1 ANALISIS EKSTRAKSI CIRI SINYAL DOMAIN WAKTU DAN FREKUENSI PADA SUARA SENJATA API

1.1 Mata Kuliah Analisis Sinyal Digital 119

- Dosen Pengampu : Haris Suhendar, M.Sc, Dr. Bambang Heru Iswanto, M.Si
- Nama : Difa Farhani Hakim (1306620040)
- Kelas : Fisika - B
- Angkatan : 2020

2 Import Modul

In [1]:
```
# !python -m pip install spafe
# !pip install spafe
```

In [2]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

import librosa
from IPython.display import Audio
import spafe.features.pncc as pncc_lib
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

import sys
import os
```

3 Load dan Zip Dataset

Source dataset kaggle: https://www.kaggle.com/datasets/emrahaydemr/gunshot-audio-dataset/ (https://www.kaggle.com/datasets/emrahaydemr/gunshot-audio-dataset/)

Link download dataset: https://unjac-my.sharepoint.com/:u:/g/personal/difafarhanihakim_1306620040_mhs_unj_ac_id/EaRFSObt2S4B6bf2Dlik4RSo-nc2sHOJKw?e=vcViRJ (https://unjac-my.sharepoint.com/:u:/g/personal/difafarhanihakim_1306620040_mhs_unj_ac_id/EaRFSObt2S4B6bf2Dlik4RSo-nc2sHOJKw?e=vcViRJ)

In [3]:
```
path = "dataset/"
dir_list = os.listdir(path)
print(dir_list)
```

```
['AK-12', 'AK-47', 'IMI Desert Eagle', 'M16', 'M249', 'M4', 'MG-42', 'MP
5', 'Zastava M92']
```

In [4]:
```
num_sample = 10
path_list = []
path_list2 = []
for i in dir_list:
    path2 = os.listdir(path +'/'+ i)
    path2.sort()
    for j in path2[:num_sample]:
        path3 = i +'/'+ j
        path_list.append(i)
        path_list2.append(path3)

print(path_list[:5])
print(path_list2[:5])
```

```
['AK-12', 'AK-12', 'AK-12', 'AK-12', 'AK-12']
['AK-12/3 (1).wav', 'AK-12/3 (10).wav', 'AK-12/3 (11).wav', 'AK-12/3 (12).
wav', 'AK-12/3 (13).wav']
```

In [5]:
```
data = []
waktu = []
sr = []
no = 0
for i in path_list2:
    data_temp, sr_temp = librosa.load(path + '/' +  i)
    waktu_temp = librosa.get_duration(y=data_temp, sr=sr_temp)
    if no % num_sample == 0:
        print(i)
        display(Audio(data=data_temp,rate=sr_temp))

    data.append(data_temp)
    sr.append(sr_temp)
    waktu.append(waktu_temp)
    no += 1
print('Data loaded:',no)
```

```
AK-12/3 (1).wav
```

0:00 / 0:00

```
AK-47/1 (1).wav
```

0:00 / 0:00

```
IMI Desert Eagle/2 (1).wav
```

0:00 / 0:00

```
M16/5 (1).wav
```

0:00 / 0:00

```
M249/6 (1).wav
```

0:00 / 0:00

```
M4/4 (1).wav
```

0:00 / 0:00

```
MG-42/7 (1).wav
```

0:00 / 0:00

```
MP5/8 (1).wav
```

0:00 / 0:00

Zastava M92/9 (1).wav

0:00 / 0:00

Data loaded: 90

In [6]:
```
data_load = {
    'Kelas' : path_list,
    'Nama File' : path_list2,
    'Waktu' : waktu,
    'Sample Rate': sr,
       }
df_load = pd.DataFrame(data_load)
df_load
```

Out[6]:

	Kelas	Nama File	Waktu	Sample Rate
0	AK-12	AK-12/3 (1).wav	2.0	22050
1	AK-12	AK-12/3 (10).wav	2.0	22050
2	AK-12	AK-12/3 (11).wav	2.0	22050
3	AK-12	AK-12/3 (12).wav	2.0	22050
4	AK-12	AK-12/3 (13).wav	2.0	22050
...	...	...	...	...
85	Zastava M92	Zastava M92/9 (14).wav	1.0	22050
86	Zastava M92	Zastava M92/9 (15).wav	1.0	22050
87	Zastava M92	Zastava M92/9 (16).wav	1.0	22050
88	Zastava M92	Zastava M92/9 (17).wav	1.0	22050
89	Zastava M92	Zastava M92/9 (18).wav	1.0	22050

90 rows × 4 columns

4 Ekstraksi Ciri Energy, ZCR, Energy of Entropy, MFCC, dan PNCC

In [7]:
```
def energy1(data):
    output = np.sum(np.abs(data**2))
    return output

def zcr1(data):
    N = len(data)
    data2 = data[0:-1]
    data2 = np.insert(data2, 0, 0)
    a = np.sign(data)
    b = np.sign(data2)
    zc = np.sum(np.abs(a - b))/2
    zcr_ = zc/N
    return zcr_

def entropy_energy(frame, k=10):
    E_short = energy1(frame)
    frame_len = len(frame)
    sub_win_len = int(frame_len/k)
    e_j = []
    for i in range(0,frame_len,sub_win_len):
        E_sub = energy1(frame[i:i+sub_win_len])
        E_divide = E_sub/E_short
        e_j.append(E_divide)

    H_i = - np.sum(e_j * np.log2(e_j))
    return H_i
```

In [8]:
```
energy_list = []
zcr_list = []
eoe_list = []
mfcc_list = []
pncc_list = []
N = len(data)
for i in range(N):
    data_input = data[i]
    N_data = len(data[i])

    energy_0 = energy1(data_input)
    zcr_0 = zcr1(data_input)
    eoe_0 = entropy_energy(data_input,k=10)
    mfcc_0 = np.mean(librosa.feature.mfcc(y=data_input, n_mfcc=13, hop_
    pncc_0 = np.mean(pncc_lib.pncc(sig=data_input, nfft=N_data+1, fs=N_

    energy_list.append(energy_0)
    zcr_list.append(zcr_0)
    eoe_list.append(eoe_0)
    mfcc_list.append(mfcc_0)
    pncc_list.append(pncc_0)
```

4.1 Membuat List Ciri

In [9]:
```
data_pd = {
    'Kelas' : path_list,
    'Nama File' : path_list2,
    'Energy': energy_list,
    'ZCR': zcr_list,
    'EoE': eoe_list,
    'MFCC': mfcc_list,
    'PNCC': pncc_list,
       }
df = pd.DataFrame(data_pd)
df
```

Out[9]:

	Kelas	Nama File	Energy	ZCR	EoE	MFCC	PNCC
0	AK-12	AK-12/3 (1).wav	1044.180786	0.153866	2.775766	-36.055309	0.230036
1	AK-12	AK-12/3 (10).wav	1446.640869	0.191803	3.095031	0.845041	0.235556
2	AK-12	AK-12/3 (11).wav	225.114731	0.128583	1.058286	-11.705831	0.169762
3	AK-12	AK-12/3 (12).wav	1088.407227	0.170918	2.739360	-37.543819	0.229652
4	AK-12	AK-12/3 (13).wav	1200.696777	0.154932	2.874003	-13.768671	0.232243
...	...	...	...	...	...	...	...
85	Zastava M92	Zastava M92/9 (14).wav	537.527954	0.114036	3.302777	5.702017	0.309107
86	Zastava M92	Zastava M92/9 (15).wav	865.876953	0.128458	3.234779	6.980283	0.316747
87	Zastava M92	Zastava M92/9 (16).wav	557.040955	0.159977	3.317926	1.940660	0.307975
88	Zastava M92	Zastava M92/9 (17).wav	576.020325	0.171361	3.313111	1.680705	0.303595
89	Zastava M92	Zastava M92/9 (18).wav	877.008118	0.147959	3.237194	8.706468	0.312235

90 rows × 7 columns

In [10]:
```
unique_classes = np.unique(df['Kelas'])
num_unique = len(unique_classes)
print('Total number of outputs : \n', unique_classes)
print('Output classes : ', num_unique)
```

```
Total number of outputs :
 ['AK-12' 'AK-47' 'IMI Desert Eagle' 'M16' 'M249' 'M4' 'MG-42' 'MP5'
 'Zastava M92']
Output classes :  9
```

5 Principle Component Analysis

In [11]:
```
target = df['Kelas']
df_fit = df.drop(['Kelas','Nama File'], axis='columns')
```

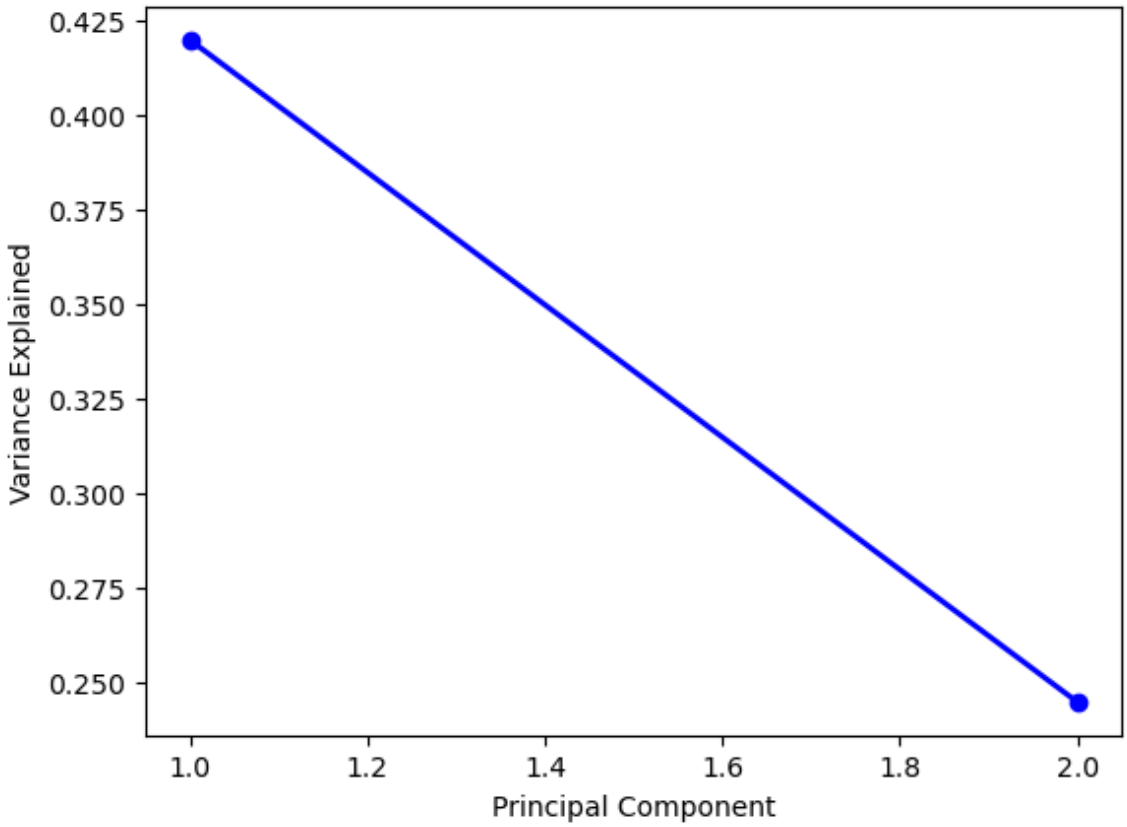
In [12]:
```
df_fit_standard = StandardScaler().fit_transform(df_fit)
pca_standard = PCA(n_components = 2)
data_PCA_standard = pca_standard.fit_transform(df_fit_standard)

```

5.1 Scree Plot

In [13]:
```
PC_values = np.arange(pca_standard.n_components_) + 1
plt.plot(PC_values, pca_standard.explained_variance_ratio_,
         'o-', linewidth=2, color='blue')

plt.xlabel('Principal Component')
plt.ylabel('Variance Explained')
plt.show()
```



In [14]:
```
print('Explained variation per principal component: {}'.format(pca_stan
```

```
Explained variation per principal component: [0.41966535 0.24463532]
```

Dari hasil tersebut menandakan bahwa PCA1 mewakili sampel data sebesar 41.96%, sedangkan pada PCA2 mewakili sampel data sebesar 24.46%. Dari kedua PCA tersebut diperoleh bahwa 33.58% informasi data telah hilang.

5.2 Hasil PCA pada ekstraksi ciri

In [15]:
```
pca_comp = pca_standard.components_.transpose()
list_ciri = ['Energy','ZCR','EoE','MFCC','PNCC']
data_comp = {
    'Ciri': list_ciri,
    'PCA1': pca_comp[:,0],
    'PCA2': pca_comp[:,1],
       }
df_comp = pd.DataFrame(data_comp)
df_comp
```

Out[15]:

	Ciri	PCA1	PCA2
0	Energy	-0.361424	0.385011
1	ZCR	-0.272838	0.688400
2	EoE	-0.600188	0.094492
3	MFCC	-0.371671	-0.471795
4	PNCC	-0.544580	-0.382560

In [16]:
```
df_comp.sort_values(by=['PCA1'])
```

Out[16]:

	Ciri	PCA1	PCA2
2	EoE	-0.600188	0.094492
4	PNCC	-0.544580	-0.382560
3	MFCC	-0.371671	-0.471795
0	Energy	-0.361424	0.385011
1	ZCR	-0.272838	0.688400

Pada PCA1 mengukur variasi berdasarkan EoE dan PNCC, sehingga bertambahnya nilai EoE dan PNCC akan memberi dampak pada ciri yang lain. Selain itu, Pada PCA1 memiliki korelasi keterbalikan dengan semua ciri ditandai dengan nilai PCA1 yang negatif pada setiap ciri.

In [17]:
```
df_comp.sort_values(by=['PCA2'])
```

Out[17]:

	Ciri	PCA1	PCA2
3	MFCC	-0.371671	-0.471795
4	PNCC	-0.544580	-0.382560
2	EoE	-0.600188	0.094492
0	Energy	-0.361424	0.385011
1	ZCR	-0.272838	0.688400

Pada PCA2 mengukur variasi berdasarkan MFCC dan ZCR. PCA2 memiliki korelasi pada rendahnya ciri MFCC dan tingginya ciri ZCR.

5.3 Hasil PCA pada setiap sampel data

In [18]:
```
data_PCA = {
    'Nama File' : path_list2,
    'PCA1': data_PCA_standard[:,0],
    'PCA2': data_PCA_standard[:,1],
       }
df_PCA = pd.DataFrame(data_PCA)
df_PCA
```

Out[18]:

	Nama File	PCA1	PCA2
0	AK-12/3 (1).wav	1.169666	1.374381
1	AK-12/3 (10).wav	-0.528279	0.571548
2	AK-12/3 (11).wav	3.189182	0.295713
3	AK-12/3 (12).wav	1.162742	1.631653
4	AK-12/3 (13).wav	0.366680	0.615526
...	...	...	...
85	Zastava M92/9 (14).wav	-1.257534	-1.404310
86	Zastava M92/9 (15).wav	-1.486232	-1.277042
87	Zastava M92/9 (16).wav	-1.359299	-0.731162
88	Zastava M92/9 (17).wav	-1.344481	-0.547245
89	Zastava M92/9 (18).wav	-1.567243	-1.073528

90 rows × 3 columns

In [19]:
```
df_PCA.sort_values(by=['PCA1'])
```

Out[19]:

	Nama File	PCA1	PCA2
49	M249/6 (18).wav	-2.942781	1.218338
11	AK-47/1 (10).wav	-2.549370	2.143396
12	AK-47/1 (11).wav	-2.451161	2.311179
62	MG-42/7 (100).wav	-2.423194	1.097180
17	AK-47/1 (16).wav	-2.409883	3.249113
...	...	...	...
16	AK-47/1 (15).wav	2.084632	0.727010
70	MP5/8 (1).wav	2.236434	-0.575000
27	IMI Desert Eagle/2 (15).wav	2.368623	-1.504423
2	AK-12/3 (11).wav	3.189182	0.295713
28	IMI Desert Eagle/2 (16).wav	3.514673	-0.835984

90 rows × 3 columns

Pada PCA 1 nilai terendah ada pada M249 dan AK-47, sedangkan nilai tertinggi ada pada IMI Desert Eagle dan AK-12. Hal tersebut menandakan bahwa variasi pada PCA1 bergantung pada kedua tipe jenis senjata api tersebut.

In [20]:
```
df_PCA.sort_values(by=['PCA2'])
```

Out[20]:

	Nama File	PCA1	PCA2
80	Zastava M92/9 (1).wav	-1.990476	-1.833612
52	M4/4 (100).wav	-0.491922	-1.778495
32	M16/5 (100).wav	-0.491922	-1.778495
84	Zastava M92/9 (13).wav	-1.192943	-1.539727
27	IMI Desert Eagle/2 (15).wav	2.368623	-1.504423
...	...	...	...
12	AK-47/1 (11).wav	-2.451161	2.311179
19	AK-47/1 (18).wav	-1.448744	2.345370
14	AK-47/1 (13).wav	-0.643670	2.777835
18	AK-47/1 (17).wav	-2.393353	2.845548
17	AK-47/1 (16).wav	-2.409883	3.249113

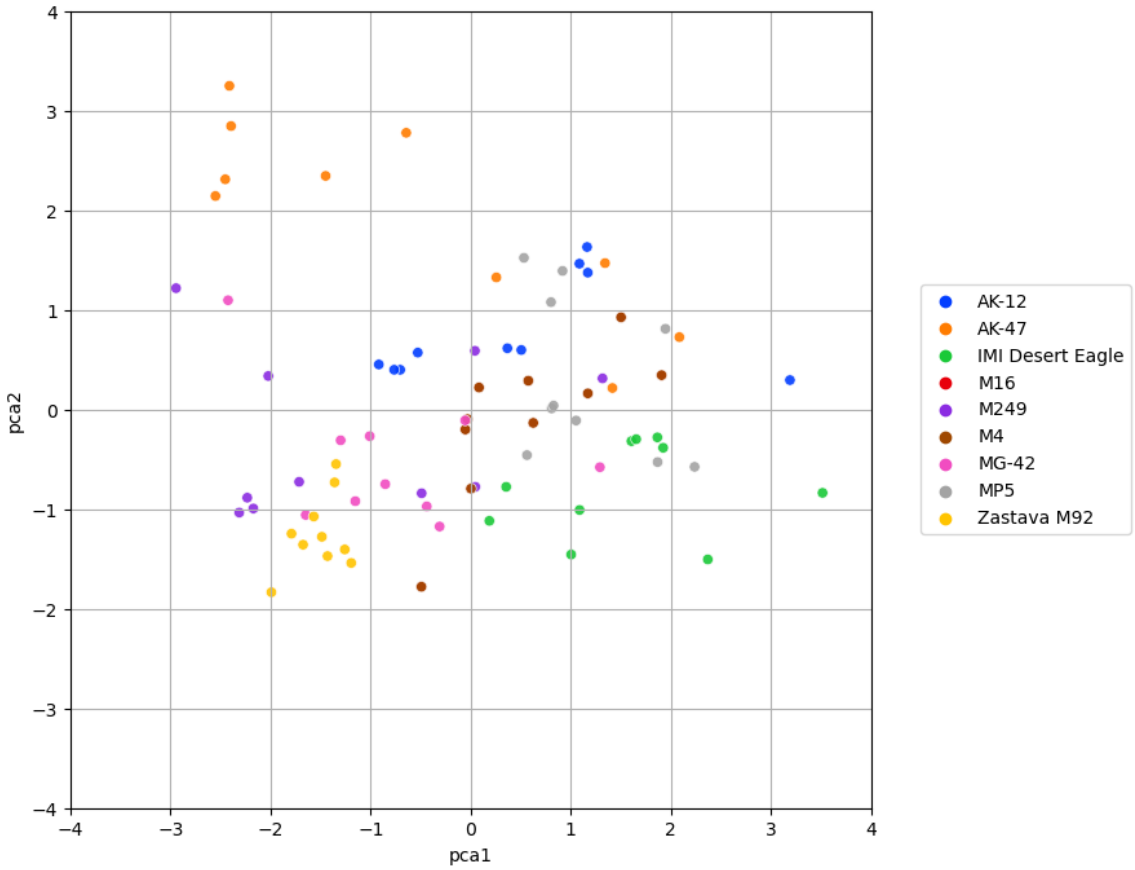
90 rows × 3 columns

Pada PCA2 nilai terendah ada pada Zastava, sedangkan nilai tertinggi ada pada AK-47. Hal tersebut menandakan bahwa variasi pada PCA2 bergantung pada tipe senjata api yang memiliki kesamaan dengan Zastava atau dengan AK-47.

In [21]:
```
pc12_standard = pd.DataFrame(data = data_PCA_standard, columns = ['pca1
pc12_standard['target'] = target
```

5.4 Plot Persebaran Data

In [22]:
```
plt.figure(figsize=(8,8))
splot = sns.scatterplot(
    x = "pca1", y="pca2",
    hue = "target",
    palette = sns.color_palette("bright", num_unique),
    data = pc12_standard,
    alpha = 0.9
)
splot.legend(loc='center left', bbox_to_anchor=(1.05, 0.5), ncol=1)
splot.set_xlim(-4,4)
splot.set_ylim(-4,4)
plt.grid()
```



Dari hasil grafik yang diperoleh bahwa senjata api bertipe Zastava M92 memiliki sampel yang sangat berdekatan satu sama lain, kemudian diikuti dengan M4, MG42, dan IMI Desert Eagle. Pada tipe senjata yang lain, sampel memiliki outlier yang cukup jauh dibandingkan kumpulannya, seperti pada AK-47 yang memiliki outlier terjauh.

6 Referensi

- Buku Giannakopoulos - Introduction to Audio Analysis
- PNCC : https://spafe.readthedocs.io/en/latest/features/pncc.html (https://spafe.readthedocs.io/en/latest/features/pncc.html)
- PCA :
  - https://www.datacamp.com/tutorial/principal-component-analysis-in-python (https://www.datacamp.com/tutorial/principal-component-analysis-in-python)

- https://www.datacamp.com/tutorial/pca-analysis-r (https://www.datacamp.com/tutorial/pca-analysis-r)
- https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html (https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html)
- https://online.stat.psu.edu/stat505/lesson/11 (https://online.stat.psu.edu/stat505/lesson/11)