

# ml1

November 9, 2024

```
[1]: import pandas as pd
     from sklearn.model_selection import train_test_split
```

```
[36]: df=pd.read_csv("heart.csv")
```

```
[5]: df.shape
```

```
[5]: (303, 15)
```

```
[6]: df.head(5)
```

```
[6]:   Unnamed: 0  Age  Sex  ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  \
0           1   63   1    typical    145   233   1         2    150
1           2   67   1  asymptomatic    160   286   0         2    108
2           3   67   1  asymptomatic    120   229   0         2    129
3           4   37   1  nonanginal    130   250   0         0    187
4           5   41   0  nontypical    130   204   0         2    172

      ExAng  Oldpeak  Slope  Ca      Thal  AHD
0         0      2.3     3  0.0    fixed  No
1         1      1.5     2  3.0    normal  Yes
2         1      2.6     2  2.0  reversable  Yes
3         0      3.5     3  0.0    normal  No
4         0      1.4     1  0.0    normal  No
```

```
[7]: df.tail(5)
```

```
[7]:   Unnamed: 0  Age  Sex  ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  \
298        299   45   1    typical    110   264   0         0    132
299        300   68   1  asymptomatic    144   193   1         0    141
300        301   57   1  asymptomatic    130   131   0         0    115
301        302   57   0  nontypical    130   236   0         2    174
302        303   38   1  nonanginal    138   175   0         0    173

      ExAng  Oldpeak  Slope  Ca      Thal  AHD
298         0      1.2     2  0.0  reversable  Yes
299         0      3.4     2  2.0  reversable  Yes
```

300	1	1.2	2	1.0	reversible	Yes
301	0	0.0	2	1.0	normal	Yes
302	0	0.0	1	NaN	normal	No

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

```
[10]: Unnamed: 0      0
      Age           0
      Sex           0
      ChestPain      0
      RestBP         0
      Chol           0
      Fbs            0
      RestECG        0
      MaxHR          0
      ExAng          0
      Oldpeak        0
      Slope          0
      Ca             4
      Thal           2
      AHD            0
      dtype: int64
```

```
[13]: count = (df['Fbs']==0).sum()
      print(count)
```

258

```
[15]: count = (df==0).sum()
      print(count)
```

```
Unnamed: 0      0
Age            0
Sex            97
ChestPain       0
RestBP          0
Chol            0
Fbs            258
RestECG        151
MaxHR           0
ExAng          204
Oldpeak         99
Slope           0
Ca             176
Thal            0
AHD             0
      dtype: int64
```

```
[16]: df.dtypes
```

```
[16]: Unnamed: 0      int64
Age              int64
Sex              int64
ChestPain        object
RestBP           int64
Chol             int64
Fbs             int64
RestECG          int64
MaxHR            int64
ExAng            int64
Oldpeak          float64
Slope            int64
Ca              float64
Thal             object
AHD             object
dtype: object
```

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

```
[17]:
```

	Unnamed: 0	Age	Sex	RestBP	Chol	Fbs \
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	152.000000	54.438944	0.679868	131.689769	246.693069	0.148515
std	87.612784	9.038662	0.467299	17.599748	51.776918	0.356198
min	1.000000	29.000000	0.000000	94.000000	126.000000	0.000000
25%	76.500000	48.000000	0.000000	120.000000	211.000000	0.000000
50%	152.000000	56.000000	1.000000	130.000000	241.000000	0.000000
75%	227.500000	61.000000	1.000000	140.000000	275.000000	0.000000
max	303.000000	77.000000	1.000000	200.000000	564.000000	1.000000

  

	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca
count	303.000000	303.000000	303.000000	303.000000	303.000000	299.000000
mean	0.990099	149.607261	0.326733	1.039604	1.600660	0.672241
std	0.994971	22.875003	0.469794	1.161075	0.616226	0.937438
min	0.000000	71.000000	0.000000	0.000000	1.000000	0.000000
25%	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000
50%	1.000000	153.000000	0.000000	0.800000	2.000000	0.000000
75%	2.000000	166.000000	1.000000	1.600000	2.000000	1.000000
max	2.000000	202.000000	1.000000	6.200000	3.000000	3.000000

```
[18]: df['Age'].mean()
```

```
[18]: np.float64(54.43894389438944)
```

```
[20]: df['Chol'].min()
```

```
[20]: np.int64(126)
```

```
[27]: subset = df[['Age', 'Sex', 'ChestPain', 'RestBP', 'Chol']]
      train_data, test_data = train_test_split(subset, test_size=0.25, random_state=42)
      print(train_data.shape)
      print(test_data.shape)
```

```
(227, 5)
```

```
(76, 5)
```

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

Matplotlib is building the font cache; this may take a moment.

```
[30]: from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, \
      ↪ recall_score, f1_score
```

```
[31]: TP = 45
      FP = 55
      FN = 5
      TN = 395
```

```
[32]: conf_matrix = [[TP, FP], [FN, TN]]
      print("Confusion Matrix:")
      print(conf_matrix)
```

Confusion Matrix:

```
[[45, 55], [5, 395]]
```

```
[35]: accuracy = (TP + TN) / (TP + TN + FP + FN)
      print("\nAccuracy:", accuracy)

      precision = TP / (TP + FP)
      print("Precision:", precision)

      # Recall
      recall = TP / (TP + FN)
      print("Recall:", recall)

      # F1 Score
      f1 = 2 * (precision * recall) / (precision + recall)
      print("F1 Score:", f1)
```

Accuracy: 0.88

Precision: 0.45

Recall: 0.9  
F1 Score: 0.6

[ ]: