Meaningful Predictive Modeling

August 24, 2020

0.0.1 Dataset Description

Our dataset is from a study of heart disease that has been open to the public for many years. The study collects various measurements on patient health and cardiovascular statistics, and of course makes patient identities anonymous.

There are 14 columns in the dataset, where the patient_id column is a unique and random identifier. The remaining 13 features are described in the section below.

1. age
2. sex
3. chest pain type (4 values)
4. resting blood pressure
5. serum cholestoral in mg/dl
6. fasting blood sugar > 120 mg/dl
7. resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved
9. exercise induced angina
10. oldpeak = ST depression induced by exercise relative to rest
11. the slope of the peak exercise ST segment
12. number of major vessels (0-3) colored by flourosopy
13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

0.0.2 Attributes types

Real: 1,4,5,8,10,12

Ordered:11,

Binary: 2,6,9

Nominal:7,3,13

Data is provided courtesy of the Cleveland Heart Disease Database via the UCI Machine Learning repository.

Aha, D., and Dennis Kibler. "Instance-based prediction of heart-disease presence with the Cleveland database." University of California 3.1 (1988): 3-2.

0.0.3 Import Libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import datetime
import sklearn

from sklearn.linear_model import LinearRegression, LogisticRegression

from sklearn.model_selection import cross_val_score, train_test_split,␣
↪GridSearchCV, RandomizedSearchCV
from sklearn.preprocessing import LabelEncoder, StandardScaler, MinMaxScaler,␣
↪OneHotEncoder
from sklearn.metrics import confusion_matrix, classification_report,␣
↪mean_absolute_error, mean_squared_error,r2_score
from sklearn.metrics import plot_confusion_matrix, plot_precision_recall_curve,␣
↪plot_roc_curve, accuracy_score
from sklearn.metrics import auc, f1_score, precision_score, recall_score,␣
↪roc_auc_score

%matplotlib inline
sns.set_style('dark')
sns.set(font_scale=1.2)

import warnings
warnings.filterwarnings('ignore')

pd.set_option('display.max_columns',100)
#pd.set_option('display.max_rows',None)
```

[2]:
```
df = pd.read_csv("heart.csv")
```

[3]:
```
df
```

[3]:
```
   age  sex  chestpain  bloodpressure  serum  bloodsugar  \
0   70    1          4            130    322           0
1   67    0          3            115    564           0
```

```
2     57    1          2            124    261           0
3     64    1          4            128    263           0
4     74    0          2            120    269           0
..    …  …        …             …     …          …
265   52    1          3            172    199           1
266   44    1          2            120    263           0
267   56    0          2            140    294           0
268   57    1          4            140    192           0
269   67    1          4            160    286           0

     electrocardiographic  heartrate  angina  depression  slope  vessels  \
0                       2        109       0         2.4      2        3
1                       2        160       0         1.6      2        0
2                       0        141       0         0.3      1        0
3                       0        105       1         0.2      2        1
4                       2        121       1         0.2      1        1
..                    …        …     …          …     …      …
265                     0        162       0         0.5      1        0
266                     0        173       0         0.0      1        0
267                     2        153       0         1.3      2        0
268                     0        148       0         0.4      2        0
269                     2        108       1         1.5      2        3

     thal  disease
0       3        1
1       7        0
2       7        1
3       7        0
4       3        0
..    …      …
265     7        0
266     7        0
267     3        0
268     6        0
269     3        1

[270 rows x 14 columns]
```

0.0.4 Exploratory Data Analysis

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270 entries, 0 to 269
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
```

```
 0   age                   270 non-null    int64
 1   sex                   270 non-null    int64
 2   chestpain             270 non-null    int64
 3   bloodpressure         270 non-null    int64
 4   serum                 270 non-null    int64
 5   bloodsugar            270 non-null    int64
 6   electrocardiographic  270 non-null    int64
 7   heartrate             270 non-null    int64
 8   angina                270 non-null    int64
 9   depression            270 non-null    float64
 10  slope                 270 non-null    int64
 11  vessels               270 non-null    int64
 12  thal                  270 non-null    int64
 13  disease               270 non-null    int64
dtypes: float64(1), int64(13)
memory usage: 29.7 KB
```

[5]:
```
df.describe()
```

[5]:

	age	sex	chestpain	bloodpressure	serum	\
count	270.000000	270.000000	270.000000	270.000000	270.000000	
mean	54.433333	0.677778	3.174074	131.344444	249.659259	
std	9.109067	0.468195	0.950090	17.861608	51.686237	
min	29.000000	0.000000	1.000000	94.000000	126.000000	
25%	48.000000	0.000000	3.000000	120.000000	213.000000	
50%	55.000000	1.000000	3.000000	130.000000	245.000000	
75%	61.000000	1.000000	4.000000	140.000000	280.000000	
max	77.000000	1.000000	4.000000	200.000000	564.000000	

	bloodsugar	electrocardiographic	heartrate	angina	depression	\
count	270.000000	270.000000	270.000000	270.000000	270.00000	
mean	0.148148	1.022222	149.677778	0.329630	1.05000	
std	0.355906	0.997891	23.165717	0.470952	1.14521	
min	0.000000	0.000000	71.000000	0.000000	0.00000	
25%	0.000000	0.000000	133.000000	0.000000	0.00000	
50%	0.000000	2.000000	153.500000	0.000000	0.80000	
75%	0.000000	2.000000	166.000000	1.000000	1.60000	
max	1.000000	2.000000	202.000000	1.000000	6.20000	

	slope	vessels	thal	disease
count	270.000000	270.000000	270.000000	270.000000
mean	1.585185	0.670370	4.696296	0.444444
std	0.614390	0.943896	1.940659	0.497827
min	1.000000	0.000000	3.000000	0.000000
25%	1.000000	0.000000	3.000000	0.000000
50%	2.000000	0.000000	3.000000	0.000000
75%	2.000000	1.000000	7.000000	1.000000

```
max       3.000000     3.000000     7.000000     1.000000
```

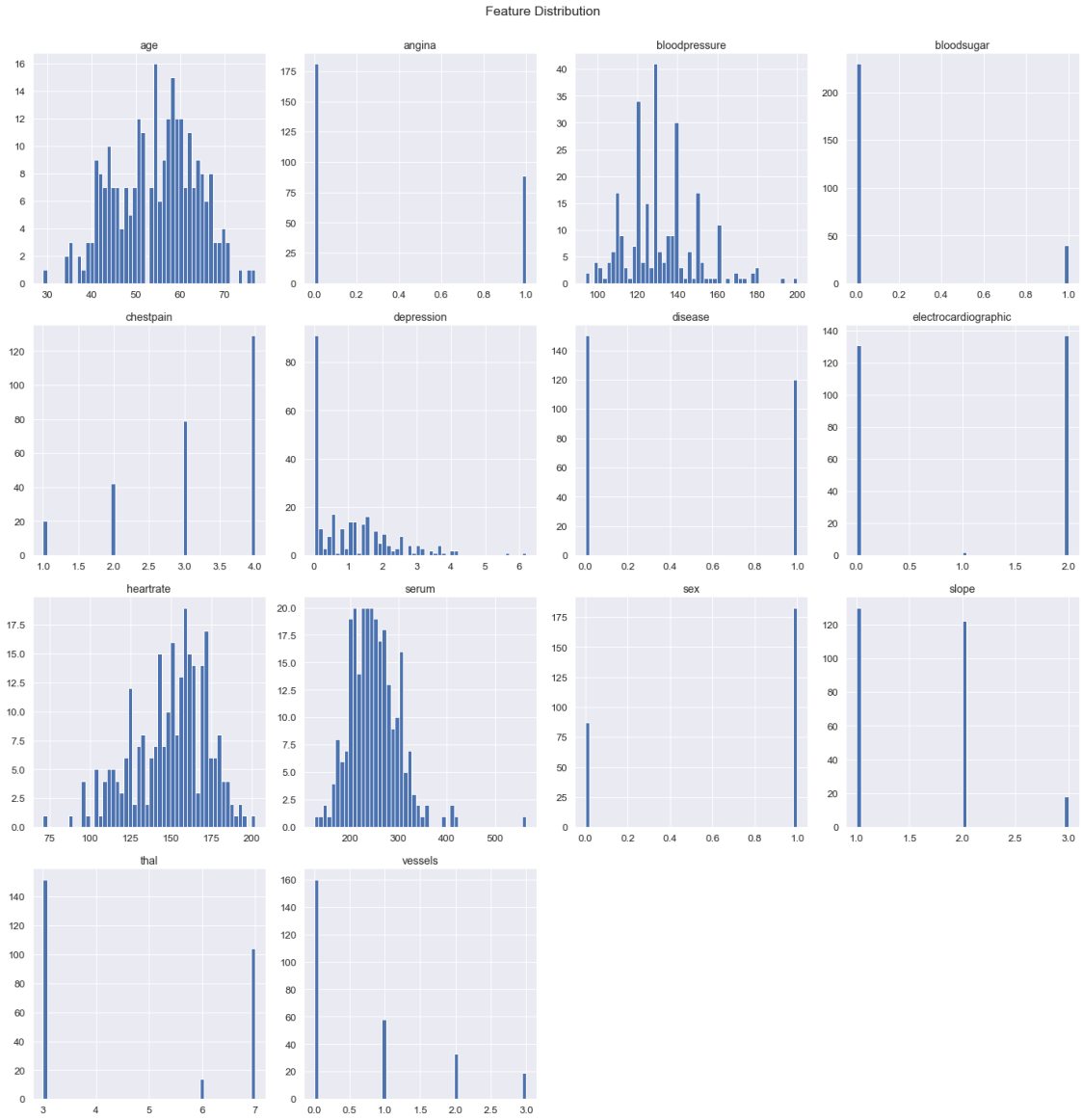
[6]:
```
df.columns
```

[6]:
```
Index(['age', 'sex', 'chestpain', 'bloodpressure', 'serum', 'bloodsugar',
       'electrocardiographic', 'heartrate', 'angina', 'depression', 'slope',
       'vessels', 'thal', 'disease'],
      dtype='object')
```

0.0.5 Data Visualization

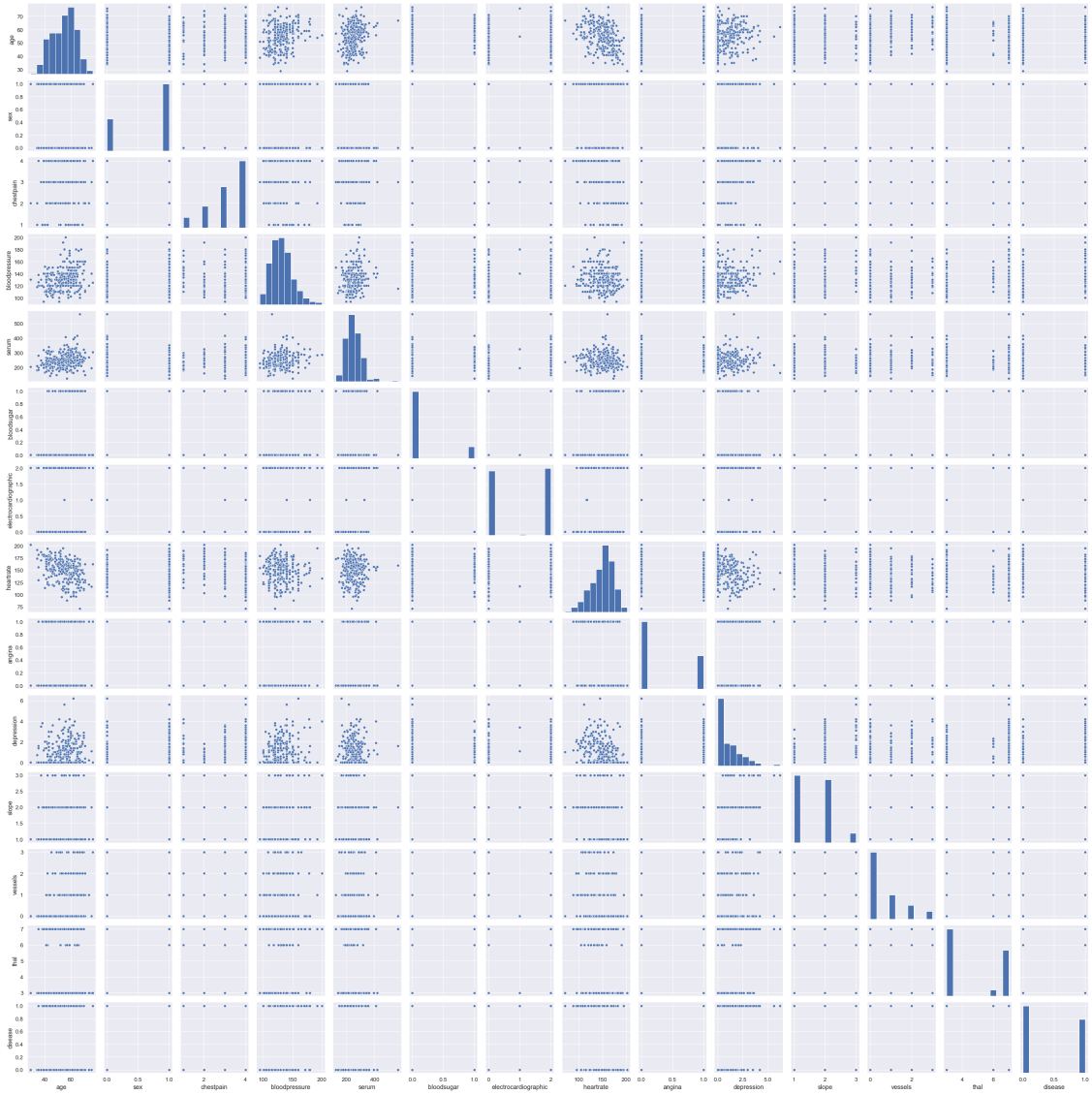
0.0.6 Univariate Data Exploration

[7]:
```
df.hist(bins=50, figsize=(20,20))

plt.suptitle('Feature Distribution', x=0.5, y=1.02, ha='center',␣
↪fontsize='large')

plt.tight_layout()

plt.show();
```



[8]:
```
plt.figure(figsize=(20,20))
plt.suptitle('Pairplots of features', x=0.5, y=1.02, ha='center',␣
 ↪fontsize='large')
sns.pairplot(df)
plt.show()
```

```
<Figure size 1440x1440 with 0 Axes>
```



0.0.7 Correlation

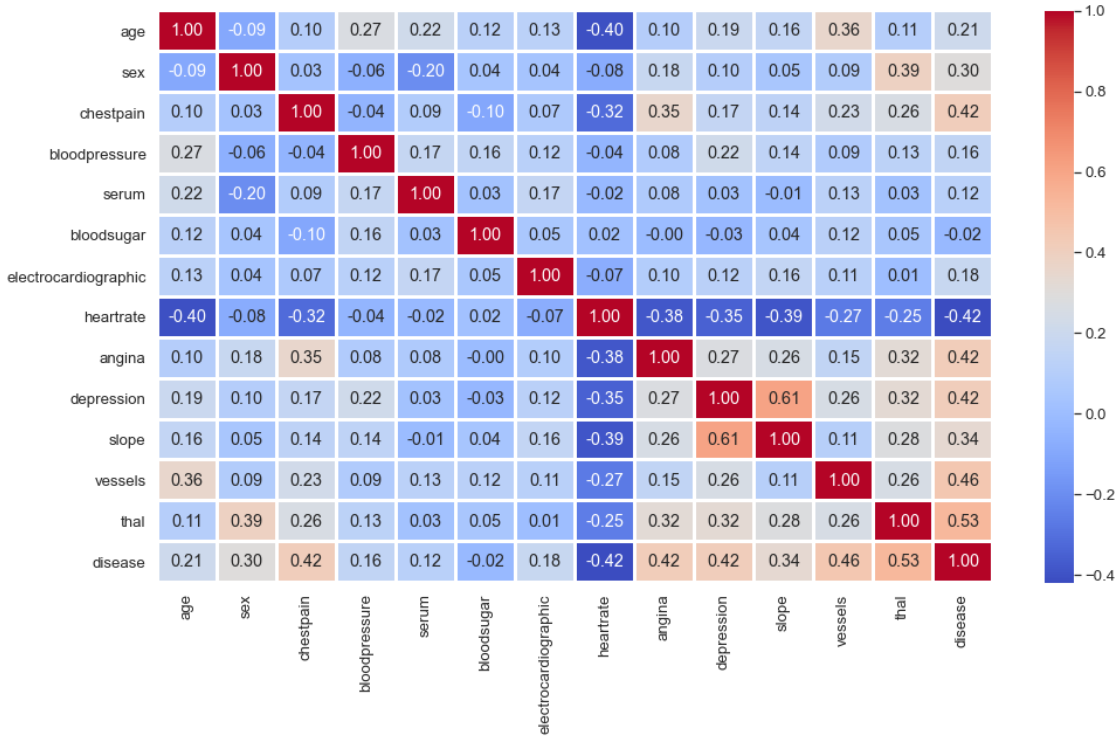
[9]:
```
df.corr()
```

[9]:

	age	sex	chestpain	bloodpressure	serum	\
age	1.000000	-0.094401	0.096920	0.273053	0.220056	
sex	-0.094401	1.000000	0.034636	-0.062693	-0.201647	
chestpain	0.096920	0.034636	1.000000	-0.043196	0.090465	
bloodpressure	0.273053	-0.062693	-0.043196	1.000000	0.173019	
serum	0.220056	-0.201647	0.090465	0.173019	1.000000	
bloodsugar	0.123458	0.042140	-0.098537	0.155681	0.025186	
electrocardiographic	0.128171	0.039253	0.074325	0.116157	0.167652	
heartrate	-0.402215	-0.076101	-0.317682	-0.039136	-0.018739	

```
angina                 0.098297  0.180022   0.353160       0.082793  0.078243
depression             0.194234  0.097412   0.167244       0.222800  0.027709
slope                  0.159774  0.050545   0.136900       0.142472 -0.005755
vessels                0.356081  0.086830   0.225890       0.085697  0.126541
thal                   0.106100  0.391046   0.262659       0.132045  0.028836
disease                0.212322  0.297721   0.417436       0.155383  0.118021

                       bloodsugar  electrocardiographic  heartrate    angina  \
age                      0.123458              0.128171  -0.402215  0.098297
sex                      0.042140              0.039253  -0.076101  0.180022
chestpain               -0.098537              0.074325  -0.317682  0.353160
bloodpressure            0.155681              0.116157  -0.039136  0.082793
serum                    0.025186              0.167652  -0.018739  0.078243
bloodsugar               1.000000              0.053499   0.022494 -0.004107
electrocardiographic     0.053499              1.000000  -0.074628  0.095098
heartrate                0.022494             -0.074628   1.000000 -0.380719
angina                  -0.004107              0.095098  -0.380719  1.000000
depression              -0.025538              0.120034  -0.349045  0.274672
slope                    0.044076              0.160614  -0.386847  0.255908
vessels                  0.123774              0.114368  -0.265333  0.153347
thal                     0.049237              0.007337  -0.253397  0.321449
disease                 -0.016319              0.182091  -0.418514  0.419303

                       depression     slope   vessels      thal   disease
age                      0.194234  0.159774  0.356081  0.106100  0.212322
sex                      0.097412  0.050545  0.086830  0.391046  0.297721
chestpain                0.167244  0.136900  0.225890  0.262659  0.417436
bloodpressure            0.222800  0.142472  0.085697  0.132045  0.155383
serum                    0.027709 -0.005755  0.126541  0.028836  0.118021
bloodsugar              -0.025538  0.044076  0.123774  0.049237 -0.016319
electrocardiographic     0.120034  0.160614  0.114368  0.007337  0.182091
heartrate               -0.349045 -0.386847 -0.265333 -0.253397 -0.418514
angina                   0.274672  0.255908  0.153347  0.321449  0.419303
depression               1.000000  0.609712  0.255005  0.324333  0.417967
slope                    0.609712  1.000000  0.109498  0.283678  0.337616
vessels                  0.255005  0.109498  1.000000  0.255648  0.455336
thal                     0.324333  0.283678  0.255648  1.000000  0.525020
disease                  0.417967  0.337616  0.455336  0.525020  1.000000
```

[10]:
```
plt.figure(figsize=(16,9))
sns.heatmap(df.corr(),cmap="coolwarm",annot=True,fmt='.2f',linewidths=2)
plt.show()
```



[ ]:

0.0.8 Treat Data Types

[11]:
```
df.head()
```

[11]:
```
   age  sex  chestpain  bloodpressure  serum  bloodsugar  \
0   70    1          4            130    322           0
1   67    0          3            115    564           0
2   57    1          2            124    261           0
3   64    1          4            128    263           0
4   74    0          2            120    269           0

   electrocardiographic  heartrate  angina  depression  slope  vessels  thal  \
0                     2        109       0         2.4      2        3     3
1                     2        160       0         1.6      2        0     7
2                     0        141       0         0.3      1        0     7
3                     0        105       1         0.2      2        1     7
4                     2        121       1         0.2      1        1     3

   disease
0        1
1        0
```

```
2        1
3        0
4        0
```

[12]:
```
df.columns
```

[12]:
```
Index(['age', 'sex', 'chestpain', 'bloodpressure', 'serum', 'bloodsugar',
       'electrocardiographic', 'heartrate', 'angina', 'depression', 'slope',
       'vessels', 'thal', 'disease'],
      dtype='object')
```

[13]:
```
df[['sex','chestpain','bloodsugar','electrocardiographic','angina','slope',
       'thal']] =␣
↪df[['sex','chestpain','bloodsugar','electrocardiographic','angina','slope',
       'thal']].astype("object")
```

[14]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270 entries, 0 to 269
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   age                   270 non-null    int64
 1   sex                   270 non-null    object
 2   chestpain             270 non-null    object
 3   bloodpressure         270 non-null    int64
 4   serum                 270 non-null    int64
 5   bloodsugar            270 non-null    object
 6   electrocardiographic  270 non-null    object
 7   heartrate             270 non-null    int64
 8   angina                270 non-null    object
 9   depression            270 non-null    float64
 10  slope                 270 non-null    object
 11  vessels               270 non-null    int64
 12  thal                  270 non-null    object
 13  disease               270 non-null    int64
dtypes: float64(1), int64(6), object(7)
memory usage: 29.7+ KB
```

0.0.9 Feature Scaling/Transform

[15]:
```
df2 = pd.get_dummies(df,drop_first=True)
```

[16]:
```
df2
```

[16]:

	age	bloodpressure	serum	heartrate	depression	vessels	disease	\
0	70	130	322	109	2.4	3	1	
1	67	115	564	160	1.6	0	0	
2	57	124	261	141	0.3	0	1	
3	64	128	263	105	0.2	1	0	
4	74	120	269	121	0.2	1	0	
..	…	…	…	…	…	…	…	
265	52	172	199	162	0.5	0	0	
266	44	120	263	173	0.0	0	0	
267	56	140	294	153	1.3	0	0	
268	57	140	192	148	0.4	0	0	
269	67	160	286	108	1.5	3	1	

	sex_1	chestpain_2	chestpain_3	chestpain_4	bloodsugar_1	\
0	1	0	0	1	0	
1	0	0	1	0	0	
2	1	1	0	0	0	
3	1	0	0	1	0	
4	0	1	0	0	0	
..	…	…	…	…	…	
265	1	0	1	0	1	
266	1	1	0	0	0	
267	0	1	0	0	0	
268	1	0	0	1	0	
269	1	0	0	1	0	

	electrocardiographic_1	electrocardiographic_2	angina_1	slope_2	\
0	0	1	0	1	
1	0	1	0	1	
2	0	0	0	0	
3	0	0	1	1	
4	0	1	1	0	
..	…	…	…	…	
265	0	0	0	0	
266	0	0	0	0	
267	0	1	0	1	
268	0	0	0	1	
269	0	1	1	1	

	slope_3	thal_6	thal_7
0	0	0	0
1	0	0	1
2	0	0	1
3	0	0	1
4	0	0	0
..	…	…	…
265	0	0	1

```
266        0       0       1
267        0       0       0
268        0       1       0
269        0       0       0

[270 rows x 19 columns]
```

[17]:
```
df2.columns
```

[17]:
```
Index(['age', 'bloodpressure', 'serum', 'heartrate', 'depression', 'vessels',
       'disease', 'sex_1', 'chestpain_2', 'chestpain_3', 'chestpain_4',
       'bloodsugar_1', 'electrocardiographic_1', 'electrocardiographic_2',
       'angina_1', 'slope_2', 'slope_3', 'thal_6', 'thal_7'],
      dtype='object')
```

[18]:
```
df2 = df2[['age', 'bloodpressure', 'serum', 'heartrate', 'depression',␣
↪'vessels',
          'sex_1', 'chestpain_2', 'chestpain_3', 'chestpain_4',
          'bloodsugar_1', 'electrocardiographic_1', 'electrocardiographic_2',
          'angina_1', 'slope_2', 'slope_3', 'thal_6', 'thal_7', 'disease']]
```

[19]:
```
df2
```

[19]:
```
     age  bloodpressure  serum  heartrate  depression  vessels  sex_1  \
0     70            130    322        109         2.4        3      1
1     67            115    564        160         1.6        0      0
2     57            124    261        141         0.3        0      1
3     64            128    263        105         0.2        1      1
4     74            120    269        121         0.2        1      0
..   ...            ...    ...        ...         ...      ...    ...
265   52            172    199        162         0.5        0      1
266   44            120    263        173         0.0        0      1
267   56            140    294        153         1.3        0      0
268   57            140    192        148         0.4        0      1
269   67            160    286        108         1.5        3      1

     chestpain_2  chestpain_3  chestpain_4  bloodsugar_1  \
0              0            0            1             0
1              0            1            0             0
2              1            0            0             0
3              0            0            1             0
4              1            0            0             0
..           ...          ...          ...           ...
265            0            1            0             1
266            1            0            0             0
267            1            0            0             0
268            0            0            1             0
```

```
269           0            0           1            0

     electrocardiographic_1  electrocardiographic_2  angina_1  slope_2  \
0                         0                       1         0        1
1                         0                       1         0        1
2                         0                       0         0        0
3                         0                       0         1        1
4                         0                       1         1        0
..                      ...                     ...       ...      ...
265                       0                       0         0        0
266                       0                       0         0        0
267                       0                       1         0        1
268                       0                       0         0        1
269                       0                       1         1        1

     slope_3  thal_6  thal_7  disease
0          0       0       0        1
1          0       0       1        0
2          0       0       1        1
3          0       0       1        0
4          0       0       0        0
..       ...     ...     ...      ...
265        0       0       1        0
266        0       0       1        0
267        0       0       0        0
268        0       1       0        0
269        0       0       0        1

[270 rows x 19 columns]
```

0.0.10 Create and save processed dataset

[20]:
```
#df2.to_csv("heartrain.csv",index=False)
```

0.0.11 Train Test Split

[21]:
```
df = pd.read_csv("heartrain.csv")
```

[22]:
```
df.shape
```

[22]:
```
(270, 19)
```

[23]:
```
df.columns
```

[23]:
```
Index(['age', 'bloodpressure', 'serum', 'heartrate', 'depression', 'vessels',
       'sex_1', 'chestpain_2', 'chestpain_3', 'chestpain_4', 'bloodsugar_1',
```

```
       'electrocardiographic_1', 'electrocardiographic_2', 'angina_1',
       'slope_2', 'slope_3', 'thal_6', 'thal_7', 'disease'],
      dtype='object')
```

[24]:
```
df = df[['age','serum', 'heartrate', 'depression', 'vessels',
       'sex_1', 'chestpain_2', 'chestpain_3', 'chestpain_4', 'bloodsugar_1',
       'electrocardiographic_1', 'electrocardiographic_2', 'angina_1',
       'slope_2', 'slope_3', 'thal_6', 'thal_7', 'disease','bloodpressure']]
```

[25]:
```
df
```

[25]:

	age	serum	heartrate	depression	vessels	sex_1	chestpain_2	\
0	70	322	109	2.4	3	1	0	
1	67	564	160	1.6	0	0	0	
2	57	261	141	0.3	0	1	1	
3	64	263	105	0.2	1	1	0	
4	74	269	121	0.2	1	0	1	
..	…	…	…	…	…	…	…	
265	52	199	162	0.5	0	1	0	
266	44	263	173	0.0	0	1	1	
267	56	294	153	1.3	0	0	1	
268	57	192	148	0.4	0	1	0	
269	67	286	108	1.5	3	1	0	

	chestpain_3	chestpain_4	bloodsugar_1	electrocardiographic_1	\
0	0	1	0	0	
1	1	0	0	0	
2	0	0	0	0	
3	0	1	0	0	
4	0	0	0	0	
..	…	…	…	…	
265	1	0	1	0	
266	0	0	0	0	
267	0	0	0	0	
268	0	1	0	0	
269	0	1	0	0	

	electrocardiographic_2	angina_1	slope_2	slope_3	thal_6	thal_7	\
0	1	0	1	0	0	0	
1	1	0	1	0	0	1	
2	0	0	0	0	0	1	
3	0	1	1	0	0	1	
4	1	1	0	0	0	0	
..	…	…	…	…	…	…	
265	0	0	0	0	0	1	
266	0	0	0	0	0	1	
267	1	0	1	0	0	0	

```
268                     0          0         1         0        1        0
269                     1          1         1         0        0        0

     disease  bloodpressure
0          1            130
1          0            115
2          1            124
3          0            128
4          0            120
..       …              …
265        0            172
266        0            120
267        0            140
268        0            140
269        1            160

[270 rows x 19 columns]
```

[26]:
```
X = df.iloc[:,0:18]
y = df.iloc[:,18]
```

[27]:
```
X.values, y.values
```

[27]:
```
(array([[ 70., 322., 109., …,   0.,   0.,   1.],
        [ 67., 564., 160., …,   0.,   1.,   0.],
        [ 57., 261., 141., …,   0.,   1.,   1.],
        …,
        [ 56., 294., 153., …,   0.,   0.,   0.],
        [ 57., 192., 148., …,   1.,   0.,   0.],
        [ 67., 286., 108., …,   0.,   0.,   1.]]),
 array([130, 115, 124, 128, 120, 120, 130, 110, 140, 150, 135, 142, 140,
        134, 128, 112, 140, 140, 110, 140, 120, 130, 115, 112, 132, 130,
        138, 120, 112, 110, 128, 160, 120, 170, 144, 130, 140, 160, 130,
        122, 152, 124, 130, 101, 126, 140, 118, 110, 160, 150, 136, 128,
        140, 140, 130, 105, 138, 120, 174, 120, 150, 130, 120, 150, 145,
        150, 140, 136, 118, 108, 120, 120, 156, 140, 106, 142, 104,  94,
        120, 120, 146, 120, 150, 130, 110, 148, 128, 178, 126, 150, 140,
        130, 124, 110, 125, 110, 120, 100, 140, 120, 108, 120, 130, 165,
        130, 124, 100, 150, 140, 112, 180, 110, 158, 135, 120, 134, 120,
        200, 150, 130, 120, 122, 152, 160, 125, 160, 120, 136, 134, 117,
        108, 112, 140, 120, 150, 142, 152, 125, 118, 132, 145, 138, 140,
        125, 192, 123, 112, 110, 132, 112, 112, 120, 108, 130, 130, 105,
        140, 128, 120, 178, 120, 150, 130, 128, 110, 180, 110, 130, 138,
        138, 160, 140, 100, 120, 118, 138, 140, 150, 125, 129, 120, 134,
        110, 102, 130, 130, 132, 108, 140, 160, 140, 145, 108, 126, 124,
        135, 100, 110, 140, 125, 118, 125, 125, 140, 160, 152, 102, 105,
        125, 130, 170, 125, 122, 128, 130, 130, 135,  94, 120, 120, 110,
```

```
       135, 150, 130, 138, 135, 130, 132, 150, 118, 145, 118, 115, 128,
       130, 160, 138, 120, 138, 120, 180, 140, 130, 140, 140, 130, 110,
       155, 140, 145, 120, 130, 112, 110, 150, 160, 150, 132, 140, 150,
       120, 130, 120, 130, 110, 172, 120, 140, 140, 160], dtype=int64))
```

[28]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=0)
```

[29]:
```
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

[29]:
```
((216, 18), (54, 18), (216,), (54,))
```

0.0.12 Feature Scaling

[30]:
```
X_train
```

[30]:

	age	serum	heartrate	depression	vessels	sex_1	chestpain_2	\
5	65	177	140	0.4	0	1	0	
22	43	303	181	1.2	0	1	0	
196	58	248	122	1.0	0	0	0	
110	55	327	117	3.4	0	0	0	
12	44	235	180	0.0	0	1	0	
..	…	…	…	…	…	…	…	
251	44	219	188	0.0	0	1	1	
192	54	309	156	0.0	0	1	1	
117	56	288	133	4.0	2	0	0	
47	44	197	177	0.0	1	1	0	
172	67	299	125	0.9	2	1	0	

	chestpain_3	chestpain_4	bloodsugar_1	electrocardiographic_1	\
5	0	1	0	0	
22	0	1	0	0	
196	0	1	0	0	
110	0	1	0	1	
12	1	0	0	0	
..	…	…	…	…	
251	0	0	0	0	
192	0	0	0	0	
117	0	1	1	0	
47	0	1	0	0	
172	0	1	0	0	

	electrocardiographic_2	angina_1	slope_2	slope_3	thal_6	thal_7	\
5	0	0	0	0	0	1	
22	0	0	1	0	0	0	
196	1	0	1	0	0	0	
110	0	1	1	0	0	0	

```
12                        1         0         0         0       0       0
..                      …         …         …         …       …       …
251                       1         0         0         0       0       0
192                       0         0         0         0       0       1
117                       1         1         0         1       0       1
47                        1         0         0         0       0       0
172                       1         1         1         0       0       0

     disease
5          0
22         0
196        0
110        1
12         0
..       …
251        0
192        0
117        1
47         1
172        1

[216 rows x 18 columns]
```

[31]:
```
minmax = MinMaxScaler()
```

[32]:
```
X_train_scaled = minmax.fit_transform(X_train)
```

[33]:
```
X_test_scaled = minmax.transform(X_test)
```

[34]:
```
X_train_scaled
```

[34]:
```
array([[0.76595745, 0.11643836, 0.45614035, …, 0.        , 1.        ,
        0.        ],
       [0.29787234, 0.40410959, 0.81578947, …, 0.        , 0.        ,
        0.        ],
       [0.61702128, 0.27853881, 0.29824561, …, 0.        , 0.        ,
        0.        ],
       …,
       [0.57446809, 0.36986301, 0.39473684, …, 0.        , 1.        ,
        1.        ],
       [0.31914894, 0.16210046, 0.78070175, …, 0.        , 0.        ,
        1.        ],
       [0.80851064, 0.39497717, 0.3245614 , …, 0.        , 0.        ,
        1.        ]])
```

[35]:
```
X_test_scaled
```

```
[35]: array([[ 0.72340426,  0.24429224,  0.54385965,  0.37096774,  0.        ,
                1.        ,  0.        ,  0.        ,  0.        ,  1.        ,
                0.        ,  1.        ,  0.        ,  0.        ,  1.        ,
                1.        ,  0.        ,  0.        ],
              [ 0.36170213,  0.11643836,  0.63157895,  0.22580645,  0.        ,
                0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
                0.        ,  1.        ,  1.        ,  0.        ,  1.        ,
                0.        ,  0.        ,  0.        ],
              [ 0.74468085,  0.40410959,  0.29824561,  0.32258065,  0.66666667,
                0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
                0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
                0.        ,  0.        ,  0.        ],
              [ 0.87234043,  0.32648402,  0.21052632,  0.46774194,  0.33333333,
                1.        ,  0.        ,  1.        ,  0.        ,  0.        ,
                0.        ,  0.        ,  1.        ,  1.        ,  0.        ,
                0.        ,  1.        ,  1.        ],
              [ 0.46808511,  0.11187215,  0.30701754,  0.09677419,  0.        ,
                1.        ,  0.        ,  1.        ,  0.        ,  0.        ,
                0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
                0.        ,  0.        ,  0.        ],
              [ 0.85106383,  0.25799087,  0.55263158,  0.29032258,  0.66666667,
                0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
                0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
                0.        ,  0.        ,  0.        ],
              [ 0.53191489,  0.35844749,  0.93859649,  0.        ,  0.33333333,
                1.        ,  1.        ,  0.        ,  0.        ,  0.        ,
                0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
                0.        ,  1.        ,  1.        ],
              [ 0.27659574,  0.26027397,  0.92982456,  0.12903226,  0.        ,
                1.        ,  0.        ,  1.        ,  0.        ,  1.        ,
                0.        ,  0.        ,  0.        ,  0.        ,  1.        ,
                0.        ,  1.        ,  0.        ],
              [ 0.68085106,  0.26712329,  0.42982456,  0.16129032,  0.        ,
                1.        ,  0.        ,  1.        ,  0.        ,  1.        ,
                0.        ,  0.        ,  1.        ,  1.        ,  0.        ,
                0.        ,  0.        ,  0.        ],
              [ 0.89361702,  0.05251142,  0.3245614 ,  0.25806452,  0.        ,
                0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
                0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
                0.        ,  0.        ,  0.        ],
              [ 0.46808511,  0.41552511,  0.47368421,  0.24193548,  0.33333333,
                0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
                0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
                0.        ,  0.        ,  0.        ],
              [ 0.59574468,  0.18493151,  0.70175439,  0.        ,  0.        ,
                1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
                0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
```

```
  0.        ,  1.        ,  0.        ],
[ 0.53191489,  0.33561644,  0.56140351,  0.08064516,  0.33333333,
  1.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  1.        ,
  0.        ,  0.        ,  0.        ],
[ 0.59574468,  0.37214612,  0.31578947,  0.16129032,  1.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  1.        ,
  0.        ,  1.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.65957447,  0.38127854,  0.71929825,  0.19354839,  0.66666667,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.44680851,  0.29223744,  0.62280702,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.61702128,  0.19406393,  0.6754386 ,  0.        ,  0.        ,
  1.        ,  0.        ,  1.        ,  0.        ,  1.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 1.0212766 ,  0.40639269,  0.64912281,  0.        ,  1.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  1.        ],
[ 0.70212766,  0.35388128,  0.13157895,  0.22580645,  0.33333333,
  1.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.40425532,  0.23515982,  0.70175439,  0.16129032,  0.        ,
  1.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  1.        ,
  0.        ,  1.        ,  1.        ],
[ 0.46808511,  0.39269406,  0.29824561,  0.67741935,  1.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.25531915,  0.16438356,  0.70175439,  0.        ,  0.33333333,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.63829787,  0.25799087,  0.47368421,  0.19354839,  0.33333333,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.34042553,  0.18721461,  0.52631579,  0.48387097,  0.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
```

```
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.31914894,  0.37442922,  0.57017544,  0.        ,  0.33333333,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  1.        ],
[ 0.25531915,  0.20091324,  0.70175439,  0.32258065,  0.        ,
  1.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.53191489,  0.25799087,  0.63157895,  0.19354839,  0.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.80851064,  0.34474886,  0.73684211,  0.        ,  0.33333333,
  0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.70212766,  0.61187215,  0.60526316,  0.19354839,  0.        ,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.80851064,  0.25342466, -0.14912281,  0.16129032,  0.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  1.        ],
[ 0.78723404,  0.40182648,  0.55263158,  0.06451613,  0.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.80851064,  0.22146119,  0.47368421,  0.0483871 ,  0.66666667,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.65957447,  0.1826484 ,  0.38596491,  0.38709677,  0.66666667,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.25531915,  0.17808219,  0.73684211,  0.22580645,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.29787234,  0.49086758,  0.42105263,  0.48387097,  0.        ,
  0.        ,  0.        ,  0.        ,  1.        ,  1.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.70212766,  0.31278539,  0.07894737,  0.19354839,  0.33333333,
```

```
  0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.53191489,  0.3196347 ,  0.18421053,  0.35483871,  0.33333333,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.78723404,  0.23287671,  0.6754386 ,  0.16129032,  0.66666667,
  0.        ,  0.        ,  0.        ,  1.        ,  1.        ,
  0.        ,  0.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.31914894,  0.21461187,  0.71929825,  0.        ,  0.        ,
  1.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.25531915,  0.10502283,  0.61403509,  0.        ,  0.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.36170213,  0.28082192,  0.49122807,  0.12903226,  0.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.46808511,  0.38584475,  0.60526316,  0.09677419,  0.        ,
  0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.46808511,  0.29680365,  0.53508772,  0.08064516,  0.        ,
  0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.68085106,  0.41324201,  0.50877193,  0.16129032,  0.        ,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.5106383 ,  0.24657534,  0.63157895,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.63829787,  0.33561644,  0.3245614 ,  0.        ,  0.        ,
  1.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  1.        ],
[ 0.4893617 ,  0.24429224,  0.51754386,  0.01612903,  1.        ,
  1.        ,  0.        ,  0.        ,  1.        ,  1.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ],
```

```
[ 0.53191489,  0.24200913,  0.6754386 ,  0.25806452,  0.        ,
  1.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  0.        ],
[ 0.80851064,  0.23515982,  0.35964912,  0.41935484,  0.66666667,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.82978723,  0.33789954,  0.54385965,  0.25806452,  0.        ,
  1.        ,  0.        ,  1.        ,  0.        ,  1.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.57446809,  0.64611872,  0.54385965,  0.30645161,  0.66666667,
  0.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  1.        ],
[ 0.61702128,  0.26027397,  0.57894737,  0.09677419,  0.        ,
  1.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  1.        ,  1.        ,  1.        ,  0.        ,
  0.        ,  1.        ,  0.        ],
[ 0.65957447,  0.43835616,  0.63157895,  0.        ,  0.33333333,
  0.        ,  0.        ,  1.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ],
[ 0.4893617 ,  0.23744292,  0.63157895,  0.        ,  0.33333333,
  1.        ,  0.        ,  0.        ,  1.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  1.        ]])
```

0.0.13 Section 1 : Regression (Predict Blood Pressure)

[36]:
```
lr = LinearRegression()
```

[37]:
```
model = lr.fit(X_train_scaled,y_train)
```

[38]:
```
y_pred = model.predict(X_test_scaled)
```

[39]:
```
y_pred[:10]
```

[39]:
```
array([162.48588562, 136.12316588, 133.4370176 , 132.48577756,
       119.31611785, 145.88854269, 135.07170234, 138.83091643,
       125.09789119, 131.10441131])
```

[40]:
```
y_test[:10]
```

[40]:
```
64     145
135    142
153    130
189    160
253    110
198    140
144    192
180    120
256    150
15     112
Name: bloodpressure, dtype: int64
```

0.0.14 Model Evaluation

[41]:
```
mse = mean_squared_error(y_test,y_pred)
mse
```

[41]:
```
355.08887178634967
```

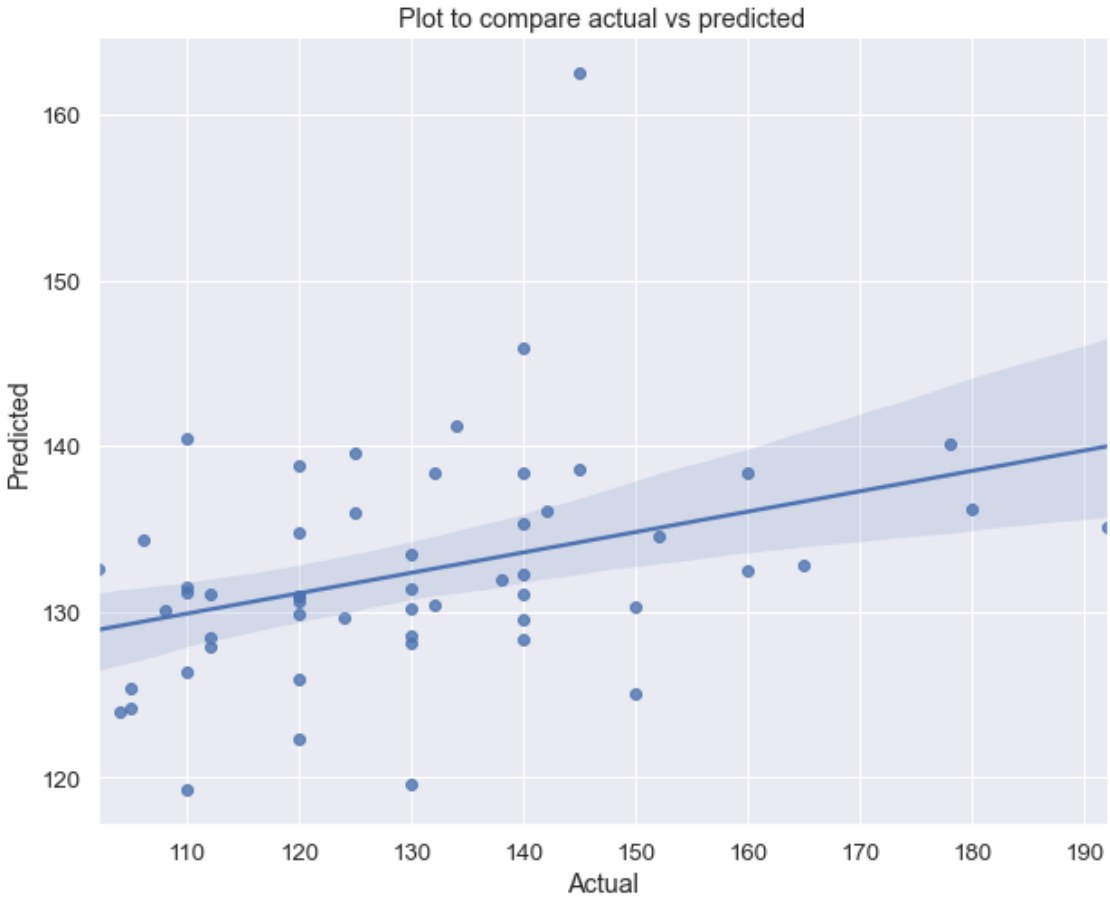
[42]:
```
rmse = np.sqrt(mse)
rmse
```

[42]:
```
18.843801946166533
```

[43]:
```
r2score = r2_score(y_test,y_pred)
r2score
```

[43]:
```
0.12561622123336702
```

[44]:
```
fig, ax = plt.subplots(figsize=(10,8))
sns.regplot(x=y_test, y=y_pred, ax=ax)
plt.title("Plot to compare actual vs predicted")
plt.ylabel("Predicted")
plt.xlabel("Actual")
plt.show()
```



0.0.15 Cross-Validation

[45]:
```
cv = cross_val_score(lr,X,y,cv=5,verbose=1,scoring='r2')
```

```
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done   5 out of   5 | elapsed:    0.0s finished
```

[46]:
```
cv.mean()
```

[46]:
```
-0.002751760815410864
```

0.0.16 Section 2: Classification (Predict Heart Disease)

[47]:
```
df = pd.read_csv("heartrain.csv")
```

[48]:
```
df
```

[48]:

	age	bloodpressure	serum	heartrate	depression	vessels	sex_1	\
0	70	130	322	109	2.4	3	1	
1	67	115	564	160	1.6	0	0	
2	57	124	261	141	0.3	0	1	
3	64	128	263	105	0.2	1	1	
4	74	120	269	121	0.2	1	0	
..	…	…	…	…	…	…	…	
265	52	172	199	162	0.5	0	1	
266	44	120	263	173	0.0	0	1	
267	56	140	294	153	1.3	0	0	
268	57	140	192	148	0.4	0	1	
269	67	160	286	108	1.5	3	1	

	chestpain_2	chestpain_3	chestpain_4	bloodsugar_1	\
0	0	0	1	0	
1	0	1	0	0	
2	1	0	0	0	
3	0	0	1	0	
4	1	0	0	0	
..	…	…	…	…	
265	0	1	0	1	
266	1	0	0	0	
267	1	0	0	0	
268	0	0	1	0	
269	0	0	1	0	

	electrocardiographic_1	electrocardiographic_2	angina_1	slope_2	\
0	0	1	0	1	
1	0	1	0	1	
2	0	0	0	0	
3	0	0	1	1	
4	0	1	1	0	
..	…	…	…	…	
265	0	0	0	0	
266	0	0	0	0	
267	0	1	0	1	
268	0	0	0	1	
269	0	1	1	1	

	slope_3	thal_6	thal_7	disease
0	0	0	0	1
1	0	0	1	0
2	0	0	1	1
3	0	0	1	0
4	0	0	0	0
..	…	…	…	…
265	0	0	1	0

```
266        0       0       1        0
267        0       0       0        0
268        0       1       0        0
269        0       0       0        1

[270 rows x 19 columns]
```

[49]:
```
X = df.iloc[:,0:18]
y = df.iloc[:,18]
```

[50]:
```
X.values, y.values
```

[50]:
```
(array([[ 70., 130., 322., …,   0.,   0.,   0.],
        [ 67., 115., 564., …,   0.,   0.,   1.],
        [ 57., 124., 261., …,   0.,   0.,   1.],
        …,
        [ 56., 140., 294., …,   0.,   0.,   0.],
        [ 57., 140., 192., …,   0.,   1.,   0.],
        [ 67., 160., 286., …,   0.,   0.,   0.]]),
 array([1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0,
        0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0,
        1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1,
        0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0,
        0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0,
        1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1,
        0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0,
        0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1,
        1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0,
        0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
        1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0,
        0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0,
        1, 0, 0, 0, 0, 1], dtype=int64))
```

[51]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=0, stratify=y)
```

[52]:
```
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

[52]:
```
((216, 18), (54, 18), (216,), (54,))
```

0.0.17 Feature Scaling

[53]:
```
X_train
```

[53]:
```
     age  bloodpressure  serum  heartrate  depression  vessels  sex_1  \
222   57            150    168        174         1.6        0      1
10    59            135    234        161         0.5        0      1
```

```
131   66            112    212        132         0.1        1      1
176   51            140    298        122         4.2        3      1
180   42            120    240        194         0.8        0      1
..   …             …    …        …          …       …    …
171   69            140    254        146         2.0        3      1
11    53            142    226        111         0.0        0      1
23    47            112    204        143         0.1        0      1
60    57            150    126        173         0.2        1      1
216   63            135    252        172         0.0        0      0

     chestpain_2  chestpain_3  chestpain_4  bloodsugar_1  \
222            0            1            0             0
10             0            0            1             0
131            0            0            1             0
176            0            0            1             0
180            0            1            0             1
..           …          …          …           …
171            0            1            0             0
11             0            0            1             0
23             0            0            1             0
60             0            1            0             1
216            0            1            0             0

     electrocardiographic_1  electrocardiographic_2  angina_1  slope_2  \
222                       0                       0         0        0
10                        0                       0         0        1
131                       0                       1         1        0
176                       0                       0         1        1
180                       0                       0         0        0
..                      …                     …       …      …
171                       0                       1         0        1
11                        0                       1         1        0
23                        0                       0         0        0
60                        0                       0         0        0
216                       0                       1         0        0

     slope_3  thal_6  thal_7
222        0       0       0
10         0       0       1
131        0       0       0
176        0       0       1
180        1       0       1
..       …     …     …
171        0       0       1
11         0       0       1
23         0       0       0
60         0       0       1
```

```
216        0       0       0

[216 rows x 18 columns]
```

[54]:
```
minmax = MinMaxScaler()
```

[55]:
```
X_train_scaled = minmax.fit_transform(X_train)
```

[56]:
```
X_test_scaled = minmax.transform(X_test)
```

[57]:
```
X_train_scaled
```

[57]:
```
array([[0.58333333, 0.57142857, 0.09589041, …, 0.        , 0.        ,
        0.        ],
       [0.625     , 0.41836735, 0.24657534, …, 0.        , 0.        ,
        1.        ],
       [0.77083333, 0.18367347, 0.19634703, …, 0.        , 0.        ,
        0.        ],
       …,
       [0.375     , 0.18367347, 0.17808219, …, 0.        , 0.        ,
        0.        ],
       [0.58333333, 0.57142857, 0.        , …, 0.        , 0.        ,
        1.        ],
       [0.70833333, 0.41836735, 0.28767123, …, 0.        , 0.        ,
        0.        ]])
```

[58]:
```
X_test_scaled
```

[58]:
```
array([[0.1875    , 0.26530612, 0.23972603, 0.84732824, 0.61290323,
        0.        , 1.        , 0.        , 0.        , 0.        ,
        0.        , 0.        , 0.        , 1.        , 1.        ,
        0.        , 0.        , 1.        ],
       [0.47916667, 0.40816327, 0.17123288, 0.66412214, 0.12903226,
        0.33333333, 1.        , 1.        , 0.        , 0.        ,
        0.        , 0.        , 0.        , 0.        , 0.        ,
        0.        , 0.        , 0.        ],
       [0.29166667, 0.28571429, 0.19863014, 0.71755725, 0.03225806,
        0.        , 0.        , 0.        , 1.        , 0.        ,
        0.        , 0.        , 0.        , 0.        , 1.        ,
        0.        , 0.        , 0.        ],
       [0.79166667, 0.06122449, 0.39497717, 0.41221374, 0.14516129,
        0.66666667, 1.        , 0.        , 0.        , 1.        ,
        0.        , 0.        , 1.        , 1.        , 1.        ,
        0.        , 0.        , 0.        ],
       [0.52083333, 0.26530612, 0.30136986, 0.58015267, 0.06451613,
        0.        , 1.        , 0.        , 1.        , 0.        ,
        0.        , 0.        , 1.        , 0.        , 1.        ,
```

```
 0.        , 0.        , 1.        ],
[0.27083333, 0.55102041, 0.26940639, 0.81679389, 0.12903226,
 0.66666667, 1.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.41666667, 0.36734694, 0.3196347 , 0.76335878, 0.09677419,
 0.        , 1.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.625     , 0.57142857, 0.19634703, 0.65648855, 0.25806452,
 0.        , 1.        , 0.        , 1.        , 0.        ,
 1.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.27083333, 0.08163265, 0.3173516 , 0.38931298, 0.09677419,
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.72916667, 0.26530612, 0.2739726 , 0.19083969, 0.35483871,
 0.33333333, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 1.        , 0.        ,
 1.        , 0.        , 0.        ],
[0.52083333, 0.14285714, 0.32191781, 0.73282443, 0.        ,
 0.        , 0.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.6875    , 0.26530612, 0.35388128, 0.24427481, 0.22580645,
 0.33333333, 1.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.64583333, 0.36734694, 0.28995434, 0.55725191, 0.22580645,
 0.33333333, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        ],
[0.27083333, 0.36734694, 0.12328767, 0.60305344, 0.        ,
 0.        , 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.52083333, 0.3877551 , 0.36986301, 0.67175573, 0.        ,
 0.33333333, 0.        , 1.        , 0.        , 0.        ,
 1.        , 0.        , 1.        , 1.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.39583333, 0.36734694, 0.29680365, 0.60305344, 0.        ,
 0.66666667, 1.        , 0.        , 0.        , 1.        ,
 1.        , 0.        , 1.        , 1.        , 0.        ,
 0.        , 0.        , 1.        ],
[0.64583333, 0.46938776, 0.38127854, 0.75572519, 0.19354839,
 0.66666667, 1.        , 0.        , 0.        , 1.        ,
```

```
 0.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.5625    , 0.26530612, 0.15296804, 0.69465649, 0.30645161,
 0.        , 1.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.52083333, 0.16326531, 0.25799087, 0.41984733, 0.4516129 ,
 0.33333333, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 1.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.52083333, 0.57142857, 0.24200913, 0.71755725, 0.25806452,
 0.        , 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 1.        ],
[0.41666667, 0.24489796, 0.05251142, 0.41984733, 0.12903226,
 1.        , 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.70833333, 0.46938776, 0.15753425, 0.82442748, 0.        ,
 0.66666667, 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.5       , 0.29591837, 0.35616438, 0.18320611, 0.32258065,
 0.66666667, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 1.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.77083333, 0.67346939, 0.2739726 , 0.3740458 , 0.        ,
 1.        , 1.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        , 1.        , 1.        ,
 0.        , 1.        , 0.        ],
[0.79166667, 0.67346939, 0.3652968 , 0.28244275, 0.24193548,
 1.        , 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 1.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.625     , 0.44897959, 0.33105023, 0.84732824, 0.        ,
 0.        , 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.58333333, 0.26530612, 0.52054795, 0.70229008, 0.09677419,
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.75      , 0.44897959, 0.35616438, 0.78625954, 0.22580645,
 0.33333333, 1.        , 0.        , 0.        , 0.        ,
 1.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.625     , 0.81632653, 0.28082192, 0.54961832, 0.        ,
```

```
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 1.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.25      , 0.18367347, 0.32420091, 0.77099237, 0.        ,
 0.        , 0.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 1.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.58333333, 0.34693878, 0.23515982, 0.60305344, 0.06451613,
 0.33333333, 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.5625    , 1.08163265, 0.36986301, 0.47328244, 0.64516129,
 0.66666667, 0.        , 0.        , 0.        , 1.        ,
 1.        , 0.        , 1.        , 1.        , 0.        ,
 1.        , 0.        , 1.        ],
[0.4375    , 0.26530612, 0.21232877, 0.66412214, 0.25806452,
 0.        , 0.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.54166667, 0.36734694, 0.31050228, 0.64122137, 0.        ,
 0.        , 1.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.77083333, 0.67346939, 0.23287671, 0.51145038, 0.37096774,
 0.        , 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 1.        , 0.        ],
[0.58333333, 0.59183673, 0.33789954, 0.12977099, 0.19354839,
 0.33333333, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 1.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.66666667, 0.57142857, 0.26712329, 0.50381679, 0.16129032,
 0.        , 1.        , 0.        , 1.        , 0.        ,
 1.        , 0.        , 0.        , 1.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.52083333, 0.31632653, 0.33561644, 0.61832061, 0.08064516,
 0.33333333, 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 1.        , 0.        , 0.        ],
[0.25      , 0.36734694, 0.20091324, 0.74045802, 0.32258065,
 0.        , 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.6875    , 0.36734694, 0.23972603, 0.57251908, 0.29032258,
 1.        , 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        ],
```

```
[0.66666667, 0.44897959, 0.0913242 , 0.41221374, 0.58064516,
 0.33333333, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 1.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.45833333, 0.46938776, 0.41552511, 0.54198473, 0.24193548,
 0.33333333, 0.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.47916667, 0.18367347, 0.23744292, 0.67938931, 0.        ,
 0.33333333, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.35416667, 0.07142857, 0.16210046, 0.64885496, 0.        ,
 0.        , 1.        , 1.        , 0.        , 0.        ,
 1.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 1.        ],
[0.6875    , 0.46938776, 0.61187215, 0.65648855, 0.19354839,
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 0.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.41666667, 0.36734694, 0.32648402, 0.70229008, 0.        ,
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.60416667, 0.42857143, 0.44063927, 0.61832061, 0.        ,
 0.66666667, 0.        , 1.        , 0.        , 0.        ,
 1.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.70833333, 0.52040816, 0.24429224, 0.60305344, 0.37096774,
 0.        , 1.        , 0.        , 0.        , 0.        ,
 1.        , 0.        , 1.        , 0.        , 0.        ,
 1.        , 1.        , 0.        ],
[0.625     , 0.67346939, 0.33561644, 0.41221374, 0.        ,
 0.        , 1.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 1.        , 0.        , 0.        ,
 0.        , 0.        , 0.        ],
[0.16666667, 0.36734694, 0.28310502, 0.88549618, 0.56451613,
 0.        , 1.        , 0.        , 1.        , 0.        ,
 0.        , 0.        , 0.        , 0.        , 0.        ,
 1.        , 0.        , 0.        ],
[0.72916667, 0.16326531, 0.19406393, 0.55725191, 0.29032258,
 0.        , 1.        , 0.        , 0.        , 0.        ,
 0.        , 0.        , 1.        , 1.        , 1.        ,
 0.        , 0.        , 0.        ],
[0.85416667, 0.52040816, 0.10958904, 0.41221374, 0.41935484,
 0.        , 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 1.        , 0.        ,
```

```
 1.        , 0.        , 1.        ],
[0.60416667, 0.34693878, 0.20547945, 0.45801527, 0.35483871,
 1.        , 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        , 1.        , 1.        ,
 0.        , 0.        , 1.        ],
[0.52083333, 0.26530612, 0.14155251, 0.32061069, 0.22580645,
 0.33333333, 1.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 0.        , 0.        , 1.        ,
 0.        , 0.        , 1.        ]])
```

[59]:
```
logic = LogisticRegression(random_state=0)
```

[60]:
```
model2 = logic.fit(X_train_scaled,y_train)
```

[61]:
```
y_pred2 = model2.predict(X_test_scaled)
```

You can do comparison with test data actual value with predicted data

[62]:
```
y_pred2[:10]
```

[62]:
```
array([1, 0, 0, 1, 1, 0, 0, 0, 0, 1], dtype=int64)
```

[63]:
```
y_test[:10]
```

[63]:
```
160    1
128    0
212    0
172    1
218    0
85     0
263    0
254    0
183    0
133    1
Name: disease, dtype: int64
```

0.0.18 Cross-Validation

[64]:
```
cv = cross_val_score(model2,X,y,cv=5,verbose=1,scoring='accuracy')
```

```
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done   5 out of   5 | elapsed:    0.1s finished
```

[65]:
```
cv.mean()
```

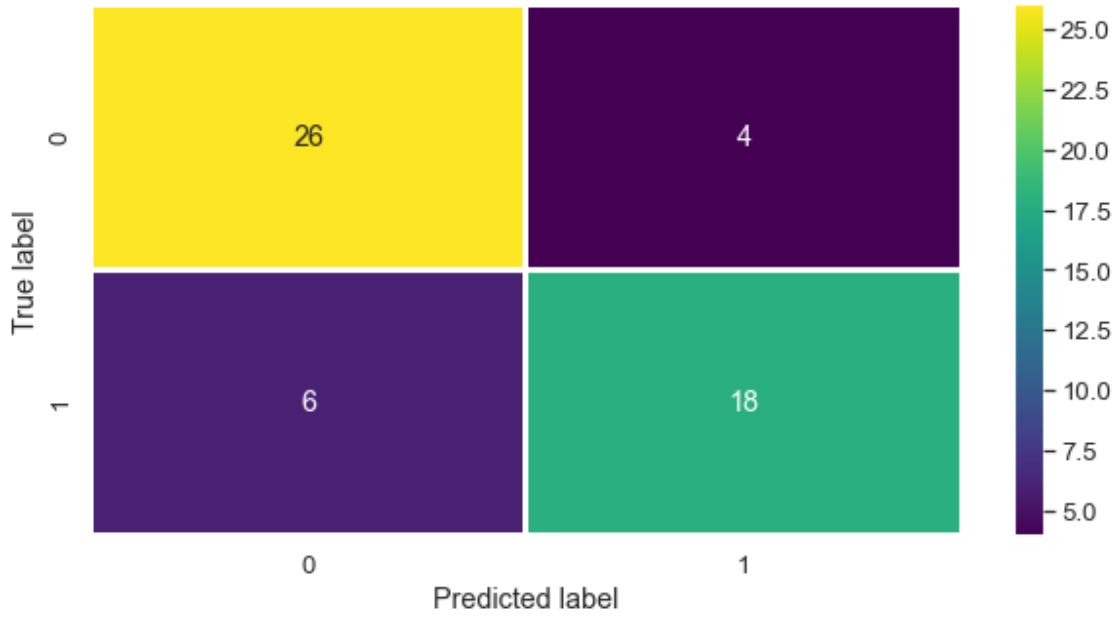
[65]:
```
0.8555555555555555
```

0.0.19 Model Evaluation

[66]:
```
cm = confusion_matrix(y_test,y_pred2)
cm
```

[66]:
```
array([[26,  4],
       [ 6, 18]], dtype=int64)
```

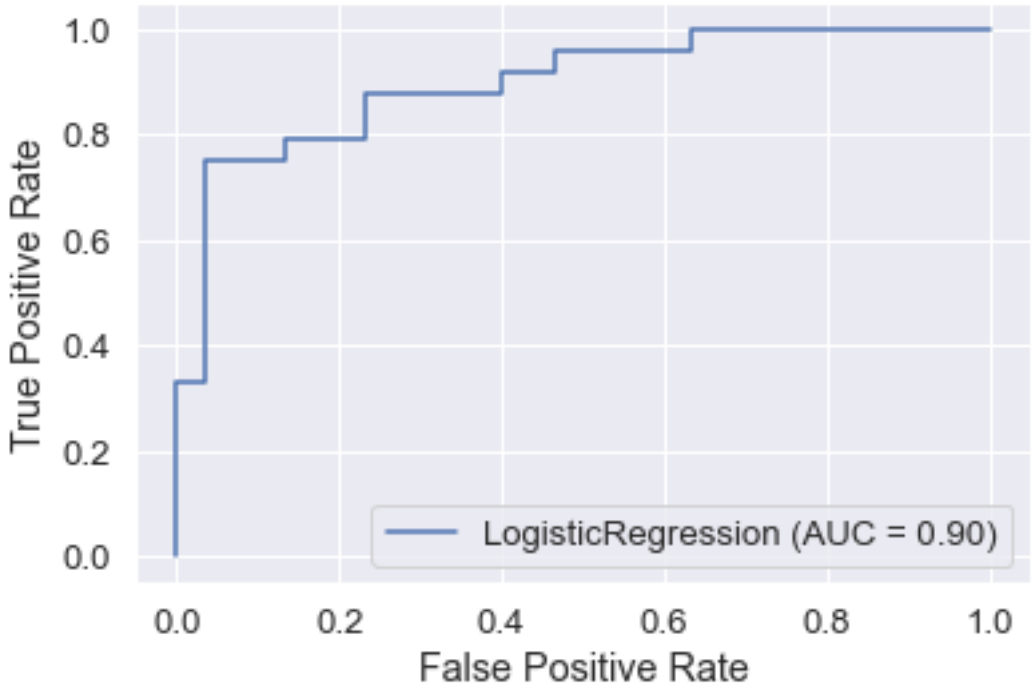
[67]:
```
fig , ax = plt.subplots(figsize=(10,5))
sns.heatmap(cm, annot=True,fmt='.4g',linewidths=2, cmap='viridis')
plt.ylabel('True label')
plt.xlabel('Predicted label')
plt.show()
```



[68]:
```
print(classification_report(y_test,y_pred2))
```

```
              precision    recall  f1-score   support

           0       0.81      0.87      0.84        30
           1       0.82      0.75      0.78        24

    accuracy                           0.81        54
   macro avg       0.82      0.81      0.81        54
weighted avg       0.82      0.81      0.81        54
```

[69]:
```
plot_roc_curve(model2,X_test_scaled,y_test)
plt.show()
```



[ ]: