

Clustering K-Means dan Agglomerative		
Praktikan	Aslab	
Nama: xxxx	Annur Hangga Prihadi	065001800028
Nim: xxxx	Faiz Kumara	065001800003

PRAKTIKUM 9

DATA SAINS DAN ANALITIK

Topik pertemuan praktikum ke-sembilan adalah mengolah data produksi ikan di semua Provinsi yang ada di Indonesia menggunakan Clustering untuk menentukan masing-masing kebutuhan bisnis klasterisasi produksi.

Source Code:

Clustering K-Means:

https://github.com/hanggaa/PrakDSDA/blob/main/Prak_9_Klastering_KMeans.ipynb

Clustering Agglomerative:

https://github.com/hanggaa/PrakDSDA/blob/main/Prak_9_Klastering_Agglomerative.ipynb

Latihan 1

Clustering K-Means

1. Memasang library yang dibutuhkan

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.cluster import KMeans
```

2. Membaca data

```
In [2]: dataset = pd.read_csv("C:/Users/hangg/Downloads/Random Aslab/DSDA/Material/Produksi Bandeng Provinsi.csv", sep=";")
```

3. Mengelompokkan data

```
In [3]: avg_vol = dataset.groupby("Provinsi")["Volume Produksi"].mean()
```

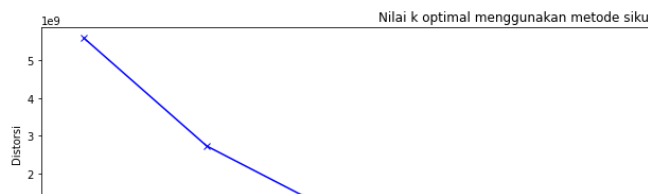
```
In [4]: print('Rata-rata Volume Produksi Bandeng:\n', avg_vol)
```

```
Rata-rata Volume Produksi Bandeng:
Provinsi
ACEH                22184.470909
BALI                 99.979000
BANTEN             12171.280000
BENGKULU            243.381000
```

4. Metode siku

```
In [5]: distorsi = []
K = range(2,10)
for k in K:
    kmeanModel = KMeans(n_clusters=k)
    kmeanModel.fit(avg_vol.values.reshape(-1, 1))
    distorsi.append(kmeanModel.inertia_)

In [6]: plt.figure(figsize=(16,4))
plt.plot(K, distorsi, 'bx-')
plt.xlabel('Nilai k')
plt.ylabel('Distorsi')
plt.title('Nilai k optimal menggunakan metode siku')
plt.show()
```



5. Implementasi Clustering

```
In [7]: clustering = KMeans(n_clusters=4, init="random", n_init=1)
clusters=clustering.fit_predict(avg_vol.values.reshape(-1, 1))

In [8]: print('\nHasil clustering:\n', clusters)

Hasil clustering:
[2 1 0 1 1 1 0 1 3 3 3 0 2 0 0 1 1 1 0 1 1 1 1 0 3 0 2 1 1 0 1]
```

6. Membuat dataframe hasil clustering

```
In [9]: avg_vol=pd.DataFrame({'Provinsi':avg_vol.index, 'Rata-Rata Vol':avg_vol.values, 'Cluster':clusters})

In [10]: centroid_perdata=avg_vol.groupby('Cluster')['Rata-Rata Vol'].transform('mean')
centroid=np.unique(centroid_perdata)

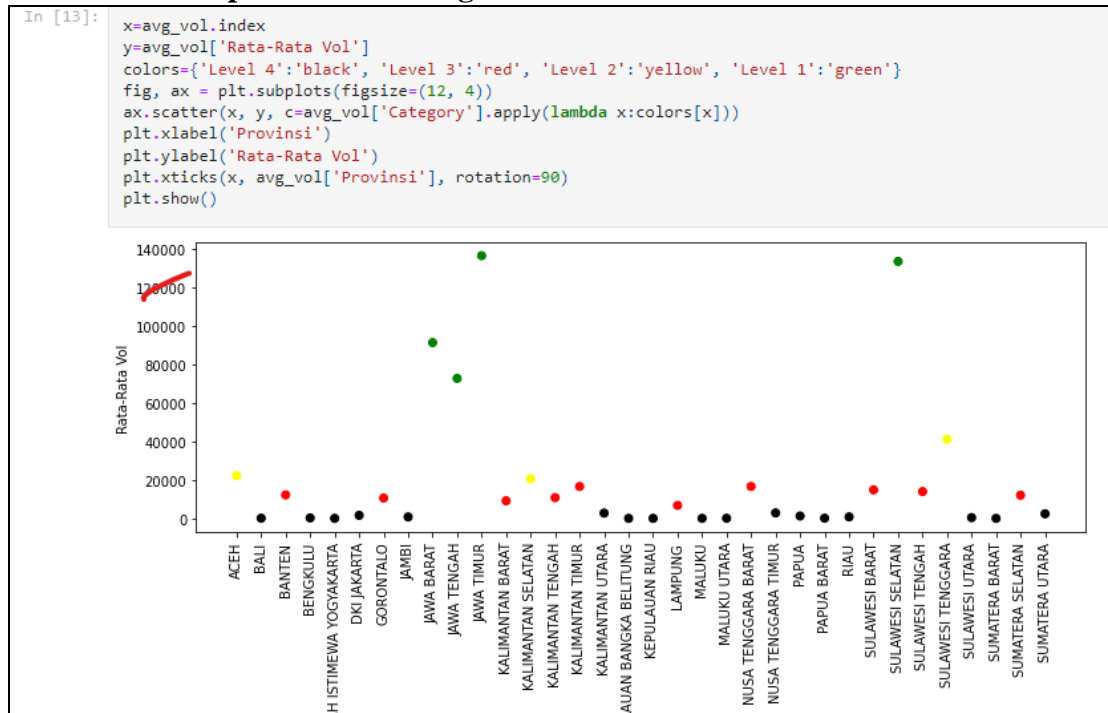
In [11]: sorted_centroid=np.sort(centroid)
Level4, Level3, Level2, Level1 = sorted_centroid[0], sorted_centroid[1], sorted_centroid[2], sorted_centroid[3]
category=centroid_perdata.map({Level4:'Level 4', Level3:'Level 3', Level2:'Level 2', Level1:'Level 1'})
avg_vol['Category']=category
```

7. Mencetak interpretasi Clustering

```
In [12]: print(avg_vol[["Provinsi", "Category"]])
```

	Provinsi	Category
0	ACEH	Level 2
1	BALI	Level 4
2	BANTEN	Level 3
3	BENGKULU	Level 4
4	DAERAH ISTIMEWA YOGYAKARTA	Level 4

8. Visualisasi interpretasi Clustering



Clustering Agglomerative

1. Memasang library yang dibutuhkan

```
In [1]: import pandas as pd
import numpy as np
from sklearn.cluster import AgglomerativeClustering
from matplotlib import pyplot as plt
from matplotlib.pyplot import figure
```

2. Membaca data

```
In [2]: dataset = pd.read_csv("C:/Users/hangg/Downloads/Random Aslab/DSDA/Material/Produksi Bandeng Provinsi.csv", sep=";")
```

3. Mengelompokkan data

```
In [3]: avg_vol = dataset.groupby("Provinsi")["Volume Produksi"].mean()
```

```
In [4]: print('Rata-rata Volume Produksi Bandeng:\n', avg_vol)
```

Rata-rata Volume Produksi Bandeng:	
Provinsi	
ACEH	22184.470909
BALI	99.979000
BANTEN	12171.280000
BENGKULU	243.381000

4. Implementasi Clustering

```
In [5]: clustering=AgglomerativeClustering(n_clusters=4, linkage='average')
clusters=clustering.fit_predict(avg_vol.values.reshape(-1, 1))

In [6]: print('\nHasil clustering:\n', clusters)

Hasil clustering:
[1 1 1 1 1 1 1 1 0 0 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 3 1 1 1 1]
```

5. Membuat dataframe hasil clustering

```
In [9]: avg_vol=pd.DataFrame({'Provinsi':avg_vol.index, 'Rata-Rata Vol':avg_vol.values, 'Cluster':clusters})

In [10]: centroid_perdata=avg_vol.groupby('Cluster')['Rata-Rata Vol'].transform('mean')
centroid=np.unique(centroid_perdata)

In [11]: sorted_centroid=np.sort(centroid)
Level4, Level3, Level2, Level1 = sorted_centroid[0], sorted_centroid[1], sorted_centroid[2], sorted_centroid[3]
category=centroid_perdata.map({'Level4':Level 4', Level3:'Level 3', Level2:'Level 2', Level1:'Level 1'})
avg_vol['Category']=category
```

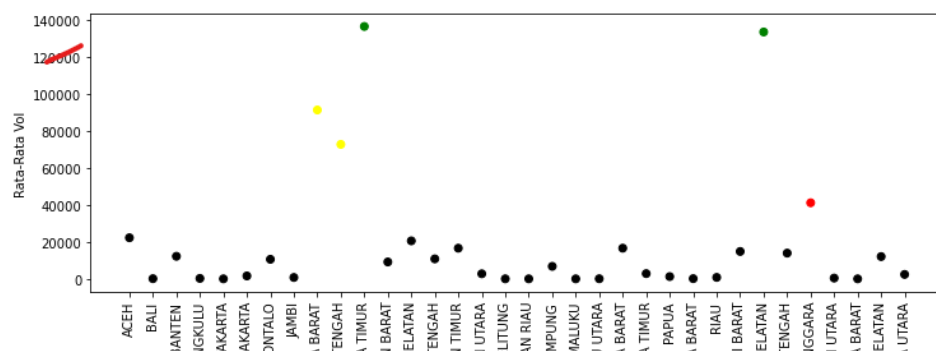
6. Mencetak interpretasi Clustering

```
In [10]: print(avg_vol[["Provinsi", "Category"]])
```

	Provinsi	Category
0	ACEH	Level 4
1	BALI	Level 4
2	BANTEN	Level 4
3	BENGKULU	Level 4
4	DAERAH ISTIMEWA YOGYAKARTA	Level 4

7. Visualisasi interpretasi Clustering

```
In [11]: x=avg_vol.index
y=avg_vol['Rata-Rata Vol']
colors={'Level 4':'black', 'Level 3':'red', 'Level 2':'yellow', 'Level 1':'green'}
fig, ax = plt.subplots(figsize=(12, 4))
ax.scatter(x, y, c=avg_vol['Category'].apply(lambda x:colors[x]))
plt.xlabel('Provinsi')
plt.ylabel('Rata-Rata Vol')
plt.xticks(x, avg_vol['Provinsi'], rotation=90)
plt.show()
```



Latihan 2

1. Diberikan dataset Produksi Lele Provinsi di Google Classroom, silahkan olah dataset tersebut menggunakan metode clustering K-Means dan Agglomerative

Clue:

Ubah bagian yang ditandai dengan dataset yang tersedia

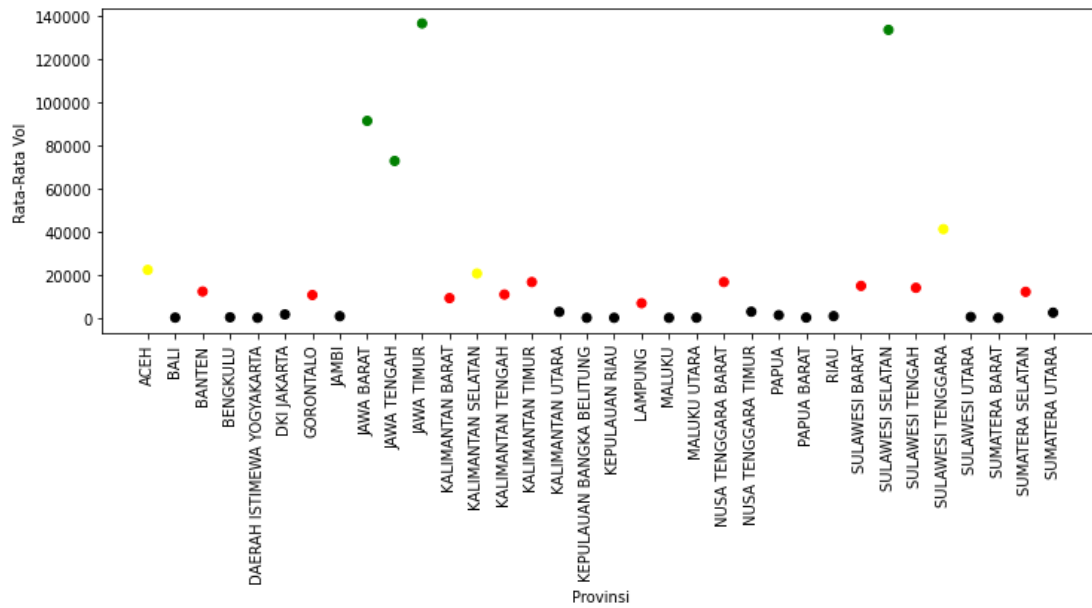
```
dataset = pd.read_csv("C:/Users/hangg/Downloads/Random Aslab/DSDA/Materia/Produksi Bandeng Provinsi.csv", sep=";")
```

Lampiran Screenshot hasil interpretasi Clustering dan visualisasinya

Contoh K-Means

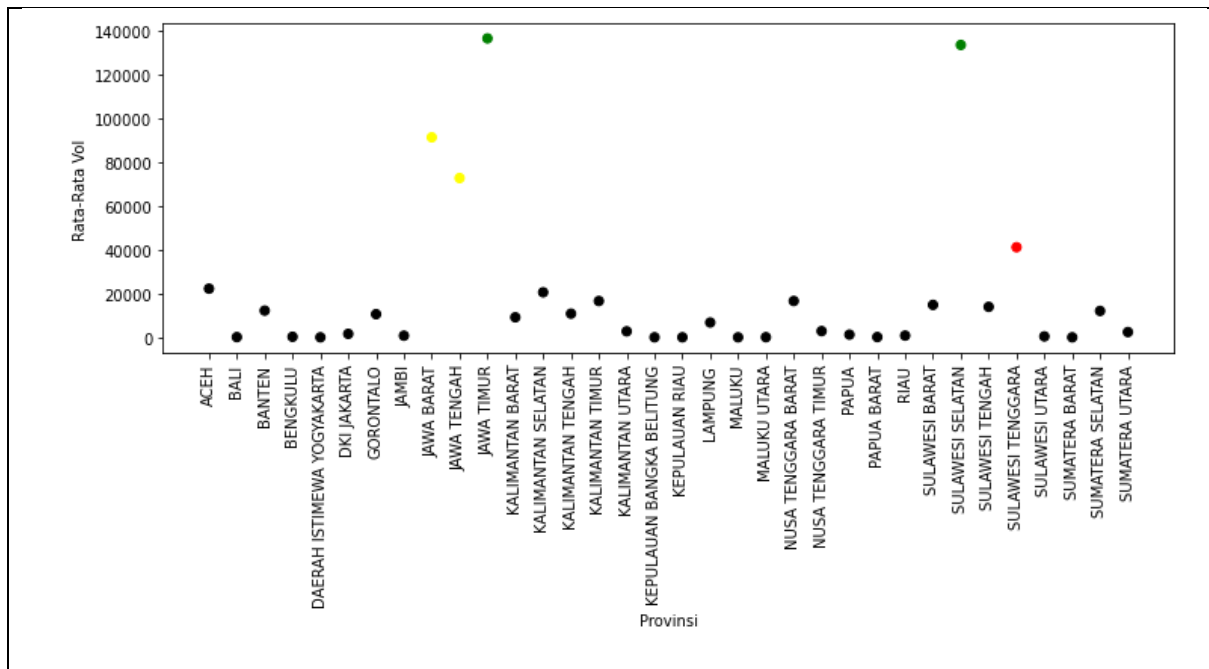
```
In [12]: print(avg_vol[["Provinsi", "Category"]])
```

	Provinsi	Category
0	ACEH	Level 2
1	BALI	Level 4
2	BANTEN	Level 3
3	BENGKULU	Level 4
4	DAERAH ISTIMEWA YOGYAKARTA	Level 4
5	DKI JAKARTA	Level 4
6	GORONTALO	Level 3
7	JAMBI	Level 4
8	JAWA BARAT	Level 1
9	JAWA TENGAH	Level 1
10	JAWA TIMUR	Level 1
11	KALIMANTAN BARAT	Level 3
12	KALIMANTAN SELATAN	Level 2
13	KALIMANTAN TENGAH	Level 3
14	KALIMANTAN TIMUR	Level 3
15	KALIMANTAN UTARA	Level 4
16	KEPULAUAN BANGKA BELITUNG	Level 4
17	KEPULAUAN RIAU	Level 4
18	LAMPUNG	Level 3
19	MALUKU	Level 4
20	MALUKU UTARA	Level 4



Agglomerative

	Provinsi	Category
0	ACEH	Level 4
1	BALI	Level 4
2	BANTEN	Level 4
3	BENGKULU	Level 4
4	DAERAH ISTIMEWA YOGYAKARTA	Level 4
5	DKI JAKARTA	Level 4
6	GORONTALO	Level 4
7	JAMBI	Level 4
8	JAWA BARAT	Level 2
9	JAWA TENGAH	Level 2
10	JAWA TIMUR	Level 1
11	KALIMANTAN BARAT	Level 4
12	KALIMANTAN SELATAN	Level 4
13	KALIMANTAN TENGAH	Level 4
14	KALIMANTAN TIMUR	Level 4
15	KALIMANTAN UTARA	Level 4
16	KEPULAUAN BANGKA BELITUNG	Level 4
17	KEPULAUAN RIAU	Level 4
18	LAMPUNG	Level 4
19	MALUKU	Level 4
20	MALUKU UTARA	Level 4



Bagaimana penjelasan dari hasil clustering yang telah anda olah di atas?

Ketik penjelasan disini