



National University of Modern Languages

Artificial Intelligence - Lab

Assignment # 2

BSSE - 5 - Morning

Submitted By:

Muhammad Umair – 12093

Talha Malik – 12089

Submitted To:

Sir Faiq

TASK 1

Implement the decision Tree classifier to detect breast cancer.

Code:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

data = pd.read_csv(r'E:\NUML\Semester Data\Semester 5\AI\AI Lab\12093_Muhammad Umair_A2\dataR2.csv')

print(data.head())
print("\nCancer data set dimensions : {}".format(data.shape))
x = data.drop('Classification', 'columns')
y = data['Classification']
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=5)
scaler = StandardScaler()
scaler.fit(x_train)
x_train=scaler.transform(x_train)
x_test=scaler.transform(x_test)

from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion= 'entropy', random_state=0)
classifier.fit(x_train, y_train)
y_pred_dt = classifier.predict(x_test)
print('Accuracy: ',accuracy_score(y_test, y_pred_dt))
cm = confusion_matrix(y_test, y_pred_dt)
print('Confusion Matrix: \n',cm)
```

```
plt.title("Heatmap of Confusion Matrix", fontsize = 15)
sns.heatmap(cm, annot= True)
plt.show()
print(classification_report(y_test, y_pred_dt))
```

Output:

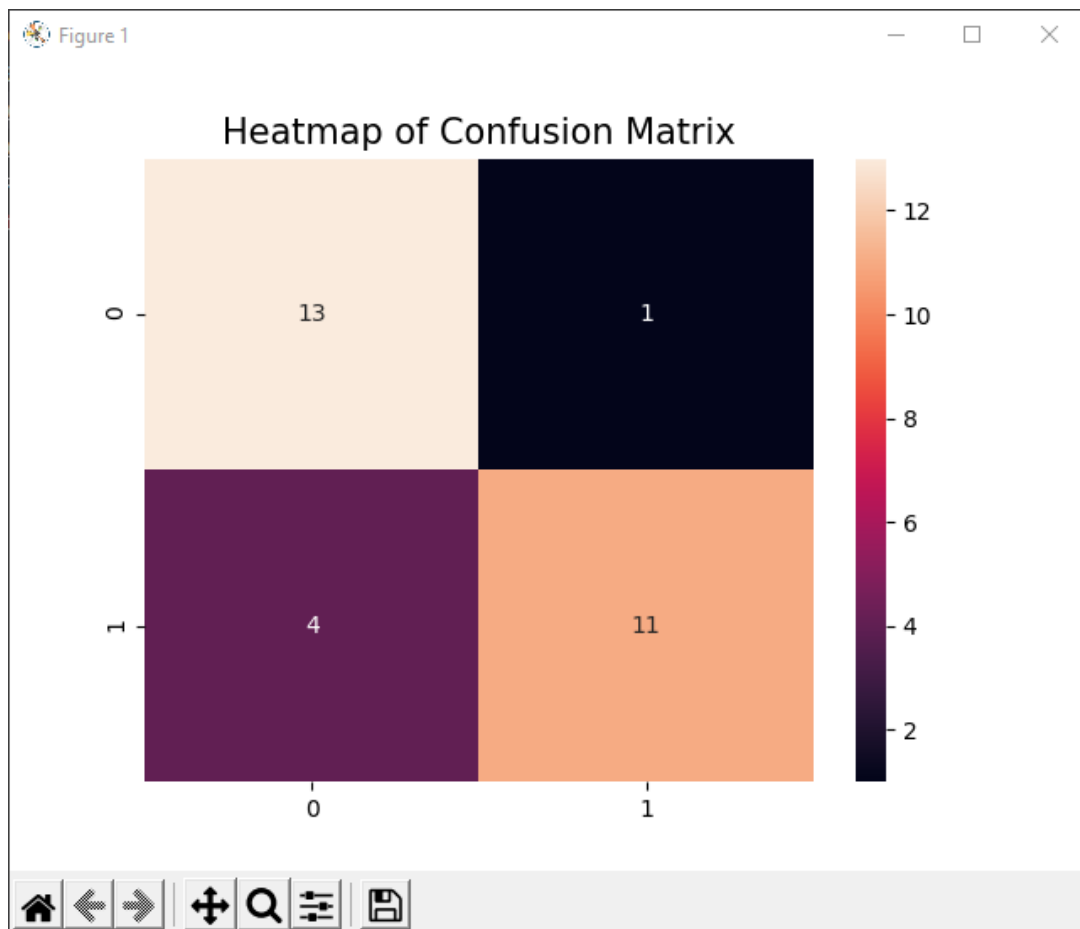
```
PS E:\NUNML\Semester Data\Semester 5\AI\AI Lab\12093_Muhammad Umair_A2> & C:/Users/muham/AppData/Local/Programs/Python/Python39/python.exe "e:/NUNML/Semester Data/Semester 5/AI/AI Lab/12093_Muhammad Umair_A2/A2_Q1.py"
Age      BMI      Glucose  Insulin  HOMA    Leptin  Adiponectin  Resistin  MCP.1  Classification
0  48  23.500000  70  2.707  0.467409  8.8071  9.702400  7.99585  417.114  1
1  83  20.690495  92  3.115  0.706897  8.8438  5.429285  4.06405  468.786  1
2  82  23.124670  91  4.498  1.009651  17.9393  22.432040  9.27715  554.697  1
3  68  21.367521  77  3.226  0.612725  9.8827  7.169560  12.76600  928.220  1
4  86  21.111111  92  3.549  0.805386  6.6994  4.819240  10.57635  773.920  1

Cancer data set dimensions : (116, 10)
Accuracy: 0.8275862068965517
Confusion Matrix:
[[13  1]
 [ 4 11]]
      precision    recall  f1-score   support

     1       0.76      0.93      0.84        14
     2       0.92      0.73      0.81        15

 accuracy          0.84
 macro avg          0.84      0.83      0.83        29
 weighted avg          0.84      0.83      0.83        29

PS E:\NUNML\Semester Data\Semester 5\AI\AI Lab\12093_Muhammad Umair_A2> █
```



TASK 2

**Implement the logistic regression to detect breast cancer. Note:
Standardize the data before training the model**

Code:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

data = pd.read_csv(r'E:\NUML\Semester Data\Semester 5\AI\AI Lab\12093_Muhammad Umair_A2\dataR2.csv')

print(data.head())
print("\nCancer data set dimensions : {}".format(data.shape))
x = data.drop('Classification', 'columns')
y = data['Classification']
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=5)
scaler = StandardScaler()
scaler.fit(x_train)
x_train=scaler.transform(x_train)
x_test=scaler.transform(x_test)

from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(x_train, y_train)
y_pred_lr = classifier.predict(x_test)
print('Accuracy: ',accuracy_score(y_test, y_pred_lr))
cm = confusion_matrix(y_test, y_pred_lr)
```

```

print('Confusion Matrix: \n',cm)

plt.title("Heatmap of Confusion Matrix", fontsize = 15)

sns.heatmap(cm, annot= True)

plt.show()

print(classification_report(y_test, y_pred_lr))

```

Output:

```

PS E:\NML\Semester Data\Semester 5\AI\AI Lab\12093_Muhammad Umair_A2> & C:/Users/muham/AppData/Local/Programs/Python/Python39/python.exe "e:/NML/Sem
ester Data/Semester 5/AI/AI Lab/12093_Muhammad Umair_A2/A2_Q2.py"
Age      BMI      Glucose  Insulin   HOMA     Leptin   Adiponectin  Resistin  MCP.1  Classification
0  48  23.500000  70  2.707  0.467409  8.8071  9.702400  7.99585  417.114  1
1  83  20.690495  92  3.115  0.706897  8.8438  5.429285  4.06405  468.786  1
2  82  23.124670  91  4.498  1.009651  17.9393  22.432040  9.27715  554.697  1
3  68  21.367521  77  3.226  0.612725  9.8827  7.169560  12.76600  928.220  1
4  86  21.111111  92  3.549  0.805386  6.6994  4.819240  10.57635  773.920  1

Cancer data set dimensions : (116, 10)
Accuracy: 0.7931034482758621
Confusion Matrix:
[[10  4]
 [ 2 13]]

      precision    recall  f1-score   support

     1         0.83     0.71     0.77         14
     2         0.76     0.87     0.81         15

 accuracy          0.79         29
 macro avg         0.80     0.79     0.79         29
 weighted avg         0.80     0.79     0.79         29

PS E:\NML\Semester Data\Semester 5\AI\AI Lab\12093_Muhammad Umair_A2>

```

