Nama : Chairunnisa Fauzia Samu

NIM : 221810220

Absen : 7

Dosen : Ibnu Santoso

Makul : DMKM

Tanggal : 23 Oktober 2020

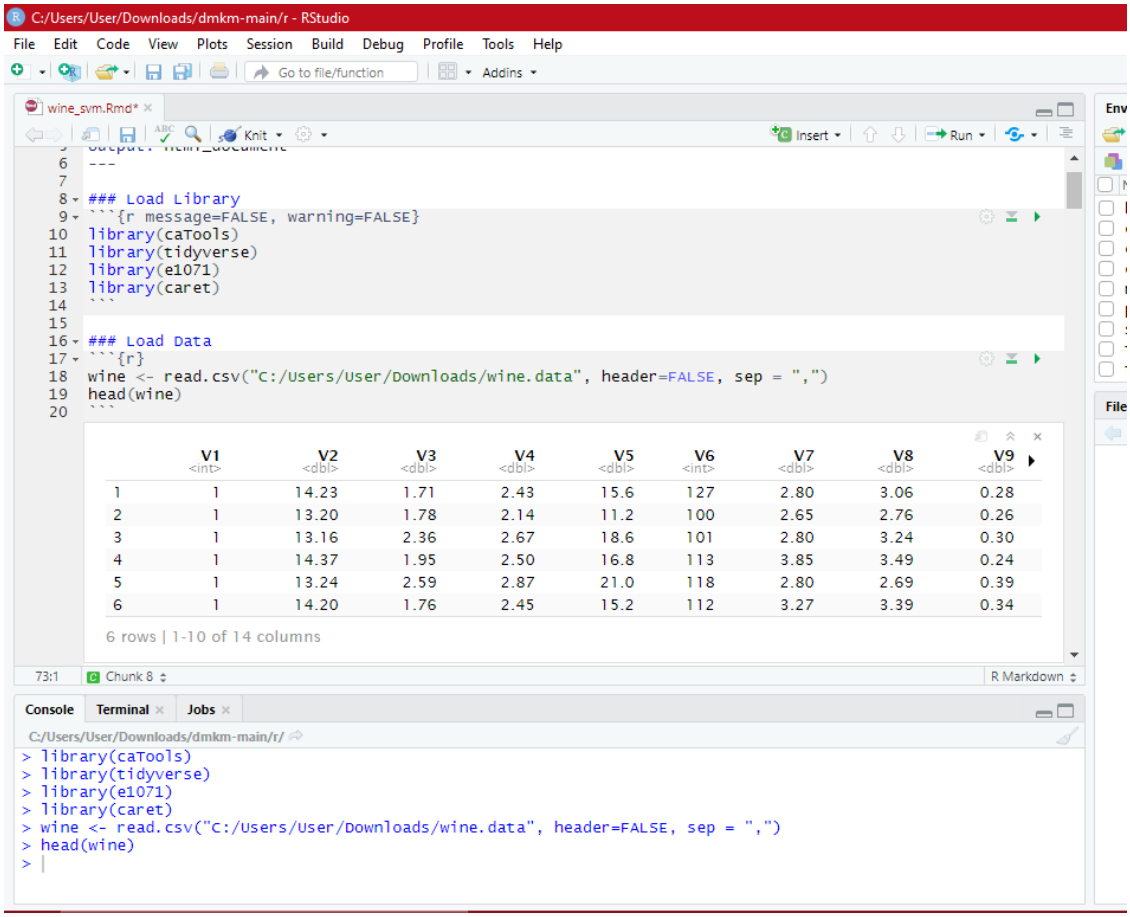
Dataset yang digunakan: https://archive.ics.uci.edu/ml/datasets/Wine

Metode: Support Vector Machine

Tools: RStudio

1. Screenshoot proses pengerjaan

Langkah pertama adalah load library dan dataset yang akan digunakan



Variabel yang akan digunakan sebagai target klasifikasi adalah V1, sehingga perlu diubah tipe datanya menjadi faktor:

C:/Users/User/Downloads/dmkm-main/r - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

wine_svm.Rmd*

5	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39
6	1	14.20	1.76	2.45	15.2	112	3.27	3.39	0.34

6 rows | 1-10 of 14 columns

```
21
22 ### Mengubah atribut target menjadi faktor
23 ```{r}
24 wine$V1 = factor(wine$V1)
25 str(wine)
26 ```
```

```
'data.frame':   178 obs. of  14 variables:
 $ V1 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
 $ V2 : num  14.2 13.2 13.2 14.4 13.2 ...
 $ V3 : num  1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ...
 $ V4 : num  2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 2.17 2.27 ...
 $ V5 : num  15.6 11.2 18.6 16.8 21 15.2 14.6 17.6 14 16 ...
 $ V6 : int  127 100 101 113 118 112 96 121 97 98 ...
 $ V7 : num  2.8 2.65 2.8 3.85 2.8 3.27 2.5 2.6 2.8 2.98 ...
 $ V8 : num  3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...
 $ V9 : num  0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...
 $ V10: num  2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...
 $ V11: num  5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...
 $ V12: num  1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...
 $ V13: num  3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...
 $ V14: int  1065 1050 1185 1480 735 1450 1290 1295 1045 1045 ...
```

26:1 Chunk 3 R Markdown

Console Terminal Jobs

C:/Users/User/Downloads/dmkm-main/r/

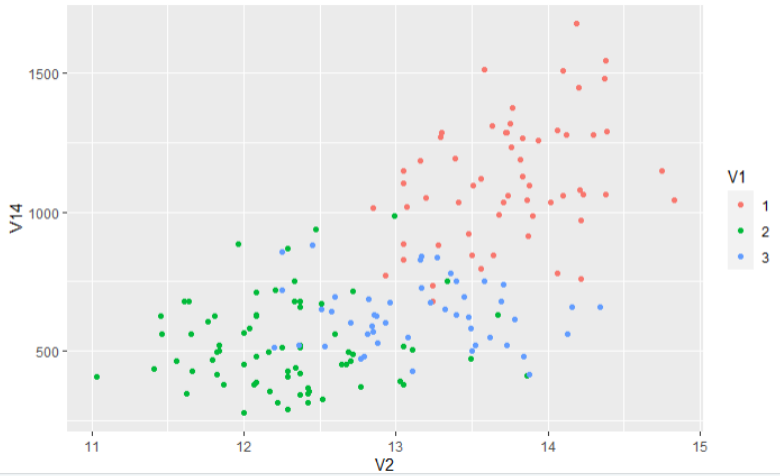
```
 $ V8 : num  3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...
 $ V9 : num  0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...
 $ V10: num  2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...
 $ V11: num  5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...
 $ V12: num  1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...
 $ V13: num  3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...
 $ V14: int  1065 1050 1185 1480 735 1450 1290 1295 1045 1045 ...
> 
```

C:/Users/User/Downloads/dmkm-main/r - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

wine_svm.Rmd*

```
27
28 ### Scatter Plot
29 Scatterplot untuk  melihat hubungan V2, V14, dan V1 dari data wine.
30 ```{r}
31 qplot(V2, V14, color = V1, data=wine)
32 ```
```



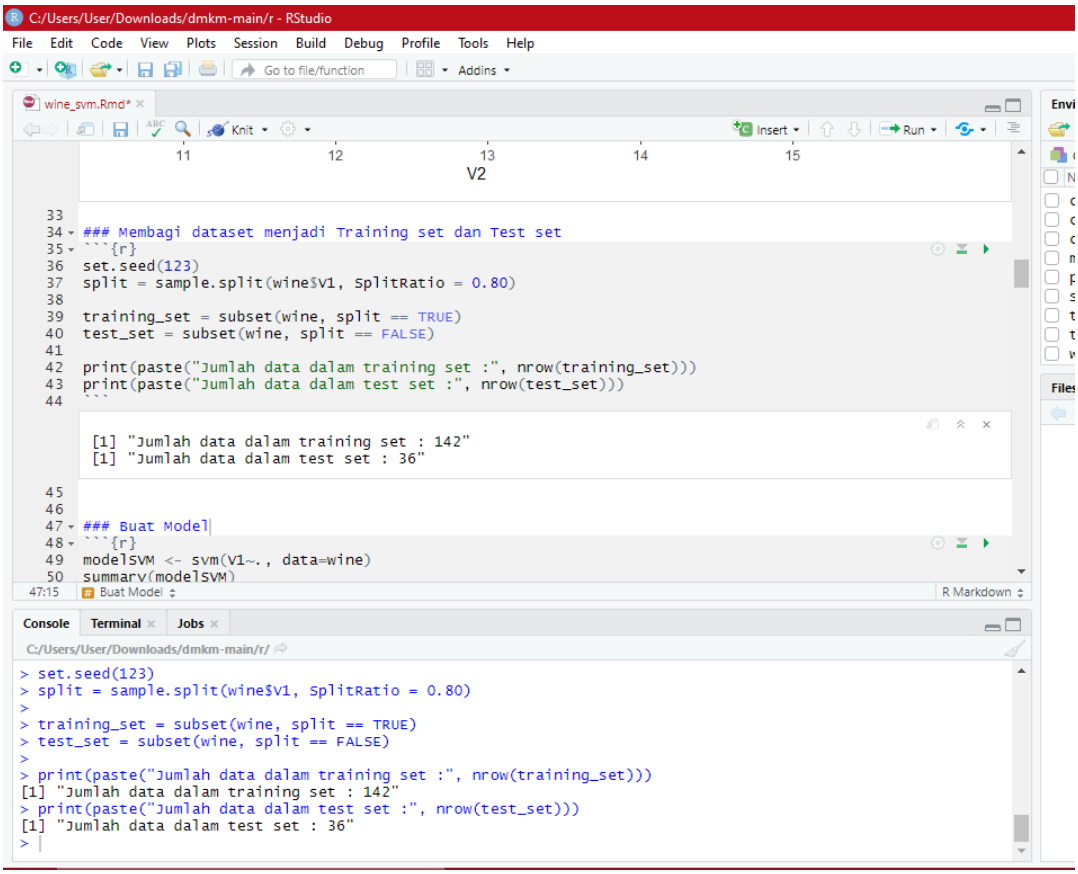
32:1 Chunk 4 R Markdown

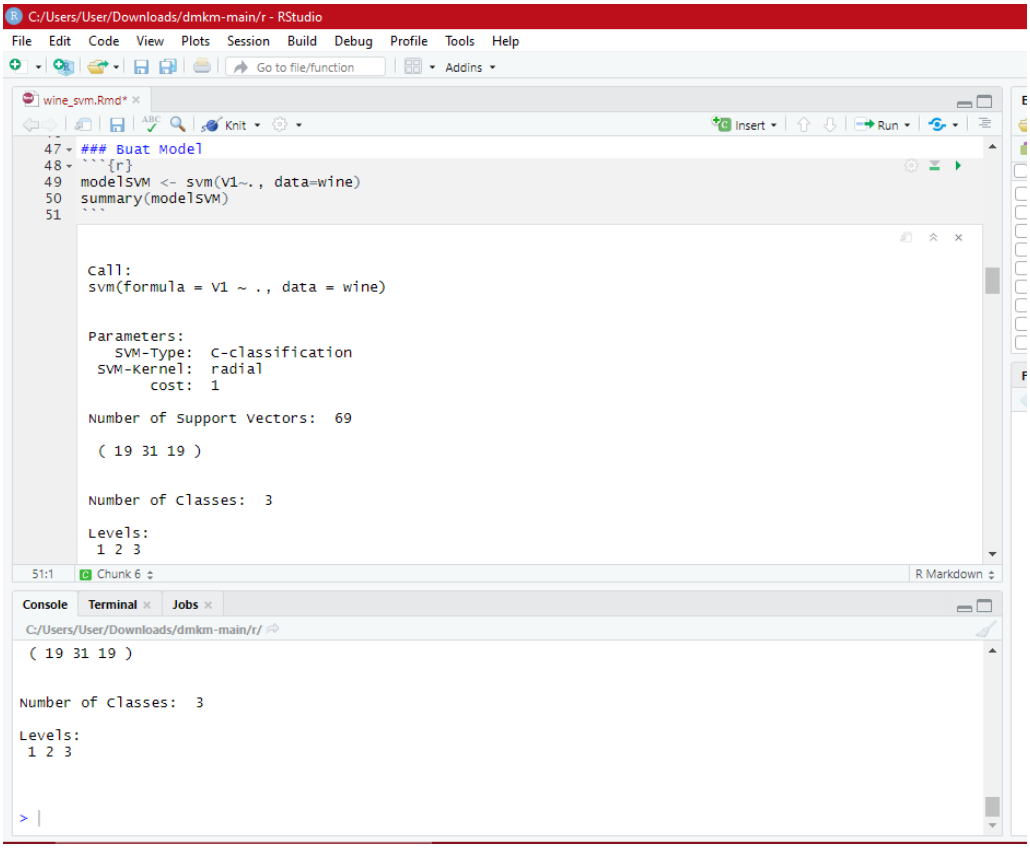
Console Terminal Jobs

C:/Users/User/Downloads/dmkm-main/r/

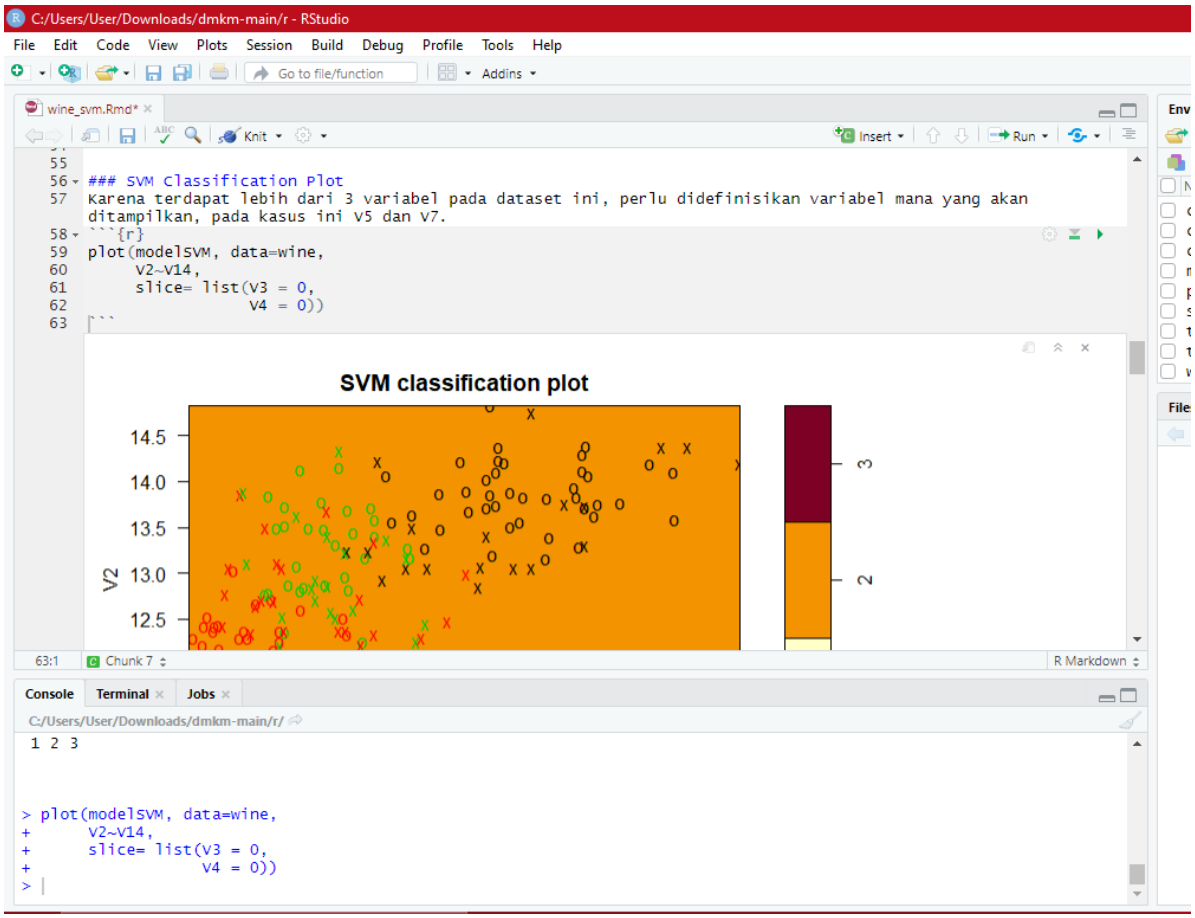
```
> qplot(V2, V14, color = V1, data=wine)
> 
```

Selanjutnya data dibagi menjadi training set dan test set

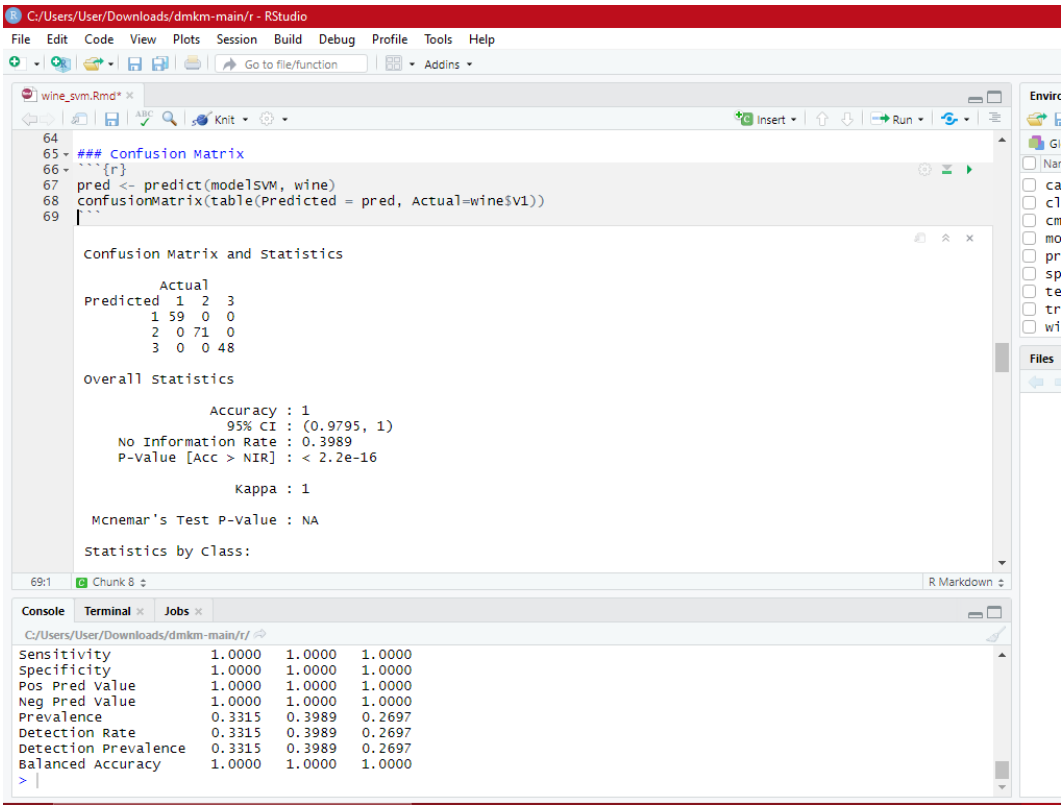




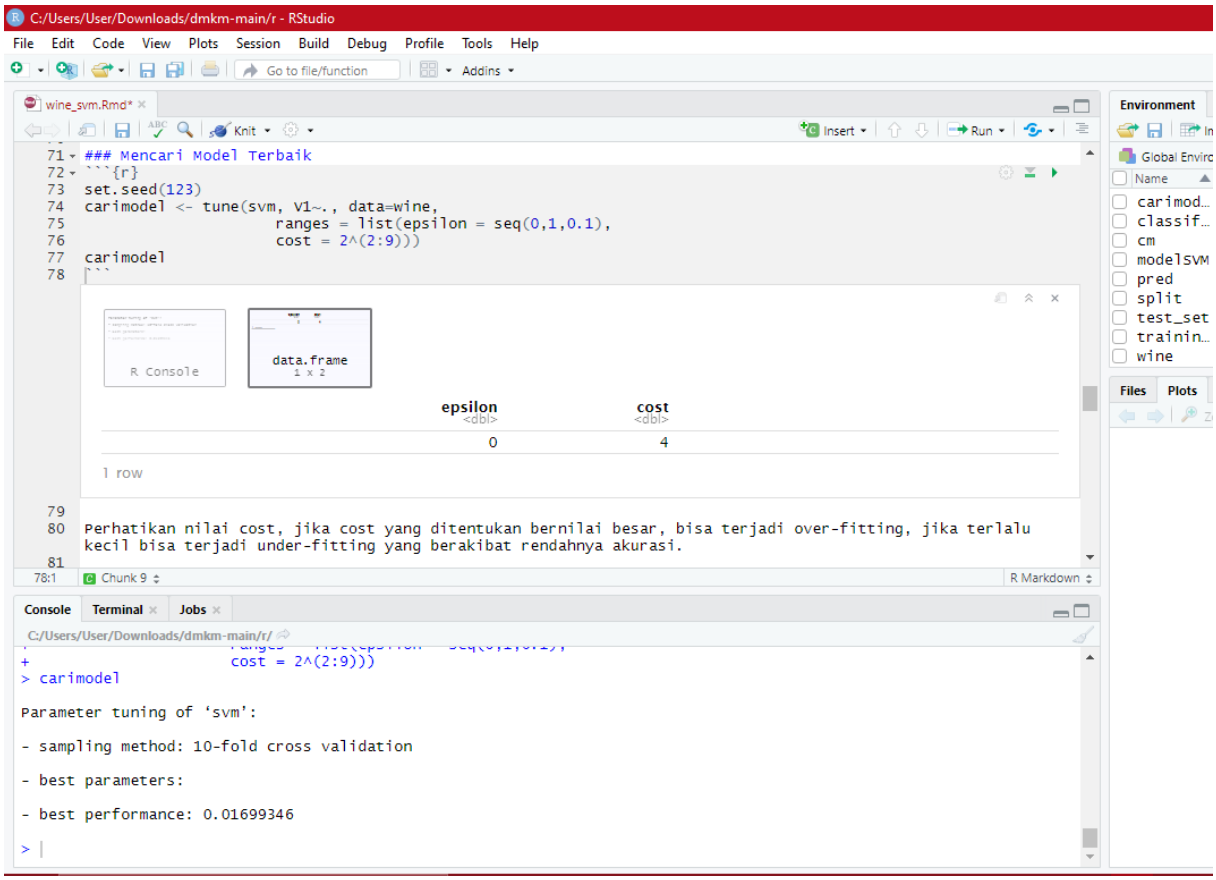
Membuat SVM classification plot:

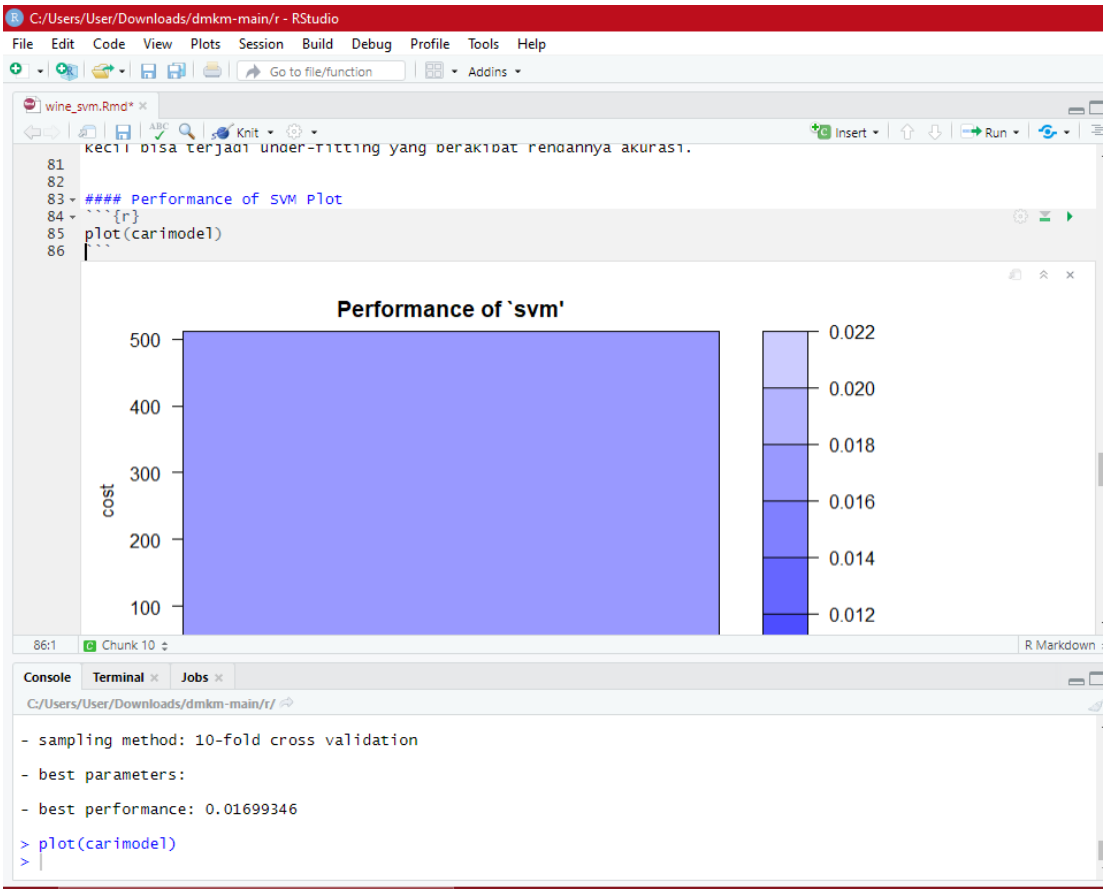


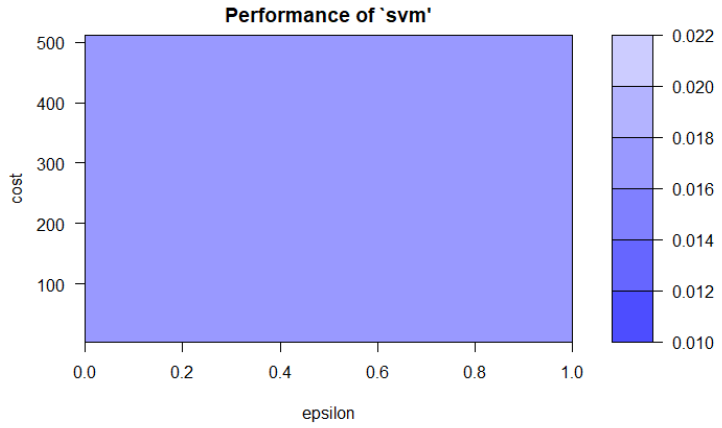
Membuat confusion matrix:

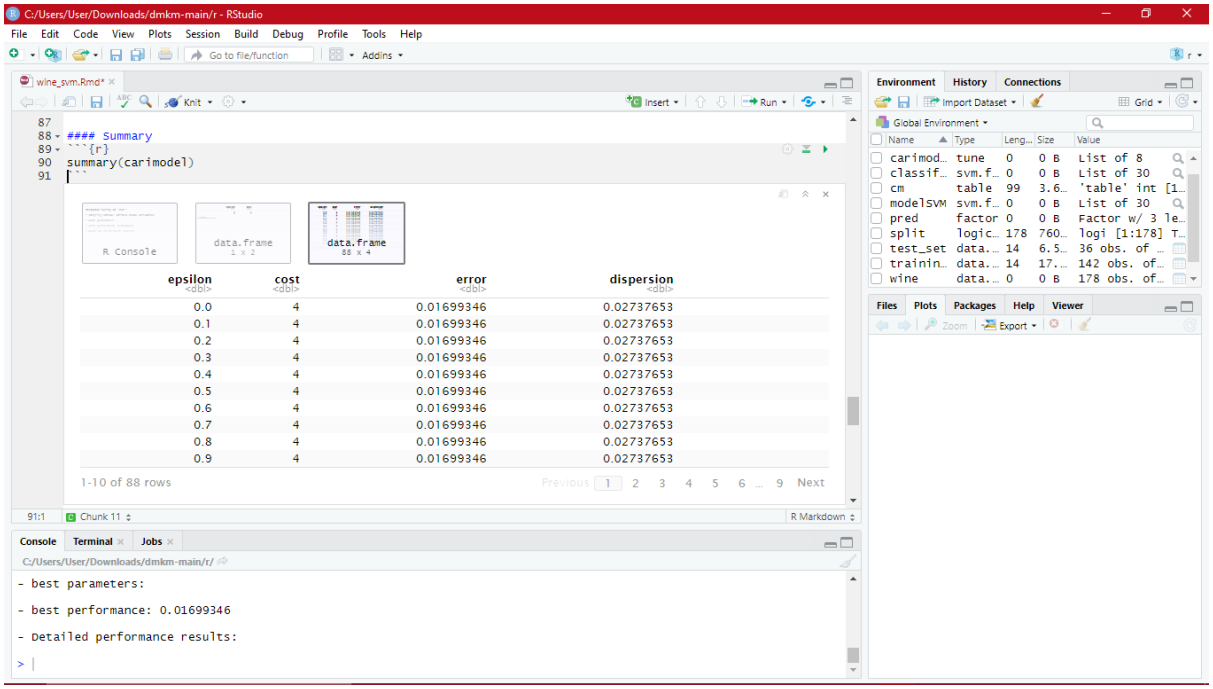


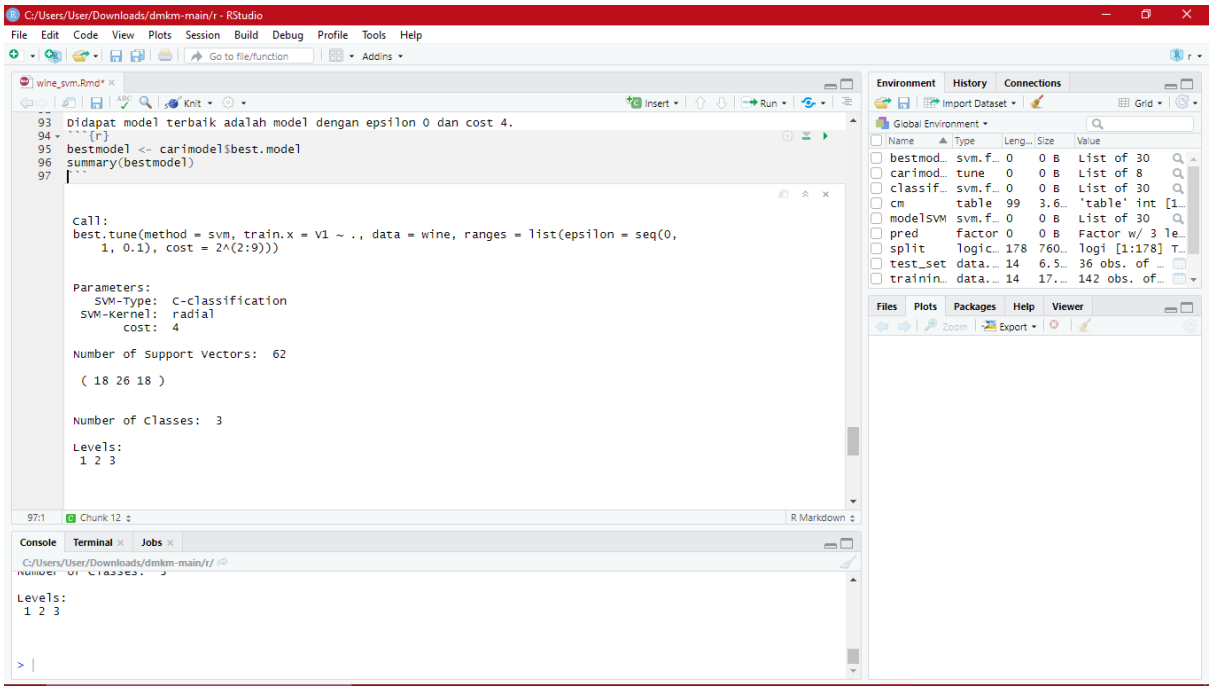
Mencari model terbaik:

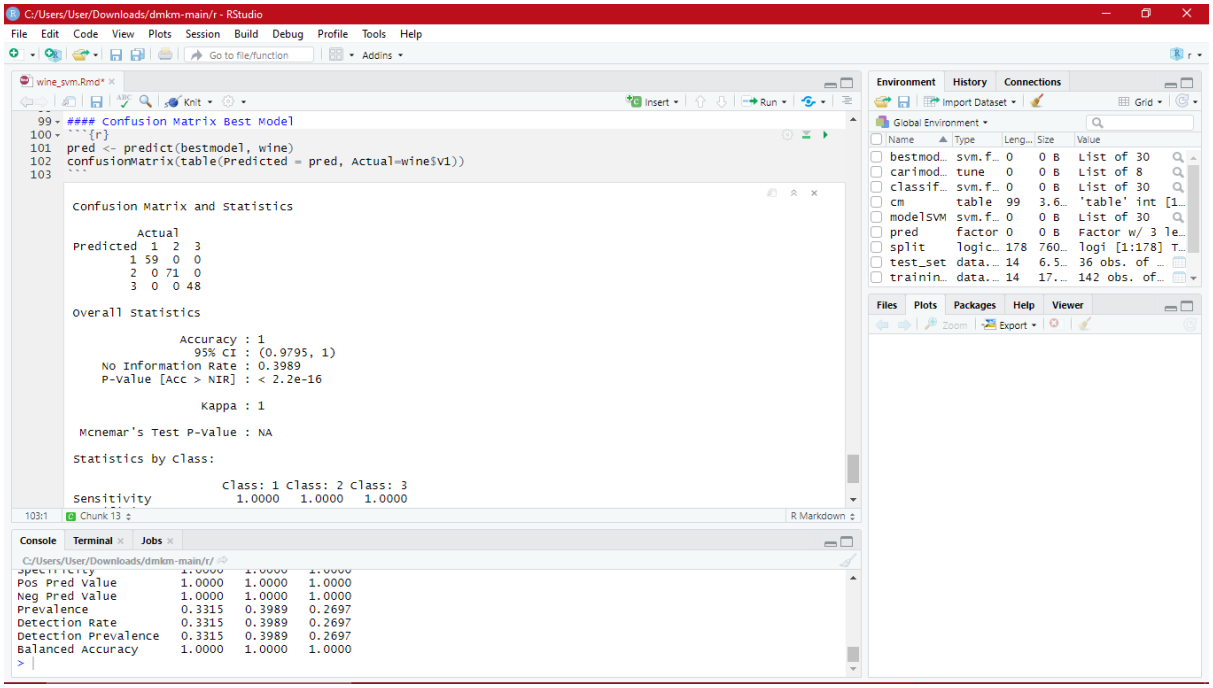





Performance of `svm'
cost
epsilon
500
400
300
200
100
0.0
0.2
0.4
0.6
0.8
1.0
0.022
0.020
0.018
0.016
0.014
0.012
0.010


C:/Users/User/Downloads/dmkm-main/r - RStudio
wine_svm.Rmd*
Summary
```{r}
summary(carimodel)
```
data.frame
88 x 4
epsilon cost error dispersion
0.0 4 0.01699346 0.02737653
0.1 4 0.01699346 0.02737653
0.2 4 0.01699346 0.02737653
0.3 4 0.01699346 0.02737653
0.4 4 0.01699346 0.02737653
0.5 4 0.01699346 0.02737653
0.6 4 0.01699346 0.02737653
0.7 4 0.01699346 0.02737653
0.8 4 0.01699346 0.02737653
0.9 4 0.01699346 0.02737653
1-10 of 88 rows
- best parameters:
- best performance: 0.01699346
- Detailed performance results:

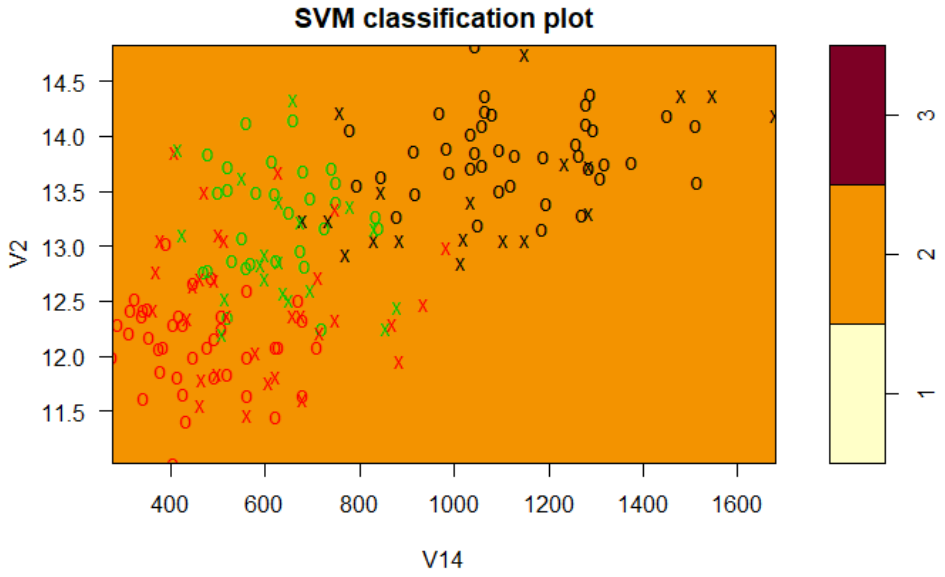

C:/Users/User/Downloads/dmkm-main/r - RStudio
wine_svm.Rmd*
Didapat model terbaik adalah model dengan epsilon 0 dan cost 4.
```{r}
bestmodel <- carimodel$best.model
summary(bestmodel)
```
Call:
best.tune(method = svm, train.x = V1 ~ ., data = wine, ranges = list(epsilon = seq(0,
1, 0.1), cost = 2^(2:9)))
Parameters:
SVM-Type: C-classification
SVM-Kernel: radial
cost: 4
Number of Support Vectors: 62
(18 26 18)
Number of Classes: 3
Levels:
1 2 3



Confusion matrix dan statistic evaluasi dari best model yang didapatkan:

```
Confusion Matrix and Statistics

          Actual
Predicted  1  2  3
        1 59  0  0
        2  0 71  0
        3  0  0 48

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9795, 1)
    No Information Rate : 0.3989
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 1 Class: 2 Class: 3
Sensitivity            1.0000   1.0000   1.0000
Specificity            1.0000   1.0000   1.0000
Pos Pred Value         1.0000   1.0000   1.0000
Neg Pred Value         1.0000   1.0000   1.0000
Prevalence             0.3315   0.3989   0.2697
Detection Rate         0.3315   0.3989   0.2697
Detection Prevalence   0.3315   0.3989   0.2697
Balanced Accuracy      1.0000   1.0000   1.0000
```

2. Model yang didapat dan interpretasinya



3. Confusion matrix

```
Confusion Matrix and Statistics

          Actual
Predicted  1  2  3
        1 59  0  0
        2  0 71  0
        3  0  0 48

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9795, 1)
    No Information Rate : 0.3989
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1

 Mcnemar's Test P-Value : NA
```

4. Evaluasi model

```
Statistics by Class:

                     Class: 1 Class: 2 Class: 3
Sensitivity            1.0000   1.0000   1.0000
Specificity            1.0000   1.0000   1.0000
Pos Pred Value         1.0000   1.0000   1.0000
Neg Pred Value         1.0000   1.0000   1.0000
Prevalence             0.3315   0.3989   0.2697
Detection Rate         0.3315   0.3989   0.2697
Detection Prevalence   0.3315   0.3989   0.2697
Balanced Accuracy      1.0000   1.0000   1.0000
```