Assignment6

#ASSIGNMENT NO:-6

Data Analytics III 1. Implement Simple Naïve Bayes classification algorithm using Python/R on iris.csv dataset. 2. Compute Confusion matrix to find TP, FP, TN, FN, Accuracy, Error rate, Precision, Recall on the given dataset.

Import libraries and create alias for Pandas, Numpy

[1]:
```
import pandas as pd
import numpy as np
```

Import the Iris Dataset

[2]:
```
from google.colab import files
files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving Iris.csv to Iris.csv
```

[2]:

```
{'Iris.csv': b'Id,SepalLengthCm,SepalWidthCm,PetalLengthCm,PetalWidthCm,Species\
n1,5.1,3.5,1.4,0.2,Iris-setosa\n2,4.9,3.0,1.4,0.2,Iris-
setosa\n3,4.7,3.2,1.3,0.2,Iris-setosa\n4,4.6,3.1,1.5,0.2,Iris-
setosa\n5,5.0,3.6,1.4,0.2,Iris-setosa\n6,5.4,3.9,1.7,0.4,Iris-
setosa\n7,4.6,3.4,1.4,0.3,Iris-setosa\n8,5.0,3.4,1.5,0.2,Iris-
setosa\n9,4.4,2.9,1.4,0.2,Iris-setosa\n10,4.9,3.1,1.5,0.1,Iris-
setosa\n11,5.4,3.7,1.5,0.2,Iris-setosa\n12,4.8,3.4,1.6,0.2,Iris-
setosa\n13,4.8,3.0,1.4,0.1,Iris-setosa\n14,4.3,3.0,1.1,0.1,Iris-
setosa\n15,5.8,4.0,1.2,0.2,Iris-setosa\n16,5.7,4.4,1.5,0.4,Iris-
setosa\n17,5.4,3.9,1.3,0.4,Iris-setosa\n18,5.1,3.5,1.4,0.3,Iris-
setosa\n19,5.7,3.8,1.7,0.3,Iris-setosa\n20,5.1,3.8,1.5,0.3,Iris-
setosa\n21,5.4,3.4,1.7,0.2,Iris-setosa\n22,5.1,3.7,1.5,0.4,Iris-
setosa\n23,4.6,3.6,1.0,0.2,Iris-setosa\n24,5.1,3.3,1.7,0.5,Iris-
setosa\n25,4.8,3.4,1.9,0.2,Iris-setosa\n26,5.0,3.0,1.6,0.2,Iris-
setosa\n27,5.0,3.4,1.6,0.4,Iris-setosa\n28,5.2,3.5,1.5,0.2,Iris-
setosa\n29,5.2,3.4,1.4,0.2,Iris-setosa\n30,4.7,3.2,1.6,0.2,Iris-
setosa\n31,4.8,3.1,1.6,0.2,Iris-setosa\n32,5.4,3.4,1.5,0.4,Iris-
setosa\n33,5.2,4.1,1.5,0.1,Iris-setosa\n34,5.5,4.2,1.4,0.2,Iris-
```



```
setosa\n35,4.9,3.1,1.5,0.1,Iris-setosa\n36,5.0,3.2,1.2,0.2,Iris-
setosa\n37,5.5,3.5,1.3,0.2,Iris-setosa\n38,4.9,3.1,1.5,0.1,Iris-
setosa\n39,4.4,3.0,1.3,0.2,Iris-setosa\n40,5.1,3.4,1.5,0.2,Iris-
setosa\n41,5.0,3.5,1.3,0.3,Iris-setosa\n42,4.5,2.3,1.3,0.3,Iris-
setosa\n43,4.4,3.2,1.3,0.2,Iris-setosa\n44,5.0,3.5,1.6,0.6,Iris-
setosa\n45,5.1,3.8,1.9,0.4,Iris-setosa\n46,4.8,3.0,1.4,0.3,Iris-
setosa\n47,5.1,3.8,1.6,0.2,Iris-setosa\n48,4.6,3.2,1.4,0.2,Iris-
setosa\n49,5.3,3.7,1.5,0.2,Iris-setosa\n50,5.0,3.3,1.4,0.2,Iris-
setosa\n51,7.0,3.2,4.7,1.4,Iris-versicolor\n52,6.4,3.2,4.5,1.5,Iris-
versicolor\n53,6.9,3.1,4.9,1.5,Iris-versicolor\n54,5.5,2.3,4.0,1.3,Iris-
versicolor\n55,6.5,2.8,4.6,1.5,Iris-versicolor\n56,5.7,2.8,4.5,1.3,Iris-
versicolor\n57,6.3,3.3,4.7,1.6,Iris-versicolor\n58,4.9,2.4,3.3,1.0,Iris-
versicolor\n59,6.6,2.9,4.6,1.3,Iris-versicolor\n60,5.2,2.7,3.9,1.4,Iris-
versicolor\n61,5.0,2.0,3.5,1.0,Iris-versicolor\n62,5.9,3.0,4.2,1.5,Iris-
versicolor\n63,6.0,2.2,4.0,1.0,Iris-versicolor\n64,6.1,2.9,4.7,1.4,Iris-
versicolor\n65,5.6,2.9,3.6,1.3,Iris-versicolor\n66,6.7,3.1,4.4,1.4,Iris-
versicolor\n67,5.6,3.0,4.5,1.5,Iris-versicolor\n68,5.8,2.7,4.1,1.0,Iris-
versicolor\n69,6.2,2.2,4.5,1.5,Iris-versicolor\n70,5.6,2.5,3.9,1.1,Iris-
versicolor\n71,5.9,3.2,4.8,1.8,Iris-versicolor\n72,6.1,2.8,4.0,1.3,Iris-
versicolor\n73,6.3,2.5,4.9,1.5,Iris-versicolor\n74,6.1,2.8,4.7,1.2,Iris-
versicolor\n75,6.4,2.9,4.3,1.3,Iris-versicolor\n76,6.6,3.0,4.4,1.4,Iris-
versicolor\n77,6.8,2.8,4.8,1.4,Iris-versicolor\n78,6.7,3.0,5.0,1.7,Iris-
versicolor\n79,6.0,2.9,4.5,1.5,Iris-versicolor\n80,5.7,2.6,3.5,1.0,Iris-
versicolor\n81,5.5,2.4,3.8,1.1,Iris-versicolor\n82,5.5,2.4,3.7,1.0,Iris-
versicolor\n83,5.8,2.7,3.9,1.2,Iris-versicolor\n84,6.0,2.7,5.1,1.6,Iris-
versicolor\n85,5.4,3.0,4.5,1.5,Iris-versicolor\n86,6.0,3.4,4.5,1.6,Iris-
versicolor\n87,6.7,3.1,4.7,1.5,Iris-versicolor\n88,6.3,2.3,4.4,1.3,Iris-
versicolor\n89,5.6,3.0,4.1,1.3,Iris-versicolor\n90,5.5,2.5,4.0,1.3,Iris-
versicolor\n91,5.5,2.6,4.4,1.2,Iris-versicolor\n92,6.1,3.0,4.6,1.4,Iris-
versicolor\n93,5.8,2.6,4.0,1.2,Iris-versicolor\n94,5.0,2.3,3.3,1.0,Iris-
versicolor\n95,5.6,2.7,4.2,1.3,Iris-versicolor\n96,5.7,3.0,4.2,1.2,Iris-
versicolor\n97,5.7,2.9,4.2,1.3,Iris-versicolor\n98,6.2,2.9,4.3,1.3,Iris-
versicolor\n99,5.1,2.5,3.0,1.1,Iris-versicolor\n100,5.7,2.8,4.1,1.3,Iris-
versicolor\n101,6.3,3.3,6.0,2.5,Iris-virginica\n102,5.8,2.7,5.1,1.9,Iris-
virginica\n103,7.1,3.0,5.9,2.1,Iris-virginica\n104,6.3,2.9,5.6,1.8,Iris-
virginica\n105,6.5,3.0,5.8,2.2,Iris-virginica\n106,7.6,3.0,6.6,2.1,Iris-
virginica\n107,4.9,2.5,4.5,1.7,Iris-virginica\n108,7.3,2.9,6.3,1.8,Iris-
virginica\n109,6.7,2.5,5.8,1.8,Iris-virginica\n110,7.2,3.6,6.1,2.5,Iris-
virginica\n111,6.5,3.2,5.1,2.0,Iris-virginica\n112,6.4,2.7,5.3,1.9,Iris-
virginica\n113,6.8,3.0,5.5,2.1,Iris-virginica\n114,5.7,2.5,5.0,2.0,Iris-
virginica\n115,5.8,2.8,5.1,2.4,Iris-virginica\n116,6.4,3.2,5.3,2.3,Iris-
virginica\n117,6.5,3.0,5.5,1.8,Iris-virginica\n118,7.7,3.8,6.7,2.2,Iris-
virginica\n119,7.7,2.6,6.9,2.3,Iris-virginica\n120,6.0,2.2,5.0,1.5,Iris-
virginica\n121,6.9,3.2,5.7,2.3,Iris-virginica\n122,5.6,2.8,4.9,2.0,Iris-
virginica\n123,7.7,2.8,6.7,2.0,Iris-virginica\n124,6.3,2.7,4.9,1.8,Iris-
virginica\n125,6.7,3.3,5.7,2.1,Iris-virginica\n126,7.2,3.2,6.0,1.8,Iris-
virginica\n127,6.2,2.8,4.8,1.8,Iris-virginica\n128,6.1,3.0,4.9,1.8,Iris-
```

virginica\n129,6.4,2.8,5.6,2.1,Iris-virginica\n130,7.2,3.0,5.8,1.6,Iris-
virginica\n131,7.4,2.8,6.1,1.9,Iris-virginica\n132,7.9,3.8,6.4,2.0,Iris-
virginica\n133,6.4,2.8,5.6,2.2,Iris-virginica\n134,6.3,2.8,5.1,1.5,Iris-
virginica\n135,6.1,2.6,5.6,1.4,Iris-virginica\n136,7.7,3.0,6.1,2.3,Iris-
virginica\n137,6.3,3.4,5.6,2.4,Iris-virginica\n138,6.4,3.1,5.5,1.8,Iris-
virginica\n139,6.0,3.0,4.8,1.8,Iris-virginica\n140,6.9,3.1,5.4,2.1,Iris-
virginica\n141,6.7,3.1,5.6,2.4,Iris-virginica\n142,6.9,3.1,5.1,2.3,Iris-
virginica\n143,5.8,2.7,5.1,1.9,Iris-virginica\n144,6.8,3.2,5.9,2.3,Iris-
virginica\n145,6.7,3.3,5.7,2.5,Iris-virginica\n146,6.7,3.0,5.2,2.3,Iris-
virginica\n147,6.3,2.5,5.0,1.9,Iris-virginica\n148,6.5,3.0,5.2,2.0,Iris-
virginica\n149,6.2,3.4,5.4,2.3,Iris-virginica\n150,5.9,3.0,5.1,1.8,Iris-
virginica\n'}

Initialize the data frame

[3]:
```
df=pd.read_csv("Iris.csv")
```

[4]:
```
df.head()
```

[4]:
```
   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      Species
0   1            5.1           3.5            1.4           0.2  Iris-setosa
1   2            4.9           3.0            1.4           0.2  Iris-setosa
2   3            4.7           3.2            1.3           0.2  Iris-setosa
3   4            4.6           3.1            1.5           0.2  Iris-setosa
4   5            5.0           3.6            1.4           0.2  Iris-setosa
```

[ ]:
```
df.tail()
```

[7]:
```
df.describe()
```

[7]:
```
               Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
count  150.000000     150.000000    150.000000     150.000000    150.000000
mean    75.500000       5.843333      3.054000       3.758667      1.198667
std     43.445368       0.828066      0.433594       1.764420      0.763161
min      1.000000       4.300000      2.000000       1.000000      0.100000
25%     38.250000       5.100000      2.800000       1.600000      0.300000
50%     75.500000       5.800000      3.000000       4.350000      1.300000
75%    112.750000       6.400000      3.300000       5.100000      1.800000
max    150.000000       7.900000      4.400000       6.900000      2.500000
```

Identification of Null Values

[6]:
```
df.isnull()
```

[6]:
```
        Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
0    False          False         False          False         False    False
1    False          False         False          False         False    False
2    False          False         False          False         False    False
```



```
3    False          False         False          False         False    False
4    False          False         False          False         False    False
..     …              …             …              …             …        …
145  False          False         False          False         False    False
146  False          False         False          False         False    False
147  False          False         False          False         False    False
148  False          False         False          False         False    False
149  False          False         False          False         False    False

[150 rows x 6 columns]
```

[8]:
```
df.isna()
```

[8]:
```
        Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
0    False          False         False          False         False    False
1    False          False         False          False         False    False
2    False          False         False          False         False    False
3    False          False         False          False         False    False
4    False          False         False          False         False    False
..     …              …             …              …             …        …
145  False          False         False          False         False    False
146  False          False         False          False         False    False
147  False          False         False          False         False    False
148  False          False         False          False         False    False
149  False          False         False          False         False    False

[150 rows x 6 columns]
```

[9]:
```
df.isnull().sum()
```

[9]:
```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

[10]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Id             150 non-null    int64
 1   SepalLengthCm  150 non-null    float64
```

```
 2   SepalWidthCm   150 non-null    float64
 3   PetalLengthCm  150 non-null    float64
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

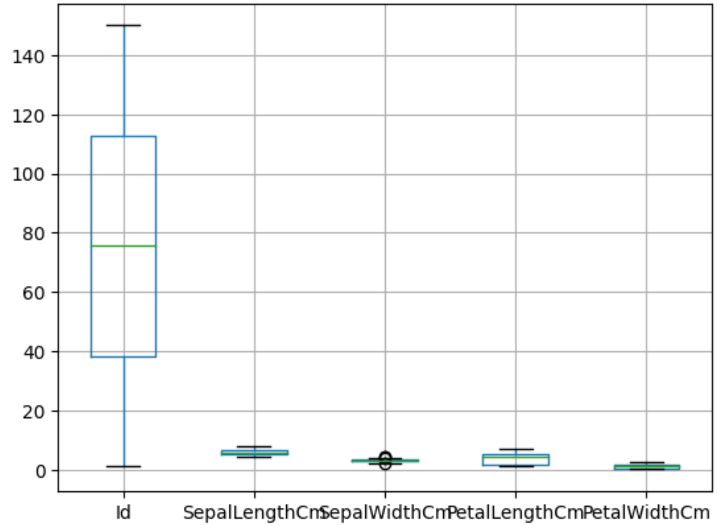
Detecting Outliers

Detecting outliers using Boxplot:

[12]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

[13]:
```
df.boxplot()
```

[13]: `<Axes: >`



Use Naive Bayes algorithm(Train the Machine) to Create Model

[11]:
```
X = df.drop(['Species'], axis = 1)
Y = df['Species']
```


[14]:
```
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(X, Y, test_size =0.
↪2,random_state = 0)
```

[15]:
```
from sklearn.naive_bayes import GaussianNB
gaussian = GaussianNB()
```

[16]:
```
gaussian.fit(xtrain, ytrain)
```

[16]: GaussianNB()

Predict the y__pred for all values of train__x and test__x

[17]:
```
y_pred_xtest = gaussian.predict(xtest)
```

[18]:
```
y_pred_xtrain = gaussian.predict(xtrain)
```

[19]:
```
print(xtrain)
print("------------\n")
print(xtest)
print("------------\n")
print(ytrain)
print("------------\n")
print(ytest)
print("------------\n")
print(y_pred_xtest)
print("------------\n")
print(y_pred_xtrain)
```

```
      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
137  138            6.4           3.1            5.5           1.8
84    85            5.4           3.0            4.5           1.5
27    28            5.2           3.5            1.5           0.2
127  128            6.1           3.0            4.9           1.8
132  133            6.4           2.8            5.6           2.2
..   …              …             …              …             …
9     10            4.9           3.1            1.5           0.1
103  104            6.3           2.9            5.6           1.8
67    68            5.8           2.7            4.1           1.0
117  118            7.7           3.8            6.7           2.2
47    48            4.6           3.2            1.4           0.2

[120 rows x 5 columns]
------------

      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
114  115            5.8           2.8            5.1           2.4
62    63            6.0           2.2            4.0           1.0
```

```
33   34           5.5           4.2           1.4           0.2
107  108          7.3           2.9           6.3           1.8
7    8            5.0           3.4           1.5           0.2
100  101          6.3           3.3           6.0           2.5
40   41           5.0           3.5           1.3           0.3
86   87           6.7           3.1           4.7           1.5
76   77           6.8           2.8           4.8           1.4
71   72           6.1           2.8           4.0           1.3
134  135          6.1           2.6           5.6           1.4
51   52           6.4           3.2           4.5           1.5
73   74           6.1           2.8           4.7           1.2
54   55           6.5           2.8           4.6           1.5
63   64           6.1           2.9           4.7           1.4
37   38           4.9           3.1           1.5           0.1
78   79           6.0           2.9           4.5           1.5
90   91           5.5           2.6           4.4           1.2
45   46           4.8           3.0           1.4           0.3
16   17           5.4           3.9           1.3           0.4
121  122          5.6           2.8           4.9           2.0
66   67           5.6           3.0           4.5           1.5
24   25           4.8           3.4           1.9           0.2
8    9            4.4           2.9           1.4           0.2
126  127          6.2           2.8           4.8           1.8
22   23           4.6           3.6           1.0           0.2
44   45           5.1           3.8           1.9           0.4
97   98           6.2           2.9           4.3           1.3
93   94           5.0           2.3           3.3           1.0
26   27           5.0           3.4           1.6           0.4
------------

137     Iris-virginica
84     Iris-versicolor
27         Iris-setosa
127     Iris-virginica
132     Iris-virginica
            …
9          Iris-setosa
103     Iris-virginica
67     Iris-versicolor
117     Iris-virginica
47         Iris-setosa
Name: Species, Length: 120, dtype: object
------------

114     Iris-virginica
62     Iris-versicolor
33         Iris-setosa
107     Iris-virginica
7          Iris-setosa
100     Iris-virginica
40         Iris-setosa
86     Iris-versicolor
76     Iris-versicolor
71     Iris-versicolor
134     Iris-virginica
51     Iris-versicolor
73     Iris-versicolor
54     Iris-versicolor
63     Iris-versicolor
37         Iris-setosa
78     Iris-versicolor
90     Iris-versicolor
45         Iris-setosa
16         Iris-setosa
121     Iris-virginica
66     Iris-versicolor
24         Iris-setosa
8          Iris-setosa
126     Iris-virginica
22         Iris-setosa
44         Iris-setosa
97     Iris-versicolor
93     Iris-versicolor
26         Iris-setosa
Name: Species, dtype: object
------------

['Iris-virginica' 'Iris-versicolor' 'Iris-setosa' 'Iris-virginica'
 'Iris-setosa' 'Iris-virginica' 'Iris-setosa' 'Iris-versicolor'
 'Iris-versicolor' 'Iris-versicolor' 'Iris-virginica' 'Iris-versicolor'
 'Iris-versicolor' 'Iris-versicolor' 'Iris-versicolor' 'Iris-setosa'
 'Iris-versicolor' 'Iris-versicolor' 'Iris-setosa' 'Iris-setosa'
 'Iris-virginica' 'Iris-versicolor' 'Iris-setosa' 'Iris-setosa'
 'Iris-virginica' 'Iris-setosa' 'Iris-setosa' 'Iris-versicolor'
 'Iris-versicolor' 'Iris-setosa']
------------
```

~~['Iris-virginica' 'Iris-versicolor' 'Iris-setosa' 'Iris-virginica'~~
~~'Iris-virginica' 'Iris-versicolor' 'Iris-setosa' 'Iris-versicolor'~~
~~'Iris-versicolor' 'Iris-versicolor' 'Iris-virginica' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-setosa' 'Iris-setosa' 'Iris-versicolor'~~
~~'Iris-virginica' 'Iris-virginica' 'Iris-virginica' 'Iris-virginica'~~
~~'Iris-versicolor' 'Iris-virginica' 'Iris-versicolor' 'Iris-versicolor'~~
~~'Iris-virginica' 'Iris-virginica' 'Iris-versicolor' 'Iris-virginica'~~
~~'Iris-versicolor' 'Iris-virginica' 'Iris-versicolor' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-versicolor' 'Iris-versicolor' 'Iris-versicolor'~~

~~'Iris-versicolor' 'Iris-virginica' 'Iris-setosa' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-versicolor' 'Iris-setosa' 'Iris-setosa'~~
~~'Iris-versicolor' 'Iris-setosa' 'Iris-virginica' 'Iris-versicolor'~~
~~'Iris-setosa' 'Iris-versicolor' 'Iris-virginica' 'Iris-versicolor'~~
~~'Iris-setosa' 'Iris-virginica' 'Iris-virginica' 'Iris-virginica'~~
~~'Iris-virginica' 'Iris-setosa' 'Iris-setosa' 'Iris-virginica'~~
~~'Iris-virginica' 'Iris-setosa' 'Iris-virginica' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-virginica' 'Iris-setosa' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa'~~
~~'Iris-versicolor' 'Iris-virginica' 'Iris-virginica' 'Iris-setosa'~~
~~'Iris-setosa' 'Iris-setosa' 'Iris-versicolor' 'Iris-versicolor'~~
~~'Iris-setosa' 'Iris-setosa' 'Iris-versicolor' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-versicolor' 'Iris-virginica' 'Iris-versicolor'~~
~~'Iris-setosa' 'Iris-virginica' 'Iris-setosa' 'Iris-virginica'~~
~~'Iris-setosa' 'Iris-setosa' 'Iris-virginica' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-versicolor' 'Iris-versicolor' 'Iris-versicolor'~~
~~'Iris-virginica' 'Iris-virginica' 'Iris-versicolor' 'Iris-versicolor'~~
~~'Iris-setosa' 'Iris-versicolor' 'Iris-virginica' 'Iris-virginica'~~
~~'Iris-setosa' 'Iris-versicolor' 'Iris-versicolor' 'Iris-versicolor'~~
~~'Iris-versicolor' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa'~~
~~'Iris-virginica' 'Iris-versicolor' 'Iris-virginica' 'Iris-setosa']~~

Evaluate the performance of Model for train_y and test_y

[20]:
```
from sklearn.metrics import␣
↪precision_score,confusion_matrix,accuracy_score,recall_score,classification_report
```

Confusion Matrix

[21]:
```
cm= confusion_matrix(ytest, y_pred_xtest)
cm
```

[21]:
```
array([[11,  0,  0],
       [ 0, 13,  0],
       [ 0,  0,  6]])
```

Accuracy Score

[25]:
```
print ("Accuracy : ", accuracy_score(ytest, y_pred_xtest))
```

```
Accuracy :  1.0
```

Error Rate

[26]:
```
error_rate = 1- accuracy_score(ytest, y_pred_xtest)
error_rate
```

[26]:
```
0.0
```

Classification Report

[27]:
```
print("classification report: ",classification_report(ytest, y_pred_xtest))
```

```
classification report:                   precision    recall  f1-score   support

    Iris-setosa       1.00      1.00      1.00        11
Iris-versicolor       1.00      1.00      1.00        13
 Iris-virginica       1.00      1.00      1.00         6

       accuracy                           1.00        30
      macro avg       1.00      1.00      1.00        30
   weighted avg       1.00      1.00      1.00        30
```