



# Pengenalan Pola

Klasifikasi – Naïve Bayes

PTIIK - 2014



# Course Contents

1

Naïve Bayes Classifier

2

Fase Training

3

Fase Testing

4

Studi Kasus dan Latihan

## ➤ Naïve Bayes Classifier

- Metode klasifikasi ini diturunkan dari penerapan **teorema Bayes** dengan asumsi **independence** (saling bebas)
- **Naïve Bayes Classifier** adalah metode pengklasifikasian dengan model peluang, dimana diasumsikan bahwa setiap atribut contoh (data sampel) bersifat **saling lepas** satu sama lain berdasarkan **atribut kelas**.

# ➤ Naïve Bayes Classifier

- Apa yang dapat kita lakukan apabila data kita yaitu  $d$  mempunyai beberapa atribut atau fitur  $a_1, a_2, \dots, a_t$  ?
- Asumsi Naïve Bayes : Atribut yang mendeskripsikan contoh data adalah independen pada hipotesis klasifikasi

$$P(\mathbf{d} | h) = P(a_1, \dots, a_T | h) = \prod_t P(a_t | h)$$

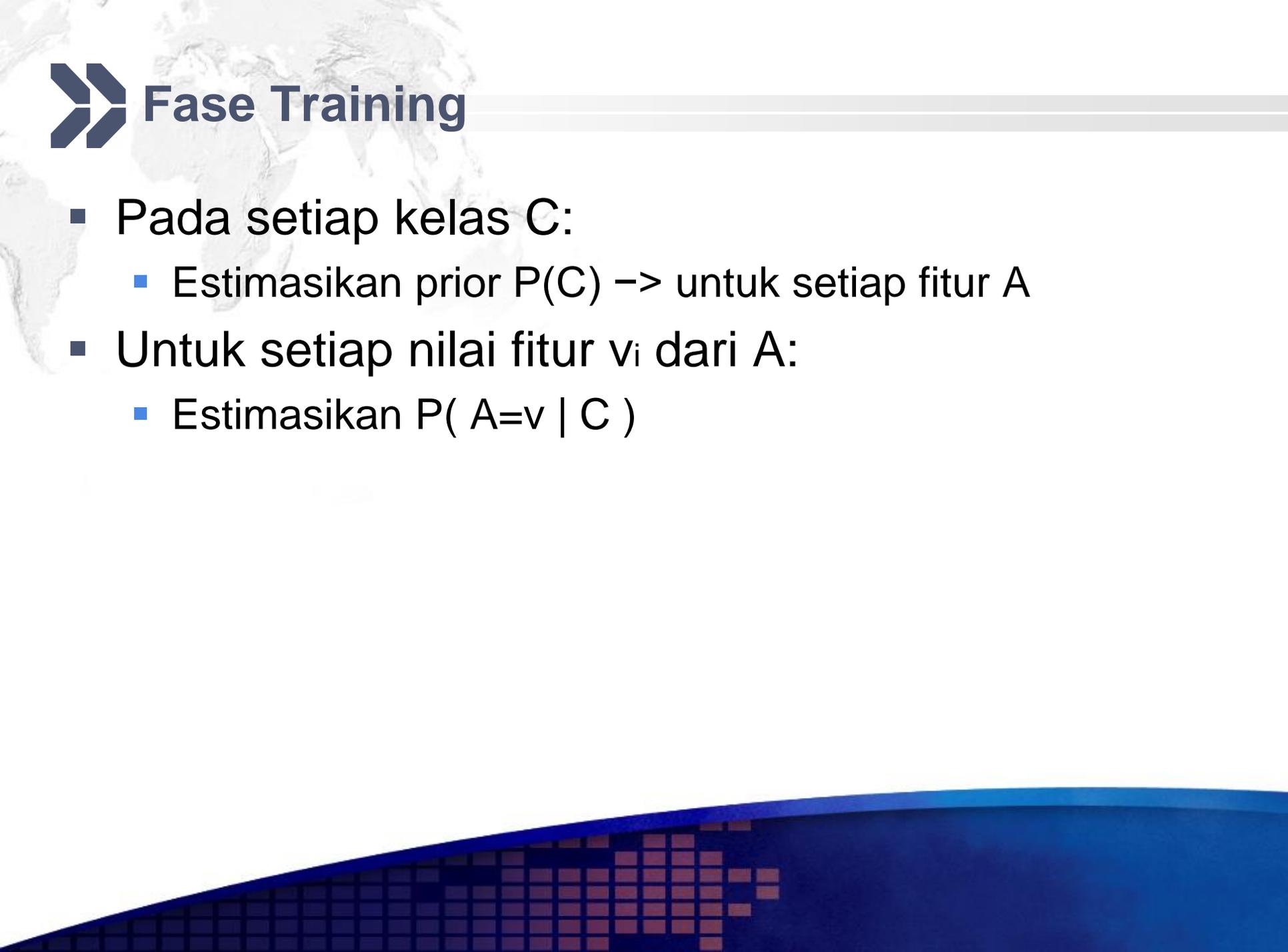
- Asumsi penyederhanaan, kemungkinan tidak sesuai realita
- Namun bekerja dengan baik secara praktis
- Beberapa aplikasi:
  - Diagnosa medis
  - Klasifikasi Teks



- Asumsikan diantara fitur  $A_i$  adalah independen pada suatu kelas:

- $P(A_1, A_2, \dots, A_n | C_j) = P(A_1 | C_j) P(A_2 | C_j) \dots P(A_n | C_j)$

$$P(C | A_1 A_2 \dots A_n) = \frac{P(A_1 A_2 \dots A_n | C) P(C)}{P(A_1 A_2 \dots A_n)}$$
$$= \frac{\left( \prod_{i=1}^n P(A_i | C) \right) P(C)}{P(A_1 A_2 \dots A_n)}$$



## ➤ Fase Training

- Pada setiap kelas  $C$ :
  - Estimasi prior  $P(C)$  → untuk setiap fitur  $A$
- Untuk setiap nilai fitur  $v_i$  dari  $A$ :
  - Estimasi  $P(A=v_i | C)$

## ➤ Fase Testing

- Beberapa fitur ( $v_1, v_2, \dots, v_k$ )
  - Pilih kelas yang memaksimalkan:

$$\left( \prod_{i=1}^n P(A_i = v_i | C) \right) P(C)$$



## Naïve Bayes - Data Kontinyu

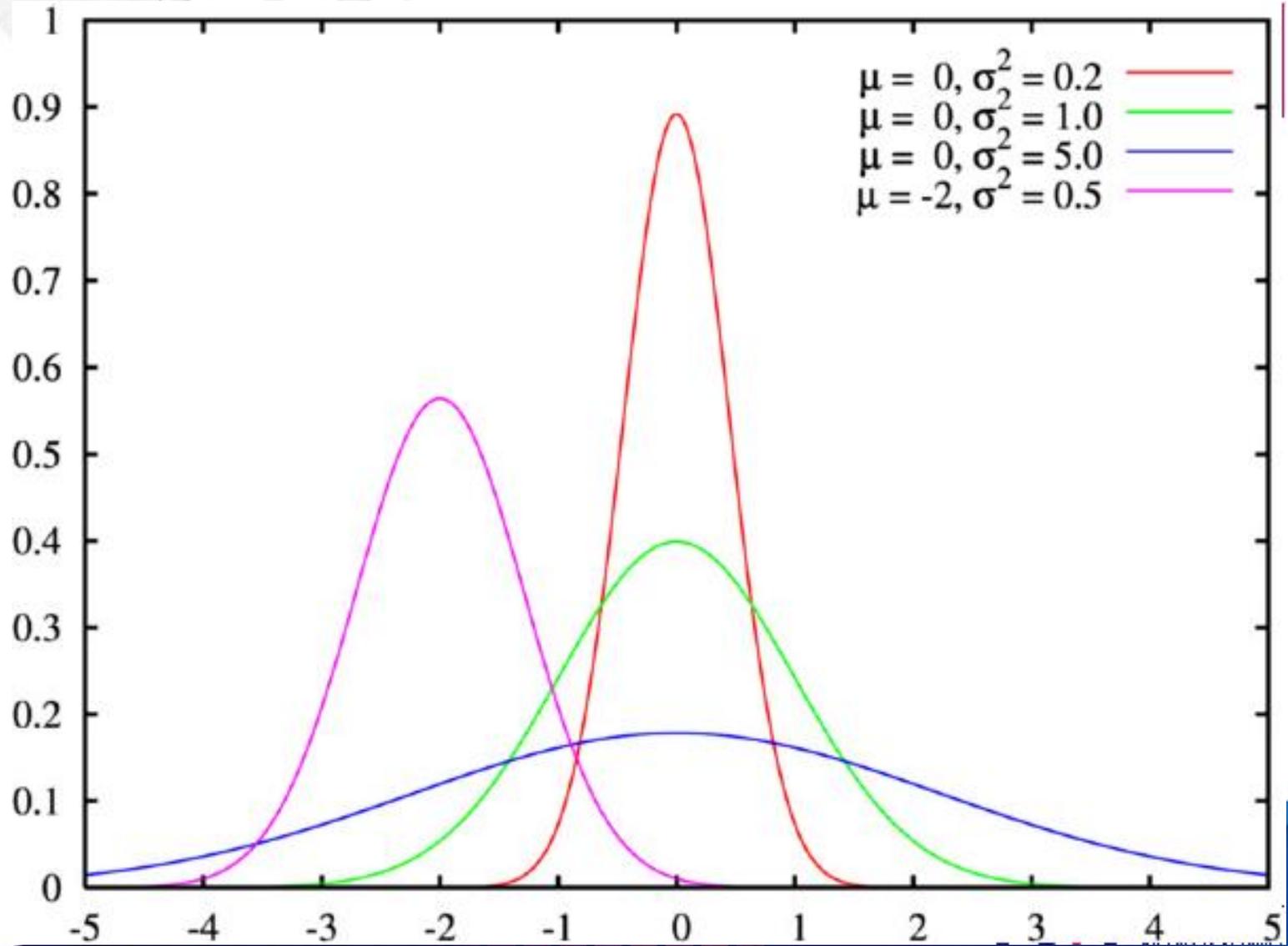
- Naive bayes classifier untuk menangani atribut bertipe kontinyu, salah satu caranya adalah menggunakan distribusi Gaussian.
- Distribusi ini dikarakterisasi dengan dua parameter yaitu mean ( $\mu$ ), dan variansi( $\sigma^2$ ).
- Untuk setiap kelas  $y_j$ , peluang kelas bersyarat untuk atribut  $X_i$  dinyatakan dengan persamaan distribusi Gaussian.

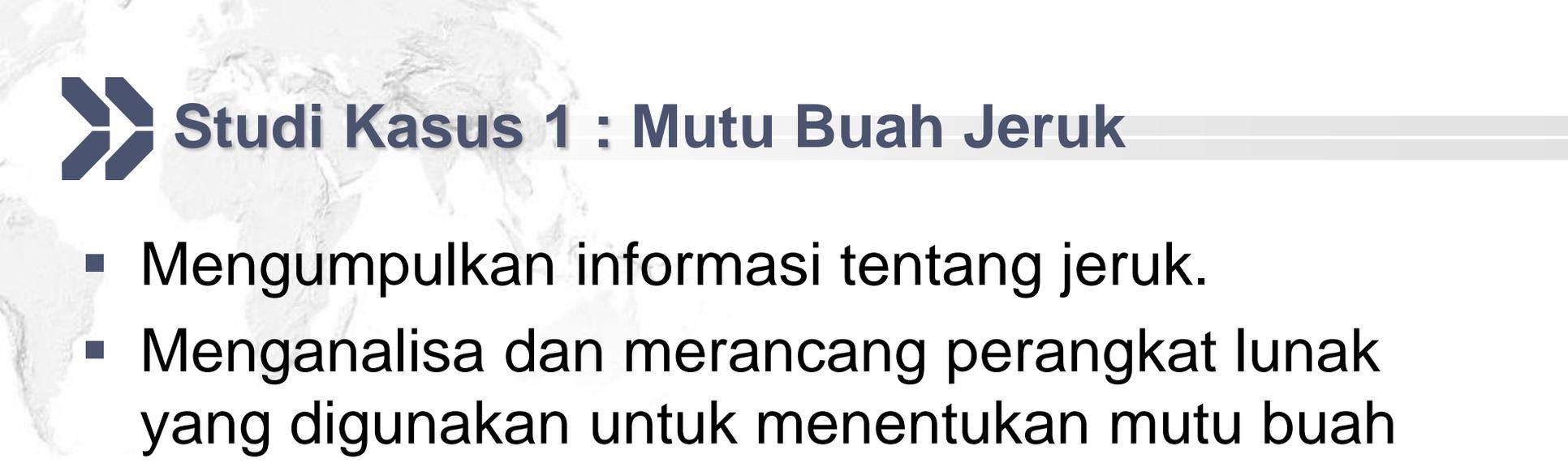
## ➤ Naïve Bayes - Data Kontinyu

- Fungsi densitas mengekspresikan probabilitas relatif.
- Data dengan mean  $\mu$  dan standar deviasi  $\sigma$ , fungsi densitas probabilitasnya adalah:

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- $\mu$  dan  $\sigma$  dapat diestimasi dari data, untuk setiap kelas.
- Gunakan  $\varphi_{\mu,\sigma}(x)$  untuk  $P(X | C)$





## ➤ Studi Kasus 1 : Mutu Buah Jeruk

- Mengumpulkan informasi tentang jeruk.
  - Menganalisa dan merancang perangkat lunak yang digunakan untuk menentukan mutu buah jeruk manis berdasarkan warna RGB dan diameter dengan menggunakan metode naive bayes.
- 



## **Untuk menghitung peluang, langkah-langkah yang dilakukan sebagai berikut :**

- Membuat image menjadi skala keabuan (gray- scale).
- Binarisasi. Pada tahap ini berguna untuk merubah nilai piksel image menjadi 0 dan 1. Nilai 0 untuk menggambarkan latar belakang, nilai 1 untuk objek buah jeruk.
- Max Filter. Tahap ini diperlukan untuk menghilangkan nilai piksel 1 tetapi bukan bagian dari buah jeruk.
- Temukan Parameter. Tahap ini digunakan menghitung nilai rata-rata red, green, blue, dan menghitung diameter buah jeruk.



- Penentuan Mutu Buah Jeruk.
- Bila parameter rata-rata  $R, G, B$  dan diameter  $D$  sudah diketahui, maka untuk menentukan mutu buah jeruk memiliki langkah-langkah sbb:
  - p1: Hitung Peluang  $(R, G, B, D)$  pada kelas A
  - p2: Hitung Peluang  $(R, G, B, D)$  pada kelas B
  - p3: Hitung Peluang  $(R, G, B, D)$  pada kelas C
  - p4: Hitung Peluang  $(R, G, B, D)$  pada kelas BS



- if  $((p1 \geq p2) \text{ and } (p1 \geq p3) \text{ and } (p1 \geq p4))$  then Output 'Quality A with probability  $p1$ '
- if  $((p2 \geq p1) \text{ and } (p2 \geq p3) \text{ and } (p2 \geq p4))$  then Output 'Quality B with probability  $p2$ '
- if  $((p3 \geq p1) \text{ and } (p3 \geq p2) \text{ and } (p3 \geq p4))$  then Output 'Quality C with probability  $p3$ '
- if  $((p4 \geq p1) \text{ and } (p4 \geq p2) \text{ and } (p4 \geq p3))$  then Output 'Quality BS with probability  $p4$ '



### Naive Bayes Classifier

Iterative



Open  
Hitung Parameter  
Test  
Close

Quality BS with probability : 0.0000000000

Threshold Object  
386

Mean R	Mean G	Mean B	Diameter	Quality
165	175	70	135	BS

D:\Program\Naive Bayes-Journal\evaluasi new\10B.bmp

Id	R	G	B	D	Mutu
1	82	78	36	119	A
2	103	92	45	116	A
3	88	88	60	111	A
4	86	85	51	108	A
5	140	109	16	101	B
6	151	116	17	99	B
7	157	113	28	92	C
8	172	122	21	89	C
9	178	121	18	93	C

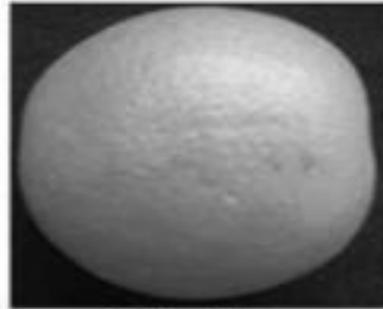


Keterangan :

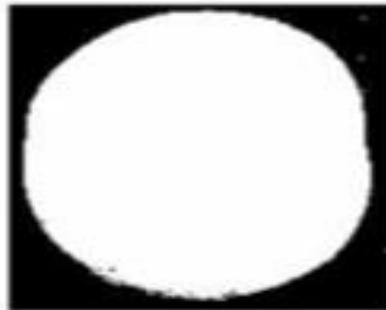
- Display vektor pembelajaran
- Display gambar jeruk yang akan diuji mutunya.
- Pada gambar tersebut terdapat dua garis vertikal yang digunakan untuk menghitung diameter buah jeruk (jarak antara dua garis vertikal).
- Display untuk mengetahui mutu jeruk yang sedang ditampilkan. Pada image yang ditampilkan, mutu yang dihasilkan adalah BS.



Gb. 4.a



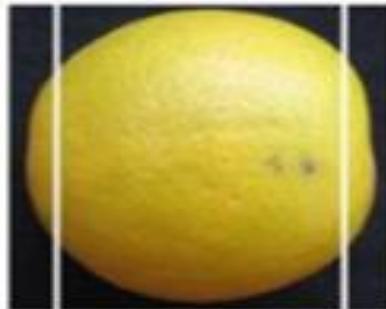
Gb. 4.b



Gb. 4.c



Gb. 4.d



Gb. 4.e

**Keterangan**

Gb. 4.a Gambar asal

Gb. 4.b Gray-Scale

Gb. 4.c Binarisasi

Gb. 4.d Max Filter

Gb. 4.e Find Parameter



## Studi Kasus 2 : Bermain Tenis

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
Day1	Sunny	Hot	High	Weak	No
Day2	Sunny	Hot	High	Strong	No
Day3	Overcast	Hot	High	Weak	Yes
Day4	Rain	Mild	High	Weak	Yes
Day5	Rain	Cool	Normal	Weak	Yes
Day6	Rain	Cool	Normal	Strong	No
Day7	Overcast	Cool	Normal	Strong	Yes
Day8	Sunny	Mild	High	Weak	No
Day9	Sunny	Cool	Normal	Weak	Yes
Day10	Rain	Mild	Normal	Weak	Yes
Day11	Sunny	Mild	Normal	Strong	Yes
Day12	Overcast	Mild	High	Strong	Yes
Day13	Overcast	Hot	Normal	Weak	Yes
Day14	Rain	Mild	High	Strong	No



Klasifikasi fakta baru berupa  $\mathbf{x}=(a_1, \dots, a_T)$  sebagaimana:

$$h_{Naive\ Bayes} = \arg \max_h P(h)P(\mathbf{x} | h) = \arg \max_h P(h) \prod_t P(a_t | h)$$

- Untuk melakukannya berdasarkan pada sampel pelatihan, kita perlu untuk mengestimasi parameter-parameter dari sampel pelatihan:
  - Untuk setiap kelas (hypothesis)  $h$

$$\hat{P}(h) := \text{estimate } P(h)$$

- Untuk setiap nilai fitur  $a_t$  dari setiap contoh fakta

$$\hat{P}(a_t | h) := \text{estimate } P(a_t | h)$$



Berdasarkan pada contoh tabel klasifikasi fakta x berikut :

$x = (\text{Outl} = \text{Sunny}, \text{Temp} = \text{Cool}, \text{Hum} = \text{High}, \text{Wind} = \text{strong})$

- That means: Play tennis or not?

$$h_{NB} = \arg \max_{h \in [\text{yes}, \text{no}]} P(h)P(\mathbf{x} | h) = \arg \max_{h \in [\text{yes}, \text{no}]} P(h) \prod_t P(a_t | h)$$

$$= \arg \max_{h \in [\text{yes}, \text{no}]} P(h)P(\text{Outlook} = \text{sunny} | h)P(\text{Temp} = \text{cool} | h)P(\text{Humidity} = \text{high} | h)P(\text{Wind} = \text{strong} | h)$$

- Working:

$$P(\text{PlayTennis} = \text{yes}) = 9 / 14 = 0.64$$

$$P(\text{PlayTennis} = \text{no}) = 5 / 14 = 0.36$$

$$P(\text{Wind} = \text{strong} | \text{PlayTennis} = \text{yes}) = 3 / 9 = 0.33$$

$$P(\text{Wind} = \text{strong} | \text{PlayTennis} = \text{no}) = 3 / 5 = 0.60$$

*etc.*

$$P(\text{yes})P(\text{sunny} | \text{yes})P(\text{cool} | \text{yes})P(\text{high} | \text{yes})P(\text{strong} | \text{yes}) = 0.0053$$

$$P(\text{no})P(\text{sunny} | \text{no})P(\text{cool} | \text{no})P(\text{high} | \text{no})P(\text{strong} | \text{no}) = \mathbf{0.0206}$$

$\Rightarrow \text{answer} : \text{PlayTennis}(x) = \text{no}$



## Latihan

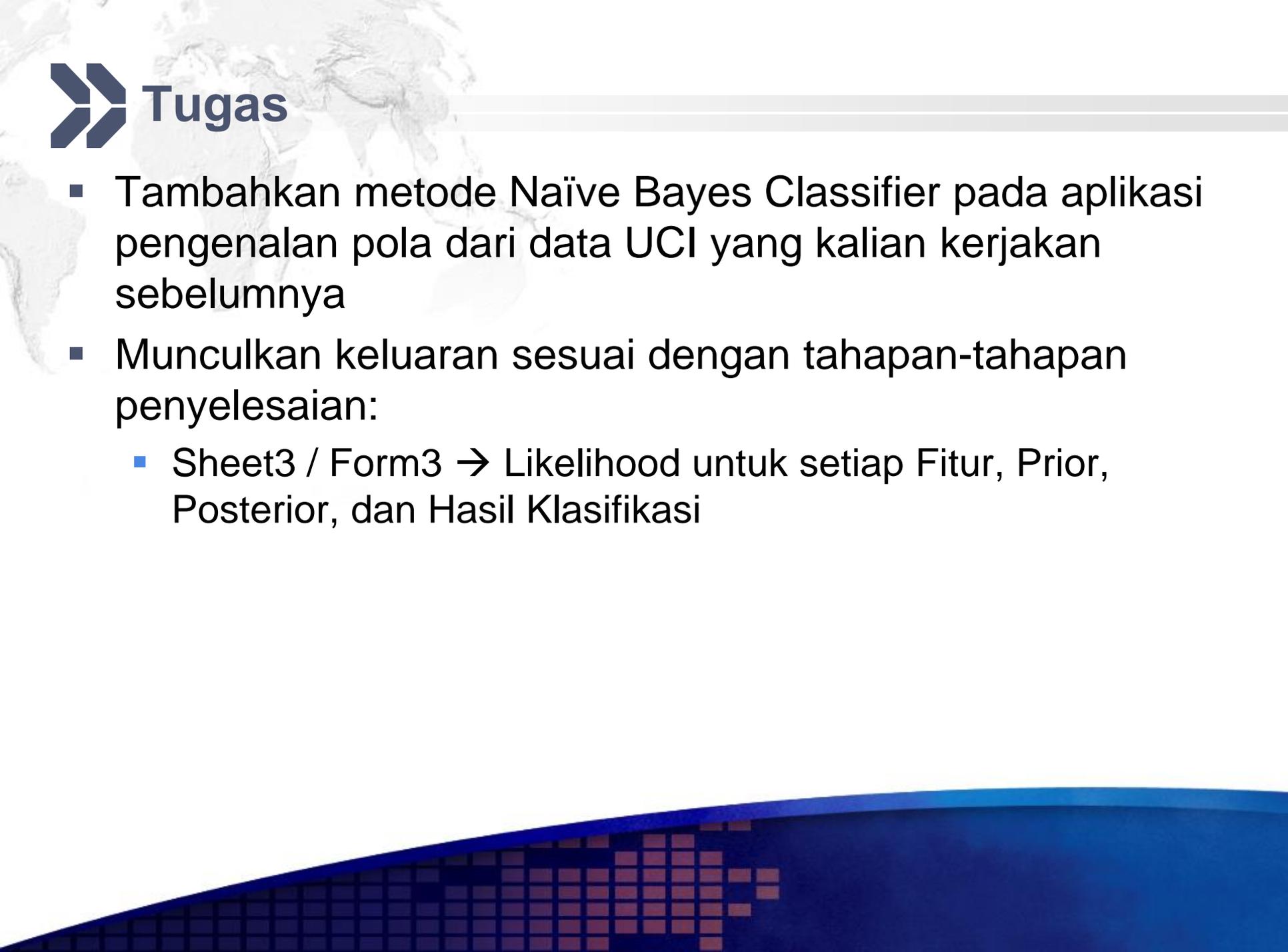
Coba Lakukan Tes Data :

X = (refund = no, marital status = divorced, income = 120k)  
masuk kelas “evade” **No** atau **Yes**?

Gunakan formula berikut untuk menghitung likelihood data Kuantitatif (continue)

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

<i>Tid</i>	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



## ➤ Tugas

- Tambahkan metode Naïve Bayes Classifier pada aplikasi pengenalan pola dari data UCI yang kalian kerjakan sebelumnya
- Munculkan keluaran sesuai dengan tahapan-tahapan penyelesaian:
  - Sheet3 / Form3 → Likelihood untuk setiap Fitur, Prior, Posterior, dan Hasil Klasifikasi



**Thank You !**

afif.supianto@ub.ac.id  
081 331 834 734 / 088 160 127 40