

Klasifikasi

PENGANTAR KECERDASAN BUATAN



Copyright © Universitas Indonesia

Dr. Prima Dewi Purnamasari
Program Studi Teknik Komputer FTUI




Halo!
Selamat datang
kembali di kuliah
pengantar kecerdasan
buatan!

UNIVERSITAS
INDONESIA

Universitas Indonesia

Dr. Prima Dewi Purnamasari



Sebelumnya:
Regresi &
gradient descent



Dr. Prima Dewi Purnamasari



Sekarang:
Klasifikasi



UNIVERSITAS
INDONESIA

© Universitas Indonesia

Dr. Prima Dewi Purnamasari

**Capaian
pembelajaran**

**Mampu menjelaskan metode
klasifikasi dalam pembelajaran mesin**

UNIVERSITAS
INDONESIA

Copyright © Universitas Indonesia

Klasifikasi

Mempelajari hubungan antara sekumpulan variabel independent (**disebut fitur**) dan variabel dependen (**disebut kelas/target**)

Atribut target dalam klasifikasi adalah variabel **kategorikal** atau nilai diskrit.

Bagaimana cara klasifikasi bekerja ?

Copyright © Universitas Indonesia

Bunga Iris

R.A. Fisher's tahun 1936



Iris Versicolor



Iris Setosa



Iris Virginica

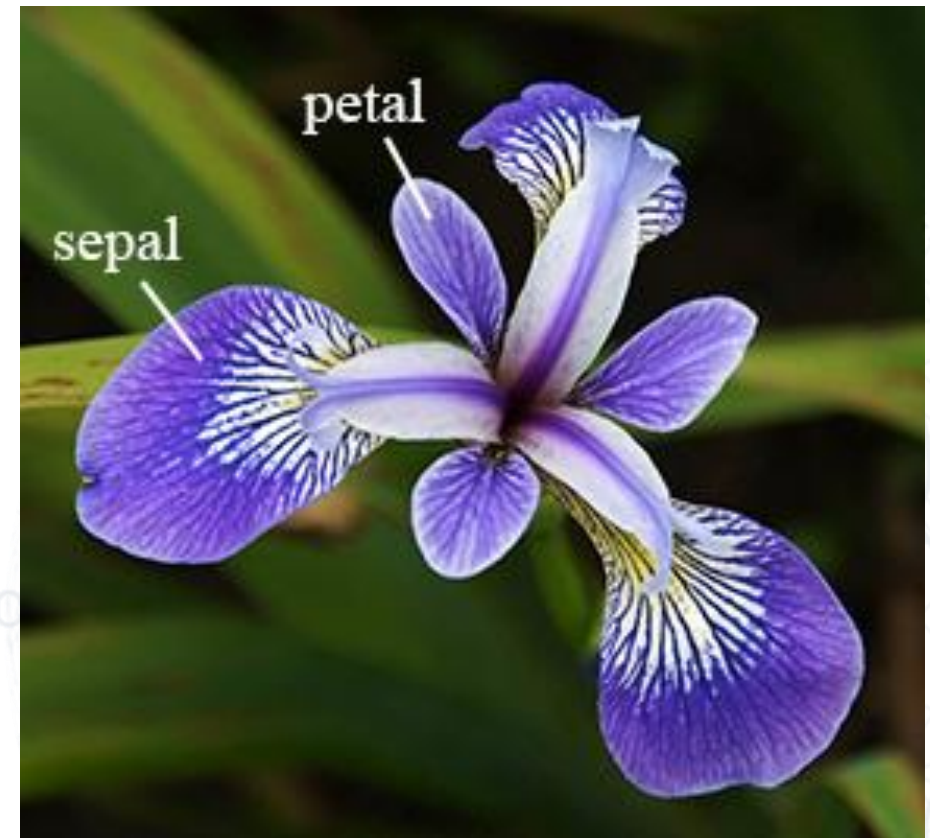
sekilas mirip baik bentuk maupun warnanya

Contoh data dari
Fisher's Iris

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
51	7	3.2	4.7	1.4	Iris-versicolor
52	6.4	3.2	4.5	1.5	Iris-versicolor
53	6.9	3.1	4.9	1.5	Iris-versicolor
54	5.5	2.3	4	1.3	Iris-versicolor
55	6.5	2.8	4.6	1.5	Iris-versicolor
101	6.3	3.3	6	2.5	Iris-virginica
102	5.8	2.7	5.1	1.9	Iris-virginica
103	7.1	3	5.9	2.1	Iris-virginica
104	6.3	2.9	5.6	1.8	Iris-virginica
105	6.5	3	5.8	2.2	Iris-virginica

Bunga IRIS dapat dibedakan berdasarkan ciri:

- Sepal length
- Sepal width
- Petal length
- Petal width



Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
51	7	3.2	4.7	1.4	Iris-versicolor
52	6.4	3.2	4.5	1.5	Iris-versicolor
53	6.9	3.1	4.9	1.5	Iris-versicolor
54	5.5	2.3	4	1.3	Iris-versicolor
55	6.5	2.8	4.6	1.5	Iris-versicolor
101	6.3	3.3	6	2.5	Iris-virginica
102	5.8	2.7	5.1	1.9	Iris-virginica
103	7.1	3	5.9	2.1	Iris-virginica
104	6.3	2.9	5.6	1.8	Iris-virginica
105	6.5	3	5.8	2.2	Iris-virginica

Bunga IRIS dapat dibedakan berdasarkan ciri:

- Sepal length
- Sepal width
- Petal length
- Petal width

FITUR (*feature*)

4 fitur

KELAS

3 kelas

Tujuan Algoritma Klasifikasi

Jika diberikan informasi

- Sepal length
- Sepal width
- Petal length
- Petal width



Dapat diketahui jenis Iris:

- Setosa
- Versicolor
- Virginica

Tujuan Algoritma Klasifikasi

Jika diberikan
informasi
FITUR



Dapat diketahui
KELAS

Copyright © Universitas Indonesia

Algoritma Klasifikasi Populer

Decision Trees

Naïve Bayes

K-nearest
neighbor

Logistic
regression

Support Vector
Machines

Neural
Networks

Copyright

K-Nearest Neighbor (KNN)

Copyright © Universitas Indonesia



KNN

Metode untuk mengklasifikasi sebuah sampel data (*data testing*) berdasarkan kemiripannya dengan sampel data lain yang sudah memiliki label (*data training*)

Sampel yang mirip disebut “*neighbor*”

Data Fisher's Iris



UNIVERSITAS
INDONESIA

4 fitur

3 kelas

masing-masing kelas: 50 sampel

total 150 sampel

Copyright © Universitas Indonesia



MEMODELKAN MESIN

Membagi data yang
dimiliki menjadi:

data training
dan
data testing

40 sampel → data training

10 sampel → data testing

40 sampel → data training

10 sampel → data testing

40 sampel → data training

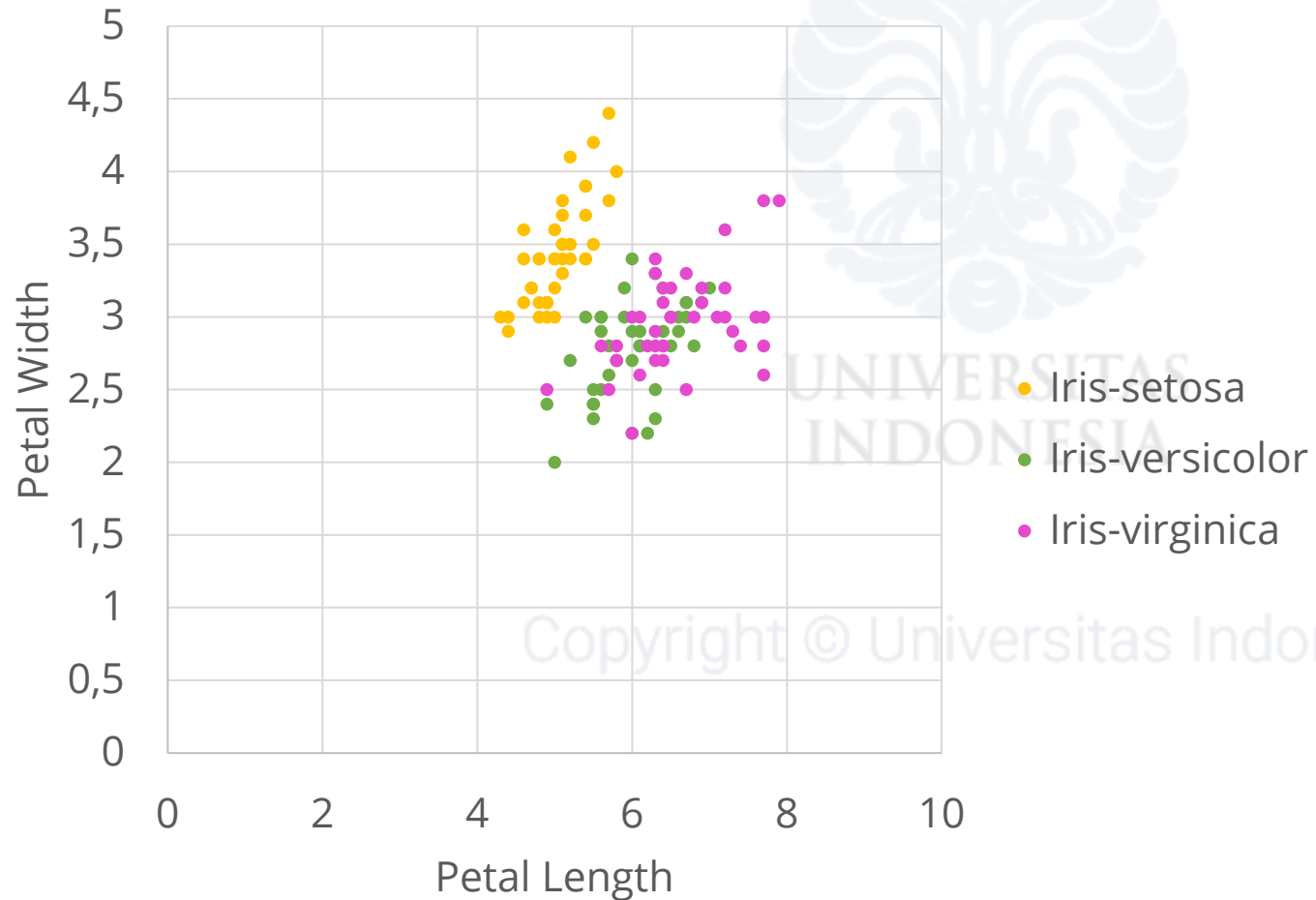
10 sampel → data testing

SETOSA

VERSICOLOR

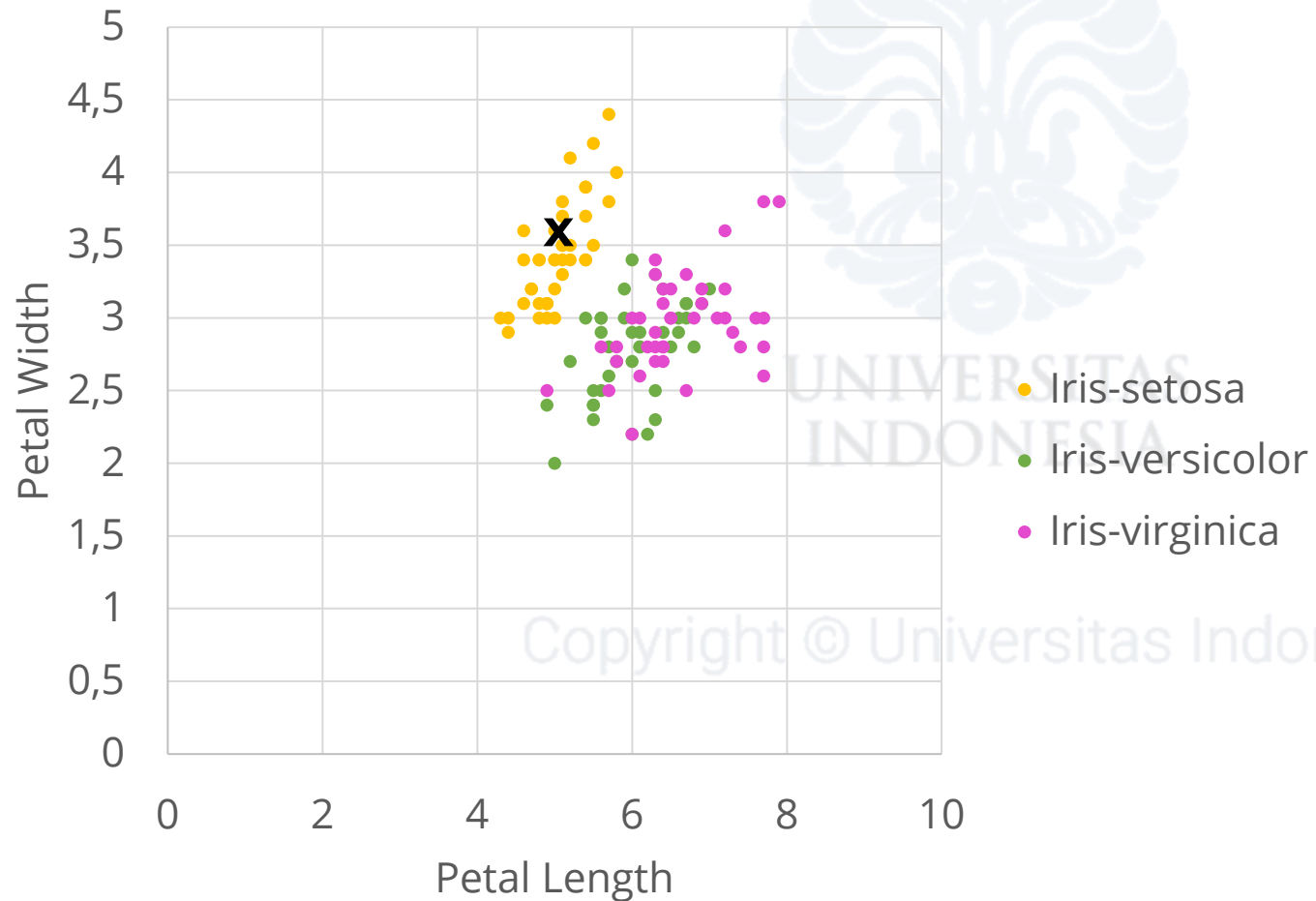
VIRGINICA

Untuk memudahkan visualisasi, ambil 2 fitur pertama: sepal length & sepal width



Copyright © Universitas Indonesia

Untuk memudahkan visualisasi, ambil 2 fitur pertama: sepal length & sepal width



Titik x akan diklasifikasikan sebagai Setosa karena jika dihitung jaraknya paling dekat ke *neighbor* sampel data training Setosa

Menghitung jarak 2 titik

Euclidean distance

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Manhattan distance

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Copyright © Universitas Indonesia

dengan:

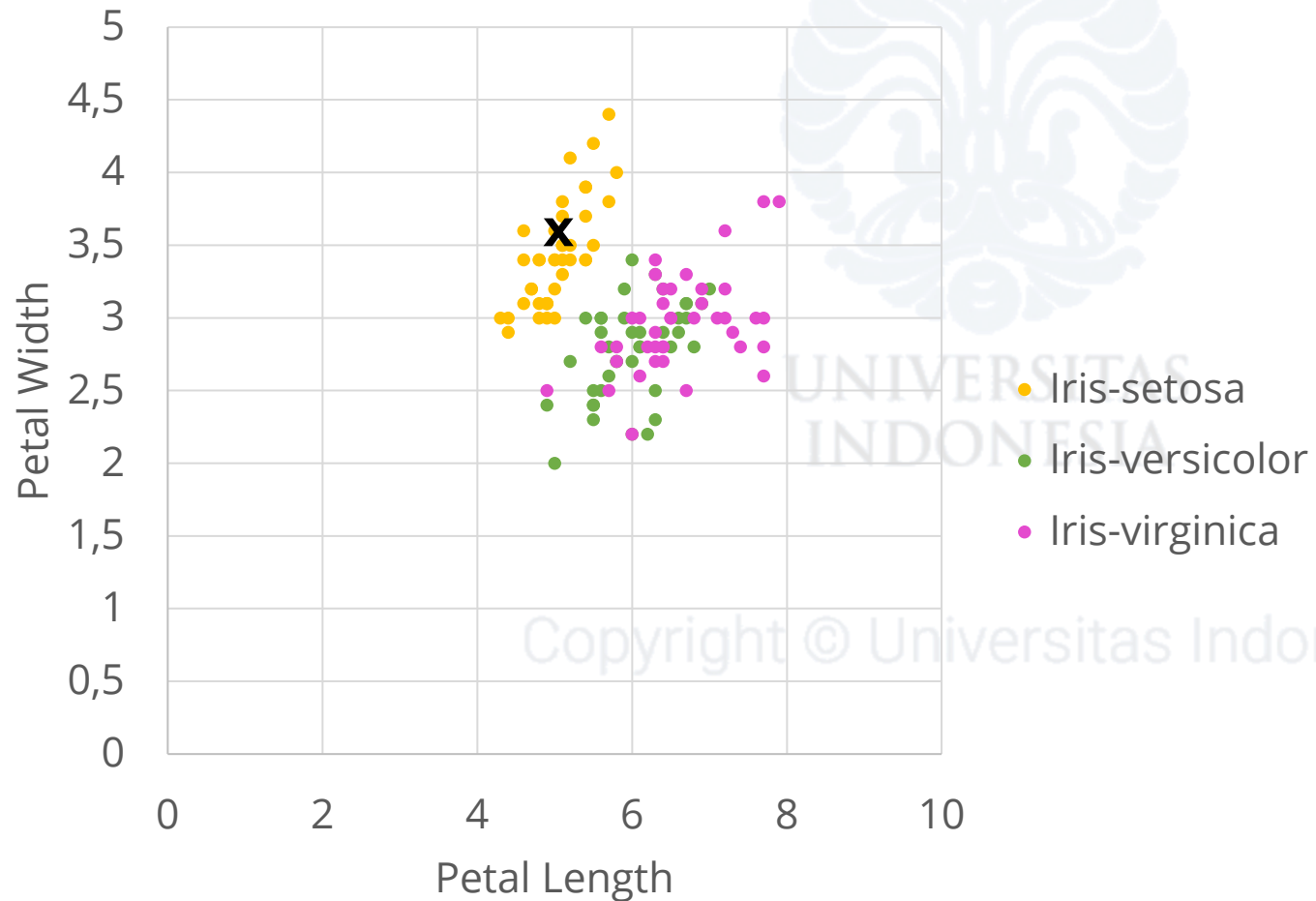
x = titik pertama

y = titik kedua

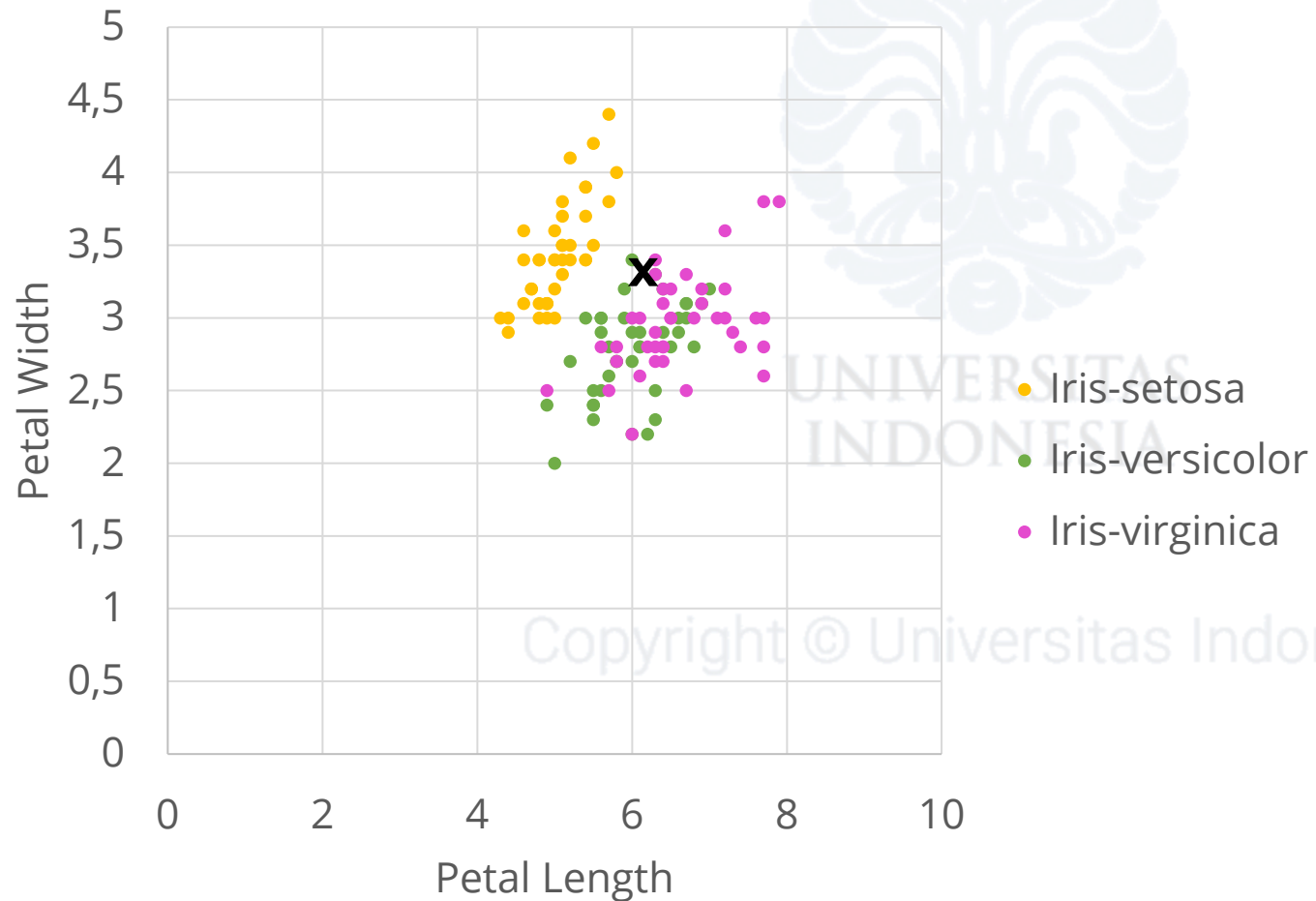
Algoritma KNN

1. Pilih jumlah *neighbor* yang diinginkan $\rightarrow K$
2. Ambil 1 titik baru yang ingin diketahui kelasnya
3. Hitung jarak dari semua data training ke titik baru tersebut
4. Urutkan titik data training dari jarak terdekat ke terjauh terhadap titik baru
5. Ambil sejumlah k titik di urutan teratas
6. Mayoritas kelas dari titik terdekat akan menjadi kelas untuk titik baru ini

Lihat contoh sebelumnya



Lihat contoh sebelumnya



Misal $k = 4$

Titik x

- 3 neighbor pink
- 1 neighbor hijau

kelas = pink (Virginica)

Copyright © Universitas Indonesia

Bagaimana cara memilih nilai K yang tepat?

- Menghitung akurasi untuk setiap K yang dipilih
- Akurasi (paling sederhana) → *Recognition rate*

$$\text{Recognition rate (\%)} = \frac{\sum \text{data yang diklasifikasi dengan benar}}{\sum \text{data}}$$


Copyright © Universitas Indonesia

Mulai dari $k = 1$, lalu tingkatkan k , dan lihat k mana yang terbaik untuk model yang dibuat

Bagaimana cara
kerja algoritma
klasifikasi yang
lain?

Copyright © Universitas Indonesia





Selanjutnya:
Pengelompokan

UNIVERSITAS
INDONESIA

Universitas Indonesia