

A background network diagram with various sized nodes (circles) in blue, dark blue, and grey, connected by thin grey lines. Some nodes are highlighted with larger concentric circles.

# K-NEAREST NEIGHBORS (K-NN)

---

Matematika III, Teknik Informatika, Politeknik Negeri Malang  
Ganjil 2022

# APA ITU K-NN?

“Salah satu algoritma untuk klasifikasi data suatu obyek berdasarkan sejumlah K-data training yang memiliki jarak paling dekat (*nearest neighbor*) dengan objek tersebut”

(Wahyono, 2018, “*Python for Machine Learning*”)

# APA ITU K-NN?

Jauh dekatnya tetangga biasanya dihitung berdasarkan jarak *Euclidean*(d) sebagai berikut :

$$d_{ij} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

# APA ITU K-NN?

Proses atau tahapan dari perhitungan algoritma k-Nearest Neighbor adalah melakukan proses normalisasi dan menentukan Euclidean Distance dan memasukkan nilai k guna memperoleh hasil terdekat dari nilai itu sendiri.

# ALGORITMA PERHITUNGAN K-NN

1. Menentukan parameter K sebagai banyaknya jumlah tetangga terdekat dengan objek baru.
2. Menghitung jarak antar objek/data baru terhadap semua objek/data yang telah di training.
3. Urutkan hasil perhitungan tersebut.
4. Tentukan tetangga terdekat berdasarkan jarak minimum ke K.
5. Tentukan kategori dari tetangga terdekat dengan objek/data.
6. Gunakan kategori mayoritas sebagai klasifikasi objek/data baru.

# CONTOH PERHITUNGAN

1. Diberikan data training berjumlah dua kelas Bad dan Good :

X	Y	Kategori
7	6	Bad
6	6	Bad
6	5	Bad
1	3	Good
2	4	Good
2	2	Good
3	5	?

2. Tentukan nilai, misalnya kita buat jumlah tetangga terdekat  $K=3$

X	Y	Euclidean Distance (3,5)
7	6	$\sqrt{(7-3)^2 + (6-5)^2} = \sqrt{(4)^2 + (1)^2} = \sqrt{17} = 4.12$
6	6	$\sqrt{(6-3)^2 + (6-5)^2} = \sqrt{(3)^2 + (1)^2} = \sqrt{10} = 3.16$
6	5	$\sqrt{(6-3)^2 + (5-5)^2} = \sqrt{(3)^2 + (0)^2} = \sqrt{9} = 3$
1	3	$\sqrt{(1-3)^2 + (3-5)^2} = \sqrt{(-2)^2 + (-2)^2} = \sqrt{8} = 2.82$
2	4	$\sqrt{(2-3)^2 + (4-5)^2} = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2} = 1.41$
2	2	$\sqrt{(2-3)^2 + (2-5)^2} = \sqrt{(-1)^2 + (-3)^2} = \sqrt{10} = 3.16$

3. Urutkan jarak dari data baru dengan data training dan menentukan tetangga terdekat berdasarkan jarak minimum K

X	Y	Euclidean Distance (3,5)	Urutan Jarak	Apakah termasuk 3-NN ?
7	6	4.12	5	Tidak ( $K > 3$ )
6	6	3.16	4	Tidak ( $K > 3$ )
6	5	3	3	Ya ( $K = 3$ )
1	3	2.82	2	Ya ( $K < 3$ )
2	4	1.41	1	Ya ( $K < 3$ )
2	2	3.16	4	Tidak ( $K > 3$ )

4. Tentukan kategori dari tetangga terdekat. Kita perhatikan baris 3, 4, dan 5 pada gambar sebelumnya (diatas). Kategori Ya diambil jika nilai  $K \leq 3$ . Jadi baris 3, 4, dan 5 termasuk kategori Ya dan sisanya Tidak.

X	Y	Euclidean Distance (3,5)	Urutan Jarak	Apakah termasuk 3-NN ?	Kategori Ya untuk KNN
7	6	4.12	5	Tidak ( $K > 3$ )	-
6	6	3.16	4	Tidak ( $K > 3$ )	-
6	5	3	3	Ya ( $K = 3$ )	Bad
1	3	2.82	2	Ya ( $K < 3$ )	Good
2	4	1.41	1	Ya ( $K < 3$ )	Good
2	2	3.16	4	Tidak ( $K > 3$ )	-

5. Gunakan kategori mayoritas yang sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi data yang baru.

X	Y	Kategori
7	6	Bad
6	6	Bad
6	5	Bad
1	3	Good
2	4	Good
2	2	Good
3	5	Good

Data yang kita miliki pada baris 3, 4 dan 5 kita punya 2 kategori Good dan 1 kategori Bad. Dari jumlah mayoritas (Good > Bad) tersebut kita simpulkan bahwa data baru (X=3 dan Y=5) termasuk dalam kategori Good.



# CONTOH KASUS DALAM PENELITIA N

## PENERAPAN ALGORITMA KLASIFIKASI NEAREST NEIGHBOR (K-NN) UNTUK MENDETEKSI PENYAKIT JANTUNG

MEI LESTARI

mei\_6s@yahoo.co.id

Program Studi Teknik Informatika, Fakultas Teknik, Matematika dan IPA

Universitas Indraprasta

Jln. Jl. Nangka No. 58 Tanjung Barat Jagakarsa, Jakarta Selatan

**Abstrak.** Data WHO menyatakan bahwa sebanyak 7,3 juta penduduk dunia meninggal dikarenakan penyakit jantung. Meskipun penyakit jantung merupakan penyakit yang tidak menular, penyakit ini merupakan jenis penyakit yang mematikan nomor satu di dunia. Penyakit jantung disebut juga dengan penyakit jantung koroner, penyakit ini terjadi bila darah ke otot jantung terhenti/tersumbat, sehingga mengakibatkan kerusakan berat pada jantung (Rajkumar & Reena, September 2010). Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) diharapkan dapat mengatasi masalah efektifitas dan akurasi dalam mendeteksi penyakit jantung. Pada penelitian ini digunakan algoritma K-NN dengan  $k = 9$  pada 100 data pasien penyakit jantung. Hasil penelitian diperoleh nilai akurasi sebesar 70% serta nilai AUC sebesar 0.875 yang masuk kedalam klasifikasi baik, sehingga algoritma K-NN dapat digunakan dan diterapkan untuk mendeteksi penyakit jantung.

URL :

[https://journal.lppmunindra.ac.id/index.php/Faktor\\_Exacta/article/download/290/276](https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/download/290/276)