



Subject : Data Mining
Lecturer : Hironimus Leong, S.Kom, M.Kom
Date/Time : November 2nd 2020

Buatlah SQL untuk menghitung algoritma K-Means. Tulis dalam 1 file SQL. Hasil output dalam soal ini hanya contoh, jika Anda mempunyai format output lain yang lebih efisien dan lebih sederhana, silahkan menggunakannya tidak harus sama dengan yang ada dalam soal ini.

NOMOR 1 (10 pt):

Data dibuat acak sesuai dengan permintaan user. Nilai X dan nilai Y acak, dengan nilai X dan Y berada di range nilai 30 sampai dengan 80 (tidak boleh di luar range ini). Contoh data acak sebanyak 25 dengan range nilai 30 SD 80:

datake	x	y
1	49.00	55.00
2	48.00	45.00
3	51.00	40.00
4	66.00	33.00
5	38.00	63.00
6	69.00	78.00
7	53.00	53.00
8	76.00	41.00
9	48.00	37.00
10	63.00	73.00
11	48.00	39.00
12	73.00	68.00
13	40.00	67.00
14	37.00	52.00
15	73.00	77.00

NOMOR 2 (10 PT):

Nilai K untuk jumlah centroid bisa ditentukan oleh user, dengan nilai X dan Y adalah nilai acak. Contoh nilai K adalah 3 dimana setiap titik adalah acak.

centroid	x	y
1	32.00	45.00
2	40.00	57.00
3	72.00	21.00

Nomor 3 (30 PT):

Buatlah proses untuk hitungan iterasi pertama, menghitung kedekatan data dengan metode Euclidean Distance, kemudian tentukan data masuk ke centroid yang mana.

iterasi	datake	x	y	jarak	centroid
1	1	49.00	55.00	C1: (32.00,45.00)=19.72; C2: (40.00,57.00)=19.72; C3: (72.00,21.00)=41.05;	2
1	2	48.00	45.00	C1: (32.00,45.00)=16.00; C2: (40.00,57.00)=16.00; C3: (72.00,21.00)=33.94;	2
1	3	51.00	40.00	C1: (32.00,45.00)=19.65; C2: (40.00,57.00)=20.25; C3: (72.00,21.00)=28.32;	1
1	4	66.00	33.00	C1: (32.00,45.00)=36.06; C2: (40.00,57.00)=36.06; C3: (72.00,21.00)=36.06;	3
1	5	38.00	63.00	C1: (32.00,45.00)=18.97; C2: (40.00,57.00)=18.97; C3: (72.00,21.00)=54.04;	2
1	6	69.00	78.00	C1: (32.00,45.00)=49.58; C2: (40.00,57.00)=49.58; C3: (72.00,21.00)=57.08;	2
1	7	53.00	53.00	C1: (32.00,45.00)=22.47; C2: (40.00,57.00)=22.47; C3: (72.00,21.00)=37.22;	2
1	8	76.00	41.00	C1: (32.00,45.00)=44.18; C2: (40.00,57.00)=44.18; C3: (72.00,21.00)=44.18;	3
1	9	48.00	37.00	C1: (32.00,45.00)=17.89; C2: (40.00,57.00)=21.54; C3: (72.00,21.00)=28.84;	1
1	10	63.00	73.00	C1: (32.00,45.00)=41.77; C2: (40.00,57.00)=41.77; C3: (72.00,21.00)=52.77;	2
1	11	48.00	39.00	C1: (32.00,45.00)=17.09; C2: (40.00,57.00)=19.70; C3: (72.00,21.00)=30.00;	1
1	12	73.00	68.00	C1: (32.00,45.00)=47.01; C2: (40.00,57.00)=47.01; C3: (72.00,21.00)=47.01;	2
1	13	40.00	67.00	C1: (32.00,45.00)=23.41; C2: (40.00,57.00)=23.41; C3: (72.00,21.00)=56.04;	2
1	14	37.00	52.00	C1: (32.00,45.00)=8.60; C2: (40.00,57.00)=8.60; C3: (72.00,21.00)=46.75;	2
1	15	73.00	77.00	C1: (32.00,45.00)=52.01; C2: (40.00,57.00)=52.01; C3: (72.00,21.00)=56.01;	2

Nomor 4 (50 PT):

Hitunglah titik baru centroid..... kemudian lakukan proses looping sampai seluruh iterasi berakhir dimana hasil hitungan centroid baru di iterasi ke n harus sama nilainya dengan hasil hitungan centroid di iterasi sebelumnya n-1

Contoh hasil output untuk hitungan rata-rata nilai centroid yang baru:

centroid	x	y
1	49.00	38.67
2	54.30	63.10
3	71.00	37.00

Contoh hasil looping untuk iterasi 2 dan seterusnya sampai selesai. Syarat iterasi berhenti ketika klasifikasi data ke centroid sudah stabil (tidak berubah):

iterasi	datake	x	y	jarak	centroid
2	1	49.00	55.00	C1: (49.00,38.67)=16.33; C2: (54.30,63.10)=16.33; C3: (71.00,37.00)=28.43;	2
2	2	48.00	45.00	C1: (49.00,38.67)=6.41; C2: (54.30,63.10)=19.17; C3: (71.00,37.00)=24.35;	1
2	3	51.00	40.00	C1: (49.00,38.67)=2.40; C2: (54.30,63.10)=23.33; C3: (71.00,37.00)=20.22;	1
.....dst					
2	13	40.00	67.00	C1: (49.00,38.67)=29.73; C2: (54.30,63.10)=29.73; C3: (71.00,37.00)=43.14;	2
2	14	37.00	52.00	C1: (49.00,38.67)=17.94; C2: (54.30,63.10)=20.55; C3: (71.00,37.00)=37.16;	1
2	15	73.00	77.00	C1: (49.00,38.67)=45.22; C2: (54.30,63.10)=45.22; C3: (71.00,37.00)=45.22;	3

..... dst

iterasi	datake	x	y	jarak	centroid
7	1	49.00	55.00	C1: (56.17,39.17)=17.38; C2: (46.33,53.33)=17.38; C3: (59.33,71.00)=19.04;	2
7	2	48.00	45.00	C1: (56.17,39.17)=10.04; C2: (46.33,53.33)=10.04; C3: (59.33,71.00)=28.36;	2
7	3	51.00	40.00	C1: (56.17,39.17)=5.24; C2: (46.33,53.33)=14.12; C3: (59.33,71.00)=32.10;	1
.....dst					
7	13	40.00	67.00	C1: (56.17,39.17)=32.19; C2: (46.33,53.33)=32.19; C3: (59.33,71.00)=32.19;	3
7	14	37.00	52.00	C1: (56.17,39.17)=23.07; C2: (46.33,53.33)=23.07; C3: (59.33,71.00)=29.32;	2
7	15	73.00	77.00	C1: (56.17,39.17)=41.40; C2: (46.33,53.33)=41.40; C3: (59.33,71.00)=41.40;	3

centroid	x	y
1	57.80	38.00
2	46.75	51.25
3	59.33	71.00

SOAL INI BERLAKU DARI JAM 08.00 SD 18.00, TIDAK MUNGKIN ADA JAWABAN YANG SAMA ANTARA SATU DENGAN YANG LAIN KARENA SQL SETIAP ORANG BERBEDA_BEDA LOGIKA DAN ALUR PROGRAMNYA.

TIDAK ADA TOLERANSI COPY PASTE YA.... BUATLAH SENDIRI, JIKA TERDAPAT KECURANGAN, MAKA SELURUH NILAI DI NOL (0) KAN