

Laporan Praktikum Pembelajaran Mesin SVM



**Disusun oleh :
Kelompok 3**

- | | |
|------------------------------------|----------------|
| 1. Gede Ranga Wira Aditya | (082011633048) |
| 2. Muhammad Rahmadhani Ferdiansyah | (082011633068) |
| 3. Arya Danu Triatmodjo | (082011633069) |
| 4. Mukhamad Ikhsanudin | (082011633086) |

Kelas I3

PROGRAM STUDI S1 SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS AIRLANGGA

SURABAYA

2021/2022

COBA SEMUA SOURCE CODE SVM

MATLAB

```

1  clc;clear
2  data = xlsread('datasbmbaru1.xlsx');
3  datatraining = data(1:floor(0.7*length(data)),1:4);
4  kelastraining = data(1:floor(0.7*length(data)),5);
5  datatesting = data(floor(0.7*length(data))+1:end,1:4);
6  kelastraining = data(floor(0.7*length(data))+1:end,5);
7  a = templateSVM('Standardize',1,'KernelFunction','polynomial');
8  traini = fitcecoc(datatraining,kelastraining,'Learners',a);
9  hasil = predict(traini,datatesting);
10 cek = [hasil kelastraining]
  
```

Workspace

Name	Value	Size	Class
a	1x1 FitTem...	1x1	classr
cek	7x2 double	7x2	double
data	23x5 double	23x5	double
datatest...	7x4 double	7x4	double
datatrain...	16x4 double	16x4	double
hasil	[2,1,2,3,2,1]	7x1	double
kelastrai...	[2,1,2,3,2,1]	7x1	double
kelastra...	16x1 double	16x1	double
traini	1x1 Classifi...	1x1	Classi

data =

1083	761	843	1080	1
464	342	404	666	2
685	260	452	930	1
592	388	560	750	1
423	160	429	613	2
1014	655	1110	1338	1
612	260	695	853	1
645	420	656	797	1
120	59	102	211	3
743	342	515	1126	1
460	249	516	698	2
594	463	422	843	1

Workspace

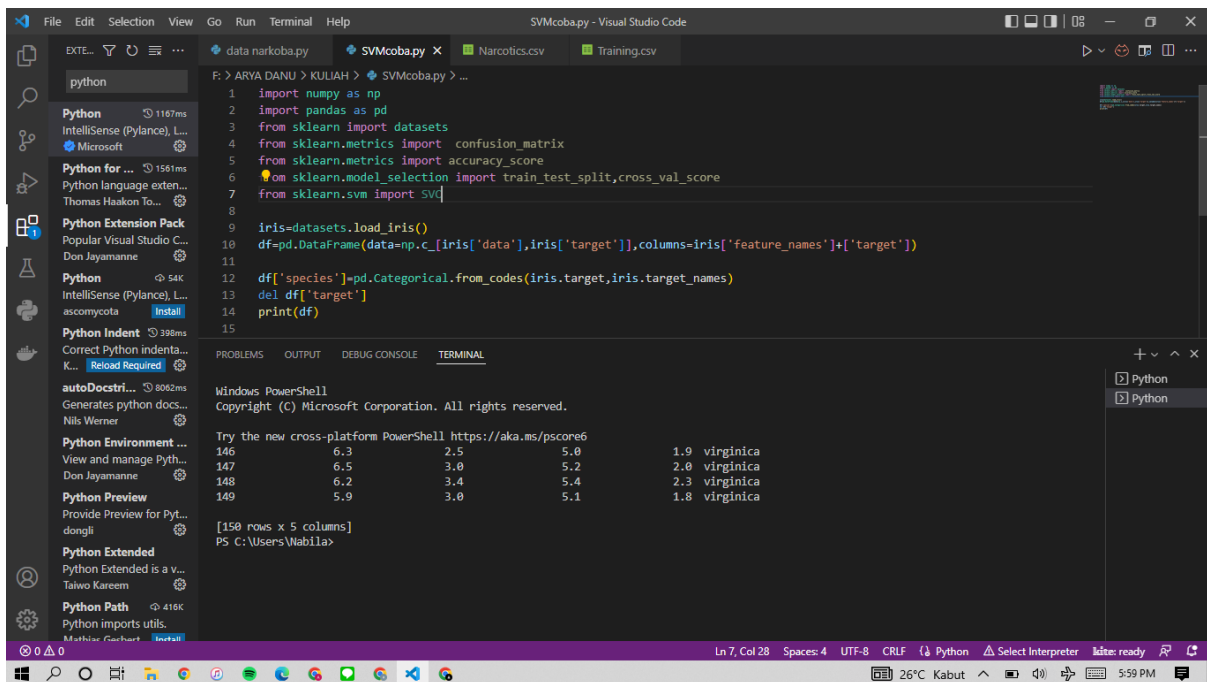
Name	Value	Size	Class
a	1x1 FitTem...	1x1	classr
cek	7x2 double	7x2	double
data	23x5 double	23x5	double
datatest...	7x4 double	7x4	double
datatrain...	16x4 double	16x4	double
hasil	[2,1,2,3,2,1]	7x1	double
kelastrai...	[2,1,2,3,2,1]	7x1	double
kelastra...	16x1 double	16x1	double
traini	1x1 Classifi...	1x1	Classi

cek =

2	2
1	1
2	2
3	3
2	2
2	2
1	1

PYTHON

A. Import Library, Load dataset iris and change dataset



```
1 import numpy as np
2 import pandas as pd
3 from sklearn import datasets
4 from sklearn.metrics import confusion_matrix
5 from sklearn.metrics import accuracy_score
6 from sklearn.model_selection import train_test_split, cross_val_score
7 from sklearn.svm import SVC
8
9 iris = datasets.load_iris()
10 df = pd.DataFrame(data=np.c_[iris['data'], iris['target']], columns=iris['feature_names'] + ['target'])
11
12 df['species'] = pd.Categorical.from_codes(iris.target, iris.target_names)
13 del df['target']
14 print(df)
```

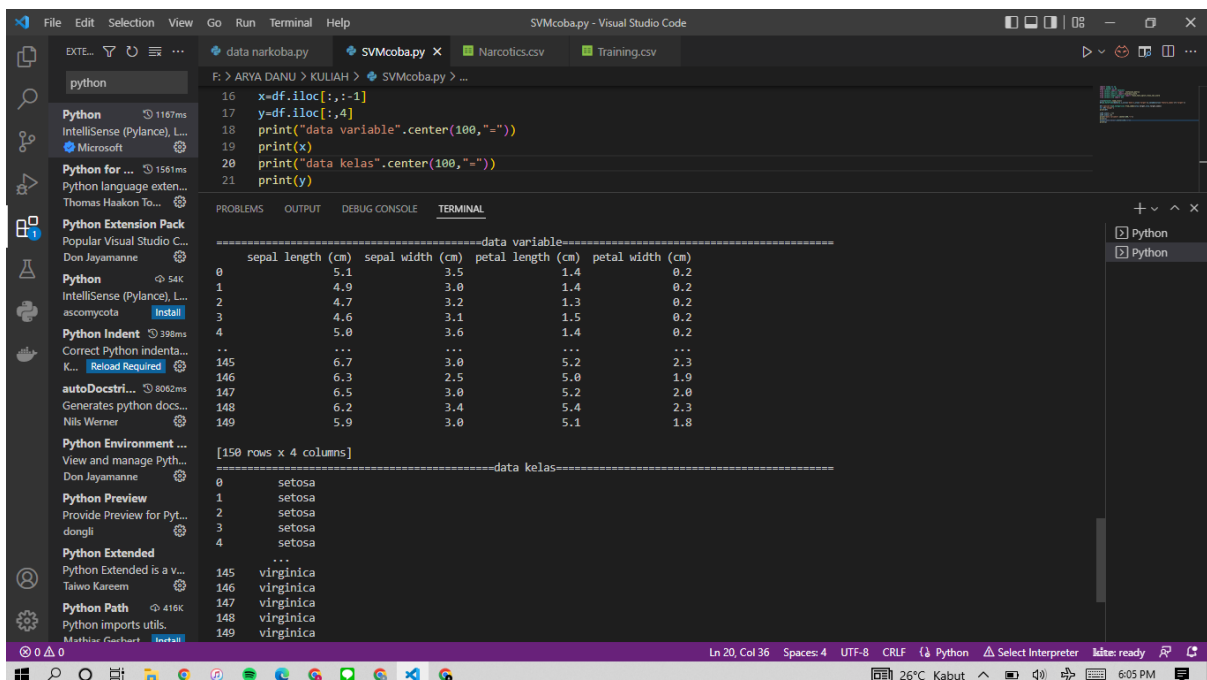
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell <https://aka.ms/pscore6>

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

[150 rows x 5 columns]
PS C:\Users\Wabila>

B. Separate Variables (X) and class instance (Y)



```
16 x = df.iloc[:, :-1]
17 y = df.iloc[:, 4]
18 print("data variable".center(100, "="))
19 print(x)
20 print("data kelas".center(100, "="))
21 print(y)
```

-----data variable-----

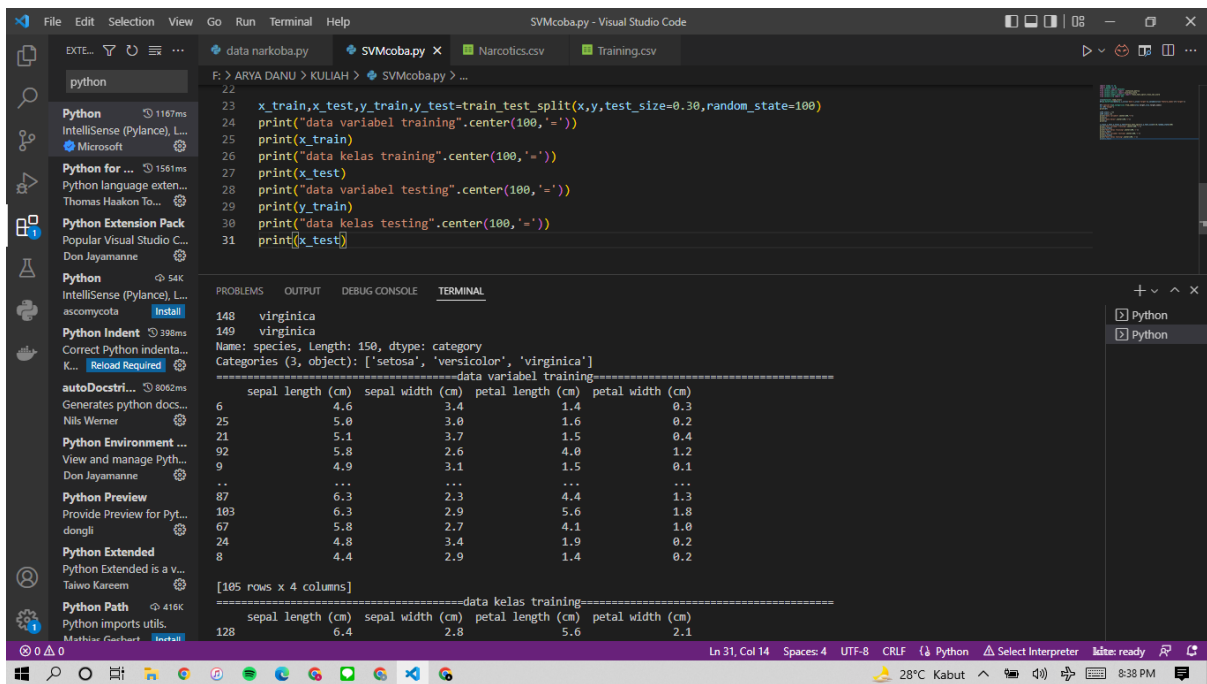
	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
..
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

[150 rows x 4 columns]

-----data kelas-----

	species
0	setosa
1	setosa
2	setosa
3	setosa
4	setosa
..	...
145	virginica
146	virginica
147	virginica
148	virginica
149	virginica

C. Split data, 70% training and 30% testing



```
22
23 x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,random_state=100)
24 print("data variabel training".center(100,'-'))
25 print(x_train)
26 print("data kelas training".center(100,'-'))
27 print(x_test)
28 print("data variabel testing".center(100,'-'))
29 print(y_train)
30 print("data kelas testing".center(100,'-'))
31 print(x_test)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

148 virginica
149 virginica
Name: species, Length: 150, dtype: category
Categories (3, object): ['setosa', 'versicolor', 'virginica']

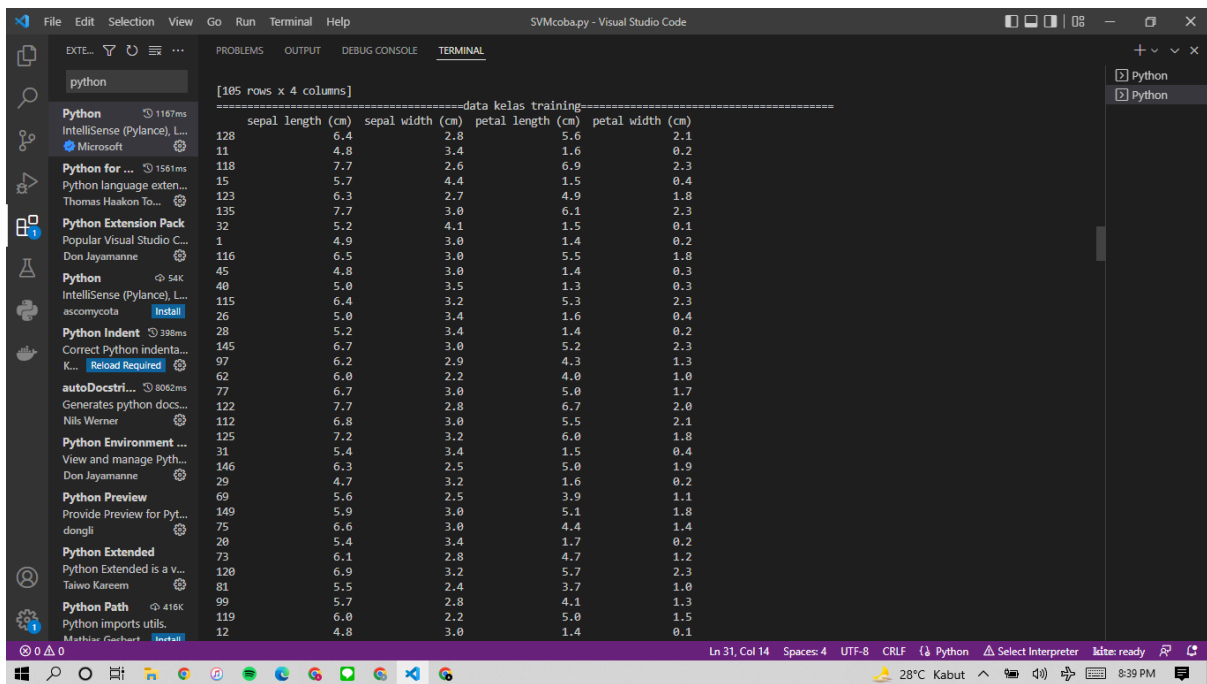
=====data variabel training=====

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
6	4.5	3.4	1.4	0.3
25	4.9	3.0	1.6	0.2
21	5.1	3.7	1.5	0.4
92	5.8	2.6	4.0	1.2
9	4.9	3.1	1.5	0.1
..
87	6.3	2.3	4.4	1.3
103	6.3	2.9	5.6	1.8
67	5.8	2.7	4.1	1.0
24	4.8	3.4	1.9	0.2
8	4.4	2.9	1.4	0.2

[105 rows x 4 columns]

=====data kelas training=====

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
128	6.4	2.8	5.6	2.1



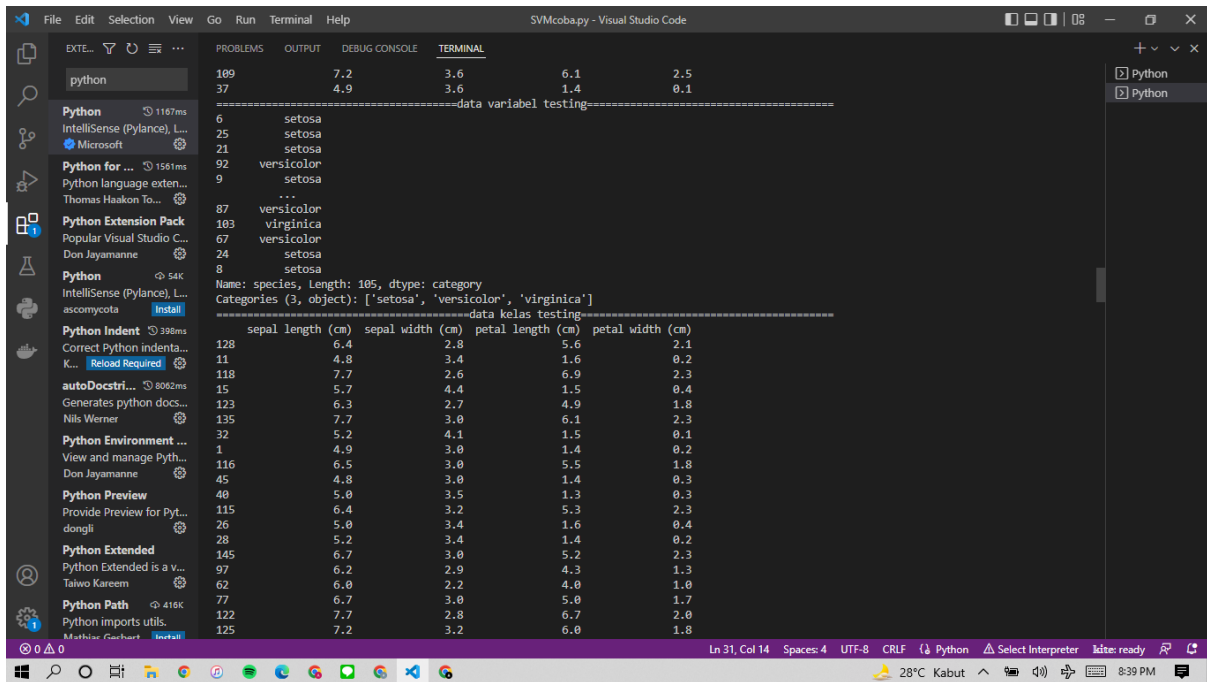
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

[105 rows x 4 columns]

=====data kelas training=====

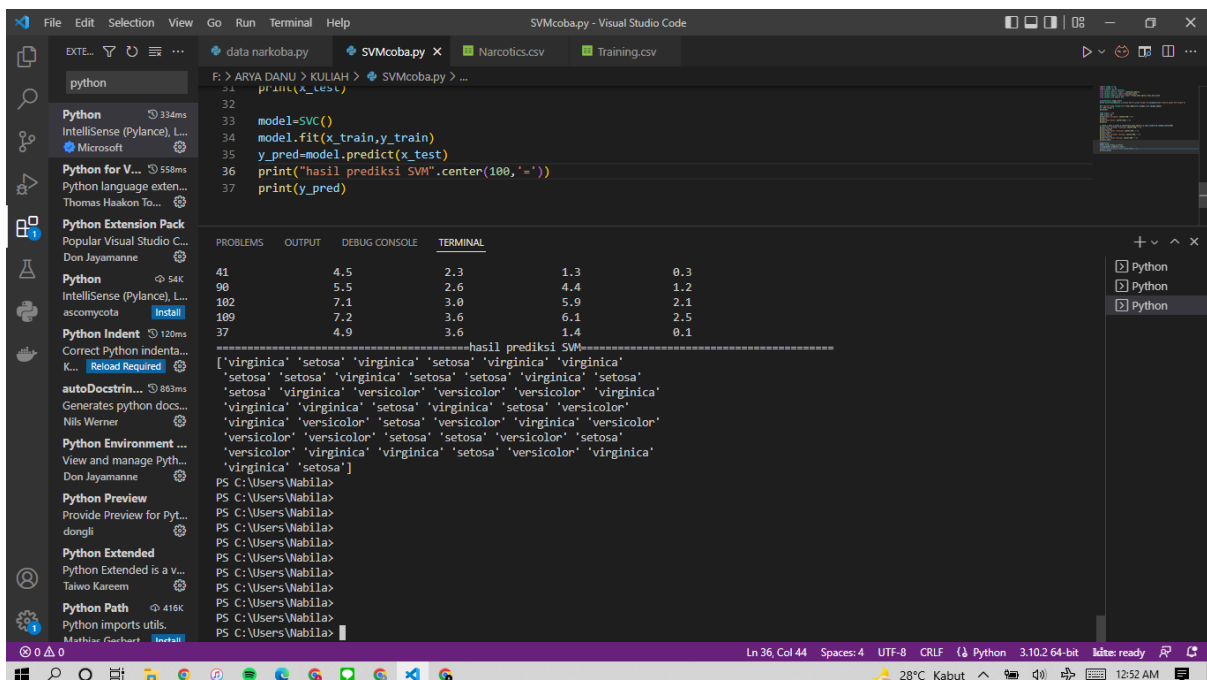
	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
128	6.4	2.8	5.6	2.1
11	4.8	3.4	1.6	0.2
118	7.7	2.6	6.9	2.3
15	5.7	4.4	1.5	0.4
123	6.3	2.7	4.9	1.8
135	7.7	3.0	6.1	2.3
32	5.2	4.1	1.5	0.1
1	4.9	3.0	1.4	0.2
116	6.5	3.0	5.5	1.8
45	4.8	3.0	1.4	0.3
40	5.0	3.5	1.3	0.3
115	6.4	3.2	5.3	2.3
26	5.0	3.4	1.6	0.4
28	5.2	3.4	1.4	0.2
145	6.7	3.0	5.2	2.3
97	6.2	2.9	4.3	1.3
62	6.0	2.2	4.0	1.0
77	6.7	3.0	5.0	1.7
122	7.7	2.8	6.7	2.0
112	6.8	3.0	5.5	2.1
125	7.2	3.2	6.0	1.8
31	5.4	3.4	1.5	0.4
146	6.3	2.5	5.0	1.9
29	4.7	3.2	1.6	0.2
69	5.6	2.5	3.9	1.1
149	5.9	3.0	5.1	1.8
75	6.6	3.0	4.4	1.4
20	5.4	3.4	1.7	0.2
73	6.1	2.8	4.7	1.2
120	6.9	3.2	5.7	2.3
81	5.5	2.4	3.7	1.0
99	5.7	2.8	4.1	1.3
119	6.0	2.2	5.0	1.5
12	4.8	3.0	1.4	0.1

Training data

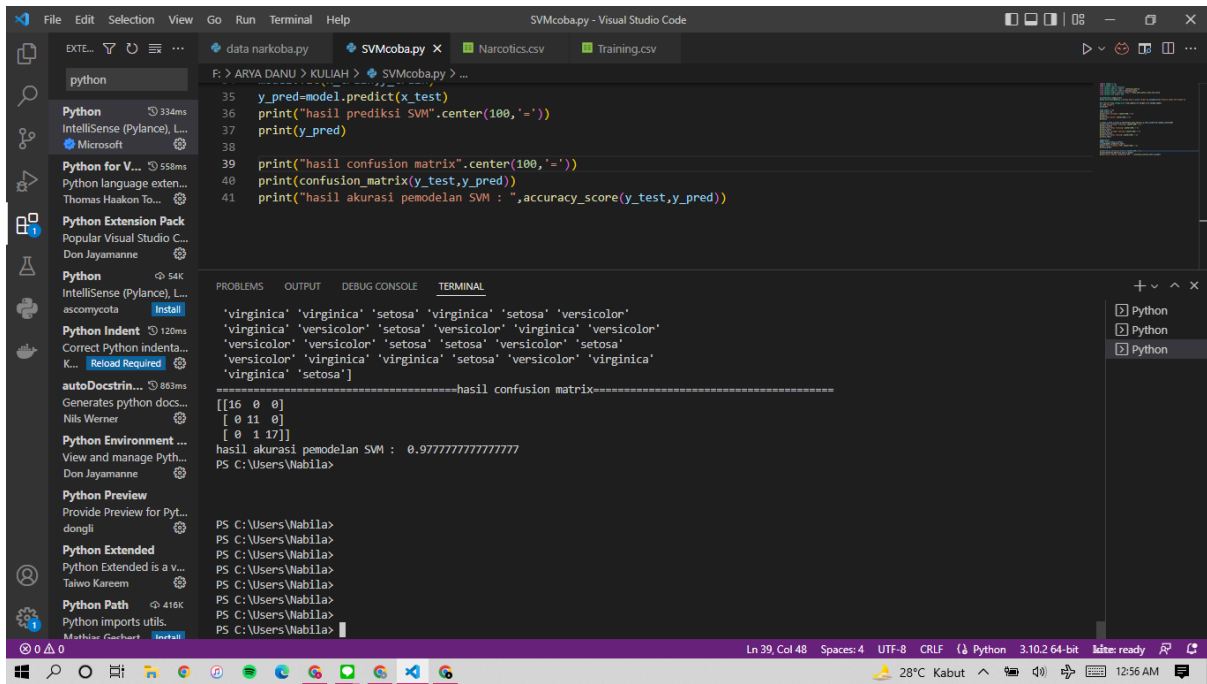


Testing Data

D. SVM Modelling



E. Calculate accuracy matrix



COBA CODE DENGAN DATA SENDIRI

MATLAB

	A	B	C	
	RATING	EFFECTIV...	SIDEEFFE...	
	Number ▾	Number ▾	Number ▾	
1	RATING	EFFECTIV...	SIDE EFFE...	▲
2	1	5	2	
3	10	5	5	
4	3	2	4	
5	2	2	2	
6	1	1	2	▼

	A	B	C	
	RATING	EFFECTIV...	SIDEEFFE...	
	Number ▾	Number ▾	Number ▾	
996	1	1	2	▲
997	9	5	4	
998	10	4	5	
999	10	5	5	
1000	1	1	1	
1001	7	4	3	▼

```
tugasLandPrice.m × SVMcoba.m × SVMdata.m × +
1  clc;clear
2  data = xlsread('Narcotics.xlsx')
3  datatraining = data(1:floor(0.7*length(data)),1:2);
4  kelastraining = data(1:floor(0.7*length(data)),3);
5  datatesting = data(floor(0.7*length(data))+1:end,1:2);
6  kelastesting = data(floor(0.7*length(data))+1:end,3);
7  a = templateSVM('Standardize',1,'KernelFunction','polynomial');
8  traini = fitcecoc(datatraining,kelastraining,'Learners',a);
9  hasil = predict(traini,datatesting);
10 cek = [hasil kelastesting]
11
```

```
Command Window
7      4      3

cek =

     2     2
     5     5
     4     4
     2     2
     2     2
     4     4
     5     5
     5     5
     2     1
     3     3
```

```
Command Window
5      5
2      1
3      3
2      2
5      5
4      4
2      2
2      2
4      4
5      5
5      5
2      1
3      3

>>
```

The dataset used this time is Narcotics.xlsx. This dataset consists of 1001 instances and the variables are divided into 2 data variables and 1 class variable. The comparison between training data and testing data is 7:3. After the SVM classifier has been carried out, a comparison/checking is carried out using 20 instances between the predicted data and the actual data. From these results found 2 differences in data. So the accuracy of the 20 data is 90%.

PYTHON

A. Import Library and Load dataset

```
1 import pandas as pd
2 from sklearn.metrics import confusion_matrix
3 from sklearn.metrics import accuracy_score
4 from sklearn.model_selection import train_test_split
5 from sklearn.svm import SVC
6
7 data=pd.read_csv('Social_Network_Ads.csv')
8 df=pd.DataFrame(data)
9 print(df)
```

There are 2 types of libraries used, namely pandas and sklearn. The dataset used is Social Network Ads which is the result of a survey from several people, whether these people will buy items that are sold based on several attributes.

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	0	19.0	19000	0
1	15810944	0	35.0	20000	0
2	15668575	1	26.0	43000	0
3	15603246	1	27.0	57000	0
4	15804002	0	19.0	76000	0
..
395	15691863	1	46.0	41000	1
396	15706071	0	51.0	23000	1
397	15654296	1	50.0	20000	1
398	15755018	0	36.0	33000	0
399	15594041	1	49.0	36000	1

[400 rows x 5 columns]

B. Separate Variables (X) and class instance (Y)

```
11 x=df.iloc[:,1:3]
12 y=df.iloc[:,4]
```

Attribute variables used are gender, age, and salary. While the target variable is a categorical attribute between true and false.

C. Split data, 70% training and 30% testing

```
14 x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=100)
```

The dataset is divided where the training data gets a share of 70% and the testing data gets a share of 30%.

D. SVM Modelling

```

16     model=SVC()
17     model.fit(x_train, y_train)
18     y_pred=model.predict(x_test)
19     print("HASIL PREDIKSI SVM")
20     print(y_pred)

```

```

HASIL PREDIKSI SVM
[0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 1 1 1 1 0
 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 1 0 1 1 0 1 0 1 0 0 0 0 1 0 0 0 0 1 1 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 1 1 1 0 1
 0 0 0 1 0 0 1 0 1]

```

Furthermore, predictions are made using SVM modeling and the results formed are in the form of a 2-dimensional array.

E. Calculate confusion matrix and accuracy score

```

21     print("HASIL CONFUSION MATRIX")
22     print(confusion_matrix(y_test, y_pred))
23     print("HASIL AKURASI PEMODELAN SVM:", accuracy_score(y_test, y_pred))

```

```

HASIL CONFUSION MATRIX
[[72  3]
 [14 31]]
HASIL AKURASI PEMODELAN SVM: 0.8583333333333333

```

The final results obtained are in the form of a confusion matrix and an accuracy score of: $(72+31)/(72+31+14+3) = 0.8583$.