

Jawaban belum tentu benar !!
Jadikan referensi jangan jadikan acuan

PEMBAHASAN UAS PM 2019

1. A) Data clustering menggunakan metode K-Means Clustering ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

- Tentukan jumlah cluster
- Alokasikan data ke dalam cluster secara random
- Hitung centroid/rata-rata dari data yang ada di masing-masing cluster
- Alokasikan masing-masing data ke centroid/rata-rata terdekat
- Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan

B) Cluster terbaik adalah dengan jumlah $k = 4$. Karena pada cluster $k=4$ terjadi penurunan SSE secara drastis dan setelahnya tidak ada perubahan signifikan.

2. Complete LinkAge

Complete Linkage = Mengisi jarak masing-masing data dengan nilai maksimal

	1	2	3	4	5
1	0	9	3	6	11
2	9	0	7	5	10
3	3	7	0	9	2
4	6	5	9	0	8
5	11	10	2	8	0

Mencari jarak dua kelompok terkecil

Iterasi 1 jarak terkecil = 2

Maka kelompok iterasi 1 adalah (3,5)

Kemudian mengisi tabel dengan nilai maksimal D

Jawaban belum tentu benar !!
 Jadikan referensi jangan jadikan acuan

	1	2	3,5	4
1	0	9	11	6
2	9	0	10	5
3,5	11	10	0	9
4	6	5	9	0

Mencari jarak dua kelompok terkecil

Iterasi 2 jarak terkecil = 5

Maka kelompok iterasi 1 adalah (2,4)

Kemudian mengisi tabel dengan nilai maksimal D

	1	2,4	3,5
1	0	9	11
2,4	9	0	10
3,5	11	10	0

Mencari jarak dua kelompok terkecil

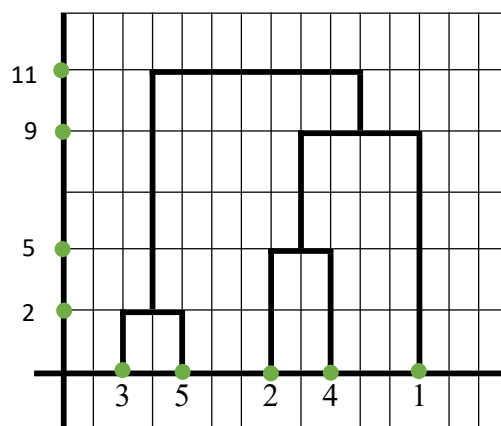
Iterasi 3 jarak terkecil = 9

Maka kelompok iterasi 1 adalah (1,2,4)

Kemudian mengisi tabel dengan nilai maksimal D

	2,4,1	3,5
2,4,1	0	11
3,5	11	0

Maka Dendrogram didapatkan seperti berikut



Jawaban belum tentu benar !!
Jadikan referensi jangan jadikan acuan

3. A)

B)

Cluster Cohesion: Measures how closely related are objects in a cluster

Example: SSE

Cluster Separation: Measure how distinct or well-separated a cluster is from other clusters

Example: Squared Error

Cohesion is measured by the within cluster sum of squares (SSE)

$$WSS = \sum_i \sum_{x \in C_i} (x - m_i)^2$$

Separation is measured by the between cluster sum of squares

$$BSS = \sum |C_i| (m - m_i)^2$$

Where $|C_i|$ is the size of cluster i

Cohesion: menentukan kualitas clustering dengan mengukur seberapa dekat masing-masing objek di dalam suatu cluster yang dihasilkan, semakin dekat maka semakin baik

Separation: menentukan kualitas cluster dengan mengukur seberapa jauh separasi antar cluster

4. A) Saya tidak setuju, karena semakin banyak fitur belum tentu hasil yang didapatkan akan maksimal. Hal tersebut disebabkan karena adanya curse dimensionality. Jadi banyaknya fitur tidak membuat hasil yang didapatkan akan lebih baik, namun bahkan bisa saja performa yang didapatkan malah menurun.

B)

5.