practical-no-8-1

April 10, 2024

[27]:
```
# Aim: To perfrom and find the accuracy of K-Nearest Neighbors Algorithm ie.KNN␣
↪Classifier
```

[28]:
```
# Name:Shardul T Rushesary
# Roll No:66
# Section:A
```

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from sklearn.model_selection import train_test_split
import warnings
warnings.filterwarnings('ignore')
```

[3]:
```
import os
```

[6]:
```
os.getcwd()
```

[6]: 'C:\\Users\\shard\\OneDrive\\Desktop'

[7]:
```
os.chdir("C:\\Users\\shar\\OneDrive\\Desktop")
```

[8]:
```
df=pd.read_csv("framingham.csv")
```

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

[9]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	\
0	1	39	4.0	0	0.0	0.0	0	
1	0	46	2.0	0	0.0	0.0	0	
2	1	48	1.0	1	20.0	0.0	0	
3	0	61	3.0	1	30.0	0.0	0	
4	0	46	3.0	1	23.0	0.0	0	

	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	\
0	0	0	195.0	106.0	70.0	26.97	80.0	77.0	
1	0	0	250.0	121.0	81.0	28.73	95.0	76.0	

```
2              0         0    245.0  127.5   80.0  25.34       75.0     70.0
3              1         0    225.0  150.0   95.0  28.58       65.0    103.0
4              0         0    285.0  130.0   84.0  23.10       85.0     85.0

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

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

[10]:
```
      male  age  education  currentSmoker  cigsPerDay  BPMeds  \
4235     0   48        2.0              1        20.0     NaN
4236     0   44        1.0              1        15.0     0.0
4237     0   52        2.0              0         0.0     0.0
4238     1   40        3.0              0         0.0     0.0
4239     0   39        3.0              1        30.0     0.0

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP    BMI  \
4235                0             0         0    248.0  131.0   72.0  22.00
4236                0             0         0    210.0  126.5   87.0  19.16
4237                0             0         0    269.0  133.5   83.0  21.47
4238                0             1         0    185.0  141.0   98.0  25.60
4239                0             0         0    196.0  133.0   86.0  20.91

      heartRate  glucose  TenYearCHD
4235       84.0     86.0           0
4236       86.0      NaN           0
4237       80.0    107.0           0
4238       67.0     72.0           0
4239       85.0     80.0           0
```

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

[11]:
```
              male          age    education  currentSmoker   cigsPerDay  \
count  4240.000000  4240.000000  4135.000000    4240.000000  4211.000000
mean      0.429245    49.580189     1.979444       0.494104     9.005937
std       0.495027     8.572942     1.019791       0.500024    11.922462
min       0.000000    32.000000     1.000000       0.000000     0.000000
25%       0.000000    42.000000     1.000000       0.000000     0.000000
50%       0.000000    49.000000     2.000000       0.000000     0.000000
75%       1.000000    56.000000     3.000000       1.000000    20.000000
max       1.000000    70.000000     4.000000       1.000000    70.000000

            BPMeds  prevalentStroke  prevalentHyp     diabetes      totChol  \
```

```
       male          age    education  currentSmoker   cigsPerDay  \
count  4187.000000      4240.000000   4240.000000  4240.000000  4190.000000
mean      0.029615         0.005896      0.310613     0.025708   236.699523
std       0.169544         0.076569      0.462799     0.158280    44.591284
min       0.000000         0.000000      0.000000     0.000000   107.000000
25%       0.000000         0.000000      0.000000     0.000000   206.000000
50%       0.000000         0.000000      0.000000     0.000000   234.000000
75%       0.000000         0.000000      1.000000     0.000000   263.000000
max       1.000000         1.000000      1.000000     1.000000   696.000000

              sysBP        diaBP          BMI    heartRate      glucose  \
count  4240.000000  4240.000000  4221.000000  4239.000000  3852.000000
mean    132.354599    82.897759    25.800801    75.878981    81.963655
std      22.033300    11.910394     4.079840    12.025348    23.954335
min      83.500000    48.000000    15.540000    44.000000    40.000000
25%     117.000000    75.000000    23.070000    68.000000    71.000000
50%     128.000000    82.000000    25.400000    75.000000    78.000000
75%     144.000000    90.000000    28.040000    83.000000    87.000000
max     295.000000   142.500000    56.800000   143.000000   394.000000

        TenYearCHD
count  4240.000000
mean      0.151887
std       0.358953
min       0.000000
25%       0.000000
50%       0.000000
75%       0.000000
max       1.000000
```

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4240 entries, 0 to 4239
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   male             4240 non-null   int64
 1   age              4240 non-null   int64
 2   education        4135 non-null   float64
 3   currentSmoker    4240 non-null   int64
 4   cigsPerDay       4211 non-null   float64
 5   BPMeds           4187 non-null   float64
 6   prevalentStroke  4240 non-null   int64
 7   prevalentHyp     4240 non-null   int64
 8   diabetes         4240 non-null   int64
 9   totChol          4190 non-null   float64
 10  sysBP            4240 non-null   float64
```

```
 11  diaBP            4240 non-null   float64
 12  BMI              4221 non-null   float64
 13  heartRate        4239 non-null   float64
 14  glucose          3852 non-null   float64
 15  TenYearCHD       4240 non-null   int64
dtypes: float64(9), int64(7)
memory usage: 530.1 KB
```

[13]:
```
df.isna().sum()
```

[13]:
```
male                0
age                 0
education         105
currentSmoker       0
cigsPerDay         29
BPMeds             53
prevalentStroke     0
prevalentHyp        0
diabetes            0
totChol            50
sysBP               0
diaBP               0
BMI                19
heartRate           1
glucose           388
TenYearCHD          0
dtype: int64
```

[14]:
```
df
```

[14]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	\
0	1	39	4.0	0	0.0	0.0	
1	0	46	2.0	0	0.0	0.0	
2	1	48	1.0	1	20.0	0.0	
3	0	61	3.0	1	30.0	0.0	
4	0	46	3.0	1	23.0	0.0	
…	…	…	…	…	…	…	
4235	0	48	2.0	1	20.0	NaN	
4236	0	44	1.0	1	15.0	0.0	
4237	0	52	2.0	0	0.0	0.0	
4238	1	40	3.0	0	0.0	0.0	
4239	0	39	3.0	1	30.0	0.0	

	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	\
0	0	0	0	195.0	106.0	70.0	26.97	
1	0	0	0	250.0	121.0	81.0	28.73	
2	0	0	0	245.0	127.5	80.0	25.34	

```
3                  0             1          0    225.0  150.0   95.0  28.58
4                  0             0          0    285.0  130.0   84.0  23.10
…                  …             …          …      …    …      …     …
4235               0             0          0    248.0  131.0   72.0  22.00
4236               0             0          0    210.0  126.5   87.0  19.16
4237               0             0          0    269.0  133.5   83.0  21.47
4238               0             1          0    185.0  141.0   98.0  25.60
4239               0             0          0    196.0  133.0   86.0  20.91

      heartRate  glucose  TenYearCHD
0          80.0     77.0           0
1          95.0     76.0           0
2          75.0     70.0           0
3          65.0    103.0           1
4          85.0     85.0           0
…           …        …             …
4235       84.0     86.0           0
4236       86.0      NaN           0
4237       80.0    107.0           0
4238       67.0     72.0           0
4239       85.0     80.0           0

[4240 rows x 16 columns]
```

1 Missing Value Treatment

[15]:
```
df['glucose'].fillna(value = df['glucose'].mean(),inplace=True)
```

[16]:
```
df['education'].fillna(value = df['education'].mean(),inplace=True)
```

[17]:
```
df['heartRate'].fillna(value = df['heartRate'].mean(),inplace=True)
```

[18]:
```
df['BMI'].fillna(value = df['BMI'].mean(),inplace=True)
```

[19]:
```
df['cigsPerDay'].fillna(value = df['cigsPerDay'].mean(),inplace=True)
```

[20]:
```
df['totChol'].fillna(value = df['totChol'].mean(),inplace=True)
```

[21]:
```
df['BPMeds'].fillna(value = df['BPMeds'].mean(),inplace=True)
```

[22]:
```
df.isna().sum()
```

[22]:
```
male               0
age                0
education          0
currentSmoker      0
```

```
cigsPerDay         0
BPMeds             0
prevalentStroke    0
prevalentHyp       0
diabetes           0
totChol            0
sysBP              0
diaBP              0
BMI                0
heartRate          0
glucose            0
TenYearCHD         0
dtype: int64
```

[23]:
```
#Splitting the dependent and independent variables.
x = df.drop("TenYearCHD",axis=1)
y = df['TenYearCHD']
```

[24]:
```
x #checking the features
```

[24]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	\
0	1	39	4.0	0	0.0	0.000000	
1	0	46	2.0	0	0.0	0.000000	
2	1	48	1.0	1	20.0	0.000000	
3	0	61	3.0	1	30.0	0.000000	
4	0	46	3.0	1	23.0	0.000000	
…	…	…	…	…	…	…	
4235	0	48	2.0	1	20.0	0.029615	
4236	0	44	1.0	1	15.0	0.000000	
4237	0	52	2.0	0	0.0	0.000000	
4238	1	40	3.0	0	0.0	0.000000	
4239	0	39	3.0	1	30.0	0.000000	

	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	\
0	0	0	0	195.0	106.0	70.0	26.97	
1	0	0	0	250.0	121.0	81.0	28.73	
2	0	0	0	245.0	127.5	80.0	25.34	
3	0	1	0	225.0	150.0	95.0	28.58	
4	0	0	0	285.0	130.0	84.0	23.10	
…	…	…	…	…	…	…	…	
4235	0	0	0	248.0	131.0	72.0	22.00	
4236	0	0	0	210.0	126.5	87.0	19.16	
4237	0	0	0	269.0	133.5	83.0	21.47	
4238	0	1	0	185.0	141.0	98.0	25.60	
4239	0	0	0	196.0	133.0	86.0	20.91	

	heartRate	glucose

```
0          80.0   77.000000
1          95.0   76.000000
2          75.0   70.000000
3          65.0  103.000000
4          85.0   85.000000
…           …         …
4235       84.0   86.000000
4236       86.0   81.963655
4237       80.0  107.000000
4238       67.0   72.000000
4239       85.0   80.000000

[4240 rows x 15 columns]
```

2 Train Test Split

[25]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.
↪2,random_state=42)
```

[26]:
```
y_train
```

[26]:
```
1427    0
3257    0
3822    0
1263    0
3575    0
       ..
3444    0
466     0
3092    0
3772    0
860     0
Name: TenYearCHD, Length: 3392, dtype: int64
```

[27]:
```
y_test
```

[27]:
```
1350    1
1434    0
2500    0
1128    0
4144    1
       ..
1844    0
4178    0
4193    1
2897    0
```

```
910     0
Name: TenYearCHD, Length: 848, dtype: int64
```

[28]:
```
x_train
```

[28]:
```
      male  age  education  currentSmoker  cigsPerDay  BPMeds  \
1427     0   53        3.0              1        20.0     0.0
3257     0   64        4.0              1         6.0     0.0
3822     0   38        3.0              0         0.0     0.0
1263     0   49        1.0              0         0.0     0.0
3575     1   56        2.0              1        20.0     0.0
…    …   …         …              …          …      …
3444     0   36        1.0              1         5.0     0.0
466      0   57        3.0              1        15.0     0.0
3092     0   60        2.0              0         0.0     0.0
3772     1   39        2.0              1        10.0     0.0
860      0   35        2.0              0         0.0     0.0

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP    BMI  \
1427                0             0         0    221.0  131.0   89.0  24.09
3257                0             1         0    239.0  143.0   84.0  20.06
3822                0             0         0    185.0  100.0   72.0  22.15
1263                0             0         0    270.0  126.5   67.5  26.56
3575                0             0         0    186.0  116.0   67.0  24.62
…                 …           …       …      …    …     …    …
3444                0             1         0    222.0  147.0   94.0  26.79
466                 0             0         0    250.0  125.0   74.0  21.08
3092                0             1         0    298.0  133.0   89.0  25.09
3772                0             0         0    215.0  102.0   64.5  24.50
860                 0             0         0    248.0  107.0   73.0  20.64

      heartRate  glucose
1427       90.0     95.0
3257       55.0     73.0
3822       85.0     83.0
1263       70.0     77.0
3575       70.0     83.0
…         …      …
3444       76.0     71.0
466        80.0     72.0
3092       83.0     81.0
3772       68.0     62.0
860        90.0     80.0

[3392 rows x 15 columns]
```

[29]:
```
x_test
```

[29]:
```
      male  age  education  currentSmoker  cigsPerDay  BPMeds  \
1350     0   49        3.0              1        10.0     0.0
1434     1   43        1.0              1        25.0     0.0
2500     1   45        1.0              1         1.0     0.0
1128     0   63        3.0              1        10.0     0.0
4144     1   59        2.0              0         0.0     0.0
…    …   …         …              …         …      …
1844     1   35        3.0              1        15.0     0.0
4178     1   41        3.0              1        30.0     0.0
4193     0   63        1.0              0         0.0     0.0
2897     0   45        1.0              0         0.0     0.0
910      1   39        1.0              0         0.0     0.0

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP    BMI  \
1350                0             0         0    260.0  123.0   80.0  23.10
1434                0             0         0    201.0  121.0   82.0  23.84
2500                0             1         0    277.0  140.0   84.0  28.74
1128                0             1         0    236.0  189.0  103.0  27.91
4144                0             0         0    237.0  131.5   84.0  24.17
…                 …             …       …       …    …     …    …
1844                0             0         0    196.0  107.5   66.5  22.64
4178                0             0         0    210.0  132.5   85.0  28.62
4193                0             1         0    306.0  195.0  105.0  27.96
2897                0             0         0    290.0  124.0   72.5  24.24
910                 0             0         0    224.0  108.0   66.0  28.57

      heartRate  glucose
1350       63.0     65.0
1434       70.0     91.0
2500       69.0     74.0
1128       60.0     74.0
4144       90.0     94.0
…          …      …
1844       45.0     79.0
4178       68.0     70.0
4193       75.0     87.0
2897       92.0     87.0
910        90.0     97.0

[848 rows x 15 columns]
```

3 KNN Classifier

[30]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=5, p=2, metric='minkowski')
knn.fit(x_train, y_train)
```

```
acc = knn.score(x_test,y_test)*100
print(acc)
```

```
84.19811320754717
```