



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,  
RISET DAN TEKNOLOGI  
**UNIVERSITAS DIPONEGORO**  
FAKULTAS SAINS DAN MATEMATIKA  
DEPARTEMEN ILMU KOMPUTER/  
INFORMATIKA

Jalan Prof. Jacob Rais  
Kampus Universitas Diponegoro Tembalang  
Semarang Kode Pos 50275  
Telp (024) 7474754  
Laman : [if@live.undip.ac.id](mailto:if@live.undip.ac.id)  
Pos-el: [fsm\[at\]undip.ac.id](mailto:fsm[at]undip.ac.id)

**UJIAN AKHIR SEMESTER GASAL 2022/2023**

Mata Kuliah	:	Pembelajaran Mesin
Kelas	:	A/B/C
Pengampu	:	Dr. Retno Kusumaningrum, S.Si, M.Kom. Rismiyati, B.Eng, M.Cs
Departemen	:	Informatika
Program Studi	:	S1 Informatika
Hari / Tanggal	:	Selasa, 13 Desember 2022
Jam / Ruang	:	07.30-09.10/ E101(62), E103 (51), A105(55)
Sifat Ujian	:	Buku tertutup

**Petunjuk Pengerjaan :**

Jawablah soal-soal berikut pada lembar jawab! Tulis penjelasan anda jangan langsung hasil akhir!

```
df.head()
```

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

**1. {Bobot 20%}**

(a) Jika akan dilakukan clustering terhadap sebuah data set dengan menggunakan K-means. Contoh 5 data teratas yang ada pada data set ditunjukkan pada tabel di atas. Apakah normalisasi diperlukan dalam kasus ini? Jelaskan jawaban anda! (harus menjelaskan apa itu normalisasi dan kenapa iya atau tidak diperlukan!) (10)

(b) Jelaskan cara kerja K-means Clustering dengan terlebih dahulu menuliskan algoritma tersebut dengan kata-kata anda sendiri. (10)

**2. {Bobot 20%}**

Diketahui 6 buah objek yang akan dikluster menggunakan algoritma *agglomerative hierrarchical clustering (AHC)*. Dengan menggunakan matriks jarak pada tabel berikut, lakukanlah clustering dan gambarkan dendogramnya menggunakan **complete** linkage AHC.

	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

3. {Bobot 20%}

Clustering dilakukan terhadap 8 objek A B C D E F G H. Evaluasi internal akan dilakukan dengan menghitung silhouette coefficient. Data perhitungan jarak rata-rata masing-masing objek dengan objek lain ditunjukkan pada tabel di bawah. Hitunglah silhouette coefficient dari hasil clustering tersebut!

Berdasar perhitungan tersebut, menurut anda, apakah cluster yang dihasilkan baik? Jelaskan dengan menyebutkan alasan anda.

	Cluster	jarak rata-rata terhadap object pada cluster yang sama	jarak antara object dengan object pada cluster lain
A	1	1.50	4.00
B	1	2.00	5.00
C	2	0.75	3.00
D	2	5.00	6.00
E	2	1.75	5.00
F	1	0.50	4.00
G	2	2.00	3.50
H	1	5.10	5.00

4. {Bobot 20%}

Diketahui 6 buah data dengan 2 atribut yang direpresentasikan dalam matriks A, akan dilakukan reduksi dimensi menggunakan algoritma *principal component analysis* (PCA).

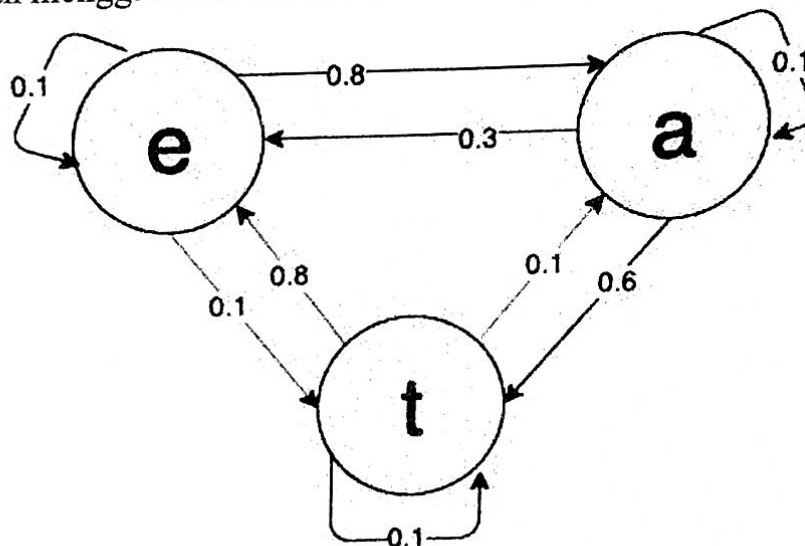
$$A = \begin{bmatrix} 3 & 2 \\ 4 & 4 \\ 1 & 6 \\ 2 & 2 \\ 5 & 4 \\ 7 & 6 \end{bmatrix}$$

Eigenvalue dan eigenvector yang didapat dari matriks kovarian dari data tersebut adalah  
Eigenvalue = (2.52, 5.34) dan eigenvector =  $\begin{bmatrix} -0.49 & 0.87 \\ 0.87 & 0.49 \end{bmatrix}$

Reduksilah data pada matriks A menjadi 1 dimensi menggunakan *principal component Analysis*.  
Tulis langkah serta analisa anda dengan jelas!

(Representasikan dalam matriks ukuran 6 x 1). {Bobot: 20%}

5. Dengan menggunakan markov chain



Jika probabilitas huruf pertama =e adalah 50 %, a=30% dan t=20%, berapakah probabilitas terbentuknya kata **ate** dan **eta** ) ? {Bobot: 20%}

Apa kesimpulan yang bisa anda dapatkan dari hasil perhitungan tadi.

---

~When the going gets tough, the tough gets going~

Selamat mengerjakan dan semoga sukses.