



Minggu ke-9

## Praktikum Cluster Analysis

Ali Ridho Barakbah, Entin Martian

Knowledge Engineering Research Group  
Department of Information and Computer Engineering  
Politeknik Elektronika Negeri Surabaya



```
import pandas as pd
from sklearn.cluster import KMeans

dataset = pd.read_csv('ruspini.csv')
data = dataset.loc[:,['X', 'Y']]

clustering = KMeans(n_clusters=4, init="random", n_init=1)
clusters=clustering.fit_predict(data)

print('\nHasil clustering:\n', clusters)

print('\nSSE = : \n', clustering.inertia_)
```

[illegible]

SSE = :  
12881.05123614663

# Transaction Dataset

InvoiceNo	StockCode	Qty	InvoiceDate	CustomerID	Country
537626	22725	830	12/7/10 14:57	12347	Iceland
537626	22729	948	12/7/10 14:57	12347	Iceland
537626	22195	695	12/7/10 14:57	12347	Iceland
542237	22725	636	1/26/11 14:30	12347	Iceland
542237	22729	536	1/26/11 14:30	12347	Iceland
542237	47559	919	1/26/11 14:30	12347	Iceland
542237	21154	803	1/26/11 14:30	12347	Iceland
542237	21035	532	1/26/11 14:30	12347	Iceland
...	...		...	...	...

# Assignment #

---

1. dataset  $\leftarrow$  transaction.csv, dan tampilkan
2. country  $\leftarrow$  berapa kemunculan tiap negeri pada dataset, dan tampilkan
3. transaksi  $\leftarrow$  hitunglah banyaknya rata-rata jumlah barang (Qty) per transaksi pada tiap negara (1 kode InvoiceNo = 1 transaksi)
4. cluster\_i[1-10], cluster\_val[1-10]  $\leftarrow$  lakukan clustering pada transaksi dengan K-Means, dengan k=3, sebanyak 10 kali. Setiap kali selesai clustering, lakukan cluster analysis dengan SSE.
5. cluster  $\leftarrow$  ambil cluster\_i yang mempunyai cluster\_val terkecil
6. centroid  $\leftarrow$  tentukan posisi centroid dari setiap cluster
7. sorted  $\leftarrow$  lakukan pengurutan posisi centroid secara ascending
8. Indeks terdepan dari centroid setelah pengurutan, mengindikasikan cluster transaksi rendah. Indeks terakhir dari centroid setelah pengurutan, mengindikasikan cluster transaksi tinggi. Indeks di antaranya, mengindikasikan cluster transaksi sedang. Tampilkan negara mana saja yang transaksinya rendah, sedang dan tinggi.
9. Visualisasi dengan warna yang berbeda untuk hasil cluster (no. 8), dimana sumbu x=urutan country dan sumbu y=transaksi



# Pengumpulan Tugas

---

- Buatlah coding dengan Bahasa pemrograman/tools apapun untuk semua assignment
- Buatlah laporan dalam slide ppt. Laporan terdiri dari screenshot coding dan hasil running untuk setiap assignment.
- Simpan laporan dalam file pdf dengan format penamaan:  
DM\_M9\_NRP\_namadepan.pdf
- Deadline upload di ethol: Kasmis, 17 Oktober 2024 pk 18.00