LAPORAN

PRAKTIKUM DATA WAREHOUSING DAN DATA MINING

(MODUL 9)

“ Tugas 9 ”

Disusun oleh :

NAMA : CINDI DILA APRILIANA

NIM : L200200106

KELAS : D

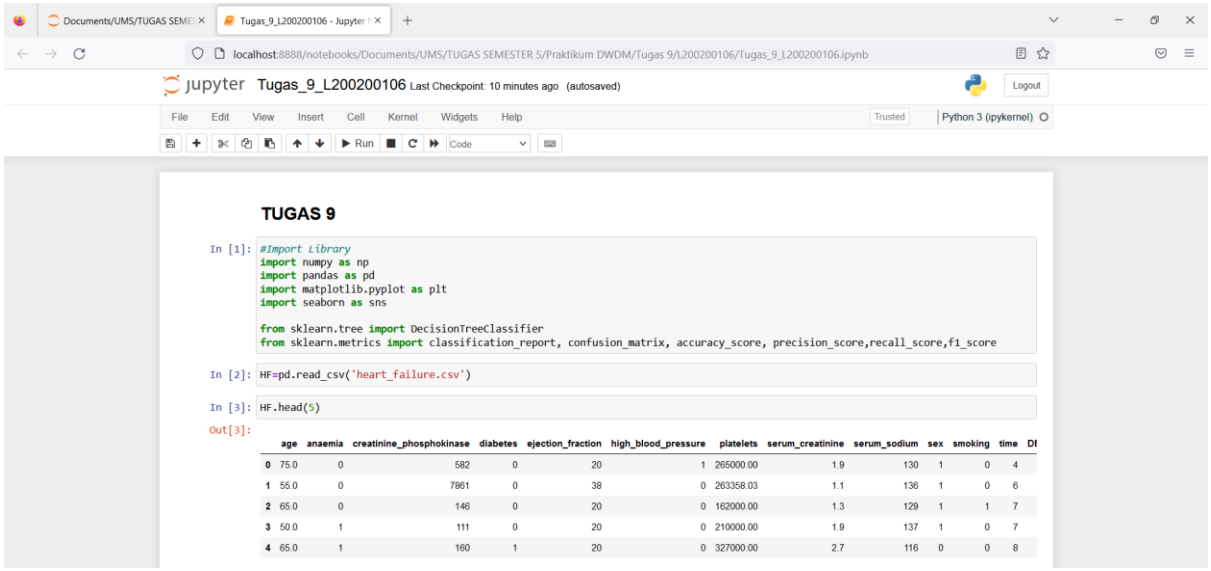
INFORMATIKA

FAKULTAS KOMUNIKASI DAN INFORMATIKA

UNIVERSITAS MUHAMMADIYAH SURAKARTA

TAHUN 2022/2023

1. Eksplorasi data dari dataset heart-failure.csv



TUGAS 9

In [1]:
```
#Import Library
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
```

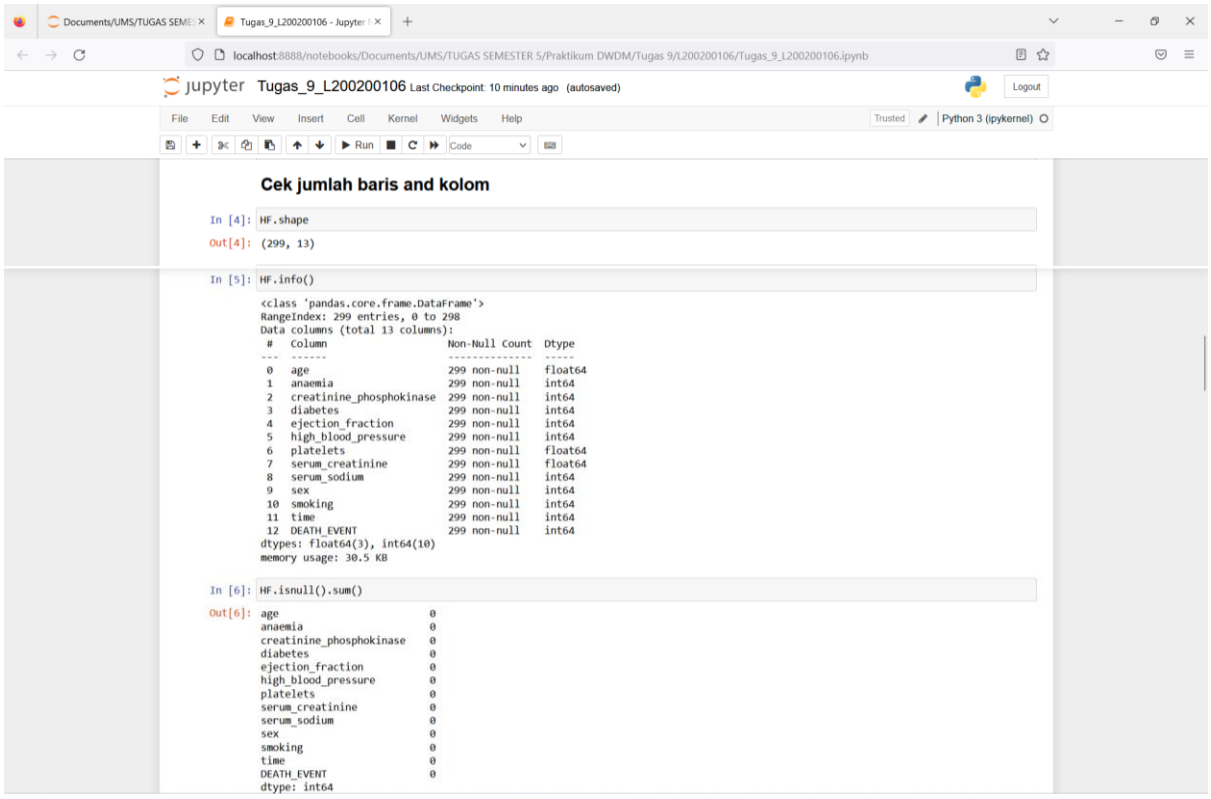
In [2]:
```
HF=pd.read_csv('heart_failure.csv')
```

In [3]:
```
HF.head(5)
```

Out[3]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DI
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	

2. Memeriksa apakah dataset memilki nilai kosong(null) atau tidak.



Cek jumlah baris and kolom

In [4]:
```
HF.shape
```

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

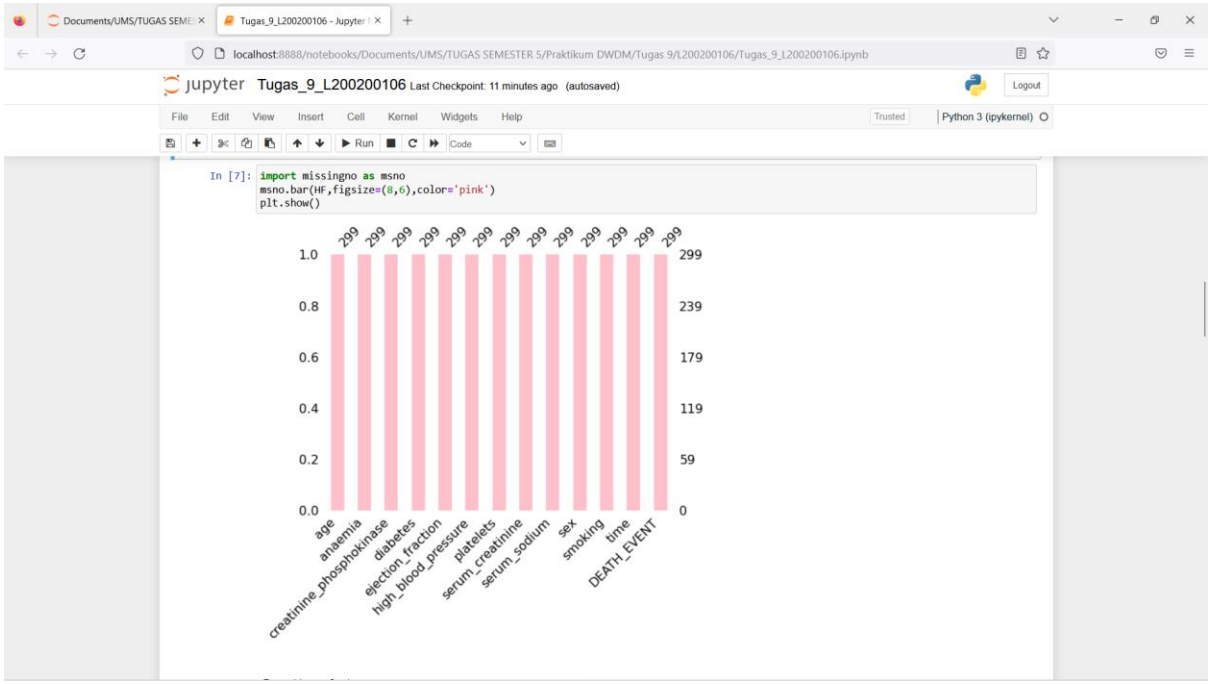
In [5]:
```
HF.info()
```

```
<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 [6]:
```
HF.isnull().sum()
```

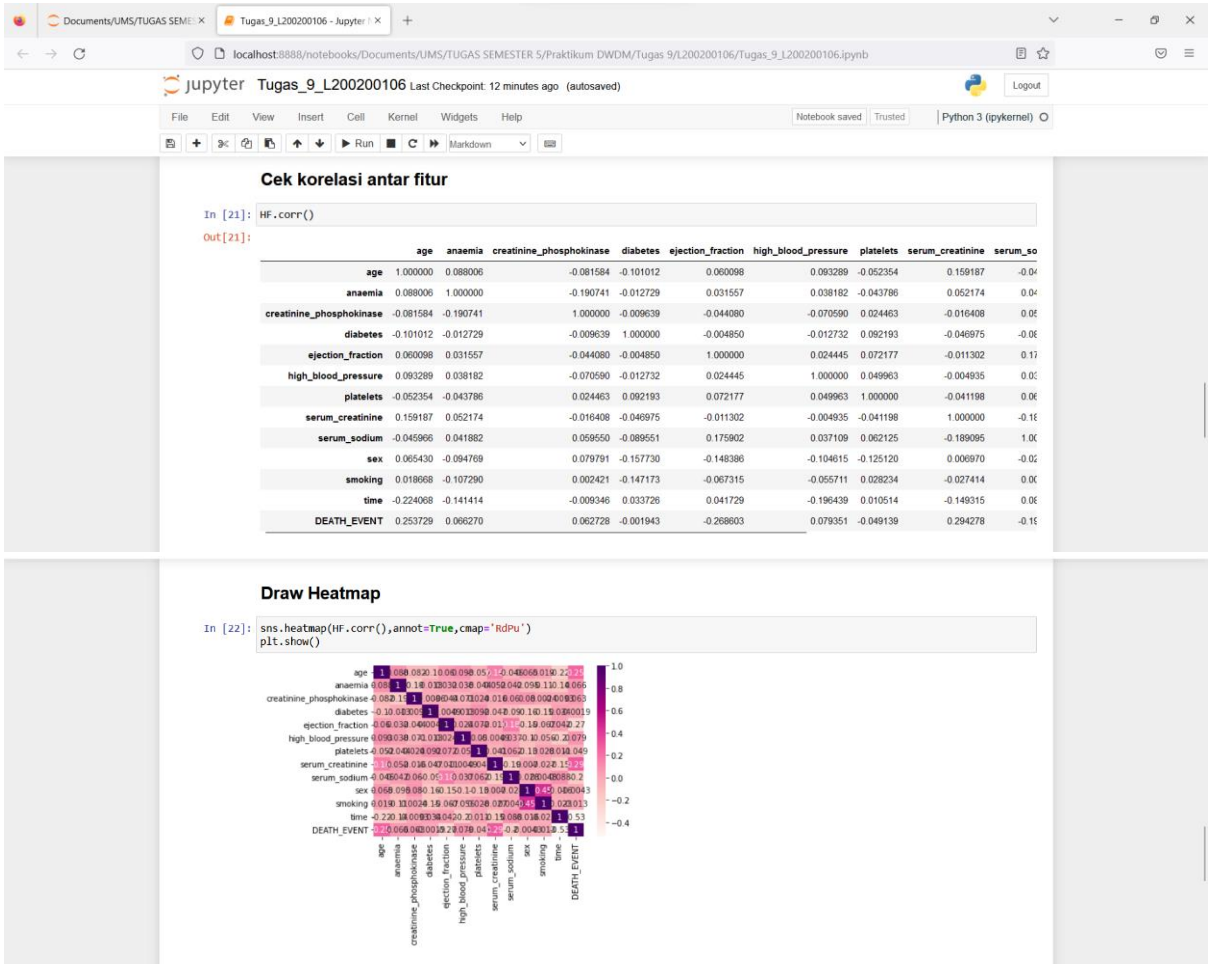
Out[6]:
```
age                         0
anaemia                     0
creatinine_phosphokinase    0
diabetes                    0
ejection_fraction           0
high_blood_pressure         0
platelets                   0
serum_creatinine            0
serum_sodium                0
sex                         0
smoking                     0
time                        0
DEATH_EVENT                 0
dtype: int64
```

3. Cek data kosong menggunakan grafik



4. Melakukan visualisasi data

5. Mencari korelasi pada setiap fitur yang digunakan untuk melakukan klasifikasi



Cek korelasi antar fitur

```
In [21]: HF.corr()
```

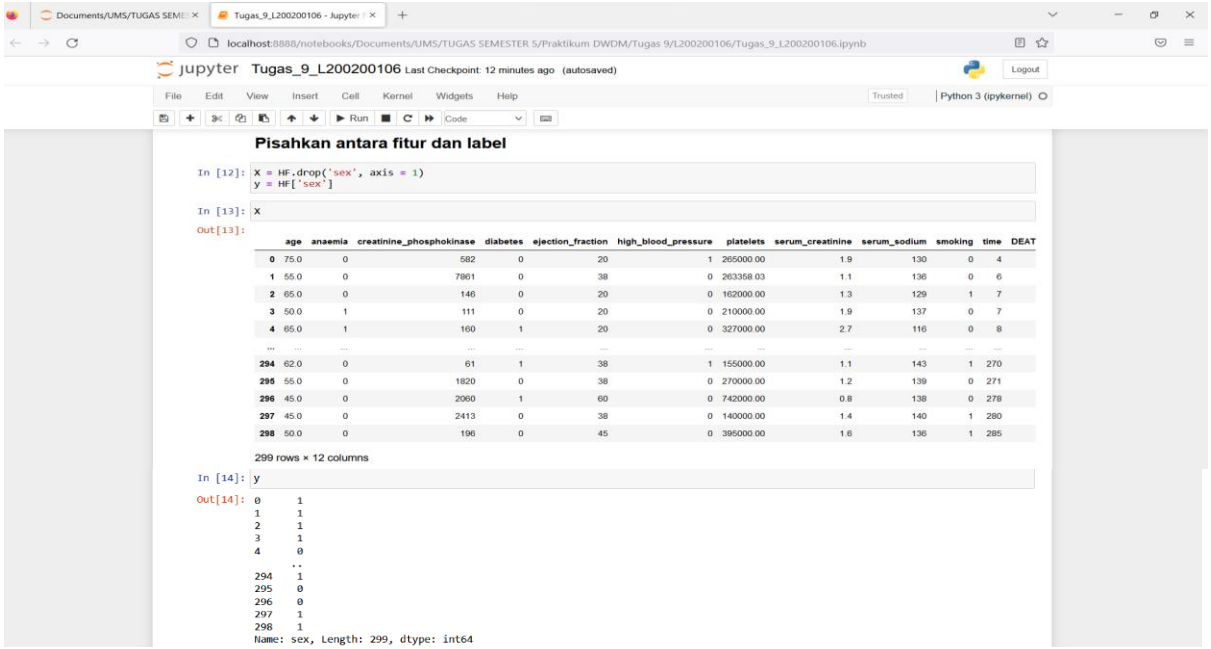
Out[21]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_so
age	1.000000	0.088006	-0.081584	-0.101012	0.060098	0.093289	-0.052354	0.159187	-0.0
anaemia	0.088006	1.000000	-0.190741	-0.012729	0.031557	0.038182	-0.043786	0.052174	0.0
creatinine_phosphokinase	-0.081584	-0.190741	1.000000	-0.009639	-0.044080	-0.070590	0.024463	-0.016408	0.0
diabetes	-0.101012	-0.012729	-0.009639	1.000000	-0.004850	-0.012732	0.092193	-0.046975	-0.0
ejection_fraction	0.060098	0.031557	-0.044080	-0.004850	1.000000	0.024445	0.072177	-0.011302	0.1
high_blood_pressure	0.093289	0.038182	-0.070590	-0.012732	0.024445	1.000000	0.049963	-0.004935	0.0
platelets	-0.052354	-0.043786	0.024463	0.092193	0.072177	0.049963	1.000000	-0.041198	0.0
serum_creatinine	0.159187	0.052174	-0.016408	-0.046975	-0.011302	-0.004935	-0.041198	1.000000	-0.1
serum_sodium	-0.045966	0.041882	0.059550	-0.089551	0.175902	0.037109	0.062125	-0.189095	1.0
sex	0.065430	-0.094769	0.079791	-0.157730	-0.148386	-0.104615	-0.125120	0.006970	-0.0
smoking	0.018668	-0.107290	0.002421	-0.147173	-0.067315	-0.055711	0.028234	-0.027414	0.0
time	-0.224068	-0.141414	-0.009346	0.033726	0.041729	-0.196439	0.010514	-0.149315	0.0
DEATH_EVENT	0.253729	0.066270	0.062728	-0.001943	-0.268603	0.079351	-0.049139	0.294278	-0.1

Draw Heatmap

```
In [22]: sns.heatmap(HF.corr(),annot=True,cmap='RdPu')
         plt.show()
```


6. Cek atribut yang sudah terpisah dengan menampilkan variable X yang merupakan fitur dan variable y sebagai label



Pisahkan antara fitur dan label

```
In [12]: X = HF.drop('sex', axis = 1)
         y = HF['sex']
```

```
In [13]: X
```

Out[13]:

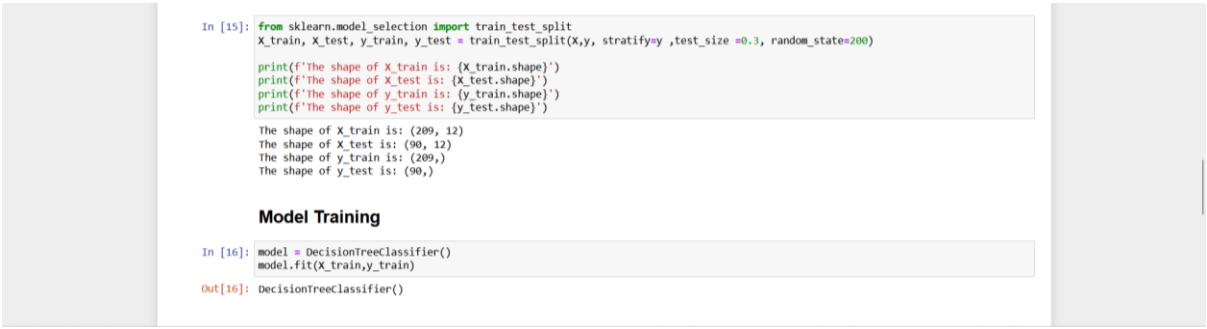
	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	smoking	time	DEAT
0	75.0	0	582	0	20	1	265000.00	1.9	130	0	4	
1	55.0	0	7861	0	38	0	263358.03	1.1	136	0	6	
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	7	
3	50.0	1	111	0	20	0	210000.00	1.9	137	0	7	
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	8	
...	...	...	...	...	...	...	...	...	...	...	...	
294	62.0	0	61	1	38	1	155000.00	1.1	143	1	270	
295	55.0	0	1820	0	38	0	270000.00	1.2	139	0	271	
296	45.0	0	2060	1	60	0	742000.00	0.8	138	0	278	
297	45.0	0	2413	0	38	0	140000.00	1.4	140	1	280	
298	50.0	0	196	0	45	0	395000.00	1.6	136	1	285	

299 rows × 12 columns

```
In [14]: y
```

```
Out[14]: 0      1
         1      1
         2      1
         3      1
         4      0
               ..
         294    1
         295    0
         296    0
         297    1
         298    1
         Name: sex, Length: 299, dtype: int64
```

7. Pembuatan model dengan menggunakan algoritma naive bayes

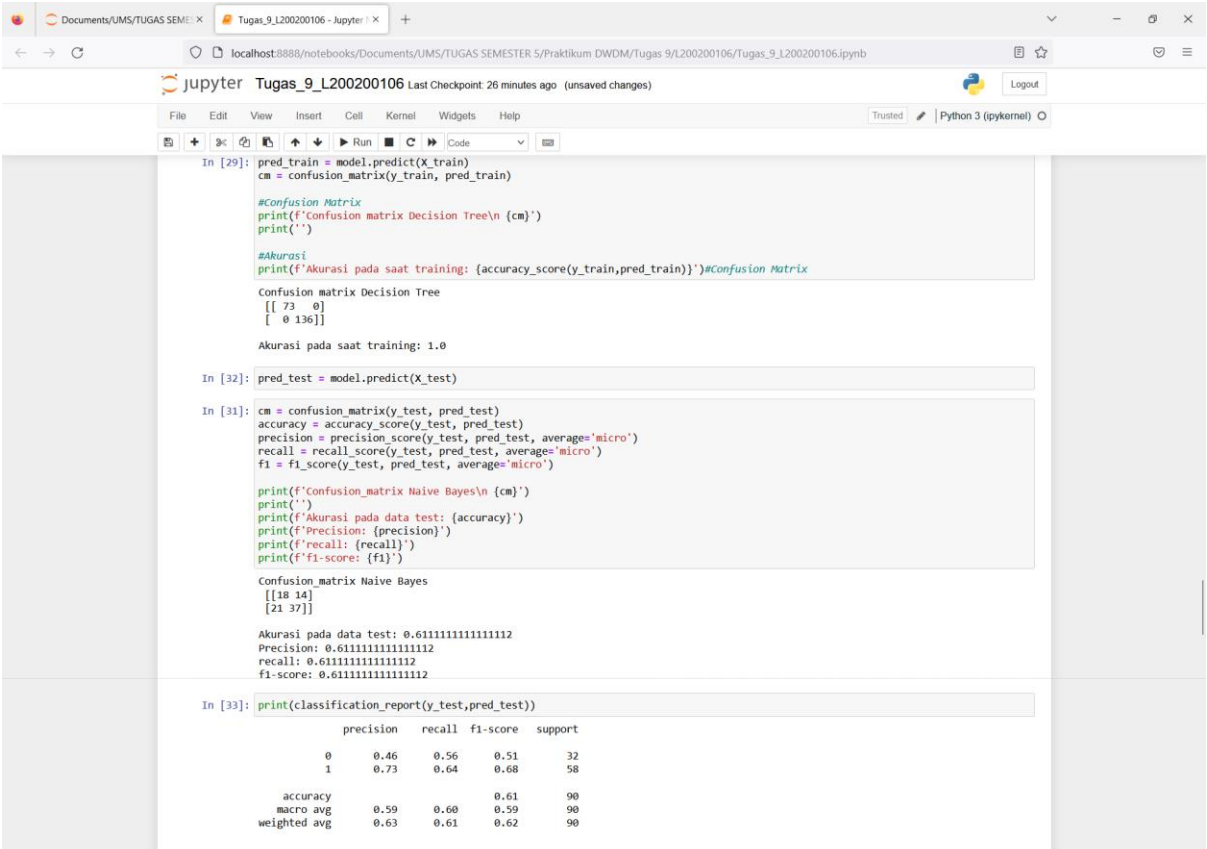


```
In [15]: from sklearn.model_selection import train_test_split
         X_train, X_test, y_train, y_test = train_test_split(X,y, stratify=y ,test_size =0.3, random_state=200)

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

         The shape of X_train is: (209, 12)
         The shape of X_test is: (90, 12)
         The shape of y_train is: (209,)
         The shape of y_test is: (90,)
```

Model Training

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

Out[16]: DecisionTreeClassifier()
```

8. Melakukana training model dan Menguji model yang dibuat dengan data testing



```
In [29]: pred_train = model.predict(X_train)
         cm = confusion_matrix(y_train, pred_train)

         #Confusion Matrix
         print(f'Confusion matrix Decision Tree\n {cm}')
         print('')

         #Akurasi
         print(f'Akurasi pada saat training: {accuracy_score(y_train,pred_train)}')#Confusion Matrix

         Confusion matrix Decision Tree
          [[ 73   0]
          [  0 136]]

         Akurasi pada saat training: 1.0

In [32]: pred_test = model.predict(X_test)

In [31]: 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(f'Confusion_matrix Naive Bayes\n {cm}')
         print('')
         print(f'Akurasi pada data test: {accuracy}')
         print(f'Precision: {precision}')
         print(f'recall: {recall}')
         print(f'f1-score: {f1}')

         Confusion_matrix Naive Bayes
          [[18 14]
          [21 37]]

         Akurasi pada data test: 0.6111111111111112
         Precision: 0.6111111111111112
         recall: 0.6111111111111112
         f1-score: 0.6111111111111112

In [33]: print(classification_report(y_test,pred_test))

                       precision    recall  f1-score   support

                    0       0.46      0.56      0.51        32
                    1       0.73      0.64      0.68        58

             accuracy                           0.61        90
            macro avg       0.59      0.60      0.59        90
         weighted avg       0.63      0.61      0.62        90
```