363241  1.68595471 ...  1.0870843  -0.24388967
   0.28118999]

 [ 1.57988811  0.45618695  1.56650313 ...  1.95500035  1.152255
   0.20139121]

 ...

 [ 0.70228425  2.0455738   0.67267578 ...  0.41406869 -1.10454895
  -0.31840916]

 [ 1.83834103  2.33645719  1.98252415 ...  2.28998549  1.91908301
   2.21963528]

 [-1.80840125  1.22179204 -1.81438851 ... -1.74506282 -0.04813821
  -0.75120669]]

pca=PCA(n_components=2)

pca.fit(scaled_data)

x_pca=pca.transform(scaled_data)

scaled_data.shape
Out[24]: (569, 30)

x_pca.shape
Out[26]: (569, 2)

scaled_data
x_pca
Out[29]:

[[ 9.19283683  1.94858307]

```

```
[ 2.3878018 -3.76817174]

[ 5.73389628 -1.0751738 ]

...

[ 1.25617928 -1.90229671]

[10.37479406  1.67201011]

[-5.4752433  -0.67063679]]
```

Sebar data 2D menggunakan nilai dari 2 komponen utama

```
plt.figure(figsize=(8,6))

plt.scatter(x_pca[:,0],x_pca[:,1],c=cancer['target'])

plt.xlabel('First principle component')

plt.ylabel('Second principle component')

plt.show()
```

