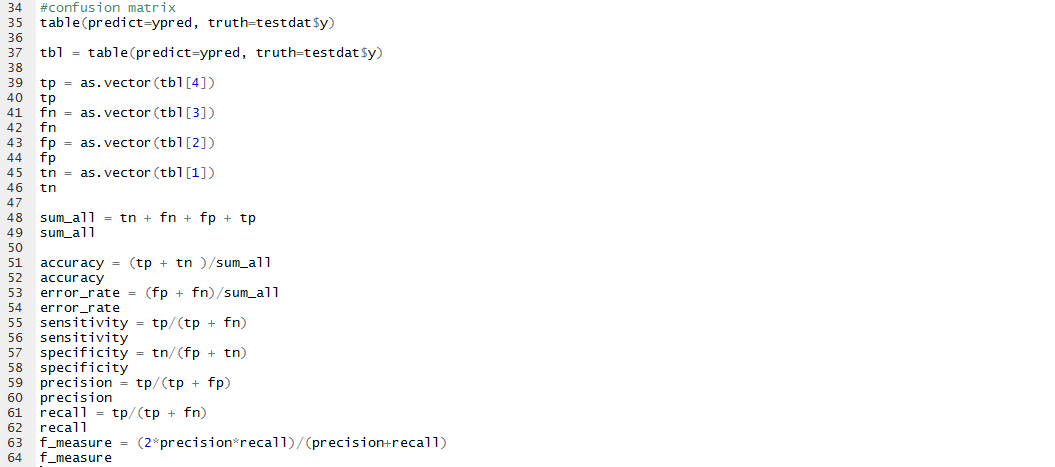
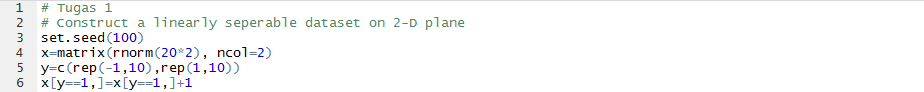
# Tugas 1 :

## Screen shot code program setelah line 34:



## Berikut ini adalah penjelasan dari setiap kode program.

Membangun dataset yang tergolong ke dalam dua kelas. Didefinisikan 20 data dengan 10 data berkelas positif (1) dan 10 data berkelas negatif (-1).



*Plot* menunjukkan bahwa data yang kita *generate* di *step* sebelumnya tidak terpisah  
secara linear.



Selanjutnya, kita simpan data tersebut dalam *data frame* yang secara sederhana kita sebut *data matrix.*



Memanggil *library* e1071 yang merupakan *library* libsvm di R.



Memanggil fungsi **svm()** dan melakukan pemilihan *kernel function* dan *error parameter* C (*cost function*). Argumen **scale=FALSE** berfungsi untuk memberitahu fungsi **svm()** untuk tidak menskalakan masing-masing fitur supaya memiliki mean nol atau standar deviasi satu.



Mengidentifikasi identitas support vector



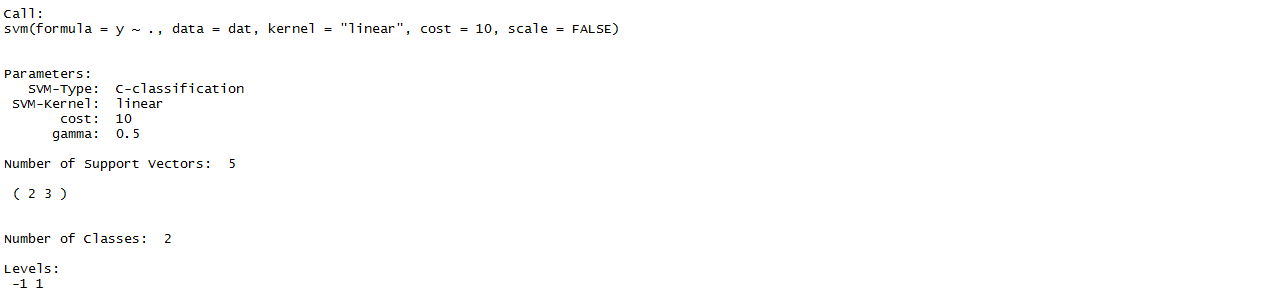
Hasilnya:



Menampilkan informasi ringkas mengenai model yang telah dibuat dengan menggunakan summary().



Hasilnya:



Menemukan *optimal tuning parameter*



Menerapkan *cross-validation* dengan menggunakan fungsi *tune() untuk* melakukan *10-fold cross-validation* pada sebuah *set model* dengan cara membandingkan beberapa *model* yang diperoleh dengan nilai *cost* yang berbedabeda



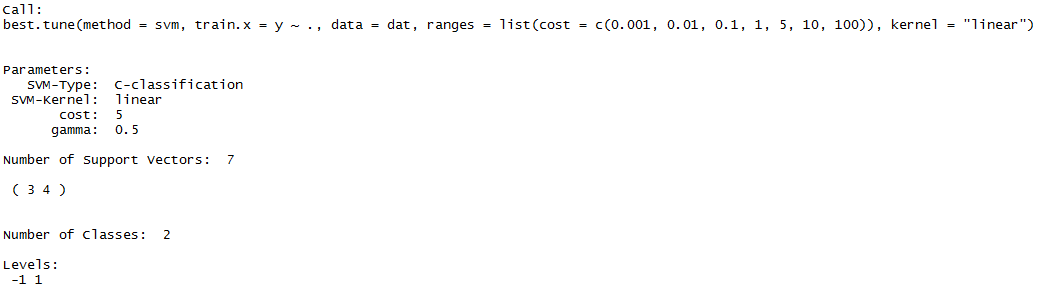
Memilih dan menyimpan model yang terbaik:



Menampilkan informasi ringkas mengenai model yang telah dibuat dengan menggunakan summary().



Hasilnya:



Mendefininisikan *testing set menggunakan 20 testing set*



Menggunakan fungsi predict untuk memprediksi label kelas masing masing testing set.



Menampilkan hasil prediksi terhadap 20 testing data dalam confusion matrix.



Hasilnya :



Mendefenisikan *variable* tbl sebagai *confusion matrix.*

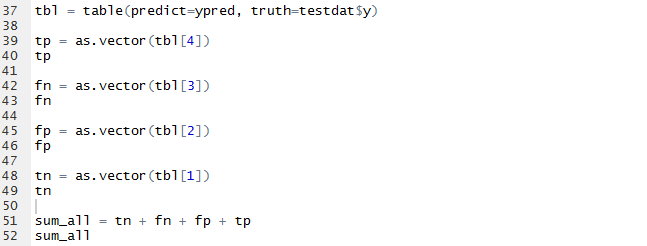
Mendefenisikan *variable* tp untuk memperoleh nilai *true positive* diperoleh dari nilai *truth* 1, dan *predict* 1 yaitu nilai pada kotak ke-4.

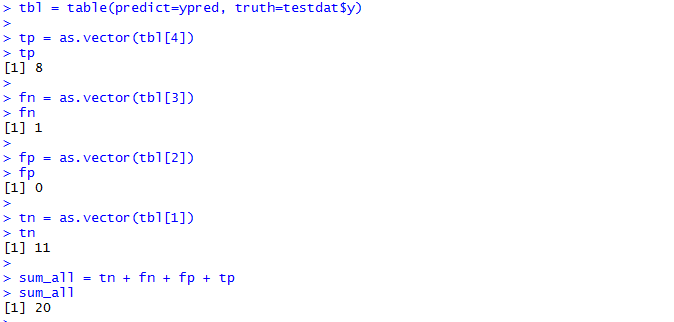
Mendefenisikan *variable* fn untuk memperoleh nilai *false negative* diperoleh dari nilai *truth* 1, dan *predict* -1 yaitu nilai pada kotak ke-3.

Mendefenisikan *variable* tp untuk memperoleh nilai *false positive* diperoleh dari nilai *truth* -1, dan *predict* 1 yaitu nilai pada kotak ke-2.

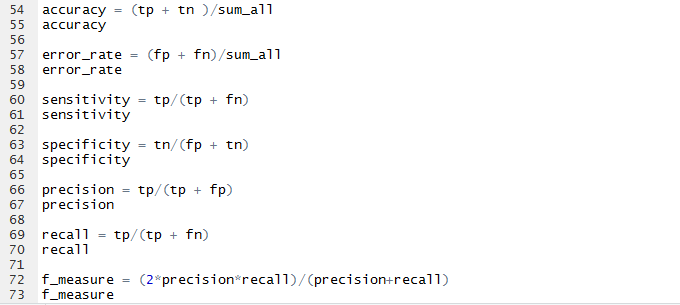
Mendefenisikan *variable* tp untuk memperoleh nilai *true neagative* diperoleh dari nilai *truth* 1, dan *predict* 1 yaitu nilai pada kotak ke-1.

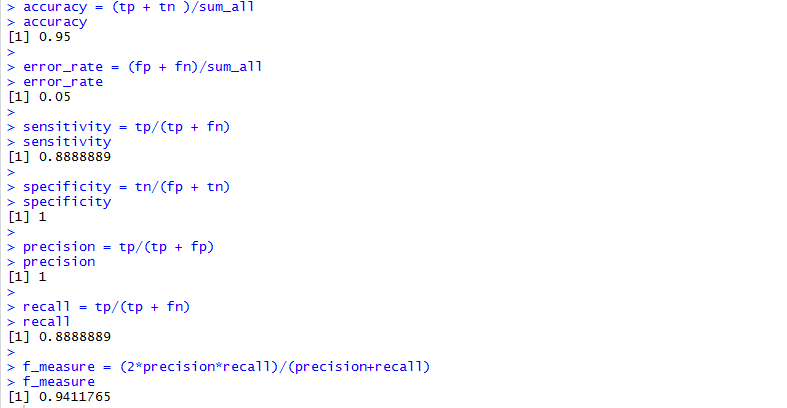
Mendefenisikan *variable* sum\_all untuk memperoleh nilai dari penjumlahan tp + tn + fp + fn.



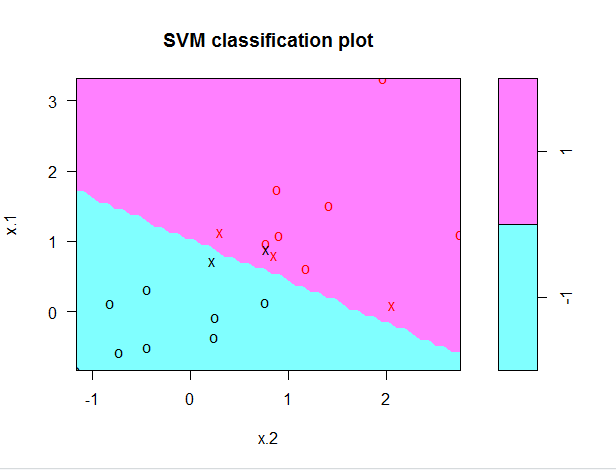
Hasilnya:

Menghitung *accuracy, error rate, sensitivity, specificity, precision, recall,* dan *f\_measure:*



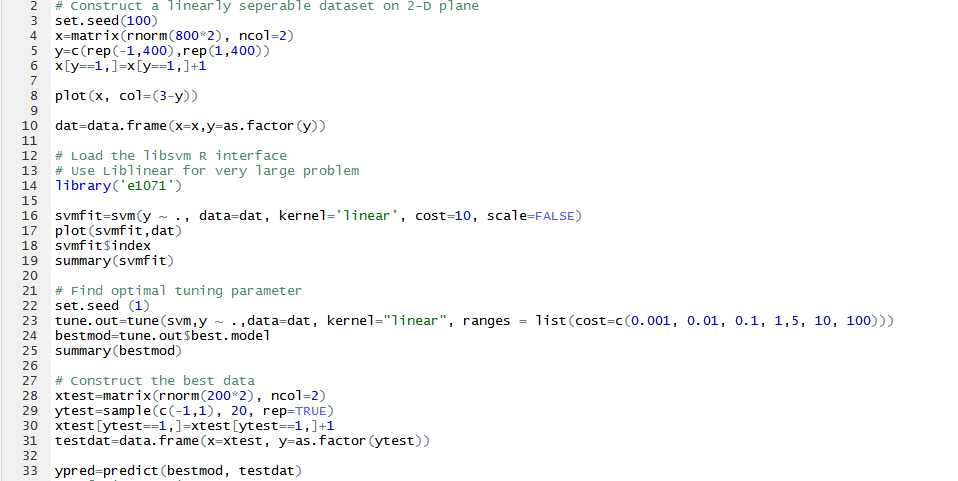
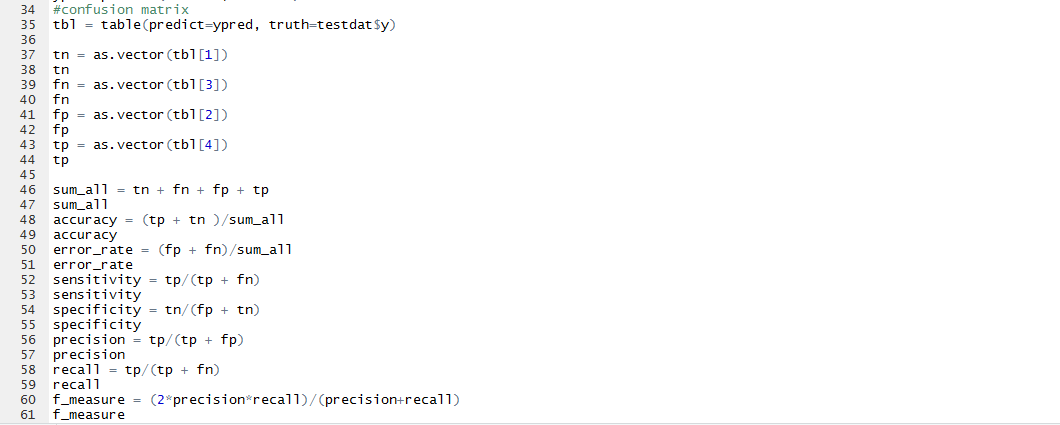
Hasilnya: 

## Hasil visualisasi SVM



# Tugas 2

## *Screen shot code program* dengan menggunakan 800 dataset:

## Berikut ini adalah penjelasan dari kode program.

Membangun dataset yang tergolong ke dalam dua kelas. Didefinisikan 800 data dengan 400 data berkelas positif (1) dan 400 data berkelas negatif (-1).



*Plot* menunjukkan bahwa data yang kita *generate* di *step* sebelumnya tidak terpisah  
secara linear.



Selanjutnya, kita simpan data tersebut dalam *data frame* yang secara sederhana kita sebut *data matrix.*



Memanggil *library* e1071 yang merupakan *library* libsvm di R.



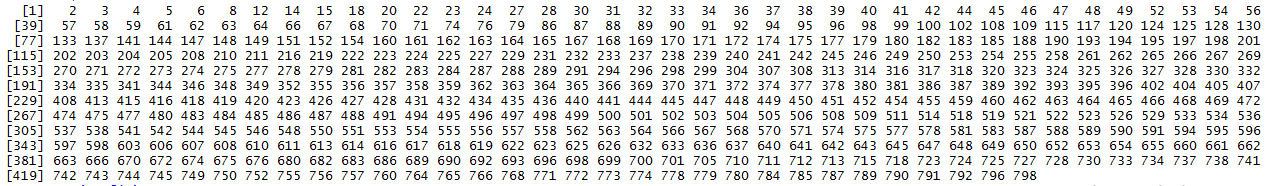
Memanggil fungsi **svm()** dan melakukan pemilihan *kernel function* dan *error parameter* C (*cost function*). Argumen **scale=FALSE** berfungsi untuk memberitahu fungsi **svm()** untuk tidak menskalakan masing-masing fitur supaya memiliki mean nol atau standar deviasi satu



Mengidentifikasi identitas *support vector*



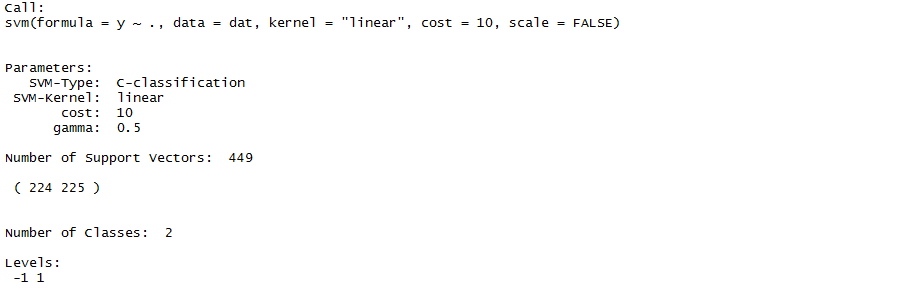
Hasilnya:



Menampilkan informasi ringkas mengenai model yang telah dibuat dengan menggunakan summary().



Hasilnya:



Menemukan *optimal tuning parameter.*



Menerapkan *cross-validation* dengan menggunakan fungsi *tune()* untuk melakukan 10*-fold cross*-*validation* pada sebuah set model dengan cara membandingkan beberapa model yang diperoleh dengan nilai *cost* yang berbedabeda



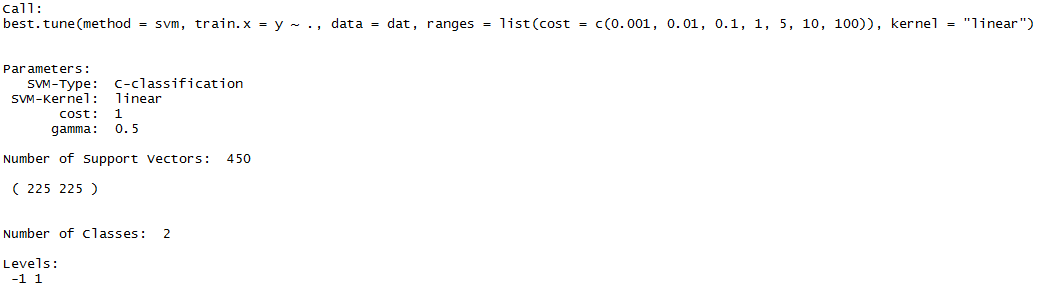
Memilih dan menyimpan model yang terbaik:



Menampilkan informasi ringkas mengenai model yang telah dibuat dengan menggunakan summary().



Hasilnya:



Mendefininisikan *testing set* menggunakan 200 *testing set*



Menggunakan fungsi *predict* untuk memprediksi label kelas masing masing testing set.



Menampilkan hasil prediksi terhadap 200 testing data dalam confusion matrix:



Hasilnya :



Mendefenisikan *variable* tbl sebagai *confusion matrix.*

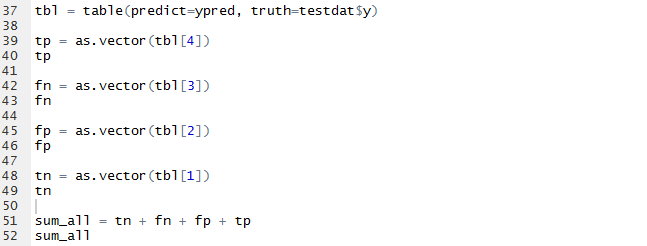
Mendefenisikan *variable* tp untuk memperoleh nilai *true positive* diperoleh dari nilai *truth* 1, dan *predict* 1 yaitu nilai pada kotak ke-4.

Mendefenisikan *variable* fn untuk memperoleh nilai *false negative* diperoleh dari nilai *truth* 1, dan *predict* -1 yaitu nilai pada kotak ke-3.

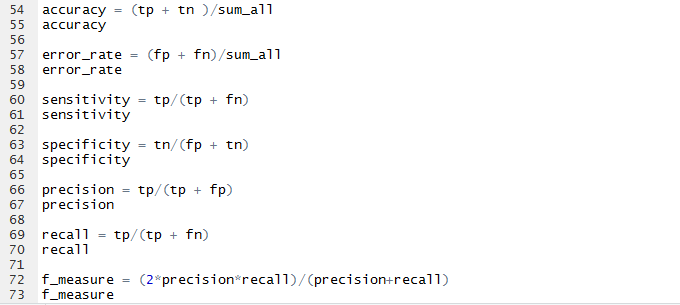
Mendefenisikan *variable* tp untuk memperoleh nilai *false positive* diperoleh dari nilai *truth* -1, dan *predict* 1 yaitu nilai pada kotak ke-2.

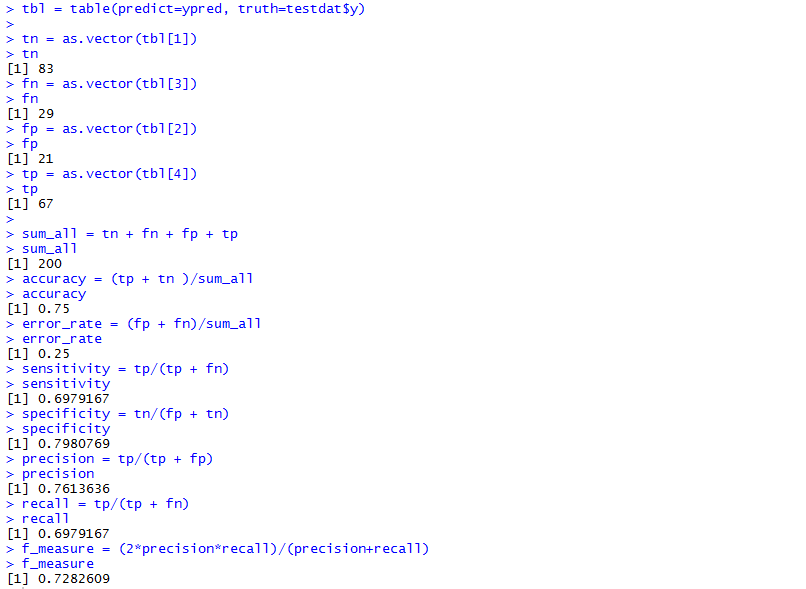
Mendefenisikan *variable* tp untuk memperoleh nilai *true neagative* diperoleh dari nilai *truth* 1, dan *predict* 1 yaitu nilai pada kotak ke-1.

Mendefenisikan *variable* sum\_all untuk memperoleh nilai dari penjumlahan tp + tn + fp + fn.

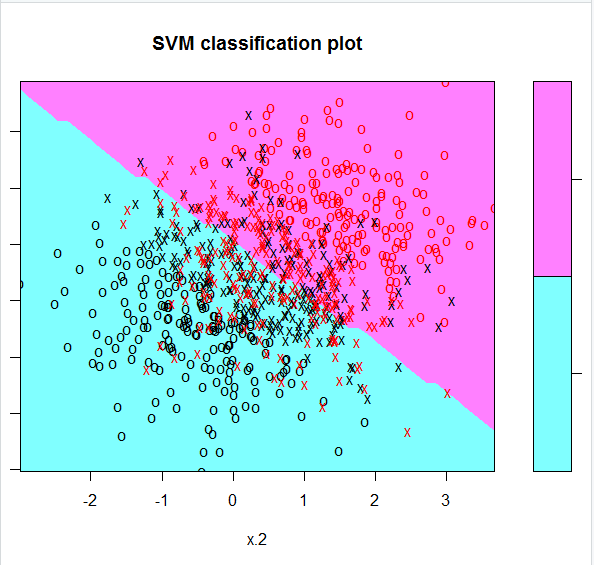


Menghitung *accuracy, error rate, sensitivity, specificity, precision, recall,* dan *f\_measure:*



Hasilnya:

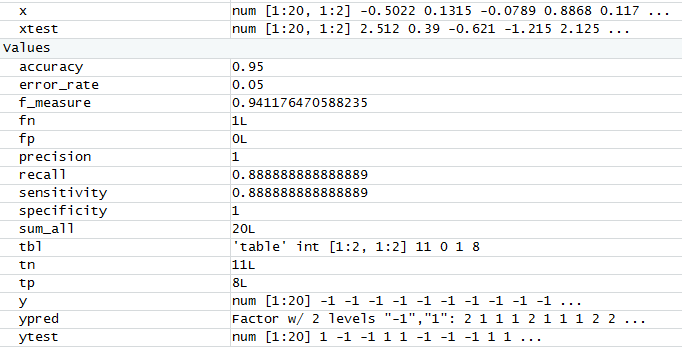
## Hasil visualisasi SVM:



# Hasil Analisa Perbandingan Hasil dengan 20 Dataset dan 800 Dataset

Hasil visualisasi SVM juga menunjukkan bahwa support vector yang diperoleh dengan menggunakan 800 dataset lebih banyak dibandingkan dengan hanya menggunakan 20 dataset. Nilai *accuracy, precision, sensitivity, specificity, f-measure* yang diperoleh dengan menggunakan 20 dataset lebih tinggi dibandingkan dengan menggunakan 800 dataset, sementara *error rate* yang diperoleh dengan menggunakan 20 dataset lebih kecil daripada *error rate* yang diperoleh dengan menggunakan 800 dataset.

Hasil *run code program* 20 dataset:



Hasil *run code program* 800 dataset:

