

221810547_3SI2_DMKM

Raka Artian Prawardana

221810547_3SI2_25_DMKM

27/10/2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Install Packages

```
library(tidyverse)
```

```
library(e1071)
```

```
library(caret)
```

mengambil data

```
data <- read.csv("E:/UTS_DMKM/data.txt")
```

```
head(data)
```

```
##   Class Age Sex Steroid Antivirals Fatigue Malaise Anorexia LiverBig Liver
Firm
## 1     2  30   2     1         2         2         2         2         1
2
## 2     2  50   1     1         2         1         2         2         1
2
## 3     2  78   1     2         2         1         2         2         2
2
## 4     2  31   1    NA         1         2         2         2         2
2
## 5     2  34   1     2         2         2         2         2         2
2
## 6     2  34   1     2         2         2         2         2         2
2
##   SpleenPalpable Spiders Ascites Varices Bilirubin AlkPhosphate Sgot Album
in
## 1              2      2      2      2      1.0          85      18      4
.0
## 2              2      2      2      2      0.9        135      42      3
```

```
.5
## 3          2          2          2          2          0.7          96    32    4
.0
## 4          2          2          2          2          0.7          46    52    4
.0
## 5          2          2          2          2          1.0          NA   200    4
.0
## 6          2          2          2          2          0.9          95    28    4
.0
##   Prottime Histologi
## 1      NA          1
## 2      NA          1
## 3      NA          1
## 4      80          1
## 5      NA          1
## 6      75          1
```

```
str(data)
```

```
## 'data.frame':    155 obs. of  20 variables:
## $ Class          : int  2 2 2 2 2 2 1 2 2 2 ...
## $ Age            : int  30 50 78 31 34 34 51 23 39 30 ...
## $ Sex            : int  2 1 1 1 1 1 1 1 1 1 ...
## $ Steroid        : int  1 1 2 NA 2 2 1 2 2 2 ...
## $ Antivirals     : int  2 2 2 1 2 2 2 2 2 2 ...
## $ Fatigue        : int  2 1 1 2 2 2 1 2 1 2 ...
## $ Malaise        : int  2 2 2 2 2 2 2 2 2 2 ...
## $ Anorexia       : int  2 2 2 2 2 2 1 2 2 2 ...
## $ LiverBig       : int  1 1 2 2 2 2 2 2 2 2 ...
## $ LiverFirm      : int  2 2 2 2 2 2 2 2 1 2 ...
## $ SpleenPalpable : int  2 2 2 2 2 2 1 2 2 2 ...
## $ Spiders        : int  2 2 2 2 2 2 1 2 2 2 ...
## $ Ascites        : int  2 2 2 2 2 2 2 2 2 2 ...
## $ Varices        : int  2 2 2 2 2 2 2 2 2 2 ...
## $ Bilirubin      : num  1 0.9 0.7 0.7 1 0.9 NA 1 0.7 1 ...
## $ AlkPhosphate   : int  85 135 96 46 NA 95 NA NA NA NA ...
## $ Sgot           : int  18 42 32 52 200 28 NA NA 48 120 ...
## $ Albumin        : num  4 3.5 4 4 4 4 NA NA 4.4 3.9 ...
## $ Prottime       : int  NA NA NA 80 NA 75 NA NA NA NA ...
## $ Histologi      : int  1 1 1 1 1 1 1 1 1 1 ...
```

mengatasi NA

```
sum(is.na(data))
```

```
## [1] 167
```

```
data <- na.omit(data)
```

mengubah type data yang sesuai

```

data$Class <- as.factor(data$Class)
data$Age <- as.numeric(data$Age)
data$Sex <- as.factor(data$Sex)
data$Steroid <- as.factor(data$Steroid)
data$Antivirals <- as.factor(data$Antivirals)
data$Fatigue <- as.factor(data$Fatigue)
data$Malaise <- as.factor(data$Malaise)
data$Anorexia <- as.factor(data$Anorexia)
data$LiverBig <- as.factor(data$LiverBig)
data$LiverFirm <- as.factor(data$LiverFirm)
data$SpleenPalpable <- as.factor(data$SpleenPalpable)
data$Spiders <- as.factor(data$Spiders)
data$Ascites <- as.factor(data$Ascites)
data$Varices <- as.factor(data$Varices)
data$Bilirubin <- as.numeric(data$Bilirubin)
data$AlkPhosphate <- as.numeric(data$AlkPhosphate)
data$Sgot <- as.numeric(data$Sgot)
data$Albumin <- as.numeric(data$Albumin)
data$Protime <- as.numeric(data$Protime)
data$Histologi <- as.factor(data$Histologi)
str(data)

## 'data.frame':    80 obs. of  20 variables:
##  $ Class      : Factor w/ 2 levels "1","2": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Age        : num  34 39 32 41 30 38 40 38 38 22 ...
##  $ Sex        : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 2 ...
##  $ Steroid     : Factor w/ 2 levels "1","2": 2 1 2 2 2 1 1 2 1 2 ...
##  $ Antivirals  : Factor w/ 2 levels "1","2": 2 1 1 1 2 2 2 2 1 1 ...
##  $ Fatigue     : Factor w/ 2 levels "1","2": 2 2 1 1 1 1 1 2 2 1 ...
##  $ Malaise     : Factor w/ 2 levels "1","2": 2 2 2 2 2 1 2 2 2 2 ...
##  $ Anorexia    : Factor w/ 2 levels "1","2": 2 2 2 2 2 1 2 2 2 2 ...
##  $ LiverBig    : Factor w/ 2 levels "1","2": 2 1 2 2 2 2 2 2 1 2 ...
##  $ LiverFirm   : Factor w/ 2 levels "1","2": 2 1 1 1 1 2 1 2 1 2 ...
##  $ SpleenPalpable: Factor w/ 2 levels "1","2": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Spiders     : Factor w/ 2 levels "1","2": 2 2 1 2 2 2 2 2 2 2 ...
##  $ Ascites     : Factor w/ 2 levels "1","2": 2 2 2 2 2 1 2 2 2 2 ...
##  $ Varices     : Factor w/ 2 levels "1","2": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Bilirubin   : num  0.9 1.3 1 0.9 2.2 2 0.6 0.7 0.7 0.9 ...
##  $ AlkPhosphate : num  95 78 59 81 57 72 62 53 70 48 ...
##  $ Sgot        : num  28 30 249 60 144 89 166 42 28 20 ...
##  $ Albumin     : num  4 4.4 3.7 3.9 4.9 2.9 4 4.1 4.2 4.2 ...
##  $ Protime     : num  75 85 54 52 78 46 63 85 62 64 ...
##  $ Histologi    : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 2 1 1 ...
##  - attr(*, "na.action")= 'omit' Named int [1:75] 1 2 3 4 5 7 8 9 10 15 ...
##  ..- attr(*, "names")= chr [1:75] "1" "2" "3" "4" ...

```

membuat model SVM

```

modelSVM <- svm(Histologi~., data=data)
summary(modelSVM)

```

```
##
## Call:
## svm(formula = Histologi ~ ., data = data)
##
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: radial
##         cost:  1
##
## Number of Support Vectors:  60
##
## ( 32 28 )
##
##
## Number of Classes:  2
##
## Levels:
##  1 2
```

Membuat Prediksi dari model SVM yang telah dibuat

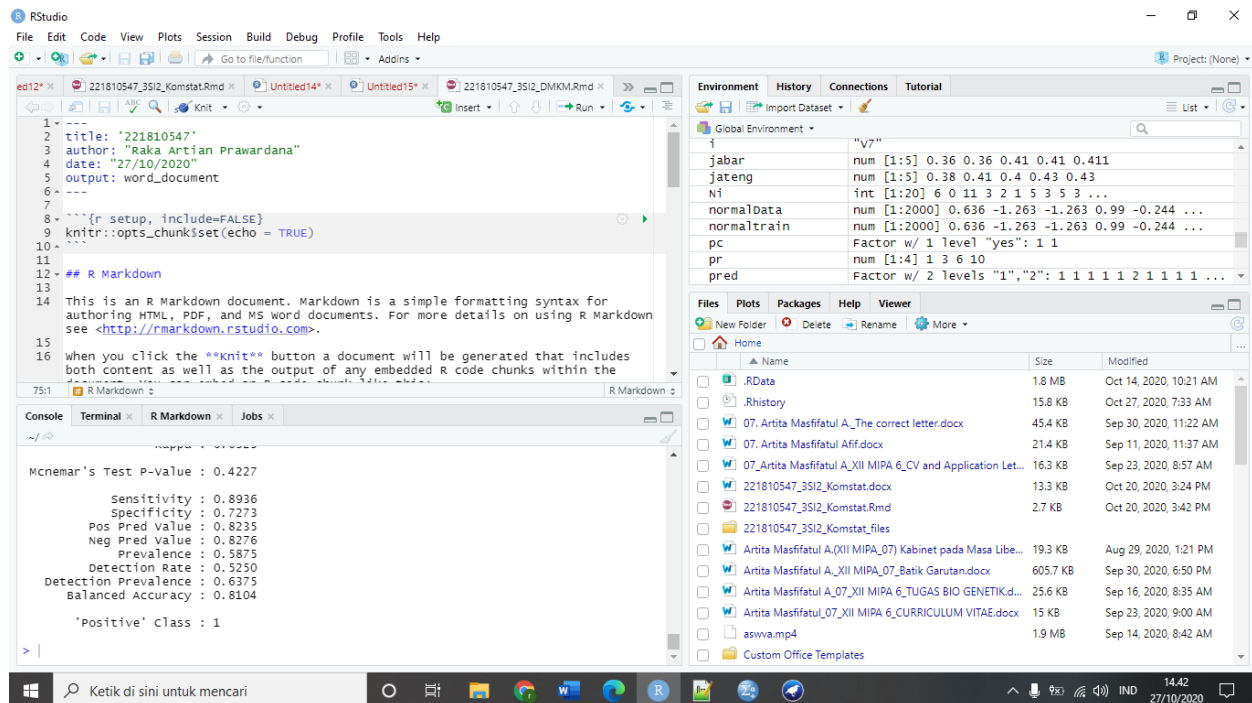
```
pred <- predict(modelSVM, data)
confusionMatrix(table(Predicted = pred, Actual=data$Histologi))

## Confusion Matrix and Statistics
##
##           Actual
## Predicted  1  2
##           1 42  9
##           2  5 24
##
##              Accuracy : 0.825
##              95% CI : (0.7238, 0.9009)
##      No Information Rate : 0.5875
##      P-Value [Acc > NIR] : 4.97e-06
##
##              Kappa : 0.6323
##
##  Mcnemar's Test P-Value : 0.4227
##
##              Sensitivity : 0.8936
##              Specificity : 0.7273
##              Pos Pred Value : 0.8235
##              Neg Pred Value : 0.8276
##              Prevalence : 0.5875
##              Detection Rate : 0.5250
##      Detection Prevalence : 0.6375
##              Balanced Accuracy : 0.8104
##
```

```
## 'Positive' Class : 1
##
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

1. Screenshot dari R



2. Output model

```
modelSVM <- svm(Histologi~., data=data)
summary(modelSVM)

##
## Call:
## svm(formula = Histologi ~ ., data = data)
##
##
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: radial
##             cost: 1
##
## Number of Support Vectors: 60
##
## ( 32 28 )
```

```
##
##
## Number of Classes: 2
##
## Levels:
## 1 2
```

Diatas merupakan model SVM. Karena adanya NA maka yang support untuk menggunakan model SVM berjumlah 60

3. Confusion matrix (Histologi [1 : No, 2 : Yes])

```
##           Actual
## Predicted 1  2
##           1 42  9
##           2  5 24
```

Diatas adalah confusion matrix dari prediksi dan actual

4. akurasi, sensitivity, specificity, precision, recall, dan F-1 score

```
##           Accuracy : 0.825
##           95% CI : (0.7238, 0.9009)
## No Information Rate : 0.5875
##
## P-Value [Acc > NIR] : 4.97e-06
##
##           Kappa : 0.6323
##
## Mcnemar's Test P-Value : 0.4227
##
##           Sensitivity : 0.8936
##           Specificity : 0.7273
##           Pos Pred Value : 0.8235
##           Neg Pred Value : 0.8276
##           Prevalence : 0.5875
##           Detection Rate : 0.5250
##           Detection Prevalence : 0.6375
##           Balanced Accuracy : 0.8104
```

5. interpretasi

		Predicted	
		No	Yes
Actual	No	True Negative (TN)	False Positive (FP)
	Yes	False Negative (FN)	True Positive (TP)

Akurasi = 0.825

Akurasi model adalah 0.825 artinya dari total prediksi, dapat diprediksi secara benar oleh model sebesar 82,5%

Sensitifity = 0.8936

banyaknya prediksi positif yang benar per total positif sebenarnya (aktual) sebesar 0.8936

Precision = 24/(24+9) = 0,727272

Precision = Precision, adalah banyaknya prediksi positif yang benar per total prediksi positif sebesar 0,727272

Recall = banyaknya prediksi positif yang benar per total positif sebenarnya (aktual).

Neg Pred Value : 0.8276

Interpretasi : total prediksi bahwa tidak sakit hepatitis dibagi total tidak sakit hepatitis sebenarnya

Pos Pred Value : 0.8235

Interpretasi : total prediksi bahwa sakit hepatitis dibagi total sakit hepatitis sebenarnya

F1 = rata-rata harmonic dari precision dan recall

$$F\text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$$

F1 = 2/((1/pres)+(1/recall)) = 1,287

rata-rata harmonic dari precision dan recall sebesar 1,287