stml-9

October 11, 2024

[ ]:
```
import pandas as pd
data = pd.read_csv('/content/breast_cancer_survival.csv')
data.head()
```

[ ]:
```
   Age  Gender  Protein1  Protein2  Protein3  Protein4 Tumour_Stage  \
0   42  FEMALE   0.95256   2.15000  0.007972 -0.048340           II
1   54  FEMALE   0.00000   1.38020 -0.498030 -0.507320           II
2   63  FEMALE  -0.52303   1.76400 -0.370190  0.010815           II
3   78  FEMALE  -0.87618   0.12943 -0.370380  0.132190            I
4   42  FEMALE   0.22611   1.74910 -0.543970 -0.390210           II

                      Histology ER status PR status HER2 status Surgery_type  \
0  Infiltrating Ductal Carcinoma  Positive  Positive    Negative        Other
1  Infiltrating Ductal Carcinoma  Positive  Positive    Negative        Other
2  Infiltrating Ductal Carcinoma  Positive  Positive    Negative   Lumpectomy
3  Infiltrating Ductal Carcinoma  Positive  Positive    Negative        Other
4  Infiltrating Ductal Carcinoma  Positive  Positive    Positive   Lumpectomy

  Date_of_Surgery Date_of_Last_Visit Patient_Status
0       20-May-18          26-Aug-18          Alive
1       26-Apr-18          25-Jan-19           Dead
2       24-Aug-18          08-Apr-20          Alive
3       16-Nov-18          28-Jul-20          Alive
4       12-Dec-18          05-Jan-19          Alive
```

[ ]:
```
data.isnull().sum()
```

[ ]:
```
Age                   0
Gender                0
Protein1              0
Protein2              0
Protein3              0
Protein4              0
Tumour_Stage          0
Histology             0
ER status             0
PR status             0
```

```
HER2 status            0
Surgery_type           0
Date_of_Surgery        0
Date_of_Last_Visit    17
Patient_Status        13
dtype: int64
```

[ ]:
```
data.replace('FEMALE',0, inplace=True)
data.replace('MALE',1, inplace=True)
data.replace('Positive',1, inplace=True)
data.replace('Negative',0, inplace=True)
data.replace('Dead',0, inplace=True)
data.replace('Alive',1, inplace=True)
```

```
<ipython-input-3-709dcaf1cf2f>:2: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  data.replace('MALE',1, inplace=True)
<ipython-input-3-709dcaf1cf2f>:3: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  data.replace('Positive',1, inplace=True)
<ipython-input-3-709dcaf1cf2f>:4: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  data.replace('Negative',0, inplace=True)
<ipython-input-3-709dcaf1cf2f>:6: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  data.replace('Alive',1, inplace=True)
```

[ ]:
```
data.head()
```

[ ]:

	Age	Gender	Protein1	Protein2	Protein3	Protein4	Tumour_Stage	Histology	ER status	PR status	HER2 status
0	42	0	0.95256	2.15000	0.007972	-0.048340	II	Infiltrating Ductal Carcinoma	1	1	0
1	54	0	0.00000	1.38020	-0.498030	-0.507320	II	Infiltrating Ductal Carcinoma	1	1	0
2	63	0	-0.52303	1.76400	-0.370190	0.010815	II				
3	78	0	-0.87618	0.12943	-0.370380	0.132190	I				
4	42	0	0.22611	1.74910	-0.543970	-0.390210	II				

```
2  Infiltrating Ductal Carcinoma          1          1            0
3  Infiltrating Ductal Carcinoma          1          1            0
4  Infiltrating Ductal Carcinoma          1          1            1

  Surgery_type Date_of_Surgery Date_of_Last_Visit  Patient_Status
0        Other       20-May-18          26-Aug-18             1.0
1        Other       26-Apr-18          25-Jan-19             0.0
2   Lumpectomy       24-Aug-18          08-Apr-20             1.0
3        Other       16-Nov-18          28-Jul-20             1.0
4   Lumpectomy       12-Dec-18          05-Jan-19             1.0
```

[ ]:
```
data['Surgery_type'].unique()
```

[ ]:
```
array(['Other', 'Lumpectomy', 'Modified Radical Mastectomy',
       'Simple Mastectomy'], dtype=object)
```

[ ]:
```
data['Histology'].unique()
```

[ ]:
```
array(['Infiltrating Ductal Carcinoma', 'Infiltrating Lobular Carcinoma',
       'Mucinous Carcinoma'], dtype=object)
```

[ ]:
```
data.replace('II',2, inplace=True)
data.replace('III',3, inplace=True)
data.replace('I',1, inplace=True)
```

```
<ipython-input-7-fd5d96a82175>:3: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  data.replace('I',1, inplace=True)
```

[ ]:
```
data.replace('Infiltrating Ductal Carcinoma',1, inplace=True)
data.replace('Infiltrating Lobular Carcinoma',2, inplace=True)
data.replace('Mucinous Carcinoma',3, inplace=True)
```

```
<ipython-input-8-ecb44d251b39>:3: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  data.replace('Mucinous Carcinoma',3, inplace=True)
```

[ ]:
```
data.replace('Other',0, inplace=True)
data.replace('Lumpectomy',1, inplace=True)
data.replace('Modified Radical Mastectomy',2, inplace=True)
data.replace('Simple Mastectomy',3, inplace=True)
```

```
<ipython-input-9-f9216a2b26c7>:4: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
```

```
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  data.replace('Simple Mastectomy',3, inplace=True)
```

[ ]:
```
data.head()
```

[ ]:
```
   Age  Gender  Protein1  Protein2  Protein3  Protein4  Tumour_Stage  \
0   42       0   0.95256   2.15000  0.007972 -0.048340             2
1   54       0   0.00000   1.38020 -0.498030 -0.507320             2
2   63       0  -0.52303   1.76400 -0.370190  0.010815             2
3   78       0  -0.87618   0.12943 -0.370380  0.132190             1
4   42       0   0.22611   1.74910 -0.543970 -0.390210             2

   Histology  ER status  PR status  HER2 status  Surgery_type Date_of_Surgery  \
0          1          1          1            0             0       20-May-18
1          1          1          1            0             0       26-Apr-18
2          1          1          1            0             1       24-Aug-18
3          1          1          1            0             0       16-Nov-18
4          1          1          1            1             1       12-Dec-18

  Date_of_Last_Visit  Patient_Status
0          26-Aug-18             1.0
1          25-Jan-19             0.0
2          08-Apr-20             1.0
3          28-Jul-20             1.0
4          05-Jan-19             1.0
```

[ ]:
```
x=data.drop(['Patient_Status','Date_of_Surgery','Date_of_Last_Visit'],axis=1)
y=data['Patient_Status']
```

[ ]:
```
x.head()
```

[ ]:
```
   Age  Gender  Protein1  Protein2  Protein3  Protein4  Tumour_Stage  \
0   42       0   0.95256   2.15000  0.007972 -0.048340             2
1   54       0   0.00000   1.38020 -0.498030 -0.507320             2
2   63       0  -0.52303   1.76400 -0.370190  0.010815             2
3   78       0  -0.87618   0.12943 -0.370380  0.132190             1
4   42       0   0.22611   1.74910 -0.543970 -0.390210             2

   Histology  ER status  PR status  HER2 status  Surgery_type
0          1          1          1            0             0
1          1          1          1            0             0
2          1          1          1            0             1
3          1          1          1            0             0
4          1          1          1            1             1
```

[ ]:
```
y.count()
```

[ ]:
```
321
```

[ ]:
```
y.head()
```

[ ]:
```
0    1.0
1    0.0
2    1.0
3    1.0
4    1.0
Name: Patient_Status, dtype: float64
```

[ ]:
```
y.isnull().sum()
y.fillna(0,inplace=True)
```

[ ]:
```
from imblearn.over_sampling import SMOTE
smote=SMOTE()
x,y=smote.fit_resample(x,y)
```

[ ]:
```
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=32)
```

[ ]:
```
y.isnull().sum()
```

[ ]:
```
0
```

[ ]:
```
from sklearn.svm import SVC
model=SVC()
model.fit(x_train,y_train)
```

[ ]:
```
SVC()
```

[ ]:
```
from sklearn.metrics import␣
↪accuracy_score,confusion_matrix,classification_report
y_pred=model.predict(x_test)
print(accuracy_score(y_test,y_pred))
```

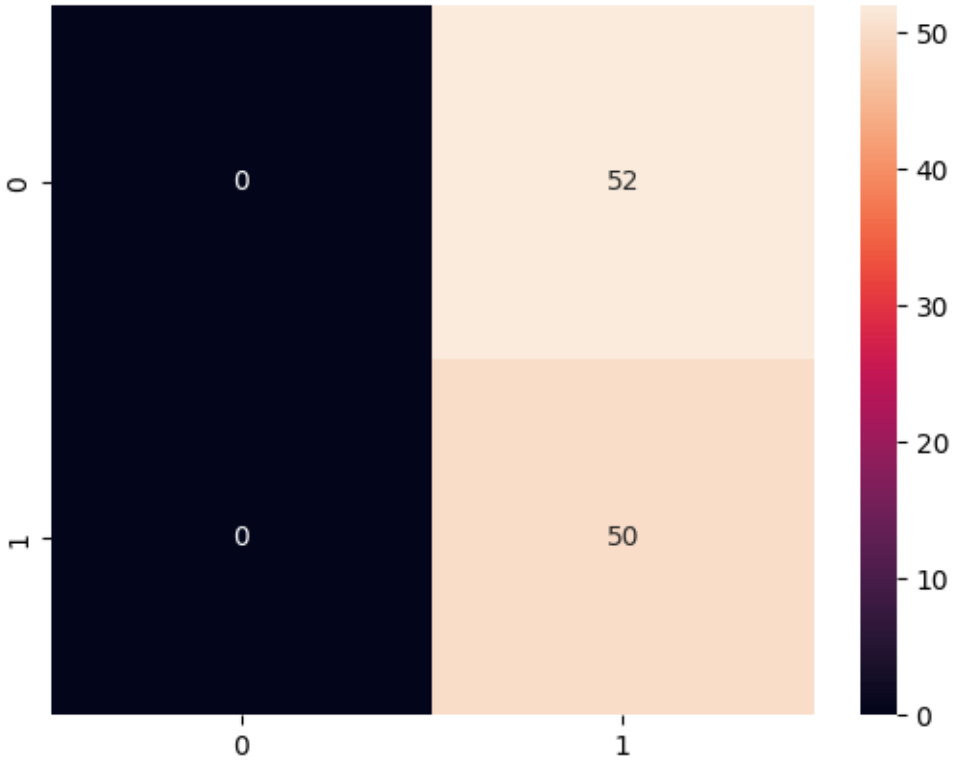
```
0.49019607843137253
```

[ ]:
```
print(confusion_matrix(y_test,y_pred))
```

```
[[ 0 52]
 [ 0 50]]
```

[39]:
```
import seaborn as sns
sns.heatmap(confusion_matrix(y_test,y_pred),annot=True)
```

[39]: <Axes: >



[38]:
```
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

         0.0       0.00      0.00      0.00        52
         1.0       0.49      1.00      0.66        50

    accuracy                           0.49       102
   macro avg       0.25      0.50      0.33       102
weighted avg       0.24      0.49      0.32       102
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
```

```
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```

[ ]:
```
from sklearn.linear_model import LogisticRegression,Perceptron
model=LogisticRegression()
model.fit(x_train,y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:469:
ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
```

[ ]:
```
LogisticRegression()
```

[ ]:
```
yp = model.predict(x_test)
print(accuracy_score(y_test,yp))
```

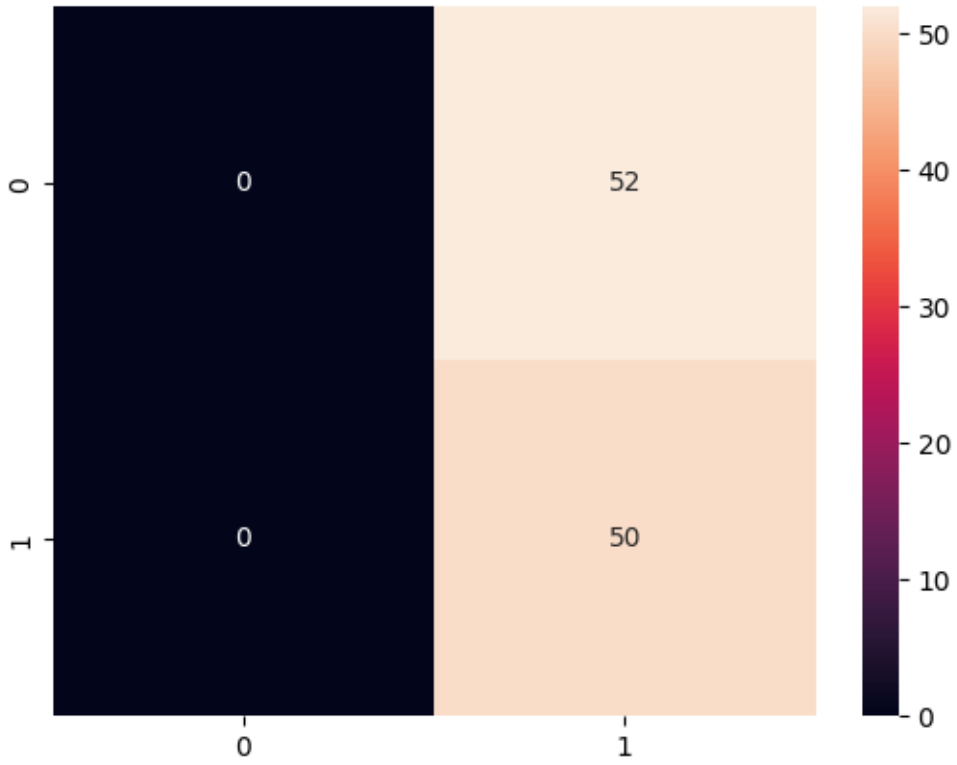
```
0.6862745098039216
```

[ ]:
```
print(confusion_matrix(y_test,yp))
```

```
[[35 17]
 [15 35]]
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test,y_pred),annot=True)
```

[ ]:
```
<Axes: >
```



[ ]:
```
print(classification_report(y_test,yp))
```

```
              precision    recall  f1-score   support

         0.0       0.70      0.67      0.69        52
         1.0       0.67      0.70      0.69        50

    accuracy                           0.69       102
   macro avg       0.69      0.69      0.69       102
weighted avg       0.69      0.69      0.69       102
```

[ ]:
```
pm= Perceptron()
pm.fit(x_train,y_train)
```

[ ]:
```
Perceptron()
```

[ ]:
```
yp = pm.predict(x_test)
print(accuracy_score(y_test,yp))
```

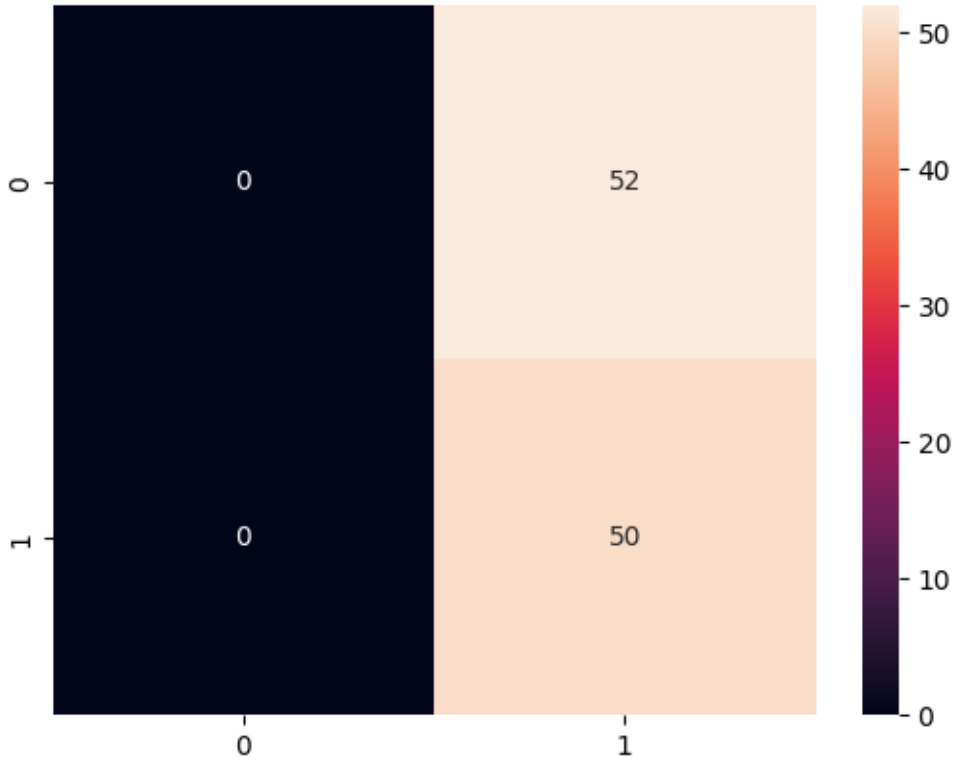
```
0.5098039215686274
```

[ ]:
```
print(confusion_matrix(y_test,yp))
```

```
[[52  0]
 [50  0]]
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test,y_pred),annot=True)
```

[ ]: <Axes: >



[ ]:
```
print(classification_report(y_test,yp))
```

```
              precision    recall  f1-score   support

         0.0       0.51      1.00      0.68        52
         1.0       0.00      0.00      0.00        50

    accuracy                           0.51       102
   macro avg       0.25      0.50      0.34       102
weighted avg       0.26      0.51      0.34       102

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
```

```
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```