

PENGANTAR PEMBELAJARAN MESIN & CONTOH KLASIFIKASI MENGGUNAKAN K NEAREST NEIGHBOR (KNN)

Margaretha Sulistyoningsih, Ph.D

APA ITU PEMBELAJARAN MESIN

- ❖ Pembelajaran Mesin : Mengekstraksi pengetahuan dari data [1].
- ❖ Merupakan pertemuan dari statistik, kecerdasan buatan, dan ilmu komputer.
- ❖ Penerapan pembelajaran mesin dalam kehidupan sehari-hari: Rekomendasi otomatis untuk film, Rekomendasi buku, atau barang yang dijual, seperti pada Amazon, dll.

PEMBELAJARAN MESIN[1]

- ❖ Dalam bidang kesehatan, untuk deteksi dan treatment penyakit (misal kanker).
- ❖ Dalam bidang biologi: Analisa DNA.
- ❖ Dalam bidang Astronomi: Menemukan planet-planet yang jauh.
- ❖ Dalam mata kuliah ini, akan digunakan dalam pengenalan sesuatu/seseorang.

MENGAPA MESIN LEARNING? [1]

- Anda sudah belajar dan praktikum Prolog. Logika diperlukan untuk menulis program Prolog dan membuat Sistem Berbasis Pengetahuan dan Sistem Pakar. Jika ada logika dalam aturan yang berubah, maka perlu perubahan besar-besaran dalam program.
- Merancang aturan dibutuhkan pengetahuan yang dalam, bahkan oleh seorang pakar.
- Bagaimana jika mesin bisa belajar sendiri dari data?

PEMBELAJARAN MESIN

➤ Jenis-jenis pembelajaran mesin:

- Supervised Learning.
- Unsupervised Learning

Apakah itu?

~ Nilai Keaktifan.

UNSUPERVISED LEARNING

- ❖ Anda tidak melakukan pelabelan, tapi data tersebut belajar sendiri untuk mengelompokkan dirinya sendiri berdasarkan kriteria-kriteria tertentu.
- ❖ Contoh: Pengklusteran (*Clustering*).

SUPERVISED LEARNING MENGGUNAKAN KNN (K NEAREST NEIGHBOR)



Data Latih



KNN (K NEAREST NEIGHBOR)



Data Testing:
Ikan apakah ini?

Masuk Klasifikasi apakah, ikan ini?
Apakah ikan nila? Atau ikan lele, atau ikan pindang?

DATA LATIH

Tabel 1. Data Latih Klasifikasi Jenis Ikan

No	Panjang (cm)	Tinggi (cm)	Berat	Sisik	Kumis	Warna	Label
1.	25	4	0.15	Tidak	Ya	Hitam	lele
2.	20	3	0.13	Ya	Tidak	Merah	nila
3.	27	4	0.145	Tidak	Ya	Hitam	lele
4.	20.5	7.5	0.155	Ya	Tidak	Merah	nila
5.	26	4	0.15	Tidak	Ya	Hitam	lele
6.	18	3	0.06	Tidak	Tidak	Abu-abu	pindang
7.	23	7	0.16	Ya	Tidak	Merah	nila
8.	18	3	0.06	Tidak	Tidak	Abu-abu	pindang
9.	27	4	0.12	Tidak	Ya	Hitam	lele
10.	16	2.5	0.06	Tidak	Tidak	Abu-abu	pindang

DATA TESTING

Tabel 2. Data TESTING. Jenis ikan apakah, ikan dengan ciri-ciri di bawah ini?

No	Panjang	Tinggi	Berat	Sisik	Kumis	Warna	Klasifikasi
1.	22.5	7	0.2	Ya	Tidak	Hitam	Ikan apakah, ini?

**PREPROCESSING DATA: JIKA ADA
DATA KATEGORIKAL, DIUBAH
DULU MENJADI DATA NUMERIKAL.**

LANGKAH-LANGKAH

Jawab:

Langkah 1: Ubah ciri yang berupa categorical data menjadi numerical data karena KNN bekerja dengan cara mengurangkan jarak antara data-data numerik.

Ya = 1

Tidak = 0

Hitam = 1

Bukan Hitam (Merah dan abu-abu) = 0

LANGKAH-LANGKAH KNN

Tabel 3. Data Training yang sudah diganti data kategorikal nya menjadi data numerical

No	Panjang	Tinggi	Berat	Sisik	Kumis	Warna	Label
1.	25	4	0.15	0	1	1	lele
2.	20	3	0.13	1	0	0	nila
3.	27	4	0.145	0	1	1	lele
4.	20.5	7.5	0.155	1	0	0	nila
5.	26	4	0.15	0	1	1	lele
6.	18	3	0.06	0	0	0	pindang
7.	23	7	0.16	1	0	0	nila
8.	18	3	0.06	0	0	0	pindang
9.	27	4	0.12	0	1	1	lele
10.	16	2.5	0.06	0	0	0	pindang

LANGKAH-LANGKAH KNN

Tabel 4. Data Testing yang sudah diganti data kategorikal nya
menjadi data numerical

No	Panjang	Tinggi	Berat	Sisik	Kumis	Warna	Klasifikasi
1.	22.5	7	0.2	1	0	1	Ikan apakah, ini?

LANGKAH KE-1: MENENTUKAN JARAK DARI TIAP CIRI PADA DATA TESTING KE TIAP CIRI PADA DATA LATIH.

Anda dapat menggunakan formula perhitungan jarak seperti City Block atau Euclidean Distance. Penjelasan apa itu City Block ada pada slide selanjutnya.

LANGKAH-LANGKAH KNN

Langkah 2:

Menggunakan city block atau jarak euclidean, jarak dari setiap ciri pada data testing ke data training dihitung. Disini digunakan city block.

Jarak

total= $|panjang\ pada\ data\ testing - panjang\ data\ training| +$
 $|tinggi\ pada\ data\ testing - tinggi\ data\ training| +$
 $|berat\ pada\ data\ testing - berat\ data\ training| +$
 $|sisik\ pada\ data\ testing - sisik\ data\ training| +$
 $|kumis\ pada\ data\ testing - kumis\ data\ training|$

CONTOH PERHITUNGAN JARAK CIRI-CIRI PADA DATA TESTING DENGAN DATA LATIH NO. 1

Jarak total =

$$|25-22.5| + |4-7| + |0.15-0.2| + |0-1| + |1-0| + \\ |1-1|$$

$$= 2.5 + 3 + 0.05 + 1 + 1 + 0$$

$$= 7.55$$

Jadi jarak total ciri-ciri pada data latih-1 dengan ciri-ciri pada data testing adalah 7.55.

Ulangi cara di atas untuk menghitung jarak ciri-ciri pada data testing dengan ciri-ciri pada data latih ke-2 sampai data latih ke-10.

Tabel 5. Tabel jarak antara tiap ciri data testing dengan data latih, beserta Jarak Totalnya.

No	Jarak ciri Panjang	Jarak ciri Tinggi	Jarak ciri Besar	Jarak ciri Sisik	Jarak ciri Kumis	Jarak ciri Warna	Jarak Total	Klasifikasi Jenis ikan
1.	2.5	3	0.05	1	1	0	7.55	lele
2.	2.5	4	0.07	0	0	1	7.57	nila
3.	4.5	3	0.055	1	1	0	9.555	lele
4.	2	0.5	0.045	0	0	1	3.545	nila
5.	3.5	3	0.05	1	1	0	8.55	lele
6.	4.5	4	0.14	1	0	1	10.64	pindang
7.	0.5	0	0.04	0	0	1	1.54	nila
8.	4.5	4	0.14	1	0	1	10.64	pindang
9.	4.5	3	0.08	1	1	0	9.58	lele
10.	6.5	4.5	0.14	1	0	1	13.14	pindang

LANGKAH KE-2: MENGURUTKAN (SORTING)

Urutkan (*sort*) hasil perhitungan jarak antara data testing dengan data latih (*training data*).

Pengurutan dimulai dengan data latih yang total jaraknya paling kecil.
Artinya dimulai dari data latih yang paling mirip dengan data testing.

Hasil Pengurutan

Tabel 6. Hasil pengurutan

No	Jarak ciri Panjang	Jarak ciri Tinggi	Jarak ciri Berat	Jarak ciri Sisik	Jarak ciri Kumis	Jarak ciri Warna	Jarak Total	Klasifikasi Jenis ikan
7.	0.5	0	0.04	0	0	1	1.54	nila
4.	2	0.5	0.045	0	0	1	3.545	nila
1.	2.5	3	0.05	1	1	0	7.55	lele
2.	2.5	4	0.07	0	0	1	7.57	nila
5.	3.5	3	0.05	1	1	0	8.55	lele
3.	4.5	3	0.055	1	1	0	9.555	lele
9.	4.5	3	0.08	1	1	0	9.58	lele
6.	4.5	4	0.14	1	0	1	10.64	pindang
8.	4.5	4	0.14	1	0	1	10.64	pindang
10.	6.5	4.5	0.14	1	0	1	13.14	pindang

HASIL PENGURUTAN

- + Dari hasil pengurutan tersebut, data nomor 7 adalah data yang jaraknya paling dekat dengan data testing. Artinya ikan pada data testing mirip sekali dengan ikan pada nomor 7 yaitu ikan nila. Akan tetapi, untuk lebih jelas apa klasifikasi ikan tersebut, mari kita lihat slide selanjutnya.

LANGKAH KE-3: KLASIFIKASI

Penentuan klasifikasi data testing: Apakah ikan pada data testing adalah ikan nila, ikan lele, atau ikan pindang.

Untuk 1 NN, klasifikasinya adalah:

Ambil jarak terkecil, yaitu data ke 7 dengan jarak: 1.54

Klasifikasi ikannya adalah ikan NILA.

Ya, untuk 1NN, sudah kita dapatkan jawabannya!!! Ikan pada data testing adalah ikan NILA.

Jika yang diminta 3NN, maka ambil 3 data latih dengan jarak terkecil dari tabel 6. Kita dapatkan rangking-1 adalah NILA (data ke-7), rangking-2 adalah NILA (Data ke-4), dan rangking ke-3 adalah Lele (Data ke-1).

Maka kita lakukan voting. Karena terdapat 2 ikan nila, dan 1 ikan lele pada pengambilan 3NN, maka yang menang adalah ikan NILA.

Ya, dengan 3NN juga menghasilkan klasifikasi ikan pada data testing adalah ikan NILA.

Perhitungan yang sama dapat dilakukan untuk 5NN.

Tabel 7. Hasil sorting 5 data

Berdasarkan Jarak Total

Data latih dan data testing

Jarak Total	Jenis ikan
1.54	Nila (Data ke-7)
3.545	Nila (Data ke-4)
7.55	Ilele (Data ke-1)
7.57	nila (Data ke-2)
8.55	Ilele (Data ke-5)

KLASIFIKASI

- + Klasifikasi berdasarkan 5NN: Karena dari hasil voting: Ikan Nila jumlahnya ada 3, dan ikan Lele ada 2, sehingga klasifikasi ikan pada data testing adalah ikan Nila.
- + Jika ada kasus di mana Urutan ikan dengan jarak terdekat adalah: Nila, Nila, Lele, Lele, Pindang, dimana Jumlah Nila ada 2, Lele ada 2, dan kemudian 1 ikan pindang, maka klasifikasi ikan pada data latih dapat ditentukan dengan menghitung nilai rata-rata jarak data testing dengan dari Nila, dan Lele yang terseleksi sebagai tetangga terdekat. Rata-rata jarak terkecil menentukan klasifikasi ikan pada data testing.

**SELAMAT! ANDA SUDAH BERHASIL
MENENTUKAN KLASIFIKASI JENIS
IKAN PADA DATA TESTING!**

Tutup PPT ini, dan latihan lah tanpa melihat langkah-langkahnya. Latihan dengan data latih pada slide 11, dan data testing pada slide 12.

ACKNOWLEDGMENT

- + Terimakasih kepada Dr. A. Rita Widiarti atas ide untuk menggunakan ciri ikan lele, nila dan pindang sebagai bahan Pengenalan Pola dan Pembelajaran Mesin.
- + Data yang digunakan dalam modul ini adalah data dari kelas penulis saat berkolaborasi dengan Dr. Rita, dalam mengajar mata kuliah Pengenalan Pola.

REFERENSI

1. Andreas C. Mueller & Sarah Guido,
“Introduction to Machine Learning with
Python, A Guide for Data Scientists”, O’Reilly,
2017.