TUGAS BESAR DATA MINING
Breast Cancer Coimbra

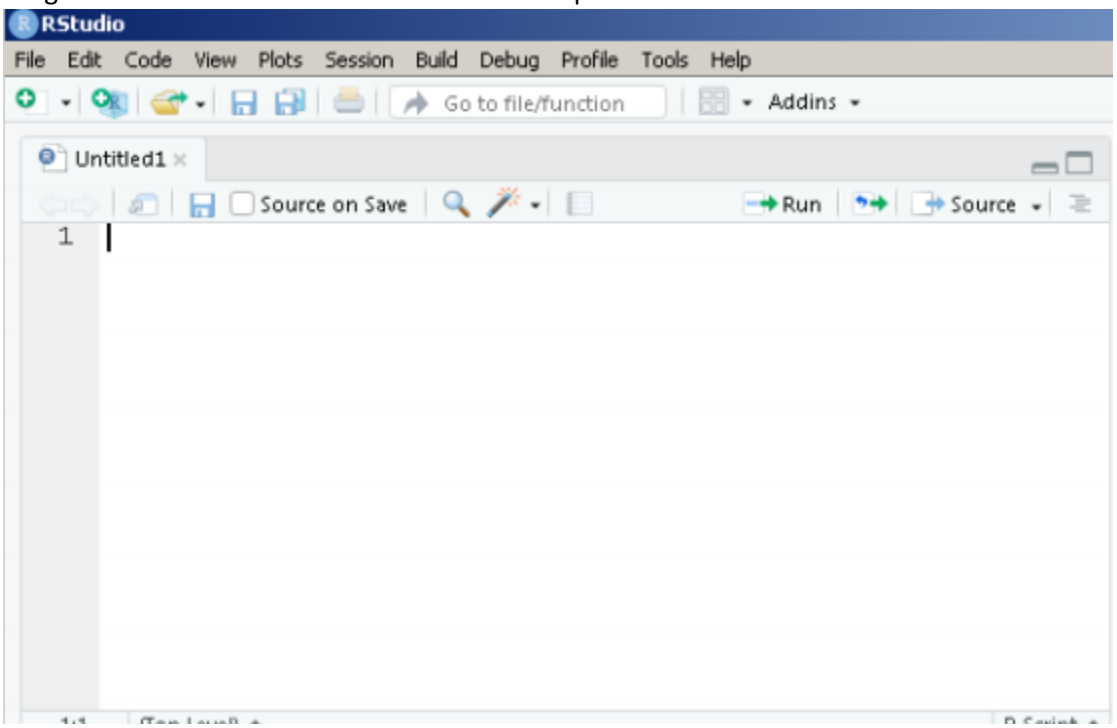
Disusun oleh:

Moh Awaludin Djaeni S	**3311801043**
Iqbal Nur Ramadhani	**3311801044**
Elpin peng sung	**3311801046**
Owen Mora Situmorang	**3311801052**

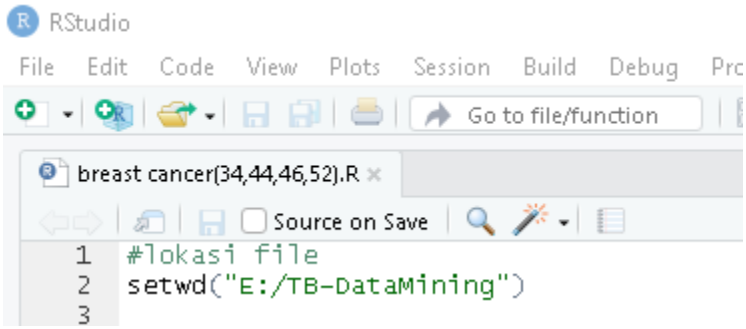
PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
POLITEKNIK NEGERI BATAM
BATAM

2019

Langkah 1 - Kita buka Rstudio dan membuat script baru

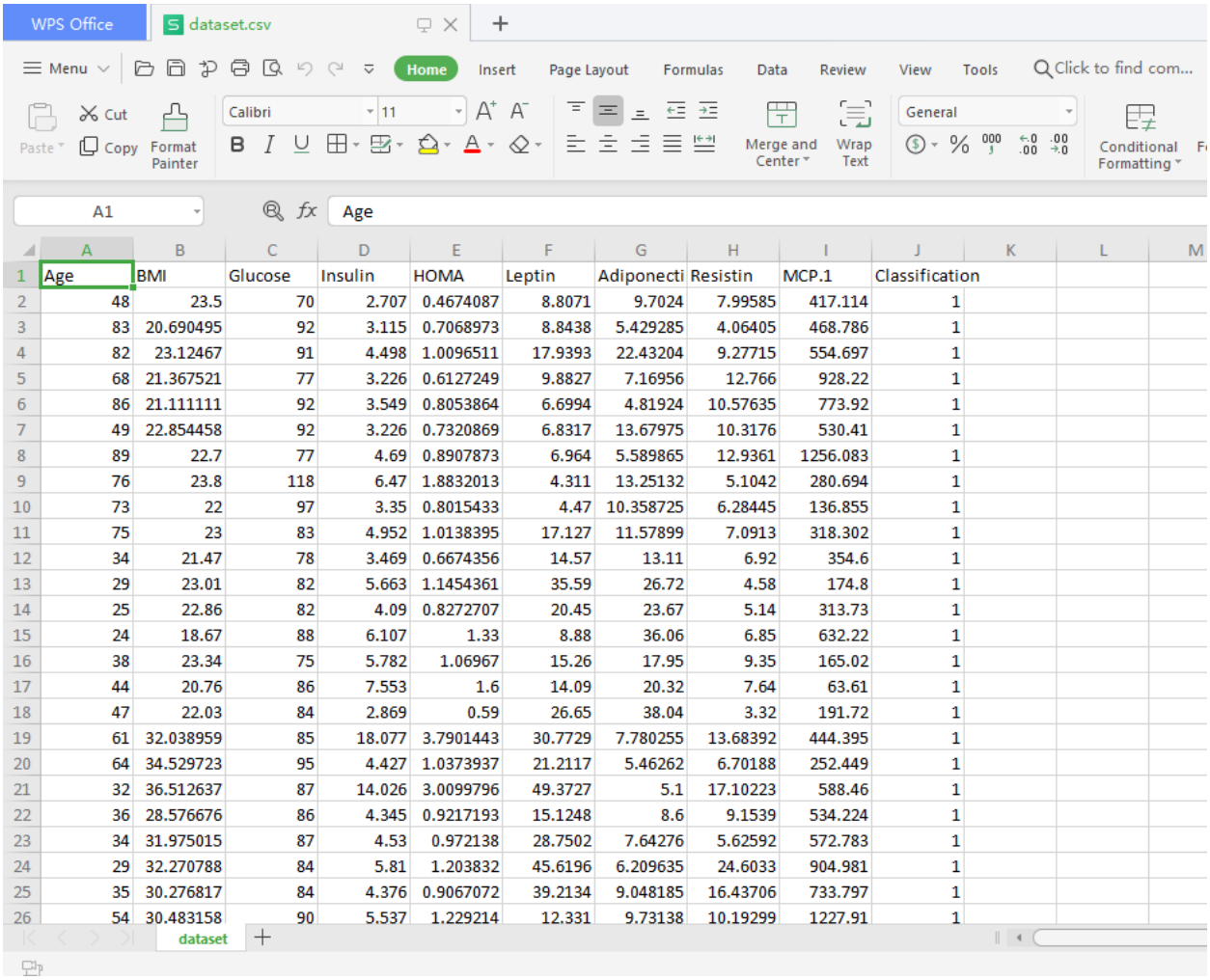


Langkah 2 - Atur lokasi directory




```
> #lokasi file
> setwd("E:/TB-DataMining")
```

Langkah 3 - Kemudian baca dataset dari sebuah file



	A	B	C	D	E	F	G	H	I	J
1	Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponecti	Resistin	MCP.1	Classification
2	48	23.5	70	2.707	0.4674087	8.8071	9.7024	7.99585	417.114	1
3	83	20.690495	92	3.115	0.7068973	8.8438	5.429285	4.06405	468.786	1
4	82	23.12467	91	4.498	1.0096511	17.9393	22.43204	9.27715	554.697	1
5	68	21.367521	77	3.226	0.6127249	9.8827	7.16956	12.766	928.22	1
6	86	21.111111	92	3.549	0.8053864	6.6994	4.81924	10.57635	773.92	1
7	49	22.854458	92	3.226	0.7320869	6.8317	13.67975	10.3176	530.41	1
8	89	22.7	77	4.69	0.8907873	6.964	5.589865	12.9361	1256.083	1
9	76	23.8	118	6.47	1.8832013	4.311	13.25132	5.1042	280.694	1
10	73	22	97	3.35	0.8015433	4.47	10.358725	6.28445	136.855	1
11	75	23	83	4.952	1.0138395	17.127	11.57899	7.0913	318.302	1
12	34	21.47	78	3.469	0.6674356	14.57	13.11	6.92	354.6	1
13	29	23.01	82	5.663	1.1454361	35.59	26.72	4.58	174.8	1
14	25	22.86	82	4.09	0.8272707	20.45	23.67	5.14	313.73	1
15	24	18.67	88	6.107	1.33	8.88	36.06	6.85	632.22	1
16	38	23.34	75	5.782	1.06967	15.26	17.95	9.35	165.02	1
17	44	20.76	86	7.553	1.6	14.09	20.32	7.64	63.61	1
18	47	22.03	84	2.869	0.59	26.65	38.04	3.32	191.72	1
19	61	32.038959	85	18.077	3.7901443	30.7729	7.780255	13.68392	444.395	1
20	64	34.529723	95	4.427	1.0373937	21.2117	5.46262	6.70188	252.449	1
21	32	36.512637	87	14.026	3.0099796	49.3727	5.1	17.10223	588.46	1
22	36	28.576676	86	4.345	0.9217193	15.1248	8.6	9.1539	534.224	1
23	34	31.975015	87	4.53	0.972138	28.7502	7.64276	5.62592	572.783	1
24	29	32.270788	84	5.81	1.203832	45.6196	6.209635	24.6033	904.981	1
25	35	30.276817	84	4.376	0.9067072	39.2134	9.048185	16.43706	733.797	1
26	54	30.483158	90	5.537	1.229214	12.331	9.73138	10.19299	1227.91	1

```
4  #pembacaan dataset
5  dataset <- read.csv("dataset.csv", sep = ",")
6
```

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

Langkah 4 - Instalasi package(c50,printr,dan recipes) dan Gunakan package tersebut

```
7  #library
8  install.packages("C50")
9  install.packages("printr")
10 install.packages("recipes")
11 library(C50)
12 library(printr)
13
```

```
> 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/ASUS/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/C50_0.1.2.zip'
Content type 'application/zip' length 604275 bytes (590 KB)
downloaded 590 KB

package 'C50' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\ASUS\AppData\Local\Temp\RtmpkTrfqX\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/ASUS/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/printr_0.1.zip'
Content type 'application/zip' length 37934 bytes (37 KB)
downloaded 37 KB

package 'printr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\ASUS\AppData\Local\Temp\RtmpkTrfqX\downloaded_packages
```

```
> install.packages("recipes")
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/ASUS/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/recipes_0.1.7.zip'
Content type 'application/zip' length 1564060 bytes (1.5 MB)
downloaded 1.5 MB

package 'recipes' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\ASUS\AppData\Local\Temp\RtmpkTrfqX\downloaded_packages
```

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

Langkah 5 - Pembuatan model menggunakan algoritma c50 lalu melihat model tersebut

```
15  #pembuatan model
16  dataset$Classification = as.factor(dataset$Classification)
17  model <- C5.0(Classification ~., data=dataset)
18
```

```
> #pembuatan model
> dataset$Classification = as.factor(dataset$Classification)
> model <- C5.0(Classification ~., data=dataset)
>
```

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

```
> #melihat hasil model
> model

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

Classification Tree
Number of samples: 116
Number of predictors: 9

Tree size: 9

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

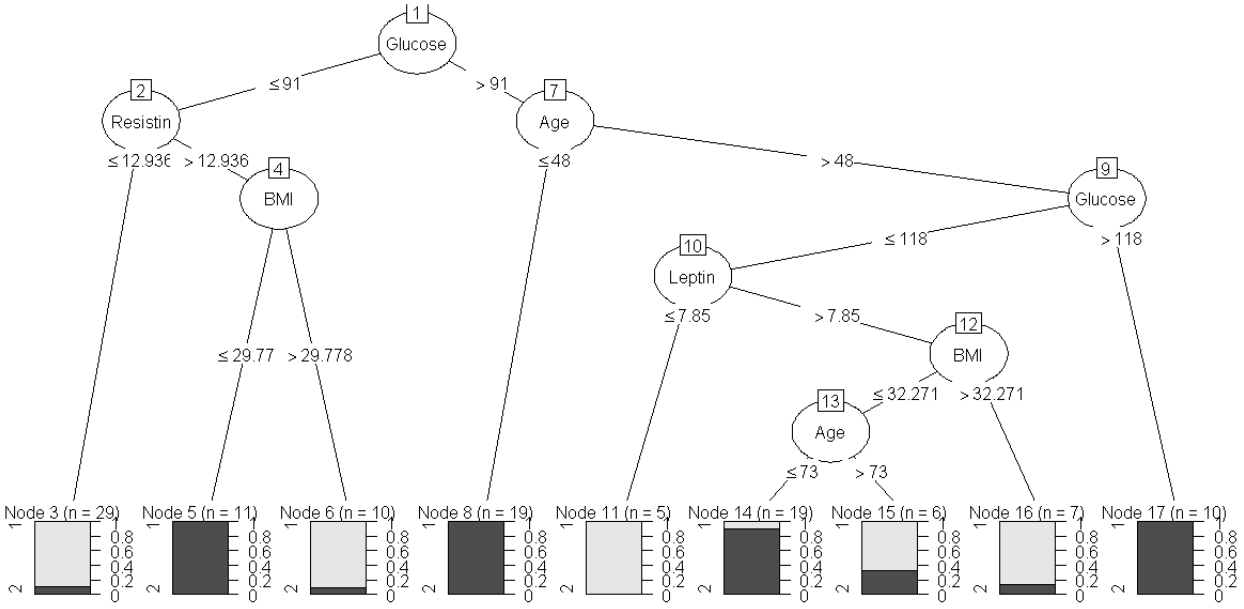
```
> summary(model)

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


C5.0 [Release 2.07 GPL Edition]         Tue Dec 03 21:32:06 2019
-------------------------------

Class specified by attribute `outcome'

Read 116 cases (10 attributes) from undefined.data

Decision tree:

Glucose <= 91:
:...Resistin <= 12.9361: 1 (29/3)
:   Resistin > 12.9361:
:   :...BMI <= 29.77778: 2 (11)
:       BMI > 29.77778: 1 (10/1)
Glucose > 91:
:...Age <= 48: 2 (19)
    Age > 48:
    :...Glucose > 118: 2 (10)
        Glucose <= 118:
        :...Leptin <= 7.85: 1 (5)
            Leptin > 7.85:
            :...BMI > 32.27079: 1 (7/1)
                BMI <= 32.27079:
                :...Age <= 73: 2 (19/2)
                    Age > 73: 1 (6/2)


Evaluation on training data (116 cases):

           Decision Tree
         ----------------
```

```
  Decision Tree
----------------
Size      Errors

   9    9( 7.8%)   <<


 (a)   (b)    <-classified as
----  ----
  50     2    (a): class 1
   7    57    (b): class 2


Attribute usage:

100.00% Glucose
 56.90% Age
 45.69% BMI
 43.10% Resistin
 31.90% Leptin


Time: 0.0 secs
```

Langkah 6 - Menampilkan pohon yang sudah dibangun

```
23  #gambar model
24  plot(model)
```

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



Langkah 7 - Menjadikan dataset, sebagai data testing lalu memprediksi dataset tersebut

```
26 #membuat dataset
27 datatesting <- dataset[,1:9]
28
29 #prediksi
30 predictions <- predict(model, datatesting)
31
```

```
> #membuat dataset
> datatesting <- dataset[,1:9]
> #prediksi
> predictions <- predict(model, datatesting)
```

Langkah 8 - Lalu Membandingkan hasil prediksi dengan dataset yang sudah di buat sebelumnya, maka hasilnya sebagai berikut :

```
32 #bandingkan hasil prediksi dengan dataset
33 table(predictions, dataset$Classification)
34
```

```
> #bandingkan hasil prediksi dengan dataset
> table(predictions, dataset$Classification)

predictions  1  2
          1 50  7
          2  2 57
>
```