



## **Lembar Kerja Praktikum 3**

### **Mata Kuliah Topik dalam Data Mining Terapan (KOM623)**

**Semester Genap Tahun Akademik 2020/2021**

**Nama : Siti Nur Hasanah**  
**NIM : G64170036**

---

#### **Topik: Klasifikasi Bayes**

```
library(EBImage)

x <- readImage(system.file('images','shapes.png', package='EBImage'))

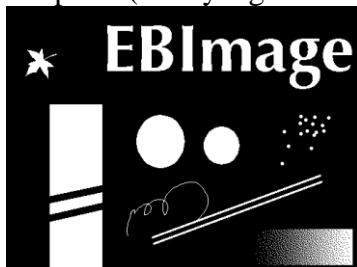
x <- x[110:312,1:130]

y <- bwlabel(x)

display(y, title='Binary')
```

#### **Question 1. Jelaskan apa maksud dari potongan kode di atas?**

Kode di atas melakukan pembacaan salah satu citra yang ada di *library* EbImage.  
Output x (citra yang dibaca)



Berikutnya disimpan hanya pixel pada indeks tertentu saja, yaitu pixel-pixel pada baris 110-312, kolom 1-130.

Outputnya:



#Unduh citra buah (cth:alpukat) dari internet,dan baca melalui R

```
original_image <- readImage(file.choose())
```

```
display(original_image)
```

```
r = channel(original_image,"r")
```

```
g = channel(original_image,"g")
```

```
b = channel(original_image,"b")
```

```
new_image = 0.2126*r+0.7152*g+0.0722*b
```

```
display(new_image)
```

**Question 2. Jelaskan apa maksud dari potongan kode di atas? Copy hasil display ke dalam box ini.**

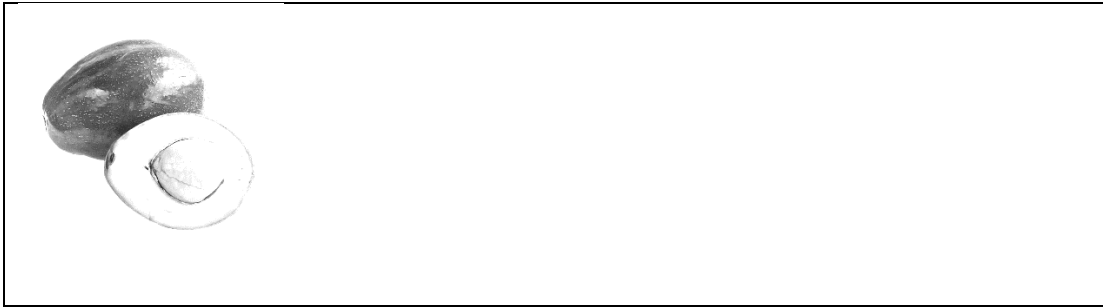
Kode ini membaca citra input yang dipilih dari folder pengguna.

Output original\_image:



Berikutnya dilakukan proses konversi citra RGB ke grayscale. Pertama diambil nilai masing-masing channel R, G dan B pada original\_image. Setelah itu dilakukan konversi dari citra RGB ke grayscale dengan menggunakan persamaan berikut  $new\_image = 0.2126*r + 0.7152*g + 0.0722*b$ .

Output:



```
Dataimage <- new_image@.Data
```

```
Subdata1 <- Dataimage[110:312,130:200]
```

```
display(Subdata1)
```

```
Subdata2<- Dataimage[c(1:40, 100:150, 350:400 ), c(1:40, 100:150, 250:300)]
```

```
display(Subdata2)
```

**Question 3. Jelaskan apa maksud dari potongan kode di atas? Copy hasil display ke dalam box di bawah ini**

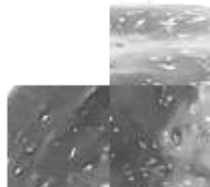
Pada Subdata1 diambil nilai matriks dari data image pada indeks baris 110-312, kolom 130-200.

Output Subdata1



Pada Subdata2, dibuat list yang berisi piksel-piksel pada beberapa posisi. Subdata2 akan menghasilkan satu gambar, yang terdiri dari potongan gambar-gambar pada posisi-posisi yang dipilih. Ada 9 potongan gambar yang dijadikan 1, yaitu piksel pada posisi [1:40, 1:40], [1:40, 100:150], [1:40, 250:300], [100:150, 1:40], [100:150, 100:150], [100:150, 250:300], [350:400, 1:40], [350:400, 100:150] dan [350:400, 250:300].

Output Subdata2



```

# Unduh citra buah lain, dan lakukan langkah yang sama dengan sebelumnya
# Ekstrak nilai citra dengan nama Dataimage2

Dataimage2 <- Dataimage2[1:dim(Dataimage)[1], 1:dim(Dataimage)[2]]

obs1 <- as.vector(t(Dataimage))
obs2 <- as.vector(t(Dataimage2))

obs_gabung <- rbind(obs1,obs2)

dataset_buah <- as.data.frame(obs_gabung)

klas<- c("alpokat", "apel")

dataset_buah_baru<-cbind(dataset_buah, klas)

dim(dataset_buah_baru)

dataset_buah_baru[1,640001]

dataset_buah_baru[2,640001]

```

**Question 4. Jelaskan apa maksud dari potongan kode di atas? Copy hasil display ke dalam box di bawah ini**

Pada potongan kode ini, membuat data frame dari dua buah vector piksel citra buah. Data frame ini berukuran [2,640001], baris pertama merupakan vector dari data citra buah pertama (buah alpukat), baris kedua merupakan vector dari citra buah kedua (buah apel). Terakhir dibuat dataset baru dengan menambahkan kelas (label) masing-masing citra buah.

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D1	Cerah	Normal	Pelan	Ya
D2	Cerah	Normal	Pelan	Ya
D3	Hujan	Tinggi	Pelan	Tidak
D4	Cerah	Normal	Kencang	Ya
D5	Hujan	Tinggi	Kencang	Tidak
D6	Cerah	Normal	Pelan	Ya

Banyaknya data "Berolah-raga = Ya" adalah 4 dari 6 data maka dituliskan  $P(\text{Berolah-raga}) = 4/6$ . Banyaknya data dengan atribut Cuaca Cerah dan target Berolah-raga Ya, adalah 4 dari 6 data, maka dituliskan  $P(\text{cuaca=cerah dan Olahraga=ya}) = 4/6$ .

Question 5. Dari informasi tersebut, maka probabilitas cuaca cerah pada saat olahraga adalah?

$$\begin{aligned}
 5). P(\text{Berolahraga} = Y_a) &= \frac{4}{6} \\
 P(\text{Cuaca} = \text{cerah} \cap \text{Berolahraga} = Y_a) &= \frac{4}{6} \\
 P(\text{Cuaca} = \text{cerah} | \text{Berolahraga} = Y_a) &= \frac{P(\text{Cuaca} = \text{cerah} \cap \text{Berolahraga} = Y_a)}{P(\text{Berolahraga} = Y_a)} \\
 &= \frac{4/6}{4/6} \\
 &= 1
 \end{aligned}$$

Diketahui hasil survey yang dilakukan sebuah lembaga kesehatan menyatakan bahwa 30% penduduk di dunia menderita sakit paru-paru. Dari 90% penduduk yang sakit paru-paru 60% adalah perokok, dan dari penduduk yang tidak sakit paru-paru 20% adalah perokok.

Fakta hasil survey didefinisikan sebagai berikut:

Y : Sakit paru-paru

X : Perokok

$\sim Y$  : Tidak sakit Paru

$\sim X$  : Bukan Perokok

Question 6. Berdasar hasil survey tersebut, apakah dapat dinyatakan bahwa besar kemungkinan akan menderita sakit paru-paru jika orang tersebut perokok??  $P(Y=\{\text{paru, tidak paru}\} | \{X\})$ ?

b)  $Y = \text{sakit paru-paru}$   $X = \text{perokok}$   
 $-Y = \text{tidak sakit paru-paru}$   $-X = \text{bukan perokok}$

$P(Y = \{\text{paru, tidak paru}\} | \{X\})$

$$P(Y) = 0,9$$

$$P(-Y) = 0,1$$

$$P(X|Y) = 0,6$$

$$P(-X|Y) = 0,4$$

$$P(X|-Y) = 0,2$$

$$P(-X|-Y) = 0,8$$

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

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

Membandingkan  $P(Y|X)$  dan  $P(-Y|X)$

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

$$= \frac{0,6 \cdot 0,9}{0,2 \cdot 0,1}$$

$$\frac{P(Y|X)}{P(-Y|X)} = \frac{0,54}{0,02}$$

Kesimpulan:

Jadi berdasarkan hasil survey tersebut, dapat dinyatakan bahwa besar kemungkinan seseorang akan menderita sakit paru jika seseorang tersebut merokok.

#	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

Question 7. Apakah keputusan Berolah-raga, bila Cuaca cerah, Temperatur Tinggi, dan Kecepatan angin kencang??

7). Apakah keputusan berolahraga, bila cuaca cerah, temperatur tinggi, dan kecepatan angin kencang?

Misal:  $Y$  = Berolahraga  
 $X_1$  = cuaca  
 $X_2$  = temperatur  
 $X_3$  = Kecepatan angin

$$P(Y = Y_a) = \frac{4}{6} \quad P(Y = \text{Tidak}) = \frac{2}{6}$$

$$P(X_1 = \text{cerah} | Y = Y_a) = 1 \quad P(X_1 = \text{Hujan} | Y = Y_a) = 0$$

$$P(X_1 = \text{cerah} | Y = \text{Tidak}) = 0 \quad P(X_1 = \text{Hujan} | Y = \text{Tidak}) = 1$$

$$P(X_2 = \text{Normal} | Y = Y_a) = 1 \quad P(X_2 = \text{Tinggi} | Y = Y_a) = 0$$

$$P(X_2 = \text{Tinggi} | Y = \text{Tidak}) = 1 \quad P(X_2 = \text{Tinggi} | Y = \text{Tidak}) = 1$$

$$P(X_3 = \text{Pelan} | Y = Y_a) = \frac{3}{4} \quad P(X_3 = \text{Kencang} | Y = Y_a) = \frac{1}{4}$$

$$P(X_3 = \text{Pelan} | Y = \text{Tidak}) = \frac{1}{2} \quad P(X_3 = \text{Kencang} | Y = \text{Tidak}) = \frac{1}{2}$$

$$P(Y = Y_a | X_1 = \text{cerah}, X_2 = \text{tinggi}, X_3 = \text{kencang})$$

$$= \frac{P(X_1 = \text{cerah} | Y = Y_a) \cdot P(X_2 = \text{tinggi} | Y = Y_a) \cdot P(X_3 = \text{kencang} | Y = Y_a)}{P(Y = Y_a)}$$

$$= \frac{1 \cdot 0 \cdot \frac{1}{4} \cdot \frac{4}{6}}{1}$$

$$= 0$$

$$P(Y = \text{Tidak} | X_1 = \text{cerah}, X_2 = \text{tinggi}, X_3 = \text{kencang})$$

$$= \frac{P(X_1 = \text{cerah} | Y = \text{Tidak}) \cdot P(X_2 = \text{tinggi} | Y = \text{Tidak}) \cdot P(X_3 = \text{kencang} | Y = \text{Tidak})}{P(Y = \text{Tidak})}$$

$$= \frac{0 \cdot 1 \cdot \frac{1}{2} \cdot \frac{2}{6}}{1}$$

$$= 0$$

Kesimpulan:

Berdasarkan nilai tersebut, tidak dapat disimpulkan apakah seseorang akan berolahraga atau tidak jika cuaca cerah, temperatur tinggi, dan kecepatan angin kencang.

		Humidity									Mean	StDev
Play Golf	yes	86	96	80	65	70	80	70	90	75	79.1	10.2
	no	85	90	70	95	91					86.2	9.7

Question 8. Tentukan Likelihood untuk Keputusan Bermain Golf atau tidak, untuk Humidity misalkan nilainya adalah 74?

(8).

$P(\text{humidity} \rightarrow \text{bermain golf} = \text{yes})$

$$= \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(\pi_i - \mu_y)^2}{2\sigma_y^2}\right)$$

$$= \frac{1}{\sqrt{2\pi \cdot (10.2)^2}} \exp\left(-\frac{(74 - 79.1)^2}{2(10.2)^2}\right)$$

$$= \frac{1}{10.2\sqrt{2\pi}} \cdot \exp\left(-\frac{(-5.1)^2}{208.08}\right)$$

$$= \frac{1}{10.2\sqrt{2\pi}} \exp(0.0245)$$

$$= (0.039121906) \exp(0.0245)$$

$$= 0.040092$$
  

$P(\text{humidity} = 74 \mid \text{bermain golf} = \text{no})$

$$= \frac{1}{\sqrt{2\pi(9.7)^2}} \exp\left(-\frac{(74 - 86.2)^2}{2(9.7)^2}\right)$$

$$= 1.692376116 \times 10^{-3} \exp\left(-\frac{12.2}{18.418}\right)$$

$$= 0.0018057$$

Kesimpulan:

Berdasarkan hasil perhitungan tersebut nilai likelihood untuk bermain golf saat humidity 74 lebih besar dibandingkan nilai likelihood untuk tidak bermain golf saat humidity 74. Maka kesimpulannya adalah seseorang akan bermain golf saat humidity = 74.



**Question 9. Dari data set di atas, buat potongan program untuk split dataset menjadi data\_training dan data\_testing, dengan proporsi training adalah 80%, testing adalah 20%**

```
# mengambil dataset dari url
url = https://vincentarelbundock.github.io/Rdatasets/csv/MASS/Pima.te.csv
#membuat dataframe berdasarkan URL
df <- read.table(url, sep=',', header=TRUE)
# split data
# train : test
####----- 80:20 -----####
set.seed(1234)
ind <- sample(2, nrow(df), replace=TRUE, prob = c(0.8, 0.2))
train_df <- df[ind==1,]
test_df <- df[ind==2,]
View(train_df)
View(test_df)
```

Output data train

	X	npreg	glu	bp	skin	bmi	ped	age	type
1	1	6	148	72	35	33.6	0.627	50	Yes
2	2	1	85	66	29	26.6	0.351	31	No
3	3	1	89	66	23	28.1	0.167	21	No
4	4	3	78	50	32	31.0	0.248	26	Yes
6	6	5	166	72	19	25.8	0.587	51	Yes
7	7	0	118	84	47	45.8	0.551	31	Yes
8	8	1	103	30	38	43.3	0.183	33	No
9	9	3	126	88	41	39.3	0.704	27	No
10	10	9	119	80	35	29.0	0.263	29	Yes
11	11	1	97	66	15	23.2	0.487	22	No
12	12	5	109	75	26	36.0	0.546	60	No

Output data test

	X	npreg	glu	bp	skin	bmi	ped	age	type
5	5	2	197	70	45	30.5	0.158	53	Yes
14	14	10	122	78	31	27.6	0.512	45	No
16	16	9	102	76	37	32.9	0.665	46	Yes
26	26	5	88	66	21	24.4	0.342	30	No
28	28	1	73	50	10	23.0	0.248	21	No
29	29	0	105	64	41	41.5	0.173	22	No
39	39	7	83	78	26	29.3	0.767	36	No
40	40	0	101	65	28	24.6	0.237	22	No
60	60	3	170	64	37	34.5	0.356	30	Yes
61	61	8	84	74	31	38.3	0.457	39	No
72	72	17	163	72	41	40.9	0.817	47	Yes

**Question 10. Apa fungsi dari cross validation? Menggunakan data set di atas, buat potongan program untuk split dataset menjadi data\_training dan data\_testing, menggunakan k-cross validation dengan k = 8.**

```
library(cvTools)
set.seed(1234)
F2<-cvFolds(nrow(df), K = 8, type = "random")
F2

for(i in 1:8){
  trainIdx <- which(F2[["which"]]==i, arr.ind=TRUE)
  trainData <- df[-trainIdx, ]
  testData <- df[trainIdx,]}
View(trainData)
View(testData)
```

Cross validation adalah salah satu metode untuk memisahkan data training dan data testing. Tujuan dilakukannya cross validation adalah menemukan kombinasi data terbaik sehingga waktu komputasi lebih berkurang namun tetap menjaga keakuratan estimasi.

---

## PENGENALAN POLA PHYTON

---

Link github: <https://github.com/annasnrhs/Praktikum-Pengenalan-Pola--KOM622->