

TUGAS 03

Untuk dataset berikut ini, buat program di python untuk mengcluster data-data tersebut. Lampirkan source code python dan hasil pengelompokan nya.

Objects	X	Y	Z
OB-1	1	4	1
OB-2	1	2	2
OB-3	1	4	2
OB-4	2	1	2
OB-5	1	1	1
OB-6	2	4	2
OB-7	1	1	2
OB-8	2	1	1

Terima kasih.

NAMA : MEIGI RAHMAN
NIM : 202420049

TUGAS 03

Untuk dataset berikut ini, buat program di python untuk mengcluster data-data tersebut. Lampirkan source code python dan hasil pengelompokannya.

Objects	X	Y	Z
OB-1	1	4	1
OB-2	1	2	2
OB-3	1	4	2
OB-4	2	1	2
OB-5	1	1	1
OB-6	2	4	2
OB-7	1	1	2
OB-8	2	1	1

Terima kasih.

Source code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler

d = pd.read_csv("d:\dataset.csv", sep=";")
print(d.head)

d = d.drop(['Objects'], axis=1)
d_x = d.iloc[:,0:4]

x_array = np.array(d_x)
print(x_array)

scaler = MinMaxScaler()
x_scaled = scaler.fit_transform(x_array)
x_scaled

kmeans = KMeans(n_clusters=2, random_state=123)
kmeans.fit(x_scaled)
print(kmeans.cluster_centers_)
print(kmeans.labels_)

d["Cluster"] = kmeans.labels_
print(d.head)

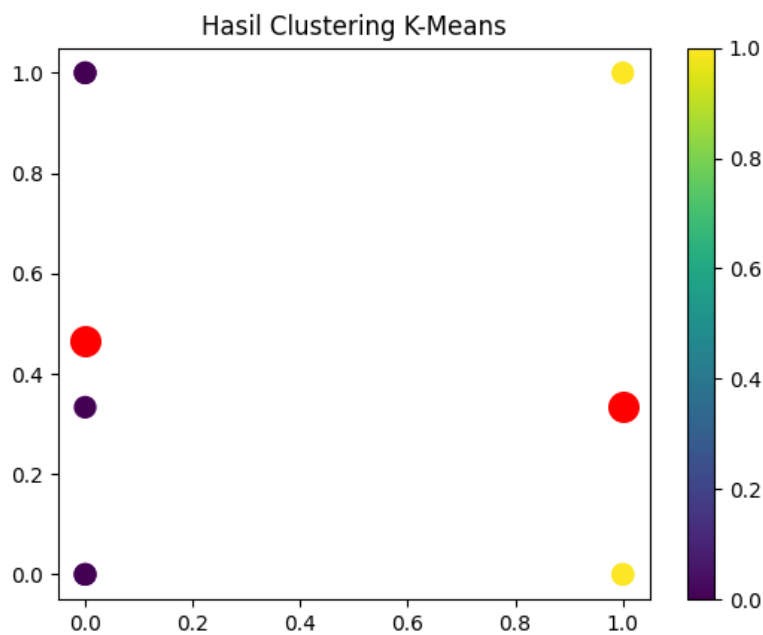
output = plt.scatter(x_scaled[:,0], x_scaled[:,1], s=100, c=d.Cluster, marker="o", alpha=1, )
centers = kmeans.cluster_centers_
plt.scatter(centers[:,0], centers[:,1], c='red', s=200, alpha=1, marker="o");
plt.title("Hasil Clustering K-Means")
plt.colorbar(output)
plt.show()
```

NAMA : MEIGI RAHMAN
NIM : 202420049

Output

```
=== RESTART: C:/Users/Shadow/AppData/Local/Programs/Python/Python39/meigi.py ===  
<bound method NDFrame.head of   Objects  X  Y  Z  
0   OB-1  1  4  1  
1   OB-2  1  2  2  
2   OB-3  1  4  2  
3   OB-4  2  1  2  
4   OB-5  1  1  1  
5   OB-6  2  4  2  
6   OB-7  1  1  2  
7   OB-8  2  1  1>  
[[[1 4 1]  
 [1 2 2]  
 [1 4 2]  
 [2 1 2]  
 [1 1 1]  
 [2 4 2]  
 [1 1 2]  
 [2 1 1]]  
 [[0.         0.46666667 0.6         ]  
 [1.         0.33333333 0.66666667]]  
 [0 0 0 1 0 1 0 1]  
<bound method NDFrame.head of   X  Y  Z  Cluster  
0  1  4  1      0  
1  1  2  2      0  
2  1  4  2      0  
3  2  1  2      1  
4  1  1  1      0  
5  2  4  2      1  
6  1  1  2      0  
7  2  1  1      1>  
|
```

Figure 1



x=0.552 y=0.986

Nama : Meigi Rahman

Nim : 202420049

A ESSAY

1. Menurut Anda apa yang dimaksud dgn Data Mining ? Jelaskan !

Data Mining merupakan suatu proses penambangan data atau proses mengekstrak informasi dari kumpulan data yang sangat besar untuk mengidentifikasi tren, pola, dan data yang berguna menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence terkini yang outputnya data tsb digunakan untuk membuat keputusan.

2. Selain ukuran data yang semakin besar secara eksponensial, alasan apa lagi yang membuat Anda memerlukan data mining ?

Tanpa data mining, kemungkinan proses pengambilan keputusan dari suatu masalah akan lebih sulit dilakukan karena tidak ada data yang bisa menjadi dasar pertimbangan. Sejumlah besar data yang didapat digunakan untuk menghasilkan informasi dan knowledge yang berguna. Informasi yang didapat tsb dapat digunakan pada banyak bidang, seperti pada manajemen bisnis, control produksi, bidang kesehatan, dll

3. Menurut Anda aplikasi apa yang menunjukkan keberhasilan data mining ?

Keberhasilan data mining telah terbukti di beberapa bidang seperti

- Keuangan

Financial Crimes Enforcement Network yang berada di Amerika Serikat memakai data mining untuk mengumpulkan triliunan berbagai subyek, seperti rekening bank, properti, hingga transaksi keuangan lainnya.

Tujuannya untuk mendeteksi transaksi keuangan yang mencurigakan

- Olahraga

Pengumpulan data juga diterapkan pada IBM Advanced Scout untuk analisis statistik permainan NBA sehingga dapat mencapai keunggulan dalam bersaing untuk tim

 - Asuransi

Data mining juga diterapkan di perusahaan asuransi salah satunya Australian Health Insurance Commission untuk identifikasi layanan kesehatan

 4. Sebutkan dan jelaskan secara singkat besungan data mining!

Data mining sangatlah berguna baik bagi pengguna pada level perusahaan atau individu. Data mining pada bisnis digunakan untuk mencari informasi spesifik dari volume data yang besar untuk menemukan solusi bagi masalah bisnis perusahaan. Data mining memiliki kemampuan untuk mengubah data mentah menjadi informasi yang dapat membantu perusahaan tumbuh dengan membuat keputusan berlandaskan pada data.

Data mining digunakan pada berbagai domain termasuk penambangan data untuk citra gambar, teks, penambangan data untuk medsos, website, audio dan video.

 5. Menurut Anda, apakah etika dalam data mining sangat penting?

Tentu, terkadang kita perlu bekerja dengan data yang sifatnya sensitif sehingga perlu berhati-hati dalam melakukan analisis terhadap data agar hasilnya tidak menimbulkan kesalahpahaman. Maka dari itu dibuat Kode Etik agar orang-orang yang bekerja dengan data dapat memperlakukan data yang didapat dengan hati-hati.

B Pilihan Ganda 1. Manakah yang benar mengenai teknik searching : d) Semua jawaban benar 2. Kapankah waktu yang tepat untuk melakukan data reduction : b) Pada saat preprocessing pada rangkaian proses KDD (Knowledge Discovery Databases) 3. Pada saat data preprocessing terdapat beberapa alat dan metode yang berbeda yang digunakan untuk preprocessing antara lain sampling yaitu : a) Menyeleksi subset representatif dari populasi data yang besar 4. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan Hal tsb dilakukan pada tahap a) Interpretation 5. Kebutuhan akan data mining dikarenakan : d) semua jawaban benar

Nama : Mohammad Ilham
Nim : 202420046
Mata Kuliah : Advanced Database

TUGAS 3

Untuk dataset berikut ini, buat program di python untuk mengcluster data-data tersebut. Lampirkan source code python dan hasil pengelompokannya.

Objects	X	Y	Z
OB-1	1	4	1
OB-2	1	2	2
OB-3	1	4	2
OB-4	2	1	2
OB-5	1	1	1
OB-6	2	4	2
OB-7	1	1	2
OB-8	2	1	1

JAWABAN

Disini saya menggunakan Google Colab.

#import package

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
```

#import data

```
from google.colab import files
files.upload()
```

#load data set

```
data = pd.read_csv('t3.csv')
data.head()
```

#melihat informasi dari setiap variabel

```
data.info()
```

```

#menentukan variable mana yang akan di cluster kan
#disini saya mengambil variable X, Y dan Z
data_x = data.iloc[:, 1:4]
data_x.head()

#melihat persebaran data
data.plot(x='X', y='Y', style='o')
plt.title('Persebaran data X dan Y')
plt.xlabel('X')
plt.ylabel('Y')
plt.show()

#merubah variable yang sebelumnya berbentuk data frame menjadi array
x_array = np.array(data_x)
print(x_array)

#melakukan standarisasi
scaler =MinMaxScaler()
x_scaled = scaler.fit_transform(x_array)
x_scaled

#menentukan dan mengkonfigurasi fungsi k-means
kmeans = KMeans(n_clusters = 5, random_state=123)

#menentukan klaster dari data
kmeans.fit(x_scaled)

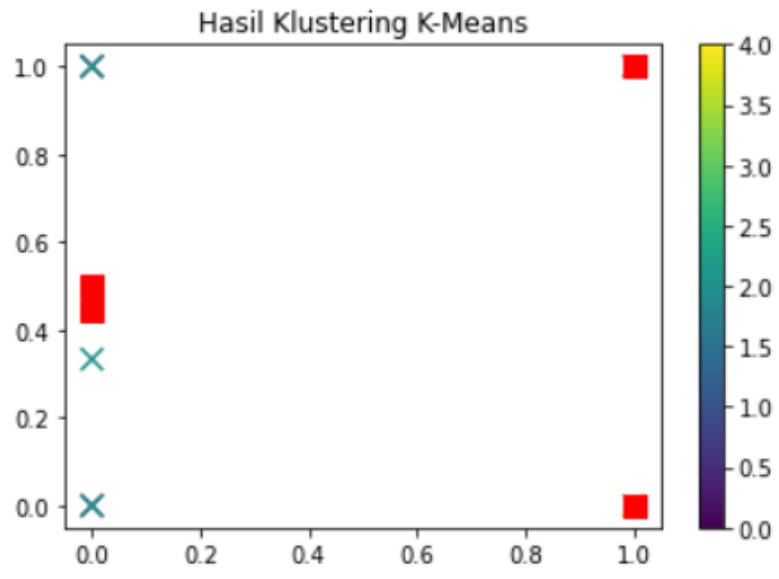
#mencari nilai pusat dari masing-masing kluster
print(kmeans.cluster_centers_)

#menampilkan hasil kluster
print(kmeans.labels_)

#menambahkan kolom "kluster" dalam data frame driver
data["kluster"] = kmeans.labels_

#memvisualisasikan hasil cluster agar nampak hasil nya dengan jelas
output =plt.scatter(x_scaled[:,0], x_scaled[:,1], s = 100, c = data.kluster, marker = "x", alpha =1, )
centers =kmeans.cluster_centers_
plt.scatter(centers[:,0], centers[:,1], c='red', s=100, alpha=1, marker ="s");
plt.title("Hasil Klustering K-Means")
plt.colorbar (output)
plt.show()

```

Nama : Achmad Nopransyah
NIM : 202420050
Program Studi : Magister Teknik Informatika

Soal

Kerjakan tugas yang ada di bagian lampiran. Kumpulkan dalam format ms word sebelum deadline. Terima kasih

Untuk dataset berikut ini, buat program di python untuk mengcluster data-data tersebut. Lampirkan source code python dan hasil pengelompokkannya.

Objects	X	Y	Z
OB-1	1	4	1
OB-2	1	2	2
OB-3	1	4	2
OB-4	2	1	2
OB-5	1	1	1
OB-6	2	4	2
OB-7	1	1	2
OB-8	2	1	1

Terima kasih.

JAWAB

Program python

```
In [11]: model.labels_
```

```
Out[11]: array([1, 2, 1, 0, 2, 1, 2, 0])
```

```
In [12]: import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

data = pd.read_excel('data.xlsx', sheet_name='Sheet1')
data.head()
target = data['objects']
data = data.drop('objects', axis=1)
```

```
In [13]: kmeans = KMeans(n_clusters=3)
kmeans
```

```
Out[13]: KMeans(n_clusters=3)
```

```
In [14]: model = kmeans.fit(data)
centroids = kmeans.cluster_centers_
```

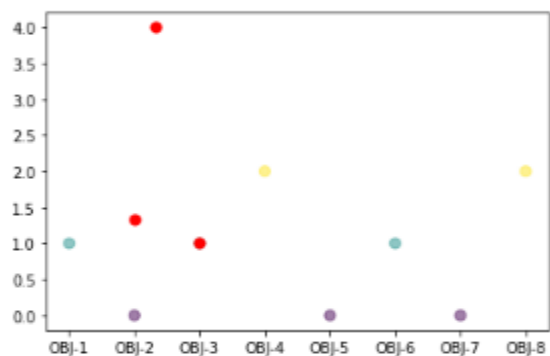
```
In [15]: pd.crosstab(target,model.labels_)
```

```
Out[15]:
```

col_0	0	1	2
objects			
OBJ-1	0	1	0
OBJ-2	1	0	0
OBJ-3	0	1	0
OBJ-4	0	0	1
OBJ-5	1	0	0
OBJ-6	0	1	0
OBJ-7	1	0	0
OBJ-8	0	0	1

Hasil Cluster

```
In [16]: plt.scatter(target, model.labels_, c= model.labels_.astype(float), s=50, alpha=0.5)
plt.scatter(centroids[:, 0], centroids[:, 1], c='red', s=50)
plt.show()
```

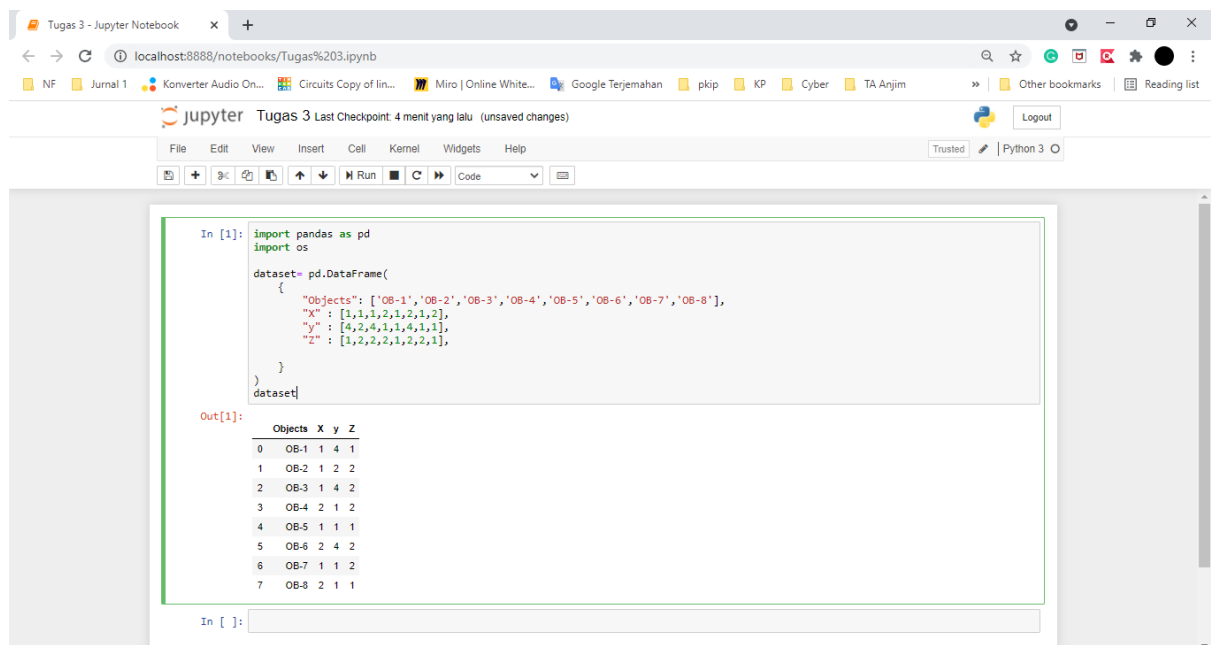


TUGAS 03

Untuk dataset berikut ini, buat program di python untuk mengcluster data-data tersebut. Lampirkan source code python dan hasil pengelompokannya.

Objects	X	Y	Z
OB-1	1	4	1
OB-2	1	2	2
OB-3	1	4	2
OB-4	2	1	2
OB-5	1	1	1
OB-6	2	4	2
OB-7	1	1	2
OB-8	2	1	1

jawaban



```
In [1]: import pandas as pd
import os

dataset= pd.DataFrame(
    {
        "Objects": ['OB-1','OB-2','OB-3','OB-4','OB-5','OB-6','OB-7','OB-8'],
        "X": [1,1,1,2,1,2,1,2],
        "Y": [4,2,4,1,1,4,1,1],
        "Z": [1,2,2,2,1,2,2,1],
    }
)
dataset

Out[1]:
```

	Objects	X	Y	Z
0	OB-1	1	4	1
1	OB-2	1	2	2
2	OB-3	1	4	2
3	OB-4	2	1	2
4	OB-5	1	1	1
5	OB-6	2	4	2
6	OB-7	1	1	2
7	OB-8	2	1	1

```
In [ ]:
```

Source code:

```
import pandas as pd
import os
```

```
dataset= pd.DataFrame(
    {
        "Objects": ['OB-1','OB-2','OB-3','OB-4','OB-5','OB-6','OB-7','OB-8'],
        "X" : [1,1,1,2,1,2,1,2],
        "y" : [4,2,4,1,1,4,1,1],
        "Z" : [1,2,2,2,1,2,2,1],
    }
)
dataset
```

TUGAS 03

Untuk dataset berikut ini, buat program di python untuk mengcluster data-data tersebut. Lampirkan source code python dan hasil pengelompokannya.

Objects	X	Y	Z
OB-1	1	4	1
OB-2	1	2	2
OB-3	1	4	2
OB-4	2	1	2
OB-5	1	1	1
OB-6	2	4	2
OB-7	1	1	2
OB-8	2	1	1

Terima kasih.

Jawab :

1. *Import package* dari *sklearn* (scikit-learn) untuk melakukan *clustering*

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="darkgrid")
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
```

2. Membaca data *clusterxyz* yang sudah di *upload*

```
data = pd.read_csv("sample_data/clusterxyz.csv")
data.head()
```

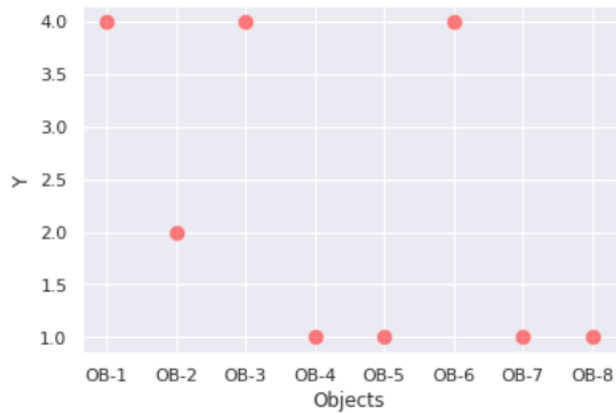
```
  Objects  X  Y  Z
0  OB-1   1  4  1
1  OB-2   1  2  2
2  OB-3   1  4  2
3  OB-4   2  1  2
4  OB-5   1  1  1
```

3. Melihat Sebaran Data

Melihat Sebaran Data

```
sns.scatterplot(x="Objects", y="Y", data=data, s=100, color="red", alpha = 0.5)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fbc9b025810>
```



4. Menentukan variabel kluster

#menentukan variabel kluster

```
data_x = data.iloc[:, 1:4]
```

```
data_x.head()
```

```
   x  y  z
0  1  4  1
1  1  2  2
2  1  4  2
3  2  1  2
4  1  1  1
```

5. Membuat Array

#membuat array

```
x_array = np.array(data_x)
```

```
print(x_array)
```

```
[[1 4 1]
 [1 2 2]
 [1 4 2]
 [2 1 2]
 [1 1 1]
 [2 4 2]
 [1 1 2]
 [2 1 1]]
```

6. Menstandarkan ukuran variabel, Membuat Fungsi K-Means, Menentukan Kluster Data

```
# menstandarkan ukuran variabel
scaler = MinMaxScaler()
x_scaled = scaler.fit_transform(x_array)

# membuat fungsi k-means
kmeans = KMeans(n_clusters = 5, random_state=123)

# Menentukan kluster dari data
kmeans.fit(x_scaled)

KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
        n_clusters=5, n_init=10, n_jobs=None, precompute_distances='auto',
        random_state=123, tol=0.0001, verbose=0)
```

7. Menampilkan Pusat Kluster

```
# Menampilkan pusat cluster
print(kmeans.cluster_centers_)

[[0.         0.5         0.         ]
 [1.         1.         1.         ]
 [0.         0.44444444 1.         ]
 [1.         0.         1.         ]
 [1.         0.         0.         ]]
```

8. Menampilkan Hasil Kluster

```
# Menampilkan hasil kluster
print(kmeans.labels_)

[0 2 2 3 0 1 2 4]
```

9. Menambahkan kolom Kluster

	Objects	X	Y	Z	kluster
0	OB-1	1	4	1	0
1	OB-2	1	2	2	2
2	OB-3	1	4	2	2
3	OB-4	2	1	2	3
4	OB-5	1	1	1	0