

Nama Anggota:

1. Evriska Dayanti-3311901003
2. Ayusda Renjani-3311901019
3. Gabriella Joice Sitompul-3311901029

Kelas : IF 3A Reguler

## Tugas Besar Data Mining

### Klasifikasi C5.0

Tujuan:

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

Deskripsi:

Pada tugas besar ini kelompok kami menggunakan metode klasifikasi C5.0 dan memakai dataset Cryotherapy Dataset yang memiliki Kumpulan data memberikan informasi yang berkaitan dengan pasien, yang karakteristiknya seperti Jumlah Kutil, Area kutil, jenis kelamin dan usia, dll. Digunakan untuk menentukan tingkat ekstremitas kanker, yaitu 0 jinak atau ganas.

**Dataset: Cryotherapy.csv** Terdiri dari:

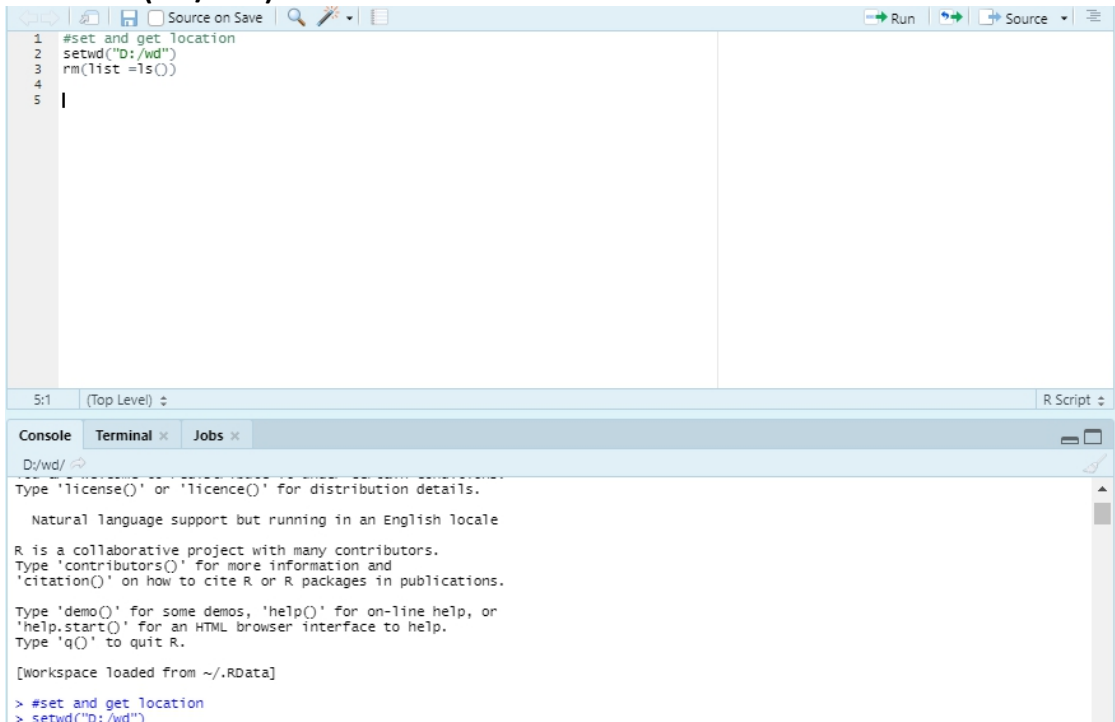
- 90 data

- 7 variabel

	A	B	C	D	E	F
1	sex,"age","Time","Number_of_Warts","Type","Area","Result_of_Treatment"					
2	1,"35","12","5","1","100","0"					
3	1,"29","7","5","1","96","1"					
4	1,"50","8","1","3","132","0"					
5	1,"32","11.75","7","3","750","0"					
6	1,"67","9.25","1","1","42","0"					
7	1,"41","8","2","2","20","1"					
8	1,"36","11","2","1","8","0"					
9	1,"59","3.5","3","3","20","0"					
10	1,"20","4.5","12","1","6","1"					
11	2,"34","11.25","3","3","150","0"					
12	2,"21","10.75","5","1","35","0"					
13	2,"15","6","2","1","30","1"					
14	2,"15","2","3","1","4","1"					
15	2,"15","3.75","2","3","70","1"					
16	2,"17","11","2","1","10","0"					
17	2,"17","5.25","3","1","63","1"					
18	2,"23","11.75","12","3","72","0"					
19	2,"27","8.75","2","1","6","0"					
20	2,"15","4.25","1","1","6","1"					
21	2,"18","5.75","1","1","80","1"					
22	1,"22","5.5","2","1","70","1"					
23	2,"16","8.5","1","2","60","1"					
24	1,"28","4.75","3","1","100","1"					
25	2,"40","9.75","1","2","80","0"					
26	1,"30","2.5","2","1","115","1"					
27	2,"34","12","3","3","95","0"					
28	1,"20","0.5","2","1","75","1"					
29	2,"35","12","5","3","100","0"					
30	2,"24","9.5","3","3","20","0"					
31	2,"19","8.75","6","1","160","1"					
32	1,"35","9.25","9","1","100","1"					
33	1,"29","7.25","6","1","96","1"					
34	1,"50","8.75","11","3","132","0"					
35	1,"37","4.75","1","1","75","0"					

## Tugas Besar Data Mining:

- Pengaturan lokasi direktori  
**setwd("D:/wd")**



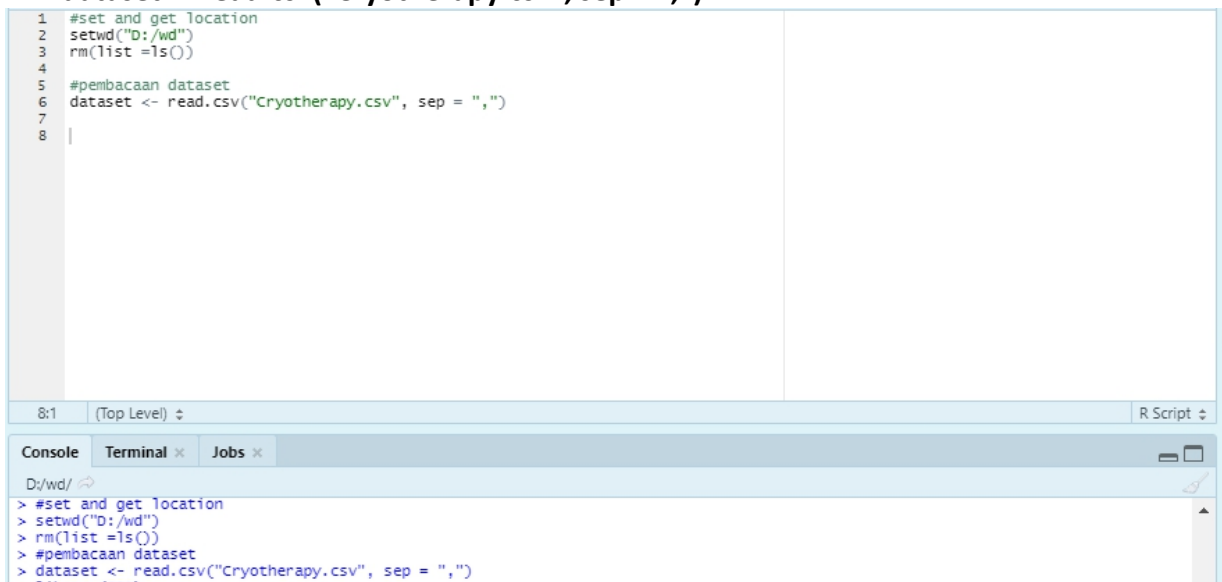
The screenshot shows the RStudio interface. The source editor on the left contains the following R code:

```
1 #set and get location
2 setwd("D:/wd")
3 rm(list = ls())
4
5 |
```

The console at the bottom displays the output of the code execution:

```
D:/wd/
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
> #set and get location
> setwd("D:/wd")
```

- Membaca dataset  
**dataset <- read.csv("Cryotherapy.csv", sep = ";")**



The screenshot shows the RStudio interface. The source editor on the left contains the following R code:

```
1 #set and get location
2 setwd("D:/wd")
3 rm(list = ls())
4
5 #membaca dataset
6 dataset <- read.csv("Cryotherapy.csv", sep = ";")
7
8 |
```

The console at the bottom displays the output of the code execution:

```
D:/wd/
> #set and get location
> setwd("D:/wd")
> rm(list = ls())
> #membaca dataset
> dataset <- read.csv("Cryotherapy.csv", sep = ";")
```

- Instalasi package  
**install.packages("C50")**  
**install.packages("printr")**

```

1 #set and get location
2 setwd("D:/wd")
3 rm(list = ls())
4
5 #pembacaan dataset
6 dataset <- read.csv("Cryotherapy.csv", sep = ",")
7
8 #library
9 install.packages("C50")
10 install.packages("printr")
11 library(C50)
12 library(printr)

```

11:1 (Top Level) R Script

Console Terminal Jobs

D:/wd/

```

> install.packages("C50")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version
of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/PERSONAL/Documents/R/win-library/4.0'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/C50_0.1.3.1.zip'
Content type 'application/zip' length 541679 bytes (528 KB)
downloaded 528 KB

package 'C50' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:/Users/PERSONAL/AppData/Local/Temp/RtmpWCM17b/downloaded_packages
> install.packages("printr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version
of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/PERSONAL/Documents/R/win-library/4.0'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/printr_0.1.zip'
Content type 'application/zip' length 37811 bytes (36 KB)
downloaded 36 KB

package 'printr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:/Users/PERSONAL/AppData/Local/Temp/RtmpWCM17b/downloaded_packages
> |

```

- Menggunakan package  
**library(C50)**  
**library(printr)**

```

C:/Users/PERSONAL/AppData/Local/Temp/RtmpWCM17b/downloaded_packages
> library(C50)
> library(printr)
> |

```

- Melihat package yang sudah terinstall

Files Plots Packages Help Viewer				
Install Update				
Name	Description	Version		
<input type="checkbox"/> broom	Vectors of Binary Data ('BLOBS') 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"/> praise	Praise Users	1.0.0	⊕	⊗
<input type="checkbox"/> prettyunits	Pretty, Human Readable Formatting of Quantities	1.1.1	⊕	⊗
<input checked="" type="checkbox"/> printr	Automatically Print R Objects to Appropriate Formats According to the 'knitr' Output Format	0.1	⊕	⊗

- Pembuatan model decision tree menggunakan algoritman C5.0

```
> model <- C5.0(Result_of_Treatment ~., data=dataset)
Error: C5.0 models require a factor outcome
> |
```

- Terjadi error karena outputnya bukan factor, factor adalah tipe data. Sehingga untuk melihat tipe kita ketikkan **class(dataset\$buys\_computer)**

```
> #jika model error cek class/tipe dari kolom Result_of_Treatment
> class(dataset$Result_of_Treatment)
[1] "integer"
> |
```

- Setelah di cek tipe datanya, ternyata tipe datanya adalah character. Sehingga kita harus mengonversinya ke factor dengan mengetikkan **dataset\$buys\_computer <- as.factor(dataset\$buys\_computer)**. Selanjutnya jalankan perintah untuk membuat model. Pembuatan model sudah berhasil.

```
8 #library
9 install.packages("C50")
10 install.packages("printr")
11 library(C50)
12 library(printr)
13
14 #Membuat model decision tree menggunakan C5.0
15 model <- C5.0(Result_of_Treatment ~., data=dataset)
16
17 #jika model error, lanjutkan langkah berikut
18 #cek class/tipe dari kolom Result_of_Treatment
19 class(dataset$Result_of_Treatment)
20
21 #mengubah tipe class ke factor
22 dataset$Result_of_Treatment <- as.factor(dataset$Result_of_Treatment)
23 model <- C5.0(Result_of_Treatment ~., data=dataset)
24
25
26:1 (Top Level)
R Script
```

```
D:\wd\
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/PERSONAL/Documents/R/win-library/4.0'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/printr_0.1.zip'
Content type 'application/zip' length 37811 bytes (36 KB)
downloaded 36 KB
package 'printr' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:/Users/PERSONAL/AppData/Local/Temp/RtmpWOMI7b/downloaded_packages
> library(C50)
> library(printr)
> #Membuat model decision tree menggunakan C5.0
> model <- C5.0(Result_of_Treatment ~., data=dataset)
Error: C5.0 models require a factor outcome
> #jika model error cek class/tipe dari kolom Result_of_Treatment
> class(dataset$Result_of_Treatment)
[1] "integer"
> #mengubah tipe class ke factor
> dataset$Result_of_Treatment <- as.factor(dataset$Result_of_Treatment)
> model <- C5.0(Result_of_Treatment ~., data=dataset)
> |
```

- Melihat model  
**Model**

```
> model

Call:
C5.0.formula(formula = Result_of_Treatment ~ ., data = dataset)

Classification Tree
Number of samples: 90
Number of predictors: 6

Tree size: 8

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

### summary(model)

```
> summary(model)

Call:
C5.0.formula(formula = Result_of_Treatment ~ ., data = dataset)

C5.0 [Release 2.07 GPL Edition]          Tue Jan 12 13:29:37 2021
-----

Class specified by attribute 'outcome'

Read 90 cases (7 attributes) from undefined.data

Decision tree:

Time <= 8:
:...age <= 41: 1 (39)
: age > 41: 0 (4)
Time > 8:
:...age <= 16: 1 (4)
 age > 16:
:...Type > 2: 0 (19)
   Type <= 2:
:...Area <= 10: 0 (9)
    Area > 10:
:...Area <= 20: 1 (3)
     Area > 20:
:...Area <= 96: 0 (9)
      Area > 96: 1 (3/1)

Evaluation on training data (90 cases):

      Decision Tree
      -----
      Size      Errors
      8      1( 1.1%)  <<

      (a)  (b)  <-classified as
      ----  ----
      41    1   (a): class 0
          48   (b): class 1

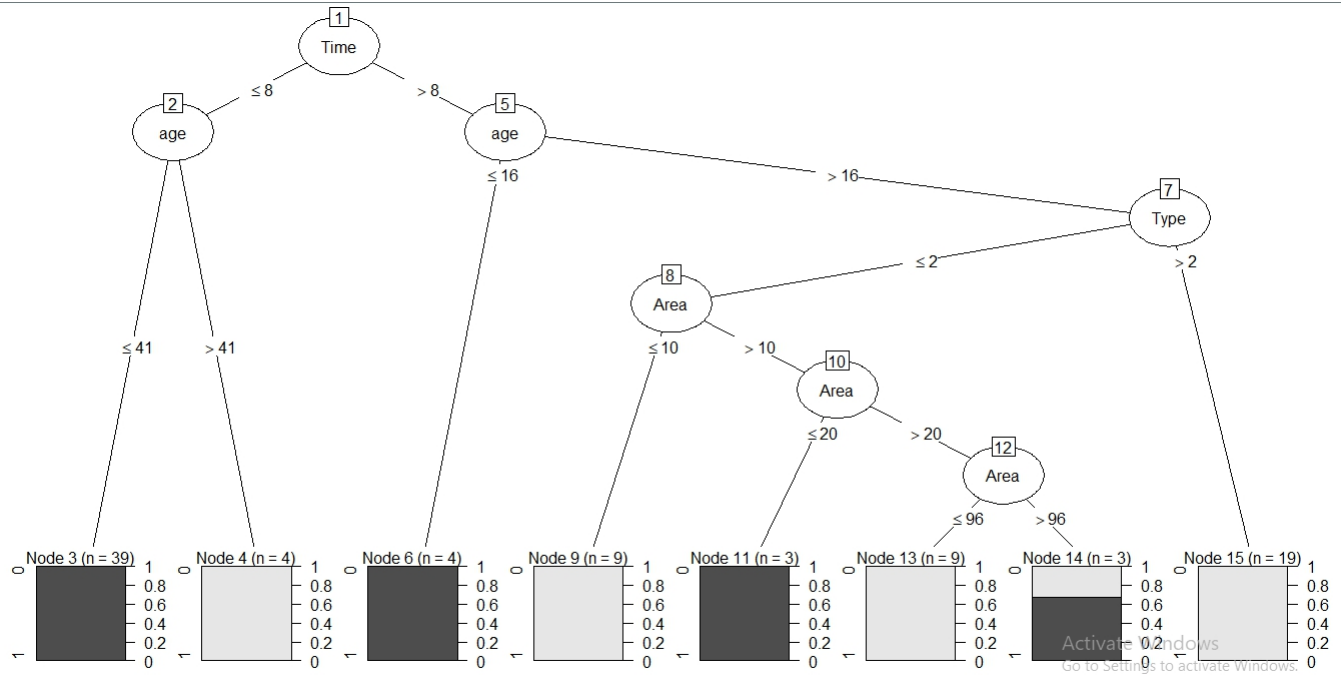
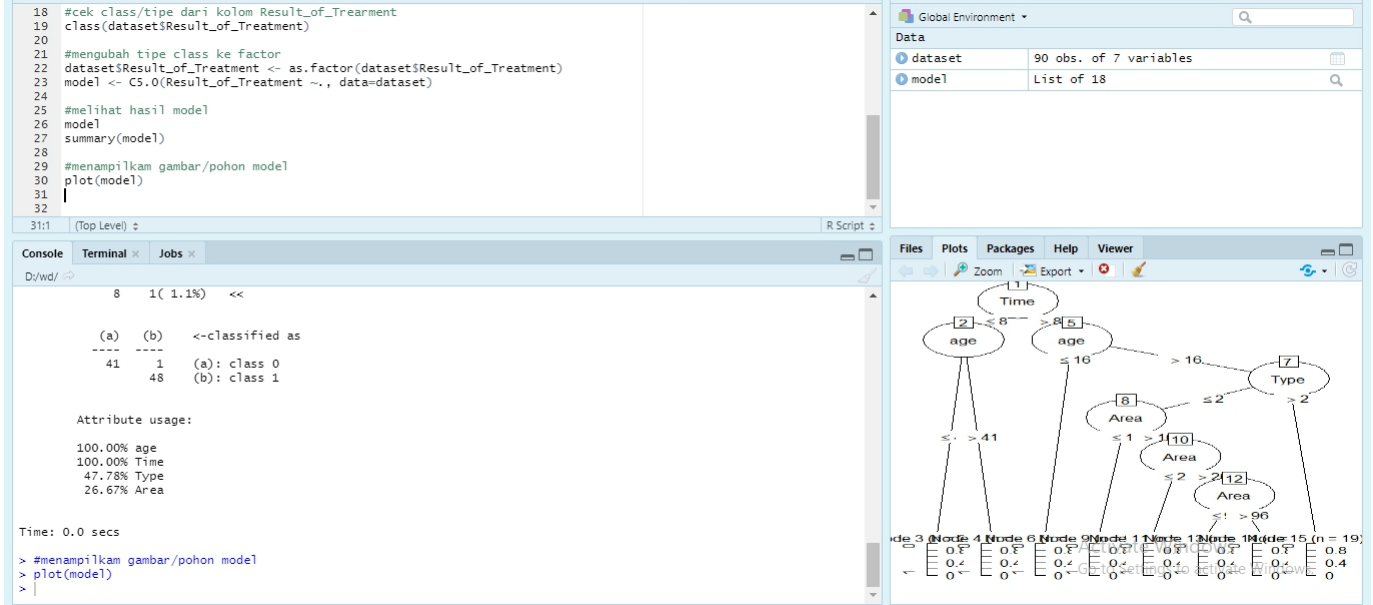
Attribute usage:

100.00% age
100.00% Time
 47.78% Type
 26.67% Area

Time: 0.0 secs

> |
```

- Menampilkan pohon yang sudah dibangun  
**plot(model)**



- Menjadikan dataset sebagai data testing. Namun hanya kolom 1, 2, 3, 4,5,6 saja dan tanpa label.  
**datatesting <- dataset[,1:6]**

	sex	age	Time	Number_of_Warts	Type	Area
1	1	35	12.00	5	1	100
2	1	29	7.00	5	1	96
3	1	50	8.00	1	3	132
4	1	32	11.75	7	3	750
5	1	67	9.25	1	1	42
6	1	41	8.00	2	2	20
7	1	36	11.00	2	1	8
8	1	59	3.50	3	3	20
9	1	20	4.50	12	1	6
10	2	34	11.25	3	3	150
11	2	21	10.75	5	1	35
12	2	15	6.00	2	1	30
13	2	15	2.00	3	1	4
14	2	15	3.75	2	3	70
15	2	17	11.00	2	1	10
16	2	17	5.25	3	1	63
17	2	23	11.75	12	3	72
18	2	27	8.75	2	1	6
19	2	15	4.25	1	1	6
20	2	18	5.75	1	1	80
21	1	22	5.50	2	1	70

**predictions <- predict(model, datatesting)**

**R Script:**

```

24 #melihat hasil model
25 model
26 summary(model)
27 #menampilkan gambar/pohon model
28 plot(model)
29 #membuat dataset
30 datatesting <- dataset[,1:6]
31 #prediksi
32 predictions <- predict(model, datatesting)
33
34
35
36
37
38

```

**Console Output:**

```

D:\wd\ 41      1      (a): class 0
        48      1      (b): class 1

Attribute usage:
100.00% age
100.00% Time
47.78% Type
26.67% Area

Time: 0.0 secs

> #menampilkan gambar/pohon model
> plot(model)
> #membuat dataset
> datatesting <- dataset[,1:6]
> View(datatesting)
> #prediksi
> predictions <- predict(model, datatesting)
>

```

**Decision Tree Plot:**

```

graph TD
    Node2((2)) -->|age ≤ 41| Node3((3))
    Node2 -->|age > 41| Node4((4))
    Node5((5)) -->|age ≤ 16| Node6((6))
    Node5 -->|age > 16| Node7((7))
    Node8((8)) -->|Area ≤ 1| Node9((9))
    Node8 -->|Area > 1| Node10((10))
    Node10 -->|Area ≤ 2| Node11((11))
    Node10 -->|Area > 2| Node12((12))
    Node12 -->|Area ≤ 96| Node13((13))
    Node12 -->|Area > 96| Node14((14))
    Node15((15)) -->|Type = 2| Node16((16))
    Node15 -->|Type ≠ 2| Node17((17))

```

**Environment Panel:**

Object	Class	Attributes
dataset	data.frame	90 obs. of 7 variables
datatesting	data.frame	90 obs. of 6 variables
model	model	List of 18
predictions	Factor w/ 2 levels "0","1"	2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1

- Membandingkan hasil prediksi dengan dataset  
**table(predictions, dataset\$buys\_computer)**

```
> table(predictions, dataset$Result_of_Treatment)
```

```
predictions  0  1
            0 41  0
            1  1 48
```

```
> |
```

- rules yang dihasilkan

```
> rulemodel <- C5.0(Result_of_Treatment ~ ., data = dataset, rules = TRUE)
> rulemodel
```

```
Call:
C5.0.formula(formula = Result_of_Treatment ~ ., data = dataset, rules = TRUE)
```

```
Rule-Based Model
Number of samples: 90
Number of predictors: 6
```

```
Number of Rules: 8
```

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

```
> |
```

```
> summary(rulemodel)
```

```
Call:
C5.0.formula(formula = Result_of_Treatment ~ ., data = dataset, rules = TRUE)
```

```
C5.0 [Release 2.07 GPL Edition] Tue Jan 12 13:40:59 2021
```

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

```
Read 90 cases (7 attributes) from undefined.data
```

```
Rules:
```

```
Rule 1: (19, lift 2.0)
      Time > 8
      Type > 2
      -> class 0 [0.952]
```

```
Rule 2: (16, lift 2.0)
      age > 16
      Time > 8
      Area > 20
      Area <= 96
      -> class 0 [0.944]
```

```
Rule 3: (9, lift 1.9)
      Time > 8
      Area <= 10
      -> class 0 [0.909]
```

```
Rule 4: (9, lift 1.9)
      age > 41
      -> class 0 [0.909]
```

```
Rule 5: (39, lift 1.8)
      age <= 41
      Time <= 8
      -> class 1 [0.976]
```



```

Rule 6: (15, lift 1.8)
  age <= 16
  -> class 1 [0.941]

Rule 7: (5, lift 1.6)
  Type <= 2
  Area > 10
  Area <= 20
  -> class 1 [0.857]

Rule 8: (11/1, lift 1.6)
  Type <= 2
  Area > 96
  -> class 1 [0.846]

Default class: 1

Evaluation on training data (90 cases):

      Rules
-----
No      Errors
  8      1( 1.1%)  <<

(a)  (b)  <-classified as
----  ----
  41      1  (a): class 0
           48 (b): class 1

Attribute usage:
84.44% Time
72.22% age
45.56% Area
38.89% Type

```

### Rule yang dihasilkan :

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>➤ <b>rule 1 :</b><br/>time &gt; 8<br/>type &gt; 2<br/>class -&gt; 0 (jinak)</li> <li>➤ <b>rule 2 :</b><br/>age &gt; 16<br/>Time &gt; 8<br/>Area &gt; 20<br/>Area &lt;= 96<br/>class -&gt; 0 (jinak)</li> <li>➤ <b>rule 3 :</b><br/>Time &gt; 8<br/>Area &lt;= 10<br/>class -&gt; 0 (jinak)</li> <li>➤ <b>rule 4 :</b><br/>age &gt; 41<br/>Area &lt;= 10<br/>Class -&gt; 0 (Jinak)</li> </ul> | <ul style="list-style-type: none"> <li>➤ <b>rule 5 :</b><br/>age &lt;= 41<br/>Time &lt;= 8<br/>class -&gt; 1 (Ganas)</li> <li>➤ <b>rule 6 :</b><br/>age &lt;= 16<br/>class -&gt; 1 (Ganas)</li> <li>➤ <b>rule 7 :</b><br/>Type &lt;= 2<br/>Area &gt;<br/>Area &lt;= 20<br/>class -&gt; 1 (Ganas)</li> <li>➤ <b>rule 8 :</b><br/>Type &lt;= 2<br/>Area &gt;96<br/>class -&gt; 1 (Ganas)</li> </ul> |
|---|---|

**Kode penuh:**

```
#set and get location
setwd("D:/wd")
rm(list = ls())

#pembacaan dataset
dataset <- read.csv("Cryotherapy.csv", sep = ",")

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

#Membuat model decision tree menggunakan C5.0
model <- C5.0(Result_of_Treatment ~., data=dataset)

#jika model error cek class/tipe dari kolom Result_of_Treatment
class(dataset$Result_of_Treatment)

#mengubah tipe class ke factor
dataset$Result_of_Treatment <- as.factor(dataset$Result_of_Treatment)
model <- C5.0(Result_of_Treatment ~., data=dataset)

#melihat hasil model
model
summary(model)

#menampilkan gambar/pohon model
plot(model)

#membuat dataset
datatesting <- dataset[,1:6]

#prediksi
predictions <- predict(model, datatesting)

#membandingkan hasil prediksi dari datatesting dengan dataset
table(predictions, dataset$Result_of_Treatment)

#Mengetahui rule model
rulemodel <- C5.0(Result_of_Treatment ~., data = dataset, rules = TRUE)
rulemodel
summary(rulemodel)
```