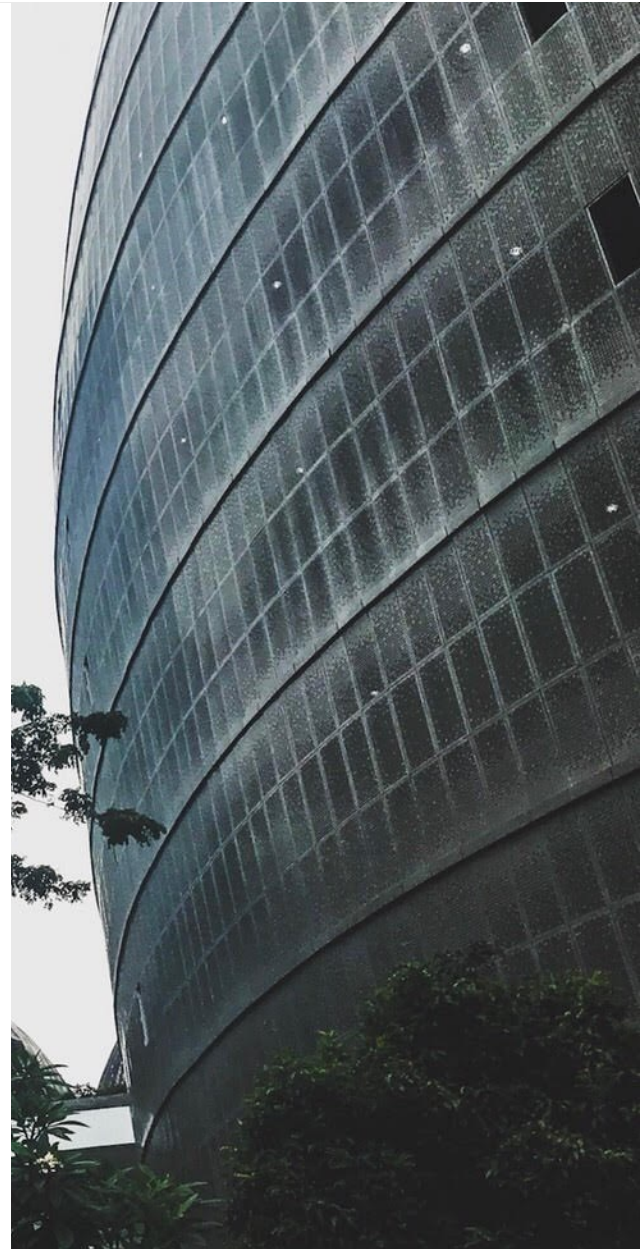


MODUL PRAKTIKUM

IS411 – DATA MODELING
PROGRAM SARJANA S1 SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

Gedung B Lantai 5, Kampus UMN
Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten-15811 Indonesia
Telp: +62-21.5422.0808 (ext. 1803), email: ict.lab@umn.ac.id, web: umn.ac.id

MODUL 11

K-MEANS



DESKRIPSI TEMA

1. K-Means Algorithm using Python

CAPAIAN PEMBELAJARAN MINGGUAN (SUB-CAPAIAN PEMBELAJARAN)

Students are able to illustrate data modeling problems using classification techniques (C4)

PENUNJANG PRAKTIKUM

1. Anaconda Navigator
 2. Jupyter Notebook
- (+ Perlengkapan Apd/Alat Pelindung Diri Yang Harus Digunakan, Jika Ada)

LANGKAH-LANGKAH PRAKTIKUM

Import Library

1. Lakukan langkah yang sama seperti minggu sebelumnya untuk meng-import library Numpy dan Pandas dan juga matplotlib.pyplot

Import Data

2. Gunakan dataset **customer.csv** dan masukkan ke dalam dataframe menggunakan Pandas. Berikan nama **X**
3. Tampilkan informasi dataset dan isi data.

Menggunakan metode elbow untuk menentukan angka cluster yang tepat

Clustering adalah salah satu teknik analisis dan eksplorasi data yang paling sering digunakan untuk mendapatkan pola suatu struktur data. Clustering juga dapat dianalogikan sebagai tugas mengidentifikasi subkelompok dalam data sedemikian rupa sehingga titik data dalam subkelompok yang sama (cluster) sangat mirip sedangkan titik data dalam cluster yang berbeda sangat berbeda. Dengan kata lain, clustering adalah suatu Teknik untuk kelompok data yang homogen sedemikian rupa sehingga titik data di setiap cluster semirip mungkin menurut ukuran kesamaan seperti jarak berbasis euclidean atau jarak berbasis korelasi. Keputusan tentang ukuran kemiripan dapat ditentukan melalui jumlah centroid masing-masing kelompok dan jaraknya. Analisis clustering dapat dilakukan berdasarkan fitur yang telah ditentukan untuk menemukan subkelompok berdasarkan fitur atau berdasarkan sampel.

4. Menggunakan metode elbow untuk menentukan angka cluster yang tepat.

```

10 # Menggunakan metode elbow untuk menentukan angka cluster yang tepat
11 from sklearn.cluster import KMeans
12 wcss = []
13 for i in range(1, 11):
14     kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
15     kmeans.fit(X)
16     wcss.append(kmeans.inertia_)
17 plt.plot(range(1, 11), wcss)
18 plt.title('Elbow Method')
19 plt.xlabel('Cluster Number')
20 plt.ylabel('WCSS')
21 plt.show()

```

5. Menjalankan K-Means Clustering ke dataset.

```

23 # Menjalankan K-Means Clustering ke dataset
24 kmeans = KMeans(n_clusters = 5, init = 'k-means++', random_state = 42)
25 y_kmeans = kmeans.fit_predict(X)
26

```

6. Visualisasi hasil clusters.

```

27 # Visualisasi hasil clusters
28 plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'blue', label = 'Cluster 1')
29 plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'red', label = 'Cluster 2')
30 plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'magenta', label = 'Cluster 3')
31 plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Cluster 4')
32 plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c = 'green', label = 'Cluster 5')
33 plt.scatter(kmeans.cluster_centers[:, 0], kmeans.cluster_centers[:, 1], s = 300, c = 'yellow', label = 'Centroids')
34 plt.title('Consumers Cluster')
35 plt.xlabel('Yearly Salary')
36 plt.ylabel('Yearly expense rating (1-100)')
37 plt.legend()
38 plt.show()

```

Challenge

1. Sebuah department store ingin menganalisa kelompok perilaku belanja dari pelanggannya, dengan data set yang diberikan buatlah analisa:

- gunakan kolom ke 2 hingga 4 sebagai input features
- dengan metode elbow, analisa jumlah cluster yang tepat
- gunakan K-Means clustering dan analisa hasilnya

2. Sebuah department store ingin menganalisa kelompok perilaku belanja dari pelanggannya, dengan data set yang diberikan buatlah analisa:

- gunakan kolom ke 2 hingga 4 sebagai input features

- b. dengan dendogram diagram, analisa jumlah cluster yang tepat
 - c. gunakan Hierachical clustering dan analisa hasilnya
 - d. bandingkan hasilnya dengan menggunakan metode K-Means
3. Generate data set sebanyak 500 titik data pelatihan bola dengan label yang sesuai
 - a. lakukan normalisasi fitur pada proses pelatihan data,
 - b. gunakan DBSCAN dari library sklearn
 - c. buatlah untuk kasus spherical dan non spherical data
 - d. Pada kasus non spherical data, uji coba dengan metode K Means dan bandingkan hasilnya.

PENGUMPULAN

1. File yang dikumpulkan terdiri dari:
 - a. File project (.ipynb)
 - b. File PDF berisi screenshot output dan jawaban (jika ada pertanyaan)
2. File di-compress (.zip) dan diberi nama **KODEMATAKULIAH_KELAS_NIM_NAMA_WEEK KE-XX.zip**
(contoh: IS5411_A_13110310017_Monika Evelin Johan_Week-01.zip).

REFERENSI

Deitel, P., & Deitel, H. (2020). Intro to Python for Computer Science and Data Science. Pearson Education.