

MACHINE LEARNING

Nama : Hanif Ridal Warits

NPM : 41155050210060

Kelas : Informatika A2 – 2021

Tugas Pertemuan 9

Clustering (Studi Kasus K-Means) ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

RAM Disk Gemini

CLUSTERING WITH K-MEANS

[1] print('Hanif Ridal Warits - 41155050210060')

Hanif Ridal Warits - 41155050210060

STUDI KASUS K-MEANS

Tabel Data Konsumen

No	Gaji	Pengeluaran
1	2500	1750
2	3800	4200
3	3900	3800
4	4350	5500
5	4400	3200
6	5500	5450
7	5600	5950
8	5750	4100
9	6850	6050
10	6900	8500
11	7250	9500
12	7350	6050
13	7500	8500
14	7800	9500

15	8200	8300
16	8500	6500
17	8550	8400
18	8750	6000
19	9100	10500
20	9100	8500

Sebuah perusahaan melakukan penelitian terhadap data-data konsumen yang dimilikinya. Perusahaan tersebut akan melakukan pengelompokan data ke dalam beberapa *cluster* berdasarkan kriteria besaran gaji yang diterima dan pengeluaran per bulannya. Berikut adalah data-data 20 konsumen dari perusahaan tersebut.

Studi kasus kali ini akan melakukan clustering terhadap data-data konsumen di atas ke dalam beberapa kelompok, dimana masing-masing kelompok memiliki tingkat kemiripan maksimum. Tujuan penelitian ini adalah agar perusahaan dapat memetakan jenis produk yang sesuai dengan karakteristik konsumen.

LANGKAH 1 KONEKSI KE DATABASE

[3] from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[5] # cek isi GDrive
!ls drive/'My Drive'/'Colab Notebooks'

'Clustering (Studi Kasus K-Means)'
decision-tree-classifier-tutorial.ipynb
konsumen.csv
Problem_A1.ipynb
Problem_A2.ipynb
Problem_A3.ipynb
Problem_A4.ipynb
Problem_A5.ipynb
Problem_B1.ipynb
Problem_B2.ipynb
Problem_B3.ipynb

04

✓

🔍

🔗

{x}

🔍

📁

Problem B4.ipynb
Problem B5.ipynb
Problem_C1.ipynb
Problem_C2.ipynb
Problem_C3.ipynb
Problem_C4.ipynb
Problem_C5.ipynb
'SKLearn 02 | Sample Dataset pada Scikit-Learn | Belajar Machine Learning Dasar.ipynb'
'SKLearn 03 | Machine Learning Workflow dengan Scikit-Learn | Belajar Machine Learning Dasar.ipynb'
'SKLearn 04 | Data Preprocessing dengan Scikit-Learn | Belajar Machine Learning Dasar.ipynb'
'SKLearn 05 | Simple Linear Regression dengan Scikit-Learn | Belajar Machine Learning Dasar.ipynb'
'SKLearn 06 Classification dengan KNN (K Nearest Neighbours).ipynb'
'SKLearn 08 | Multiple Linear Regression & Polynomial Regression | Belajar Machine Learning Dasar.ipynb'
'SKLearn 12 | Logistic Regression pada Binary Classification Task | Belajar Machine Learning Dasar.ipynb'
'SKLearn 13 - Naive Bayes Classification.ipynb'
'SKLearn 14 Classification Task dengan Support Vector Machine (SVM).ipynb'
'SKLearn 15 - Decision Tree.ipynb'
'SKLearn 16 - Random Forest.ipynb'
Untitled
Untitled2.ipynb

LANGKAH 2 IMPORT LIBRARY YANG DIPERLUKAN

Matplotlib = library yang digunakan untuk membuat grafik plot sesuai kebutuhan.

NumPy = digunakan untuk kebutuhan scientific.

Pandas = library yang digunakan untuk manipulasi data seperti membuat tabel, mengubah dimensi data, mengecek data, dsb

Sklearn = library untuk berbagai metode dan algoritma yang digunakan dalam machine learning

05

✓

🔍

🔗

{x}

🔍

📁

Import Library yang akan digunakan
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans

LANGKAH 3 MENYIAPKAN DATASET

Menyiapkan data dan memanggil dataset
dataset = pd.read_csv('drive/My Drive/Colab Notebooks/konsumen.csv')
dataset.keys()

Index(['gaji', 'pengeluaran'], dtype='object')

Menampilkan 5 baris data pertama dari dataset tersebut
dataku = pd.DataFrame(dataset)
dataku.head()

	gaji	pengeluaran
0	2500	1750
1	3800	4200
2	3900	3800
3	4350	5500
4	4400	3200

Next steps: [Generate code with dataku](#) [View recommended plots](#) [New interactive sheet](#)

06

✓

🔍

🔗

{x}

🔍

📁

Konversi ke data array
X = np.asarray(dataset)
print(X)

[[2500 1750]
[3800 4200]
[3900 3800]
[4350 5500]
[4400 3200]
[5500 5450]
[5600 5950]
[5750 4100]
[6850 6050]
[7250 9500]
[7350 6050]
[7500 8500]
[7800 9500]
[8200 8300]
[8500 6500]
[8550 8400]
[8750 6000]
[9100 10500]
[9100 8500]]

07

✓

🔍

🔗

{x}

🔍

📁

Menampilkan data dalam bentuk scatter plot
plt.scatter(X[:, 0], X[:, 1], label='True Position')
plt.xlabel("Gaji")
plt.ylabel("Pengeluaran")
plt.title("Grafik Konsumen")

08

✓

🔍

🔗

{x}

🔍

📁

plt.legend()
plt.show()

LANGKAH 4 MENGGUNAKAN LIBRARY K-MEANS

```
[12] # Mengaktifkan K-Means dengan jumlah K = 2
kmeans = KMeans(n_clusters=2)
kmeans.fit(X)
```

```
KMeans
```

```
KMeans(n_clusters=2)
```

```
[13] # Menampilkan nilai centroid yang di-generate oleh algoritma
print(kmeans.cluster_centers_)
```

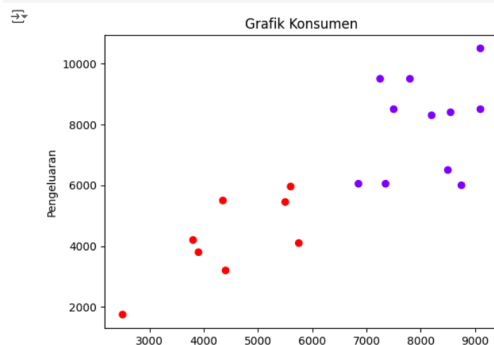
```
[[8086.363636 7981.818182]
 [4475.      4244.875   ]]
```

```
[14] # Menampilkan label data point
print(kmeans.labels_)
```

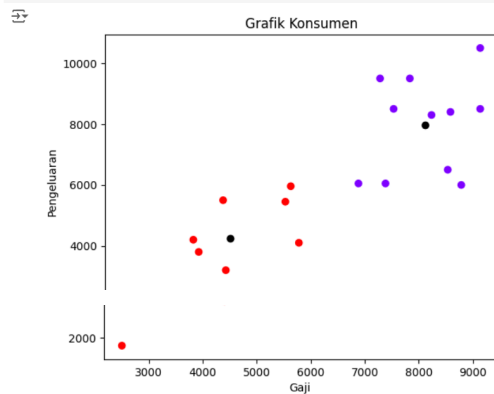
```
[1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0]
```

LANGKAH 5 MENAMPILKAN OUTPUT

```
[15] # Plot data point
# Memvisualisasikan bagaimana telah di-klasterisasi
plt.scatter(X[:, 0], X[:, 1], c=kmeans.labels_, cmap='rainbow')
plt.xlabel("Gaji")
plt.ylabel("Pengeluaran")
plt.title("Grafik Konsumen")
plt.show()
```



```
[16] # Plot data point
# Memvisualisasikan bagaimana telah di-klasterisasi.
# Menampilkan centroids dengan warna hitam.
plt.scatter(X[:, 0], X[:, 1], c=kmeans.labels_, cmap='rainbow')
plt.scatter(kmeans.cluster_centers[:, 0], kmeans.cluster_centers[:, 1], color='black')
plt.xlabel("Gaji")
plt.ylabel("Pengeluaran")
plt.title("Grafik Konsumen")
plt.show()
```



```
[17] print('Hanif Ridal Warits - 41155050210060')
```

```
Hanif Ridal Warits - 41155050210060
```