

Machine Learning

Metode Bayesian

Entin Martiana

Knowledge Engineering Research Group

Soft Computing Laboratory

Department of Information and Computer Engineering

Politeknik Elektronika Negeri Surabaya



Politeknik Elektronika Negeri Surabaya
Departemen Teknik Informatika dan Komputer

Konten

- Mengapa Metode Bayes
- Probabilitas Bersyarat
- Probabilitas Bersyarat dalam Data
- MAP Hypothesis

Tujuan Instruksi Umum

Mahasiswa mampu menyelesaikan masalah – masalah menggunakan metode mesin pembelajaran yang tepat berdasarkan supervised, unsupervised dan reinforcement learning, baik secara individu maupun berkelompok/kerjasama tim.

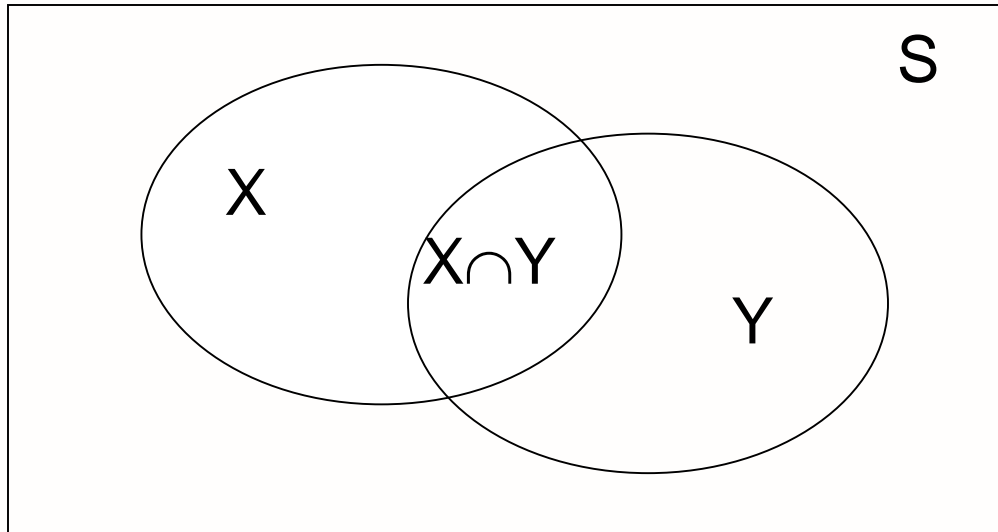
Tujuan Instruksi Khusus

- Memahami metode naive Bayes
- Memahami metode Bayes Gaussian
- Mampu menerapkan Bayesian

Mengapa Metode Bayes

- Metode Find-S tidak dapat digunakan untuk data yang tidak konsisten dan data yang bias, sehingga untuk bentuk data semacam ini salah satu metode sederhana yang dapat digunakan adalah metode bayes.
- Metode Bayes ini merupakan metode yang baik di dalam mesin pembelajaran berdasarkan data training, dengan menggunakan probabilitas bersyarat sebagai dasarnya.

Probabilitas Bersyarat



$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)}$$

Probabilitas X di dalam Y adalah probabilitas interseksi X dan Y dari probabilitas Y , atau dengan bahasa lain $P(X|Y)$ adalah prosentase banyaknya X di dalam Y

Probabilitas Bersyarat Dalam Data

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Banyaknya data berolah-raga=ya adalah 4 dari 6 data maka dituliskan

$$P(\text{Olahraga}=\text{Ya}) = 4/6$$

Banyaknya data cuaca=cerah dan berolah-raga=ya adalah 4 dari 6 data maka dituliskan

$$P(\text{cuaca}=\text{cerah dan Olahraga}=\text{Ya}) = 4/6$$

$$P(\text{cuaca} = \text{cerah} | \text{olahraga} = \text{ya}) = \frac{4/6}{4/6} = 1$$

Probabilitas Bersyarat Dalam Data

#	Cuaca	Temperatur	Berolahraga
1	cerah	normal	ya
2	cerah	tinggi	ya
3	hujan	tinggi	tidak
4	cerah	tinggi	tidak
5	hujan	normal	tidak
6	cerah	normal	ya

Banyaknya data berolah-raga=ya adalah 3 dari 6 data maka dituliskan

$$P(\text{Olahraga}=\text{Ya}) = 3/6$$

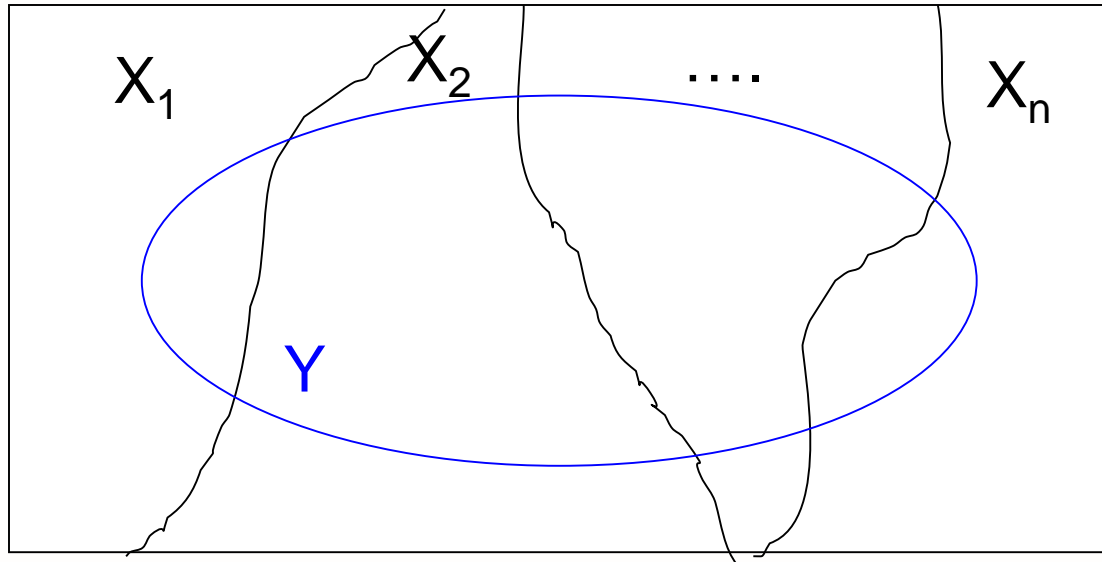
Banyaknya data cuaca=cerah, temperatur=normal dan berolah-raga=ya adalah 4 dari 6 data maka dituliskan

$$P(\text{cuaca}=\text{cerah}, \text{temperatur}=\text{normal}, \text{Olahraga}=\text{Ya}) = 2/6$$



$$P(\text{cuaca} = \text{cerah}, \text{temperatur} = \text{normal} \mid \text{olahraga} = \text{ya}) = \frac{2/6}{3/6} = \frac{2}{3}$$

Metode Bayes



$$P(X_k | Y) = \frac{P(Y | X_k)}{\sum_i P(Y | X_i)}$$

Keadaan Posteriror (Probabilitas X_k di dalam Y) dapat dihitung dari keadaan prior (Probabilitas Y di dalam X_k dibagi dengan jumlah dari semua probabilitas Y di dalam semua X_i)

MAP Hypotesis

MAP (*Maximum A propri Probability*) Hypothesis menyatakan hipotesa yang diambil berdasarkan nilai probabilitas berdasarkan kondisi prior yang diketahui.

$$\begin{aligned}h_{\text{MAP}} &= \operatorname{argmax}_{h \in H} P(h | D) \\ &= \operatorname{argmax}_{h \in H} \frac{P(D | h) P(h)}{P(D)} \\ &= \operatorname{argmax}_{h \in H} P(D | h) P(h)\end{aligned}$$

Contoh MAP Hypotheses

Diketahui hasil survey yang dilakukan sebuah lembaga kesehatan menyatakan bahwa 90% penduduk di dunia menderita sakit paru-paru. Dari 90% penduduk yang sakit paru-paru ini 60% adalah perokok, dan dari penduduk yang tidak menderita sakit paru-paru 20% perokok.

Fakta ini bisa didefinisikan dengan: X =sakit paru-paru dan Y =perokok.

Maka : $P(X) = 0.9$

$$P(\sim X) = 0.1$$

$$P(Y|X) = 0.6 \rightarrow P(\sim Y|X) = 0.4$$

$$P(Y|\sim X) = 0.2 \rightarrow P(\sim Y|\sim X) = 0.8$$

Dengan metode bayes dapat dihitung:

$$P(Y|X).P(X) = (0.6) \cdot (0.9) = 0.54$$

$$P(Y|\sim X) P(\sim X) = (0.2) \cdot (0.1) = 0.02$$

$$P(\{Y\}|X) = 0.54/(0.54+0.02) = 0.96$$

$$P(\{Y\}|\sim X) = 0.02/(0.54+0.02) = 0.04$$

Bila diketahui seseorang merokok, maka dia menderita sakit paru-paru karena $P(\{Y\}|X)$ lebih besar dari $P(\{Y\}|\sim X)$. HMAP diartikan mencari probabilitas terbesar dari semua instance pada atribut target atau semua kemungkinan keputusan.



Bayes Theorem

- Goal: To determine the most probable hypothesis, given the data D plus any initial knowledge about the prior probabilities of the various hypotheses in H .
- Prior probability of h , $P(h)$: it reflects any background knowledge we have about the chance that h is a correct hypothesis (before having observed the data).
- Prior probability of D , $P(D)$: it reflects the probability that training data D will be observed given no knowledge about which hypothesis h holds.
- Conditional Probability of observation D , $P(D|h)$: it denotes the probability of observing data D given some world in which hypothesis h holds.



Bayes Theorem

- **Posterior probability of h , $P(h/D)$:** it represents the probability that h holds given the observed training data D . It reflects our confidence that h holds after we have seen the training data D and it is the quantity that Machine Learning researchers are interested in.
- **Bayes Theorem** allows us to compute $P(h/D)$:

$$P(h | D) = P(D | h)P(h) / P(D)$$

Maximum A Posteriori (MAP) Hypothesis and Maximum Likelihood

- **Goal:** To find the most probable hypothesis h from a set of candidate hypotheses H given the observed data D .
- **MAP Hypothesis, $h_{MAP} = \operatorname{argmax}_{h \in H} P(h|D)$**
$$= \operatorname{argmax}_{h \in H} P(D|h)P(h)/P(D)$$
$$= \operatorname{argmax}_{h \in H} P(D|h)P(h)$$
- If every hypothesis in H is equally probable a priori, we only need to consider the likelihood of the data D given h , $P(D|h)$. Then, h_{MAP} becomes the **Maximum Likelihood**,

$$h_{ML} = \operatorname{argmax}_{h \in H} P(D|h)$$



Bayes Optimal Classifier

- One great advantage of Bayesian Decision Theory is that it gives us a lower bound on the classification error that can be obtained for a given problem.
- **Bayes Optimal Classification:** The most probable classification of a new instance is obtained by combining the predictions of all hypotheses, weighted by their posterior probabilities:

$$\mathit{argmax}_{v_j \in V} \sum_{h_i \in H} P(v_h/h_i)P(h_i/D)$$

where V is the set of all the values a classification can take and v_j is one possible such classification.

Unfortunately, Bayes Optimal Classifier is usually too costly to apply!

==> Naïve Bayes Classifier



HMAP Dari Data Training

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Asumsi:

$Y = \text{berolahraga},$

$X_1 = \text{cuaca},$

$X_2 = \text{temperatur},$

$X_3 = \text{kecepatan angin}.$

Fakta menunjukkan:

$$P(Y=\text{ya}) = 4/6 \rightarrow P(Y=\text{tidak}) = 2/6$$

HMAP Dari Data Training

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Apakah bila cuaca cerah dan kecepatan angin kencang, orang akan berolahraga?

Fakta: $P(X1=cerah | Y=ya) = 1$, $P(X1=cerah | Y=tidak) = 0$
 $P(X3=kencang | Y=ya) = 1/4$, $P(X3=kencang | Y=tidak) = 1/2$

HMAP dari keadaan ini dapat dihitung dengan:

$$P(X1=cerah, X3=kencang | Y=ya) \\ = \{ P(X1=cerah | Y=ya) \cdot P(X3=kencang | Y=ya) \} \cdot P(Y=ya) \\ = \{ (1) \cdot (1/4) \} \cdot (4/6) = 1/6$$

$$P(X1=cerah, X3=kencang | Y=tidak) \\ = \{ P(X1=cerah | Y=tidak) \cdot P(X3=kencang | Y=tidak) \} \cdot P(Y=tidak) \\ = \{ (0) \cdot (1/2) \} \cdot (2/6) = 0$$

KEPUTUSAN ADALAH BEROLAHRAGA = YA

Naïve Bayes Algorithm

- Naïve Bayes Algorithm (for discrete input attributes) has two phases

– **1. Learning Phase:** Given a training set S ,

Learning is easy, just create probability tables.

For each target value of c_i ($c_i = c_1, \dots, c_L$)

$\hat{P}(C = c_i) \leftarrow$ estimate $P(C = c_i)$ with examples in S ;

For every attribute value x_{jk} of each attribute X_j ($j = 1, \dots, n; k = 1, \dots, N_j$)

$\hat{P}(X_j = x_{jk} | C = c_i) \leftarrow$ estimate $P(X_j = x_{jk} | C = c_i)$ with examples in S ;

Output: conditional probability tables; for $N_j \times L$ elements

– **2. Test Phase:** Given an unknown instance

$$X' = (a'_1, \dots, a'_n)$$

Look up tables to assign the label c^* to X' if

$$[\hat{P}(a'_1 | c^*) \cdots \hat{P}(a'_n | c^*)] \hat{P}(c^*) > [\hat{P}(a'_1 | c) \cdots \hat{P}(a'_n | c)] \hat{P}(c), \quad c \neq c^*, c = c_1, \dots, c_L$$

Classification is easy, just multiply probabilities



Tennis Example

- Example: Play Tennis

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



PlayTennis: training examples

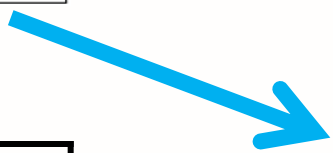
The learning phase for tennis example

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$P(\text{Play}=\text{Yes}) = 9/14$


$P(\text{Play}=\text{No}) = 5/14$

We have four variables, we calculate for each we calculate the conditional probability table



Outlook	Play=Yes	Play=No
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	Play=Yes	Play=No
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5



Humidity	Play=Yes	Play=No
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=Yes	Play=No
Strong	3/9	3/5
Weak	6/9	2/5

The *test phase* for the tennis example

- **Test Phase**

- Given a **new instance of variable values**,

$x' = (\text{Outlook}=\textit{Sunny}, \text{Temperature}=\textit{Cool}, \text{Humidity}=\textit{High}, \text{Wind}=\textit{Strong})$

- Given calculated Look up tables

$$P(\text{Outlook}=\textit{Sunny} \mid \text{Play}=\textit{Yes}) = 2/9$$

$$P(\text{Temperature}=\textit{Cool} \mid \text{Play}=\textit{Yes}) = 3/9$$

$$P(\text{Humidity}=\textit{High} \mid \text{Play}=\textit{Yes}) = 3/9$$

$$P(\text{Wind}=\textit{Strong} \mid \text{Play}=\textit{Yes}) = 3/9$$

$$P(\text{Play}=\textit{Yes}) = 9/14$$

$$P(\text{Outlook}=\textit{Sunny} \mid \text{Play}=\textit{No}) = 3/5$$

$$P(\text{Temperature}=\textit{Cool} \mid \text{Play}=\textit{No}) = 1/5$$

$$P(\text{Humidity}=\textit{High} \mid \text{Play}=\textit{No}) = 4/5$$

$$P(\text{Wind}=\textit{Strong} \mid \text{Play}=\textit{No}) = 3/5$$

$$P(\text{Play}=\textit{No}) = 5/14$$

- **Use the MAP rule to calculate Yes or No**

$$P(\text{Yes} \mid x'): [P(\textit{Sunny} \mid \textit{Yes})P(\textit{Cool} \mid \textit{Yes})P(\textit{High} \mid \textit{Yes})P(\textit{Strong} \mid \textit{Yes})]P(\text{Play}=\textit{Yes}) = 0.0053$$

$$P(\text{No} \mid x'): [P(\textit{Sunny} \mid \textit{No})P(\textit{Cool} \mid \textit{No})P(\textit{High} \mid \textit{No})P(\textit{Strong} \mid \textit{No})]P(\text{Play}=\textit{No}) = 0.0206$$

Given the fact $P(\text{Yes} \mid x') < P(\text{No} \mid x')$, we label x' to be "No".

An Example of the Naïve Bayes Classifier

The weather data, with counts and probabilities

	outlook		temperature			humidity			windy		play		
	yes	no	yes	no	yes	no	yes	no	yes	no			
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

A new day

outlook	temperature	humidity	windy	play
sunny	cool	high	true	?



- Likelihood of yes

$$= \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14} = 0.0053$$

- Likelihood of no

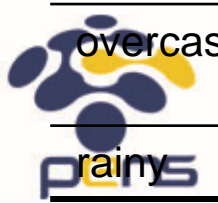
$$= \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14} = 0.0206$$

- Therefore, the prediction is No

The Naive Bayes Classifier for Data Sets with Numerical Attribute Values

- One common practice to handle numerical attribute values is to assume normal distributions for numerical attributes.

The numeric weather data with summary statistics													
outlook			temperature			humidity			windy		play		
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2	3		83	85		86	85	false	6	2	9	5
overcast	4	0		70	80		96	90	true	3	3		
rainy	3	2		68	65		80	70					
				64	72		65	95					
				69	71		70	91					
				75			80						
				75			70						
				72			90						
				81			75						
sunny	2/9	3/5	mean	73	74.6	mean	79.1	86.2	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	std dev	6.2	7.9	std dev	10.2	9.7	true	3/9	3/5		
rainy	3/9	2/5											



- Let x_1, x_2, \dots, x_n be the values of a numerical attribute in the training data set.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$$

$$f(w) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(w-\mu)^2}{\sigma^2}}$$

- For examples,

$$f(\text{temperature} = 66 \mid \text{Yes}) = \frac{1}{\sqrt{2\pi}(6.2)} e^{-\frac{(66-73)^2}{2(6.2)^2}} = 0.0340$$

- Likelihood of Yes = $\frac{2}{9} \times 0.0340 \times 0.0221 \times \frac{3}{9} \times \frac{9}{14} = 0.000036$

- Likelihood of No = $\frac{3}{5} \times 0.0291 \times 0.038 \times \frac{3}{5} \times \frac{5}{14} = 0.000136$

Kelemahan Metode Bayes

- Metode Bayes hanya bisa digunakan untuk persoalan klasifikasi dengan *supervised learning*.
- Metode Bayes memerlukan pengetahuan awal untuk dapat mengambil suatu keputusan. Tingkat keberhasilan metode ini sangat tergantung pada pengetahuan awal yang diberikan.

Beberapa Aplikasi Metode Bayes

- Menentukan diagnosa suatu penyakit berdasarkan data-data gejala (sebagai contoh hipertensi atau sakit jantung).
- Mengenali buah berdasarkan fitur-fitur buah seperti warna, bentuk, rasa dan lain-lain
- Mengenali warna berdasarkan fitur indeks warna RGB
- Mendeteksi warna kulit (*skin detection*) berdasarkan fitur warna chrominant
- Menentukan keputusan aksi (olahraga, art, psikologi) berdasarkan keadaan.
- Menentukan jenis pakaian yang cocok untuk keadaan-keadaan tertentu (seperti cuaca, musim, temperatur, acara, waktu, tempat dan lain-lain)
- Menentukan ekspresi (sedih, gembira, dll) dari kalimat yang diucapkan



Latihan Soal

WAKTU	PAKET	FREKWEKSI	PRIORITAS	GANGGUAN
PENDEK	BESAR	SEDANG	RENDAH	GANGGUAN
PENDEK	KECIL	RENDAH	TINGGI	GANGGUAN
PANJANG	BESAR	SEDANG	TINGGI	NORMAL
PANJANG	KECIL	TINGGI	RENDAH	NORMAL
PENDEK	BESAR	TINGGI	TINGGI	GANGGUAN
PANJANG	KECIL	RENDAH	TINGGI	GANGGUAN
PANJANG	KECIL	TINGGI	RENDAH	GANGGUAN
PANJANG	KECIL	SEDANG	RENDAH	NORMAL
PANJANG	BESAR	TINGGI	TINGGI	NORMAL
PANJANG	KECIL	SEDANG	RENDAH	GANGGUAN
PENDEK	BESAR	SEDANG	TINGGI	NORMAL
PANJANG	BESAR	RENDAH	TINGGI	NORMAL

1. Implementasikan fase training untuk data training di atas!
2. Berilah fasilitas untuk memasukkan data test!
3. Berapa persen besarnya error yang terjadi jika data tersebut dimasukkan pada fase training?



Latihan Soal

USIA	KELAMIN	MEROKOK	OLAHRAGA	JANTUNG
TUA	PRIA	TIDAK	YA	TIDAK
TUA	PRIA	YA	YA	TIDAK
MUDA	PRIA	YA	TIDAK	TIDAK
TUA	PRIA	TIDAK	TIDAK	TIDAK
MUDA	WANITA	TIDAK	TIDAK	YA
MUDA	PRIA	TIDAK	YA	YA
MUDA	PRIA	TIDAK	YA	TIDAK
TUA	WANITA	TIDAK	TIDAK	YA
MUDA	PRIA	YA	TIDAK	TIDAK
TUA	PRIA	YA	TIDAK	TIDAK
MUDA	PRIA	YA	YA	YA
TUA	PRIA	YA	TIDAK	TIDAK
MUDA	PRIA	TIDAK	TIDAK	TIDAK
TUA	PRIA	TIDAK	YA	TIDAK
MUDA	PRIA	YA	TIDAK	TIDAK

1. Implementasikan fase training untuk data training di atas!
2. Berilah fasilitas untuk memasukkan data test!
3. Berapa persen besarnya error yang terjadi jika data tersebut dimasukkan pada fase training?



Referensi

- Modul Ajar Machine Learning, Entin Martiana, Ali Ridho Barakbah, Nur Rosyid Mubtadaí, Politeknik Elektronika Negeri Surabaya, 2013.
- Machine Learning, Tom Mitchell, McGraw-Hill. 2008.



bridge to the future

<http://www.eepis-its.edu>