

# K-Nearest Neighbor dan K-means

Pertemuan 8

# K-Nearest Neighbor - 1

- Konsep dasar dari K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga terdekatnya dalam data pelatihan.
- Penghitungan jarak dilakukan dengan konsep Euclidean.
- Jumlah kelas yang paling banyak dengan jarak terdekat tersebut akan menjadi kelas dimana data evaluasi tersebut berada.

# K-Nearest Neighbor - 2

- Algoritma
  1. Tentukan parameter  $K$  = jumlah tetangga terdekat.
  2. Hitung jarak antara data yang akan dievaluasi dengan semua data pelatihan.
  3. Urutkan jarak yang terbentuk (urut naik) dan tentukan jarak terdekat sampai urutan ke- $K$ .
  4. Pasangkan kelas ( $C$ ) yang bersesuaian.
  5. Cari jumlah kelas terbanyak dari tetangga terdekat tersebut, dan tetapkan kelas tersebut sebagai kelas data yang dievaluasi.



# Contoh

Jika ada 20 data sampel yang digunakan sebagai basis pengetahuan untuk mengetahui status gizi berdasarkan tinggi badan, berat badan, nilai persen lemak, tekanan darah sistolik, tekanan darah diastolic, detak jantung, lingkar lengan atas, lingkar lengan bawah, lingkar perut dan lingkar panggul seperti pada Tabel 1.

# Tabel 1.

<i>Tinggi Badan</i>	<i>Berat Badan</i>	<i>% Lemak</i>	<i>Sistolik</i>	<i>Diastolik</i>	<i>Detak Jantung</i>	<i>Lengan Atas</i>	<i>Lengan Bawah</i>	<i>Lingkar Perut</i>	<i>Lingkar Panggul</i>	<i>Status Gizi</i>
163	59	25.4	102	68	69	23	14	74	94	Normal
170	125	42.9	113	73	79	40	19	112	135	Obesitas
172	75	31	114	78	86	26	16	79	100	Normal
166	58	19	128	69	69	26	16	72	92	Normal
167	50	16.5	111	80	89	21	13	71	88	Kurus
168	50	10.4	100	72	97	23	14	62	87	Kurus
173	56	18.4	114	73	62	24	15	66	93	Normal
168	73	22.7	107	73	81	30	18	77	96	Normal
177	60	17.4	107	71	104	24	15	71	90	Normal
168	52	13.9	122	82	101	22	15	68	84	Kurus
159	58	23.4	118	75	64	26	15	70	85	Normal
167	75	30.7	126	81	65	30	16	91	103	Obesitas
170	72	26.1	113	72	83	28	16	85	98	Normal
172	68	22.8	110	65	68	27	15	79	94	Normal
165	73	29.1	105	67	75	28	18	83	100	Obesitas
169.5	55	15.2	112	77	70	22	14	75	92	Kurus
160	54	15.7	138	104	78	27	15	73	86	Normal
173	56	17.9	120	76	97	25	14	72	88	Normal
162	54	18	108	70	76	24	15	71	88	Normal
169	79	22	123	76	70	29	17	84	101	Obesitas

# Ingin diketahui status gizi seseorang

tinggi badan 175 cm,  
berat badan 67 kg,  
% lemak 23%,  
tekanan darah sistolik 97,  
tekanan darah diastolik 67,  
detak jantung 79/menit,  
lingkar lengan atas 25 cm,  
Lingkar lengan bawah 15 cm,  
lingkar perut 75 cm,  
lingkar panggul 95 cm.

# Algoritma 1

- Tentukan nilai  $K$ , misal  $K = 5$

# Algoritma 2. hitung jarak p ke d

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

dengan:

$x_1$  = sampel data

$x_2$  = data uji

$i$  = variabel data

$d$  = jarak

$p$  = dimensi data

$$\begin{aligned} d_1 &= \sqrt{(163-175)^2 + (59-67)^2 + (25.4-23)^2 + (102-97)^2 + (68-67)^2 + (69-79)^2 + (23-25)^2 + (14-15)^2 +} \\ &\quad \sqrt{(74-75)^2 + (94-95)^2} = 18.6215 \\ d_2 &= \sqrt{(170-175)^2 + (126-67)^2 + (42.9-23)^2 + (118-97)^2 + (78-67)^2 + (79-79)^2 + (40-25)^2 + (19-16)^2 +} \\ &\quad \sqrt{(112-75)^2 + (195-95)^2} = 26.9477 \end{aligned}$$



## Tabel 2.setelah ditambahkan jarak data uji

Tabel 2. Tabel data sebelum ditambahkan jarak terhadap data pengujian

<i>Tinggi Badan</i>	<i>Berat Badan</i>	<i>% Lemak</i>	<i>Sistolik</i>	<i>Diastolik</i>	<i>Detak Jantung</i>	<i>Lengan Atas</i>	<i>Lengan Bawah</i>	<i>Lingkar Perut</i>	<i>Lingkar Panggul</i>	<i>Status Gizi</i>	<i>Jarak</i>
163	59	25.4	102	68	69	23	14	74	94	Normal	18.6215
170	125	42.9	113	73	79	40	19	112	135	Obesitas	85.9477
172	75	31	114	78	86	26	16	79	100	Normal	26.2679
166	58	19	128	69	69	26	16	72	92	Normal	35.5387
167	50	16.5	111	80	89	21	13	71	88	Kurus	30.7449
168	50	10.4	100	72	97	23	14	62	87	Kurus	29.4747
173	56	18.4	114	73	62	24	15	66	93	Normal	25.6351
168	73	22.7	107	73	81	30	18	77	96	Normal	18.976
177	60	17.4	107	71	104	24	15	71	90	Normal	18.503
168	52	13.9	122	82	101	22	15	68	84	Kurus	38.5462
159	58	23.4	118	75	64	26	15	70	85	Normal	32.6827
167	75	30.7	126	81	65	30	16	91	103	Obesitas	40.8692
170	72	26.1	113	72	83	28	16	85	98	Normal	23.6561
172	68	22.8	110	65	68	27	15	79	94	Normal	17.4367
165	73	29.1	105	67	75	28	18	83	100	Obesitas	21.0763
169.5	55	15.2	112	77	70	22	14	75	92	Kurus	26.0594
160	54	15.7	138	104	78	27	15	73	86	Normal	60.7148
173	56	17.9	120	76	97	25	14	72	88	Normal	30.3317
162	54	18	108	70	76	24	15	71	88	Normal	25.671
169	79	22	123	76	70	29	17	84	101	Obesitas	34.2783

# Algoritma 3 dan 4. Urutkan jarak mulai terkecil, ambil C yang bersesuaian

Tabel 1. Data setoran diurutkan berdasarkan jarak terkecil

<i>Tinggi Badan</i>	<i>Berat Badan</i>	<i>% Lemak</i>	<i>Sistolik</i>	<i>Diastolik</i>	<i>Detak Jantung</i>	<i>Lengan Atas</i>	<i>Lengan Bawah</i>	<i>Lingkar Perut</i>	<i>Lingkar Panggul</i>	<i>Status Gizi</i>	<i>Jarak</i>
172	68	22.8	110	65	68	27	15	79	94	Normal	17.4367
177	60	17.4	107	71	104	24	15	71	90	Normal	18.503
163	59	25.4	102	68	69	23	14	74	94	Normal	18.6215
168	73	22.7	107	73	81	30	18	77	96	Normal	18.976
165	73	29.1	105	67	75	28	18	83	100	Obesitas	21.0763
170	72	26.1	113	72	83	28	16	85	98	Normal	23.6561
173	56	18.4	114	73	62	24	15	66	93	Normal	25.6351
162	54	18	108	70	76	24	15	71	88	Normal	25.671
169.5	55	15.2	112	77	70	22	14	75	92	Kurus	26.0594
172	75	31	114	78	86	26	16	79	100	Normal	26.2679
168	50	10.4	100	72	97	23	14	62	87	Kurus	29.4747
173	56	17.9	120	76	97	25	14	72	88	Normal	30.3317
167	50	16.5	111	80	89	21	13	71	88	Kurus	30.7449
159	58	23.4	118	75	64	26	15	70	85	Normal	32.6827
169	79	22	123	76	70	29	17	84	101	Obesitas	34.2783
166	58	19	128	69	69	26	16	72	92	Normal	35.5387
168	52	13.9	122	82	101	22	15	68	84	Kurus	38.5462
167	75	30.7	126	81	65	30	16	91	103	Obesitas	40.8692
160	54	15.7	138	104	78	27	15	73	86	Normal	60.7148
170	125	42.9	113	73	79	40	19	112	135	Obesitas	85.9477

# Algoritma 5.

Cari jumlah kelas terbanyak dari tetangga terdekat tersebut, dan tetapkan kelas tersebut sebagai kelas data yang dievaluasi.

<i>Tinggi Badan</i>	<i>Berat Badan</i>	<i>% Lemak</i>	<i>Sistolik</i>	<i>Diastolik</i>	<i>Detak Jantung</i>	<i>Lengan Atas</i>	<i>Lengan Bawah</i>	<i>Lingkar Perut</i>	<i>Lingkar Panggul</i>	<i>Status Gizi</i>	<i>Jarak</i>
172	68	22.8	110	65	68	27	15	79	94	Normal	17.4367
177	60	17.4	107	71	104	24	15	71	90	Normal	18.503
163	59	25.4	102	68	69	23	14	74	94	Normal	18.6215
168	73	22.7	107	73	81	30	18	77	96	Normal	18.976
165	73	29.1	105	67	75	28	18	83	100	Obesitas	21.0763

Jadi : Data yang diuji termasuk kelas Normal

# K-Means

- Konsep dasar dari **K-Means** adalah pencarian pusat cluster secara iteratif.
- Pusat cluster ditetapkan berdasarkan jarak setiap data ke pusat cluster.
- Proses clustering dimulai dengan mengidentifikasi data yang akan dicluster,  $x_{ij}$  ( $i=1,\dots,n$ ;  $j=1,\dots,m$ ) dengan  $n$  adalah jumlah data yang akan dicluster dan  $m$  adalah jumlah variabel.

# K-Means

- Pada awal iterasi, pusat setiap cluster ditetapkan secara bebas (sembarang),  $c_{kj}$  ( $k=1,...,K$ ;  $j=1,...,m$ ).
- Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster.
- Untuk melakukan penghitungan jarak data ke- $i$  ( $X_i$ ) pada pusat cluster ke- $k$  ( $C_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean, yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2}$$

# K-Means

- Suatu data akan menjadi anggota dari cluster ke-J apabila jarak data tersebut ke pusat cluster ke-J bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lainnya.
- Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap cluster.
- Nilai pusat cluster yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada cluster tersebut, dengan rumus:

$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke } k$$

# K-Means

- Algoritma:
  - Tentukan jumlah cluster ( $K$ ), tetapkan pusat cluster sembarang.
  - Hitung jarak setiap data ke pusat cluster.
  - Kelompokkan data ke dalam cluster yang dengan jarak yang paling pendek.
  - Hitung pusat cluster.
  - Ulangi langkah 2 - 4 hingga sudah tidak ada lagi data yang berpindah ke cluster yang lain.

# Pengelompokan mahasiswa berdasarkan BMI dan Ukuran kerangka

- Persamaan BMI

$$BMI = \frac{\text{berat badan (kg)}}{(\text{tinggi badan (m)})^2} \dots\dots\dots(1)$$

Dengan batas pengelompokan :

<18,5	: berat kurang
18,5 – 22,9	: berat normal
23 – 24,9	: obesitas ringan
25 – 29,9	: obesitas sedang
>= 30	: obesitas berat

- Persamaan ukuran kerangka

$$\text{ukuran rangka} = \frac{\text{tinggi badan(cm)}}{\text{lingkar lengan bawah(cm)}} \dots\dots(2)$$

Dengan batas pengelompokan

Laki-laki :

< 9,6	: kerangka kecil
9,6 – 10,4	: kerangka sedang
>10,4	: kerangka besar

Perempuan :

<10,1	: kerangka kecil
10,1 – 11,0	: kerangka sedang
>11,0	: kerangka besar



# Diketahui data dan hitung BMI & UK

**Tabel 1.** Contoh data mahasiswa

Mhs-ke	Tb	Bb	LLB
1	163	59	14
2	170	125	19
3	164	53	15
4	166	58	16
5	167	50	13
6	168	50	14
7	173	56	15
8	168	73	18
9	177	60	15
10	168	52	15
11	159	58	15
12	167	75	16
13	170	72	16
14	172	68	15
15	165	73	18
16	169,5	55	14
17	160	54	15
18	173	56	14
19	162	54	15
20	169	79	17

Ket :

Tb : Tinggi badan.

Bb : Berat badan.

LLB : Lingkar lengan bawah.

**Tabel 2.** Hasil perhitungan nilai BMI dan ukuran kerangka setiap data

Mhs ke-	BMI	Ukuran Kerangka
1	22,21	11.64
2	43,25	8.95
3	19,71	10.93
4	21,05	10.38
5	17,93	12.85
6	17,72	12.00
7	18,71	11.53
8	25,86	9.33
9	19,15	11.80
10	18,42	11.20
11	22,94	10.60
12	26,89	10.44
13	24,91	10.63
14	22,99	11.47
15	26,81	9.17
16	19,14	12.11
17	21,09	10.67
18	18,71	12.36
19	20,58	10.80
20	27,66	9.94

# 1. Tetapkan jumlah cluster dan pusat cluster sembarang

- Jumlah cluster = 3
- Tentukan pusat cluster secara acak, misalkan kita tentukan  $c1 = (20,9)$ ;  $c2 = (23,10)$ ; dan  $c3 = (27,11)$

## 2. Hitung jarak setiap data yang ada terhadap setiap pusat cluster.

---

menghitung jarak data mahasiswa pertama dengan pusat cluster pertama adalah :

$$d_{11} = \sqrt{(22,21 - 20)^2 + (11,64 - 9)^2} = 4,97$$

Jarak data mahasiswa pertama dengan pusat cluster kedua :

$$d_{12} = \sqrt{(22,21 - 23)^2 + (11,64 - 10)^2} = 2,04$$

Jarak data mahasiswa pertama dengan pusat cluster ketiga :

$$d_{13} = \sqrt{(22,21 - 27)^2 + (11,64 - 11)^2} = 1,91$$

# Masukkan ke tabel perhitungan jarak

**Tabel 3.** Hasil perhitungan jarak setiap data

Mhs ke	BMI	Uk. rangka	C1	C2	C3
1	22,21	11,64	4,97	2,04	1,91
2	43,25	8,95	25,25	22,28	19,36
3	19,71	10,93	2,58	1,60	4,29
4	21,05	10,38	3,34	0,38	3,02
5	17,93	12,85	3,85	4,19	6,35
6	17,72	12,00	3,01	3,85	6,36
7	18,71	11,53	2,63	2,76	5,32
8	25,86	9,33	7,87	4,91	2,50
9	19,15	11,80	3,03	2,58	4,91
10	18,42	11,20	2,24	2,84	5,58
11	22,94	10,60	5,19	2,03	1,13
12	26,89	10,44	9,01	5,91	2,95
13	24,91	10,63	7,10	3,96	0,99

14	22,99	11,47	5,56	2,47	1,12
15	26,81	9,17	8,82	5,87	3,36
16	19,14	12,11	3,31	2,81	4,98
17	21,09	10,67	3,51	0,67	2,93
18	18,71	12,36	3,43	3,29	5,46
19	20,58	10,80	3,14	0,91	3,43
20	27,66	9,94	9,71	6,66	3,81

3. Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat clusternya.

- Misalkan untuk data pertama, jarak terkecil diperoleh pada cluster ketiga, sehingga data pertama akan menjadi anggota dari cluster ketiga. Demikian juga untuk data kedua, jarak terkecil ada pada cluster ketiga, maka data tersebut akan masuk pada cluster ketiga. Posisi cluster selengkapnya dapt dilihat pada Tabel 4

**Tabel 4.** Posisi cluster pada iterasi pertama

Mhs ke	BMI	Uk.rangka	C1	C2	C3
1	22,21	11,64			*
2	43,25	8,95			*
3	19,71	10,93		*	
4	21,05	10,38		*	
5	17,93	12,85	*		
6	17,72	12,00	*		
7	18,71	11,53	*		
8	25,86	9,33			*
9	19,15	11,80		*	
10	18,42	11,20	*		
11	22,94	10,60			*
12	26,89	10,44			*
13	24,91	10,63			*
14	22,99	11,47			*
15	26,81	9,17			*
16	19,14	12,11		*	
17	21,09	10,67		*	
18	18,71	12,36		*	
19	20,58	10,80		*	
20	27,66	9,94			*

## 4. Hitung pusat cluster baru

- Untuk cluster pertama, ada 4 data yaitu data ke-5, 6, 7 dan data ke-10, sehingga:
  - $C11 = (17,93+17,72+18,71+18,42) / 4 = 18,19$
  - $C12 = (12,85+12,00+11,53+11,20) / 4 = 11,89$
- Untuk cluster kedua, ada 7 data yaitu data ke-3, 4, 9, 16, 17, 18 dan data ke-19, sehingga :
  - $C21 = (19,71+21,05+19,15+19,14+21,09 +18,71+20,58) / 7 = 19,92$
  - $C22 = (10,93+10,38+11,8+12,11+10,67+12,36 +10,8) / 7 = 11,29$
- Untuk cluster ketiga, ada 9 data yaitu data ke-1, 2, 8, 11, 12, 13, 14, 15 dan data ke-20, sehingga
  - $C31 = (22,21+43,25+25,86+22,94+26,89+24,91 +22,99+26,81+27,66) / 9 = 27,06$
  - $C32 = (11,64+8,95+9,33+10,6+10,44+10,63+11,47+9,17+9,94) / 9 = 10,24$

## 5. Ulangi langkah 2 hingga posisi data sudah tidak mengalami perubahan.

**Tabel 5.** Hasil cluster pada iterasi ke-2

Mhs-ke	BMI	Uk.rangka	C1	C2	C3
1	22,21	11,64		*	
2	43,25	8,95			*
3	19,71	10,93		*	
4	21,05	10,38		*	
5	17,93	12,85	*		
6	17,72	12,00	*		
7	18,71	11,53	*		
8	25,86	9,33			*
9	19,15	11,80		*	
10	18,42	11,20	*		
11	22,94	10,60		*	
12	26,89	10,44			*
13	24,91	10,63			*
14	22,99	11,47		*	
15	26,81	9,17			*
16	19,14	12,11	*		
17	21,09	10,67		*	
18	18,71	12,36	*		
19	20,58	10,80		*	
20	27,66	9,94			*

**Tabel 6.** Hasil cluster pada iterasi ke-3

Mhs-ke	BMI	Uk.rangka	C1	C2	C3
1	22,21	11,64		*	
2	43,25	8,95			*
3	19,71	10,93		*	
4	21,05	10,38		*	
5	17,93	12,85	*		
6	17,72	12,00	*		
7	18,71	11,53		*	
8	25,86	9,33			*
9	19,15	11,80	*		
10	18,42	11,20	*		
11	22,94	10,60		*	
12	26,89	10,44			*
13	24,91	10,63		*	
14	22,99	11,47		*	
15	26,81	9,17			*
16	19,14	12,11	*		
17	21,09	10,67		*	
18	18,71	12,36	*		
19	20,58	10,80		*	
20	27,66	9,94			*



# Dan seterusnya hingga tidak berubah

**Tabel 7.** Hasil cluster pada iterasi ke-8

Mhs-ke	BMI	Uk.rangka	C1	C2	C3
1	22,21	11,64	*		
2	43,25	8,95			*
3	19,71	10,93	*		
4	21,05	10,38	*		
5	17,93	12,85	*		
6	17,72	12,00	*		
7	18,71	11,53	*		
8	25,86	9,33		*	
9	19,15	11,80	*		
10	18,42	11,20	*		
11	22,94	10,60		*	
12	26,89	10,44		*	
13	24,91	10,63		*	
14	22,99	11,47		*	
15	26,81	9,17		*	
16	19,14	12,11	*		
17	21,09	10,67	*		
18	18,71	12,36	*		
19	20,58	10,80	*		
20	27,66	9,94		*	

**Tabel 8.** Hasil cluster pada iterasi ke-9

Mhs-ke	BMI	Uk.rangka	C1	C2	C3
1	22,21	11,64	*		
2	43,25	8,95			*
3	19,71	10,93	*		
4	21,05	10,38	*		
5	17,93	12,85	*		
6	17,72	12,00	*		
7	18,71	11,53	*		
8	25,86	9,33		*	
9	19,15	11,80	*		
10	18,42	11,20	*		
11	22,94	10,60		*	
12	26,89	10,44		*	
13	24,91	10,63		*	
14	22,99	11,47		*	
15	26,81	9,17		*	
16	19,14	12,11	*		
17	21,09	10,67	*		
18	18,71	12,36	*		
19	20,58	10,80	*		
20	27,66	9,94		*	

- Karena pada iterasi ke-8 dan ke-9 (Tabel 7 & 8) posisi cluster tidak berubah, maka iterasi dihentikan dan hasil akhir yang diperoleh adalah 3 cluster :
- Cluster pertama memiliki pusat (19,53; 11,52) yang dapat diartikan sebagai kelompok mahasiswa dengan BMI normal dan kerangka besar.
- Cluster kedua memiliki pusat (25,44; 10,22) yang dapat diartikan sebagai kelompok mahasiswa dengan BMI obesitas sedang dan kerangka sedang.
- Cluster ketiga memiliki pusat (43,25; 8,95) yang dapat diartikan sebagai kelompok mahasiswa dengan BMI obesitas berat dan kerangka kecil.