LAPORAN

PRAKTIKUM DATA WAREHOUSING DAN DATA MINING

(MODUL 8 & 9)

Disusun oleh :

NAMA : DIEGO ALFIANDRO

NIM : L200200039

KELAS : D

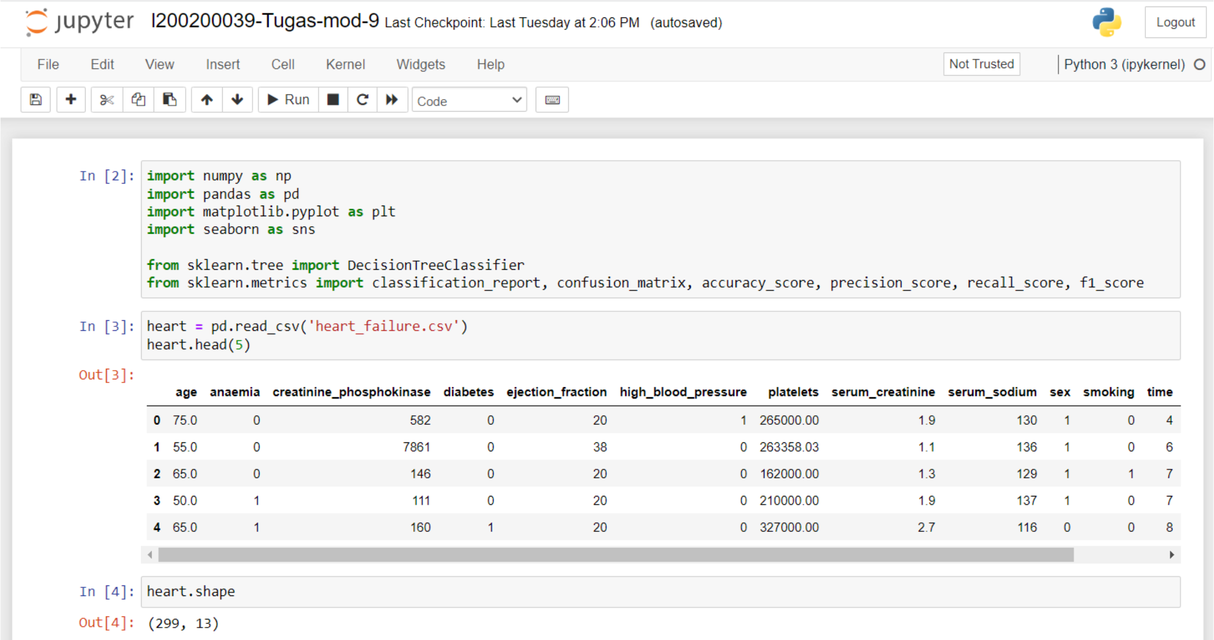
INFORMATIKA

FAKULTAS KOMUNIKASI DAN INFORMATIKA

UNIVERSITAS MUHAMMADIYAH SURAKARTA

TAHUN 2022/2023

➢



In [2]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, precision_score, recall_score, f1_score
```

In [3]:
```python
heart = pd.read_csv('heart_failure.csv')
heart.head(5)
```

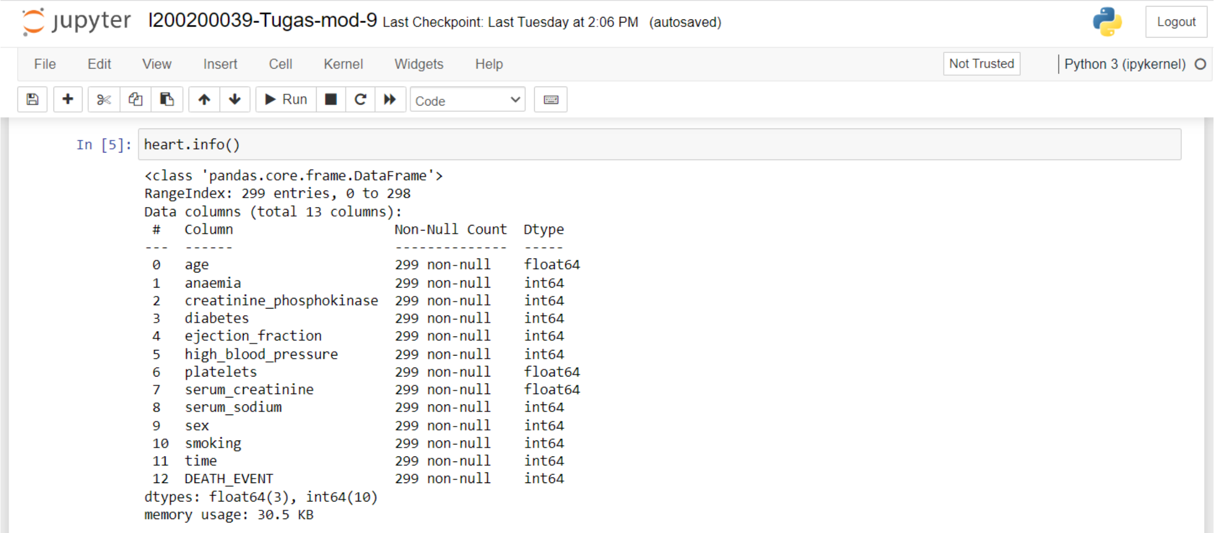
Out[3]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	4
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	6
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	7
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	7
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	8

In [4]:
```python
heart.shape
```

Out[4]:
```
(299, 13)
```

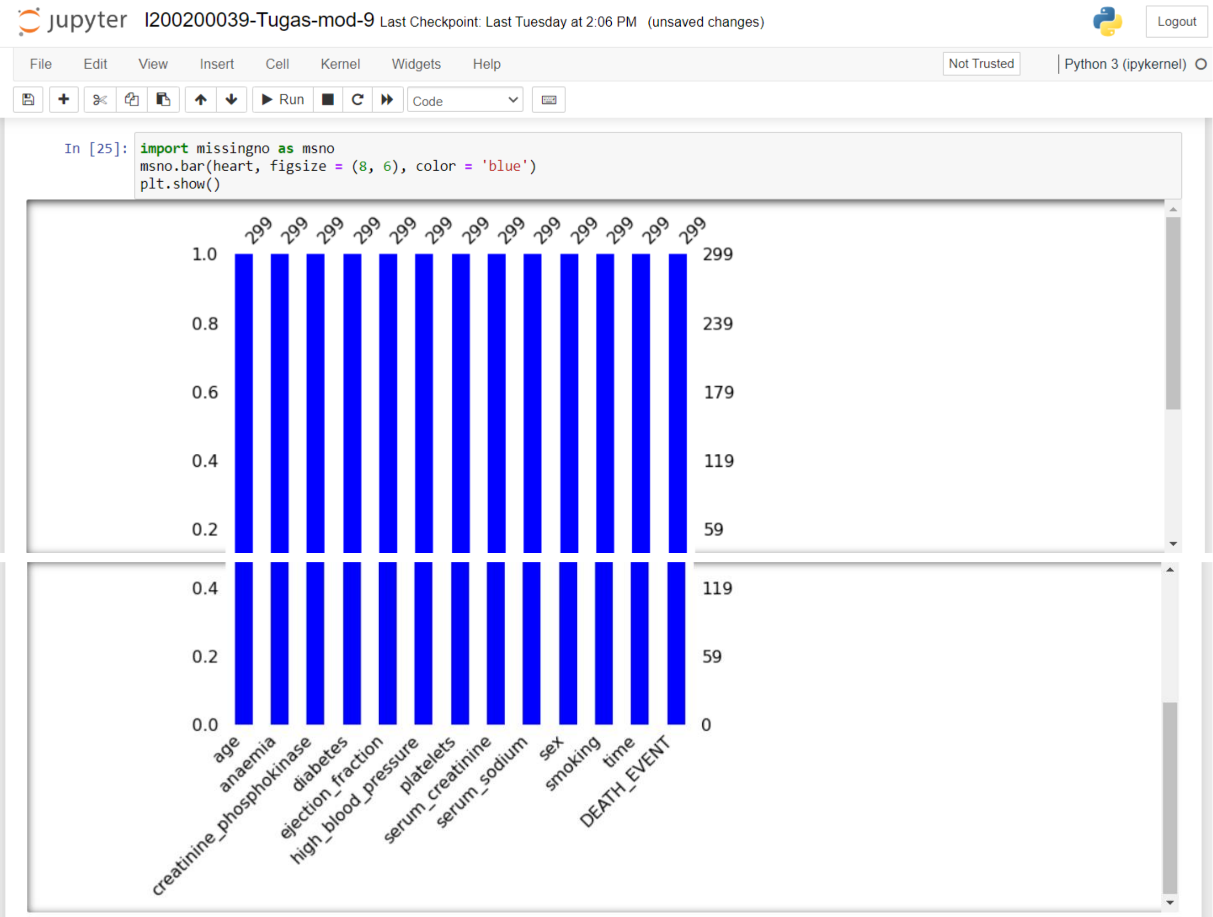
➢



In [5]:
```python
heart.info()
```

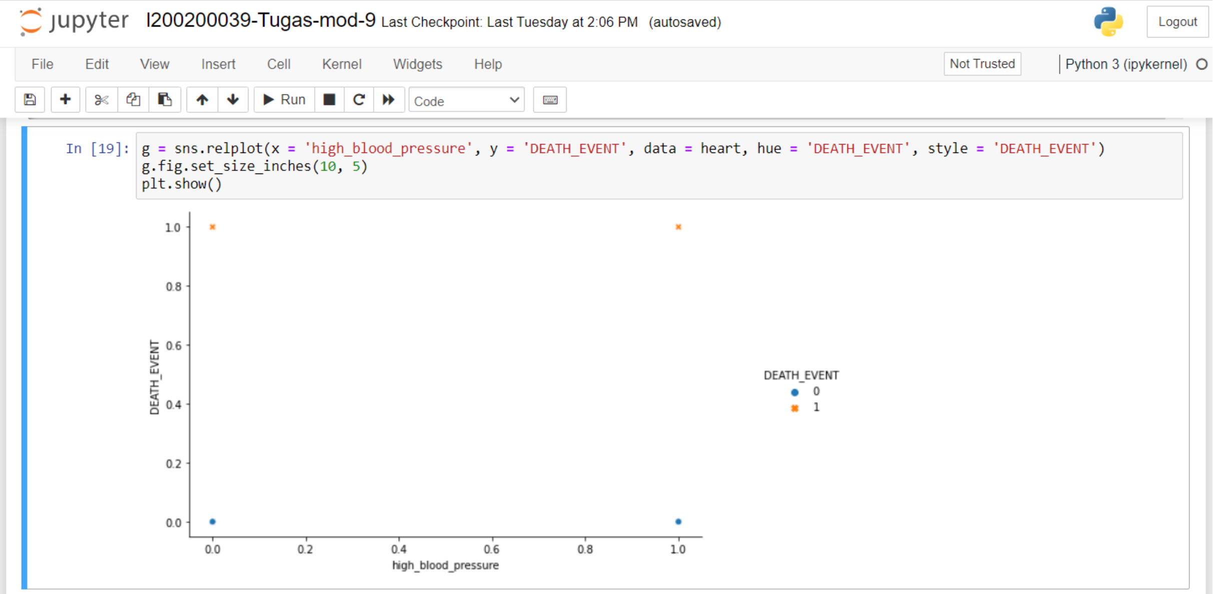
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 13 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   age                       299 non-null    float64
 1   anaemia                   299 non-null    int64  
 2   creatinine_phosphokinase  299 non-null    int64  
 3   diabetes                  299 non-null    int64  
 4   ejection_fraction         299 non-null    int64  
 5   high_blood_pressure       299 non-null    int64  
 6   platelets                 299 non-null    float64
 7   serum_creatinine          299 non-null    float64
 8   serum_sodium              299 non-null    int64  
 9   sex                       299 non-null    int64  
 10  smoking                   299 non-null    int64  
 11  time                      299 non-null    int64  
 12  DEATH_EVENT               299 non-null    int64  
dtypes: float64(3), int64(10)
memory usage: 30.5 KB
```

➢



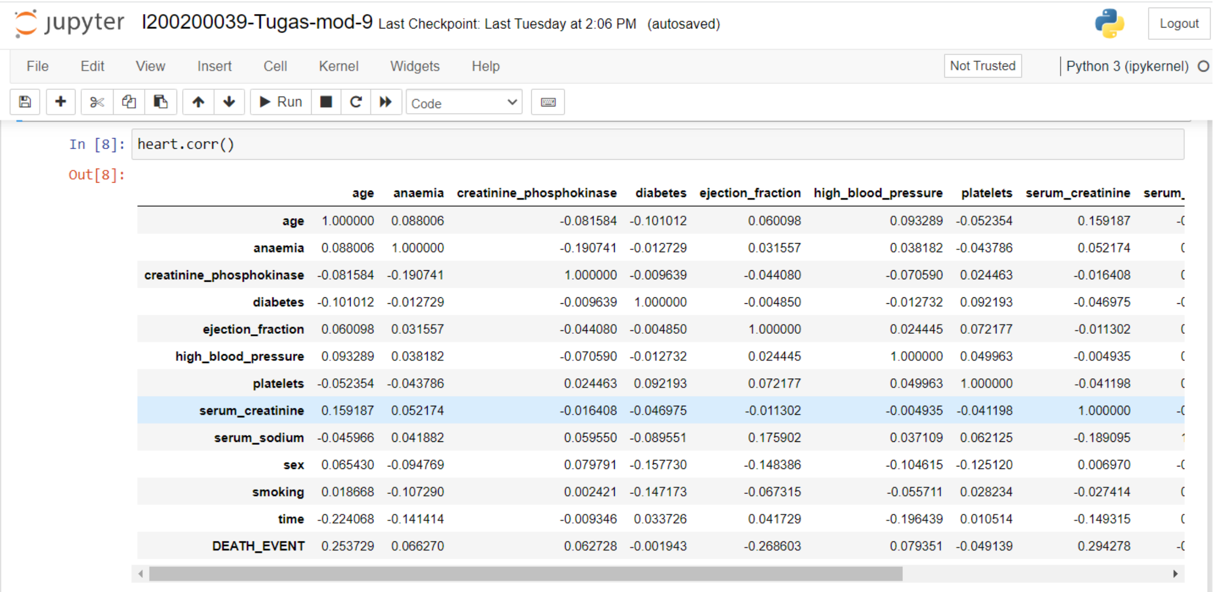
```
In [25]: import missingno as msno
         msno.bar(heart, figsize = (8, 6), color = 'blue')
         plt.show()
```

➢



```
In [19]: g = sns.relplot(x = 'high_blood_pressure', y = 'DEATH_EVENT', data = heart, hue = 'DEATH_EVENT', style = 'DEATH_EVENT')
         g.fig.set_size_inches(10, 5)
         plt.show()
```

➢

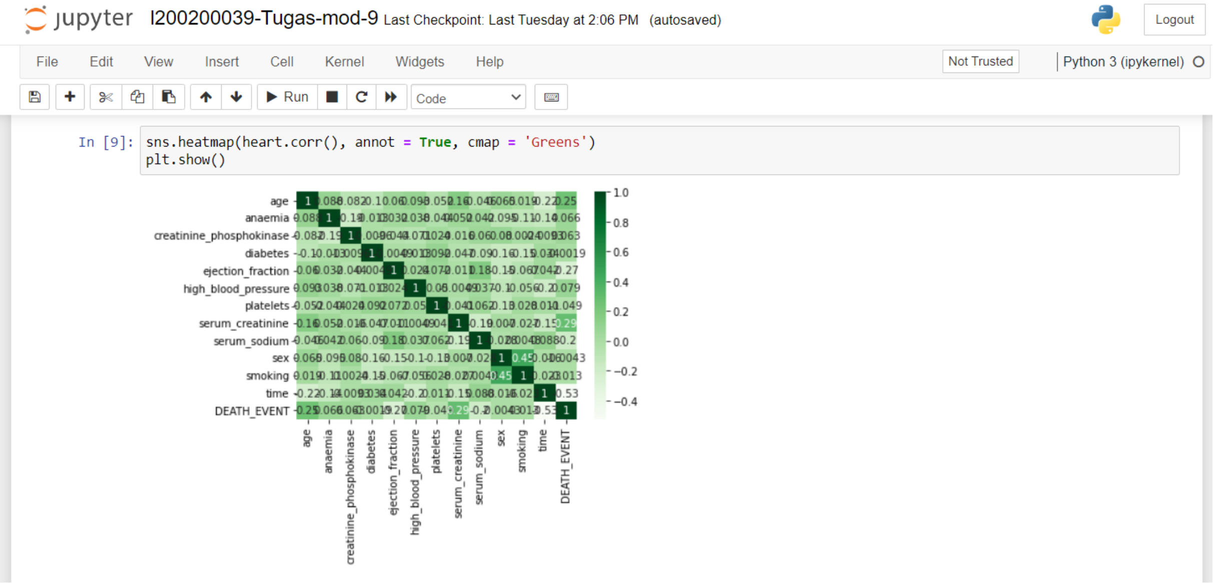


```
In [8]: heart.corr()
```

Out[8]:

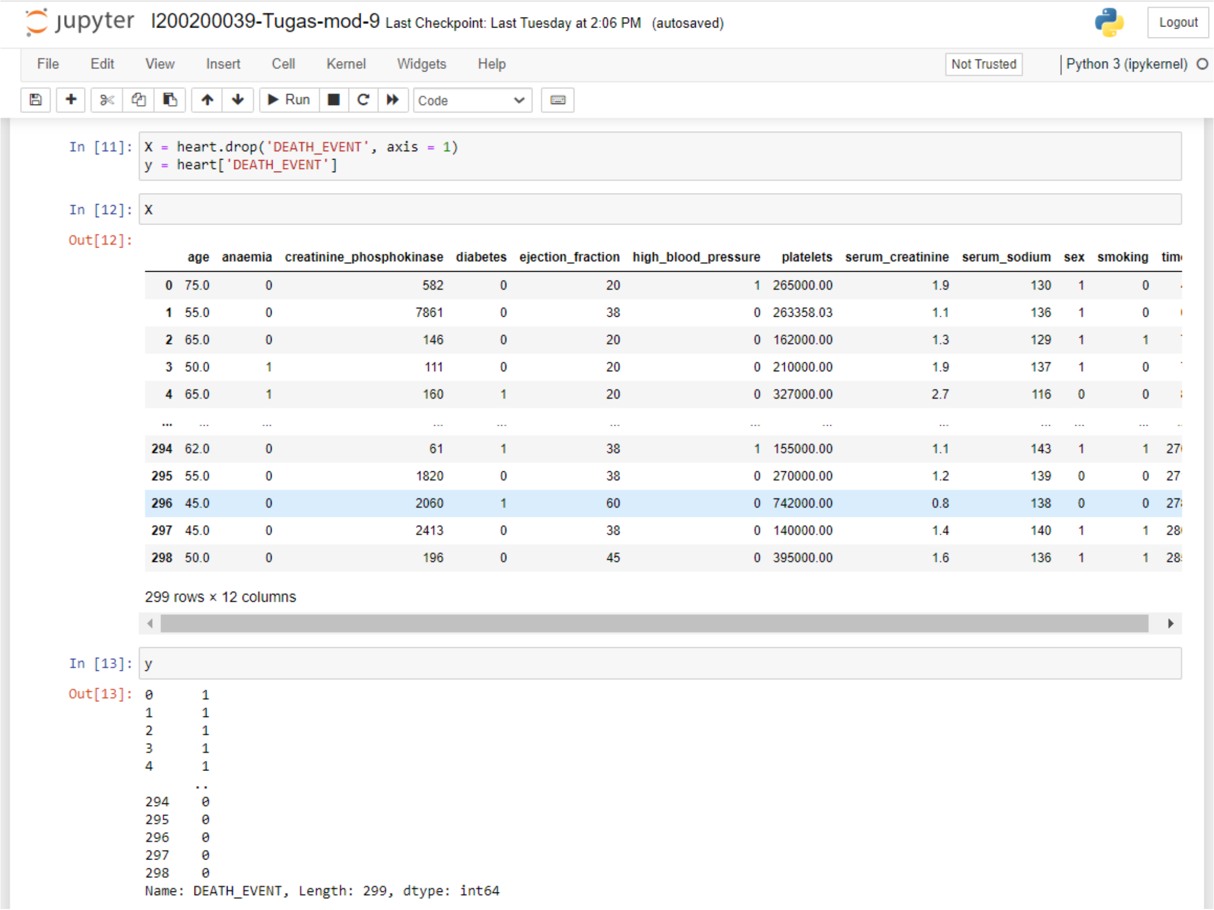
	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_
age	1.000000	0.088006	-0.081584	-0.101012	0.060098	0.093289	-0.052354	0.159187	-0
anaemia	0.088006	1.000000	-0.190741	-0.012729	0.031557	0.038182	-0.043786	0.052174	0
creatinine_phosphokinase	-0.081584	-0.190741	1.000000	-0.009639	-0.044080	-0.070590	0.024463	-0.016408	0
diabetes	-0.101012	-0.012729	-0.009639	1.000000	-0.004850	-0.012732	0.092193	-0.046975	-0
ejection_fraction	0.060098	0.031557	-0.044080	-0.004850	1.000000	0.024445	0.072177	-0.011302	0
high_blood_pressure	0.093289	0.038182	-0.070590	-0.012732	0.024445	1.000000	0.049963	-0.004935	0
platelets	-0.052354	-0.043786	0.024463	0.092193	0.072177	0.049963	1.000000	-0.041198	0
serum_creatinine	0.159187	0.052174	-0.016408	-0.046975	-0.011302	-0.004935	-0.041198	1.000000	-0
serum_sodium	-0.045966	0.041882	0.059550	-0.089551	0.175902	0.037109	0.062125	-0.189095	1
sex	0.065430	-0.094769	0.079791	-0.157730	-0.148386	-0.104615	-0.125120	0.006970	-0
smoking	0.018668	-0.107290	0.002421	-0.147173	-0.067315	-0.055711	0.028234	-0.027414	0
time	-0.224068	-0.141414	-0.009346	0.033726	0.041729	-0.196439	0.010514	-0.149315	0
DEATH_EVENT	0.253729	0.066270	0.062728	-0.001943	-0.268603	0.079351	-0.049139	0.294278	-0

➢



```
In [9]: sns.heatmap(heart.corr(), annot = True, cmap = 'Greens')
        plt.show()
```

➢



In [11]:
```
X = heart.drop('DEATH_EVENT', axis = 1)
y = heart['DEATH_EVENT']
```

In [12]:
```
X
```

Out[12]:

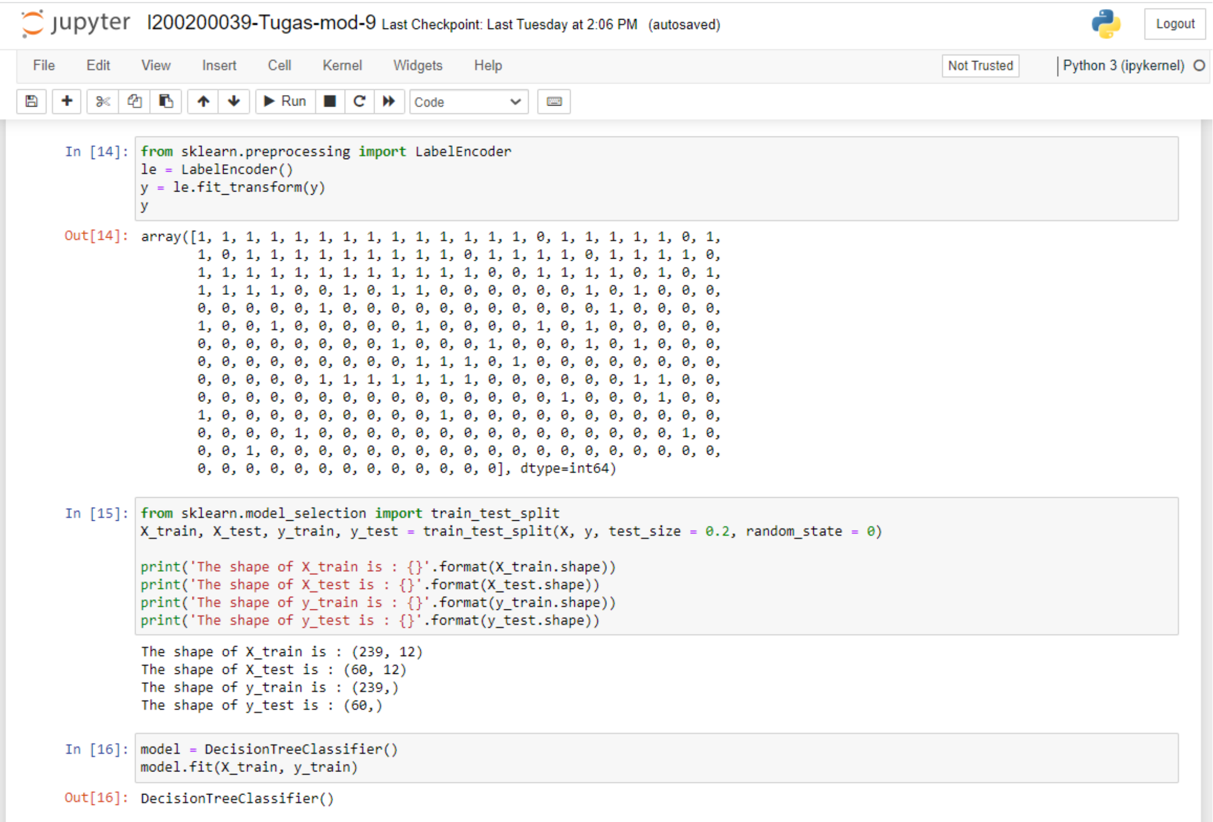
	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	tim
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...
294	62.0	0	61	1	38	1	155000.00	1.1	143	1	1	27
295	55.0	0	1820	0	38	0	270000.00	1.2	139	0	0	27
296	45.0	0	2060	1	60	0	742000.00	0.8	138	0	0	27
297	45.0	0	2413	0	38	0	140000.00	1.4	140	1	1	28
298	50.0	0	196	0	45	0	395000.00	1.6	136	1	1	28

299 rows × 12 columns

In [13]:
```
y
```

Out[13]:
```
0      1
1      1
2      1
3      1
4      1
      ..
294    0
295    0
296    0
297    0
298    0
Name: DEATH_EVENT, Length: 299, dtype: int64
```

➢


jupyter I200200039-Tugas-mod-9 Last Checkpoint: Last Tuesday at 2:06 PM (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Not Trusted
Python 3 (ipykernel)
Run
Code

In [14]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
y = le.fit_transform(y)
y
```

Out[14]:
```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1,
       1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1,
       1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 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, 0, 1, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int64)
```

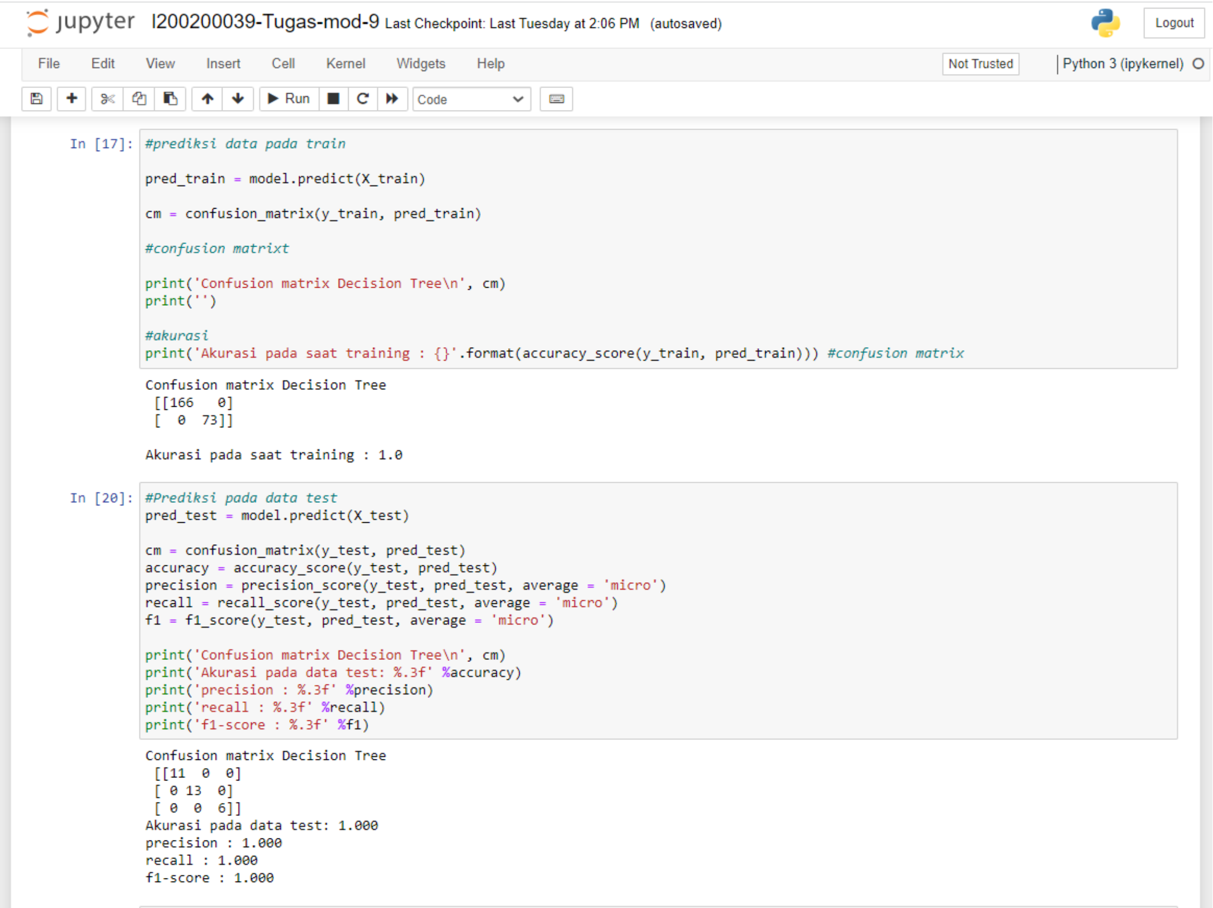
In [15]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)

print('The shape of X_train is : {}'.format(X_train.shape))
print('The shape of X_test is : {}'.format(X_test.shape))
print('The shape of y_train is : {}'.format(y_train.shape))
print('The shape of y_test is : {}'.format(y_test.shape))
```

```
The shape of X_train is : (239, 12)
The shape of X_test is : (60, 12)
The shape of y_train is : (239,)
The shape of y_test is : (60,)
```

In [16]:
```
model = DecisionTreeClassifier()
model.fit(X_train, y_train)
```

Out[16]:
```
DecisionTreeClassifier()
```

➢


jupyter I200200039-Tugas-mod-9 Last Checkpoint: Last Tuesday at 2:06 PM (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Not Trusted
Python 3 (ipykernel)
Run
Code

In [17]:
```
#prediksi data pada train

pred_train = model.predict(X_train)

cm = confusion_matrix(y_train, pred_train)

#confusion matrixt

print('Confusion matrix Decision Tree\n', cm)
print('')

#akurasi
print('Akurasi pada saat training : {}'.format(accuracy_score(y_train, pred_train))) #confusion matrix
```

```
Confusion matrix Decision Tree
 [[166   0]
 [  0  73]]

Akurasi pada saat training : 1.0
```

In [20]:
```
#Prediksi pada data test
pred_test = model.predict(X_test)

cm = confusion_matrix(y_test, pred_test)
accuracy = accuracy_score(y_test, pred_test)
precision = precision_score(y_test, pred_test, average = 'micro')
recall = recall_score(y_test, pred_test, average = 'micro')
f1 = f1_score(y_test, pred_test, average = 'micro')

print('Confusion matrix Decision Tree\n', cm)
print('Akurasi pada data test: %.3f' %accuracy)
print('precision : %.3f' %precision)
print('recall : %.3f' %recall)
print('f1-score : %.3f' %f1)
```

```
Confusion matrix Decision Tree
 [[11  0  0]
 [ 0 13  0]
 [ 0  0  6]]
Akurasi pada data test: 1.000
precision : 1.000
recall : 1.000
f1-score : 1.000
```