

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

Tugas Besar Data Mining Klasifikasi C5.0

Tujuan :

- ✓ Menerapkan algoritma C5.0 alat:
- ✓ R dan R Studio

Deskripsi

Pada tugas besar ini kelompok kali menggunakan metode klasifikasi C5.0 dan memakai dataset mammographic_masses Dataset yang memiliki kumpulan data memberikan informasi yang berkaitan dengan skrining kanker payudara.

Dataset Mammographic mass.csv Terdiri dari :

-961 data

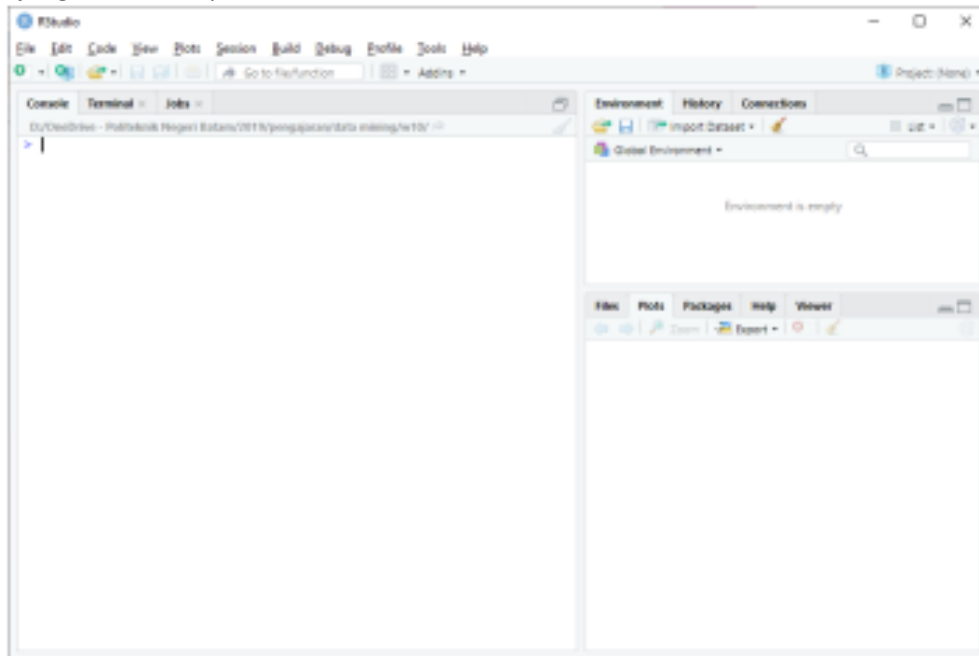
-5 variable

1	BI_RADS,age,shape,margin,density,severity			
2	5,67,3,5,3,1			
3	4,43,1,1,?,1			
4	5,58,4,5,3,1			
5	4,28,1,1,3,0			
6	5,74,1,5,?,1			
7	4,65,1,?,3,0			
8	4,70,?,?,3,0			
9	5,42,1,?,3,0			
10	5,57,1,5,3,1			
11	5,60,?,5,1,1			
12	5,76,1,4,3,1			
13	3,42,2,1,3,1			
14	4,64,1,?,3,0			
15	4,36,3,1,2,0			
16	4,60,2,1,2,0			
17	4,54,1,1,3,0			
18	3,52,3,4,3,0			
19	4,59,2,1,3,1			
20	4,54,1,1,3,1			
21	4,40,1,?,?,0			
22	?,66,?,?,1,1			
23	5,56,4,3,1,1			

1. Buka R studio

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa



2. Buat script baru
3. Pengaturan untuk mengatur working directory

#set and get location

Getwd()

Setwd("D:/UAD_DM")

```
#set and get location  
getwd()  
setwd("D:/UAS_DM")
```

4. Instalasi dan penggunaan library

#Importing library packages

install.packages("C50")

install.packages("printr")

```
#importing library packages  
install.packages("C50")  
install.packages("printr")
```

library(C50)

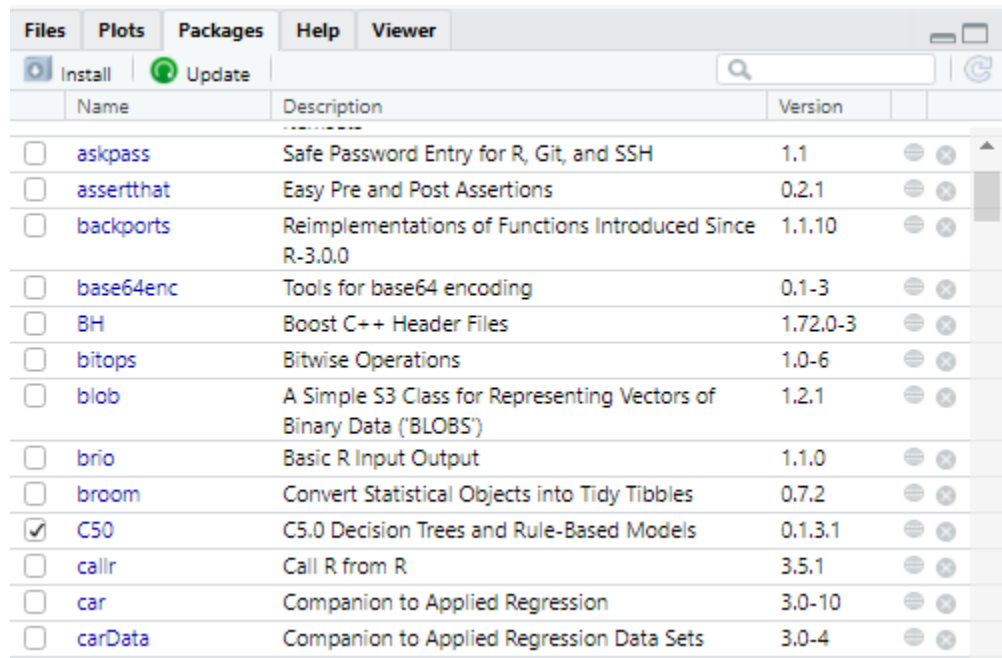
library(printr)

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

```
library(C50)
library(printr)
```

Melihat package yang sudah terinstall



Files	Plots	Packages	Help	Viewer
Install Update		Search		
	Name	Description	Version	
<input type="checkbox"/>	askpass	Safe Password Entry for R, Git, and SSH	1.1	
<input type="checkbox"/>	assertthat	Easy Pre and Post Assertions	0.2.1	
<input type="checkbox"/>	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.10	
<input type="checkbox"/>	base64enc	Tools for base64 encoding	0.1-3	
<input type="checkbox"/>	BH	Boost C++ Header Files	1.72.0-3	
<input type="checkbox"/>	bitops	Bitwise Operations	1.0-6	
<input type="checkbox"/>	blob	A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')	1.2.1	
<input type="checkbox"/>	brio	Basic R Input Output	1.1.0	
<input type="checkbox"/>	broom	Convert Statistical Objects into Tidy Tibbles	0.7.2	
<input checked="" type="checkbox"/>	C50	C5.0 Decision Trees and Rule-Based Models	0.1.3.1	
<input type="checkbox"/>	callr	Call R from R	3.5.1	
<input type="checkbox"/>	car	Companion to Applied Regression	3.0-10	
<input type="checkbox"/>	carData	Companion to Applied Regression Data Sets	3.0-4	

5. Membaca dataset dari file .csv

#Reading the file

```
datasets <- read.csv("mass.csv",na.strings = "?", sep = ",")
```

```
#reading the file
datasets <- read.csv("mass.csv",na.strings = "?", sep = ",")
```

#melihat struktur pada sebuah object

str(datasets)

```
> #melihat struktur pada sebuah object
> str(datasets)
'data.frame': 961 obs. of 6 variables:
 $ BI_RADS : int 5 4 5 4 5 4 4 5 5 5 ...
 $ age     : int 67 43 58 28 74 65 70 42 57 60 ...
 $ shape   : int 3 1 4 1 1 1 NA 1 1 NA ...
 $ margin  : int 5 1 5 1 5 NA NA NA 5 5 ...
 $ density : int 3 NA 3 3 NA 3 3 3 3 1 ...
 $ severity: int 1 1 1 0 1 0 0 0 1 1 ...
```

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

#menghilangkan baris yang memiliki NA

datasets <- na.omit(datasets)

str(datasets)

```
#menghilangkan baris yang memiliki NA
datasets <- na.omit(datasets)
str(datasets)
```

```
> datasets <- na.omit(datasets)
> str(datasets)
'data.frame': 830 obs. of 6 variables:
 $ BI_RADS : int 5 5 4 5 5 3 4 4 4 3 ...
 $ age      : int 67 58 28 57 76 42 36 60 54 52 ...
 $ shape    : int 3 4 1 1 1 2 3 2 1 3 ...
 $ margin   : int 5 5 1 5 4 1 1 1 1 4 ...
 $ density  : int 3 3 3 3 3 3 2 2 3 3 ...
 $ severity : int 1 1 0 1 1 1 0 0 0 0 ...
 - attr(*, "na.action")= 'omit' Named int [1:131] 2 5 6 7 8 10 13 20 21 23 ...
 ..- attr(*, "names")= chr [1:131] "2" "5" "6" "7" ...
```

6. Pembuatan model

#pembuatan model decision tree menggunakan algoritma C5.0

datasets\$severity <- as.factor(datasets\$severity)

model <- C5.0(severity ~., data=datasets)

```
#pembuatan model
datasets$severity <- as.factor(datasets$severity)
model <- C5.0(severity ~., data=datasets)
```

7. Melihat model

#melihat hasil model

model

```
#melihat hasil model
model
summary(model)
```

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

```
> model
```

```
Call:
```

```
C5.0.formula(formula = severity ~ ., data = datasets)
```

```
Classification Tree
```

```
Number of samples: 830
```

```
Number of predictors: 5
```

```
Tree size: 4
```

```
Non-standard options: attempt to group attributes
```

summary(model)

```
> summary(model)
```

```
Call:
```

```
C5.0.formula(formula = severity ~ ., data = datasets)
```

```
C5.0 [Release 2.07 GPL Edition]
```

```
Fri Jan 15 16:33:06 2021
```

```
-----  
Class specified by attribute 'outcome'
```

```
Read 830 cases (6 attributes) from undefined.data
```

```
Decision tree:
```

```
BI_RADS > 4: 1 (326/33)
```

```
BI_RADS <= 4:
```

```
  :...shape <= 2: 0 (334/33)
```

```
    shape > 2:
```

```
      :...age <= 69: 0 (143/55)
```

```
        age > 69: 1 (27/5)
```

```
Evaluation on training data (830 cases):
```

```
Decision Tree
```

```
-----  
Size      Errors
```

```
4  126(15.2%)  <<
```

```
(a)  (b)  <-classified as
```

```
-----
```

```
389   38  (a): class 0
```

```
88   315  (b): class 1
```

```
Attribute usage:
```

```
100.00% BI_RADS
```

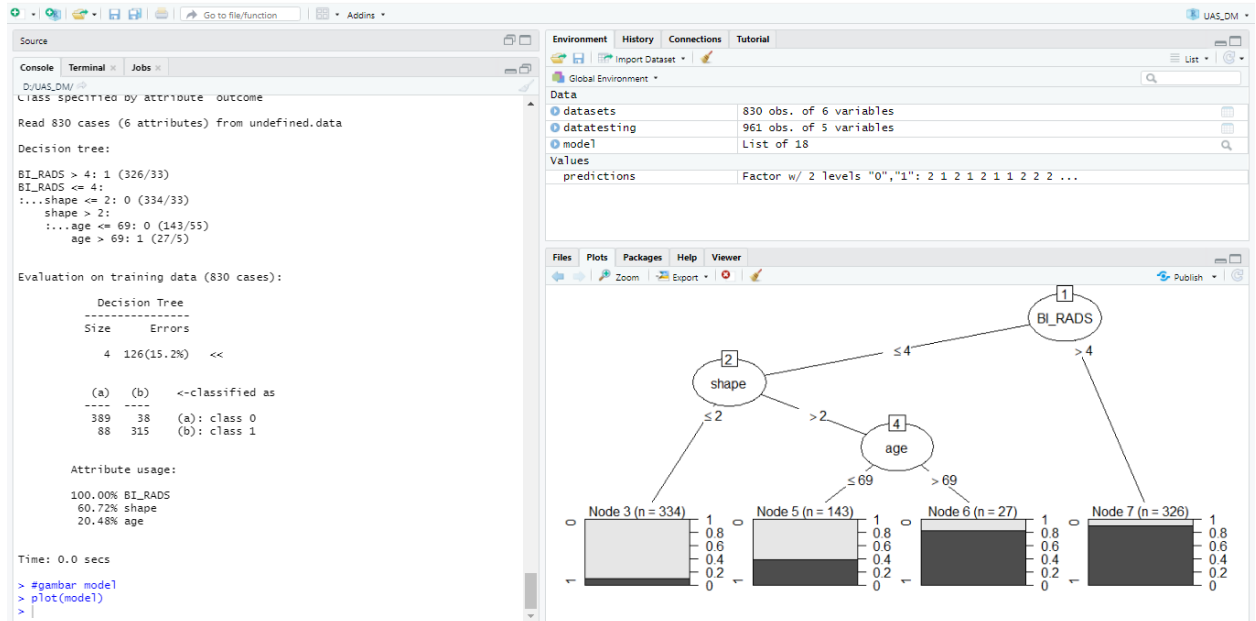
8. Gambar

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

#gambar model
plot(model)

```
#gambar model  
plot(model)
```



9. Membuat dataset

#membuat dataset
datatesting <- datasets[,1:5]

```
#membuat dataset  
datatesting <- datasets[,1:5]
```

```
> #membuat dataset  
> datatesting <- datasets[,1:5]
```

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa
10. Pridiksi

```
#prediksi  
predictions <- predict(model, datatesting)
```

```
#prediksi  
predictions <- predict(model, datatesting)  
  
> #prediksi  
> predictions <- predict(model, datatesting)
```

11. Membandingkan dataset

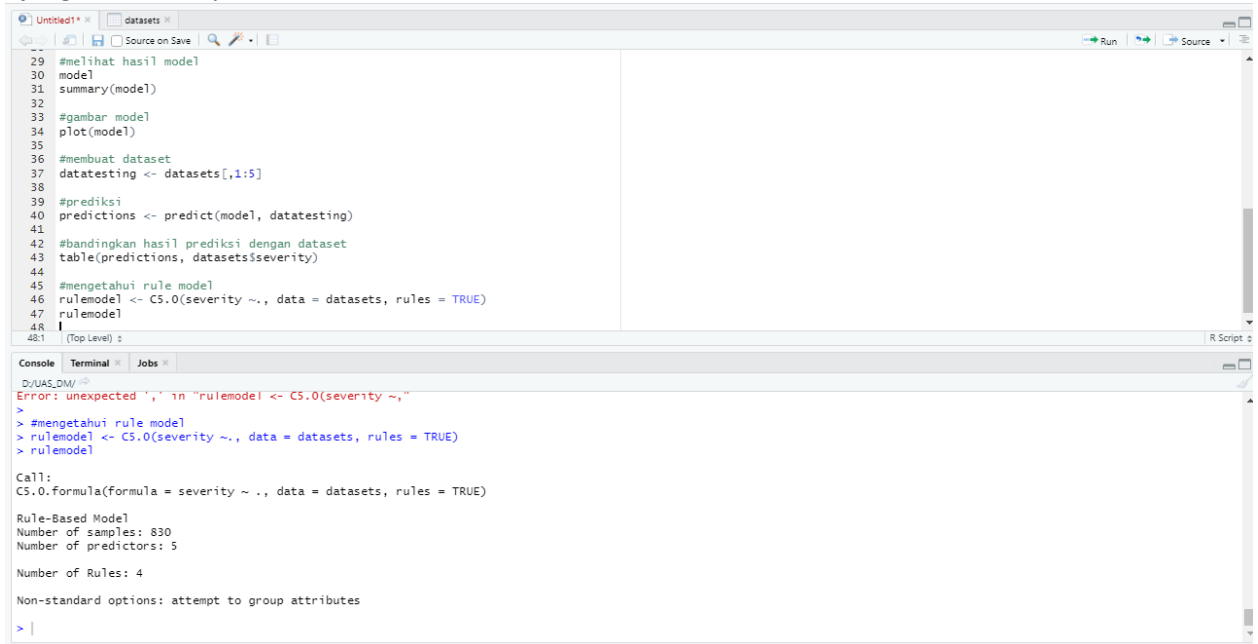
```
#bandingkan hasil prediksi dengan dataset  
table(predictions, datasets$severity)
```

```
#bandingkan hasil prediksi dengan dataset  
table(predictions, datasets$severity)  
  
> #bandingkan hasil prediksi dengan dataset  
> table(predictions, datasets$severity)  
  
predictions    0    1  
      0 389   88  
      1   38 315  
  
> |
```

12. Rules ynag dihasilkan

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa



The screenshot shows the RStudio environment. The top pane displays an R script with the following code:

```
29 #melihat hasil model
30 model
31 summary(model)
32
33 #gambar model
34 plot(model)
35
36 #membuat dataset
37 datatesting <- datasets[,1:5]
38
39 #prediksi
40 predictions <- predict(model, datatesting)
41
42 #bandingkan hasil prediksi dengan dataset
43 table(predictions, datasets$severity)
44
45 #mengetahui rule model
46 rulemodel <- C5.0(severity ~., data = datasets, rules = TRUE)
47 rulemodel
48 |
48:1 (Top Level) :
```

The bottom pane shows the console output, which includes an error message and the summary of the C5.0 model:

```
Error: unexpected ',' in "rulemodel <- C5.0(severity ~,"
>
> #mengetahui rule model
> rulemodel <- C5.0(severity ~., data = datasets, rules = TRUE)
> rulemodel

Call:
C5.0.formula(formula = severity ~ ., data = datasets, rules = TRUE)

Rule-Based Model
Number of samples: 830
Number of predictors: 5

Number of Rules: 4

Non-standard options: attempt to group attributes
> |
```

```
> summary(rulemodel)
```

```
Call:
C5.0.formula(formula = severity ~ ., data = datasets, rules = TRUE)
```

```
C5.0 [Release 2.07 GPL Edition]          Fri Jan 15 16:51:20 2021
```

```
-----
Class specified by attribute 'outcome'
```

```
Read 830 cases (6 attributes) from undefined.data
```

```
Rules:
```

```
Rule 1: (334/33, lift 1.7)
      BI_RADS <= 4
      shape <= 2
      -> class 0 [0.899]
```

```
Rule 2: (460/84, lift 1.6)
      BI_RADS <= 4
      age <= 69
      -> class 0 [0.816]
```

```
Rule 3: (115/10, lift 1.9)
      age > 69
      shape > 2
      -> class 1 [0.906]
```

```
Rule 4: (326/33, lift 1.8)
      BI_RADS > 4
      -> class 1 [0.896]
```

```
Default class: 0
```


Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

Evaluation on training data (830 cases):

Rules		
No	Errors	
4	126(15.2%)	<<

(a)	(b)	<-classified as
389	38	(a): class 0
88	315	(b): class 1

Attribute usage:

96.75% BI_RADS
69.28% age
54.10% shape

Time: 0.1 secs

Rules yang di hasilkan

✓ **Rule 1 :**
BI_RADS <= 4
Shape <= 2
Class -> 0 (benign)

✓ **Rule 2 :**
BI_RADS <= 4
age <= 69
class -> 0 (benign)

✓ **Rule 3 :**
age > 69
shape > 2
class -> 1 (malignant)

✓ **Rule 4 :**
BI_RADS > 4
-> class 1 (malignant)

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

```
#set and get location
getwd()
setwd("D:/UAS_DM")

#importing library packages
install.packages("C50")
install.packages("printr")

library(C50)
library(printr)

#reading the file
datasets <- read.csv("mass.csv", na.strings = "?", sep = ",")

#melihat struktur pada sebuah object
str(datasets)

#pembuatan model
datasets$severity <- as.factor(datasets$severity)
model <- C5.0(severity ~., data=datasets)
```

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa

```
#menghilangkan baris yang memiliki NA
datasets <- na.omit(datasets)
str(datasets)
```

```
#melihat hasil model
model
summary(model)
```

```
#gambar model
plot(model)
```

```
#membuat dataset
datatesting <- datasets[,1:5]
```

```
#prediksi
predictions <- predict(model, datatesting)
```

```
#bandingkan hasil prediksi dengan dataset
table(predictions, datasets$severity)
#mengetahui rule model
rulemodel <- C5.0(severity ~., data = datasets, rules = TRUE)
rulemodel
```

```
summary(rulemodel)
```

Nama Anggota :

1. Huznul Rachel Pratama (3311901009)
2. Umi Nasintha (3311901016)
3. Ajeng Kartika Puspa