tn1saueea

October 17, 2024

[1]:
```
#Aim: To Perform Operation on SVM Classifier
```

[3]:
```
# Name : Ritika Junekar
# Roll no : 30
# Sec: C
# Subject : ET1
```

[5]:
```
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')
```

[7]:
```
import os
```

[9]:
```
os.getcwd()
```

[9]:
```
'C:\\Users\\USER'
```

[11]:
```
os.chdir('C:\\Users\\USER\\DESKTOP')
```

[13]:
```
df=pd.read_csv("framingham.csv")
```

[15]:
```
df.head()
```

[15]:

	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
```

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

[17]:
```
      male  age  education  currentSmoker  cigsPerDay  BPMeds  \
4233     1   50        1.0              1         1.0     0.0
4234     1   51        3.0              1        43.0     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

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP    BMI  \
4233                0             1         0    313.0  179.0   92.0  25.97
4234                0             0         0    207.0  126.5   80.0  19.71
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

      heartRate  glucose  TenYearCHD
4233       66.0     86.0           1
4234       65.0     68.0           0
4235       84.0     86.0           0
4236       86.0      NaN           0
4237       80.0    107.0           0
```

[19]:
```
df.shape
```

[19]:
```
(4238, 16)
```

[21]:
```
df.size
```

[21]:
```
67808
```

[23]:
```
df.info
```

[23]:
```
<bound method DataFrame.info of       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
…        …   …           …                  …          …       …
4233       1   50          1.0                1          1.0      0.0
4234       1   51          3.0                1         43.0      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

      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
…                 …             …         …      …      …     …      …
4233                0             1         0    313.0  179.0   92.0  25.97
4234                0             0         0    207.0  126.5   80.0  19.71
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

      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
…          …        …              …
4233       66.0     86.0           1
4234       65.0     68.0           0
4235       84.0     86.0           0
4236       86.0      NaN           0
4237       80.0    107.0           0

[4238 rows x 16 columns]>
```

[25]:
```
df.describe
```

[25]:
```
<bound method NDFrame.describe of       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
…    …   …         …                …        …        …
4233     1   50        1.0              1         1.0     0.0
4234     1   51        3.0              1        43.0     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

      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
…                 …             …       …       …    …      …     …
4233                0             1         0    313.0  179.0   92.0  25.97
4234                0             0         0    207.0  126.5   80.0  19.71
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

      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
…           …       …           …
4233       66.0     86.0           1
4234       65.0     68.0           0
4235       84.0     86.0           0
4236       86.0      NaN           0
4237       80.0    107.0           0

[4238 rows x 16 columns]>
```

[27]:
```
df
```

[27]:
```
      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
…    …   …         …                …        …        …
4233     1   50        1.0              1         1.0     0.0
4234     1   51        3.0              1        43.0     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

      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
…                 …             …         …        …      …     …      …
4233                0             1         0    313.0  179.0   92.0  25.97
4234                0             0         0    207.0  126.5   80.0  19.71
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

      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
…           …        …           …
4233       66.0     86.0           1
4234       65.0     68.0           0
4235       84.0     86.0           0
4236       86.0      NaN           0
4237       80.0    107.0           0

[4238 rows x 16 columns]
```

[29]:
```
df.isna().sum()
```

[29]:
```
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
```

[31]:
```
df.isnull()
```

[31]:
```
       male    age  education  currentSmoker  cigsPerDay  BPMeds  \
0     False  False      False          False       False   False
1     False  False      False          False       False   False
2     False  False      False          False       False   False
3     False  False      False          False       False   False
4     False  False      False          False       False   False
…     …    …        …              …           …     …
4233  False  False      False          False       False   False
4234  False  False      False          False       False   False
4235  False  False      False          False       False    True
4236  False  False      False          False       False   False
4237  False  False      False          False       False   False

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP    BMI  \
0               False         False     False    False  False  False  False
1               False         False     False    False  False  False  False
2               False         False     False    False  False  False  False
3               False         False     False    False  False  False  False
4               False         False     False    False  False  False  False
…               …           …       …      …    …    …    …
4233            False         False     False    False  False  False  False
4234            False         False     False    False  False  False  False
4235            False         False     False    False  False  False  False
4236            False         False     False    False  False  False  False
4237            False         False     False    False  False  False  False

      heartRate  glucose  TenYearCHD
0         False    False       False
1         False    False       False
2         False    False       False
3         False    False       False
4         False    False       False
…         …      …          …
4233      False    False       False
4234      False    False       False
4235      False    False       False
4236      False     True       False
4237      False    False       False

[4238 rows x 16 columns]
```

[33]:
```
df.isnull().any()
```

[33]:
```
male               False
age                False
education           True
currentSmoker      False
cigsPerDay          True
BPMeds              True
prevalentStroke    False
prevalentHyp       False
diabetes           False
totChol             True
sysBP              False
diaBP              False
BMI                 True
heartRate           True
glucose             True
TenYearCHD         False
dtype: bool
```

1 Missing Value Treatment

[36]:
```
df['glucose'].fillna(value = df['glucose'].mean(),inplace=True)

df['education'].fillna(value = df['education'].mean(),inplace=True)
df['heartRate'].fillna(value = df['heartRate'].mean(),inplace=True)

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

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

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

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

[38]:
```
df.isna().sum()
```

[38]:
```
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
```

[40]:
```
#Splitting the dependent and independent variables.
x = df.drop("TenYearCHD",axis=1)
y = df['TenYearCHD']
```

[42]:
```
x
```

[42]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	\
0	1	39	4.0	0	0.0	0.00000	
1	0	46	2.0	0	0.0	0.00000	
2	1	48	1.0	1	20.0	0.00000	
3	0	61	3.0	1	30.0	0.00000	
4	0	46	3.0	1	23.0	0.00000	
…	…	…	…	…	…	…	
4233	1	50	1.0	1	1.0	0.00000	
4234	1	51	3.0	1	43.0	0.00000	
4235	0	48	2.0	1	20.0	0.02963	
4236	0	44	1.0	1	15.0	0.00000	
4237	0	52	2.0	0	0.0	0.00000	

	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	
…	…	…	…	…	…	…	…	
4233	0	1	0	313.0	179.0	92.0	25.97	
4234	0	0	0	207.0	126.5	80.0	19.71	
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	

	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

```
…          …          …
4233       66.0   86.000000
4234       65.0   68.000000
4235       84.0   86.000000
4236       86.0   81.966753
4237       80.0  107.000000

[4238 rows x 15 columns]
```

[44]:
```
#Splitting the dependent and independent variables.
x = df.drop("TenYearCHD",axis=1)
y = df['TenYearCHD']
```

[46]:
```
x #checking the features
```

[46]:
```
      male  age  education  currentSmoker  cigsPerDay   BPMeds  \
0        1   39        4.0              0         0.0  0.00000
1        0   46        2.0              0         0.0  0.00000
2        1   48        1.0              1        20.0  0.00000
3        0   61        3.0              1        30.0  0.00000
4        0   46        3.0              1        23.0  0.00000
…    …   …         …              …          …       …
4233     1   50        1.0              1         1.0  0.00000
4234     1   51        3.0              1        43.0  0.00000
4235     0   48        2.0              1        20.0  0.02963
4236     0   44        1.0              1        15.0  0.00000
4237     0   52        2.0              0         0.0  0.00000

      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
…                 …           …       …      …    …     …     …
4233                0             1         0    313.0  179.0   92.0  25.97
4234                0             0         0    207.0  126.5   80.0  19.71
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

      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
```

```
…          …          …
4233       66.0   86.000000
4234       65.0   68.000000
4235       84.0   86.000000
4236       86.0   81.966753
4237       80.0  107.000000

[4238 rows x 15 columns]
```

[48]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.
↪2,random_state=42)
```

[50]:
```
y_train
```

[50]:
```
3252    0
3946    0
1261    0
2536    0
4089    0
       ..
3444    0
466     0
3092    0
3772    0
860     0
Name: TenYearCHD, Length: 3390, dtype: int64
```

2 KNN

[53]:
```
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)
```

```
83.13679245283019
```

[ ]: